



## Research article

# Physiological records-based situation awareness evaluation under aviation context: A comparative analysis

Jun Chen<sup>a,b</sup>, Anqi Chen<sup>a</sup>, Bingkun Jiang<sup>a</sup>, Xinyu Zhang<sup>a,\*</sup><sup>a</sup> School of Electronics and Information, Northwestern Polytechnical University, Xi'an, 710072, China<sup>b</sup> Chongqing Institute for Brain and Intelligence, Guangyang Bay Laboratory, Chongqing, 400064, China

## ARTICLE INFO

## Keywords:

Situation awareness  
Electroencephalogram  
Brain electrical activity mapping  
Convolutional neural network  
Multi-class classification  
Aviation decision-making

## ABSTRACT

Situational Awareness (SA) assessment is of paramount importance in various domains, with particular significance in the military for safe aviation decision-making. It involves encompassing perception, comprehension, and projection levels in human beings. Accurate evaluation of SA statuses across these three levels is crucial for mitigating human false-positive and false-negative rates in monitoring complex scenarios in the aviation context. This study proposes a comprehensive comparative analysis by involving two types of physiological records: electroencephalogram (EEG) signals and brain electrical activity mapping (BEAM) images. These two modalities are leveraged to automate precise SA evaluation using both conventional machine learning and advanced deep learning techniques. Benchmarking experiments reveal that the BEAM-based deep learning models attain state-of-the-art performance scores of 0.955 for both SA perception and comprehension levels, respectively. Conversely, the EEG signals-based manual feature extraction, selection, and classification approach achieved a superior accuracy of 0.929 for the projection level of SA. These findings collectively highlight the potential of deploying diverse physiological records as valuable computational tools for enhancing SA evaluation throughout aviation decision-making safety.

## 1. Introduction

Situation awareness (SA) is the understanding of the current situation given a complex environmental setting, which is also a critical component within the information processing perceptions in human beings [1]. The SA assessment has been considered indispensable in a variety of domains [2,3], particularly vital in military operations, for safe aviation decision-making [4,5].

The widely acknowledged concept of SA assessment was formulated by Mica Endsley, the former Chief Scientist of the United States Air Force, in the late 80s [6]. SA was characterized by three distinctive yet intricate levels: *Level 1* involves the perception of environmental factors, *Level 2* indicates the comprehension of those factors with respect to the aviator's objectives, and *Level 3* interprets the projection of forecasting the operating system for timely decision-making. This definition elucidates the overall status of an aviator's comprehension of critical elements given a complex environment within a volume of time and space [6].

The three-level SA assessment plays a pivotal role in aviation as it is highly correlated with fatigue and cognition, ensuring the seamless and smooth execution of missions for aviators, air traffic controllers, and drone operators. However, attaining accurate SA is a formidable challenge, particularly in the context of the cooperative mission mode of crewed aircraft and crewless aerial vehicles. The

\* Corresponding author.

E-mail address: [xinyu.zhang@nwpu.edu.cn](mailto:xinyu.zhang@nwpu.edu.cn) (X. Zhang).

workload of aviators increases sharply, which is bound to affect their SA status assessment significantly [7]. Moreover, human errors were crucially attached to SA status, which was identified in 88% of accidents by commercial airlines [7]. Therefore, precise SA assessment is imperative for reducing human errors within the intricate realm of complex aviation systems.

In recent years, the implementation of machine learning-driven techniques in SA assessment was extensively analysed in the existing literature body [6,7]. However, deploying deep learning-driven biomedical image-based approaches has yielded unprecedented detection rates across varied domains [8–10], yet it has not been adequately applied for SA evaluation within the aviation decision-making domain. To address such an existing limitation, this paper sought to evaluate the SA status confronting pilots and unmanned system operators. More specifically, real-time electroencephalogram (EEG) signals are leveraged to evaluate the three-level SA status through conventional machine learning techniques. In the meantime, deep convolutional neural networks are employed to process brain electrical activity mapping (BEAM) images for SA evaluation. EEG signals are obtained by amplifying and recording the brain's spontaneous biopotentials from the scalp through sophisticated electronic instruments. EEG signals are decomposed by the Independent Component Analysis (ICA) method to obtain subcomponents, each with a specific IC activation and terrain. BEAM is the abbreviation for brain electrical activity mapping. It is a two-dimensional flat figure formed by a spherical scalp showing the power values in each frequency band of the brain wave in different colours. It objectively reflects the spatial distribution of potential changes in various brain regions. BEAM images are the topological graphs of independent components obtained by ICA processing of EEG signals in different frequency bands. Each row of the EEG signal matrix represents the change in the potential difference between each channel and the reference channel with time. After ICA decomposition, each row of data represents the change of independent components over time after spatial filtering from channel data. Therefore, the results of ICA decomposition provide the temporal and spatial attributes of independent components, while BEAM shows the distribution of different components in the scalp. Therefore, it is worth digging into the performance of the two modalities on the SA evaluation task. Thus, a comparative analysis is conducted with the two modalities through the two distinctive approaches.

### 1.1. Related works

Extensive studies adopted various techniques to detect SA status, aiming to elevate personnel performance and prevent potential human errors. For instance, the SA global assessment technique (SAGAT) is the most well-known approach proposed by Endsley [11], and it has been widely applied in varied situations [12–16]. There are other types of SA assessment criteria, such as the SA rating technique (SART) [17–20], the situation-present assessment method (SPAM) [21,22], and the SA behavioural rating scale (SABRS) [23], to name a few. However, those traditional approaches are majorly built upon subjectivity, as they require the operators or observers to fill in questionnaires during or after tasks, sometimes even demanding interruption of missions. Moreover, many of those questionnaires were answered after finishing the required tasks instead of real-time measurement, leading to increased error rates.

Due to those limitations, more advanced techniques were employed for precise SA evaluation. For instance, Kwok and Viridi [24] proposed an artificial intelligence (AI) technique which adopted computer vision and behavioural learning modules for SA assessment. Other scholars tend to apply more objective tools in assisting evaluate SA automatically, such as eye-trackers, physiological collection equipment, and many more [25,26]. For example, Zhou et al. [27] combined eye-tracking and self-reported data to predict SA through an ensemble model, and they have reached a mean absolute error (MAE) of 0.096. Kastle et al. [28] utilised EEG signals to investigate their hidden correlations with SA status, attaining an accuracy rate of 67% on test data. Yiu et al. [29] utilised Bayesian neural networks to process EEG signals and have obtained an accuracy rate of 66.5% for SA assessment. Chen et al. [30] adopted EEG signals for fatigue identification, reaching an accuracy rate of 75.26% for the three-class classification task. Similar studies were conducted by de Winter et al. [31], Li et al. [32], Li et al. [33], Feng et al. [34], Zhang et al. [35], Ye et al. [36], and [37,38] all demonstrating superior performance on SA assessment with sensor data than classic inference questionnaires.

Nevertheless, those techniques mainly require manual processing of feature extraction, selection, and classification approaches. With the selected “hand-crafted” features, the SA evaluation was implemented on the basis of machine learning classifiers. Such a procedure is time-consuming, and each of the stages confronts with varying degrees of uncertainty, aggravating the final error rates. Deep learning-based techniques, on the other hand, in particular, deep convolutional neural networks (CNN) have outstanding

**Table 1**  
Literature review summary on intelligent SA evaluation.

	Type	Feature Selection	Method	Data	No. Subject	Performance
[25]	Binary class (Low/High)	–	Brain network topology	EEG	109	EER = 0.04
[26]	Binary class	–	D-UDA	EEG	13	ACC = 0.98
[27]	Binary class	SHAP	LightGBM	Eye-track data	32	RMSE = 0.12
[28]	Binary class	ICA	Random forest, Boosted trees	EEG	32	ACC = 0.67
[29]	Binary class	ANOVA, Mutual Information	Bayesian neural network	EEG	30	ACC = 0.67
[30]	Multi-class	–	Graph neural network	EEG	10	ACC = 0.75
[32]	Binary class	–	Graph neural network	EEG	27	ACC = 0.71
[34]	Binary class	ANOVA, PCA	Bayesian Model	EEG	48	ACC = 0.71
[35]	Binary class	DNN	Autoencoder, LSTM	EEG, EOG	–	RMSE = 0.08
[36]	Binary class	Logistic-ARFE	Logistic regression	EEG	–	ACC = 0.90
[37]	Binary class	–	MATCN-GT	EEG	23	ACC = 0.94
[38]	Binary class	EEMD + PSD	PSO-H-ELM	EEG	6	ACC = 0.97

performance in interpreting images by automatically selecting significant features [39]. For instance, Liu et al. [8] proposed a CNN-based multi-modal image fusion model to enhance SA for target detection. Wang et al. [9] generated a CNN-LSTM model to predict post fault actions (PFA) in electric power systems. Table 1 summarises some related works on SA evaluation through the utilisation of computerised methods using physiological data. Nevertheless, based on the literature review analysis, the automatic SA evaluation for aviation decision-making is still absent in existing studies. Accordingly, this paper aims to have a comparative analysis of the conventional machine learning process and advanced deep CNN implementation on SA evaluation through two types of physiological records, more specifically, EEG signals and BEAM images.

