*The Aeronautical Journal* (2024), **128**, pp. 2561–2576
doi:10.1017/aer.2024.36

ROYAL AERONAUTICAL SOCIETY

**RESEARCH ARTICLE**

# $\beta$-wave-based exploration of sensitive EEG features and classification of situation awareness

C. Feng[1,2], S. Liu[1], X. Wanyan[1], Y. Dang[1], Z. Wang[1] and C. Qian[1]

[1]School of Aeronautic Science and Engineering, Beihang University, Beijing, China
[2]Tianmushan Laboratory, Hangzhou, China
**Corresponding author:** X. Wanyan; Email: wanyanxiaoru@buaa.edu.cn

**Received:** 30 September 2023; **Revised:** 16 March 2024; **Accepted:** 25 March 2024

**Keywords:** situation awareness; measurement and classification; EEG features; interpretability; flight safety

## Abstract

The purpose of this study was to explore the electroencephalogram (EEG) features sensitive to situation awareness (SA) and then classify SA levels. Forty-eight participants were recruited to complete an SA standard test based on the multi-attribute task battery (MATB) II, and the corresponding EEG data and situation awareness global assessment technology (SAGAT) scores were recorded. The population with the top 25% of SAGAT scores was selected as the high-SA level (HSL) group, and the bottom 25% was the low-SA level (LSL) group. The results showed that (1) for the relative power of $\beta_1$ (16–20Hz), $\beta_2$ (20–24Hz) and $\beta_3$ (24–30Hz), repeated measures analysis of variance (ANOVA) in three brain regions (Central Central-Parietal, and Parietal) $\times$ three brain lateralities (left, midline, and right) $\times$ two SA groups (HSL and LSL) showed a significant main effect for SA groups; post hoc comparisons revealed that compared with LSL, the above features of HSL were higher. (2) for most ratio features associated with $\beta_1 \sim \beta_3$, ANOVA also revealed a main effect for SA groups. (3) EEG features sensitive to SA were selected to classify SA levels with small-sample data based on the general supervised machine learning classifiers. Five-fold cross-validation results showed that among the models with easy interpretability, logistic regression (LR) and decision tree (DT) presented the highest accuracy (both 92%), while among the models with hard interpretability, the accuracy of random forest (RF) was 88.8%, followed by an artificial neural network (ANN) of 84%. The above results suggested that (1) the relative power of $\beta_1 \sim \beta_3$ and their associated ratios were sensitive to changes in SA levels; (2) the general supervised machine learning models all exhibited good accuracy (greater than 75%); and (3) furthermore, LR and DT are recommended by combining the interpretability and accuracy of the models.

## Nomenclature

| | |
|---|---|
| EEG | Electroencephalogram |
| SA | Situation Awareness |
| MATB | Multi-Attribute Task Battery |
| HSL | High-SA Level |
| LSL | Low-SA Level |
| SAGAT | Situation Awareness Global Assessment Technology |
| ANOVA | Analysis Of VAriance |
| DT | Decision Tree |
| LR | Logistic Regression |
| RF | Random Forest |
| ANN | Artificial Neural Network |
| HF | Human Factors |
| ECG | Electrocardiographic |
| HRV | Heart Rate Variability |
| PSD | Power Spectral Density |

| | |
|---|---|
| SVM | Support Vector Machine |
| ANN | Artificial Neural Network |
| BNN | Bayesian Neural Network |
| LDA | Linear Discriminant Analysis |
| GNB | Gaussian Naive Bayes |
| KNN | K-Nearest Neighbors |
| SD | Standard Deviation |
| EOG | Electrooculogram |
| ICA | Independent Component Analysis |
| FFT | Fast Fourier Transform |
| $\delta$ | Delta |
| $\theta$ | Theta |
| $\alpha$ | Alpha |
| $\sigma_1$ | Sigma1 |
| $\sigma_2$ | Sigma2 |
| $\beta_1$ | Beta1 |
| $\beta_2$ | Beta2 |
| $\beta_3$ | Beta3 |
| PCA | Principal Component Analysis |
| PCs | Principal Components |

## 1.0 Introduction

Situation awareness (SA) is commonly defined as 'the perception of elements in the environment within a volume of time and space, the comprehension of their meaning, and the projection of their status in the near future' [1]. Seventy to eighty percent of aviation accidents can be attributed to human factors (HF) in the chain of causation [2, 3], among which 88% was found to be due to problems with SA [4]. Real-time measurement and accurate classification of the operator's SA are critical issues in SA research, and related research can be used to support the design of triggering mechanisms for dynamic human-machine function allocation in complex industrial systems [5, 6], which can serve to improve the effectiveness of human-system integration and ensure human factors safety. Considering that operators may exhibit undesirable SA levels during the execution of a task, and there is still a relative lack of research on the exploration of sensitive physiological features and state classification of SA levels, this direction has received extensive and continuous attention from researchers and practitioners in the field of HF [7, 8].

The measurement of SA can be categorised into inquiry method, subjective scoring, modeling and physiological measures [9]. Among them, the situation awareness global assessment technique (SAGAT) of the inquiry method is commonly regarded as the gold standard of SA measurement. However, SAGAT suffers from the drawback of requiring task interruption or freezing, which prevents its adoption in real scenarios and frontline operational environments [10]. On the other hand, subjective scoring and modeling methods suffer from shortcomings such as high subjectivity and poor scalability. Accompanied by the recent development of machine learning techniques for classifying critical human states (e.g., SA) and considering that physiological measures have the advantages of real-time signals and high sampling rates, the method has received increasing attention [11].

Typical physiological measures include electrocardiographic (ECG), eye movement, electroencephalogram (EEG), etc [12]. Heart rate variability (HRV) of ECG has a strong relationship with executive functioning (especially working memory), which is often considered an essential component of SA [13, 14]. Previous studies have found that HRV decreases during task execution but less in SA-training groups [14]. In addition, individuals with higher resting HRV typically have higher SA [14]. Thus, HRV has the potential to predict and assess SA from the perspective of cognitive workload [10, 15]. Eye tracking is essential for the visual perception of information, and researchers have noted

https://doi.org/10.1017/aer.2024.36 Published online by Cambridge University Press

its potential to characterise vision-related SA [16, 17]. An air traffic control study suggested that the attention allocation pattern (i.e. widely or narrowly allocated) affects the operator's SA and that more fixations on a target aircraft lead to higher SA for the aircraft [18]. Additionally, in SA-related studies, eye movement measures are often considered to be correlated with direct SA measures [12]. Compared with the aforementioned physiological measures, EEG is becoming one of the most popular methods for measuring SA because of its high temporal resolution, objective measurement of the operator's neural activity, and strong connection with cognitive processes related to SA [7, 19].

The literature suggests that $\beta$ absolute or relative power may be associated with or senstive to SA. Kästle et al. [8] observed a correlation between $\beta$ band (absolute power) and SA [8]. Feng et al. [7] highlighted that $\beta$ relative power, rather than $\beta$ absolute power is senstive to SA [7]. Absolute power provides information about the overall strength of neural activity within a specific frequency range but does not account for changes in the overall EEG spectrum. In contrast, relative power focuses on the contribution of activity within a specific frequency range relative to the overall EEG activity, making it inherently relative. The utility of $\beta$ absolute or relative power in assessing SA requires further investigation. Furthermore, the $\beta$ band can be subdivided into more refined categories such as $\beta_1$, $\beta_2$ and $\beta_3$ [20, 21], and research on their sensitivity to SA is still lacking. Although several EEG features (e.g., $\beta$ waves) in various brain regions have been noted to be associated with or sensitive to SA [7, 8], there is still a relative lack of SA-related EEG studies [12], particularly in regard to the sensitivity of more refined $\beta_1$, $\beta_2$ and $\beta_3$ bands.

After the sensitive EEG features for SA were explored, they were further combined with machine learning algorithms to classify SA levels. Li et al. [22] selected power spectral density (PSD) from 1,674 samples of 11 participants as inputs and used the response time of performance as the label to carry out fatigue-related classification of high and low SA levels [22]. Compared with the results of classification algorithms such as logistic regression (LR) and support vector machine (SVM), it was found that the subject-matching framework proposed in the study obtained the best accuracy (74.32%) [19]. In addition, Kästle et al. [8] used PSD from 2,880 samples of 32 participants as inputs, high and low SA groups divided by SAGAT scores as outputs and adopted random forest (RF) and boosted trees for model construction and reached an accuracy 67% [8]. Feng et al. [7] used SA-sensitive EEG features based on $\beta$-wave from 24 sample points of 46 participants as inputs, high and low SA groups classified by SAGAT scores as outputs and the model of Bayesian classification algorithm achieved an accuracy of 70.8% [7]. Li et al. [23] used EEG and eye tracking data from 2,240 samples of 28 participants as inputs, and high and low SA were classified by positional deviation or angular error of performance as labels [23]. SVM, LR, RF, artificial neural network (ANN), and Bayesian neural network (BNN) were further used for classification. The results showed that the best accuracy for high and low SA classification was 76.1%, and the best accuracy for recognising low SA associated with high workload reached 82.7% [23].

