Reliability Engineering and System Safety 260 (2025) 110962

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

Assessing human situation awareness reliability considering fatigue and mood using EEG data: A Bayesian neural network-Bayesian network approach

Song Ding [a], Lunhu Hu [b], Xing Pan [a,*], Dujun Zuo [c], Liuwang Sun [a]

[a] School of Reliability & Systems Engineering, Beihang University, Beijing, China
[b] School of Mechanical Engineering, Inner Mongolia University of Technology, China
[c] Institute of System Engineering, China Academy of Engineering Physics, Mianyang, China

ARTICLE INFO

Keywords:
Situation awareness reliability
Bayesian neural network
Bayesian network
Fatigue
Mood

ABSTRACT

Situation awareness (SA) assessment is the process of acquiring and maintaining SA, which serves as a crucial indicator of operator task performance and behavioral safety in human-machine interaction. SA reliability is the evaluation of how well SA is established, and it is also the goal of SA assessment. Nonetheless, current SA assessment models rarely consider the influence of human physiological states, such as fatigue and mood, and rely heavily on subjective data. To address these deficiencies, this paper proposes a SA assessment model based on a Bayesian Neural Network (BNN) and Bayesian Network (BN), with a focus on examining the impact of fatigue and mood on the SA reliability. Firstly, fatigue and mood state classification models are developed using EEG data based on a BNN, and the uncertainty is assessed. Secondly, a BN model for SA reliability evaluation is proposed, where the uncertainty of BNN outputs is used as the prior probability, and conditional probability tables are established based on experimental statistics. Finally, a SA experiment is conducted using a civil aviation scenario based on the SAGAT platform to validate the proposed model. This model overcomes the limitations of previous approaches by leveraging objective physiological data and experimental statistics to infer the influence of physiological states on the SA reliability.

1. Introduction

Operators nowadays are required to make effective and immediate decisions in the face of increasingly complex operating systems [1]. Situation awareness (SA) is considered a prerequisite for operators to make effective decisions and perform correct operations, and it has been used in human factors research to explain the extent to which operators in complex human-computer systems are aware of what is going on in the system and the environment [2–4]. Maintaining proper SA among operators is essential for minimizing human errors, ultimately enhancing human reliability, which directly contributes to the improved overall reliability and safety of the human-computer system [5–7]. SA is defined as the perception of elements over the time and in the space, the understanding of their significance, and future prediction [8]. Largely, SA is regarded as a distinct knowledge state, while the process to achieve SA is termed situation assessment and refers to the methods for establishing, acquiring, and sustaining SA. This process may vary significantly depending on individual and contextual factors. In brief, SA assessment serves as the process of forging SA, while SA represents the culmination of this process [9,10]. The development of SA assessment models forms the foundational framework for evaluating operators' SA proficiency [11].

Many SA assessment models have been developed, and these models can be broadly categorized into two types: descriptive models and computational models. The descriptive models of SA assessment can elucidate the impact of both internal and external factors on SA and enhance the comprehension of the SA assessment process. A prime example of such models is Endsley's Three-Level Model of SA [8], which delineates the hierarchical progression of SA assessment grounded in information processing: from perception (Level 1 SA) to comprehension (Level 2 SA) and to projection (Level 3 SA). In this model, the achievement of SA is influenced by memory, workload, task complexity, and interactions between them [12]. Many other descriptive models have been put forth by scholars, including Bedny and Meister's

* Corresponding author.
E-mail address: panxing@buaa.edu.cn (X. Pan).

https://doi.org/10.1016/j.ress.2025.110962
Received 11 September 2024; Received in revised form 15 February 2025; Accepted 20 February 2025
Available online 21 February 2025

Perceptual-Action Loop [13] and Adams' Event Flow Model [14], among others. These models primarily encapsulate the fundamental principles and common characteristics governing how individuals process information and interact with their surroundings to achieve SA.

While the descriptive models offer a systematic description of the development of SA, they fall short of predicting how SA will develop under specific conditions. In contrast, the computational models of SA assessment can anticipate the progression of SA assessment in particular circumstances, which is essential for the design of human decision-making systems [15]. Examples of such models are relatively fewer: Kim and Seong [16] proposed an analytical computational approach for nuclear power plant operator SA assessment based on Bayesian inference. You and Guo [5] proposed a dynamic Bayesian network based reliability assessment method for short-term multi-round SA considering round dependencies, the effectiveness of the method was verified by Boeing 737–8 (MAX) accident. Schneider and Halekotte [17] proposed a systematic approach for SA assessment under emergency response based on Bayesian networks. All these models share a fundamental assumption that SA is essentially a kind of diagnostic inference. Consequently, they model SA assessment as an integrated diagnostic inference process, with the situation acting as the hypothesized cause, the event (high or low SA) as the effect, and the sensory (or sensor) data as the symptom (detected effect). Once the events are detected, Bayesian logic is employed to trace the situation-event relationships backward, evaluating their impact on the development of SA. Through inference based on the network of situation-event relationships, the occurrence of future events is better predicted with updated situation likelihood assessment. In summary, BN, functioning as causal inference networks, find widespread application in SA assessment models to capture the effects of various uncertainties on SA formation during human-computer interaction [11]. Nonetheless, these computational BN-based models grapple with two notable shortcomings.

On one hand, existing computational models of SA assessment have limited applicability as most of them are developed for specific safety scenarios such as the nuclear industry [16], maritime traffic [7,18] and hazardous chemicals production [19,20]. These models primarily focus on the inference process of how an operator develops SA under specific conditions, incorporating equipment and environment as contextual elements. However, they often neglect the operator factors, which may significantly impact both the process and outcome of situation assessment. In particular, operator factors like fatigue [21] and mood [22] require close attention. Fatigue significantly reduces operator performance, leading to decreased accuracy and alertness as fatigue accumulates. A recent 15-year accident report from the U.S. Air Force indicated that 24 % of Class A accidents were fatigue-related [23], a figure comparable to the 23 % observed in the civil aviation industry [24]. Mood plays a critical role in reflecting an individual's psychological state, and studies have shown that mood significantly affects operators' decision-making and alertness, thereby reducing performance [25,26]. Negative mood, in particular, is one of the contributing factors to traffic accidents [27]. In high-risk domains, it is essential to strictly control these factors to prevent accidents stemming from excessive fatigue and mood swings [28]. Therefore, more comprehensive consideration of the impact of fatigue and mood on the SA process is needed to enhance the accuracy and applicability of the assessment model.

On the other hand, it is not always possible to acquire sufficient data to obtain prior probabilities for BN [29]. Many existing SA assessment models choose assumptions and expert judgments [30], but it can be challenging to apply such subjective evaluation data in large-scale SA analysis. Another choice is objective evaluation. For instance, physiological measurements, such as data from electroencephalography (EEG) [31,32] sensors, can serve as objective indicators of an operator's mood [33] and overall functional state (such as workload [34], mental fatigue [35], and alertness [38]). The typical research paradigm involves segmenting EEG signals for feature extraction, followed by the use of machine learning or deep learning algorithms (particularly deep learning methods) to develop physiological state classification models [36,37]. However, the impact of uncertainty is an issue that these approaches tend to overlook, such as the aleatoric uncertainty encoded by the randomness of the deep learning process and the epistemic uncertainty caused by insufficient data [38]. Without quantifying uncertainty, models may become overly confident in their predictions, which is intolerable for safety-critical applications and could lead to dangerous outcomes, such as in autonomous driving [39] or medical diagnosis [40].

In fact, the purpose of establishing the computational model for SA assessment is to evaluate the operator's ability to correctly establish SA as a knowledge state within a specified time and under specific conditions. This is conceptually similar to the concept of human reliability, which refers to the operator's ability to complete a specified task within a given time and under specific conditions, with the "specified task" here being the establishment of correct SA. From this perspective, the computational model of SA assessment is essentially evaluating the reliability of SA. This study formally introduces the definition of SA reliability: The operator's ability to perceive environmental elements in time and space, understand their significance, and predict future states within a specified time window. Therefore, in this study, we refer to the computational model for SA assessment as the SA reliability evaluation model.

In summary, while existing studies have made significant progress in SA reliability evaluation, several critical challenges remain, such as the general neglect of the impact of fatigue and mood on SA, as well as the lack of objectivity in the models. To address the above issues, this paper proposes a new SA reliability model based on Bayesian Neural Networks (BNNs) and BN. BNNs are stochastic neural networks trained using Bayesian methods. Wang and Yeung [41] refer to them as a combination of deep learning for perception and traditional Bayesian models for reasoning. BNNs are considered a promising approach for handling uncertainty in deep learning [42]. BNNs can output the probability distribution of prediction uncertainty for samples, allowing them to be directly used as prior probabilities in BN models. The proposed BN model shows full consideration of the fatigue and mood in the formation of operator SA and constructs a causal conceptual model for SA reliability evaluation. Specifically, it accurately establishes the prior probability distribution of BN based a BNN using EEG data. In addition, the conditional probability table (CPT) for the BN is determined based on experimental statistics, allowing for the construction of a complete BN for SA reliability evaluation. Finally, the model and the method are validated by the SA experiment based on civil aviation. The main contributions of this work are as follows:

(1) A Bayesian Neural Network-Bayesian Network (BNN-BN) framework for SA reliability evaluation was developed, integrating physiological data with experimental data to ensure the objectivity of the assessment.
(2) Based on EEG data, a neural network is used to estimate the uncertainty of operator fatigue and mood, which is then applied as prior probabilities in the BN.
(3) A SA reliability evaluation model was developed based on a BN, accounting for the uncertainties in fatigue and mood. This approach enhances the reliability of the resulting probability data, improving both data quality and analysis accuracy.
(4) Considering the impact of fatigue and mood on SA reliability, a SA experiment is designed based on the OPEN-MATB platform, and a SAGAT question bank is proposed.

The rest of the paper is organized as follows: Section 2 introduces a new SA reliability evaluation model based on BNN-BN. Section 3 presents the SA experiment based on civil aviation. Section 4 gives the results of this study. And Section 5 discusses the results together with the proposed methodology. Section 6 summarizes the paper.

2. A new SA reliability evaluation model based on BNN-BN

2.1. Framework of the methodology

This paper first uses a BNN to model the uncertainty distribution of fatigue and mood states reflected in EEG data samples (expressed in discrete probabilities). Based on this, a BN model for SA reliability evaluation is constructed, with fatigue and mood as the root nodes to comprehensively examine their impact on the SA reliability. The probability distributions output by the BNN serve as prior probabilities for the BN model, and the CPT for the BN is obtained based on experimental statistics. Finally, the impact of fatigue and mood on SA reliability is investigated through a SA experiment set in the context of civil aviation, with the methodological framework illustrated in Fig. 1.

