aerospace

MDPI

Article

Research on Pilot Workload Identification Based on EEG Time Domain and Frequency Domain

Weiping Yang [1], Yixuan Li [2], Lingbo Liu [3], Haiqing Si [3,*], Haibo Wang [3,*], Ting Pan [2], Yan Zhao [3] and Gen Li [2]

[1] Aviation Industries Corporation of China Xi'an Flight Automatic Control Research Institute, Xi'an 710065, China; yangwp001@avic.com
[2] College of Civil Aviation, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China; liyixuan@nuaa.edu.cn (Y.L.); nuaapanting@nuaa.edu.cn (T.P.)
[3] College of General Aviation and Flight, Nanjing University of Aeronautics and Astronautics, Liyang 213300, China; liulingbo@nuaa.edu.cn (L.L.)
* Correspondence: sihaiqing@126.com (H.S.); nhwanghaibo@nuaa.edu.cn (H.W.)

Abstract

Pilot workload is a critical factor influencing flight safety. This study collects both subjective and objective data on pilot workload using the NASA-TLX questionnaire and electroencephalogram acquisition systems during simulated flight tasks. The raw EEG signals are denoised through preprocessing techniques, and relevant EEG features are extracted using time-domain and frequency-domain analysis methods. One-way ANOVA is employed to examine the statistical differences in EEG indicators under varying workload levels. A fusion model based on CNN-Bi-LSTM is developed to train and classify the extracted EEG features, enabling accurate identification of pilot workload states. The results demonstrate that the proposed hybrid model achieves a recognition accuracy of 98.2% on the test set, confirming its robustness. Additionally, under increased workload conditions, frequency-domain features outperform time-domain features in discriminative power. The model proposed in this study effectively recognizes pilot workload levels and offers valuable insights for civil aviation safety management and pilot training programs.

Keywords: pilot workload identification; EEG data analysis; fusion model; CNN-Bi-LSTM model; pilot cognitive state

1. Introduction

With the rapid growth of the aviation industry, understanding human performance, cognitive processes, and physiological factors has become essential for improving pilot training and flight safety. Operating a civil aircraft involves a complex, multitasking environment that imposes varying levels of workload on pilots. According to statistics, over 60% of commercial aviation accidents are attributed to human error [1], primarily due to workload imbalances during flight operations [2]. Excessive workload can result in pilot fatigue, spatial disorientation, and delayed reactions, thereby increasing the risk of operational errors [3]. In response, researchers are focused on improving the safety of the human–machine–environment system to mitigate the consequences of pilot performance degradation. A critical step in this effort is the systematic collection of both subjective and objective indicators of pilot workload and the development of accurate, effective models for workload identification.

Check for updates

Academic Editor: Miroslav Kelemen

Received: 16 December 2025
Revised: 16 January 2026
Accepted: 21 January 2026
Published: 23 January 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1.1. Literature Review

Flying is a human–machine–environment interaction that primarily involves mental activity, combining both cognitive and physical tasks [4]. During flight operations, pilots must process external information received through sensory input and coordinate physical actions to maintain stable aircraft control. Pilot workload refers to the physiological and psychological demands placed on pilots per unit of time during flight. Pilot workload is a multidimensional construct that reflects the overall demands imposed on pilots during flight operations. It typically encompasses physiological demands, psychological demands, and cognitive load. Physiological demands refer to bodily responses such as neural and cardiovascular activity. Psychological demands describe subjective experiences including perceived stress, frustration, and effort. Cognitive load, in contrast, specifically refers to the mental resources required for information processing, attention allocation, decision-making, and task management. In this study, pilot workload is treated as an integrative concept, within which cognitive load represents a core component that is indirectly reflected through EEG features and subjective NASA-TLX assessments. Both excessive and insufficient workload levels can pose risks to flight safety. In recent years, with advances in ergonomics and aviation psychology, researchers worldwide have explored pilot workload monitoring through both subjective and objective measures.

Subjective assessment remains the most widely used method for evaluating pilot workload. Since Cooper et al. [5] introduced the Aircraft Operating Characteristics Scale (Cooper–Harper Scale), several prominent institutions have developed additional tools: the U.S. Air Force School of Aerospace Medicine [6] created the Subjective Workload Assessment Technique (SWAT), the Royal Aeronautical Establishment [7] proposed the Bedford Workload Rating Scale, and NASA developed the NASA [8] Task Load Index (NASA-TLX). Wang et al. applied the Bedford Scale to assess the subjective workload of flight trainees performing traffic pattern tasks, demonstrating its advantage in evaluating time pressure [9]. Shihan Luo et al. utilized the NASA-TLX to measure the subjective psychological workload of flight trainees operating the Cessna 21 simulator under three conditions: daytime crosswind, nighttime crosswind, and daytime no-wind scenarios [10]. However, due to the inherent subjectivity of scale-based assessments, most researchers now incorporate objective data to more comprehensively evaluate pilot workload.

In the objective assessment of pilot workload, Zhang et al. developed the Mental Workload Assessment–Information Theory (MWA-IT) model to estimate peak pilot workload under specific scenarios [11]. This model integrates eye movement data and applies information theory, considering factors such as flight experience, environmental conditions, and task complexity to minimize the adverse effects of excessive workload. Othman et al. investigated the correlation between pupil dilation and pilot workload during human–computer interaction using an eye tracker and the NASA Task Load Index [12]. Their findings revealed that increased workload was associated with a significant decrease in the standard deviation of pupil diameter and a prolonged reaction time, suggesting that pupil dynamics can serve as a reliable workload indicator. To monitor workload and enhance pilot performance, Alaimo et al. employed electrocardiogram (ECG) sensors to record heart rate across different flight stages [13]. Their results indicated that elevated workload and stress levels corresponded with changes in heart rate, supporting its use as a workload assessment metric. Socha et al. explored the influence of cockpit layout changes on novice pilots' workload and aircraft handling ability, using heart rate variability (HRV) as the primary physiological indicator [14]. The results showed that abrupt modifications to cockpit displays significantly increased pilot workload during early training. Mohanavelu et al. analyzed pilot workload under varying visibility conditions using ECG-based physiological indicators [15]. They found that HRV features varied notably across flight phases and

conditions, reflecting workload changes. After a comprehensive comparison of multiple physiological signals, including eye tracking and ECG, Houssein et al. concluded that EEG signals are more sensitive and reliable for detecting variations in pilot workload [16].

EEG signals offer high temporal resolution—accurate to the millisecond—enabling rapid and direct reflection of changes in brain states [17]. Ji et al. analyzed EEG features of pilots during left and right turns and employed support vector machines (SVM) to classify their psychological workload [18]. The results showed a significant increase in workload during turning maneuvers compared to the cruising phase. Cho et al. evaluated EEG spectral features of 12 pilots operating an Airbus A320 flight simulator under fully automatic, semi-automatic, and manual control modes [19]. Their findings indicated that cognitive workload was highest during manual flight. Zhang et al. examined EEG data collected from pilots flying with and without a head-up display (HUD) [20]. The results demonstrated that HUD usage helped reduce pilots' psychological workload. Verkennis et al. collected EEG data from 52 subjects performing flight tasks in a simulator and classified them based on NASA-TLX evaluations [21]. Their study emphasized the critical role of parietal directional connectivity in distinguishing workload levels. Wang et al. proposed a local and global network (LGNet) to classify two levels of cognitive workload using EEG data collected during simulated flight, achieving an average classification accuracy of 91.19% [22]. Compared with studies that explicitly focus on cognitive load identification, such as Wang et al., the present work adopts a broader workload framework that integrates cognitive, physiological, and subjective dimensions, while still emphasizing EEG-based indicators closely related to cognitive processing. Lee et al. extracted EEG signal features from simulated flight tasks and input them into a spatiotemporal convolutional neural network (CNN) model, achieving a workload classification accuracy exceeding 86% [23].

1.2. Contribution

In summary, subjective rating scales have irreplaceable advantages in characterizing pilots' perceived workload and subjective experience; however, they cannot support real-time monitoring. Behavioral indicators can reflect task performance, but they are easily influenced by task design and environmental factors. Physiological indicators, by contrast, are able to capture subtle workload-related changes even before performance degradation occurs and are therefore regarded as a key pathway for developing adaptive human–machine systems [24]. Nevertheless, most existing studies focus on a single physiological indicator or a single flight scenario. Systematic comparisons of EEG time-domain and frequency-domain features across different workload levels remain limited, and the correspondence between subjective workload ratings and EEG feature patterns has not yet been sufficiently established.

Accordingly, at the levels of experimental design and data collection, this study constructed a simulated standard five-leg flight task that incorporates three typical flight scenarios: clear weather, heavy fog, and single-engine failure. By manipulating visual conditions and task complexity, pilots were induced into low, medium, and high workload states. The NASA Task Load Index (NASA-TLX) was employed to subjectively grade workload levels for each scenario, thereby obtaining an EEG dataset with explicitly labeled workload states. The design of this dataset balances the ecological validity of flight tasks with the controllability of psychological workload, providing an experimental foundation for systematic analysis of pilot workload. At the feature analysis level, this study systematically extracted multiple EEG time-domain and frequency-domain features and compared their distribution differences under three workload levels at both whole-brain and brain-region scales. Feature combinations that are significantly sensitive to workload variations were identified, and the cognitive processing and alertness regulation

mechanisms reflected by these features were further discussed. At the model construction level, this study proposes a workload recognition model based on a hybrid convolutional neural network and bidirectional long short-term memory network (CNN–BiLSTM). Statistically selected time-domain and frequency-domain features were used as model inputs, and multi-channel feature representations were constructed according to different brain regions. By jointly leveraging spatial distribution characteristics and temporal sequence dependencies, the proposed model classifies low, medium, and high workload levels and explores the underlying neural mechanisms by which different brain regions contribute to workload modulation.