Based on the background mentioned above, it can be found that the current studies are still deficient in the following aspects. First, on SA-sensitive features, although some studies have pointed out that there is a correlation between the $\beta$-wave features of EEG and SA or that they are sensitive to SA, there still exists a gap for exploring the more refined $\beta_1$, $\beta_2$ and $\beta_3$ wave features. Second, current SA classification models still differ in selecting input features and labels, which may bring considerations about the models' interpretability and rationality. For inputs of classification models, PSD features or SA-sensitive EEG features [7, 8, 19, 22] were often adopted, which may lead to differentiated classification results. Among them, the selection of various EEG features for SA classification remains to be further explored.

In addition, researchers chose performance features or SAGAT scores in a psychological paradigm [8], and performance metrics in a driving context [19, 22, 23] as output labels to generate relatively large-sample data for SA classification. Considering the unclear relationship between SA and performance, the latter may raise considerations about the rationality of labeling. For small-sample data in an aviation task scenario, less research concentrates on selecting SA-sensitive EEG features as model inputs and adopting SAGAT scores as labels for classification modeling. Last, supervised machine learning

https://doi.org/10.1017/aer.2024.36 Published online by Cambridge University Press

classification algorithms have been employed individually or partially in constructing SA classification models, with corresponding classification accuracies ranging from 67% to 82.7%, regardless of their interpretability. There is a lack of comparisons of the classification accuracies of generalised algorithms with different interpretability, including algorithms with easy interpretability (LR, decision tree (DT), linear discriminant analysis (LDA), linear kernel function based SVM, gaussian naive Bayes (GNB)) and algorithms with hard interpretability (RF, ANN, k-nearest neighbors (KNN)) [24].

The research content in this paper mainly includes (1) exploring the EEG features sensitive to SA; (2) applying typical supervised machine learning algorithms on classifying SA levels based on the above sensitive EEG features; and (3) comparing the classification accuracies of algorithms with different interpretability. Based on the above scenario, the following hypotheses are proposed:

Hypothesis 1 (H1): The absolute power of $\beta_1$, $\beta_2$ and $\beta_3$ is sensitive to changes in SA levels.

Hypothesis 2 (H2): The relative power of $\beta_1$, $\beta_2$ and $\beta_3$ and the corresponding ratio features are sensitive to changes in SA levels.

Hypothesis 3 (H3): SA-sensitive EEG features associated with $\beta_1$, $\beta_2$ and $\beta_3$ were selected to carry out small-sample classification of SA levels, and all of the above generalised supervised machine learning algorithms can obtain good classification accuracies.

## 2.0 Methods

### *2.1 Participants*

Forty-eight graduate students ($23.15 \pm 0.72$ years, mean $\pm$ standard deviation (SD)) with aviation backgrounds were selected as participants (including nine females) from Beihang University. They all had right-hand dominance, good health, normal or corrected-to-normal vision and at least 8 hours of sleep the night before the test. A comprehensive evaluation of participants' sleep status was conducted in the current study, including: (a) pre-experiment inquiries on time to fall asleep, sleep duration, emotion and arousal states, diet, medication, etc.; (b) naturalistic observations during the experiment of the current state, drowsiness, expressions, body behaviours (e.g., sitting posture, yawning), task performance, etc.; and (c) post-experiment documentation of subjective status reporting, eye movement and EEG sampling status, and the aforementioned details. The Biological and Medical Ethics Committee at Beihang University gave its approval to this study (approval number: BM20220176).

### *2.2 Experimental platform and EEG equipment*

System monitoring, target tracking, scheduling display and resource management are four common subtasks in multi-attribute task battery (MATB) II [25]. Keyboard, mouse and flying joystick controls are available for interacting with the platform, as shown in Fig. 1. EEG data from 32 channels, including F7, FT7, T3, TP7, T5, A1, FP1, F3, FC3, C3, CP3, P3, O1, FZ, FCZ, CZ, CPZ, PZ, OZ, FP2, F4, FC4, C4, CP4, P4, O2, F8, FT8, T4, TP8, T6 and A2, were recorded using the Neuroscan Neuamps amplifier (Compumedics limited, Victoria, Australia). Ag/AgCl was used for all electrodes. The recorded wave was between 0.1 and 200Hz, and 1,000Hz was used for sampling. It was kept under 5kΩ for the electrode impedance. The left mastoid A1 was used as the reference point, and ground (GND) served as the ground. The electrooculogram (EOG)'s vertical and horizontal channels were concurrently recorded. Typical EEG procedures include: (a) ensuring participants have washed and blow-dried their hair before the experiment or instructing them to do so promptly; (b) selecting reference electrodes; (c) locating eye movement recording positions by placing electrodes 1cm from the outer corners of both eyes for horizontal movement and above the eyebrows and below the lower edge of the eye sockets for vertical movement; (d) choosing an electrode cap based on the participant's head circumference, worn from front to back with the Cz point at the centre; (e) applying approximately 0.5ml of conductive gel to the ground and reference electrodes, followed by other electrode points; (f) configuring data collection parameters and initiating EEG recording once satisfactory impedance is achieved.

https://doi.org/10.1017/aer.2024.36 Published online by Cambridge University Press

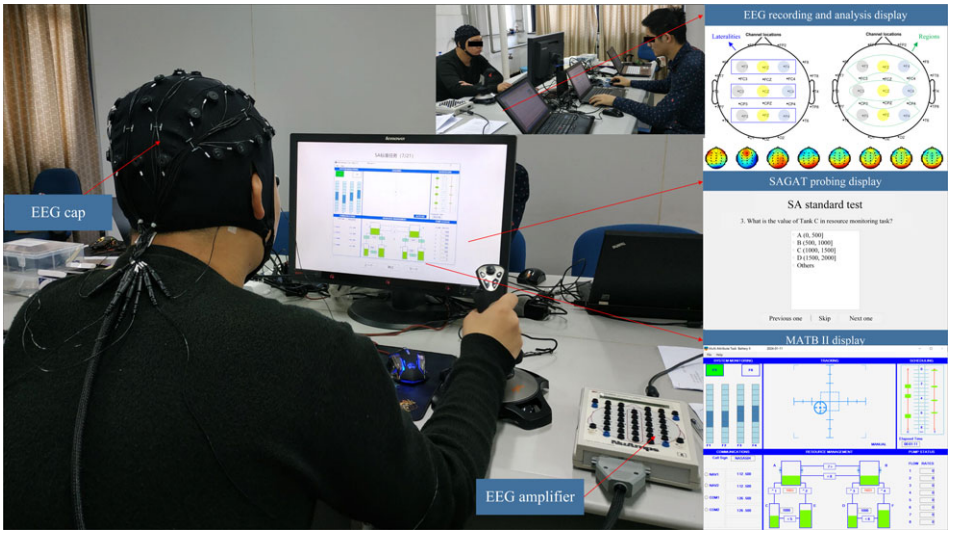


***Figure 1.*** *Platform, measurement and analysis system displays and experimental scenarios.*

### *2.3 Experimental design*

On the above basis, 48 participants were recruited to complete an SA standard test in an aviation task scenario based on MATB II, and the corresponding EEG data and SAGAT scores were recorded. The population with the top 25% of SAGAT scores was selected as the high-SA level (HSL) group, and the bottom 25% was selected as the low-SA level (LSL) group. The HSL group included 12 males and no females, whereas the LSL group comprised 7 males and 5 females in terms of gender distribution. The SA level was the independent variable in our experimental design, whereas the EEG features were the dependent variable. High- and low-SA groups were categorised based on SAGAT scores on the MATB II tasks, thus attaining the SA levels.

The SAGAT freezing interface occupied the entire screen, and participants responded to questions using a mouse, as shown in Fig. 1. The MATB II, derived from the primary flight task [25], underwent a redesign and simplification, as follows: (a) in the tracking monitoring task, a two-dimensional random-input compensatory task displayed the circular reticle's movement, with manual programming for mode switching. Participants were required to press the left arrow on the keyboard when transitioning from 'auto-pilot' to 'manual'. (b) For the system monitoring task, participants identified any scale displaying a steady 'extreme' position at the upper or lower end by clicking on the corresponding scale. (c) In the resource monitoring task, participants responded to pump failures by clicking on the corresponding pump using the left mouse button. (d) In the communication monitoring task, participants monitored communication task quantities on the schedule line and pressed the right arrow on the keyboard when the upper edge of the communication block reached the zero lines (located on the left of the scheduling display, marked in green).

