



OPEN

A multi-source physiological data-driven method for assessing the hazard perception level of operators

Aomeiqian Qi¹✉, Hongrui Liu^{2,3}, Jizu Li¹✉, Benyi Li⁴, Min Yu¹, Ting Lei¹, Yiman He¹, Yong Yan¹ & Yanyu Guo¹

With the continuous advancement of coal mine intelligence, unmanned operations at the working face have become an inevitable trend, and the safety monitoring center plays a central role in ensuring mine safety. Its operation relies not only on system monitoring but also on operators' timely perception and judgment of hazard information. Therefore, the hazard perception level of operators is critical for accident prevention and safety management. In this context, an experiment is conducted to assess the hazard perception levels of safety monitoring center operators. Physiological signals, including electroencephalogram (EEG), electrodermal activity (EDA), and heart rate variability (HRV), are collected during hazard perception tasks. Using these physiological signals in conjunction with machine learning and deep learning methods, a high-accuracy hazard perception assessment model is developed, and its performance is evaluated using reliable metrics. The results show that certain features derived from all three physiological signals exhibit distinct responses corresponding to different hazard perception levels. The LightGBM model achieves the highest performance, with an accuracy of 99.89%. Furthermore, the study demonstrates the feasibility of high-accuracy hazard perception assessment based on the combined use of EEG, EDA, and HRV signals. These findings provide a practical pathway for future coal mine safety management, enhance the capacity for accident prevention in intelligent coal mines, and offer strong support for safe production.

Keywords Coal mine safety, Hazard perception, Multi-source physiological data, Machine learning

In recent years, with the synergistic development of mechanization, automation, and intelligent technologies, the coal mining industry has been rapidly advancing toward intelligent transformation¹. Intelligent coal mines have largely realized unmanned operations at working faces through remote monitoring. As the hub for mine safety information processing and the core of decision-making, the safety monitoring center plays a crucial role in ensuring both safety and efficient operations². Operators in the safety monitoring center are primarily responsible for real-time hazard monitoring, anomaly detection, and emergency response. Their hazard perception level significantly determines their performance, directly impacting the probability of mining accidents and the efficiency of emergency responses. A high hazard perception level enables individuals to identify potential environmental hazards and make appropriate decision responses³. Major accidents such as the 24 November 2012 Xiangshui Coal Mine gas accident in Guizhou and the 6 January 2019 Dapu'an Coal Mine roof collapse accident in Yunnan are both caused by operator errors in monitoring centers. Furthermore, investigations into the Pike River coal mine multiple explosions in New Zealand on 19 November 2010 and the Eynez coal mine fire in Turkey on 13 May 2014 both indicate that insufficient operator hazard perception and delayed response constitute key contributing factors. Therefore, identifying and monitoring operators' hazard perception levels is essential for reducing accident risks and enhancing coal mine safety.

Current research on hazard perception primarily focuses on the driving domain^{4,5} and the construction domain⁶. Beyond using response time, accuracy, and composite scores in hazard perception tests as standard

¹School of Economics and Management, Taiyuan University of Technology, Taiyuan 030024, China. ²College of Physics and Optoelectronics Engineering, Taiyuan University of Technology, Taiyuan 030024, China. ³Key Laboratory of Advanced Transducers and Intelligent Control System, Ministry of Education, and Shanxi Province, Taiyuan University of Technology, Taiyuan 030024, China. ⁴Jinneng Holding Coal Industry Group Company Limited, Datong 037001, China. ✉email: 2023310288@link.tyut.edu.cn; lijizu@tyut.edu.cn

measures of hazard perception level^{3,7,8}, some scholars have attempted to explore physiological signals collected through wearable biosensors. Electroencephalogram (EEG), due to its ability to effectively detect and evaluate brain activity, is commonly applied to monitor various psychological states, including cognitive load⁹, hazard recognition¹⁰, and hazard perception¹¹. Electrodermal activity (EDA), as an indicator of sympathetic nervous system activation, is widely employed to assess arousal levels and cognitive changes¹². Research indicates that when individuals perceive hazardous stimuli, the sympathetic nervous system is rapidly activated, leading to an increase in skin conductance level^{13,14}. Heart rate variability (HRV) is extensively applied in the evaluation of heart rate, autonomic nervous system function, and emotional states. Although research indicates that HRV exhibits lower performance in risk assessment compared with EEG and EDA, combining it with EEG and EDA signals can significantly improve hazard recognition accuracy¹⁵. However, each physiological signal has its own advantages and limitations. Single-dimensional physiological indicators are insufficient for accurately identifying hazard perception levels. Currently, there remains a lack of systematic and comprehensive assessment methods capable of effectively assessing hazard perception levels based on multi-source physiological data.

To achieve more accurate evaluation and recognition, machine learning and deep learning techniques have been widely applied to the analysis of multi-source physiological data, effectively improving target recognition and precise classification^{16,17}. In the maritime domain, Li et al.¹⁸ integrated ECG and EDA signals with machine learning methods to achieve high-precision classification of mental workload levels. In the construction domain, Jeon and Cai¹¹ collected EEG signals in a VR environment and developed an EEG-based classifier to identify construction-related hazards at worksites, with the CatBoost model achieving an accuracy of 95.1%. Furthermore, deep learning has attracted increasing scholarly attention due to its ability to capture potential high-order correlations, thereby enhancing model sensitivity and stability. For instance, Zhang et al.¹⁹ employed multi-source physiological data and proposed a Regularized Deep Fusion of Kernel Machine framework for emotion recognition. Umer et al.²⁰ applied LSTM, Bi-LSTM, and GRU to monitor construction workers' fatigue. Despite the growing integration of machine learning and deep learning with multi-source physiological data, most studies remain focused on emotion and fatigue recognition. There is still a need for further exploration into the assessment of operators' hazard perception levels in high-risk industries, particularly among operators in intelligent coal mine safety monitoring centers.

Therefore, this study aims to integrate EEG, EDA, and HRV signals to construct a model for assessing the hazard perception levels of safety monitoring center operators during remote monitoring. By systematically introducing and comparing 12 models, including AdaBoost, Decision Tree, Extra Trees, Gradient Boosting, KNN, LightGBM, Naive Bayes, Random Forest, Ridge, SVM, XGBoost, and 1D-CNN, this study proposes a low-intrusion and high-accuracy hazard perception assessment model. The proposed method provides a feasible pathway for the quantitative assessment of hazard perception levels among operators in high-risk industries such as coal mining, enabling dynamic monitoring and intelligent intervention, thereby enhancing accident prevention capacity in intelligent coal mines and supporting safe production.

The remainder of this study is structured as follows. Section 2 provides a comprehensive review of the literature on hazard perception. Section 3 outlines the experimental procedure, including participants, equipment, and experimental design. Section 4 presents the details of data preprocessing and the development of the hazard perception assessment model. Section 5 discusses the research findings in depth and demonstrates the effectiveness of the proposed model. Section 6 summarizes the research findings, implications, limitations, and future research directions. Finally, Sect. 7 concludes the study.

Related work

Hazard perception assessment

Hazard perception refers to an individual's ability to sense, identify, and predict dangers or unsafe factors in their environment²¹. The concept of hazard perception originated in industrial safety and human factors engineering, serving as the first line of defense in accident prevention. It plays a vital role in anticipating and responding to potential hazards, particularly in high-risk areas such as coal mine safety, traffic safety, and emergency management. The assessment of hazard perception is crucial for accident prevention, safety management, and risk control. Three primary methods are used to measure hazard perception: subjective questionnaire, behavioral measurement, and physiological signal analysis. The subjective method typically relies on questionnaires to obtain individuals' self-reported hazard perception levels, such as the standardized Hazard Perception Questionnaire²². While offering operational simplicity, subjective questionnaires are susceptible to individual biases and cognitive differences. Behavioral measurement methods indirectly assess hazard perception by observing changes in task performance during hazardous scenarios. In the traffic domain, researchers have investigated pedestrians' hazard perception by measuring response times and accuracy in response to hazards^{23,24}. In the construction domain, hazard perception has been assessed through workers' visual search, attention distribution, and gait parameters⁶. Although behavioral measures can objectively reflect hazard perception levels in contexts close to real-world situations, they are easily influenced by testing scenarios and individual states, and their evaluation dimensions remain limited. Considering the limitations of subjective questionnaires and behavioral measurement, physiological signals have been increasingly employed. For instance, Jeon and Cai¹¹ used EEG signals to perform multi-class classification of construction hazards. Chong et al.²⁵ employed EDA signals to investigate workers' ability to identify safety risks. Abdullahi et al.²⁶ used HR and HRV to evaluate the effects of fatigue on participants' hazard perception performance.

Machine learning and deep learning in hazard perception assessment

The integration of machine learning and deep learning into the analysis of physiological data can significantly enhance the ability to identify individuals' mental states²⁷. In the high-risk environment of intelligent coal mine safety monitoring centers, operators are required to maintain a high hazard perception level while performing

multiple concurrent tasks. Notably, in work contexts characterized by high cognitive load and heavy information processing, a decline in operators' hazard perception levels may lead to the neglect of hazard signals²⁸, thereby increasing the probability of coal mine accidents. Machine learning provides an effective method of analyzing multi-source physiological data. For example, Liang and Lin²⁹ integrated machine learning with EEG and EDA signals, and successfully classified hazardous and safe drivers in a hazard perception experiment using the LDA method. In the construction domain, researchers have integrated EDA, PPG, and ST physiological signals with four supervised machine learning methods to identify construction workers' risk perception levels, with the GSVM model achieving 81.2% accuracy in distinguishing between low and high risk perception³⁰. Compared with traditional machine learning, deep learning offers stronger capabilities for automatic feature extraction. Consequently, researchers have begun to introduce deep learning into hazard classification research. For instance, Duorinaah et al.¹⁵ recorded EEG, EDA, PPG, and ST signals under different risk conditions and trained 15 supervised machine learning models and 3 deep learning models, ultimately achieving a risk classification accuracy of 99.8%. Beyond risk classification, Hao et al.³¹ employed a combined machine learning and deep learning method to evaluate mental workload in high-risk tasks. Moreover, while substantial studies have been conducted in traffic and construction domains, investigations into hazard perception among coal mine operators are limited. Given that coal mine safety monitoring center operators face large volumes of real-time, multi-source dynamic data, their hazard perception levels directly determine coal mine safety. Therefore, the introduction of machine learning and deep learning is particularly critical. By integrating these models with physiological signal data, the assessment framework overcomes the limitations of traditional reliance on managerial observation and subjective assessment, enables a faster and more direct understanding of operators' states, and reveals latent associations between multi-source physiological data and behavioral performance.

