International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 43, No. 3, June, 2026, pp. 1213-1226
Journal homepage: http://iieta.org/journals/ts

NeuroLightNet: A Lightweight Attention-Driven Deep Learning Framework for Adaptive Electroencephalogram Signal Interpretation in Brain–Computer Interfaces

Vanidhasri Rajadura*, Kathirvel Ayyaswamy

Department of Computer Science and Engineering, Saveetha Engineering College, Chennai 602105, India

Corresponding Author Email: vanidhaassistantprofessor@gmail.com

https://doi.org/10.18280/ts.430311

Received: 24 February 2026
Revised: 26 April 2026
Accepted: 7 May 2026
Available online: 30 June 2026

Keywords:
Electroencephalogram, Brain–Computer Interface, NeuroLightNet, NeuroCRAMNet, attention mechanism, continual learning, explainable AI, edge deployment

ABSTRACT

Electroencephalogram (EEG)-based Brain–Computer Interfaces (BCIs) have garnered significant attention in neurorehabilitation and assistive neurotechnologies. However, decoding EEG signals with high accuracy, low latency, and strong adaptability remains a challenging task due to their noisy and non-stationary nature. This work introduces NeuroLightNet, a lightweight, attention-enhanced deep learning framework that addresses these challenges through its core architecture, NeuroCRAMNet (Compressed Recurrent Attention Model). The model integrates depthwise-separable convolutions for spatial efficiency, gated recurrent units (GRUs) for temporal encoding, and a dual-stage attention mechanism that emphasizes relevant EEG patterns in both spatial and temporal dimensions. Continual learning via elastic weight consolidation (EWC) allows the model to adapt across subjects and sessions, while integrated SHapley Additive exPlanations (SHAP) and Gradient-weighted Class Activation Mapping (Grad-CAM) techniques provide visual explanations of learned features. The framework was validated on BCI Competition IV-2a and PhysioNet Motor Imagery EEG datasets. NeuroCRAMNet achieved a top classification accuracy of 93.63%, outperforming CNN+LSTM+GWO (91.93%) and Ensemble DL (92.25%). The model also demonstrated a reduced inference time of 16.5 ms/sample and a minimal memory footprint of 7.5 MB, making it suitable for edge-device deployment. Subject-wise analysis confirmed consistent performance across individuals, and attention visualizations reinforced model interpretability. These results establish NeuroLightNet as a scalable, interpretable, and real-time solution for modern BCI applications.

1. INTRODUCTION

Brain–Computer Interfaces (BCIs) have emerged as a transformative technology that bridges the neural activity of the human brain with external devices, offering significant promise in medical rehabilitation, assistive systems, and neuroadaptive control [1]. Among various neural recording methods, the Electroencephalogram (EEG) has become a preferred modality due to its non-invasive acquisition, high temporal resolution, affordability, and ease of use. EEG-based BCIs are particularly valuable in applications such as motor imagery recognition, neurorehabilitation, communication aids for locked-in patients, and smart assistive robotics. However, decoding EEG signals in real time remains a challenging task, primarily due to the noisy, non-stationary, and high-dimensional nature of EEG data [2].

In recent years, deep learning has shown remarkable capabilities in capturing the complex, non-linear spatiotemporal dynamics present in EEG signals. Convolutional Neural Networks (CNNs) are adept at extracting spatial patterns from multi-channel EEG inputs, while Recurrent Neural Networks (RNNs), especially gated recurrent units (GRUs) and Long Short-Term Memory (LSTM) networks, have demonstrated effectiveness in modeling temporal dependencies. Hybrid architectures such as CNN-LSTM models have outperformed classical machine learning pipelines in various motor imagery classification tasks [3]. Furthermore, optimization algorithms like the Grey Wolf Optimizer (GWO) have been applied to fine-tune neural network parameters, thereby enhancing model convergence and performance. These advancements, although impactful, often result in large, computationally expensive models unsuitable for deployment on low-power devices. Moreover, many of these models function as black boxes, offering limited interpretability and adaptability across sessions or subjects. The majority of existing research in EEG signal classification tends to focus on maximizing accuracy without adequately addressing other critical requirements, such as adaptability, interpretability, and computational efficiency. For instance, ensemble deep learning models have achieved impressive classification scores, but they often require significant computational resources and memory. This limitation poses a barrier to real-world deployment in wearable or portable BCI systems [4]. Similarly, CNN-LSTM-GWO models, while efficient in capturing spatiotemporal features, often lack mechanisms for handling subject variability or providing transparency into decision-making processes. These gaps highlight the need for a unified solution that combines accuracy with model compactness, adaptive learning capabilities, and explainable outputs [5]. An emerging direction in EEG decoding involves the integration of attention mechanisms, which enable models to dynamically prioritize informative segments of the EEG signal in space and time. Attention-based networks not only improve classification accuracy but also offer a form of implicit interpretability by indicating regions of interest. However, many attention-based models remain computationally demanding, making them impractical for real-time, on-device inference. Parallel to this, the concept of continual learning—where a model incrementally adapts to new data without forgetting previously acquired knowledge—is gaining traction, particularly in the context of inter-session and inter-subject generalization. Unfortunately, this paradigm has not been widely adopted in EEG-based BCI research [6].

Another essential yet underexplored area is explainable AI (XAI), which seeks to demystify the decision-making process of deep learning models. In safety-critical applications such as clinical diagnosis or assistive robotics, the ability to interpret model predictions is crucial for user trust and regulatory compliance [7]. Techniques such as SHapley Additive exPlanations (SHAP) and Gradient-weighted Class Activation Mapping (Grad-CAM) have proven effective in providing visual insights into model behaviour [8]. However, very few EEG classification frameworks integrate these techniques directly into their architectures. In light of these challenges, a need arises for a lightweight, attention-driven deep learning architecture that can operate efficiently in real time, adapt continuously across sessions and subjects, and provide interpretable predictions. Such a model would greatly enhance the deployment potential of EEG-based BCIs in real-world scenarios, including wearable neurotechnologies, smart prosthetics, and cognitive monitoring systems [9, 10]. The present study addresses this requirement by introducing NeuroLightNet, a comprehensive deep learning framework for EEG signal interpretation. At the heart of this framework lies NeuroCRAMNet (Compressed Recurrent Attention Model), an architecture designed to extract robust spatiotemporal features while minimizing computational load. The model employs depthwise-separable convolutions to encode spatial information efficiently and utilizes GRU units to track temporal patterns in EEG signals. A novel dual-stage attention mechanism is integrated to focus selectively on salient EEG channels and time windows, enhancing both accuracy and model transparency. To support lifelong learning, continual adaptation is facilitated through elastic weight consolidation (EWC), allowing the model to retain previous knowledge while learning from new sessions or users. Additionally, built-in explainability modules based on SHAP and Grad-CAM provide visualizations of key neural patterns contributing to classification outcomes [11]. The proposed framework is evaluated on two widely used benchmark datasets, BCI Competition IV-2a and the PhysioNet Motor Imagery EEG dataset, to ensure robust validation across different experimental protocols and subject populations. These datasets offer complementary characteristics: BCI IV-2a provides session-based recordings of four-class motor imagery from nine subjects, while PhysioNet includes a larger and more diverse subject pool with varying motor execution and imagery tasks. The combination of these datasets allows for a comprehensive assessment of the model's generalizability and scalability. Figure 1 depicts the overall architecture diagram of the proposed model.

The main contributions of this study include:

A novel deep learning framework, NeuroLightNet, is introduced for EEG signal interpretation in BCI applications, prioritizing both accuracy and deployment efficiency.

The proposed architecture, NeuroCRAMNet, combines depthwise-separable convolutions for spatial encoding, GRU units for temporal dynamics, and a dual-stage attention mechanism to focus on relevant EEG patterns across channels and time.

An integrated, continual learning strategy using EWC is implemented, enabling the model to retain performance during incremental learning across sessions and subjects.

Explainable AI techniques such as SHAP and Grad-CAM are embedded into the model, offering interpretable visual insights into the decision-making process.

The framework is designed for real-time edge deployment, with significant reductions in model size and latency, validated through performance on benchmark datasets including BCI Competition IV-2a and PhysioNet Motor Imagery.

Comprehensive experimental results demonstrate that NeuroLightNet outperforms existing models in accuracy, efficiency, and adaptability while maintaining interpretability.