2.2. EEG data preprocessing and feature extraction

2.2.1. EEG data preprocessing

Raw EEG data is affected by various sources of noise, including but not limited to muscle activity (EMG), heartbeat (ECG), eye movements, and environmental electromagnetic interference. Preprocessing steps can be used to remove these non-brain sources of noise. The preprocessing of EEG data typically includes three steps: selecting reference electrodes, filtering, and Independent Component Analysis (ICA).

(1) The first step involves selecting the reference electrodes. Commonly used reference electrodes include the mastoid reference, earlobe reference, average reference, and common average reference [43].
(2) Next, filtering is applied to the EEG signals to isolate signals relevant to the research objectives. Typically, a low-pass filter is set at 0.1 Hz, while a high-pass filter is set at 40 Hz. When using a band-pass filter, the frequency range is set from 0.1 Hz to 40 Hz.
(3) Finally, data segmentation and ICA are performed. Data segmentation is used to extract features from time segments of interest, with 6 s chosen for this study. The goal of ICA is to eliminate artifacts from EMG, ECG, and eye movements that interfere with the EEG signals.

2.2.2. EEG data feature extraction

EEG data feature extraction is a crucial step for deriving meaningful features from preprocessed EEG data that represent brain activity patterns and can be used as input for subsequent deep learning. Commonly used EEG signal features include time-domain features, frequency-domain features, and other relevant features.

The most direct and important type of EEG data features are time-domain features. Common time-domain features include the signal's maximum and minimum values, peak value, mean, root mean square, square root amplitude, average amplitude, variance, mean square error, root mean square error, skewness, and kurtosis.

Frequency-domain features of EEG data, also known as brain rhythms. The brain rhythms of interest to researchers can be categorized by frequency from low to high as δ, θ, α, and β waves. By analyzing these rhythms in the frequency domain, the absolute power of each rhythm can be extracted as a feature. Additionally, the ratio of power between different rhythms can reflect various physiological states.

In addition to time-domain and frequency-domain features, other features are also used in EEG-related research, such as renyi entropy, shannon entropy, minimum entropy, collision entropy, hartley entropy, log energy entropy, tsallis entropy, and approximate entropy [44]. Furthermore, other nonlinear features, such as the largest Lyapunov exponent and complexity measures like co-complexity, are also applied in EEG analysis.

In this study, 21 time-domain features, 13 frequency-domain features, and 9 other features were selected, totaling 43 features for the analysis of fatigue and mood. Detailed information on the features can be found in Appendix A.

2.3. Probability distribution of fatigue and mood based on BNN

2.3.1. Bayesian neural network

This study aims to use a BN model to account for the impact of fatigue and mood on the SA reliability. However, obtaining the prior probabilities for fatigue and mood presents a challenge. Traditional machine learning and deep learning methods typically output labels indicating an operator's state, which neglects the presence of uncertainty. BNN are considered promising approaches for addressing uncertainty in deep learning, as they combine deep learning with traditional Bayesian inference to output probability distributions for

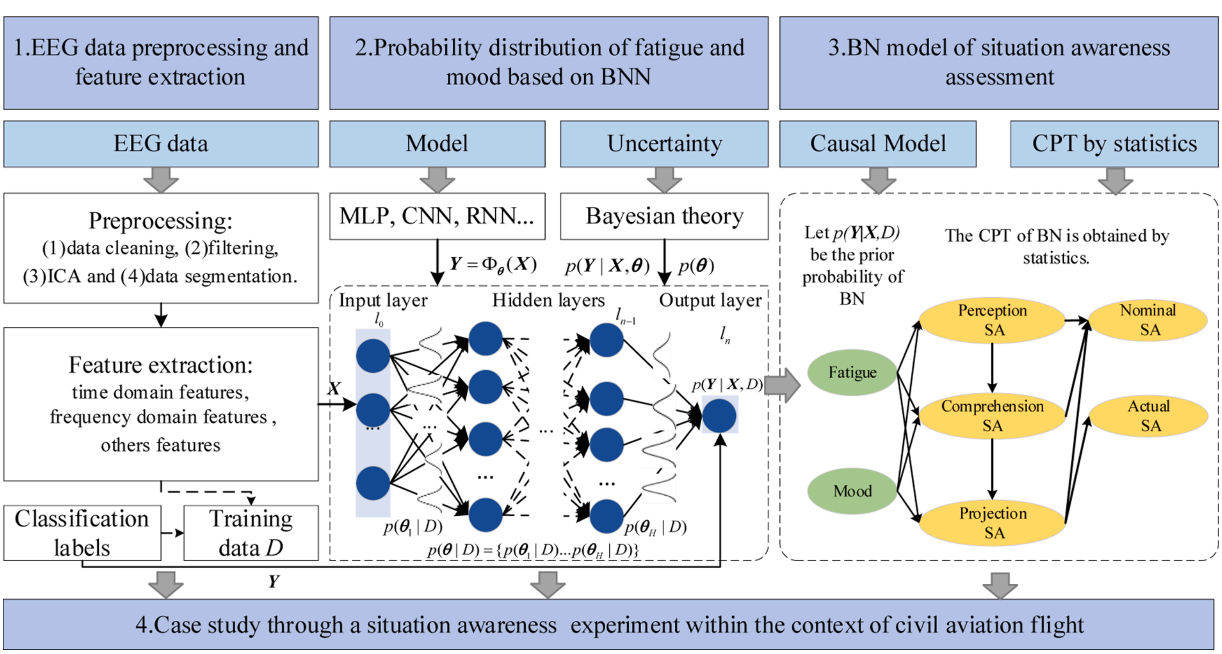

Fig. 1. The framework of the methodology.

prediction samples. Therefore, this study uses a BNN based on EEG data to obtain the probability distributions of fatigue and mood for each sample, which are then used as prior probabilities in the BN model.

BNNs combine deep learning with traditional Bayesian inference, so the first step in designing a BNN is to choose a deep neural network architecture, represented as $\boldsymbol{Y} = \Phi_{\boldsymbol{\theta}}(\boldsymbol{X})$, also known as the function model, where $\boldsymbol{Y}$ represents the label, $\boldsymbol{X}$ represents the features, and $\boldsymbol{\theta}$ represents the parameters. A traditional artificial neural network typically consists of an input layer ($\boldsymbol{l}_0$), a series of hidden layers ($\boldsymbol{l}_i, i = 0, 1\ldots n,$), and an output layer ($\boldsymbol{l}_n$). In the simplest feedforward neural network, each layer $\boldsymbol{l}$ represents a linear transformation followed by a nonlinear activation function s:

$$\boldsymbol{l}_0 = \boldsymbol{X} \tag{1}$$

$$\boldsymbol{l}_i = s_i(\boldsymbol{w}_i\boldsymbol{l}_{i-1} + \boldsymbol{b}_i)\ \forall i \in [1, n] \tag{2}$$

$$\boldsymbol{Y} = \boldsymbol{l}_n \tag{3}$$

where $\boldsymbol{\theta} = (\boldsymbol{w}, \boldsymbol{b})$ represents the model's parameters, s_i represents the activation function of layer $\boldsymbol{l}_i$.Deep learning is essentially the process of regressing these parameters from the data $D = \{(X_1, Y_1), \ldots, (X_n, Y_n)\}$. The standard approach is to use the backpropagation algorithm to obtain point estimates for each parameter. While this simplifies the deployment of software packages, it often lacks interpretability and can lead to overconfidence in predictions for unseen data.

A BNN can be defined as any artificial neural network trained using Bayesian inference. It simulates multiple possible model parameters $\boldsymbol{\theta}$ and their associated distributions $p(\boldsymbol{\theta})$ by assigning random weights to the network, providing a better understanding of the uncertainty related to the underlying process. If different models agree, uncertainty is low; otherwise, uncertainty is high [45]. This process can be represented as follows:

$$\boldsymbol{\theta} = p(\boldsymbol{\theta}) \tag{4}$$

$$\boldsymbol{Y} = \Phi_{\boldsymbol{\theta}}(\boldsymbol{X}) + \boldsymbol{\varepsilon} \tag{5}$$

where $\boldsymbol{\varepsilon}$ represents random noise, used to account for the fact that the function $\Phi_{\boldsymbol{\theta}}$ is only an approximation.

In addition to selecting a function model, a stochastic model must also be chosen to determine the prior distribution of the model parameters $\boldsymbol{\theta}$ and the prior confidence in the model's predictive ability $p(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\theta})$. Let the training dataset be D, the training labels be D_Y, and the training features be D_X. Using Bayes' theorem, the Bayesian posterior of the model parameters can be expressed as follows:

$$p(\boldsymbol{\theta}|D) = \frac{p(D_Y|D_X, \boldsymbol{\theta})p(\boldsymbol{\theta})}{\int_{\theta} p(D_Y|D_X, \boldsymbol{\theta}')p(\boldsymbol{\theta}')d\boldsymbol{\theta}'} \propto p(D_Y|D_X, \boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{6}$$

accurately estimating $p(\boldsymbol{\theta}|D)$ is challenging, as it is a high-dimensional, highly non-convex probability distribution, particularly when calculating the evidence $\int_{\theta} p(D_Y|D_X, \boldsymbol{\theta}')p(\boldsymbol{\theta}')d\boldsymbol{\theta}'$. Therefore, two widely adopted methods have been introduced: (1) Variational Inference [46] and (2) Markov Chain Monte Carlo [47]. This study does not provide a detailed explanation of the parameter inference process, and interested readers are encouraged to refer to the literature [46,47].

The prediction output of a BNN is $p(\boldsymbol{Y}|\boldsymbol{X}, D)$, which quantifies the model's uncertainty in its predictions and is referred to as the marginal distribution. Given $p(\boldsymbol{\theta}|D)$ and $p(\boldsymbol{Y}|\boldsymbol{X}, D)$ it can be computed as follows:

$$p(\boldsymbol{Y}|\boldsymbol{X}, D) = \int_{\theta} p(\boldsymbol{Y}|\boldsymbol{X}, \boldsymbol{\theta}')p(\boldsymbol{\theta}'|D)d\boldsymbol{\theta}' \tag{7}$$

in practical applications, indirect sampling is performed using Eq. (5), and the final prediction results are summarized using statistics computed via the Monte Carlo method. A large number of weight sets $\boldsymbol{\theta}_i \in \Theta$ are sampled from the posterior to compute the corresponding outputs $\boldsymbol{Y}_i \in Y$. For classification tasks, the average model prediction provides the relative probability for each classification, which can be considered a measure of output uncertainty:

$$\widehat{\boldsymbol{p}} = \frac{1}{\Theta}\sum_{\theta_i \in \Theta} \Phi_{\theta_i}(\boldsymbol{X}) \tag{8}$$

The final prediction result is the most likely class:

$$\widehat{Y} = \mathrm{arg}\max_{i} p_i \in \widehat{\boldsymbol{p}} \tag{9}$$

2.3.2. Parameters used in the fatigue and mood BNN model

Based on the data used in this study, a multilayer perceptron (MLP) was chosen as the architecture for the BNN, consisting of an input layer, a hidden layer, and an output layer. The ReLU function was used as the activation function, allowing the model to handle nonlinear relationships. In the BNN model, the input and hidden layers of the MLP are Bayesian linear layers, where the weights are estimated as probability distributions. For the basic architecture, the standard approach is to use a normal prior with a mean of 0 and a diagonal covariance matrix $\sigma\boldsymbol{I}$ on the network's weights [48].

$$p(\boldsymbol{\theta}) = \mathscr{N}(0, \sigma\boldsymbol{I}) \tag{10}$$

it is worth noting that while a normal prior is preferred in practice due to its mathematical properties, there is no theoretical evidence suggesting it is superior to other formulations. The optimizer used is the Adam optimizer, with a learning rate set to 0.01.

