scientific reports

Check for updates

OPEN

ACXNet hybrid deep learning model for cross task mental workload estimation using EEG neural manifolds

G. Abinaya[1]✉ & K. Dinakaran[2]

Mental workload is an interdisciplinary construct that significantly influences human performance, particularly in tasks requiring sustained attention and cognitive processing. Effective mental workload assessment is critical for preventing cognitive overload, which can lead to errors and reduced efficiency in high-stakes environments. The approach leverages topographic neural manifolds (spatial electrode arrangements) and temporal neural manifolds (time-series patterns) to capture comprehensive brain activity representations.Traditional methods rely on subjective reports or task performance, but physiological signals like EEG provide a more objective and continuous means of monitoring cognitive states. Therefore, this paper proposes a hybrid novel approach ACXNet which integrates autoencoder, CNN and XGBoost to learn features of EEG from an individual cross task performance without prior subject-specific calibration or task specific pre-labeled .training data. Utilizing the STEW (Simultaneous Task EEG Workload) dataset, containing recordings from 48 participants experiencing different levels of cognitive demands. Unsupervised feature extraction was carried out using an autoencoder. Subsequently, a CNN was employed to capture the spatial-temporal dependencies in the data, and XGBoost was utilized for efficient mental workload classification. This research adopts a binary classification approach to differentiate between low and high mental workload during SIMKAP and No task. The ACXNet model proposed in this study outperforms the existing methods with an average accuracy of 92.10% for SIMKAP task and 89.94% for No task condition. These findings show that ACXNet significantly improves the robustness and precision of mental workload estimation, providing a scalable solution adaptable to real-world applications, opening new avenues for the development of intelligent systems in human-computer interaction, healthcare, and beyond.

Keywords Mental Workload, Neural Manifolds, Brain-Computer Interface (BCI), Cognitive Computing, XGBoost, CNN

Mental workload (MWL) refers to the relationship between the complexity of a task and an individual's cognitive capacity to manage its demands[1–4]. It is a critical factor in determining human performance, especially in environments that require sustained attention, quick decision-making, and the ability to operate under pressure. Accurate estimation and management of MWL are essential for optimizing human-computer interactions, enhancing user experience, and ensuring safety[5] and efficiency of operations in complex systems such as aviation, healthcare, and military applications.

The assessment of MWL is typically performed using three main categories of metrics: self-report measures, physiological (and neurophysiological) measures, and primary task performance measures[1]. Self-report measures, such as the NASA-TLX or the Rating Scale Mental Effort (RSME), involve participants rating their perceived workload. These methods are easy to administer and provide direct insight into how people feel during tasks, though they can be subjective and prone to bias. Task performance measures focus on how well a person performs the primary task, looking at metrics like response time, accuracy, and efficiency. While these measures give a direct indication of how workload affects performance, they are often task-specific and provide limited information about the underlying mental processes. In contrast, physiological and neurophysiological measures, including heart rate variability (HRV)[6], pupil dilation, and skin conductance, offer objective, continuous ways

[1]Information Technology, Saveetha Engineering College, Chennai, India. [2]Computer Science and Engineering, Vel Tech Multi Tech Dr. Rangarajan Dr. Sakunthala Engineering College, Chennai, India. ✉email: abinaya.g05@gmail.com

to assess MWL by monitoring the body's response to cognitive demands. Among these, electroencephalogram (EEG) signals are particularly useful for tracking changes in brain activity, offering a high-temporal-resolution, non-invasive method for real-time cognitive load estimation.

EEG can track changes in brain wave patterns associated with different levels of cognitive load, such as increases in frontal theta or decreases in parietal alpha waves, making it a reliable tool for real-time workload assessment. EEG's ability to provide continuous, non-invasive monitoring of brain function, coupled with its high temporal resolution, makes it one of the most effective ways to measure mental workload in both research and practical applications.

Electroencephalogram (EEG) signals[7–9] have a strong correlation with brain fatigue and cognitive states, making them valuable for mental workload assessment. EEG captures electrical signals produced when large groups of neurons fire synchronously, particularly within the cerebrum's specialized regions including the frontal, parietal, temporal, and occipital lobes[10,11].The resulting postsynaptic potentials provide insights into brain function across frequency ranges of 1-80 Hz, enabling the study of cognitive states and behavioral characteristics. The cerebral cortex is particularly significant for memory, attention, perception, and consciousness - all critical aspects of mental workload. Each lobe is specialized for processing different kinds of information, and when groups of neurons within the cerebrum fire synchronously, the resulting postsynaptic potentials can be recorded by EEG. Effective classification of EEG signals requires extracting meaningful features from this complex data to accurately analyze mental states and understand how different workload levels impact brain activity.

Table 1 summarizes the different types of EEG brain waves, their frequency ranges, and the corresponding brain states they are typically associated with. These waves are instrumental in understanding various cognitive and mental workload levels.

In the context of this work, topographic neural manifolds refer to the spatial distribution patterns of electrical activity across different electrode locations on the scalp, capturing the geographic organization of brain responses. Temporal neural manifolds represent the time-series evolution of these electrical patterns, encoding the dynamic changes in brain states over time. Together, these manifolds provide a comprehensive representation of both the spatial configuration and temporal dynamics of neural activity during cognitive tasks.

However, extracting meaningful features from this complex data remains a significant challenge, requiring advanced techniques to accurately interpret mental states. To address these challenges, this paper introduces ACXNet, a novel hybrid deep learning model that integrates autoencoders[12,13], Convolutional Neural Networks (CNNs)[14–16], and Extreme Gradient Boosting (XGBoost)[17–19] to enhance the accuracy of EEG-based mental workload estimation. A key contribution of this work is the use of autoencoders[20,21] for dimensionality reduction, which effectively compresses EEG data while preserving its critical features, thereby reducing computational complexity. Additionally, the integration of CNNs[7] enables automatic learning of spatial feature hierarchies, which is crucial for capturing the complex patterns present in EEG signals. XGBoost[17] is employed for feature selection and classification, offering robustness and improving the interpretability of the model[15]. The proposed approach is validated using the STEW dataset, demonstrating its effectiveness in real-world scenarios of mental workload estimation[21]. Overall, the proposed methodology addresses the inherent complexity of EEG signal processing, delivering both efficiency and improved classification accuracy, while also enhancing the interpretability of the results for practical cognitive workload analysis.

Related work

The field of mental workload estimation using EEG has experienced significant advancement through the integration of sophisticated signal processing techniques and machine learning algorithms. Accurate estimation

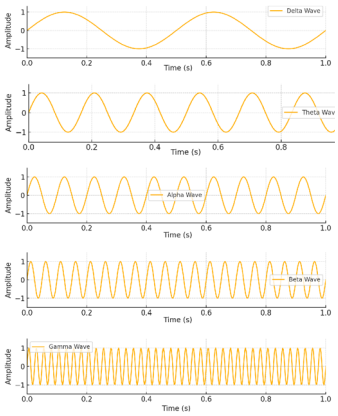
Wave Type	Frequency Range (Hz)	Waveform	Associated Brain State
Delta Wave δ	0.5 - 4 Hz		Restorative sleep, unconsciousness
Theta Wave θ	4 - 8 Hz		Light sleep, relaxation
Alpha Wave α	8 - 13 Hz		Calm, relaxed, awake
Beta Wave β	13 - 30 Hz		Active thinking, focus
Gamma Wave γ	30 - 80 Hz		Higher mental activity

Table 1. Characteristics of EEG Brain Waves and Corresponding Mental States.
This table presents the frequency ranges and mental states associated with five EEG wave types—delta, theta, alpha, beta, and gamma. These waves are fundamental for interpreting mental workload and cognitive activity in various conditions

of mental workload is essential for optimizing performance and safety in real-world domains such as healthcare, aviation, and human-computer interaction[22]. This section reviews key contributions from existing literature, highlighting methodologies and innovations that have shaped this evolving field.

Early research in mental workload estimation primarily relied on handcrafted features extracted from EEG signals, including power spectral density (PSD), statistical measures, and frequency domain analysis[23]. These features were typically used with conventional classifiers such as Support Vector Machines (SVM), Random Forest (RF), Linear Discriminant Analysis (LDA), and K-nearest neighbors. Ke et al.[24] employed SVM for cross-task mental workload classification using electroencephalography based on feature selection, reporting decent performance on selected EEG features. However, SVM's limitations became apparent in non-linear scenarios, as the model struggled to generalize across different tasks or subjects. Similarly, RF approaches demonstrated reasonable performance but required significant computational resources due to the complexity of decision tree structures[25]. Hongquan et al.[25] investigated the categorization of mental workload using EEG independent component features; nevertheless, their approach was limited for real-time applications. In their thorough study of physiological measurements for mental workload assessment, Charles and Nixon[3] brought to light the difficulties that traditional statistical methods have in capturing the intricate dynamics of brain activity.

To overcome the limitations of conventional methods, researchers used deep learning techniques for extracting complex, higher-level features from raw EEG data. Convolutional Neural Networks (CNNs) have gained prominence due to their ability to automatically identify spatial information from multi-channel EEG signals. This capability is particularly advantageous in mental workload estimation, as demonstrated by Pang et al.[26], who used stochastic configuration networks (SCN) for subject-specific mental workload classification using EEG, achieving accuracies around 87%. Jin et al.[15] enhanced CNN models through deep transfer learning via the DS-CNN framework, which enabled stable EEG pattern identification across tasks. Their approach allowed knowledge transfer from one task to another, enhancing performance in multi-task scenarios with reported accuracies exceeding 90%. Wang et al.[16] developed LGNet for learning local-global EEG representations for cognitive workload classification in simulated flights, demonstrating the effectiveness of hierarchical feature learning in complex cognitive tasks.

Recent advances have also explored attention mechanisms in deep learning architectures. Ji et al.[11] proposed a dynamic residual network with attention mechanisms for cross-task cognitive workload recognition using neurophysiological signals, achieving improved performance through better feature selection and temporal dependency modeling. Thekkekara et al.[14] demonstrated the effectiveness of attention-based CNN-BiLSTM models in related signal processing tasks, showing the potential of attention mechanisms for complex signal processing tasks. Autoencoders are widely used in EEG-based mental workload estimation for unsupervised feature learning and denoising. Chakladar et al.[27] utilized deep BLSTM-LSTM networks combined with evolutionary algorithms for EEG-based mental workload estimation, achieving approximately 91% accuracy. Their approach proved effective for capturing temporal dependencies in EEG data while filtering out irrelevant noise. Yu and Chen[12] developed a robust operators' cognitive workload recognition method based on masked denoising autoencoders, demonstrating enhanced robustness in noisy environments with high classification accuracy. Chakladar et al.[13] later explored cognitive workload estimation using variational autoencoders combined with attention-based deep models, achieving improved performance through better feature representation and selection.

Yang et al.[20] proposed an ensemble deep learning classifier based on denoising autoencoders for assessing cognitive mental workload via EEG signals, highlighting ensemble methods' value in increasing robustness. Ying and Chen[21] investigated recognition of mental workload based on hybrid autoencoders, showing the potential of combining different autoencoder architectures for enhanced feature extraction.

Gradient boosting techniques, particularly XGBoost, have gained attention for their scalability and precision in classification tasks. Wang et al.[19] showed that ensemble XGBoost models are more accurate and efficient than traditional neural networks in identifying disorders of consciousness using EEG connectivity. Their approach leveraged XGBoost's ability to reduce bias and variance, achieving accuracies consistently exceeding 90%. Tseng and Tang[18] used feature selection and image segmentation to detect brain tumors using optimized XGBoost algorithms, proving its applicability across many biomedical domains. Nguyen and Hoang[2] achieved strong performance in challenging classification scenarios by combining deep convolutional neural networks with metaheuristic-optimized extreme gradient boosting machines. The integration of XGBoost with deep learning models has shown particular promise. Sugiharti et al.[28] explored the integration of CNNs and extreme gradient boosting for medical image analysis, demonstrating improved accuracy through hybrid architectures. These studies collectively highlight the potential of combining gradient boosting with deep learning for enhanced model performance and interpretability.