2. Methods

This section describes the experimental protocol and data processing methods. Sections 2.1 and 2.2 present the detailed implementation of the pilot workload experiment and the procedures for acquiring subjective and objective data. Sections 2.3–2.5 describe the processing and analysis workflows applied to the collected subjective and objective data. Finally, Sections 2.6–2.9 provide the specific structure and training parameters of the CNN–BiLSTM model used for pilot workload recognition in this study.

2.1. Subjects

A total of 33 participants were recruited for this study, all of whom were male flight trainees from the School of General Aviation and Flight at Nanjing University of Aeronautics and Astronautics. Participants ranged in age from 19 to 24 years, were right-handed, and had normal vision (E-chart 5.0 or better) and normal hearing. Prior to participating in the experiment, all participants had completed 72 h of flight control training on a Cessna 172 flight simulator and had achieved at least third prize in national flight simulation competitions. To ensure the reliability of the experimental data, participants were instructed to abstain from alcohol, caffeine, and other substances that might affect central nervous system function on the day prior to the experiment and to maintain a well-rested mental state. All participating pilots were fully informed of the study's objectives, procedures, and methodologies prior to the experiment. Each participant provided written informed consent, confirming their voluntary involvement and authorizing the use of their data for research purposes. To ensure participant confidentiality, all collected data were anonymized. Furthermore, data were securely stored on a protected server with strict access controls and appropriate security measures in place to prevent unauthorized access, data leakage, or misuse.

This experiment was conducted in accordance with the principles of the Declaration of Helsinki and was approved by the Institutional Review Board of the School of General Aviation and Flight at Nanjing University of Aeronautics and Astronautics.

2.2. Experiments

2.2.1. Equipment

This experiment utilized a Cessna 172 flight simulator equipped with a Garmin G1000 avionics system (Garmin Ltd., Olathe, KS, USA) to replicate a realistic flight control environment. Flight control operations included the use of the flight control panel, throttle lever, mixture control lever, and pitch trim lever, with flight tasks completed by referencing standard flight instruments. Pilot workload data were collected using a semi-dry EEG cap with water-based electrodes, capable of multi-channel EEG acquisition with high temporal resolution. EEG data collection was synchronized with task events using the ErgoLAB human–machine interaction platform, enabling real-time recording and annotation of EEG signals. The experimental setup is illustrated in Figure 1.

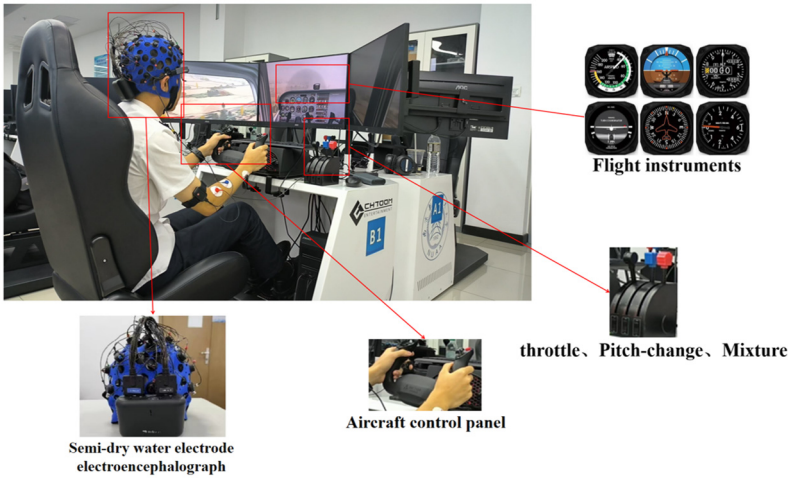
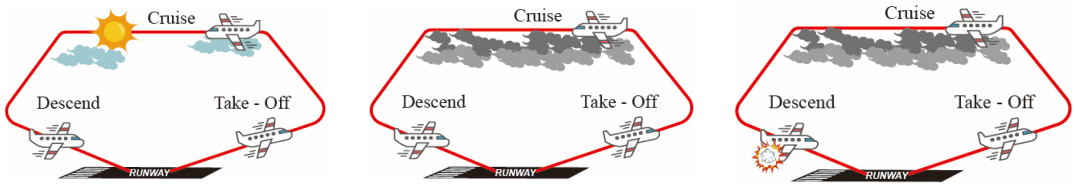


Figure 1. EEG acquisition equipment and simulated flight laboratory.

2.2.2. Experimental Design

This study designed three representative simulated flight scenarios, as illustrated in Figure 2. By varying visual meteorological conditions and task complexity, pilots were induced to experience different levels of workload—categorized as low, medium, and high. The experimental task followed the "standard five-leg flight" pattern, which includes key flight phases: takeoff, climb, cruise, descent, and landing. Each participant completed one full five-leg mission under each scenario. The selected flight conditions were: clear weather (dominant visibility greater than 10 km with light cumuliform cloud cover) and heavy fog (dominant visibility of 800 m with extensive stratus cloud cover). In the single-engine failure scenario, the pilot was required to execute high-difficulty tasks such as gliding and landing without engine power, resulting in a high workload state. This engine failure scenario was implemented under foggy conditions to further induce maximum cognitive and operational demand.



(**a**) Clear weather (**b**) Heavy fog (**c**) Single-engine failure

Figure 2. Simulated flight experiment scene.

Before the formal experiment, all participants received training on the standard five-leg flight pattern. They were instructed to obtain at least 8 h of sleep on the night before the experiment and to participate in the experiment between 9:00 and 11:00 a.m. on the following day.

The entire experiment lasted approximately 45 min. First, each participant sat quietly for 5 min to reduce potential physiological and psychological disturbances caused by prior physical activity. The participant then performed the low-workload task (clear-weather flight scenario), which lasted about 10 min. Upon completion of the task, the participant completed the NASA-TLX questionnaire separately for the takeoff, downwind, and final approach phases. Next, the participant sat quietly for another 5 min to wash out the effect of the low-workload task. The participant then performed the medium-workload task (heavy-fog flight scenario), which also lasted about 10 min. After this task, the NASA-TLX was again completed separately for the takeoff, downwind, and final approach phases.

Finally, the participant sat quietly for 5 min to wash out the effect of the medium-workload task, and then performed the high-workload task (single-engine-failure flight scenario), which lasted about 10 min. At the end of this task, the participant once more completed the NASA-TLX questionnaire for the takeoff, downwind, and final approach phases.

Because administering the NASA-TLX during flight execution may interrupt task performance, and task interruption and changes in participants' emotional state can have negative effects [25,26], the NASA-TLX questionnaires for the takeoff, downwind, and final approach phases were all completed immediately after the corresponding flight tasks had been finished.

2.3. Data Collection

In this study, a flight simulation experiment was conducted to acquire NASA-TLX ratings, eye-movement data, electrocardiogram (ECG), respiration, and other physiological data from 33 flight trainees. The ratio between the training set and the test set was approximately 75% to 25%. Specifically, data from 25 trainees were randomly assigned to the training set, and data from the remaining 8 trainees were assigned to the test set. The composition of the dataset is summarized in Table 1.

Table 1. Dataset schedule.

Dataset	Label	Clear-Weather Condition	Heavy-Fog Condition	Single-Engine-Failure Condition
Training	Low workload	25	25	25
	Medium workload	25	25	25
	High workload	25	25	25
Test	Low workload	8	8	8
	Medium workload	8	8	8
	High workload	8	8	8

2.4. Data Preprocessing

2.4.1. Subjective Data Preprocessing

This study adopted the NASA Task Load Index (NASA-TLX), developed by the National Aeronautics and Space Administration, to subjectively assess participant workload. NASA-TLX is a multidimensional subjective workload assessment tool with good reliability and validity, capable of comprehensively evaluating an individual's subjective experience during task execution across multiple psychophysical dimensions. The scale comprises six dimensions: mental demand, physical demand, temporal demand, performance, effort, and frustration. After each experimental session, each participant was required to immediately rate the completed flight task on all six dimensions and to perform pairwise comparisons among the dimensions to assign weights to each factor. A weighted overall workload score was then calculated. For each participant, comprehensive workload scores for the takeoff, downwind, and final approach phases were calculated under low, medium, and high workload conditions, as illustrated in Figure 3.

It should be noted that workload and fatigue are conceptually distinct. Workload reflects the immediate demands imposed on an individual during task execution, whereas fatigue represents a cumulative and longer-term state resulting from prolonged or sustained high workload. Therefore, in this study, the NASA-TLX was used to assess immediate task-related workload associated with the flight tasks, rather than pilots' fatigue. Because the NASA-TLX cannot be administered during the execution of flight tasks, participants were required to complete the questionnaire immediately after each simulated flight to ensure accurate recall of their perceived workload.

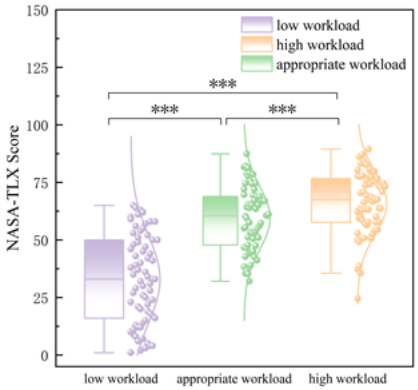


Figure 3. NASA-TLX Workload Rating Results (*** indicates statistical significance at $p < 0.001$).

As shown in Figure 3, the subjective workload scores increase significantly with the rising difficulty of the flight tasks. The lowest scores are observed under the low-load condition, followed by moderate scores under the medium-load condition, and the highest scores appear under the high-load condition, demonstrating a clear hierarchical distribution of perceived workload.

2.4.2. EEG Data Preprocessing

As a low-amplitude, easily disturbed, and non-invasive physiological signal, EEG is highly susceptible to contamination by various physiological and non-physiological artifacts during acquisition, necessitating careful preprocessing. To extract more representative and stable EEG features, this study adopts a multi-stage preprocessing pipeline that integrates signal filtering, independent component analysis (ICA) for artifact removal, and wavelet packet denoising.

