



OPEN Multivariate synchrosqueezing transform and time-frequency attention for mental workload classification from EEG signals

Zahed Nouri¹, Asghar Charmin¹, Hashem Kalbkhani²✉ & Saeed Barghandan¹

Mental workload (MWL) indicates cognitive effort during a task and is a key marker of mental state. Understanding MWL is vital in neuroergonomics, cognitive neuroscience, human-machine interaction, and intelligent systems. This paper introduces a new multi-stage approach for classifying mental workload using EEG signals. The method involves four key steps: preprocessing the signals to eliminate noise; applying the multivariate synchrosqueezing transform (MSST) for precise time-frequency analysis; extracting deep features with a new convolutional neural network (CNN) architecture that includes a time-frequency attention module (CNN-TFAN); and finally, reducing feature dimensionality for classification. The approach was tested on two public datasets, STEW and MAT. Results indicated that the optimized combination of semisupervised discriminant analysis (SDA) for feature reduction and support vector machine (SVM) for classification yielded the best results, with accuracy rates of 97.1% on STEW and 98.6% on MAT. Additional analysis revealed that MSST outperformed other time-frequency methods and that the deeper, attention-enhanced network architecture significantly improved classification accuracy. These findings demonstrate the effectiveness and robustness of the proposed framework for mental workload analysis.

Keywords Deep learning, EEG signals, Mental workload, Multivariate synchrosqueezing transform, Time-frequency attention

Motivations

Mental workload (MWL) indicates the cognitive effort needed to complete a task and is a vital factor in neuroergonomics, cognitive neuroscience, and human-machine interaction. Accurate estimation of MWL allows adaptive systems to enhance safety and performance in high-stakes environments like driving, air-traffic management, medical monitoring, and intelligent tutoring, settings where overload and underload can impair decision-making and raise risks^{1,2}. EEG is beneficial for MWL monitoring because it offers a direct, time-sensitive measurement of brain activity. Nonetheless, classifying MWL using EEG remains difficult due to issues such as strong nonstationarity, nonlinear dynamics, and inter-individual variability³. Moreover, MWL signatures often span multiple frequency bands and channels, necessitating methods that can analyze both spectro-temporal patterns and dependencies across channels.

A key technical challenge in MWL classification systems is the quality of the signal representation they rely on. Traditional spectral techniques, such as FFT/PSD, summarize frequency information globally but often overlook transient or rapidly changing workload effects. Even advanced time-frequency methods such as STFT and CWT are limited by a fundamental compromise between time and frequency resolution, often resulting in blurred signals from energy smearing. This hampers the ability to distinguish workload states clearly^{4,5}, leading downstream models to rely on less accurate or poorly localized features, which can affect their robustness and ability to generalize.

A second limitation relates to the common way multichannel EEG is modeled. Many existing methods rely on handcrafted features or analyze channels separately, thereby neglecting the inter-channel relationships that indicate functional coupling and distributed neural activity. Overlooking these spatial dependencies can lead to features that are less consistent across different subjects or experimental conditions⁶. Although deep learning, especially CNN-based methods, has advanced automatic feature extraction from EEG³, typical CNNs often treat EEG inputs as static 2D patterns and may not adequately highlight the most relevant time-frequency

¹Department of Electrical Engineering, Ahar Branch, Islamic Azad University, Ahar, Iran. ²Department of Electrical Engineering, Urmia University of Technology, Urmia, Iran. ✉email: h.kalbkhani@uut.ac.ir

regions unless specific mechanisms are incorporated^{7–10}. Therefore, there is a clear need for a framework that (i) offers a sharper, more informative time–frequency representation for nonstationary EEG, (ii) maintains the multichannel structure rather than analyzing channels separately, and (iii) adaptively concentrates the learning on the most discriminative spectro-temporal features.

Motivated by these gaps, this study introduces a multi-stage MWL classification framework. It combines the multivariate synchrosqueezing transform (MSST)¹¹ which provides detailed multichannel time–frequency analysis, using a CNN model augmented with a time–frequency attention module (CNN-TFAN) to improve feature learning. Additionally, it uses semi-supervised discriminant analysis (SDA) and the support vector machine (SVM) classifier to support reliable decision-making. This approach directly addresses the issues of smeared time–frequency representations and unstructured multichannel modeling, aiming to enhance resolution, interpretability, and stability in classification on standard MWL benchmarks.

Related works

With this motivation, the following section reviews (i) EEG-based MWL assessment studies, (ii) time–frequency and decomposition-based representations for cognitive-state classification, and (iii) deep and attention-based architectures designed to improve robustness and generalization in multichannel EEG learning.

In recent years, EEG signals have become increasingly crucial for measuring cognitive load in human-computer interaction, intelligent education, and mental performance tracking. Early research indicates that variations in the brain's electrical activity can reliably reflect a person's cognitive load in both real-world and laboratory environments. For example, real-time EEG monitoring of remote maintenance operators showed that changes in power in specific frequency bands can effectively indicate mental workload¹². Studies with *n*-back tasks have also demonstrated a clear link between cognitive difficulty and EEG responses¹³. Additionally, examining the role of cognitive load in creating intelligent tutoring systems underscores the need for accurate, real-time assessment of this metric within advanced learning frameworks^{14,15}.

To enhance the accuracy of cognitive load measurement, special focus has been given to processing and extracting relevant features from EEG data. Using adaptive and fixed wavelet transforms alongside multi-domain optimization has notably improved the distinction between different cognitive load levels^{16,17}. The redundant adaptive discrete wavelet transform (RADWT) has proven effective in increasing frequency resolution and detection precision¹⁸. Other techniques include variational mode decomposition (VMD)-based spectral analysis combined with feature selection via the LightGBM algorithm¹⁹, singular spectrum analysis (SSA) and circulant SSA paired with metaheuristic algorithms²⁰, and traditional wavelet methods for assessing cognitive load²¹. Moreover, comprehensive analytical reviews have compared various EEG preprocessing techniques for cognitive load detection²², highlighting the importance of initial processing quality in achieving better classification results. Signal decomposition methods were also widely used for EEG classification, especially in motor imagery (MI) BCIs^{23,24}, where multivariate VMD/EWT frameworks have reported strong accuracy by separating task-relevant oscillatory components. However, mental workload EEG often exhibits subtler, more nonstationary patterns distributed across bands, making direct comparisons with MI-focused results nontrivial. Methodologically, decomposition approaches explicitly split signals into a limited set of modes, requiring mode selection.

Alongside progress in feature extraction, machine learning, and deep learning have become key approaches for modeling cognitive load. Early on, statistical models like the hidden Markov model (HMM) were used to develop generalizable pipelines²⁵. Later, more sophisticated models such as bidirectional long short-term memory (BLSTM)-long short-term memory (LSTM) networks, combined with evolutionary algorithms²⁶, hybrid CNN-LSTM models²⁷, and brain connectivity-based models¹ were adopted. Multi-task deep networks, for example, EEGMeNet, showed strong results in joint learning and classification stability²⁸. Additionally, unsupervised clustering methods²⁹ and transfer learning approaches, especially for cross-session and cross-subject applications, have been introduced to better simulate real-world conditions^{2,30}. Some studies have integrated EEG with other data, such as eye-tracking, which has led to enhanced classification accuracy³¹.

Research in this area often uses time-frequency representations and attention mechanisms to isolate critical signal information. For instance, the attention-based recurrent fuzzy network (ARFN) model achieved high accuracy in classifying cognitive load by utilizing fuzzy recurrent attention³². Transformer-based models such as MST-Net have demonstrated strong performance by leveraging multi-scale time-frequency features³³. Applying BLSTM on these features has also improved the detection of load levels on an individual basis³⁴. Furthermore, IoT-driven hybrid models³⁵ and semi-supervised methods focusing on time-frequency analysis have been developed for real-time cognitive load assessment. Channel-wise feature optimization and spatial pattern techniques have enhanced model generalizability^{36,37}. Additionally, functional brain connectivity analysis has facilitated multi-class classification of cognitive load³⁸. In addition to these efforts, new methods have explored the theoretical aspects of cognitive load and how to incorporate them into educational system design^{15,39}. The integration of AI-based tools with technologies such as functional near-infrared spectroscopy (fNIRS) has expanded opportunities to monitor and analyze cognitive load⁴⁰. Furthermore, there is a focus on improving the reproducibility of results and developing clear evaluation standards within the EEG and cognitive load fields, which continue to pose significant challenges².

Recent studies have shown that advanced information-theoretic features and attention-based deep learning models can significantly improve EEG analysis of cognitive and emotional states⁴¹. In particular, mutual information-based features have been effective at distinguishing complex mental conditions. For example, a framework combining normalized mutual information with a self-optimized Gaussian kernel radial-basis-function extreme learning machine was developed to decode different states, yielding strong results through adaptive hyperparameter tuning and meaningful feature extraction. These findings highlight the importance of capturing nonlinear relationships in EEG signals to improve classification accuracy. Beyond feature-driven methods, graph-based deep learning models are increasingly popular for explicitly modeling connections

between EEG channels. Graph attention convolutional neural networks were used in⁴² to detect driver fatigue, where mutual information was used to construct connectivity-aware graphs that guide attention mechanisms in an end-to-end system. These approaches emphasize the importance of spatial dependencies and adaptive attention in representing distributed brain activity across channels. Additionally, reviews of EEG-based emotion and cognitive state detection highlight key challenges, including non-stationarity, feature robustness, and attention modeling⁴³. Overall, there is a clear trend toward integrating advanced time–frequency analysis, attention mechanisms, and deep learning architectures to make EEG systems more accurate, generalizable, and easier to interpret.