Cross-task mental workload recognition remains a major challenge in the field, as models must generalize across different cognitive tasks without task-specific calibration. Ke et al.[24] conducted early work in cross-task mental workload recognition using electroencephalography based on feature selection and support vector machine regression. Their approach addressed the fundamental challenge of creating models that can operate across diverse cognitive tasks. Ji et al.[11] extended this line of work by employing a dynamic residual network with attention mechanisms, which demonstrated improved generalization across different neurophysiological signal patterns. Taori et al.[29] contributed to cognitive workload classification with their innovative pipeline interface using Hidden Markov Models, focusing on generalization through novel preprocessing approaches. Panwar et al.[30] introduced EEG-CogNet, a deep learning framework for cognitive state assessment using EEG brain connectivity, addressing cross-task challenges through advanced network architectures. Compared to task-specific CNNs, these models exhibited greater robustness across cognitive conditions. These studies collectively demonstrate the evolution from task-specific to more generalizable cross-task approaches in mental workload estimation.

Multimodal approaches to mental workload measurement are being investigated more and more in current research. Using electroencephalography, electrodermal activity, and video signals, Mehmood et al.[5] examined multimodal integration for data-driven classification of mental weariness. Their research showed the benefits of integrating several physiological indicators for a more reliable evaluation of mental state. Ding et al.[9] investigated the use of machine learning and multimodal techniques to assess and identify mental workload during simulated computer tasks. improving precision by integration of signals. Safari et al.[31] examined how to classify mental effort using machine learning and brain connectivity on EEG data, emphasizing the significance of connection patterns in evaluating cognitive state. In order to demonstrate the usefulness of mental workload assessment in educational contexts, Gupta et al.[10] created machine learning-based decision support systems for real-time mental workload estimation during online education utilizing wearable physiological monitoring devices.

Hybrid techniques that integrate many deep learning architectures have garnered more attention in recent research. In order to achieve superior performance through architectural variety, Nguyen et al.[32] presented mental workload assessment utilizing EEG signals by merging multi-space deep models. The usefulness of transfer learning techniques was demonstrated by Yin et al.[17], who investigated mental workload estimation using transfer dynamical autoencoders in deep learning frameworks. Wu et al.[7] investigated fatigue detection using deep learning networks with EEG signals, contributing to understanding of cognitive state detection through advanced neural architectures. These hybrid approaches leverage complementary strengths across models and reflect a growing trend toward more generalizable, interpretable systems.

Research gaps and motivation

Despite significant advances in EEG-based mental workload estimation, current approaches face several limitations. Traditional methods rely heavily on handcrafted features that may not fully capture brain dynamics, while deep learning models often require large labeled datasets and struggle with cross-task generalizability. Additionally, model interpretability remains challenging, making it difficult to connect predictions with underlying neural activity.

There is also a critical need for robust feature extraction and denoising methods, especially in real-world and noisy environments where labeled data are scarce. Current cross-task approaches, while promising, still require further development to achieve robust performance across diverse cognitive tasks without extensive task-specific calibration.

This research addresses these limitations by proposing ACXNet, a novel hybrid approach that integrates autoencoders for efficient feature extraction and noise reduction, CNNs for hierarchical spatial-temporal feature learning, and XGBoost for robust classification with enhanced interpretability. The proposed approach aims to overcome existing limitations by enhancing generalization across different tasks, subjects, and noise conditions while providing real-time mental workload estimation capabilities essential for practical applications.

Proposed work

The suggested model ACXNet improves the categorization of mental workload states from EEG data by utilizing a hybrid strategy that combines an autoencoder, CNNs, and XGBoost. The proposed hybrid approach for mental workload estimation is presented in Figure 1.

The autoencoder is strategically positioned as the first processing step to address the high-dimensional nature of EEG data (3584 features from 14 channels $\times$ 256 samples). Rather than using traditional dimensionality reduction techniques that may lose critical information, the autoencoder learns an optimal compressed representation (128 features) that preserves the most relevant characteristics for mental workload classification. This unsupervised feature learning approach maintains the intrinsic structure of the EEG signals while reducing computational complexity for subsequent CNN processing. The reconstruction loss (MSE) ensures that essential information is retained during compression. This process is crucial for lowering the EEG data's complexity while maintaining the key characteristics required for accurate categorization. The reduced feature set is then fed into a CNN, which excels at detecting spatial patterns and features within the data through its convolutional layers. The CNN's role is to capture and enhance the intricate patterns that characterize different mental workload states. After the CNN processes the features, the resulting data is passed to an XGBoost classifier. XGBoost applies gradient boosting techniques to refine the classification process, improving accuracy and performance. This model leverages the strengths of each component: the autoencoder for efficient feature extraction, the CNN for advanced pattern recognition, and XGBoost for robust classification. The proposed approach aims to outperform conventional methods by integrating various techniques to deliver a more accurate and dependable assessment of mental workload.

The following is a summary of the entire workflow for estimating mental workload:

Step 1: Normalize the data and eliminate artifacts from the raw EEG signals.
Step 2: Use the autoencoder to extract compressed, high-level features from the preprocessed EEG signals.
Step 3: Pass the extracted features through the CNN to learn spatial and temporal patterns.
Step 4: Use XGBoost for classifying the learned features into different levels of mental workload. This process is further detailed in Algorithm 1.

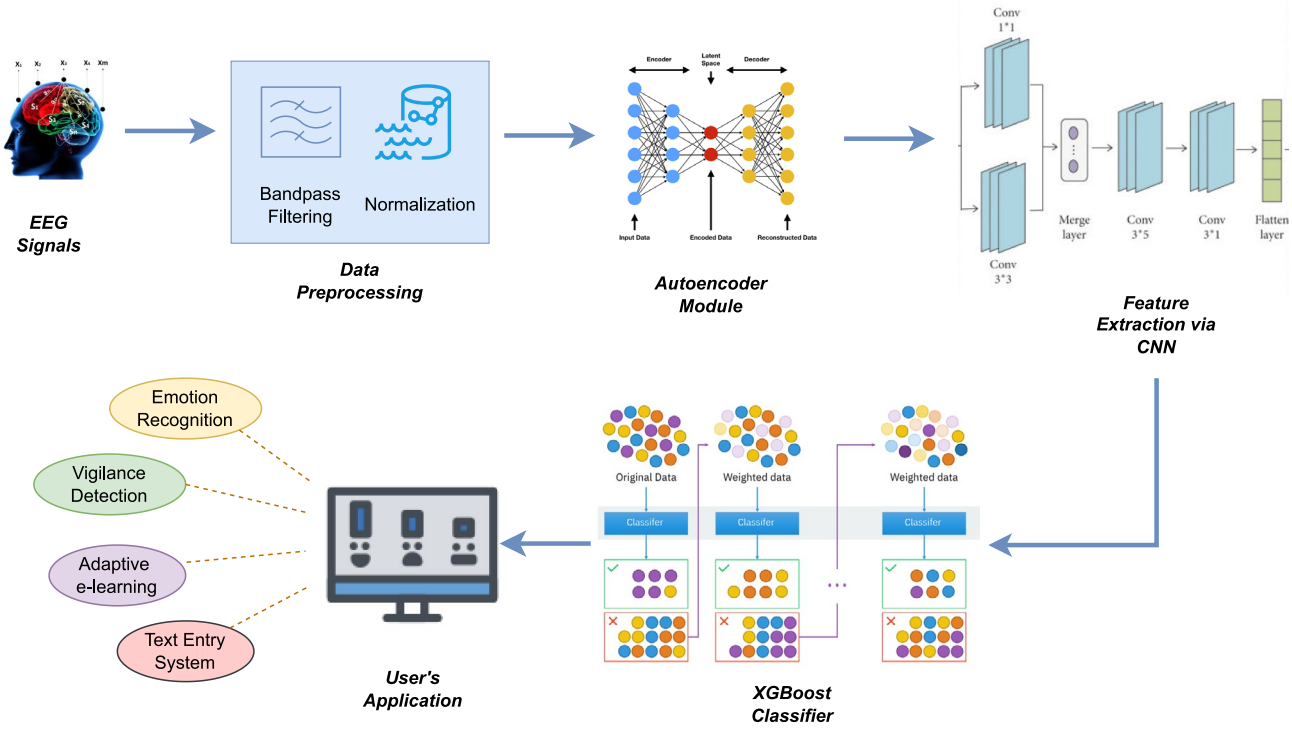


Fig. 1. Architecture of the ACXNet Model for EEG-Based Mental Workload Classification. Input: Raw EEG (14 channels × 256 samples) → Preprocessing → Autoencoder (Input: 3584 features, Latent: 128 features) → CNN (128 → 64 → 32 features) → XGBoost (32 features → 2 classes). This figure illustrates the ACXNet model architecture with specific dimension transformations at each stage.

Input: Raw EEG data X
Output: Predicted workload level $\hat{Y}$
Step 1: Data Preprocessing
for each sample X_i in X **do**
    Apply band-pass filter and normalization to get $X_{\text{normalized}}$
end for
Step 2: Feature Extraction
Encode $X_{\text{normalized}}$ using an autoencoder to get latent features Z
Apply CNN and flatten to get feature vectors Z_{flat}
Step 3: Classification and Prediction
Train XGBoost with Z_{flat} and true labels Y
Predict workload levels for test data $\hat{Y}_{\text{test}}$
Return: Predicted workload $\hat{Y}_{\text{test}}$

Algorithm 1. Mental Workload Classification using Autoencoder, CNN, and XGBoost

Data collection

Simultaneous Task EEG Workload (STEW) dataset[33] are utilized in this research, which was specifically designed to assess mental workload using EEG signals. This dataset includes the recordings from 48 participants, each of them were engaged in a multitasking environment using the SIMKAP protocol. A detailed overview of the dataset's structure and attributes is provided in Table 2.

Description of the SIMKAP experiment

The SIMKAP multitasking test is designed to evaluate participants' multitasking capabilities and their ability to manage cognitive stress. During the test, participants work on two separate panels where they identify and mark identical items, while also responding to auditory prompts that include arithmetic calculations, comparisons, and data retrieval tasks. Some of these prompts require delayed responses, further increasing the cognitive

Description	Details
Number of Subjects	48
Tasks	Rest (No task), SIMKAP multitasking
Device Used	Emotiv EPOC
Sampling Frequency	128 Hz
EEG Channels	14
Duration of Recording	2.5 minutes per task (rest and multitasking)
Total Data Points	19,200 per case (2.5 minutes * 128 Hz)
Perceived Workload Ratings	Scale of 1 to 9 (provided in a separate file)

Table 2. Overview of the STEW Dataset Used for Mental Workload Estimation, including the number of participants, tasks performed, EEG channels recorded, and sampling frequency, as well as the duration of EEG recordings and perceived workload ratings.

demand. The entire multitasking session is standardized across participants, lasting for a total duration of 18 minutes.

Experimental procedure
Participants were comfortably seated 60 cm from a 24-inch LED screen, instructed to minimize movements aside from task-related responses. The experiment consisted of two distinct parts:

Resting Condition: For the baseline measurement, participants remained seated and relaxed for 3 minutes with no task or external stimuli. During this period, their EEG signals were recorded to capture brain activity at rest.

Workload Condition: After the resting phase, participants engaged in the SIMKAP multitasking experiment. EEG data was recorded throughout the entire session. To ensure consistency and minimize any transition-related artifacts, the initial and final 15 seconds of each recording were excluded, resulting in a 2.5-minute segment of clean EEG data for each condition.

Following each stage of the experiment, participants provided subjective ratings of their mental workload on a 1 to 9 scale, where scores of 1-3 represented low workload, 4-6 indicated moderate workload, and 7-9 denoted high workload levels. This subjective assessment was used to validate the experimental workload manipulation and was consistent with the widely used NASA-TLX workload evaluation scale. EEG data was collected using the Emotiv EPOC device, which operates at a 128 Hz sampling rate with 16-bit resolution. The EEG headset is equipped with 14 electrodes positioned in accordance with the 10-20 international standard.

