Mühendislik Bilimleri ve Tasarım Dergisi
13(2), 466-479, 2025
e-ISSN: 1308-6693

Journal of Engineering Sciences and Design
DOI: 10.21923/jesd.1669626

Araştırma Makalesi		Research Article

CLASSIFICATION OF COGNITIVE WORKLOAD FROM EEG SIGNALS USING MULTIDIMENSIONAL FEATURES WITH MACHINE LEARNING AND DEEP LEARNING

Yavuz Bahadır KOCA[1*]

[1] Afyon Kocatepe University, Faculty of Engineering, Electrical Engineering Department, Afyonkarahisar, Türkiye

Keywords
*EEG,
Cognitive Workload,
Machine Learning,
Deep Learning,
Feature Extraction,
Biomedical.*

Abstract
This study aims to classify cognitive workload levels from EEG signals. EEG signals from 48 subjects under resting and task cognitive load conditions were analyzed. Noise and artifacts were removed by applying band-pass and notch filtering methods in the 1-50 Hz band on the EEG data. Then, the EEG data were segmented with the windowing technique in 256 and 512 sample sizes, and a total of 309 features based on time, frequency, and complexity were extracted. Using the extracted feature set, logistic regression (LR), support vector machines (SVM), k-nearest neighbor (k-NN), random forest (RF), XGBoost machine learning (ML) algorithms and deep neural networks (DNN), one-dimensional convolutional neural networks (1D-CNN) and long short-term memory (LSTM) deep learning (DL) methods were applied for multi-class classification. In the experimental results, the highest success was obtained in the XGBoost model with a 99.4% accuracy rate and 0.990 Cohen's kappa value, and in DL methods, a 98.75% accuracy rate and 0.981 Kappa value in the LSTM model. This study reveals that integrating multidimensional features obtained from EEG signals with both ML algorithms and DL models provides high accuracy in cognitive workload classification.

MAKİNE ÖĞRENMESİ VE DERİN ÖĞRENME İLE ÇOK BOYUTLU ÖZELLİKLER KULLANILARAK EEG SİNYALLERİNDEN KOGNİTİF İŞ YÜKÜNÜN SINIFLANDIRILMASI

Anahtar Kelimeler
*EEG,
Kognitif İş Yükü,
Makine Öğrenmesi,
Derin Öğrenme,
Özellik Çıkarımı,
Biyomedikal.*

Öz
Bu çalışmada EEG sinyallerinden kognitif iş yükü seviyelerinin sınıflandırılması amaçlanmıştır. 48 deneğe ait dinlenme ve görev kognitif yük koşullarındaki EEG sinyalleri analiz edilmiştir. EEG verileri üzerinde 1-50 Hz bandında bant geçiren ve çentik filtreleme yöntemleri uygulanarak gürültü ve artefaktlar temizlenmiştir. Daha sonra, EEG verileri 256 ve 512 örnek boyutlarında pencereleme tekniğiyle segmente edilerek zaman, frekans ve karmaşıklık temelli toplam 309 öznitelik çıkarılmıştır. Elde edilen öznitelik seti kullanılarak, çok sınıflı sınıflandırma işlemi için lojistik regresyon, destek vektör makineleri, k-en yakın komşu, rastgele orman, XGBoost makine öğrenmesi algoritmaları ile derin sinir ağları (DNN), tek boyutlu konvolüsyonel sinir ağları (1D-CNN) ve uzun kısa süreli bellek (LSTM) gibi derin öğrenme yöntemleri uygulanmıştır. Deneysel sonuçlarda en yüksek başarı, 99.4% doğruluk oranı ve 0,990 kohen kappa değeri ile XGBoost modelinde, derin öğrenme yöntemlerinde ise 98.75% doğruluk oranı ve 0,981 kappa değeri ile LSTM modelinde elde edilmiştir. Bu çalışma, EEG sinyallerinden elde edilen çok boyutlu özelliklerin hem makine öğrenmesi algoritmaları hem de derin öğrenme modelleriyle entegrasyonunun kognitif iş yükü sınıflandırmasında yüksek doğruluk sağladığını ortaya koymaktadır.

Alıntı / Cite
Koca, Y.B. (2025). Classification of Cognitive Workload from EEG Signals Using Multidimensional Features with Machine Learning and Deep Learning, Journal of Engineering Sciences and Design, 13(2), 466-479.

Yazar Kimliği / Author ID (ORCID Number)	Makale Süreci / Article Process	
Y.B. Koca, 0000-0002-0317-1417	**Başvuru Tarihi / Submission Date**	03.04.2025
	Revizyon Tarihi / Revision Date	18.04.2025
	Kabul Tarihi / Accepted Date	27.04.2025
	Yayım Tarihi / Published Date	27.06.2025

* İlgili yazar / Corresponding author: ybkoca@gmail.com, ybkoca@aku.edu.tr +90-272-211-1267

CLASSIFICATION OF COGNITIVE WORKLOAD LEVELS FROM EEG SIGNALS USING MULTIDIMENSIONAL FEATURES WITH MACHINE LEARNING AND DEEP LEARNING

Yavuz Bahadır Koca[1†]

[1] Afyon Kocatepe University, Faculty of Engineering, Electrical Engineering Department, Afyonkarahisar, Türkiye

Highlights

- Cognitive workload levels were classified by extracting features based on time, frequency, and complexity from EEG signals.
- Among the machine learning methods, the XGBoost algorithm, 99.4% showed the highest classification performance. In deep learning methods, the LSTM model, 98.75% accuracy modeled time-dependent relationships in EEG signals and exhibited high performance.
- Channel-based fractal analysis results revealed that EEG signal complexity increased significantly in the frontal and temporal regions as cognitive load increased.

Graphical Abstract

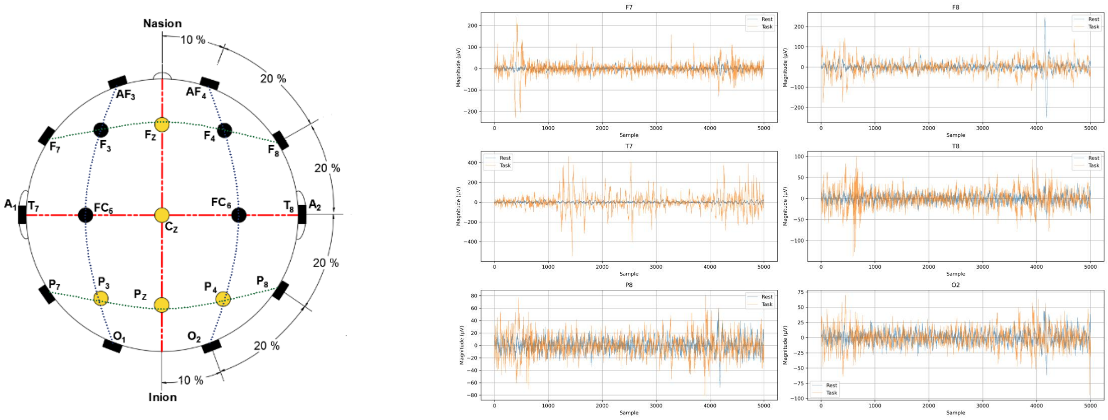


Figure. EEG electrode placement and multidimensional neural responses in rest and cognitive tasks

Purpose and Scope

The purpose of this study is to provide effective classification of cognitive workload levels from EEG signals. EEG-based analyses can provide significant benefits for improving performance, especially in occupations requiring attention and concentration, and in daily life by accurately determining people's cognitive load levels.

Design/methodology/approach

The research was conducted on EEG recordings obtained from the STEW dataset. EEG data were first cleaned with band-pass and notch filtering methods, then analyzed by dividing into small sections with the windowing method. Time, frequency and complexity features were extracted for each window. Machine learning methods and deep learning techniques were used for classification.