(1) Signal filtering

To effectively retain neural activity within the target frequency bands while suppressing irrelevant noise, this study applied finite impulse response (FIR) band-pass filtering using the ErgoLAB platform (Kingfar International Inc., Beijing, China). The high-pass cutoff frequency was set to 2 Hz, and the low-pass cutoff frequency was set to 60 Hz, ensuring the preservation of critical information in the δ (0.5–4 Hz), θ (4–8 Hz), α (8–13 Hz), and β (13–30 Hz) frequency bands. In addition, a notch filter at 50 Hz was employed to remove power-line interference from the power supply system, thereby improving the signal-to-noise ratio and overall signal stability. Figure 4 illustrates the EEG waveforms before and after filtering. After preprocessing, the signals exhibit reduced fluctuations and enhanced rhythmicity.

(2) ICA anti-counterfeiting

To further enhance data purity, this study applies independent component analysis (ICA) to perform blind source separation on EEG signals. ICA is a signal unmixing algorithm that maximizes statistical independence between components. It decomposes mixed signals into a set of statistically independent source signals and is widely used for signal denoising and feature extraction in EEG processing.

Within the ErgoLAB platform, the filtered EEG data were decomposed into 31 independent components (ICs), and each component was automatically identified and classified using the ICLabel lug-in (version 1.3). The classification threshold for ICLabel was set at 80%; components with a probability greater than 80% of being non-EEG sources were removed. The remaining valid EEG components were retained and used to reconstruct the cleaned EEG signal. Figure 5 displays the EEG waveforms and IC distributions before and

after ICA processing, illustrating effective artifact separation and marked improvements in signal rhythmicity and morphology.

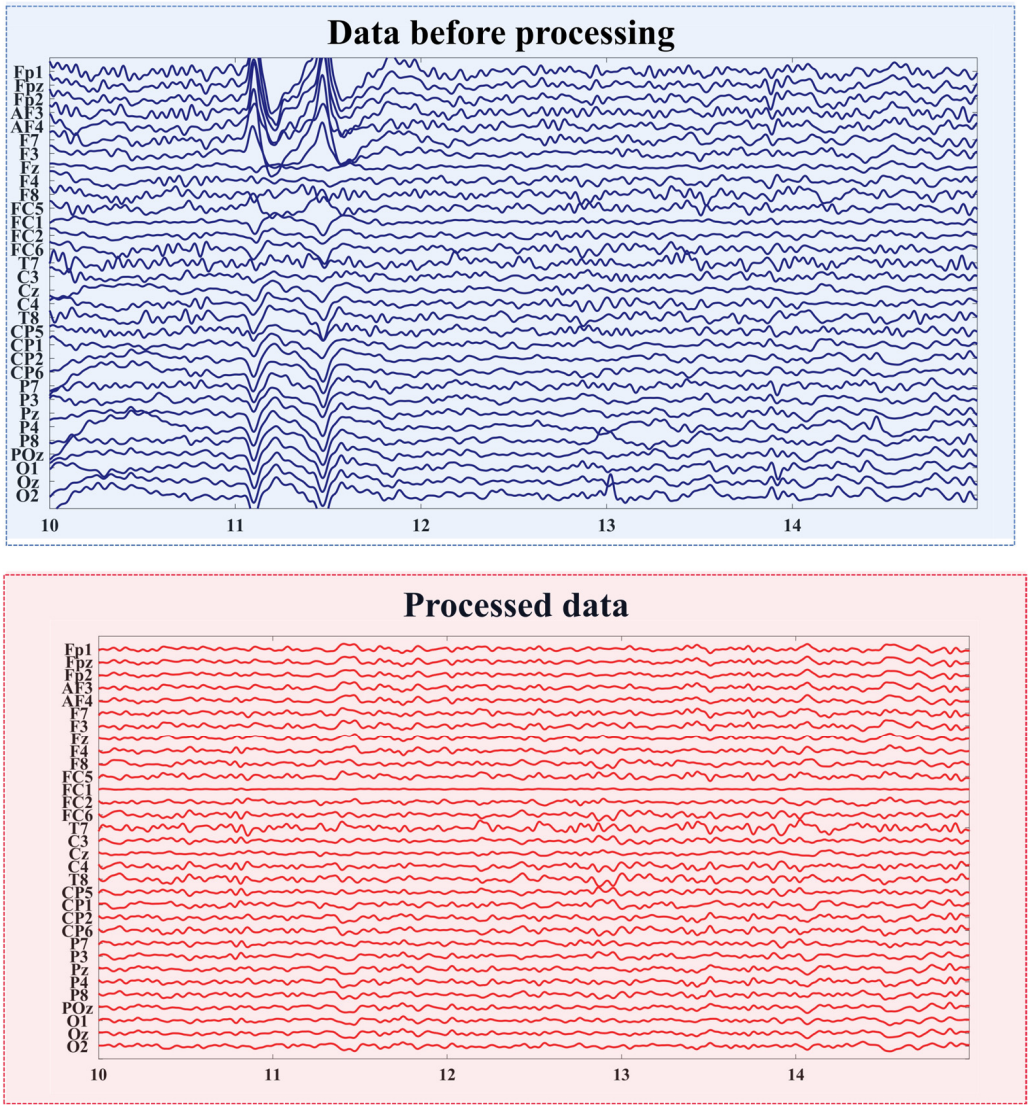


Figure 4. EEG signal filtering.

(3) Wavelet packet denoising

After removing systematic artifacts and independent interference sources through frequency-domain filtering and ICA, residual high-frequency noise—caused by environmental disturbances or poor electrode contact—may still be present in the EEG signals. To address this, the present study introduces the Wavelet Packet Transform (WPT) technique for further denoising. Based on the theory of wavelet multiscale decomposition, WPT enables simultaneous refinement and decomposition of both low- and high-frequency signal components across multiple scales. Compared to traditional wavelet transform methods, which recursively decompose only the low-frequency components, WPT offers a significant advantage in preserving local discontinuities and high-frequency transient features. This makes it particularly suitable for processing non-stationary signals such as EEG, which exhibit strong mid-frequency variability and complex rhythmic patterns. The denoising effect is illustrated in Figure 6.

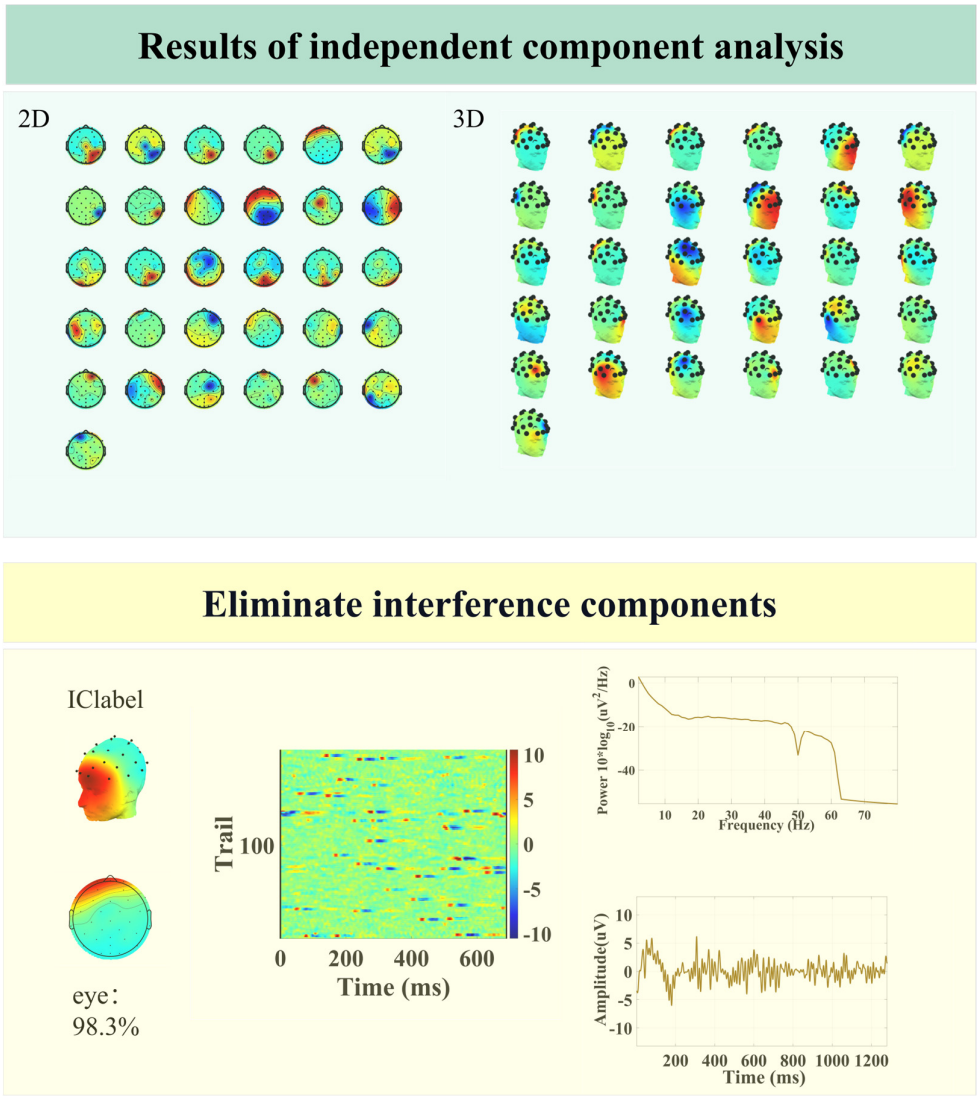


Figure 5. Independent component analysis results.