Research gaps

A systematic review of existing studies reveals that current research on hazard perception still has several limitations. First, compared with previous research, relatively few studies have focused on individuals' own hazard perception levels, while most have concentrated on identifying external environmental risks^{32–34}. Second, existing risk-related studies have primarily concentrated on the traffic and construction domains, with limited attention given to the coal mining industry. Due to its unique environmental conditions, coal mining exhibits significantly higher accident probabilities than many other industries³⁵, placing stricter requirements on the hazard perception levels of safety monitoring center operators. Third, although existing methods for hazard perception assessment include subjective questionnaires, behavioral measurement, and physiological signal analysis, each method has its own advantages and limitations. Few studies combine multiple methods, resulting in insufficient comprehensiveness in hazard perception research.

Based on these limitations, this study adopts a multidimensional method by combining subjective methods (questionnaires) with objective methods (task performance and physiological signal) to identify operators' hazard perception levels. EEG, EDA, and HRV signals of the operators are measured in real time. Hazard perception levels are determined through the Hazard Perception Questionnaire and an E-prime hazard perception experiment. Hazard perception assessment model will be developed using 12 machine learning models and deep learning models, with the optimal model identified via model evaluation metrics and confusion matrices. In addition, by comparing the performance of models under different combinations of physiological data, the optimal combination is identified. This study focuses on detecting operators' hazard perception levels during prolonged working periods, with the aim of enabling timely identification of insufficient risk perception caused by fatigue, reduced attention, or accumulated mental workload, thereby advancing the refinement of safety management.

Experiment design

The overall procedure of the hazard perception level assessment model constructed in this study is illustrated in Fig. 1. The process comprises three steps: (1) Various coal mine accidents are simulated at the safety monitoring center, and EEG, EDA, and HRV physiological data are recorded. (2) Data preprocessing. The recorded physiological signals are preprocessed, and relevant features are extracted for each category of physiological data. (3) Hazard perception level assessment. Both traditional machine learning models and deep learning models are employed to classify hazard perception levels, and their accuracy and performance are compared and discussed.

Participants

The determination of the minimum sample size is achieved through a priori power analysis using G*Power 3.1.9.7³⁶. This study employs multiple paired t-tests to examine physiological differences across hazard perception levels, and the corresponding parameter settings are presented in Table 1. Effect sizes used in prior physiological signal research include 0.4³⁷, 0.5³⁸, 0.8³⁹, and 0.9⁴⁰. Accordingly, an effect size of 0.7 is selected for this study. Considering that pairwise comparisons among the three hazard perception levels involve multiple testing, the Bonferroni correction is applied, adjusting the significance threshold to $\alpha = 0.017$, and the statistical power is set to 0.80. Based on these parameters, the minimum sample size required for this study is 21 participants.

A total of 26 participants are recruited from three intelligent coal mine safety monitoring centers. All participants undergo vision screening and complete the Patient Health Questionnaire for mental health assessment⁴¹. Three participants are excluded due to elevated mental health risk (PHQ-9 ≥ 10). Ultimately, 23 participants complete the experimental procedures. The sample comprised participants with diverse genders, ages, and lengths of work experience to ensure adequate representativeness and comprehensiveness. Their ages ranged from 23 to 50 years ($M = 35.6$, $SD = 9.2$), and their work experience range from less than 1 year to more than 10 years ($M = 7.24$, $SD = 4.23$). Detailed sample characteristics are presented in Table 2.

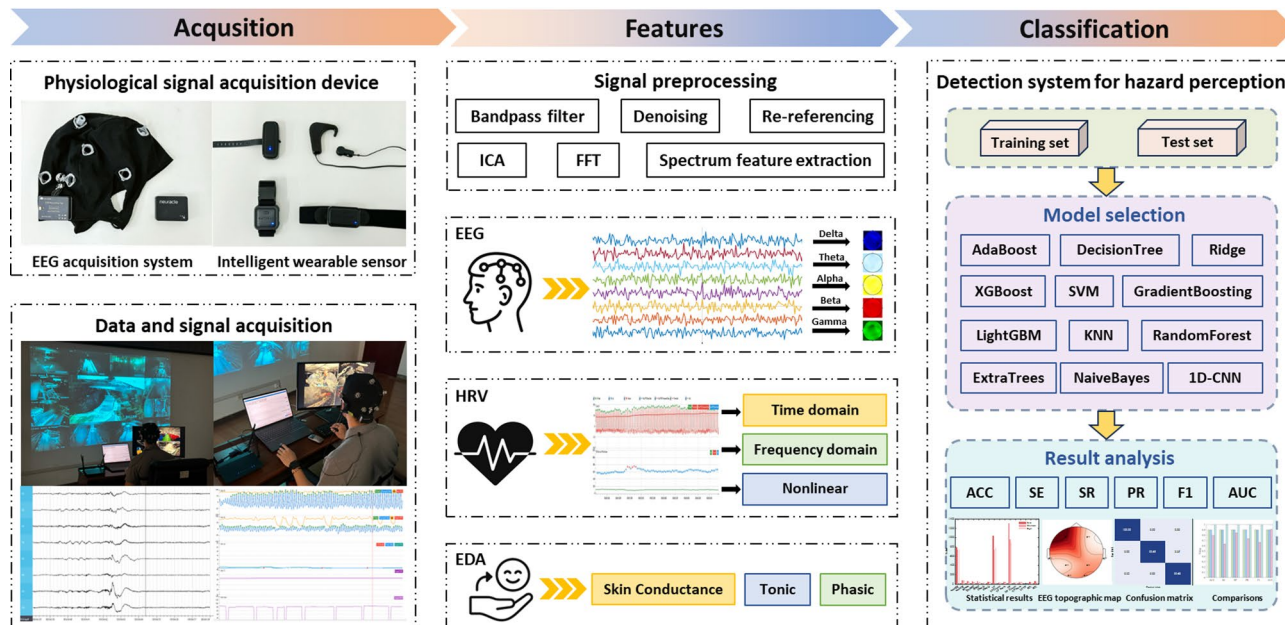


Fig. 1. The overall framework of the hazard perception assessment model for operators.

Variable name	Variable value
Test family	t tests
Statistical test	Means: Difference between two dependent means (matched pairs)
Type of power analysis	A priori: Compute required sample size - given α , power, and effect size
Effect size dz	0.7
α error probability	0.017
Power (1- β error probability)	0.80

Table 1. Variables used in G*Power.

Characteristic	Category	Frequency	Percentage (%)
Gender	Male	15	65.2
	Female	8	34.8
Age	20 ≤ Age < 30	5	21.7
	30 ≤ Age < 40	11	47.8
	40 ≤ Age < 50	7	30.4
Work experience	Experience < 1	2	8.7
	1 ≤ Experience < 5	6	26.1
	5 ≤ Experience < 10	8	34.8
	Experience ≥ 10	7	30.4

Table 2. Sample characteristics of participants.

Prior to the experiment, participants are required to have at least seven hours of sleep and to refrain from consuming alcohol, coffee, or other central nervous system stimulants within 24 h before testing, and to sign informed consent forms. The experiment is carried out in accordance with relevant guidelines and regulations, including the Declaration of Helsinki, and is approved by the Institutional Review Board of Taiyuan University of Technology (TYUT2025121102).

Experimental environment and setup

In the safety monitoring center, operators analyze and evaluate hazard signals at the working face through multi-screen displays on small-pitch LED screens. In this experiment, E-prime 3.0 software is used to simulate a coal

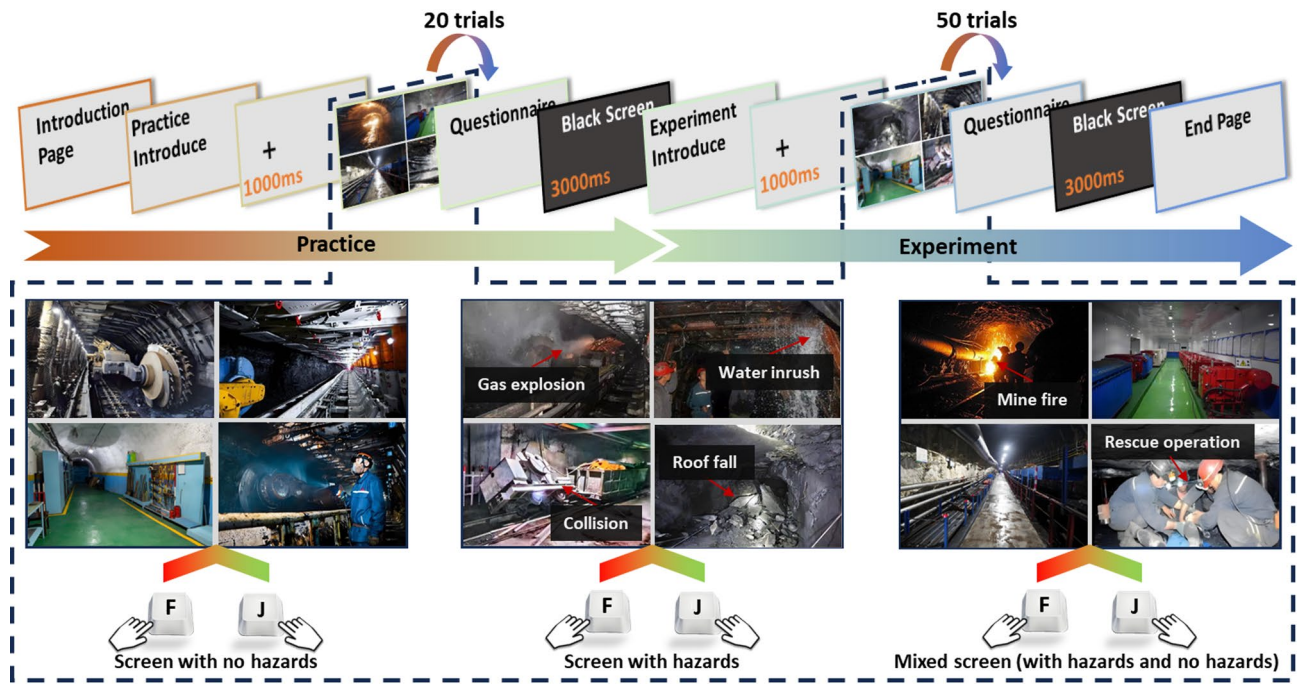


Fig. 2. Sample illustration of the hazard perception experiment implemented in E-prime.

Coding	Items	Response	Factor loadings	Related metrics
HP1	Spotting hazards quickly	[1] [2] [3] [4] [5]	0.832	Cronbach's $\alpha=0.876$ CR=0.877 AVE=0.545
HP2	Spotting hazards with enough time to react	[1] [2] [3] [4] [5]	0.740	
HP3	Spotting numerous hazards at a time in multi-screen monitoring?	[1] [2] [3] [4] [5]	0.844	
HP4	Reacting to more than one potential hazard at a time	[1] [2] [3] [4] [5]	0.796	
HP5	Spotting hazards under high workload?	[1] [2] [3] [4] [5]	0.720	
HP6	Spotting hazards under low workload?	[1] [2] [3] [4] [5]	0.778	

Table 3. Hazard perception questionnaire and results of reliability and validity analysis. Note: All items are rated on a scale from 1 (strongly disagree) to 5 (strongly agree), with a midpoint of 3 (neutral).

mine safety monitoring center, presenting coal mine images across four small screens (as shown in Fig. 2). All images are obtained from coal mining enterprises or online coal mine resources, ensuring their authenticity and representativeness. A total of 286 images are used, including 147 hazardous images and 139 safe images. Hazardous images refer to scenarios in underground coal mine operations where actual or potential safety hazards are present, such as personnel without safety gear, excessive methane concentrations, unstable roof conditions, equipment sparks, excessive coal buildup on conveyors, water accumulation in tunnels, obstructed ventilation, or improper operations that could trigger accidents. Such images contain obvious cues of unsafe behaviors, environmental abnormalities, or accident precursors, providing strong indicators for hazard perception. Images without safety hazards are defined as safe images. These coal mine images are presented in two-dimensional forms to better simulate actual monitoring operations. To avoid learning effects, each image is judged by participants only once.

