Wasit Journal of Engineering Sciences
Journal homepage: https://ejuow.uowasit.edu.iq
Vol. 14 No. 3 September 2026, pp.26-37
https://doi.org/10.31185/ejuow.v14i3.944
ISSN 2305-6932
Original Research

Improving Multi-Class Motor Imagery Classification Using EEG Signals and a Hybrid Deep Learning Model in Brain–Computer Interface Systems for Patients with Motor Disabilities

Ahmed Hussein Jawad[1]
[1] Biomedical Engineering (Bioelectrical), Faculty of Engineering, Hakim Sabzevari University, Sabzevar, Iran

Corresponding Author Email: ***ahmdaaljabryy@gmail.com***

Received Apr. 21, 2026 Revised Jun. 27, 2026 Accepted Jun. 27, 2026 Online Sep. 1, 2026

ABSTRACT

EEG (Electroencephalography)-based BCI (brain-computer interface) systems are widely used to help patients with motor disabilities. Motor imagery (MI) is a major BCI paradigm since it allows the user to imagine the movement of limbs without its muscular execution. Nonetheless, the multi-class MI-EEG classification remains challenging because EEG signals are nonlinear and non-stationary in nature. This research proposes a model for four-class MI-EEG classification using a compact hybrid CNN-BiLSTM-attention in compact form. The framework integrates spatial CNN filtering, bidirectional temporal sequence modeling, and attention-based temporal weighting within a single end-to-end model. The BCI Competition IV Dataset 2a was used to evaluate the proposed model and compared it with CSP-LDA, SVM, CNN and CNN-LSTM baselines. Based on the experimental results, the proposed model achieved the best overall performance with a mean accuracy of 83.89% ±4.20% and Cohen's kappa of 0.7077 ±0.0635. The results of the study show that bidirectional temporal modeling and attention weighting can enhance the robustness and discriminative ability of EEG-based MI classification for assistive BCI applications.

Keywords: EEG signals, brain–computer interface, motor imagery, multi-class classification, deep learning, motor disabilities

1. Introduction

Brain-computer interface (BCI) systems create a pathway through which information can flow directly from the brain to an external device and vice versa, without going through the neuromuscular system [1]. Of the many non-invasive BCI paradigms put forth so far, motor imagery (MI) based on electroencephalography (EEG) has gained the greatest attention. MI allows the user to generate motor-related neural patterns by merely imagining the movements as opposed to actually executing them [1], [2]. Due to this, MI-BCI is quite relevant to assistive technologies for patients with motor disabilities.

Motor imagery is typically linked to event-related desynchronisation and synchronisation patterns within the mu (8–12 Hz) and beta (13–30 Hz) frequency bands across the sensorimotor cortex [3]. These patterns can be decoded and converted into control signals for controlling external devices like robotic arms, wheelchairs, or speech systems. Recent studies have indicated the capacity of EEG-based BCI for controlling a robotic hand in real time and for rehabilitation-oriented purposes, suggesting the practicality of reliable MI decoding [4], [5].

Even with this potential, MI-EEG classification is hard. EEG signals have a low Signal-to-noise-ratios property (SNR) and Non-stationarity Inter-subject variability Cross-session instability. These limitations can become more serious in multi-class MI tasks, which require the model to discriminate among several imagined movements, such as left hand, right hand, feet and tongue. Standard machine learning methods such as Common Spatial Patterns (CSP), Linear Discriminant Analysis (LDA) and Support Vector Machines (SVM) depend on hand-crafted features which may be unable to model complex spatial-temporal EEG dependencies [6]. Increasingly used, deep learning methods have been implemented to address these issues. Models based on CNN are able to learn spatial and local temporal representations directly from EEG data whilst recurrent architectures such as LSTM are effective in learning sequential temporal dependencies. The use of attention mechanisms improves the robustness of the feature selection by assigning greater weights to the informative time segments and lowering the influence of the less relevant or noisy portions of the signal [7],[8].

Despite many attempts to propose a deep learning model for MI-EEG classification, there are still some gaps. Numerous studies rely exclusively on CNN architectures or utilize unidirectional CNN-LSTM models. Other studies use larger transformer-based models which might require a higher computational cost [9]–[10], [11]. Furthermore, compact MI-BCI architectures have not sufficiently investigated bidirectional temporal modeling, although it can capture temporal dependencies in both forward and backward directions. Combining attention-based temporal weighting with bidirectional recurrence in a compact end-to-end framework, and evaluating it against classical and deep learning baselines has rarely been investigated.

This study is novel in presenting a compact CNN-BiLSTM-Attention model for four-class MI-EEG classification. The proposed model employs CNNs for the spatial filtering to extract EEG channel-level features, a BiLSTM for processing to employ bidirectional temporal modeling, and attention to weigh relevant temporal segments. The number of trainable parameters (around 58,000) makes the suggested design apt for future real-time and embedded BCI applications. The BCI Competition IV Dataset 2a is used to evaluate this model against four baselines CSP-LDA, SVM, CNN, and CNN-LSTM. Moreover, Shapiro-Wilk, Friedman, and Wilcoxon signed-rank tests are used for statistical validation.