1.2. Contributions

This work facilitates the implementation of deep learning techniques in the aviation domain for precise and automatic SA evaluation. Overall, this paper makes the following four contributions:

- This work is the first of its kind, which proposes a comparative analysis between classic machine learning approaches and deep learning techniques for SA assessment. The two procedures were evaluated and compared to identify the better approach for precise and automatic SA detection.
- Two distinctive types of physiological records were employed in this study. Biomedical EEG signals were incorporated with classic machine learning approaches for manual feature extraction and selection, and BEAM images were deployed for deep CNN implementation. Their results were analysed and compared to ameliorate future research recommendations.
- Existing studies majorly apply binary classification tasks to evaluate SA status (i.e., normal or abnormal), whereas the precise three-level SA assessment was neglected. Hence, this study implements a multi-class classification task for a more precise three-level SA evaluation.

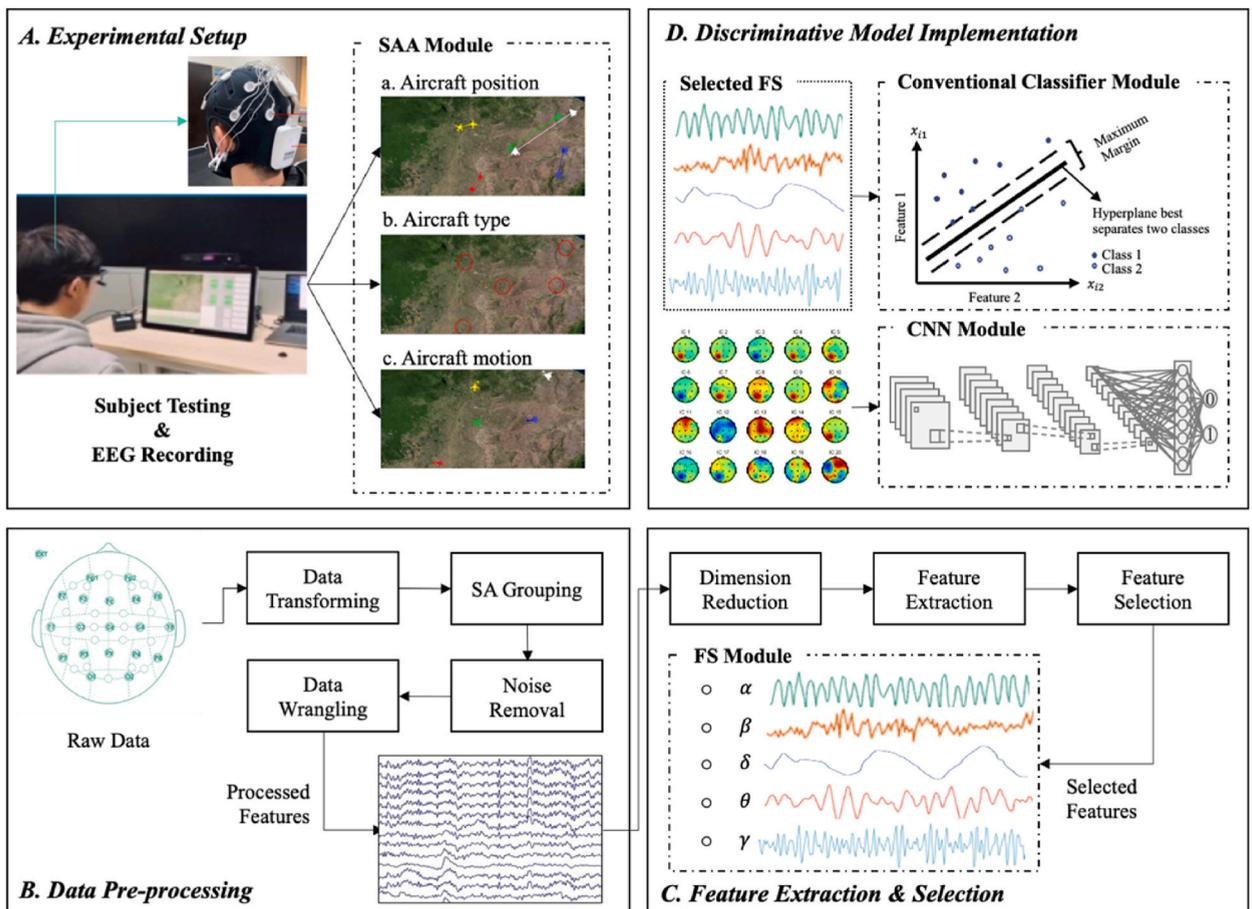


Fig. 1. SA assessment framework.

## 2. Methodology

With the objective of conducting comparative analysis on different machine learning approaches, this section presents the proposed SA assessment framework in Fig. 1, which consists of four main components: A) the experimental setup process for SA data collection; B) the indispensable pre-processing steps for EEG and BEAM data preparation; C) the feature extraction and selection process for choosing appropriate “hand-crafted” features; and D) the discriminative model implementation including conventional classifiers and CNN models for accurate and automatic SA evaluation.

### 2.1. Experimental setup

Fig. 1 Component A demonstrates the proposed experimental setup module. In order to collect data and assign labels accordingly for the three levels of SA, the overall experimental design adhered to the guidelines outlined in the Psychology Experiment Building Language (PEBL) [40], the SAGAT designing principles [6], the work conducted by Kästle [28], and considering the key personnel (i.e., pilots and operators) working environment. More specifically, the simulation system is designed with reference to the Vigilant Spirit Control Station system developed by the U.S. Air Force Research Laboratory. The experiments involved simulating three tasks to collect the three levels of SA status, including identifying positions for each aircraft, identifying types for each aircraft, and predicting motions for each aircraft. A touchable 27-inch display forms a human-computer interaction system utilised to collect all task performance records. Throughout the data collection phase, the participant would be wearing an EEG cap, and each of the pre-established experiments would be interrupted at a random time interval, then, he/she would be queried a series of questions regarding the tasks they were actively engaged in just now.

Five aircrafts move continuously on the designed application platform, with diversified shapes and colours. Following the SAGAT guidelines, those aircraft would disappear after a random time interval. At which, participants were supposed to answer a set of related questions to assess their SA. Then, their performance would be analysed for the labelling purpose. Details can be summarized as follows:

- To evaluate the perception of participants (Level 1 of SA), they were expected to target and define the last positions of the five aircraft before disappearing.
- To assess the comprehension of participants (Level 2 of SA), they were required to illustrate the type of each aircraft given a particular position.
- To assess the projection of participants (Level 3 of SA), they were queried to predict the motion directions of each aircraft.

With those three proposed tasks (exhibited in Fig. 2), the participants’ biomedical EEG signals were recorded throughout the simulation procedure for further processing and labelling purposes.

### 2.2. Data pre-processing

Fig. 1 Components A and B showcase the adopted data collection equipment used in this research, which is the NE wireless EEG system. In accordance with the 10–20 international system guidelines, the conventional EEG collection typically involves electrode counts of 16, 32, 62, 64, or 128 leads. However, insights gleaned from the experiments conducted by Montoya-Martínez et al. [41] revealed that higher correlations could be obtained when reducing the number of electrodes for EEG collection. Their findings indicated that a 20-channel configuration delivered the most optimal performance. Drawing inspiration from their research, this study adopted the 20-lead system selected based on the standard 64-lead EEG electrode distribution for data collection (refer to Fig. 3 and Table 2 for the chosen 20-lead EEG electrode system).