The subjective assessment in this experiment is conducted using the Hazard Perception Questionnaire (HPQ)²². The HPQ is a self-assessment tool designed to evaluate individuals' ability to identify and assess potential hazards in specific contexts. It is originally developed for traffic safety applications and has since been extended to industrial safety and other high-risk operational scenarios. The questionnaire adopts a Likert scale, ranging from 1 (strongly disagree) to 5 (strongly agree).

In this study, the questionnaire is modified according to actual coal mine scenarios to better fit with the coal mine safety context, as detailed in Table 3. To ensure the reliability and validity of the questionnaire, a pre-survey is conducted prior to the formal experiment, involving 60 coal mining practitioners with relevant professional backgrounds. A total of 53 valid questionnaires are retrieved. Subsequently, reliability and validity analyses are performed using SPSS 27.0 and Mplus 8.0. As shown in Table 3, these results indicate that the revised questionnaire demonstrates satisfactory reliability and validity.

In addition to subjective method, this study collects participants' physiological signals in real time during the experiment. The Neusen W-series wireless EEG acquisition system, produced by Neuracle, is employed to

record EEG signals (as shown in Fig. 3a). This system enables high-precision wireless data synchronization with a sampling frequency of 1000 Hz. Eight EEG channels are used in this study: F3, F4, T3, T4, P3, P4, O1, and O2, corresponding to the frontal region (F3 and F4), temporal region (T3 and T4), parietal region (P3 and P4), and occipital region (O1 and O2). EDA signals are recorded using the ErgoLAB wearable human factors recorder (as shown in Fig. 3a). This wearable physiological monitoring system is a multi-channel, distributed platform for collecting multiple human physiological parameters, with a sampling frequency of 64 Hz. All acquisition modules are connected wirelessly to the ErgoLAB 3.0 system, enabling the integrated acquisition of multi-source physiological data. The specific experimental setup is illustrated in Fig. 3b. The computing environment is configured as follows: the software is PyTorch 3.10.12, the CPU is AMD 9950X, and the GPU is NVIDIA GeForce RTX 5090.

Experimental procedure

Upon arrival at the laboratory, participants are informed of the experimental procedure and are instructed to remain quiet and avoid interacting with each other throughout the session. Subsequently, they are fitted with an EEG cap and physiological sensors. E-prime 3.0 is used to record participants' accuracy, reaction time, and Hazard Perception Questionnaire scores. During the formal experimental phase, participants are seated in front of the display screen at 60–70 cm, with their gaze fixed on its center. The experimental program comprises the following steps:

- (1) *Instructions* This page introduced the task and keypress rules. The instructions are presented as follows: Dear participant, welcome to this experiment! This study aims to simulate the work of a safety monitoring center to assess your hazard perception level during tasks. During the experiment, you will be presented with multiple monitoring screens showing real coal mine operation scenes in the form of images. Some of these scenes contain hazards, while others represent normal operating conditions. In each trial, please make a judgment based on the content of the monitoring screen: if you believe the scene contains a hazard, press the “F” key on the keyboard; if you believe the scene is safe, press the “J” key. Please remain attentive and respond as quickly as possible while ensuring accuracy when judging whether the presented image contains a hazard. Afterward, please use the numeric keys 1–5 to answer the subjective questionnaire. At this stage, participants are allowed to ask questions.
- (2) *PracticeProc* This phase includes the recognition of 20 hazardous images and one subjective questionnaire, designed to help participants become familiar with the task requirements.

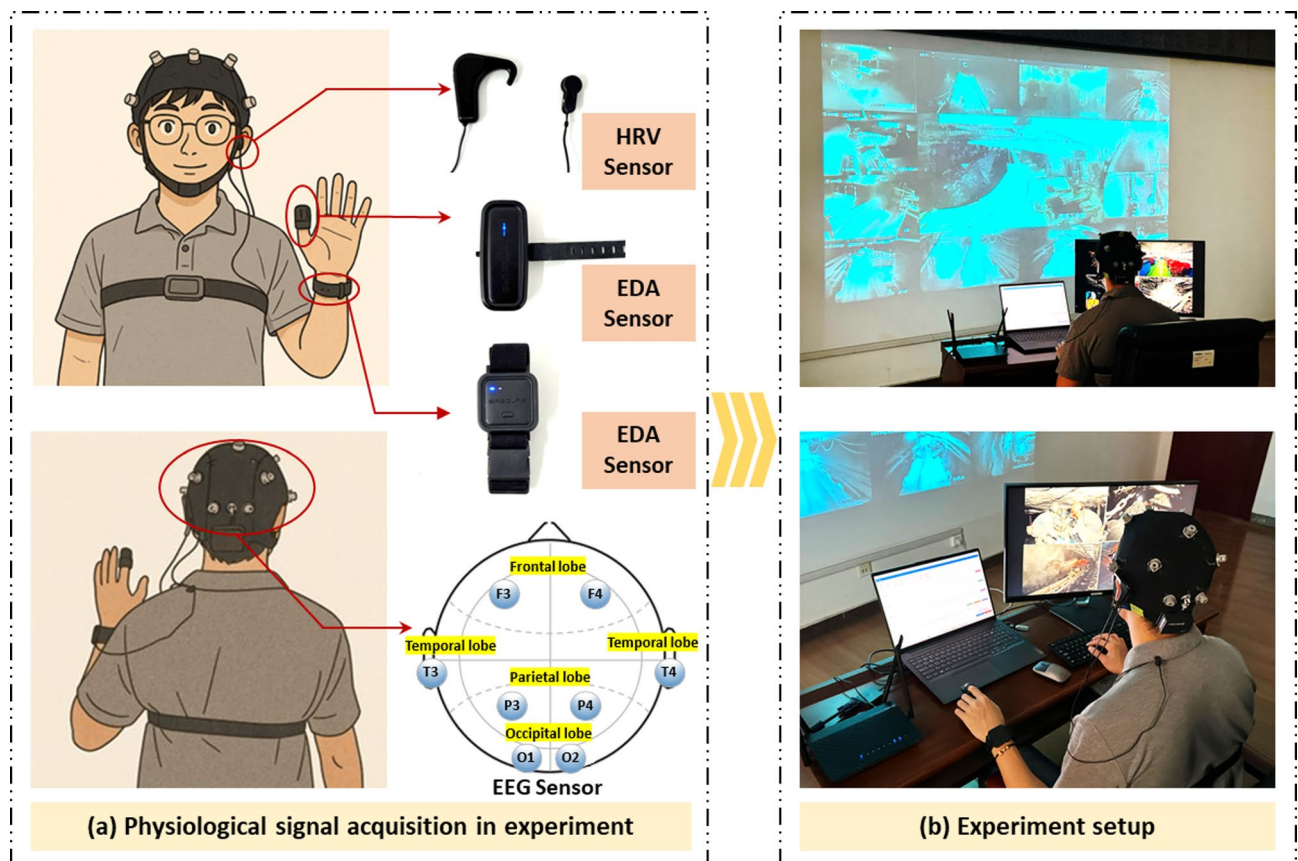


Fig. 3. Participant conducting experiments with all setup devices. (a) Physiological signal acquisition in experiment. (b) Experimental setup in a real scenario.

- (a) *Fixation Point* Before the formal start of each trial, a “+” symbol is briefly presented at the center of the screen to prompt participants to focus their attention.
 - (b) *Image Stimuli* Participants view the image stimuli and press the “F” key if they judged the scene to be hazardous, or the “J” key if they judged it to be safe. The practice phase comprises 20 trials.
 - (c) *Hazard Perception Questionnaire*. Participants use the numeric keys 1–5 to provide a self-assessment of their hazard perception level during the current practice task.
 - (d) *Black Screen Interval*. A blank screen is presented before the start of the next trial to help participants regain attention.
- (3) *MainProc*. This phase comprises 50 trials, during which participants follow the same procedure as in the practice phase.
 - (4) *End Page*. A thank-you message, and an exit prompt are presented.

Method

Data collection and processing

In the experiment, three types of data are collected: subjective data, task performance data, and physiological signal data, as shown in Table 4. Subjective questionnaire data and task performance data are used to quantify and determine the hazard perception level of each participant, thereby providing a basis for subsequent physiological signal data analysis. Specifically, the Hazard Perception Questionnaire score is denoted as HPQ , the average action time as RT , and the overall accuracy as ACC . Because shorter reaction time indicates higher hazard perception levels, RT is first reverse coded, as shown in Eq. 1. Subsequently, HPQ , RT' , and ACC are standardized using the z-score method. Based on the mean value of the three standardized indicators, the Hazard Perception Composite Index ($HPCI$) is constructed, as presented in Eq. 2. In this equation, Z_{HPQ} , $Z_{RT'}$, and Z_{ACC} represent the z-score standardized results of HPQ , RT' , and ACC , respectively.

$$RT' = -RT \quad (1)$$

$$HPCI = \frac{Z_{HPQ} + Z_{RT'} + Z_{ACC}}{3} \quad (2)$$

In the final stage of label assignment, the hazard perception level is determined using a quantile-based labeling method based on the $HPCI$ distribution. Specifically, the 33rd percentile (P33) and 67th percentile (P67) are used as threshold values, where $HPCI \leq P33$ is assigned to the low hazard perception level, $P33 < HPCI \leq P67$ to the moderate hazard perception level, and $HPCI > P67$ to the high hazard perception level. This strategy ensures objectivity in label assignment and prevents severe class imbalance across levels, which could otherwise bias model classification and undermine generalization performance.

Since the collected physiological signals are susceptible to environmental noise, preprocessing is required before analysis. EEG signals are preprocessed using the open-source MATLAB toolbox EEGLAB⁴² for filtering, re-referencing, and independent component analysis (ICA). First, a 1–49 Hz band-pass filter is applied to effectively remove low and high frequency noise. Subsequently, re-referencing is performed to further improve the consistency and comparability of the signals. Finally, independent component analysis (ICA) is used to identify and remove artifacts such as eye blinks, eye movements, and electromyographic activity. EDA and HRV signals are preprocessed using the ErgoLAB software platform. High-pass and low-pass filtering are applied to remove baseline drift and high-frequency noise, respectively. For EDA, the signals are further decomposed into skin conductance level (SCL) and skin conductance response (SCR) components which reflect the activity of the autonomic nervous system. For HRV, the signals are band-pass filtered at 0.5–8 Hz to extract pulse peaks, from which inter-beat interval sequences are derived to calculate heart rate and heart rate variability parameters.