Recent progress in EEG-based brain–computer interface (BCI) research has emphasized that achieving reliable subject-independent generalization remains a key challenge. The large-cohort studies, such as those by Sadiq et al.⁴⁴ and Yu et al.⁴⁵ reveal that effective BCI systems must handle significant inter-subject variability, nonstationarity, and differences in neural patterns among users. These studies indicate that scalable and adaptable BCI frameworks rely on thoughtfully designed signal representations and learning architectures, especially when tested on datasets involving many subjects using subject-independent protocols. While these studies offer valuable insights, it's essential to recognize that classifying mental workload differs from motor and mental imagery paradigms, both in experimental setup and dataset availability. Unlike motor imagery research—where extensive datasets with 60 or more participants are available—public mental workload datasets with standardized protocols are scarce. Consequently, datasets like STEW and MAT have become standard benchmarks in this area and are frequently cited in recent research to facilitate fair comparisons and reproducibility.

Contributions

This paper outlines a framework for classifying mental workload using EEG signals. It combines multivariate time–frequency analysis with attention-based deep learning, offering several advancements over earlier methods. The key contributions of this work are as follows:

- **Development of a multi-stage framework for MWL classification:** A comprehensive method was devised, consisting of four key phases: signal preprocessing, time–frequency analysis via the MSST, deep feature extraction with a CNN integrated with a TFAN module, and ultimately, optimized joint dimensionality reduction and classification. This approach improves the model's accuracy and robustness across diverse data scenarios.
- **Application and extension of the MSST for EEG signals of WML:** This study introduces a multivariate SST, unlike traditional methods like STFT, CWT, or single-channel SST, to simultaneously model spatial dependencies across EEG channels. This approach improves the clarity of the time–frequency representation and provides a deeper insight into inter-regional brain dynamics under various mental workload conditions.
- **Design of a time–frequency attention network (TFAN) module:** A dual-branch attention module (temporal–spectral dual attention) is introduced, allowing the network to adaptively concentrate on the most important regions in time and frequency domains. This module comprises temporal and spectral attention branches, each detecting key patterns with distinct convolutional filters. Incorporating this module into the CNN architecture has significantly enhanced classification accuracy compared to models lacking attention mechanisms.
- **Optimization of the dimensionality reduction and classification:** This study employs semi-supervised discriminant analysis (SDA) for reducing feature dimensions. Bayesian optimization was used to fine-tune the method's parameters to maximize class separability. Ultimately, a support vector machine (SVM) served as the main classifier, showing the highest accuracy among the compared traditional algorithms.
- **Comprehensive evaluation on two Public datasets (STEW and MAT):** The proposed model was tested on two validated EEG datasets. It achieved 97.1% accuracy on the STEW dataset and 98.6% on the MAT dataset. These results significantly outperform traditional time–frequency methods and deep learning models lacking attention mechanisms.

The rest of this paper is organized as follows. Section II describes the publicly used datasets in this research. The proposed method is explained in detail in Section III. Section IV contains the results obtained during performance analysis, and Section V concludes this paper.

Dataset

In this paper, two publicly available multivariate EEG datasets collected for cognitive load analysis are utilized. The MAT and STEW datasets are explained below.

STEW dataset

The open-access simultaneous task EEG workload (STEW) dataset is a valuable resource for studying multitasking workload and analyzing brain activity during different cognitive tasks⁴⁶. Researchers can utilize this dataset to develop and evaluate algorithms and models for classifying and predicting mental workload. The multitasking workload experiment utilized the SIMKAP multitasking test, in which participants identified and marked matching items on two panels while simultaneously answering auditory questions of various types, such as arithmetic, comparison, or data retrieval. The experiment comprised two phases: in the first, participants remained inactive for 2.5 min, representing a “low” mental workload; in the second, they completed the SIMKAP test for 2.5 min and were monitored for brain activity, representing a “high” mental workload⁴⁶.

This dataset includes multivariate EEG signals from 48 male subjects. The signals were recorded at a sampling rate of 128 Hz with 16-bit A/D resolution using the Emotiv EPOC EEG headset. It features 14 channels: AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, and AF4, aligned with the 10–20 international system. Figure 1 presents some recordings from the STEW dataset⁴⁶.

MAT dataset

The National Technical University of Ukraine supplied the Mental Arithmetic Tasks (MAT) dataset, which focuses on arithmetic tasks involving the consecutive subtraction of two numbers. Researchers used this dataset to study brain activity across various cognitive neuroscience functions. This dataset features EEG recordings from 36 students aged 18 to 26. The EEG data were collected using 23 electrodes positioned across the scalp according to the 10–20 system¹⁷. Each recording contains artifact-free EEG segments. Resting state segments last 3 min, whereas mental counting segments are 1 min long^{47,48}.

Proposed method

Here, we outline the proposed method for classifying mental workload. As shown in Fig. 2, the proposed method classifies the multivariate EEG signals in four steps: (1) preprocessing, (2) time-frequency analysis, (3) extracting deep features, and (4) joint feature reduction and classification. In the following, each step will be explained in detail.

Preprocessing

In the preprocessing step, several filters are employed to reduce artifacts and produce clean signals. The high-pass filter, with a cutoff frequency of 1 Hz, removed line noise from the signals in the STEW dataset. Also, data averaging was implemented to reduce artifacts in the subspace reconstruction and facilitate re-referencing⁴⁶. For the MAT dataset, high-pass, low-pass, and power-line notch filters with cutoff frequencies of 0.5 Hz, 45 Hz, and 50 Hz were applied to minimize artifacts¹⁹.

Beyond traditional filtering techniques, hybrid EEG denoising approaches, such as multiscale PCA (MSPCA)^{7,23,24,49}, which integrate wavelet multiresolution analysis with PCA-based component selection, are widely employed to reduce multiscale artifacts while retaining task-related activity. MSPCA preprocessing has been shown to enhance robustness across various EEG classification systems, especially in motor imagery and clinical EEG contexts. In this study, we adhere to the standard preprocessing protocols specified for the STEW and MAT benchmarks to ensure consistency and facilitate a fair comparison with previous research.

Time-frequency representation

Linear projection-based time-frequency algorithms, such as the STFT and CWT, face challenges when analyzing nonlinear and nonstationary signals. Reassignment techniques can address this issue by enhancing

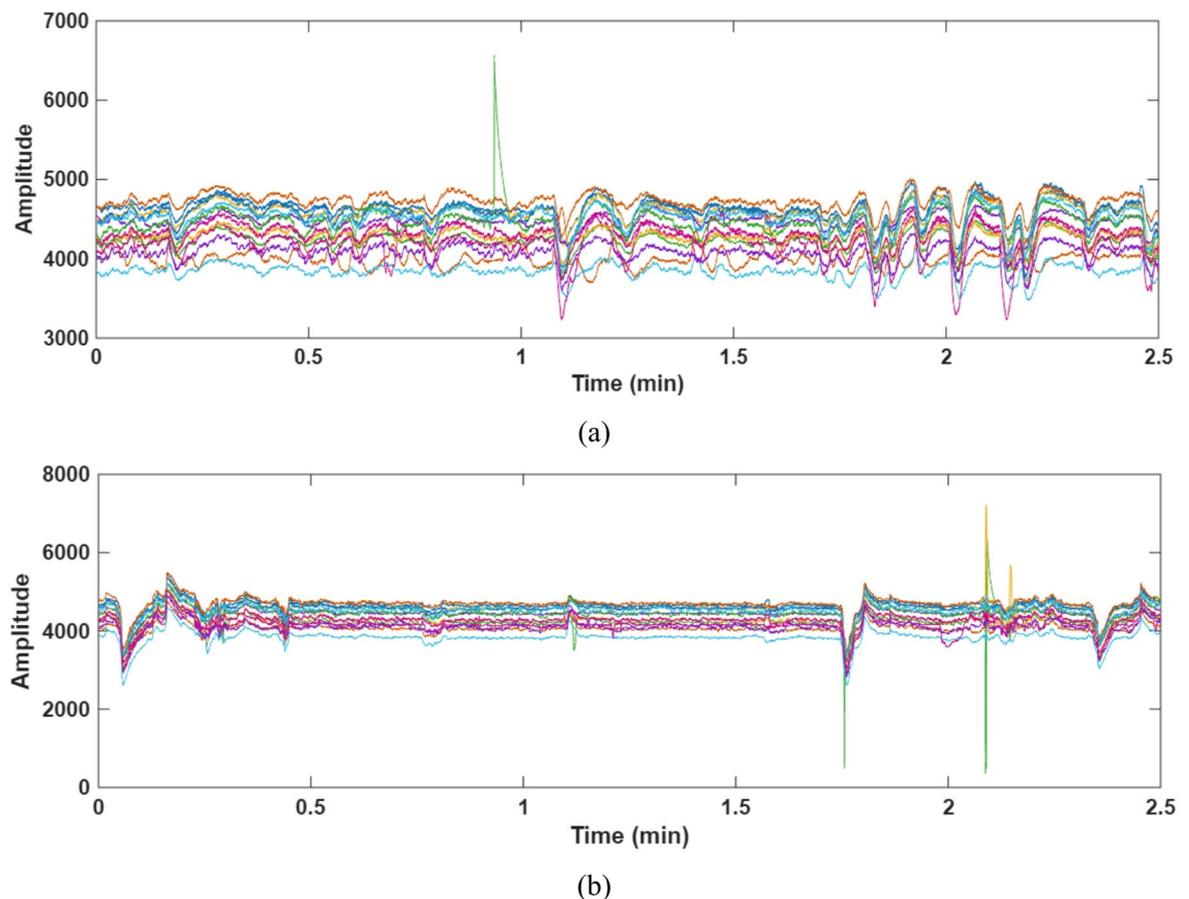


Fig. 1. Recorded EEG signals from the STEW dataset in (a) low and (b) high mental work load scenarios.