To mitigate overfitting, the loss function comprises two components: the cross-entropy loss, which measures classification performance, and the Kullback-Leibler (KL) divergence loss, which quantifies the discrepancy between the model's weight posterior distribution and the prior distribution.

$$Loss = \underbrace{-\frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{C} Y_j^{(i)}\log\left(p_j^{(i)}\right)}_{L(Y,p)} + \beta\underbrace{\sum_{X_i \in X}\mathrm{Pri}(X_i)\log\left(\frac{\mathrm{Pri}(X_i)}{Pos(X_i)}\right)}_{D_{\mathrm{KL}}(\mathrm{Pri}\|Pos)} \tag{11}$$

where $L(Y,p)$ denotes the cross-entropy loss function, where N is the number of samples, C is the number of classes, $Y_j^{(i)}$ represents the probability that sample i belongs to class j. If sample i actually belongs to class j, then $Y_j^{(i)} = 1$; otherwise, $Y_j^{(i)} = 0$. The term $p_j^{(i)}$ represents the model's predicted probability that sample i belongs to class j. $D_{KL}(\mathrm{Pri} \parallel Pos)$ refers to the KL divergence, which quantifies the difference between the model's weight posterior distribution $Pos(X_i)$ and the prior distribution $\mathrm{Pri}(X_i)$. β is a weight coefficient used to balance the relative importance of the cross-entropy and KL divergence losses, with a default value set to 0.01.

The proposed BNN model is implemented using the PyTorch library and the Torchbnn library in Python.

2.4. BN model for SA reliability evaluation

2.4.1. Bayesian network

BN is a kind of directed acyclic graph, in which nodes represent random variables of influencing factors or events and edges between nodes represent their causality (from "cause" to "result") [49]. If the random variables in BN are represented by x, then the probability of each node is expressed by its parent nodes $PA(x)$, as shown in Eq. (12). The probability of each node is calculated based on the joint probability distribution of BNs.

$$P(x_1, \cdots, x_i) = \prod_{i=1}^{n} P(x_i|PA(x_i)) \tag{12}$$

BN can combine multiple information sources (e.g., physiological data and experimental data) to model the influence of uncertainties on cognitive process with a strong basis in cognitive theory [50,51]. The construction of a BN generally consists of two critical steps: (1) defining the BN's logical structure, and (2) estimating the parameters.

2.4.2. Logical structure of the proposed BN model

The basis of the proposed BN-based SA reliability model is the causal conceptual model of SA, which demonstrates the hierarchical relationships and logical structure underlying SA. As Fig. 2 illustrates, this model considers factors in two dimensions:

a) Levels of SA: SA can be divided into three key levels, namely perception, comprehension, and projection. Each of these levels is influenced differently, so the SA at different levels is modeled separately to increase the accuracy of the model.
b) Performance shaping factors (PSFs): This study primarily investigates the influence of physiological factors on the SA reliability, with a specific focus on fatigue and mood, both of which are widely acknowledged to significantly affect human cognitive process. Other factors, such as environmental conditions, equipment, and task-related variables, are beyond the scope of this research.

In existing SA reliability models, the three levels of SA are closely related. In other words, if one cannot perceive the surrounding environment correctly, he cannot have a correct understanding of the current scenario, and cannot predict the future state correctly. However, in practice, it is still possible to form a correct understanding and prediction even with incorrect perception, and incorrect understanding does not necessarily lead to incorrect prediction. Existing SA disruption measurement methods do not always lead to the accurate judgment of SA, so there is a great possibility for such discordance to occur. It would be problematic to disregard the possibility, and treating the discordance entirely as anomalous state data is a deviation from the actual situation. To solve this problem, this paper proposes the concepts of nominal and actual SA. The actual SA is directly determined by the projection SA, while the nominal SA is jointly determined by the SA of the perception, comprehension, and projection level. The relationships between these concepts are shown in Fig. 3.

Based on the PSFs (fatigue and mood) affecting SA and the interactions (causality) between the factors, a causal conceptual BN model for SA reliability evaluation is developed, as shown in Fig. 4.

2.4.3. The parameter estimation of the proposed BN

Conducting SA reliability evaluation using a BN requires the precise determination of its parameters. Typically, these parameters consist of the prior probabilities for the root nodes and the CPTs for the remaining nodes. In the proposed BN model, parameter estimation involves the prior probability distributions for the 'fatigue' and 'mood' nodes, alongside the CPTs for the other nodes. One of the key advantages of BN is their capacity for uncertainty reasoning. However, constrained by data scarcity, many BN models rely on expert knowledge to define prior probability distributions and CPTs, introducing cognitive uncertainty and reducing both the reliability and applicability of the model in real-world scenarios. To address this limitation, the parameter estimation in this study is derived from objective EEG data and experimental results, aiming to construct a model that more accurately represents real-world conditions. The SAGAT technique measures the success rate of SA establishment by pausing the task at predetermined time points, during which participants are asked to quickly answer questions about their current perception of the situation. Their responses are then compared with the correct answers. This success rate is represented by the Eq. (13):

$$P(sa = R) = \frac{1}{N} \cdot \sum_{j=1}^{N} \sigma(j) \tag{13}$$

$$whereby : \sigma(j) = \begin{cases} 1 \ if \ the \ jth \ item \ is \ correct \\ \quad 0 \ otherwise \end{cases}$$

where $sa \in \{SA_p, SA_c, SA_F, ASA, NSA\}$, $P(sa = R)$ represents the probability of correct for nodes related to SA. When using the SAGAT technique, time is frozen to measure the operator's success rate in establishing SA under the current task conditions. This process incorporates three elements from the concept of SA reliability: specified time, specified conditions, and the establishment of correct SA. Therefore, our model can be regarded as an assessment of the reliability of each level of SA, considering factors such as fatigue and mood influences.

Initially, the outputs $p(\boldsymbol{Y}_F|\boldsymbol{X}, D)$ and $p(\boldsymbol{Y}_M|\boldsymbol{X}, D)$ from the BNN serve as the prior probabilities for the 'fatigue' and 'mood' nodes. Subsequently, multiple SA experiments were conducted on a multi-attribute task platform. These experiments involved obtaining fatigue and mood labels from participants using standardized scales. The SAGAT was then employed to evaluate participants' SA reliability across different levels.

Finally, the CPTs of each node were determined based on statistical analysis of experimental data. For instance, the conditional probability $P(SA_C = R|SA_P = R, F = N, M = P)$ was computed using the following equation:

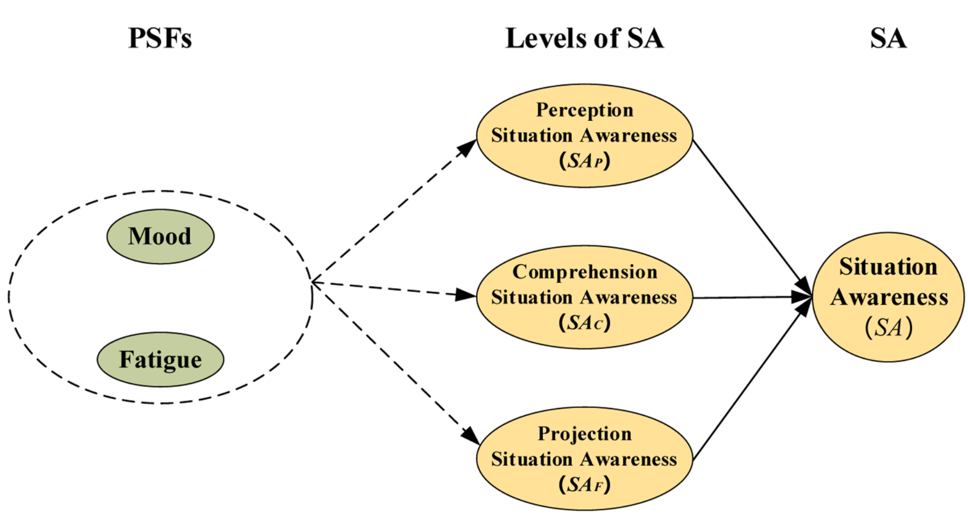

Fig. 2. Logical structure of SA reliability evaluation.

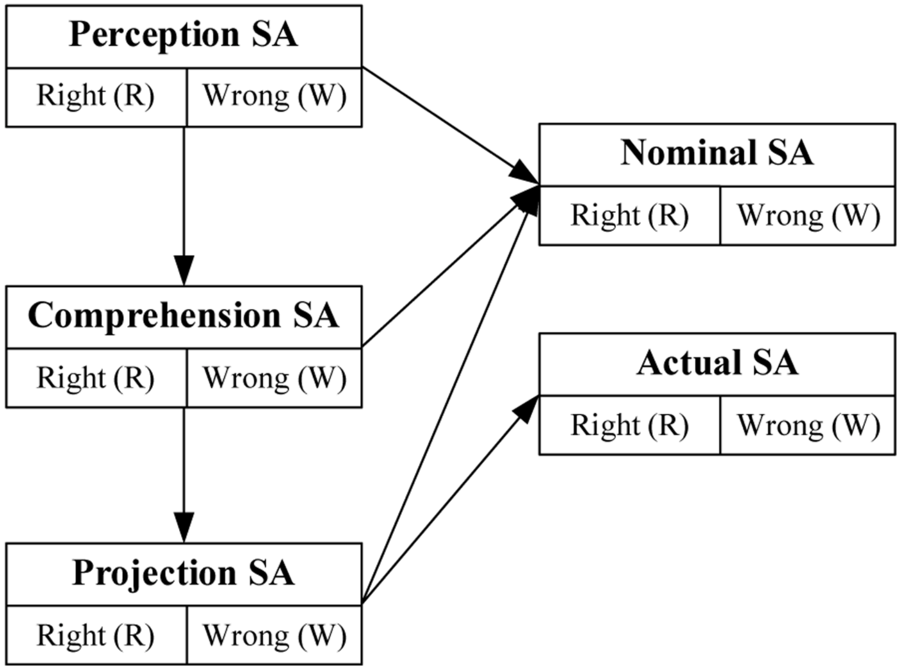

Fig. 3. Nominal SA and Actual SA.