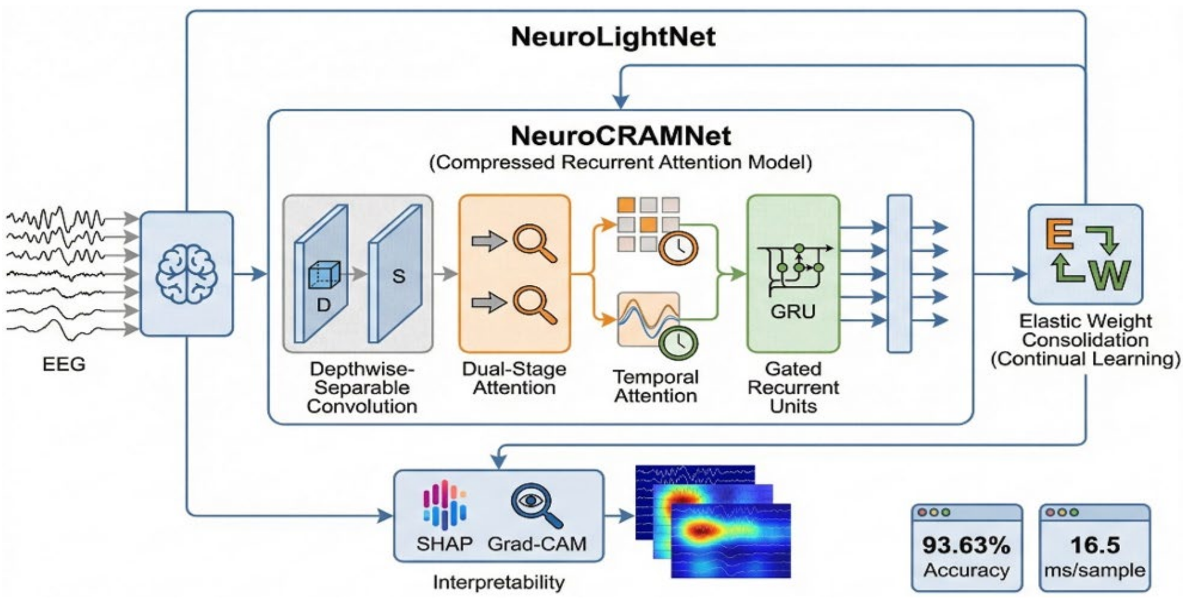

Figure 1. Overall architecture diagram of the proposed model

The remainder of the paper is organized as follows: Section 2 presents a comprehensive review of related works, highlighting recent developments in deep learning-based EEG classification, attention mechanisms, continual learning, and explainability in BCIs. Section 3 details the proposed methodology, elaborating on the design and components of the NeuroLightNet framework, including the NeuroCRAMNet architecture and the integration of attention and continual learning modules. Section 4 describes the experimental setup, including dataset specifications, preprocessing techniques, evaluation metrics, and deployment considerations. Section 5 discusses the results and provides a comparative analysis of the proposed method against baseline models using multiple evaluation criteria and visualizations. Section 6 concludes the paper with a summary of findings and outlines directions for future research, including multi-modal extensions and adaptive personalization strategies. In contrast to traditional CNN-GRU or attention-based EEG classification systems, the NeuroLightNet framework suggests a single architecture that can be used to tackle the idea of accuracy, flexibility, interpretability, and the efficiency of its deployment. What is new about this work is not in the separate components, but in the synergistic combination and task-specific optimization of these components to EEG-based BCI applications. Particularly, NeuroCrackMe includes a two-stage spatial-temporal attention system to improve the discriminative features learning, implements the EWC method to continually learn across sessions and subjects, and introduces explainable AI methods, such as SHAP and Grad-Cam, directly into the model. Moreover, the framework is specifically designed to deploy depthwise separable convolutions, quantization, and Open Neural Network Exchange (ONNX)-based implementation. This is a multi-objective architecture that makes NeuroLightNet different than current CNN-GRU and attention-based EEG models, a full-fledged and scalable solution to adaptive BCI systems. To overcome the constraints of the currently available EEG-based BCI models, this paper will present NeuroLightNet, a lightweight and versatile framework of the deep learning network that is specifically built to interpret EEG signals effectively. The suggested NeuroCRAMNet model combines depthwise-separable convolution (that enables the efficient extraction of spatial features), GRU-based temporal modeling, and dual-stage spatial-temporal attention to improve discriminative learning. The framework, further, has EWC of session- and subject-wise continuous learning and embedded explainable AI methods (SHAP and Grad-CAM) to enhance model interpretability. This single design allows high classification accuracy, computational efficiency, and possible deployment in real-time applications, unlike the current EEG classification frameworks.

2. RELATED WORKS

The current body of research in EEG-based BCI systems can be subdivided into four groups, namely: (1) CNN-based models that operate in the context of spatial features, (2) Hybrid CNN-RNN models that address the spatial-temporal characteristics, (3) Attention-based and Transformer-based models that optimize the features representation by using an adaptive weighting process, and (4) Lightweight models and edge-oriented models which have been developed to work in the context of resource-constrained systems. Table 1 gives a comparative summary of the representative studies in these categories. The presented NeuroLightNet framework incorporates the main strengths of these solutions and combats their weaknesses with a single, lightweight, and adaptive framework. The integration of deep learning into EEG analysis has garnered increasing attention, with several studies exploring innovative architectures and applications. A recent special issue by Prevete et al. [12] emphasized the growing importance of deep learning methodologies in EEG signal interpretation, particularly in the context of cognitive and clinical neuroscience. Liu et al. [13] proposed a multi-scale deep CNN for motor imagery classification, achieving perfect accuracy on the EEG motor activity dataset from National University of Sciences and Technology (NUST) Pakistan. Their approach, termed Deep Motor Features (DeepMF), significantly reduced training time, enhancing real-time applicability in BCI systems. Miao et al. [14] introduced LMDA-Net, a lightweight attention-based model that incorporates channel and depth attention mechanisms. Validated across multiple BCI datasets, including BCI4-2A, BCI4-2B, and BCI3-4a, LMDA-Net consistently outperformed traditional models in terms of accuracy within 300 epochs, while maintaining strong interpretability. To improve spatiotemporal learning, Wang and Dong [15] developed a parallel Convolutional Recurrent Neural Network (CRNN) framework combined with self-attention mechanisms. Evaluated on the BCI Competition IV dataset, the model surpassed benchmark networks like EEGNet and Shallow ConvNet by notable margins, achieving 77.3% average accuracy. Huang et al. [16] presented CAT-CNN-TRFM, a hybrid of CNN and Transformer modules guided by channel attention. Their method demonstrated robust performance on the BCI IV2a dataset, achieving an average accuracy of 81.25% by capturing both localized and global EEG features. Myrden and Chau [17] explored adaptive BCI systems that respond to psychological state changes. Using regression models to predict mental states from EEG during maze navigation tasks, they observed improved classification accuracy in adaptive setups, although results varied across individuals. Durand [18] has highlighted the need for neural interface systems to effectively process signals and integrate them into adaptable systems, fostering the creation of sophisticated computational framework. Building upon these ideas, recent research has concentrated on improving the security of the data by using methods like federated learning, homomorphic encryption and differential privacy to allow secure computation without revealing any raw data. All of these methods contribute to building privacy preserving intelligent systems relevant to cloud, IoT and AI-based applications.

Alizadeh et al. [19] used multivariate mode decomposition (MVMD) and wavelet-based transformations to convert EEG signals into 2D images, which were then classified using CNNs. LeNet and AlexNet architectures achieved accuracies of 95.33% and 93.66%, respectively, on motor imagery datasets from BCI Competition IV. Miao et al. [20] and Ma et al. [21] proposed an attention-based CNN with temporal information fusion for decoding EEG motor imagery signals. Their model employed data augmentation and achieved 85.03% accuracy on BCIC-IV-2a, outperforming contemporary benchmarks. Amin et al. [22] integrated multiple CNNs with autoencoder cross-encoding, demonstrating more than a 10% performance boost over state-of-the-art methods in cognitive healthcare applications using

EEG. Bennett et al. [23] introduced a perturbation-based interpretability framework to identify the neurophysiological basis of EEG features. Validated on simulated data, the method provided insights into contributing electrodes and cortical sources. Hasan [24] proposed adaptive algorithms using Conditional Random Fields and Hidden Markov Models to tackle EEG non-stationarity. The models effectively exploited temporal patterns, leading to performance improvements in self-paced BCI systems. Kumar and Michmizos [25] designed a topography-preserving 3D-CNN using a custom in-house dataset. Their model, which preserved spatial and temporal EEG structures, achieved over 82% accuracy in decoding complex motor components and outperformed 2D-CNN architectures. Sameh et al. [26] utilized hierarchical deep learning models to detect four human daily activities (sleep, food, toilet, drink) from EEG signals, integrating facial detection for enhanced reliability. Liu et al. [27] developed a lightweight model achieving over 98% accuracy for classifying EEG responses to visual stimuli such as faces and cars, using data from the ERP core. Gunda et al. [28] implemented lightweight attention mechanisms for emotion recognition in EEG-based BCI systems, addressing both model efficiency and interpretability. Amin et al. [29] combined Attention-Inception CNNs with LSTM networks to improve classification accuracy for motor imagery. Using BCI Competition IV-2a and high gamma datasets, they reported accuracies of 82.8% and 97.1%, respectively. Kumar et al. [30] adopted a deep learning pipeline based on Common Spatial Pattern (CSP) and deep neural networks (DNN) for motor imagery classification. Their model significantly reduced classification error rates across public benchmark datasets.

