A Passive Brain-Computer Interface for Predicting Pilot Workload in Virtual Reality Flight Training

Evy van Weelden
Department of Cognitive Science & Artificial Intelligence
Tilburg University
Tilburg, the Netherlands
e.vanweelden@uvt.nl

Carl W. E. van Beek
Department of Cognitive Science & Artificial Intelligence
Tilburg University
Tilburg, the Netherlands
c.w.e.vanbeek@uvt.nl

Maryam Alimardani
Department of Cognitive Science & Artificial Intelligence
Tilburg University
Tilburg, the Netherlands
m.alimardani@uvt.nl

Travis J. Wiltshire
Department of Cognitive Science & Artificial Intelligence
Tilburg University
Tilburg, the Netherlands
t.j.wiltshire@uvt.nl

Wietse D. Ledegang
Department of Human Performance
TNO
Soesterberg, the Netherlands
wietse.ledegang@tno.nl

Eric L. Groen
Department of Human Performance
TNO
Soesterberg, the Netherlands
eric.groen@tno.nl

Max M. Louwerse
Department of Cognitive Science & Artificial Intelligence
Tilburg University
Tilburg, the Netherlands
m.m.louwerse@uvt.nl

2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS) | 979-8-3503-1579-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICHMS59971.2024.10555679

***Abstract*—Quantifying workload is necessary for effective and personalized flight training of student pilots: their workload must not be too low (risk of boredom) nor too high (overload). Passive brain-computer interfaces (pBCIs) allow for measurement of an individual's workload from their brain activity, however, the performance of pBCIs remains sub-optimal due to individual differences and lack of data for classifier training. In this study, we addressed this problem by combining EEG and behavioral data from six novice military pilots who performed a flight task in Virtual Reality in order to develop calibration-free pBCIs for workload assessment. Three pBCI classifiers were trained on EEG spectral power features from theta (4-8 Hz), alpha (8-13 Hz) and beta (13-30 Hz) bands, and an additional behavioral feature derived from pilots' control inputs on the (joy)stick. The models reached average classification accuracies of 0.82, 0.78, and 0.78. The key feature driving the models' performance was EEG theta power from several regions of the brain. The pilots' control inputs (i.e., behavioral feature) did not contribute to the model performance, however, it moderately correlated with several EEG theta power features. The results demonstrate the feasibility of a subject-independent pBCI for calibration-free classification of workload in pilots as well as the importance of theta power at frontal and centro-parietal areas as a metric for real-time monitoring of workload. The use of behavioral control inputs together with fewer but highly predictive EEG features warrants further research.**

Keywords—workload, aviation, Virtual Reality (VR), mental state prediction, passive brain-computer interface (pBCI)

I. Introduction

In recent years, progress has been made in the development of passive brain-computer interfaces (pBCIs) for mental state detection of operators for the purpose of improving operational safety and training [1], [2]. These pBCIs do not require active and conscious inputs from the pilots, and allow for adaptation of training environments and tasks to their (neurophysiological) responses [2]. Workload, which refers to the mental resources needed by an individual to perform and complete a given task [3]-[5], is a critical mental state during pilot training. An excessive increase in workload can negatively affect an individual's task performance and training outcomes [4], [5]. Therefore, pBCIs could be used in flight training programs to provide adaptive feedback to student pilots, or automatically adapt the difficulty of flight tasks to the student's level of proficiency [6].

Previous studies have attempted to predict workload in both novice (non-pilot) subjects and licensed pilots during simulated and real flight using electroencephalography (EEG) [7]-[10]. However, monitoring workload based on EEG data comes with serious methodological challenges. For instance, a ground truth must be established to generate distinctive classes from a mental state of interest [11], which usually depends on (potentially unreliable) subjective measures. More importantly, a problem that BCIs studies currently face is the (lack of) generalizability of findings [1], [2], [11]. BCIs are often highly subject-dependent. Some efforts have been made to automatically tailor BCI classifiers to each individual, e.g., by recording and using a lengthy set of calibration data per subject [12], or by designing subject-independent classifiers that are trained with EEG data from multiple subjects [13]. However, these methods come with their own disadvantages such as requiring a large amount of resources for the calibration of the classifier to each user, or in the case of a calibration-free approach, not working for all subjects [13].

The problem of subject-dependency of pBCI classifiers may be mitigated in the case of homogeneous groups of participants such as novice pilots with the same level of experience, gender, and age. In such cases, calibration of the classifier or correction for individual differences based on resting-state signals may not be needed. Determining whether baseline correction or subject-specific calibration is needed for accurate prediction of workload among a homogeneous group of pilots would therefore be desirable for effective development and application of pBCIs in simulated flight training environments. The addition of certain behavioral features as a supplementary source of information, such as

This research is funded by the MasterMinds project, part of the RegionDeal Mid- and West-Brabant, and co-funded by the Ministry of Economic Affairs and Municipality of Tilburg.