The raw signals and event information were obtained from the NE EEG equipment in the European Data Format (EDF). With the collected raw EEG signals, pre-processing techniques were applied to normalize the data and prepare them for further analysis. Detailed procedures are explained in Fig. 4 and Algorithm 1.

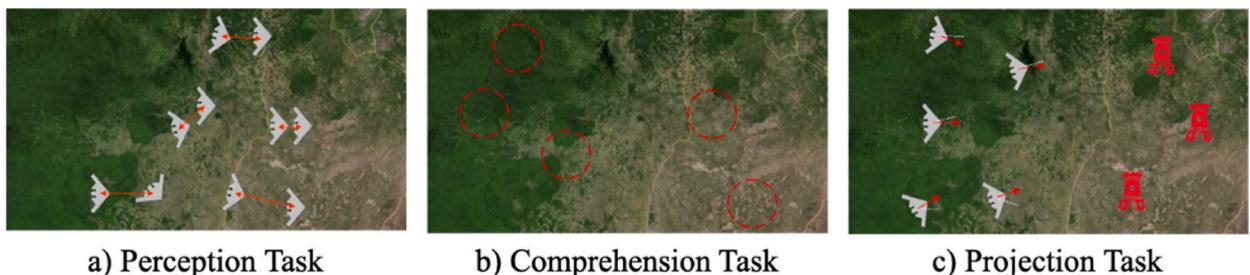


Fig. 2. Three-levels of SA collection tasks.

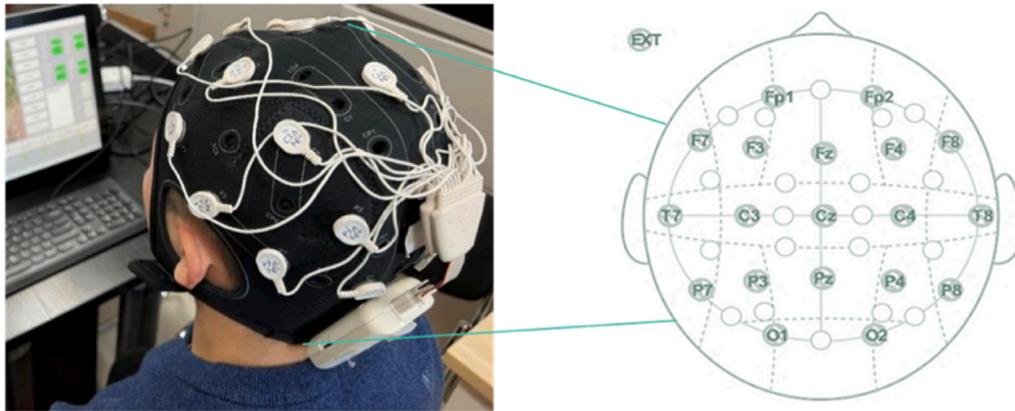


Fig. 3. 20-Lead EEG electrode system utilised.

**Table 2**  
Electrodes and corresponding brain regions and functions.

Index	Electrode	Brain region	Functions		
1	P7	Parietal Lobe	Process sensations such as touching, pressuring, temperature, and pain, in the meantime, regulate or allocate spatial attention		
2	P4				
4	PZ				
5	P3				
6	P8				
3	CZ			Paracentral Lobe	Responsible for thinking, calculating, and anything related to individual needs and emotions
11	C4				
15	C3				
7	O1	Occipital Lobe	Involved in the integration of visual information		
8	O2	Temporal Lobe	Process auditory stimulation, (hearing, language, memory)		
9	T8				
18	T7				
10	F8	Frontal Lobe	Related to body movement, pronunciation, language, and advanced thinking activities		
12	F4				
13	FP2				
14	FZ				
16	F3				
17	FP1				
19	F7				
20	FPZ			-	Ground, reference signal

- **Step 1:** Data transforming. Filtering data to obtain signals in different frequency bands using a 50 Hz high-pass filter and 0.1 Hz low-pass filter.
- **Step 2:** Independent component analysis (ICA). Separate the multi-channel multivariate EEG signals into subcomponents from the filtered signals.
- **Step 3:** Fourier transform (FFT) and Final power spectrum (PSD). EEG records of each component were calculated.
- **Step 4:** Band-pass filtering. The corresponding EEG data of the selected channel was filtered through band-pass filtering.
- **Step 5:** Remove damaged segments. The initial 3–4 s signals of the experiments were eliminated, as the testing subjects were not yet immersed in the experimental states.
- **Step 6:** 3-sigma principle. The outliers were eliminated using the 3-sigma principle.
- **Step 7:** Data wrangling. Data cleaning, duplicates removal, and addressing missing data.

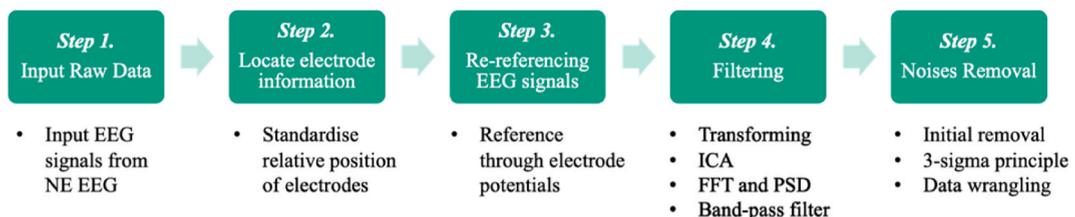


Fig. 4. 20-Lead EEG electrode system utilised.

**Algorithm 1.** Pseudo-code for EEG pre-processing.

**Algorithm 1:** Pseudo-code for EEG signals pre-processing

```

Inputs:
 $X = X_1^t, X_2^t, X_3^t, \dots, X_n^t$ ;  $X$  is the raw corrupted continuous EEG signals inputted from the NE system,  $t$  is the time interval,  $n$ 
denotes the overall lasting time
 $X_t$  is the collected signals at a time interval,  $X_t^i$  is the collected signals given a specific time interval,  $X_t^n$  is the signals at the
last time interval
Initialisation:
Set  $i = 0$ ;  $k = 1$ ,  $K = 20$ ;  $i$  is initialisation,  $k$  is the  $k$ th electrode in  $K$ ,  $K$  is the total number of electrodes of the EEG cap
 $new\_X = []$ ; empty array for storing temporarily processed EEG signals
 $New\_X = []$ ; empty array for storing the final processed EEG signals
for  $k$  in  $K$  electrodes do
    while  $i \leq len(X)$  do
        Map  $X_t^i$  with  $k$  electrodes
         $i = i + 1$ 
    end
end
for  $k$  in  $K$  electrodes do
    while  $i \leq len(X)$  do
        Re-reference  $X_t^i$  with  $k$  electrodes
         $i = i + 1$ 
    end
end
for  $k$  in  $K$  electrodes do
    Filter  $X$  in  $k$  electrode based on high-pass and low-pass filters
    if  $X_t^i$  in the filtering range then
        Apply ICA to extract frequency bands  $\alpha, \beta, \delta, \theta, \gamma$ 
        Add  $X_t$  into  $new\_X$ 
    end
    while  $i \leq len(new\_X)$  do
        Apply FFT and PSD calculation
        Apply band-pass filtering
        Add  $X_t^i$  into  $New\_X$ 
        while  $i$  in  $New\_X$  do
            Remove corrupted signals
            Apply 3-sigma principle
            Apply data wrangling
             $i = i + 1$ 
        end
         $i = i + 1$ 
    end
end
end
Output: Clean EEG signals  $\rightarrow New\_X$ 

```

Afterwards, the SA was conducted to label the transformed EEG signals. In order to label the data for Level 1 of SA, the SA state of the observer was judged based on distance errors between the predicted aircraft position and its actual position. The smaller the sum of the position error was, the better the SA status of the observer was. Details on labelling the Level 1 SA status can be found in Table 3.

To label the records for Level 2 of SA, the state was judged based on the number of correctly identified aircraft types. The higher number of correctly selected aircraft was, the better the SA of the observer was. Details on Level 2 SA status identification rules can be found in Table 4.