Findings

In the analyses performed, the best classification results were obtained with the XGBoost algorithm from machine learning methods with an accuracy rate of 99.4%. From deep learning methods, the LSTM model showed high performance with an accuracy rate of 98.75%. In addition, in channel-based analyses, it was determined that the complexity of EEG signals in the frontal and temporal regions increased significantly as cognitive load increased.

Originality

The originality of this study is that it classifies cognitive load levels with high accuracy using a multidimensional and comprehensive feature set from EEG data. Thanks to detailed channel-based examinations, it is shown which brain regions are more active under load. In this respect, a different analysis in the field of cognitive load assessment is presented with the combined evaluation of machine learning and deep learning methods.

[†] Corresponding author: ybkoca@gmail.com, ybkoca@aku.edu.tr, +90-272-211-1267

1. Introduction

Cognitive load refers to the balance between an individual's capacity to process information and the cognitive demands they encounter. It is very important in various fields such as education, health, and technology (Sweller, 1988; Sweller et al., 2019). When this balance is disrupted, an individual's decision-making ability, attentional process, and information processing skills can be negatively affected. Long-term intense cognitive load can manifest as cognitive fatigue, leading to decreased performance, increased probability of making errors, and loss of motivation (Taddeini et al., 2025).

Last decades, information exposure is increasing day by day, and this significantly affects people's mental activities (Wang, 2024). As the complexity of decision-making processes increases, how the human mind works under cognitive load and how mental fatigue occurs as a result of this becomes increasingly important (Mundlos et al., 2024). Especially, increasing technology use has a significant effect on cognitive load and mental fatigue. Long periods spent in front of a smartphone, tablet, or computer shorten the attention span of individuals and lead to the rapid depletion of mental resources. Increased screen exposure increases cognitive load due to constantly divided attention and digital multitasking habits, which can make mental fatigue chronic in the long term (Amalakanti et al., 2024; Sheng, 2025; Skowronek et al., 2023). In addition, excessive use of digital devices disrupts sleep patterns, disrupts brain resting processes, and prevents cognitive functions such as attention, memory, and arithmetic ability from working efficiently (Holding et al., 2021). In this context, correct and healthy screen use strategies are of great importance for human health in terms of managing cognitive load and mental fatigue.

Cognitive fatigue is a condition associated with physical and cognitive performance disorders that occur as a result of long-term cognitive strain (Weiler et al., 2025). It can lead to serious consequences in both individual and professional contexts. For example, long-term cognitive deterioration can reduce employee productivity, make learning processes difficult, and negatively affect daily decision-making processes (Chen et al., 2016; Kunasegaran et al., 2023; Mizuno et al., 2011). Moreover, when cognitive fatigue is examined at neurophysiological and psychological levels, it is also important with its energy consumption in the brain, decrease in cortical activity, and negative effects on working memory (Lorist et al., 2005).

This study aims to classify cognitive workload levels with high accuracy using comprehensive and diverse features obtained from EEG signals. In this context, unlike previous studies, time, frequency, and complexity-based multi-dimensional analyses (Shannon entropy, Higuchi fractal dimension, Hjorth parameters, Hurst coefficient) were performed on EEG data. In addition, the temporal changes of EEG signals and their differences according to brain regions were determined in detail. Thus, it shows which parts of the brain regions are more active under high cognitive load. This approach makes a significant contribution to the accurate and reliable measurement of cognitive workload.

2. Literature Survey

Electroencephalography (EEG), offers the opportunity to evaluate cognitive load by detecting brain activity associated with cognitive tasks and analyzing EEG signals according to frequency bands (Park and Chung, 2020; Zafar et al., 2017). EEG, is a non-invasive technique and is a physiological signal measurement method used to analyze significant changes in brain waves. EEG signals are divided into alpha, beta, gamma and theta frequency bands, each associated with different cognitive processes. Delta waves are generally associated with deep sleep and are less pronounced in awake states. However, increased delta activity can be seen depending on the individual's psychological state. This manifests itself as an inappropriate state of arousal and decreased attention (Howells et al., 2018). Alpha waves are associated with relaxed, awake states and are most pronounced in the occipital region of the brain (Ono et al., 2023). Beta waves are associated with active thinking, problem solving and motor control (Borra et al., 2023). They are unsuccessful in inhibiting the response under high cognitive demand tasks (Taddeini et al., 2025). Gamma waves are associated with higher-level cognitive functions, including perception and consciousness (Archila-Meléndez et al., 2020). For example, gamma band power is effective in distinguishing single-task and multi-task scenarios and achieves high classification accuracy in detecting cognitive workload (Korkmaz et al., 2024).

Studies show that cognitive fatigue negatively affects cognitive functions and impairs decision-making processes. In aviation, cognitive fatigue can cause deterioration of performance and reactions to unpredictable events. It can be evaluated that this situation weakens cognitive functions and affects decision-making mechanisms (Hamann and Carstengerdes, 2023). So, in recent years, machine learning (ML) algorithms with artificial intelligence-supported analyses have made a significant difference in measuring cognitive fatigue and cognitive load and have offered innovative approaches. With technological developments, various methods have been developed in this context to detect cognitive fatigue. In studies, how ML and deep learning (DL) techniques are integrated with EEG data, methods for automatically estimating cognitive workload and cognitive fatigue are examined with different algorithms and analyses.

Using EEG signals, individuals' cognitive workload levels can be classified as low, medium, and high. Lim et al. (2018) performed cognitive load classification with EEG signals to measure the cognitive load of people during multitasking. Akman (2021) performed classification with Katz and Higuchi's fractal dimension algorithms, feature extraction, support vector machines (SVM), k-nearest neighbor (k-NN), and quadratic discriminant analysis (QDA) methods. In a study conducted using EEG signals and DL algorithms, cognitive fatigue of construction workers was determined with an accuracy rate of 88.85% (Wang et al., 2023). In addition, studies on cognitive fatigue detection with ML using fMRI data have achieved an accuracy rate of 73% (Zadeh et al., 2020). Kamrud et al. (2021) developed cross-participant EEG models to show that cognitive load can serve as a consistent physiological marker across different individuals and tasks. Zhou et al. (2022) discussed how EEG data are used in cognitive load estimation systems through preprocessing, feature extraction, and classification steps.

Examining studies on the relationship between EEG-based functional connectivity metrics and cognitive load, Safari et al. (2024) have investigated cognitive fatigue using hierarchical feature selection and support vector-based connectivity metrics. Gupta et al. (2021) have employed DL methods to classify cognitive load through brain connectivity analysis in EEG data. Li et al. (2023) developed EEG-based fatigue detection systems with a brain cognitive dynamic recognition network and showed how different brain regions are related to cognitive load. In studies on cognitive fatigue detection, Zeng et al. (2021) addressed the difficulties of estimating cognitive fatigue in different individuals using an EEG-based transfer learning method. Li et al. (2024) developed a new method for the assessment of cognitive overload using EEG signals. Karmakar et al. (2024) aimed to increase user experience by performing real-time cognitive load detection with EEG signals. Shafiei et al. (2024) combined EEG and eye-tracking data with the XGBoost model to estimate cognitive load in surgical tasks. Zhou et al. (2025) improved cognitive load detection in air traffic control operators by using an EEG-based hybrid DL model. Khan et al. (2024) performed cognitive load analysis using functional infrared spectroscopy data instead of EEG and stated that DL models showed superior performance. Among the studies examining EEG spectral analysis in cognitive load measurement, Chikhi et al. (2022) stated that frontal theta bands are a sensitive marker for cognitive load estimation. (Roy et al., 2019) examined the application of DL models in EEG and emphasized such as CNN and RNN provide successful results for EEG analysis. But they explained there are some problems, such as reproducibility. So, cognitive load and cognitive fatigue negatively affect both cognitive and physical performance.