979-8-3503-1579-0/24/$31.00 ©2024 European Union

pilot control inputs, could possibly improve calibration-free prediction.

The current study focuses on this research gap and explores the possibility of a calibration-free pBCIs for prediction of workload in a group of novice military pilots who performed fight tasks in a Virtual Reality (VR) simulator. Unlike previous work that relied on subjective ratings from pilots or instructors as an estimate of their workload, e.g., [10], we used a dual-task paradigm to form a ground truth for low and high workload classes. Using this approach, we hope to overcome the challenges previously described, i.e., dependency on individual differences and subjective measures [11].

We trained three (offline) pBCI classifiers using EEG signals and control stick data from the pilots: 1) a subject-dependent model trained with baseline-corrected EEG features (conventional model in the literature), 2) a subject-independent classifier trained with absolute EEG spectral features (calibration-free model) and 3) a subject-independent model trained with absolute EEG features and additional behavioral feature (multi-modal calibration-free model). We compared the performance of the three models and further investigated the EEG features that were selected as the most performing ones in the optimization of the models.

II. Methods

A. Participants

In this study, six male student pilots participated, who had recently graduated from the Elementary Military Flight School of the Royal Netherlands Air Force (M_{age} = 25.00, SD_{age} = 6.36) with an average number of 34.33 (SD = 4.50) flight hours. Note that the participants were all recently graduated student pilots, all male, all of approximately similar age with a comparable number of flight hours.

The study was designed in accordance with the (revised) Helsinki Declaration and was approved by the Research Ethics and Data Management Committee of Tilburg School of Humanities and Digital Sciences and the Netherlands Organisation for Applied Scientific Research (TNO) institutional ethics committee.

B. Materials

A VR flight simulator of the single propellor and fixed-wing PC-7 aircraft was used (multiSIM B.V., the Netherlands)

Fig. 1. Experimental set-up. A) Image taken from the simulated (VR) environment. B) Participants conducted the experimental tasks while seated in a PC-7 cockpit mock-up, wearing the VR-headset and EEG equipment.

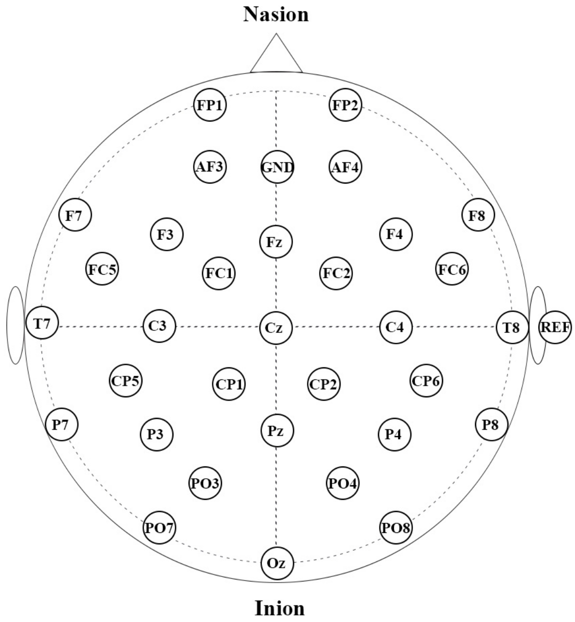

Fig. 2. Scalp map with electrode sites. The 32 electrodes were placed according to the 10-20 system. F = Frontal, FP = Pre-frontal, AF = Anterior-frontal, FC = Frontocentral, C = Central, CP = Centroparietal, T = Temporal, P = Parietal, PO = Parieto-occipital, O = Occipital, GND = ground electrode, REF = reference electrode.

(Fig. 1A), featuring a cockpit mock-up including stick and pedals with control loading (Fig. 1B).

Flight parameters and pilot control input were recorded with varying sample rates up to 500 Hz. The VR environment was displayed using the Varjo Aero headset (Varjo Technologies Oy, Finland). Neurophysiological signals were recorded with a wireless 32-channel EEG system at 250 Hz with a notch filter at 50 Hz (g.Nautilus PRO, g.tec medical engineering GmbH, Austria), see Fig. 2 for a scalp map with the electrode sites.

