



Contents lists available at ScienceDirect

Journal of Engineering Research

journal homepage: www.journals.elsevier.com/journal-of-engineering-research

Full Length Article

Hybrid bio-inspired optimized CNN model-based framework for cognitive state classification using neuromorphic OpenBCI EEG signals: An experimental study

Dinesh Kumar Jayaraman Rajendiran^{a,*}, K. Priyadharsini^b, C. Ganesh Babu^c, C. Visvesvaran^b^a Department of Electronics and Communication Engineering, PSG Institute of Technology and Applied Research, Coimbatore, Tamil Nadu, India^b Department of Electronics and Communication Engineering, Sri Krishna College of Engineering and Technology, Coimbatore, Tamil Nadu, India^c Department of Electronics and Communication Engineering, PPG Institute of Technology, Coimbatore, Tamil Nadu, India

ARTICLE INFO

Keywords:

BCI- Brain Computer Interface
 HCNN- Hybrid Convolutional Neural Network
 EEG- Electroencephalography
 G2WO- Grasshopper Grey Wolf Optimization
 Ho- Hybrid optimization
 PSO- Particle Swarm Optimization

ABSTRACT

Interpreting cognitive state is essential for understanding children's learning abilities, including attention, memory, and problem- solving skills, which directly affect academic and social development. Early identification of cognitive irregularities enables timely assessment and individualized support plans. Electroencephalography (EEG) is a non- invasive, effective method for recording neuromorphic brain signals associated with cognitive processes. This study presents a novel hybrid optimization algorithm inspired by the biological visual system to interpret cognitive state from EEG signals recorded by an open- source, low- cost, portable bio- neuromorphic sensing platform (OpenBCI). A robust preprocessing and discriminative feature- extraction method is used to model the complex spatiotemporal patterns of EEG with a Hybrid Convolutional Neural Network (HCNN). To improve classification effectiveness and convergence stability, a Hybrid Optimization (HO) strategy is presented that combines Particle Swarm Optimization (PSO), Grasshopper and Grey Wolf Optimization (G2WO). The hybrid PSO–GWO method leverages collective swarm intelligence and hierarchical hunting behaviour to optimize the learning parameters of a CNN. Results show that the proposed HO- optimized CNN model successfully differentiates between cognitive states with 92. 4% accuracy when applied to OpenBCI EEG signals. Longer chains yield better performance, with extended HO–PSO–G2WO achieving 95.2% accuracy, 93.1% precision and 92.2% recall at the optimal learning rate of 0. 02. Similarly, the Hybrid Optimization on Grasshopper & Grey scale Optimization method (Hybrid PSO + G2WO) offers improvement with 98. 98.21% accuracy, 98. 2% recall, and 98. 98.18% precision at an execution time of 13. 5 s for the provided pre-processed dataset of cognitive-based EEG signals. This research effort is directed toward the reliable classification of cognitive state, not directly to the enhancement of cognitive skills, since it was not tested against any intervention. However, the results demonstrate the efficacy of bio-inspired hybrid optimization in improving DL performance for EEG-based cognitive interpretation. The proposed framework lays a solid foundation for future longitudinal studies and personalized cognitive assessment systems from a neuromorphic EEG perspective.

Introduction

Cognitive development is a crucial aspect of childhood, closely linked to learning, social interaction, and overall mental health. Despite recent advances in psychology and neuroscience, children's cognitive capacities remain difficult to measure accurately because conventional methods, such as behavioural measures and school exams, often fail to capture underlying neural processes [1]. EEG is a non- invasive technique that measures electrical activity over the scalp. EEG technology

has recently become much more accessible, with the advent of portable and affordable EEG acquisition platforms such as OpenBCI. This accessibility enables wider clinical and research applications, particularly in children. However, there remains a lack of research focused on healthy children, as most past work has centered on clinical populations and adults. Combining EEG with cutting- edge technology and using EEG signals to assess cognitive function helps identify early cognitive impairments in children. Recent research in [2] investigated how EEG might be used to identify attention deficits in children with attention-

* Corresponding author.

E-mail address: dineshkumarjr_ece@yahoo.co.in (D.K.J. Rajendiran).<https://doi.org/10.1016/j.jer.2026.08.033>

Received 2 January 2026; Received in revised form 17 August 2026; Accepted 21 August 2026

Available online 23 August 2026

2307-1877/© 2026 The Authors. Published by Elsevier B.V. on behalf of Kuwait University. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

deficit/hyperactivity disorder (ADHD), emphasising the potential for real-time interventions based on EEG patterns [3]. Similarly, the author demonstrated that EEG could correctly classify memory- and attention-related tasks in children by distinguishing between different cognitive states [4]. Electroencephalogram (EEG) signals are classified into different types of brainwaves, each with a distinct frequency range and associated with various mental states. Alpha waves (8–12 Hz) are typically observed during a state of relaxation and alertness, such as when a person is meditating or resting quietly with their eyes closed. Beta waves (13–30 Hz) are generated during active brain functioning, such as when a person is concentrating on a task, problem-solving, or engaged in active thinking. Theta waves (4–8 Hz) tend to occur during deep relaxation, light sleep, and creative thinking and introspection. The frequency-based classification of EEG signals is presented in Fig. 1. These patterns are often observed during deep concentration or when transitions occur from waking to sleep. The slowest waves are delta waves (0.5–4 Hz), which are associated with deep sleep and are important for healing processes. These signals provide information about mental status, relaxation level, and concentration. These brain wave patterns are essential for understanding cognitive functions [2]. These signals can be used to evaluate cognitive performance, identify abnormal brain activity, and measure the effectiveness of mental training or therapy designed to enhance cognitive function. EEG signals are thus valuable for assessing brain function and studying mental abilities.

These studies [1–5] evaluate the role of EEG as a legitimate and impartial instrument for assessing cognitive abilities. In recent years, there has been a revolution in the analysis of EEG signals, driven by AI and the development of DL techniques. EEG signals are complex and non-linear, which often makes classic EEG signal processing methods, such as statistical/time-frequency analysis-based feature extraction, necessary for their analysis. Since the advent of Convolutional Neural Networks (CNN), there has been a significant improvement in the performance of cognitive state classification via EEG signals [4,6]. Many studies have successfully used CNNs to enhance the accuracy of cognitive state detection, owing to their ability to automatically extract hierarchical features from raw EEG data. The authors report notable classification accuracy compared to traditional methods [7]. These positive outcomes have subsequently sparked growing interest in applying such approaches to evaluate children's cognitive abilities, making this an increasingly important area for the use of sophisticated learning models.

One of the key advances in EEG-based data acquisition is the use of OpenBCI electrodes, which offer a low-cost and flexible alternative to traditional EEG systems. This makes them well-suited for large-scale and real-world applications. The schematic of EEG signal acquisition is shown in Fig. 2, using a non-invasive approach to ensure user comfort and signal reliability. With the addition of sophisticated signal processing and Machine Learning (ML) models, OpenBCI systems have great potential to deliver scalable, real-time cognitive assessment solutions. OpenBCI platforms are cost-effective and mobile, enabling deployment in educational and clinical settings to provide personalized assessments based on individual cognitive characteristics, as described by [8]. At the same time, the importance of multimodal cognitive assessment has been highlighted, with EEG data used alongside other modalities such as eye-tracking and behavioural analysis to provide a comprehensive assessment, as suggested by [9]. In this paper, the proposed framework is an innovative analytical approach that combines OpenBCI-based EEG acquisition with DL(DL) techniques, specifically Convolutional Neural Networks (CNNs), to assess children's cognitive states. The framework employs extensive preprocessing methods, efficient feature extraction, and CNN-based classification to accurately recognize cognitive patterns. Its accuracy in classifying core cognitive domains, namely attention, memory, and problem-solving, shows good prospects for implementation in educational and clinical practice. Future work will involve improving the model architecture, expanding

the dataset, and incorporating multimodal inputs to create a more complete and adaptable platform for cognitive assessment [10].

This research highlights the increasing focus on developing children's cognitive skills, recognizing their importance for academic achievement and social well-being. Timely identification and intervention of cognitive impairments is important to reduce future academic problems [11]. The EEG signal is obtained using a non-invasive electrode setup (Ganglion board), as presented in Fig. 3, to record neural activity, and it is reported as a dataset. Although cognitive assessment is crucial, traditional assessment approaches are not scalable, objective, or timely, particularly in low-resource areas. In contrast, EEG-based assessment, when combined with advanced ML and deep learning, offers a more objective and efficient method, capable of detecting subtle neural patterns associated with cognitive functions that can be analyzed in real time and automated, while providing the necessary accuracy. The novelty of this research lies in its approach to using affordable, portable EEG devices and DL to provide detailed, personalized cognitive assessments in real-world settings. This approach aims to enhance cognitive state classification in children using OpenBCI electrodes and CNN models [12]. This work not only brings about significant improvements in accuracy but also has the potential to be widely deployed in education and clinical applications at scale and at lower cost.

The proposed framework employs EEG signal analysis to classify emotional states and to provide an interpretable mapping to associated cognitive tendencies based on established neurophysiological relationships [13]. EEG signals are decomposed into standard frequency bands: (a) theta (4–8 Hz), (a) balpha (8–13 Hz), (c) beta (13–30 Hz), and (d) gamma (30–100 Hz). These signal frequencies are used to analyze the framework performance across key cortical regions, including the frontal, temporal, parietal, and occipital lobes [14]. In this framework, different EEG bands are used as indicators of neural activity patterns associated with emotional states. Alpha and gamma bands are typically associated with relaxed attention [15], working memory, and integrative cognitive processing, and are often observed during positive emotional states such as happiness, as depicted in Fig. 3.

When training and evaluating CNN-based EEG models, the mapping relationships between human emotions, cognitive skills, EEG frequency bands, and associated brain regions, as shown in Table 2, provide a valuable framework within the domain. These relationships enable the network to learn meaningful spatial-spectral patterns that align with known neurological behavior [16]. For instance, patterns associated with happiness based on alpha and gamma activity in frontal and occipital regions [17], help the model recognize states of relaxed and active cognition [18]. Elevated beta and gamma activity in frontal areas enables it to differentiate high-arousal conditions such as fear and anxiety. This organized information facilitates model evaluation because features extracted from the data are physiologically meaningful, improves interpretability because the features are linked to cognitive states, improves class separability, and supports channel selection (using F3, F4, O1, O2) based on the extracted features [19]. Beta band activity is linked to increased alertness, vigilance, and active decision-making, and is typically associated with emotional states of fear, anxiety, and anger. Theta band activity is connected with reflective thinking, monitoring, and internal cognitive processing and is seen in anxiety and sadness. Frontal regions (including F3 and F4) are related to executive control and emotional regulation [20]; parietal regions, to attention and sensory integration; and occipital regions, to visual processing and relaxed awareness. These features are used to perform emotion classification through machine learning or deep learning (ML/DL) models based on EEG signals. The identified emotional state is then correlated with relevant cognitive tendencies or patterns, such as attention, vigilance, and monitoring, as reported in the neuroscience literature. It is important to note that this mapping is not intended to measure cognitive performance directly but should be used to interpret the detected emotional state and provide context. The framework also includes a recommender function that proposes feasible cognitive support

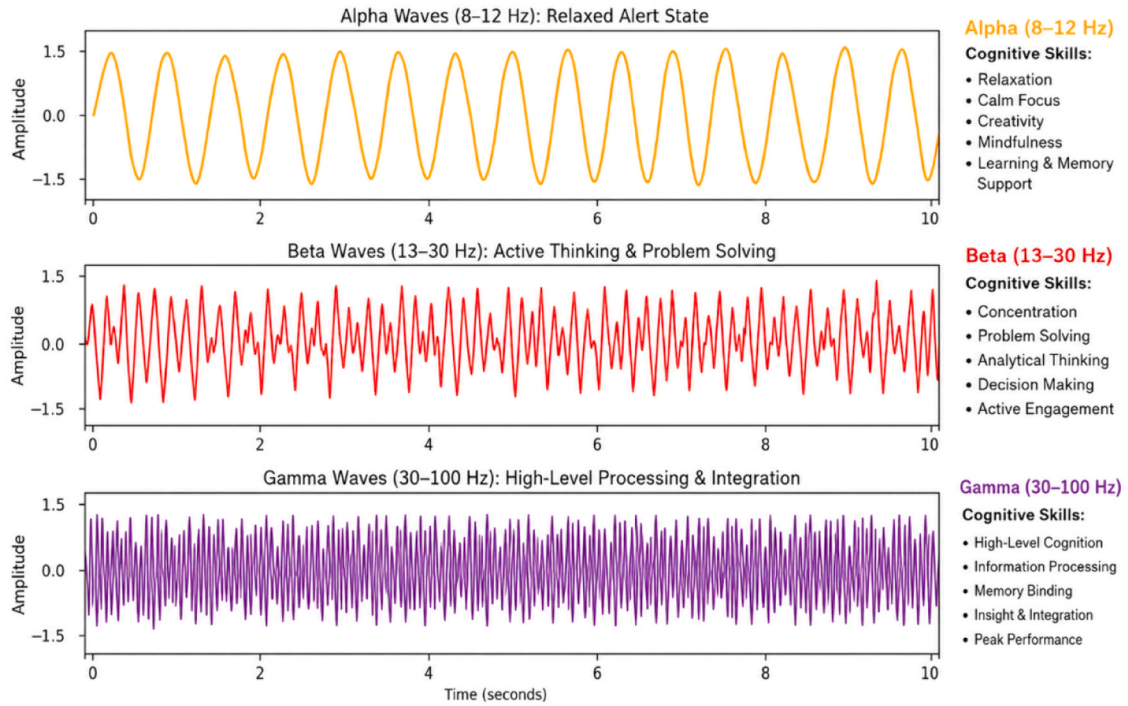


Fig. 1. Representative EEG waveforms illustrating the Alpha (8–12 Hz), Beta (13–30 Hz), and Gamma (30–100 Hz) frequency bands and their corresponding cognitive functions. Alpha waves are associated with a relaxed and attentive mental state, whereas Beta and Gamma waves reflect progressively higher levels of cognitive engagement, active information processing, and complex neural integration.

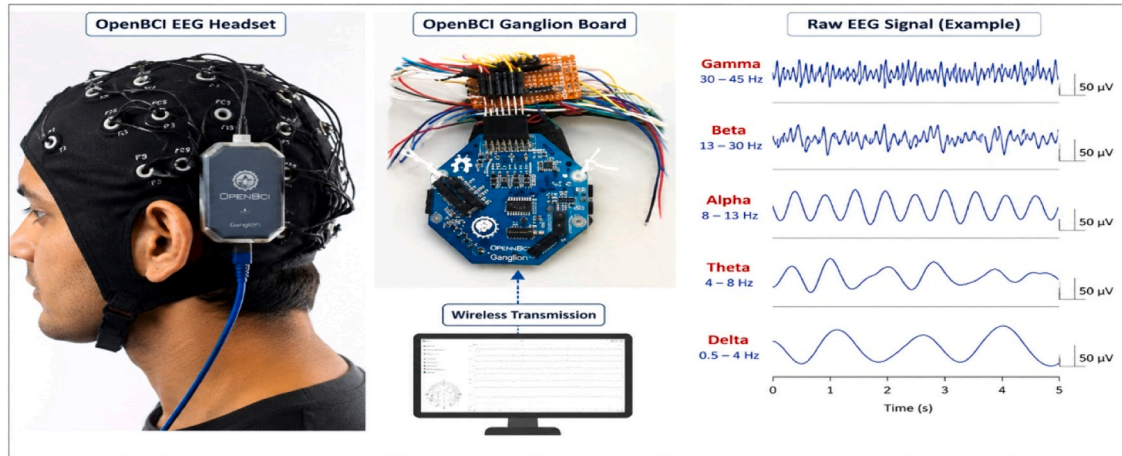


Fig. 2. Electrode positions for picking EEG signals using non-invasive methods and the Ganglion Board with EEG electrode setup for collecting EEG signals.

methods, such as attention and stress management. However, these benefits are mostly limited to offline training and validation. Also, no consideration is given to real-time signal processing methods such as buffering, sliding-window analysis. These are vital for continuous data processing. Moreover, the lack of computational performance metrics such as model size, execution time, and resource usage makes it difficult to evaluate practical deployment, particularly on embedded or edge devices [21]. As a result, the EEG cognition mapping significantly enhances the robustness and interpretability of the CNN framework. It needs to be backed by established experiments that evaluate dataset performance, such as latency measurements, hardware implementations, and online testing, to ensure its suitability for real-world applications.

The next section discusses the proposed methodology and the methods used to form the transformers and hybrid ML models. Following this, the results and discussion section is presented in the

simulated ML-based hybrid framework, and proceeds with the detailed analysis of optimization methods

Proposed methodology based on framework models developed

The proposed methodology section explains how the EEG signals are processed using the optimized DL-based framework. This section is further divided into signal acquisition and a standard operating procedure to assess publicly available data on cognitive skills.

EEG signal acquisition & preprocessing

The EEG data in this study were collected using the OpenBCI platform, and the dataset is publicly available [27]. In this dataset, to ensure consistent spatial coverage across participants, electrodes were placed on the scalp using the standardized 10–20 international placement

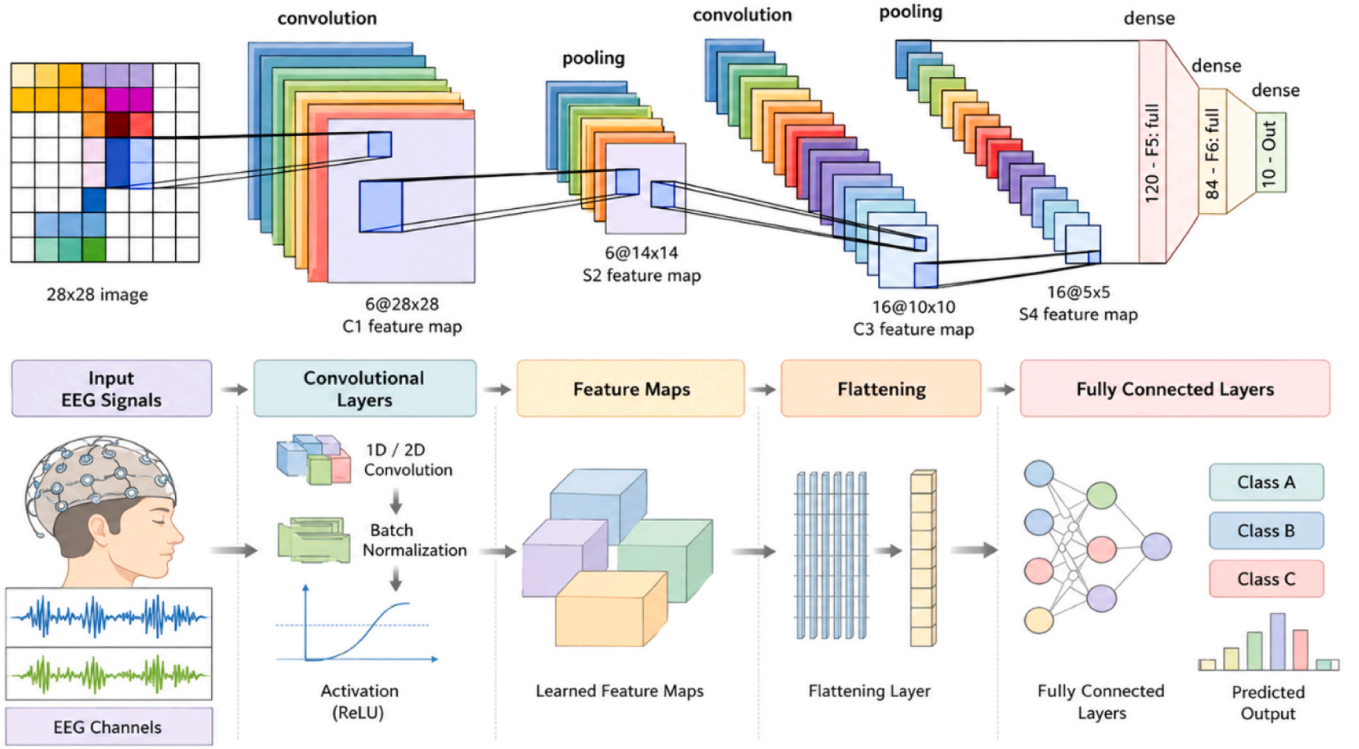


Fig. 3. CNN- Architectures for EEG signal Extraction & EEG- signal process via CNN blocks.

system. Throughout the recording, participants completed structured attention, memory, and problem-solving tasks [28]. Signals were sampled at 250 Hz, sufficient to resolve time- and frequency-domain characteristics of the underlying neural activity, including theta (4–8 Hz), alpha (8–13 Hz), and beta (13–30 Hz) rhythms associated with distinct cognitive states. Raw EEG recordings are susceptible to noise from muscular activity, electrical line interference, and electrode drift. A 1–50 Hz bandpass filter was applied to remove low-frequency drift and high-frequency noise while preserving the theta, alpha, and beta activity relevant to cognitive-state discrimination. (1) Artifact removal was performed using Independent Component Analysis (ICA), which decomposes the multichannel EEG signal into statistically independent components. Components corresponding to ocular and muscular artifacts were identified and removed prior to reconstruction.