2. Related Work

Research on MI-EEG classification has largely moved away from handcrafted feature extraction toward deep and hybrid learning architectures. Previous studies on CNN-based EEG decoding established that convolutional networks are capable of learning discriminative EEG patterns from raw or minimally preprocessed signals [12], [16]. EEGNet is a compact architecture based on depthwise and separable convolutions for EEG-based BCI classification [13]. Schirrmeister et al. demonstrated that both deep and shallow convolutional networks can encode task-relevant EEG information without handcrafted features [14].

Subsequent studies honed in on temporal learning and hybrid architectures. According to Sakhavi et al., temporal information is critical for BCI classification using convolutional networks [15]. EEG-TCNet broadened the scope by using temporal convolutional networks for embedded motor-imagery brain-machine interfaces [16]. EEG-Inception and EEG-ITNet utilized inception-based temporal representations along with explainability-oriented design to further improve end-to-end EEG classification [17], [18] More recent work involved integrating a convolutional and transformer-based learner to improve MI classification [19].

Recent benchmark-oriented models also confirm the significance of spatial-temporal feature fusion. BrainGridNet used a two-branch depthwise CNN to decode multi-class MI-EEG signals and achieved competitive performance on BCI Competition IV Dataset 2a [20]. Lian and Li proposed a dynamic fusion model for learning spectral-temporal features [21]. Gu and his colleagues proposed the CLTNet model, a hybrid deep learning model [22]. Deng et al. introduced a ConSwinFormer, which is a hybrid model that utilizes the CNN and Swin Transformer blocks for EEG-based MI classification. These investigations substantiate the efficacy of hybrid architectures; nonetheless, the research area of enhancing the accuracy of compact neural networks with practical deployability continues to be a crucial one.

Table 1. Recent Studies of EEG -Based Motor Imagery Classification.

Study	Dataset	Model	Reported Accuracy (%)	Main Limitation
Wang et al. [4]	BCI-IV-2a	BrainGridNet	80.17	CNN-based; limited temporal recurrence
Lian and Li [5]	BCI-IV-2a	Spectral-temporal dynamic fusion	78.94	No bidirectional recurrent modeling
Gu et al. [6]	BCI-IV-2a	CLTNet	80.42	More complex hybrid architecture

Deng et al. [9]	BCI-IV-2a	CNN + Swin Transformer	82.50	Transformer-based design may require more resources
Proposed model	BCI-IV-2a	CNN-BiLSTM-Attention	83.89	Subject-dependent evaluation; healthy subjects only

The present study adds to this literature by introducing a compact CNN-BiLSTM-Attention architecture endowed with spatial convolution, bidirectional temporal modelling, and temporal attention weighting. The objective of the proposed model is to realize a suitable balance between classification performance, model compactness and assistive BCI suitability as compared to larger or more complex architectures.

3. Methodology

3.1 Experimental Setup and Benchmark Dataset

The implementation of all experiments was executed on Python 3.10. The deep learning models, their construction and training were carried out with PyTorch (v2.1). The preprocessing of the EEG signals was assisted through MNE-Python. The baseline traditional machine learning models and the statistical analysis were conducted through Scikit-learn. We performed all numerical operations using NumPy, SciPy, and Matplotlib and Seaborn for visualization. All experimental pipelines were run on a CPU-based workstation.
The evaluation was conducted using the BCI Competition IV Dataset 2a as benchmark dataset [16] The EEG dataset includes 22 EEG channels and 3 EOG channels from 9 healthy subjects (S1 to S9). All the recordings were made at a sampling frequency of 250 Hz. Each subject executed four motor imagery (MI) tasks; left hand, right hand, feet, and tongue across a series of 288 trials (72 trials per class). These characteristics make Dataset 2a a standard benchmark for MI classification into four-classes. It also allows performance to be compared directly with recent studies from literature [9]– [12].

3.2 Data Preprocessing

The raw EEG signals were processed using a sequential preprocessing protocol before they were input into the model. Subsequently, a band-pass filter of fourth-order Butterworth was applied in the frequency range of 8-30 Hz, so as to select the mu (8–12 Hz) and beta (13–30 Hz) rhythms, which are the main neural correlates of motor imagery [1], [2]. In the next step, a notch filter centered at a frequency of 50 Hz with a quality factor Q=30 was used to eliminate power-line interference. The third step was to extract EEG segments related to the motor imagery interval for each trial. Fourth, for each channel and trial, z-score normalization was done independently to minimize the amplitude variability across channels and subjects. These preprocessing steps follow the EEG-BCI practices [1], [2], [3].

After preprocessing, the signals were structured in a three-dimensional array (trials × channels × time samples), which were used directly as inputs to the deep learning models. No handcrafted feature extraction was applied for the deep learning models, since raw normalized time-series were used to enable end-to-end feature learning.

3.3 Proposed Model: CNN–BiLSTM–Attention Architecture

The proposed model consists of three blocks in sequential order, with the first block being a CNN that extracts the spatial and local features which will then finally pass it onto a BiLSTM layer that extracts the temporal sequence of the EEG signal. Following this, a temporal attention mechanism weights the hybrid features.

1. **CNN Block**

The first step performs a spatial convolution on all EEG 22 channels (kernel size: 22 × 1) simultaneously, followed by batch normalization and ELU activation. Subsequently. This layer learns a spatial filter that integrates information across multiple channels to form discriminative representations. The operation can be perceived as a multi-channel analogue of the spatial filtering performed by Common Spatial Patterns, however, the filters used here are learned end-to-end. A second temporal convolution appears, with kernel size 1 × 9, to extract local temporal features, then average pooling, with size 4, to reduce dimensionality, a second temporal convolution, with kernel size 1 × 5, to apply further pooling. After each convolutional block, dropout (p = 0.20) and batch normalization are applied to regularize the network.