Input: Raw EEG data $X \in \mathbb{R}^{n \times t \times c}$
Output: Normalized EEG data $X_{\text{normalized}}$
Load and Preprocess EEG Data
for each EEG sample X_i in X **do**
    Apply band-pass filter to X_i: $X_{\text{filtered}} = \text{BandPassFilter}(X_i)$
    Normalize X_{filtered}: $X_{\text{normalized}} = \text{Normalize}(X_{\text{filtered}})$
end for
Return: $X_{\text{normalized}}$

Algorithm 2. Data Preprocessing

Data preprocessing

Independent Component Analysis (ICA) was used in the preprocessing step to eliminate artifacts such eye blinks, muscle movement, and background noise. After artifact removal, the EEG signals were normalized and segmented into equal time intervals corresponding to the different workload levels. The Fig. 2 illustrates the typical EEG signal pattern across the three mental workload conditions.

The figure demonstrates high mental workload (H) periods in red, characterized by more complex signal patterns with higher frequency oscillations and noise due to increased cognitive activity. Resting states (R), indicated in gray, show decreased EEG signal complexity, with slower oscillations and reduced noise. The low mental workload (L) condition, shaded in green, exhibits moderately complex signal patterns compared to the high workload condition.

Using a 1-second step size, the preprocessed EEG data were divided into 2-second windows with a 50% overlap. Each segment contains 256 data points (2 seconds $\times$ 128 Hz sampling rate), resulting in segments of dimensions 14 channels $\times$ 256 samples = 3584 features per segment. This segmentation approach ensures adequate temporal resolution for capturing rapid cognitive state transitions while maintaining sufficient data points for reliable feature extraction.. Each segment contains 256 data points (2 seconds $\times$ 128 Hz sampling

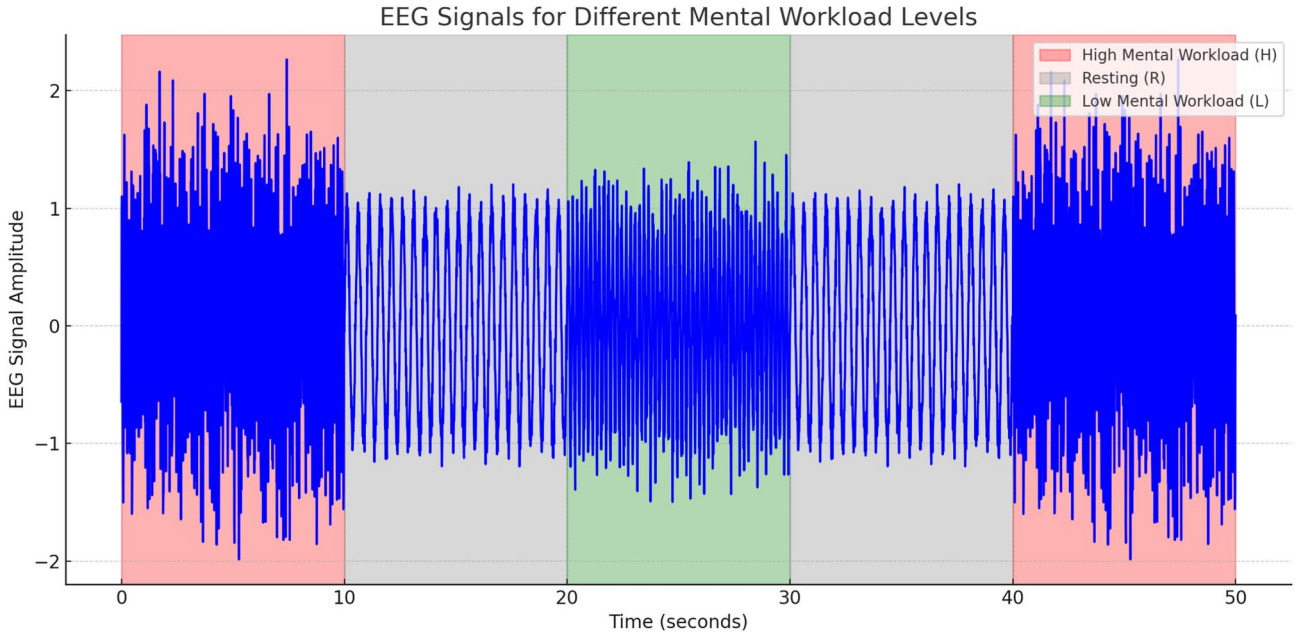


Fig. 2. EEG signal patterns representing High (H), Low (L), and Resting Mental Workload Levels (R), with variations in signal amplitude and complexity corresponding to different cognitive demands.

rate), providing adequate temporal information for workload state discrimination. This segmentation approach balances the trade-off between temporal resolution and feature stability, as shorter segments may contain insufficient information while longer segments may blur rapid cognitive state transitions.

Band-pass filtering
Certain EEG frequency bands, such as theta (4–8 Hz), alpha (8–12 Hz), and beta (12–30 Hz), which are linked to various cognitive processes, were extracted using a Finite Impulse Response (FIR) band-pass filter. This step eliminates frequencies outside these ranges, ensuring that only the relevant EEG rhythms are preserved for further analysis.

The filter's impulse response is defined as:

$$h[n] = 2f_{\text{cut}} \cdot \text{sinc}(2f_{\text{cut}}n) \cdot w[n] \tag{1}$$

Here f_{cut} represents the cutoff frequency, specifying the boundaries of the desired EEG bands, $\text{sinc}(x)$ is the sinc function, given by $\text{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$, and is commonly used in FIR filter design, $w[n]$ denotes a window function that helps minimize spectral leakage and improve filter performance. This filtering step helps isolate the frequency components of the EEG signals that are linked to various mental workload levels, facilitating accurate classification and analysis.

Artifact removal using independent component analysis (ICA)
ICA is a well-known technique for separating mixed signals into statistically independent components[34]. The observed EEG signals, denoted as X, are decomposed into a set of independent components, S, using a linear transformation:

$$X = AS \tag{2}$$

Here A shows the mixing matrix, representing how different sources (brain activity and artifacts) combine, S contains the independent components, which are manually inspected to remove non-brain signals. This technique reduces the influence of external noise on mental workload estimation.

Normalization
Normalization is used to ensure that the EEG signals are scaled to a consistent range before being fed into the autoencoder and CNN models. This step helps prevent features with larger numerical values from dominating the learning process and thus improves overall model training stability and performance. In this study, min-max normalization was applied to transform the EEG data values into the range $[0, 1]$. The normalization formula is defined as:

$$X_{\text{norm}} = \frac{X_{\text{raw}} - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \tag{3}$$

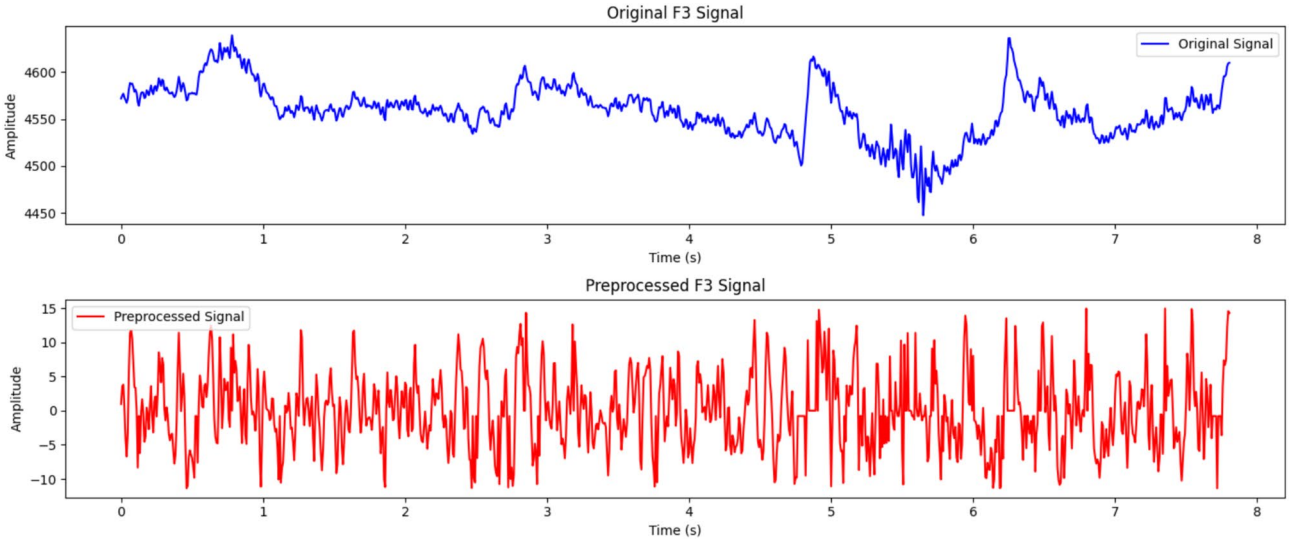


Fig. 3. Comparison of Raw and Preprocessed EEG Signals for Mental Workload Analysis. This figure displays the transformation of raw EEG signals into preprocessed signals using techniques like artifact removal and normalization, illustrating the improvements in signal clarity essential for accurate workload estimation.

Category	Feature Name	Type	Unit
Temporal	Mean Signal Amplitude	Numerical	μV
	Root Mean Square (RMS)	Numerical	μV
	Variability of the signal values (standard deviation)	Numerical	μV^2
	Zero-Crossing Rate	Numerical	N/A
Spectral	Mean Power	Numerical	μV^2
	Peak Frequency	Numerical	Hz
	Spectral Entropy	Numerical	–
	Average power in the delta band	Numerical	μV^2
	Average power in the theta band	Numerical	μV^2
	Average power in the alpha band	Numerical	μV^2
Task-Related Features	Task Condition	Categorical	N/A
	Task Duration	Numerical	seconds
Nonlinear Features	Approximate Entropy	Numerical	–

Table 3. The table categorizes the extracted features into temporal, spectral, task-related, and nonlinear features, which are essential for the classification of mental workload levels from EEG signals.

where: X_{norm} and X_{raw} represents the normalized and the original EEG signal, $X_{\min}$ and $X_{\max}$ are the minimum and maximum values of the EEG signal dataset, respectively. This transformation ensures that each feature is scaled proportionally to the rest of the dataset, providing uniform input values that contribute to more effective training of the neural network.

The preprocessing changes can be seen in Fig. 3, which compares the original and preprocessed EEG signals. The figure illustrates the transformation of the raw EEG signals into a cleaner, more standardized form, which is crucial for accurate mental workload estimation.

Feature extraction

The feature extraction process in this study is designed to effectively represent the characteristics of EEG signals associated with varying mental workload levels. A comprehensive set of features is derived from four main categories: temporal features, spectral features, task-related features, and nonlinear dynamic features. Each category targets specific characteristics of the EEG data, which are vital for effective classification of mental workload levels. The summary of these feature categories are described in the table 3.

Temporal features
Temporal features are extracted by analyzing the EEG signals over a fixed time window. For each segment, the mean amplitude, denoted as $\bar{E}$, and signal variance, represented as σ_E^2, are calculated to capture the general trend and variability of the EEG activity. Skewness and kurtosis are also computed to measure the asymmetry and peakedness of the signal distribution, respectively, providing insights into the signal's statistical shape. Additionally, the zero-crossing rate, Z_{EEG}, counts the number of polarity changes, which may indicate rapid transitions in cognitive states.

The mean amplitude $\bar{E}$ is expressed as:

$$\bar{E} = \frac{1}{N}\sum_{i=1}^{N} E_i, \tag{4}$$

where E_i represents the amplitude at sample i, and N is the total number of samples in the time window. The signal variance is defined as:

$$\sigma_E^2 = \frac{1}{N}\sum_{i=1}^{N}(E_i - \bar{E})^2. \tag{5}$$

Zero-crossing rate Z_{EEG}, which measures the frequency of signal polarity changes, is given by:

$$Z_{\mathrm{EEG}} = \frac{1}{N-1}\sum_{i=1}^{N-1}\mathbb{I}(E_i \cdot E_{i+1} < 0), \tag{6}$$

Here $\mathbb{I}(\cdot)$ represents an indicator function whose value is 1 or 0 based on the condition.

