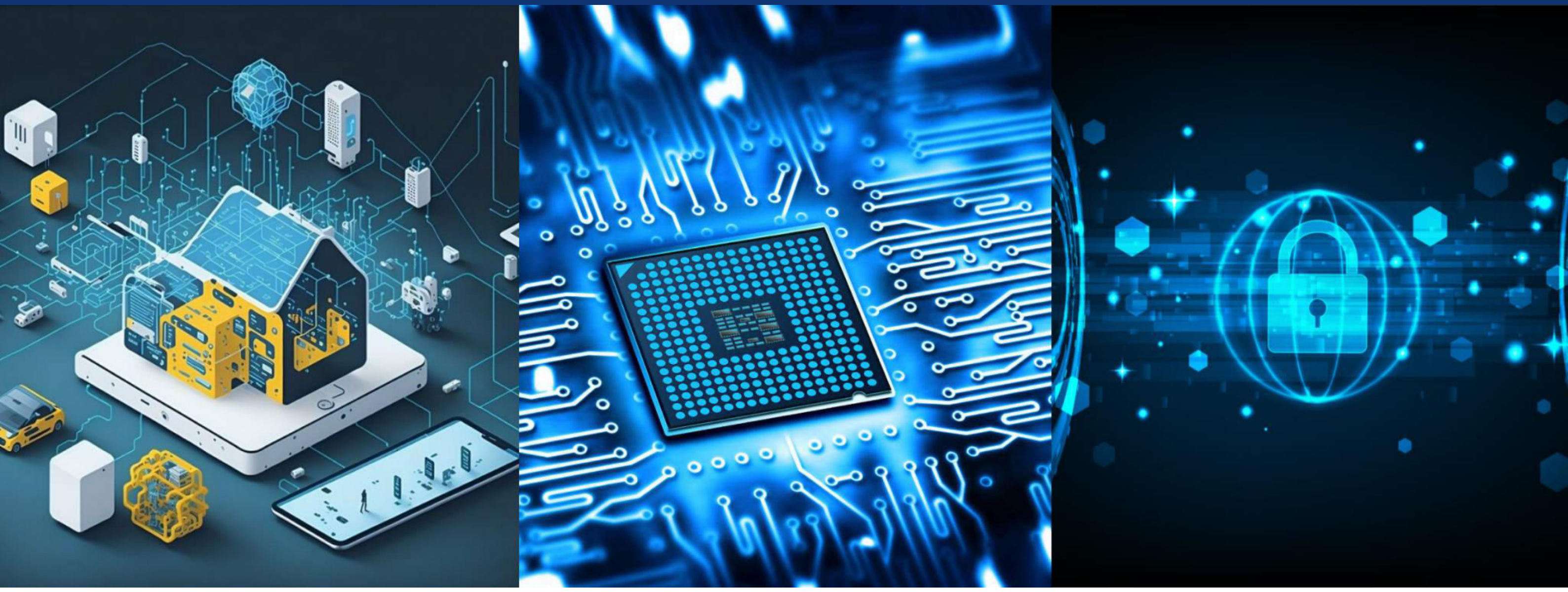