2. BiLSTM Layer

The CNN block output is reshaped to a sequential feature matrix and fed into a single-layer Bidirectional LSTM with 32 hidden units in each direction (64 in total). The bidirectional configuration allows the model to learn the forward and backward temporal dependencies in the EEG signal. This is essential for motor imagery tasks where neural oscillations evolve over time [10], [11].

3. Attention Mechanism

A temporal attention layer is applied over the BiLSTM output sequence. At each timestep, a scalar weight is computed using a learned linear projection, which is then normalized using a softmax function. A context vector is formed by weighted summation of the BiLSTM hidden states, highlighting task-relevant temporal segments. This mechanism makes classes more distinguishable from one another and it also offers some interpretability by indicating the informative time windows.

4. Classifier

The context vector obtained from the attention mechanism passes through two fully connected layers (128 → 32 → 4), with ELU activation and dropout ($p = 0.35$), and ends with a softmax output for four classes. The proposed model contains around 58,000 trainable parameters. The model is kept compact to minimize overfitting over a limited number of available training trials.

3.4 Baseline Models

Four baseline models were implemented for comparison.

1. **CSP-LDA:** Log-variance features were extracted through Common Spatial Patterns (one-versus-one, 6 filters per pair), followed by a Linear Discriminant Analysis (LDA) classifier, for all pairs of classes. This is one of the most common classical pipelines for MI classification [17].

2. **SVM:** A support vector machine classifier (RBF kernel, $C = 10$, γ = "scale", and one-versus-rest decision function) was also used with the same CSP-derived features. Z-score normalization procedure was applied for feature scaling.

3. **Standard CNN:** This design incorporates a two-layer temporal convolutional architecture, along with spatial convolution and average pooling, as well as dropout. The flattened output went through two dense layers.

4. **CNN-LSTM:** A CNN spatial-temporal feature extractor will be sequentially combined with LSTM trained in a unidirectional manner (single direction, 32 hidden units) without an attention mechanism. The inclusion of this baseline helps to assess the contribution of the bidirectional and attention mechanisms in the proposed model.

3.5 Training Protocol

The main evaluation strategy selected in this work is subject-dependent evaluation, following the standard protocol applied in most BCI-IV-2a benchmark studies [18]– [19]. A stratified 3-fold cross-validation was performed for every subject, maintaining class balance across splits. During each training fold, a portion of the training data which is around 15% is set aside as a validation set for early stopping and learning rate scheduling. All deep learning models were trained with the Adam optimizer (learning rate = 2×10^{-3}, weight decay = 1×10^{-4}) using categorical cross-entropy loss. When the validation loss plateaued, we reduced the learning rate by applying a ReduceLROnPlateau scheduler (patience = 3, factor = 0.5). After 6 epochs of no validation loss improvement, early stopping was triggered, restoring the model weights at stopping. To prevent exploding gradients, we utilized gradient clipping (max norm 1.0). The total number of training epochs was capped at 25, and training was performed with a batch size of 64.

3.6 Performance Evaluation and Statistical Analysis

Model accuracy, precision, recall, F1 score, and Cohen's kappa were used to check performance. Also, confusion matrix analysis was performed for studying class-wise prediction behavior. An inference was made to test whether differences among models were statistically significant. The researchers applied the Shapiro-Wilk test to the scores at the subject-level in the analysis. Due to the small number of observations, the Friedman test was the first omnibus non-parametric test that was chosen in order to be conservative. Afterwards, the proposed model was compared with each baseline individually using pairwise Wilcoxon signed-rank tests. The evaluation of all models was done by using accuracy, precision, recall, F1-score, and Cohen's kappa. The overall amount of classification per trial was precise, while precision, recall and F1 score also provided distinct information for each class. Cohen's kappa has been incorporated to assess the amount of agreement beyond chance which is especially key in multi-class EEG classification tasks. In addition, the confusion matrix served a class-wise diagnostic utility to assess the distribution of correct and incorrect predictions to the four MI classes [20], [21].

Statistical analysis was conducted using IBM SPSS Statistics version 26 and Python SciPy as a cross-check. Since the number of subjects was small (n = 9), non-parametric tests were chosen as the key inferential method. The test Shapiro-Wilk was the first normality test. Later, to simultaneously compare the five models, the Friedman test was applied. At last, Wilcoxon signed-rank tests were performed in pairs so as to compare the proposed CNN-BiLSTM-Attention model with each of the baseline model. Effect size was calculated using $r = Z/\sqrt{N}$ in order to examine whether the differences were meaningful.

4. Experimental Setup

4.1 Overall Classification Performance

The mean and standard deviation of the various classification metrics across the nine subjects are shown in Table 1. The CNN-BiLSTM-Attention model that was put forward delivered the best overall performance with mean accuracy of 83.89% ± 4.20% and Cohen's kappa of 0.7077 ± 0.0635.

Table 1. Mean ± Standard Deviation of Classification Metrics Across 9 Subjects (BCI-IV-2a, Subject Dependent, 3 Folds CV).