In this study, physiological signals data are divided by participant into a training set (80%) and a testing set (20%), ensuring that data from the same participant do not appear in both sets simultaneously. On this basis, a sliding window segmentation approach with a window length of 5s and a step size of 1s is applied to the continuous physiological signals, which increases the number of valid samples while maintaining physiological validity. A total of 13,708 samples are obtained, including 10,966 samples in the training set and 2742 samples in the test set, which are subsequently used for model training and testing.

Feature extraction

EEG signals are a direct reflection of cortical neural activity. Frequency-domain features are extracted using the fast Fourier transform (FFT) and are divided into five frequency bands: δ (1–4 Hz), θ (4–8 Hz), α (8–13 Hz), β (13–30 Hz), and γ (30–45 Hz). These features are subsequently used for feature analysis and classification modeling. In addition, previous research has shows that power band ratios, including α/β , θ/β , $(\alpha + \theta)/\beta$, and $(\alpha$

Data type	Description
Subjective data	Hazard perception questionnaire
Task performance data	Average action time; Overall accuracy
Physiological signal data	Multi-source physiological data, including EEG, EDA and HRV

Table 4. Collected data used in experiments.

+ θ)/($\alpha + \beta$) are commonly employed as indicators of mental states such as fatigue or alertness⁴³. Therefore, these ratios are also included as feature indices in this study.

EDA, as a key physiological signal reflecting sympathetic nervous system activity, is often evaluated through multidimensional feature analysis to assess individual states. Its features are generally categorized into three types: overall skin conductance (SC), tonic component, and phasic component⁴⁴. SC features include meanSC, maxSC, minSC, stdSC, and varSC, which primarily reflect the overall level of skin conductance and its fluctuations. Tonic features include meanTonic, maxTonic, minTonic, stdTonic, and varTonic, which describe low-frequency, slow-varying trends and reveal baseline levels and overall arousal states during prolonged tasks. Phasic features include meanPhasic, maxPhasic, minPhasic, stdPhasic, and varPhasic, which capture short-term rapid responses and reflect operators' immediate reactivity to sudden stimuli.

HRV is a physiological signal that reflects autonomic nervous system regulation through the analysis of variations in inter-beat intervals. Feature extraction based on HRV enables effective evaluation of an individual's cardiac functional status, mental workload, and emotional regulation capacity. Common HRV analysis methods are typically categorized into time-domain, frequency-domain, and nonlinear analysis³¹. In time-domain analysis, the variability of inter-beat intervals is measured directly in the time domain. Common indices include meanRR, meanHR, SDNN, RMSSD, SDD, as well as the proportion of successive interval differences exceeding 20 ms and 50 ms (PNN20 and PNN50). Frequency-domain analysis estimates power spectral density to reflect autonomic nervous activity across different frequency components. Typical indices include low-frequency power (LF), high-frequency power (HF), normalized LF and HF power (LF Power Norm, HF Power Norm), their corresponding percentages (LF Power Percent, HF Power Percent), and the LF/HF ratio, which is widely used to evaluate the balance between sympathetic and parasympathetic nervous system activity. In addition, nonlinear analysis reveals the complex dynamic characteristics of heart rate regulation, often calculated based on the Poincaré plot. Related indices such as SD1 and SD2 provide additional feature descriptors. All physiological features and their descriptions are summarized in Table 5.

Classification models

In this study, 12 classification models are employed, including LightGBM, 1D-CNN, Random Forest, Gradient Boosting, KNN, XGBoost, Extra Trees, Decision Tree, SVM, AdaBoost, Naive Bayes, and Ridge, which comprise both machine learning and deep learning models. The study aims to comprehensively compare their assessment performance and thereby identify the optimal model for assessing operators' hazard perception levels.

- (1) AdaBoost. An iterative boosting algorithm that combines multiple weak classifiers with weight adjustments to progressively minimize classification errors, thereby enhancing generalization ability and robustness.
- (2) DecisionTree. A hierarchical model based on feature splitting that intuitively represents the mapping between features and classes, characterized by high interpretability and fast training speed.
- (3) ExtraTrees. An ensemble method that introduces randomization in feature selection and split thresholds, resulting in a low-bias, high-variance model that improves stability and computational efficiency.
- (4) GradientBoosting. A boosting algorithm that builds an ensemble by iteratively fitting residuals, capable of effectively capturing complex nonlinear relationships and achieving high accuracy in classification tasks.
- (5) KNN. A lazy learning algorithm based on distance metrics, which classifies samples through majority voting among neighboring instances. It is simple and intuitive, making it suitable for non-parametric modeling.
- (6) LightGBM. An efficient gradient boosting method that employs histogram-based feature splitting and a leaf-wise growth strategy, providing superior efficiency and accuracy in high-dimensional data.
- (7) NaiveBayes. A probabilistic classifier based on Bayes' theorem and the assumption of conditional independence among features. It offers rapid training speed and low computational cost, making it suitable as a baseline model.
- (8) RandomForest. An ensemble of multiple decision trees constructed with incorporated randomness, which enhances robustness and generalization ability while effectively mitigating overfitting.
- (9) SVM. A classification algorithm that constructs an optimal hyperplane by maximizing the margin between classes, well suited for high-dimensional feature spaces and effective under small-sample conditions.
- (10) XGBoost. An optimized gradient boosting framework that incorporates regularization and parallel computation, combining high accuracy with efficiency, and demonstrating advantages in complex classification tasks.
- (11) Ridge. A linear regression model with a penalty term, using L2 regularization to suppress multicollinearity and overfitting, thereby providing improved stability.
- (12) 1D-CNN. A deep learning model capable of automatically extracting local and hierarchical features from time-series data, demonstrating strong performance in sequence classification tasks such as physiological signal analysis.

To train the models and optimize their hyperparameters, five-fold cross-validation is applied within the training set. Specifically, the training set is further partitioned into five mutually exclusive subsets, with four subsets used for model training and the remaining subset used for validation in each fold. This procedure is repeated five times, and the average performance across the five folds is used to guide model selection and hyperparameter tuning. Finally, the selected model is evaluated once on the independent test set to assess its generalization performance on unseen participants.

Six evaluation metrics (Accuracy, Sensitivity, Specificity, Precision, F1-Score, and AUC) is used to assess the model's performance. Accuracy is the percentage of correctly classified samples (see Eq. 3). Sensitivity is the

Signals	Indexes	Description
EEG	δ	Frequency ranges from 1 to 4 Hz
	θ	Frequency ranges from 4 to 8 Hz
	α	Frequency ranges from 8 to 13 Hz
	β	Frequency ranges from 13 to 30 Hz
	γ	Frequency ranges from 30 to 45 Hz
	α/β	Ratio of α to β
	θ/β	Ratio of θ to β
	$(\alpha + \theta)/\beta$	Ratio of $\alpha + \theta$ to β
	$(\alpha + \theta)/(\alpha + \beta)$	Ratio of $\alpha + \theta$ to $\alpha + \beta$
EDA	meanSC	Mean of Skin Conductance Data
	maxSC	Maximum Skin Conductance Data
	minSC	Minimum of Skin Conductance Data
	stdSC	Standard deviation of Skin Conductance Data
	varSC	Variances of Skin Conductance Data
	meanTonic	Mean of Tonic Data
	maxTonic	Maximum of Tonic Data
	minTonic	Minimum of Tonic Data
	stdTonic	Standard deviation of Tonic Data
	VarTonic	Variances of Tonic Data
	meanPhasic	Mean of Phasic Data
	maxPhasic	Maximum of Phasic Data
	minPhasic	Minimum of Phasic Data
	stdPhasic	Standard deviation of Phasic Data
VarPhasic	Variances of Phasic Data	
HRV	meanRR	Mean of NN intervals
	meanHR	Mean of HR
	SDNN	Standard deviation of NN intervals
	RMSSD	Square root of the mean of the sum of the squares of differences between adjacent NN intervals
	SDSD	Standard deviation of differences between adjacent NN intervals
	PNN50	NN50 count divided by the total number of all NN intervals
	PNN20	NN20 count divided by the total number of all NN intervals
	LF Power	Power of the low frequency band (0.04–0.15 Hz)
	LF Power Percent	Ratio of LF Power
	LF Power Norm	LF power in normalized
	HF Power	Power of the high frequency band (0.15–0.4 Hz)
	HF Power Percent	Ratio of HF Power
	HF Power Norm	HF power in normalized
	LF/HF	Ratio of LF to HF
	SD1	Std. deviation of the major axis of the ellipse in Poincaré plot
SD2	Std. deviation of the short axis of the ellipse in Poincaré plot	

Table 5. All extracted features and their descriptions.

percentage of correctly identified positive samples (see Eq. 4). Specificity is the percentage of negative samples correctly identified as negative (see Eq. 5). Precision is the percentage of samples predicted as positive that are positive (see Eq. 6). F1-Score is the harmonic means of Precision and Sensitivity (see Eq. 7). AUC is the area under the ROC curve, representing the overall percentage of correct classifications by the model, with values closer to 1 indicating superior performance. In this study, TP denotes the number of samples correctly classified as positive, TN denotes the number of samples correctly classified as negative, FP denotes the number of negative samples incorrectly classified as positive, and FN denotes the number of positive samples incorrectly classified as negative.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (4)$$

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (5)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad (6)$$

$$F1 = \frac{2TP}{2TP + FP + FN} \quad (7)$$

Data analysis and results

Subjective data and task performance data

Based on the subjective data and task performance data, the participants' Hazard Perception Questionnaire scores, average action time, and overall accuracy are presented in Fig. 4. A one-way ANOVA is conducted to examine overall differences among the low, moderate, and high levels. For *HPQ*, significant differences are observed between the Low ($M=16.74$, $SD=2.73$), Moderate ($M=19.87$, $SD=2.67$), and High ($M=22.74$, $SD=3.33$) hazard perception levels ($F=24.175$, $p<0.001$). For *RT*, significant differences are observed between Low ($M=855.02$, $SD=81.80$), Moderate ($M=751.14$, $SD=96.02$), and High ($M=723.80$, $SD=111.83$) hazard perception levels ($F=11.637$, $p<0.001$). Regarding *ACC*, significant differences exist between Low ($M=0.85$, $SD=0.04$), Moderate ($M=0.87$, $SD=0.03$), and High ($M=0.90$, $SD=0.04$) hazard perception levels ($F=10.225$, $p<0.001$).

These results demonstrate that participants with higher hazard perception levels exhibit higher subjective Hazard Perception Questionnaire scores, shorter average action times, and higher overall accuracy. Moreover, *HPQ*, *RT*, and *ACC* differ significantly across hazard perception levels, supporting the validity of the label assignment.