### *2.4 Experimental procedure*

The experiment consisted of two stages: training and the formal experiment. In the training stage, all participants received adequate training and took an assessment test, of which the answers' accuracy was required to be above 90% before they were allowed to participate in the formal experiment. Each participant had to perform the resting test with EEG sensors first. Before the experiment, all participants completed an informed consent form after receiving information about the procedure and instructions. In the formal stage, the computerised freezing technique-based SAGAT measurements were used to

https://doi.org/10.1017/aer.2024.36 Published online by Cambridge University Press

assess participants' SA levels during MATB II tasks [10]. Specifically, at some random time during the experiment, the MATB II interface was frozen and covered by the given inquiry interface. Six questions were asked at each interval of presentation freezing, which lasted around five minutes. Questions were generally shown for less than a minute, and there were more than two minutes between each freezing [26]. No questions would be asked in the first three minutes, which guaranteed the participant to gather information in a secure environment [26]. The participant used mouse clicks to provide an answer to the pop-up query that had been chosen from the SAGAT database.

## 3.0 Results

### *3.1 Quantitative β-based EEG analysis of different SA groups*

#### *3.1.1 Data analysis methods*

The average of the bilateral mastoid electrodes offline was used as the new reference for the EEG signal. The bandpass filtering of 1–30Hz was used, and the analysis was carried out using the toolbox MATLAB R2020a eeglab v2023.0 (accessible via https://sccn.ucsd.edu/eeglab/download/daily/). Independent component analysis (ICA) was used to evaluate the EEG signals. The Adjust 1.1.1 toolbox, proven effective in mitigating EEG artifacts as indicated by various studies [27, 28], was employed to remove artifacts, including horizontal eye movement, vertical eye movement, eye blink and generic discontinuity, as defined in Mognon et al. [29]. Access the toolbox at https://www.nitrc.org/frs/?group_id=739&release_id=3054#. As shown in Fig. 2, Fast Fourier transform (FFT) was carried out after segmenting the EEG data. After that, eight waves were created by averaging the segmented data: 1–4Hz for the delta ($\delta$) wave, 4–8Hz for the theta ($\theta$) wave, 8–12Hz for the alpha ($\alpha$) wave, 12–14Hz for the sigma1 ($\sigma_1$), 14–16Hz for the sigma2 ($\sigma_2$), 16–20Hz for the beta1 ($\beta_1$), 20–24Hz for the beta2 ($\beta_2$) and 24–30Hz for the beta3 ($\beta_3$) wave [21, 30, 31]. The total band power was calculated from the integration of the power across the defined spectrum. Microvolts squared per octave ($\mu V^2$) served as the unit of power analysis.

The EEG acquisition device was used throughout the experiment to capture the EEG data in real-time. The EEG spectral features were the eight bands' absolute power ($\mu V^2$), relative power (%) and $\beta_1 \sim \beta_3$ related ratios. The ratio of a wave's absolute power to the total wave's absolute power (1–30Hz) is known as the relative power of that wave. In addition, $\beta_1 \sim \beta_3$ related ratio feature selection was conducted according to the slow and fast wave features. Since δ oscillations are closely associated with states of diminished consciousness [32], and no significant main effects were found in the subsequent absolute and relative power, they were not conceived in this study. Based on the above, 72 $\beta_1 \sim \beta_3$ related ratio features in three categories were selected, including single, double and triple denominator categories. After EEG processing, subsequent steps involving data preprocessing, classifier application and performance evaluation are further conducted. Detailed descriptions for these processes are provided in the subsequent sections.

Previous studies showed that the relative power of the $\beta$ wave in C and P regions is sensitive to changes in SA. Therefore, in our investigation, electrode placements in three different brain regions – central: C3, Cz, C4; central-parietal: CP3, CPz, CP4; and parietal: P3, Pz, P4 – were taken into account for analysis [7, 8]. Additionally, studies frequently fixed on the differences between central, right and left brain literalities [33, 34]; so three lateralities (left: C3, CP3, P3; central: Cz, CPz, Pz; right: C4, CP4, P4) were also investigated. Statistical analysis was conducted using Statistical Package for the Social Sciences (SPSS) Statistics 25.0 (IBM, United States), and the confidence levels for statistical tests were set to $\alpha = 0.05$. Repeated-measures analysis of variance (ANOVA) was adopted in three brain regions (central, central-parietal and parietal) × three brain lateralities (left, middle and right) × two SA levels (LSL and HSL) to investigate the interaction and main effects of SA levels on the features. When interaction and main effects were significant, the Bonferroni method was used for simple effects analysis and post hoc analysis. Sphericity was tested using the Mauchly method. If the hypothesis was not accepted, it was adjusted using the Greenhouse-Geisser method [35, 36].

https://doi.org/10.1017/aer.2024.36 Published online by Cambridge University Press

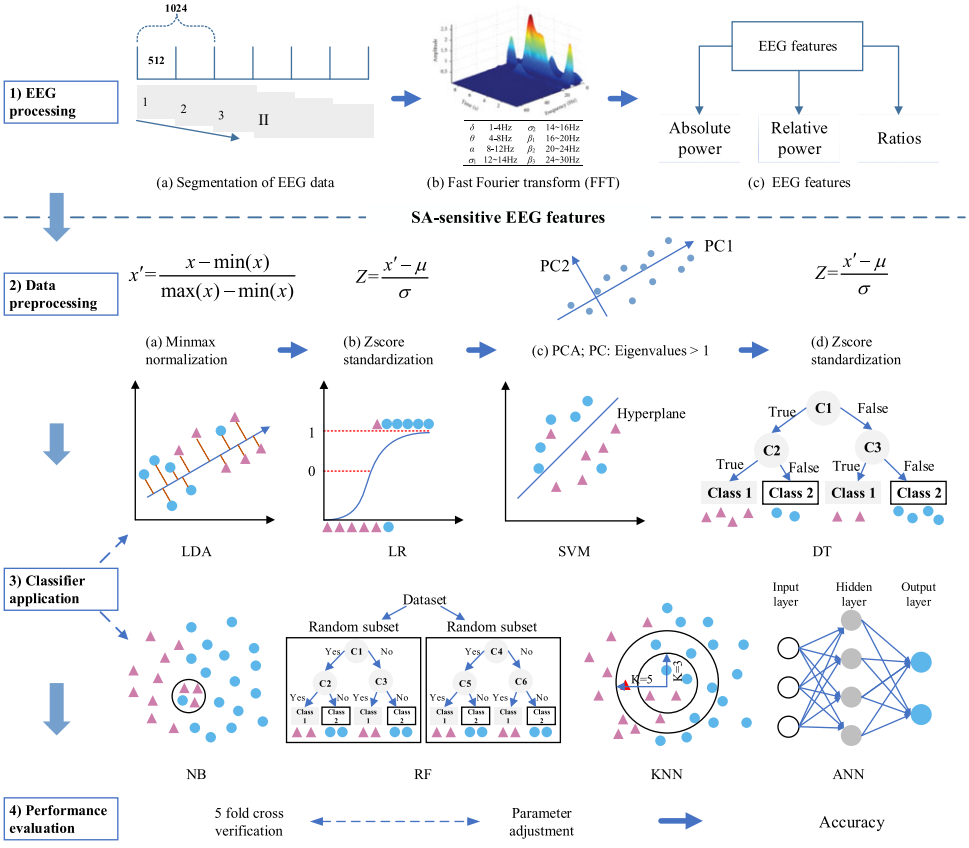


***Figure 2.*** *Steps for EEG feature processing and classification of SA levels.*

*3.1.2 Absolute power of different SA levels*

The topographical maps in Fig. 3 have clarified the average values of absolute and relative power of EEG spectrum features for all participants at two SA levels. Brighter hues (in this case, red) signify electrode sites with higher activations, while blue denotes electrode sites with lower activations.

The absolute power of EEG features under the LSL and HSL were selected for ANOVA, and the results indicated no significant interaction effect ($ps > 0.05$). For the $\theta$ absolute power, the main effect of SA levels was significant ($F(1, 22) = 4.351$, $p = 0.049$, $\eta^2 = 0.165$). Post hoc comparisons showed that the $\theta$ absolute power for HSL was significantly lower than that for LSL. As indicated in Table 1, the main effects of SA levels for the absolute power in the other waves were also insignificant ($ps > 0.05$).