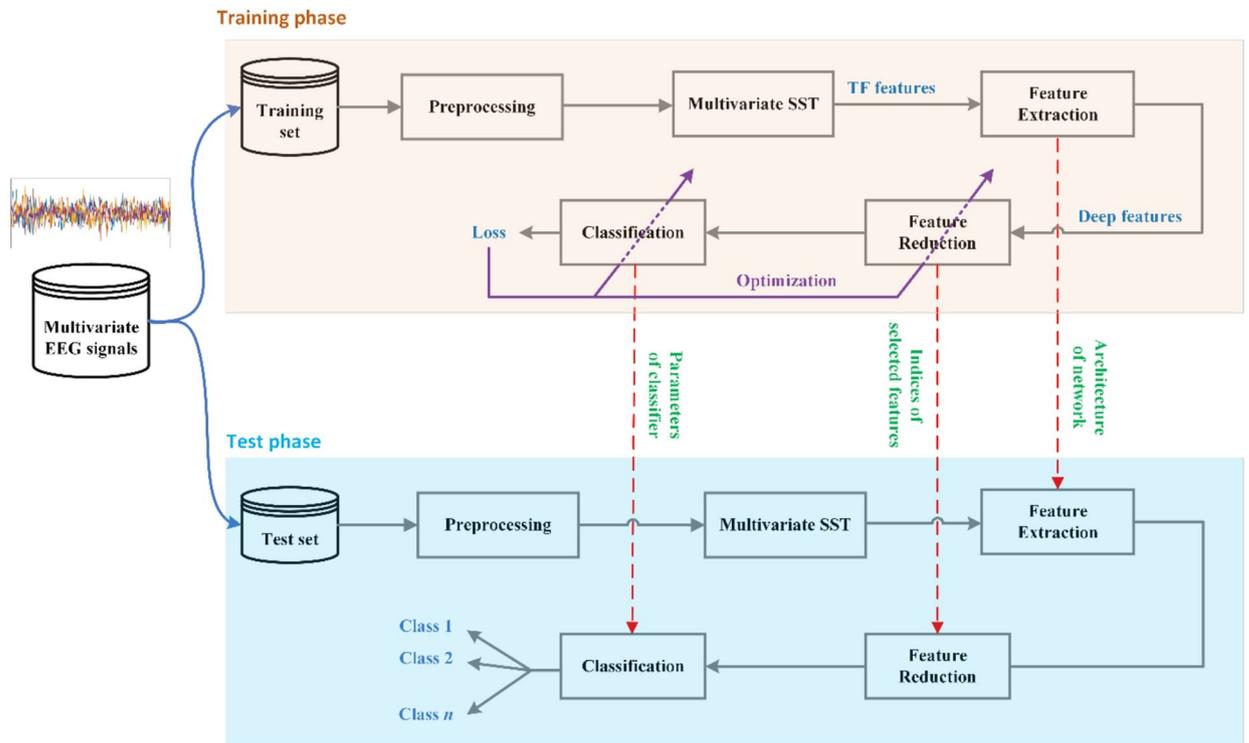


Fig. 2. The general steps of the proposed method for mental workload classification.

the localization of the time-frequency representation (TFR). One transform derived from such TFRs, suitable for creating highly localized TFRs for nonlinear and nonstationary signals, is the synchrosqueezing transform (SST)⁵⁰, which is based on the continuous wavelet transform. SST refines the representation by concentrating the energy of coefficients along the instantaneous frequency curves of the modulated oscillations. As a post-processing technique using the CWT, SST enhances time-frequency analysis of signals by focusing energy along instantaneous frequency trajectories. SST effectively reallocates the coefficients of the CWT, providing a more precise and more localized time-frequency representation. This approach overcomes the limitations of traditional Fourier analysis.

The multivariate SST (MSST) is designed to identify common oscillatory patterns across multiple data channels, which is crucial for signal analysis that involves understanding variable interactions. It calculates joint instantaneous frequencies for multivariate data, offering insights into the relationships among different signals. MSST is a powerful tool for analyzing complex multivariate signals, improving comprehension of their temporal behavior and dependencies. Its focus on relevant oscillatory features across channels makes it essential for many scientific and engineering fields. The MSST produces a concise time-frequency representation of multichannel signals. It enhances traditional SST by effectively analyzing multivariate signals with evolving oscillatory patterns. This makes it particularly useful in biomedical engineering, finance, or any domain dealing with multivariate time series data. For example, in biomedical signals like EEG or ECG, multiple channels represent various physiological measurements. MSST helps to better understand their temporal dynamics and interrelations. Its ability to highlight significant oscillatory features across channels makes it especially valuable for EEG analysis¹¹.

The continuous wavelet transform for the nonlinear and nonstationary signal $x(t)$, $X(a, b)$, is defined as¹¹:

$$X(a, b) = \int a^{-0.5} x(t) \psi\left(\frac{t-b}{a}\right) dt \tag{1}$$

where $\psi(t)$ is the mother wavelet. The scale factor a shifts the mother wavelet in the frequency domain and changes its bandwidth. For the set of wavelet coefficients $X(a, b)$, the SST with the frequency resolution $\Delta\omega$, $S(\omega_l, b)$, is defined as¹¹:

$$S(\omega_l, b) = \sum_{a_k: |\omega_x(a_k, b) - \omega_l| \leq \Delta\omega/2} X(a_k, b) a^{-1.5} \Delta a_k \tag{2}$$

where the set of frequency bins is denoted by a_k .

For a multivariate signal $x_N(t)$ with N channels and the corresponding SST coefficients for each channel $S_n(\omega_l, b)$, $n = 1, \dots, N$, (the SST coefficients $S_n(\omega_l, b)$ have been normalised with the constant R_ψ), and a given set of oscillatory scales, $\{\omega_k\}$, $k = 1, \dots, K$, obtained using a multivariate extension of a method proposed in¹¹, the instantaneous frequency $\Omega_k^n(b)$ for each frequency band, k , is given by¹¹:

$$\Omega_k^n(b) = \frac{\sum_{\omega \in \omega_k} |S_n(\omega, b)|^2 \omega}{\sum_{\omega \in \omega_k} |S_n(\omega, b)|^2} \quad (3)$$

Also, the instantaneous amplitude $A_k^n(b)$ for each frequency band is calculated as¹¹:

$$A_k^n(b) = \sqrt{\sum_{\omega \in \omega_k} |S_n(\omega, b)|^2} \quad (4)$$

To estimate the multivariate instantaneous frequency for a given frequency band k , the instantaneous frequencies across the N channels are combined using the joint instantaneous frequency. As a result, the multivariate instantaneous frequency band $\Omega_k^{multi}(b)$ is given by¹¹:

$$\Omega_k^{multi}(b) = \frac{\sum_{n=1}^N (A_k^n(b))^2 \Omega_k^n(b)}{\sum_{n=1}^N (A_k^n(b))^2} \quad (5)$$

Also, the instantaneous amplitude $A_k^{multi}(b)$ for each frequency band is obtained as¹¹:

$$A_k^{multi}(b) = \sqrt{\sum_{n=1}^N (A_k^n(b))^2} \quad (6)$$

After determining the joint instantaneous amplitude and frequency for each frequency band, the multivariate TFR, $\mathbf{T}_k^{multi}(\omega, b)$, for each oscillatory scale k , $k = 1, \dots, K$, is calculated as¹¹:

$$\mathbf{T}_k^{multi}(\omega, b) = A_k^{multi}(b) \delta(\omega - \Omega_k^{multi}(b)) \quad (7)$$

where $\delta(\cdot)$ is the Dirac delta function. For a multivariate signal $x_N(t)$ with N channels, the MSST can be summarised as follows:

- Apply the SST channel-wise to obtain the coefficients $S_n(\omega_l, b)$.
- A set of partitions along the frequency axis for the time-frequency domain is determined, and the instantaneous frequency $\Omega_k^n(b)$ and amplitude $A_k^n(b)$ is calculated for each frequency bin k .
- The multivariate instantaneous frequency $\Omega_k^{multi}(b)$ and amplitude $A_k^{multi}(b)$ is calculated.
- The multivariate synchrosqueezed coefficients, $\mathbf{T}_k^{multi}(\omega, b)$, is calculated.

Deep feature extraction

Time-frequency attention module

The structure of the utilized time-frequency attention (TFATN) module is shown in Fig. 3. To enable the model to dynamically focus on the most salient regions of the time-frequency representation, we employ a dual-path attention mechanism. A time-attention branch and a frequency-attention branch concurrently process the input features. The time attention branch utilizes convolutional kernels elongated along the frequency axis, 3×3 and 3×5 , and average pooling across the time dimension, *PoolingT*, to generate a vector of weights, identifying the importance of each frequency band. Symmetrically, the frequency attention branch uses kernels elongated along the time axis, 3×3 and 5×3 , and average pooling across the frequency dimension, *PoolingF*, to determine the significance of each time step. These two attention vectors are then multiplied to form a comprehensive 2D time-frequency attention map. This map is applied to the original input via element-wise multiplication, effectively re-weighting the features to amplify key information and suppress noise. Finally, the attended feature map is concatenated with the original input and passed through a final convolution, producing a refined output that has learned to focus on the most critical spectro-temporal information.