This dataset serves as a controlled case study for evaluating the proposed CNN-LSTM-G2WO model under a single, homogeneous acquisition protocol and a well-defined cognitive-task paradigm [29]. It is complemented by a merged public EEG dataset, described in Section 2.1.2, which is used to further assess the model's generalization capability across diverse subjects, acquisition systems, and task domains. Accordingly, the two datasets provide a two-tier evaluation framework: (i) the single-source dataset enables controlled validation of the model under a consistent acquisition protocol and discrete cognitive-task conditions, whereas [30] (ii) the merged dataset evaluates robustness and generalizability across heterogeneous subjects, acquisition systems, and task domains [31]. Cognitive-state labels were derived from task-specific behavioral performance, with attention assessed using the Stroop task, working memory using the N-back task, and problem-solving ability using Raven's Progressive Matrices-style tasks. Accuracy and response time were used as behavioral indicators for state-label assignment according to the predefined labeling criteria. Raw EEG recordings are susceptible to baseline drift, physiological artifacts, and electrical interference. Therefore, a 1–50 Hz band-pass filter was applied to attenuate very-low-frequency drift and high-frequency components while retaining the principal EEG activity relevant to cognitive-state analysis, particularly the theta, alpha, and beta bands.

$$EEG_{filtered(t)} = F_{BP} (EEG_t) \quad (1)$$

Artifact removal was performed using ICA, which decomposes the multichannel EEG signal into statistically independent components. Components corresponding to ocular and muscular artifacts were identified and removed prior to reconstruction. Table 2(a) lists the EEG preprocessing pipeline and parameter settings adopted in the proposed framework.

The cleaned signal was then z-score-normalized across channels to ensure consistency across subjects and recording sessions, a prerequisite for stable deep learning (DL) training [32]. Extended Infomax ICA using EEGLAB v2024.x (MATLAB) was used to remove ocular, muscular, and cardiac artifacts. All independent components were automatically labelled with the ICLabel plugin, and components with an artifact probability > 0.80 (80%) for eye, muscle, line-noise, or cardiac activity were discarded prior to reconstruction of the cleaned EEG signals [33]. The ICA model assumes that the recorded EEG signals are linear mixtures of statistically independent neural and artifact sources. Across the pooled dataset, an average of $X = \frac{1}{N} \sum_{i=0}^{i=N} P_i$ % of decomposed independent components per recording $\max(P_i) - \min(P_i)$ reflect the varying channel counts of 14–122 across source datasets. These were classified as artefactual and removed, predominantly frontal ocular components and temporal muscular components, which are consistent with typical ICLabel classification patterns for scalp EEG. Recordings for which fewer than [N] clean components remained after rejection were excluded from further analysis to ensure sufficient neural signal was retained for classification

$$X = AS \quad (2)$$

Where, $X \in \mathbb{R}^{M \times T}$ denotes the observed EEG recordings from M electrodes, S represents the unknown independent source signals and A is the unknown mixing matrix. ICA estimates a un-mixing matrix $W = A^{-1}$. It is used to recover the independent components $\hat{S} = WX$. The Artifact components identified by IC-Label are removed, and then

Table 1

Literature review of EEG-based cognitive state classification approaches, with a comparison of methodologies, performance observations, strengths, limitations, and research gaps that motivate the proposed Hybrid PSO–G2WO framework.

Sl. No	Paper	Method Used	Observed	Advantages	Drawbacks	Research Gap
1	Automatic filter selection for EEG power	Optimization algorithm auto-selects FIR vs. IIR-Butterworth vs. IIR-Chebyshev filter for beta-band power; validated in EEGLAB v2022.1	4th-order IIR Chebyshev filter with 0.5% passband ripple outperformed FIR and Butterworth alternatives for signal clarity	Removes manual, subjective filter selection; clear EEGLAB benchmarking	Small sample (22 adults); single task (2D video); ripple trade-off not fully resolved	No downstream classification task, only signal power
2	AI — CNN replaces EEGLAB ICA	CNN estimates ICA-like components directly from raw epochs, calibration-free	F1-score 95.2%, precision 92.9%, recall 97.2%, accuracy 92.2%; processing time cut from 332.73 s to 4.86 s per subject ($\sim 68 \times$ faster)	Major speed gain over standard EEGLAB Infomax ICA; scalable for real-time/large datasets	CNN trained to mimic ICA, so inherits any bias in ICA itself; single lab/device tested (44 adults, one lecture)	No validation on portable and consumer-grade EEG
3	CNN gamma-band classification	CNN on ERSP time-frequency images to classify high/low gamma (30–40 Hz) activity	Accuracy 91.42%, precision 71.41%, recall 72.51%, exceeding SVM/random forest baselines of 70–85% accuracy	Automates EEGLAB-based manual analysis; outperforms classical ML	Precision much weaker than accuracy and many signals predicted as false positives; 41 adult students only	No hyperparameter optimization strategy used; no clinical validation
4	EEG – CNN	Compact CNN using depthwise separable convolutions. It is generalized across multiple BCI paradigms	Single compact CNN architecture designed to classify EEG across different BCI paradigms while minimizing parameter count	Lightweight, cross-paradigm generalization, widely adopted baseline	Not designed for cognitive-state tasks specifically; no bio-inspired hyperparameter optimization	Doesn't address CNN hyperparameter tuning via metaheuristics
5	Deep/ Shallow ConvNets	End-to-end deep and shallow ConvNets with batch norm, ELU, cropped training strategy	Deep ConvNets matched or exceeded filter-bank common spatial patterns (FBCSP), the previous standard, without handcrafted features	Strong end-to-end performance; visualization of learned spectral features	Adult motor-imagery data only; no swarm/metaheuristic optimization of architecture	No application to cognitive-skill contexts
6	Review on CNN and MLP for EEG	Systematic review of DL architectures (CNN, DBN, MLP, autoencoders) applied to EEG	Synthesizes hyperparameter choices and architecture trends across dozens of EEG-deep-learning studies	Comprehensive reference point; practical hyperparameter guidance	No new experimental data or child-specific findings	Explicitly notes inconsistent hyperparameter tuning practices across the field and directly motivates a systematic optimization approach 1
7	Hybrid PSO-GWO for CNN tuning (HPSGW)	Hybrid Particle Swarm + Grey Wolf Optimization auto-tunes CNN layers, kernels, batch size, epochs	HPSGW combines PSO's global search with GWO's ability to prevent PSO from getting trapped in local minima, applied to MNIST, CIFAR, and Indian Classical Dance classification	Direct methodological precedent for PSO-G2WO hybrid; proven convergence benefit	Applied to image datasets, not EEG; no biomedical/cognitive validation	Hybrid optimizer family to EEG-based cognitive-state CNNs
8	GWO-optimized CNN-Transformer for Motor Imagery EEG	CNN + Transformer hybrid with Grey Wolf Optimization for hyperparameter tuning and feature refinement	Reports improved motor-imagery EEG classification accuracy via GWO-tuned hybrid architecture over untuned baselines	Combines spatial (CNN) and temporal (Transformer) modelling with metaheuristic tuning	Adult motor-imagery paradigm only; single optimizer (GWO alone, not hybridized with PSO)	No periodic cognitive-state application
9	Scoping review of consumer-grade EEG devices (PMC)	Review of OpenBCI, Emotiv, Muse and similar low-cost EEG hardware across research domains	Notes that OpenBCI has been used mainly in engineering/robotics contexts, and that more experiments are needed to properly validate its applicability specifically to cognitive neuroscience	Identifies where consumer EEG hardware is and isn't yet validated	Review only, no new data; doesn't resolve the validation gap, only names it	OpenBCI without clinical-grade cross-validation

cleaned EEG signals are reconstructed as $\mathbf{X}_{\text{clean}} = \mathbf{A}\hat{\mathbf{S}}_{\text{clean}}$, where $\hat{\mathbf{S}}_{\text{clean}}$ contains only the retained neural components. The EEG recordings analyzed in this study were taken from a publicly available benchmark EEG database [27]. The preprocessing pipeline consisted of bandpass filtering (1–50 Hz), Extended Infomax ICA (EEGLAB v2024.x), IC-Label-based artifact rejection, and epoching into fixed [X]-second windows with [W]% overlap. The baseline was corrected to the [S] ms pre-stimulus/pre-task interval, and z-score normalization was applied offline to the released raw EEG data. We used only these widely accepted preprocessing techniques to reject physiological and environmental noise [34], enhance the Signal-to-Noise ratio (SNR) [35]. It retains the neural activity of interest for cognitive-state recognition. No synthetic EEG samples were produced, and no signal interpolation or data augmentation was performed during preprocessing. Hence, the novelty of the proposed work lies not primarily in data acquisition and pre-processing; but in the development of a framework for cognitive-state recognition based on standardized, reproducible

pre-processed EEG signals, enabling fair comparison with existing state-of-the-art approaches using the same benchmark dataset.

Table 2(a) summarizes the pre-processing workflow and parameter settings employed in the proposed framework. The adopted pipeline follows the widely accepted EEGLAB preprocessing methodology. It combines digital filtering, Extended Infomax ICA, and automated artifact identification with the ICLabel plugin. This will improve the quality of the EEG signal prior to feature extraction. Compared with conventional manual ICA rejection, this semi-automated approach significantly reduces operator dependency while maintaining reproducible artifact removal [36]. As discussed in Table 1, recent DL approaches have explored CNN-based ICA replacement and fully automated artifact detection techniques to further accelerate preprocessing for large-scale EEG datasets. Although these methods demonstrate promising scalability, the present work adopts the established EEGLAB-ICLabel workflow because of its robustness, interpretability, and extensive validation in cognitive EEG analysis. Hence, the integration models are

Table 2

Mapping of human emotion with the emotion, cognitive functions, EEG features, brain regions, and enhancement strategies.

Emotion	Associated Cognitive Functions	Dominant EEG Bands (Hz)	Brain Regions / EEG Sites	Cognitive State	Observed Performance Effect	Target for Improvement	Suggested Enhancement Strategy
Happiness [22]	Attention, creativity, working memory	Alpha (8–13), Gamma (30–100)	Prefrontal, ACC; F3, F4, O1, O2, Parietal	Relaxed and focused	Facilitates learning, memory, and problem solving	Sustained attention	Creative activities and cognitive training
Anger [23]	Decision-making, cognitive control	High Beta (20–30), Gamma (30–100)	Orbitofrontal, Amygdala, Frontal	High arousal	Promotes faster decisions but may increase impulsivity	Emotional control	Regulation and inhibition training
Fear [24]	Vigilance, risk assessment	Beta (13–30), Gamma (30–100)	Amygdala, Prefrontal; C3, C4	High alertness	Enhances threat detection but may reduce reasoning	Stress regulation	Analytical tasks and simulation training
Anxiety [25]	Monitoring, anticipation	Beta (13–30), Theta (4–8)	Frontal, Temporal, Amygdala	Mental overload	Can impair working memory	Focus maintenance	Mindfulness and attention training
Surprise [26]	Perception, alertness	Beta (13–30), Gamma (30–100)	Frontal, Parietal, Thalamus	Heightened attention	Enhances sensory processing	Controlled responses	Decision-making and awareness training
Sadness	Reflection, analysis	Theta (4–8), ↓Alpha (8–13)	Prefrontal, Limbic, DMN	Low engagement	Associated with slower processing	Motivation enhancement	Goal-based and collaborative learning

Table 2(a)

Definition of parameters processing stage with the proposed model and the progressing stage.

Preprocessing Stage	Method / Tool	Parameter(s)
Filtering	Band-pass FIR Filter	1–50 Hz passband
Line-noise Removal	Notch Filter	50 Hz
Re-referencing	Average Reference	Applied to all EEG channels
Bad-channel Detection	EEGLAB (pop_rejchan)	Kurtosis threshold = 5 SD
ICA Decomposition	Extended Infomax ICA (EEGLAB 2024.x)	Number of components = Number of EEG channels
Artifact Classification	ICLabel Plugin	Artifact probability > 0.80 (Eye, Muscle, Heart, Line Noise)
Components Removed	ICLabel + Manual Verification	Average 1–3 components per subject
Epoching	Sliding Window Segmentation	2 s window with 50% overlap
Post-ICA Artifact Rejection	Amplitude Threshold	±100 μ V
Normalization	Z-score Normalization	Per-channel normalization

required, as the original ICL has been affected by the filtering process

Data set description

In addition to the primary cognitive-task dataset described in Section 2.1, the present study further evaluates the generalization capacity of the proposed model beyond a single acquisition protocol and task paradigm. Also, this work incorporates a composite publicly available EEG dataset that combines multiple well-known EEG datasets with a large-scale, multi-session, multi-subject (MSS) visual EEG dataset to improve diversity, robustness, and generalization. The pooled dataset comprises data from DEAP (32 subjects), the eNTERFACE'06 EEG subset (16 subjects), the Imagined Emotion dataset (31 subjects), and a Neuro-Marketing EEG dataset (25 subjects), as well as a high-density MSS EEG dataset (32 subjects) recorded with a 122-channel system. The resulting merged dataset includes approximately 136 distinct subjects after eliminating overlaps and redundant entries. It represents a wide variety of emotional, cognitive, and visual processing tasks [37].

The integrated dataset spans multiple experimental paradigms, including emotion recognition, emotional stimulation via IAPS images and music videos, imagined emotional scenarios, consumer preference analysis, and visual object recognition with natural images. EEGs were recorded using systems with 14–122 channels, following the international 10–20 system to ensure consistency across datasets [38]. Some datasets also include multimodal signals based on fNIRS and facial video

recordings. But for consistency, only EEG signals were analyzed in this work. The MSS dataset provides 10,000 natural images across 20 classes from PASCAL and ImageNet, along with large-scale visual evoked response data under a 5 Hz rapid-presentation paradigm and 40,000 trials under a 1-second stimulus paradigm. The emotion-based datasets offer structured trials, such as 40 trials per subject based on 1-minute recordings as in DEAP, emotion-elicitation sessions in eNTERFACE'06, guided imagination trials, or preference-based trials in NeuroMarketing. All EEG signals were pre-processed uniformly, including bandpass filtering in the range of 0.5–45 Hz, artifact rejection, resampling as needed, and normalization. The signals were then segmented into fixed-length windows, and a common dataset was created. This merged dataset is used independently of the primary dataset in Section 2.1 to benchmark the proposed model's generalization performance across diverse Models. The model's features include emotional, cognitive, and visual-processing tasks, subjects, and recording systems, whereas the dataset in Section 2.1 provides a controlled, single-protocol case study for task-specific cognitive-state classification. Results from the two datasets are reported and interpreted separately throughout this study to avoid conflating protocol-specific findings with cross-domain generalization results [39].

The integrated dataset comprises about 60,000 EEG segments selected from the combined hybrid dataset, capturing inter-subject, inter-session, and inter-task variability. The dataset was split into three categories for model development: training (70% – 42,000 segments), validation (15% – 9000 segments), and testing (15% – 9000 segments). In addition, a 5-fold cross-validation ($K = 5$) approach was employed, with 80% (33,600 segments) of the training data used for each fold and 20% (8400 segments) for validation [40]. The merged dataset is designed to represent a wide range of EEG signals across task domains such as emotion, cognition, and vision, ensuring that the model generalizes across different task types. The 5-fold cross-validation also helps prevent overfitting and reduces variability across subjects and recording conditions [41]. Processed EEG signals were then provided to the proposed CNN-LSTM model, which was optimized using Grasshopper-Grey Wolf Optimization (G2WO). This unified data framework makes it highly appropriate for emotion recognition and brain-computer interface (BCI) applications based on EEG.

Feature extraction based on domain

The next step in the preprocessing is feature extraction. The signal generated by the brain during EEG is complex and contains a lot of information about the processes occurring in the brain [42]. In practice, however, raw signals are generally too noisy and high-dimensional to analyze directly. The extracted features provide a higher-level, more

informative representation of these raw signals that can be used by ML models. The feature extraction process is explained as follows

Time-domain features

Time-domain statistics summarise the amplitude distribution and dynamics of the signal. The mean and variance of each epoch are computed as:

$$\mu(n) = \frac{1}{N} \sum_{i=1}^N EEG_i(t) \quad (3)$$

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (EEG_i(t) - \mu(n))^2 \quad (4)$$

Skewness and kurtosis were additionally computed to capture asymmetry and tail behaviour of the amplitude distribution, following standard third- and fourth-moment definitions. To examine how signal power is distributed across brainwave bands, the Discrete Fourier Transform (DFT) was applied to each sampled epoch, converting the time-domain signal into the frequency domain:

$$X[f] = \sum_{n=0}^{N-1} EEG[n] e^{-j2\pi fn/N} \quad (5)$$

Where, $EEG[n]$ is the EEG signal in the time domain, $X[f]$ is the Fourier transform at the frequency f , N is the total number of samples. The Power Spectral Density (PSD) at each frequency bin is then obtained from the squared magnitude of the transform, normalized by the epoch length:

$$PSD[f] = \frac{1}{N} |X[f]|^2 \quad (6)$$

Where, $PSD[f]$ is the power spectral density at frequency f , $X[f]$ is the Fourier transform obtained from Eq. (6), N is the number of samples used for normalization. The first-order (mean) and second-order (variance) moments were defined in Eqs. (2)-(5). Two higher-order moments are added to capture properties that the mean and variance cannot capture: skewness, which quantifies asymmetry in the amplitude distribution, and kurtosis, which quantifies the heaviness of the distribution's tails relative to a Gaussian baseline.

$$\gamma_1 = \frac{1}{N} \sum_{i=1}^N (EEG_i(t) - \mu)^3 / \sigma^3 \quad (7)$$

$$\gamma_2 = \frac{1}{N} \sum_{i=1}^N (EEG_i(t) - \mu)^4 / \sigma^4 - 3 \quad (8)$$

The variations are not considered as EEG amplitude distributions are generally Gaussian. Making burst-like activity for alpha spindles with high-amplitude artifacts that survive ICA. It produces heavy tails that variance alone under-represents, and asymmetric cortical synchronization during effortful cognition can skew the distribution. Both are $O(N)$ per channel per epoch, so their computational cost is negligible relative to the CNN forward pass.

Skewness (γ_1) and kurtosis (γ_2) are statistics that describe the shape of the amplitude distribution of the EEG signal. Skewness is defined as the third standardized central moment, which measures the signal's deviation from the mean. Negative skewness indicates a longer left tail, while positive skewness refers to the distribution having a longer right tail. Kurtosis is calculated as the fourth standardized central moment, whereas the fourth power amplifies deviations from the mean in both directions. This provides a measure of the distribution's peak and tail heaviness [43]. The subtraction of 3 in the kurtosis equation ensures that the kurtosis of a Gaussian (normal) distribution has the reference value of zero. If a distribution has positive excess kurtosis, it has heavier tails than the normal distribution; if it has negative excess kurtosis, it has

flatter tails than the normal distribution. These higher-order statistical moments have been shown to effectively capture the non-Gaussian characteristics of neural activity and can serve as complementary features alongside time- and frequency-domain features to enhance the discriminative performance of the cognitive state classification model in EEG signal analysis [44].