C. Procedure

Fig. 3 displays the experimental procedure of the study, which has previously been reported in detail in a behavioral study [14]. All participants performed twelve trials of a speed change task (i.e., deceleration from 180 to 110 Knots Indicated Air Speed) in the VR simulator. Each trial had a length of 210 seconds. Six of the twelve trials included an additional auditory N-back task to increase workload. This N-back task (2-back) required participants to remember the last two letters of an auditory sequence of continuously changing letters, and respond to letters that were identical to letters two trials back with the use of a button on the throttle. 'N-back' trials were categorized as high workload, and 'No N-back' trials were categorized as low workload. The order of six 'N-back' and six 'No N-back' trials was randomized using a block randomization method. Prior to this experiment, the participants joined another study in which they performed three flight tasks (i.e., straight-and-level, level turn, and a speed change) using the same set-up and additional N-back task, hence the participants were familiar with the tasks and there should be no novelty or learning effect present.

For a duration of 180 seconds, just before the onset of the trials, baseline EEG was recorded in which the participants were instructed to look at a black cross on a white background in the VR environment.

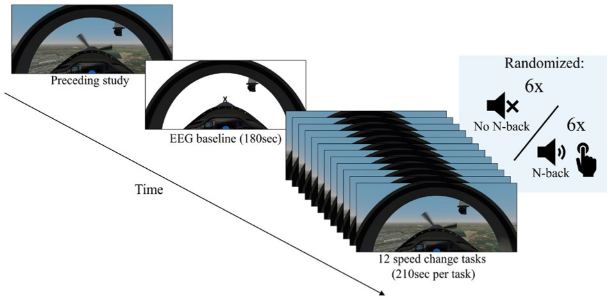

Fig. 3. Schematic outline of experimental procedure.

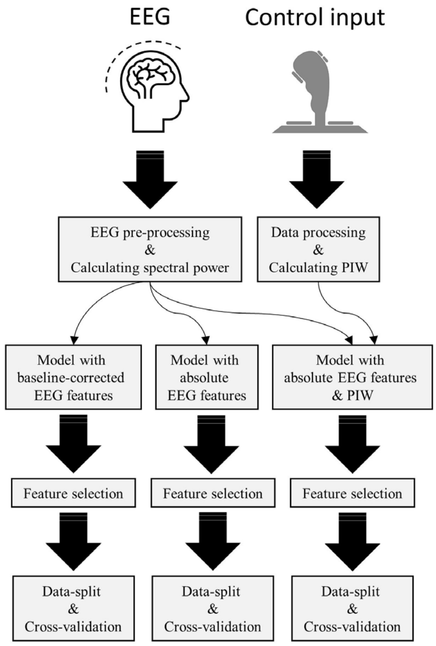

Fig. 4. Schematic processing pipeline. Includes data processing steps, feature extraction, feature selection and classification (cross-validation). PIW = Pilot Inceptor Workload.

D. Data pre-processing and feature extraction

All pre-processing steps were performed using MATLAB version R2022b and the Signal Processing Toolbox [15]. Code has been made available on Github [16]. The processing pipeline is schematically illustrated in Fig. 4.

1) Pre-processing of EEG data and feature extraction

EEGLAB v2022.1 [17] was then used to bandpass filter the raw EEG data (cutoff frequencies of 0.5 Hz and 45 Hz), to manually remove artifacts, and to conduct Independent Component Analysis (ICA) for automatic removal of eye-movement and other undetected artifacts.

Next, the EEGLAB "spectopo" function was used to compute spectral power values in theta (4-8 Hz), alpha (8-13 Hz) and beta (13-30 Hz) bands for all EEG channels per trial (i.e., absolute values) and for the baseline recording. For the calculation of baseline-corrected EEG features, the absolute spectral power values were baseline-corrected using a percentage-based method, similar to the method described in [18], see equation (1).

$$\textit{Baseline-corrected spectral power} = \frac{\textit{Absolute spectral power} - \textit{Baseline spectral power}}{\textit{Baseline spectral power}} \quad (1)$$

Additionally, EEG Engagement Indices [4], [19], were calculated using Equation (2) with the use of the absolute and the baseline-corrected spectral power values for each trial respectively:

$$\text{EEG Engagement Index} = \frac{\beta}{\alpha+\theta} \quad (2)$$

in which α is alpha power, β is beta power, and θ is theta power. This index has previously been reported to serve as an indicator of workload [4].

2) Behavioral data and feature extraction

The pre-processing of the behavioral data and feature extraction have previously been described in detail in [14]. The behavioral feature used in this study is called Pilot Inceptor Workload (PIW) [20], which reflects one's effort to control the aircraft with the use of the inceptor (stick). Accelerations in stick inputs and/or increasing amounts of inputs on the stick in a given period of time, increase the overall value of PIW [14], [20].