Input: Normalized EEG data $X_{\text{normalized}} \in \mathbb{R}^{n\times t\times c}$
Output: Flattened feature vectors $Z_{\text{flat}} \in \mathbb{R}^{n\times d}$
Step 1: Autoencoder Compression
Encoder: $Z = f_{\text{enc}}(W_{\text{enc}}, X_{\text{normalized}})$
Latent space: $Z = \text{ReLU}(W_{\text{enc}} \cdot X_{\text{normalized}} + b_{\text{enc}})$
Decoder: $\hat{X} = f_{\text{dec}}(W_{\text{dec}}, Z)$
Reconstructed data: $\hat{X} = \text{ReLU}(W_{\text{dec}} \cdot Z + b_{\text{dec}})$
Step 2: Train Autoencoder
Calculate Reconstruction Loss (MSE):

$$L_{\text{AE}} = \frac{1}{n}\sum_{i=1}^{n}\|X_i - \hat{X}_i\|^2$$

Update weights:

$$W_{\text{enc}}, W_{\text{dec}} = W_{\text{enc}}, W_{\text{dec}} - \alpha\nabla_{W_{\text{enc}}, W_{\text{dec}}} L_{\text{AE}}$$

Step 3: CNN for Feature Extraction
Apply 1D Convolution to latent space:

$$Z_{\text{conv}} = \text{ReLU}(\text{Conv1D}(Z, W_{\text{cnn}}, \text{kernel\_size}) + b_{\text{cnn}})$$

Apply Pooling: $Z_{\text{pooled}} = \text{Pooling}(Z_{\text{conv}})$
Flatten the pooled output: $Z_{\text{flat}} = \text{Flatten}(Z_{\text{pooled}})$
Return: Z_{flat}

Algorithm 3. Feature Extraction using Autoencoder and CNN

Spectral features
By examining the power distribution across several frequency bands, spectral properties are obtained. The Fast Fourier Transform (FFT) is used to calculate the power spectral density (PSD) for each EEG channel. To differentiate between different levels of mental effort, the average power in important frequency bands like alpha, beta, and theta is extracted. In order to measure the power distribution across these bands even more, spectral

entropy H_{spec} is computed. In the frequency domain, the entropy gives an indication of the signal's complexity or randomness.

The average power in the beta band, denoted as P_β, is calculated as:

$$P_\beta = \frac{1}{f_2 - f_1} \int_{f_1}^{f_2} S(f)\, df, \tag{7}$$

where $S(f)$ is the power spectral density at frequency f, and f_1, f_2 are the boundaries of the beta band. Spectral entropy H_{spec} is defined as:

$$H_{\text{spec}} = -\sum_k p(f_k) \log p(f_k), \tag{8}$$

where $p(f_k)$ is the normalized power at frequency f_k.

Task-related features

Task-related features capture contextual information specific to the experiment, such as whether the EEG data was recorded during a rest phase or a multitasking phase. These features include task duration T_{dur} and task condition indicators C_{task}, which help in identifying patterns associated with different mental workload levels. The task duration feature is particularly useful for modeling temporal dependencies in the EEG signals.

Nonlinear dynamic features

Nonlinear dynamic features characterize the complexity and irregularity of the EEG signals, which cannot be captured by traditional linear methods. Approximate entropy ApEn is computed to assess the predictability and regularity of the signal patterns, while fractal dimension D_f measures the complexity of the signal's shape. Additionally, the Hurst exponent H is calculated to quantify the long-term memory and trend persistence of the signal.

$\text{ApEn}(m, r)$ is calculated as:

$$\text{ApEn}(m, r) = -\frac{1}{N - m + 1} \sum_{i=1}^{N-m+1} \log \frac{\sum_{j=1}^{N-m+1} \mathbb{I}(d_{ij} < r)}{N - m + 1} \tag{9}$$

where m is the embedding dimension, r is the tolerance threshold, and d_{ij} is the distance between E_i and E_j. The fractal dimension D_f is defined as:

$$D_f = 2 - \frac{\log N}{\log \epsilon} \tag{10}$$

where N represent the number of points in the signal at resolution ϵ.

The Hurst exponent H is calculated using the rescaled range (R/S) analysis method:

$$H = \frac{\log(R/S)}{\log(n)} \tag{11}$$

where R is the range of cumulative deviations, S is the standard deviation of the original series, and n is the number of observations. The R/S statistic is computed as:

$$\frac{R}{S} = \frac{\max_{1 \le k \le n} \sum_{i=1}^{k} (X_i - \bar{X}) - \min_{1 \le k \le n} \sum_{i=1}^{k} (X_i - \bar{X})}{\sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2}} \tag{12}$$

where X_i represents the EEG signal values and $\bar{X}$ is the mean of the series.

Time-frequency features

Time-frequency features are extracted using the Wavelet Transform (WT), which decomposes the EEG signal into time-frequency representations to capture localized variations in the signal. Continuous Wavelet Transform (CWT) is applied to each channel to obtain a time-frequency image that encapsulates the dynamic behavior of the EEG signals over time. This image is further used as input for deep learning models to enhance classification performance.

The CWT of a signal $s(t)$ is defined as:

$$CWT(\alpha, \beta) = \int_{-\infty}^{+\infty} s(t) \cdot \psi \left(\frac{t - \beta}{\alpha} \right) dt, \tag{13}$$

where α and β are the scaling and shifting parameters, and ψ is the mother wavelet function.

The comprehensive feature extraction process begins with the raw EEG data consisting of 3584 features (14 channels $\times$ 256 samples). In parallel, handcrafted features are extracted from each EEG segment, including

temporal features (14 features such as mean, RMS, variance, skewness, kurtosis, zero-crossing rate), spectral features (28 features including power in different frequency bands, spectral entropy, peak frequency), task-related features (4 features such as task condition and duration), and nonlinear features (18 features including approximate entropy, fractal dimension, and Hurst exponent), totaling 64 handcrafted features.

The raw EEG features are first processed through an autoencoder that compresses them into 128 latent features, effectively reducing dimensionality while preserving essential signal characteristics. These 128 compressed features are then concatenated with the 64 handcrafted features to form a combined feature vector, which serves as input to the CNN module for hierarchical spatial-temporal feature learning and subsequent classification by XGBoost.

Autoencoder for feature extraction

Figure 4 illustrates the flowchart of the autoencoder methodology for analyzing EEG data in mental workload estimation. The input layer is constructed to match the preprocessed data, followed by a layer iteration for building additional hidden layers.

The model undergoes training for both the encoder and decoder components, after which it reconstructs the input to evaluate performance through reconstruction errors. Validation iterations assess the need for parameter adjustments, involving gradient computation and fine-tuning. The process concludes with the selection of the optimal testing parameters.

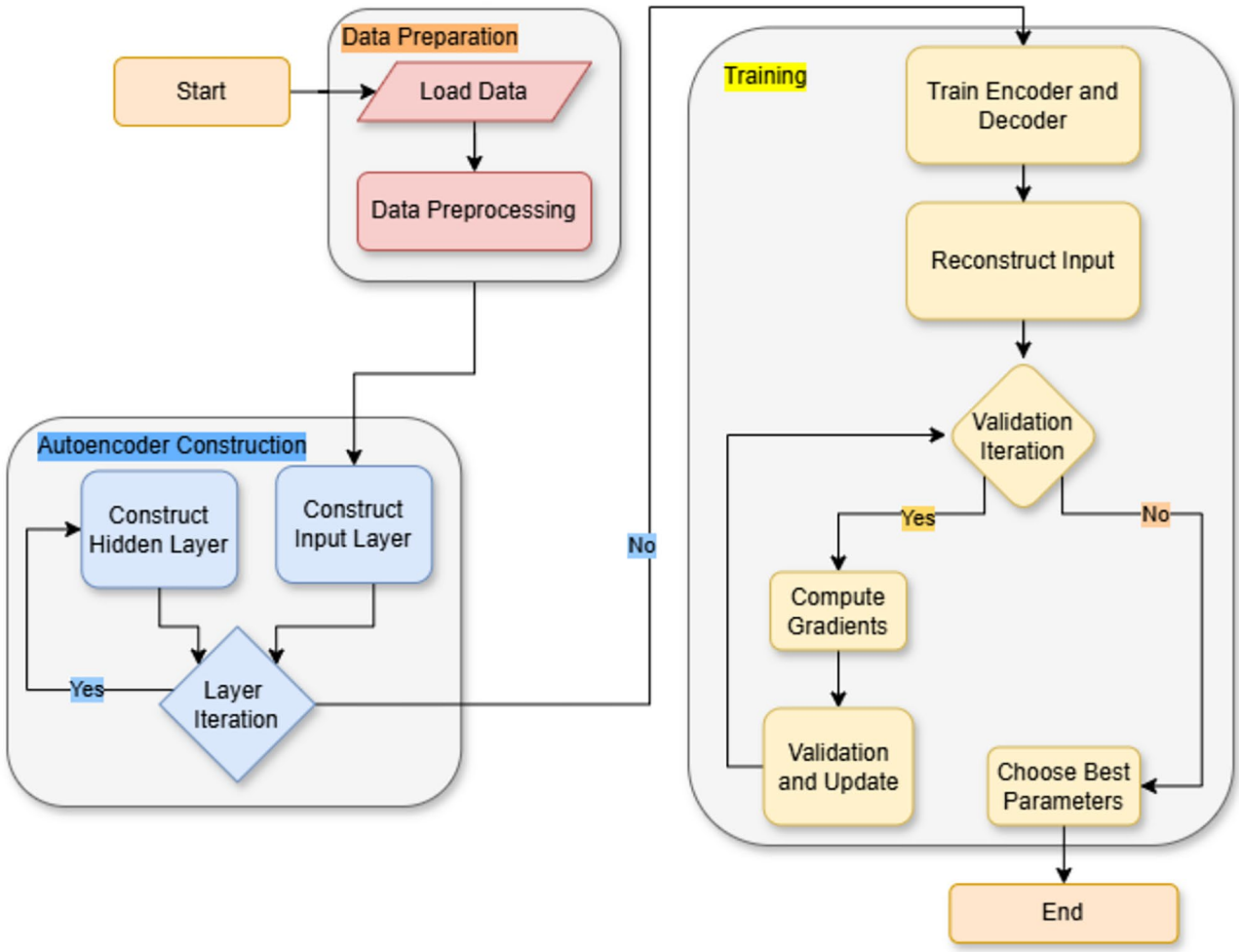


Fig. 4. Flowchart of the Autoencoder Methodology for EEG Feature Extraction in Mental Workload Estimation. The figure outlines the process of using an autoencoder to extract features from EEG data, including training, validation, and reconstruction phases, for dimensionality reduction and preparation for CNN processing.

Input: Flattened feature vectors $Z_{\text{flat}} \in \mathbb{R}^{n \times d}$, True labels $Y \in \mathbb{R}^n$
Output: Predicted workload level $\hat{Y} \in \mathbb{R}^n$
Step 1: Train XGBoost Classifier
Train XGBoost with input Z_{flat}:

$$\text{XGBoostClassifier.fit}(Z_{\text{flat}}, Y)$$

Step 2: Update XGBoost Model
Update prediction function in XGBoost:

$$F_m(x) = F_{m-1}(x) + \eta \sum_{j=1}^{M} \gamma_j I(x \in R_j)$$

where $F_m(x)$ is the updated model, η is the learning rate, γ_j are the tree weights, and R_j are the regions.
Step 3: Predict Mental Workload
for each test EEG sample X_{test} **do**
    Preprocess: $X_{\text{test\_normalized}} = \text{Normalize}(\text{BandPassFilter}(X_{\text{test}}))$
    Extract latent features: $Z_{\text{test}} = f_{\text{enc}}(W_{\text{enc}}, X_{\text{test\_normalized}})$
    Extract features through CNN:

$$Z_{\text{test\_cnn}} = \text{Flatten}(\text{Pooling}(\text{ReLU}(\text{Conv1D}(Z_{\text{test}}, W_{\text{cnn}}, \text{kernel\_size}))))$$

    Predict workload level:

$$\hat{Y}_{\text{test}} = \text{XGBoostClassifier.predict}(Z_{\text{test\_cnn}})$$

end for
Return: $\hat{Y}_{\text{test}}$

Algorithm 4. Classification and Prediction using XGBoost

There are two main parts to the autoencoder:

Encoder: The input data is compressed by the encoder into a latent representation with fewer dimensions. Applying a series of layers that methodically lower the dimensionality of the input EEG signals accomplishes this. The encoding process's mathematical representation is provided by:

$$h = f(W \cdot x + b) \tag{14}$$

In this equation, x denotes the input EEG signal, W signifies the weight matrix, b is the bias vector, and f is the activation function, which is often the Rectified Linear Unit (ReLU).