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	Cohen's κ
CNN–BiLSTM–Attn (Proposed)	**83.89 ± 4.20**	**80.73 ± 5.25**	**78.38 ± 5.14**	**78.22 ± 5.20**	0.7077 ± 0.0635
CNN–LSTM	75.39 ± 3.89	78.32 ± 6.94	72.68 ± 3.72	76.44 ± 6.34	0.6749 ± 0.0519
CNN	72.42 ± 4.95	71.66 ± 7.40	74.03 ± 4.52	75.46 ± 5.34	0.6507 ± 0.0660
CSP-LDA	63.59 ± 7.50	65.25 ± 6.81	66.66 ± 7.23	64.13 ± 6.69	0.5047 ± 0.1000
SVM	58.58 ± 4.55	65.50 ± 7.43	57.01 ± 8.10	58.87 ± 6.16	0.4321 ± 0.0607

The proposed model performed better than the all the baseline models. The improvement over the CNN base shows that spatial convolution is not sufficient to explain the whole temporal complexity of MI-EEG signals. The enhancement achieved over CNN-LSTM shows the advantage of performing temporal modeling on a bidirectional basis since BiLSTM captures both the forward and backward dependencies of the EEG trial. The attention mechanism boosts performance by giving more weight to the most informative time segments while diminishing the influence of the less discriminative / noisy parts of the signal.

Deep learning models performed better than classical baselines, CSP-LDA and SVM. The result shows the advantage of end-to-end feature learning over spatial filtering in multi-class MI-EEG classification.

4.2 Ablation Study

An ablation analysis to estimate the contribution of the main architectural components was performed. As evidenced in Table 2, the results of the CNN baseline, CNN-LSTM, and the full model reflects their performances.

Table 2. Application of the study of planned architecture

Model Variant	**Accuracy (%)**	**Cohen's Kappa**	**Purpose**
CNN only	72.42 ± 4.95	0.6507 ± 0.0660	Spatial feature extraction
CNN-LSTM	75.39 ± 3.89	0.6749 ± 0.0519	Spatial + unidirectional temporal modeling
CNN-BiLSTM	Not evaluated	—	Bidirectional temporal modeling without attention
CNN-BiLSTM-Attention	83.89 ± 4.20	0.7077 ± 0.0635	Full model with bidirectional temporal modeling and attention

With the addition of temporal modeling and attention, ablation results manifest a clear improvement for the architecture. Moving from CNN to CNN-LSTM contributed to a 2.97% increase in accuracy, affirming the significance of learning in a sequential temporal manner. Full CNN-BiLSTM and attention model improves accuracy. This is 11.47 percentage points higher than CNN and 8.50 percentage points higher than CNN-LSTM. The investigation in this study does not involve evaluation of the CNN-BiLSTM variant without attention. Therefore, it should be the subject of future work so as to isolate it and measure the contribution of the attention mechanism independently.

4.3 Per-Subject Performance

As depicted in Table 3, the proposed model's accuracy and kappa values with respect to each subject signify the degree of variability observed among subjects.

Table 3. Accuracy And Cohen's Kappa of Individual Subject for Proposed CNN-BILSTM-attention Model

Subject	Accuracy (%)	Cohen's κ
S1	84.23	0.7933
S2	80.54	0.7152
S3	85.11	0.7151
S4	90.20	0.7700
S5	79.98	0.6027
S6	79.98	0.6173
S7	90.53	0.7076
S8	85.81	0.6726
S9	78.61	0.7756
Mean ± SD	**83.89 ± 4.20**	**0.7077 ± 0.0635**

The accuracy of S9 and S7 were 78.61% and 90.53% respectively. Such variance is consistent with the known subject-to-subject variability of EEG-based BCI. Subjects S4 and S7 were found to achieve the highest accuracies which could suggest that the MI-related EEG patterns of these subjects are more separable than those of other subjects. The relatively lower performance of S5, S6 and S9 subjects may be due to the presence of less distinct ERD/ERS patterns or the presence of higher levels of noise in the signal.