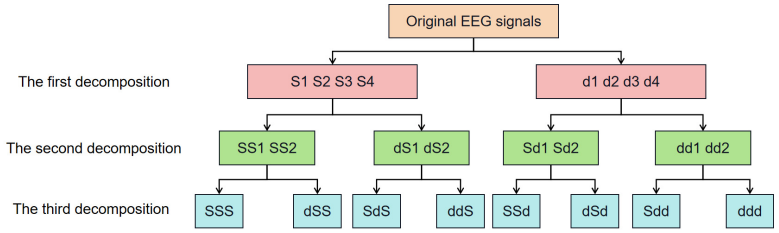


Figure 6. Wavelet packet multi-layer recursive decomposition.

This study employs the built-in wpdec function in MATLAB (R2021a, The MathWorks, Natick, MA, USA) to perform a three-level wavelet packet decomposition of the EEG signal. Default wavelet basis functions are used to construct the decomposition tree, ultimately dividing the original signal into eight subbands containing both low- and high-frequency components. To determine an appropriate denoising threshold, the ddencmp function is applied to automatically calculate the optimal thresholding strategy. This is followed by the use of the wdencmp function to perform threshold-based denoising on the decomposed signal. To avoid signal distortion associated with hard thresholding, a soft thresholding approach is adopted. The results are illustrated in Figure 7. Compared to the original noisy signal, the denoised waveform exhibits a marked reduction in high-frequency noise

amplitude while preserving the underlying neural rhythmic features, thereby improving overall signal smoothness and interpretability.

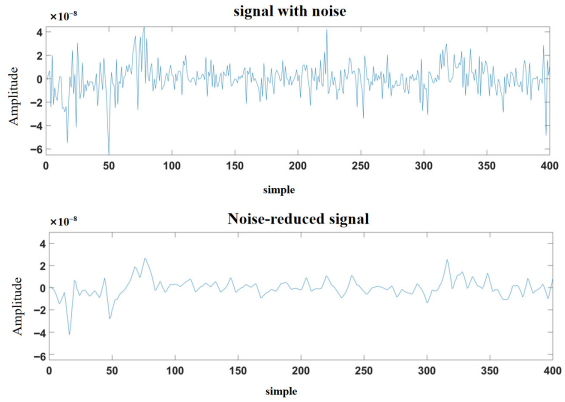


Figure 7. Comparison of before and after wavelet packet denoising.

2.5. Feature Extraction of EEG Data

EEG signals were continuously recorded throughout the entire experiment, starting from the first 5 min resting baseline period and ending after the completion of the final high-workload flight task. To ensure comparability across different workload conditions, the EEG data from the three 10 min flight tasks were selected as the source for workload feature extraction and classification modeling. The low-workload condition corresponded to the EEG recorded during the entire standard five-leg flight under clear-weather conditions; the medium-workload condition corresponded to the EEG recorded during the entire standard five-leg flight under heavy-fog conditions; and the high-workload condition corresponded to the EEG recorded during the entire single-engine-failure five-leg flight under heavy-fog conditions. The engine failure event was pre-programmed in the simulator script and was automatically triggered during the third leg of the flight. Therefore, the EEG segment for the high-workload condition captured the complete process in which the pilot transitioned from normal flight to engine-failure detection, decision-making, and emergency handling. The three 5 min resting baseline EEG segments were primarily used for baseline comparison and signal quality assessment and were not included in the training or testing of the workload recognition models.

2.5.1. Time Domain Analysis

The time-domain characteristics of EEG signals involve analyzing the signal's temporal sequence to extract instantaneous features along the time axis. These characteristics intuitively reflect the behavioral patterns and dynamic properties of EEG activity and are commonly used to assess signal stability and temporal structure. In this study, five standard time-domain metrics are employed as feature indicators: root mean square (RMS), waveform factor, peak factor, pulse factor, and margin factor, as defined in Equation (1).

$$\begin{cases} RMS = \sqrt{\dfrac{1}{N}\sum\limits_{i=1}^{N} x_i^2} \\ S = \dfrac{x_{rms}}{x_{arv}} \\ CF = \dfrac{x_{peak}}{x_{rms}} \\ f = \dfrac{X_{peak}}{X_{rms}} \\ L = \dfrac{x_{peak}}{\sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2}} \end{cases} \quad (1)$$

where RMS denotes the root mean square of the EEG signal, N represents the number of sampling points, x_i is the amplitude of the EEG signal at the i-th sampling point, S is the waveform factor, X_{rms} is the root mean square of the signal, X_{arv} is the rectified average value of the signal, CF is the peak factor, X_{peak} is the maximum instantaneous value (i.e., the peak) within a given time window, f is the pulse factor, and L is the margin factor.

2.5.2. Frequency Domain Analysis

While time-domain analysis reveals the transient characteristics of EEG signals over time, it does not intuitively reflect the energy distribution or dynamic variations in neural activity across different frequency bands. To capture the spectral characteristics of pilots' neural rhythms under varying workload conditions, this study further analyzes the preprocessed EEG signals in the frequency domain using power spectral density (PSD). PSD quantifies the average distribution of signal power across frequency bands and serves as a key indicator for evaluating EEG rhythmic energy. This study adopts the Welch's averaged periodogram method, which applies segmented discrete Fourier transforms (DFTs) to the signal and averages the resulting periodograms to reduce spectral variance, as expressed in Equation (2).

$$\begin{cases} P_i(k) = \frac{1}{M}|X_i(k)|^2 \\ PSD(k) = \frac{1}{L}\sum_{i=1}^{L} P_i(k) \end{cases} \quad (2)$$

where $P_i(k)$ denotes the periodogram of the i-th segment, M is the length of each segment, $X_i(k)$ represents the discrete Fourier transform (DFT) result of the i-th segment, L is the total number of segments into which the signal is divided, and $\frac{1}{L}$ is the averaging factor used to compute the final power spectral density estimate by averaging the PSDs of all segments.

2.6. Convolutional Neural Networks

The convolutional neural network (CNN) is a deep learning model specifically designed to process data with a grid-like structure. The convolutional layer is the core component of CNNs; it operates by sliding a set of learnable convolutional kernels across the input data to perform local connections and weighted summations, thereby generating feature maps. These layers are capable of identifying spatial patterns and extracting localized features from the input. The pooling layer subsequently downsamples the feature maps to reduce dimensionality, lower computational complexity, and retain the most salient features. The fully connected layer applies a linear transformation using weights and biases and introduces nonlinearity via activation functions, enabling global analysis and decision-making based on the features extracted by the convolution and pooling layers. The structural diagram of the CNN architecture is shown in Figure 8.

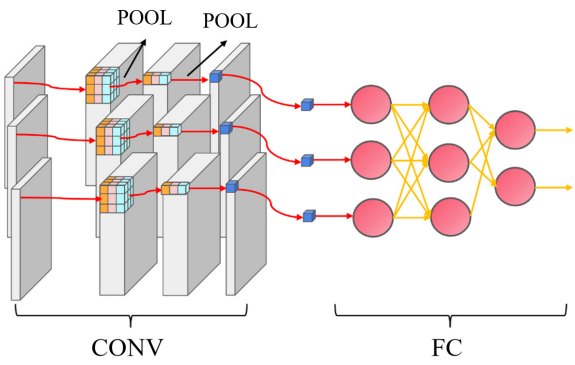


Figure 8. CNN Schematic.

2.7. Bidirectional Long Short-Term Memory Network

The bidirectional long short-term memory (Bi-LSTM) network is a neural network model designed for processing time series data with strong predictive performance. A Bi-LSTM consists of two independent LSTM layers that process the input sequence in forward and reverse directions, respectively. The outputs from both directions are concatenated to form a combined hidden representation. This architecture allows the model to access both past and future contextual information at each time step, making it well-suited for capturing long-term dependencies in sequential data. Furthermore, the bidirectional information flow enables Bi-LSTM to more effectively capture contextual relationships within the sequence, thereby enhancing model performance. The structure of the Bi-LSTM model is illustrated in Figure 9.

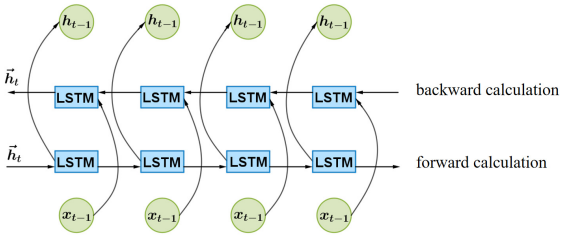


Figure 9. Bi-LSTM Schematic.

2.8. CNN-Bi-LSTM Model Construction

The CNN-Bi-LSTM model integrates the strengths of convolutional neural networks (CNN) and bidirectional long short-term memory networks (Bi-LSTM) to create a more powerful deep learning architecture. CNNs excel at extracting spatial features from input data, while Bi-LSTMs are capable of capturing long-term dependencies and temporal dynamics, such as trends and periodicity in time series data. By combining these two architectures, the CNN-Bi-LSTM model enables comprehensive analysis of data with spatiotemporal characteristics, improves the accuracy of feature extraction, and significantly enhances the model's ability to understand and predict complex patterns. The structural architecture of the CNN-Bi-LSTM model is illustrated in Figure 10.