With respect to label Level 3 SA status, the observer was judged based on the angular errors between the predicted motion and the actual motion of the target aircraft. The smaller the sum of the angular error was, the better the SA status of the observer was. Judgmental rules can be found in Table 5. The processed signals were further input into the feature selection Component C module for correlation analysis.

2.3. Feature extraction and selection

The characteristic dimension of the processed EEG samples was with 100 dimensions, including 100 characteristic indicators of

**Table 3**  
Judgmental rules for SA Level 1 perception.

Sum of position errors	SA Status Labels
$[1000, +\infty)$	Status 1
$[250, 1000)$	Status 2
$[0, 250)$	Status 3

**Table 4**  
Judgmental rules for SA Level 2 comprehension.

Sum of correct selections	SA Status Labels
0 – 2	Status 1
3 – 4	Status 2
5	Status 3

**Table 5**  
Judgmental rules for SA Level 3 projection.

Sum of angular errors	SA Status Labels
$[400, +\infty)$	Status 1
$[50, 400)$	Status 2
$[0, 50)$	Status 3

EEG data with 20 electrodes  $\times$  5 bands (including Alpha, Beta, Delta, Theta, and Gamma waves). With such high dimensions, the potential difficulties might be brought by directly classifying them, such as the “curse of dimensionality”, “empty space”, or even “algorithm failure” [42]. To avoid such potentials, the 100-dimensional EEG records were initially processed with the principal component analysis (PCA) [43] to obtain a new principal component for dimension reduction in the feature extraction stage.

Feature extraction and selection techniques were applied through the Shapiro-Wilk test [44] to evaluate the sample for normality. Afterwards, with the objective to better select critical features, a set of feature selection algorithms were compared on the basis of the EEG data analysis procedure [45], including fisher score [46], mutual information [47], and sensitivity analysis. Accordingly, the extracted and selected mutual features contain five components in each electrode, including the  $\alpha$ ,  $\beta$ ,  $\delta$ ,  $\theta$ , and  $\gamma$  waves, ready for SA evaluation.

#### 2.4. Discriminative model implementation

In the Component D discriminative model implementation stage, the classic machine learning classifiers were compared to the advanced deep learning CNN models to ultimately evaluate the three-levels of SA.

With the selected salient EEG signal features, a series of machine learning models were adopted in the conventional classifier (CC) module, including the most commonly used algorithms: K-Nearest Neighbour (KNN), Decision Tree (DT), Support Vector Machines (SVM), Random Forest (RF), AdaBoost (AdaB), and Back Propagation (BP) network. In order to conduct the comparative analysis of the two alternative approaches, the image-based CNN models were applied to process BEAM images. More specifically, the collected EEG signals from the 20-lead electrodes were imported into the EEGLAB software [48]. We used EEGLAB software embedded Topoplot function to construct BEAM images from EEG signals with the independent components of EEG signal after ICA processing. With the obtained BEAM images, a set of highly acknowledged CNN models were deployed in the CNN module, including VGG [49], ResNet [50], Inception [51], DenseNet [52], and Xception [53]. This study selects the most widely used classifiers from different research domains for comprehensive analysis, including VGG-19, ResNet50, InceptionV3, DenseNet121, and Xception. Transfer learning was applied to use the pre-trained models from Keras applications on our dataset. The performance from the two approaches was compared for future research recommendations.



**Fig. 5.** Raw data acquisition process and operational interface.

### 3. Experiment

With the ethics obtained from the Medical and Experimental Human and Animal Ethics Committee of Northwestern Polytechnical University, this section describes the data acquisition and parameters setting to better demonstrate the conduction of experiments.

#### 3.1. Dataset acquisition

A total of 25 subjects were involved in the data acquisition stage, including 20 male and 5 female operators, with an average age of  $23.5 (\pm 2.12)$ . All the participants are graduate research students from Northwestern Polytechnical University. Before conducting the pre-established experimental simulations, all the subjects were selected under rigorous conditions that they have to be 1) in good health, 2) right-handed, 3) with normal vision, 4) with normal hearing, 5) adequate sleep, and 6) avoid strenuous exercise before experiments.

In order to preserve the anonymity of the participants, their physiological records have all been de-identified during the data acquisition stage by removing their names, addresses, and contact information. Informed consents were obtained from all participants for this research.

Fig. 5 illustrates the raw biomedical EEG signals acquisition process, where the participant will conduct the pre-mentioned three tasks for each SA level using the deployed human-computer interaction simulation system. The proposed simulation system was designed following the guidelines proposed by the U.S. Air Force Research Laboratory, including the battlefield situation information module, formation and weapon information module, single machine detailed information module, battlefield information module, and interactive control module.

#### 3.2. Parameters setting

During the data acquisition stage, the NE Wireless EEG equipment has the sampling frequency set to 500 Hz, and the EEG acquisition electrodes use dry electrodes.

With the band-pass filter, five different waves were filtered using 8–13 Hz for  $\alpha$  waves, 13–30 Hz for  $\beta$  waves, 0–4 Hz for  $\delta$  waves, 4–8 Hz for  $\theta$  waves, and 30–50 Hz for  $\gamma$  waves (details can be viewed in Fig. 6).

The maximum offset between data points recorded from the EEG equipment and SA simulations was 2 ms, which was based on the temporal resolution of the EEG device. EEG data covering time  $t$  were isolated before the task was performed and only while the target was moving across the screen. Assuming participants' SA statuses were independent between each run, the results of each test were considered as a separate sample. As a result, a total of 2500 labelled samples (25 participants  $\times$  100 runs) were obtained for EEG signals, including 100 runs for each SA level. Fig. 7 details the dataset acquisition process.

The data obtained from the EEG acquisition system is composed of EEG signals and various noises. All the non-EEG signals are collectively called artefacts. Artefacts in EEG recordings can be divided into subject-related and technology-related categories. The subject-related artefacts mainly include electrooculography and electromyography. The technology-related artefacts mainly include electrical noise caused by the surrounding environment, that is, circuit noise. Therefore, the generated BEAM images from those EEG signals contain a large portion of images related to artefacts, rather than to EEG components, thus, those BEAM images have to be eliminated. As a result, a total of 660-sized  $224 \times 224$  BEAM images were applied (see Fig. 8), and the specific distribution over different classes is outlined in Table 6.

In the classifier implementation stage, all experiments were conducted under the same computational environment. All the SA evaluation experiments adopted the Adam optimizer. Initially, the learning rate was set to  $1 \times 10^{-3}$ , and it was gradually updated using the gradient descents algorithm during fine-tuning; eventually, the learning rate of  $1 \times 10^{-6}$  showed the most stable performance for VGG models and was fixed thereafter, and  $1 \times 10^{-5}$  was the most stable learning rate for the rest of models. The 10-fold stratified cross-validation was applied in the experiments to reach consistency and address the data imbalance issue [54]. During each training iteration, the batch size was set to 5.

Additionally, the categorical cross-entropy (CCE) was adopted as a loss function for the classification tasks. CCE is generally applied in multi-class classification tasks as it can be weighted based on different classes. In this case, CCE is appropriate as SA status was labelled under three levels so that it can adapt the penalty of a probabilistic false-negative rate for each class [55]. The CCE is calculated using Eq. (1).

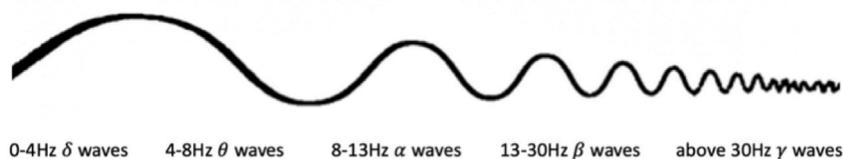


Fig. 6. EEG frequency band.