The goal of the time-attention branch is to determine which frequency bands are most important, regardless of the specific time point. The input first passes through two parallel convolutional layers. The 3×5 kernel is wider than it is tall. This shape is effective at capturing features across different frequencies at a specific moment in time. The 3×3 kernel captures more general local features. Each is followed by batch normalization (BN) and a rectified linear unit (ReLU) activation function (BN-ReLU). The outputs from the two convolutional blocks are concatenated along the channel dimension. Then, a 1×1 convolution is used to efficiently reduce the number of channels and create a compact representation of the combined features. Pooling is the key step and is applied across the time dimension. This collapses the time dimension, yielding a single vector in which each element corresponds to a frequency band and represents its overall importance. The vector is passed through another 1×1 Conv and then a Sigmoid function—the sigmoid squashes the values between 0 and 1, producing the final attention weights. A value near 1 indicates the corresponding frequency band is critical, while a value near 0 indicates it's less important. The mentioned procedure can be described mathematically as follows:

$$X_{time} = \text{Concat}(\text{Conv}_{3 \times 3}(X), \text{Conv}_{3 \times 5}(X)) \quad (8)$$

$$X_{time,pool} = \text{Pool}_{(1,none)}(\text{BNReLU}(\text{Conv}_{1 \times 1}(X_{time}))) \quad (9)$$

$$A_{time,pool} = \sigma(\text{Conv}_{1 \times 1}(X_{time,pool})) \quad (10)$$

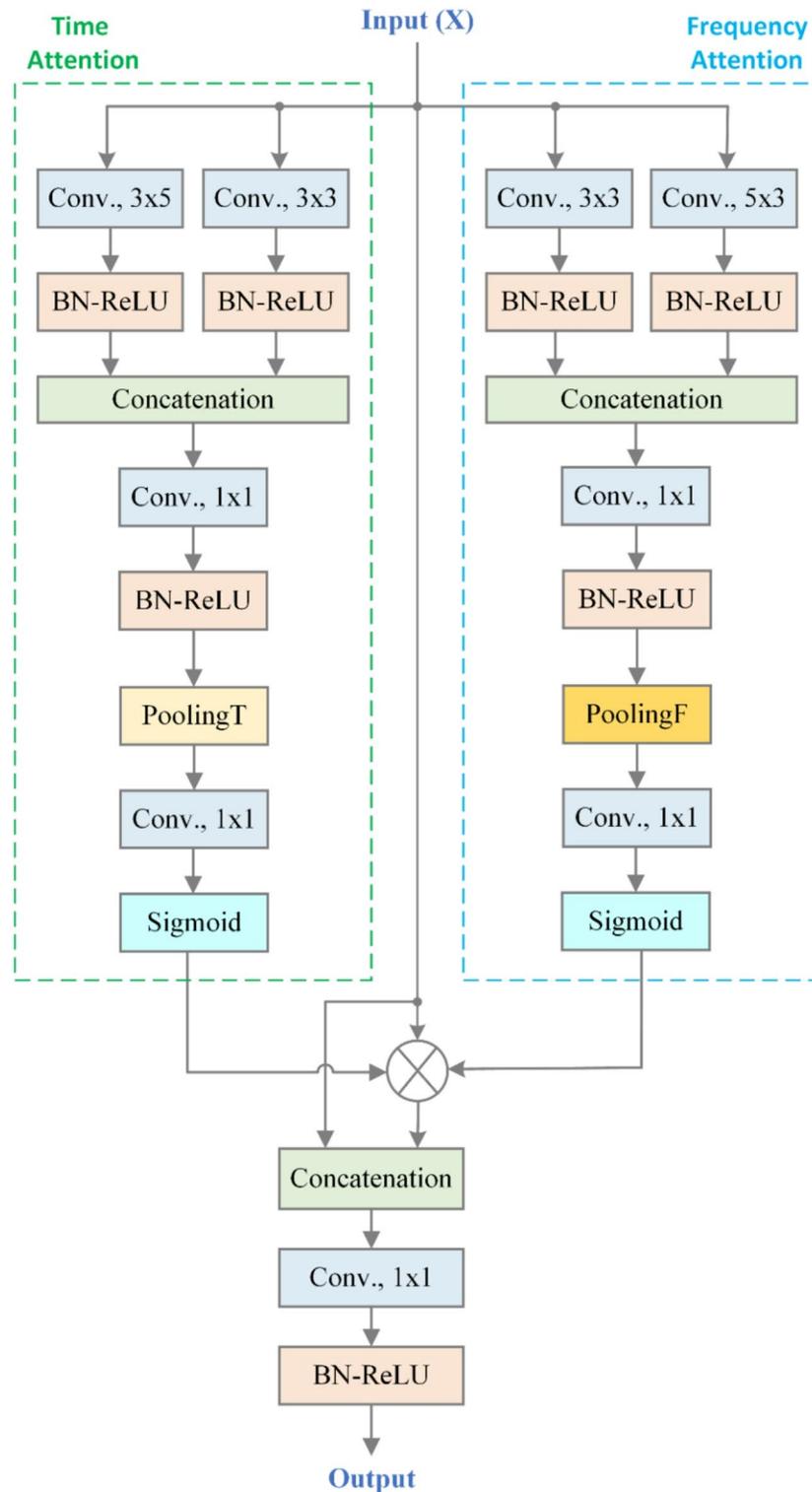


Fig. 3. The structure of the utilized time-frequency attention (TFATN).

Symmetrically, the goal of the frequency-attention branch is to figure out which time steps are most important, regardless of the specific frequency. The input passes through two parallel convolutional layers (Conv., 3×3 & Conv., 5×3). Here, the 5×3 kernel is taller than it is wide. This shape is effective at capturing features and patterns across different time steps within a specific frequency band. Same as the time branch, the outputs are concatenated and then reduced using a 1×1 convolution. Then, the pooling is applied across the frequency dimension. This collapses the frequency dimension, resulting in a single vector where each element corresponds to a time step and represents its overall importance across all frequencies. A Sigmoid function again creates

the final attention weights (between 0 and 1) for each time step. The similar equations as (8)-(10) are valid for frequency-attention branch and $A_{frequency, pool}$ is the output of this branch.

To apply the attention, the time attention vector, indicating which frequencies matter, and the frequency attention vector, indicating which times matter, are multiplied together. This creates a 2D time-frequency attention map. This map assigns a weight to each point in the original input, indicating its combined importance. This 2D attention map is then multiplied element-wise with the original time-frequency. This step re-weights the input; important time-frequency points are amplified, while unimportant ones are suppressed. The re-weighted (attended) input is concatenated with the original input. This is a residual-style connection, ensuring that the model doesn't lose original information while learning the attention. A final 1×1 convolution processes the concatenated data to produce the final, refined Output. This output is a feature map that has been enhanced to focus on the most relevant parts of the original signal.

$$X_{TFAN} = Conv_{1 \times 1} (Concat (X \odot X_{time, pool} \odot X_{freq, pool}, X)) \tag{11}$$

CNN-TFAN architecture

The architecture of the proposed CNN with time-frequency attention (CNN-TFAN) for deep feature extraction is illustrated in Fig. 4. This architecture consists of three sequential processing blocks that hierarchically refine and extract features. Each block in this architecture comprises three main components. A 2D convolutional layer identifies local patterns in the input feature maps using 3×3 kernels. A Time-Frequency Attention Network (TFATN) Module positioned immediately after the convolutional layer adaptively focuses on salient and essential information in both the time and frequency dimensions by reweighting different regions of the feature map. A Max-Pooling Layer using a 2×2 window reduces the spatial dimensions of the feature maps. This process decreases computational complexity and helps the model achieve spatial invariance, making it less sensitive to minor shifts in patterns. The input time-frequency map is passed sequentially through these three blocks. The number of filters in the convolutional layers progressively decreases throughout the network: starting with 32 in the first block, then 16 in the second, and 8 in the final block. This design creates an information bottleneck, compelling the network to learn a compact and efficient representation of the data. The output of the final block is a set of deep features, which constitutes a high-level, compressed representation of the key information present in the input signal. This feature vector obtained from the flatten layer is subsequently utilized for the final classification task.