3. Material and Method

In this study, the simultaneous task EEG workload (STEW) dataset containing EEG signals collected under cognitive workload was used. The dataset included EEG recordings of 48 participants at rest and under high cognitive load conditions. A study was carried out on 45 people due to the lack of data of three of the participants. The participants were conducted on university students who did not have a neurological or psychiatric disorder and had not participated in the EEG study before. The study was conducted with the approval of the Ethics Committee of Nanyang Technological University (IRB-2014-04-026). EEG signals were obtained with the Emotiv EPOC device consisting of 14 electrodes (AF3, F7, F3, FC5, T7, P7, O1, O2, P8, T8, FC6, F4, F8, AF4). The sampling frequency is 128 Hz, and the resolution is 16 bits. Participants were subjected to 2 separate EEG analyses: resting and multitasking test. It performed a multi-task test using the SIMKAP module of the Vienna Test System. In the study, EEG data were collected in a three-minute resting state and an eighteen-minute visual matching and auditory arithmetic test. In the resting state, the first and last 15 seconds were omitted, and a total of 2.5 minutes of clean EEG data was obtained for analysis. In the multitasking test, the first parts were not used due to slow-paced temporary processes, and the last 2.5 minutes of the EEG recording were recorded for evaluation. At the end of each segment, participants rated their cognitive load level on a scale of 1 to 9 (Lim et al., 2018). Accordingly, scores 1 to 3 were classified as "low load", scores 4 to 6 as "medium load", and scores 7 to 9 as "high load" categories. Figure 1 shows the layout of the Emotiv EPOC 14-electrode system.

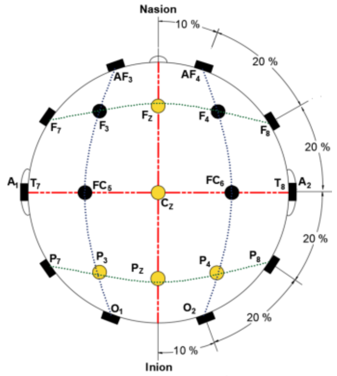


Figure 1. Emotiv EPOC 14 Electrode Placement

A flowchart for an overview of the analysis used in this study is given in Figure 2. The process started with the acquisition of the raw EEG signal. After filtering and segmentation, features were extracted from the time, frequency, and complexity domains. These features were then combined into a hybrid feature vector and used to train machine learning and deep learning models for multi-class classification. The classification results were finally evaluated using performance metrics such as accuracy, F1 score, and Cohen's Kappa.

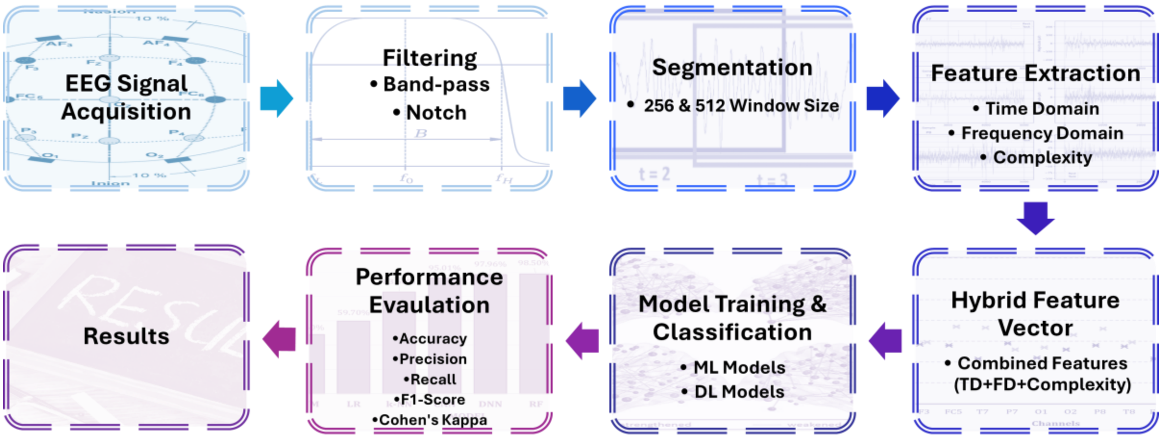


Figure 2. Workflow of the EEG-based cognitive workload classification

3.1. Data Preprocessing

In this study, first, filtration and artifact removal procedures were performed on EEG signals in the STEW data set, respectively. In order to select the relevant frequency, range from EEG signals, the Butterworth bandpass filtering method was used. The filtering process reduces noise and artifacts at undesirable low (<1 Hz) and high (>50 Hz) frequencies in the EEG signal, preserving the 1–50 Hz frequency band required for analysis. All neuro-physiologically significant EEG frequency bands were made available as delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-30 Hz), and gamma (30-50 Hz). Thus, the quality of EEG data and the accuracy of the analysis have been increased. In addition, the notch filtering method was used to eliminate 50 Hz network interference caused by the electrical network during EEG recordings. In the pre-processing phase of EEG signals, 512 and 256-dimensional windowing procedures were performed. Then, the feature extraction process was started for each window. In this way, the data set was enriched by combining the features obtained with different window sizes and it was ensured that the features in EEG signals at different time resolutions were included in the analysis. Thus, it is aimed to increase the model performance in the separation of EEG signals between states.

3.2. Feature Extraction

EEG signals reflect the temporal and frequency characteristics of electrical activity in the brain. The attributes extracted from these signals provide information about numerous parameters such as the individual's cognitive state, attention level, mental load, and relaxation level. In this context, after the pre-processing steps for the collected EEG signals, feature extraction processes were applied sequentially. After the signals were separated into fixed-length windows for analysis, the attributes of each window were extracted. It is aimed to use these attributes as parameters with high distinctiveness in analyzing mental states. In this context, time domain properties, frequency domain properties, Shannon entropy, Hjorth parameters, Hurst exponential coefficients and Higuchi Fractal Dimension (HFD) analysis were applied respectively.

3.2.1. Time Domain Properties

Mean, variance, root mean square (RMS), and zero-crossing rate (ZCR) calculations were made in the study as time domain features. With the average, the calculation of the center value of the signal for each channel is made. Variance also shows the measure of variability in signal amplitude. With Equations (1) and (2), the mean and variance calculations are determined, respectively. Both of these calculations are considered basic identifiers for classification algorithms. Here, too, it was used to determine the basic statistical distribution of EEG signals. The root mean square (RMS) is calculated by taking the square root of the mean of the squares of the signal. It is obtained by Equation (3). It reflects the energy of the signal. It is especially effective in the analysis of active brain activities in EEG signals. High RMS values indicate that the amplitude of the signal is high and therefore more neural activity is present. A low RMS, on the other hand, indicates a more sluggish signal and is usually associated with a resting state. Therefore, while the RMS value is expected to increase during the task, this value may decrease at rest. Finally, the Zero-Crossing Rate (ZCR) was calculated. ZCR indicates the oscillation level of the signal. This is determined by calculating how many times the signal crosses the zero axis. It provides noise and frequency information. The ZCR value is determined by Equation (4). This value gives an idea of the density of the high-frequency components in the signal. The higher the ZCR, the faster, the more active signal; a lower ZCR reflects a slower and stationary state. This trait tends to increase during tasks that require particular attention. The expression 1 in the ZCR formula is 1 when there is a zero transition, and 0 in the opposite cases.(Stancin et al., 2021; Subasi, 2007).