For this study, the control input in longitudinal direction (pitch, see Fig. 5) was included as a feature, because it was predictive of workload (classes based on 'N-back' and 'No N-back') and varied between the ('N-back' and 'No N-back') trials in our previous study [14]. Therefore, we hypothesized that this, computationally simple, behavioral measure could potentially increase the prediction accuracy in an EEG-based model.

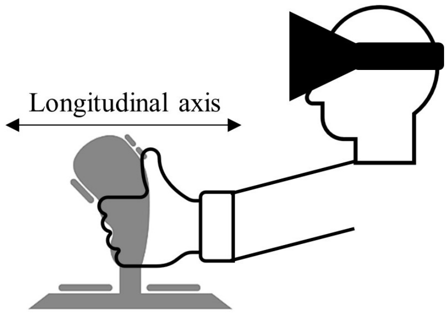

Fig. 5. Schematic representation of the longitunidal axis along the stick (forward-aft).

E. Feature selection

Our architecture, i.e., the pipeline of our feature selection and classifiers, is based on the methods described by [7]. The scikit-learn library for Python [21] was used to create a feature selection algorithm, as well as our classification models. Three classification models were created, trained and tested: (Model 1) subject-dependent pBCI trained with baseline-corrected EEG features, (Model 2) subject-independent pBCI trained with absolute EEG features, and (Model 3) multi-modal subject-independent pBCI trained with absolute EEG features and PIW behavioral feature.

The input features for the three models consisted of alpha, beta, theta spectral power values and the EEG Engagement Indices extracted from all 32 EEG channels, resulting in a total of 128 features (4 features × 32 channels) for Model 1 and

Model 2, and 129 features (128 + 1 PIW behavioral feature) for Model 3. This number of features was significantly high compared to the small size of the dataset. Therefore, to avoid the curse of dimensionality [22], the feature space had to be significantly reduced, which was performed using Recursive Feature Elimination (RFE). The RFE made use of a Support Vector Machine (SVM) classifier (kernel = ‘linear’, C = 1, gamma = 0.1) to indicate the most important features. Using RFE feature selection algorithm, we selected the ten best performing features per model, because a larger number could cause overfitting and the resulting models would be more difficult to interpret.

F. Classification of workload

A stacking classification algorithm was used to predict two levels of workload (low and high). The first level of the stacking classifier consisted of an ensemble model, containing a SVM, Random Forest (RF), and Logistic Regression (LR). The second, final, level was selected to be an SVM. Grid search with K-fold cross validation (k = 5) was used to optimize the hyperparameters for all individual classifiers (i.e., for the two SVMs, the RF, and the LR) in the stacking model. This resulted in different hyperparameters per iteration of the model.

Each model was trained and tested for a total of ten iterations, since the dataset was relatively small, and variability in the models’ performance metrics was expected. Each iteration, the train-test-split function was used for splitting the data into 70% train and 30% test data. Hyperparameters were tuned on the training data. The stacking classifier predicted the level of workload (low and high) on the remaining test data.

III. Results

A. Features and model performance

Table 1 gives the mean performance metrics (based on ten iterations) and the best-performing features for each workload classification approach. All three classifiers performed above chance-level for the classification of workload. The subject-dependent pBCI based on baseline-corrected EEG features resulted in the highest performance with an accuracy of 0.82, thus outperforming the subject-independent classifiers trained with either absolute EEG features or multi-modal EEG and behavioral features. Also, the variability of the performance metrics obtained by the subject-dependent classifier was overall lower, making it more reliable in comparison to the other two models that were trained and tested using absolute EEG data. However, the other two classifiers still achieved satisfactory prediction accuracy above the 70% threshold [23] for a subject-independent workload estimation task.

The ten features selected by the RFE algorithm as the most prominent features for each of the three models are also reported in Table 1. In Model 3, the PIW behavioral feature included in the initial feature space was not selected by the RFE algorithm and hence the ten selected features were equal to the EEG features of Model 2. This means that PIW was a less important predictor than the selected EEG features in this model, or that the behavioral feature may have correlated with and masked by redundant EEG features.