Optimized feature reduction and classification

Feature reduction

Linear discriminant analysis (LDA) is the supervised and linear version of SDA and considers only labeled samples. Its objective function is as follows⁵¹:

$$a_{opt} = \operatorname{argmax}_a \frac{a^T S_b a}{a^T S_w a} \tag{12}$$

where S_w and S_b denote the intra- and inter-class scatter matrices, respectively, and computed as⁵¹:

$$S_b = \sum_{k=1}^{n_c} n^{(k)} (\mu^{(k)} - \mu) (\mu^{(k)} - \mu)^T \tag{13}$$

$$S_w = \sum_{k=1}^{n_c} \sum_{i=1}^{n^{(k)}} (x_i^{(k)} - \mu^{(k)}) (x_i^{(k)} - \mu^{(k)})^T \tag{14}$$

where $n^{(k)}$ denotes the number of training samples for the class C_k , μ and $\mu^{(k)}$ are the total sample mean vector and the mean vector of class C_k , respectively. Also, $x_i^{(k)}$ is the sample i in class C_k . The total scatter matrix S_t , is defined as $S_t = \sum_{i=1}^N (x_i - \mu) (x_i - \mu)^T$, hence we have $S_t = S_b + S_w$. Thus, the objective function equals⁵¹:

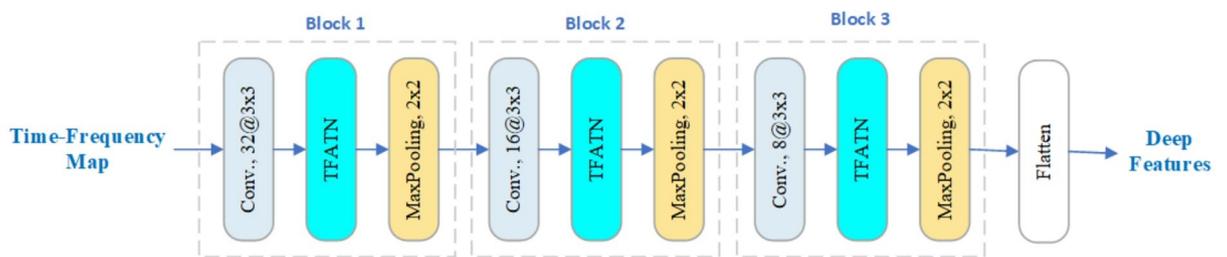


Fig. 4. The architecture of the proposed CNN-TFAN for deep feature extraction.

$$\mathbf{a}_{opt} = \operatorname{argmax}_a \frac{\mathbf{a}^T \mathbf{S}_b \mathbf{a}}{\mathbf{a}^T \mathbf{S}_t \mathbf{a}} \quad (15)$$

When training samples are insufficient, overfitting can happen. Regularizers are commonly employed to prevent this. The optimization problem in such cases is defined as follows⁵¹:

$$\max \frac{\mathbf{a}^T \mathbf{S}_b \mathbf{a}}{\mathbf{a}^T \mathbf{S}_t \mathbf{a} + J(\mathbf{a})} \quad (16)$$

The regulation coefficient controls the balance between the model's complexity and the empirical loss. Also, the $J(\mathbf{a})$ denotes the learning complexity of the hypothesis family, and considering a natural regularizer, we have⁵¹:

$$J(\mathbf{a}) = \sum_{ij} (\mathbf{a}^T x_i - \mathbf{a}^T x_j)^2 S_{ij} = 2 \sum_i \mathbf{a}^T x_i D_{ii} x_i^T \mathbf{a} - 2 \sum_{ij} \mathbf{a}^T x_i S_{ij} x_j^T \mathbf{a} = 2 \mathbf{a}^T \mathbf{X} (\mathbf{D} - \mathbf{S}) \mathbf{X}^T \mathbf{a} = 2 \mathbf{a}^T \mathbf{X} \mathbf{L} \mathbf{X}^T \mathbf{a} \quad (17)$$

Considering the $N_p(x_i)$ as the set of p nearest neighbors of x_i , the weight matrix, \mathbf{S} , is defined as⁵¹:

$$S_{ij} = \begin{cases} 1, & \text{if } x_i \in N_p(x_j) \text{ or } x_j \in N_p(x_i) \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

The diagonal matrix \mathbf{D} is defined as $D_{ii} = \sum_j S_{ij}$. Also, the Laplacian matrix is defined as $\mathbf{L} = \mathbf{D} - \mathbf{S}$. Hence, the objective function of SDA can be formulated as⁵¹:

$$\max_a \frac{\mathbf{a}^T \mathbf{S}_b \mathbf{a}}{\mathbf{a}^T (\mathbf{S}_t + \mathbf{X} \mathbf{L} \mathbf{X}^T) \mathbf{a}} \quad (19)$$

The objective function reaches its maximum when using the projective vector \mathbf{a} , which is determined by the largest eigenvalue solution to the generalized eigenvalue problem⁵¹.

$$\mathbf{S}_b \mathbf{a} = \lambda (\mathbf{S}_t + \alpha \mathbf{X} \mathbf{L} \mathbf{X}^T) \mathbf{a} \quad (20)$$

Considering $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{nz}]$, where nz is the number of non-zero eigenvalues, and samples are embedded as⁵¹:

$$\mathbf{x} \rightarrow \mathbf{z} = \mathbf{A}^T \mathbf{x} \quad (21)$$

The performance of SDA depends on the regulation parameter. This paper uses Bayesian optimization to identify the optimal value of this parameter, yielding the highest classification accuracy.

Classification

This paper assessed and reported the accuracy of various classifiers. The most commonly used classifiers, including SVM⁵², kNN⁵³, decision tree⁵⁴, and random forest⁵⁵, are evaluated individually. Their performance varies with specific hyperparameters, so an optimization process was used to determine the optimal values. Table 1 outlines these hyperparameters and the optimizer used for their fine-tuning. It should be noted that for the decision tree, the split criterion is Gini's index, and height-balancing is achieved when the heights of the subtrees differ by no more than one.

Bayesian optimization

Bayesian optimization is a powerful approach for optimizing functions that are expensive to evaluate, particularly in machine learning and hyperparameter tuning. Unlike conventional techniques relying on exhaustive search or gradients, it builds a probabilistic model to guide the search process more efficiently. This method is beneficial when the objective function is nonconvex, costly to compute, or lacks analytical gradients. The Bayesian optimization process begins with evaluating an initial set of randomly chosen points on the objective function. A Gaussian Process is then fitted to model this function. An acquisition function guides the selection of the following sampling point, after which the objective function is evaluated. The model is updated with the new data, and the cycle repeats until the process converges or the evaluation budget is exhausted. This iterative approach allows Bayesian optimization to efficiently identify the best solution with fewer evaluations than brute-force methods. Its main advantage is its sample efficiency, making it suitable for expensive functions. It can also

Classifier	Hyperparameters
kNN	Number of neighbors, distance metric (Euclidean, Mahalanobis, cubic, cosine), weighting scheme (equal, inverse, squared inverse)
SVM	Kernel type (linear, quadratic, cubic, Gaussian), box constraint, kernel scale (only for Gaussian)
Decision tree	Maximum number of splits
Random forest	Minimum number of leaf sizes and number of predictors to sample

Table 1. The hyperparameters of classifiers.

Parameter	Value
Optimizer	Stochastic gradient descent with momentum (SGD)
Loss function	Cross-entropy
Number of epochs	150
Batch size	32
Learning rate	10^{-4}
Momentum	0.85
Dropout ratio	0.30

Table 2. Parameters used for tuning the CNNs.

Dataset	Classifier							
	kNN		SVM		Decision tree		Random forest	
	wo	woo	wo	woo	wo	woo	wo	woo
STEW	94.8	90.7	97.1	93.9	94.1	90.8	95.2	92.1
MAT	96.4	92.6	98.6	94.9	95.5	91.9	96.9	93.5

Table 3. Classification accuracy of different classifiers. The terms “wo” and “woo” respectively denote “with optimization” and “without optimization”.

handle noisy and black-box functions, even when gradient information is unavailable or unreliable. By balancing exploration and exploitation through modeling uncertainty, it effectively optimizes complex search spaces⁵⁶.

Results

Here, we present results demonstrating the effectiveness of the proposed schemes for classifying mental workload from EEG signals.

Simulation setup and performance metrics

The presented method is implemented and tested on a system configuration of an Intel Core i7 CPU and 32 GB of RAM. For both datasets, the recordings of 80% of subjects are used for training, and the remaining 20% are used to test the trained model. The parameters used in the tuning process for the CNN-TFAN mentioned are given in Table 2.

The sensitivity, precision, specificity, and accuracy are used to evaluate the performance of the proposed method. These metrics are defined as follows¹⁹.

$$Sens. = \frac{TP}{TP + FN} \quad (22)$$

$$Prec. = \frac{TP}{TP + FP} \quad (23)$$

$$Spec. = \frac{TN}{TN + FP} \quad (24)$$

$$Acc. = \frac{TN + TP}{TN + TP + FN + FP} \quad (25)$$

where TP , TN , FP , and FN respectively denote the true positive, true negative, false positive, and false negative. The ‘no task’ or ‘rest’ class is the negative class, and the positive class denotes the ‘task’ class.

Performance analysis

Table 3 compares the performance of four classifiers, including k NN, SVM, decision tree, and random forest, on the STEW and MAT datasets. The results show that SVM outperformed the other models significantly on both datasets, achieving 97.1% accuracy on STEW and 98.6% on MAT. The overall ranking of the classifiers was consistent across both datasets, with random forest and k NN following SVM. The decision tree consistently performed the weakest, with accuracies of 94.1% and 95.5% for STEW and MAT, respectively. These results identify SVM as the most effective and suitable classifier for the proposed framework in mental workload classification.

SVM’s exceptional performance mainly stems from two key attributes. It aims to find a hyperplane that not only separates clusters but also maximizes the margin, the distance between the nearest data points of each cluster (support vectors). This “margin maximization” technique enhances generalization and reduces overfitting. When EEG signals are processed through the SST transform and a deep network, their features often exhibit complex nonlinear relationships. The kernel trick allows SVM to map these features into a higher-dimensional

		Predicted Class		<i>Sens.</i>	<i>Prec.</i>	<i>Spec.</i>	<i>Acc.</i>
		Negative	Positive				
Actual Class	Negative	95.8% (TN)	4.2% (FP)	98.4%	95.9%	95.8%	97.1%
	Positive	1.6% (FN)	98.4% (TP)				

Table 4. Confusion matrix for the STEW dataset.

		Predicted Class		<i>Sens.</i>	<i>Prec.</i>	<i>Spec.</i>	<i>Acc.</i>
		Negative	Positive				
Actual Class	Negative	98.1% (TN)	1.9% (FP)	99.1%	98.1%	98.1%	98.6%
	Positive	0.9% (FN)	99.1% (TP)				

Table 5. Confusion matrix for the MAT dataset.

space where linear separation is feasible. This makes SVM particularly effective for modeling intricate decision boundaries.