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i \tag{1}$$

$$\sigma^2 = \frac{1}{N}\sum_{i=1}^{N} (x_i - \mu)^2 \tag{2}$$

$$RMS = \sqrt{\frac{1}{N}\sum_{i=1}^{N} x_i^2} \tag{3}$$

$$ZCR = \frac{1}{N-1}\sum_{i=1}^{N-1} 1_{\{x_i \, . \, x_{i+1} < 0\}} \tag{4}$$

3.2.2. Frequency Domain Characteristics

EEG signals can be analyzed by separating them into different frequency bands. Each band represents a specific cognitive state. The delta band is associated with deep sleep and unconscious processes. The Theta band shows an increase in meditation and low attention states. The alpha band usually occurs during moments of rest with the eyes closed and represents a relaxed but alert state. On the other hand, the beta band rises during alertness, focus, and cognitive tasks. During the mission, an increase in beta power and a decrease in alpha and theta powers are expected. With Equation (5) the calculated power values for each band obtained by frequency analysis can be determined. Here, $X(f)$ is the Fourier transform of the signal, and is the lower and upper bounds of the respective frequency $f_1 f_2$.

$$P_{band} = \sum_{f=f_1}^{f_2} |X\,(f)|^2 \tag{5}$$

3.2.3. Entropy-Based and Complexity-Based Features

The basic Shannon entropy was used in the study. Shannon entropy is a parameter that measures the irregularity and randomness of the signal. It measures the amount of information contained in the signal. High entropy indicates a more chaotic and complex signal, while low entropy indicates a more orderly and predictable structure.

Entropy is expected to increase as the brain performs more intensive cognitive operations during mental tasks. On the other hand, entropy remains at lower levels during moments of rest. The equation is calculated by (6). Here p_i are the probabilities obtained from the amplitude histogram of the signal.

$$H = -\sum_{i=1}^{N} p_i \, \log_2(p_i) \qquad (6)$$

Hjorth parameters consist of three components: Activity, Mobility and Complexity, which measure the variability and complexity of the EEG signal over time. Activity represents the variance of the signal and reflects the signal strength, similar to RMS. Mobility shows how fast the signal is changing; This value corresponds to the frequency information and may increase in situations that require attention. Complexity, on the other hand, is the ratio of the second derivative of the signal to the first derivative; The higher this ratio, the more complex the structure of the signal. Complexity usually increases under cognitive load. The calculations for the Hjorth parameters are presented in Equation (7).

$$\begin{aligned} Activity &= Var\,(x(t)) \\ Mobility &= \sqrt{\frac{Var\,(x'(t))}{Var\,\big(x(t)\big)}} \qquad (7) \\ Complexity &= \frac{Mobility\,\big(x'(t)\big)}{Mobility\,(x(t))} \end{aligned}$$

The Hurst exponential coefficient reveals long-term correlations by measuring the long-range dependency of the signal. The equation is calculated by (8). Here, the length of the time series is expressed by N, the interval of cumulative deviations is expressed by R, and the standard deviation of the time series is denoted by S. The HFD size is a criterion that determines the geometric complexity of the signal. A higher fractal size indicates that the signal has a more complex structure and contains more cognitive activity. The lower fractal dimension, on the other hand, reflects more static and regular brain activity. It usually takes values between 1.0 and 2.0 and is expected to increase during the task. Here, $L(k)$k is the length of the signal obtained by sampling in length. In Equations (8) and (9), the calculation of the hurst exponential coefficient and the calculation of the HFD are given, respectively.

$$H = \frac{\log(R/S)}{\log(N)} \qquad (8)$$

$$HFD = \lim_{k \to 0} \frac{\log L\,(k)}{\log\left(\frac{1}{k}\right)} \qquad (9)$$

3.3. Classification Analysis

This study constructs a hybrid feature structure by combining various features extracted from the EEG signals. Specifically, time-domain features (mean, variance, RMS, ZCR), frequency-domain features (delta, theta, alpha, beta, gamma), and complexity-related metrics (Shannon entropy, Hjorth parameters, Hurst exponent, and Higuchi fractal dimension) are combined to construct a comprehensive feature vector. Here, it is aimed to capture both the statistical features, spectral characteristics, and nonlinear dynamics of the EEG signals. The obtained hybrid feature vector contained 309 features. It was uniformly used as the input for all ML and DL models.

In this study, ML algorithms and DL models were applied to evaluate the performance of time, frequency and entropy-based hybrid features extracted from EEG data in classifying cognitive workload levels (Low, Medium, High). The models were tested on a multi-class structure and compared with various performance metrics. Logistic Regression (LR) was applied with a multinomial structure. The model evaluated cognitive workload levels with metrics such as ROC-AUC, accuracy and complexity matrix. Random Forest (RF), a community learning algorithm consisting of 100 trees, was used. The model has a strong structure against overfitting, and it also provides the opportunity to examine the importance levels of the features. XGBoost, this powerful algorithm based on Boosting, has been used to maximize the separation between classes. k-NN was applied with a value of k=5 and the classification of data points according to their close neighbors was carried out. SVM are operated on multi-class structure with RBF core function. To reduce the effect of the class imbalance problem observed in the dataset, balancing strategies were applied to the models. In this context, the *class_weight='balanced'* parameter was enabled. This method automatically determined weights according to the frequency of the classes and ensured that minority classes were represented adequately in the training process. In this way, the weighting approach provided balance while also reducing the risk of overfitting. In addition, for the DL models, stratified sampling was used during data splitting to maintain the class ratios between the training and test sets.

The Deep Neural Network (DNN) uses a network structure that contains two hidden layers with 128 and 64 neurons following the input layer and arranged by Dropout (0.3). The output layer is configured with *softmax* activation for three classes. The model was trained over 50 epochs and validation accuracy was monitored. One-Dimensional Convolutional Neural Network (1D-CNN), was used to capture the temporal patterns of the EEG feature matrix. First, a Conv1D layer with 32 filters was applied, followed by MaxPooling1D. This was followed by a second Conv1D with 64 filters and a re-pooling layer. The network is flattened and connected to a fully connected layer with 64 neurons and finally to the *softmax* output layer. The model has been trained over 50 epochs. In Long Short-Term Memory (LSTM), EEG attributes were reshaped as one-time step sequences and a two-layer LSTM network was implemented. The first LSTM layer was configured with 64 units with the parameter *return_sequences=True*; followed by the addition of a second layer of LSTM with 32 units. Dropout (0.3) is applied between the two layers, and the last output layer is configured with *softmax* function. The model is trained over 50 epochs. The detailed architectures of the implemented DL models, including layer configurations, activation functions, epochs, and dropout rates, are summarized in Table 1.

Table 1. DL architectures used for EEG-based cognitive workload classification

MODEL	ARCHITECTURE	ACTIVATION	EPOCH	DROPOUT
DNN	Dense(128) → Dropout(0.3) → Dense(64) → Dropout(0.3) → Dense(3)	ReLU / Softmax	50	0.3
1D-CNN	Conv1D(32) → MaxPool → Conv1D(64) → Flatten → Dense(64) → Dense(3)	ReLU / Softmax	50	0.3
LSTM	LSTM(64) → Dropout(0.3) → LSTM(32) → Dense(3)	Tanh / Softmax	50	0.3

3.4. Performance Criteria

In this study, multiple performance measures were used to evaluate cognitive workload classification models. These criteria determine model success not only in terms of overall accuracy. At the same time, it was selected to take into account the power of discrimination, general disequilibrium situations and random prediction probabilities for each class.

Accuracy is the ratio of the number of samples that the model correctly guesses in all classes to the total number of samples. It is often used as a key performance indicator, but it can be misleading in an unbalanced class distribution. Precision is the ratio of true positive predictions for each class to the total sample estimated to belong to that class. It is especially preferred in cases where false positives are significant. Sensitivity is the ratio of true positive predictions for each class to all of the actual samples of that class. It stands out especially in cases where missed samples are important. The F1-Score is the harmonic mean of the Precision and Recall values. It offers a balanced assessment, especially in data sets where there is a class imbalance. In this study, macro mean (equal weighted average of all classes) was used. The ROC-AUC measures how well the model can discriminate between classes. Since there is more than one class, the macro average ROC-AUC value was calculated. With Cohen's Kappa, it is a coefficient of consistency that excludes random agreement between the actual labels and the model predictions. Improves reliability in data sets with class imbalance. As the coefficient κ approaches 1, it expresses high consistency, and as it approaches 0, it expresses luck-based results. All of these metrics were calculated using the Python programming language, through functions in the scikit-learn and TensorFlow/Keras libraries.