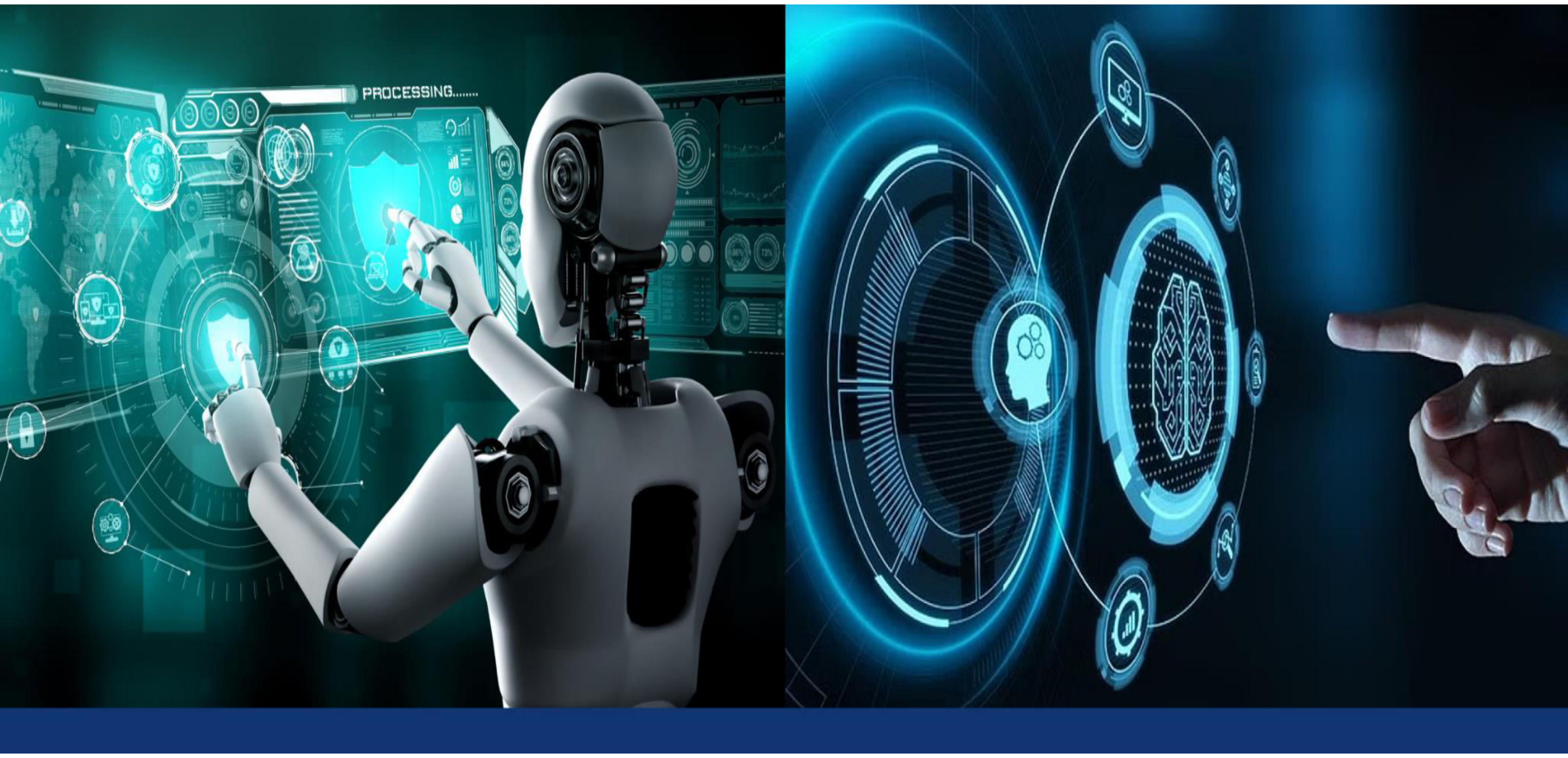
IJIRCCE