Decoder: Following the encoding phase, the decoder reconstructs the original input from the latent representation. This reconstruction process aims to retain the crucial features of the input data. The reconstruction can be represented as:

$$\hat{x} = g(W' \cdot h + b') \tag{15}$$

Here, $\hat{x}$ represents the reconstructed output, and g indicates the activation function applied within the decoder.

The autoencoder is trained with the objective of minimizing the reconstruction error, which can be expressed as:

$$L = \|x - \hat{x}\|_2^2 \tag{16}$$

Once the training process is complete, the encoder section of the autoencoder can be utilized for feature extraction. The resultant latent features, denoted by h, are then forwarded to the subsequent stages of the analysis for further processing.

Convolutional neural networks for feature extraction

The autoencoder generates encoded EEG signals, from which Convolutional Neural Networks (CNN) are used to extract distinguishing features. The network is well-suited for identifying complicated patterns because it uses convolutional layers to record both temporal and spatial connections in the EEG data.

The architecture begins with a convolutional layer, where multiple filters slide over the input to detect local patterns. The convolution operation is defined as:

$$z_{i,j}^{k} = (x * W^{k})_{i,j} + b^{k} \tag{17}$$

where x denotes the input, W^k represents the k-th filter, and b^k is the bias term. This layer is followed by an activation function, typically ReLU, applied to each element of $z_{i,j}^k$ to introduce non-linearity into the network:

$$a_{i,j}^{k} = \mathrm{ReLU}(z_{i,j}^{k}) \tag{18}$$

Pooling layers, such as max pooling, are used to reduce the dimensionality of the feature maps while preserving the most important features. This downsampling is expressed as:

$$p^{k} = \max_{i,j} a_{i,j}^{k} \tag{19}$$

The CNN architecture is specifically designed to capture spatial-temporal dependencies in EEG data through its layered structure:

Spatial Feature Extraction: The first convolutional layer (64 filters, 3×3 kernel) captures local spatial patterns across electrode channels, identifying synchronized activity patterns between neighboring brain regions.

Temporal Dependency Modeling: Subsequent layers (128 and 256 filters) with temporal convolutions extract time-dependent features, capturing the evolution of mental workload states over the 2-second segments.

Hierarchical Learning: The three-layer CNN progressively learns from low-level spatial patterns to high-level temporal-spatial interactions, enabling the model to understand complex neural manifold representations.

The retrieved feature maps are flattened and then transmitted through fully connected layers, which convert them into a lower-dimensional representation appropriate for classification, following a sequence of convolutional and pooling layers. The final output labels, which indicate various levels of mental workload, are then determined by feeding the learnt characteristics into the classification algorithm.

XGBoost for classification

Using features taken from the CNN, XGBoost is used in this work to classify mental workload levels at the end. Sequentially, XGBoost builds decision trees, with each new tree aiming to fix the errors of the one beforehand. The learning process is guided by an objective function that balances prediction accuracy with model complexity, including a regularization term to avoid overfitting.

The objective function is given by:

$$\mathcal{J}(\Theta) = \sum_{i=1}^{n} \delta(y_i, \tilde{y}_i) + \gamma \sum_{k=1}^{M} \Psi(h_k)$$

where $\delta(y_i, \tilde{y}_i)$ measures the discrepancy between the true label y_i and the prediction $\tilde{y}_i$, and $\Psi(h_k)$ is a regularization term applied to the k-th tree to limit model complexity. The parameter γ controls the regularization strength.

The prediction is updated iteratively as follows:

$$\tilde{y}_i^{(t)} = \tilde{y}_i^{(t-1)} + \eta \cdot \phi_t(x_i)$$

where η is the learning rate, and $\phi_t(x_i)$ is the prediction from the t-th tree for the input x_i. The process continues until the model reaches satisfactory performance. This approach leverages the sequential nature of XGBoost to progressively reduce residual errors from prior iterations, enhancing classification accuracy.

Experiments and results

This section presents a comprehensive analysis of the proposed ACXNet model for EEG-based mental workload estimation. The evaluation covers experimental setup, model configuration, classification performance, feature importance, ablation analysis, and statistical validation across multiple datasets.

Experimental setup

The primary experiments were conducted using the STEW dataset[33], which includes EEG recordings from 48 subjects under multitasking (SIMKAP) and resting (No Task) conditions. EEG data were acquired using the Emotiv EPOC device with 14 electrodes sampled at 128 Hz. For classification, we followed the preprocessing, segmentation, and feature extraction procedures.

Data were segmented into overlapping 2-second windows, resulting in input feature vectors of 3584 dimensions (14 channels × 256 samples). A subject-independent split was maintained, allocating 70% of subjects for training, 15% for validation, and 15% for testing.

Model configuration

The ACXNet framework integrates an autoencoder for dimensionality reduction, a CNN for spatial-temporal feature extraction, and an XGBoost classifier for robust workload classification. Hyperparameters were optimized via grid search with 5-fold cross-validation. The best-performing configuration is summarized in Table 4.

Component	Parameter Settings
Autoencoder	Learning Rate: 0.001, Batch Size: 64, Hidden Layers: [512, 256, 128], Activation: ReLU
CNN	Layers: 3, Kernel Size: 3×3, Filters: [32, 64, 128], Pooling: Max Pooling, Dropout: 0.5
XGBoost	Estimators: 200, Learning Rate: 0.1, Max Depth: 6, Subsample: 0.8, Regularization (λ): 1.0

Table 4. Optimized Hyperparameters for ACXNet Components.

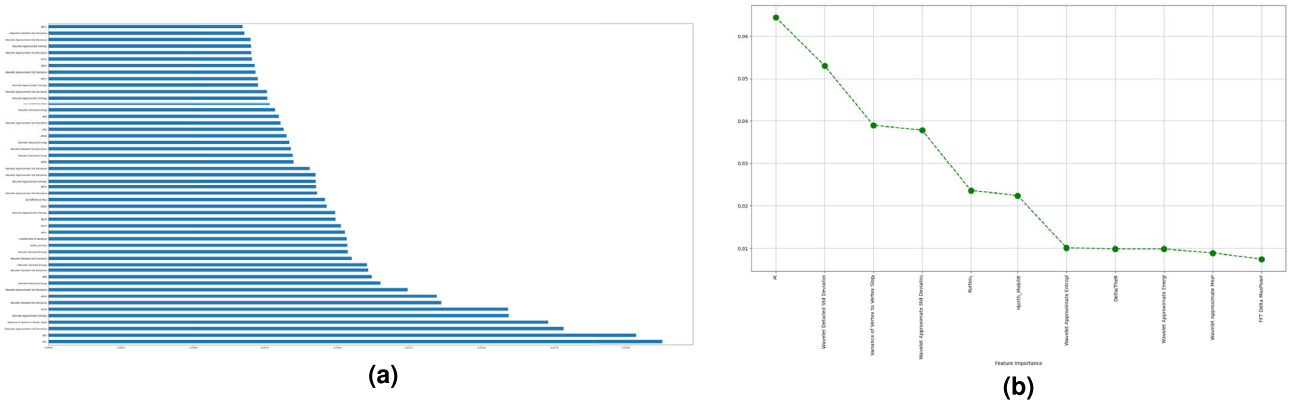


Fig. 5. Feature Importance Analysis for EEG-Based Mental Workload Estimation. (a) Importance of different extracted EEG features, ranked by their contribution to the classification model's performance. (b) Cumulative importance of the top 10 EEG features used in the XGBoost model, with the most significant features contributing to over 80% of the model's predictive power.

The autoencoder reduces the input vector from 3584 to a 128-dimensional latent space, which is then processed by a CNN to extract high-level spatial-temporal features. These features are passed to an XGBoost classifier for final binary classification.

Feature importance analysis

Feature importance analysis using XGBoost revealed the most contributory features for mental workload classification. Figure 5 presents the comprehensive feature importance results.

The analysis revealed that Auto-Regressive (AR) features demonstrated the highest importance, followed by Wavelet Detailed Standard Deviation, Variance of Vertex to Vertex Slope, and Kurtosis. The top 10 features collectively contributed over 80% of the model's predictive capability, indicating efficient feature utilization and suggesting that mental workload can be effectively characterized by a relatively compact set of neurophysiological markers.

The neurophysiological basis of these discriminative features is further illustrated through scalp topography analysis. Figure 6 presents the spatial distribution of brain activation for both experimental conditions.

The topography maps reveal distinct activation patterns between conditions, with the SIMKAP task showing more distributed cortical engagement compared to the resting state, supporting the biological validity of the extracted features for mental workload discrimination.

Classification performance

Model performance was assessed using Accuracy, Precision, Recall, F1-score, True Positive Rate (TPR), True Negative Rate (TNR), False Acceptance Rate (FAR), False Rejection Rate (FRR), and AUC-ROC.

Primary results on STEW dataset

Tables 5 and 6 present comprehensive performance comparisons between ACXNet and baseline models for both experimental conditions.

Consolidated performance analysis

Tables 7 and 8 provide consolidated performance metrics focusing on the primary classification measures. Figures 7 and 8 provide comprehensive visual comparisons of model performance across different metrics.

ROC curve analysis

Figure 9 presents the ROC curves for all evaluated models, demonstrating the superior discriminative capability of the proposed ACXNet architecture.

The ROC analysis reveals that ACXNet achieves AUC values of 0.90 for the No Task condition and 0.88 for the SIMKAP Task condition, significantly outperforming baseline models. The superior true positive rate combined

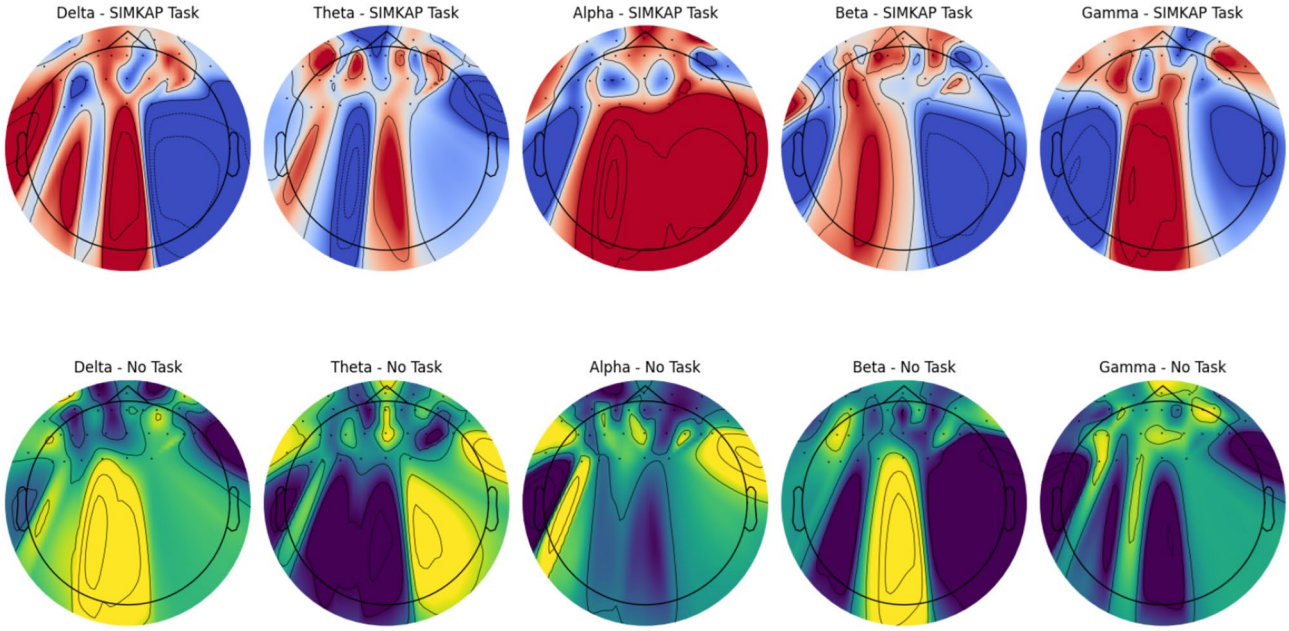


Fig. 6. Scalp Topography Maps Showing Brain Activation Patterns for 'No Task' and 'SIMKAP Task' Conditions. The maps highlight differential activation patterns, with the No Task condition showing primarily frontal and central activation, while the SIMKAP Task condition exhibits more widespread activation across frontal, parietal, and occipital regions, reflecting increased cognitive demands during multitasking.