4. Experimental Results

In this section, the performance of multidimensional features obtained from EEG signals in the classification of cognitive workload levels is analyzed. In this context, performance comparisons were made by applying both ML algorithms and DL approaches in the study. For each model, evaluation was carried out with metrics such as accuracy, recall, specificity, F1-score and Cohen's Kappa, and the most successful methods were determined. In addition, temporal EEG signals were analyzed on a channel basis and the differences between task and rest states were visualized. These signal patterns obtained after filtration processes are important in terms of showing the effect of cognitive load especially in the frontal, temporal and occipital regions. In this context, the effect of cognitive workload increase on EEG signals is presented with Figure 3 for a subject who was subjected to testing. Here, time series comparisons of task and rest conditions were made over filtered EEG data. For this purpose, F7, F8, T7, T8, P8 and O2 channels representing frontal, temporal, parietal and occipital regions were examined.

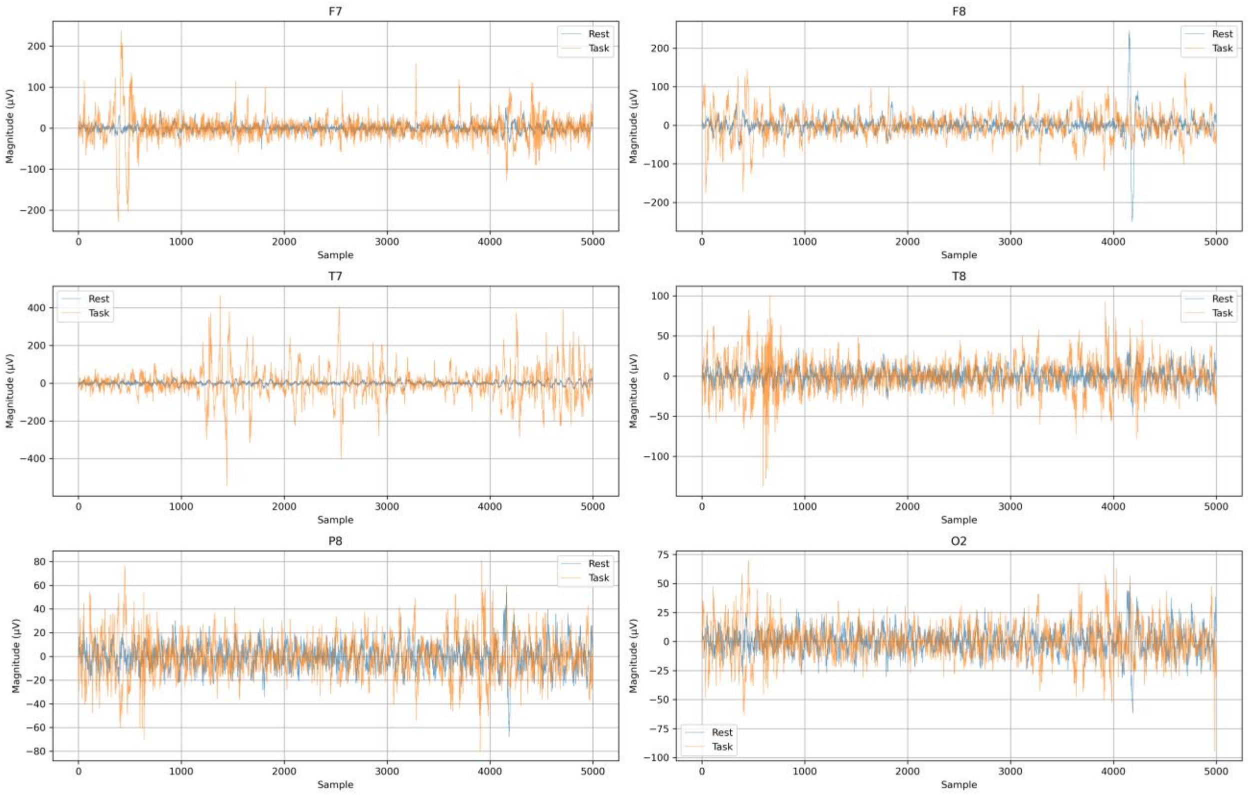


Figure 3. Effect of cognitive workload on EEG signals in channel-based on time analysis

In the time series graphs obtained after the filtering process, amplitude increases of around 300–400 μV ± were observed in the F7 and T7 channels under the duty condition. This indicates that increased cognitive effort and attention are concentrated in the frontal and left temporal regions. The O2 channel produced more erratic and high-amplitude signals during the task. This indicates an increase in visual information processing activity in the right occipital cortex. Increases in amplitude were also observed in the P8 and F8 channels, but these changes remained moderate. In the resting state, the signals in all channels exhibited a more regular, low-amplitude and stable structure. Then, Frequency analysis, HFD and Shannon Entropy were calculated for each channel. HFD and entropy were evaluated in terms of reflecting the structural complexity of the signal and information content changes. The calculation of HFD and Shannon entropy analysis of task and rest for the EEG channel on a channel basis is shown in Figure 4.

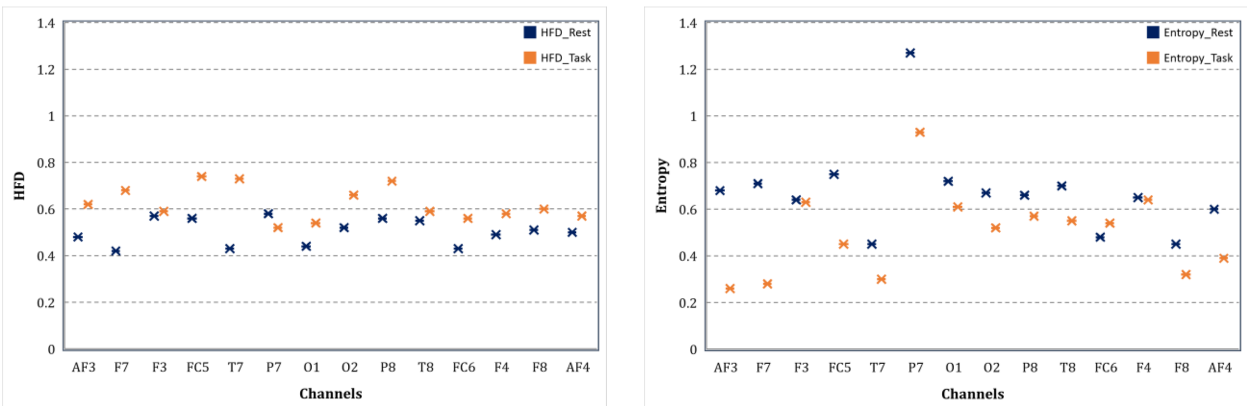


Figure 4. Distribution of average HFD and Entropy values in channels under resting and task conditions (ID: 2).

In the study, a multi-analysis approach was adopted using ML and DL methods. In this study, it was aimed to classify cognitive workload levels by applying ML and DL based classification algorithms on EEG signals collected under rest and multitasking conditions. Signals obtained after feature extraction of data; accuracy, precision, recall and F1-score. In this context, cognitive workload classification was performed on a data set consisting of 13,230 samples using 309 attributes obtained from EEG data. When the class distribution is examined; An uneven distribution of approximately 6174, 3381 and 3675 specimens was observed, including Low, Medium and High. Table 2 presents the performance results of ML and DL algorithms.

