

EEG-MFTNet: An Enhanced EEGNet Architecture with Multi-Scale Temporal Convolutions and Transformer Fusion for Cross-Session Motor Imagery Decoding

Panagiotis Andrikopoulos, Siamak Mehrkanoon*

Department of Information and Computing Sciences, Utrecht University, Utrecht, Netherlands
panosandrikos3@gmail.com, s.mehrkanoon@uu.nl

Abstract—Brain-computer interfaces (BCIs) enable direct communication between the brain and external devices, providing critical support for individuals with motor impairments. However, accurate motor imagery (MI) decoding from electroencephalography (EEG) remains challenging due to noise and cross-session variability. This study introduces EEG-MFTNet, a novel deep learning model based on the EEGNet architecture, enhanced with multi-scale temporal convolutions and a Transformer encoder stream. These components are designed to capture both short and long-range temporal dependencies in EEG signals. The model is evaluated on the SHU dataset using a subject-dependent cross-session setup, outperforming baseline models, including EEGNet and its recent derivatives. EEG-MFTNet achieves an average classification accuracy of 58.9% while maintaining low computational complexity and inference latency. The results highlight the model’s potential for real-time BCI applications and underscore the importance of architectural innovations in improving MI decoding. This work contributes to the development of more robust and adaptive BCI systems, with implications for assistive technologies and neurorehabilitation.

I. INTRODUCTION

Brain-computer interfaces (BCIs) enable direct communication between brain activity and external devices, bypassing neuromuscular control to assist individuals with motor impairments [1], [2]. Applications span rehabilitation and assistive robotics, enhancing autonomy through control of prosthetics, wheelchairs, and exoskeletons [3]. A typical BCI pipeline consists of signal acquisition, feature extraction and classification, followed by command execution [2]. Brain signals can be recorded using either invasive or non-invasive methods. Non-invasive techniques, such as functional magnetic resonance imaging (fMRI), magnetoencephalography (MEG), and electroencephalography (EEG), are generally safer and more practical for everyday use. While fMRI and MEG offer high spatial resolution, EEG is often preferred for real-time motor imagery (MI) BCIs due to its low cost, portability, and high temporal resolution [4], [5].

MI involves mentally simulating physical movements, which triggers neural activity that can be captured using EEG [2], [6]. Several EEG rhythms are associated with different cognitive

and motor processes, including delta (≤ 4 Hz), theta (4–8 Hz), alpha and mu (8–13 Hz), beta (13–30 Hz), and gamma (≥ 31 Hz). In MI tasks, specific changes in the mu and beta bands, referred to as event-related desynchronization (ERD) and synchronization (ERS), are commonly used as indicators of motor cortex activation [7], [8]. However, accurate MI-EEG classification is hindered by signal variability and noise due to emotional, physiological, and environmental factors [6]. Traditional approaches, such as Common Spatial Patterns (CSP) [9] and Filter Bank CSP (FBCSP) [10], extract handcrafted features and typically employ classifiers like linear discriminant analysis (LDA) or support vector machines (SVM). While effective in some settings, these methods require expert-driven parameter tuning and are highly sensitive to both inter-subject and intra-subject variability [11].

Recent deep learning approaches have improved MI classification by learning features directly from raw EEG, reducing dependence on manual preprocessing and increasing robustness [2], [12]. Still, challenges remain in managing cross-session and cross-subject variability without increasing computational complexity. This work addresses these limitations by enhancing an end-to-end model for robust MI-EEG decoding. Specifically, the contributions of this work are as follows:

- **EEG-MFTNet Architecture:** We propose a novel motor imagery EEG decoding model that extends EEGNet by incorporating (i) a multi-scale temporal convolution block to capture patterns across different temporal resolutions, and (ii) a Transformer encoder stream to model long-range temporal dependencies.
- **Cross-Session Robustness:** The model is evaluated under a subject-dependent cross-session protocol on the SHU [13] dataset, achieving an average accuracy of 58.9%, which represents a +5% absolute improvement over the baseline EEGNet, while maintaining low computational cost and real-time inference capability.
- **Component Analysis:** An ablation study highlights the complementary benefits of the convolutional and Transformer modules, with each contributing to the overall performance gain.

