*Wasit Journal of Engineering Sciences*
Journal homepage: https://ejuow.uowasit.edu.iq

Vol. 14 No. 3 September 2026, pp.14-25
https://doi.org/10.31185/ejuow.v14i3.938

ISSN 2305-6932
*Original Research*

# Subject-Independent Four-Class Motor Imagery EEG Classification Using a Hybrid CNN–BiLSTM Network with Channel Attention

**Saif Rasool Abdulhussein** [1]
[1] Biomedical Engineering (Bioelectrical), Faculty of Engineering, Hakim Sabzevari University, Sabzevar, Iran

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

Received Apr. 10, 2026 | Revised Aug. 4, 2026 | Accepted Aug. 4, 2026 | Online Sep. 1, 2026

## ABSTRACT

The classification of MI-EEG is complex due to the noise, non-stationarity and inter-subject variation of these signals. Some of these challenges are magnified in subject-agnostic settings, where a model must generalise to an unseen user without subject-specific tuning. This study presents a hybrid deep-learning architecture that combines convolutional neural networks (CNNs), bidirectional long short-term memory (BiLSTM), and squeeze-and-excitation feature-channel attention for subject-independent four-class MI-EEG classification. The CNN extracts local spatiotemporal representations, the BiLSTM models forward and backward temporal dependencies, and the attention module recalibrates informative feature maps before classification. The framework was evaluated on BCI Competition IV Dataset 2a using leave-one-subject-out cross-validation across nine subjects. It achieved a mean accuracy of 69.75% and Cohen's kappa of 0.5967, exceeding EEGNet by 6.94 percentage points and ShallowConvNet by 9.72 percentage points. Ablation analysis showed that removing attention reduced accuracy by 2.47 percentage points. Although the proposed model has higher computational cost and latency than EEGNet, it provides a favorable accuracy–complexity trade-off for subject-independent multi-class BCI applications in which decoding performance is prioritized over minimal latency.

**Keywords**: Motor imagery EEG; subject-independent classification; brain–computer interface; CNN–BiLSTM; channel attention.

## 1. Introduction

Brain-computer interface (BCI) systems interface directly with brain activity to control external devices without using peripheral neuromuscular output. Of all the modalities of neuroimaging, electroencephalography (EEG) is especially appropriate for BCI development, as it is non-invasive, has high temporal resolution and can be acquired by relatively portable and low-cost equipment. Mental imagery (MI) is one of the most intensely studied EEG-based BCI paradigms in which a user imagines executing a movement without actually doing it. MI-EEG decoding can assist communication and motor-neurorehabilitation applications in people with impaired voluntary movement [1], [2].

MI-EEG classification that is reliable remains a challenge despite this potential. The signal-to-noise ratio of EEG signals is low, and they are non-stationary, vulnerable to physiological and environmental noise (artifacts), and vary considerably across recording sessions and subjects. These features constrain the extraction of stable discriminative patterns, particularly when the task is four motor-imagery classes rather than a binary decision. Recurrent methodological issues such as inconsistent evaluation protocols and poor cross-subject generalization are common across the literature on MI decoding based on deep learning [3].

Consequently, there is a fundamental difference between subject-dependent and subject-independent classification. By subject dependence systems we mean systems that need to be trained and calibrated on each single user. They can produce good within subject performance. However, they require dedicated recording sessions followed by retraining before being put into practice. Instead, subject-independent systems try to classify recordings from unseen subjects with no calibration. This setting may be operationally attractive but is considerably more challenging because the model will have to learn representations that are robust with respect to differences in neural patterns and signal distributions across subjects [4], [5].

In multi-class settings, this challenge is more severe. The BCI Competition IV Dataset 2a includes four MI tasks which are left hand, right hand, feet and tongue of 9 subjects. Consequently, it is a standard benchmark foe four-class MI decoding evaluation. According to the leave-one-subject-out (LOSO) protocol, to classify all the trials from a held-out subject, the subject's data should not be used for training. This provides a strict calibration-free generalization test [6].

By leveraging the capacity of deep learning to learn hierarchical representations from multichannel EEG, reliance on handcrafted feature engineering is minimized. Convolutional neural networks (CNNs) can learn local temporal and spatial patterns, and EEG-specific models, such as EEGNet achieve competitive decoding with a very small number of parameters [7], [8]. Importantly, long-range distributed correlations across an MI epoch may not be fully captured by purely convolutional architectures.

Bidirectional long short-term memory (BiLSTM) networks help to model forward and backward dependency in a feature sequence. Consequently, hybrid CNN-recurrent architectures are appealing, as convolutional layers can encapsulate local spatiotemporal patterns while the recurrent encoder can combine their temporal context [9]. Moreover, attention mechanisms reject weak or redundant learned features and modulate more informative ones. The principle has been applied in studies concerned with the selection of channels and the re-weighting of features in MI-EEG decoding [10], [16].

In light of these factors, this research proposes a hybrid CNN–BiLSTM network that integrates squeeze-and-excitation feature-channel attention for subject-independent four-class MI-EEG classification. The novelty does not lie in claiming that CNN, BiLSTM or attention are new, but in the systematic integration and evaluation of complementary modules. The research work (1) examines the architecture under a stringent nine-fold LOSO protocol (2) quantifies the contribution of the CNN, BiLSTM and attention modules through ablation analysis (3) employs subject-wise statistical comparisons and (4) reports computational complexity and inference latency along with accuracies. The joint evaluation considers the trade-off between the proposed model's decoding performance and latency of compact architecture EEGNet.