Table 2. Performance results of ML and DL algorithms

MODEL	ACCURACY	PRECISION	RECALL	F1-SCORE	COHEN'S KAPPA
LR	0.597	0.572	0.556	0.55	0.362
RF	0.985	0.983	0.98	0.981	0.976
XGBoost	0.994	0.993	0.992	0.992	0.99
k-NN	0.836	0.832	0.819	0.824	0.741
SVM	0.486	0.7	0.36	0.271	0.046
DNN	0.9796	0.9747	0.9748	0.9746	0.968
CNN	0.9501	0.9485	0.9458	0.9471	0.922
LSTM	0.9875	0.9848	0.9845	0.9847	0.981

In the classification made with ML methods, RF and XGBoost stand out as the most successful algorithms. The LR model was applied in a multi-class structure and the overall accuracy of the results obtained was around 59%. In the classification report, a clear distinction was made between the low and high classes, but the recall value remained low in the medium class. The ROC-AUC score indicates a moderate performance of 73.8%. The RF model, consisting of 100 trees, has achieved a very successful distinction between the classes in the dataset. While the accuracy rate obtained was observed as 98.5%, the ROC-AUC score was close to 100%. In the complexity matrix, there is almost no margin of error between classes; This indicates the high generalization ability of the model. The XGBoost algorithm has shown high performance, similar to RF; the accuracy was recorded as 99.4% and the ROC-AUC was 100%. In the classification report, high precision, recall and F1-score values were obtained in all classes. The accuracy of the results obtained with the k-NN model (k=5) was determined as approximately 83.6%. In the performance of the model, it was observed that the confusion between classes was more pronounced; The ROC-AUC value was around 93.8%. The SVM model with RBF core has underperformed compared to other methods. While the overall accuracy was found to be 48.6%, there was serious confusion, especially between the Low and Medium classes. The ROC-AUC score was below the desired level with 67.2%. Figure 5 shows the accuracy performance results of different models. Comparative analysis provides insights into how each model performs in terms of accuracy.

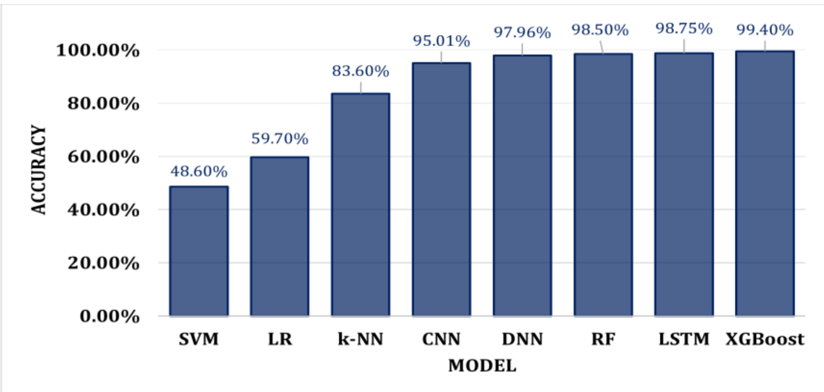
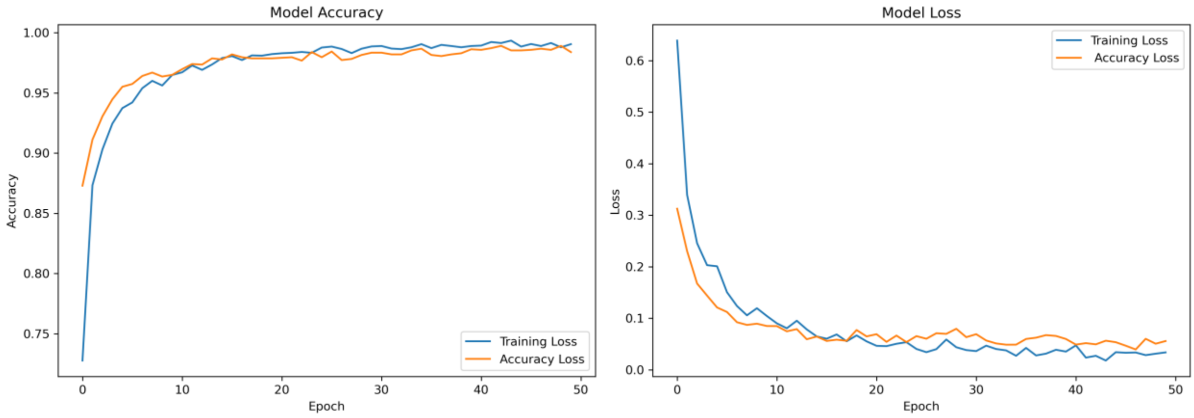
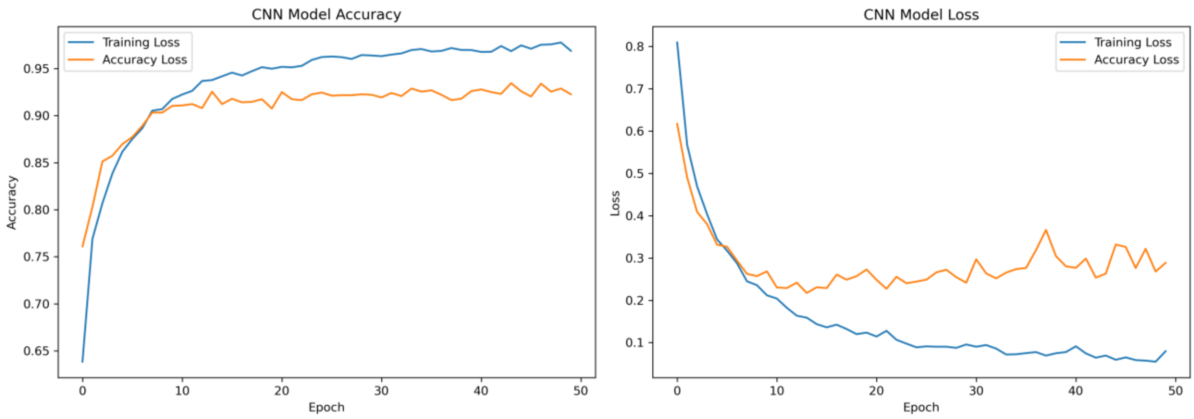


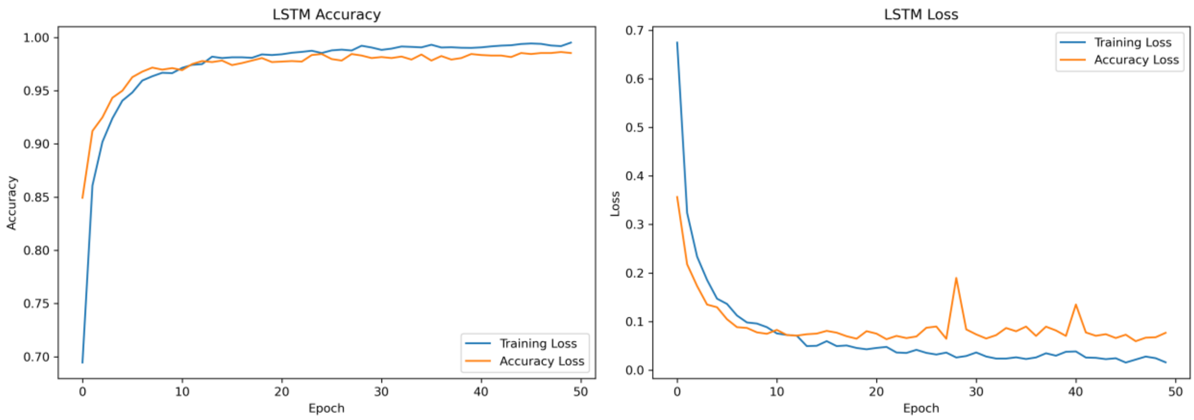
Figure 5. Models' accuracy performance results



(a)



(b)



(c)