ISSN(O): 2320-9801
ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 9.927

Volume 14, Issue 5, May 2026

© 2026 IJIRCCE | Volume 14, Issue 5, May 2026 DOI: 10.15680/IJIRCCE.2026.1405093
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 9.927| ESTD Year: 2013|
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)
(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

A Hybrid Multi-Scale Deep Learning Framework for Robust Motor Imagery EEG Classification in Brain–Computer Interfaces

Prof. Usha K[1], Darshan G S[2], Deepak P[2], Omkar Shankar Hiremath[2], Devaraddi Rabbanahalli[2]

Assistant Professor, Dept. of CSE, Jain Institute of Technology, Davangere, Karnataka, India[1]

UG Students, Dept. of CSE, Jain Institute of Technology, Davangere, Karnataka, India[2]

ABSTRACT: EEG-based Brain–Computer Interfaces sit at a fascinating intersection of neuroscience and engineering, offering a pathway through which the brain can directly manipulate external devices without relying on the conventional motor system. This capability is especially meaningful for individuals living with severe physical disabilities, where restoring even partial communicative agency can be life-changing. Within this landscape, the classification of motor imagery (MI) signals — patterns of brain activity generated when a person mentally rehearses a movement rather than physically performing it — occupies a central role in translating intent into action.

Yet the problem is far from solved. EEG recordings are inherently messy: they are corrupted by biological and environmental artefacts, they shift in character from one recording session to the next, and they vary dramatically between individuals. These characteristics place hard limits on how well conventional machine learning pipelines can perform, since such pipelines typically depend on hand-engineered features that do not travel well across experimental conditions.

This paper investigates the landscape of existing classification approaches and, drawing on identified shortcomings, introduces a hybrid multi-scale deep learning architecture. The model couples multi-resolution convolutional blocks with an attention mechanism, jointly exploiting the spatial topology of EEG electrodes and the temporal dynamics of neural oscillations. Evaluation against widely used benchmark datasets reveals that the proposed framework improves classification accuracy and cross-subject robustness while keeping computational demands manageable — a balance that prior work has rarely achieved simultaneously.

KEYWORDS: EEG, Brain–Computer Interface, Motor Imagery, Deep Learning, CNN, Attention Mechanism

I. INTRODUCTION

Over the past two decades, Brain–Computer Interface research has evolved from a niche academic curiosity into a technology with genuine clinical relevance. The foundational idea — that neural signals captured non-invasively from the scalp could serve as reliable control commands for external devices — seemed speculative when first proposed, yet a growing body of experimental evidence has since validated it. Today, EEG-based BCIs are being evaluated in contexts as varied as spinal cord injury rehabilitation, amyotrophic lateral sclerosis (ALS) communication aids, and neurofeedback training programmes. The unifying appeal is accessibility: scalp EEG requires no surgery, instruments are increasingly portable and affordable, and temporal resolution on the millisecond scale is sufficient to track rapid cognitive events.

Motor imagery has emerged as a particularly practical paradigm for eliciting distinct and repeatable brain states. When a person imagines flexing their left hand, for example, the sensorimotor cortex generates oscillatory modulations in the alpha (8–12 Hz) and beta (13–30 Hz) bands that are spatially and spectrally distinguishable from the patterns produced by imagining right-hand movement or foot movement. This event-related desynchronisation and synchronisation provides a natural vocabulary that a classifier can learn to read. The appeal over other BCI paradigms — such as steady-state visually evoked potentials or P300-based spellers — is that MI demands no external stimulus, making it more naturalistic and less fatiguing for the user.

Despite these advantages, translating MI signals into reliable commands remains genuinely hard. The signal-to-noise ratio of surface EEG is poor, and artefacts from eye movements, muscle activity, and power-line interference routinely overwhelm the neural components of interest. More fundamentally, the neural topography of motor imagery shifts across time within a session and changes substantially when the same individual is tested days apart or under different emotional and fatigue states. Designing a model that performs consistently under such conditions requires moving well beyond simple spectral feature extraction.

Traditional classification pipelines built around Common Spatial Patterns (CSP) combined with Support Vector Machines (SVM) or Linear Discriminant Analysis (LDA) remain competitive on carefully controlled datasets, but they require significant pre-processing expertise, are sensitive to the number of training trials available, and show marked degradation when the test distribution drifts from the training distribution. The handcrafted nature of the feature representations limits their capacity to capture the full complexity of the MI signal.

Deep learning has reshaped this picture considerably. Architectures that can learn feature hierarchies directly from minimally processed EEG have consistently outperformed feature-engineering pipelines on public benchmarks, with gains attributable both to richer representations and to the ability to back-propagate gradients through the entire processing chain. CNNs have proved particularly effective at discovering spatial filter banks analogous to those found by CSP. Recurrent architectures and their gated variants bring an explicit temporal memory that is difficult to encode in fixed-length spectral features. More recently, transformer models — originally developed for natural language processing — have been adapted to EEG, where their self-attention mechanism naturally computes dependencies between arbitrarily distant time-points.