Results analysis of physiological indexes

Figure 5a,b present the variations in EDA and HRV features, across low, moderate, and high hazard perception levels of operators. For the EDA features, meanSC and meanTonic show an increasing trend, while stdPhasic shows a decreasing trend. Indicators such as maxSC, stdSC, varSC, maxTonic, stdTonic, VarTonic, meanPhasic, and maxPhasic initially increase and then decrease, whereas minSC, minTonic, minPhasic, and VarPhasic initially decrease and then increase. For the HRV indices, meanHR shows an increasing trend, while meanRR, SDNN, RMSSD, SDDSD, PNN50, PNN20, HF Power, and SD1 all show decreasing trends. In contrast, LF Power, LF Power Percent, HF Power Percent, and HF Power Norm initially decrease and then increase, while LF Power Norm, LF/HF, and SD2 initially increase and then decrease. Thus, operators' hazard perception levels can be inferred by observing these feature variations.

Table 6 presents the variations in key physiological features across different hazard perception levels together with the corresponding paired-sample t-test results. To improve the robustness of statistical rigor, Bonferroni correction is applied to all paired-sample t-test results. Four main findings are observed. First, in terms of EDA, stdSC and stdTonic show an increasing trend from low hazard perception level to moderate hazard perception level, followed by a slight decrease from moderate to high hazard perception level. When hazard perception level rises from low to moderate, the sympathetic nervous system is significantly activated, leading to intensified fluctuations in skin conductance signals. However, as operators reach the high hazard perception level, the body may remain under sustained tension, thereby reducing response flexibility. Significant differences are found between low and moderate hazard perception levels as well as between low and high hazard perception levels ($p<0.001$). This indicates that as hazard perception level increases, sympathetic nervous system excitability is enhanced, resulting in intensified fluctuations in electrodermal activity.

Second, stdPhasic does not show significant differences across different hazard perception levels. This may be attributed to the experimental design, which involved a series of continuous hazard perception tasks. With the continuous presentation of stimuli, participants' instantaneous responses may gradually diminish. The phasic component primarily reflects short-term, event-driven electrodermal responses and is typically more sensitive under sudden stimuli or acute stress conditions. In contrast, in continuous or progressive hazardous scenarios, its fluctuations remain relatively limited, resulting in non-significant outcomes.

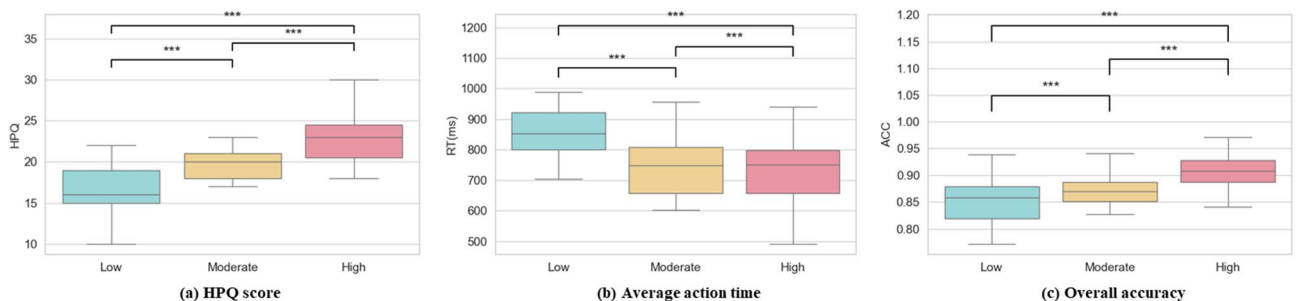


Fig. 4. Comparison of subjective data and task performance data across hazard perception levels. (a) HPQ score. (b) Average action time. (c) Overall accuracy.

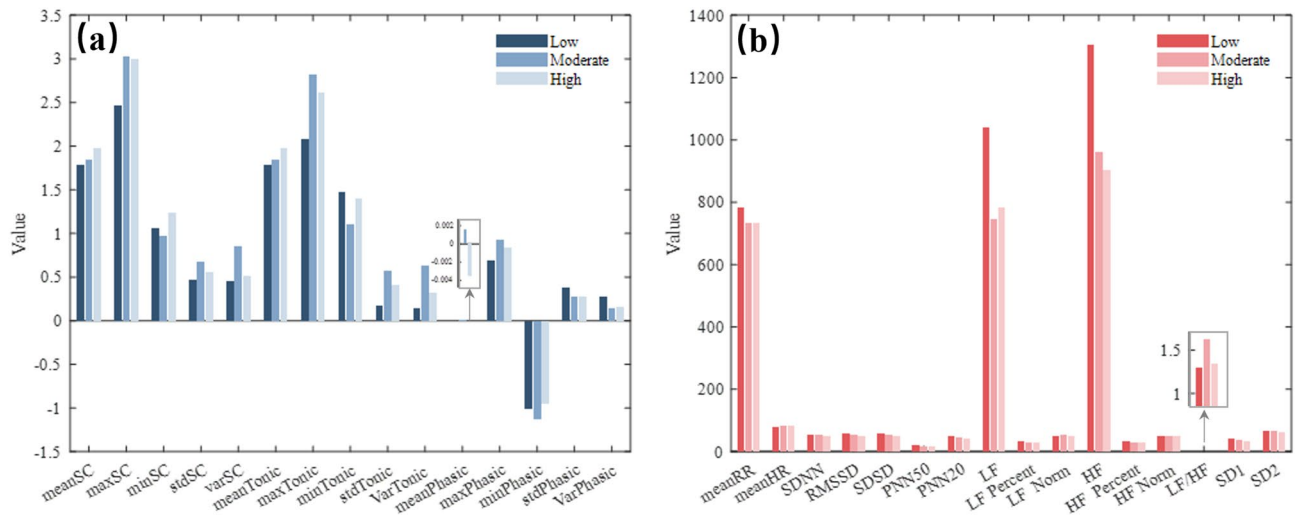


Fig. 5. Statistical results of physiological features under varying hazard perception levels. **(a)** The results of EDA features. **(b)** The results of HRV features.

Physiological feature	HP level	Mean	Diff	Standard Deviation	Paired t-test P-Value		
					Low vs. moderate	Low vs. high	Moderate vs. high
stdSC (μS)	Low	0.466	—	56.563	<0.001*	<0.001*	0.154
	Moderate	0.675	0.209	0.634			
	High	0.555	-0.120	0.457			
stdTonic (μS)	Low	0.179	—	0.340	<0.001*	<0.001*	0.067
	Moderate	0.569	0.390	0.566			
	High	0.408	-0.161	0.403			
stdPhasic (μS)	Low	0.374	—	0.377	0.0612	0.169	0.658
	Moderate	0.277	-0.097	0.250			
	High	0.273	-0.004	0.278			
meanHR (bpm)	Low	77.285	—	7.935	<0.001*	<0.001*	0.879
	Moderate	82.626	5.341	8.093			
	High	82.862	0.235	8.461			
SDNN (ms)	Low	55.668	—	43.881	0.396	0.034	0.0985
	Moderate	55.282	-0.386	37.889			
	High	50.024	-5.258	41.726			

Table 6. Changes in key physiological features in response to varying hazard perception levels and paired t-test results. Note: Bonferroni correction significance threshold * $p < 0.0167$.

Third, the mean heart rate of operators at low hazard perception level is 77.285 bpm, which increases to 82.626 bpm (+ 5.341 bpm) at moderate hazard perception level, and further increases to 82.862 bpm (+ 0.235 bpm) at high hazard perception level. This indicates that an increase in heart rate is associated with higher hazard perception levels. Significant differences are observed between low and moderate hazard perception levels as well as between moderate and high hazard perception levels ($p < 0.001$), whereas no statistically significant difference is found between moderate and high hazard perception levels ($p = 0.879$).

Finally, SDNN does not show significant differences across the three hazard perception levels. A plausible explanation is related to the temporal properties of SDNN. As a long-term HRV indicator, SDNN reflects overall HRV and requires relatively long ECG segments to achieve sufficient statistical stability. In contrast, the hazard perception tasks in our study involve transient physiological responses. Therefore, SDNN may not be sensitive to short-term fluctuations in hazard perception.

In this study, EEG frequency bands are extracted from each channel, and the average band power across all channels is calculated for each hazard perception level. Figure 6 shows EEG topographic maps of the delta, theta, alpha, beta, and gamma frequency bands. Distinct dynamic changes in EEG frequency bands are observed with variations in hazard perception level. Specifically, theta, alpha, and beta bands continuously increase as hazard perception level rises, particularly in the frontal F3 region. This finding is consistent with previous studies,

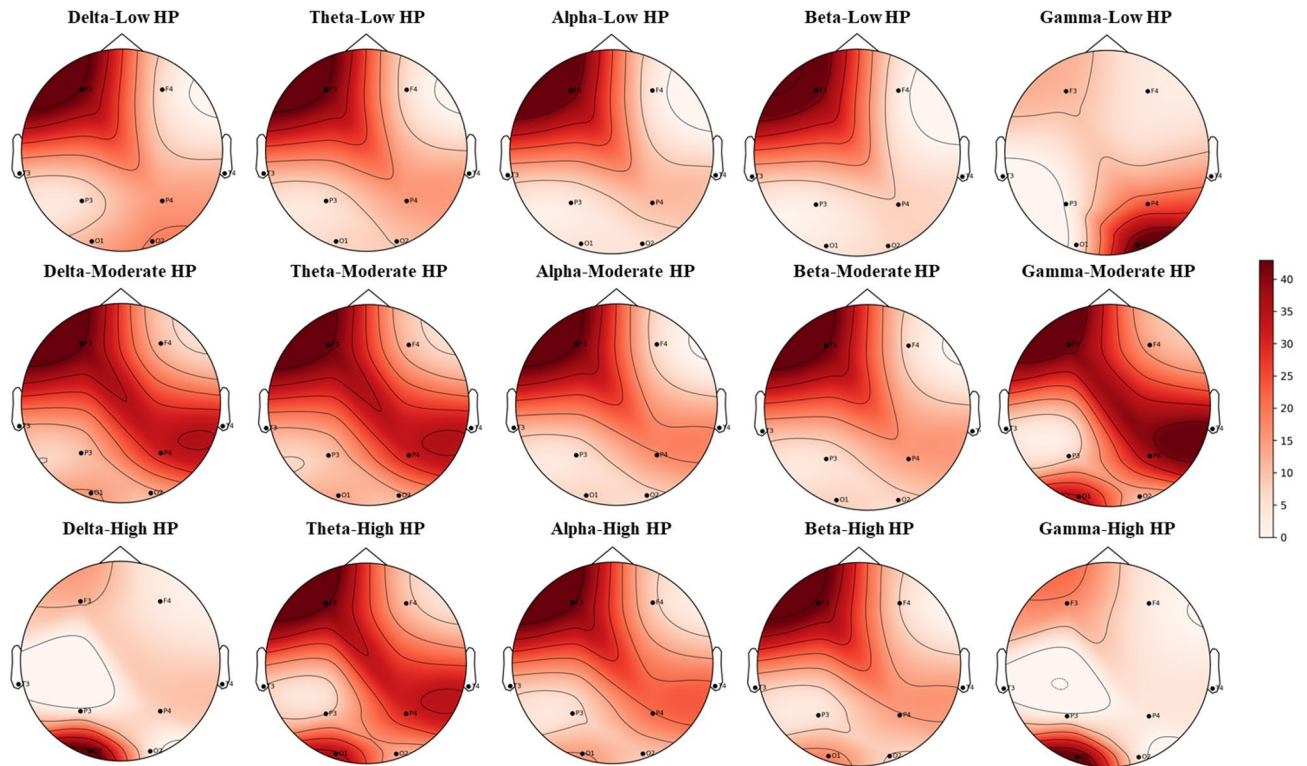


Fig. 6. EEG topographic map for theta, alpha, beta, and gamma bands.