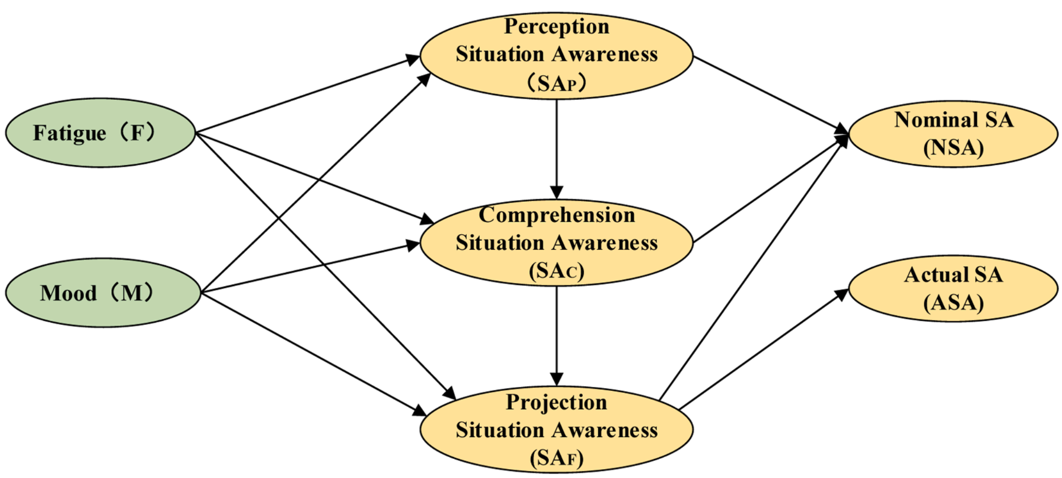

Fig. 4. Causal conceptual model for SA reliability evaluation of BN.

$$P(SA_C = R|SA_P = R, F = N, M = P) = \frac{Sample(SA_C = R, SA_P = R, F = N, M = P)}{Sample(SA_P = R, F = N, M = P)} \tag{14}$$

where $Sample(SA_C = R, SA_P = R, F = N, M = P)$ denotes the number of occurrences of the event $(SA_C = Right) \cap (SA_P = Right) \cap (F = No) \cap (M = Positive)$; and $Sample(SA_P = R, F = N, M = P)$ represents the number of occurrences of the event $(SA_P = Right) \cap (F = No) \cap (M = Positive)$, which can be obtained from experimental statistical data.

In addition to the nodes discussed above, the CPTs of two nodes also need to be determined: "Nominal SA" and "Actual SA". These two nodes have strong logical relationships, so they can be determined by directly choosing the appropriate logic gates.

"Nominal SA" should be strictly defined as deterministic nodes corresponding only to {0,1}. "Nominal SA" can be considered correct only when the "Perception SA", " Comprehension SA", and " Projection SA" are all correct. Therefore, the "and" logic is chosen to describe the "Nominal SA" node, and its Boolean logic relationship of the node is shown in Eq. (15).

$$B(NSA) = B(SA_P) \times B(SA_C) \times B(SA_F) \tag{15}$$

where $B(NSA)$=1 denotes Nominal SA is right and $B(NSA)$=0 denotes Nominal SA is wrong.

The actual SA is directly determined by the projection SA.

The proposed BN model is implemented using GeNIe 4.0 software.

3. Experient

3.1. Participants

18 students from Beihang University participated in the experiment, including 9 males and 9 females, with an average age of 22 ± 2 years. All participants had normal or corrected-to-normal vision, were right-handed, and were in good physical health. Informed consent was obtained from all participants prior to the commencement of the study.

3.2. Experimental apparatus and task

To measure the participants' EEG signals, the experiment employed the Bitbrain wireless portable EEG measurement system equipped with water-based electrodes. This system comprises 32 channels arranged in accordance with the standard 10–20 system and operates with a sampling frequency of 256 Hz. Data transmission is conducted wirelessly via Bluetooth, allowing for real-time monitoring of EEG dynamics.

The experiment utilizes a multi-attribute task to design the SA experiment. The multi-attribute task is implemented using OPEN-MATB, which is derived from the Multi-Attribute Task Battery (MATB) developed by NASA. It provides a set of benchmarking tasks that can be used for a wide range of laboratory research on human performance and workload assessment. It has been used in several fields to simulate various tasks faced by pilots during flight and explore various theories [52,53], and studies have been conducted to demonstrate the persuasiveness of the MATB platform [54]. In this case study, the tasks included the communication task, the tracking task, the system monitoring task, and the resource management task. The task interface is shown in Fig. 5.

3.3. Experimental design

The degradation of SA among pilots is a major factor affecting flight safety. According to statistical data, 35.1 % of minor incidents and 51.6 % of major accidents are attributed to pilot decision-making errors, with a significant proportion linked to inadequate SA or SA errors rather than actual decision-making faults [55]. Fatigue and mood are pivotal factors contributing to the loss of SA or SA errors in pilots. However, there is a scarcity of research investigating the impact of fatigue and mood on the SA reliability in pilots. Accordingly, this study designed a series of experimental tasks based on the typical phases of civil aviation operations—takeoff, level flight, and landing—using the OPEN-MATB platform.

The task mapping for the typical flight scenarios is depicted in Fig. 6, illustrating the tasks that pilots are required to perform during the takeoff, level flight, and landing phases. During the takeoff phase, pilots engage in communication with the control tower while simultaneously managing tracking, system monitoring, and fuel resource management tasks. In the level flight phase, pilots can opt for autopilot or manual tracking, yet they must continue to manage fuel resources. For instance, in the event of a fuel system malfunction, pilots may need to allocate additional cognitive resources to coordination. During the landing phase, pilots sustain communication with the control tower while performing manual tracking and system monitoring tasks.

Prior to the formal experiment, participants engaged in a 5-minute pre-experiment to familiarize themselves with the task procedures. Following this, they completed a 12-minute formal experimental session, during which EEG data were continuously recorded. The experimental protocol incorporated two interruptions to facilitate the assessment of participants' SA using the SAGAT.

3.4. Measurement of fatigue, mood and SA

The primary objective of this study is to investigate the impact of fatigue and mood on the SA reliability. To this end, participants' fatigue and mood states are systematically induced, measured, and recorded to serve as labels for the BNN model. Furthermore, during designated interruption intervals, participants' SA reliability are rigorously assessed to evaluate the impact of these variables on SA reliability.

Fatigue was induced through controlled sleep deprivation. Due to the individual differences in susceptibility to sleep deprivation, participants' fatigue levels were assessed using the Stanford Sleepiness Scale (SSS) [56]. Based on the SSS, participants' fatigue was classified into three categories: no fatigue, mild fatigue, and severe fatigue.

Mood states were induced by exposing participants to auditory stimuli over an extended duration. Prior to the experimental procedures, participants' mood states were evaluated using the Profile of Mood States (POMS) [57]. The results from the POMS assessment enabled the classification of mood states into two distinct categories: positive and negative.

SA reliability was assessed using the SAGAT [58]. As discussed before, this technique involves interrupting the task to prompt operators to quickly respond to questions regarding their current understanding of the situation. Their responses are then compared to the actual situation to objectively evaluate their SA. Responses that match the actual

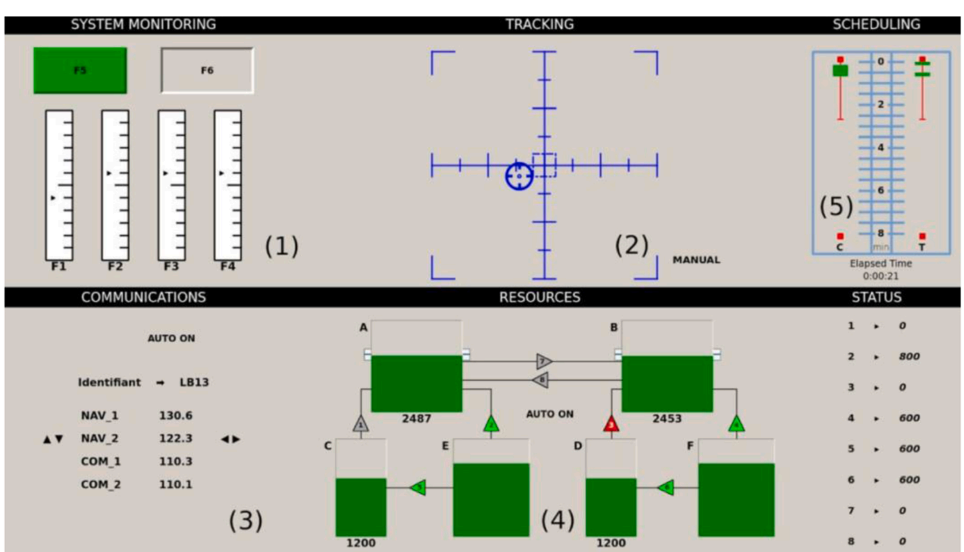

Fig. 5. MATB task interface for the case study.

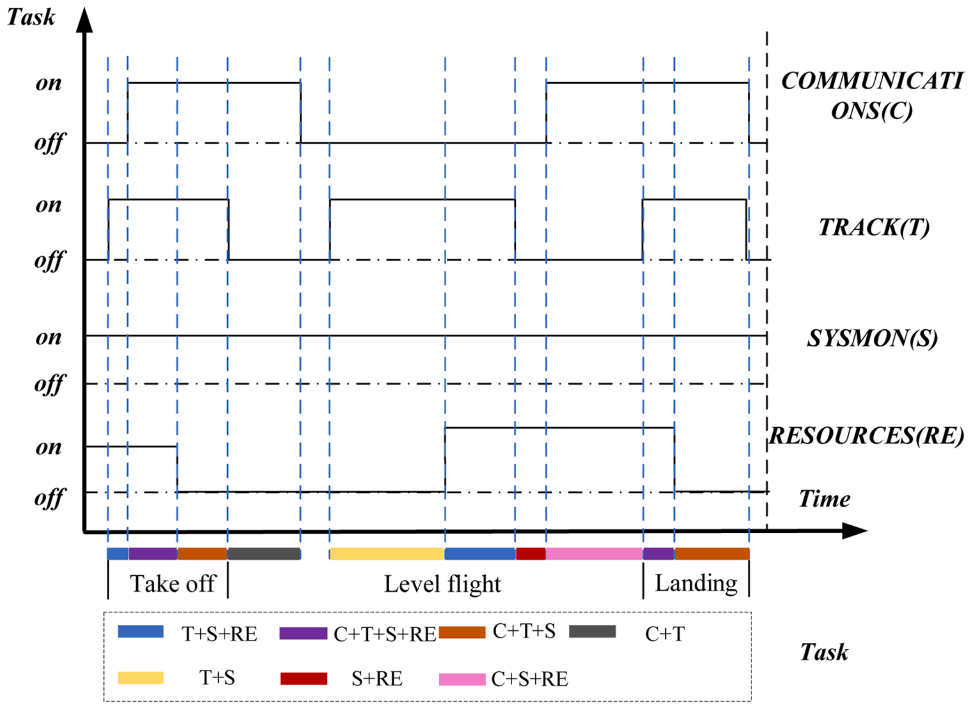

Fig. 6. Task mapping or the typical flight task.

situation are deemed right, while discrepancies are considered wrong.