B. Post-hoc test for multicollinearity

To explore why the behavioral feature was not selected by RFE, a post-hoc test for multicollinearity between PIW and absolute EEG features was conducted in Rstudio [24] using

TABLE I. PERFORMANCE METRICS OF THE CLASSIFIERS

	Models		
	Subject-dependent pBCI (trained with baseline-corrected EEG spectral features)	***Subject-independent pBCI (trained with absolute EEG spectral features)***	***Multi-modal subject-independent pBCI (trained with absolute EEG spectral and behavioral features)***
Features selected by RFE	Alpha PO8, Beta PO4, Beta AF3, EEG Engagement FC2, **Theta FP1**, **Theta CP6**, Beta P7, Beta PO8, **Theta Oz**, **EEG Engagement T8**	Beta P8, Alpha P8, **Theta Oz**, **EEG Engagement T8**, Beta F8, Theta T7, EEG Engagement P8, **Theta CP6**, Beta FP2, **Theta FP1**	Beta P8, Alpha P8, **Theta Oz**, **EEG Engagement T8**, Beta F8, Theta T7, EEG Engagement P8, **Theta CP6**, Beta FP2, **Theta FP1**
Mean accuracy (*SD*)	0.82 (0.06)	0.78 (0.07)	0.78 (0.13)
Mean F1 score (*SD*)	0.82 (0.07)	0.76 (0.09)	0.74 (0.22)
Mean precision (*SD*)	0.82 (0.09)	0.78 (0.11)	0.71 (0.22)
Mean Recall (*SD*)	0.83 (0.11)	0.76 (0.14)	0.78 (0.25)

Note. SD = standard deviation. Bold faced features overlap between the three models. Selected features are in order of importance, with the first feature being the most important and the last feature being the least important of the selected features in the corresponding model. P = Parietal, O = Occipital, T = Temporal, F = Frontal, CP = Centroparietal, FP = Pre-frontal, PO = Parieto-occipital, AF = Anterior-frontal, FC = Frontocentral.

the corrplot package [25], and the Hmisc package [26]. Pearson’s correlations were computed for PIW and the overlapping features from the EEG models, i.e., Theta FP1, Theta CP6, Theta Oz, and EEG Engagement T8. We observed a negative correlation between PIW and Theta FP1, $r = -0.36$, $p = 0.002$, and between PIW and Theta CP6, $r = -0.41$, $p < 0.001$. The correlation matrix is displayed in Fig. 6. Other, stronger, correlations were found as well such as a positive correlation between Theta FP1 and Theta CP6 ($r = 0.74$, $p < 0.001$). Still, these variables have been selected by RFE to be included as features in the models. In other words, multicollinearity was likely not an issue, and the EEG features were more important than PIW in our models. This is in line with our previous work [14], where the explanatory power of PIW as a predictor of workload was significant, but weak.

IV. Discussion

In the current study, we examined the possibility of a calibration-free multi-modal pBCI for prediction of workload in a homogeneous group of novice pilots. EEG signals and behavioral stick control inputs were collected when the pilots performed flight tasks with high and low workload in a VR

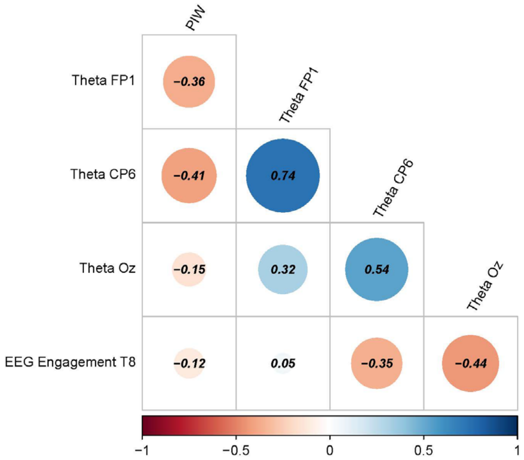

Fig. 6. Correlation matrix of features of interest showing r values. The matrix shows the negative (red) and positive (blue) relationships between variables. All correlations with r values greater than ± 0.3 were significant with $p < 0.05$.

simulation. The results from three models showed that it is possible to obtain a satisfactory mean accuracy of 78% in a subject-independent classification approach. The addition of a behavioral feature from pilot's stick control input to this model did not improve the classification results. These results indicate that baseline-correction and subject-specific classification approaches may not always be necessary for accurate classifications of workload in homogeneous subject groups, such as the sample used in this study. Our models showed little difference between the precision and recall scores, indicating that the models were largely unbiased towards predicting (true) positives and negatives. Additionally, more research is required to identify effective behavioral features that would enhance the prediction outcome in a multi-modal approach.

The three models reported in this study performed well in comparison to models from the existing pBCI literature, e.g., the study of Fraser et al. [8] reached an accuracy of 0.76 for an offline binary classification of workload using non-pilot subjects in a VR flight simulation. Multiple pBCI approaches were used, in which spectral power values of the theta, alpha and beta band at frontal and parietal sites were the most predictive of workload [8]. However, [8] reached this accuracy by removing some of the initially included electrodes and subjects, and separating sequential data.