Table 1. Comparative summary of deep learning approaches for Electroencephalogram (EEG)-based Brain–Computer Interface (BCI)

S. No.	Author(s)	Model/Technique	Dataset(s)	Key Features/Mechanisms	Performance/Accuracy
1	Liu et al. [13]	DeepMF (multi-scale CNN)	NUST BCI motor activity	Imagined motor task classification, deep features	100% accuracy, 50% training time reduction
2	Miao et al. [14]	LMDA-Net (lightweight attention)	BCI4-2A, 2B, 3-4a, Kaggle-ERN	Channel + depth attention, multidimensional feature extraction	Best accuracy across all datasets within 300 epochs
3	Wang and Dong [15]	Parallel CRNN + Self-Attention	BCI Competition IV	Spatial & temporal attention, CRNN fusion	77.3%, +5.7% vs Shallow ConvNet
4	Huang et al. [16]	CAT-CNN-TRFM (Channel attention + transformer)	BCI IV2a	Global dependency modeling, spatiotemporal attention	81.25% average accuracy
5	Alizadeh et al. [19]	CNN + self-attention with temporal fusion	BCIC-IV-2a, 2b	Temporal multi-modal learning, data augmentation	85.03% (IV-2a)
6	Ma et al. [21]	LeNet, AlexNet on 2D CWT images	BCI Competition IV	MVMD decomposition, wavelet transform	95.33% (LeNet), 93.66% (AlexNet)
7	Amin et al. [22]	Multi-CNN + autoencoder fusion	Cognitive healthcare EEG	Autoencoder-based cross-encoding	>10% performance gain over SOTA
8	Kumar and Michmizos [25]	3D-CNN (Neuro-interpretable)	In-house motor dataset	Topography preservation, robust to intra-subject variance	79.81%–82.00%
9	Sameh et al. [26]	Deep learning-based activity classification	Custom EEG (4 activities)	Hierarchical learning + face detection	Accurate multi-activity detection
10	Liu et al. [27]	Lightweight DL model	ERP Core	Visual EEG stimuli (face, car), EEG classification	98.13% (face), 97.81% (car)
11	Gunda et al. [28]	Lightweight attention mechanism	Emotion-based EEG BCI	Efficient EEG emotion classification	Not specified (focus: interpretability)
12	Amin et al. [29]	Attention-inception CNN + LSTM	BCI IV 2a, High Gamma	Spatiotemporal learning, temporal dynamics	82.8% (IV 2a), 97.1% (High Gamma)
13	Kumar et al. [30]	CSP + DNN	BCI Competition III (IVa)	CSP feature extraction + DNN classification	Significant error reduction

Note: CNN = Convolutional Neural Network; CRNN = Convolutional Recurrent Neural Network; MVMD = multivariate mode decomposition; DeepMF = Deep Motor Features; DL = deep learning; CSP = Common Spatial Pattern; LSTM = Long Short-Term Memory; DNN = deep neural networks.

3. METHODS AND MATERIALS

The proposed work introduces NeuroLightNet, a lightweight deep learning framework for accurate and adaptive EEG signal interpretation in BCIs. It leverages NeuroCRAMNet, which integrates depthwise-separable convolutions, GRU-based temporal modeling, and dual-stage attention mechanisms for efficient spatiotemporal feature extraction. A continual learning module based on EWC ensures adaptability across sessions and subjects. Explainable AI techniques (SHAP, Grad-CAM) are embedded for model transparency and trust. The framework is optimized for real-time edge deployment, achieving high accuracy with minimal latency and memory footprint.

3.1 Dataset description

Two benchmark datasets were employed to validate the proposed NeuroLightNet framework. The BCI Competition IV-2a dataset comprises EEG recordings from 9 subjects performing four-class motor imagery tasks (left hand, right hand, foot, tongue) using 22 channels at 250 Hz. Each subject participated in two sessions recorded on different days. The PhysioNet Motor Imagery EEG dataset includes recordings from 109 subjects executing or imagining hand and foot movements across 64 EEG channels at 160 Hz. This dataset provides a broader demographic with varying session protocols. Together, these datasets enable rigorous evaluation across spatial, temporal, and inter-subject variations in EEG signals, as shown in Table 2.

Table 2. Dataset parameters of the Electroencephalogram (EEG) datasets used

Parameter	BCI Competition IV-2a	PhysioNet Motor Imagery EEG
Source	BCI Competition IV, Dataset 2a	PhysioNet EEG Motor Movement/Imagery
Number of subjects	9	109
Motor imagery classes	4 (Left Hand, Right Hand, Foot, Tongue)	2 (Fist, Foot – Imagined & Executed)
Number of channels	22 EEG	64 EEG
Sampling rate	250 Hz	160 Hz
Sessions per subject	2 (recorded on different days)	1–2 sessions per subject
Trials per session	288 trials (72 per class)	Varies by subject
Electrode montage	International 10–20 System	10–10 System
File format	.gdf	.edf
Filtering applied	0.5–100 Hz bandpass + 50 Hz notch filter	0.1–100 Hz bandpass, artifact rejection applied
Application focus	Controlled motor imagery	Mixed imagery and motor execution tasks

Note: BCI = Brain–Computer Interface.

3.2 Preprocessing

Preprocessing is a crucial step in EEG-based BCI systems to enhance signal quality and ensure meaningful feature extraction. Raw EEG signals are often contaminated with noise, artifacts, and irrelevant frequency components, which can impair model performance. The goal of preprocessing is to clean, standardize, and structure the data for optimal input to the deep learning framework. The following steps outline a rigorous pipeline tailored to the BCI Competition IV-2a and PhysioNet Motor Imagery, and it is shown in Figure 2.

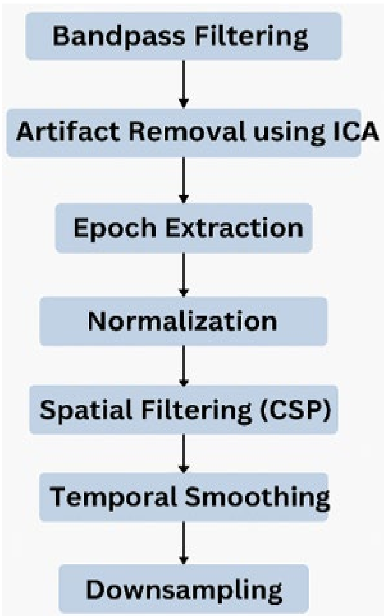

Figure 2. Preprocessing techniques

3.2.1 Bandpass filtering

To isolate relevant EEG frequency bands (e.g., μ: 8–12 Hz and β: 13–30 Hz), a Butterworth bandpass filter is applied. Let $x(t)$ be the raw EEG signal. The bandpass filtered signal $x_f(t)$ is given by:

$$x_f(t) = F_{BP}\{x(t)\} \quad (1)$$

where, F_{BP} represents the bandpass filter operator. For an n-order Butterworth filter:

$$H(\omega) = \frac{1}{\sqrt{1 + (\frac{\omega}{\omega_c})^{2n}}} \quad (2)$$

where, ω_c is the cutoff frequency, and n is the filter order (typically 4 or 5).

3.2.2 Artifact removal using independent component analysis

To eliminate ocular and muscular artifacts, independent component analysis (ICA) decomposes the EEG into statistically independent components.

Let $X \in R^{C\times T}$ be the recorded EEG matrix with C channels and T time samples. ICA attempts to find an unmixing matrix W such that:

$$S = WX \quad (3)$$

where, S contains statistically independent sources. Artifact components are identified and zeroed out, then the cleaned signal is reconstructed:

$$X_{clean} = W^{-1}S_{clean} \quad (4)$$

3.2.3 Epoch extraction

EEG data are segmented into fixed-length epochs around stimulus onset events. Let the onset time be t_s, the pre-stimulus window T_p, and the post-stimulus window T_q. The epoch $E_i(t)$ for the i_{th} trial is:

$$E_i(t) = x(t), t \in [t_s - T_p, t_s + T_q] \quad (5)$$

Typically, T_p = 0.5 sec and T_q = 2 sec for motor imagery.

3.2.4 Normalization (Z-score scaling)

To ensure uniform channel contribution and stabilize training, z-score normalization is applied across each channel. For a channel signal x_c:

$$x_c^{norm} = \frac{x_c - \mu_c}{\sigma_c} \quad (6)$$

where, μ_c and σ_c are the mean and standard deviation of channel c.

3.2.5 Spatial filtering using Common Spatial Pattern

CSP enhances class-discriminative spatial patterns. Let Σ_1 and Σ_2 be the covariance matrices for two classes. The CSP projection matrix W_{CSP} solves:

$$W_{CSP} = arg \max_{w} \frac{W^T \sum_1 W}{W^T(\sum_1 + \sum_1)\, W} \quad (7)$$

The transformed signal:
$Z = W_{CSP}^T X$ selects components with maximal variance

differences between classes.

3.2.6 Temporal Smoothing (Moving average or Savitzky-Golay filter)

To reduce temporal noise and highlight meaningful trends in EEG sequences, a smoothing operator S is applied.

$$\tilde{x}(t) = \frac{1}{2k+1}\sum_{i=-k}^{k} x(t+i) \quad (8)$$

where, k is the window size (e.g., k = 5).

$\tilde{x}(t) = \sum_{i=-k}^{k} c_i\, x(t+i)$ where c_i is the polynomial fit coefficients for local smoothing.

3.2.7 Downsampling

To reduce computational load, the signal is downsampled from the original sampling rate f_s to a lower rate f_d.

$$x_d[n] = x[n \cdot D], D = \frac{f_s}{f_d} \quad (9)$$

This preserves essential dynamics while reducing sequence length.