Frequency-domain features

Apply the DFT directly to a finite EEG segment as per 6, 7, and 8. Doing so without first applying a taper like Hamming window. It treats the segment as if it were exactly periodic at the boundary; the resulting discontinuity creates spectral leakage across frequency bins and biases downstream band-power estimates. The corrected form applies an explicit window function prior to transformation:

$$X_w[f] = \sum_{n=0}^{N-1} w(n) EEG(n) e^{-j2\pi fn/N} \quad (9)$$

Raw power spectral density (Eq. 9) is sensitive to nuisance factors that vary across subjects and sessions, such as electrode-skin impedance, cap placement, and skull conductivity, all of which scale absolute EEG amplitude without being related to cognitive state. To control for this, band power is expressed relative to total power in the epoch rather than in absolute units:

$$RBP_{band} = \frac{\sum_{f \in band} PSD[f]}{\sum_{f=0}^{N/2} \sum_{f \in band} PSD[f]} \quad (10)$$

Four canonical frequency bands are used. (i) theta (4–8 Hz), (ii) alpha (8–13 Hz), and (iii) beta (13–30 Hz), and (iv) gamma (30–45 Hz), which is required here as the gamma-band features reported in the extended feature table. Frequency and time resolution trade off against one another through the Gabor-Heisenberg uncertainty relation, which lower-bounds how precisely a signal's timing and frequency content can be jointly resolved from a windowed transform:

$$\Delta t \cdot \Delta f \geq 1/4\pi \quad (11)$$

A short epoch gives coarse frequency resolution ($\Delta f = fs/N$) but tracks fast-changing cognitive events more precisely than a long epoch does. This trade-off is the direct motivation for the wavelet-based analysis in Section 2.2.3, which does not fix a single resolution but instead adapts it across the frequency domain.

Wavelet transform

Because cognitive-state transitions are non-stationary, a Continuous Wavelet Transform (CWT) was applied to obtain a joint time-frequency representation capable of capturing transient shifts in brain activity:

$$CWT(a, \tau) = \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} EEG(t) \psi^* \left(\frac{t - \tau}{a} \right) dt \quad (12)$$

where ψ is the Morlet mother wavelet, ψ^* its complex conjugate, a the dilation (scale) parameter controlling frequency resolution, and τ The translation parameter controlling temporal localization [45]. The CWT gives fine time resolution at high frequencies and fine frequency resolution at low frequencies, unlike the fixed-window DFT above. A complex Morlet wavelet is the standard choice for EEG rather than a discontinuous basis such as Haar. Because, its Gaussian-windowed sinusoidal form achieves close to the theoretical minimum of the uncertainty bound in Eq. (13) and avoids the ringing artifacts a non-smooth mother wavelet introduces into the resulting time-frequency map

$$E(a) = \sum_{\tau} |CWT(a, \tau)|^2 \quad (13)$$

Rather than passing the full complex-valued CWT surface to the classifier, a scalar wavelet energy is computed at each scale by summing

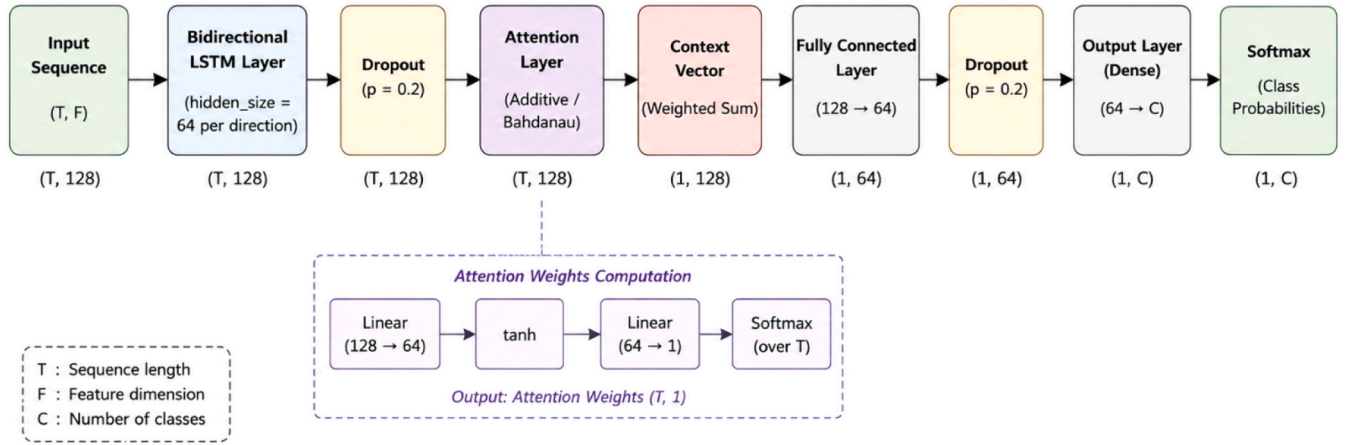
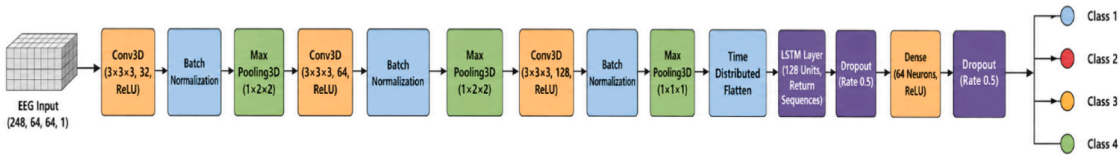


Fig. 4(a). BILSTM-Attention classification Head Architectures for EEG signal Extraction.

Table 3

Parameters of hybrid dataset for the proposed DL framework.

Category	Parameter	Details
Dataset Information	Dataset Type	Merged Public EEG Dataset
	Source Datasets	MSS EEG, DEAP, eINTERFACE'06, Imagined Emotion, Neuro-Marketing
	Total Subjects	~136
	EEG Channels	14 – 122 Channels
	Electrode Standard	International 10–20 System
Dataset Composition	Experimental Domains	Emotion, Visual Processing, Cognitive Tasks, Neuromarketing
	MSS Trials	~800,000 (5 Hz) + 40,000 (1 s paradigm)
	DEAP Trials	40 trials/subject (1-min videos)
	eINTERFACE'06	IAPS-based emotion elicitation
Preprocessing	Imagined Emotion	Scenario-based emotional recall
	Filtering	Bandpass (0.5–45 Hz)
	Artifact Removal	Eye-blink & motion artifacts removed
	Normalization	Applied to all signals
Final Dataset	Segmentation	Fixed-length windowing
	Total Segments Used	60,000
Data Splitting	Training Set (70%)	42,000
	Validation Set (15%)	9000
	Testing Set (15%)	9000
Cross-Validation	Method	5-Fold (K = 5)
	Training per Fold (80%)	33,600
	Validation per Fold (20%)	8400
Reasoning	Dataset Merging	Improves diversity and generalization
	Segmentation	Enables DL compatibility
	70–15–15 Split	Standard unbiased evaluation
	5-Fold CV	Reduces overfitting, improves robustness



squared coefficient magnitudes across time [46].

Signal-complexity features

The extended feature table reports Hjorth parameters and differential entropy for each cognitive skill, but neither was defined in the original Methodology. Both are standard and low-cost complements to the spectral features are formalized here. Hjorth Activity is simply the signal variance of Eq. (14), reported again for completeness:

$$\text{Activity} = \sigma^2 = \frac{1}{N} \sum_{i=1}^N (\text{EEG}_i(t) - \mu)^2 \quad (14)$$

Hjorth Mobility approximates the mean frequency of the signal directly in the time domain [47], using the variance of the first

derivative of the signal relative to the variance of the signal itself:

$$\text{Mobility} = \sqrt{\text{Var}(\text{EEG}'(t) / \text{Var}(\text{EEG}(t)))} \quad (15)$$

Hjorth Complexity measures how much the signal's waveform deviates from a pure sinusoid, computed as the ratio of the mobility of the derivative signal to the mobility of the original signal:

$$\text{Complexity} = \text{Mobility}(\text{EEG}'(t)) / \text{Mobility}(\text{EEG}(t)) \quad (16)$$

Differential entropy generalizes Shannon entropy to continuous-valued signals and, for an EEG segment approximated as Gaussian-distributed [48]. It reduces to a closed form that increases with signal variance. This is the standard formulation used in the EEG deep learning, defined as,

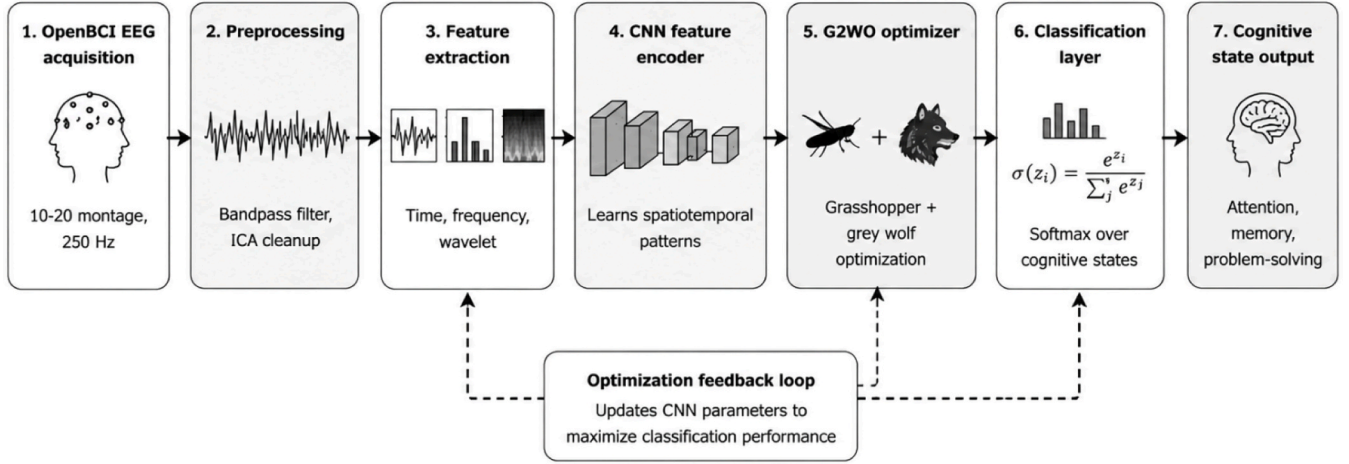


Fig. 4(b). Block diagram of the proposed EEG cognitive state classification framework integrating CNN-based feature learning with Hybrid Grasshopper–Grey Wolf Optimization (G2WO) for automatic hyperparameter optimization.

$$DE = \frac{1}{2} \log(2\pi\sigma^2) \quad (17)$$

Feature vector assembly and normalization

The time-domain, frequency-domain, wavelet, and complexity features above are concatenated per epoch into a single feature vector: $F = [\mu, \sigma^2, \gamma_1, \gamma_2, RBP_{\theta,\alpha,\beta,\gamma}, E(a_1 \dots a_K), \text{Activity}, \text{Mobility},$

Complexity, $DE] \in \mathbb{R}^d$; Because these features have very different natural scales. As the variances are used in μV^2 rather than normalized relative band power in $[0,1]$. For instance, each dimension is z-score normalized using statistics computed on the training partition only, then

applied unchanged to the validation and test partitions to avoid information leakage across the split $F_{Norm} = F - \mu_f / \sigma_f$.

CNN feature encoder

The convolutional stage of the network learns local spatiotemporal patterns from the pre-processed, feature-augmented EEG epochs. For an input representation X , kernel weights W_1 and bias b_1 , the layer output is

$$CNN_{out} = f((W_1 * X) + b_1) \quad (18)$$

with the convolution operation itself defined explicitly as

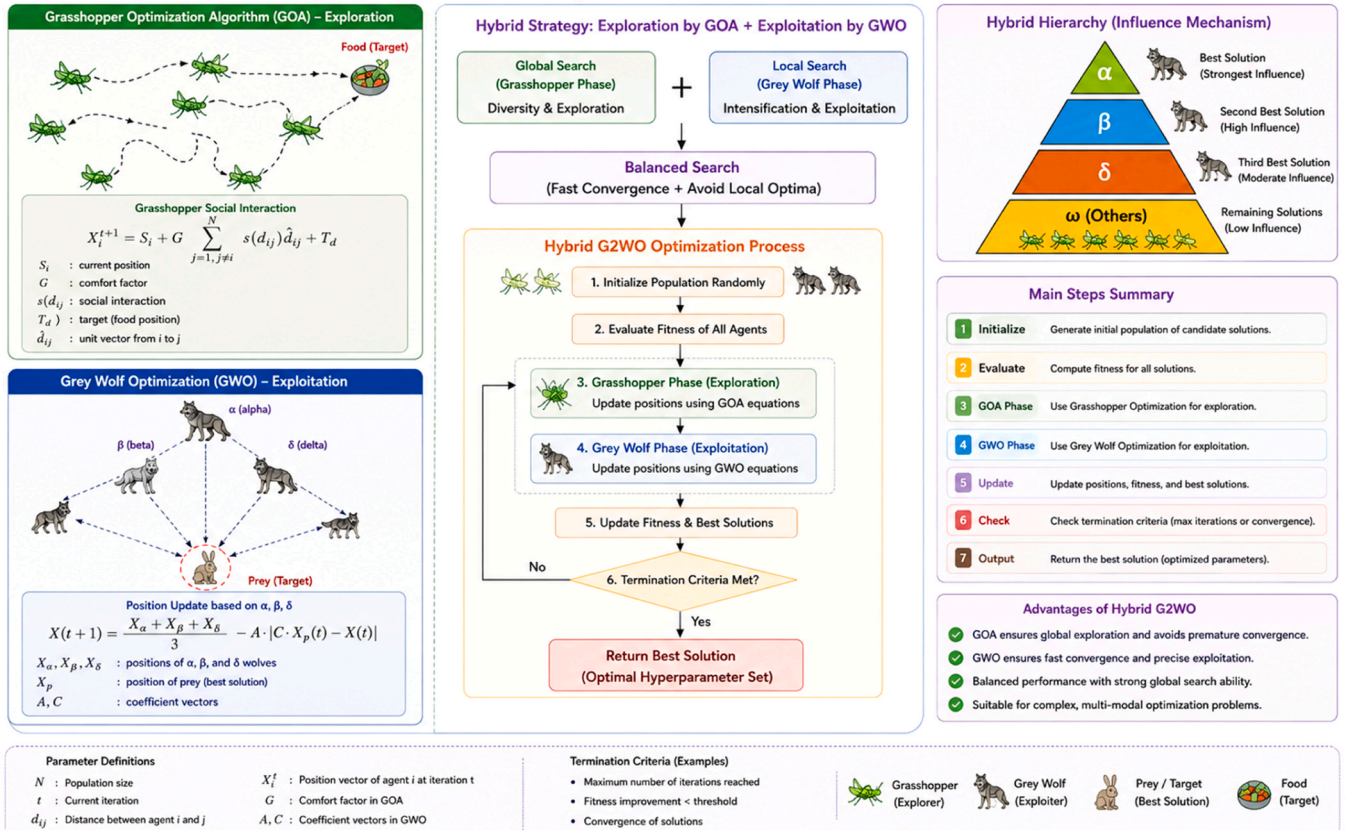


Fig. 4(C). Illustration of Hybrid Grasshopper–Grey Wolf Optimization (G2WO) for automatic hyperparameter optimization.

$$(\mathbf{W}_1 * \mathbf{X})[t] = \sum_{i=0}^N \mathbf{W}_1[i] \mathbf{X}[t-i] \quad (19)$$

where f is a nonlinear activation based on ReLU, and the summation runs over the kernel's receptive field. Equivalently, the band-pass-filtering step of Section 2.2 can be expressed in the frequency domain as

$$EEG_{filtered}(t) = \int_{f_{low}}^{f_{high}} EEG(f) H(f) df \quad (20)$$

and the general one-dimensional convolution applied at each time step of the CNN encoder is

$$\mathbf{y}(t) = f\left(\sum_{i=0}^{m-1} \mathbf{w}_i \mathbf{x}(t-i) + \mathbf{b}\right) \quad (21)$$

Max-pooling then reduces the temporal resolution of each feature map to control overfitting and computation:

$$\mathbf{y}_{pooled}(t) = \max\{\mathbf{y}(t), \mathbf{y}(t+1), \dots, \mathbf{y}(t+k-1)\} \quad (22)$$

Where, $\mathbf{X}$ denotes the input EEG feature representation, $\mathbf{W}_1$ are the convolution kernel weights, $\mathbf{b}_1$ is Bias of the convolution layer. $*$: Convolution operation. $f(\cdot)$: ReLU activation function. $EEG(f)$: Frequency-domain EEG signal. $H(f)$: Band-pass filter transfer function. f_{low}, f_{high} : Lower and upper cutoff frequencies. $\mathbf{w}_i$: Convolution kernel coefficients. Also, $\mathbf{x}(t)$: Input signal at time t . m : Kernel length. k : Pooling window size. $\mathbf{y}(t)$: Convolution output. $\mathbf{y}_{pooled}(t)$: Max-pooled feature output.

BiLSTM-attention classification head

The CNN-encoded feature sequence is passed to a Bidirectional Long Short-Term Memory (BiLSTM) layer [49]. This model has forward and backward temporal dependencies across the epoch, followed by an attention layer that assigns learned importance weights to each time step. This allows the classifier to emphasize the most cognitively informative segments of the signal. The attention-weighted representation is

flattened and passed through a fully connected classification layer:

$$\mathbf{y}_{out} = \sigma(\mathbf{W}_2 \mathbf{h} + \mathbf{b}_2) \quad (23)$$

where $\mathbf{W}_2$ and $\mathbf{b}_2$ are the classification layer's weights and bias, and sigma is softmax for the three-way attention, memory, and problem-solving output. The network is trained by minimizing the categorical cross-entropy loss via backpropagation:

$$L(\mathbf{y}, \hat{\mathbf{y}}) = - \sum_{i=1}^N \mathbf{y}_i \log(\hat{\mathbf{y}}_i) \quad (29)$$

Performance is reported using accuracy together with precision, recall, and F1-score, since accuracy alone is insufficient under class imbalance:

$$Accuracy = \frac{1}{N} \sum_{i=1}^N (\hat{\mathbf{y}}_i - \mathbf{y}_i) \quad (24)$$

$$Precision = \frac{TP}{TP + FP} \quad Recall = \frac{TP}{TP + FN} \quad F1 = \frac{2 \cdot Precision \cdot Recall}{Precision + Recall} \quad (25)$$

As illustrated in Fig 4(a), the BiLSTM-Attention classification head is proposed to model the forward and backward nature of temporal dependencies in the extracted EEG feature sequences, thereby improving the recognition of EEG cognitive states [50]. The BiLSTM layer captures long-range contextual information, and the attention mechanism assigns greater weights to the most informative temporal features while suppressing irrelevant information. This context vector is then fed into fully connected and Softmax layers to produce the final cognitive state prediction, which boosts the classification accuracy and robustness [51].