The features that were selected by the RFE algorithm in our three models (i.e., Theta FP1, Theta CP6, Theta Oz, and EEG Engagement T8) were mostly associated with spectral power of the theta band. In line with our findings, theta spectral power has previously been associated with pilots' workload and performance in simulated flight tasks, commonly at frontal electrode sites [27]-[30], but also at central sites [30]. Parieto-occipital sites were of importance as well, likely because this region is generally associated with visual processing, attention [31], and visuo-spatial awareness [32]. Further, EEG Engagement Indices derived from temporal sites, which contain theta power as an element, have previously worked effectively in an adaptive system for a multi-tasking test battery that contains tasks analogous to piloting tasks [33]. However, contrary to our results, the temporal EEG Engagement Indices did not outperform the same indices derived from frontal or parietal recording sites [33]. As temporal EEG Engagement Index was an important feature in our study, the exact role of the temporal lobe in flight-related tasks remains unclear. We speculate that it may be related to the involvement of the medial temporal lobe in spatial navigation, as evidenced by intracranial theta oscillations [34], and single-neuron activity [35] in humans. Alternatively, the importance of the selected features could also be the result of differences in working memory load or auditory processing in relation to the presence or absence of the N-back task. For instance, theta coupling of the prefrontal-temporal network plays an important role in other working memory tasks, e.g., [36].

Additionally, our post-hoc tests showed support for an association between stick control input and theta band power in frontal and centro-parietal areas. Previously, Hebbar et al. [37] also found evidence of a relationship between theta band power and the percentage of time that stick inputs changed in a simulated flight task, as well as for the power of the lower frequencies of the beta band and the same measure of stick control input. However, these relationships in [37] were positive, as opposed to the negative correlations observed in our study. Considering that increased PIW is indicative of increased workload [14], and increases in theta power also reflect increases of workload [3], we would have expected a positive relationship between theta power and PIW. Perhaps our contradictory finding is because the absolute EEG spectral powers behave differently than the baseline-corrected powers, which are more commonly used in EEG studies. However it is unclear whether [37] has followed this practice and hence more research is needed to validate this assumption. Nonetheless, considering the lack of practicality of using a large number of EEG sensors for real-time workload assessment during (VR) flight training, the use of pilots' behavioral control inputs warrants further research. PIW could still be a useful addition to EEG-based models if fewer sensors with high predictive power are selected. By combining fewer EEG sensors with a less time-consuming behavioral feature such as PIW, both classification accuracy and time efficiency could be improved.

While the performance of our classifiers was satisfactory (accuracy > 0.75 in all three models), our study was limited by multiple factors. First, the N-back task was used to increase workload instead of applying manipulations to the task demands of the flight task. The N-back task required auditory processing, while the flight task required visuo-spatial processing. It may therefore be possible that our model classified the presence of auditory load instead of the overall task load. To be able to generalize these findings for future application in simulated flight training scenarios, multiple types of flight tasks [7], [10], as well as various training environments (VR or non-VR) [8]-[10], should be employed in follow-up studies. Secondly, the 'N-back' and 'No N-back' conditions only allowed for binary classifications of workload. As workload is commonly visualized on a continuous scale, multiclass classifications are generally more appropriate for pBCIs [9]. In this respect, a study with more variations in workload would be beneficial. Lastly, our dataset was relatively small, with a total of 72 observations from six pilots. In our sample, the addition of the EEG baseline correction did not considerably increase prediction accuracy, but this result may be different for a larger heterogeneous sample, such as one with more inter-subject variability. To build upon our current findings, we recommend conducting a study that features a higher number of participants, preferably with additional flight tasks.

Even though we cannot yet generalize our findings to real training applications, our findings show that it is possible to develop a subject-independent pBCI system for accurate prediction of workload in a homogeneous group of novice male military pilots. Our results also support the promise of EEG-based pBCIs in real-time monitoring of workload during flight training tasks, with theta power at frontal and centro-parietal areas being an important feature in the assessment of workload in military pilots. Such pBCIs will provide more insight into pilots' training trajectory as well as the ability to automatically adapt training systems to become more efficient and personalized for each user.

Acknowledgment

We would like to thank the Royal Netherlands Airforce and multiSIM B.V. for their help with this study.

References

[1] P. Aricò, G. Borghini, G. Di Flumeri, N. Sciaraffa, A. Colosimo, & F. Babiloni, "Passive BCI in operational environments: insights, recent advances, and future trends," IEEE Trans Biom Eng, vol. 64, no. 7, pp. 1431-1436, July 2017, doi: 10.1109/TBME.2017.2694856.