NO.	Classifiers	Accuracy	Sensitivity	Specificity	Precision	F1-Score	AUC
1	LightGBM	0.9989	0.9978	0.9992	0.9989	0.9983	0.9999
2	1D-CNN	0.9907	0.9821	0.9821	0.9886	0.9853	0.9991
3	RandomForest	0.9554	0.9465	0.9406	0.9406	0.9434	0.9970
4	GradientBoosting	0.9400	0.9400	0.9440	0.9439	0.9373	0.9985
5	KNN	0.9211	0.8458	0.9275	0.9275	0.8799	0.9888
6	XGBoost	0.8530	0.7023	0.8617	0.9165	0.7666	0.9915
7	ExtraTrees	0.8187	0.6450	0.8239	0.9279	0.7175	0.9999
8	DecisionTree	0.7658	0.5446	0.7725	0.9123	0.6037	0.7713
9	SVM	0.7597	0.5224	0.5224	0.8360	0.5615	0.8434
10	AdaBoost	0.6976	0.4142	0.7081	0.8480	0.4141	0.7530
11	NaiveBayes	0.6251	0.3485	0.3485	0.3752	0.3234	0.5337
12	Ridge	0.4041	0.3845	0.3845	0.3723	0.3518	0.5369

Table 7. Comparison of test performance across 12 models.

which have report that elevated theta and beta power in the frontal region are associated with higher hazard perception^{45,46}. The delta band shifts from the frontal F3 region to the occipital O1 region as hazard perception level increases, while the gamma band shifts from the occipital O2 region to the occipital O1 region, showing a pattern of initial increase followed by a decrease in average power. This suggests that high hazard perception may inhibit activities related to information integration and higher-order cognitive processes. Overall, the results demonstrate that EEG band power features differ across hazard perception levels, indicating that EEG frequency band power can serve as an indicator for assessing operators' hazard perception levels.

Model performance evaluation

Table 7 presents the assessment performance of 12 models based on 3 types of physiological signals, ranked in descending order of accuracy. First, the best-performing method is the gradient boosting model LightGBM, which achieves a testing accuracy of 99.89%, sensitivity of 99.78%, specificity of 99.92%, precision of 99.89%, F1-score of 99.83%, and an AUC close to 1, indicating excellent performance in hazard perception assessment. Second, it is found that among the six tree-based models (LightGBM, Random Forest, Gradient Boosting, XGBoost, Extra Trees, and Decision Tree), their rankings are 1, 3, 4, 6, 7, and 8, respectively. Their performance

is generally superior to that of traditional machine learning models (e.g., SVM, Naive Bayes, Ridge). This advantage derives from the ability of tree-based models to flexibly model complex nonlinear relationships and to perform particularly well in scenarios with high-dimensional data and strong feature interactions, thereby demonstrating strong robustness and adaptability⁴⁷. Such characteristics enable tree-based models to exhibit superior performance in complex physiological signal analysis. In addition, the deep learning model 1D-CNN achieves an Accuracy of 99.07% and an AUC close to 1, demonstrating significant advantages in overall assessment capability. Through convolutional operations, 1D-CNN can automatically extract local and hierarchical features from physiological signals, effectively capturing latent patterns in temporal information. This enables it to overcome the limitations of traditional methods in feature selection and feature interaction modeling⁴⁸.

The six best-performing models (LightGBM, 1D-CNN, Random Forest, Gradient Boosting, KNN, and XGBoost) are selected based on their assessment performance, and their confusion matrices are shown in Fig. 7. LightGBM demonstrates the best assessment performance, with correct recognition rates exceeding 99% for low, moderate, and high hazard perception level samples. The 1D-CNN model also shows superior performance, achieving correct recognition rates above 96% across all three classes. The Random Forest model performs well for low and moderate hazard perception level samples but shows a decline in recognition accuracy for high hazard perception level samples. GradientBoosting, KNN, and XGBoost exhibit relatively limited recognition ability. Overall, LightGBM demonstrates both high accuracy and robustness in hazard perception level assessment.

Figure 8a–f presents the performance of the six models in hazard perception level assessment under dual and triple physiological signal combinations. Four types of physiological signal combinations, namely EEG and EDA, EEG and HRV, EDA and HRV, as well as EEG, EDA, and HRV, are used to retrain the six models (LightGBM, 1D-CNN, Random Forest, Gradient Boosting, KNN, and XGBoost) to determine the optimal combination for hazard perception assessment. Overall, the results show that EEG + EDA + HRV achieves the best assessment performance across all models, indicating that the complementary information from EEG, EDA, and HRV significantly enhances hazard perception assessment capability. Specifically, in the LightGBM model, $ACC^{EEG+EDA}_{LightGBM} = 89.88\%$, $ACC^{EEG+HRV}_{LightGBM} = 89.90\%$, $ACC^{EDA+HRV}_{LightGBM} = 79.61\%$, $ACC^{EEG+EDA+HRV}_{LightGBM} = 99.89\%$. In the 1D-CNN model, $ACC^{EEG+EDA}_{1D-CNN} = 88.35\%$, $ACC^{EEG+HRV}_{1D-CNN} = 87.50\%$, $ACC^{EDA+HRV}_{1D-CNN} = 72.36\%$, $ACC^{EEG+EDA+HRV}_{1D-CNN} = 99.07\%$. In the RandomForest model, $ACC^{EEG+EDA}_{RandomForest} = 88.16\%$, $ACC^{EEG+HRV}_{RandomForest} = 87.34\%$, $ACC^{EDA+HRV}_{RandomForest} = 54.90\%$, $ACC^{EEG+EDA+HRV}_{RandomForest} = 95.54\%$. In the GradientBoosting model, $ACC^{EEG+EDA}_{GradientBoosting} = 84.67\%$, $ACC^{EEG+HRV}_{GradientBoosting} = 84.01\%$, $ACC^{EDA+HRV}_{GradientBoosting} = 69.80\%$, $ACC^{EEG+EDA+HRV}_{GradientBoosting} = 94.00\%$. In the KNN model, $ACC^{EEG+EDA}_{KNN} = 88.38\%$, $ACC^{EEG+HRV}_{KNN} = 78.95\%$, $ACC^{EDA+HRV}_{KNN} = 65.14\%$, $ACC^{EEG+EDA+HRV}_{KNN} = 92.11\%$. In the XGBoost model, $ACC^{EEG+EDA}_{XGBoost} = 76.55\%$, $ACC^{EEG+HRV}_{XGBoost} = 75.48\%$, $ACC^{EDA+HRV}_{XGBoost} = 68.49\%$.

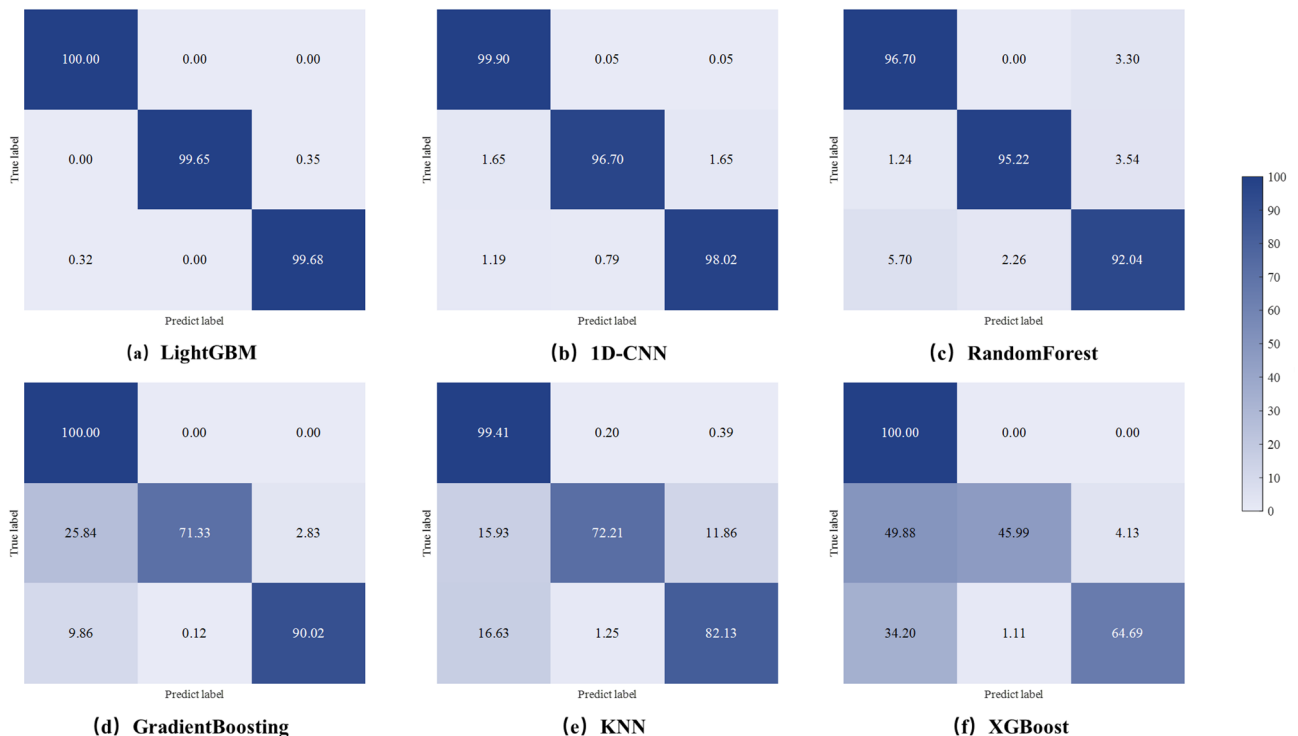


Fig. 7. Confusion matrix of six best-performing models.