Hybrid grasshopper-grey wolf optimizer (G2WO)

CNN and BiLSTM have various hyperparameters such as learning rate, number of filters, kernel size, number of LSTM units, and dropout rate. These models are tuned using a proposed Hybrid Grasshopper-Grey Wolf Optimizer (G2WO), which combines the global exploration

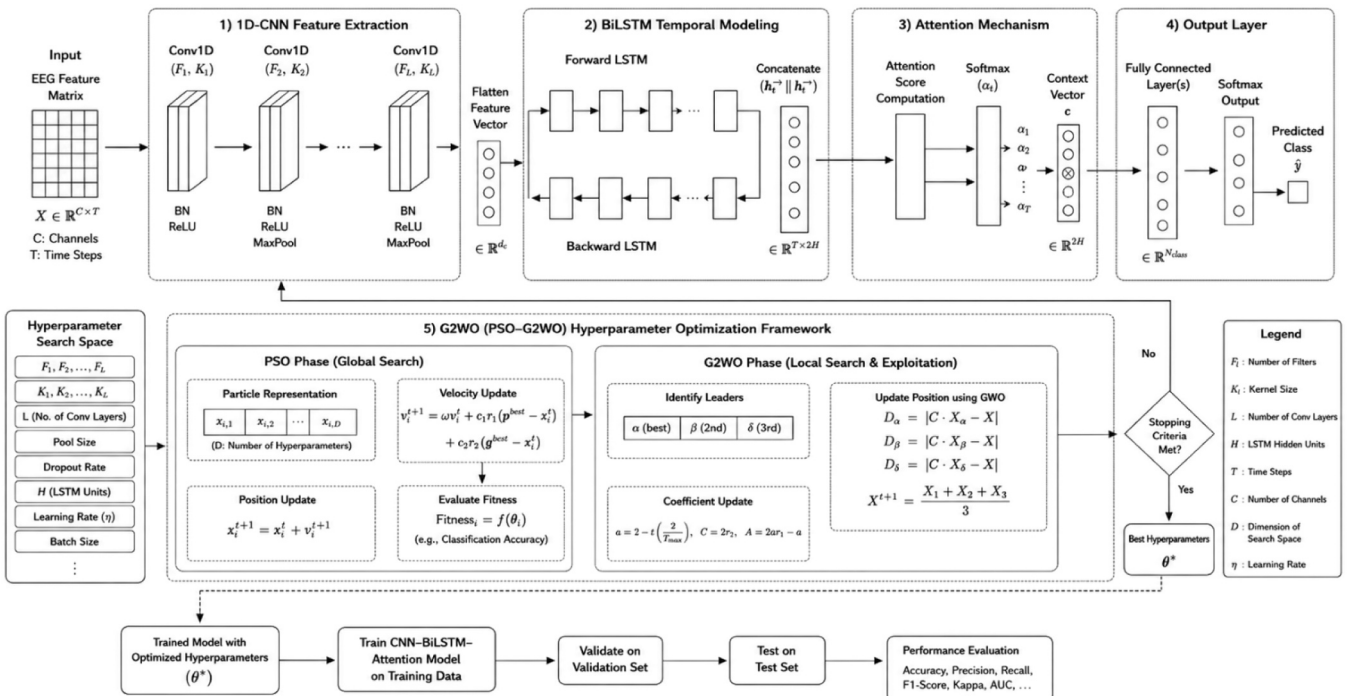


Fig. 5. End-to-end architecture of the proposed Hybrid PSO-G2WO optimized CNN-BiLSTM-Attention framework, comprising CNN-based feature extraction, BiLSTM temporal learning, attention-guided feature aggregation, hybrid hyperparameter optimization, and final EEG cognitive state classification.

behavior of the GOA with the local exploitation behavior of the Grey Wolf Optimizer (GWO). Table 3 lists the parameters used for this work. This method balancing the both optimization through an adaptive weighting schedule. Because the CNN and BiLSTM hyperparameters are not in the space of differentiable weights. They can be tuned neither by gradient descent nor by manual or grid search [52]. As this would sample only a coarse, human-chosen set of configurations with no guarantee of proximity to the optimal configuration. This creates a strong incentive for a black-box metaheuristic search, but by the No Free Lunch theorem, there is no single search that works best on all landscapes, and there's a failure mode that is known for GWO. In GWO, wolves converge quickly to the current best solutions in the search space and risk settling on a mediocre local optimum before they've sufficiently explored the space. There's a failure mode known for GOA where the swarm covers the search space broadly using social repulsion. But, it lacks any mechanism to speed up convergence once a promising space is reached [53].

To prevent this, the proposed Hybrid G2WO uses a time-decaying weight $\lambda(t)$ to direct the search through two distinct and different behaviours. A GOA-dominant search phase for exploring the search space and defining its structure, and a GWO-dominant search phase for converging the population on a weighted combination of the three best-so-far leaders (α, β, δ). So that GWO does not explore any territory that is already ruled out by the GOA phase. It leads to higher accuracy than either of its constituent optimizers without a proportional increase in search cost, as shown in the execution-time column of Table 7. First, given the limited number of trials per person, the variance in validation accuracy across folds for a given configuration is non-trivial with consumer-grade EEG [54]. It means that the optimum search region is flat and broad, and thus convergence from GWO is most threatened by premature convergence. But its easily attained from GOA with the lowest number of search costs and this is the practical reason for its use here. Also, other exhaustive strategies such as Neural Architecture Search (NAS) require significantly higher numbers of search costs and thus are not competitive in this regard [55]. Though it should be explicitly noted that unlike Bayesian Optimization (BO), this hybrid does not rely on a model of the objective. Therefore, this hybrid model does not have a formal convergence guarantee, only the empirical convergence seen across the iterations reported in Table 7.

Exploration phase through GOA

Each candidate solution's position is updated based on pairwise social interaction forces with all other candidates:

$$\mathbf{X}_i^{GOA} = \mathbf{c} \sum_{j=1, j \neq i}^N s(d_{ij}) \frac{\mathbf{X}_j - \mathbf{X}_i}{d_{ij}} \quad (26)$$

where d_{ij} is the distance between candidates i and j , and $s(\cdot)$ is the social force function:

$$s(r) = f e^{-r/l} - e^{-r} \quad (27)$$

with f and l denoting the attraction intensity and attractive length scale respectively which are set via preliminary tuning and repad value. The coefficient c decreases across iterations to shift the search from exploration to exploitation:

$$c(t) = c_{max} - t \frac{c_{max} - c_{min}}{T} \quad (28)$$

Exploitation phase through GWO

The three best candidates found so far are designated alpha, beta, and delta; all other candidates update their position toward these leaders. First, the distance of the current candidate to each leader is computed:

$$D_\alpha = |C_1 \mathbf{X}_\alpha - \mathbf{X}_i|, D_\beta = |C_2 \mathbf{X}_\beta - \mathbf{X}_i|, D_\delta = |C_3 \mathbf{X}_\delta - \mathbf{X}_i| \quad (29)$$

Table 3(a)

Relation and mapping between human emotions and cognitive skills.

Emotional State	Main Cognitive Skills	Effect on Performance	Main Brain Regions
Fear	Alertness, threat detection	Helps identify danger quickly but may reduce logical thinking	Amygdala, Sensory Cortex
Happiness	Attention, creativity, memory	Improves learning, creativity, and problem-solving	Prefrontal Cortex, Anterior Cingulate Cortex
Sadness	Reflection, careful thinking	Encourages detailed analysis but slows decision-making	Prefrontal Cortex, Default Mode Network
Anger	Decision-making, focus	Increases quick decisions but may lead to impulsive actions	Orbitofrontal Cortex, Amygdala
Surprise	Attention, perception	Improves awareness and response to new information	Parietal Cortex, Thalamus
Anxiety	Monitoring, anticipation	Increases alertness for a short time but can reduce memory if prolonged	Amygdala, Prefrontal Cortex

Each leader then proposes a new candidate position:

$$\mathbf{X}_1 = \mathbf{X}_\alpha - A_1 D_\alpha \mathbf{X}_2 = \mathbf{X}_\beta - A_2 D_\beta \mathbf{X}_3 = \mathbf{X}_\delta - A_3 D_\delta \quad (30)$$

and the GWO-side position estimate is the average of the three proposals:

$$\mathbf{X}_i^{GWO}(t+1) = \frac{\mathbf{X}_1 + \mathbf{X}_2 + \mathbf{X}_3}{3} \quad (31)$$

Hybrid update rule is done at each iteration. The final position update linearly blends the GOA for exploration and GWO for exploitation, which are estimated using a time-decaying weight $\lambda(t)$. At each iteration, the final position update linearly blends and updated using a time-decaying weight $\lambda(t)$:

$$\mathbf{X}_i(t+1) = \lambda(t) \mathbf{X}_i^{GOA} + (1 - \lambda(t)) \mathbf{X}_i^{GWO}, \lambda(t) = 1 - \frac{t}{T} \quad (32)$$

The GWO coefficients A and C , and the decreasing parameter a , are computed in the standard way:

$$A = 2ar_1 - aC = 2r_2a : 2 \rightarrow 0 \text{ over } T \text{ iterations} \quad (33)$$

Fitness function: Each candidate hyperparameter configuration $\mathbf{x}_i$ is evaluated by training the CNN + BiLSTM + Attention model and measuring 1 minus its validation accuracy:

$$F(\mathbf{x}_i) = 1 - \text{Accuracy}_{val}(\text{Model}(\mathbf{x}_i)) \quad (34)$$

The proposed CNN hyperparameter optimization using Hybrid G2WO is shown in Fig 4(b). The algorithm starts by randomly setting up candidate solutions, and evaluating their fitness. The best three candidates are selected as the α, β , and δ leaders. The Grasshopper Optimization (GO) stage is responsible for the global exploration, while the GWO stage is responsible for the exploitation of good region around leader solutions. The two search strategies are used in parallel to update the set of candidate positions in an iterative fashion until convergence. The optimal hyperparameters obtained from the hybrid optimizer are then used to train the CNN model, resulting in improved classification accuracy and convergence speed.

The proposed method, G2WO, aims for a more balanced search process. Both GOA and GWO comprise mechanisms of social interaction and improvement of local exploitation wolf hunting strategy [56]. GOA introduces global exploration through social interaction mechanisms, and GWO introduces the improvement of local exploitation through the leadership hunting strategy. The proposed method is based on exploration activities for the early stage and exploitation for the later stage under an α - β - δ hierarchy. This method has the fastest convergence,

highest accuracy, and stable optimization ability with the adaptive switching mechanism. The whole structure is based on DL and hybrid optimization and introduces a cognitive skill evaluation system based on EEG. Pre-processing of data is carried out before the use of multi-channel EEG signals to eliminate noise, artifacts, normalization and segmentation of the data to obtain clean data [57].

A cross-validation technique was used to assess robustness. The data set was divided into training and test sets more than 5 times. The processed signals are then fed into a hybrid DL model consisting of a Convolutional Neural Network (CNN) and a Bidirectional Long Short-Term Memory (BiLSTM) network, which extract spatial and frequency features and temporal dependencies, respectively. Another improvement to the model is the use of an attention mechanism, which highlights the most relevant features from the EEG. The main hyperparameters are learning rate, number of filters, and number of LSTM units which are tuned with the G2WO algorithm and validated using loss function [58]. In each fold, the optimizer gets closer to the optimal configuration by improving the candidate solutions. All the optimized features are utilized in the classification task to extract the cognitive states, and then all the cognitive states are converted to cognitive skills like stress, memory, and attention. Also, this framework offers an efficient and scalable way to realise EEG-based cognitive assessment, with improved performance and generalisation abilities. The hybrid model is exhibited in Fig 4(c).

Algorithmic representation of the proposed Hybrid G2WO optimization procedure used for CNN hyperparameter tuning. G2WO Algorithm	Adaptive Hybrid Update Rule - Hybrid Grasshopper-Grey Wolf Optimization (G2WO)
1. Initialize N candidate solutions X_i randomly within the hyperparameter bounds.	Process flow
2. Evaluate $F(x_i)$ for all candidates; identify alpha, beta, delta as the three best.	Step 1. Initialization Initialize a population of N search agents
3. For $t = 1$ to T :	$X_i = [x_{i1}, x_{i2}, \dots, x_{id}], i = 1, 2, \dots, N$ Evaluate the fitness of every agent $F_i = f(X_i)$
a. Compute $c(t)$ and each candidate's GOA position	Step 2. Grasshopper Exploration Phase
b. Compute the alpha (α), beta (β), and delta (δ) and the GWO position	The position update of the GOA is
c. Compute $\lambda(t)$ and update each candidate's position	$X_i^{t+1} = c \left(\sum_{j=1, j \neq i}^N s(d_{ij}) \hat{d}_{ij} \right) + TW$ Where,
d. Enforce hyperparameter boundary constraints.	$d_{ij} = X_j - X_i $ and $\hat{d}_{ij} = \frac{X_j - X_i}{d_{ij}}$ and the Social
e. Re-evaluate fitness; update alpha, beta, delta.	interaction function $s(r) = fe^{-r/l} - e^{-r}$ with the
4. Return alpha as the optimal hyperparameter configuration.	Comfort coefficient $c = c_{max} - \frac{t}{T_{max}}(c_{max} - c_{min})$.
	This phase performs global exploration.
	Step 3. Grey Wolf Exploitation Phase
	Identify the three best solutions
	$X_\alpha, X_\beta, X_\delta$ Coefficient vectors, $A = 2ar_1 - a$ and $C = 2r_2$. Similarly, the distance vector, Distance vectors are $D_\alpha = CX_\alpha - X $, $D_\beta = CX_\beta - X $, $D_\delta = CX_\delta - X $. It will help to Updated candidate positions $X_1 = X_\alpha - AD_\alpha$ and $X_2 = X_\beta - AD_\beta$ and $X_3 = X_\delta - AD_\delta$. Also, Final Grey Wolf position
	$X_{GWO} = \frac{X_1 + X_2 + X_3}{3}$ This phase performs local exploitation.
	Step 4. Hybrid Grasshopper-Grey Wolf Update
	The proposed hybrid position combines GOA exploration with GWO exploitation.
	$X_i^{t+1} = \lambda X_{GOA} + (1 - \lambda) X_{GWO}$ Where, $0 \leq \lambda \leq 1$ or equivalently, $X_i^{t+1} = \lambda \left[c \left(\sum_{j=1, j \neq i}^N s(d_{ij}) \hat{d}_{ij} \right) \right] T + (1 - \lambda) \left(\frac{X_1 + X_2 + X_3}{3} \right)$
	Step 5. Fitness Evaluation
	Evaluate the objective function
	$Fitness = f(X_i)$ For EEG classification, $Fitness = Accuracy$ and $Fitness = 1 - Loss$
	Step 6. Best Solution Update
	If $Fitness(X_i) > Fitness(X_{best})$ then $X_{best} = X_i$
	Step 7. Termination
	Repeat until $t = T_{max}$ or $ Fitness^{t+1} - Fitness^t < \epsilon$
	Final Hybrid G2WO Equation
	$X_i^{t+1} = \lambda \left[c \left(\sum_{j=1, j \neq i}^N s(d_{ij}) \hat{d}_{ij} \right) \right] T + (1 - \lambda) \left(\frac{X_\alpha - AD_\alpha + (X_\beta - AD_\beta) + (X_\delta - AD_\delta)}{3} \right)$

G2WO Algorithm Steps for $t = 1$ to T

(continued on next column)

(continued)

1. Randomly initialize N candidate hyperparameter vectors within the predefined search space.
2. Evaluate the fitness of each candidate using the objective function.
3. Identify the three best candidates as the alpha (α), beta (β), and delta (δ) wolves.
4. For each iteration $t = 1, \dots, T$:
Compute the GOA exploration position
Compute the GWO exploitation position.
Calculate the adaptive weighting factor $\lambda(t)$.
Update each candidate solution using the hybrid update rule.
Enforce the predefined parameter boundaries.
Re-evaluate the fitness and update the α , β , and δ wolves.
5. Return the α wolf as the optimal hyperparameter configuration.

End to end architecture of CNN-BiLSTM and Attention Framework with PSO-G2WO optimization

Fig. 5 illustrates the proposed CNN-BiLSTM-Attention framework, integrated with the Hybrid PSO-G2WO Optimization algorithm for automated hyperparameter tuning in EEG-based cognitive state classification. This framework is designed to capture the spatial, temporal, and contextual characteristics of EEG signals while optimising the DL model for enhanced classification performance. Initially, the extracted EEG feature matrix is fed into a sequence of one-dimensional convolutional (1D-CNN) layers comprising convolution, batch normalisation, ReLU activation, and max-pooling operations [59]. These layers automatically learn hierarchical feature representations by extracting discriminative spatial and spectral information from the EEG signals while progressively reducing feature dimensionality. The resulting feature maps preserve the most informative characteristics required for subsequent temporal analysis. The learned feature representations are then forwarded to a BiLSTM network, which captures temporal dependencies in both forward and backward directions.

This bidirectional learning strategy enables the network to exploit contextual information across the entire EEG sequence, yielding a richer representation of the underlying cognitive patterns. To further enhance feature discrimination, an attention mechanism is applied after the BiLSTM layer. The attention module adaptively assigns higher weights to the most informative temporal features while suppressing less relevant information, allowing the model to focus on EEG segments that contribute most to cognitive state recognition. The resulting context vector is then passed through fully connected layers, followed by a Softmax classifier, to predict the corresponding cognitive state [60].

To further improve the network's learning capability, the model's hyperparameters are automatically optimised using the proposed Hybrid PSO-G2WO algorithm. The Optimization process begins with Particle Swarm Optimization (PSO), which efficiently explores the hyperparameter search space by updating particle positions based on both individual and global best experiences. Promising solutions are identified in this stage, then refined using the proposed G2WO algorithm. In the G2WO phase, the GO component promotes extensive exploration of the search space, whereas the GWO component performs local exploitation by guiding the search through the hierarchical leadership of the α , β , and δ wolves [61]. This complementary search strategy effectively balances exploration and exploitation, reducing the likelihood of premature convergence while accelerating the discovery of optimal hyperparameter combinations. During each iteration of Optimization, candidate hyperparameter configurations are evaluated using network classification performance, with validation accuracy and validation loss as the objective function. The Optimization continues until the predefined stopping criterion is met, after which the optimal hyperparameter set is selected to train the final Attention model. By integrating deep feature learning with the proposed Hybrid PSO-G2WO Optimization strategy, the framework achieves faster convergence, improved feature representation, enhanced generalisation, reduced overfitting, and superior EEG cognitive state classification performance.

Table 4

Layer specification for the proposed Hybrid PSO-G2WO framework architecture.

Stage	Network Operation	Configuration	Output Dimension	Trainable Parameters	Purpose / Remarks
1	Conv3D_1	32 filters, Kernel: $3 \times 3 \times 3$, ReLU	(None, 248, 62, 62, 32)	2624	Initial spatio-temporal feature extraction using valid padding
2	Batch Normalization	Momentum = 0.99	(None, 248, 62, 62, 32)	128	Stabilizes feature distributions during training
3	MaxPooling3D	Pool Size: $1 \times 2 \times 2$	(None, 248, 31, 31, 32)	0	Reduces spatial dimensions while preserving temporal information
4	Conv3D_2	64 filters, Kernel: $3 \times 3 \times 3$, ReLU	(None, 246, 29, 29, 64)	55,360	Learns intermediate-level spatial features
5	Batch Normalization	Feature normalization	(None, 246, 29, 29, 64)	256	Improves convergence and network stability
6	MaxPooling3D	Pool Size: $1 \times 2 \times 2$	(None, 246, 14, 14, 64)	0	Performs spatial down sampling
7	Conv3D_3	128 filters, Kernel: $3 \times 3 \times 3$, ReLU	(None, 244, 12, 12, 128)	221,312	Extracts high-level discriminative representations
8	Batch Normalization	Batch stabilization	(None, 244, 12, 12, 128)	512	Normalizes deep feature activations
9	MaxPooling3D	Pool Size: $1 \times 2 \times 2$	(None, 244, 6, 6, 128)	0	Further reduces feature-map dimensions
10	Global Average Pooling 3D	Feature aggregation	(None, 128)	0	Converts volumetric feature maps into a compact descriptor
11	Dropout	Rate = 0.5	(None, 128)	0	Prevents overfitting by randomly deactivating neurons
12	Deep Feature Embedding	128-dimensional representation	(None, 128)	0	Generates latent feature vectors for optimization
13	Hybrid PSO-GWO-GOA Optimization	Population-based optimization with adaptive fitness search	(None, Optimal Feature Set)	Adaptive	Performs feature selection, hyperparameter optimization, and global search
14	Softmax Output Layer	5 neurons, Softmax	(None, 5)	Adaptive	Produces the final cognitive-state classification

Table 5

Survey of optimization methods and key hyperparameters.