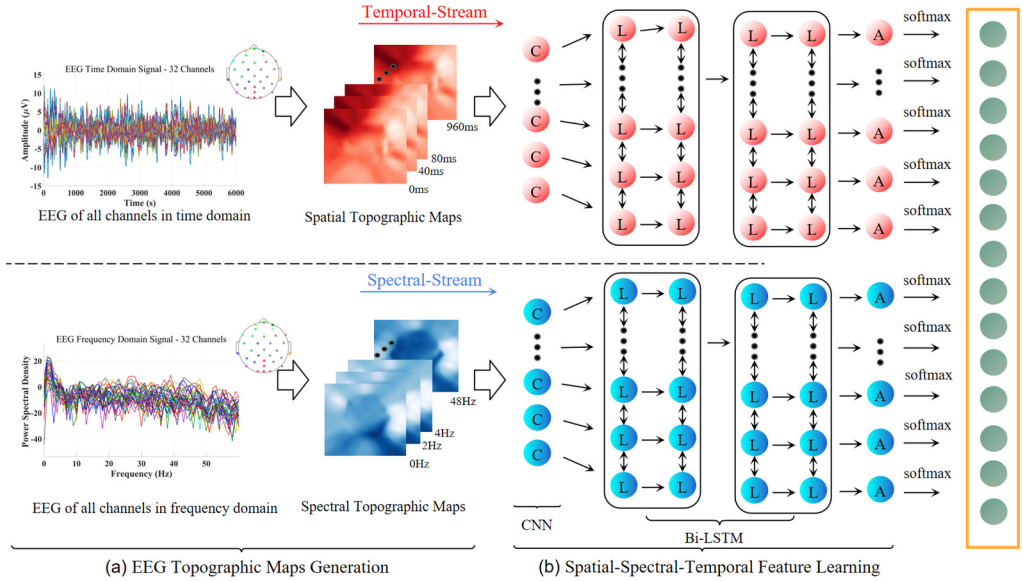


Figure 10. CNN-Bi-LSTM structure diagram.

The model in this study was implemented using Python 3.11 (Python Software Foundation). The convolutional layer contains 32 filters with a kernel size of 5 and uses the

ReLU activation function. The Bi-LSTM layer consists of 64 hidden units and employs the sigmoid activation function. The learning rate was set to 0.001, and the model was trained for 400 iterations. The detailed model parameters are summarized in Table 2. All model computations were performed on a server equipped with an Intel Core i9-13900K processor, 64 GB of RAM, and an NVIDIA RTX 4090 graphics card with 24 GB of video memory.

Table 2. CNN-Bi-LSTM model pa.ameters.

	Parameter	Value
Input Layer	Number of input nodes	10
CNN layer	Convolutional layer filters	32
	Convolutional layer kernel size	5
	activation function	relu
	Convolutional layer padding	1
	Pooling layer pool_size	2
Bi-LSTM Layer	Number of Bi-LSTM hidden units	64
	activation function	sigmoid
Output Layer	Number of output nodes	1
	Loss Function	binary_crossentropy
	batch_size	128
	Learning Rate	0.001
	epoch	400

2.9. Model Training

A subject-level data partition strategy was adopted in this study. Feature sequence samples obtained from all participants under the three flight scenarios were randomly divided into training and test sets, with approximately 75% of the data assigned to the training set and 25% to the test set. A portion of the training set was further separated as a validation set, which was used for hyperparameter tuning and for determining the early stopping criterion during model training. After training was completed, the model performance on the test set was evaluated in terms of accuracy, precision, recall, and F1-score. In addition, confusion matrices were plotted to analyze the detailed classification performance across different workload levels.

In the whole-brain model training experiments, all significant features extracted from all electrodes were combined into a single feature set and fed into the CNN–BiLSTM model to obtain the classification performance of the whole-brain model. In the feature-comparison experiments, only time-domain features, only frequency-domain features, or fused time–frequency features were used as inputs while keeping the model architecture unchanged, in order to compare the contributions of different feature combinations to workload recognition performance. In the brain-region comparison experiments, electrodes were grouped according to the frontal, parietal, temporal, and occipital lobes, and features extracted from each brain region were used as inputs to construct four CNN–BiLSTM models with identical architectures but different input sources. These models were used to analyze the differential responses of each brain region to changes in workload. All three levels of experiments followed the same training and validation pipeline, ensuring a clear correspondence between the methodological design and the reported results.

3. Result

All statistical analyses in this study were performed using SPSS 22.0. Descriptive statistics were conducted for the NASA-TLX subscales, overall workload scores, and EEG time-domain and frequency-domain features. For the subjective scale, Pearson correlation coefficients between the six NASA-TLX dimensions and the overall workload score were

calculated to examine the contribution of each dimension to the global workload construct. With respect to workload level differences, one-way analysis of variance (one-way ANOVA) was used to compare the differences among the three workload conditions.

3.1. Subjective Data Extraction

3.1.1. Reliability and Validity Analysis

Reliability and validity are fundamental for evaluating the accuracy and stability of scale-based assessments. The results show that the Cronbach's alpha coefficient for the NASA-TLX scale is 0.811, which exceeds the commonly accepted threshold of 0.6, indicating good internal consistency reliability. The scale's significance level is less than 0.001, and the Kaiser-Meyer-Olkin (KMO) sampling adequacy index is 0.832, which is greater than 0.7, suggesting strong construct validity. Therefore, the NASA-TLX can be considered a reliable and valid subjective tool for assessing pilot workload.

3.1.2. Correlation Analysis

Correlation analysis can be used to explore the internal relationships among the six dimensions of the NASA-TLX. In this study, Pearson correlation analysis was adopted, where the correlation coefficient ranges from −1 to 1, and values closer to 1 or −1 indicate a stronger linear relationship between two variables. A total of 297 NASA-TLX records, obtained for the takeoff, downwind, and final approach phases under the three workload conditions for all participants, were subjected to Pearson correlation analysis. The results are shown in Table 3.

Table 3. Results of correlation analysis of NASA-TLX scale.

Dimensions	Effort	Mental Demands	Task Performance	Time Requirements	Frustration Level	Physical Burden
Mental demands	0.767 **					
Task Performance	0.764 **	0.824 **				
Time requirements	−0.167	0.011	−0.293 **			
Frustration level	0.732 **	0.708 **	0.807 **	−0.302 **		
Physical burden	0.660 **	0.651 **	0.741 **	−0.416 **	0.703 **	
Total load fraction	0.886 **	0.933 **	0.912 **	−0.065	0.859 **	0.747 **

Note: ** indicates statistical significance at $p < 0.001$.

The correlation matrix in Table 3 reveals that, with the exception of temporal demand, all the other dimensions exhibit significant positive correlations with the overall workload score. Among these, mental demand (r = 0.933, $p < 0.01$), performance (r = 0.912, $p < 0.01$), effort (r = 0.886, $p < 0.01$), and frustration (r = 0.859, $p < 0.01$) show the strongest associations, suggesting that these factors play a critical role in pilots' subjective perception of workload. Additionally, the physical demand dimension shows a moderately strong positive correlation with overall workload (r = 0.747, $p < 0.01$), indicating that physical workload is also a non-negligible factor during flight operations.

It is noteworthy that temporal demand exhibits generally low correlations with the other NASA-TLX dimensions. In particular, it shows significant negative correlations with task performance (r = −0.293, $p < 0.01$), frustration (r = −0.302, $p < 0.01$), and physical demand (r = −0.416, $p < 0.01$). These results suggest that, in this flight simulation task, pilots' subjective assessment of time pressure may not align with other cognitive and emotional workload dimensions. This discrepancy may be influenced by individual differences in flight experience or in the perception of task pacing.

3.2. EEG Feature Analysis

To examine whether the differences in EEG time-domain features across varying workload conditions are statistically significant, this study conducted a one-way analysis of variance (ANOVA) on five time-domain indicators using SPSS software. Dunnett's post hoc test was employed to perform pairwise comparisons between experimental groups. The analysis was based on three experimental scenarios corresponding to different workload levels: clear weather, low-visibility fog, and single-engine failure. The extracted time-domain features included root mean square (RMS), waveform factor, peak factor, pulse factor, and margin factor. The results are illustrated in Figure 11.

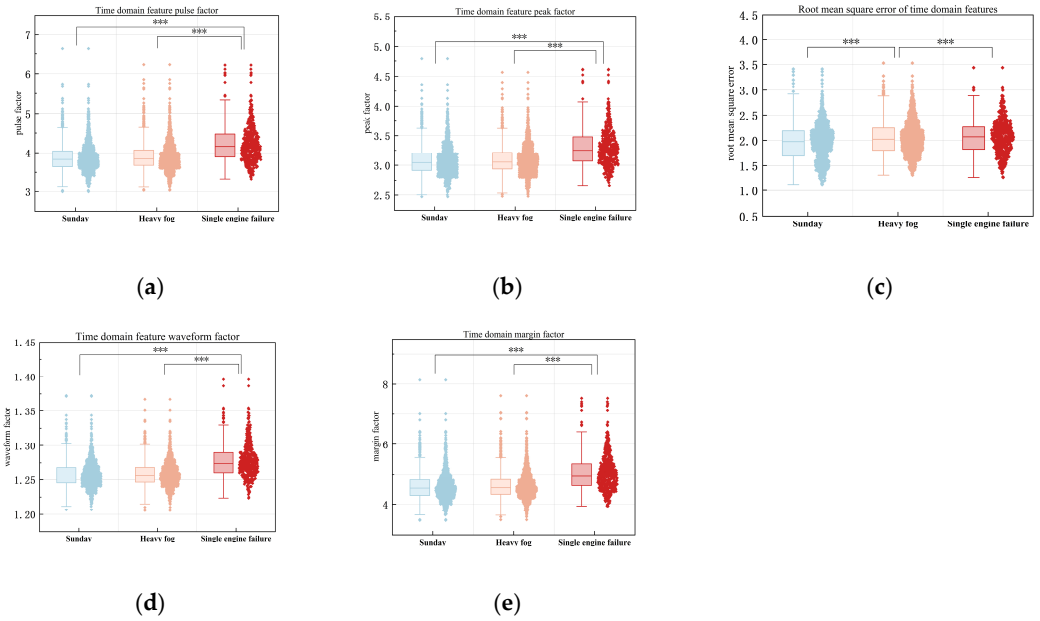


Figure 11. Time domain characteristic results under different workloads. (*** indicates statistical significance at $p < 0.001$). (**a**) Pulse Factor Difference Plot. (**b**) Peak Factor Difference Plot. (**c**) Root Mean Square Error Difference Plot. (**d**) Form Factor Difference Plot. (**e**) Margin Factor Difference Plot.