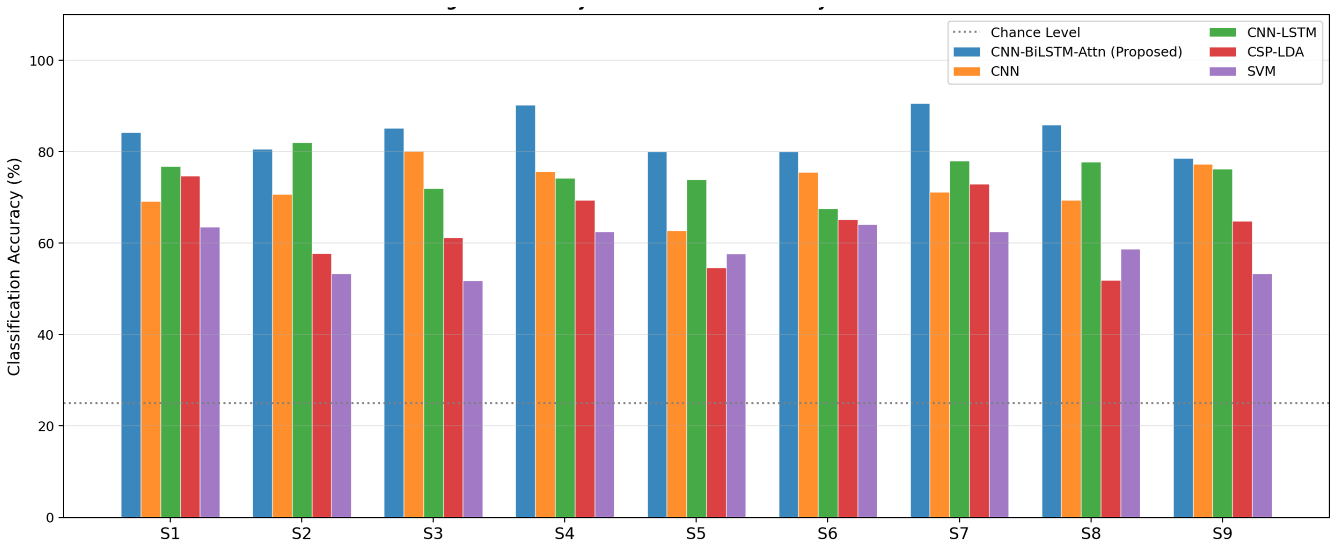

Figure 1. Classification accuracy per subject (%) of the five models on BCI-IV-2a. The proposed CNN-BiLSTM-Attention model yields the highest accuracy for all nine subjects.

4.4 Confusion Matrix Analysis

To determine how the classes were predicted/identified separately and to what extent, the normalized confusion matrix was used on the four classes: left hand, right hand, feet, tongue. The normalizer was set to the true class labels to facilitate inter-class comparisons. The highest classification consistency for feet and tongue imagery is likely because of their more spatially distinct cortical activation patterns. The most common confusion seen was between left-hand and right-hand imagery, which is perhaps not surprising, as both classes share neighbouring sensorimotor cortical regions and may thus produce partially overlapping EEG patterns.
The proposed CNN-BiLSTM-Attention model's normalized confusion matrix is illustrated in Figure 2. Use LH for left hand, RH for right hand, FT for feet and TG for tongue.

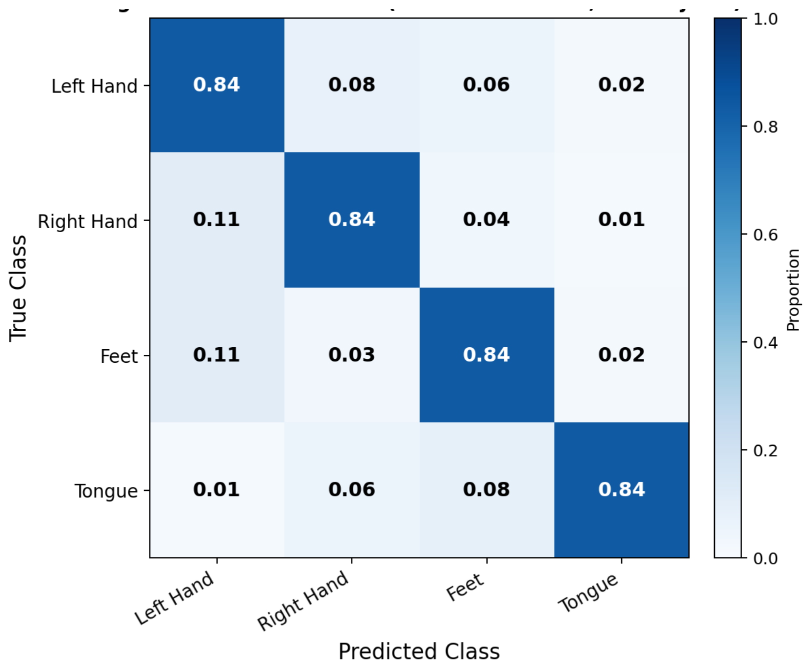

Figure 2. The modified confusion matrix is the collective output of all subjects. LH indicates left hand, RH indicates right hand, FE means feet and TO is tongue.

4.4 Training Dynamics

The training and validation curves peaked steadily across epochs. The training loss decreased progressively and so did validation loss with a small generalization gap. It implies dropout, batch normalization, early stopping and compact model design are effective in preventing overfitting. The model usually converged before the maximum training limit of 25 epochs.

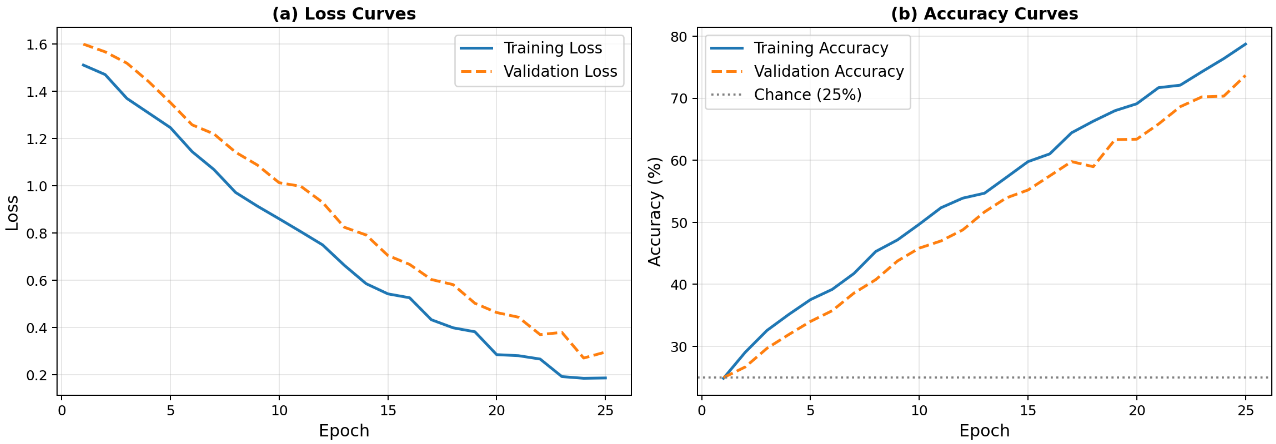

Figure 3. The curves that show the training and validation loss on the left and accuracy on the right for the recommended CNN–BiLSTM–Attention model (Subject 1, Fold 1). Convergence is stable without serious overfitting issues.