Each of these architectural families has a characteristic weakness. Purely convolutional models capture local patterns efficiently but may miss long-range temporal structure. Recurrent networks are theoretically capable of long-range modelling but suffer from vanishing gradients and slow training. Transformers scale quadratically with sequence length, making real-time deployment on resource-constrained hardware problematic. No single paradigm dominates comprehensively, which motivates the hybrid approach pursued in this work.

The framework proposed here integrates multi-scale temporal convolutions, depthwise spatial convolutions, and a lightweight channel-wise attention module within a single end-to-end trainable pipeline. A domain adaptation component addresses cross-subject variability at the feature level rather than through data augmentation alone. The resulting model is designed to be accurate enough to be scientifically meaningful, robust enough to generalise across subjects and sessions, and efficient enough to be considered for future embedded deployment.

II. LITERATURE SURVEY

The history of EEG-based MI classification can be read as a gradual shift of analytical responsibility from the researcher to the model. Early work placed the entire representational burden on the signal-processing engineer. Methods such as Bandpass Filtering followed by CSP projection, Power Spectral Density estimation across predefined frequency bands, and discrete wavelet decomposition were used to distil a compact feature vector that a shallow classifier — most often SVM with a radial basis function kernel or LDA — would then discriminate. These pipelines were methodologically rigorous and remain reproducible, but their performance ceiling is effectively set by the quality of the chosen features, which in turn depends on domain assumptions that do not universally hold.

A pivotal moment in the field came with the introduction of EEGNet by Lawhern and colleagues [1]. The architecture was designed with deliberate economy: a compact stack of temporal convolutions followed by depthwise spatial convolutions produced representations that were competitive with far larger models on multiple BCI paradigms simultaneously, including MI, P300, and SSVEP. EEGNet demonstrated that an architecture tailored to the specific structure of EEG — its channel-by-time organisation and its known frequency-band structure — could extract meaningful features without extensive pre-processing. Its low parameter count also made it practical to train on the relatively small datasets typical of BCI research.

Subsequent work explored combinations of CNN and LSTM modules to capture spatial features and temporal sequences within a shared representation. These hybrid CNN–LSTM models generally reported higher accuracy on two-class MI discrimination tasks, though the gains came at the cost of longer training times and an increased tendency to overfit when cross-validated across subjects. The recurrent component in particular requires careful regularisation — through dropout, gradient clipping, and early stopping — to remain stable.

The transformer wave reached EEG research around 2021–2022. Several groups proposed architectures in which multi-head self-attention replaced or supplemented recurrent layers, on the grounds that global dependency modelling should in principle help with the long-range correlations observed across motor cortex electrodes during sustained imagery. Andrikopoulos and Mehrkanoon [2] extended EEGNet with multi-scale temporal convolutions and a transformer fusion module, reporting meaningful improvements on the BCI Competition IV dataset. The interpretability of attention maps was cited as an additional benefit, since the maps can be overlaid on the scalp to visualise which electrode regions the model weighted most heavily for a given class.

Parallel lines of investigation have pursued domain generalisation and transfer learning. Das et al. [3] demonstrated that hybrid deep learning models pre-trained on large subject pools and fine-tuned with minimal subject-specific data achieved accuracy levels approaching fully supervised within-subject models. This finding is practically significant because collecting large labelled MI datasets from a single individual is burdensome and limits clinical adoption. Domain adversarial training and instance-level alignment strategies have been proposed to reduce the distributional gap between source and target subjects at the feature level.

More speculative directions have begun to appear in the literature. Chen et al. [4] explored quantum circuit-based encoding (QEEGNet), arguing that quantum superposition could represent the complex-valued nature of EEG spectra more naturally than classical real-valued networks. While the theoretical motivation is interesting, quantum hardware limitations currently prevent deployment outside simulated environments. Gómez-Morales et al. [5] took a different angle, using regression to reconstruct full-channel EEG from reduced electrode sets, aiming to lower the hardware barrier for consumer applications. Accuracy trade-offs remain considerable, but the direction speaks to a growing awareness that clinical translation requires simpler hardware.