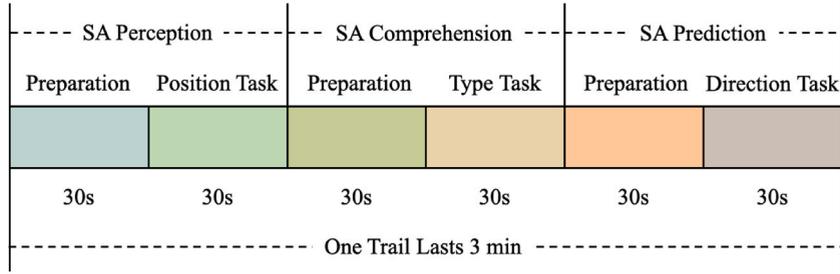


Fig. 7. Detailed setup for SA tasks during data acquisition.

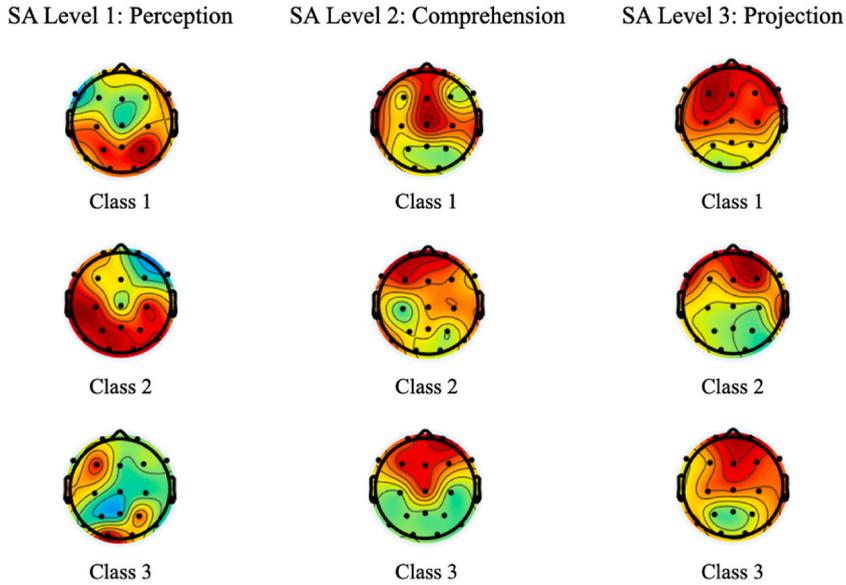


Fig. 8. Sample BEAM demonstration.

**Table 6**  
BEAM image distribution in different SA levels.

SA Levels	SA Status	No. of Images	Total Number
Level 1	1	60	220
	2	140	
	3	20	
Level 2	1	40	220
	2	40	
	3	140	
Level 3	1	40	220
	2	100	
	3	80	

$$CCE = -\log\left(\frac{e^{s_p}}{\sum_j^C e^{s_j}}\right) \tag{1}$$

With the one-hot encoded labels for each target, CCE will produce a vector indicating the probability of each class. Specifically,  $s_p$  denotes the predicted score for the given class,  $s_j$  is the inferred score for each class in  $C$ , and  $C$  is the total number of classes, in this case, 3 for the three levels of SA statuses. And the study deployed accuracy, precision, recall, and F1 scores for model evaluation.

Additionally, this work is highly reproducible as all the algorithmic protocols are publicly available through the GitHub link: <https://github.com/Amyyy-z/SA-Assessment>.

## 4. Experimental results and analysis

With the pre-defined steps for data pre-processing and feature selection, salient features were extracted for SA evaluation. This section interprets those features and the corresponding discriminative model performance.

### 4.1. Salient features analysis

**Table 7** demonstrates the selected salient features under three levels of SA. The first level of SA perception is to grasp the battlefield information through cognition, thus, the parietal and occipital lobes areas are more prominent, and the  $\beta$  and  $\gamma$  bands are the dominating frequency bands. The parietal region functions are to regulate or allocate spatial attention, and the occipital region is associated with the integration of visual information. More specifically, the  $\beta$  band indicates whether the aviator is in a state of excitement or stress, and it is significant when the participant is with tension and anticipation [56]. The  $\gamma$  band is associated with cognitive and advanced information processing functions, such as learning, memory, and information processing.

The understanding of the second level of SA comprehension is based on the perception of the first level, including integrating, memorizing, and understanding the information of the perceived elements of the battlefield situation. Therefore, the parietal lobe, the paracentral lobe (i.e., related to thinking and calculation), and the occipital lobe are more prominent, with the  $\beta$  and  $\gamma$  bands more outstanding.

The projection level is mainly based on the perception level, including information understanding for judgmental decisions about the future, mainly referring to the thinking and inference processes. Therefore, salient regions, including the parietal lobe, the paracentral lobe, and the occipital lobe, with the  $\beta$ ,  $\gamma$ , and  $\theta$  bands. The  $\theta$  band can usually be found in healthy and alert infants, children, and adults with drowsiness, and it is also critical when encountering operation difficulties, and it is also detectable in deep meditative states.

It should be noticed that the  $\alpha$  and  $\delta$  bands were not selected as salient during the feature selection phase, and this is due to the characteristics of these two bands. The  $\alpha$  band is usually detectable in healthy, awake adults when they are relaxed or mentally inactive [56]. The  $\delta$  band represents prolonged activity and generally can be detected in infants up to 1-year-old or healthy adults with deep sleep stage [56], therefore, the two bands were filtered out in this case.

### 4.2. Classic discriminative model performance

With the selected salient EEG features, **Table 8** demonstrates the conventional machine learning-based discriminative model classification performance. When designing the data collection experiments, the three stages of SA status were outlined in each level, and the third stages of SA status among all three levels are the most critical ones, therefore, are interpreted and exhibited.

For the perception level, the SVM algorithm demonstrated the optimal performance, achieving an accuracy of 0.893. The BP network performed as the second-best, with an accuracy of 0.857, whereas the least accuracy was obtained by the RF algorithm with an accuracy of 0.778. For the comprehension level of SA, the SVM model again achieved the highest accuracy of 0.893 when correctly identifying the aircraft type, the BP network obtained the same accuracy rate, whereas the AdaBoost generated the worst accuracy rate of 0.750. As for the projection level, SVM, AdaBoost, and BP models outperformed other models and reached an accuracy of 0.929 when predicting the motion of aircraft, yet the KNN algorithm produced the worst accuracy of 0.821 (see **Fig. 9**).

### 4.3. CNN performance

**Table 9** exhibits the CNN model performance when incorporating with BEAM images. For the first level of SA, the DenseNet model achieved the best performance with an accuracy of 0.955, which also outperforms the SVM classifier using EEG signals. The second-best model was VGG which generated an accuracy of 0.909 for the perception level of SA, again outperformed the classic machine learning algorithms. For the second level of SA, the VGG model obtained the most promising performance, with an accuracy of 0.955. Xception obtained the second-best performance with an accuracy of 0.909. The other models produced an accuracy of 0.864 as the bottom line, yet it is still superior to the KNN, AdaBoost, RF, and DT algorithms. As for the third level of SA, the DenseNet and Xception models produced the most optimal result, reaching an accuracy of 0.773, which is much lower than the results produced by classic machine learning models. In generally, the CNN-based performance obtained satisfying results for the perception and comprehension layers of SA, whereas they cannot compete with conventional EEG-based classifiers when assessing the projection layer. Details are illustrated in **Fig. 10**.

## 5. Discussion

Although there exist few studies on adopting EEG signals for SA assessment, there is still a paucity of work on using BEAM images for analysis, let alone the comparative study between the two approaches. This study bridges this gap and interprets the performance of both techniques.