*3.1.3 Relative power results of different SA levels*

Figure 4 depicts descriptive results of the relative power in eight waves for various SA levels, in which * represents that $p \leq 0.001$. As shown in Table 2, ANOVA results demonstrated a significant interaction effect of the $\beta_3$ relative power for the laterality × SA level ($F(1.951, 42.911) = 4.313$, $p = 0.020$, $\eta^2 = 0.164$). As the result of simple effect analysis, for HSL, $\beta_3$ relative power on the right side of the brain was significantly higher than that on the left side and the midline of the brain ($p = 0.017$, $p < 0.001$), with no significant difference between the left side and the midline of the brain ($p > 0.05$). For LSL, $\beta_2$ relative power on the left side of the brain was significantly higher than that on the midline of the brain ($p = 0.026$), with no other significant difference ($ps > 0.05$). The interaction effects of the relative power

https://doi.org/10.1017/aer.2024.36 Published online by Cambridge University Press

***Table 1.*** *ANOVA results of absolute power for the two SA levels*

| Waves ($\mu V^2$) | Region × laterality × SA level | | | Region × SA level | | | Laterality × SA level | | | SA level | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ***F*** | ***p*** | $\eta^2$ | ***F*** | ***p*** | $\eta^2$ | ***F*** | ***p*** | $\eta^2$ | ***F*** | ***p*** | $\eta^2$ |
| δ | 1.063 | 0.358 | 0.046 | 0.987 | 0.338 | 0.043 | 1.714 | 0.198 | 0.072 | 3.6 | 0.071 | 0.141 |
| $\theta$ | 0.396 | 0.693 | 0.018 | 0.018 | 0.915 | 0.001 | 2.022 | 0.155 | 0.084 | 4.351 | **0.049*** | 0.165 |
| α | 0.230 | 0.815 | 0.010 | 0.760 | 0.400 | 0.033 | 0.061 | 0.940 | 0.003 | 0.885 | 0.357 | 0.039 |
| $\sigma_1$ | 0.202 | 0.837 | 0.009 | 0.290 | 0.608 | 0.013 | 0.180 | 0.806 | 0.008 | 0.715 | 0.407 | 0.031 |
| $\sigma_2$ | 0.138 | 0.904 | 0.006 | 0.473 | 0.506 | 0.021 | 0.399 | 0.648 | 0.018 | 0.166 | 0.688 | 0.007 |
| $\beta_1$ | 0.534 | 0.581 | 0.024 | 0.035 | 0.875 | 0.002 | 0.790 | 0.432 | 0.035 | 0.115 | 0.737 | 0.005 |
| $\beta_2$ | 0.071 | 0.905 | 0.003 | 0.305 | 0.621 | 0.014 | 2.056 | 0.157 | 0.085 | 2.003 | 0.171 | 0.083 |
| $\beta_3$ | 0.418 | 0.594 | 0.019 | 0.926 | 0.361 | 0.040 | 0.844 | 0.405 | 0.037 | 3.723 | 0.067 | 0.145 |

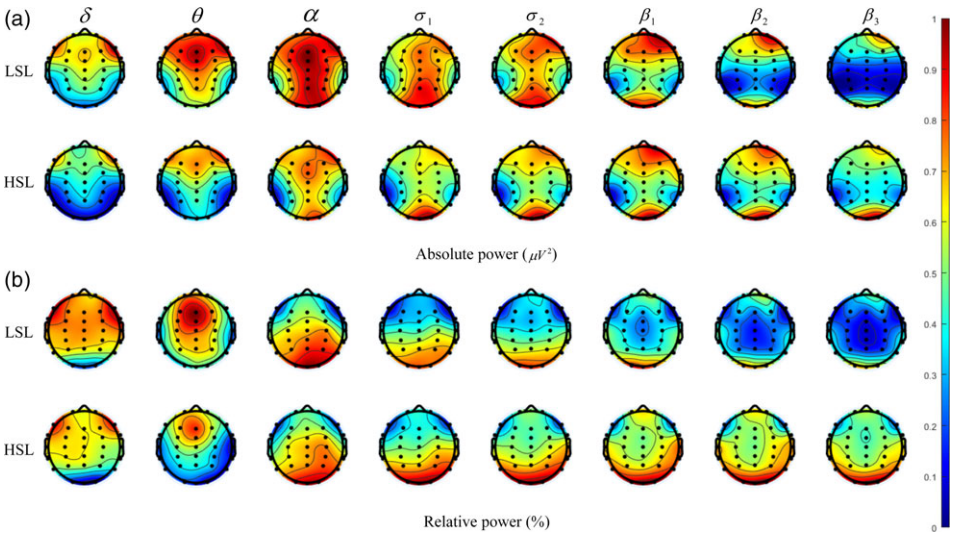


***Figure 3.*** *Topographical representation of average values of absolute and relative power of EEG spectrum features (scores on the color bar indicate the logarithm for the spectral power represented in each topography).*

in the other waves were not significant. In addition, for $\theta$, $\beta_1$, $\beta_2$ and $\beta_3$ relative power, the main effects of SA levels were significant ($p = 0.027$, $p = 0.027$, $p = 0.007$ and $p = 0.003$). Post hoc comparisons revealed that $\beta_1$, $\beta_2$ and $\beta_3$ relative power for LSL were significantly lower than those for HSL; and $\theta$ relative power for LSL was significantly higher than that for HSL.

*3.1.4 Ratios results of different SA levels*

For the ratio features, ANOVA showed that the main effects of SA levels were all significant ($ps < 0.05$) except six features, including $\alpha/\beta_1, \sigma_1/\beta_1, \sigma_2/\beta_1, (\alpha+\sigma_1)/\beta_1, (\sigma_1+\sigma_2)/\beta_1, (\alpha+\sigma_1+\sigma_2)/\beta_1$ and $\alpha/(\beta_1+\beta_2)$. Detailed table format of ratios can be found in a Supplementary Materials.

### *3.2 Classification of SA levels*

An independent sample t-test was conducted for the above typical EEG features. Considering the large amount of corresponding SA-sensitive features, this study set $p \leq 0.001$ as the selection criteria and obtained 57 model input features, as shown in Table 3. Furthermore, supervised learning algorithms

https://doi.org/10.1017/aer.2024.36 Published online by Cambridge University Press