## 2. Related Work

Early MI-EEG systems typically used pipelines where features were extracted with some handcrafted method, especially common spatial pattern filters, followed by LDA or SVM. While these approaches are still considered important baselines, their performance largely depends on frequency-band selection, spatial-filter estimation, and preprocessing choices. They also respond sensitively to EEG non-stationarity and subject-specific variability [3].

Subsequent focus was on the learned representation as an end–ends convolutional model. DeepConvNet showed that hierarchical convolutional filters can directly decode EEG and EEGNet adopted depthwise and separable convolutions to achieve a compact structure suitable for EEG [7], [8]. The above methods illustrate a central design tension; larger, deeper models can increase representational capacity while compact architectures provide substantially lower latency and memory demand.

Cross-subject classification adds a further domain-shift problem. Kwon et al. showed that deep convolutional representations can improve subject-independent BCI decoding, and Zhang et al. proposed a hybrid neural framework for learning transferable EEG representations [4], [5]. Recent domain-generalization work has attempted to reduce subject-related distribution differences through knowledge distillation, spectral feature fusion, and correlation alignment [12]. Transfer-learning frameworks have also been evaluated under LOSO protocols to improve adaptation and reduce between-subject variability [13].

Hybrid and multi-scale networks aim to combine complementary EEG characteristics. Tang et al. used multi-scale convolutions and feature enhancement to capture patterns at different temporal resolutions [11]. Recurrent

components provide an additional mechanism for representing sequential structure. Hou et al. combined attention-guided BiLSTM modeling with deep feature extraction, supporting the use of bidirectional recurrence when discriminative information is distributed over time [9].

Attention-based approaches are motivated by the unequal relevance of channels and learned feature maps. Effective channel-selection methods can reduce redundant input information [10], while feature-reweighting modules can suppress noisy temporal and channel-related representations within a CNN [16]. In the present architecture, squeeze-and-excitation attention is applied to learned CNN feature maps. Consequently, its weights are interpreted as feature-channel importance and not as a direct physiological localization of individual EEG electrodes [17].

The most recent literature increasingly combines attention with transfer learning, meta-learning, domain adaptation, or Transformer encoders. Meta-transfer learning has been used to improve cross-subject MI decoding with limited calibration data [14], while spatiotemporal Transformer models have been developed for zero-calibration EEG classification across multiple benchmark datasets [15]. These approaches highlight the value of long-range dependency modeling but may introduce additional optimization and computational requirements.

The remaining need is therefore not simply for another high-capacity architecture, but for an evaluation that jointly considers generalization, component contribution, and computational demand. The present work proposes a single four-class LOSO architecture, that blends convolutional feature extraction, BiLSTM sequence modeling and feature-channel attention. We also explicitly compare accuracy, statistical significance, parameter count, FLOPs, model size and inference latency.

## 3. 3. Methodology

### 3.1. Dataset Description

The experiments incorporated BCI Competition IV Dataset 2a, the public benchmark for 4-class motor-imagery classification [6]. The dataset comprises of EEG recordings from nine healthy subjects conducted over two sessions on separate days. The analysis of the current study used the labeled 288-trial session for each subject, as was the fold size discussed below. Every session consisted of six runs (48 trials). The researchers gave respondents four balanced classes for imagery: left hand, right hand, feet, and tongue.

Scalp electrodes recorded EEG from 22 channels according to the international 10–20 system. Three monopolar electrooculography (EOG) channels were used to monitor ocular activity. The sampling rate of signals was 250Hz. The subject in each class provided 72 trials which provided a chance level accuracy of 25% for the four-class task.

**Table 1.** Class distribution in the analyzed session of BCI Competition IV Dataset 2a.

| Class | Code | Trials per subject |
|---|---|---|
| Left hand | LH | 72 |
| Right hand | RH | 72 |
| Feet | FT | 72 |
| Tongue | TG | 72 |
| Total | — | 288 |

### 3.2. EEG Preprocessing

Prior to model training a uniform preprocessing pipeline applied. Initially, a Butterworth Filter of 4th order was applied on each EEG channel and this was done from 0.5 Hz to 40 Hz bandwidth. This range maintains the μ (8–12 Hz) and β (13–30 Hz) rhythms usually linked to motor imagery, and mitigates slow drift and high-frequency noise.

In correcting ocular artifacts for this EEG study, independent component analysis was used with the EOG channels as references. Any epoch where the residual amplitudes in any EEG channel exceed ±100μV, was rejected. Next, you applied z-score normalization channel-wise. For every LOSO fold, we estimated the mean

and standard deviation using the training subjects exclusively and then applied them to the validation and held-out test data.

$$x_{\text{norm},c,t} = \frac{x_{c,t} - \mu_c}{\sigma_c + \varepsilon} \qquad (1)$$

Here, $\mu_c$ and $\sigma_c$ denote the training-set mean and standard deviation of EEG channel c, and ε is a small numerical constant. Motor-imagery epochs were extracted from 0.5 to 2.5 s after cue onset. At 250 Hz, each epoch contained 500 samples and was represented as a 22 × 500 matrix.