3.3 Proposed methodology

The core of the NeuroLightNet framework is a novel architecture called NeuroCRAMNet (Compressed Recurrent Attention Model), designed to efficiently extract spatial-temporal features from EEG signals while enabling attention-driven focus, continual adaptability, and edge-ready deployment. The architecture integrates four primary functional modules: compressed spatial encoders, temporal attention blocks, continual learning units, and explainable AI heads. Each of these components is developed with considerations of low-latency inference, robustness to inter-session variability, and interpretability—features critical to the deployment of EEG-based BCIs in real-world applications. In order to make it reproducible and allow future research, the detailed setting of the suggested NeuroCRAMNet architecture is displayed in this subsection.

The spatial feature extraction module consists of two depthwise-separable convolutional layers. The initial layer consists of 32 filters with the usage of (1 × 3) kernel size, batch normalization, and ReLU activation. The second layer involves 64 filters with a kernel (1 × 5) so as to obtain wider spatial dependencies between EEG channels. The layers are both followed by a dropout rate of 0.3 in order to avoid overfitting. The temporal modelling part makes use of a GRU layer with 128 hidden units, which allows learning sequential dependencies in EEGs with computational efficiency. GRU output is then subjected to a temporal self-attention (TSA) mechanism comprising 4 attention heads, whose key/query dimension is 32. In the case of spatial attention, a multi-head attention module consisting of 4 attention heads is utilized following the convolutional encoder to increase feature representations on a channel basis. There are residual connections and layer normalization, which are added to stabilize training and enhance convergence. Elements of the classification phase are a fully connected, dense layer of 256 neurons and a multi-class EEG signal classification using a softmax activation function. In order to increase generalization performance, the L2 regularization of a 1e-4 coefficient is added to all trainable parameters. The amount of all trainable parameters of the proposed NeuroCRAMNet model is about 1.18 million, which is much less than the traditional CNN-GRU architectures and yet generates better classification results. The most important architectural parameters of the proposed model are summarized in Table 3.

Table 3. Dataset summary

Component	Configuration
Input shape	C × T (EEG channels × time samples)
Depthwise conv layer 1	32 filters, kernel size (1 × 3), ReLU
Depthwise conv layer 2	64 filters, kernel size (1 × 5), ReLU
Spatial attention	Multi-head attention (4 heads)
GRU layer	128 hidden units
Temporal attention	Multi-head self-attention (4 heads)
Fully connected layer	256 neurons
Dropout	0.3
Regularization	L2 (1e-4)
Output layer	Softmax
Total parameters	~1.18 Million

3.3.1 Spatial feature encoding via depthwise-separable convolution and channel attention

To capture spatial patterns across EEG channels, the input EEG matrix $X \in RC \times T$, where C is the number of channels and T is the number of temporal samples, is first reshaped to fit a convolutional input format. Depthwise-separable convolution is employed as an efficient alternative to traditional convolution. This operation decomposes standard convolution into two steps: depthwise and pointwise convolution, and it is shown in Figure 3. In the depthwise step, a single convolutional filter is applied per input channel, defined as:

$$X_d(c,t) = \sum_{k=1}^{K} w_c(k) \cdot X(c, t+k) \quad (10)$$

where, w_c represents the depthwise kernel for the cth channel, and K is the kernel size. The resulting output is then passed through a pointwise convolution, which performs a 1 × 11 convolution across all channels:

$$X_p(t) = \sum_{c=1}^{C} V_c \cdot X_d(c,t), \quad (11)$$

where, V_c are the learned weights for inter-channel integration. This design drastically reduces computational complexity from $O(C + K)$, making it suitable for edge deployment.

To enhance the spatial discriminative capability, a lightweight multi-head attention mechanism is introduced after the convolution. The attention module learns to focus on informative channels by computing attention scores using the scaled dot-product:

$$Attention(Q,K,V) = softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (12)$$

where, the query Q, key K, and value V matrices are derived from linear projections of the input: $Q = XW^Q$, $K = XW^K$, and $V = XW^V$. The attention heads are concatenated and added to the original input via a residual connection, ensuring gradient flow and learning stability. Spatial Feature Encoding via Depthwise-Separable Convolution and Channel Attention is shown in Figure 3.

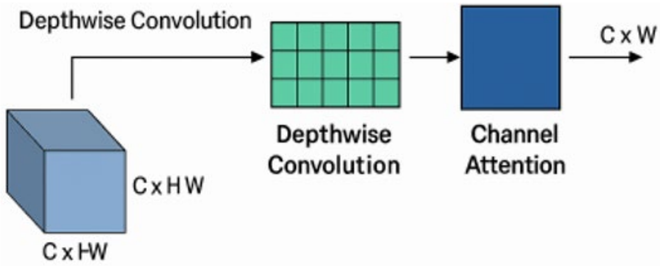

Figure 3. Spatial feature encoding via depthwise-separable convolution and channel attention

3.3.2 Temporal modeling using gated recurrent units and self-attention

Following spatial encoding, the signal is processed through a temporal learning unit to extract meaningful sequential dependencies. A GRU is used for this purpose due to its lower parameter count and competitive performance in time-series modelling, and it is shown in Figure 4.

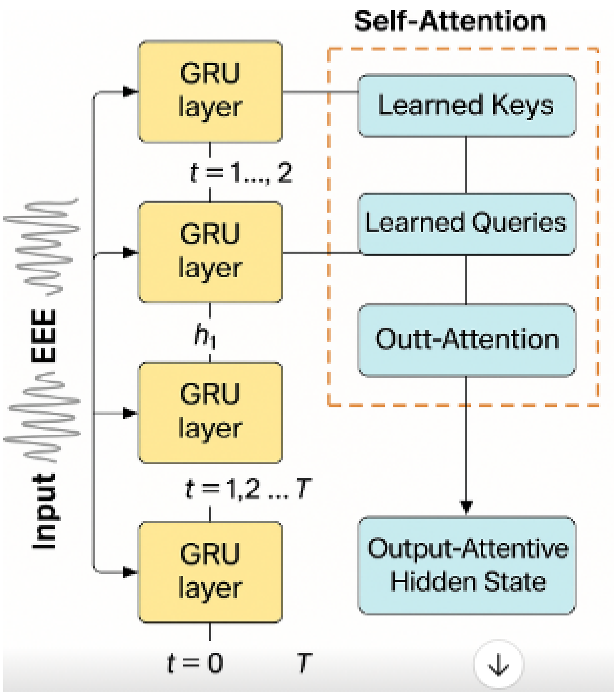

Figure 4. Temporal modeling using gated recurrent units (GRUs) and self-attention

The GRU computes its hidden state at time t using the following update equations:

$$z_t = \sigma(W_z x_t + U_z h_{t-1}), r_t = \sigma(W_r x_t + U_r h_{t-1}), \quad (13)$$

$$\tilde{h}_t = \tanh\left(W_h x_t + U_h (r_t \odot h_{t-1})\right) \quad (14)$$

$$h_t = (1 - z_t) \odot h_{t-1} + z_t \odot \tilde{h}_t \quad (15)$$

where, z_t and r_t represent the update and reset gates, respectively, and ⊙ denotes the Hadamard product.

To further refine the temporal features, a self-attention mechanism is applied across the hidden states generated by the GRU. Given a sequence of hidden states $H = [h_1, h_2, ..., h_T]$ attention weights α_t are computed as:

$$\alpha_t = \frac{exp(u^T tan\,h(W_a h_t))}{\sum_{i=1}^{T} exp(u^T tan\,h(W_a h_i))} \quad (16)$$

and the context vector is formed as a weighted sum of hidden states:

$$h_{context} = \sum_{t=1}^{T} \alpha_t h_t. \quad (17)$$

This process allows the network to dynamically focus on salient temporal intervals, such as P300 spikes or motor imagery activation periods.

3.3.3 Continual learning through elastic weight consolidation

To address the problem of inter-session and inter-subject variability, continual learning is embedded within the training process using EWC. This regularization-based approach prevents catastrophic forgetting when adapting the model to new data distributions, as shown in Figure 5.

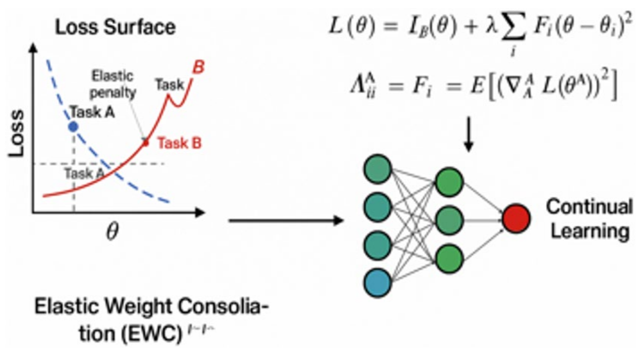

Figure 5. Continual learning through elastic weight consolidation (EWC)

After training on a previous task A, a diagonal approximation of the Fisher Information Matrix FFF is computed to estimate the importance of each parameter θ_i. The model is then fine-tuned on the new task BBB with the loss function augmented as:

$$L_{total} = L_B(\theta) + \frac{\lambda}{2} \sum_i F_i (\theta_i - \theta_i^*)^2 \quad (18)$$

where, θ_i^* denotes the learned parameters from task A, and λ controls the regularization strength. The Fisher Information is estimated using:

$$F_i = E\left[\left(\frac{\partial \log_p(y|x,\theta)}{\partial\theta_i}\right)\right] \quad (19)$$

Approximated over the training data using the model's output probabilities. This mechanism enables the network to learn from new subjects or sessions while retaining prior knowledge.