In this study, EEG data, including those obtained during interruption periods, were segmented into 6-second time slices, yielding a total of 3768 samples. These samples were then used to construct the dataset for the BNN model, which focuses on analyzing fatigue and mood.

4. Results

Based on the experimental data collected, we developed and trained two BNNs to assess the participants' fatigue and mood states, respectively. The CPTs for the nodes in the proposed BN were subsequently computed. The output from the BNN model was used to establish the prior probabilities for the fatigue and mood nodes within the BN framework. Following this, Bayesian inference and sensitivity analyses were performed to examine the influence of fatigue and mood on SA reliability.

4.1. BNN models of fatigue and mood

The BNN model proposed in this study was trained on EEG data from participants, with an 80 % training and 20 % testing data split. Model performance was assessed by calculating accuracy, recall, precision, and F1-score for each category, as detailed in Table 1 and Table 2. The results demonstrate that the BNN models for fatigue and mood achieved high accuracy rates of 95 % and 98 %, respectively. Furthermore, the other performance metrics were also commendable, with all values surpassing 0.9. These outcomes affirm the robustness of the BNN model's uncertainty outputs, making them highly reliable for use as prior probabilities in subsequent BN model.

To quantify uncertainty, multiple forward propagations were conducted using the trained BNN. Specifically, for each test sample, 1000 forward propagations were performed, and the results were averaged to derive the final uncertainty. We selected five samples from the test sets for fatigue and mood to illustrate the model's uncertainty. As presented in Fig. 7 and Fig. 8, the BNN effectively assessed the uncertainty in predicting fatigue and mood from EEG data. For example, certain samples demonstrated low uncertainty: sample 3 was classified with 100 % certainty as experiencing mild fatigue. Conversely, other samples exhibited higher uncertainty, for instance, sample 4 had a 58.19 % probability of being categorized as negative.

Table 1
The results of the fatigue BNN model

Category	Accuracy	Precision	Recall	F1-score
No	0.95	0.94	0.93	0.93
Mild		0.90	0.93	0.92
Severe		0.98	0.96	0.97

Table 2
The results of the mood BNN model

Category	accuracy	Precision	Recall	F1-score
Positive	0.98	0.99	0.98	0.99
Negative		0.98	1.00	0.99

4.2. The impact of fatigue and mood on the SA reliability

Using Eq. (13), we derived the CPTs for each node from the experimental statistics, as shown in Tables 3–6.

The CPTs for "Nominal SA" and "Actual SA" can be derived from the logical gates, as detailed in Table 6 and Table 7.

Based on the CPTs above and the prior probabilities for the fatigue and mood nodes derived from the BNN model outputs, we subsequently conducted posterior probability inference using the BN. This approach allowed us to compute the SA reliability across different levels for the EEG data of sample 1. The results are presented in Fig. 9.

Furthermore, we conducted a comparative analysis of BN posterior inference results with and without incorporating uncertainties associated with the fatigue and mood nodes. In the comparison, we utilized the classification results from a standard MLP, wherein probabilities for specific fatigue and mood states were assigned as 100 %, resulting in what is termed the MLP-BN model. As detailed in Fig. 10, the proposed BNN-BN model, which accounts for uncertainties in fatigue and mood, offers a more nuanced assessment by mitigating potential

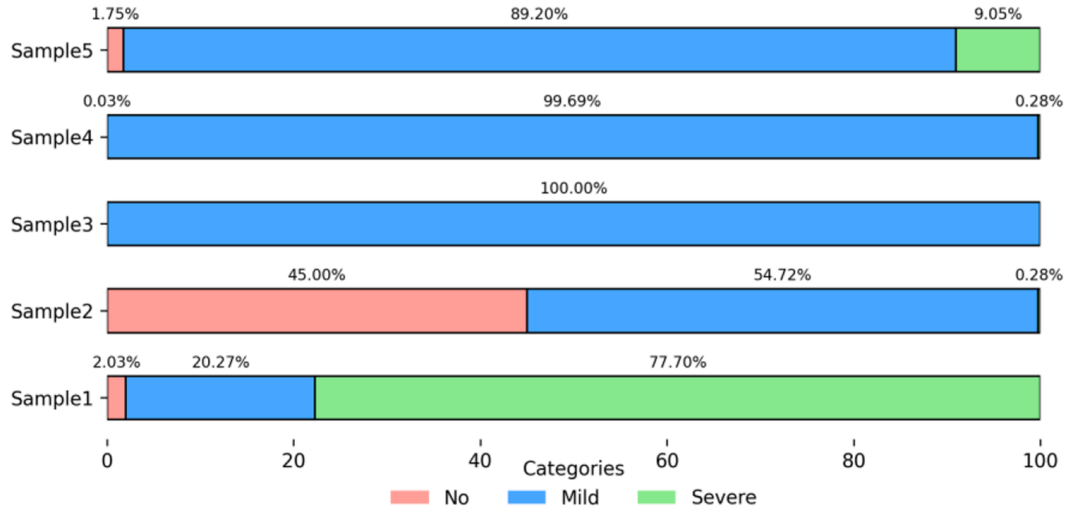

Fig. 7. Uncertainty of fatigue Samples on the test Set (Part).

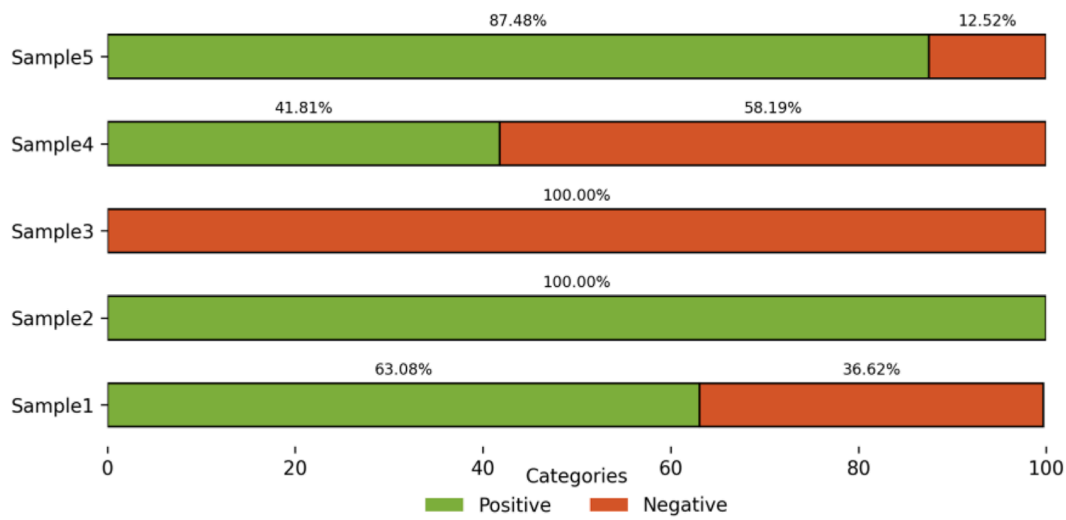

Fig. 8. Uncertainty of mood Samples on the test Set (Part).

Table 3
The CPT of node Perception SA

Fatigue	Mood	Perception SA	
		Right	Wrong
No	Positive	0.831168831	0.168831169
	Negative	0.761904762	0.238095238
Mild	Positive	0.732142857	0.267857143
	Negative	0.674603175	0.325396825
Severe	Positive	0.642857143	0.357142857
	Negative	0.592592593	0.407407407

overestimation or underestimation of the physiological states' impact on SA during BN posterior inference. For example, in the posterior inference for sample 4, $P(SA_P = R) = 67.46$ %, whereas the BNN-BN model estimates a probability of 69.85 %. This disparity suggests that the MLP-BN model may underestimate the influence of physiological states on SA reliability.

4.3. Sensitivity analysis of the established SA model

BN sensitivity analysis involves assessing how minor variations in input parameters (such as prior probabilities and CPTs) influence the output parameters (posterior probabilities). Parameters exhibiting high sensitivity significantly impact the inference outcomes. The GeNIe software utilizes the sensitivity analysis algorithm developed by Kjaerulff and van der Gaag [59], which is designed to perform fundamental sensitivity assessments within BN. Broadly speaking, for a given set of target nodes, this algorithm efficiently computes the full set of

Table 4
The CPT of node Comprehension SA

Perception SA	Fatigue	Mood	Comprehension SA	
			Right	Wrong
Right(R)	No	Positive	0.918181818	0.0818182
		Negative	0.883333333	0.1166667
	Mild	Positive	0.9125	0.0875
		Negative	0.85	0.15
	Severe	Positive	0.8625	0.1375
		Negative	0.811111111	0.1888889
Wrong(W)	No	Positive	0.159090909	0.8409091
		Negative	0.104166667	0.8958333
	Mild	Positive	0.09375	0.90625
		Negative	0.125	0.875
	Severe	Positive	0.09375	0.90625
		Negative	0.111111111	0.8888889

Table 5
The CPT of node Projection SA

Comprehension SA	Fatigue	Mood	Projection SA	
			Right	Wrong
Right(R)	No	Positive	0.939393939	0.0606061
		Negative	0.925925926	0.0740741
	Mild	Positive	0.944444444	0.0555556
		Negative	0.87037037	0.1296296
	Severe	Positive	0.875	0.125
		Negative	0.851851852	0.1481481
Wrong(W)	No	Positive	0.151515152	0.8484848
		Negative	0.138888889	0.8611111
	Mild	Positive	0.125	0.875
		Negative	0.222222222	0.7777778
	Severe	Positive	0.166666667	0.8333333
		Negative	0.111111111	0.8888889

Table 6
The CPT of node Nominal SA

Perception SA	Comprehension SA	Projection SA	Nominal SA	
			Right	Wrong
Right	Right	Right	1	0
		Wrong	0	1
	Wrong	Right	0	1
		Wrong	0	1
	Right	Right	0	1
		Wrong	0	1
Wrong	Wrong	Right	0	1
		Wrong	0	1
	Right	Right	0	1
		Wrong	0	1
	Wrong	Right	0	1
		Wrong	0	1

Table 7
The CPT of node Actual SA

Projection SA	Actual SA	
	Right	Wrong
Right	1	0
Wrong	0	1

derivatives of the posterior probability distributions at these nodes with respect to each numerical parameter in the BN. These derivatives indicate the significance of the precision of the network's numerical parameters for computing the posterior probabilities of the targets. A large derivative implies that even minor changes in the parameter can lead to substantial variations in the posterior probabilities. Conversely, a small derivative suggests that significant changes in the parameter will have minimal impact on the posterior probabilities.