The second-ranked model, random forest, is an ensemble made up of many decision trees. By aggregating the outputs of these trees, each trained on a different subset of data and features, it effectively addresses the overfitting issue common to single decision trees. Nevertheless, its decision boundaries are a blend of axis-aligned boundaries from its individual trees, which may be less effective for specific data types than the hyperplane optimized by an SVM. The lower performance of *k*NN and decision tree classifiers is understandable. Decision trees tend to overfit the training data and rely on simple, axis-aligned boundaries that can't effectively separate classes with complex relationships. Similarly, *k*NN struggles in high-dimensional spaces due to the curse of dimensionality, where measuring distance between samples becomes less meaningful. This reduces its ability to distinguish between samples, leading to decreased performance.

The classification accuracy comparison clearly demonstrates that the optimization step significantly improves all classifiers across both datasets. Specifically, optimization boosts accuracy by 3.1% to 4.1%, underscoring its crucial role. These results strongly suggest that the optimization process is a vital part of the proposed framework, essential for maximizing the models' discriminant capacity and achieving optimal outcomes.

A comprehensive evaluation of the proposed classifier on the STEW dataset was conducted using the confusion matrix shown in Table 4. The analysis of the matrix indicates a well-balanced ability to distinguish between the two classes, correctly identifying 98.4% of true positives and 95.8% of true negatives. Error rates were minimal, with a false-positive rate of 4.2% and a false-negative rate of 1.6%. Key metrics support these results; a high sensitivity of 98.4% highlights excellent positive detection, while a specificity of 95.8% confirms accurate negative classification. The precision of 95.9% further validates the robustness of positive predictions. Overall, the high accuracy, balanced sensitivity and specificity, and low error rates demonstrate that the proposed model is a reliable and effective classifier for this task.

The model's performance on the MAT dataset, as shown in Table 5, not only confirms the high performance observed on the STEW dataset but also shows significant improvements across all metrics. Specifically, the false positive rate has decreased from 4.2% to 1.9%, and the false negative rate has decreased from 1.6% to 0.9%. This reduction in errors has directly led to improvements in key metrics, including specificity from 95.8% to 98.1% and precision from 95.9% to 98.1%, while the model's very high sensitivity has also increased to 99.1%. Overall, these results show that the model achieved more accurate and balanced classification on the MAT dataset and further demonstrated its reliability and robustness by significantly reducing errors of both types.

The accuracy of frequency bands

To assess how various EEG frequency bands contribute to mental workload classification, the model's accuracy was calculated separately for the delta, theta, alpha, beta, and gamma bands, as well as for the entire signal across the STEW and MAT datasets. As shown in Fig. 5, the highest classification accuracy in both datasets was achieved using the full signal. This indicates that features from different frequency bands provide complementary and valuable information, and combining them is crucial for optimal performance.

The Alpha band emerged as the most effective and informative, achieving the highest accuracy among the bands. Conversely, the Gamma band offered the least relevant information for classification, resulting in the lowest accuracy across both datasets. The performance hierarchy, from strongest to weakest, is consistent across both datasets: alpha, beta, delta, theta, and gamma. Additionally, the analysis shows that the model performs better across all conditions on the MAT dataset than on the STEW dataset, demonstrating the robustness of the proposed model.

The effect of deep network architecture

To identify the most effective architecture for the CNN-TFAN network and assess the impact of the TFAN module, the model's accuracy was evaluated across different numbers of convolutional blocks. The findings, shown in Table 6, indicate that increasing the network depth by adding more blocks enhances classification accuracy. Specifically, in both the STEW and MAT datasets, the three-block architecture consistently outperforms the one- and two-block models. The beneficial influence of the TFAN module is also evident, as its inclusion in

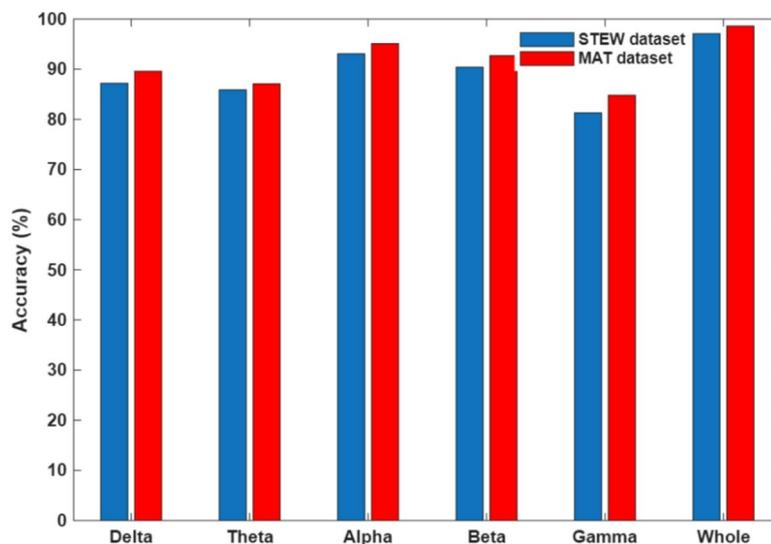


Fig. 5. The contributions of different frequency bands of the EEG signal to the MWL classification.

Dataset	The number of convolutional blocks					
	With the TFAN module			Without the TFAN module		
	1	2	3	1	2	3
STEW	90.2	94.7	97.1	87.1	89.8	94.5
MAT	91.7	95.6	98.6	88.6	90.9	95.2

Table 6. The effect of the number of convolutional blocks on the accuracy of mental workload classification.

Method	Dataset		Method	Dataset	
	STEW	MAT		STEW	MAT
STFT	88.2	89.6	Hilbert transform	91.3	92.4
Spectrogram	88.9	90.2	SST	93.4	94.2
FAWT	89.5	90.9	FSST	94.7	95.5
CWT	89.8	91.1	WSST	95.9	96.3
Stockwell	90.4	91.7	MSST	97.1	98.6

Table 7. Performance comparison between different time-frequency methods.

all structures significantly boosts accuracy. For example, in the three-block setup, applying the TFAN module increased accuracy from 94.5% to 97.1% on the STEW dataset and from 95.2% to 98.6% on the MAT dataset. These results demonstrate the TFAN module's effectiveness in discriminant feature extraction and confirm that a deeper architecture with three convolutional blocks provides superior performance for this task.

The effect of time-frequency analysis

To determine the best feature extraction technique, we evaluated the performance of ten different time-frequency analysis methods, with full results shown in Table 7. This comparison reveals a distinct performance ranking. Traditional methods such as the short-time Fourier transform (STFT), spectrograms, and the continuous wavelet transform (CWT) achieved classification accuracies between 88% and 92%—meanwhile, SST-based methods performed notably better, surpassing 93% accuracy. Ultimately, MSST was identified as the top performer, attaining the highest accuracy across both datasets.

The apparent advantage of SST-based methods comes from their functional approach to addressing the limitations of traditional techniques. Classic methods like STFT and CWT face an inherent trade-off between time and frequency resolution due to the uncertainty principle, leading to 'energy smearing' and a blurry signal representation. The synchrosqueezing technique, a robust post-processing method, sharpens this blurry image by reallocating energy in the time-frequency plane to reflect accurate instantaneous frequencies. Univariate versions, such as WSST, process each EEG channel individually. In contrast, MSST's strength lies in its multivariate approach; it disambiguates across all channels simultaneously, allowing it to model and extract

Method	Accuracy		Method	Accuracy	
	STEW	MAT		STEW	MAT
PCA	95.4	97.1	UMAP	96.2	97.7
LDA	96.1	97.6	t-SNE	95.6	96.5
KPCA (RBF)	95.9	97.3	SDA	97.1	98.6

Table 8. Performance comparison between different time–frequency methods.

Reference / Year	Method	Accuracy (%)
Ref. [57] /2021	SWT + Optimized-KNN (MAT)	94.00
Ref. [58]/2022	Ci-SSA + KNN (MAT)	95.28
Ref. [60]/2023	Ci-SSA + BHHO + KNN (MAT, STEW)	96.88
Ref. [59]/2023	Pooling function + SVM (MAT)	96.42
Ref [19]/2024	VMD + LightGBM (MAT)	97.22
	VMD + LightGBM (STEW)	95.51
Ref [61]/2021	Graph features + non-parametric Wilcoxon test + SVM (STEW)	89.6
Ref [1]/2024	Effective connectivity + Top-AUC + Forward feature selection + SVM (STEW)	89.53
This work / 2025	MSST + CNN + TFAN + SDA + SVM (MAT)	98.86
	MSST + CNN + TFAN + SDA + SVM (STEW)	97.1

Table 9. Performance comparison with other works in WML classification.

the brain network's interdependencies and spatiotemporal dynamics. This information, often overlooked by univariate methods, is vital for accurate classification of complex brain activity.

The effect of feature reduction

To better understand how feature reduction impacts classification results, we tested several dimensionality reduction techniques on deep features from the CNN-TFAN network. Specifically, we examined PCA, LDA, KPCA (RBF), UMAP, t-SNE, and compared them with SDA. Using the same SVM classifier and evaluation, the accuracy on the STEW and MAT datasets is shown in Table 8. LDA generally outperforms PCA by improving accuracy from 95.4% to 96.1% on STEW and from 97.1% to 97.6% on MAT, because PCA only considers variance, not class labels, whereas LDA separates classes after projection. However, LDA doesn't match SDA, which leverages the data's structure beyond labeled data. Nonlinear methods like KPCA with RBF kernels score 95.9% on STEW and 97.3% on MAT, showing benefits but sensitivity to kernel setup. UMAP outperforms others with 96.2% on STEW and 97.7% on MAT, likely due to maintaining local relationships. t-SNE, which focuses on visualization, performs worse at classification, with scores of 95.6% on STEW and 96.5% on MAT. Overall, SDA consistently produces the most effective reduced feature space, achieving 97.1% on STEW and 98.6% on MAT. This is because SDA preserves class separation by maintaining discriminative structure and leveraging the data's geometry through graph-based regularization.