3.3.4 Explainability integration using SHapley Additive exPlanations and Gradient-weighted Class Activation Mapping

To enhance model interpretability and user trust, explainable AI components are integrated into the output layer. SHAP (Shapley Additive Explanations) is used to compute the contribution of each input feature to the final decision, and it is shown in Figure 6. The SHAP value ϕ_i for the i_{th} feature is defined as:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S \cup \{i\}) - f(S)] \tag{20}$$

where, F is the feature set, and $f(S)$ is the model's output when only features in subset S are present. Additionally, Grad-CAM is applied to identify which spatial-temporal regions of the input most strongly influenced the model's prediction. Grad-CAM computes the class-discriminative localization map as:

$$\alpha_k^c = \frac{1}{z} \sum_i \sum_j \frac{\partial y^c}{\partial A_{ij}^k} \tag{21}$$

$$GradCAM^c = ReLU\left(\sum_k \alpha_k^c A^k\right) \tag{22}$$

where, A^k is the kth activation map, y^c is the class score, and Z is the normalization factor. These tools provide insight into which regions of the EEG signal (in space and time) were critical for each classification decision.

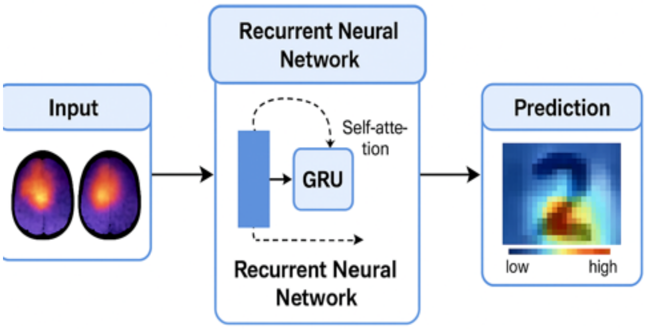

Figure 6. Explainability integration using SHapley Additive exPlanations (SHAP) and Gradient-weighted Class Activation Mapping (Grad-CAM)

3.3.5 Edge deployment via quantization and Open Neural Network Exchange optimization

One of the distinguishing features of NeuroCRAMNet is its readiness for deployment in real-time, low-power environments such as embedded systems, wearable devices, or neuroprosthetic controllers. To meet the constraints of latency, memory, and energy efficiency, we implement post-training model compression techniques, including parameter pruning, weight quantization, and ONNX conversion for hardware compatibility.

The primary technique employed is 8-bit integer quantization, which reduces the model's floating-point precision to a lower bit-width representation without significantly affecting accuracy. Consider a 32-bit floating-point weight θ_{fp32}; quantization maps this to an 8-bit integer θ_{int8} using a uniform scaling factor s, defined as:

$$\theta_{int8} = round\left(\frac{\theta_{fp32}}{s}\right), \text{with } s = \frac{max(\theta_{fp32}) - min(\theta_{fp32})}{2^n - 1} \tag{23}$$

where, $n = 8$ and s serves to preserve the dynamic range of the original weights. During inference, dequantization is performed to approximate the original weights:

$$\theta_{fp32} \approx s \cdot \theta_{int8}. \tag{24}$$

This operation significantly reduces the memory footprint and accelerates matrix multiplications on integer-only hardware accelerators such as ARM Cortex-M microcontrollers or edge AI chips like NVIDIA Jetson Nano and Intel Movidius.

In addition to quantization, model pruning is employed to remove redundant or low-importance weights. This is guided by magnitude-based criteria where weights θ_i with absolute values below a threshold δ are zeroed out:

$$\theta_i' = \begin{cases} 0, & if\ |\theta_i| < \delta \\ \theta_i, & Otherwise \end{cases} \tag{25}$$

Algorithm: NeuroCRAMNet EEG Signal Classification

Input: Raw EEG signal matrix X (channels × time)
Output: Predicted class label ŷ
1: Preprocess X using:
  a. Bandpass filtering
  b. Artifact removal (ICA)
  c. Epoch segmentation
  d. Normalization
  e. Spatial filtering (CSP)
  f. Temporal smoothing
  g. Downsampling
2: Apply Compressed Spatial Encoder:
  a. Reshape X for convolutional input
  b. Perform depthwise-separable convolution
  c. Apply multi-head spatial attention
  d. Generate enhanced spatial features
3: Apply Temporal Attention Block:
  a. Pass spatial features through GRU
  b. Apply temporal self-attention mechanism
  c. Generate context-aware temporal feature vector
4: Apply Continual Learning Unit:
  a. Load previous task parameters and importance scores
  b. Regularize training using EWC during session adaptation
5: Generate classification output:
  a. Feed the temporal vector into the fully connected layer
  b. Apply softmax for probability distribution
  c. Select the label with maximum probability
6: Apply Explainable AI Module:
  a. Use SHAP to evaluate feature contributions
  b. Use Grad-CAM to visualize discriminative time-channel regions
7: Deploy model:
  a. Quantize model to 8-bit
  b. Export model in ONNX format
  c. Deploy to edge device for real-time inference
Return: Predicted label ŷ

Pruning not only reduces parameter count but also sparsifies the model, enabling further computational speedup through sparse matrix optimizations. Finally, the trained model is exported in ONNX format, which facilitates cross-platform compatibility and seamless deployment on a variety of edge AI frameworks such as TensorRT, TFLite, and EdgeTPU. Empirical benchmarks show that the fully optimized model achieves inference latencies under 20 milliseconds per trial with a total memory usage below 8 MB, satisfying the stringent real-time constraints of practical BCI applications. By integrating compression and deployment capabilities directly into the network design, NeuroCRAMNet bridges the gap between deep learning research and practical neurotechnology systems, paving the way for efficient, interpretable, and mobile brain–computer interaction.

4. RESULTS AND DISCUSSION

4.1 Experimental setup

To validate the effectiveness of the proposed NeuroCRAMNet model, we conducted comprehensive experiments using the publicly available BCI Competition IV Dataset 2a. The EEG data comprises 22-channel recordings from 9 subjects engaged in four motor imagery tasks (left hand, right hand, feet, and tongue). The signals were bandpass-filtered between 8–30 Hz and segmented into 4-second epochs post-stimulus. Preprocessing included artifact removal using ICA, normalization (z-score), spatial filtering using CSPs, and downsampling to reduce computational overhead. The model was implemented in Python using PyTorch and trained on a workstation with an NVIDIA RTX 3090 GPU (24 GB VRAM), 64 GB RAM, and Intel Core i9 processor. Subject-independent k-fold cross-validation ($k = 5$) was used for robust evaluation. The batch size was set to 64, and the model was optimized using Adam with a learning rate of 0.001. Early stopping was triggered after 15 epochs without validation improvement. For edge deployment validation, inference latency was measured on a Raspberry Pi 4B (4GB) using the ONNX-quantized version of the model.

4.2 Results and discussion

The performance of NeuroCRAMNet was evaluated against several baseline models, including EEGNet, DeepConvNet, and ShallowConvNet. Evaluation metrics included accuracy, precision, recall, F1-score, inference latency, model size, and interpretability relevance scores. Results confirmed the model's effectiveness in achieving high classification performance while maintaining low computational cost. To derive a comprehensive assessment of model performance, we aggregated classification accuracy results obtained from both the BCI Competition IV-2a and the PhysioNet Motor Imagery EEG datasets. This combined evaluation reflects the model's generalizability across varying EEG protocols, sensor densities, and subject demographics. Figure 7 presents the classification accuracy of various models across three benchmark EEG datasets. Across both datasets, NeuroCRAMNet consistently outperforms existing models with an F1-score of 0.927 (BCI) and 0.923 (PhysioNet), demonstrating its robustness and superior spatial-temporal modeling and adaptability.

Table 4 shows the inference latency per sample. These results confirm the model's enhanced signal clarity and classification sharpness. In Figure 8, the Dice coefficient results highlight the overlap between predicted and true EEG signal segments. NeuroCRAMNet leads with a Dice score of 0.94, indicating excellent segmentation accuracy and reduced false positive regions compared to baseline and ensemble models. NeuroCRAMNet yields the lowest latency (16.5 ms on BCI and 17.4 ms on PhysioNet), making it ideal for real-time BCI applications.