**Fig. 11** displays the performance comparison with the best-performing models from the two approaches for the three levels of SA assessment. Specifically, the 3 bars represent three levels of SA, and the dashed line indicates the decrease of the projection level from the EEG-based method to the BEAM-based method. For the perception level, the DenseNet model obtained the best performance, with an average accuracy of 0.955. The VGG model generates an average accuracy of 0.909, which is also above SVM classifier. Xception

**Table 7**  
Salient Features for Three-Levels of SA

Level 1 SA: Perception
<i>P7-β, P4-β, PZ-β, P5-β, P8-β, O1-β, O2-β, CZ-β, PZ-γ, P8-γ, O1-γ, O2-γ, P4-θ, CZ-θ</i>
Level 2 SA: Comprehension
<i>P7-β, P8-β, O1-β, O2-β, P4-γ, PZ-γ, P5-γ, O1-γ, O2-γ, CZ-γ, C3-γ, C4-γ, P4-θ, P5-θ, CZ-θ</i>
Level 3 SA: Projection
<i>P4-β, P5-β, CZ-β, P7-γ, PZ-γ, P5-γ, P8-γ, O1-γ, O2-γ, CZ-γ, C4-γ, CZ-θ, C3-θ</i>

**Table 8**  
EEG salient features for three-levels of SA assessment.

Model	SA Level	Measurement			
		Accuracy	Precision	Recall	F1
KNN	Level 1: Perception	0.821	0.824	0.821	0.821
	Level 2: Comprehension	0.857	0.880	0.857	0.856
	Level 3: Projection	0.821	0.848	0.821	0.824
SVM	Level 1: Perception	<b>0.893</b>	0.920	0.893	0.892
	Level 2: Comprehension	<b>0.893</b>	0.899	0.893	0.894
	Level 3: Projection	<b>0.929</b>	0.940	0.929	0.929
AdaBoost	Level 1: Perception	0.786	0.856	0.786	0.780
	Level 2: Comprehension	0.750	0.841	0.750	0.757
	Level 3: Projection	<b>0.929</b>	0.944	0.929	0.930
DT	Level 1: Perception	0.786	0.814	0.786	0.781
	Level 2: Comprehension	0.786	0.793	0.786	0.785
	Level 3: Projection	0.893	0.920	0.893	0.896
RF	Level 1: Perception	0.778	0.689	0.778	0.728
	Level 2: Comprehension	0.786	0.786	0.786	0.784
	Level 3: Projection	0.857	0.890	0.857	0.849
BP	Level 1: Perception	0.857	0.857	0.857	0.857
	Level 2: Comprehension	<b>0.893</b>	0.922	0.893	0.891
	Level 3: Projection	<b>0.929</b>	0.932	0.929	0.928



**Fig. 9.** CC model performance.

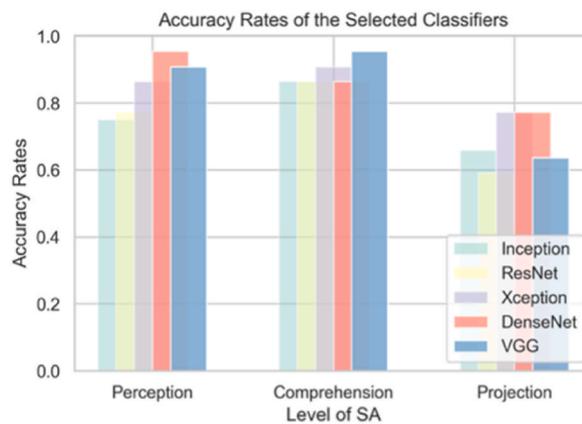
generates an average accuracy of 0.864, which is slightly lower than the SVM model. With the comprehension level, the VGG model is in the leading point, with an accuracy of 0.955, and the Xception model gains the second-best position with an accuracy of 0.909. For the projection level, three models achieved comparable performance, all from the classic machine learning approaches, being the SVM, BP, and AdaBoost models, reaching a weighted average of 0.929. Based on the results, this study provides a comprehensive analysis and investigations on the three-level SA assessment, and the evaluated models demonstrated promising performance, outperforming the accuracy obtained by Kästle et al. [28] and Li et al. [5].

The deep learning models eliminated the manual feature extraction and selection procedures, thus, being more efficient than the classic machine learning approaches. And notably, the BEAM-based CNN models produced superior classification results compared to the classic EEG-based discriminative models when evaluating the first and second levels of SA. Hence, the use of advanced CNN models in detecting the perception and comprehension levels of SA is quite feasible and advisable.

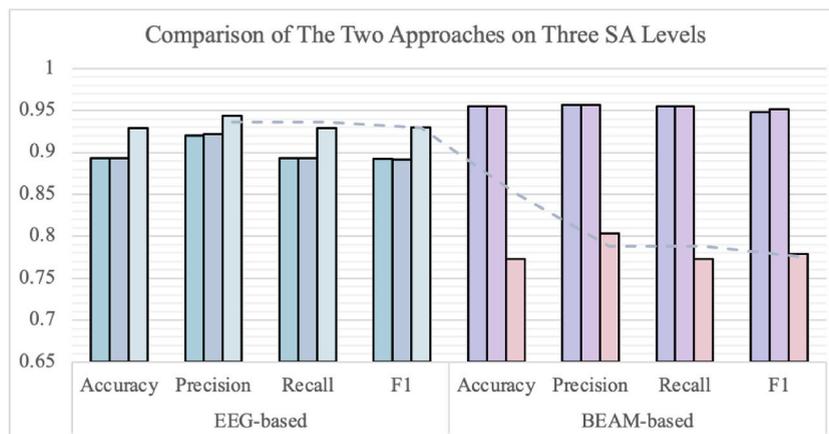
However, based on the dashed line from Fig. 11, the CNN performance was significantly reduced when evaluating the third level of

**Table 9**  
BEAM images with CNN for three-levels of SA assessment.

Model	SA Level	Measurement			
		Accuracy	Precision	Recall	F1
VGG	Level 1: Perception	0.909	0.828	0.909	0.866
	Level 2: Comprehension	<b>0.955</b>	0.957	0.955	0.951
	Level 3: Projection	0.636	0.636	0.636	0.583
ResNet	Level 1: Perception	0.773	0.736	0.773	0.702
	Level 2: Comprehension	0.864	0.750	0.864	0.802
	Level 3: Projection	0.591	0.501	0.591	0.525
Inception	Level 1: Perception	0.750	0.671	0.750	0.684
	Level 2: Comprehension	0.864	0.760	0.864	0.805
	Level 3: Projection	0.659	0.695	0.659	0.641
Xception	Level 1: Perception	0.864	0.884	0.864	0.855
	Level 2: Comprehension	0.909	0.915	0.909	0.909
	Level 3: Projection	<b>0.773</b>	0.803	0.773	0.779
DenseNet	Level 1: Perception	<b>0.955</b>	0.957	0.955	0.948
	Level 2: Comprehension	0.864	0.846	0.864	0.844
	Level 3: Projection	<b>0.773</b>	0.768	0.773	0.767



**Fig. 10.** CNN model performance.



**Fig. 11.** CC & CNN model performance comparison.

SA from 0.929 to 0.773. The reason behind it might be due to the characteristics of EEG signals, which were established continuously with a time-to-event mechanism [57]. More specifically, the first two tasks were designed to emphasize the current cognition of operators in order to finish performing the tasks. However, the third task was built upon aviators' judgements of current understanding and took a step further for making a prediction, and this is where EEG signals are more solid and accountable than BEAM images. With

respect to the physiological records involved in this study, it is more convincing that EEG signals and BEAM images both can generate competitive performance when interpreting different levels of SA. Accordingly, in our future research, we plan to make use of both modalities and integrate them into a multi-modal evaluation tool for automatic SA evaluation.

## 6. Conclusion

This study is the first of its kind, which deploys two types of physiological modalities for a comprehensive of three-level SA evaluation. The EEG signals were extracted and analysed with a series of conventional machine learning approaches, and the BEAM images were interpreted through more advanced deep learning-based CNN models. A comparative analysis was conducted among the two approaches in this research for future relevant research recommendations. The results demonstrated that BEAM-based models produced better evaluation results for the perception and comprehension levels of SA, with an accuracy rate of 0.955 reached. The EEG-based classifiers generated higher accuracy for the projection level of SA, being 0.929.

The future plan would focus on acquiring more data samples to establish more robust algorithms for precise SA evaluation while considering more complex situations. For example, future work will consider the workload impact on SA status, such as insufficient information, information overload, and emergency confrontation. In the meantime, we would like to deploy a multi-modality form of the evaluation model to achieve a more accurate three-level SA assessment by including more physiological data types like electro-myography and eye-tracking records.

## Funding

The authors would like to acknowledge the National Natural Science Foundation of China (No. 61305133) and the Aeronautical Science Foundation of China (Grant No. 2020Z023053002) to provide the fundings for conducting the research.

## Ethics statement

This study was conducted with the ethics approval obtained from the Medical and Experimental Human and Animal Ethics Committee of Northwestern Polytechnical University (Project ID: 202102055) and written consents were obtained from all the participants.