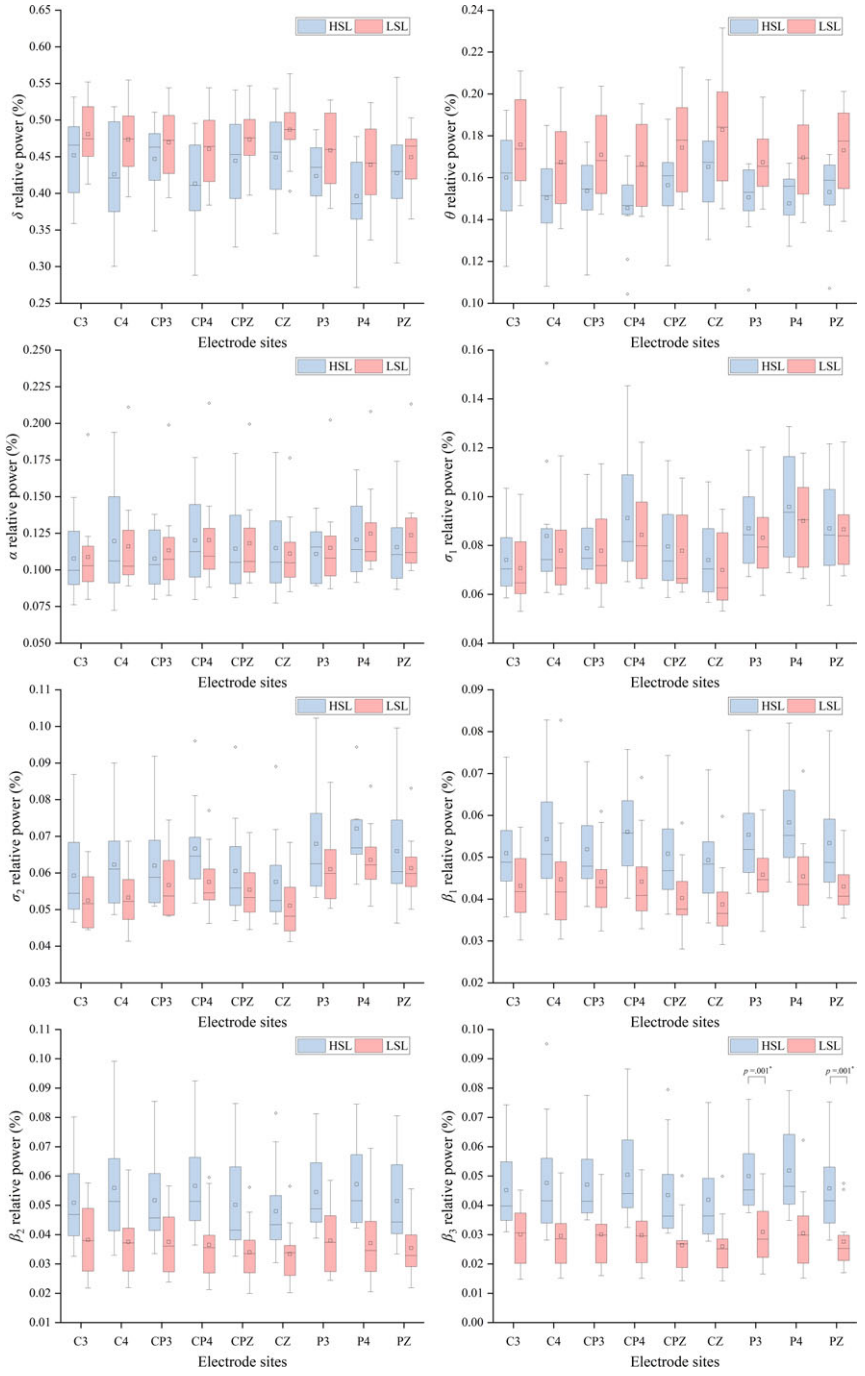


***Figure 4.*** *Relative power for the two SA levels. (*: $p \leq 0.001$).*

based on these sensitive features were applied to classify of SA levels, including two main steps: data preprocessing and data processing, as shown in Fig. 2.

*3.2.1 Data preprocessing*

As shown in Fig. 2, Data preprocessing was first conducted in the following steps: (1) min-max normalisation, i.e., $x' = (x - \min(x))/(\max(x) - \min(x))$; (2) Z-score standardisation, i.e., $Z = (x' - \mu)/\sigma$;

https://doi.org/10.1017/aer.2024.36 Published online by Cambridge University Press

***Table 2.*** *ANOVA results of relative power for the two SA levels*

| Waves (%) | Region × laterality × SA level | | | Region × SA level | | | Laterality × SA level | | | SA level | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | ***F*** | ***p*** | $\eta^2$ | ***F*** | ***p*** | $\eta^2$ | ***F*** | ***p*** | $\eta^2$ | ***F*** | ***p*** | $\eta^2$ |
| δ | 1.978 | 0.125 | 0.082 | 0.244 | 0.656 | 0.011 | 2.880 | 0.067 | 0.116 | 2.744 | 0.112 | 0.111 |
| $\theta$ | 0.402 | 0.701 | 0.018 | 0.213 | 0.667 | 0.010 | 0.756 | 0.474 | 0.033 | 5.628 | **0.027*** | 0.204 |
| α | 1.252 | 0.294 | 0.054 | 1.725 | 0.202 | 0.073 | 0.319 | 0.719 | 0.014 | 0.034 | 0.856 | 0.002 |
| $\sigma_1$ | 0.827 | 0.438 | 0.036 | 0.132 | 0.759 | 0.006 | 0.551 | 0.544 | 0.024 | 0.286 | 0.598 | 0.013 |
| $\sigma_2$ | 0.803 | 0.472 | 0.035 | 0.209 | 0.671 | 0.009 | 2.053 | 0.146 | 0.085 | 2.484 | 0.129 | 0.101 |
| $\beta_1$ | 0.987 | 0.386 | 0.043 | 0.976 | 0.352 | 0.042 | 1.431 | 0.250 | 0.061 | 5.655 | **0.027*** | 0.204 |
| $\beta_2$ | 0.530 | 0.635 | 0.024 | 1.256 | 0.281 | 0.054 | 4.313 | **0.020*** | 0.164 | 8.803 | **0.007*** | 0.286 |
| $\beta_3$ | 0.289 | 0.765 | 0.013 | 1.765 | 0.197 | 0.074 | 1.929 | 0.161 | 0.081 | 11.079 | **0.003*** | 0.335 |

***Table 3.*** *Selection of model inputs*

| EEG features | *Electrode sites* |
|---|---|
| $\beta_1/\beta_2$ | CP4 |
| $\beta_1/\beta_3$ | P3 |
| $\theta/\beta_1, \theta/\beta_2, (\theta+\alpha)/\beta_1, (\theta+\alpha+\sigma_1)/\beta_2, (\theta+\alpha+\sigma_1)/(\beta_1+\beta_2)$ | P4 |
| $\beta_1/(\beta_2+\beta_3)$ | CP4, C4 |
| $(\theta+\alpha)/\beta_2$ | CP4, P4 |
| $\beta_3$ relative power | P3, P4 |
| $\beta_2/\beta_3, (\theta+\alpha+\sigma_1+\sigma_2)/\beta_3, (\theta+\alpha+\sigma_1+\sigma_2+\beta_1)/\beta_3,$ | |
| $(\theta+\alpha+\sigma_1+\sigma_2+\beta_1+\beta_2)/\beta_3$ | P3, Pz |
| $(\theta+\alpha)/(\beta_1+\beta_2)$ | P4, Pz |
| $(\sigma_2+\beta_1+\beta_2)/\beta_3$ | CP3, P3, Pz |
| $\theta/(\beta_1+\beta_2)$ | CP4, Pz, P4 |
| $(\theta+\alpha)/\beta_3, (\theta+\alpha+\sigma_1)/\beta_3, (\theta+\alpha)/(\beta_2+\beta_3), (\theta+\alpha)/(\beta_1+\beta_2+\beta_3)$ | P3, P4, Pz |
| $(\beta_1+\beta_2)/\beta_3$ | CP3, CPz, P3, Pz |
| $\theta/\beta_3$ | CP4, CPz, P3, Pz |
| $\theta/(\beta_2+\beta_3), \theta/(\beta_1+\beta_2+\beta_3)$ | CP4, P3, Pz, P4 |

(3) principal component analysis (PCA) is a method employed for reducing data dimensionality by representing correlated datasets as uncorrelated principal components (PCs) (Ayesha et al., 2020), only PCs with eigenvalues exceeding 1 were selected here.; (4) Z-score standardisation for the PCs.

*3.2.2 Data processing*

As shown in Fig. 2, classifiers, including LDA, LR, GNB, DT, KNN, RF and ANN, were adopted to classify SA levels through the sklearn package in Python. The model inputs were the PCs of the preprocessed SA-sensitive EEG features. As a comparison, the PCs for the other seven categories of EEG features were also computed, including (1) absolute power; (2) relative power; (3) ratios; (4) absolute power versus relative power; (5) relative power versus ratios; (6) absolute power versus ratios; and (7) absolute power, relative power and ratios. The main steps were as follows: (1) found the optimal parameter settings for different classification models based on the results of five-fold cross-validation; (2) adopted the optimal parameter settings to calculate the classification results for each fold of cross-validation, including accuracy, precision, recall, F1 score, etc; (3) averaged the classification results of five-fold cross-validation to obtain the mean values; and (4) repeated the process 10 times to get the final classification accuracies of the above classifiers, as shown in

https://doi.org/10.1017/aer.2024.36 Published online by Cambridge University Press

Table 4. The interpretability of a classification model is contingent upon its underlying algorithms and discriminatory methods. Comprehensive information on this aspect can be found in relevant literature [24, 37] as well as in the MATLAB documentation on machine learning, accessible at https://ww2.mathworks.cn/help/releases/R2020a/stats/choose-a-classifier.html#bu62uj8.

## 4.0 Discussion

This study explored EEG features sensitive to SA and further classified different SA levels. The results found that the difference between LSL and HSL was insignificant for the absolute power of $\beta_1$, $\beta_2$ and $\beta_3$. The relative power of $\beta_1$, $\beta_2$ and $\beta_3$, as well as most ratios associated with $\beta_1$, $\beta_2$ and $\beta_3$, was sensitive to changes in SA levels, which is evidenced by the fact that these metrics were higher in HSL than in LSL. Furthermore, SA-sensitive EEG features were screened according to the criterion of $ps \leq 0.001$ in the independent samples T-test, and the screened EEG features were sent into the classifiers for small-sample classification of SA levels. The results showed that LR, DT, LDA, SVM, GNB, RF, ANN and KNN models achieve good accuracy (greater than 75%). Last but not least, for the classification of SA levels in this study, LR and DT are recommended by combining the interpretability and accuracy of the models.