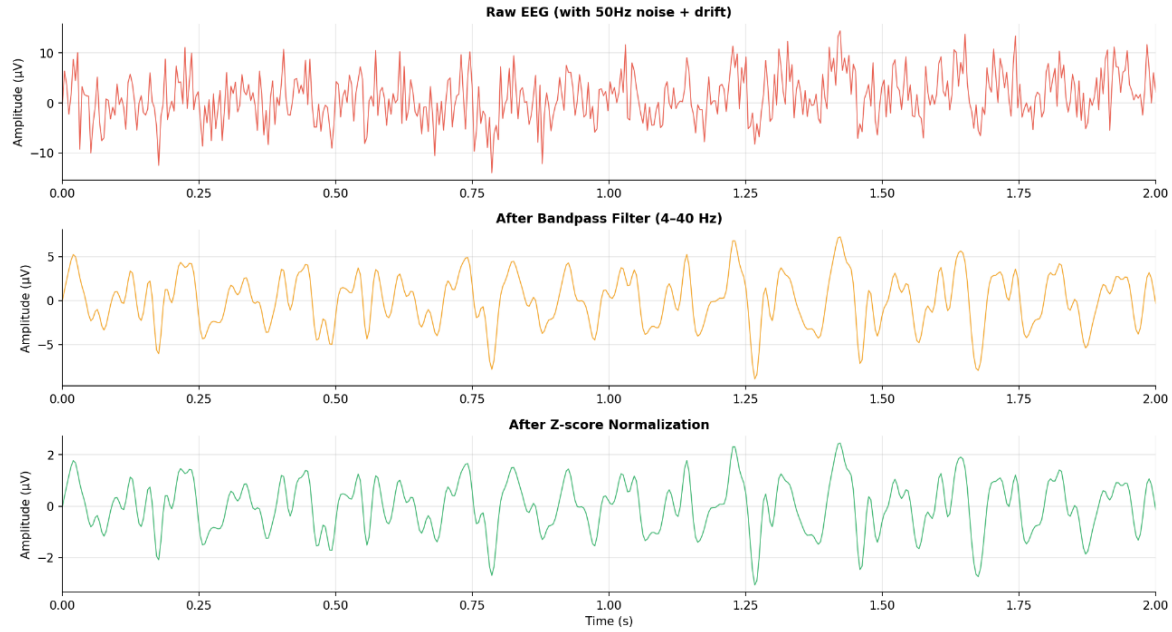


**Figure 1.** An illustration of the preprocessing pipeline showing a representative raw EEG segment, the signal after 0.5–40 Hz band-pass filtering, and the normalised signal used for epoch extraction.

### 3.3. Proposed CNN–BiLSTM Architecture

The proposed framework merges convolutional feature extraction, a squeeze-and-excitation feature-channel attention mechanism, bidirectional temporal modeling, and a four-class classifier. The input tensor is of shape B × 1 × 22 × 500; here, B is the batch size.

#### 3.3.1. CNN Feature Extractor

The CNN is made of three blocks in sequence. The first block obtains 32 temporal filters of kernel size 1 × 25. It then applies batch normalization together with an exponential linear unit (ELU). The second block is a depthwise spatial convolution with kernel size 22 × 1. The convolution resulting in 64 feature maps learns cross-channel spatial filters. Followed by batch normalization, ELU activation, average pooling of 1 × 4, and dropout of 0.25. The third block employs a separable temporal convolution with a kernel size of 1 × 15 and generates 64 output feature maps. The output is then subjected to batch normalization, followed by an ELU activation, average pooling of 1 × 4, and dropout of 0.25. This design captures both temporal and spatial structure while controlling the parameter growth from fully-connected convolutions.

#### 3.3.2. Feature-Channel Attention

A squeeze-and-excitation (SE) module is applied to the 64 learned CNN feature maps [17]. Let F_c(t) denote feature map c at temporal index t and let T′ be the temporal length after convolution and pooling. Global temporal averaging produces the channel descriptor

$$z_c = \frac{1}{T'}\sum_{t=1}^{T'} F_c\,(t) \qquad (2)$$

The descriptor vector is passed through a bottleneck excitation function:

$$\mathbf{s} = \sigma\big(\mathbf{W}_2\,\delta(\mathbf{W}_1\mathbf{z})\big) \qquad (3)$$

The first learnable matrix reduces the C-dimensional descriptor to C/r units, and the second expands it back to C units. Here, C = 64 is the number of feature maps, r = 16 is the reduction ratio, δ denotes ReLU, and σ denotes the sigmoid function. The recalibrated feature maps are

$$\tilde{F}_c(t) = s_c F_c(t) \qquad (4)$$

The SE module therefore assigns larger weights to more informative learned feature maps and suppresses weak or redundant activations. These weights should not be interpreted as direct electrode-level importance.

### 3.3.3. BiLSTM Encoder and Classification Head

The attention-recalibrated feature sequence is passed to a two-layer BiLSTM encoder. Each direction contains 128 hidden units, producing a 256-dimensional output at each sequence step. A dropout rate of 0.30 is applied between recurrent layers. The final sequence representation is processed by a fully connected layer from 256 to 128 units, ELU activation, dropout of 0.30, and a four-unit output layer. Softmax converts the logits into class probabilities. The complete model contains approximately 487,000 trainable parameters.