*Corresponding author.

Code available at: <https://github.com/pandrikopoul/EEG-MFTNet>

- **Interpretability:** We provide an in-depth interpretability analysis using Gradient $\times$ Input and electrode deletion tests, demonstrating that the model focuses on class-specific electrodes and identifies meaningful neurophysiological patterns, thereby enhancing transparency and trustworthiness.
- **Reproducibility:** To support future research, the full implementation of the proposed model and experimental setup is made publicly available, addressing the current lack of reproducible implementations for this dataset.

The remainder of this paper is organized as follows. Section II reviews relevant literature on BCI systems and motor imagery classification. Section III describes the proposed model architecture and its principal components. Section IV details the dataset and preprocessing procedures. Section V presents the experimental setup, results, performance comparisons, and the ablation study. Section VI discusses the interpretability of the proposed model. Finally, Section VII summarizes the main findings and outlines directions for future research.

II. RELATED WORK

MI classification using EEG signals has seen significant progress through deep learning, addressing challenges like inter-subject variability and session-to-session inconsistency. Deep learning models have enabled end-to-end learning directly from raw EEG data. Early models like Deep and Shallow ConvNets [14] demonstrated the feasibility of CNNs for EEG decoding. EEGNet [15] further advanced this by introducing a compact architecture using depthwise and separable convolutions. Its success inspired several extensions: EEG-TCNet [16] added Temporal Convolutional Networks (TCNs) for long-range modeling, while MI-EEGNet [17] and EEG-ITNet [18] introduced multi-scale and inception-style modules. Notably, EEG-ITNet also integrates TCNs with residual connections to enhance temporal feature extraction. To address temporal asymmetry in TCNs, EEG-CDILNet [19] uses circular dilated convolutions for balanced temporal modeling. EEG-GENet [20] incorporates graph embeddings to capture spatial relationships between electrodes. Lightweight models like EEG-SimpleConv [21] show that simpler architectures can still perform competitively, making them suitable for real-time applications. Other innovations include MFRC-Net [11], which balances performance and efficiency via multi-scale residual blocks, and MAFNet [22], which uses attention mechanisms to extract multi-domain features.

While the majority of prior research on MI classification has centered on EEG, MEG-based models also offer architectural insights relevant to EEG classification. AA-EEGNet [23], an attention-enhanced variant of EEGNet, improves cross-subject generalization by emphasizing salient spatiotemporal features. Similarly, DS-GTF [24] fuses Graph Attention Networks with a Transformer encoder to capture spatial and temporal dependencies. Additionally, recently proposed architectures, although developed for other tasks such as sleep staging [25] and sound localization [26], offer valuable insights for MI classification, as they share similar architectural and signal processing challenges.

Although the aforementioned techniques have demonstrated encouraging classification results, considerable potential for advancement remains. Many recent models still rely on specific preprocessing pipelines or engineered input representations that may limit their adaptability. For instance, 3D-CNN-GAN [27] requires the transformation of raw EEG signals into temporal-frequency-phase (TFPF) maps, a step that introduces additional complexity and may not generalize well across datasets. Others, like MI-EEGNet [17], introduce complex multi-branch architectures that, while effective, increase computational demands and may hinder deployment on resource-constrained devices. Distinctively, our model operates without any preprocessing steps, relying entirely on its architectural design to enhance classification performance. Moreover, the proposed approach prioritizes low latency and computational efficiency, thereby increasing its feasibility for deployment in practical real-time BCI systems.