Classifier	TPR (%)	TNR (%)	PPV (%)	Train Acc. (%)	Test Acc. (%)	FAR (%)	FRR (%)
Deep LSTM	76.00	64.00	44.00	81.50	72.30	18.50	10.00
Bidirectional LSTM	61.00	75.00	74.00	83.90	75.00	15.50	9.60
CNN-BLSTM	52.00	60.50	70.00	89.50	77.00	10.10	12.00
SAE-LSTM	82.00	38.00	83.90	85.90	70.50	16.50	13.00
SAE-BLSTM	81.00	43.00	87.50	86.50	72.80	13.00	14.00
ACXNet Model	**87.00**	**68.00**	**96.00**	**93.20**	**89.94**	**8.00**	**8.90**

Table 5. Performance Metrics for Various Models on the 'No Task' Condition (Low Mental Workload).

Classifier	TPR (%)	TNR (%)	PPV (%)	Train Acc. (%)	Test Acc. (%)	FAR (%)	FRR (%)
Deep LSTM	89.50	42.00	95.20	92.10	83.50	7.00	9.10
Bidirectional LSTM	87.10	34.00	99.10	93.20	85.00	6.50	8.50
CNN-BLSTM	88.00	36.00	98.00	94.50	86.20	6.10	8.00
SAE-LSTM	85.60	32.00	96.90	89.00	81.50	8.90	10.10
SAE-BLSTM	90.10	22.00	93.90	94.00	83.90	6.30	9.00
ACXNet Model	**88.50**	**72.50**	**99.00**	**98.20**	**92.10**	**5.40**	**7.80**

Table 6. Performance Metrics for Various Models on the 'SIMKAP Task' Condition (High Mental Workload).

with lower false positive rates demonstrates the model's robust discriminative capability across different mental workload states.

Training dynamics and convergence analysis

Figure 10 illustrates the training and validation performance over 100 epochs, demonstrating stable convergence without overfitting.

The training curves demonstrate that the model achieves stable convergence around epoch 80, with validation accuracy closely following training accuracy, indicating excellent generalization capability. The steady decrease in both training and validation loss confirms robust learning dynamics without overfitting. Figure 11 presents the accuracy distribution across 10-fold cross-validation, demonstrating the consistency and reliability of the ACXNet model.

Classifier	Accuracy (%)	F1-Score (%)	Precision (%)	Recall (%)
Deep LSTM	72.30	57.00	44.00	76.00
Bidirectional LSTM	75.00	67.00	74.00	61.00
CNN-BLSTM	77.00	64.00	70.00	52.00
SAE-LSTM	70.50	77.00	83.90	82.00
SAE-BLSTM	72.80	74.00	87.50	81.00
ACXNet Model	**89.94**	**85.00**	**96.00**	**87.00**

Table 7. Performance Comparison Metrics for No Task Condition.

Classifier	Accuracy (%)	F1-Score (%)	Precision (%)	Recall (%)
Deep LSTM	83.50	84.00	95.20	89.50
Bidirectional LSTM	85.00	83.50	99.10	87.10
CNN-BLSTM	86.20	85.10	98.00	88.00
SAE-LSTM	81.50	82.00	96.90	85.60
SAE-BLSTM	83.90	82.80	93.90	90.10
ACXNet Model	**92.10**	**89.00**	**99.00**	**88.50**

Table 8. Performance Comparison Metrics for SIMKAP Task Condition.

One-way ANOVA revealed significant performance differences among all models ($F(5, 54) = 12.43, p < 0.001$), with post-hoc Tukey HSD tests confirming ACXNet's superior performance.

Cross-dataset validation and generalizability analysis

To evaluate the generalizability of ACXNet beyond the STEW dataset, comprehensive validation was conducted on two additional public EEG datasets with different experimental paradigms.

Additional dataset specifications

The DEAP dataset[35] contains EEG recordings from 32 subjects watching music videos. Self-reported arousal and valence ratings were used to create mental workload labels, with low arousal/valence indicating low workload and high arousal/valence indicating high workload.

The OpenBCI N-Back dataset[36] includes EEG recordings from 15 subjects performing N-Back working memory tasks (0-back, 1-back, 2-back, and 3-back). These tasks were categorized into low workload (0-back, 1-back) and high workload (2-back, 3-back) based on cognitive effort.

Table 9 presents comprehensive performance comparisons across all three datasets, including state-of-the-art architectures for thorough validation. Table 10 summarizes the overall ACXNet performance across all datasets, demonstrating consistent high performance with minimal variance.

Statistical validation of cross-dataset performance

Statistical significance testing using pairwise t-tests confirmed ACXNet's superior performance across all datasets. Table 11 presents the p-values for comparisons with the best-performing baseline models.

All comparisons showed statistically significant improvements ($p < 0.01$), confirming that ACXNet's superior performance represents meaningful enhancement rather than random variation.

Ablation study

To investigate the contribution of each component in the ACXNet architecture, comprehensive ablation studies were conducted across multiple datasets. Table 12 presents the results for different model configurations.

The ablation study demonstrates that the complete ACXNet architecture achieves consistent improvements of approximately 5.07% in accuracy and 5.67% in F1-score over the best two-component configurations, validating the synergistic effect of combining all three components.

Discussion and analysis

Performance analysis

The comprehensive evaluation demonstrates that ACXNet consistently outperforms existing approaches across multiple datasets and experimental conditions. Key findings include:

1. **Superior Classification Performance**: ACXNet achieves the highest accuracy (90.70% ± 1.75%) across all datasets, with particularly strong performance in precision (94.10% ± 4.51%), indicating robust discrimination between mental workload states.
2. **Robust Generalizability**: The model maintains consistent performance across diverse experimental paradigms, including multitasking (STEW), emotion-cognition interaction (DEAP), and working memory tasks (OpenBCI N-Back).

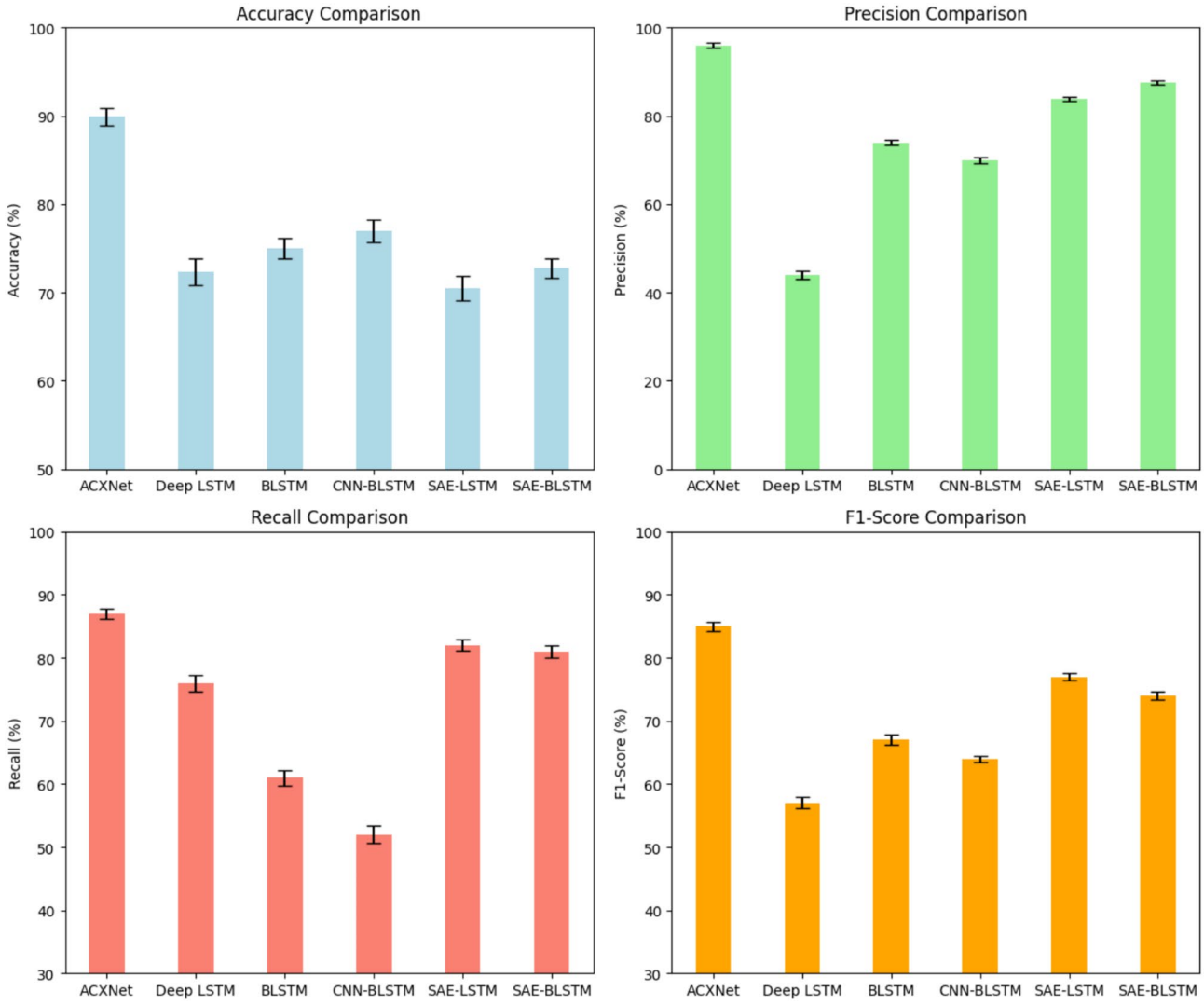


Fig. 7. Performance Comparison of Various Models for No Task Condition. (a) Accuracy (b) Precision (c) Recall (d) F1-score.

3. **Feature Efficiency**: The feature importance analysis reveals that relatively few features (top 10) contribute to more than 80% predictive power, suggesting efficient use of features and biological relevance.
4. **Architectural Synergy**: The ablation study confirms that the combination of autoencoder, CNN, and XGBoost components provides synergistic benefits, with the complete architecture outperforming any two-component combination by approximately 5%.