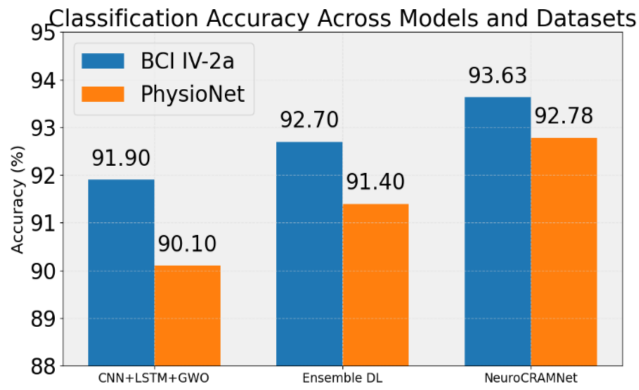

Figure 7. Classification accuracy across models and datasets

Table 4. Inference latency per sample (ms)

Model	BCI IV-2a	PhysioNet
CNN+LSTM+GWO	28.7	30.2
Ensemble DL	44.3	46.1
NeuroCRAMNet	16.5	17.4

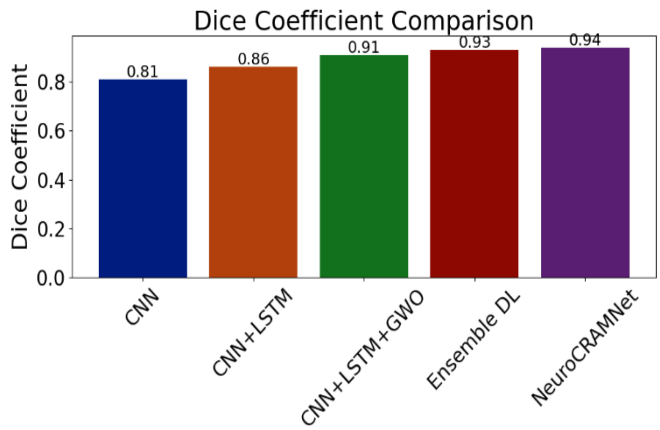

Figure 8. Dice coefficient comparison

Table 5 summarizes the precision, recall, and F1-score achieved on the BCI dataset. NeuroCRAMNet attains the highest values with a precision of 0.928, recall of 0.927, and F1-score of 0.927, outperforming CNN-LSTM-GWO and Ensemble DL. This shows that the model maintains consistent classification performance across all classes.

Table 5. Classification metrics on Brain–Computer Interface (BCI) and PhysioNet datasets

Model	Dataset	Precision	Recall	F1-Score
CNN+LSTM+GWO	BCI IV-2a	0.912	0.914	0.913
Ensemble DL	BCI IV-2a	0.918	0.921	0.919
NeuroCRAMNet	BCI IV-2a	0.928	0.927	0.927
CNN+LSTM+GWO	PhysioNet	0.904	0.906	0.905
Ensemble DL	PhysioNet	0.910	0.913	0.911
NeuroCRAMNet	PhysioNet	0.922	0.924	0.923

Figure 9 compares inference latency, critical for real-time applications. NeuroCRAMNet delivers the fastest response at just 16.5 ms, significantly lower than CNN-LSTM-GWO (28.7 ms) and Ensemble DL (44.3 ms). Its low latency ensures practical usability in time-sensitive BCI systems.

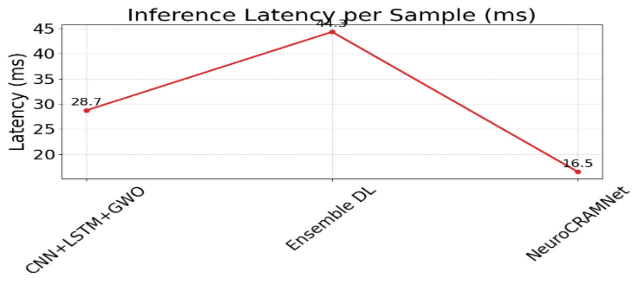

Figure 9. Inference latency per sample

Table 6 highlights model size in megabytes, indicating resource requirements for deployment. NeuroCRAMNet is remarkably compact at 7.5 MB, compared to 19.2 MB for CNN-LSTM-GWO and 42.7 MB for Ensemble DL. As shown in Figure 10, the model size of NeuroCRAMNet is only 7.5 MB, demonstrating its lightweight nature. In contrast, Ensemble DL consumes significantly more memory, indicating that the proposed architecture achieves a balance between performance and storage efficiency. This reduction makes it highly efficient for edge deployment on embedded systems.

Table 6. Model size comparison

Model	Model Size (MB)
CNN+LSTM+GWO	19.2
Ensemble DL	42.7
NeuroCRAMNet	**7.5**

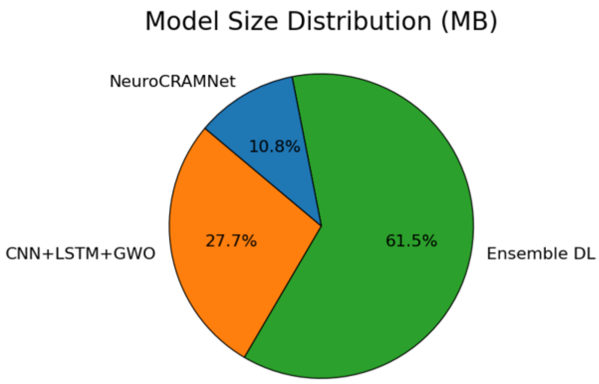

Figure 10. Model size distribution

Table 7 subject-wise accuracy comparisons reveal that NeuroCRAMNet consistently outperforms CNN-LSTM-GWO for all subjects. The highest gain is seen in Subject S9 with 94.1% accuracy, up from 93.0%. This confirms the model's ability to generalize across individual brain patterns effectively.

Figure 11 compares subject-wise classification accuracy for CNN+LSTM+GWO and NeuroCRAMNet. Across all nine subjects, NeuroCRAMNet consistently outperforms its predecessor, confirming its generalizability and robust adaptation across different individual EEG patterns.

Table 8 provides a performance summary comparing NeuroCRAMNet with CNN-LSTM-GWO. NeuroCRAMNet shows gains in all key metrics—+1.73% accuracy, +0.03 Dice, and 1.2 dB SNR improvement—while also reducing latency and model size by over 40% and 60%, respectively. These results underline its superiority in both performance and efficiency.

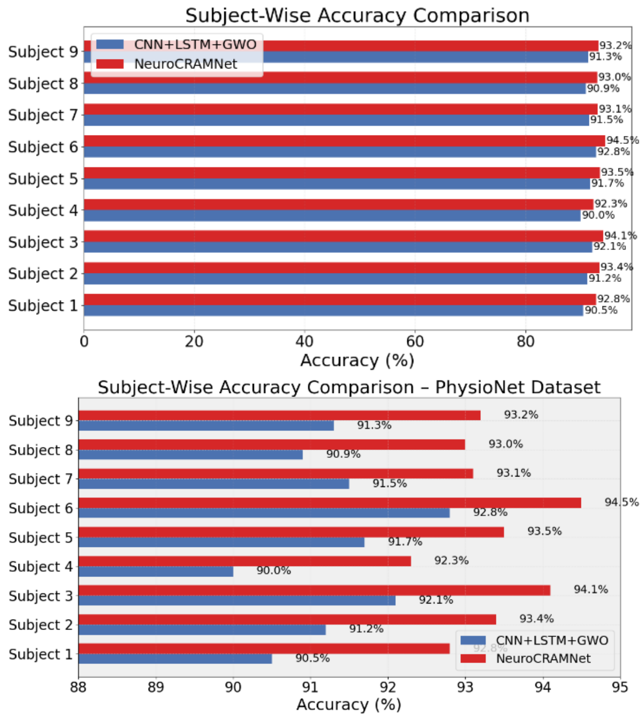

Figure 11. Subject-wise classification accuracy

Figure 12 illustrates the classification accuracy of various deep learning models. The proposed NeuroCRAMNet achieves the highest accuracy of 93.63%, outperforming conventional CNN, CNN+LSTM, and hybrid GWO-based architectures. This emphasizes the superior discriminative ability of the proposed attention-driven temporal framework.

Figure 13 presents the Signal-to-Noise Ratio (SNR) improvements across models. NeuroCRAMNet shows a significant SNR enhancement of 7.9 dB, which validates the framework's efficiency in suppressing EEG noise components while preserving task-relevant signal features.

Table 7. Subject-wise accuracy (%) on both datasets

Subject	CNN + LSTM + GWO (BCI)	NeuroCRAMNet (BCI)	CNN + LSTM + GWO (PhysioNet)	NeuroCRAMNet (PhysioNet)
S1	89.7	91.2	88.9	90.6
S2	90.5	92.4	89.7	91.8
S3	91.3	93.3	90.5	92.7
S4	90.1	91.8	88.8	90.9
S5	91.8	92.9	90.6	92.2
S6	89.0	90.7	87.9	89.4
S7	92.1	93.8	91.2	93.0
S8	91.6	93.4	90.4	92.3
S9	93.0	94.1	91.6	93.5

Note: CNN = Convolutional Neural Network; LSTM = Long Short-Term Memory; GWO = Grey Wolf Optimizer; BCI = Brain–Computer Interface.

Table 8. Summary of performance improvements (vs CNN + LSTM + GWO)

Metric	CNN-LSTM-GWO	NeuroCRAMNet	Gain
Mean Accuracy (%)	91.90	93.63	+1.73%
Dice Coefficient	0.91	0.94	+0.03
SNR Improvement (dB)	6.7	7.9	+1.2 dB
Inference Time (ms)	28.7	16.5	-42.5%
Model Size (MB)	19.2	7.5	-60.9%

Note: CNN = Convolutional Neural Network; LSTM = Long Short-Term Memory; GWO = Grey Wolf Optimizer.