III. PROPOSED MODEL

In this work, we propose **EEG-MFTNet** (EEG-based Multi-Scale Temporal Convolutions and Transformer Fusion Network), a lightweight architecture extending EEGNet [15] with two major enhancements: (1) a multi-scale temporal convolution block inspired by MI-EEGNet [17] and EEG-ITNet [18], and (2) a Transformer encoder branch, motivated by the dual-stream design in [24]. These components operate in parallel during early temporal processing and are fused to provide a richer feature representation, as shown in Figure 1.

Each EEG trial is represented as $X \in \mathbb{R}^{C \times T \times 1}$, where C is the number of electrodes and T the number of time samples (e.g., 32 and 1000 in the SHU dataset). The signal is processed through two parallel branches:

- **Multi-Scale Temporal Convolutions:** To capture patterns at varying temporal resolutions, we use six parallel convolutional branches with kernel sizes $k \in \{5, 9, 13, 29, 61, 125\}$. Each branch uses a $(1 \times k)$ kernel, outputs 8 feature maps, and incorporates batch normalization, ELU activation, and spatial dropout (rate = 0.5). The outputs of all six branches (48 channels total) are concatenated and scaled with learnable weights, enabling the model to adaptively emphasize the most informative temporal scales.
- **Transformer Encoder:** In parallel, the input is reshaped to $\mathbb{R}^{T \times C}$, treating each time step as a token and the electrodes as features. A single Transformer encoder block, comprising two attention heads, a feedforward layer (dim = 32), GELU activations, dropout (0.2), and a residual connection, captures long-range dependencies across time. The output is reshaped back to $\mathbb{R}^{C \times T \times 1}$ for fusion with the convolutional branch.

Fusion: The outputs of both branches are merged using trainable scalar weights and concatenated along the channel axis, yielding a fused representation of shape $32 \times 1000 \times 49$. Layer normalization is subsequently applied to enhance training stability.

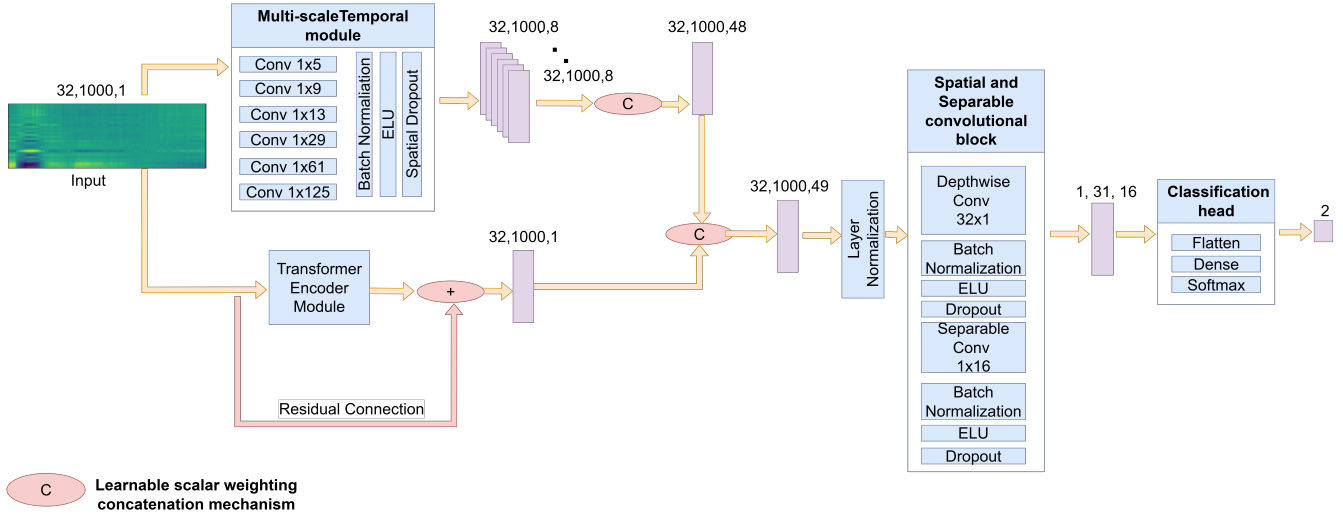


Fig. 1: EEG-MFTNet’s architecture. The numbers displayed above the input and each component’s output (purple boxes) denote the corresponding output dimensions.