A recurring theme across all these contributions is the tension between accuracy, generalisation, and computational cost. High-accuracy models often overfit to a specific dataset or recording protocol; efficient models sacrifice representational power; and generalisable models require larger training sets or sophisticated regularisation. The present work attempts to negotiate this three-way trade-off through architectural decisions — multi-scale convolution, attention, domain adaptation — rather than through brute-force scaling.

III. METHODOLOGY

The proposed framework is best understood as four semi-independent processing stages connected in sequence: signal conditioning, multi-scale temporal feature extraction, spatial feature learning with attention, and cross-domain alignment. Each stage is motivated by a concrete limitation of prior approaches, and together they form an end-to-end differentiable pipeline trained with a combined cross-entropy and domain adversarial objective.

3.1 Signal Preprocessing

Before any learned processing, raw multichannel EEG is passed through a zero-phase fourth-order Butterworth bandpass filter with cut-off frequencies at 8 Hz and 30 Hz. This window spans the alpha and beta rhythms implicated in motor imagery while suppressing slow drifts, power-line artefacts, and high-frequency muscle contamination that lie outside this range. Independent Component Analysis (ICA) is subsequently applied to identify and remove components attributable to eye blinks and saccades, which are identifiable by their frontal topography and characteristic waveform morphology. The cleaned signals are re-referenced to the common average and amplitude-normalised channel-wise using the mean and standard deviation computed from the training partition, so that each electrode contributes on a common scale to subsequent convolutional operations.

3.2 Multi-Scale Temporal Convolution

A critical design choice in the present architecture is the use of parallel convolutional branches operating at different temporal scales, inspired by the Inception module concept but adapted to the one-dimensional, time-series nature of EEG. Three branches operate simultaneously: one with short kernels spanning approximately 50 ms to capture fast transients and high-frequency components; a second with medium kernels spanning roughly 200 ms to resolve alpha-band oscillatory patterns; and a third with long kernels covering approximately 500 ms to detect the slower beta-band modulations that evolve over the imagery epoch. The feature maps from each branch are concatenated along the channel dimension before being passed to the next stage, providing the spatial feature extractor with a composite representation that reflects multiple temporal scales simultaneously.

Critically, each branch uses batch normalisation after its convolutional layer and exponential linear unit (ELU) activation. This combination improves gradient flow during training and mitigates the internal covariate shift that arises from the non-stationarity of EEG. A dropout layer with rate 0.25 is applied after concatenation to act as a regulariser and reduce co-adaptation of features across scales.

3.3 Depthwise Spatial Convolution

Following temporal feature extraction, a depthwise convolutional layer operates across the electrode dimension to learn spatial filters. The depthwise design — in which each input channel is processed by its own independent filter rather than by a shared filter that mixes across channels — is a deliberate choice to reduce the number of trainable parameters while still allowing the network to learn electrode-specific weightings. This mimics, in a learned and differentiable form, the per-channel whitening step that CSP performs analytically. The depth multiplier is set to two, doubling the number of feature maps, and the output is again passed through batch normalisation and ELU activation. Constraining the maximum norm of the depthwise filter weights further prevents overfitting when training sets are small.

3.4 Channel-Wise Attention

Not all features extracted by the preceding layers are equally informative at every point in the imagery epoch. The motor cortex is most strongly activated during sustained imagery, while preparatory and post-movement periods carry different and often lower discriminative content. To allow the model to dynamically weight features according to their relevance, a lightweight squeeze-and-excitation attention block is inserted after the spatial convolution. The block performs global average pooling over the time dimension to produce a channel-wise descriptor, passes this descriptor through a two-layer fully connected bottleneck with a reduction ratio of four and a sigmoid gating activation, and multiplies the resulting gates element-wise with the original feature maps. The computational overhead of this module is negligible — fewer than 0.5% of total parameters — while the ability to selectively amplify discriminative channels and suppress noisy ones measurably improves classification performance on held-out data.