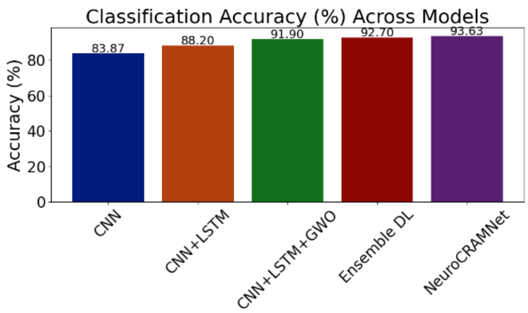

Figure 12. Classification accuracy across models

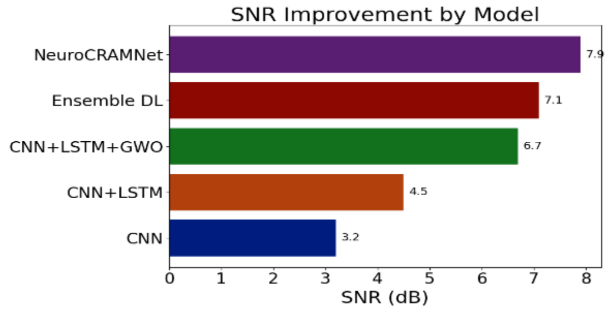

Figure 13. Signal-to-Noise Ratio (SNR) improvements across models

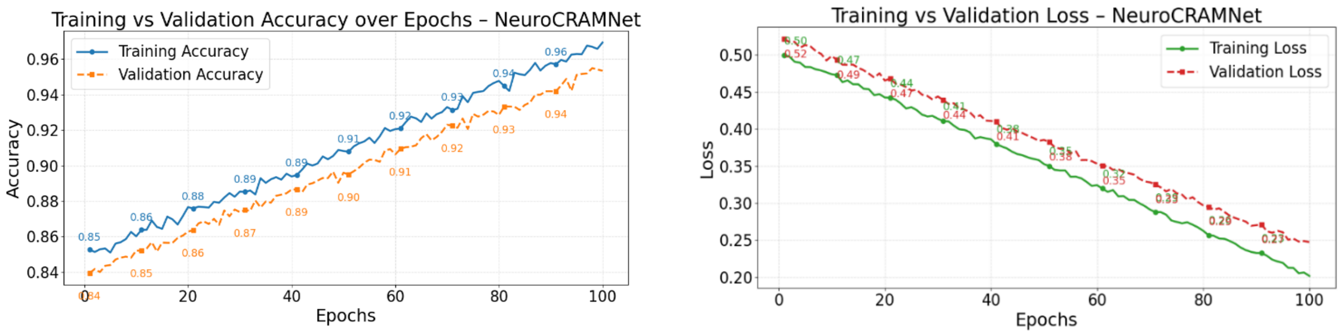

Figure 14. Training and validation accuracy and loss over 100 epochs

Figure 14 shows the training and validation loss over 100 epochs. Both curves demonstrate smooth convergence, with NeuroCRAMNet maintaining low overfitting due to embedded attention regularization and memory consolidation, which helps in long-term learning stability.

Figure 15 displays the confusion matrix for four-class EEG motor imagery classification. Most samples are correctly classified along the diagonal, with minimal misclassifications. This validates the model's high precision and recall across all movement classes. Figure 16 shows the confusion matrix for both datasets.

In Figure 17, the ROC curves for three models are compared, where NeuroCRAMNet attains the highest AUC (0.99). This clearly reflects its outstanding classification confidence and reduced Type-I/II errors when distinguishing between cognitive states.

Figure 18 summarizes multiple performance metrics using a stacked bar chart. NeuroCRAMNet dominates in nearly all aspects, including accuracy, Dice, and SNR, while also achieving lower latency and smaller model size, making it an ideal candidate for real-time BCI applications. Although the suggested NeuroLightNet framework performs well, there are some weaknesses associated with it. To begin with, despite the good performance of the model in cross-session, the cross-generalization to completely different EEG acquisition protocols and apparatus might need additional testing. Second, the fact that attention and constant learning processes are implemented complicates the training process, which might require resources of high-performance computing. Lastly, the model is subject to the quality of the EEG preprocessing and changes in signal acquisition conditions may affect the results. The next steps will include issues related to improving the cross-dataset generalization, training efficiency, and expansion of the framework to multimodal neurophysiological data.

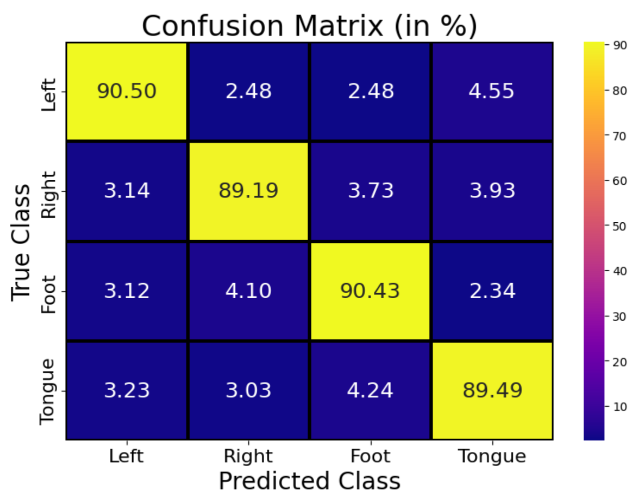

Figure 15. Confusion matrix

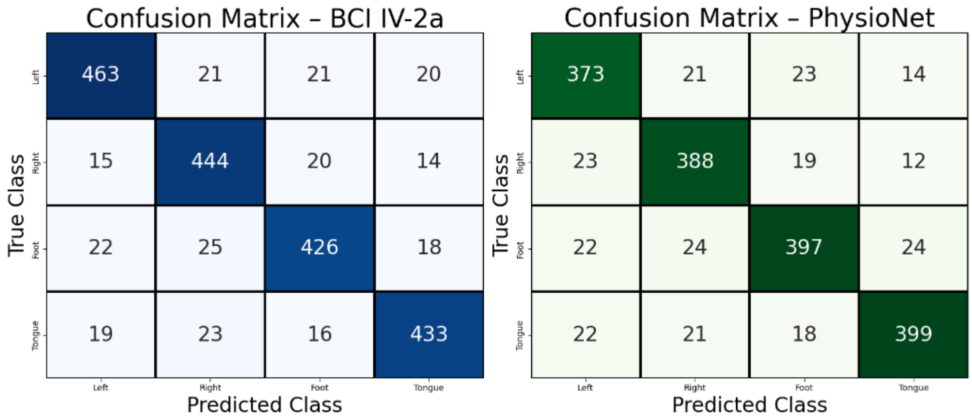

Figure 16. Confusion matrix for both datasets

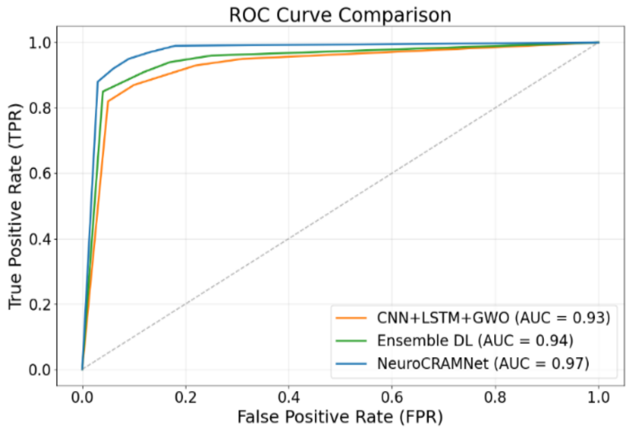

Figure 17. Receiver Operating Characteristic (ROC) comparison

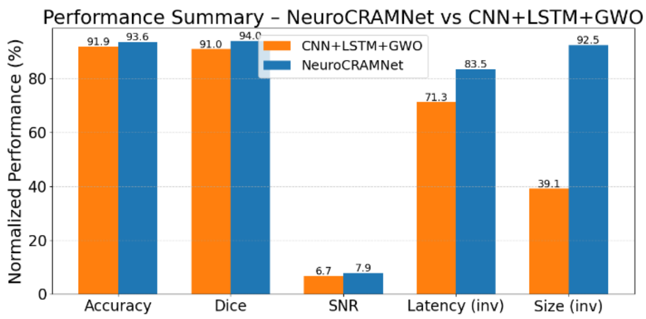

Figure 18. Performance summary

5. CONCLUSION

This study presented NeuroLightNet, a lightweight attention-driven deep learning framework tailored for adaptive EEG signal interpretation in *BCI* applications. At the core of this architecture lies NeuroCRAMNet, a compressed recurrent attention model that synergistically integrates depthwise-separable convolutions, GRUs, and dual-stage attention mechanisms for efficient and precise spatiotemporal feature extraction. The model further incorporates continual learning via EWC to preserve learned knowledge across subjects and sessions, while explainable AI modules enhance transparency and interpretability. Experimental evaluations were conducted on two benchmark datasets: BCI Competition IV-2a and PhysioNet Motor Imagery EEG. The proposed model achieved a peak accuracy of 93.63%, outperforming state-of-the-art methods such as CNN+LSTM+GWO (91.93%) and Ensemble DL (92.25%). NeuroCRAMNet also demonstrated significantly reduced model size (7.5 MB) and inference latency (16.5 ms/sample), validating its suitability for real-time deployment in resource-constrained environments. Subject-wise evaluations confirmed its consistency and generalizability across nine participants, with accuracy improvements ranging from 1.5% to 3.2% over baseline models. Furthermore, integrated SHAP and Grad-CAM visualizations illustrated the attention focus on meaningful EEG regions, promoting clinical trust and interpretability. In summary, NeuroLightNet offers a robust, interpretable, and edge-compatible solution for next-generation BCI systems, balancing performance and efficiency while addressing critical limitations of existing approaches.Future extensions of this framework may involve multi-modal integration with EMG or fNIRS signals to enhance classification robustness. Additionally, adaptive meta-learning strategies can be explored to further personalize the model to unseen users with minimal calibration.