Spatial Processing: Next, spatial relationships across electrodes are modeled using a depthwise convolution with a kernel size of $C \times 1$ and a depth multiplier of 2, yielding 96 output channels. This is followed by batch normalization, ELU activation, average pooling (1×4), and dropout (rate = 0.3). A subsequent separable convolution (1×16), consisting of a depthwise and a pointwise component, generates 16 feature maps. This is again followed by batch normalization, ELU activation, average pooling (1×8), and dropout (rate = 0.3).

Classification Head: Finally, the output is flattened and passed through a dense layer with a max-norm constraint of 0.25, generating class logits. A softmax layer then computes class probabilities, as shown in (1):

$$\hat{y} \in \mathbb{R}^N, \quad \sum_{i=1}^N \hat{y}_i = 1, \quad (1)$$

where $N = 2$ for binary MI classification. All hyperparameters, such as kernel sizes, number of filters, and dropout rates, were optimized through empirical search, with model selection and checkpointing performed based on validation-set performance.

IV. DATA DESCRIPTION

This study uses the SHU dataset [13], a large-scale, publicly available EEG dataset containing data from 25 healthy subjects, each completing five sessions across multiple days. In each session, participants performed 100 trials of left- or right-hand motor imagery, each lasting 4 seconds. EEG signals were recorded from 32 electrodes at 250 Hz, yielding 1000 time samples per trial, which are used directly as input to the model. The dataset was preprocessed by the original authors, who removed baseline drifts and applied a 0.5–40 Hz FIR band-pass filter to retain MI-relevant frequency components. The SHU dataset was selected for its emphasis on session-to-session variability, an ongoing challenge in EEG-based BCIs due to

non-stationary signals influenced by factors such as electrode shifts, fatigue, and mental state.

V. EXPERIMENTAL RESULTS AND DISCUSSION

A. Experimental Setup

The proposed model was evaluated on the SHU dataset using the subject-dependent cross-session protocol proposed by the dataset authors [13]. Each subject’s first session was used for training (with 20% held out for validation), while sessions 2–5 were used for testing. A separate model was trained per subject. Experiments were conducted on an NVIDIA GTX 1650 GPU.

To enhance model performance, several training strategies were employed. A model checkpoint callback was used to save the weights corresponding to the highest validation accuracy during training. These best weights were automatically restored upon completion of training. The proposed model, along with all baseline models, was trained for 100 epochs with a batch size of 16, using the categorical cross-entropy loss function and the AdamW optimizer (learning rate of 0.001, weight decay of $1e-4$). A learning rate scheduler was also utilized, reducing the learning rate by a factor of 0.5 if the validation loss did not improve for 5 consecutive epochs, down to a minimum of $1e-4$. This strategy was adopted to prevent overfitting and to accelerate convergence by dynamically adjusting the learning rate throughout training. The performance of the proposed model was compared against EEGNet, which served as the strongest baseline in the original SHU dataset paper [13] and formed the architectural foundation for the current work. Additional comparisons were made with two recent EEGNet derivatives: AA-EEGNet and EEG-GeNet.

To ensure a fair comparison, all baselines were trained using the same training settings as our proposed model. Additionally, all baselines used EEGNet-specific hyperparameters consistent with those reported in the SHU paper [13], except for AA-EEGNet, where dropout was set to 0.25. For non-EEGNet-

specific parameters, we followed the original configurations provided by each model’s authors, with the exception of EEG-GeNet, where the polynomial degree was set to $K = 3$ after empirical tuning. To ensure reproducibility across all experiments, a fixed random seed of 42 was applied throughout. However, it is important to note that comparisons with other SHU-based studies were excluded due to missing source code, unclear training setups, or protocol inconsistencies with the official SHU benchmark. Thus, we restrict evaluation to models with fully available implementations to ensure fair comparison.