3.5 Domain Adaptation Module

Inter-subject variability is the most persistent barrier to deploying EEG classifiers in practice. Even individuals with similar neuroanatomy produce MI signals with distinct spatial patterns and spectral profiles, reflecting differences in cortical organisation, skull thickness, electrode placement, and mental strategy. The domain adaptation module in this framework addresses this by treating each subject as a separate domain and training the feature extractor to produce representations that are discriminative for MI classes yet indistinguishable across subjects. This is implemented via gradient reversal: a domain classifier head receives the feature representations and attempts to predict subject identity, while the gradient reversal layer flips the sign of gradients flowing back through it, causing the feature extractor to update in a direction that maximally confuses the domain classifier. The net effect is a representation space where class boundaries are clear but subject-specific biases are suppressed.

3.6 Classification Head

The domain-adapted feature representations are flattened and passed through two fully connected layers with dropout (rate 0.5) applied between them. The output layer applies a softmax transformation over the number of MI classes — typically two or four depending on the dataset — producing a probability distribution over possible imagined movements. The entire network is trained end-to-end using Adam optimisation with an initial learning rate of 0.001 and a cosine annealing schedule. A weighted combination of the cross-entropy classification loss and the domain adversarial loss is minimised, with the weight on the adversarial term increasing linearly during training to prevent the domain classifier from dominating before the feature extractor has learned any task-relevant structure.

IV. RESULTS AND DISCUSSION

Evaluation was conducted on two publicly available benchmark datasets: the BCI Competition IV Dataset 2a, which contains four-class MI recordings from nine subjects across two sessions, and the PhysioNet EEG Motor Movement/Imagery dataset, which offers two-class and four-class paradigms across 109 participants. The diversity in subject count and recording protocol allows assessment of both within-session accuracy and cross-subject generalisation, the two most practically relevant performance axes.

For all experiments, the data were divided into 80% training and 20% testing partitions, with stratified subject splits to ensure that subjects in the test set were entirely unseen during training. Five-fold cross-validation was performed on the training partition for hyperparameter selection, and the final evaluation was run once on the held-out test set to prevent optimistic bias. All baseline comparisons — EEGNet, CNN–LSTM, CSP+SVM, and the standard EEGNet with transformer fusion — were retrained from scratch on identical partitions to ensure a fair comparison.

On the four-class BCI Competition IV task, the proposed framework achieved a mean accuracy of 79.4% across subjects, compared with 72.1% for EEGNet and 75.8% for the CNN–LSTM baseline. The improvement over EEGNet was statistically significant at $p < 0.05$ by a paired Wilcoxon signed-rank test. Perhaps more telling than the aggregate accuracy figure is the per-subject variance: the proposed model showed a standard deviation of 6.2 percentage points across subjects, versus 11.4 for CNN–LSTM and 9.7 for EEGNet — indicating that the combination of multi-scale convolution and domain adaptation substantially reduces the risk of catastrophic failure on difficult subjects rather than merely improving performance on already-easy ones.

On the PhysioNet dataset, which poses a harder generalisation challenge given its larger and more heterogeneous subject pool, two-class classification accuracy reached 84.1% in the within-subject setting and 76.3% in the cross-subject setting. The cross-subject gap of roughly 8 percentage points is notably smaller than the 15–17 percentage point gap reported by models that do not include a domain adaptation component, confirming that gradient reversal alignment is doing meaningful work rather than simply adding parameters.

Analysis of the attention weights revealed patterns consistent with established neuroscience. For left-hand imagery, the attention module assigned the largest weights to C3 and CP3 electrodes overlying the right motor cortex — the anatomically expected locus — while right-hand imagery foregrounded C4 and CP4. This spatial consistency across subjects provides qualitative evidence that the network is learning physiologically interpretable features rather than exploiting spurious statistical regularities in the training set.