Figure 11a–e illustrate the distribution patterns and statistical significance of various time-domain EEG features across the three workload scenarios. As shown in the figures, all feature values exhibit an increasing trend under the single-engine failure condition, accompanied by greater dispersion. This suggests heightened uncertainty and variability in the pilot's neural activity under high workload. In contrast, under the clear weather condition, the distributions of all indicators are more concentrated with smaller fluctuations, indicating that the pilot's EEG activity is more stable and consistent in low-workload situations.

To further investigate the distribution and variation in EEG spectral energy under different flight mission scenarios, this study compared the mean PSD values across four commonly used frequency bands: δ, θ, α, and β. Additionally, a set of ratio-based indices was constructed to serve as sensitive indicators of changes in EEG frequency structure. Specifically, five ratio features—θ/β, α/β, (θ + α)/β, (θ + α)/(α + β), and θ/α—were calculated to reflect the relative balance between low-frequency (θ, α) and high-frequency (β) rhythms. These indices are designed to more accurately characterize the neural rhythm regulation mechanisms of pilots under varying task loads.

This study employed one-way analysis of variance (ANOVA) to assess the statistical significance of the aforementioned EEG spectral ratio indicators across three workload conditions: clear weather, heavy fog, and single-engine failure. Post hoc pairwise comparisons were subsequently conducted to further examine group differences. The results of the one-way ANOVA are presented in Table 4.

Table 4. Average PSD of the whole brain in each frequency band under different situations.

Category	Group	Clear Weather	Heavy Fog	Single-Engine Failure	F	*p*
ESD	δ	2.00 ± 0.93 a	1.94 ± 0.96 a	1.75 ± 0.86 b	15	<0.001
	θ	1.35 ± 0.52 c	1.36 ± 0.60 b	1.43 ± 0.62 a	5	<0.001
	α	0.78 ± 0.41 c	0.81 ± 0.40 b	1.37 ± 0.97 a	277	<0.001
	β	0.91 ± 0.45 c	085 ± 0.49 b	0.73 ± 0.31 a	38	<0.001
	θ/β	1.76 ± 1.00 b	1.90 ± 0.97 b	2.15 ± 1.02 a	36	<0.001
	α/β	0.97 ± 0.44 c	1.03 ± 0.47 b	1.98 ± 1.21 a	600	<0.001
	(θ + α)/β	2.74 ± 1.30 c	2.93 ± 1.29 b	4.13 ± 1.90 a	226	<0.001
	(θ + α)/(α + β)	1.40 ± 0.41 c	1.37 ± 0.32 b	1.34 ± 0.44 a	8	<0.001
	θ/α	1.94 ± 0.86 c	1.90 ± 0.89 b	1.31 ± 0.63 a	128	<0.001

Note: Within the same row, values with different lowercase letters indicate statistically significant differences among the three workload conditions (clear weather, heavy fog, and single-engine failure) according to Duncan's post hoc multiple comparison test ($p < 0.05$), whereas values sharing the same lowercase letter are not significantly different.

As shown in Table 3, the whole-brain average power spectral density (PSD) values for the four frequency bands—δ, θ, α, and β—exhibited significant differences across the three workload conditions (all $p < 0.001$). PSD values in the δ and β bands gradually declined as workload increased, suggesting that pilots experienced heightened arousal and increased mobilization of cognitive resources under high workload conditions. In contrast, PSD values in the θ and α bands increased significantly under the single-engine failure condition, indicating enhanced neural activity associated with greater cognitive processing and attentional demands. These findings align with previous research demonstrating that α and θ wave activity tends to increase with rising cognitive workload.

Furthermore, all five frequency band ratios demonstrated significant differences across workload conditions ($p < 0.001$). The θ/β, α/β, and (θ + α)/β ratios increased markedly under high workload scenarios, suggesting a greater reliance on low-frequency neural rhythms during complex tasks. This may be associated with enhanced synchronous activity in brain regions involved in decision-making, spatial awareness, and sustained attention. Notably, the (θ + α)/(α + β) and θ/α ratios were higher under clear weather conditions compared to heavy fog and single-engine failure conditions, reflecting the relative dominance of theta wave activity under low workload. This finding indicates that pilots sustain attention even during seemingly less demanding flight conditions.

4. Discussion

4.1. Whole-Brain Model Training

Figure 12 presents the accuracy and loss curves of the CNN-Bi-LSTM model. The training and test set accuracies reached 98.2% and 98.9%, respectively. Even at an early stage of training, both accuracies exceeded 95%, indicating strong short-term learning capability. As the number of training iterations increased, test accuracy gradually stabilized, with fluctuations within 3%, and no upward trend was observed in the loss curve. These results demonstrate the model's robust generalization ability and effectiveness in identifying pilot workload over extended training periods.

Based on the basic statistical outcomes of the confusion matrix, four secondary evaluation metrics—accuracy, precision, recall, and F1 score—were derived through standard arithmetic operations on the underlying components. This study computed these metrics for the CNN-Bi-LSTM model, yielding an accuracy of 98.2%, a precision of 98.2%, a recall of 98.1%, and an F1 score of 98.4%. The corresponding confusion matrix is illustrated in Figure 13.

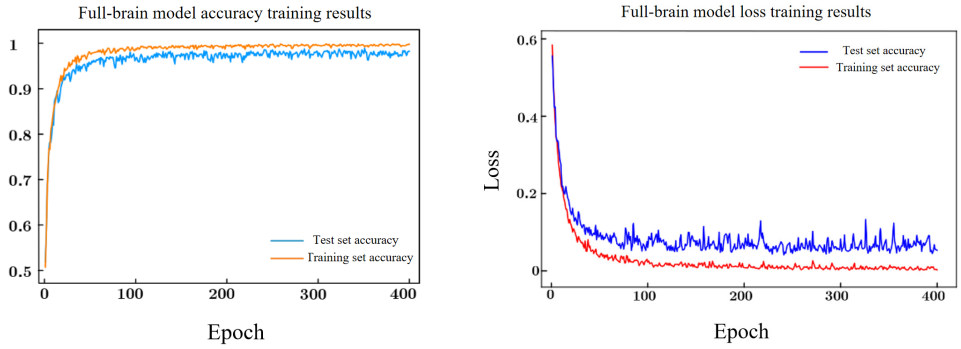


Figure 12. Accuracy Results of the Whole-Brain Model on Training and Test Sets.

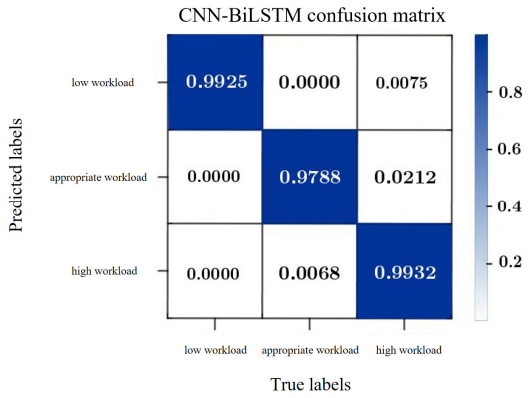


Figure 13. CNN-Bi-LSTM confusion matrix.

4.2. Comparison of EEG Characteristic Indicators

To further examine the sensitivity of pilot workload recognition to time-domain and frequency-domain EEG features, this study trained and evaluated the established CNN-Bi-LSTM model using each feature type independently. The training outcomes are presented in Figure 14.

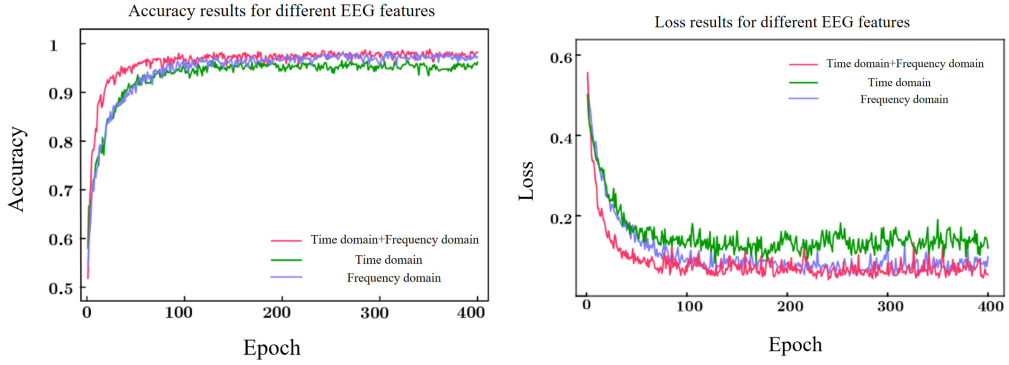


Figure 14. Training and Testing Accuracy of Models Using Different EEG Feature Types.

To further examine the effectiveness of the extracted features in characterizing pilot workload and the superiority of the proposed CNN–BiLSTM model in workload recognition, three types of workload-related EEG features obtained from the simulated flight experiment were separately used as inputs to a Random Forest (RF), Support Vector Machine (SVM), a standalone CNN model, a standalone LSTM model, and the CNN–BiLSTM model. Model performance was evaluated using four metrics: accuracy, precision, recall, and F1-score. The performance comparison of the models under different input conditions is summarized in Table 5.

Table 5. Performance comparison of different EEG feature sets.