B. Results and Comparison with State-of-the-Art

Table I summarizes the comparative classification accuracies obtained by various deep learning architectures evaluated on the SHU dataset. The models considered include EEGNet, the leading baseline reported in the original SHU dataset study [13], alongside two recently proposed variants, namely AA-EEGNet and EEG-GENet. These architectures, all based on the EEGNet family, provide relevant baselines for evaluating the performance of the proposed EEG-MFTNet model, which also builds upon and extends EEGNet’s architecture.

As indicated in Table I, the original EEGNet model attains a mean classification accuracy of 53.7%. Its derivatives, AA-EEGNet and EEG-GENet, exhibit marginal improvements, achieving 54.8% and 54.7% respectively. These modest gains reflect the incremental complexity and refined inductive biases introduced within these extended models. In comparison, EEG-MFTNet shows a noticeable improvement, achieving an average accuracy of **58.9%**. This represents an absolute increase of over 5% compared to the baseline EEGNet, and it also outperforms AA-EEGNet and EEG-GENet by 4.1% and 4.2%, respectively.

This notable improvement underscores the efficacy of explicitly capturing both short- and long-range temporal dependencies inherent in EEG signals, an aspect often overlooked in more lightweight architectures. While the standard deviation across subjects increases slightly (10.5%), this may be due to inter-subject variability, where some subjects exhibit more stable cross-session EEG patterns or patterns that are better captured by our proposed model.

Regarding computational complexity and inference speed, as shown in Table I, EEGNet remains the most computationally efficient model, with only 3,274 trainable parameters and the lowest average inference latency of 47.86 ms per trial. EEG-GENet introduces a modest increase in complexity, utilizing

7,146 parameters while maintaining a similar latency of 48.06 ms. In contrast, AA-EEGNet significantly increases model size to 517,498 parameters, over 150 times more than EEGNet, yet results in only a slight latency rise to 48.83 ms.

The proposed EEG-MFTNet balances model expressiveness with computational efficiency. It contains 16,096 parameters, substantially fewer than AA-EEGNet, and demonstrates an average inference latency of 49.63 ms per trial. Although this latency slightly exceeds that of the compared models, it remains well within acceptable bounds for near real-time operation, especially given the considerable improvement in classification accuracy observed in Table I.

In summary, the integration of multi-scale temporal convolutions with a Transformer-based attention mechanism significantly enhances the EEGNet architecture. This fusion yields EEG-MFTNet, setting a new performance benchmark among lightweight, EEGNet-inspired models for cross-session motor imagery classification. Although EEG-MFTNet introduces moderate computational overhead compared to baseline EEGNet variants, it achieves a favorable trade-off by delivering notably enhanced classification performance, thereby justifying the increased complexity for many practical BCI applications.

C. Analysis of the Results

The mean classification accuracies for Sessions 2, 3, 4, and 5 were 58.4%, 57.2%, 61.0%, and 58.8%, respectively, revealing moderate cross-session variability with the highest performance observed during Session 4.

Figure 2 reveals clear individual differences. Some participants (e.g., Subjects 6 and 20) consistently exceeded 90% accuracy, suggesting stable motor imagery patterns. Conversely, subjects such as 1, 18, 19, and 25 remained below 55%, likely due to signal variability or inconsistent imagery. Other subjects (e.g., Subject 13) showed large performance swings across sessions. These results highlight the significant impact of both individual differences and session variability on classification accuracy, underscoring the ongoing challenges in cross-session MI decoding and the need for adaptive approaches.