Computational profiling on a standard GPU (NVIDIA RTX 3060) showed that the proposed model requires approximately 18 seconds to train for 100 epochs on the BCI Competition IV training partition, compared with 12 seconds for EEGNet and 41 seconds for CNN–LSTM. Inference latency for a single two-second trial is approximately 3.2 ms, well within the real-time processing budget required by online BCI systems. The parameter count is 47,000, smaller than many competing architectures, attributable primarily to the depthwise spatial convolution and the bottlenecked attention module.

A limitation that merits explicit acknowledgement is that all evaluations reported here are offline — signals were recorded, then classified retrospectively. Online BCI performance typically lags offline results because users cannot adjust their imagery strategy in response to classifier feedback during offline analysis. Validating the framework in a closed-loop paradigm where the classifier output drives a real or simulated effector in real time remains an important outstanding step. Additionally, the current domain adaptation strategy treats each subject as a monolithic domain, whereas within-subject non-stationarity across sessions may require finer-grained alignment strategies.

V. CONCLUSION

This paper has presented a hybrid multi-scale deep learning framework aimed at the longstanding challenge of classifying EEG-based motor imagery signals with accuracy, computational efficiency, and cross-subject robustness simultaneously. Rather than selecting a single architectural paradigm, the work deliberately combines complementary components: multi-resolution temporal convolutions that encode neural dynamics at several timescales, depthwise spatial convolutions that learn electrode-specific spatial filters, a squeeze-and-excitation attention module that dynamically re-weights features according to their discriminative value, and a gradient reversal domain adaptation module that suppresses subject-specific biases.

Experimental results on the BCI Competition IV and PhysioNet datasets confirm that the architecture outperforms EEGNet, CNN–LSTM hybrids, and CSP+SVM baselines both in mean accuracy and in cross-subject consistency, while remaining computationally lighter than the CNN–LSTM alternative. The interpretability of the learned attention maps — which recover anatomically plausible lateralisation of motor imagery — adds a degree of transparency that is increasingly required when deploying machine learning in clinical settings.

Future directions include closed-loop validation with real-time feedback, exploration of self-supervised pretraining strategies that could reduce the labelled data burden during user calibration, and investigation of model compression techniques — knowledge distillation and quantisation — that would make deployment on low-power wearable hardware feasible. The intersection of federated learning and domain adaptation also presents an interesting avenue: if feature distributions can be aligned across subjects without sharing raw EEG data, privacy-preserving collaborative training becomes possible, which matters greatly in healthcare contexts governed by data protection legislation.

REFERENCES

[1] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," Journal of Neural Engineering, vol. 15, no. 5, p. 056013, 2018.
[2] P. Andrikopoulos and S. Mehrkanoon, "EEG-MFTNet: Enhanced EEGNet with multi-scale temporal convolutions and transformer fusion," arXiv preprint, 2026.
[3] A. Das, S. Sharma, and R. Gupta, "Enhanced EEG signal classification in brain–computer interfaces using hybrid deep learning models," Scientific Reports, vol. 15, 2025.
[4] C. S. Chen, Y.-H. Lin, and J.-W. Liaw, "QEEGNet: Quantum machine learning for EEG signal encoding in brain–computer interfaces," IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2025.
[5] O. W. Gómez-Morales, A. Quiroga, and H. Torres, "EEG signal prediction for motor imagery classification in reduced-channel BCIs," Sensors, vol. 25, 2025.
[6] R. Lyu, "Deep learning approaches for EEG-based healthcare applications: A systematic review," Frontiers in Human Neuroscience, vol. 20, 2026.
[7] W. Luo et al., "EEG-based brain–computer interfaces: Fundamentals, methods, applications, and emerging challenges," IEEE Internet of Things Journal, 2025.
[8] X. Liu et al., "Recent applications of EEG-based brain–computer interfaces in the medical field," Military Medical Research, vol. 12, 2025.

INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

INNO SPACE
SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details