Model	Feature	Accuracy (%)	Precision (%)	Recall (%)	F1 (%)
RF	Frequency	70.81%	73.38%	69.79%	71.54%
	Time–frequency	73.04%	75.39%	72.01%	73.66%
	Hybrid	75.39%	80.54%	73.02%	76.60%
SVM	Frequency	72.82%	76.51%	71.25%	73.79%
	Time–frequency	75.62%	78.75%	74.11%	76.36%
	Hybrid	78.41%	82.55%	76.24%	79.27%
CNN	Frequency	74.83%	77.18%	73.72%	75.41%
	Time–frequency	76.92%	80.09%	75.69%	77.83%
	Hybrid	81.99%	86.13%	79.55%	82.71%
LSTM	Frequency	77.85%	81.43%	75.99%	78.62%
	Time–frequency	84.79%	89.04%	82.06%	85.41%
	Hybrid	86.24%	89.49%	84.03%	86.67%
CNN-BiLSTM	Frequency	97.9%	97.0%	96.79%	97.42%
	Time–frequency	95.74%	95.80%	95.74%	95.73%
	Hybrid	98.2%	98.2%	98.1%	98.4%

As shown in Table 5, models trained with fused time- and frequency-domain features achieve better overall performance. When the number of training iterations is relatively small, the accuracies of the models based solely on time-domain or frequency-domain features increase more slowly than those of the model using fused features, indicating that the fused-feature model has stronger short-term learning capability. As the number of iterations increases, the model with fused time- and frequency-domain features attains higher test-set accuracy and exhibits fewer abrupt increases or decreases, suggesting that the model can more effectively learn and discriminate EEG features. In terms of the secondary performance metrics, the model based on fused time- and frequency-domain features outperforms the models using only frequency-domain or only time-domain features across all indicators. This finding indicates that the fused time–frequency feature set used in this study provides the highest feature quality, and that frequency-domain features are generally more informative than time-domain features.

4.3. Model Training Results

To investigate how different brain regions respond to variations in pilot workload, this study employed EEGLAB to segment the EEG signals and identify the electrode channels corresponding to specific brain regions. According to prior research, the frontal lobe is involved in motor control, higher-order cognition, language production, and emotional and behavioral regulation. The temporal lobe is associated with auditory processing, language comprehension, memory, olfaction, and emotional regulation. The occipital lobe primarily processes visual information and contributes to spatial localization and visual memory formation. The parietal lobe integrates sensory input and is responsible for spatial orientation, motor planning and coordination, attention regulation, and body image perception. The classification results of the models trained on EEG features from each brain region are presented in Figure 15.

The results indicate that the models based on EEG signals from the occipital and frontal lobes achieved higher test set accuracy and lower test set loss, whereas the parietal lobe model exhibited lower accuracy and higher loss. Notably, the test accuracy curves for the occipital and frontal regions showed pronounced fluctuations at early iterations; however, the amplitude of these fluctuations diminished as training progressed, demonstrating the models' strong performance in long-term learning. Additionally, to further assess the

classification performance of models trained on different brain regions, a comparative analysis using secondary metrics was conducted. The results are presented in Table 6.

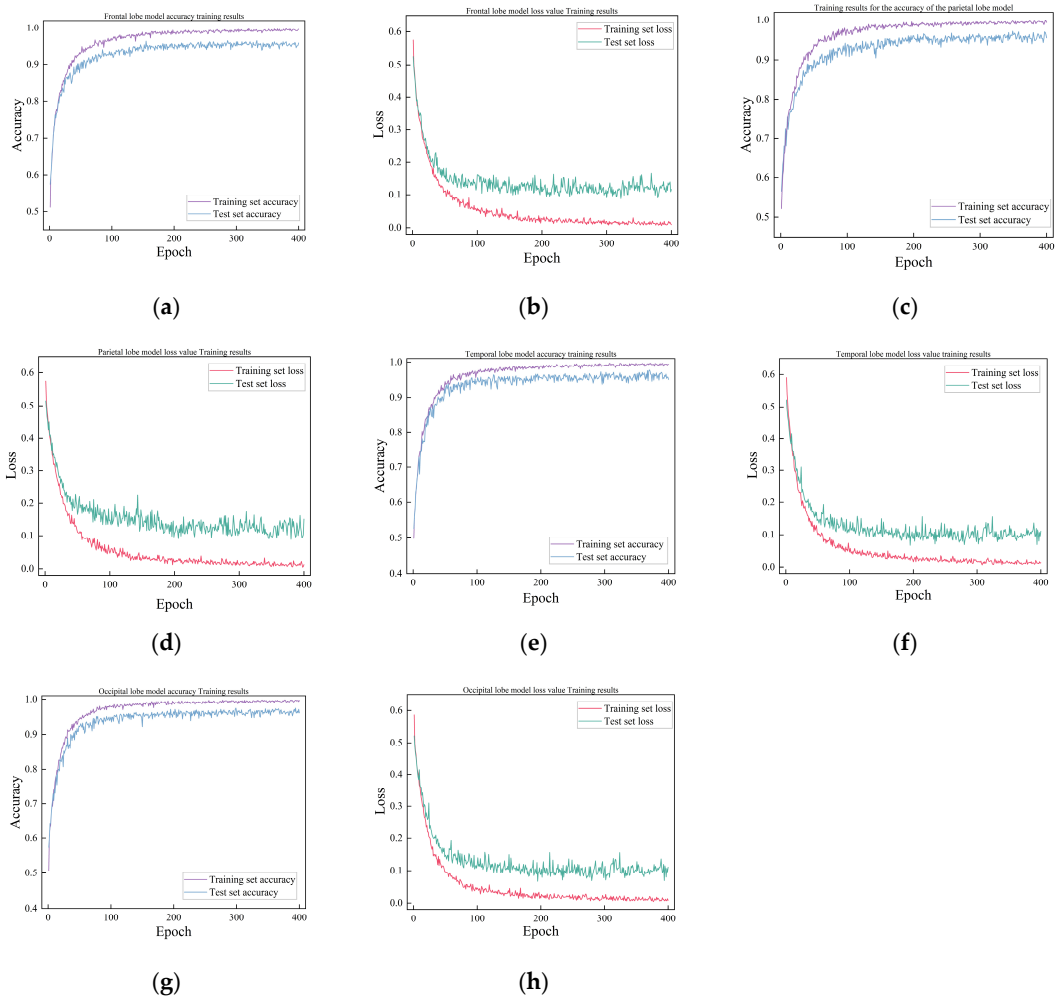


Figure 15. Model Accuracy and Loss Results for Different Brain Regions. (**a**) Frontal Lobe Model Accuracy. (**b**) Frontal Lobe Model Loss Value. (**c**) Parietal Lobe Model Accuracy. (**d**) Parietal Lobe Model Loss Value. (**e**) Temporal Lobe Model Accuracy. (**f**) Temporal Lobe Model Loss Value. (**g**) Occipital Lobe Model Accuracy. (**h**) Occipital Lobe Model Loss Value.

Table 6. Comparison of Model Performance Across Different Brain Regions.

	Accuracy	Precision	Recall	F1 Score
Frontal Lobe	97.46%	97.32%	97.55%	97.24%
Parietal Lobe	95.73%	95.44%	95.78%	95.48%
Temporal Lobe	96.14%	96.28%	96.14%	96.11%
Occipital Lobe	97.47%	97.45%	97.46%	97.36%

The data presented in Table 6 indicate that the models based on the frontal and occipital regions achieved similar levels of accuracy and precision, both outperforming the parietal region across all evaluated performance metrics. The frontal lobe is primarily responsible for complex information processing and decision-making. Under high workload conditions, pilots must rapidly interpret aircraft status and instrument readings while managing abnormal or emergency situations. This heightened demand on cognitive resources leads to increased activation of the frontal region. The occipital lobe, as the principal center for visual information processing, becomes more active as pilots continuously monitor both the

external environment and cockpit instrumentation to maintain situational awareness and ensure flight safety. In contrast, the parietal lobe, which is involved in sensory integration and spatial orientation, exhibits reduced activation under high workload conditions. This may be attributed to standard flight protocols that emphasize reliance on instrument-based navigation over visual cues to avoid spatial disorientation, thereby diminishing the engagement of the parietal region during high cognitive load.

5. Conclusions

By integrating both subjective and objective assessment methods, this study collected data from pilots under varying workload levels. EEG signals were preprocessed and subjected to time-domain and frequency-domain feature extraction. A CNN-Bi-LSTM deep learning model was then employed to train and classify the extracted features. The main conclusions are as follows:

(1) Reliability and validity analyses demonstrate that the NASA-TLX can effectively distinguish different levels of pilot workload. In addition, Pearson correlation analysis shows that the effort, mental demand, performance, frustration, and physical demand dimensions of the scale are strongly associated with overall workload, indicating that these factors play a key role in pilots' subjective perception of workload.

(2) One-way ANOVA with post hoc pairwise comparisons conducted in SPSS reveals that, in the time domain, root mean square, waveform factor, peak factor, pulse factor, and margin factor can be effectively used as EEG feature indicators. Furthermore, in the frequency domain, the ESD values of each frequency band across the whole brain were extracted and used to compute ratio features. The results indicate that δ, θ, α, β, θ/β, α/β, $(\theta+\alpha)/\beta$, $(\theta+\alpha)/(\alpha+\beta)$, and θ/α can all serve as discriminative EEG indicators of workload.

(3) A CNN–BiLSTM model was constructed and trained using the extracted EEG features. Comparative experiments with traditional machine learning models show that frequency-domain EEG features provide better recognition performance than time-domain features, while the fused time- and frequency-domain feature set outperforms any single-type EEG feature set, confirming the superiority of multimodal feature fusion in pilot workload recognition.

(4) Based on the CNN–BiLSTM models constructed for different brain regions, it is found that, as workload increases, the frontal and occipital regions exhibit more pronounced activation, whereas the parietal region shows relatively weaker responsiveness compared with the other regions. This suggests that different cortical areas make differentiated contributions to workload modulation during flight tasks.

In addition, pilot workload is influenced by the development of operational automaticity through training and experience. As pilots become more proficient, certain flight operations gradually become automated, thereby reducing the cognitive resources required to complete the tasks. This improvement in automaticity may help explain the workload-related EEG changes observed under different task conditions.