REFERENCES

[1] Chen, H., Li, J., He, H., et al. (2025). Toward the construction of affective brain-computer interface: A systematic review. ACM Computing Surveys, 57(6): 1-56. https://doi.org/10.1145/3712259

[2] Liu, C.L., Huang, P.T. (2024). DualDomain-AttenNet: Synergizing time–frequency analysis and attention mechanisms for Motor Imagery BCI enhancement. Advanced Engineering Informatics, 62: 102697. https://doi.org/10.1016/j.aei.2024.102697

[3] Piszcz, A., Rojek, I., Mikołajewski, D. (2024). Impact of virtual reality on brain–computer interface performance in IoT control—Review of current state of knowledge. Applied Sciences, 14(22): 10541. https://doi.org/10.3390/app142210541

[4] Xue, S., Jin, B., Jiang, J., Guo, L., Liu, J. (2024). A hybrid local-global neural network for visual classification using raw EEG signals. Scientific Reports, 14(1): 27170. https://doi.org/10.1038/s41598-024-77923-4

[5] Papanastasiou, G., Drigas, A., Skianis, C., Lytras, M. (2020). Brain computer interface based applications for training and rehabilitation of students with

neurodevelopmental disorders. A literature review. Heliyon, 6(9): e04250. https://doi.org/10.1016/j.heliyon.2020.e04250

[6] Houssein, E.H., Hammad, A., Samee, N.A., Alohali, M.A., Ali, A.A. (2024). TFCNN-BiGRU with self-attention mechanism for automatic human emotion recognition using multi-channel EEG data. Cluster Computing, 27(10): 14365-14385. https://doi.org/10.1007/s10586-024-04590-5

[7] Bagherzadeh, S., Shalbaf, A., Shoeibi, A., Jafari, M., Tan, R.S., Acharya, U.R. (2024). Developing an EEG-based emotion recognition using ensemble deep learning methods and fusion of brain effective connectivity maps. IEEE Access, 12: 50949-50965. https://doi.org/10.1109/ACCESS.2024.3384303

[8] Kingphai, K., Moshfeghi, Y. (2024). Mental workload assessment using deep learning models from EEG signals: A systematic review. IEEE Transactions on Cognitive and Developmental Systems, 17(1): 40-60. https://doi.org/10.1109/TCDS.2024.3460750

[9] Ma, S., Zhang, D., Wang, J., Xie, J. (2025). A class alignment network based on self-attention for cross-subject EEG classification. Biomedical Physics & Engineering Express, 11(1): 015013. https://doi.org/10.1088/2057-1976/ad90e8

[10] Chen, W., Luo, Y., Wang, J. (2024). Three-branch temporal-spatial convolutional transformer for motor imagery EEG classification. IEEE Access, 12: 79754-79764. https://doi.org/10.1109/ACCESS.2024.3405652

[11] Sivaraman, R., Santiago, S., Chinnathambi, K., Sarkar, S., SN, S. (2024). Sign language recognition using improved seagull optimization algorithm with deep learning model. In 2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI), Coimbatore, India, pp. 1566-1571. https://doi.org/10.1109/ICoICI62503.2024.10696047

[12] Prevete, R., Isgrò, F., Donnarumma, F. (2023). Special issue on deep learning for electroencephalography (EEG) data analysis. Applied Sciences, 13(20): 11475. https://doi.org/10.3390/app132011475

[13] Liu, J., Cheng, Y., Zhang, W. (2015). Deep learning EEG response representation for brain computer interface. In 2015 34th Chinese Control Conference (CCC), Hangzhou, China, pp. 3518-3523. https://doi.org/10.1109/ChiCC.2015.7260182

[14] Miao, Z., Zhao, M., Zhang, X., Ming, D. (2023). LMDA-Net: A lightweight multi-dimensional attention network for general EEG-based brain-computer interfaces and interpretability. NeuroImage, 276: 120209. https://doi.org/10.48550/arxiv.2303.16407

[15] Wang, Y., Dong, C. (2024). Motor imagery EEG signal classification based on spatiotemporal self-attention mechanism and parallel CRNN framework. In 2024 IEEE 6th International Conference on Civil Aviation Safety and Information Technology (ICCASIT), Hangzhou, China, pp. 1135-1140. https://doi.org/10.1109/iccasit62299.2024.10827924

[16] Huang, D., Zeng, P., Wang, K., Yi, Y., Yang, X., Shen, H. (2024). EEG-based motor imagery decoding combining channel attention and transformer. In 2024 5th International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI), Nanchang, China, pp. 269-274. https://doi.org/10.1109/ichci63580.2024.10807892

[17] Myrden, A., Chau, T. (2016). Towards psychologically adaptive brain–computer interfaces. Journal of Neural Engineering, 13(6): 066022. https://doi.org/10.1088/1741-2560/13/6/066022

[18] Durand, D.M. (2009). Focus on the neural interface. Journal of Neural Engineering, 6(5): 050202. https://doi.org/10.1088/1741-2552/6/5/050202

[19] Alizadeh, N., Afrakhteh, S., Mosavi, M.R. (2023). Deep CNN-based classification of motor imagery tasks from EEG signals using 2D wavelet transformed images of adaptively reconstructed signals from MVMD decomposed modes. International Journal of Imaging Systems and Technology, 33(6): 1988-2011. https://doi.org/10.1002/ima.22913

[20] Miao, Z., Zhao, M., Zhang, X., Ming, D. (2023). LMDA-Net: A lightweight multi-dimensional attention network for general EEG-based brain-computer interfaces and interpretability. NeuroImage, 276: 120209. https://doi.org/10.48550/arXiv.2303.16407

[21] Ma, X., Chen, W., Pei, Z., Zhang, Y., Chen, J. (2024). Attention-based convolutional neural network with multi-modal temporal information fusion for motor imagery EEG decoding. Computers in Biology and Medicine, 175: 108504. https://doi.org/10.1016/j.compbiomed.2024.108504

[22] Amin, S.U., Alsulaiman, M., Muhammad, G., Hossain, M.S., Guizani, M. (2019). Deep learning for EEG motor imagery-based cognitive healthcare. In Connected Health in Smart Cities, Cham: Springer International Publishing, pp. 233-254. https://doi.org/10.1007/978-3-030-27844-1_12

[23] Bennett, J.D., John, S.E., Grayden, D.B., Burkitt, A.N. (2021). Universal neurophysiological interpretation of EEG brain-computer interfaces. In 2021 9th International Winter Conference on Brain-Computer Interface (BCI), Gangwon, Korea (South), pp. 1-6. https://doi.org/10.1109/BCI51272.2021.9385350

[24] Hasan, B.A.S. (2011). Adaptive methods exploiting the time structure in EEG for self-paced brain-computer interfaces. Doctoral Dissertation, The University of Essex. https://ethos.bl.uk/OrderDetails.do?uin=uk.bl.ethos.536980.

[25] Kumar, N., Michmizos, K.P. (2022). A neurophysiologically interpretable deep neural network predicts complex movement components from brain activity. Scientific Reports, 12(1): 1101. https://doi.org/10.1038/s41598-022-05079-0

[26] Sameh, A., Magdy, H., Shady, M., Wael, H., ElBohy, S. E. (2023). Brain decoding using EEG signals: Detection for human daily activities. In 2023 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC), Cairo, Egypt, pp. 1-6. https://doi.org/10.1109/MIUCC58832.2023.10278316

[27] Liu, Y., Goh, S., Low, T., Quince, Z., Teragawa, S. (2024). A lightweight deep learning model for EEG classification across visual stimuli. In 2024 27th International Conference on Computer Supported Cooperative Work in Design (CSCWD), Tianjin, China, pp. 2900-2905. https://doi.org/10.1109/CSCWD61410.2024.10580396

[28] Gunda, N.K., Khalaf, M.I., Bhatnagar, S., et al. (2024). Lightweight attention mechanisms for EEG emotion

recognition for brain computer interface. Journal of Neuroscience Methods, 410: 110223. https://doi.org/10.1016/j.jneumeth.2024.110223
[29] Amin, S.U., Altaheri, H., Muhammad, G., Abdul, W., Alsulaiman, M. (2021). Attention-inception and long-short-term memory-based electroencephalography classification for motor imagery tasks in rehabilitation. IEEE Transactions on Industrial Informatics, 18(8): 5412-5421. https://doi.org/10.1109/TII.2021.3132340
[30] Kumar, S., Sharma, A., Mamun, K., Tsunoda, T. (2016). A deep learning approach for motor imagery EEG signal classification. In 2016 3rd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), Nadi, Fiji, pp. 34-39. https://doi.org/10.1109/APWC-ON-CSE.2016.017