## CRedit authorship contribution statement

**Jun Chen:** Writing – review & editing, Visualization, Validation, Supervision, Project administration, Funding acquisition, Formal analysis, Conceptualization. **Anqi Chen:** Resources, Writing – review & editing, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Data curation. **Bingkun Jiang:** Resources, Validation, Software, Investigation, Data curation. **Xinyu Zhang:** Writing – review & editing, Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Formal analysis, Conceptualization.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.heliyon.2024.e26409>.

## References

- [1] R. Mica, Endsley. *SITUATION AWARENESS*, Chapter 17, John Wiley & Sons, Ltd, 2021, pp. 434–455, <https://doi.org/10.1002/9781119636113.ch17>. ISBN 9781119636113.
- [2] Ting Zhang, Jing Yang, Nade Liang, Brandon J. Pitts, Kwaku Prakah-Asante, Reates Curry, Duerstock Bradley, Juan P. Wachs, Denny Yu, Physiological measurements of situation awareness: a systematic review, *Hum. Factors* 65 (5) (2023) 737–758, <https://doi.org/10.1177/0018720820969071>.
- [3] Luke S. Snyder, Yi-Shan Lin, Morteza Karimzadeh, Dan Goldwasser, S. David, Ebert, Interactive learning for identifying relevant tweets to support real-time situational awareness, *IEEE Trans. Visual. Comput. Graph.* 26 (1) (2020) 558–568, <https://doi.org/10.1109/TVCG.2019.2934614>.
- [4] Qinqiao Li, K. Kam, H. Ng, Simon C.M. Yu, Yin Yiu Cho, Mengtao Lyu, Recognising situation awareness associated with different workloads using eeg and eye-tracking features in air traffic control tasks, *Knowl. Base Syst.* 260 (2023) 110179, <https://doi.org/10.1016/j.knosys.2022.110179>.
- [5] Ruilin Li, Lipo Wang, Olga Sourina, Subject matching for cross-subject eeg-based recognition of driver states related to situation awareness, *Methods* 202 (2022) 136–143, <https://doi.org/10.1016/j.jymeth.2021.04.009>. Machine Learning Methods for Bio-Medical Image and Signal Processing: Recent Advances.
- [6] M.R. Endsley, Situation awareness global assessment technique (sagat), in: *Proceedings of the IEEE National Aerospace and Electronics Conference* vol. 3, 1988, p. 789, <https://doi.org/10.1109/NAECON.1988.195097>. –795.
- [7] Thanh Nguyen, Chee Peng Lim, Ngoc Duy Nguyen, Lee Gordon-Brown, Saeid Nahavandi, A review of situation awareness assessment approaches in aviation environments, *IEEE Syst. J.* 13 (3) (2019) 3590–3603, <https://doi.org/10.1109/JSYST.2019.2918283>.