It should be noted that this study still has several limitations, including the relatively small sample size, the homogeneity of the participant group, the use of a simulated flight environment rather than real flight operations, and the partial reliance on subjective NASA-TLX ratings for workload labeling. Therefore, the generalizability of the findings to a broader pilot population requires further verification. Future research will expand the sample to include pilots of different ages and flight experience levels (including professional airline captains) and collect data under more complex task scenarios and more realistic flight simulation environments. In addition, future work will further integrate multiple physiological modalities such as eye movements and electrocardiography to construct

multimodal fusion models, and will employ attention mechanisms, feature-importance analysis, and visualization techniques to enhance model interpretability. These efforts are expected to deepen our understanding of the neurophysiological mechanisms underlying pilot workload and to promote the practical application of EEG-based workload recognition models in flight training and adaptive assistance systems.

Author Contributions: Conceptualization, Methodology, Writing—original draft, Project administration, Funding acquisition, W.Y.; Writing—original draft, Formal analysis, Data curation, Visualization, Y.L.; Writing—original draft, Formal analysis, Formal analysis, Data curation, L.L.; Conceptualization, Project administration, Funding acquisition, Supervision, H.S.; Writing—original draft, Conceptualization, Methodology, Investigation, Visualization, Supervision, H.W.; Writing—original draft, Conceptualization, Methodology, Investigation, Visualization, T.P.; Writing—original draft, Software, Visualization, Y.Z.; Software, Writing—review and editing, G.L. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Aeronautical Science Foundation of China [Grant 2024Z071052007]; the Key Laboratory of Brain–Machine Intelligence Technology, Ministry of Education, Nanjing University of Aeronautics and Astronautics [Grant NJ2024029]; Research on Safety Risk Assessment Technology and Method of Human–Computer Intelligent Interaction in Civil Aircraft Cockpit [Grant U2033202]; the Fundamental Research Funds for the Central Universities [Grant NS2022094].

Institutional Review Board Statement: The experimental design and protocol of this study are scientifically sound, fair, and ethical. The study poses no harm or risk to participants. Recruitment of participants is conducted on a voluntary and informed basis, with full protection of participants' rights and privacy. The research does not involve any conflict of interest or violation of ethical principles or legal regulations.

Informed Consent Statement: All participating pilots were fully informed about the purpose, content, and methods of the study and signed an informed consent form, explicitly stating their voluntary participation and agreement to the use of their data for research purposes. All collected data are anonymized to protect participants' privacy. Data are stored on secure servers with appropriate security measures in place to prevent data leakage or misuse.

Data Availability Statement: All data generated or analyzed during this study are included in this manuscript.

Acknowledgments: The authors thank the editor and anonymous reviewers for their constructive comments and valuable suggestions for improving the quality of the study.

Conflicts of 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. And the Aviation Industries Corporation of China Xi'an Flight Automatic Control Research Institute has no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

1. Chaturvedi, A.K.; Craft, K.J.; Canfield, D.V.; Whinnery, J.E. Toxicological findings from 1587 civil aviation accident pilot fatalities, 1999–2003. *Aviat Space Environ. Med.* **2005**, *76*, 1145–1150. [PubMed]
2. Steiner, S.; Fakleš, D.; Gradišar, T.; Scientific Research Center Ltd. Problems of crew fatigue management in airline operations. In Proceedings of the International Conference on Traffic and Transport Engineering, ICTTE, Belgrade, Serbia, 29–30 November 2012; pp. 617–623.
3. Yu, B.T. Pilots' Multiple Physiological Signals Based Workload Research. Master's Thesis, Shanghai Jiao Tong University, Shanghai, China, 2020.
4. Chen, J.; Xue, L.; Liu, Z.C. A pilot workload evaluation method based on EEG data and physiological data. In *2020 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*; IEEE: Macau, China, 2020; pp. 1–6. [CrossRef]

5. Cooper, G.E.; Harper, R.P. *The Use of Pilot Rating in the Evaluation of Aircraft Handling Qualities*; National Aeronautics and Space Administration: Washington, DC, USA, 1969.
6. Moriasi, D.N.; Wilson, B.N.; Douglas-Mankin, K.R.; Arnold, J.G.; Gowda, P.H. Hydrologic and water quality models: Use, calibration, and validation. *Trans. ASABE* **2012**, *55*, 1241–1247. [CrossRef]
7. Shu, S.X. *Human Factors and Crew Resource Management*; Beijing University of Aeronautics and Astronautics Press: Beijing, China, 2015.
8. Mansikka, H.; Virtanen, K.; Harris, D. Comparison of NASA-TLX scale, modified Cooper–Harper scale and mean inter-beat interval as measures of pilot mental workload during simulated flight tasks. *Ergonomics* **2019**, *62*, 246–254. [CrossRef] [PubMed]
9. Wang, Z.Y.; Liu, S.; Wanyan, X.; Dang, Y.; Chen, X.; Zhang, X. Pilot workload measurement model based on task complexity analysis. *Int. J. Ind. Ergon.* **2024**, *104*, 103637. [CrossRef]
10. Luo, S.; Zhang, C.; Zhu, W.; Chen, H.; Yuan, J.; Li, Q.; Wang, T. Noncontact perception for assessing pilot mental workload during the approach and landing under various weather conditions. *Signal Image Video Process.* **2025**, *19*, 98. [CrossRef]
11. Zhang, X.; Qu, X.; Xue, H.; Zhao, H.; Tao, D. Modeling pilot mental workload using information theory. *Aeronaut. J.* **2019**, *123*, 828–839. [CrossRef]
12. Othman, N.; Abdullah, U.N.; Romli, F.I. Evaluating Mental Workload Using Pupil Dilation and Nasa-Task Load Index. In *Convergence of Ergonomics and Design: Proceedings of ACED SEANES 2020*; Springer International Publishing: Berlin/Heidelberg, Germany, 2021; pp. 253–260. [CrossRef]
13. Alaimo, A.; Esposito, A.; Milazzo, A.; Orlando, C. An aircraft pilot workload sensing system. In *European Workshop on Structural Health Monitoring*; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; pp. 883–892. [CrossRef]
14. Socha, V.; Socha, L.; Hanakova, L.; Valenta, V.; Kusmirek, S.; Lalis, A. Pilots' performance and workload assessment: Transition from analogue to glass-cockpit. *Appl. Sci.* **2020**, *10*, 5211. [CrossRef]
15. Mohanavelu, K.; Poonguzhali, S.; Ravi, D.; Singh, P.K.; Mahajabin, M.; Ramachandran, K. Cognitive Workload Analysis of Fighter Aircraft Pilots in Flight Simulator Environment. *Def. Sci. J.* **2020**, *70*, 131–139. [CrossRef]
16. Houssein, E.H.; Hammad, A.; Ali, A.A. Human emotion recognition from EEG-based brain–computer interface using machine learning: A comprehensive review. *Neural Comput. Appl.* **2022**, *34*, 12527–12557. [CrossRef]
17. Berger, H. On the electroencephalogram of man, Fourteenth report. *Electroencephalogr. Clin. Neurophysiol.* **1969**, *69*, 299.
18. Ji, L.; Yi, L.; Li, H.; Han, W.; Zhang, N.N. Detection of Pilots' Psychological Workload during Turning Phases Using EEG Characteristics. *Sensors* **2024**, *24*, 5176. [CrossRef] [PubMed]
19. Cho, Y.Y.; Kam, K.H.N.; Li, Q.; Yuan, X. How do the levels of automation in flight operations affect pilots' cognitive workload, reaction time, and EEG brain waves in cruising flights? In Proceedings of the 2024 IEEE 4th International Conference on Human-Machine Systems (ICHMS), Toronto, ON, Canada, 15–17 May 2024; Volume 19, pp. 1–6. [CrossRef]
20. Zhang, Z.W.; Mei, Y.H.; Zhou, X.Z. Assessing the Effect of HUD on Operator Mental Workload and Behavioral Performance During Simulated Flight. In Proceedings of the 2024 10th International Conference on Control, Automation and Robotics (ICCAR), Singapore, 27–29 April 2024; Volume 28, pp. 28–33. [CrossRef]
21. Verkennis, B.; van Weelden, E.; Marogna, F.L.; Alimardani, M.; Wiltshire, T.J.; Louwerse, M.M. Predicting Workload in Virtual Flight Simulations using EEG Features (Including Post-hoc Analysis in Appendix). In Proceedings of the 2025 IEEE International Conference on Artificial Intelligence and eXtended and Virtual Reality (AIxVR), Lisbon, Portugal, 27–29 January 2024.
22. Wang, Y.W.; Han, M.X.; Peng, Y.D.; Zhao, R.Q.; Fan, D.Q.; Meng, X.; Xu, H.; Niu, H.J.; Cheng, J.; Liu, T. LGNet: Learning local–global EEG representations for cognitive workload classification in simulated flights. *Biomed. Signal Process. Control* **2024**, *92*, 106046. [CrossRef]
23. Lee, D.H.; Kim, S.J.; Kim, S.H.; Lee, S.W. Decoding EEG–based Workload Levels Using Spatio–temporal Features Under Flight Environment. In Proceedings of the 2024 12th International Winter Conference on Brain-Computer Interface (BCI), Gangwon, Republic of Korea, 26–28 February 2024; pp. 1–5. [CrossRef]
24. Boumann, H.; Hamann, A.; Biella, M.; Carstengerdes, N.; Sammito, S. *Suitability of Physiological, Self-Report and Behavioral Measures for Assessing Mental Workload in Pilots*; ACM: New York, NY, USA, 2023. [CrossRef]
25. Bailey, B.P.; Konstan, J.A. On the Need for Attention-Aware Systems: Measuring Effects of Interruption on Task Performance, Error Rate, and Affective State. *Comput. Hum. Behav.* **2006**, *22*, 685–708. [CrossRef]
26. Iqbal, S.T.; Bailey, B.P. *Investigating the Effectiveness of Mental Workload as a Predictor of Opportune Moments for Interruption*; ACM: New York, NY, USA, 2005; pp. 1489–1492. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.