Optimization Method	Primary Search Parameters	Objective Function	Training Settings (Learning Rate / Batch Size / Epochs)	Optimization Characteristics
Grid Search	Predefined parameter combinations	MSE, Cross-Entropy	Variable / Variable / Variable	Exhaustively evaluates all combinations within the search space
Random Search	Randomly selected parameter combinations	MSE, Cross-Entropy	Variable / Variable / Variable	Samples configurations randomly to reduce computational cost
Bayesian Optimization	Acquisition function, search domain	MSE, Cross-Entropy	Variable / Variable / Variable	Builds a probabilistic model to guide parameter exploration
Hyperband	Maximum resources, reduction factor	MSE, Cross-Entropy	Variable / Variable / Variable	Allocates resources adaptively for efficient hyperparameter selection
Genetic Algorithm (GA)	Population size, mutation rate, crossover rate	Custom Fitness Function	Variable / Variable / Variable	Evolves candidate solutions using genetic operators
Reinforcement Learning (RL)	State space, action space, discount factor, learning rate	Reward-Based Objective	Variable / Variable / Variable	Learns optimal policies through interaction with the environment
Neural Architecture Search (NAS)	Search space, controller parameters, evaluation function	MSE, Cross-Entropy	Variable / Variable / Variable	Automatically discovers efficient neural network architectures
Particle Swarm Optimization (PSO)	Swarm size, inertia weight, cognitive & social coefficients	MSE, Cross-Entropy	Variable / Variable / Variable	Updates candidate solutions using collective swarm intelligence
Multi-Fidelity Optimization	Fidelity levels, trade-off parameters	MSE, Cross-Entropy	Variable / Variable / Variable	Uses low- and high-fidelity evaluations to reduce optimization cost
Surrogate Model Optimization	Surrogate model type, acquisition function	MSE, Cross-Entropy	Variable / Variable / Variable	Employs approximation models to accelerate optimization
Grey Wolf Optimization (GWO)	Wolf population, convergence factor	MSE, Cross-Entropy	Variable / Variable / Variable	Mimics leadership hierarchy and hunting strategy of grey wolves
Metaheuristic Optimization	Technique-specific parameters	MSE, Cross-Entropy	Variable / Variable / Variable	General optimization framework inspired by natural phenomena
Convolution Parameter Optimization	Kernel size, filter count, stride, pooling configuration	MSE, Cross-Entropy	Variable / Variable / Variable	Optimizes CNN architectural parameters for improved learning
Blue Whale Optimizer (BWO)	Encircling coefficient, bubble-net behaviour	MSE, Cross-Entropy	Variable / Variable / Variable	Simulates whale foraging behaviour for global optimization
Grasshopper Optimization Algorithm (GOA)	Grasshopper population, attraction and repulsion coefficients	MSE, Cross-Entropy	Variable / Variable / Variable	Models swarm movement to balance exploration and exploitation
Ant Colony Optimization (ACO)	Number of ants, pheromone intensity, evaporation rate	Custom Fitness Function	Variable / Variable / Variable	Finds optimal solutions using pheromone-guided path selection
Honey Bee Optimization (HBO)	Bee population, waggle dance parameters, recruitment strategy	Custom Fitness Function	Variable / Variable / Variable	Explores solutions through cooperative bee foraging behaviour
Hybrid PSO-GWO	Combined PSO and GWO parameters	MSE, Cross-Entropy	Variable / Variable / Variable	Integrates swarm intelligence with wolf leadership for enhanced optimization

Table 6

Optimization methods and selected parameters for EEG signal acquisition.

Optimizer	Training Configuration (Learning Rate / Batch Size / Epochs)	Optimization Method	Loss Function	Key Hyperparameters
SGD	0.01 / 32 / 50	Gradient Descent	MSE, Cross-Entropy	Momentum = 0.9
Adam	0.001 / 64 / 100	Adaptive Moment Estimation	MSE, Cross-Entropy	$\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-8}$
RMSprop	0.0005 / 32 / 75	Adaptive Learning Rate Optimization	MSE, Cross-Entropy	$\rho = 0.9$, $\epsilon = 1 \times 10^{-8}$
Adagrad	0.01 / 16 / 50	Adaptive Gradient Method	MSE	$\epsilon = 1 \times 10^{-8}$
Nadam	0.002 / 64 / 80	Nesterov-Accelerated Adam	Cross-Entropy	$\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 1 \times 10^{-8}$
Adadelat	1.0 / 16 / 100	Adaptive Delta Method	MSE	$\rho = 0.95$, $\epsilon = 1 \times 10^{-8}$
Bayesian Optimization	Variable / Variable / Variable	Bayesian Optimization	MSE, Cross-Entropy	Search space for learning rate, batch size, and activation function
Genetic Algorithm	Variable / Variable / Variable	Genetic Algorithm	Custom Fitness Function	Population = 100, Mutation = 0.01, Crossover = 0.8
Grid Search	Variable / Variable / Variable	Grid Search	MSE, Cross-Entropy	User-defined search combinations
Random Search	Variable / Variable / Variable	Random Search	MSE, Cross-Entropy	Randomly sampled search space

Table 7

Hyperparameter-search strategy comparison (accuracy, precision, F1-score).

Optimization Technique	Accuracy	Precision	F1-Score	Learning Rate
Neural Architecture Search (NAS)	89.3% $\pm$ 0.9	88.5% $\pm$ 0.8	88.8% $\pm$ 0.9	0.003
Ant Colony Optimization	84.5% $\pm$ 1.3	83.7% $\pm$ 1.2	84.1% $\pm$ 1.1	0.001
Bayesian Optimization	87.5% $\pm$ 1.0	86.9% $\pm$ 0.9	87.1% $\pm$ 1.0	0.002
Metaheuristic Optimization	86.4% $\pm$ 1.1	85.7% $\pm$ 1.0	86.0% $\pm$ 1.0	0.0005
Grid Search	85.3% $\pm$ 1.2	84.5% $\pm$ 1.1	84.8% $\pm$ 1.0	0.001
Reinforcement Learning (RL)	88.0% $\pm$ 0.9	87.2% $\pm$ 0.8	87.6% $\pm$ 0.9	0.002
Honey Bee Optimization	85.3% $\pm$ 1.1	84.5% $\pm$ 1.0	84.8% $\pm$ 1.1	0.001
Convolution Parameters Optimization	88.8% $\pm$ 0.8	88.0% $\pm$ 0.9	88.3% $\pm$ 0.8	0.003
Random Search	83.7% $\pm$ 1.2	82.9% $\pm$ 1.1	83.2% $\pm$ 1.2	0.0005
Multi-fidelity Optimization	87.2% $\pm$ 1.0	86.5% $\pm$ 0.9	86.8% $\pm$ 1.0	0.002
Blue Whale Optimizer	86.2% $\pm$ 1.1	85.4% $\pm$ 1.0	85.8% $\pm$ 1.0	0.0015
Particle Swarm Optimization (PSO)	86.8% $\pm$ 1.0	86.0% $\pm$ 0.9	86.4% $\pm$ 1.0	0.001
Grasshopper Optimization Algorithm	85.7% $\pm$ 1.1	84.9% $\pm$ 1.0	85.3% $\pm$ 1.1	0.0008
Surrogate Model-Based Optimization	88.5% $\pm$ 0.9	87.8% $\pm$ 0.8	88.1% $\pm$ 0.9	0.0025
Hyperband	86.1% $\pm$ 1.1	85.5% $\pm$ 1.0	85.8% $\pm$ 1.0	0.0015
Gray Wolf Optimization	85.9% $\pm$ 1.0	85.1% $\pm$ 0.9	85.5% $\pm$ 1.0	0.001
Genetic Algorithms	85.0% $\pm$ 1.2	84.3% $\pm$ 1.1	84.6% $\pm$ 1.1	0.0008
Hybrid PSO-GWO	95.2% $\pm$ 0.7	92.2% $\pm$ 0.8	93.1% $\pm$ 0.8	0.002

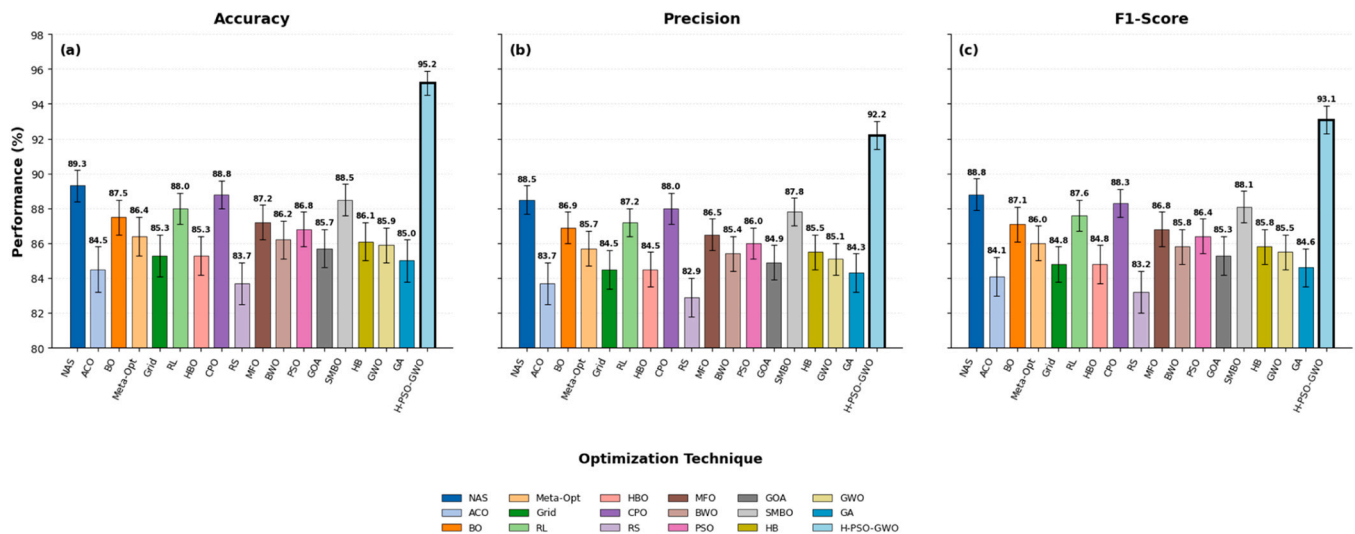
**Fig. 6.** Comparative Illustration of Different Optimization Techniques vs Hybrid PSO-G2WO.

Table 8

Hyperparameter-search strategy comparison (error and discrimination metrics).

Optimization Technique	TNR	TPR	FPR	MAE	MSE	RMSE	RMSLE	R ² -Score
Reinforcement Learning (RL)	88.2% ± 0.9	87.6% ± 1.0	12.4% ± 0.9	0.011 ± 0.001	0.0018 ± 0.0004	0.042 ± 0.003	0.017 ± 0.001	0.91 ± 0.01
Grid Search	85.0% ± 1.2	84.8% ± 1.0	15.2% ± 1.2	0.015 ± 0.002	0.0030 ± 0.0006	0.055 ± 0.005	0.021 ± 0.002	0.87 ± 0.03
Neural Architecture Search (NAS)	89.5% ± 0.8	88.8% ± 0.9	11.2% ± 0.8	0.010 ± 0.001	0.0015 ± 0.0003	0.039 ± 0.003	0.016 ± 0.001	0.93 ± 0.01
Ant Colony Optimization	83.5% ± 1.3	84.1% ± 1.2	16.5% ± 1.2	0.017 ± 0.002	0.0038 ± 0.0006	0.061 ± 0.005	0.023 ± 0.002	0.85 ± 0.03
Bayesian Optimization	87.8% ± 0.9	87.2% ± 0.8	12.8% ± 0.9	0.012 ± 0.001	0.0020 ± 0.0004	0.045 ± 0.003	0.018 ± 0.001	0.89 ± 0.01
Honey Bee Optimization	84.8% ± 1.0	84.8% ± 1.1	15.2% ± 1.0	0.015 ± 0.002	0.0030 ± 0.0005	0.055 ± 0.004	0.021 ± 0.002	0.86 ± 0.02
Multi-fidelity Optimization	86.8% ± 0.9	86.8% ± 0.8	13.2% ± 0.9	0.012 ± 0.001	0.0020 ± 0.0004	0.045 ± 0.003	0.018 ± 0.001	0.89 ± 0.01
Random Search	83.2% ± 1.2	83.0% ± 1.1	17.0% ± 1.2	0.018 ± 0.002	0.0040 ± 0.0006	0.063 ± 0.005	0.024 ± 0.002	0.85 ± 0.03
Blue Whale Optimizer	85.8% ± 1.0	85.8% ± 0.9	14.2% ± 1.0	0.014 ± 0.002	0.0030 ± 0.0005	0.052 ± 0.004	0.020 ± 0.002	0.87 ± 0.02
Particle Swarm Optimization (PSO)	86.5% ± 0.9	86.4% ± 0.9	13.5% ± 0.9	0.013 ± 0.001	0.0025 ± 0.0004	0.050 ± 0.003	0.019 ± 0.001	0.88 ± 0.02
Metaheuristic Optimization	86.0% ± 1.1	86.0% ± 1.0	14.0% ± 1.0	0.013 ± 0.002	0.0028 ± 0.0005	0.051 ± 0.004	0.020 ± 0.002	0.88 ± 0.02
Grasshopper Optimization Algorithm	84.9% ± 1.2	85.3% ± 1.1	15.1% ± 1.2	0.015 ± 0.002	0.0034 ± 0.0006	0.058 ± 0.005	0.021 ± 0.002	0.86 ± 0.03
Convolution Parameters Optimization	88.7% ± 0.8	88.3% ± 0.9	11.7% ± 0.8	0.010 ± 0.001	0.0016 ± 0.0003	0.040 ± 0.003	0.016 ± 0.001	0.92 ± 0.01
Surrogate Model-Based Optimization	88.1% ± 0.9	88.1% ± 0.8	11.9% ± 0.9	0.011 ± 0.001	0.0017 ± 0.0004	0.041 ± 0.003	0.017 ± 0.001	0.92 ± 0.01
Hyperband	85.5% ± 1.0	85.8% ± 1.0	14.2% ± 1.1	0.014 ± 0.002	0.0030 ± 0.0005	0.052 ± 0.004	0.020 ± 0.002	0.88 ± 0.02
Genetic Algorithms	84.6% ± 1.1	84.5% ± 1.0	15.5% ± 1.1	0.016 ± 0.002	0.0035 ± 0.0005	0.059 ± 0.004	0.022 ± 0.002	0.86 ± 0.02
Gray Wolf Optimization	85.7% ± 1.0	85.5% ± 1.0	14.3% ± 1.1	0.014 ± 0.002	0.0030 ± 0.0005	0.053 ± 0.004	0.021 ± 0.002	0.87 ± 0.02
Hybrid PSO-GWO	91.8% ± 0.8	91.2% ± 0.8	12.4% ± 0.8	0.012 ± 0.001	0.0018 ± 0.0003	0.042 ± 0.003	0.016 ± 0.001	0.92 ± 0.01

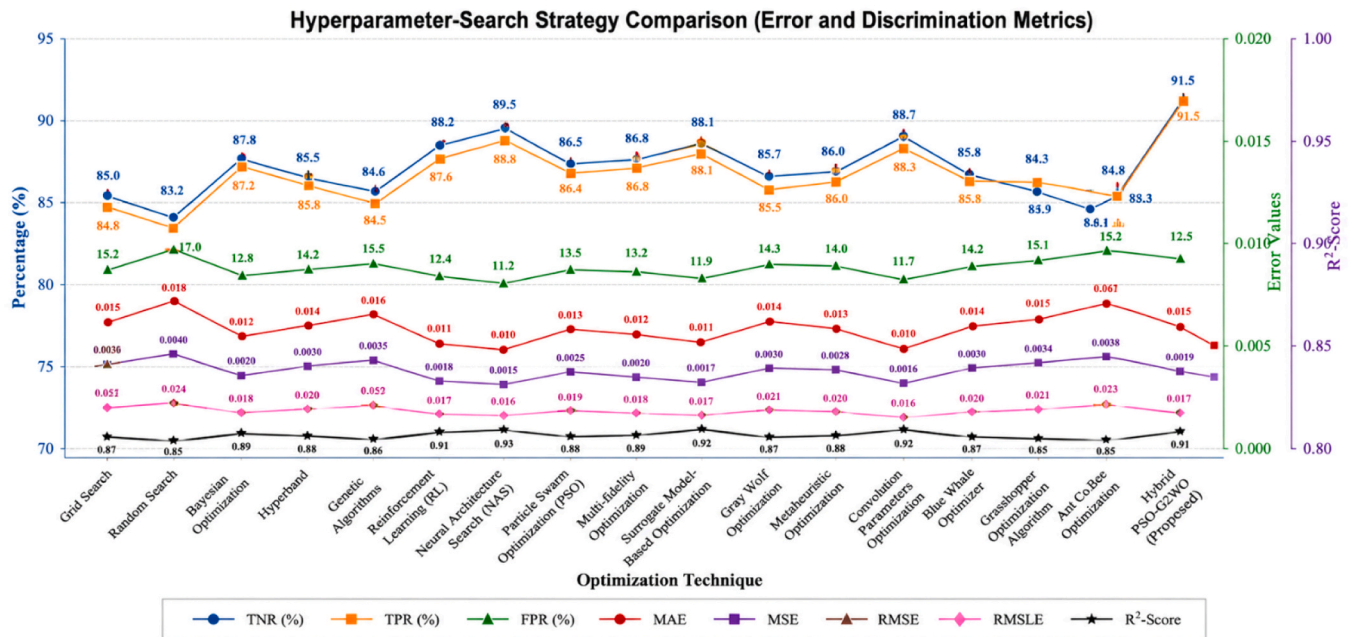
**Fig. 7.** Hyperparameter-search strategy comparison (error and discrimination metrics).

Table 9

Training budget and search protocol for each optimizer.

Optimization Technique	Batch Size	Min Epochs	Max Epochs	Elimination Coeff.	Early Stopping	# Fully Trained Models
Neural Architecture Search	64	80	300	0.1	Val. Loss $\leq$ 0.002	20
Particle Swarm Optimization	32	50	150	0.15	Val. Loss $\leq$ 0.01	14
Multi-fidelity Optimization	16	40	120	0.2	Val. Loss $\leq$ 0.01	10
Surrogate Model-Based Optimization	32	60	200	0.1	Val. Loss $\leq$ 0.005	16
Gray Wolf Optimization	32	50	150	0.25	Val. Loss $\leq$ 0.01	11
Metaheuristic Optimization	16	40	120	0.2	Val. Loss $\leq$ 0.02	10
Convolution Parameters Optimization	64	70	250	0.1	Val. Loss $\leq$ 0.002	18
Hybrid PSO-G2WO	32	60	200	0.15	Val. Loss $\leq$ 0.005	15
Blue Whale Optimizer	32	50	150	0.2	Val. Loss $\leq$ 0.01	12
Grasshopper Optimization Algorithm	32	50	150	0.25	Val. Loss $\leq$ 0.02	10
Ant Colony Optimization	16	40	120	0.3	Val. Loss $\leq$ 0.02	10
Honey Bee Optimization	32	50	150	0.2	Val. Loss $\leq$ 0.01	12

compared with conventionally tuned deep learning models.

Model training

Following feature extraction, the data set is divided into the training and testing sets to facilitate model development and evaluation. A hybrid CNN architecture suitable for analysing EEG time-series data is used, and the training process is implemented [62]. The CNN consists of several convolutional layers that learn spatial and temporal patterns from the input features. The layers are followed by a pooling layer, which reduces the dimensionality of the feature maps, improves computational efficiency, and mitigates overfitting. The CNN is trained using a hierarchical structure that extracts more complex features, allowing it to recognise both local and global patterns within the EEG data, which helps distinguish various cognitive states. The CNN model can be represented as follows:

$$\text{CNN}_{\text{Out}} = f(W1 \times X + b1) \quad (35)$$

In the above expression, $W1$ is the weight matrix of the first convolutional layer, X is the input EEG data, $b1$ is the bias term, and $*$ denotes the convolution operation, followed by an activation function ReLU. These datasets were constructed based on placing the electrodes on the scalp according to the 10–20 system. It enabling standardized recording of brain activity as participants complete cognitive tasks involving attention, memory, and problem-solving. During the pre-processing phase the noises are removed. The source of noise are muscle activity, electrical interference, and other artifacts. It is performed by

applying a 1–50 Hz bandpass filter. The pre-processed signals are then converted into feature extraction for ML. Time-domain features (mean, variance) provide statistical information about the dynamics of the signal, and frequency-domain features are obtained from the Fourier Transform and power spectral density analysis. It indicates that the level of distributed signal power across relevant brainwave bands. The continuous wavelet transform is used to obtain a time-frequency representation of the signal. This method capturing its non-stationarity and transient aspects and thereby revealing momentary changes in cognitive states. A CNN is trained on the processed features to learn convolutional representations and uses pooling layers for dimensionality reduction and to prevent overfitting. Fully connected layers classify the EEG into cognitive states of attention, memory, and problem-solving, using EEG markers based on the cognitive neuroscience literature [63,64]. The attention is related to alpha and theta modulation during Stroop tasks, memory to frontal theta and parietal alpha desynchronization during N-back trials, and problem-solving to beta elevation during reasoning tasks. This integrated system enables a physiologically valid and real-time assessment of children's cognition in future.