Comparison with other works

To validate the proposed model's effectiveness, a thorough performance comparison was performed against several recent state-of-the-art MWL classification methods. Details of this comparison are shown in Table 9. The proposed approach, integrating MSST, CNN, TFAN, SDA, and SVM, outperforms other methods on both the MAT and STEW datasets.

Sharma et al.⁵⁷ reported an accuracy of 94% using an SWT and Optimized-KNN approach. This was improved by Yedukondalu and Sharma⁵⁸ who achieved 95.28% with the Ci-SSA and kNN methods. Baygin et al.⁵⁹ reached 96.42% using a pooling function with SVM, while Yedukondalu and Sharma⁶⁰ attained 96.88% by combining Ci-SSA and BHHO with kNN. More recently, Jain et al.¹⁹ achieved a 97.22% accuracy with VMD and a LightGBM classifier. In contrast, the model proposed here significantly surpasses all previous methods, setting a new accuracy record of 98.86% on the MAT dataset.

Similarly, on the STEW dataset, the proposed model outperforms existing techniques—previous studies by Zhu et al.⁶¹ and Safari et al.¹ reported accuracies of 89.6% and 89.53%, respectively, using graph features and effective connectivity with SVM classifiers. More recent methods have achieved higher accuracy, such as the VMD-based method with the LightGBM model¹⁹, which reached 95.51%, and the method proposed in⁶⁰, which achieved 96.88%. The proposed model exceeds these results, attaining an accuracy of 97.1%. These findings consistently show that the proposed framework, i.e., MSST + CNN + TFAN + SDA + SVM, presents a more robust and accurate solution for MWL classification.

Statistical analysis

To assess the strength and reliability of our proposed framework, we conducted a statistical analysis of the classification results on both the STEW and MAT datasets. Since EEG-based deep learning models can be

sensitive to how data is split and initialized, we repeated the entire training and evaluation process multiple times with different random seeds, keeping the same subject-wise data split. For each run, we recorded the classification accuracy, which was then used for further statistical analysis.

The proposed method, which combines MSST, CNN-TFAN, SDA, and SVM, achieved an average classification accuracy of $97.1\% \pm 0.5\%$ on the STEW dataset and $98.6\% \pm 0.3\%$ on the MAT dataset. The minor standard deviations suggest that the model's performance is consistent across different runs and isn't heavily influenced by initial settings or data partitioning. To see if these improvements are statistically meaningful, the proposed approach was compared to two leading alternatives: (i) the top existing time–frequency method based on WSST, and (ii) the same CNN architecture without the TFAN module. Since the data didn't fully meet the assumptions of normality, we used the Wilcoxon signed-rank test, a nonparametric method, for these comparisons. The results showed that the proposed method significantly outperformed both baselines across both datasets, with p-values < 0.01 in all cases. This indicates that the improvements are unlikely to be due to mere chance. Effect size analysis using rank-biserial correlation showed significant effects (over 0.6), confirming that these gains are both statistically and practically important. To better understand the uncertainty, we calculated 95% confidence intervals for the average accuracy: [96.0%, 98.2%] for STEW and [97.8%, 99.3%] for MAT. These intervals further support the consistent performance of our framework across multiple tests. Overall, this analysis demonstrates that the improvements brought by the MSST–CNN-TFAN approach are reliable, statistically significant, and reproducible, strengthening the credibility of our results.

Conclusion

This paper introduced a new framework for MWL classification using EEG signals. It employs MSST to generate precise time–frequency representations that incorporate spatial dependencies across EEG channels. Deep features are then extracted using a novel CNN architecture with an integrated time–frequency attention (TFAN) module, which enables the network to focus on key regions of the representation and extract more relevant information. Using SDA for semi-supervised feature reduction and SVM for classification, the model achieved accuracies of 97.1% on the STEW dataset and 98.6% on the MAT dataset. The results demonstrate that this comprehensive framework surpasses previous methods, with each component, particularly MSST and TFAN, effectively enhancing the model's accuracy and robustness.

The next step is to evaluate the MSST–CNN-TFAN framework using more realistic protocols that don't rely on individual subjects and, where possible, on larger datasets, such as leave-one-subject-out or cross-session validation. This will help us better understand how well the system generalizes. To make the model more robust against differences between subjects and sessions, future efforts could include domain adaptation and transfer learning techniques, such as fine-tuning pre-trained CNNs on EEG time–frequency data or using feature alignment methods to reduce distribution gaps. Since signal quality can affect EEG modeling, exploring advanced preprocessing methods—such as MSPCA/wavelet–PCA denoising, ICA artifact removal, and automated artifact detection—could improve reliability, especially in noisier real-world settings. On the modeling front, the current two-step approach (SDA + SVM) might evolve into a more integrated learning framework that incorporates supervised contrastive learning, metric learning, or transformer-based classifiers, while preserving the interpretability of time–frequency attention. Improving interpretability further by visualizing attention maps over time and frequency, and analyzing the importance of specific channels and frequencies, can help connect model decisions to neurophysiological patterns of workload. Finally, for practical use, future research should aim to enhance computational efficiency through lightweight attention modules, pruning, quantization, and streaming inference, making real-time mental workload monitoring in human–machine interactions more feasible.

Data availability

The STEW dataset analyzed during the current study is available in the [Dataset Files.zip] repository, *<https://iee-dataport.org/open-access/stew-simultaneous-task-eeg-workload-dataset>*. Also, the MAT dataset is available in the [Download the ZIP file] repository, *<https://physionet.org/content/eeqmat/1.0.0/>*.

Received: 16 October 2025; Accepted: 31 December 2025

Published online: 09 January 2026

References

1. Safari, M., Shalhaf, R., Bagherzadeh, S. & Shalhaf, A. Classification of mental workload using brain connectivity and machine learning on electroencephalogram data. *Sci. Rep.* **14** (1), 9153 (2024).
2. Demirezen, G., Taşkaya, T., Temizel & Brouwer, A. M. Reproducible machine learning research in mental workload classification using EEG. *Front. Neuroergonomics*. **5**, 1346794 (2024).
3. Kingphai, K. & Moshfeghi, Y. Mental workload assessment using deep learning models from Eeg signals: A systematic review. *IEEE Trans. Cogn. Dev. Systems*. **17**, 40–69 (2025).
4. Morales, S. & Bowers, M. E. Time-frequency analysis methods and their application in developmental EEG data. *Dev. Cogn. Neurosci.* **54**, 101067 (2022).
5. Bhalerao, S. V. & Pachori, R. B. Imagined Speech–EEG detection using multivariate swarm sparse Decomposition-Based joint Time–Frequency analysis for intuitive BCI. *IEEE Trans. Human-Machine Systems*. **55**, 347–357 (2025).
6. Hassan, J., Reza, M. S., Ahmed, S. U., Anik, N. H. & Khan, M. O. EEG workload Estimation and classification: a systematic review. *Journal Neural Engineering*. **22**, 051003 (2024).
7. Sadiq, M. T., Akbari, H., Siuly, S., Li, Y. & Wen, P. Alcoholic EEG signals recognition based on phase space dynamic and geometrical features. *Chaos Solitons Fractals*. **158**, 112036 (2022).
8. Sadiq, M. T. et al. Exploiting pretrained CNN models for the development of an EEG-based robust BCI framework. *Comput. Biol. Med.* **143**, 105242 (2022).