In contrast to previous studies that indicated that $\beta$ wave activation is associated with changes in SA levels, this study further explored refined $\beta_1$, $\beta_2$ and $\beta_3$ wave activities. In the present study, the difference between SA levels for the absolute power features of $\beta_1$, $\beta_2$ and $\beta_3$ was insignificant. Therefore, H1 was rejected. That is, the absolute power of $\beta_1$, $\beta_2$ and $\beta_3$ were insensitive to changes in SA levels, which suggests that these EEG features are limited in revealing SA levels. This result enriches the findings of previous studies in $\beta$ absolute power [7]. In addition, a significant difference was found in the absolute power of the $\theta$ wave, which was manifested by the fact that the absolute power of the $\theta$ wave of LSL was higher than that of HSL. $\theta$ oscillations are strongly associated with increased working-memory load or attention demands [22, 38]. This indicates that LSL requires more memory or attention resources when facing the same operational task.

For EEG feature extraction, it has been ubiquitous to carry out relative power and ratio based on absolute power [39, 40]. The present study found that the relative power of $\beta_1$, $\beta_2$, and $\beta_3$ were higher in HSL than in LSL. Also, it found that most of $\beta_1$, $\beta_2$ and $\beta_3$ related ratio features were sensitive to SA levels. Therefore, H2 was accepted. $\beta$ wave oscillations are generally associated with higher levels of arousal or vigilance [40, 41], which suggests that LSL has lower levels of arousal or vigilance. This result can be considered as an extension of previous studies [7, 8], suggesting that the relative power of $\beta_1$, $\beta_2$ and $\beta_3$ and their associated ratio features are sensitive in revealing changes in SA levels. This finding provides an ideal theoretical basis for subsequent classification of SA levels using the aforementioned SA-sensitive EEG features, which is conducive to improving the interpretability of model inputs. Finally, a significant difference between different SA levels was also found in the $\theta$ wave relative power, which manifested as significantly higher $\theta$ relative power in LSL than in HSL, once more suggesting the lack of resources exhibited by LSL when oriented to the same task demands.

In addition, this study constructed a classification model for SA. As demonstrated in the above results, this study used SA-sensitive EEG features as inputs and selected SA levels classified by SAGAT scores as outputs for small-sample classification of SA levels in an aviation task scenario. The results of the five-fold cross-validation showed that the classification accuracies of LR, DT, LDA, SVM, GNB, RF, ANN and KNN models were greater than 75%. Therefore, H3 was accepted. From the results in Table 4, on one hand, compared with the accuracy of the previous work in selecting the sensitive EEG features related to $\beta$-wave, the above classifiers obtained higher accuracies after selecting the refined sensitive EEG features related to $\beta_1$, $\beta_2$ and $\beta_3$; On the other hand, the optimal classification accuracy using SA-sensitive EEG features as inputs was found to be higher compared to using the following data as input: (a) absolute power; (b) relative power; (c) ratios; (d) absolute power versus relative power; (e) relative power versus ratios; (f) absolute power versus ratios; and (g) absolute power, relative power and ratios. The results can provide some ideas for subsequent SA classification studies.

https://doi.org/10.1017/aer.2024.36 Published online by Cambridge University Press

https://doi.org/10.1017/aer.2024.36 Published online by Cambridge University Press

***Table 4.*** *Classification results of typical classifiers*

| Model category | Interpretability | Mean accuracy (standard deviation) (%) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Absolute power | Relative power | Ratios | Absolute power relative power | Relative power ratios | Absolute power ratios | Absolute power, relative power, and ratios | SA-sensitive EEG feature |
| LDA | Easy | 49.0 (0.0) | 51.0 (0.0) | 83.0 (0.0) | 38.0 (0.0) | **79.0** (0.0) | 65.0 (0.0) | 54.0 (0.0) | 88.0 (0.0) |
| LR | Easy | **55.0** (0.0) | 54.1 (2.7) | 79.0 (0.0) | **47.2** (4.0) | **79.0** (0.0) | 65.0 (0.0) | 52.8 (2.6) | **92.0** (0.0) |
| GNB | Easy | 28.0 (0.0) | **59.0** (0.0) | 62.0 (0.0) | 37.0 (0.0) | 54.0 (0.0) | 49.0 (0.0) | 45.0 (0.0) | 76.0 (0.0) |
| SVC-linear | Easy | 46.0 (0.0) | 51.0 (0.0) | **84.0** (0.0) | 33.0 (0.0) | **79.0** (0.0) | **75.0** (0.0) | **57.0** (0.0) | 83.0 (0.0) |
| DT | Easy | 44.9 (8.0) | 57.8 (9.6) | 59.5 (8.8) | **49.3** (9.4) | 52.8 (11.2) | 74.7 (11.6) | 48.7 (12.4) | **92.0** (0.0) |
| SVC-nonlinear | Hard | 52.0 (0.0) | 59.0 (0.0) | **76.0** (0.0) | 46.0 (0.0) | 75.0 (0.0) | 70.0 (0.0) | **79.0** (0.0) | 80.0 (0.0) |
| KNN | Hard | **46.0** (0.0) | **67.0** (0.0) | 66.0 (0.0) | 49.0 (0.0) | 66.0 (0.0) | 70.0 (0.0) | 63.0 (0.0) | 79.0 (0.0) |
| RF | Hard | 42.6 (6.8) | 51.7 (5.2) | 59.7 (4.9) | 38.7 (5.5) | 50.7 (7.4) | 68.2 (5.6) | 56.3 (6.4) | **88.8** (2.4) |
| ANN | Hard | 43.2 (5.0) | 51.2 (7.4) | 77.3 (3.8) | 36.1 (4.8) | **77.0** (8.8) | **70.0** (6.8) | 46.2 (5.1) | 84.0 (4.0) |

Moreover, considering the urgent need and application potential of cockpit transparency and human-machine trust in aviation, interpretability becomes one of the crucial references for model selection. Therefore, combining the characteristics of small-sample data makes it practical to compare the modeling effects using generalised supervised machine learning classifiers with different interpretability. The studies above show that the best classification accuracies obtained by LR and DT with easy interpretability (92%) are superior to those obtained by RF and ANN with hard interpretability (88.8% and 84%, respectively). The best classification accuracies in both types of models are better than 67% [8], 70.6%–74.32% [19, 22], 76.1%–82.7% [23] and 70.8% [7]. Analysing the reasons, on the one hand, may lie in the selection of refined SA-sensitive EEG features and the attempt at parameter tuning of multiple supervised machine learning classifiers. On the other hand, small sample data may be more suitable for using easily interpretable machine learning classifiers, while more complex network structures and hierarchical parameter settings may not be suitable for this context. For the classification of SA levels in this study, LR and DT are recommended by combining the interpretability and accuracy of the models.

Compared to prior studies on SA discrimination models, the present study focused on analysing SA levels using SA-sensitive EEG features in small-sample dataset. This included performing model construction and comparison with accuracy and interpretability. The study addresses specific considerations in the following aspects. The first is the input and output of the classifier. Differences in inputs may lead to differences in discriminatory results and trigger considerations in the interpretability of model inputs. Unlike prior studies using PSD features [8, 19, 22], SA-sensitive $\beta$-features [7] or a combination of EEG and eye movement features [23], this study selected SA-sensitive, refined $\beta_1$, $\beta_2$ and $\beta_3$ wave features as inputs. Utilising these EEG features resulted in higher classification accuracy compared to alternative feature combinations, as shown in Table 4. For output labels, the SAGAT method is employed due to its proven sensitivity and reliability in assessing SA across various experimental settings and domains [42]. In line with prior studies [7, 8], SAGAT scores are used to categorise high and low SA levels, differing from studies that used performance metrics as labels for fatigue/load-related SA [19, 22, 23]. Sleep-related fatigue, often associated with reduced SA [43, 44], needs consideration in SA-related experiments. Based on the experiment recordings of participants' statuses in our study, we indeed observed phenomena that might be related to fatigue in some participants, such as three times reporting falling asleep after midnight the day before the experiment and three times showing occasional adjustments in sitting posture or yawning. Considering the relatively short experiment duration, and limited number of fatigue related participant reactions, SA is considered to remain unaffected by sleep-related fatigue in this study. However, exploring the impact of sleep-related fatigue on SA will be quite meaningful and should be investigated in long-duration experiment tasks carefully.