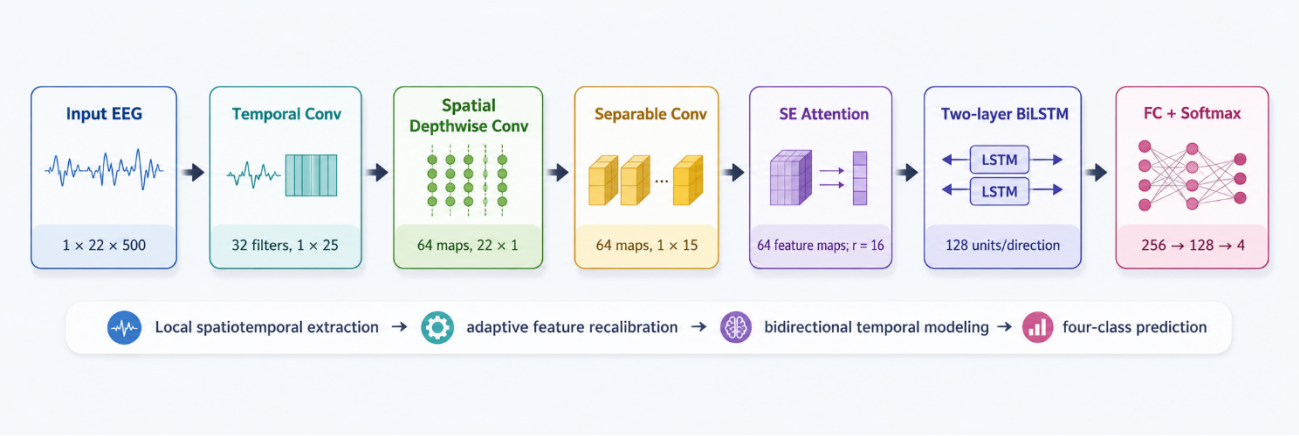


**Figure 2.** Revised architecture of the proposed CNN–BiLSTM model. SE attention recalibrates the 64 learned CNN feature maps before bidirectional temporal modeling and four-class classification.

## 3.4. Training Configuration

The network was implemented in PyTorch 2.1 and trained using the Adam optimizer with an initial learning rate of $1 \times 10^{-3}$. A ReduceLROnPlateau scheduler reduced the learning rate by a factor of 0.5 when validation loss did not improve for 10 consecutive epochs, with a minimum learning rate of $1 \times 10^{-6}$. Training used a batch size of 64, a maximum of 200 epochs, categorical cross-entropy loss, and early stopping with a patience of 20 epochs. A validation subset drawn exclusively from the eight development subjects was used for learning-rate scheduling and early stopping; the held-out subject was not used for preprocessing statistics, optimization, or model selection. Training was performed on an NVIDIA RTX 3090 GPU with mixed-precision computation.

## 3.5. Evaluation Protocol and Metrics

Subject-independent performance was evaluated using nine-fold leave-one-subject-out cross-validation. In each fold, 2,304 trials from eight subjects formed the development set, while all 288 trials from the remaining subject formed the test set. The procedure was repeated until every subject had served once as the unseen test subject. The performance values that were reported were averages across the nine held-out subjects.

Accuracy was calculated as the proportion of correctly classified trials:

$$\text{Accuracy} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{1}\,(\hat{y}_i = y_i) \qquad (5)$$

Cohen's kappa was used to correct observed agreement for chance:

$$\kappa = \frac{p_o - p_e}{1 - p_e} \qquad (6)$$

where $p_o$ is observed agreement and p_e is the agreement expected by chance. Aggregated confusion matrices were additionally examined to describe class-wise recall and residual confusion patterns.

### 3.6. Baselines, Ablation, and Statistical Analysis

The EEGNet [8] and ShallowConvNet [7] served as models for comparison. The same preprocessing and LOSO folds were used for training both baselines. They included a CNN-only model, a BiLSTM-only model, and a CNN-BiLSTM model without attention. Subject-wise accuracy was assessed using an exact two-sided Wilcoxon signed-rank test for pairwise differences. The three planned comparisons underwent Holm adjustment application. The computational analysis looked at trainable parameters, the model size, approximate floating-point operations, and implementation-specific inference latency in milliseconds per trial. All latency values have been done under the same hardware setup and should be interpreted in relative terms, not hardware-independent terms.

## 4. Results and Discussion

### 4.1. Overall Classification Performance

Table 2 provides a summary of the proposed model's LOSO performance and the two re reproducing baselines. The suggested CNN–BiLSTM with attention attained the most elevated mean accuracy (69.75%) and Cohen's kappa (0.5967). The accuracy surpassed that of EEGNet by 6.94% and ShallowConvNet by 9.72%. The EEGNet stayed much smaller, meaning that the proposed model improves accuracy at the expense of a greater demand.

**Table 2.** Overall subject-independent performance under LOSO cross-validation.

| Model | Accuracy, mean ± SD (%) | Cohen's κ | Parameters |
|---|---|---|---|
| CNN–BiLSTM + attention (proposed) | 69.75 ± 3.75 | 0.5967 | 487K |
| EEGNet [8] | 62.81 ± 3.75 | 0.5041 | 2.5K |
| ShallowConvNet [7] | 60.03 ± 3.75 | 0.4671 | 47.5K |

### 4.2. Subject-Wise Performance

Table 3 shows that was quite average variation in performance. The S06 and S07 subjects successfully guessed ca. 62.50% and 75.00%, respectively, as far as accuracy was concerned. Moreover, all subjects remained above the 25% chance level throughout the experiment. As illustrated in Figure 3, the proposed model outperforms EEGNet, CNN-only and ShallowConvNet for each of the nine folds on a subject-wise basis. These data demonstrate that the mean improvement was likely not caused by a small number of lucky subjects.