Figure 6. Training and validation performances (a) DNN model: Accuracy and loss curves indicate stable convergence after epoch 30 with minimal overfitting (b) CNN model: While training accuracy continues to increase, validation loss shows signs of overfitting after epoch 30 (c) LSTM model: High training and validation accuracy with stable loss values throughout training

In DL methods, the performances of DNN, CNN and LSTM methods were evaluated, respectively. In the DNN model, which was carried out using a multi-layer artificial neural network, the training process was continued for 50 epochs and the model gave successful results with an accuracy rate of 98.3%. In the classification report, high precision, recall and F1-score values were obtained in all classes. The ROC-AUC value is 99.88%, supporting the overall performance of the model. He focused on capturing local patterns in EEG data with the One-Dimensional Convolutional Neural Network (1D-CNN). The accuracy obtained in this model is 94.2%, and the ROC-AUC is 99.19%. In the classification report, although high achievement was shown in all grades, some minor confusion was observed. In order to model the time series structure well, the LSTM network was used, and the accuracy rate of this model was determined as 98.53%. In the complexity matrix, the model has achieved an almost perfect distinction; The ROC-AUC was at 99.89%. LSTM has performed in parallel with the CNN and DNN models in capturing dependencies over time. The training and accuracy loss graphs obtained for DNN, CNN, and LSTM are presented with Figure 6.

5. Result and Discussion

In this study, cognitive workload levels that occur during rest and multitasking were attempted to be classified using EEG data. For this purpose, both ML and DL models were tested and compared to see which one gave more successful results. Since there was an imbalance between classes in the data set, this directly affected the model's success. As a result of the tests, the XGBoost model showed the highest success (99.4% accuracy, 0.992 F1 score, 0.990 Kappa). RF and LSTM models were also quite successful. While RF reached 98.5% accuracy and 0.976 Kappa, the LSTM model attracted attention on the DL side with 98.75% accuracy and 0.981 Kappa. This shows their suitability for models such as LSTM, especially for time-varying EEG data.

Although the CNN model falls slightly behind DNN and LSTM (with over 95% accuracy and 0.922 Kappa), it can still be considered a strong alternative. It was noticed that it was a bit confused, especially between low and high workload classes, and showed a slight tendency to over-fit. This model could perform even better with more data or some adjustments. The DNN model, on the other hand, gave consistent results in the training and validation processes. It provided a balanced approach to EEG data with 97.96% accuracy and 0.968 Kappa value. The k-NN algorithm gave lower but still acceptable results compared to the others (83.6% accuracy, 0.741 Kappa). This is most likely due to the k value not being well-tuned and the poor distance calculations in high-dimensional data. Simpler ML models, such as SVM and LR, performed lower than expected. While SVM made almost random predictions with 48.6% accuracy, the LR model had difficulty distinguishing classes, especially at medium load, with 59.7% accuracy. Overall, the XGBoost and LSTM models were the most successful in distinguishing cognitive workload levels. The main factors affecting the model's success include data imbalance, the types of features used, and the ability of the model to process time-varying data.

In this study conducted with the STEW dataset, EEG-based workload classification was tested under multi-channel and different conditions. Both ML and DL models were systematically tested in this study. In addition, methods such as channel-based fractal analysis and feature extraction by dividing into time windows were also used. According to channel-based Higuchi fractal analysis, the complexity of the signal increased during the task in some EEG channels. Time-varying EEG signals were divided into segments, which contributed to the more effective operation of time-sensitive models, especially LSTM. Several different analyses and perspectives may come to the fore in terms of future studies. In this context, end-to-end DL models that work with raw EEG data and process the signal directly instead of feature engineering can be tried. In addition, it can be understood more clearly which EEG channels or features the models are based on with methods such as SHAP or LIME. Transfer learning approaches can be applied to provide generalization across different task types or individuals. Such models can be integrated into real-time applications (e.g., driving simulations, training systems) and made more functional. In addition to EEG data, multiple systems that will work with physiological signals such as heart rate and skin conductance also have a strong potential for future research. In conclusion, the combined use of ML and DL methods has yielded quite successful results in cognitive workload classifications using EEG. This approach provides a strong option in terms of both accuracy and model diversity, creating a rich foundation for future studies.

Conflict of Interest

No conflict of interest was declared by the author.

References

Akman Aydın, E., 2021. EEG Sinyalleri Kullanılarak Zihinsel İş Yükü Seviyelerinin Sınıflandırılması. Politeknik Dergisi 24, 681–689. https://doi.org/10.2339/politeknik.794655

Amalakanti, S., Mulpuri, R.P., Avula, V.C.R., Reddy, A., Jillella, J.P., 2024. Impact of smartphone use on cognitive functions: A PRISMA-guided systematic review. Medicine India 0, 1–8. https://doi.org/10.25259/medindia_33_2023

Archila-Meléndez, M.E., Valente, G., Gommer, E.D., Correia, J.M., ten Oever, S., Peters, J.C., Reithler, J., Hendriks, M.P.H., Cornejo Ochoa, W., Schijns, O.E.M.G., Dings, J.T.A., Hilkman, D.M.W., Rouhl, R.P.W., Jansma, B.M., van Kranen-Mastenbroek, V.H.J.M., Roberts, M.J., 2020. Combining Gamma With Alpha and Beta Power Modulation for Enhanced Cortical Mapping in Patients With Focal Epilepsy. Front Hum Neurosci 14. https://doi.org/10.3389/fnhum.2020.555054

Borra, D., Fantozzi, S., Bisi, M.C., Magosso, E., 2023. Modulations of Cortical Power and Connectivity in Alpha and Beta Bands during the Preparation of Reaching Movements. Sensors 23. https://doi.org/10.3390/s23073530

Chen, Z., Xu, Xianfa, Zhang, J., Liu, Y., Xu, Xianggang, Li, L., Wang, W., Xu, H., Jiang, W., Wang, Y., 2016. Application of LC-MS-based global metabolomic profiling methods to human mental fatigue. Anal Chem 88, 11293–11296. https://doi.org/10.1021/acs.analchem.6b03421

Chikhi, S., Matton, N., Blanchet, S., 2022. EEG power spectral measures of cognitive workload: A meta-analysis. Psychophysiology. https://doi.org/10.1111/psyp.14009

Gupta, A., Siddhad, G., Pandey, V., Roy, P.P., Kim, B.G., 2021. Subject-specific cognitive workload classification using eeg-based functional connectivity and deep learning. Sensors 21. https://doi.org/10.3390/s21206710

Hamann, A., Carstengerdes, N., 2023. Assessing the development of mental fatigue during simulated flights with concurrent EEG-fNIRS measurement. Sci Rep 13. https://doi.org/10.1038/s41598-023-31264-w

Holding, B.C., Ingre, M., Petrovic, P., Sundelin, T., Axelsson, J., 2021. Quantifying Cognitive Impairment After Sleep Deprivation at Different Times of Day: A Proof of Concept Using Ultra-Short Smartphone-Based Tests. Front Behav Neurosci 15. https://doi.org/10.3389/fnbeh.2021.666146

Howells, F.M., Temmingh, H.S., Hsieh, J.H., Van Dijen, A. V., Baldwin, D.S., Stein, D.J., 2018. Electroencephalographic delta/alpha frequency activity differentiates psychotic disorders: A study of schizophrenia, bipolar disorder and methamphetamine-induced psychotic disorder. Transl Psychiatry. https://doi.org/10.1038/s41398-018-0105-y

Kamrud, A., Borghetti, B., Kabban, C.S., Miller, M., 2021. Generalized deep learning eeg models for cross-participant and cross-task detection of the vigilance decrement in sustained attention tasks. Sensors 21. https://doi.org/10.3390/s21165617

Karmakar, S., Kamilya, S., Koley, C., Pal, T., 2024. A Deep Learning Technique for Real Time Detection of Cognitive Load using Optimal Number of EEG Electrodes. IEEE Trans Instrum Meas. https://doi.org/10.1109/TIM.2024.3509604