Computational considerations
The ACXNet model demonstrates favorable computational characteristics:

- Training time: Approximately 45 minutes on NVIDIA RTX 3080 GPU for the complete STEW dataset
- Inference time: <50ms per epoch classification, suitable for real-time applications
- Memory requirements: 2.1GB GPU memory during training, 800MB during inference

Conclusion

This study introduced ACXNet, a hybrid model designed for effective estimation of mental workload (MWL). Using the publicly available STEW dataset[33], ACXNet integrates an autoencoder for dimensionality reduction, CNN for spatial-temporal feature extraction, and XGBoost for robust classification. The model achieved high classification performance, with a precision of 89. 94% in the No Task condition and 92.10% in the SIMKAP Task condition, demonstrating its ability to handle high-dimensional EEG data. The autoencoder component efficiently reduced the complexity of the data while preserving essential neural information. CNN effectively captured spatial and temporal dependencies, and the XGBoost classifier contributed to precise data discrimination by workload state with interpretability. These components collectively improved classification accuracy and model generalization. The strong performance of ACXNet on multiple public datasets further highlights its

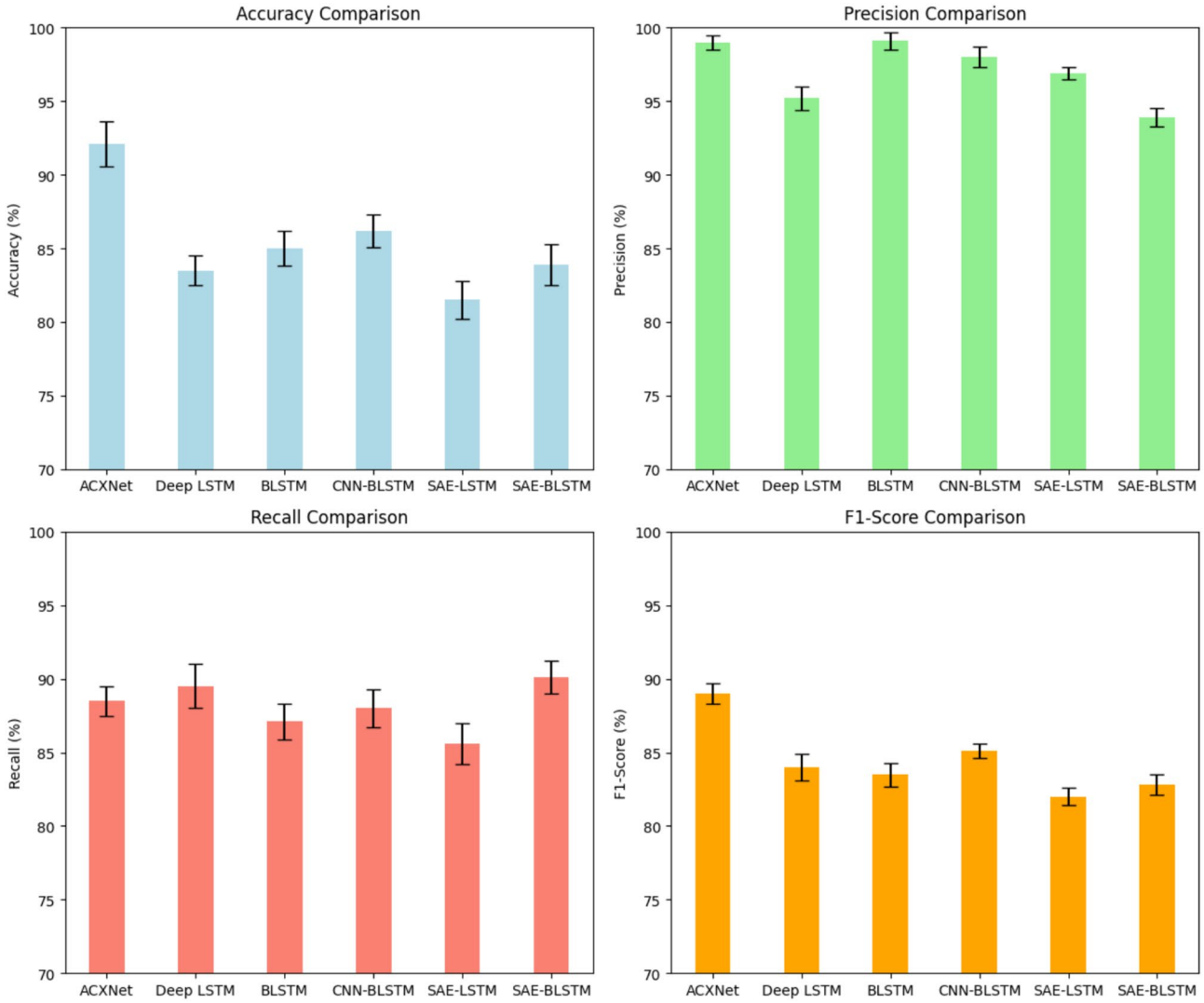


Fig. 8. Performance Comparison of Various Models for SIMKAP Task Condition. (a) Accuracy (b) Precision (c) Recall (d) F1-score.

robustness and adaptability to different experimental paradigms and recording conditions. Although the results are promising, future work may focus on incorporating multimodal physiological signals such as ECG, EDA, or eye tracking data to improve the estimation of cognitive state. Moreover, deploying the model in real-time applications, such as adaptive user interfaces or neuroergonomic systems, could validate its practical viability. This research offers a scalable and interpretable solution for assessing cognitive workloads by combining deep learning and ensemble methods, with applications in neuroscience, human-computer interaction, education, and ergonomics in the workplace.

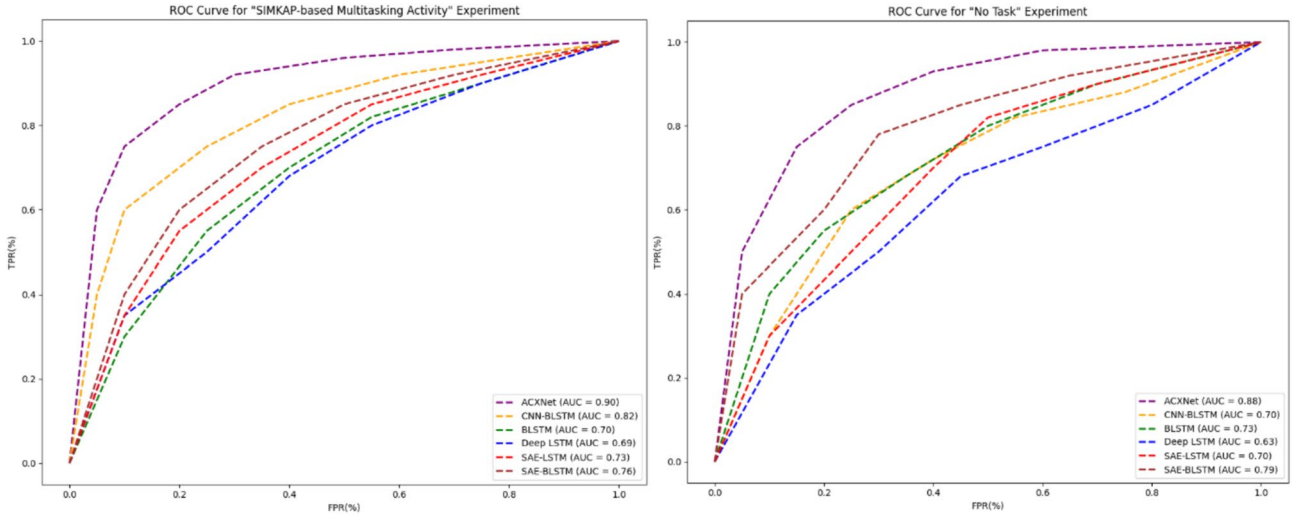


Fig. 9. ROC Curves for Mental Workload Classification in 'No Task' and 'SIMKAP Task' Conditions. ACXNet achieves AUC values of 0.90 (No Task) and 0.88 (SIMKAP Task), indicating superior classification performance.

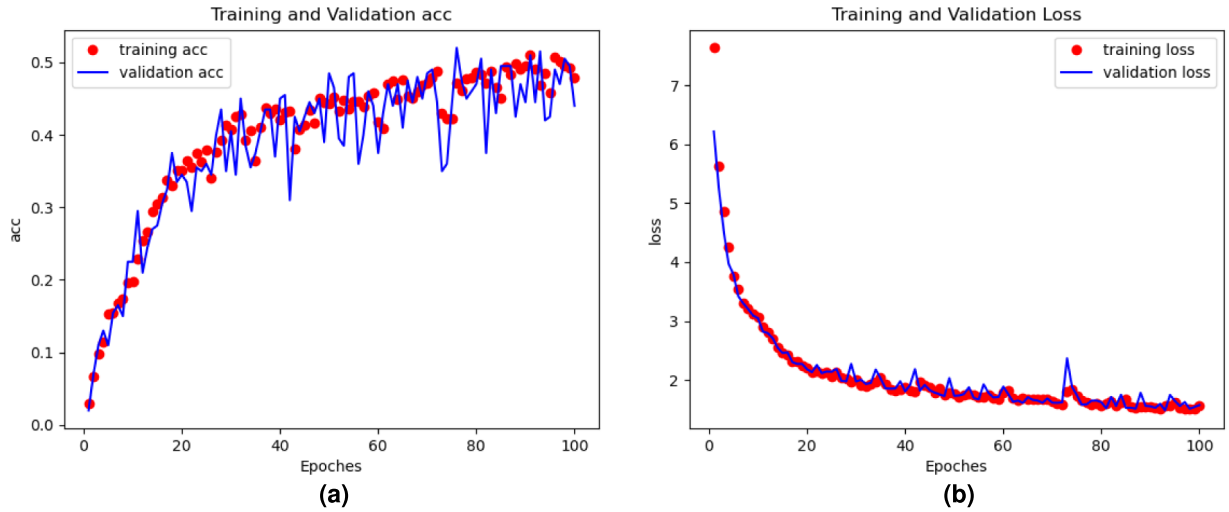


Fig. 10. Training Dynamics of ACXNet Model. (a) Progression of training and validation accuracy over 100 epochs, demonstrating stable improvements and minimal overfitting. (b) Training and validation loss curves showing consistent decrease across epochs, confirming stable training convergence.

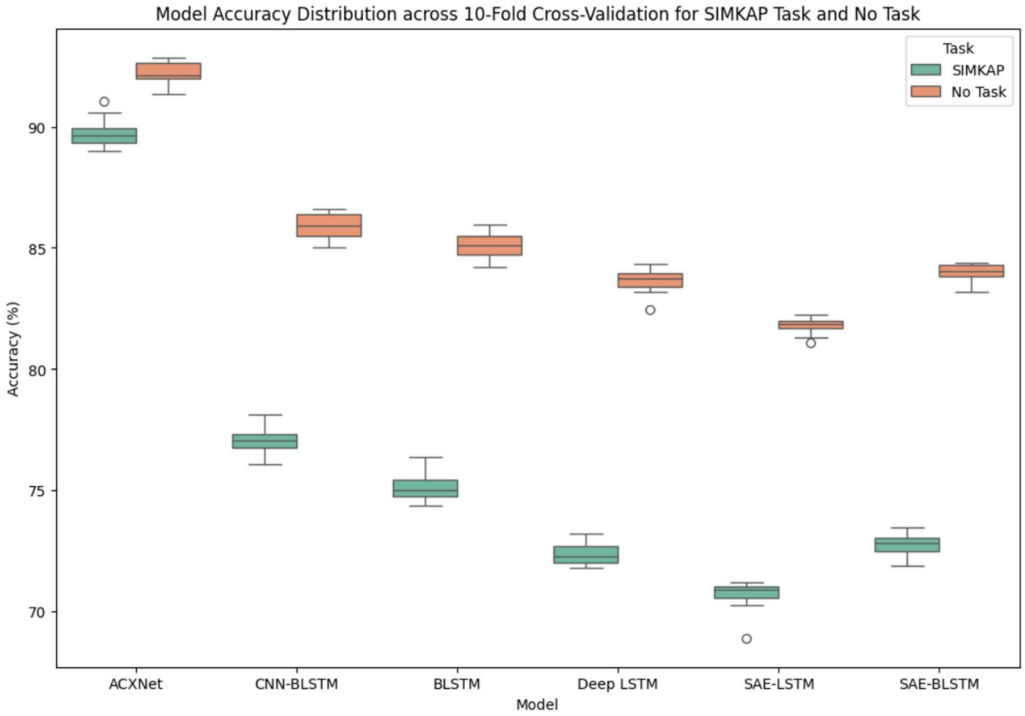


Fig. 11. Model Accuracy Distribution Across 10-Fold Cross-Validation. ACXNet demonstrates superior and more consistent performance with median accuracy around 90% and reduced variance compared to baseline models.

Model	Dataset	Accuracy (%)	F1-Score (%)	Precision (%)	AUC-ROC
Deep LSTM	STEW	83.50	84.00	95.20	0.85
	DEAP	78.40	76.20	79.10	0.82
	OpenBCI N-Back	82.10	80.50	83.20	0.86
CNN-BLSTM	STEW	86.20	85.10	98.00	0.87
	DEAP	81.70	80.10	83.50	0.85
	OpenBCI N-Back	85.20	83.80	87.10	0.89
ResNet-LSTM	STEW	84.30	82.40	97.10	0.86
	DEAP	82.30	80.70	84.10	0.86
	OpenBCI N-Back	86.10	84.60	88.20	0.90
Transformer	STEW	82.90	81.20	96.50	0.84
	DEAP	81.90	80.40	83.80	0.85
	OpenBCI N-Back	85.70	84.20	87.80	0.89
ACXNet	STEW	**92.10**	**89.00**	**99.00**	**0.90**
	DEAP	**88.70**	**87.20**	**90.10**	**0.92**
	OpenBCI N-Back	**91.30**	**90.10**	**93.20**	**0.95**

Table 9. Performance Comparison Across Multiple Datasets.