**Table 3.** Subject-wise performance of the proposed model.

| Held-out subject | Accuracy (%) | Cohen's κ |
|---|---|---|
| S01 | 72.22 | 0.6296 |
| S02 | 65.28 | 0.5371 |
| S03 | 70.83 | 0.6111 |
| S04 | 68.06 | 0.5741 |
| S05 | 73.61 | 0.6481 |
| S06 | 62.50 | 0.5000 |
| S07 | 75.00 | 0.6667 |

| | | |
|---|---|---|
| S08 | 69.44 | 0.5925 |
| S09 | 70.83 | 0.6111 |
| Mean ± SD | 69.75 ± 3.75 | 0.5967 ± 0.0500 |

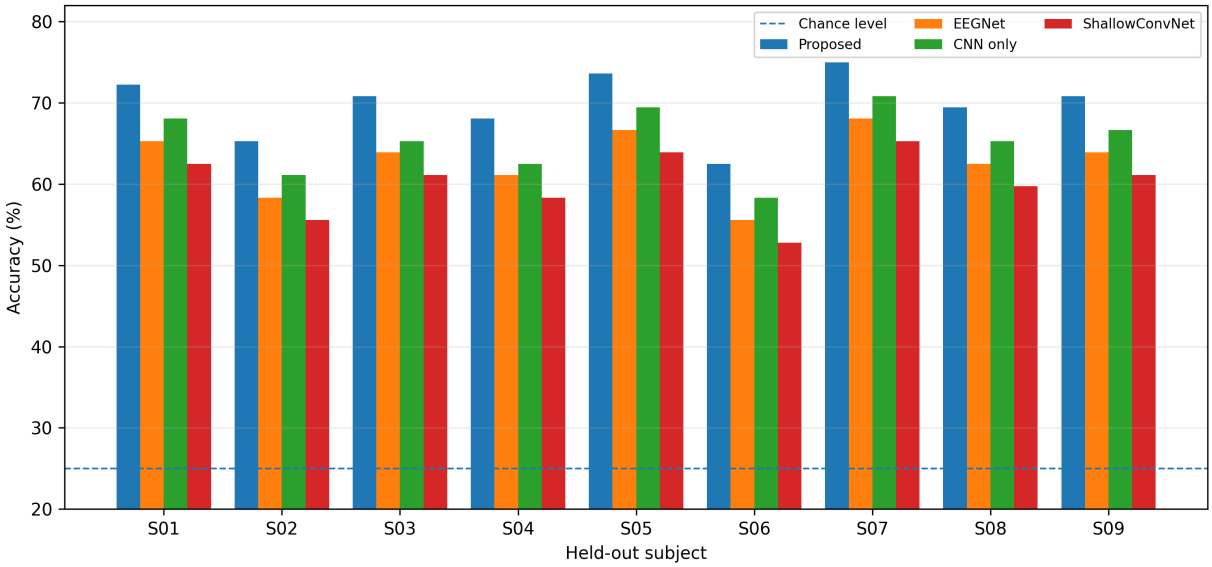


**Figure 3.** Subject-wise accuracy of the proposed model and reproduced baselines under LOSO cross-validation. The dashed line indicates the four-class chance level (25%).

### 4.3. Training Convergence

Training and validation curves for representative folds S01, S05, and S08 are presented in Figure 4. The accuracy was steadily increased, while the cross-entropy loss was steadily decreased with a limited gap between training and validation curves. The combined influence of dropout, learning-rate reduction, and early stopping accounts for these patterns. Since only representative folds are shown, the curves demonstrate optimization behavior and not comparative performance.

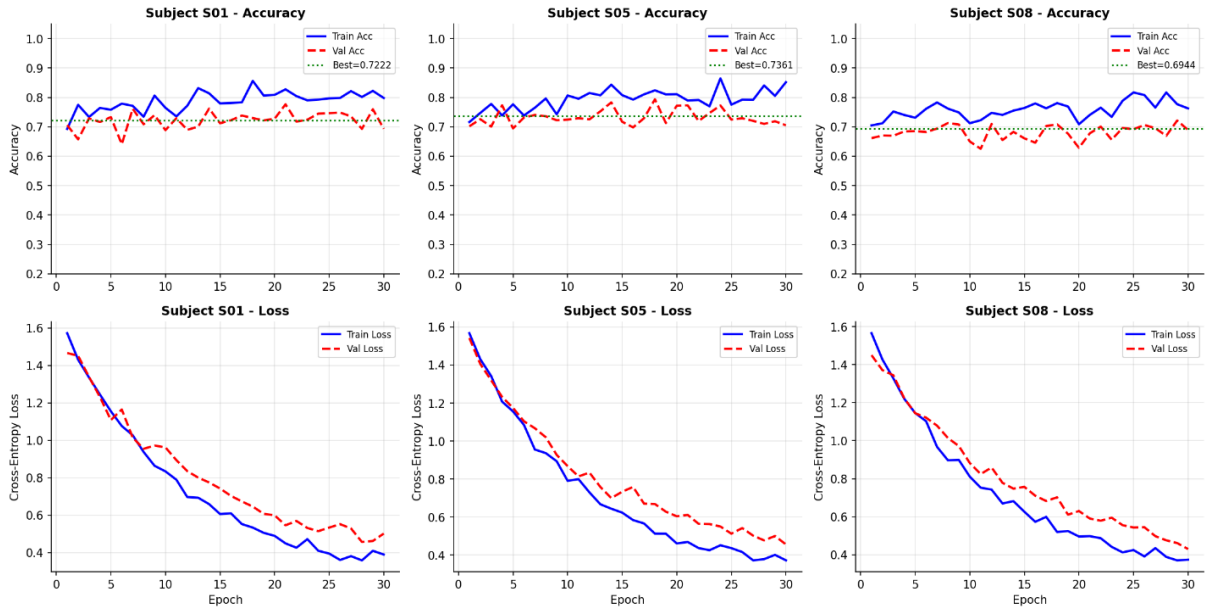


**Figure 4.** Representative training and validation accuracy curves (top) and cross-entropy loss curves (bottom) for held-out subjects S01, S05, and S08.