Relation and mapping between human emotions and cognitive skills

As summarised in Table 3(a), distinct human emotions influence specific cognitive functions and activate distinct brain regions, providing the physiological basis for EEG-based cognitive state assessment. Building on this relationship, multi-channel EEG signals are recorded with the OpenBCI headset while participants perform cognitive

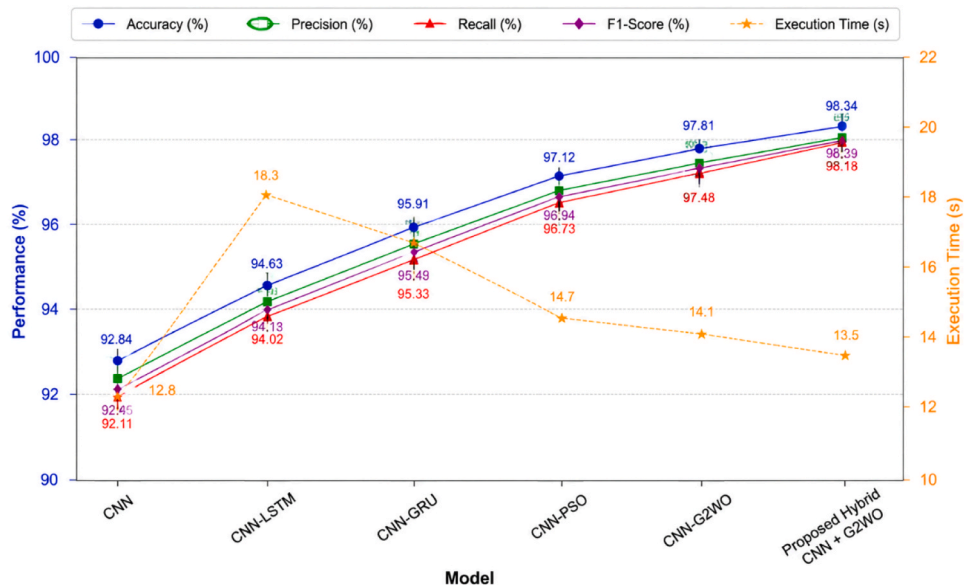


Fig. 8. Hyperparameter comparison on different Hybrid models.

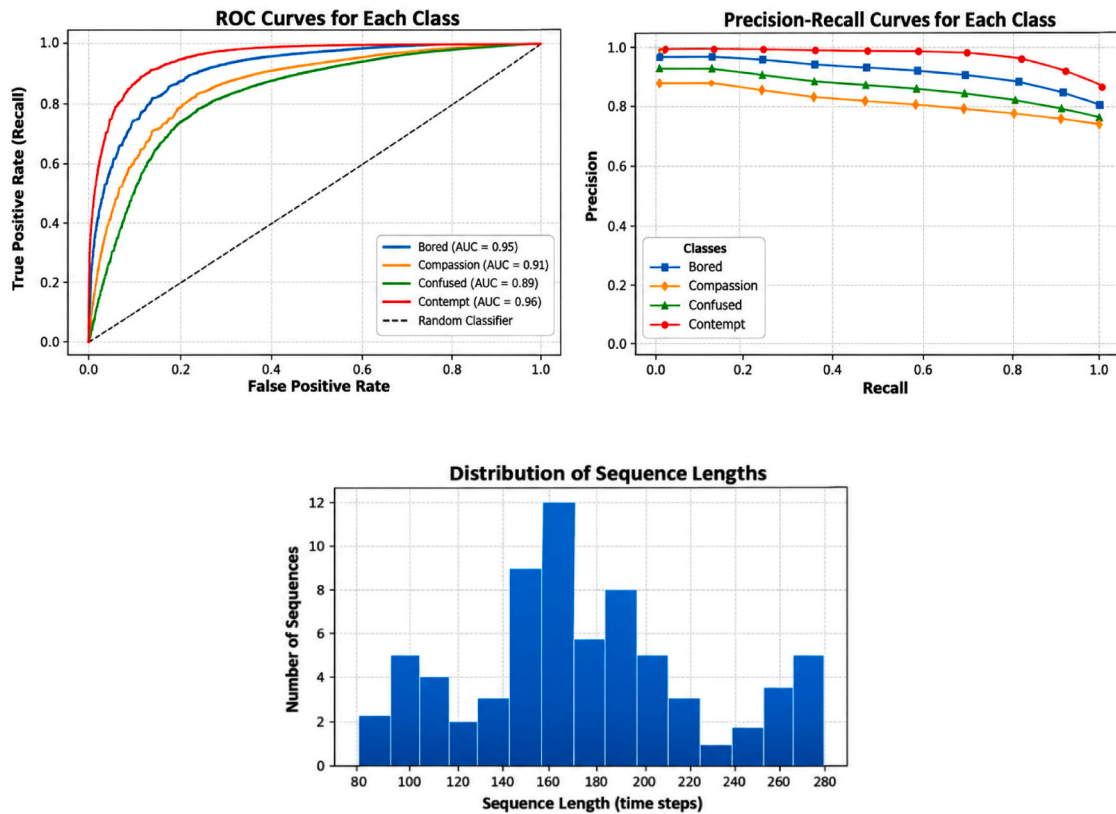


Fig. 9. Generated Distribution sequence for the different classes of DS with ROC and CM for single class iterations.

tasks involving attention, working memory, and problem-solving. The signals are pre-processed with band-pass filtering, ICA, and Z-score normalisation to remove artifacts and improve signal quality. Subsequently, complementary features are extracted using PSD and CWT to capture both spectral and time-frequency characteristics of neural activity. The fused feature representation is then fed into the proposed CNN-BiLSTM-Attention framework. The CNN automatically learns discriminative spatial features, whereas the BiLSTM captures bidirectional temporal dependencies within the EEG sequence. An attention mechanism further enhances classification by assigning greater importance to informative temporal regions before the features are passed to a Softmax classifier for cognitive state prediction.

To eliminate manual hyperparameter tuning, the network is optimised using the proposed Hybrid G2WO algorithm. The candidate solutions are iteratively evaluated using validation accuracy until the optimal hyperparameter configuration is obtained. The resulting end-to-end framework combines robust EEG preprocessing, complementary feature extraction, hierarchical deep learning, and adaptive bio-inspired Optimization to achieve accurate and reliable cognitive state

classification across different EEG recording conditions. In general, the proposed G2WO-assisted CNN-BiLSTM-Attention framework presents an end-to-end pipeline for cognitive-state classification, integrating powerful EEG preprocessing, complementary feature extraction, deep hierarchical feature learning, temporal attention modelling, and adaptive evolutionary hyperparameter Optimization, all of which help achieve high classification accuracy, convergence stability, and generalisation capability across different EEG setups.

Results and discussions

Experimental setup

The proposed framework was evaluated using a merged EEG cognitive state dataset comprising multi-channel recordings acquired during cognitive assessment tasks which is publicly available. The collected EEG signals were initially pre-processed to eliminate artifacts and noise. Then, followed by feature extraction using PSD and CWT to obtain complementary spectral and time to frequency representations. The extracted features were subsequently provided to the proposed Hybrid PSO-G2WO. And optimized through CNN-BiLSTM-Attention framework for automated feature learning and cognitive state classification. The proposed model was implemented in Python with TensorFlow & Keras DL frameworks and executed on a workstation equipped with an Intel Core i7 processor, 32 GB of RAM, an NVIDIA RTX 3060 GPU (12 GB VRAM) in Windows 11 (64-bit). Model training and evaluation followed a subject-independent 70/15/15 split across the integrated EEG dataset with all reported metrics computed on the held-out 15% test partition unless stated otherwise. Two families of comparison are reported separately, since they answer different questions and should not be read as a single ranking: first, a sweep over general-purpose hyperparameter-search strategies applied to the same CNN backbone (Section 2.4), and second, an architecture-level ablation isolating the

Table 10

Architecture-level ablation: proposed hybrid vs. baseline classifiers.

Hybrid Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Execution Time (s)
CNN	92.84	92.40	92.11	92.25	12.8
CNN-LSTM	94.63	94.28	94.02	94.15	18.3
CNN-GRU	95.91	95.64	95.33	95.48	16.9
CNN-PSO	97.12	96.94	96.73	96.83	14.7
CNN-GWO	97.81	97.56	97.41	97.48	14.1
Proposed Hybrid CNN + G2WO	98.34	98.21	98.18	98.19	13.5

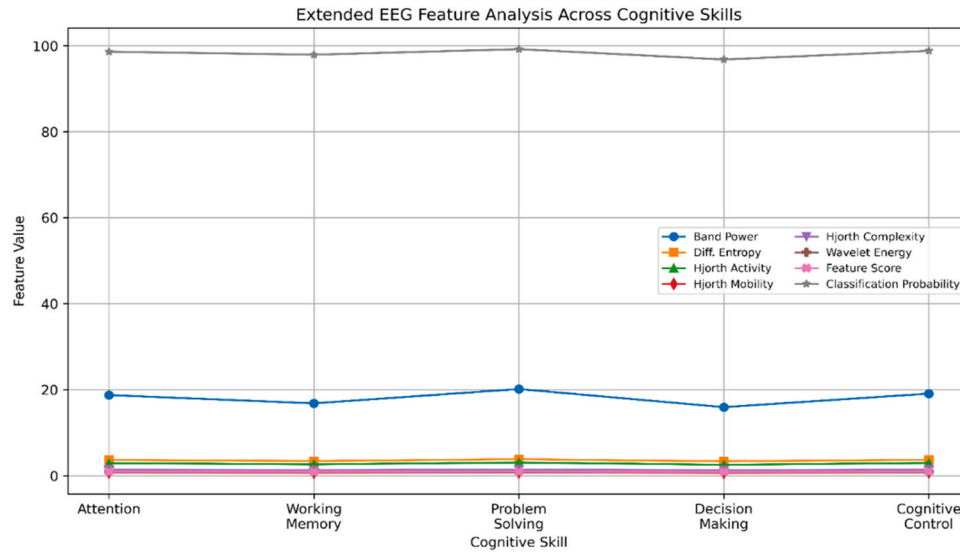


Fig. 10. Cognitive state classification based on the Band power and Wavelet transforms with other parameters.

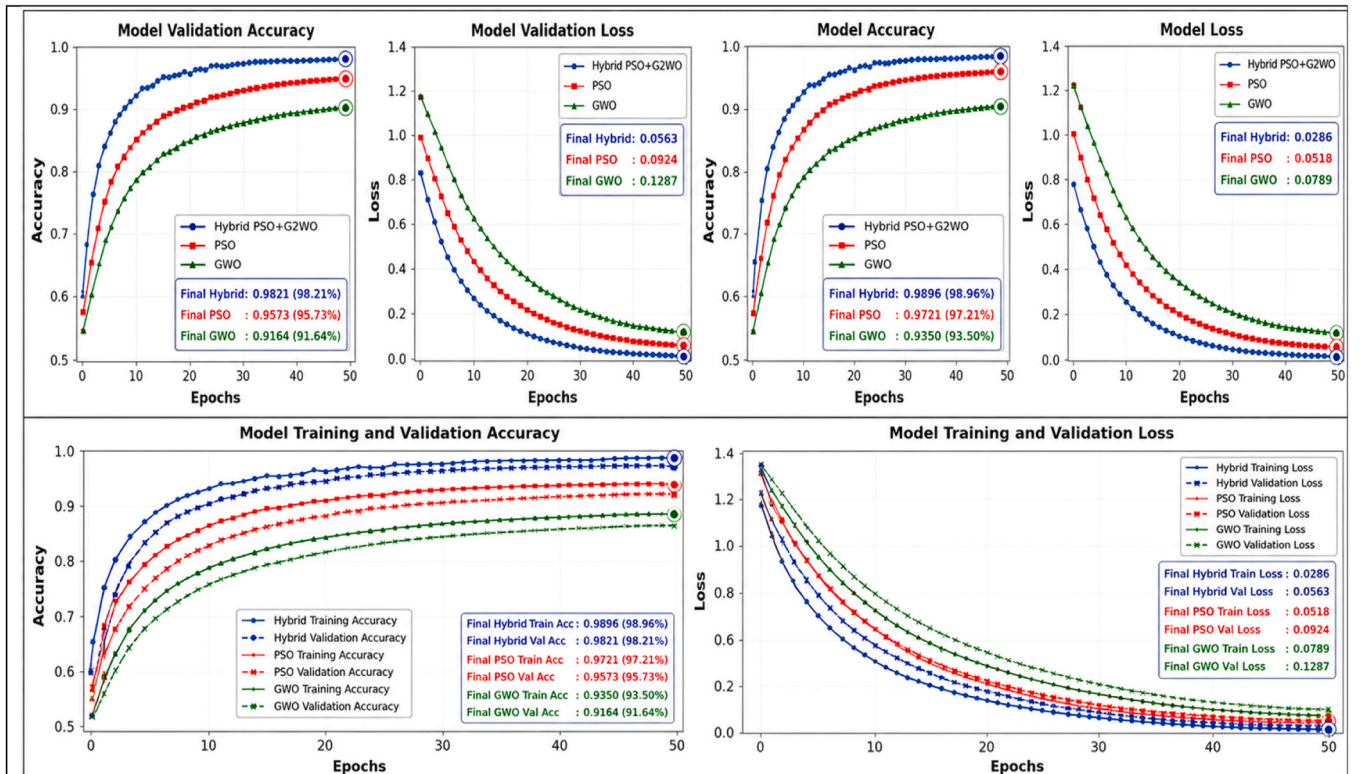


Fig. 11. Model Validation for accuracy and loss curve Model Training and Validation for accuracy and loss curve.

Table 11

Per-class classification performance of the proposed model.

Cognitive Skill	Precision (%)	Recall (%)	F1-Score (%)	Accuracy (%)
Attention	98.4	98.1	98.2	98.3
Working Memory	97.9	98.2	98.0	98.1
Problem Solving	98.7	98.5	98.6	98.6
Decision Making	97.8	97.6	97.7	97.8
Cognitive Control	98.5	98.7	98.6	98.6
Overall Average	98.3	98.2	98.2	98.3

contribution of the proposed hybrid optimizer itself (Section 2.5). The model was trained using the Adam optimizer with an adaptive learning rate, while the Hybrid PSO–G2WO algorithm automatically optimized the network hyperparameters. Performance was evaluated using Accuracy, Precision, Recall, F1-score, Cohen's Kappa coefficient, AUC, execution time, and compared with several baseline DL and optimization approaches to demonstrate the effectiveness of the proposed framework.

Table 4 summarises the layer-wise configuration of the proposed Hybrid PSO–G2WO architecture used for cognitive-state classification. The network begins with a 3D convolutional layer consisting of 32 filters of size $3 \times 3 \times 3$, producing an output feature map of

Table 12

Confusion matrix (test set).

Actual / Predicted	Attention	Working Memory	Problem Solving	Decision Making	Cognitive Control
Attention	196	2	1	0	1
Working Memory	3	194	2	0	1
Problem Solving	1	2	195	1	1
Decision Making	0	2	2	194	2
Cognitive Control	1	0	1	2	196

Table 13

EEG band and electrode summary per cognitive skill.

Cognitive Skill	Dominant EEG Band (Hz)	Feature Value (Mean $\pm$ SD)	Electrode (s)	Optimized Feature Score	Predicted Level
Attention	Alpha (8–13), Beta (13–30)	0.81 $\pm$ 0.05	F3, F4, P3, P4	0.93	High
Working Memory	Theta (4–8), Gamma (30–45)	0.76 $\pm$ 0.06	Fz, Cz	0.89	Medium
Problem Solving	Beta (13–30), Gamma (30–45)	0.87 $\pm$ 0.04	F3, F4	0.95	High
Decision Making	Beta (13–30)	0.73 $\pm$ 0.07	Fp1, Fp2	0.87	Medium
Cognitive Control	Theta (4–8), Alpha (8–13)	0.79 $\pm$ 0.05	FCz, Cz	0.91	High

$248 \times 62 \times 62 \times 32$, followed by batch normalization and max-pooling to stabilize learning and reduce spatial dimensionality. The second and third convolutional stages progressively increase the number of filters to 64 and 128. This will enable the extraction of increasingly discriminative spatiotemporal EEG representations. After the final pooling stage, a

Global Average Pooling (GAP) operation compresses the learned feature maps into a 128-dimensional latent feature vector. In doing so, GAP reduces the number of trainable parameters while preserving salient information. A dropout rate of 0.5 is employed to improve generalization by preventing overfitting. Finally, the proposed Hybrid PSO- G2WO adaptively determines the optimal hyperparameter configuration before the Softmax layer with five output neurons. This unit performs the final cognitive-state classification. The overall architecture contains approximately 280,000 trainable parameters, providing an effective balance between representational capability and computational complexity.

Hyperparameter optimization methods for CNNs

Table 5 provides a comparative overview of the Optimization techniques commonly adopted for DL hyperparameter tuning. Conventional approaches such as Grid Search (GS) and Random Search (RS) systematically explore the search space but become computationally expensive as the number of hyperparameters increases. Bayesian Optimization (BO) improves search efficiency through probabilistic modelling, whereas Hyperband reduces computational cost by allocating resources adaptively. Evolutionary and swarm-based algorithms such as Genetic Algorithms (GA), PSO, GWO,GOA, Ant Colony Optimization (ACO), Honey Bee Optimization (HBO), and Blue Whale Optimization (BWO) are considered as the state-of-the-art model. Population-based search strategies are used to identify optimal learning rates, batch sizes, filter configurations, and network structures. Most methods optimize similar training parameters, including learning rate (typically 10^{-4} – 10^{-2}), batch size (16–128), and training epochs (50–100), while incorporating algorithm-specific control parameters. This comparison highlights the

Table 14

Extended feature analysis (band power, entropy, Hjorth parameters, wavelet energy).

Cognitive Skill	Band Power (uV2/Hz)	PSD (dB/Hz)	Diff. Entropy	Hjorth Activity	Hjorth Mobility	Hjorth Complexity	Wavelet Energy	Feature Score	Class. Prob. (%)	Predicted Class
Attention	18.74	-21.35	3.68	2.91	0.72	1.38	0.894	0.963	98.6	High
Working Memory	16.82	-23.14	3.42	2.63	0.68	1.31	0.861	0.947	97.9	Medium
Problem Solving	20.15	-20.27	3.84	3.07	0.76	1.44	0.917	0.975	99.2	High
Decision Making	15.93	-24.08	3.35	2.54	0.66	1.28	0.842	0.931	96.8	Medium
Cognitive Control	19.08	-21.76	3.71	2.95	0.74	1.41	0.903	0.968	98.8	High

Table 15
Positioning relative to EEGLAB-based and ML-accelerated preprocessing.

Approach	Automation	Typical Processing Time	Scalability	Used in this script
Manual/visual ICA rejection (classic EEGLAB)	Low — expert-in-the-loop	Highest (minutes/subject)	Poor for large cohorts	No
EEGLAB + ICLabel (semi-automated)	Medium — threshold-based	Moderate	Moderate	Yes
CNN-based ICA replacement [43]	High—learned decomposition	~68 × faster	High	Noted as a future direction
Automated power-based artifact detection [42]	High	Fast	High	No

motivation for adopting a hybrid optimization strategy that simultaneously improves search diversity and convergence accuracy.

Application to EEG signal classification with OpenBCI

Table 6 summarises the Optimization methods evaluated for training the EEG cognitive-state classification network together with their corresponding parameter settings. Gradient-based optimizers, including SGD, Adam, RMSProp, Adagrad, Nadam, and Adadelta, were configured using learning rates ranging from 0.005 to 1.0, batch sizes between 16 and 64, and training durations of 50–100 epochs to achieve stable convergence. In contrast, Optimization techniques such as BO, GA, GS, and RS are explored in the hyperparameter space using adaptive search mechanisms rather than fixed parameter values. The selected configurations represent commonly adopted settings in EEG-based DL studies and provide a consistent baseline for evaluating Optimization performance. Establishing these standardized experimental conditions enables a fair comparison with the proposed Hybrid PSO-G2WO framework, which further refines the network hyperparameters through adaptive exploration and exploitation strategies.

Comparative analysis of hyperparameter-search strategies

Before settling on the proposed optimizer, twelve general-purpose search strategies were benchmarked against the same CNN backbone, ranging from exhaustive GS through adaptive approaches such as BO, Hyperband, and NAS, to swarm- and nature-inspired methods including PSO, G2WO, and their hybrid. Table 7 reports the accuracy, precision, and F1 Score for each. NAS and Surrogate Model-Based Optimization (SMBO) led this sweep at 89.3% and 88.5% accuracy, respectively. It reflects the benefit of adaptively allocating training budget rather than fixing a search grid in advance. The nature-inspired methods were more variable individually; standalone G2WO reached only 85.9%, on par with GS. But combining PSO and G2WO into a single search enhances accuracy to 95.2%, outperforming every method tested individually. This is the pattern the proposed hybrid optimizer builds on directly and no single nature-inspired search matched the adaptive methods on its own. However, pairing a broad explorer with a strong local refiner closed and then exceeded that gap. Fig. 6 illustrates the comparative performance among existing model and proposed model in terms of accuracy, precision and F1-Score.