Finally, a sensitivity analysis of the developed SA reliability evaluation model was conducted, focusing on the "Actual SA" node as the core element. Fig. 11 presents a tornado diagram for the sensitivity analysis of the NSA node. The diagram illustrates the top ten most sensitive parameters affecting the probability of NSA=Right. For each parameter, the model displays its precise location (the node and its state, which depend on the states of the parent nodes). The bar graph shows the range of changes in the target state when the parameter varies within its bounds (in this case, its current value fluctuates between [0,1]). The color of the bars indicates the direction of change in the target state, with red representing negative change and green positive change.

The results indicate that the parameter $P(SA_F = R|SA_C = R, F = Mild, M = Positive)$ has the greatest influence on $P(ASA = Right)$. Moreover, fatigue has a more significant impact on ASA than mood. In terms of sensitivity, the parameters ranked as $P(F = Severe) > P(F = No) > P(F = Mild) > P(Mood = Positive) > P(Mood = Negative)$.

5. Discussion

Our study demonstrates that integrating BNN-derived uncertainty into BN significantly improves the robustness of SA reliability assessment. Compared to traditional models relying on subjective expert judgments [30], our BNN-BN framework reduces overconfidence in predictions of SA reliability by considering the uncertainty of fatigue and emotional EEG samples (Fig. 10). This advance addresses the critical gap in existing SA models that neglect physiological state uncertainties. This methodology can be discussed in the following aspects.

(1) SA experimental design based on the OPEN-MATB platform.

Our SA experiment, designed around the OPEN-MATB platform, represents a significant advancement in simulating real-world aviation scenarios. By incorporating tasks such as communication, tracking, and resource management, we replicated the cognitive demands faced by pilots during critical flight phases. This design provides a standardized

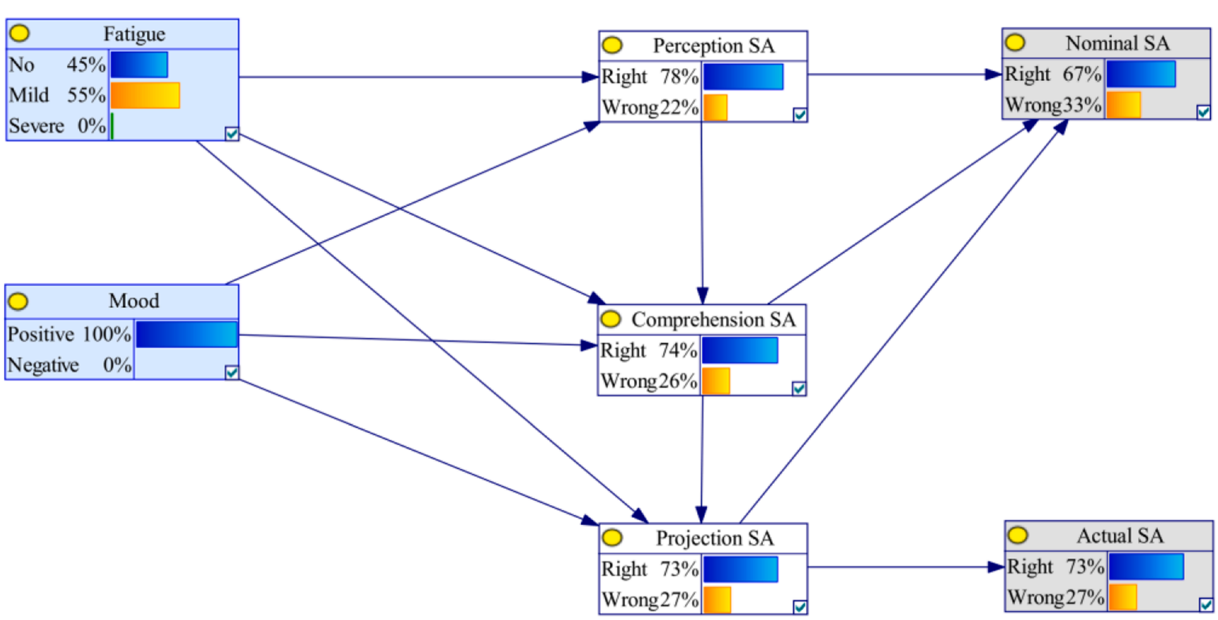

Fig. 9. The Forward Inference Results of Sample 1 with BN.

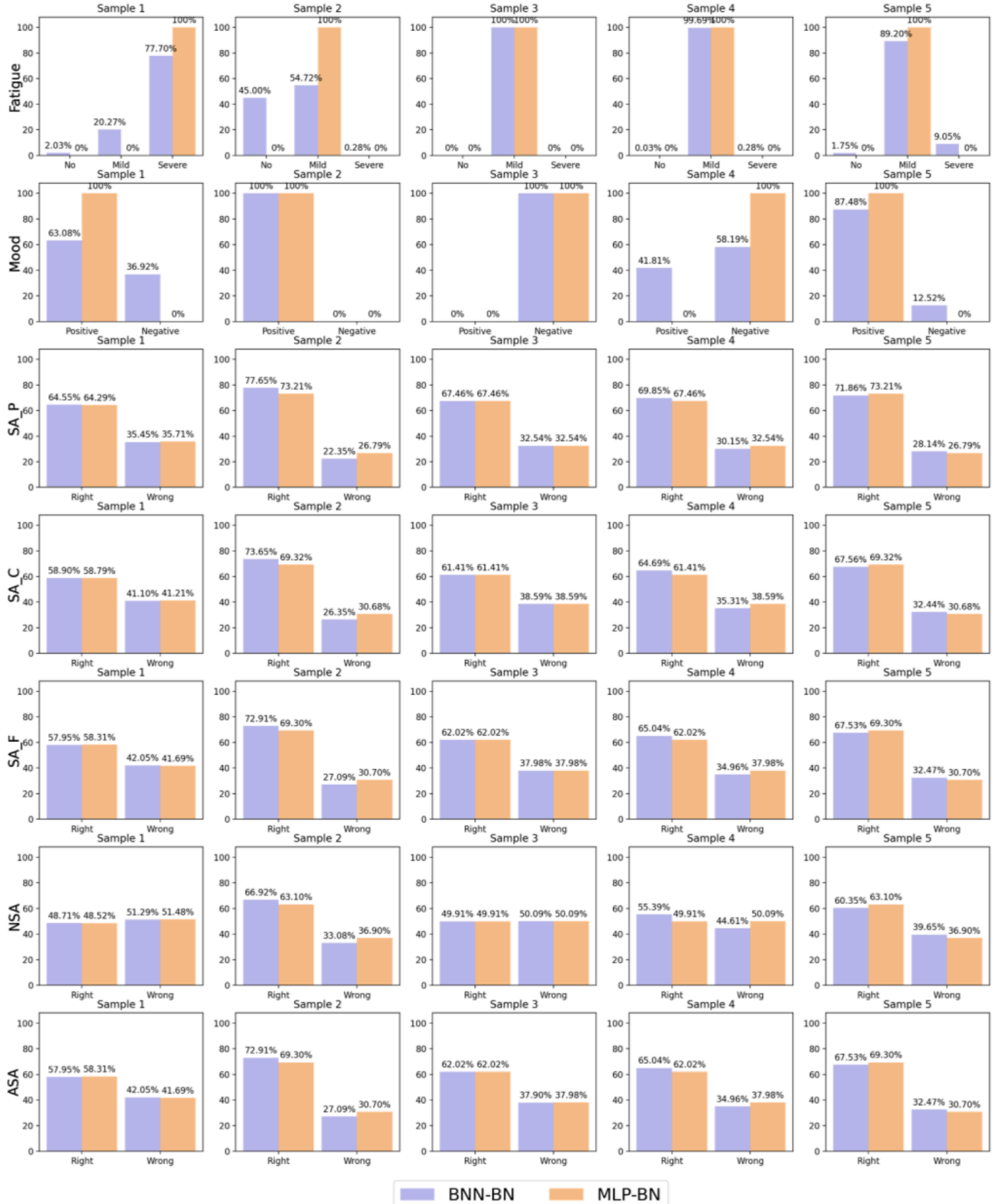

Fig. 10. Comparison of inference results between BNN-BN and NN-BN.

framework for future SA research. For instance, the SAGAT question bank developed in this study (Appendix B) can be adapted to other high-risk domains, such as nuclear power plant operations or autonomous vehicle supervision, where SA reliability is equally critical.

The proposed experimental framework allows researchers to systematically investigate the impact of various factors (e.g., workload, stress) on SA reliability. This capability is particularly valuable for designing targeted interventions, such as adaptive automation systems that adjust task complexity based on real-time operator states.

(2) Utilizing BNNs to assess the uncertainty in operators' physiological states

Traditional deep learning models, while achieving high accuracy in physiological state classification, often fail to account for prediction uncertainty, leading to overconfident and potentially hazardous outcomes [60,61]. In contrast, our BNN-based approach quantifies both aleatoric and epistemic uncertainties, providing a more reliable foundation for SA reliability assessment. For example, the BNN model's uncertainty estimates (Fig. 7 and Fig. 8) reveal that certain EEG samples

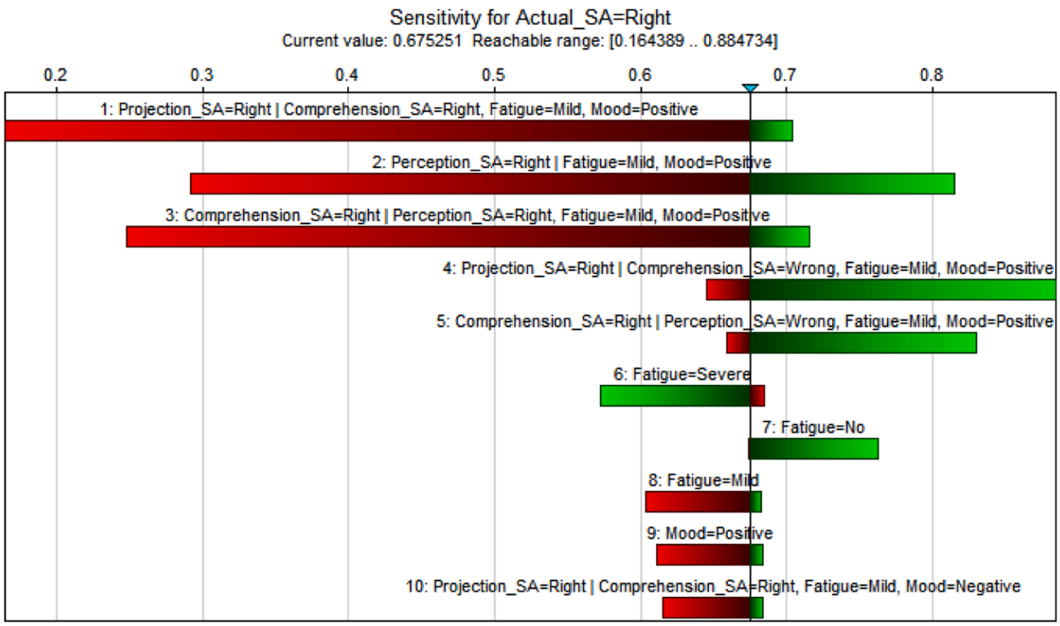

Fig. 11. Tornado plot for ASA nodal sensitivity analysis.

exhibit high ambiguity (e.g., sample 4 with 58.19 % probability of negative mood), which would be overlooked by deterministic models.