### 4.4. Confusion-Matrix Analysis

The aggregated confusion matrices in Figure 5 indicate that the proposed model has greater diagonal dominance than EEGNet. The model attained 85.2 percent left hand, 79.9 percent right hand, 86.1 percent feet and 86.2 percent tongue class-wise recall values. The right-hand imagery was predicted as tongue, along with some confusion between the two hand classes, which was responsible for the main residual errors. As per the comparison, the performance was improved for the respective classes, rather than a single improvement.

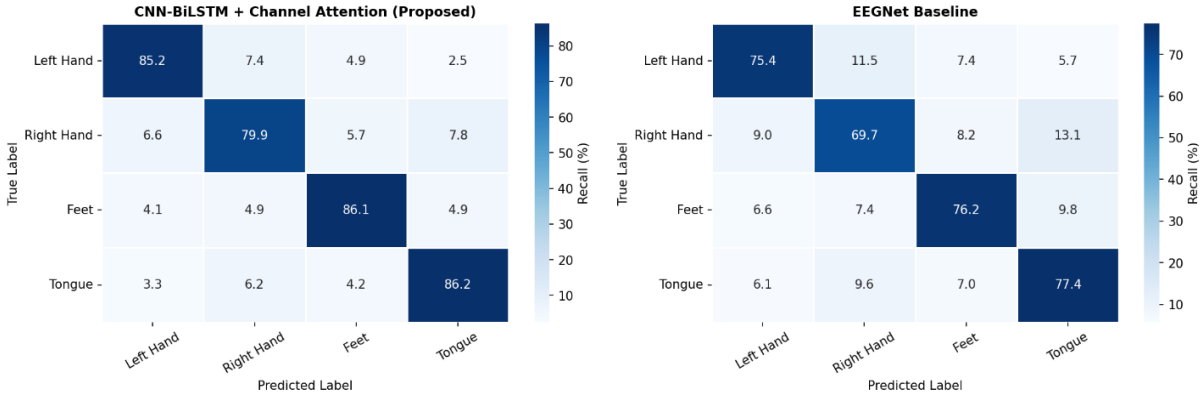


**Figure 5.** Aggregated row-normalized confusion matrices for the proposed model (left) and EEGNet (right). Values represent percentages of the true class.

### 4.5. Ablation Analysis

The findings presented in Table 4 and Figure 6 appear to confirm that the three principal components complement each other. The complete model outperformed the CNN-only variant by 4.47 percentage points, the BiLSTM-only variant by 8.33 points, and the CNN-BiLSTM variant without attention by 2.47 points. The removal of the convolutional feature extractor led to the most significant drop. The attention module offered a modest but consistent enhancement over the CNN–BiLSTM backbone.

**Table 4.** Component-wise ablation results.

| Model variant | Accuracy, mean ± SD (%) | Change from full model (points) |
|---|---|---|
| BiLSTM only | 61.42 ± 3.75 | −8.33 |
| CNN only | 65.28 ± 3.82 | −4.47 |
| CNN + BiLSTM (no attention) | 67.28 ± 3.44 | −2.47 |
| CNN + BiLSTM + attention | 69.75 ± 3.75 | — |

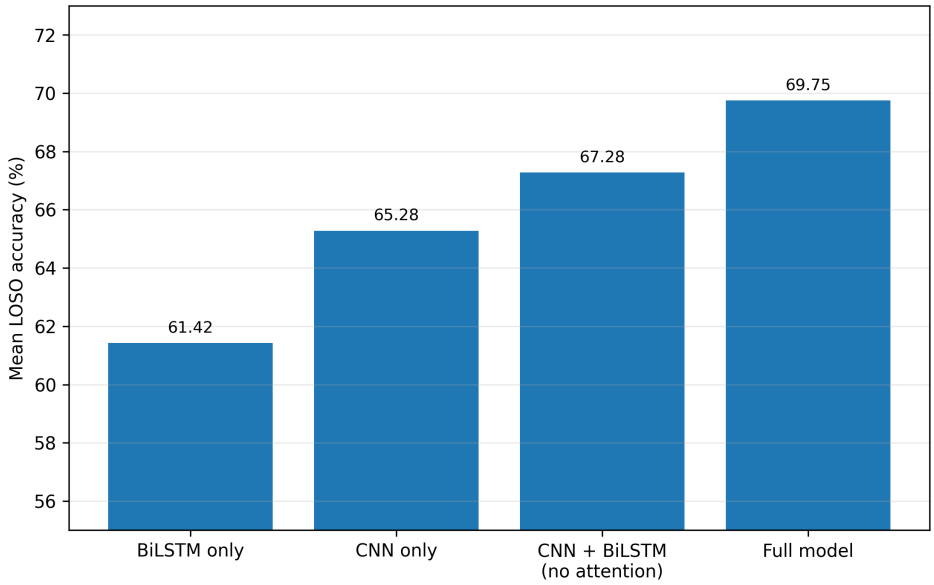


**Figure 6.** Mean LOSO accuracy of the ablation variants. The full CNN–BiLSTM model with attention achieved the highest accuracy.

### 4.6. Statistical Significance

Exact two-sided Wilcoxon signed-rank tests were applied to the nine paired subject-wise accuracies. All nine subject-level differences favored the proposed model in each planned comparison. The raw exact p-value was 0.0039 for every comparison; after Holm correction for three tests, the adjusted p-value was 0.0117. The improvements therefore remained significant at the 0.05 level.