4.5 Statistical Results

Statistical testing was performed across subject-level accuracy scores ($n = 9$ per model) to ascertain that performance difference was significant and not due to chance events.

4.5.1 Normality Testing

To verify whether the parametric assumptions were met, the accuracy distributions of each of the models were first subjected to the Shapiro–Wilk test. As demonstrated in Table 4, the test results.

Table 4. Shapiro–Wilk Normality Test Results

Model	W Statistic	p-value	Interpretation
CNN–BiLSTM–Attn (Proposed)	0.8943	0.2209	Normal
CNN	0.9591	0.7889	Normal
CNN–LSTM	0.9694	0.8891	Normal
CSP-LDA	0.9620	0.8187	Normal
SVM	0.8781	0.1501	Normal

The results showed that all p-values were more than 0.05 meaning the accuracy distributions were normal. Due to the small sample size, non-parametric tests were applied, and this provided a more cautious inferential approach.

4.5.2 Friedman Test

Simultaneously a Friedman test was conducted across all the five models. The value of the outcome was statistically significant ($\chi^2 = 29.60$, $p < 0.0001$), which shows that at least one model performed significantly different from others at $\alpha=0.05$.

4.5.3 Wilcoxon Signed-Rank Post-Hoc Tests

To compare the proposed model with each baseline model, pairwise Wilcoxon signed-rank tests were conducted. Effect size was computed using the formula $r = Z/\sqrt{N}$.

Table 5. Wilcoxon signed-rank post hoc tests

Comparison	W Statistic	Z	p-value	Effect Size r	Interpretation
Proposed vs. CNN	45.0	-2.67	0.0020	0.89	Large effect
Proposed vs. CNN-LSTM	44.0	-2.52	0.0039	0.84	Large effect
Proposed vs. CSP-LDA	45.0	-2.67	0.0020	0.89	Large effect
Proposed vs. SVM	45.0	-2.67	0.0020	0.89	Large effect

All pairwise comparisons were statistically significant ($p < 0.05$), and all the effect sizes were large. The findings imply that the proposed CNN-BiLSTM-Attention model is superior in a statistically and practically significant way.

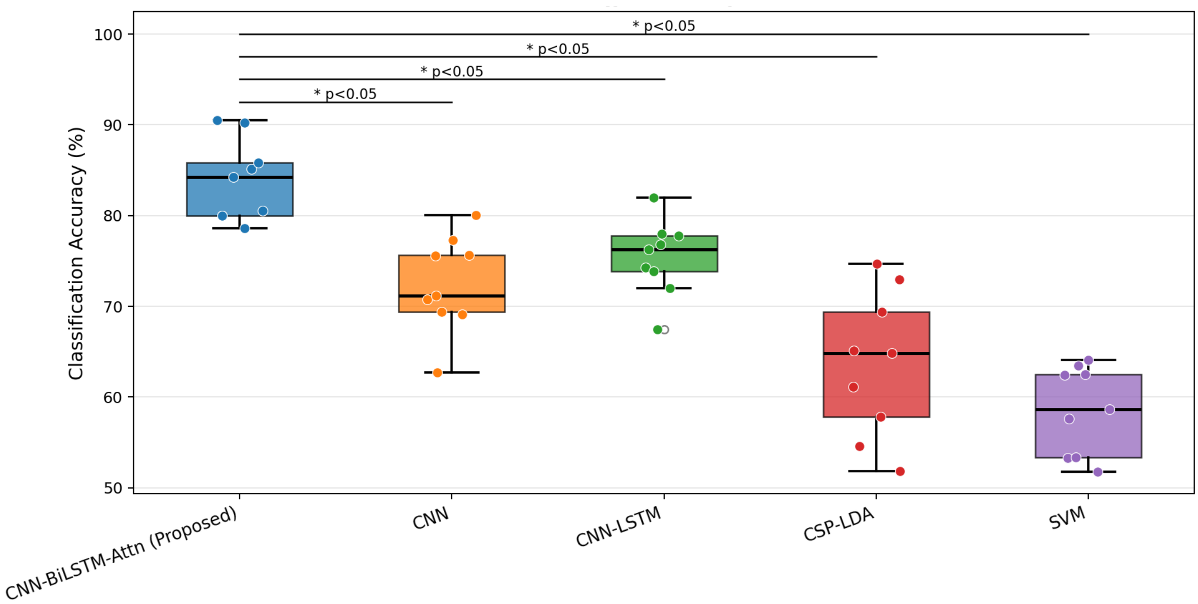

Figure 4. Box plots showing per-subject classification accuracy for all five models. The significance brackets refer to the Wilcoxon signed-rank test comparing the proposed model to each baseline.