- [8] Shuo Liu, Huan Liu, Vijay John, Liu Zheng, Erik Blasch, Enhanced situation awareness through cnn-based deep multimodal image fusion, *Opt. Eng.* 59 (5) (2020) 053103, <https://doi.org/10.1117/1.OE.59.5.053103>, 053103.
- [9] Qi Wang, Siqi Bu, Zhengyou He, Zhao Yang Dong, Toward the prediction level of situation awareness for electric power systems using cnn- lstm network, *IEEE Trans. Ind. Inf.* 17 (10) (2021) 6951–6961, <https://doi.org/10.1109/TII.2020.3047607>.
- [10] Chunhua Ye, Zhong Yin, Mengyuan Zhao, Ying Tian, Zhanquan Sun, Identification of mental fatigue levels in a language understanding task based on multi-domain eeg features and an ensemble convolutional neural network, *Biomed. Signal Process Control* 72 (2022) 103360, <https://doi.org/10.1016/j.bspc.2021.103360>.
- [11] Mica R. Endsley, Measurement of situation awareness in dynamic systems, *Hum. Factors* 37 (1) (1995) 65–84, <https://doi.org/10.1518/001872095779049499>.
- [12] David N. Hogg, K.N.U.T. Follesø, Frode Strand-Volden, Belen Torralba, Devel- opment of a situation awareness measure to evaluate advanced alarm systems in nuclear power plant control rooms, *Ergonomics* 38 (11) (1995) 2394–2413, <https://doi.org/10.1080/00140139508925275>.
- [13] Mica R. Endsley, Stephen J. Selcon, Thomas D. Hardiman, Darryl G. Croft, A comparative analysis of sagat and sart for evaluations of situation awareness, in: *Proceedings of the Human Factors and Ergonomics Society Annual Meeting vol. 42*, SAGE Publications Sage CA, Los Angeles, CA, 1998, pp. 82–86, <https://doi.org/10.1177/154193129804200119>.
- [14] Yorck Hauss, Klaus Eyferth, Securing future atm-concepts' safety by measuring situation awareness in atc, *Aero. Sci. Technol.* 7 (6) (2003) 417–427, [https://doi.org/10.1016/S1270-9638\(02\)00011-1](https://doi.org/10.1016/S1270-9638(02)00011-1).
- [15] Xu Wu, Chuanyan Feng, Xiaoru Wanyan, Yu Tian, Shoupeng Huang, Dynamic measurement of pilot situation awareness, in: *International Conference on Engineering Psychology and Cognitive Ergonomics*, Springer, 2017, pp. 306–316, <https://doi.org/10.1007/978-3-319-58475-123>.
- [16] Mica R. Endsley, A systematic review and meta-analysis of direct objective measures of situation awareness: a comparison of sagat and spam, *Hum. Factors* 63 (1) (2021) 124–150, <https://doi.org/10.1177/001872081987537>.
- [17] Wayne L. Waag, Michael R. Houck, Tools for assessing situational awareness in an operational fighter environment, *Aviat Space Environ. Med* 65 (5 Suppl) (1994) A13. -9.
- [18] Barry McGuinness, Louise Foy, A subjective measure of sa: the crew awareness rating scale (cars), in: *Proceedings of the First Human Performance, Situation Awareness, and Automation Conference vol. 16*, 2000, pp. 286–291. Savannah, Georgia.
- [19] Barry McGuinness, *Quantitative Analysis of Situational Awareness (Quasa): Applying Signal Detection Theory to True/false Probes and Self-Ratings*, Bae systems Bristol (United Kingdom) advanced technology centre, 2004. Technical report.
- [20] Richard M. Taylor, *Situational awareness rating technique (sart): the development of a tool for aircrew systems design*, in: *Situational Awareness*, Routledge, 2017, pp. 111–128.
- [21] Francis T. Durso, Carla A. Hackworth, R Truitt Todd, Jerry Crutchfield, Danko Nikolic, Carol A. Manning, Situation awareness as a predictor of performance for en route air traffic controllers, *Air Traffic Control Q* 6 (1) (1998) 1–20, <https://doi.org/10.2514/atcq.6.1.1>.
- [22] Christopher Kelly, Emmanuelle Jeannot, and Oliver Straeter. Development of a Situation Awareness Measure for Atm Simulations..
- [23] Michael D. Matthews, A Beal Scott, *Assessing Situation Awareness in Field Training Exercises*, Military Academy West Point NY Office of Military Psychology and Leadership, 2002. Technical report.
- [24] Kenneth Kwok, S.S. Virdi, Ai-based situation awareness assessment, in: *Journal of Physics: Conference Series vol. 2311*, IOP Publishing, 2022 012011, <https://doi.org/10.1088/1742-6596/2311/1/012011>.
- [25] Matteo Fraschini, Arjan Hillebrand, Matteo Demuru, Luca Didaci, Gian Luca Marcialis, An eeg-based biometric system using eigenvector centrality in resting state brain networks, *IEEE Signal Process. Lett.* 22 (6) (2015) 666–670, <https://doi.org/10.1109/LSP.2014.2367091>.
- [26] Magdiel Jimenez-Guarneros, Pilar G'omez-Gil, Custom domain adaptation: a new' method for cross-subject, eeg-based cognitive load recognition, *IEEE Signal Process. Lett.* 27 (2020) 750–754, <https://doi.org/10.1109/LSP.2020.2989663>.
- [27] Feng Zhou, X. Jessie Yang, Joost C.F. de Winter, Using eye-tracking data to predict situation awareness in real time during takeover transitions in conditionally automated driving, *IEEE Trans. Intell. Transport. Syst.* 23 (3) (2022) 2284–2295, <https://doi.org/10.1109/TITS.2021.3069776>.
- [28] Jan Luca Kastle, Bani Anvari, Jakub Krol, Helge A. Wurdemann, Correlation between situational awareness and eeg signals, *Neurocomputing* 432 (2021) 70–79, <https://doi.org/10.1016/j.neucom.2020.12.026>.
- [29] Yin Yiu Cho, Kam K.H. Ng, Xinyu Li, Xiaoge Zhang, Qinbiao Li, Hok Sam Lam, and Man Ho Chong. Towards safe and collaborative aerodrome operations: assessing shared situational awareness for adverse weather detection with eeg-enabled bayesian neural networks, *Adv. Eng. Inf.* 53 (2022) 101698, <https://doi.org/10.1016/j.aei.2022.101698>.
- [30] Jieli Chen, Xuefen Lin, Weifeng Ma, Yuchen Wang, Wei Tang, Eegbased emotion recognition for road accidents in a simulated driving environment, *Biomed. Signal Process Control* 87 (2024) 105411, <https://doi.org/10.1016/j.bspc.2023.105411>.
- [31] Joost CF de Winter, Yke Bauke Eisma, C.D.D. Cabrall, Peter A. Hancock, Neville A. Stanton, Situation awareness based on eye movements in relation to the task environment, *Cognit. Technol. Work* 21 (1) (2019) 99–111, <https://doi.org/10.1007/s10111-0180527-6>.
- [32] Ruilin Li, Zirui Lan, Jian Cui, Olga Sourina, Lipo Wang, Eeg-based recognition of driver state related to situation awareness using graph convolutional networks, in: *2020 International Conference on Cyberworlds (CW)*, 2020, pp. 180–187, <https://doi.org/10.1109/CW49994.2020.00037>.
- [33] Qinbiao Li, Yin Yiu Cho, C. Simon, M. Yu, Kam K.H. Ng, Situational awareness and flight approach phase event recognition based on psychophysiological measurements, in: *2021 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, 2021, pp. 1308–1312, <https://doi.org/10.1109/IEEM50564.2021.9673081>.
- [34] Chuanyan Feng, Shuang Liu, Xiaoru Wanyan, Hao Chen, Yuchen Min, Yilan Ma, Eeg feature analysis related to situation awareness assessment and discrimination, *Aerospace* 9 (10) (2022), <https://doi.org/10.3390/aerospace9100546>.
- [35] Jinxuan Shi, Kun Wang, Fatigue driving detection method based on time-space frequency features of multimodal signals, *Biomed. Signal Process Control* 84 (2023) 104744, <https://doi.org/10.1016/j.bspc.2023.104744>.
- [36] Yuhao Zhang, Hanying Guo, Yongjiang Zhou, Chengji Xu, Yang Liao, Recognising drivers' mental fatigue based on eeg multi-dimensional feature selection and fusion, *Biomed. Signal Process Control* 79 (2023) 104237, <https://doi.org/10.1016/j.bspc.2022.104237>.
- [37] Huijie Jia, Zhongjun Xiao, Ji Peng, End-to-end fatigue driving eeg signal detection model based on improved temporal-graph convolution network, *Comput. Biol. Med.* 152 (2023) 106431, <https://doi.org/10.1016/j.combiomed.2022.106431>.
- [38] Yun Zheng, Yuliang Ma, Jared Cammon, Songjie Zhang, Jianhai Zhang, Yingchun Zhang, A new feature selection approach for driving fatigue eeg detection with a modified machine learning algorithm, *Comput. Biol. Med.* 147 (2022) 105718, <https://doi.org/10.1016/j.combiomed.2022.105718>.
- [39] Hui Wu, Hui Zhang, Jinfang Zhang, Fanjiang Xu, Fast aircraft detection in satellite images based on convolutional neural networks, in: *2015 IEEE International Conference on Image Processing, ICIP*, 2015, pp. 4210–4214, <https://doi.org/10.1109/ICIP.2015.7351599>.
- [40] Shane T. Mueller, Brian J. Piper, The psychology experiment building language (pebl) and pebl test battery, *J. Neurosci. Methods* 222 (2014) 250–259, <https://doi.org/10.1016/j.jneumeth.2013.10.024>.
- [41] Jair Montoya-Martinez, Vanthornhout Jonas, Alexander Bertrand, Tom Francart, Effect of number and placement of eeg electrodes on measurement of neural tracking of speech, *PLoS One* 16 (2) (2021) e0246769, <https://doi.org/10.1371/journal.pone.0246769>.
- [42] Mario Koppen, The curse of dimensionality, in: *5th Online World Conference on Soft Computing in Industrial Applications (WSC5) vol. 1*, 2000, pp. 4–8.
- [43] Herve Abdi, J Williams Lynne, *Principal component analysis*, Wiley interdisciplinary reviews: *Comput. Stat.* 2 (4) (2010) 433–459.
- [44] Samuel Sanford Shapiro, B Wilk Martin, *An analysis of variance test for normality (complete samples)*, *Biometrika* 52 (3/4) (1965) 591–611.
- [45] Jun Chen, X.U.E. Lei, R.O.N.G. Jia, Xudong Gao, Real-time evaluation method of flight mission load based on sensitivity analysis of physiological factors, *Chin. J. Aeronaut.* 35 (3) (2022) 450–463, <https://doi.org/10.1016/j.cja.2021.11.010>.
- [46] Quanquan Gu, Zhenhui Li, Jiawei Han, *Generalized Fisher Score for Feature Selection*, 2012 *arXiv preprint arXiv:1202.3725*.
- [47] Benedikt Gierlichs, Lejla Batina, Pim Tuyls, Bart Preneel, Mutual information analysis: a generic side-channel distinguisher, in: *International Workshop on Cryptographic Hardware and Embedded Systems*, Springer, 2008, pp. 426–442, <https://doi.org/10.1007/978-3-540-85053-327>.

- [48] Arnaud Delorme, Makeig Scott, Eeglab: an open source toolbox for analysis of single-trial eeg dynamics including independent component analysis, *J. Neurosci. Methods* 134 (1) (2004) 9–21, <https://doi.org/10.1016/j.jneumeth.2003.10.009>.
- [49] Karen Simonyan, Andrew Zisserman, Very Deep Convolutional Networks For Largescale Image Recognition, 3rd International Conference on Learning Representations, 2015, pp.1-14.
- [50] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.
- [51] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Reed Scott, Dragomir Anguelov, Dumitru Erhan, Vanhoucke Vincent, Andrew Rabinovich, Going deeper with convolutions, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.
- [52] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, Kilian Q. Weinberger, Densely connected convolutional networks, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 4700–4708.
- [53] Francois Chollet, Xception: deep learning with depthwise separable convolutions, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1251–1258.
- [54] Swarnalatha Purushotham, B.K. Tripathy, Evaluation of classifier models using stratified tenfold cross validation techniques, in: P. Venkata Krishna, M. Rajasekhara Babu, Ezendu Ariwa (Eds.), *Global Trends in Information Systems and Software Applications*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2012, pp. 680–690, <https://doi.org/10.1007/978-3-642-29216-374>.
- [55] Yaoshiang Ho, Samuel Wookey, The real-world-weight cross-entropy loss function: modeling the costs of mislabeling, *IEEE Access* 8 (2020) 4806–4813, <https://doi.org/10.1109/ACCESS.2019.2962617>.
- [56] Thiago Gabriel Monteiro, Charlotte Skourup, Houxiang Zhang, Using eeg for mental fatigue assessment: a comprehensive look into the current state of the art, *IEEE Trans. Human-Machine Syst.* 49 (6) (2019) 599–610, <https://doi.org/10.1109/THMS.2019.2938156>.
- [57] Yue Gao, Xiangling Fu, Tianxiong Ouyang, Yi Wang, Eeg-gcn: spatio-temporal and self-adaptive graph convolutional networks for single and multi-view eeg-based emotion recognition, *IEEE Signal Process. Lett.* 29 (2022) 1574–1578, <https://doi.org/10.1109/LSP.2022.3179946>.