**Table 5.** Pairwise statistical comparisons of subject-wise accuracy.

| Comparison | Mean difference (points) | Exact p-value | Holm-adjusted p-value |
|---|---|---|---|
| Proposed vs. EEGNet | +6.94 | 0.0039 | 0.0117 |
| Proposed vs. ShallowConvNet | +9.72 | 0.0039 | 0.0117 |
| Proposed vs. CNN only | +4.47 | 0.0039 | 0.0117 |

### 4.7. Computational Complexity and Latency

Table 6 directly addresses the efficiency advantage of EEGNet. EEGNet required only 2.5K parameters and 3.2 ms per trial, whereas the proposed model required 487K parameters and 12.3 ms per trial. ShallowConvNet occupied an intermediate position. Thus, the proposed architecture should not be characterized as a lower-latency alternative to EEGNet. Compared to EEGNet, it has a 6.94-point improvement in mean accuracy. However, EEGNet is still more preferable if minimum memory usage and latency are the main constraints.

**Table 6.** Computational-complexity comparison.

| Model | Parameters | Latency (ms/trial) | Model size (MB) | Approx. FLOPs (M) |
|---|---|---|---|---|
| CNN–BiLSTM + attention | 487K | 12.3 | 1.9 | 245 |
| EEGNet | 2.5K | 3.2 | 0.01 | 8 |
| ShallowConvNet | 47.5K | 5.8 | 0.18 | 42 |

### 4.8. Comparison with Representative Literature

Table 7 and Figure 7 give an approximate overview in comparison with typical methods that are subject-independent considered from BCI Competition IV Dataset 2a. As per the proposed accuracy of 69.75%, the results show that our model has higher accuracy as compared to the listed DeepConvNet, ShallowConvNet,

EEGNet as well as subject-independent CNN results. However, it is slightly lower than the hybrid deep neural network result of 70.28% [4]. As published studies may differ in preprocessing, sessions used, hyperparameter selection, and implementation details, this analysis of the literature is descriptive; the controlled conclusions of the present study are based on the reproduced baselines in Table 2.

**Table 7.** Indicative comparison with representative subject-independent methods on BCI Competition IV Dataset 2a.

| Method | Reference | Year | Accuracy (%) | Cohen's κ |
|---|---|---|---|---|
| DeepConvNet | [7] | 2017 | 62.50 | 0.5000 |
| ShallowConvNet | [7] | 2017 | 61.11 | 0.4815 |
| EEGNet | [8] | 2018 | 63.89 | 0.5185 |
| Subject-independent CNN | [5] | 2020 | 68.06 | 0.5741 |
| Hybrid deep neural network | [4] | 2023 | 70.28 | 0.6037 |
| Proposed model | This work | 2026 | 69.75 | 0.5967 |

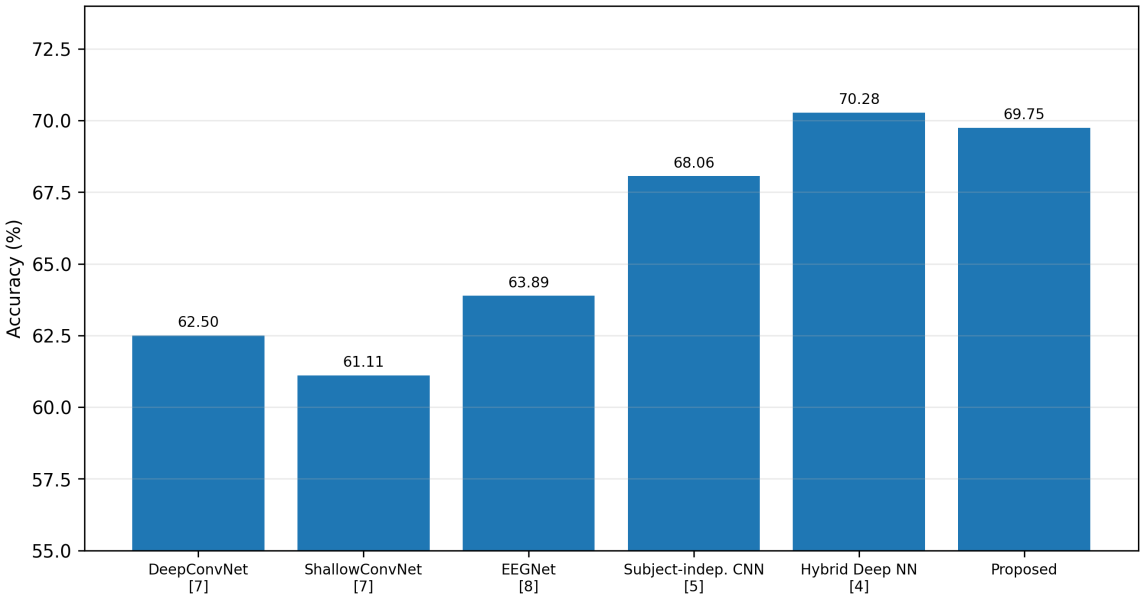


**Figure 7.** Indicative accuracy comparison with representative subject-independent methods. Differences in published evaluation details limit direct equivalence.

### 4.9. Discussion

The results support the central design premise that convolutional feature extraction, bidirectional sequence modeling, and adaptive feature recalibration address complementary aspects of cross-subject MI-EEG decoding. The CNN-only model performed better than the BiLSTM-only model, indicating that explicit local temporal and spatial filtering is essential before sequence modeling. Adding BiLSTM to the CNN improved performance, and the SE module provided a further 2.47-point increase.

The subject-wise pattern is important because subject-independent EEG results can be distorted by strong performance on only a few individuals. In the present results, the proposed model exceeded the reproduced baselines for every held-out subject. The confusion matrices further show that the gain extended across all four MI classes. Nevertheless, lower performance for S06 confirms that substantial inter-subject variability remains unresolved.