What's more, the accuracies of classifiers with different interpretability are compared. Machine learning models with good transparency or easy interpretability enhance human-machine mutual trust, thus improving situation awareness, task performance and human factors safety [45, 46]. Some studies chose only discriminative models with easy interpretability [7], others chose models with hard interpretability [8, 19] and still others used both types of models [22, 23]. In this study, classification algorithms such as LR and DT with easy interpretability were selected and compared with models such as RF and ANN with hard interpretability. In terms of accuracy, LR and DT obtained superior accuracy, outperforming 67%–82.7% of previous results. In conclusion, this study conducted a classification study of SA levels divided by SAGAT scores based on SA-sensitive EEG features with a small-sample dataset in an aviation task scenario. The results show that the selected generic supervised learning classifiers all obtained good classification accuracies, and the LR and DT are recommended by combining the interpretability and accuracy of the models.

Despite the previous findings, this study has some shortcomings. First, the platform is MATB II based on aviation task settings, which have specific differences from the actual flight scenario. It is still necessary to build task scenarios closer to actual flights to further verify the reliability of the conclusions. Secondly, considering the difference between students and real frontline operators, an improvement in subject selection is needed in the subsequent study. The results of this study can provide valuable references for the exploration of SA-sensitive features and the selection of SA classification models. The

https://doi.org/10.1017/aer.2024.36 Published online by Cambridge University Press

practical application in real-world settings still requires attention to the following aspects: (a) validating and optimising SA-sensitive EEG features in specific aviation domain; (b) considering factors like time frame and computational power for meaningful assessment of SA levels; (c) exploring the feasibility of wearable, real-time EEG measurements in flight, taking into account aviation regulations, previous practices and potential interferences in physiological measurement.

## 5.0 Conclusions

Based on the typical human-computer interaction task setting in this study, the following conclusions can be obtained:

(1) The refined relative power of $\beta_1$, $\beta_2$ and $\beta_3$ and their corresponding ratio features are sensitive to SA and can provide effective inputs for subsequent classification of SA level.

(2) Small-sample classification of SA levels was carried out by using SA-sensitive EEG features, and it was found that the general supervised learning classifier, including LR, DT, RF, ANN and other classification algorithms, can obtain good classification accuracies (greater than 75%).

(3) For the classification of SA levels in this study, LR and DT are recommended by combining the interpretability and accuracy of the models.

In this study, the exploration of SA-sensitive EEG features and the classification of SA levels were carried out under different SA levels, and the corresponding conclusions can provide some references for online monitoring and classification optimisation of SA.

**Supplementary Material.** To view supplementary material for this article, visit https://doi.org/10.1017/aer.2024.36

**Data availability statement.** The data presented in this study are available on request from the corresponding author. The data are not publicly available due to privacy.

**Acknowledgments.** The authors gratefully acknowledge the agencies NSFC, CCAC and ASFC for the financial support. In addition, the authors acknowledge the participants for their participation.

**Author contributions.** All authors have read and agreed to the published version of the manuscript. Conceptualisation, formal analysis, investigation, methodology, validation, visualisation and writing – original draft, C.F.; conceptualisation, data curation, investigation, methodology, resources and writing – review and editing, S.L.; conceptualisation, data curation, methodology, resources, supervision and writing – review and editing, project administration, funding acquisition, X.WY.; methodology, validation and writing – original draft, Y.D.; visualisation, investigation and writing – original draft Z.W.; data curation, validation and writing – original draft, C.Q.

**Funding.** This research was funded by the joint program of the National Natural Science Foundation of China and Civil Aviation Administration of China (No. U1733118), the National Natural Science Foundation of China (No. 71301005), and the Aeronautical Science Foundation of China (No. 201813300002).

**Competing interests.** The authors declare no competing interests.

## References

[1] Endsley, M.R. Situation awareness global assessment technique (SAGAT), Proceedings of the IEEE 1988 National Aerospace and Electronics Conference, 1988, pp 789–795. https://doi.org/10.1109/NAECON.1988.195097

[2] Sant'Anna, D.A.L.M.D. and Hilal, A.V.G.D. The impact of human factors on pilots' safety behavior in offshore aviation companies: A brazilian case, *Safety Sci.*, 2021, **140**, p 105272. https://doi.org/10.1016/j.ssci.2021.105272

[3] Wiegmann, D.A. and Shappell, S.A. *A Human Error Approach to Aviation Accident Analysis: The Human Factors Analysis and Classification System*, Ashgate, 2003.

[4] Endsley, M.R. A taxonomy of situation awareness errors, in R. Fuller, N. Johnston and N. McDonald (Eds), *Human Factors in Aviation Operations*, Avebury Aviation, Ashgate Publishing Ltd, 1995a, pp. 287–292.

[5] Schwarz, J. and Fuchs, S. Validating a "real-time assessment of multidimensional user state" (RASMUS) for adaptive human-computer interaction, 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 2018, pp 704–709. https://doi.org/10.1109/SMC.2018.00128

[6] Schwerd, S. and Schulte, A. Operator state estimation to enable adaptive assistance in manned-unmanned-teaming, *Cognit. Syst. Res.*, 2021, **67**, pp 73–83. https://doi.org/10.1016/j.cogsys.2021.01.002

[7] Feng, C., Liu, S., Wanyan, X., Chen, H., Min, Y. and Ma, Y. EEG feature analysis related to situation awareness assessment and discrimination, *Aerospace*, 2022, **9**, (10), p 546. https://doi.org/10.3390/aerospace9100546

[8] Kästle, J.L., Anvari, B., Krol, J. and Wurdemann, H.A. Correlation between situational awareness and EEG signals, *Neurocomputing*, 2021, **432**, pp 70–79. https://doi.org/10.1016/j.neucom.2020.12.026

[9] Gawron, V.J. *Human Performance, Workload, and Situational Awareness Measures Handbook*, 2nd ed, 2019.

[10] Smith, K., Endsley, T. and Clark, T. Physiological correlates of objective situation awareness measurements, 2023 IEEE Aerospace Conference, 2023, pp 1–6. https://doi.org/10.1109/AERO55745.2023.10115788

[11] Bracken, B., Tobyne, S., Winder, A., Shamsi, N. and Endsley, M.R. Can situation awareness be measured physiologically?, in H. Ayaz, U. Asgher and L. Paletta (Eds), *Advances in Neuroergonomics and Cognitive Engineering,* vol. **259**, Springer International Publishing, 2021, pp 31–38. https://doi.org/10.1007/978-3-030-80285-1_4

[12] Zhang, T., Yang, J., Liang, N., Pitts, B.J., Prakah-Asante, K.O., Curry, R., Duerstock, B.S., Wachs, J.P. and Yu, D. Physiological measurements of situation awareness: A systematic review, *Hum. Factors*, 2020, pp 1–22. https://doi.org/10.1177/0018720820969071

[13] Mehta, R.K., Peres, S.C., Shortz, A.E., Hoyle, W., Lee, M., Saini, G., Chan, H.-C. and Pryor, M.W. Operator situation awareness and physiological states during offshore well control scenarios, *J. Loss Prev. Process Ind.*, 2018, **55**, pp 332–337. https://doi.org/10.1016/j.jlp.2018.07.010

[14] Saus, E.-R., Johnsen, B.H., Eid, J., Riisem, P.K., Andersen, R. and Thayer, J.F. The effect of brief situational awareness training in a police shooting simulator: An experimental study, *Mil. Psychol.*, 2006, **18**, (sup1), pp S3–S21. https://doi.org/10.1207/s15327876mp1803s_2

[15] Sætrevik, B. A controlled field study of situation awareness measures and heart rate variability in emergency handling teams, *Proc. Hum. Factors Ergon. Soc. Ann. Meet.*, 2012, **56**, (1), pp 2006–2010. https://doi.org/10.1177/1071181312561419

[16] de Winter, J.C.F., Eisma, Y.B., Cabrall, C.D.D., Hancock, P.A. and Stanton, N.A. Situation awareness based on eye movements in relation to the task environment, *Cognit. Technol. Work*, 2019, **21**, (1), pp 99–111. https://doi.org/10.1007/s10111-018-0527-6

[17] van de Merwe, K., van Dijk, H. and Zon, R. Eye movements as an indicator of situation awareness in a flight simulator experiment, *Int. J. Aviat. Psychol.*, 2012, **22**, (1), pp 78–95. https://doi.org/10.1080/10508414.2012.635129

[18] Moore, K. and Gugerty, L. Development of a novel measure of situation awareness: The case for eye movement analysis, The Annual Meeting, 2010.