9. Madhavan, S., Tripathy, R. K. & Pachori, R. B. Time–frequency domain deep convolutional neural network for the classification of focal and non-focal EEG signals. *IEEE Sens. J.* **20** (6), 3078–3086 (2019).
10. Lin, S., Zeng, Y. & Gong, Y. Learning of time–frequency attention mechanism for automatic modulation recognition. *IEEE Wirel. Commun. Lett.* **11** (4), 707–711 (2022).
11. Ahrabian, A., Looney, D., Stanković, L. & Mandic, D. P. Synchrosqueezing-based time–frequency analysis of multivariate data. *Sig. Process.* **106**, 331–341 (2015).
12. Xue, Q., Vodolazskii, D., Wu, H., Song, Y. & Nando, M. Real-time cognitive load monitoring of fusion remote maintenance system operators by electroencephalogram. *Fusion Eng. Des.* **215**, 114920 (2025).
13. Pillai, P., Balasingam, B., Jaekel, A. & Biondi, F. N. Comparison of concurrent cognitive load measures during n-back tasks. *Appl. Ergon.* **117**, 104244 (2024).
14. Zammouri, A., Ait, A., Moussa & Chevallier, S. Use of cognitive load measurements to design a new architecture of intelligent learning systems. *Expert Syst. Appl.* **237**, 121253 (2024).
15. Gkintoni, E., Antonopoulou, H., Sortwell, A. & Halkiopoulos, C. Challenging cognitive load theory: the role of educational neuroscience and artificial intelligence in redefining learning efficacy. *Brain Sci.* **15** (2), 203 (2025).
16. Yedukondalu, J., Sharma, L. D. & Bhattacharyya, A. Cognitive load detection using adaptive/fixed-frequency empirical wavelet transform and multi-domain feature optimization. *Biomed. Signal Process. Control.* **110**, 108124 (2025).
17. Zhang, J., Li, J. & Wang, R. Instantaneous mental workload assessment using time–frequency analysis and semi-supervised learning. *Cogn. Neurodyn.* **14** (5), 619–642 (2020).
18. Ghasimi, A. & Shamekhi, S. Enhanced EEG-based cognitive workload detection using RADWT and machine learning. *Neuroscience*, **569**, 231–244 (2025).
19. Jain, P., Yedukondalu, J., Chhabra, H., Chauhan, U. & Sharma, L. D. EEG-based detection of cognitive load using VMD and LightGBM classifier. *Int. J. Mach. Learn. Cybernet.* **15** (9), 4193–4210 (2024).
20. Yedukondalu, J. & Sharma, L. D. Cognitive load detection using circulant singular spectrum analysis and binary Harris Hawks optimization based feature selection. *Biomed. Signal Process. Control.* **79**, 104006 (2023).
21. Üyulan, C. & Erguzel, T. T. Analysis of Time–Frequency EEG feature extraction methods for mental task classification. *Int. J. Comput. Intell. Syst.* **10** (1), 1280–1288 (2017).
22. Kyriaki, K., Koukopoulos, D. & Fidas, C. A. A comprehensive survey of Eeg preprocessing methods for cognitive load assessment. *IEEE Access.* **12**, 23466–23489 (2024).
23. Sadiq, M. T. et al. Motor imagery BCI classification based on multivariate variational mode decomposition. *IEEE Trans. Emerg. Top. Comput. Intell.* **6** (5), 1177–1189 (2022).
24. Sadiq, M. T. et al. Motor imagery EEG signals decoding by multivariate empirical wavelet transform-based framework for robust brain–computer interfaces. *IEEE access.* **7**, 171431–171451 (2019).
25. 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 (2022).
26. 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 (2020).
27. Safari, M., Shalhaf, R., Bagherzadeh, S. & Shalhaf, A. Classification of mental workload with EEG analysis by using effective connectivity and a hybrid model of CNN and LSTM. *Computer Methods Biomech. Biomedical Engineering*, **29**, 218–232 (2026).
28. Kongwudhikunakorn, S. et al. EEGMeNet: End-to-End Multi-Task neural network for Brain-Based mental workload classification. *IEEE Internet Things Journal.* **12**, 42573–42589 (2025).
29. Hassan, I. et al. Cognitive load Estimation using a hybrid Cluster-Based unsupervised machine learning technique. *IEEE Access.* **12**, 118785–118801 (2024).
30. Chaturvedi, S. & Ahirwal, M. K. Investigation of various validation and testing schemes for classification of cross-session mental workload to match real-life expectations of BCI: A transfer learning and XAI approach. *Biomed. Signal Process. Control.* **112**, 108523 (2026).
31. Aksu, Ş. H., Çakıt, E. & Dağdeviren, M. Mental workload assessment using machine learning techniques based on Eeg and eye tracking data. *Appl. Sci.* **14** (6), 2282 (2024).
32. Teymourlouei, A., Hu, M., Gentili, R. & Reggia, J. Functional connectivity methods for multi-class mental workload classification, in *46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1–4 (IEEE, 2024)
33. Li, Z. et al. MST-net: A multi-scale Swin transformer network for EEG-based cognitive load assessment. *Brain Res. Bull.* **206**, 110834 (2024).
34. Yedukondalu, J., Sharma, D. & Sharma, L. D. Subject-wise cognitive load detection using time–frequency EEG and Bi-LSTM. *Arab. J. Sci. Eng.* **49** (3), 4445–4457 (2024).
35. Murata, A. An attempt to evaluate mental workload using wavelet transform of EEG. *Hum. Factors.* **47** (3), 498–508 (2005).
36. Yedukondalu, J. et al. Cognitive load detection through EEG lead wise feature optimization and ensemble classification. *Sci. Rep.* **15** (1), 842 (2025).
37. Wang, Z., Ouyang, Y. & Zeng, H. ARFN: an attention-based recurrent fuzzy network for EEG mental workload assessment. *IEEE Trans. Instrum. Meas.* **73**, 1–14 (2024).
38. Shao, S. et al. EEG-based mental workload classification method based on hybrid deep learning model under IoT. *IEEE J. Biomedical Health Inf.* **28** (5), 2536–2546 (2023).
39. Khan, M. A. et al. Application of artificial intelligence in cognitive load analysis using functional near-infrared spectroscopy: A systematic review. *Expert Syst. Appl.* **249**, 123717 (2024).
40. Pušica, M. et al. Mental workload classification and tasks detection in multitasking: deep learning insights from EEG study. *Brain Sci.* **14** (2), 149 (2024).
41. Chen, J., Fan, F., Wei, C., Polat, K. & Alenezi, F. Decoding driving States based on normalized mutual information features and hyperparameter self-optimized Gaussian kernel-based radial basis function extreme learning machine. *Chaos Solitons Fractals.* **199**, 116751 (2025).
42. Chen, J., Cui, Y., Wei, C., Polat, K. & Alenezi, F. Driver fatigue detection using EEG-based graph attention convolutional neural networks: an end-to-end learning approach with mutual information-driven connectivity. *Applied Soft Computing*, 114097 (2025).
43. Chen, J., Cui, Y., Wei, C., Polat, K. & Alenezi, F. Advances in EEG-Based emotion recognition: Challenges, Methodologies, and future directions. *Applied Soft Computing*, 113478 (2025).
44. Sadiq, M. T. et al. Toward the development of versatile brain–computer interfaces. *IEEE Trans. Artif. Intell.* **2** (4), 314–328 (2021).
45. Yu, X., Aziz, M. Z., Sadiq, M. T., Fan, Z. & Xiao, G. A new framework for automatic detection of motor and mental imagery EEG signals for robust BCI systems. *IEEE Trans. Instrum. Meas.* **70**, 1–12 (2021).
46. Lim, W. L., Sourina, O. & Wang, L. P. Simultaneous task EEG workload data set. *IEEE Trans. Neural Syst. Rehabil. Eng.* **26** (11), 2106–2114 (2018).
47. Zyma, I. et al. Electroencephalograms during mental arithmetic task performance, *Data*, **4**(1), 14 (2019).
48. Teplan, M. Fundamentals of EEG measurement. *Meas. Sci. Rev.* **2** (2), 1–11 (2002).
49. Sadiq, M. T., Yu, X., Yuan, Z. & Aziz, M. Z. Motor imagery BCI classification based on novel two-dimensional modelling in empirical wavelet transform. *Electron. Lett.* **56** (25), 1367–1369 (2020).
50. Wang, P., Gao, J. & Wang, Z. Time–frequency analysis of seismic data using synchrosqueezing transform. *IEEE Geosci. Remote Sens. Lett.* **11** (12), 2042–2044 (2014).

51. Cai, D., He, X. & Han, J. Semi-supervised discriminant analysis, in *IEEE 11th international conference on computer vision*, 1–7 (IEEE, 2007)
52. Cortes, C. & Vapnik, V. Support-vector networks. *Mach. Learn.* **20** (3), 273–297 (1995).
53. Naseer, N., Qureshi, N. K., Noori, F. M. & Hong, K. S. Analysis of different classification techniques for two-class functional near-infrared spectroscopy-based brain-computer interface, *Computational intelligence and neuroscience*, **1**, 5480760 (2016).
54. Isa, N. M., Amir, A., Ilyas, M. & Razalli, M. Motor imagery classification in brain computer interface (BCI) based on EEG signal by using machine learning technique. *Bull. Electr. Eng. Inf.* **8** (1), 269–275 (2019).
55. Breiman, L. Random forests. *Mach. Learn.* **45** (1), 5–32 (2001).
56. Shahriari, B., Swersky, K., Wang, Z., Adams, R. P. & De Freitas, N. Taking the human out of the loop: A review of Bayesian optimization, *Proceedings of the IEEE*, **104**(1), 148–175 (2015).
57. Sharma, L. D., Chhabra, H., Chauhan, U., Saraswat, R. K. & Sunkaria, R. K. Mental arithmetic task load recognition using EEG signal and bayesian optimized K-nearest neighbor. *Int. J. Inform. Technol.* **13** (6), 2363–2369 (2021).
58. Yedukondalu, J. & Sharma, L. D. Cognitive load detection using binary salp swarm algorithm for feature selection, in *IEEE 6th Conference on Information and Communication Technology (CICT)*, 1–5 (IEEE, 2022).
59. Baygin, N. et al. Automated mental arithmetic performance detection using quantum pattern-and triangle pooling techniques with EEG signals. *Expert Syst. Appl.* **227**, 120306 (2023).
60. Yedukondalu, J. & Sharma, L. D. Cognitive load detection using Ci-SSA for EEG signal decomposition and nature-inspired feature selection. *Turkish J. Electr. Eng. Comput. Sci.* **31** (5), 771–791 (2023).
61. Zhu, G., Zong, F., Zhang, H., Wei, B. & Liu, F. Cognitive load during multitasking can be accurately assessed based on single channel electroencephalography using graph methods. *IEEE Access.* **9**, 33102–33109 (2021).

Author contributions

Zahed Nouri: Conceptualization, Investigation, Methodology, Software, Writing - Original Draft; Asghar Charmin: Conceptualization, Methodology, Validation, Supervision; Hashem Kalbkhani: Conceptualization, Software, Visualization, Writing - Review & Editing; Saeed Barghandan: Methodology, Writing - Review & Editing.

Declarations

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to H.K.

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

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

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

© The Author(s) 2026