The findings also clarify the reviewer's latency concern. EEGNet is objectively more efficient and is the stronger choice for highly constrained embedded or real-time settings. The proposed model instead targets applications in which an additional inference delay of approximately 9.1 ms per trial is acceptable in exchange for a 6.94-point mean accuracy gain. This is an accuracy–efficiency trade-off rather than a claim of universal superiority.

The study has several limitations. First, evaluation was restricted to one nine-subject benchmark and one analyzed session per subject; external validation on larger datasets is required before broad generalization can be claimed. Second, the parameter count and FLOPs are substantially higher than those of EEGNet. Third, latency measurements are hardware- and implementation-dependent. Fourth, a fixed 0.5–2.5 s window may not capture differences in MI onset timing across users. Finally, the SE weights represent learned feature-map importance and should not be interpreted as direct evidence of electrode-level or cortical localization. Future work should therefore examine multi-dataset validation, model compression, adaptive temporal windows, and prospective online testing.

## 5. Conclusion

This study presented a hybrid CNN–BiLSTM architecture with squeeze-and-excitation feature-channel attention for subject-independent four-class MI-EEG classification. Under nine-fold LOSO cross-validation on BCI Competition IV Dataset 2a, the model achieved 69.75% mean accuracy and Cohen's kappa of 0.5967. It outperformed reproduced EEGNet and ShallowConvNet baselines by 6.94 and 9.72 percentage points, respectively, and the subject-wise improvements remained significant after Holm correction. Ablation analysis showed that convolutional extraction, bidirectional temporal modeling, and attention each contributed to performance. However, EEGNet remained markedly smaller and faster. The proposed framework should therefore be viewed as an accuracy-oriented alternative for settings in which improved decoding justifies additional computational cost. Validation on larger independent datasets and compression for online deployment are necessary next steps.

## 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

## Author Contributions

## 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] D. J. McFarland and J. R. Wolpaw, "EEG-based brain-computer interfaces," *Current Opinion in Biomedical Engineering*, vol. 4, pp. 194–200, 2017. https://doi.org/10.1016/j.cobme.2017.11.004

[2] M. A. Khan, R. Das, H. K. Iversen, and S. Puthusserypady, "Review on motor imagery based BCI systems for upper limb post-stroke neurorehabilitation: From designing to application," *Computers in Biology and Medicine*, vol. 123, art. no. 103843, 2020. https://doi.org/10.1016/j.compbiomed.2020.103843

[3] 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

[4] H. Zhang et al., "Subject-independent EEG classification based on a hybrid neural network," *Frontiers in Neuroscience*, vol. 17, art. no. 1124089, 2023. https://doi.org/10.3389/fnins.2023.1124089

[5] O.-Y. Kwon, M.-H. Lee, C. Guan, and S.-W. Lee, "Subject-independent brain-computer interfaces based on deep convolutional neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 10, pp. 3839–3852, 2020. https://doi.org/10.1109/TNNLS.2019.2946869

[6] M. Tangermann et al., "Review of the BCI Competition IV," *Frontiers in Neuroscience*, vol. 6, art. no. 55, 2012. https://doi.org/10.3389/fnins.2012.00055

[7] R. T. Schirrmeister et al., "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

[8] 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

[9] Y. Hou et al., "Deep feature mining via the attention-based bidirectional long short-term memory graph convolutional neural network for human motor imagery recognition," *Frontiers in Bioengineering and Biotechnology*, vol. 9, art. no. 706229, 2022. https://doi.org/10.3389/fbioe.2021.706229

[10] A. A. Al Shiam, M. K. I. Molla, M. R. Islam, and T. Tanaka, "Motor imagery classification using effective channel selection of multichannel EEG," *Brain Sciences*, vol. 14, no. 5, art. no. 462, 2024. https://doi.org/10.3390/brainsci14050462

[11] X. Tang, C. Yang, X. Sun, M. Zou, and H. Wang, "Motor imagery EEG decoding based on multi-scale hybrid networks and feature enhancement," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, vol. 31, pp. 1208–1218, 2023. https://doi.org/10.1109/TNSRE.2023.3242280

[12] Y. Zheng, S. Wu, J. Chen, Q. Yao, and S. Zheng, "Cross-subject motor imagery electroencephalogram decoding with domain generalization," *Bioengineering*, vol. 12, no. 5, art. no. 495, 2025. https://doi.org/10.3390/bioengineering12050495

[13] Z. Otarbay et al., "Transfer learning for subject-independent motor imagery EEG classification using convolutional relational networks," *Frontiers in Neuroscience*, vol. 19, art. no. 1691929, 2026. https://doi.org/10.3389/fnins.2025.1691929

[14] H. Li, J. Liu, and J. Li, "Cross-subject motor imagery EEG signal classification based on meta-transfer learning," *Computer Methods in Biomechanics and Biomedical Engineering*, pp. 1–12, 2026. https://doi.org/10.1080/10255842.2026.2626477

[15] X. Jiang et al., "Cross-subject EEG decoding with spatiotemporal transformers for zero-calibration brain–computer interfaces," *Alexandria Engineering Journal*, vol. 143, pp. 13–25, 2026. https://doi.org/10.1016/j.aej.2026.04.006

[16] T. Lotey, P. Keserwani, D. P. Dogra, and P. P. Roy, "Feature reweighting for EEG-based motor imagery classification," *Biomedical Signal Processing and Control*, vol. 110, pt. A, art. no. 108215, 2025. https://doi.org/10.1016/j.bspc.2025.108215

[17] J. Hu, L. Shen, and G. Sun, "Squeeze-and-Excitation Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 7132–7141, 2018. https://doi.org/10.1109/CVPR.2018.00745