By integrating Bayesian inference with deep learning, our BNN framework not only achieves high classification accuracy (95 % for fatigue, 98 % for mood) but also evaluates the uncertainty of fatigue and mood. This dual capability is crucial for safety-critical applications, where overconfidence in predictions can have severe consequences.

(3) SA reliability evaluation considering the uncertainty of physiological states

Our BNN-BN framework bridges the gap between data-driven uncertainty quantification and causal reasoning, offering a novel approach to SA reliability evaluation. Unlike traditional BN models that rely on subjective expert judgments [62,63], our approach leverages objective EEG data and experimental statistics to derive prior probabilities and CPTs. This integration not only enhances model credibility but also provides a scalable solution for real-time SA assessment in dynamic environments.

The sensitivity analysis (Fig. 11) highlights that fatigue has a greater impact on Actual SA than mood. This can be attributed to the fact that fatigue directly impairs key cognitive functions such as attention, memory, decision-making, and reaction time, all of which are critical for maintaining a high SA [64,65]. While mood also influences individual behavior and performance, the mood states in this study were induced through audio stimuli, which may not exert as strong an effect on cognitive functions as more severe mood conditions such as depression or anxiety. In other words, participants might be able to mitigate the influence of mood on SA through conscious effort, whereas overcoming the detrimental effects of fatigue is far more challenging. This result underscores the need for real-time fatigue monitoring systems in aviation. Our model's ability to quantify this impact can inform the design of adaptive interventions, such as workload redistribution or rest scheduling, to mitigate fatigue-induced SA degradation. Furthermore, the framework's modular design allows for the incorporation of additional factors (e.g., environmental noise, task complexity), making it adaptable to diverse operational contexts.

(4) Limitations and Future Directions

Despite its advancements, our study has two key limitations. First, the model does not account for non-physiological factors (e.g., task complexity, environmental noise), which may influence SA reliability. Second, the participant-dependent nature of the BNN training limits generalizability to other operator cohorts (e.g., professional pilots vs. novices).

To address these limitations, future work should: (1) Expand the BN to include additional nodes for task complexity and environmental factors, leveraging multi-modal data (e.g., eye-tracking, heart rate variability); (2) Validate the framework in real-world settings, such as flight simulators or nuclear control rooms, to assess its practical utility.

6. Conclusion

SA is a critical prerequisite for effective decision-making and successful task execution, SA reliability is the evaluation of how well SA is established, and it is also the goal of SA assessment. While numerous descriptive models of SA have been proposed, SA reliability evaluation models often suffer from limited applicability and objectivity, particularly in failing to account for the effects of fatigue and mood. To address these gaps, this paper presents a novel SA reliability evaluation model using the BNN-BN approach. This approach integrates the effects of fatigue and mood on SA reliability, leveraging EEG data to innovatively capture the uncertainty in predictions of fatigue and mood states via BNNs. The uncertainty is then used as prior probabilities in a BN model, which helps to prevent overestimation or underestimation of the results. The SA experiment is designed around a typical flight scenario, taking into account the three levels of SA, with CPTs derived from experimental data. The results demonstrate that the proposed BNN-BN model is capable of directly assessing an operator's SA reliability from EEG data, providing a theoretical foundation for real-time SA reliability evaluation based on physiological data.

CRediT authorship contribution statement

Song Ding: Writing – original draft, Validation, Data curation. **Lunhu Hu:** Validation, Conceptualization. **Xing Pan:** Project administration, Funding acquisition, Conceptualization. **Dujun Zuo:** Writing – review & editing, Methodology. **Liuwang Sun:** Visualization, Software, Methodology.

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. All authors have confirmed that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Acknowledgement

This work is supported by the National Natural Science Foundation of China under Grants No 71171008.

Appendix A

The time-domain features selected from EEG signal for this study are as follows: Tables 8, 9, 10

Table 8
The time-domain features used in this study.

Number	Feature	Number	Feature	Number	Feature
1	Maximum value	8	Sample standard deviation	15	Crest factor
2	Minimum value	9	Standard error	16	Impulse factor
3	Mean value	10	Coefficient of variation	17	Margin factor
4	Arithmetic mean	11	Root mean square (RMS) value	18	Skewness factor
5	Peak value	12	Root amplitude	19	Kurtosis factor
6	Variance (effective estimate/population)	13	Skewness	20	Population standard deviation
7	Variance (unbiased/sample)	14	Kurtosis	21	Form factor

The frequency-domain features selected from EEG signal for this study are as follows:.

Table 9
The frequency-domain features used in this study.

Number	Feature	Number	Feature
22	Absolute energy of δ wave	29	Relative energy of α/(δ+θ+α+β)
23	Absolute energy of θ wave	30	Relative energy of α/(δ+θ+α+β)
24	Absolute energy of α wave	31	Relative energy of α/(δ+θ+α+β)
25	Absolute energy of β wave	32	Relative energy of α/(δ+θ+α+β)
26	Total absolute energy	33	Relative energy of (θ+α)/(β+α)
27	Relative energy of δ/(δ+θ+α+β)	34	Relative energy of θ/β
28	Relative energy of θ/(δ+θ+α+β)		

The other features selected from EEG signal for this study are as follows:.

Table 10
The other features used in this study.

Number	Feature	Number	Feature
35	Information entropy/Shannon entropy	40	First-order differential standard deviation
36	Log energy entropy	41	Second-order differential standard deviation
37	Threshold entropy	42	Mobility
38	Deterministic entropy	43	Complexity
39	Norm entropy		

Appendix B

Table 11

Table 11
The SAGAT question bank for the OPEN-MATB platform.

Number	Question	Task	Level of SA
1	Were the F1-F4 values within the normal range during the aforementioned time period?	Monitoring	Perception
2	What were the colors representing the normal states of F5 and F6 during the aforementioned time period?	Monitoring	Perception
3	Did you notice when F1-F4 moved to the upper or lower boundary during the recent time period?	Monitoring	Perception
4	What action is required when the green color of F5 disappears?	Monitoring	Comprehension
5	What action needs to be taken when the red color of F6 appears?	Monitoring	Comprehension
6	What changes occurred in the status of F1-F4 when you clicked on them?	Monitoring	Projection
7	What changes occurred in the status of F5 when you clicked on it?	Monitoring	Projection
8	What changes occurred in the status of F6 when you clicked on it?	Monitoring	Projection
9	What was the status of the tracking task during the recent period?	Tracking	Perception
10	Did you notice when the tracking task was in automatic mode?	Tracking	Perception

(continued on next page)

Table 11 (*continued*)

Number	Question	Task	Level of SA
11	Did you immediately notice when the tracking task was in manual mode?	Tracking	Perception
12	What actions did you take when the tracking task switched to manual mode?	Tracking	Comprehension
13	What actions are required when the cursor center deviates from the central box?	Tracking	Projection
14	Who was the recipient of the last voice call?	Communication	Perception
15	Were you aware of the significance of the voice call recipient being NASA504?	Communication	Comprehension
16	Were you aware of the significance when the voice call instructed you to switch the radio to the designated channel?	Communication	Projection
17	Did you notice the schedule displayed in the upper-right corner?	Scheduling	Perception
18	How many minutes into the schedule is the next voice task expected to occur?	Scheduling	Perception
19	Does the schedule indicate that the next operation will be received after the first minute?	Scheduling	Perception
20	What do the green indicators in the C and T columns represent?	Scheduling	Comprehension
21	What action should you take when the T column is about to change from red to green?	Scheduling	Projection
22	What are the current fuel levels in main tanks A and B, respectively?	Resource management	Perception
23	What are the current fuel levels in auxiliary tanks C and D, respectively?	Resource management	Perception
24	Which valve is currently indicated in red?	Resource management	Perception
25	What does it signify when the valve status is indicated in red?	Resource management	Comprehension
26	What are the consequences when a particular valve fails?	Resource management	Projection
27	What happens when all the valves are open?	Resource management	Projection
28	What actions do you need to take when the fuel levels in tanks A and B fall below 2500?	Resource management	Comprehension
29	What actions should you take when the fuel level in tank A is below 2500 and the fuel level in tank B is above 2500?	Resource management	Comprehension

Data availability

Data will be made available on request.

References

[1] Capallera M, Angelini L, Meteier Q, Abou Khaled O, Mugellini E. Human-vehicle interaction to support driver's situation awareness in automated vehicles: a systematic review. IEEE Trans Intell Veh 2022;8:2551.

[2] Pan X, Sun L, Ding S, Zuo D. Impact analysis of situation awareness based on a multitasking difficulty quantitative model. Int J Human–Comput Int 2023;1.

[3] Panakkal P, Padgett JE. More eyes on the road: sensing flooded roads by fusing real-time observations from public data sources. Reliab Eng Syst Saf 2024;251:110368.

[4] Ghosh D. Impact of situational awareness attributes for resilience assessment of active distribution networks using hybrid dynamic Bayesian multi criteria decision-making approach. Reliab Eng Syst Saf 2022;228:108772.

[5] You Q, Guo J, Zeng S, Che H. A dynamic Bayesian network based reliability assessment method for short-term multi-round situation awareness considering round dependencies. Reliab Eng Syst Saf 2024;243:109838.

[6] Wang Z, Zeng S, Guo J, Che H. A bayesian network for reliability assessment of man-machine phased-mission system considering the phase dependencies of human cognitive error. Reliab Eng Syst Saf 2021;207:107385.