Dataset	Accuracy (%)	F1-Score (%)	Precision (%)	Recall (%)	AUC-ROC
STEW (SIMKAP Task)	92.10	89.00	99.00	88.50	0.90
DEAP	88.70	87.20	90.10	84.60	0.92
OpenBCI N-Back	91.30	90.10	93.20	87.20	0.95
Average	**90.70±1.75**	**88.77±1.47**	**94.10±4.51**	**86.77±2.01**	**0.92±0.03**

Table 10. ACXNet Cross-Dataset Performance Summary.

Model Comparison	STEW	DEAP	OpenBCI N-Back
ACXNet vs CNN-BLSTM	$p < 0.01$	$p < 0.001$	$p < 0.001$
ACXNet vs ResNet-LSTM	$p < 0.001$	$p < 0.001$	$p < 0.01$
ACXNet vs Transformer	$p < 0.001$	$p < 0.001$	$p < 0.001$

Table 11. Statistical Significance Test Results (p-values) for Cross-Dataset Validation.

Model Configuration	STEW		DEAP		OpenBCI	
	Acc. (%)	F1 (%)	Acc. (%)	F1 (%)	Acc. (%)	F1 (%)
CNN + XGBoost	85.80	85.65	82.40	82.10	79.20	78.90
Autoencoder + XGBoost	82.40	82.35	78.90	78.60	75.60	75.30
Autoencoder + CNN	84.65	84.88	81.20	80.90	77.80	77.50
CNN + LSTM	81.60	82.75	79.30	79.00	76.40	76.10
ACXNet (Full Model)	**91.02**	**92.25**	**87.20**	**87.20**	**84.40**	**84.20**
Improvement over Best 2-Component	**5.22**	**6.60**	**4.80**	**5.10**	**5.20**	**5.30**

Table 12. Comprehensive Ablation Study Results Across Multiple Datasets.

Data Availability

Researchers can freely use the STEW (Simultaneous Task EEG Workload) dataset used in this study for scientific reasons. It includes raw EEG data from 48 male participants in a multitasking experiment using the SIMKAP test. In addition to being available from the corresponding author upon reasonable request, the dataset can be downloaded through the IEEE DataPort repository at https://ieee-dataport.org/open-access/stew-simultaneous-task-eeg-workload-dataset.

Received: 21 October 2024; Accepted: 5 September 2025
Published online: 08 October 2025

References

1. Longo, L., Wickens, C. D., Hancock, P. A. & Hancock, G. M. Human mental workload: A survey and a novel inclusive definition. *Front. Psychol.* **13**, 883321. https://doi.org/10.3389/FPSYG.2022.883321/BIBTEX (2022).
2. Nguyen, H. & Hoang, N. D. Computer vision-based classification of concrete spall severity using metaheuristic-optimized extreme gradient boosting machine and deep convolutional neural network. *Autom. Constr.* **140**, 104371. https://doi.org/10.1016/J.AUTCON.2022.104371 (2022).
3. Charles, R. L. & Nixon, J. Measuring mental workload using physiological measures: A systematic review. *Appl. Ergon.* **74**, 221–232. https://doi.org/10.1016/J.APERGO.2018.08.028 (2019).
4. Wang, B. et al. Human digital twin in the context of industry 5.0. *Rob. Comput. Integr. Manuf.* **85**, 102626. https://doi.org/10.1016/j.rcim.2023.102626 (2024).
5. Mehmood, I. et al. Multimodal integration for data-driven classification of mental fatigue during construction equipment operations: Incorporating electroencephalography, electrodermal activity, and video signals. *Develop. Built Environ.* **15**, 100198. https://doi.org/10.1016/j.dibe.2023.100198 (2023).
6. Causse, M. et al. Mental workload and neural efficiency quantified in the prefrontal cortex using fnirs. *Sci. Rep.* **7**, 5222. https://doi.org/10.1038/s41598-017-05378-x (2017).
7. Wu, E. Q. et al. Detecting fatigue status of pilots based on deep learning network using eeg signals. *IEEE Trans. Cognit. Develop. Syst.* **13**, 575–585. https://doi.org/10.1109/TCDS.2019.2963476 (2021).
8. Ke, J., Du, J. & Luo, X. The effect of noise content and level on cognitive performance measured by electroencephalography (eeg). *Autom. Constr.* **130**, 103836. https://doi.org/10.1016/j.autcon.2021.103836 (2021).
9. Ding, Y., Cao, Y., Duffy, V. G., Wang, Y. & Zhang, X. Measurement and identification of mental workload during simulated computer tasks with multimodal methods and machine learning. *Ergonomics* **63**, 896–908. https://doi.org/10.1080/00140139.2020.1759699 (2020) (**PMID: 32330080**).
10. Gupta, S., Kumar, P. & Tekchandani, R. A machine learning-based decision support system for temporal human cognitive state estimation during online education using wearable physiological monitoring devices. *Decis. Anal. J.* **8**, 100280. https://doi.org/10.1016/j.dajour.2023.100280 (2023).
11. Ji, Z. et al. Cross-task cognitive workload recognition using a dynamic residual network with attention mechanism based on neurophysiological signals. *Comput. Methods Programs Biomed.* **230**, 107352. https://doi.org/10.1016/J.CMPB.2023.107352 (2023).
12. Yu, X. & Chen, C. H. A robust operators' cognitive workload recognition method based on denoising masked autoencoder. *Knowl.-Based Syst.* **301**, 112370. https://doi.org/10.1016/J.KNOSYS.2024.112370 (2024).
13. Chakladar, D. D., Datta, S., Roy, P. P. & Prasad, V. A. Cognitive workload estimation using variational autoencoder and attention-based deep model. *IEEE Trans. Cognit. Develop. Syst.* **15**, 581–590. https://doi.org/10.1109/TCDS.2022.3163020 (2023).
14. Thekkekara, J. P., Yongchareon, S. & Liesaputra, V. An attention-based cnn-bilstm model for depression detection on social media text. *Expert Syst. Appl.* https://doi.org/10.1016/j.eswa.2024.123834 (2024).
15. Jin, L., Qu, H., Pang, L., Zhang, Z. & Lyu, Z. Identifying stable eeg patterns over time for mental workload recognition using transfer ds-cnn framework. *Biomed. Signal Process. Control* **89**, 105662. https://doi.org/10.1016/J.BSPC.2023.105662 (2024).
16. Wang, Y. et al. Lgnet: Learning local-global eeg representations for cognitive workload classification in simulated flights. *Biomed. Signal Process. Control* https://doi.org/10.1016/j.bspc.2024.106046 (2024).
17. Yin, Z. et al. Physiological-signal-based mental workload estimation via transfer dynamical autoencoders in a deep learning framework. *Neurocomputing* **347**, 212–229. https://doi.org/10.1016/j.neucom.2019.02.061 (2019).

18. Tseng, C.-J. & Tang, C. An optimized xgboost technique for accurate brain tumor detection using feature selection and image segmentation. *Healthcare Analytics* **4**, 100217. https://doi.org/10.1016/j.health.2023.100217 (2023).
19. Wang, F., Tian, Y.-C., Zhang, X. & Hu, F. An ensemble of xgboost models for detecting disorders of consciousness in brain injuries through eeg connectivity. *Expert Syst. Appl.* **198**, 116778. https://doi.org/10.1016/j.eswa.2022.116778 (2022).
20. Yang, S. et al. Assessing cognitive mental workload via eeg signals and an ensemble deep learning classifier based on denoising autoencoders. *Comput. Biol. Med.* **109**, 159–170. https://doi.org/10.1016/j.compbiomed.2019.04.034 (2019).
21. Ying, Z. & Lanlan, C. Recognition of mental workload based on hybrid autoencoders. *Journal of East China University of Science and Technology* **49**, 98–107. https://doi.org/10.14135/j.cnki.1006-3080.20210922003 (2023).
22. Mahdavi, N. et al. Unraveling the interplay between mental workload, occupational fatigue, physiological responses and cognitive performance in office workers. *Sci. Rep.* **14**, 17866. https://doi.org/10.1038/s41598-024-68889-4 (2024).
23. Hamann, A. & Carstengerdes, N. Investigating mental workload-induced changes in cortical oxygenation and frontal theta activity during simulated flights. *Sci. Rep.* **12**, 6449. https://doi.org/10.1038/s41598-022-10044-y (2022).
24. Ke, Y. et al. Towards an effective cross-task mental workload recognition model using electroencephalography based on feature selection and support vector machine regression. *Int. J. Psychophysiol.* **98**, 157–166. https://doi.org/10.1016/J.IJPSYCHO.2015.10.004 (2015).
25. Hongquan, et al. Mental workload classification method based on eeg independent component features. *Applied Sciences***10** (2020).
26. Pang, L. et al. Subject-specific mental workload classification using eeg and stochastic configuration network (scn). *Biomed. Signal Process. Control* **68**, 102711. https://doi.org/10.1016/J.BSPC.2021.102711 (2021).
27. Chakladar, D. D., Dey, S., Roy, P. P. & Dogra, D. P. Eeg-based mental workload estimation using deep blstm-lstm network and evolutionary algorithm. *Biomed. Signal Process. Control* **60**, 101989. https://doi.org/10.1016/J.BSPC.2020.101989 (2020).
28. Sugiharti, E., Arifudin, R., Wiyanti, D. T. & Susilo, A. B. Integration of convolutional neural network and extreme gradient boosting for breast cancer detection. *Bulletin of Electrical Engineering and Informatics* **11**, 803–813. https://doi.org/10.11591/EEI.V11I2.3562 (2022).
29. Taori, T. J., Gupta, S. S., Gajre, S. S. & Manthalkar, R. R. Cognitive workload classification: Towards generalization through innovative pipeline interface using hmm. *Biomed. Signal Process. Control* **78**, 104010. https://doi.org/10.1016/J.BSPC.2022.104010 (2022).
30. Panwar, N., Pandey, V. & Roy, P. P. Eeg-cognet: A deep learning framework for cognitive state assessment using eeg brain connectivity. *Biomed. Signal Process. Control* https://doi.org/10.1016/j.bspc.2024.106770 (2024).
31. Safari, M. et al. Classification of mental workload using brain connectivity and machine learning on electroencephalogram data. *Sci. Rep.* **14**, 9153. https://doi.org/10.1038/s41598-024-59652-w (2024).
32. Nguyen, H. H., Iyortsuun, N. K., Kim, S., Yang, H. J. & Kim, S. H. Mental workload estimation with electroencephalogram signals by combining multi-space deep models. *Biomed. Signal Process. Control* **94**, 106284. https://doi.org/10.1016/J.BSPC.2024.106284 (2024).
33. Stew: Simultaneous task eeg workload dataset, https://doi.org/10.21227/44r8-ya50 (2018).
34. Hyvärinen, A. & Oja, E. Independent component analysis: Algorithms and applications. *Neural Netw.* **13**, 411–430. https://doi.org/10.1016/S0893-6080(00)00026-5 (2000).
35. Koelstra, S. et al. Deap: A database for emotion analysis; using physiological signals. *IEEE Trans. Affect. Comput.* **3**, 18–31. https://doi.org/10.1109/T-AFFC.2011.15 (2012).
36. Hinss, M. F. et al. Open multi-session and multi-task eeg cognitive dataset for passive brain-computer interface applications. *Scientific Data***10**, 1–22. DOI: https://doi.org/10.1038/S41597-022-01898-Y;SUBJMETA=2649,376,378,443,631;KWRD=COGNITIVE+NEUROSCIENCE, NEUROPHYSIOLOGY (2023).

Author contributions

G Abinaya conceptualized the study, developed the ACXNet model, and performed data pre-processing and analysis of the Simultaneous Task EEG Workload (STEW) dataset. Dr. K Dinakaran provided guidance on the methodology, assisted in model validation, and contributed to the interpretation of results. G Abinaya drafted the manuscript and Dr. K Dinakaran reviewed and provided critical revisions. Both authors have read and approved the final version of the manuscript.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to G.A.

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.