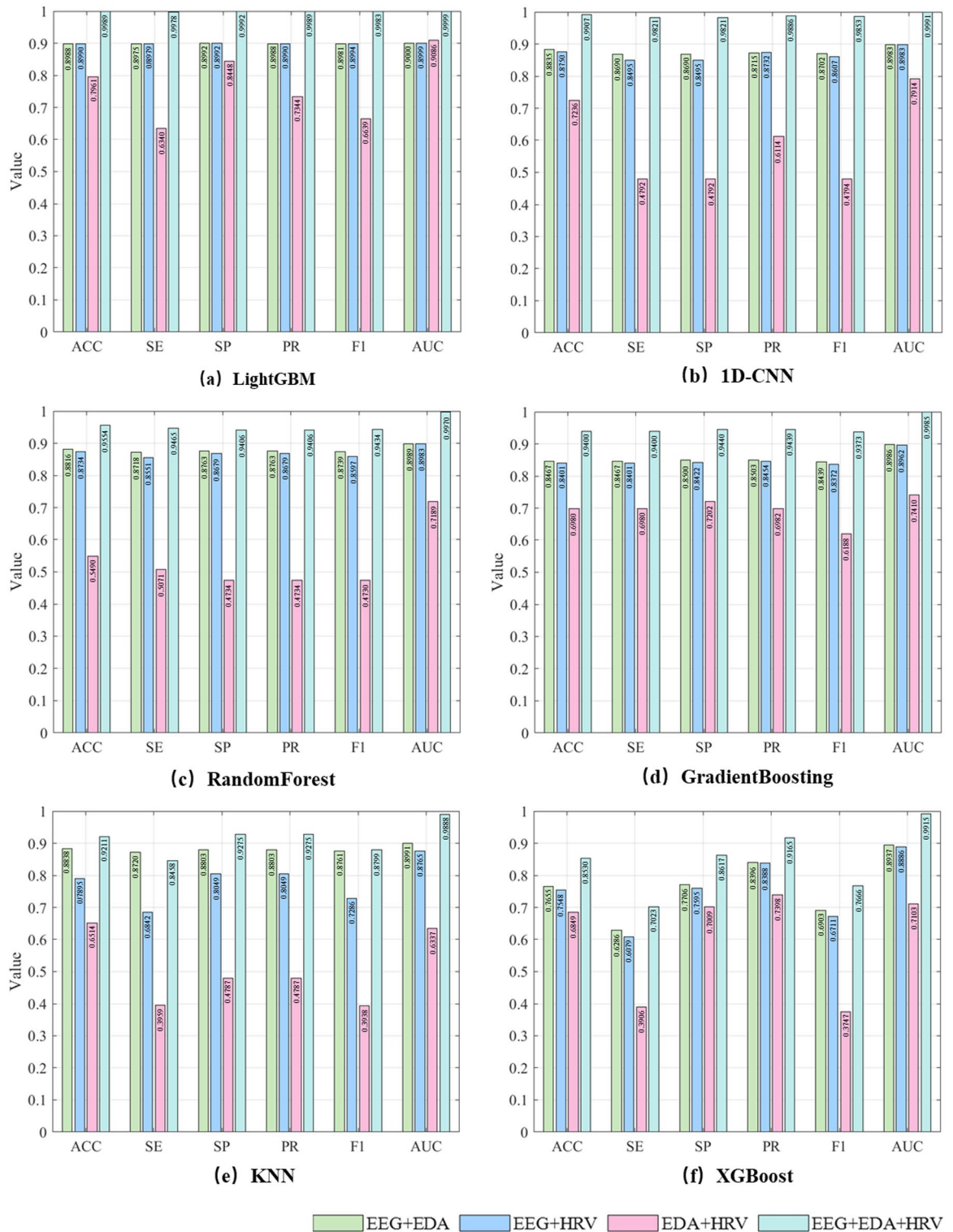


Fig. 8. Test results of different signal combinations of six best-performing models.

$ACC_{EEG+EDA+HRV}^{XGBoost} = 85.30\%$. The results indicate that among dual-signal combinations, EDA + HRV demonstrates weaker accuracy compared to other pairs, while EEG + EDA shows superior performance. When compared to dual-signal combinations, triple-signal combinations achieve the highest test accuracy of all six models. This provides crucial evidence supporting the use of multi-source physiological data for operator hazard perception assessment in this study.

References	Focus of study	Data source	Best model	Best physiological combination	Accuracy(Max)
11	Hazard identification	EEG	CatBoost	—	0.9510
51	Perceived risk	fNIRS	CatBoost	—	0.9030
30	Perceived risk	PPG, EDA, ST	Gaussian SVM	EDA + ST + PPG	0.8120
15	FFH risk	EEG, EDA, PPG, ST	CatBoost	EEG + EDA + PPG + ST	0.9980
52	Fall-risk assessment	inertial motion data	CNN-LSTM	—	0.9130
53	Emotion recognition	multisource throat physiological signals	LightGBM	—	0.9890
18	Mental workload	ECG, EDA	KNN	—	0.9310
Current study	Hazard perception	EEG, EDA, HRV	LightGBM	EEG + EDA + HRV	0.9989

Table 8. Summary of related studies using physiological signals.

Discussion

The results of this study, based on machine learning and deep learning models, demonstrate that multi-source physiological data can effectively assess the hazard perception levels of operators in safety monitoring centers. In particular, the fusion of EEG, EDA, and HRV enables high-accuracy assessment of hazard perception levels. These findings emphasize the critical role of real-time multi-source physiological data in assessing adverse operator states, thereby reducing risks and preventing accidents. As shown in Table 8, compared with previous studies, most existing research on physiological signals has focuses on cognitive load, fatigue, drowsiness, and emotion^{49,50}. The few studies addressing hazard perception and risk identification have largely concentrate on recognizing external risks rather than operators' internal hazard perception levels. Moreover, compared with multi-source physiological data fusion, single physiological signals often exhibit limitations in feature representation and fail to achieve high accuracy. Therefore, multi-source fusion not only compensates for the shortcomings of single signals but also improves model accuracy and robustness through the integration of complementary information.

This study advances the development of remote monitoring management models for high-risk industries, focusing on the niche yet critical research topic of operator hazard perception levels within intelligent coal mine safety monitoring centers. Its primary contributions are as follows: First, by training 12 models with multi-source physiological data, LightGBM achieve the best performance among all evaluated models. In addition, the study demonstrates the feasibility of real-time assessment of operators' hazard perception levels based on EEG, EDA, and HRV signals. This contributes to enhancing real-time monitoring of operator states, and when combined with the LightGBM model, enables early warnings of low hazard perception levels and attention lapses during operations, thereby reducing the risk of coal mine accidents caused by human errors and further improving mine safety. Second, the findings provide insights into the dynamic changes of physiological signals under different hazard perception levels and reveal higher sensitivity of specific brain regions during the hazard perception process. Future research can focus on physiological features of EDA and HRV that exhibit variations across different hazard perception levels, as well as the frontal F3 region of the brain. This offers a scientific reference for operator state assessment in intelligent coal mine safety monitoring centers and other high-risk operational scenarios, allowing systems to more effectively monitor critical physiological indicators and brain-region-specific signals, thereby improving the efficiency of potential hazard detection. Finally, the results provide a scientific basis for operational and safety management by supporting the construction of dynamic scheduling and risk warning systems based on real-time physiological signal monitoring. Specifically, when an operator's hazard perception level is assessed to be low, multiple dynamic adjustment measures can be implemented, including optimizing task allocation, adjusting work pace, temporarily reducing workload, or arranging short breaks, when necessary, to restore the operator's hazard perception level, thus promoting the evolution of safety management toward greater intelligence and precision. Moreover, the implications can be extended to safety management in transportation, aviation, and other high-risk industries, further advancing the development of human-machine collaborative safety management systems.

It should be noted that several challenges and limitations remain. First, although the laboratory environment replicates the actual working conditions of operators in coal mine safety monitoring centers, it may still fail to fully reflect the complexity of real coal mine operations. The collection of physiological signals may be subject to interference from external environmental factors. Therefore, before applying the proposed hazard perception intelligent platform to real-world intelligent coal mine monitoring systems, further testing and validation in real working environments are required to comprehensively evaluate its reliability and robustness. Second, this study focuses only on operators' physiological responses under different hazard perception levels. However, other dimensions, such as subjective emotions and explicit behavioral characteristics, could also serve as additional indicators of hazard perception. Future research may extend the current framework by incorporating multi-source information (e.g., facial expression recognition, emotional state analysis, eye-tracking) on top of EEG, EDA, and HRV data, thereby constructing a more comprehensive hazard perception assessment framework and enhancing the applicability and robustness of the models in complex working environments.

Conclusion

This study designs an experiment to assess the hazard perception levels of operators in coal mine safety monitoring centers, simulating real working scenario. Subjective data from the Hazard Perception Questionnaire and task performance data are collected, while EEG, EDA, and HRV data are simultaneously recorded in real time. Physiological features are extracted from each signal type, and their variations are analyzed. A total of 12 machine learning and deep learning models are applied to hazard perception level assessment, ultimately identifying the optimal assessment method based on multi-source physiological data. The main conclusions are as follows:

1. The experimental results show that changes in operators' hazard perception levels are reflected in variations of EEG, EDA, and HRV several physiological features. Corrected significant differences ($p < 0.0167$) are observed in certain HRV and EDA features across different hazard perception levels, especially between the low-moderate and low-high levels. In addition, EEG signals are crucial for assessing hazard perception levels.
2. Compared with other models, the LightGBM model demonstrates both accuracy and robustness in assessing operators' hazard perception levels. With an accuracy of 99.89%, its gradient boosting tree structure effectively captured complex nonlinear relationships within multi-source physiological data, making the model more reliable and practical for real-world applications.
3. By evaluating the performance of 6 models based on different physiological signal combinations, it is found that the EEG + EDA + HRV combination consistently achieves the best performance across all models, whereas the EDA + HRV combination yields the lowest accuracy. These results highlight the significant complementarity among multi-source physiological data and demonstrate that signal fusion substantially improves classification accuracy and model robustness.

By advancing the understanding of operators' hazard perception and contributes to the development of an intelligent platform for hazard perception assessment. Such a platform aims to effectively reduce safety risks associated with decreased hazard perception levels among operators and improve the overall safety of coal mine production. By integrating physiological signals with machine learning and deep learning models, this research provides a novel pathway for enhancing safety management in intelligent coal mines. Moreover, the proposed method can also be applied to other high-risk industries, offering safer and more intelligent management strategies.

Data availability

The datasets generated during and/or analysed during the current study are available in the <https://github.com/QAMQ1204/multi-source-physiological-git>.