Khan, M.A., Asadi, H., Zhang, L., Qazani, M.R.C., Oladazimi, S., Loo, C.K., Lim, C.P., Nahavandi, S., 2024. Application of artificial intelligence in cognitive load analysis using functional near-infrared spectroscopy: A systematic review. Expert Syst Appl. https://doi.org/10.1016/j.eswa.2024.123717

Korkmaz, O.E., Korkmaz, S.G., Aydemir, O., 2024. Detection of multitask mental workload using gamma band power features. Neural Comput Appl 36, 10915–10926. https://doi.org/10.1007/s00521-024-09627-9

Kunasegaran, K., Ismail, A.M.H., Ramasamy, S., Gnanou, J.V., Caszo, B.A., Chen, P.L., 2023. Understanding mental fatigue and its detection: a comparative analysis of assessments and tools. PeerJ 11. https://doi.org/10.7717/peerj.15744

Li, P., Zhang, Y., Liu, S., Lin, L., Zhang, H., Tang, T., Gao, D., 2023. An EEG-based Brain Cognitive Dynamic Recognition Network for representations of brain fatigue. Appl Soft Comput 146. https://doi.org/10.1016/j.asoc.2023.110613

Li, Z., Tong, L., Zeng, Y., Gao, Y., Gong, D., Yang, K., Hu, Y., Yan, B., 2024. A novel method of cognitive overload assessment based on a fusion feature selection using EEG signals. J Neural Eng 21, 066047. https://doi.org/10.1088/1741-2552/ad9cc0

Lim, W.L., Sourina, O., Wang, L.P., 2018. STEW: Simultaneous task EEG workload data set. IEEE Transactions on Neural Systems and Rehabilitation Engineering 26, 2106–2114. https://doi.org/10.1109/TNSRE.2018.2872924

Lorist, M.M., Boksem, M.A.S., Ridderinkhof, K.R., 2005. Impaired cognitive control and reduced cingulate activity during mental fatigue. Cognitive Brain Research 24, 199–205. https://doi.org/10.1016/j.cogbrainres.2005.01.018

Mizuno, K., Tanaka, M., Yamaguti, K., Kajimoto, O., Kuratsune, H., Watanabe, Y., 2011. Mental fatigue caused by prolonged cognitive load associated with sympathetic hyperactivity. Behavioral and Brain Functions 7. https://doi.org/10.1186/1744-9081-7-17

Mundlos, P., Wulf, T., Mueller, F.A., 2024. Perceived task complexity in strategic decision situations: the role of cognitive integration and cognitive load. European Business Review. https://doi.org/10.1108/EBR-08-2024-0253

Ono, H., Sonoda, M., Sakakura, K., Kitazawa, Y., Mitsuhashi, T., Firestone, E., Jeong, J.W., Luat, A.F., Marupudi, N.I., Sood, S., Asano, E., 2023. Dynamic cortical and tractography atlases of proactive and reactive alpha and high-gamma activities. Brain Commun 5. https://doi.org/10.1093/braincomms/fcad111

Park, Y., Chung, W., 2020. A Novel EEG Correlation Coefficient Feature Extraction Approach Based on Demixing EEG Channel Pairs for Cognitive Task Classification. IEEE Access 8, 87422–87433. https://doi.org/10.1109/ACCESS.2020.2993318

Roy, Y., Banville, H., Albuquerque, I., Gramfort, A., Falk, T.H., Faubert, J., 2019. Deep learning-based electroencephalography analysis: A systematic review. J Neural Eng. https://doi.org/10.1088/1741-2552/ab260c

Safari, M.R., Shalbaf, R., Bagherzadeh, S., Shalbaf, A., 2024. Classification of mental workload using brain connectivity and machine learning on electroencephalogram data. Sci Rep 14. https://doi.org/10.1038/s41598-024-59652-w

Shafiei, S.B., Shadpour, S., Mohler, J.L., 2024. An Integrated Electroencephalography and Eye-Tracking Analysis Using eXtreme Gradient Boosting for Mental Workload Evaluation in Surgery. Hum Factors. https://doi.org/10.1177/00187208241285513

Sheng, Q., 2025. Understanding the biomechanics of smartphone addiction: The physical and cognitive impacts of prolonged device use on college students. Molecular & Cellular Biomechanics 22, 650. https://doi.org/10.62617/mcb650

Skowronek, J., Seifert, A., Lindberg, S., 2023. The mere presence of a smartphone reduces basal attentional performance. Sci Rep 13. https://doi.org/10.1038/s41598-023-36256-4

Stancin, I., Cifrek, M., Jovic, A., 2021. A review of eeg signal features and their application in driver drowsiness detection systems. Sensors. https://doi.org/10.3390/s21113786

Subasi, A., 2007. EEG signal classification using wavelet feature extraction and a mixture of expert model. Expert Syst Appl 32, 1084–1093. https://doi.org/10.1016/j.eswa.2006.02.005

Sweller, J., 1988. Cognitive Load During Problem Solving: Effects on Learning. Cogn Sci 12, 257–285. https://doi.org/10.1207/s15516709cog1202_4

Sweller, J., van Merriënboer, J.J.G., Paas, F., 2019. Cognitive Architecture and Instructional Design: 20 Years Later. Educ Psychol Rev. https://doi.org/10.1007/s10648-019-09465-5

Taddeini, F., Avvenuti, G., Vergani, A.A., Carpaneto, J., Setti, F., Bergamo, D., Fiorini, L., Pietrini, P., Ricciardi, E., Bernardi, G., Mazzoni, A., 2025. Extended Cognitive Load Induces Fast Neural Responses Leading to Commission Errors. eNeuro 12. https://doi.org/10.1523/ENEURO.0354-24.2024

Wang, J., 2024. Research on the Speed of Information Transmission and User Cognition in the New Media Era. Communications in Humanities Research 40, 204–210. https://doi.org/10.54254/2753-7064/40/20242397

Wang, Y., Huang, Y., Gu, B., Cao, S., Fang, D., 2023. Identifying mental fatigue of construction workers using EEG and deep learning. Autom Constr 151. https://doi.org/10.1016/j.autcon.2023.104887

Weiler, H., Russell, S., Spielmann, J., Englert, C., 2025. Mental Fatigue: Is It Real? Journal of Applied Sport and Exercise Psychology 32, 14–26. https://doi.org/10.1026/2941-7597/a000033

Zadeh, M.Z., Babu, A.R., Lim, J.B., Kyrarini, M., Wylie, G., Makedon, F., 2020. Towards cognitive fatigue detection from functional magnetic resonance imaging data, in: Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments, PETRA '20. Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3389189.3397648

Zafar, R., Dass, S.C., Malik, A.S., 2017. Electroencephalogram-based decoding cognitive states using convolutional neural network and likelihood ratio based score fusion. PLoS One 12. https://doi.org/10.1371/journal.pone.0178410

Zeng, H., Li, X., Borghini, G., Zhao, Y., Aricò, P., Di Flumeri, G., Sciaraffa, N., Zakaria, W., Kong, W., Babiloni, F., 2021. An eeg-based transfer learning method for cross-subject fatigue mental state prediction. Sensors 21. https://doi.org/10.3390/s21072369

Zhou, Y., Huang, S., Xu, Z., Wang, P., Wu, X., Zhang, D., 2022. Cognitive Workload Recognition Using EEG Signals and Machine Learning: A Review. IEEE Trans Cogn Dev Syst. https://doi.org/10.1109/TCDS.2021.3090217

Zhou, Y., Jiang, J., Wang, L., Liang, S., Liu, H., 2025. Enhanced Cognitive Load Detection in Air Traffic Control Operators Using EEG and a Hybrid Deep Learning Approach. IEEE Access. https://doi.org/10.1109/ACCESS.2025.3530091