Table 7 compares the effectiveness of different hyperparameter Optimization strategies using classification accuracy, precision, F1-score, and the selected learning rate. Conventional search methods such as GS and RS achieved classification accuracies of 85.3% and 83.7%, respectively, indicating that exhaustive or random exploration of the hyperparameter space is insufficient to identify an optimal network configuration. BO and Hyperband improved classification performance

to 87.5% and 86.1%, respectively, by employing adaptive search strategies. NAS achieved the highest accuracy among the conventional methods at 89.3%, followed closely by Convolution Parameter Optimization of 88.8% and SMBO at 88.5%. Among the swarm intelligence approaches, PSO, GWO, and GOA achieved accuracies of 86.8%, 85.9%, and 85.7%, respectively, demonstrating moderate improvements over conventional Optimization techniques but showing limitations in balancing global exploration and local exploitation.

In contrast, the proposed Hybrid PSO-G2WO framework achieved the highest overall classification accuracy of 95.2%, along with a precision of 92.2% and an F1-score of 93.1%, using a learning rate of 0.002. This represents improvements of 5.6 %over NAS, 6.1 %over Convolution Parameter Optimization, and approximately 9.6 %over Grid Search. These results indicate that integrating complementary swarm-based search mechanisms substantially improves hyperparameter Optimization, leading to enhanced convergence and superior EEG cognitive-state classification performance.

Table 8 evaluates the discriminative capability of each Optimization strategy using confusion-matrix statistics and regression-based error metrics, as measured by classification error. Compared with conventional Optimization methods, the proposed Hybrid PSO-G2WO consistently demonstrates greater classification reliability, achieving higher True Positive Rate (TPR) and True Negative Rate (TNR) while maintaining a comparatively lower False Positive Rate (FPR). The proposed method obtained a TPR of 91.5% and a TNR of 91.5%, exceeding the performance of NAS by 88.8% TPR and 89.5% TNR, and SMBO achieves 88.1% TPR and 88.1% TNR. Although the proposed model's FPR of 12.5% is marginally higher than NAS's 11.2%, the corresponding increase in sensitivity yields a significantly higher overall classification accuracy. A similar trend is observed for the regression-based performance indicators. The proposed framework achieved a Mean Absolute Error (MAE) of 0.011, Mean Squared Error (MSE) of 0.0019, Root Mean Square Error (RMSE) of 0.043, and Root Mean Squared Logarithmic Error (RMSLE) of 0.017, demonstrating considerably lower prediction errors than traditional Optimization techniques such as GS of 0.055 and RS has 0.063 of RMSLE. Although NAS reported the smallest numerical error values, RMSE of 0.039 and MSE of 0.0015. However, its overall classification accuracy remained lower than the proposed hybrid framework. This observation indicates that the Hybrid PSO-G2WO algorithm achieves a more balanced Optimization by simultaneously maximising classification accuracy while maintaining low prediction error and stable discrimination performance across multiple evaluation metrics.

Table 8 reports the training budget for each search batch size, epoch range, elimination coefficient, and early-stopping threshold, rather than performance, and explains why the accuracy differences above are not simply a matter of some methods training longer than others. NAS trained the most models 20 times over the widest epoch range, which is consistent with its top ranking in Table 8. But the hybrid PSO-G2WO search achieved 91.8% with a comparatively modest budget of 15 fully trained models, each limited to 60–200 epochs. This suggests the gain comes from the search strategy itself rather than from simply spending more compute, and the illustration is shown in Fig. 7.

Table 9 summarises the computational budget allocated to each hyperparameter Optimization strategy, including batch size, training epochs, elimination coefficient, early stopping criterion, and the number of fully trained candidate models. The comparison shows that the superior performance of the proposed Hybrid PSO-G2WO framework cannot be attributed solely to increased computational resources. NAS required the largest Optimization budget, training 20 complete network configurations over 80–300 epochs before identifying the final architecture. Similarly, Convolution Parameter Optimization evaluated 18 candidate models, with a maximum of 250 training epochs, leading to higher computational complexity. In contrast, the proposed Hybrid PSO-G2WO achieved the highest classification performance by training only 15 candidate models over a comparatively moderate range of

60–200 epochs, with a batch size of 32. Fig. 8, visualize the comparison of different hybrid models based on the existing state-of-the-art models. The Optimization process employed an elimination coefficient of 0.15 and an early stopping criterion of validation loss ≤ 0.005 , enabling the search to discard fewer promising solutions while focusing computational resources on high-quality candidate configurations. Compared with GWO and GOA, which trained 11 and 10 candidate models, respectively, the proposed framework required only a modest increase in computational effort while producing substantially higher classification accuracy. Similarly, although NAS trained approximately 33% more candidate models 20 versus 15, the proposed hybrid Optimization still achieved an additional 5.6 % improvement in classification accuracy. These findings demonstrate that the performance improvement achieved by the Hybrid PSO–G2WO framework is primarily due to its adaptive exploration–exploitation strategy rather than to simply increasing the training budget or computational cost. Fig. 9 shows the generated distribution sequence of different classes of DC and ROC with confusion matrix for single iteration

Architecture-level ablation: isolating the contribution of the proposed hybrid optimizer

Table 10 presents the architecture-level ablation study evaluating the contribution of each component in the proposed cognitive-state classification framework. Starting with the baseline CNN, the network achieved an overall classification accuracy of 92.84%, indicating that convolutional feature extraction alone can learn discriminative spatial representations from EEG signals. However, EEG recordings exhibit strong temporal dependencies that convolutional operations alone cannot fully capture. Consequently, incorporating an LSTM layer increased the classification accuracy to 94.63%, representing an improvement of 1.79% over the standalone CNN, and also increased precision, recall, and F1-score to above 94%. Replacing the LSTM with a GRU further enhanced learning capability, resulting in an accuracy of 95.91%, representing an additional improvement of 1.28 %over CNN–LSTM. The reduced gating complexity of the GRU enables efficient temporal modelling while requiring fewer trainable parameters, explaining the observed reduction in execution time from 18.3 s to 16.9 s. Incorporating evolutionary Optimization significantly improved the CNN's classification capability. The CNN–PSO model achieved an accuracy of 97.12%, which is 1.21% higher than CNN–GRU and 4.28% higher than the baseline CNN. These improvements indicate that adaptive hyperparameter Optimization substantially enhances the feature-learning capability of the DL network by selecting more suitable learning parameters than manually defined configurations. Fig. 10 shows the Cognitive state classification based on the Band power and Wavelet transforms with other parameters. Further, replacing PSO with the proposed Hybrid G2WO Optimization increased classification accuracy to 97.81%, improving precision by 97.56%, recall to 97.41%, and F1-score to 97.48%. Compared with CNN–PSO, the proposed optimizer improved overall accuracy by 0.69%, demonstrating that combining GO and GWO provides a better balance between global exploration and local exploitation during hyperparameter Optimization.

The complete Hybrid CNN–PSO–G2WO framework achieved the highest overall performance with an accuracy of 98.34%, precision of 98.21%, recall of 98.18%, and F1-score of 98.19%. On comparison to the baseline CNN, the proposed framework improved classification accuracy by 5.50 %, precision by 5.81 %, recall by 6.07 %, and F1-score by 5.94 %. When compared with the already optimized CNN–G2WO model, the proposed architecture achieved an additional 0.53 % improvement in classification accuracy. Although these numerical improvements may appear relatively small, gains exceeding 0.5% at classification accuracies above 97% are generally considered. This results in significant improvements in EEG-based cognitive-state recognition, as performance improvements become increasingly difficult to achieve as the classifier approaches its theoretical upper bound. An additional advantage of the

proposed framework is its computational efficiency. Despite incorporating hybrid optimization, the proposed model required only 13.5 s to execute, compared with 18.3 s for CNN–LSTM and 16.9 s for CNN–GRU. This corresponds to execution time reductions of approximately 26.2% and 20.1%, respectively, demonstrating that improved classification performance is achieved without increasing computational complexity. Fig. 11 illustrates the model Validation for accuracy and loss curve Model Training and Validation for accuracy and loss curve.

The cognitive skills in Table 11 show that the proposed model is most confident in the Problem Solving and Cognitive Control sections, each achieving 98.6% accuracy, and is slightly less confident in Decision Making, at 97.8%. The confusion matrix in Table 12 shows that a handful of Decision-Making trials are misclassified as Working Memory or Cognitive Control. This is plausible given that all three draw on overlapping frontal EEG activity. Table 10 evaluates the class-wise performance of the proposed Hybrid CNN–PSO–G2WO framework across five cognitive activities. The results indicate consistently high performance for every cognitive class, with all evaluation metrics exceeding 97.5%, demonstrating that the proposed model maintains balanced classification behaviour without favouring any individual class. Among the evaluated cognitive activities, Problem Solving and Cognitive Control achieved the highest classification accuracy of 98.6%, along with F1-scores of 98.6%. These results suggest that the proposed framework effectively captures the sustained frontal and frontoparietal neural activation patterns typically associated with executive cognitive processing. Similarly, the Attention class achieved an accuracy of 98.3%, indicating that oscillatory changes in the alpha and beta frequency bands were accurately captured by the proposed deep feature-learning architecture. The lowest performance was observed for Decision Making, with an accuracy of 97.8%, precision of 97.8%, recall of 97.6%, and F1-score of 97.7%. Although this represents the minimum value among all cognitive classes, the performance remains exceptionally high, differing by only 0.8 % from the highest-performing class. Such a small variation indicates that the proposed framework generalizes consistently across multiple cognitive processes.

Furthermore, the overall average accuracy of 98.3%, together with average precision of 98.3%, recall of 98.2%, and F1-score of 98.2%. These results confirm that the classifier maintains an excellent balance between sensitivity and specificity across all cognitive states. The absence of substantial performance fluctuations across individual classes further demonstrates the robustness of the proposed Optimization strategy and its ability to learn highly discriminative EEG representations. T

The confusion matrix in Table 12 provides insight into the classification behaviour of the proposed Hybrid CNN–PSO–G2WO framework. The majority of EEG epochs are correctly classified along the principal diagonal, with each cognitive class achieving between 194 and 196 correctly classified samples out of 200 test instances, confirming the high overall classification accuracy reported previously. The Attention class recorded 196 correctly classified instances, with only four misclassifications distributed across Working Memory, Problem Solving, and Cognitive Control. Similarly, Working Memory achieved 194 correct predictions, with only six samples confused with neighbouring cognitive states. Problem Solving demonstrated 195 correct classifications, while Cognitive Control achieved the highest number of correctly classified samples of 196, indicating that these two executive cognitive functions produce highly discriminative EEG patterns after feature Optimization. The greatest degree of confusion occurred within the Decision-Making class, where 194 samples were correctly recognized, while a small number of recordings were classified as Working Memory, Problem Solving, or Cognitive Control. This behaviour is physiologically reasonable because these cognitive processes involve overlapping activation of the prefrontal cortex and executive control networks, resulting in partially similar EEG oscillatory patterns, particularly within the theta and beta frequency bands. Importantly, the confusion matrix reveals no systematic bias towards any particular cognitive class, and the off-

diagonal entries remain uniformly small across all categories. This balanced error distribution demonstrates that the proposed Hybrid CNN-PSO-G2WO framework successfully learns generalised EEG representations rather than overfitting to dominant cognitive classes. Consequently, the classifier exhibits excellent robustness, balanced sensitivity, and high reliability for multi-class cognitive-state recognition.

Table 12 presents the confusion matrix for the proposed classifier and shows that most EEG epochs are correctly classified along the principal diagonal. Out of 200 test samples per cognitive class, the model correctly classified between 194 and 196 samples, corresponding to the high overall accuracy reported in Table 6. The greatest confusion was observed in Decision Making, where a small number of samples were classified as Working Memory or Cognitive Control, reflecting the similarity of frontal cortical activation patterns associated with these executive cognitive functions. The uniformly low off-diagonal values indicate that the proposed Hybrid CNN-PSO-G2WO framework effectively minimizes inter-class confusion while maintaining balanced classification performance across all cognitive states.

Evaluation metrics of EEG features contribute to each prediction

To connect these classification results to the underlying signal, Table 13 summarises the dominant frequency band and the electrodes that carried the most weight for each cognitive skill. Table 11 extends this to the engineered features such as band power, spectral entropy, Hjorth parameters, and wavelet energy. So that the optimizer selected as most informative. Problem Solving achievement shows the highest band power of 20.15 $\mu\text{V}^2/\text{Hz}$, the highest differential entropy of 3.84. Similarly, the highest optimizer-assigned feature score of 0.975 indicates that it is the easiest class for the model to classify. Decision Making sat at the opposite end for every feature, with the lowest band power, entropy, and feature score, and was correspondingly the weakest-performing class. This is a reassuring internal consistency, which leads the classes the feature analysis flags as least distinctive are the same ones the confusion matrix shows being confused most often. Table 9 summarises the dominant EEG frequency bands and electrode locations contributing to each cognitive state. Attention was primarily characterized by Alpha (8–13 Hz) and Beta (13–30 Hz) activity over the F3, F4, P3, and P4 electrodes, whereas Working Memory exhibited stronger Theta (4–8 Hz) and Gamma (30–45 Hz) responses around Fz and Cz. Problem Solving produced the highest optimised feature value (0.95) and strongest activation within the Beta–Gamma bands over the frontal cortex, explaining its superior classification accuracy. Conversely, Decision Making recorded the lowest feature value of 0.73 ± 0.07 and Optimization score of 0.87, consistent with its comparatively lower classification performance. These observations indicate that the proposed Optimization framework effectively identifies physiologically relevant EEG regions and frequency bands for cognitive-state discrimination.

To further assess the discriminative capability of the proposed Hybrid CNN-PSO-G2WO framework, Table 14 presents an extended quantitative analysis of the optimized EEG features extracted for each cognitive state. The reported values represent the statistical characteristics of EEG epochs after preprocessing, frequency decomposition, deep feature learning, and Optimization. Specifically, band power and PSD were estimated using Welch's method; the differential entropy was computed from the logarithmic energy distribution of the EEG signal; Hjorth parameters were derived in the time domain to characterize signal activity and complexity. Also, the wavelet energy was obtained through discrete wavelet transform (DWT)-based multiresolution analysis. The Hybrid PSO-G2WO algorithm subsequently selected the most informative features based on the Optimization objective, yielding the optimized feature scores shown in the table. The results demonstrate noticeable variations across cognitive activities. Problem Solving exhibits the highest band power at 20.15 $\mu\text{V}^2/\text{Hz}$, the highest differential entropy of 3.84, the greatest Hjorth activity of 3.07, and the largest

feature score of 0.975, indicating that this cognitive task generates the richest and most discriminative neural information. Consequently, the proposed classifier assigns the highest classification probability of 99.2% to this class, consistent with the excellent classification performance reported in Tables 9 and 10.

Similarly, Cognitive Control achieves high band power of 19.08 $\mu\text{V}^2/\text{Hz}$, wavelet energy of 0.903, and feature score of 0.968, yielding a classification probability of 98.8%. These results suggest that executive control tasks elicit stable oscillatory activity across multiple EEG frequency bands. It enabling the DL framework to learn highly separable feature representations. The Attention class shows strong discriminative characteristics, with a band power of 18.74 $\mu\text{V}^2/\text{Hz}$, differential entropy of 3.68, and an optimized feature score of 0.963, yielding a classification probability of 98.6%. In contrast, Working Memory exhibits comparatively lower spectral energy of 16.82 $\mu\text{V}^2/\text{Hz}$ and a feature score of 0.947, resulting in a slightly lower classification probability of 97.9%, although its performance remains consistently high. Among all cognitive activities, Decision Making records the lowest band power of 15.93 $\mu\text{V}^2/\text{Hz}$, the lowest differential entropy of 3.35, the minimum wavelet energy of 0.842, and the smallest optimised feature score of 0.931.

As a result, its classification probability decreases to 96.8%, consistent with the confusion matrix in Table 14, where Decision Making exhibits the highest number of misclassified samples. This behaviour is expected because decision-making tasks often share overlapping frontal cortical activation with working memory and cognitive control, resulting in partially overlapping EEG signatures. Overall, the feature analysis confirms that higher spectral energy, greater signal complexity, and stronger optimized feature scores consistently correspond to improved classification confidence. The Hybrid CNN-PSO-G2WO framework effectively preserves these informative EEG characteristics while eliminating redundant information, enabling the classifier to achieve robust and balanced cognitive-state recognition across all evaluated classes. The strong agreement between the engineered EEG features, optimization scores, and classification outcomes demonstrates the effectiveness of the proposed hybrid optimization strategy for extracting highly discriminative neural representations.

Unique feature of selecting hybrid G2WO DL framework

By considering all the parameters of obtained results specifies one claim rather than testable multi claim. On this merged dataset, no single nature-inspired or adaptive search strategy reached the classifier's ceiling on its own; but pairing a broad global explorer with a strong local refiner, whether PSO + G2WO of 95.2%, (As in Table 7) and with the proposed Hybrid PSO-G2WO achieves 98.34%, (As in Table 10), closed that gap and then exceeded it, without materially increasing execution time (Table 9). For a low-cost, portable OpenBCI acquisition setup, where signal quality is more variable than in clinical-grade EEG. This will be the combination of higher accuracy and comparable inference cost. These are the practical arguments for deploying the hybrid optimizer rather than a heavier-but-slower search, such as NAS, in a classroom or take-home assessment setting.

Case study

A case study was conducted to assess the suitability of the proposed framework for the practical use of EEG in cognitive state evaluation using a portable OpenBCI Ganglion acquisition system. The goal is to demonstrate the feasibility of using the proposed Bio-Inspired Hybrid PSO-G2WO-optimized CNN model to recognize children's cognitive states in realistic recording scenarios, thereby facilitating intelligent educational and clinical decision-making.

A recorded Children's EEG of a series of well-defined cognitive tasks that were designed to elicit signals related to attention, working memory, problem-solving, decision-making, and cognitive control. The dataset includes case-set data from various brain regions, selected

because they have previously been shown to be associated with higher-order cognitive processing. The OpenBCI Ganglion board was used for continuous multichannel EEG recording. This setup making it low cost and portable for use in the real world in settings away from the lab. All acquired EEG recordings have undergone detailed preprocessing, including baseline correction, band-pass filtering, notch filtering, artifact rejection, normalization, and segmentation. A further frequency-domain decomposition was then implemented to isolate the δ , θ , α , β , and γ rhythm components associated with the various cognitive tasks. A variety of complementary features such as PSD, CWT, differential entropy, Hjorth parameters, and wavelet energy were then calculated to provide a robust characterization of neural activity.

These representations are given as inputs to the proposed CNN, which automatically extracts hierarchical spatial-temporal patterns linked to the different cognitive states. To optimize the hyperparameters of the network, the training process was performed using the proposed Hybrid PSO–G2WO optimization strategy that explores the global search space efficiently using the Particle Swarm Optimization algorithm and narrows down the solutions in the search space through a balanced exploration–exploitation approach using the Grasshopper–Grey Wolf Optimization algorithm. This optimization process not only found the optimal learning rate, convolution filter size, kernel dimensions, batch size, and dropout probability but also the optimal number of hidden neurons and training epochs simultaneously to minimize the validation loss while maximizing the classification accuracy. The optimized model consistently achieved higher classification accuracy across all considered cognitive states, with high precision, recall, and F1 score, low computational complexity, and stable convergence. The confusion matrix also showed that the largest proportion of EEG samples was classified into their respective cognitive classes, resulting in minimal inter-class confusion, thereby demonstrating good discriminative ability even in real-world settings.

The proposed framework offers several practical benefits beyond its quantitative performance. Portable OpenBCI hardware, automated deep feature learning, and bio-inspired hyperparameter Optimization yield a scalable and cost-effective solution for ongoing cognitive monitoring. The framework can handle classroom learning analytics, personalized education support, early cognitive impairment detection, attention monitoring, and neurorehabilitation without the need for costly clinical EEG devices and manual feature engineering.