[2] F. Lotte, & R. Roy, "Brain-Computer Interface contributions to neuroergonomics," Neuroergonomics: the brain at work and in everyday life, Elsevier, pp. 43-48, 2019, URL: https://hal.inria.fr/hal-01946095.

[3] E. van Weelden, M. Alimardani, T. J. Wiltshire, & M. M. Louwerse, "Aviation and neurophysiology: a systematic review," J. Appl. Erg., vol. 105, no. 1, pp. 103838, 2022, doi: j.apergo.2022.103838

[4] M. S. Young, K. A. Brookhuis, C. D. Wickens, & P. A. Hancock, "State of science: mental workload in ergonomics," Erg., vol. 58, no. 1, pp. 1-17, doi: 10.1080/00140139.2014.956151

[5] A. P. G. Martins, "A review of important cognitive concepts in aviation," Aviat., vol. 20, no. 2, pp. 65-84, 2016, doi: 10.3846/16487788.2016.1196559.

[6] E. van Weelden, M. Alimardani, T. J. Wiltshire, and M. M. Louwerse, "Advancing the Adoption of Virtual Reality and Neurotechnology to Improve Flight Training," in *2nd IEEE Int. Conf. on Hum. Mach. Syst.,* Sep. 2021, doi: 10.1109/ICHMS53169.2021.9582658

[7] H. Taheri Gorji et al., "Using machine learning methods and EEG to discriminate aircraft pilot cognitive workload during flight," Sci. Rep., vol. 13, pp. 2507, 2023, doi: 10.1038/s41598-023-29647-0.

[8] A. Fraser, K. van Benthem, & C. M. Herdman, "A passive electroencephalography brain-computer interface predicts mental workload during flight simulation," in *Int. Sym. Av. Psych*., 2021, URL: https://corescholar.libraries.wright.edu/isap_2021/73

[9] Z. Pei, H. Wang, A. Bezerianos, & J. Li, "EEG-based multiclass workload identification using feature fusion and selection," IEEE Trans. Instrum. Meas., vol. 70, 2021, doi: 10.1109/TIM.2020.3019849.

[10] F. Dehais, B. Somon, T. Mullen, & D. E. Callan, "A neuroergonomics approach to measure pilot's cognitive incapacitation in the real world with EEG," in *AHFE 2020*, pp. 111-117, 2021, doi: 10.1007/978-3-030-51041-1_16.

[11] A-M. Brouwer, T. O. Zander, J. B. F. van Erp, J. E. Korteling, & A. W. Bronkhorst, "Using neurophysiological signals that reflect cognitive or affective state: six recommendations to avoid common pitfalls," Front. Neurosci., vol. 9, no. 136, April 2015, doi: 10.3389/fnins.2015.00136.

[12] Y. Liu, H. Ayaz, & P. A. Shewokis, "Multisubject "learning" for mental workload classification using concurrent EEGm, fNIRS, and physiological measures," Front. Hum. Neursci., vol. 11, no. 389, July 2017, doi: 10.3389/fnhum.2017.00389.

[13] B. Kenny, & S. D. Power, "Toward a subject-independent EEG-based neural indicator of task proficiency during training," Front. Neuroergon., vol 1, no. 618632, Jan. 2021, doi: 10.3389/fnrgo.2020.618632.

[14] E. van Weelden et al., "Differentiating workload using pilot's stick input in a Virtual Reality flight task," 2023, arXiv:2309.09619 [cs.HC].

[15] MATLAB and Signal Processing Toolbox Release R2022b, The MathWorks, Inc., Natick, Massachusetts, USA, 2022.

[16] E. van Weelden & C. W. E. van Beek, GitHub Respository, 2023, URL: https://github.com/evyvanweelden/EEG-PIW-WL.git.

[17] A. Delorme & S. Makeig, "EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis", J. Neurosci. Methods, vol. 134, no. 1, pp. 9-21, 2004, doi: 10.1016/j.jneumeth.2003.10.009.

[18] M. X. Cohen, "Chapter 18: time-frequency power and baseline corrections," in *Analyzing neural time series data: theory and practice*, MIT press, 2014, doi: 10.7551/mitpress/9609.001.0001.

[19] L. J. Prinzel, F. G. Freeman, M. W. Scerbo, O. J. Mikulka, & A. T. Pope, "A closed-loop system for examining psychophysiological measures for adaptive task allocation," Int. J. Aviat. Psychol., vol. 10, no. 4, pp. 393-410, 2000, doi: 10.1207/S15327108IJAP1004_6.