[7] Sui Z, Wang S. Traffic advisory for ship encounter situation based on linear dynamic system. Reliab Eng Syst Saf 2025;253:110591.

[8] Endsley MR. Toward a theory of situation awareness in dynamic systems. Hum Factors 1995;37:32.

[9] Ignatious HA, El-Sayed H, Khan MA, Mokhtar BM. Analyzing factors influencing situation awareness in autonomous vehicles—A survey. Sensors 2023;23:4075.

[10] Domeh V, Obeng F, Khan F, Bose N, Sanli E. An operational risk awareness tool for small fishing vessels operating in harsh environment. Reliab Eng Syst Saf 2023;234:109139.

[11] Naderpour M, Lu J, Zhang G. An abnormal situation modeling method to assist operators in safety-critical systems. Reliab Eng Syst Saf 2015;133:33.

[12] Endsley MR. From here to autonomy: lessons learned from human–automation research. Hum Factors 2017;59(5).

[13] Bedny G, Meister D. Theory of activity and situation awareness. Int J Cogn Ergon 1999;3:63.

[14] Adams MJ, Tenney YJ, Pew RW. Situation awareness and the cognitive management of complex systems. Hum Factors 1995;37:85.

[15] Rajala M, Ritala R. Situation awareness Framework for Multi-machine environments: an application to Operator assistive collision Warning System. IEEE Trans Intell Transp Syst 2015;17:336.

[16] Kim MC, Seong PH. An analytic model for situation assessment of nuclear power plant operators based on bayesian inference. Reliab Eng Syst Saf 2006;91:270.

[17] Schneider M, Halekotte L, Comes T, Lichte D, Fiedrich F. Emergency response inference mapping (ERIMap): a Bayesian Network-based method for dynamic observation processing. Reliab Eng Syst Saf 2024:110640.

[18] Xin X, Liu K, Loughney S, Wang J, Yang Z. Maritime traffic clustering to capture high-risk multi-ship encounters in complex waters. Reliab Eng Syst Saf 2023;230:108936.

[19] Naderpour M, Lu J, Zhang G. A safety-critical decision support system evaluation using situation awareness and workload measures. Reliab Eng Syst Saf 2016;150:147.

[20] Naderpour M, Lu J, Zhang G. The explosion at institute: modeling and analyzing the situation awareness factor. Accident Anal Prevent 2014;73:209.

[21] Fan S, Yang Z. Accident data-driven human fatigue analysis in maritime transport using machine learning. Reliab Eng Syst Saf 2024;241:109675.

[22] Jeon M, Walker BN, Gable TM. The effects of social interactions with in-vehicle agents on a driver's anger level, driving performance, situation awareness, and perceived workload. Appl Ergon 2015;50:185.

[23] Gaines AR, Morris MB, Gunzelmann G. Fatigue-related aviation mishaps. Aerosp Med Hum Perform 2020;91:440.

[24] Marcus JH, Rosekind MR. Fatigue in transportation: NTSB investigations and safety recommendations. Injury Prevent 2017;23:232.

[25] Eyben F, Wöllmer M, Poitschke T, Schuller B, Blaschke C, Färber B, Nguyen-Thien N. Emotion on the road—Necessity, acceptance, and feasibility of affective computing in the car. Adv Human-Comput Int 2010;2010:263593.

[26] van den Berg NS, Reesink FE, de Haan EH, Kremer HP, Spikman JM, Huitema RB. Emotion recognition and traffic-related risk-taking behavior in patients with neurodegenerative diseases. J Int Neuropsychol Soc 2021;27:136.

[27] Mou L, Zhao Y, Zhou C, Nakisa B, Rastgoo MN, Ma L, Huang T, Yin B, Jain R, Gao W. Driver emotion recognition with a hybrid attentional multimodal fusion framework. IEEE Trans Affect Comput 2023;14:2970.

[28] Zimasa T, Jamson S, Henson B. Are happy drivers safer drivers? Evidence from hazard response times and eye tracking data. Transp Res Part F: Traffic Psychol Behav 2017;46:14.

[29] Ding S, Pan X, Zuo D, Zhang W, Sun L. Uncertainty analysis of accident causality model using Credal Network with IDM method: a case study of hazardous material road transportation accidents. Process Saf Environ Protect 2022;158:461.

[30] Liu Y, Jin X, Luo Z, Dai L, Liu Z, Li P. Methodology for dynamic reliability assessment of team situation awareness of digital nuclear power plants. Progress in Nuclear Energy 2022;144:104086.

[31] Li Q, Ng KK, Yiu CY, Yuan X, So CK, Ho CC. Securing air transportation safety through identifying pilot's risky VFR flying behaviours: an EEG-based neurophysiological modelling using machine learning algorithms. Reliab Eng Syst Saf 2023;238:109449.

[32] Yang L, Li L, Liu Q, Ma Y, Liao J. Influence of physiological, psychological and environmental factors on passenger ship seafarer fatigue in real navigation environment. Saf Sci 2023;168:106293.

[33] French HT, Clarke E, Pomeroy D, Seymour M, Clark CR. Psycho-physiological measures of situation awareness. decision making in complex environments. CRC Press; 2017. p. 291. P.

[34] Borghini G, Astolfi L, Vecchiato G, Mattia D, Babiloni F. Measuring neurophysiological signals in aircraft pilots and car drivers for the assessment of mental workload, fatigue and drowsiness. Neurosci Biobehav Rev 2014;44:58.

[35] Ahn S, Nguyen T, Jang H, Kim JG, Jun SC. Exploring neuro-physiological correlates of drivers' mental fatigue caused by sleep deprivation using simultaneous EEG, ECG, and fNIRS data. Front Hum Neurosci 2016;10:219.

[36] Xu T, Wang H, Lu G, Wan F, Deng M, Qi P, Bezerianos A, Guan C, Sun Y. E-key: an EEG-based biometric authentication and driving fatigue detection system. IEEE Trans Affect Comput 2021;14:864.

[37] Anwer S, Li H, Umer W, Antwi-Afari MF, Mehmood I, Yu Y, Haas C, Wong AYL. Identification and classification of physical fatigue in construction workers using linear and nonlinear heart rate variability measurements. J Constr Eng Manag 2023;149:4023057.

[38] Baraldi P, Mangili F, Zio E. Investigation of uncertainty treatment capability of model-based and data-driven prognostic methods using simulated data. Reliab Eng Syst Saf 2013;112:94.

[39] Rao Q, Frtunikj J. Deep learning for self-driving cars: chances and challenges. In: Proceedings of the 1st international workshop on software engineering for AI in autonomous systems. 35; 2018.

[40] Kwon Y, Won J, Kim BJ, Paik MC. Uncertainty quantification using bayesian neural networks in classification: application to biomedical image segmentation. Comput Stat Data Anal 2020;142:106816.
[41] Wang H, Yeung D. A survey on Bayesian deep learning. ACM Comput Surveys 2020;53:1.
[42] Ghahramani Z. Probabilistic machine learning and artificial intelligence. Nature 2015;521:452.
[43] Yao D. A method to standardize a reference of scalp EEG recordings to a point at infinity. Physiol Meas 2001;22:693.
[44] Ming D, Zhang M, Xi Y, Qi H, Luk KDK. Multiscale entropy analysis of attention ralated eeg based on motor imaginary potential. IEEE; 2009.
[45] MacKay David JC. A practical bayesian framework for backpropagation networks. Neural Comput 1992;4:448.
[46] Blei D.M., Kucukelbir A., Mcauliffe J.D.: Variational inference: a review for statisticians. 2016).
[47] Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 1970;97.
[48] Silvestro D., Andermann T.: Prior choice affects ability of Bayesian neural networks to identify unknowns. 2020).
[49] Pan X, Zuo D, Zhang W, Hu L, Wang H, Jiang J. Research on human error risk evaluation using extended Bayesian networks with hybrid data. Reliab Eng Syst Saf 2021;209:107336.
[50] Li J, Li H, Wang F, Cheng AS, Yang X, Wang H. Proactive analysis of construction equipment operators' hazard perception error based on cognitive modeling and a dynamic bayesian network. Reliab Eng Syst Saf 2021;205:107203.
[51] Mkrtchyan L, Podofillini L, Dang VN. Bayesian belief networks for human reliability analysis: a review of applications and gaps. Reliab Eng Syst Saf 2015; 139:1.
[52] Ferraro J, Clark L, Christy N, Mouloua M. Effects of automation reliability and trust on system monitoring performance in simulated flight tasks. Proc Human Factors Ergonom Soc Annual Meet 2018;62:1232.
[53] Nixon J, Charles R. Understanding the human performance envelope using electrophysiological measures from wearable technology. Cognit Technol Work 2017;19:655.
[54] Wickens CD, Gutzwiller RS, Vieane A, Clegg BA, Sebok A, Janes J. Time sharing between robotics and process control: validating a model of attention switching. Hum Factors 2016;58:322.
[55] Feng C, Wanyan X, Liu S, Zhuang D, Wu X. Dynamic prediction model of situation awareness in flight simulation. In: International Conference on Engineering Psychology and Cognitive Ergonomics; 2018.
[56] E, Hoddes V, Zarcone H, Smythe R, Phillips WC. Quantification of sleepiness: a new approach. Psychophysiology 1973.
[57] Mcnair D., Lorr J., Droppelman L., Systems M.H.: Profile of mood states manual. 1992).
[58] Endsley MR, Radwin RG. A systematic review and meta-analysis of direct objective measures of situation awareness: a comparison of SAGAT and SPAM. Human Fact J Human Factors Ergonom Soc 2021;63:124.
[59] Gaag LCVD. Making Sensitivity Analysis Computationally efficient. Morgan Kaufmann Publishers Inc; 2013.
[60] Mitros J., Mac Namee B.: On the validity of bayesian Neural networks for uncertainty estimation. arXiv e-prints. 2019).
[61] Kristiadi A., Hein M., Hennig P.: Being Bayesian, even just a bit, fixes overconfidence in ReLU networks. 2020).
[62] Information V.F.A., Org H.O., Com S.A.G., Soltanzadeh A., Information V.F.A., Fallah H., Information V.F.A., Org H.O., Aliabadi M.M., Information V.F.A.: Developing a Bayesian network model for improving chemical plant workers' situation awareness. Theoret. Issues Ergonom. Sci.. 2023).
[63] Lee HC, Seong PH. A computational model for evaluating the effects of attention, memory, and mental models on situation assessment of nuclear power plant operators. Reliab Eng Syst Saf 2009;94:1796.
[64] Zhou X, Han J, Qin H, Xue C. Research on multilevel situation awareness changes under the cumulative effect of mental fatigue. Cognit Technol Work 2023.
[65] Sneddon A, Mearns K, Flin R. Stress, fatigue, situation awareness and safety in offshore drilling crews. Saf Sci 2013;56:80.