Received: 25 October 2025; Accepted: 9 January 2026

Published online: 29 January 2026

References

1. Wang, G. et al. Research and practice of intelligent coal mine technology systems in China. *Int. J. Coal Sci. Technol.* **9**, 24 (2022).
2. Wang, J. & Huang, Z. The recent technological development of intelligent mining in China. *Engineering* **3**, 439–444 (2017).
3. Johnston, K. A. & Scialfa, C. T. Hazard perception in emergency medical service responders. *Accid. Anal. Prev.* **95**, 91–96 (2016).
4. Moran, C., Bennett, J. M. & Prabhakaran, P. Road user hazard perception tests: A systematic review of current methodologies. *Accid. Anal. Prev.* **129**, 309–333 (2019).
5. Dong, W. et al. Railway safety under increasing speed: train drivers' hazard perception of foreign object intrusion on railway tracks. *Int. J. Ind. Ergon.* **105**, 103684 (2025).
6. Sun, J., Chang, F., Zhou, Z. & Man, S. S. Shou Chan, A. H. A systematic review of hazard recognition and risk perception research in the construction industry. *Saf. Sci.* **186**, 106813 (2025).
7. Crundall, D., Van Loon, E., Baguley, T. & Kroll, V. A novel driving assessment combining hazard perception, hazard prediction and theory questions. *Accid. Anal. Prev.* **149**, 105847 (2021).
8. Goode, T., Kroll, V., Vernon, M., Ventsislavova, P. & Crundall, D. A comparison of cybersickness symptoms across 360-degree hazard perception and hazard prediction tests for drivers. *Appl. Ergon.* **97**, 103549 (2021).
9. Liu, Y. et al. Fusion of Spatial, Temporal, and spectral EEG signatures improves multilevel cognitive load prediction. *IEEE Trans. Human-Mach Syst.* **53**, 357–366 (2023).
10. Jeon, J. & Cai, H. Wearable EEG-based construction hazard identification in virtual and real environments: A comparative study. *Saf. Sci.* **165**, 106213 (2023).
11. Jeon, J. & Cai, H. Classification of construction hazard-related perceptions using: wearable electroencephalogram and virtual reality. *Autom. Constr.* **132**, 103975 (2021).
12. Angkan, P. et al. Multimodal Brain–Computer interface for In-Vehicle driver cognitive load measurement: dataset and baselines. *IEEE Trans. Intell. Transp. Syst.* **25**, 5949–5964 (2024).
13. Giannakakis, G. et al. Review on psychological stress detection using biosignals. *IEEE Trans. Affect. Comput.* **13**, 440–460 (2022).
14. Vos, G., Trinh, K. & Saranyai, Z. Rahimi Azghadi, M. Generalizable machine learning for stress monitoring from wearable devices: A systematic literature review. *Int. J. Med. Informatics.* **173**, 105026 (2023).
15. Duorinaah, F. X., Olatunbosun, S. O., Won, J. H., Chi, H. L. & Kim, M. K. Assessment of construction workers' fall-from-height risk using multi-physiological data and virtual reality. *Autom. Constr.* **176**, 106254 (2025).
16. Sudha, K. V. K., Jayasudha, J. S. & L. R. & Optimizing EEG - based emotion recognition with a multi-modal ensemble approach. *Results Eng.* **26**, 104886 (2025).
17. Qiu, S. et al. Multi-sensor information fusion based on machine learning for real applications in human activity recognition: State-of-the-art and research challenges. *Inform. Fusion.* **80**, 241–265 (2022).
18. Li, Z. et al. A multi-source physiological data-driven method for evaluating the mental workload of operators in remote control environment. *Saf. Sci.* **187**, 106864 (2025).
19. Zhang, X. et al. Emotion recognition from multimodal physiological signals using a regularized deep fusion of kernel machine. *IEEE Trans. Cybern.* **51**, 4386–4399 (2021).

20. Umer, W., Mehmood, I., Qarout, Y., Antwi-Afari, M. F. & Anwer, S. Deep learning-based fatigue monitoring of construction workers using physiological signals. *Autom. Constr.* **177**, 106356 (2025).
21. Wetton, M. A. et al. The development and validation of two complementary measures of drivers' hazard perception ability. *Accid. Anal. Prev.* **42**, 1232–1239 (2010).
22. White, M. J., Cunningham, L. C. & Titchener, K. Young drivers' optimism bias for accident risk and driving skill: accountability and insight experience manipulations. *Accid. Anal. Prev.* **43**, 1309–1315 (2011).
23. Bennett, J. M. et al. Development of a virtual reality pedestrian street-crossing task: the examination of hazard perception and gap acceptance. *Saf. Sci.* **181**, 106706 (2025).
24. Endriulaitienė, A. et al. Correlations among self-report, static image, and video-based hazard perception assessments: the validity of a new Lithuanian hazard prediction test. *Accid. Anal. Prev.* **173**, 106716 (2022).
25. Chong, D., Yu, A., Su, H. & Zhou, Y. The impact of emotional States on construction workers' recognition ability of safety hazards based on social cognitive neuroscience. *Front. Psychol.* **13**, 895929 (2022).
26. Ibrahim, A., Nnaji, C., Namian, M., Koh, A. & Techera, U. Investigating the impact of physical fatigue on construction workers' situational awareness. *Saf. Sci.* **163**, 106103 (2023).
27. Khare, S. K., Blanes-Vidal, V., Nadimi, E. S. & Acharya, U. R. Emotion recognition and artificial intelligence: A systematic review (2014–2023) and research recommendations. *Inform. Fusion.* **102**, 102019 (2024).
28. Lavie, N., Beck, D. M. & Konstantinou, N. Blinded by the load: attention, awareness and the role of perceptual load. *Phil Trans. R Soc. B.* **369**, 20130205 (2014).
29. Liang, B. & Lin, Y. Using physiological and behavioral measurements in a picture-based road hazard perception experiment to classify risky and safe drivers. *Transp. Res. Part. F: Traffic Psychol. Behav.* **58**, 93–105 (2018).
30. Lee, B. G., Choi, B., Jebelli, H. & Lee, S. Assessment of construction workers' perceived risk using physiological data from wearable sensors: A machine learning approach. *J. Building Eng.* **42**, 102824 (2021).
31. Hao, T. et al. Towards mental load assessment for high-risk works driven by Psychophysiological data: combining a 1D-CNN model with random forest feature selection. *Biomed. Signal Process. Control.* **96**, 106615 (2024).
32. Liao, P. C., Sun, X. & Zhang, D. A multimodal study to measure the cognitive demands of hazard recognition in construction workplaces. *Saf. Sci.* **133**, 105010 (2021).
33. Wang, J., Liang, M. & Liao, P. C. Toward an intuitive device for construction hazard recognition management: eye Fixation-Related potentials in reinvestigation of hazard recognition performance prediction. *J. Constr. Eng. Manage.* **150**, 04024028 (2024).
34. Xu, Q., Chong, H. Y. & Liao, P. Exploring eye-tracking searching strategies for construction hazard recognition in a laboratory scene. *Saf. Sci.* **120**, 824–832 (2019).
35. Zhang, C., Wang, P., Wang, E., Chen, D. & Li, C. Characteristics of coal resources in China and statistical analysis and preventive measures for coal mine accidents. *Int. J. Coal Sci. Technol.* **10**, 22 (2023).
36. Faul, F., Erdfelder, E., Buchner, A. & Lang, A. G. Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods.* **41**, 1149–1160 (2009).
37. Khadir, A., Maghareh, M., Sasani Ghamsari, S. & Beigzadeh, B. Brain activity characteristics of RGB stimulus: an EEG study. *Sci. Rep.* **13**, 18988 (2023).
38. Shirota, A. et al. Discrepancies in the time course of sleep stage Dynamics, electroencephalographic activity and heart rate variability over sleep cycles in the adaptation night in healthy young adults. *Front. Physiol.* **12**, 623401 (2021).
39. Banani Ardecani, F. & Shoghli, O. Assessing workers' neuro-physiological stress responses to augmented reality safety warnings in immersive virtual roadway work zones. *Autom. Constr.* **180**, 106565 (2025).
40. Parés-Pujolràs, E., Kim, Y. W., Im, C. H. & Haggard, P. Latent awareness: early conscious access to motor Preparation processes is linked to the readiness potential. *NeuroImage* **202**, 116140 (2019).
41. Kroenke, K., Spitzer, R. L. & Williams, J. B. W. The PHQ-9: validity of a brief depression severity measure. *J. Gen. Intern. Med.* **16**, 606–613 (2001).
42. Delorme, A. & Makeig, S. EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods.* **134**, 9–21 (2004).
43. Jap, B. T., Lal, S., Fischer, P. & Bekiaris, E. Using EEG spectral components to assess algorithms for detecting fatigue. *Expert Syst. Appl.* **36**, 2352–2359 (2009).
44. Sriram Kumar, P., Govarthan, P. K., Shaik Gadda, A. & Ganapathy, A. Fredo agastinose Ronickom, J. Deep Learning-Based automated emotion recognition using multimodal physiological signals and Time-Frequency methods. *IEEE Trans. Instrum. Meas.* **73**, 1–12 (2024).
45. Chikhi, S., Matton, N. & Blanchet, S. EEG power spectral measures of cognitive workload: A meta-analysis. *Psychophysiology* **59**, e14009 (2022).
46. Savage, S. W., Potter, D. D. & Tatler, B. W. The effects of cognitive distraction on behavioural, oculomotor and electrophysiological metrics during a driving hazard perception task. *Accid. Anal. Prev.* **138**, 105469 (2020).
47. Bhaduri, D., Toth, D. & Holan, S. H. A review of Tree-Based methods for analyzing survey data. *WIREs Comput. Stats.* **17**, e70010 (2025).
48. Kiranyaz, S. et al. 1D convolutional neural networks and applications: A survey. *Mech. Syst. Signal Process.* **151**, 107398 (2021).
49. Çetinkaya, M. & Acarman, T. Driver impairment detection using decision tree based feature selection and classification. *Results Eng.* **18**, 101025 (2023).
50. Jiang, W., Ma, C., Xu, M., Li, Y. & Zhou, Q. Investigation of the correlation between multiple physiological signals in the lower limbs during long-term driving. *Results Eng.* **28**, 107295 (2025).
51. Lee, K., Pooladvand, S., Esmaili, B. & Hasanzadeh, S. Understanding construction workers' risk perception using neurophysiological responses. *J. Comput. Civ. Eng.* **38**, 04024039 (2024).
52. Kou, J., Xu, X., Ni, X., Ma, S. & Guo, L. Fall-risk assessment of aged workers using wearable inertial measurement units based on machine learning. *Saf. Sci.* **176**, 106551 (2024).
53. Mao, J. H. et al. Machine Learning-Enabled emotion recognition by multisource throat signals. *ACS Nano.* **19**, 18397–18408 (2025).

Acknowledgements

The authors thank the support given by the Humanities and Social Sciences Research Fund of the Ministry of Education of the People's Republic of China (Grant No. 23YJA630054).

Author contributions

Conceptualization, AQ, TL; methodology, AQ, YH; software, HL, YG; validation, AQ, YY; formal analysis, AQ; investigation, HL, BL, MY; data curation, HL; writing—original draft preparation, AQ; writing—review and editing, HL, JL, MY, TL, YH, YY, YG; visualization, HL; supervision, JL; funding acquisition, JL; All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported by the Humanities and Social Sciences Research Fund of the Ministry of Education of the People's Republic of China (Grant No. 23YJA630054). The authors would like to thank for their support.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to A.Q. or J.L.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2026