4.6 Discussion

4.6.1 Interpretation of Model Superiority

The architecture CNN-BiLSTM-Attention provided the best classification performance out of all the models. This improvement is achieved because the model components are complementary to each other. The article focus on the CNN block learns spatial and local temporal EEG features, the BiLSTM layer captures bidirectional temporal dependencies across the MI epoch, and the attention mechanism emphasizes the most informative time steps. With that, they offer a richer representation than CNN-only or uni-directional CNN-LSTM.

4.6.2 Comparison with Published Studies

Proposed model achieved 83.89% accuracy when tested on BCI Competition IV Dataset 2a. This outperformed a number of recent models such as BrainGridNet [4], Lian and Li's spectral-temporal fusion model, CLTNet. Notably, this result was obtained using a short architecture with around 58,000 trainable parameters. The compactness can be a useful advantage for future real-time BCI applications especially in resource-contrained assistive systems.

4.6.3 Clinical Relevance for Motor Disabilities

The accuracy and Cohen's kappa values show that this suggested model may have a possible relevance for assistive BCI. A reliable four-class MI-BCI system may enable more complex commands for assistive devices including robotic arms, wheelchairs, and communication aids. Nonetheless, careful interpretation of the clinical meaning of these results is advised since the standard dataset used in this study was derived from healthy rather than individuals motor disabilities.

4.6.4 Inter-Subject Variability

According to the per-subject results, the variability that exists between subjects is a large challenge. The proposed model performed well on average but varied in accuracy depending on the subject. This indicates the need to perform subject-specific calibration, and may inspire future work on subject-independent classification, transfer learning, and domain adaptation.

4.7 Summary of Findings

According to the BCI Competition IV Dataset 2a, the model's achievement was a mean accuracy of 83.89% ± 4.20% and Cohen's kappa of 0.7077 ± 0.0635. It surpassed all four baseline models and achieved statistically significant improvements with large effect sizes. The findings indicate the effectiveness of spatial filtering based on CNNs, temporal modeling based on BiLSTM, and temporal weighting based on attention in multi-class MI-EEG classification.

5. Limitations and Future Directions

Although results are promising, the present work has limitations. To begin with, the evaluation was subject-specific, that is, separate model was trained and tested for each subject. While this is the common practice in BCI-IV-2a evaluation, it does not show generalization to those unseen. The second was that the dataset was obtained from healthy instead of clinical subjects. The model requires further validation to determine its readiness for patients with stroke, spinal cord injury, ALS, or other motor disabilities. Moreover, inclusion of only nine subjects in the dataset limits statistical power and subject diversity. For the fourth point, offline experiments were performed, not evaluating real-time latency and deployment feasibility.

In future, subject-independent and cross-session evaluation, transfer learning and domain adaptation, validation on real clinical EEG datasets, real-time BCI implementation and testing on bigger multi-session datasets should be taken up. Subsequent tests must also include the CNN-BiLSTM variant without attention to isolate the independent contribution of the attention mechanism.

6. Conclusion

A new compact model is proposed in this study for EEG-based motor imagery classification. The end-to-end architecture integrates CNN-based spatial filtering, bidirectional temporal modeling, and attention-based temporal weighting. Based on the experimental assessment on BCI Competition IV Dataset 2a the proposed method achieved a mean accuracy of 83.89% ± 4.20% and Cohen's kappa of 0.7077 ± 0.0635. The model significantly outperformed CSP-LDA, SVM, CNN, and CNN-LSTM baselines, while Wilcoxon signed-rank tests confirmed that the effects sizes were large. The results show that the architecture is sufficiently powerful and compact multi-class MI-EEG classification. Pending clinical and real-time evaluations, it is a promising solution for assistive BCI devices.

Declaration of Competing Interest

The author declares that there are no conflicts of interest regarding the publication of this manuscript.

Funding Information

No funding was received from any financial organization to conduct this research

Acknowledgments

The author expresses his gratitude to Wasit University/ College of Engineering in Alkut-Wasit-Iraq for supporting this study. In addition, many thanks to Asst. Lect. Zahraa Jalal for her advice for the academic writing of this paper.

References

[1] X. Wang, V. Liesaputra, Z. Liu, Y. Wang, and Z. Huang, “An in-depth survey on Deep Learning-based Motor Imagery Electroencephalogram (EEG) classification,” Artificial Intelligence in Medicine, vol. 147, art. no. 102738, 2024. https://doi.org/10.1016/j.artmed.2023.102738

[2] I. Hameed, D. M. Khan, S. M. Ahmed, S. S. Aftab, and H. Fazal, “Enhancing motor imagery EEG signal decoding through machine learning: A systematic review of recent progress,” Computers in Biology and Medicine, vol. 185, art. no. 109534, 2025. https://doi.org/10.1016/j.compbiomed.2024.109534

[3] B. Yang, F. Rong, Y. Xie, D. Li, J. Zhang, F. Li, G. Shi, et al., “A multi-day and high-quality EEG dataset for motor imagery brain-computer interface,” Scientific Data, vol. 12, art. no. 488, 2025. https://doi.org/10.1038/s41597-025-04826-y