[19] Li, R., Wang, L. and Sourina, O. Subject matching for cross-subject EEG-based recognition of driver states related to situation awareness, *Methods*, 2021, S1046202321001018. https://doi.org/10.1016/j.ymeth.2021.04.009

[20] Corace, K., Baysarowich, R., Willows, M., Baddeley, A., Schubert, N. and Knott, V. Resting state EEG activity related to impulsivity in people with prescription opioid use disorder, *Psychiatry Res. Neuroimaging*, 2022, **321**, p 111447. https://doi.org/10.1016/j.pscychresns.2022.111447

[21] Shokouh Alaei, H., Ghoshuni, M. and Vosough, I. Directed brain network analysis in anxious and non-anxious depression based on EEG source reconstruction and graph theory, *Biomed. Signal Process. Control*, 2023, **83**, p 104666. https://doi.org/10.1016/j.bspc.2023.104666

[22] Li, R., Lan, Z., Cui, J., Sourina, O. and Wang, L. EEG-based recognition of driver state related to situation awareness using graph convolutional networks, Proceedings of the 2020 International Conference on Cyberworlds (CW). 2020 International Conference on Cyberworlds (CW), Caen, France, 2020. https://doi.org/10.1109/CW49994.2020.00037

[23] Li, Q., Ng, K.K.H., Yu, S.C.M., Yiu, C.Y. and Lyu, M. Recognising situation awareness associated with different workloads using EEG and eye-tracking features in air traffic control tasks, *Knowl.-Based Syst.*, 2023, **260**, p 110179. https://doi.org/10.1016/j.knosys.2022.110179

[24] Osisanwo, F., Akinsola, J., Awodele, O., Hinmikaiye, J., Olakanmi, O. and Akinjobi, J. Supervised machine learning algorithms: Classification and comparison, *Int. J. Comput. Trends Technol. (IJCTT)*, 2017, **48,** (3), pp 128–138.

[25] Santiago-Espada, Y., Myer, R.R., Latorella, K.A. and Comstock, J.R. The Multi-Attribute Task Battery II (MATB-II) software for human performance and workload research: A user's guide (NASA/TM–2011-217164), National Aeronautics and Space Administration,Langley Research Center, 2011.

[26] Endsley, M.R. Measurement of situation awareness in dynamic systems, *Hum. Factors*, 1995b, **37**, (1), pp 65–84. https://doi.org/10.1518/001872095779049499

[27] Billeci, L., Callara, A.L., Guiducci, L., Prosperi, M., Morales, M.A., Calderoni, S., Muratori, F. and Santocchi, E. A randomized controlled trial into the effects of probiotics on electroencephalography in preschoolers with autism, *Autism*, 2023, **27**, (1), pp 117–132. https://doi.org/10.1177/13623613221082710

[28] Nakuci, J., Covey, T.J., Shucard, J.L., Shucard, D.W. and Muldoon, S.F. Single trial variability in neural activity during a working memory task reveals multiple distinct information processing sequences, *NeuroImage*, 2023, **269**, p 119895. https://doi.org/10.1016/j.neuroimage.2023.119895

[29] Mognon, A., Jovicich, J., Bruzzone, L. and Buiatti, M. ADJUST: An automatic EEG artifact detector based on the joint use of spatial and temporal features, *Psychophysiology*, 2011, **48**, (2), pp 229–240.

[30] Nayak, C.S. and Anilkumar, A.C. *EEG Normal Waveforms,* 2023.

[31] Park, Y.-M., Che, H.-J., Im, C.-H., Jung, H.-T., Bae, S.-M. and Lee, S.-H. Decreased EEG synchronization and its correlation with symptom severity in Alzheimer's disease, *Neurosci. Res.*, 2008, **62,** (2), pp 112–117. https://doi.org/10.1016/j.neures.2008.06.009

[32] Frohlich, J., Toker, D. and Monti, M.M. Consciousness among delta waves: A paradox?, *Brain*, 2021, **144**, (8), pp 2257–2277. https://doi.org/10.1093/brain/awab095

[33] Trapsilawati, F., Herliansyah, M.K., Nugraheni, A.S.A.N.S., Fatikasari, M.P. and Tissamodie, G. EEG-based analysis of air traffic conflict: Investigating controllers' situation awareness, stress level and brain activity during conflict resolution, *J. Navig.*, 2020, **73**, (3), pp 678–696. https://doi.org/10.1017/S0373463319000882


https://doi.org/10.1017/aer.2024.36 Published online by Cambridge University Press

[34] Wanyan, X., Zhuang, D., Lin, Y., Xiao, X. and Song, J.-W. Influence of mental workload on detecting information varieties revealed by mismatch negativity during flight simulation, *Int. J. Ind. Ergon.*, 2018, **64**, pp 1–7. https://doi.org/10.1016/j.ergon.2017.08.004
[35] Armstrong, R.A. Recommendations for analysis of repeated-measures designs: Testing and correcting for sphericity and use of MANOVA and mixed model analysis, *Ophthalmic Physiol. Opt.*, 2017, **37**, (5), pp 585–593. https://doi.org/10.1111/opo.12399
[36] Mishra, P., Singh, U., Pandey, C., Mishra, P. and Pandey, G. Application of student's t-test, analysis of variance, and covariance, *Ann. Card. Anaesth.*, 2019, **22**, (4), p 407. https://doi.org/10.4103/aca.ACA_94_19
[37] Uddin, S., Khan, A., Hossain, M.E. and Moni, M.A. Comparing different supervised machine learning algorithms for disease prediction, *BMC Med. Inf. Decis. Making*, 2019, **19**, (1), p 281. https://doi.org/10.1186/s12911-019-1004-8
[38] Ren, M., Xu, J., Zhao, J., Zhang, S., Wang, W., Xu, S., Zhou, Z., Chen, X., Chen, S., Li, Y. and Shan, C. The modulation of working-memory performance using gamma-electroacupuncture and Theta-electroacupuncture in healthy adults, *Evidence-Based Complementary Altern. Med.*, 2021, **2021**, pp 1–11. https://doi.org/10.1155/2021/2062718
[39] Hou, L., Chen, L. and Zhou, R. Premenstrual syndrome is associated with an altered spontaneous electroencephalographic delta/beta power ratio across the menstrual cycle, *Int. J. Psychophysiol.*, 2022, **181**, pp 64–72. https://doi.org/10.1016/j.ijpsycho.2022.08.007
[40] Yi Wen, T. and Mohd Aris, S.A. Electroencephalogram (EEG) stress analysis on alpha/beta ratio and theta/beta ratio, *Indones. J. Electr. Eng. Comput. Sci.*, 2020, **17**, (1), p 175. https://doi.org/10.11591/ijeecs.v17.i1.pp175-182
[41] Zhao, C., Zhao, M., Liu, J. and Zheng, C. Electroencephalogram and electrocardiograph assessment of mental fatigue in a driving simulator, *Accid. Anal. Prev.*, 2012, **45**, pp 83–90. https://doi.org/10.1016/j.aap.2011.11.019
[42] Endsley, M.R. A systematic review and meta-analysis of direct objective measures of situation awareness: A comparison of SAGAT and SPAM, *Hum. Factors*, 2021, **63**, (1), pp 124–150. https://doi.org/10.1177/0018720819875376
[43] Hopko, S.K., Khurana, R., Mehta, R.K. and Pagilla, P.R. Effect of cognitive fatigue, operator sex, and robot assistance on task performance metrics, workload, and situation awareness in human-robot collaboration, *IEEE Rob. Autom. Lett.*, 2021, **6**, (2), pp 3049–3056. https://doi.org/10.1109/LRA.2021.3062787
[44] Sneddon, A., Mearns, K. and Flin, R. Stress, fatigue, situation awareness and safety in offshore drilling crews, *Safety Sci.*, 2013, **56**, pp 80–88. https://doi.org/10.1016/j.ssci.2012.05.027
[45] Schmidt, P., Biessmann, F. and Teubner, T. Transparency and trust in artificial intelligence systems, *J. Decis. Syst.*, 2020, **29**, (4), pp 260–278. https://doi.org/10.1080/12460125.2020.1819094
[46] Win, A.F.T., Jacobs, N., Muttram, R.I., Olszewska, J.I., Rajabiyazdi, F., Theodorou, A., Underwood, M.A., Wortham, R.H. and Watson, E. IEEE P7001: A proposed standard on transparency, *Front. Rob. AI*, 2021, **8**, p 11.

**Cite this article:** Feng C., Liu S., Wanyan X., Dang Y., Wang Z. and Qian C. (2024). $\beta$-wave-based exploration of sensitive EEG features and classification of situation awareness. *The Aeronautical Journal*, **128**, 2561–2576. https://doi.org/10.1017/aer.2024.36