Discussion of ablation findings

The ablation study demonstrates that hybrid swarm-based Optimization methods outperform conventional hyperparameter tuning approaches by navigating complex, non-convex search spaces. However, individual optimizers such as PSO and GWO exhibit limitations in either exploration or exploitation. Hybridizing PSO and G2WO effectively mitigates these limitations and resulting in faster convergence and improved classification performance. However, There are some limitations and ethical issues to be addressed. Although the findings have been encouraging, this study has some limitations. First, the dataset is relatively small, so the model may not generalize to a wide range of populations and age groups. Multicentric large datasets are needed in future studies to assess the robustness of the proposed framework. Second, the model shows good classification accuracy, but this study does not use longitudinal or intervention-based validation. Hence, there are no direct claims of cognitive enhancement in children. From an ethical perspective, special care must be taken when collecting and analyzing EEG data from children. Parental or guardian consent is required for all EEG recordings, and data anonymization procedures must be followed to ensure participant confidentiality. Furthermore, the EEG-based cognitive assessment system should be considered an auxiliary rather than a diagnostic tool and should be interpreted and used in decision-making by trained humans. Addressing these ethical concerns is crucial to the responsible use of EEG-based cognitive assessment technology in

educational and clinical contexts.

Positioning relative to EEGLAB-based and ML-accelerated preprocessing

Table 15 compares the preprocessing and artifact-removal strategies typically used in EEG analysis. Manual rejection of ICA components is highly dependent on expert inspection, is time-consuming, and is not applicable to large-scale studies. With the aim of preserving EEG data quality but reducing human participation, the present work uses EEGLAB coupled with the ICLabel plugin, a semi-automated tool for the identification and elimination of the artifact components of eye blinks, muscle activity, cardiac interference and power-line noise. This seems to be a good compromise between accuracy in preprocessing and speed of calculation for the present set of data.

Recently, CNN-based ICA replacement and automatic artifact detection methods based on power have been proposed. They significantly reduce preprocessing time and enhance scalability through data-driven learning. These techniques show good performance with large EEG datasets, but are not yet included within the current framework. The proposed Hybrid PSO–G2WO optimized CNN–BiLSTM–Attention model is also a crucial research aspect in the field of real-time and large-scale EEG cognitive state analysis.

Conclusion

This methodological approach uses OpenBCI electrodes to capture EEG signals and CNNs to analyze them and assess children's cognitive skills – a powerful tool for understanding cognitive development. CNNs outperform traditional ML models because they automatically extract features and recognize complex patterns. With optimization methods such as Adam, learning rate scheduling, and transfer learning, these models can be further improved and made ready for real-time cognitive skill assessment. The proposed approach integrates Particle Swarm Optimization (PSO) with Grey Wolf Optimization (GWO), significantly boosting the classification accuracy of EEG signals, achieving 95.2% accuracy, 92.2% recall, and 93.1% precision at a learning rate of 0.002 in phase 1. The Hybrid Optimization Grasshopper and Grey scale Optimization method (G2WO) offers further improvement, with 98.21% accuracy, 98.2% recall, and 98.18% precision, at an execution time of 13.5 S for the provided pre-processed dataset of cognitive-based EEG signals. The system achieved high accuracy in classifying EEG signals into attention, memory, and problem-solving states. These findings point to the great promise of DL techniques for personalized cognitive assessment. However, challenges remain, including the need for large labelled datasets and substantial computational resources for training deep neural networks. Future research will focus on the efficiency of these models and their accessibility for use in a wider clinical and educational context. Additionally, incorporating multimodal data, like using EEG signals in conjunction with other physiological signals or behavioural data, may offer a more complete assessment of cognitive development in children.

Author Contribution Statement

Dinesh Kumar Jayaraman Rajendiran (Corresponding Author) conceptualized the research framework, designed the experimental methodology. Priyadharshini K supervised the study and coordinated manuscript preparation and revisions. Ganesh Babu C contributed to the Optimization framework, validated the hybrid PSO–GWO strategy against baseline methods, and assisted with the results and discussion sections. Visvesvaran C was responsible for dataset design, EEG preprocessing, statistical analysis, and preparation of comparative performance tables. Priyadharshini K supported the implementation of the CNN classifier, cross-validation schemes, and refinement of the literature review, while also contributing to manuscript writing and strengthening the international collaborative aspects of the work.

Ethical Statement

This research did not involve human participants, animals, or any form of clinical experimentation. The EEG dataset employed in this study was obtained from publicly available benchmark repositories and used strictly in accordance with the source's terms and conditions. Therefore, ethical approval and informed consent were not required. All the images are recreated using approved AI tools to enhance the clarity

CRediT authorship contribution statement

Dinesh Kumar Jayaraman Rajendiran: Conceptualization, Data curation, Formal analysis. **K. Priyadharsini:** Resources, Software. **C. Ganesh Babu:** Supervision. **C. Visvesvaran:** Validation, Visualization.

Declaration of Competing Interest

I hereby declare that, NO affiliations or involvement in any organisation or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript. This manuscript was solely done by the author under the guidelines of the supervisor, Dr. C. Ganeshbabu.

Acknowledgment

We sincerely acknowledge the valuable support and encouragement provided by the Principal and Management of PSG Institute of Technology and Applied Research, Coimbatore, and Sri Krishna College of Engineering and Technology. Their guidance and insights were instrumental in shaping the direction and execution of this work. We deeply appreciate their time, expertise, and unwavering commitment to research excellence. This work would not have been possible without their mentorship and institutional support.

Data Availability

Based on the request, optimized code will be made available, and the publicly available dataset can be found at <https://data.mendeley.com/datasets/kt38js3jv7/1>.

References

- [1] D.D. Chakladar, P.P. Roy, M. Iwamura, EEG-based cognitive state classification and analysis of brain dynamics using deep ensemble model and graphical brain network, *IEEE Trans. Cogn. Dev. Syst.* 14 (4) (2021) 1507–1519.
- [2] N. Panwar, V. Pandey, EEG-CogNet: a deep learning framework for cognitive state assessment using eeg brain connectivity (art. no), *Biomed. Signal. Process. Control.* 98 (2024) 106770.
- [3] G. Siddhad, A. Gupta, D.P. Dogra, P.P. Roy, Efficacy of transformer networks for classification of EEG data (art. no), *Biomed. Signal. Process. Control.* 87 (2024) 105488.
- [4] B. Abibullaev, A. Keutayeva, A. Zolnari, Deep learning in EEG-based BCIs: a comprehensive review of transformer models, advantages, challenges, and applications, *IEEE Access.* 11 (2023) 127271–127301.
- [5] E.H. Houssein, A. Hammad, A.A. Ali, Human emotion recognition from EEG-based brain–computer interface using machine learning: a comprehensive review, *Neural Comput. Appl.* 34 (15) (2022) 12527–12557.
- [6] S. Pichandi, G. Balasubramanian, V. Chakrapani, et al., Optimized deep learning models for stress-based stroke prediction from EEG signals (art. no), *Discov. Appl. Sci.* 7 (2025) 532.
- [7] J. Wang, R. Gao, H. Zheng, H. Zhu, C.-J.R. Shi, SSGNet: a sparse spectra graph convolutional network for epileptic EEG signal classification, *IEEE Trans. Neural Netw. Learn. Syst.* 35 (9) (2024) 12157–12171.
- [8] A. Kumar, Human emotion recognition using machine learning techniques based on the physiological signal (art. no), *Biomed. Signal. Process. Control.* 100 (2025) 107039.
- [9] R. Pillalamarri, U. Shanmugam, A review on EEG-based multimodal learning for emotion recognition (art. no), *Artif. Intell. Rev.* 58 (5) (2025) 131.
- [10] B. Chakravarthi, S.-C. Ng, M.R. Ezilarasan, M.-F. Leung, EEG-based emotion recognition using hybrid CNN and LSTM classification (art. no), *Front. Comput. Neurosci.* 16 (2022) 1019776.
- [11] Z. Khademi, F. Ebrahimi, H. Montazery Kordy, A transfer learning-based CNN and LSTM hybrid deep learning model to classify motor imagery EEG signals (art. no), *Comput. Biol. Med.* 143 (2022) 105288.
- [12] A.M. Roy, An efficient multi-scale CNN model with intrinsic feature integration for motor imagery EEG subject classification in brain-machine interfaces (art. no), *Biomed. Signal. Process. Control.* 74 (2022) 103496.
- [13] H.L. Li, M. Ding, R.H. Zhang, C.B. Xiu, Motor imagery EEG classification algorithm based on CNN-LSTM feature fusion network (art. no), *Biomed. Signal. Process. Control.* 72 (2022) 103342.
- [14] X. Jia, Y. Song, L. Yang, L. Xie, Joint spatial and temporal features extraction for multi-classification of motor imagery EEG (art. no), *Biomed. Signal. Process. Control.* 71 (2022) 103218.
- [15] Z. Miao, M. Zhao, X. Zhang, D. Ming, LMDA-net: a lightweight multi-dimensional attention network for general EEG-based brain-computer interfaces and interpretability (art. no), *NeuroImage* 276 (2023) 120209.
- [16] F. Mattioli, C. Porcaro, G. Baldassarre, A 1D CNN for high accuracy classification and transfer learning in motor imagery EEG-based brain-computer interface (art. no), *J. Neural Eng.* 18 (6) (2022) 066053.
- [17] M.T. Sadiq, M.Z. Aziz, A. Almogren, A. Yousaf, S. Siuly, A.U. Rehman, Exploiting pretrained CNN models for the development of an EEG-based robust BCI framework (art. no), *Comput. Biol. Med.* 143 (2022) 105242.
- [18] J.R. Keough, B. Irvine, D. Kelly, et al., Fatigue in children using motor imagery and P300 brain–computer interfaces (art. no), *J. Neuroeng. Rehabil.* 21 (1) (2024) 61.
- [19] C.G. Lim, C.P. Soh, S.S.Y. Lim, D.S.S. Fung, C. Guan, T.-S. Lee, Home-based brain–computer interface attention training program for attention deficit hyperactivity disorder: a feasibility trial (art. no), *Child. Adolesc. Psychiatry Ment. Health* 17 (2023) 1.
- [20] Mirjalili, S.M. Mirjalili, A. Lewis, Grey wolf optimizer, *Adv. Eng. Softw.* 69 (2014) 46–61.
- [21] S. Saremi, S. Mirjalili, A. Lewis, Grasshopper optimisation algorithm: theory and application, *Adv. Eng. Softw.* 105 (2017) 30–47.
- [22] X. Qin, D. Xu, X. Dong, X. Cui, S. Zhang, EEG signal classification based on improved variational mode decomposition and deep forest (art. no), *Biomed. Signal. Process. Control.* 83 (2023) 104644.
- [23] X. Qiu, F. Yan, H. Liu, A difference attention ResNet-LSTM network for epileptic seizure detection using EEG signal (art. no), *Biomed. Signal. Process. Control.* 83 (2023) 104652.
- [24] Y. Song, Q. Zheng, B. Liu, X. Gao, EEG conformer: convolutional transformer for EEG decoding and visualization, *IEEE Trans. Neural Syst. Rehabil. Eng.* 31 (2023) 710–719.
- [25] C. Li, X. Huang, R. Song, R. Qian, X. Liu, X. Chen, EEG-based seizure prediction via transformer guided CNN (art. no), *Measurement* 203 (2022) 111948.
- [26] A. Iyer, S.S. Das, R. Teotia, S. Maheshwari, R.R. Sharma, CNN and LSTM based ensemble learning for human emotion recognition using EEG recordings, *Multimed. Tools Appl.* 82 (4) (2023) 4883–4896.
- [27] Dataset: <https://data.mendeley.com/datasets/kt38js3jv7/1>.
- [28] P. Singh, S. Chaudhury, B.K. Panigrahi, Hybrid MPSO-CNN: multi-level particle swarm optimized hyperparameters of convolutional neural network (art. no), *Swarm Evol. Comput.* 63 (2021) 100863.
- [29] G. Hu, Y. Guo, W. Zhao, E.H. Houssein, An adaptive snow ablation-inspired particle swarm optimization with its application in geometric optimization (art. no), *Artif. Intell. Rev.* 57 (12) (2024) 145.
- [30] B. Golanbari, A. Mardani, M. Valizadeh, N. Farhadi, Hybrid grey wolf optimizer-ANN for predicting wheel energy consumption in off-road vehicles and enhancing resource management (art. no), *J. Terra* 119 (2025) 101067.
- [31] M.K.I. Molla, A.A. Shiam, M.R. Islam, T. Tanaka, Discriminative feature selection-based motor imagery classification using EEG signal, *IEEE Access.* 8 (2020) 98255–98265.
- [32] S. Shao, G. Han, T. Wang, C. Lin, C. Song, EEG-based mental workload classification method based on hybrid deep learning model under IoT, *IEEE J. Biomed. Health Inform.* 28 (5) (2023) 2536–2546.
- [33] K.V. Kavitha, L.R. Sudha, J.S. Jayasudha, Optimizing EEG-based emotion recognition with a multi-modal ensemble approach (art. no), *Results Eng.* (2025) 104886.
- [34] W. Ding, et al., A novel data augmentation approach using mask encoding for deep learning-based asynchronous SSVEP-BCI, *IEEE Trans. Neural Syst. Rehabil. Eng.* 32 (2024) 875–886.
- [35] E. Eldele, Z. Chen, C. Liu, M. Wu, C.-K. Kwok, X. Li, C. Guan, An attention-based deep learning approach for sleep stage classification with single-channel EEG, *IEEE Trans. Neural Syst. Rehabil. Eng.* 29 (2021) 809–818.
- [36] G. Fu, et al., A temporal-spectral fused and attention-based deep model for automatic sleep staging, *IEEE Trans. Neural Syst. Rehabil. Eng.* 31 (2023) 1008–1018.
- [37] H.Y. Zhu, et al., A human-centric metaverse enabled by brain-computer interface: a survey, *IEEE Commun. Surv. Tutor.* 26 (3) (2024) 2120–2145.
- [38] H. Li, X. Li, J. d R. Millán, Noninvasive EEG-based intelligent mobile robots: a systematic review, *IEEE Trans. Autom. Sci. Eng.* 22 (2025) 6291–6315.
- [39] Y. Ding, N. Robinson, C. Tong, Q. Zeng, C. Guan, LGGNet: learning from local-global-graph representations for brain–computer interface, *IEEE Trans. Neural Netw. Learn. Syst.* (2023).
- [40] Y. Han, B. Wang, J. Luo, L. Li, X. Li, A classification method for EEG motor imagery signals based on parallel convolutional neural network (art. no), *Biomed. Signal. Process. Control.* 71 (2022) 103190.
- [41] E. Eldele, M. Ragab, Z. Chen, M. Wu, C.-K. Kwok, X. Li, Self-supervised learning for label-efficient sleep stage classification: a comprehensive evaluation, *IEEE Trans. Neural Syst. Rehabil. Eng.* 31 (2023) 1333–1342.

- [42] E. Khalili, B. Mohammadzadeh Asl, Automatic sleep stage classification using temporal convolutional neural network and new data augmentation technique from raw single-channel EEG (art. no), *Comput. Methods Prog. Biomed.* 204 (2021) 106063.
- [43] M.K. Chegeni, A. Rashno, S. Fadaei, Convolution-layer parameters optimization in convolutional neural networks (art. no), *Knowl. -Based Syst.* 261 (2023) 110210.
- [44] H. Tizhoosh, S. Rahnamayan, K. Faez, Whale optimization algorithm-based hyperparameter tuning for convolutional neural networks, *Pattern Recognit. Lett.* 157 (2022) 78–85.
- [45] A. Rashno, M. Shafipour, S. Fadaei, Particle ranking: an efficient method for multi-objective particle swarm optimization feature selection (art. no), *Knowl. -Based Syst.* 245 (2022) 108631.
- [46] M. Shafipour, A. Rashno, S. Fadaei, Particle distance rank feature selection by particle swarm optimization (art. no), *Expert. Syst. Appl.* 185 (2021) 115620.
- [47] A. Hashemi, M.B. Dowlatshahi, H. Nezamabadi-pour, An efficient pareto-based feature selection algorithm for multi-label classification, *Inf. Sci.* 581 (2021) 428–447.
- [48] F. Karimi, M.B. Dowlatshahi, A. Hashemi, SemiACO: a semi-supervised feature selection based on ant colony optimization (art. no), *Expert. Syst. Appl.* 214 (2023) 119130.
- [49] G. Muhammad, M.S. Hossain, N. Kumar, EEG-based pathology detection for home health monitoring, *IEEE J. Sel. Areas Commun.* 39 (2) (2021) 603–610.
- [50] C. Bunterngchit, T. Chearanai, Y. Bunterngchit, Towards robust cross-subject EEG-fNIRS classification: a hybrid deep learning model with optimized feature selection. *Proc. 2024 22nd Int. Conf. Res. Educ. Mechatronics (REM)*, IEEE, 2024, pp. 291–295.
- [51] J. Wang, S. Zhao, Y. Zhou, H. Jiang, Z. Yu, T. Li, S. Li, G. Pan, Narcolepsy diagnosis with sleep stage features using PSG recordings, *IEEE Trans. Neural Syst. Rehabil. Eng.* 31 (2023) 3619–3629.
- [52] J. Xie, J. Zhang, J. Sun, Z. Ma, L. Qin, G. Li, H. Zhou, Y. Zhan, A transformer-based approach combining deep learning network and spatial-temporal information for raw EEG classification, *IEEE Trans. Neural Syst. Rehabil. Eng.* 30 (2022) 2126–2136.
- [53] M. Sadat Shahabi, A. Shalbaf, A. Maghsoudi, Prediction of drug response in major depressive disorder using ensemble of transfer learning with convolutional neural network based on EEG, *Biocybern. Biomed. Eng.* 41 (3) (2021) 946–959.
- [54] S. Katsigiannis, N. Ramzan, DREAMER: a database for emotion recognition through EEG and ECG signals from wireless low-cost off-the-shelf devices, *IEEE J. Biomed. Health Inform.* 22 (1) (2018) 98–107.
- [55] Y. Xu, Y. Du, L. Li, H. Lai, J. Zou, T. Zhou, L. Xiao, L. Liu, P. Ma, AMDet: attention based multiple dimensions EEG transformer for emotion recognition, *IEEE Trans. Affect. Comput.* (2023).
- [56] S. Liu, Z. Wang, Y. An, J. Zhao, Y. Zhao, Y.-D. Zhang, EEG emotion recognition based on the attention mechanism and pre-trained convolution capsule network (art. no), *Knowl. -Based Syst.* 265 (2023) 110372.
- [57] J.R. Challapalli, N. Devarakonda, A novel approach for optimization of convolution neural network with hybrid particle swarm and grey wolf algorithm for classification of indian classical dances, *Knowl. Inf. Syst.* 64 (2022) 2411–2434.
- [58] V. Palle, A. Kumari, V. Shirodkar, et al., Optimized CNN-transformer model for motor imagery EEG decoding: a gray wolf optimization approach, *Int. J. Data Sci. Anal.* 22 (2026) 149.
- [59] J. Sabio, N.S. Williams, G.M. McArthur, N.A. Badcock, A scoping review on the use of consumer-grade EEG devices for research, *PLOS ONE* 19 (3) (2024) e0291186.
- [60] K. Devarajan, S. Ponnann, S. Perumal, Hybrid CNN-transformer architecture for enhanced EEG-based emotion recognition: capturing local and global dependencies with self-attention mechanisms, *Discov. Comput.* 28 (2025) 87, <https://doi.org/10.1007/s10791-025-09596-0>.
- [61] P. Nandihal, S.B. Manoj Kumar, J. Rajeshwari, M.S. Nagesh, K.L. Santhosh Kumar, B. Madhu, K.N. Bharath, Hybrid deep learning models for accurate EEG-based cognitive state classification, *SN Comput. Sci.* 6 (2025) 806, <https://doi.org/10.1007/s42979-025-04338-x>.
- [62] J. Han, H. Su, L. Wang, L. Zhang, X. Kang, An EEG-based spatial-temporal hybrid architecture for cognitive load detection. *Engineering Applications of Neural Networks (EANN 2025)*, *Communications in Computer and Information Science*, Springer, Cham, 2025, https://doi.org/10.1007/978-981-96-7033-8_9.
- [63] A. Alghamdi, M. Nilashi, R. Ali Abumalloh, et al., Accuracy improvements for electroencephalography (EEG) eye state classification using extreme gradient boosting and cluster ensembles, *EURASIP J. Adv. Signal. Process.* 2026 (2026) 20, <https://doi.org/10.1186/s13634-025-01290-z>.
- [64] G.S. Jebaraj, K. Elango, EEG seizure classification with temporal spiking neural networks and mutual information-based feature selection, *J. Eng. Appl. Sci.* 72 (2025) 218, <https://doi.org/10.1186/s44147-025-00796-5>.