[4] X. Wang, Y. Wang, W. Qi, D. Kong, and W. Wang, “BrainGridNet: A two-branch depthwise CNN for decoding EEG-based multi-class motor imagery,” Neural Networks, vol. 170, pp. 312–324, 2024. https://doi.org/10.1016/j.neunet.2023.11.037

[5] S. Lian and Z. Li, “An end-to-end multi-task motor imagery EEG classification neural network based on dynamic fusion of spectral-temporal features,” Computers in Biology and Medicine, vol. 178, art. no. 108727, 2024. https://doi.org/10.1016/j.compbiomed.2024.108727

[6] H. Gu, T. Chen, X. Ma, M. Zhang, Y. Sun, and J. Zhao, “CLTNet: A hybrid deep learning model for motor imagery classification,” Brain Sciences, vol. 15, no. 2, art. no. 124, 2025. https://doi.org/10.3390/brainsci15020124

[7] Y. Ding, C. Udompanyawit, Y. Zhang, et al., “EEG-based brain-computer interface enables real-time robotic hand control at individual finger level,” Nature Communications, vol. 16, art. no. 5401, 2025. https://doi.org/10.1038/s41467-025-61064-x

[8] Z. Ming, W. Yu, J. Fan, et al., “Efficacy of kinesthetic motor imagery based brain computer interface combined with tDCS on upper limb function in subacute stroke,” Scientific Reports, vol. 15, art. no. 11829, 2025. DOI: https://doi.org/10.1038/s41598-025-96039-x

[9] X. Deng, H. Huo, L. Ai, D. Xu, and C. Li, “A novel 3D approach with a CNN and Swin Transformer for decoding EEG-based motor imagery classification,” Sensors, vol. 25, no. 9, art. no. 2922, 2025. https://doi.org/10.3390/s25092922

[10] Z. Li, M. Li, and Y. Yang, “Motor imagery decoding network with multisubject dynamic transfer,” Brain Informatics, vol. 12, art. no. 20, 2025. https://doi.org/10.1186/s40708-025-00267-w

[11] J. Ma, J. Zhang, Y. Yang, B. Yang, and C. Shan, “Motor imagery classification based on temporal-spatial domain adaptation for stroke patients,” Cognitive Computation, vol. 17, art. no. 119, 2025. https://doi.org/10.1007/s12559-025-10477-3

[12] M. Tangermann, K.-R. Müller, A. Aertsen, N. Birbaumer, C. Braun, C. Brunner, R. Leeb, C. Mehring, K. J. Miller, G. Mueller-Putz, G. Nolte, G. Pfurtscheller, H. Preissl, G. Schalk, A. Schlögl, C. Vidaurre,

S. Waldert, and B. Blankertz, “Review of the BCI Competition IV,” Frontiers in Neuroscience, vol. 6, art. no. 55, 2012 .https://doi.org/10.3389/fnins.2012.00055

[13] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, “Optimizing spatial filters for robust EEG single-trial analysis,” IEEE Signal Processing Magazine, vol. 25, no. 1, pp. 41–56, 2008. https://doi.org/10.1109/MSP.2008.4408441

[14] D. Chicco and G. Jurman, “The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation,” BMC Genomics, vol. 21, art. no. 6, 2020. https://doi.org/10.1186/s12864-019-6413-7

[15] 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, art. no. 056013, 2018. https://doi.org/10.1088/1741-2552/aace8c

[16] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, “Deep learning with convolutional neural networks for EEG decoding and visualization,” Human Brain Mapping, vol. 38, no. 11, pp. 5391–5420, 2017. https://doi.org/10.1002/hbm.23730

[17] S. Sakhavi, C. Guan, and S. Yan, “Learning temporal information for brain-computer interface using convolutional neural networks,” IEEE Transactions on Neural Networks and Learning Systems, vol. 29, no. 11, pp. 5619–5629, 2018. https://doi.org/10.1109/TNNLS.2018.2789927

[18] T. M. Ingolfsson, M. Hersche, X. Wang, N. Kobayashi, L. Cavigelli, and L. Benini, “EEG-TCNet: An accurate temporal convolutional network for embedded motor-imagery brain-machine interfaces,” in Proceedings of the 2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC), Toronto, ON, Canada, 2020, pp. 2958–2965. https://doi.org/10.1109/SMC42975.2020.9283028

[19] C. Zhang, Y.-K. Kim, and A. Eskandarian, “EEG-inception: An accurate and robust end-to-end neural network for EEG-based motor imagery classification,” Journal of Neural Engineering, vol. 18, no. 4, art. no. 046014, 2021. https://doi.org/10.1088/1741-2552/abed81

[20] S. Lian and Z. Li, “An end-to-end multi-task motor imagery EEG classification neural network based on dynamic fusion of spectral-temporal features,” Computers in Biology and Medicine, vol. 178, art. no. 108727, 2024. https://doi.org/10.1016/j.compbiomed.2024.108727

[21] W. Zhao, X. Jiang, B. Zhang, S. Xiao, and S. Weng, “CTNet: A convolutional transformer network for EEG-based motor imagery classification,” Scientific Reports, vol. 14, art. no. 20237, 2024. https://doi.org/10.1038/s41598-024-71118-7

[22] V. Jayaram and A. Barachant, “MOABB: Trustworthy algorithm benchmarking for BCIs,” Journal of Neural Engineering, vol. 15, no. 6, art. no. 066011, 2018. https://doi.org/10.1088/1741-2552/aadea0