[20] I. Niewind, "Pilot gain and the workload buildup flight test technique: a closer investigation of Pilot Inceptor Workload", *DLR-Interner Bericht, Report No. IB 111-2012/74*, Oct. 2012, URL: https://elib.dlr.de/88208/.

[21] F. Pedregosa, et al., "Scikit-learn: machine learning in Python," HAL, 2011, URL: https://hal.inria.fr/hal-00650905.

[22] A. Crespo Márquez, "The curse of dimensionality," in *Springer Series in Reliability Engineering: Digital Maintenance Management*, Springer, Cham, 2022, pp. 67-86, doi: 10.1007/978-3-030-97660-6_7.

[23] B. Blankertz et al., "The Berlin brain-computer interface: non-medi al uses of BCI technology", Front. Neurosci., Sec. Neuroprosthetics, vol. 4, no. 198, 2010, doi: 10.3389/fnins.2010.00198.

[24] RStudio Team. RStudio: Integrated Development Environment for R. RStudio, PBC, Boston, MA, 2023, URL: http://www.rstudio.com/.

[25] T. Wei, & V. Simko, "R package 'corrplot': Visualization of a Correlation Matrix. (Version 0.92)", 2021, URL: https://github.com/taiyun/corrplot.

[26] F. E. Harrell Jr., & C. Dupont, "Hmisc: Harrell Miscellaneous. R Package Version 4.7-0," 2019, URL: https://hbiostat.org/R/Hmisc/.

[27] K. A. Feltman, K. A. Bernhardt, & A. M. Kelley, "Measuring the domain specificity of workload using EEG: auditory and visual domains in rotary-wing simulated flight," Hum. Fac., vol. 63, no. 7, pp. 1271-1283, 2021, doi: 10. 1177/ 0018 7208 20928626.

[28] G. Borghini, et al., "Frontal EEG theta changes assess the training improvements of novices in flight simulation tasks", in *35th Ann. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Osaka, Japan, 3-7 July, 2013, pp. 6619-6622, doi: 10.1109/EMBC.2013.6611073.

[29] A. Hamann & N. Carstengerdes, "Investigating mental workload-induced changes in cortical oxygenation and frontal theta activity during simulated flights," Sci. Rep., vol. 12, April 2022, Art. no. 6449, doi: 10.1038/s41598-022-10044-y.

[30] C. Diaz-Piedra et al., "The effects of flight complexity on gaze entropy: an experimental study with fighter pilots," J. Appl. Erg., vol. 77, pp. 92-99, May 2019, doi: 10.1016/j.apergo.2019.01.012.

[31] M. Behrmann, J. J. Geng, & S. Shomstein, "Parietal cortex and attention," Curr. Opin. Neurobiol., vol. 14, pp. 212-217, no. 2, 2004, doi: 10.1016/j.conb.2004.03.012.

[32] C. Babiloni, F. Vecchio, M. Miriello, G. L. Romani, & P. M. Rossini, "Visuo-spatial consciousness and parieto-occipital areas: a high-resolution EEG study," Cereb. Cortex, vol. 16, pp. 37-46, 2006, doi: 10.1093/cercor/bhi082.

[33] F. G. Freeman, P. J. Mikulka, M. W. Scerbo, L. J. Prinzel, & K. Clouatre, "Evaluation of a psychophysiologically controlled adaptive automation system, using performance on a tracking task," Appl. Psychophysiol. Biofeedback, vol. 25, no. 2, pp. 103–115, 2000, doi: 10.1023/A:1009566809021.

[34] Z. M. Aghajan et al., "Theta oscillations in the human Medial temporal lobe during real-world ambulatory movement," Curr. Biol., vol. 27, pp. 3743-3751, Dec. 2017, doi: 10.1016/j.cub.2017.10.062.

[35] L. Kunz et al., "A neural code for egocentric spatial maps in the human medial temporal lobe," Neuron, vol. 109, no. 17, pp. 2781-2796, 2021, doi: 10.1016/j.neuron.2021.06.019.

[36] P. Sauseng et al., "Theta coupling in the human electroencephalogram during a working memory task," Neurosci. Lett., vol. 354, no. 2, pp. 123-126, 2004, doi: 10.1016/j.neulet.2003.10.002.

[37] P. A. Hebbar, K. Bhattacharya, G. Prabhakar, and P. Biswas, "Correlation between physiological and performance-based metrics to estimate pilots' cognitive workload," Front. Psychol., vol. 12, April 2021, Art. no. 555446, doi: 10.3389/fpsyg.2021.5554