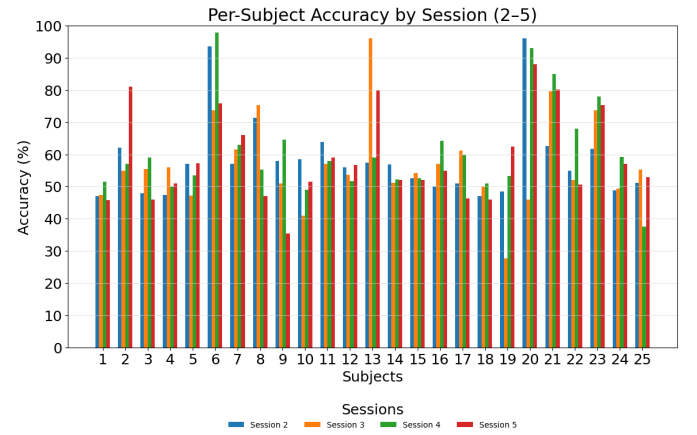


Fig. 2: Subject-wise classification accuracy of EEG-MFTNet across the four testing sessions. Each cluster of bars represents a single subject, with each bar indicating the model’s accuracy during a specific testing session.

TABLE I: Comparison of classification accuracy (averaged across all sessions and subjects), model complexity, and inference latency (per trial) for various EEGNet-based architectures.

Model	Accuracy (%)	Parameters	Latency (ms)
EEGNet	53.7 ± 6.3	3,274	47.86
EEG-GENet	54.7 ± 8.5	7,146	48.06
AA-EEGNet	54.8 ± 8.9	517,498	48.83
EEG-MFTNet	58.9 ± 10.5	16,096	49.63

The “ $\pm$ ” symbol indicates the standard deviation (in percentage points) across subjects to capture inter-subject variability.

D. Ablation Study

To elucidate the individual contributions of the Transformer module and the multi-scale temporal convolution module to the overall efficacy of the proposed model, an ablation study was conducted. This analysis involved systematically removing each component independently and evaluating the resulting impact on classification performance. The findings are summarized in Table II.

As indicated by the results, the exclusion of either the Transformer stream or the multi-scale temporal convolution module caused a marked reduction in classification accuracy. Specifically, the removal of the Transformer component decreased the average accuracy to 57.2%, whereas omitting the multi-scale temporal module led to a more pronounced decline, with accuracy dropping to 55.2%. In comparison, the complete model, which integrates both components, attained the highest average accuracy of **58.9%**. Despite the performance drops, both ablated variants still outperformed the baseline EEGNet model (53.7% accuracy). Specifically, adding only the multi-scale temporal module improved accuracy by 3.5%, while the Transformer encoder alone yielded a 1.5% gain. These results highlight the complementary contributions of both modules and their combined impact on performance. Even when used individually, each component provides a clear advantage, confirming the effectiveness of the architectural enhancements in EEG-MFTNet.

VI. MODEL INTERPRETATION

To gain insight into the model’s decision-making process, the Gradient $\times$ Input technique was applied to assess the contribution of each EEG channel. This approach has proven effective across diverse deep learning architectures, particularly in EEG-based BCI systems [28]. Subject 6, Session 4 was selected for this analysis, as it yielded the highest classification accuracy. Since checkpointing selects the weights with the highest validation accuracy and Subject 6’s training converges rapidly, the resulting predictions tend to remain close to 50% confidence. Therefore, the model was retrained for 50 epochs on Subject 6’s data using the final weights, yielding more distinct confidence scores and enabling a clearer trustworthiness analysis.

The interpretability method involved computing the element-wise product of the input and the gradient of the model’s output with respect to the input. This yielded attribution maps for each correctly classified trial, which were then averaged across time

TABLE II: Ablation study of EEG-MFTNet. Checkmarks (✓) indicate the inclusion of the corresponding component. “—” in both columns corresponds to the baseline EEGNet.

Transformer Encoder Stream	Multi-scale Conv Module	Accuracy (%)
—	—	53.7 [*]
—	✓	55.2
✓	—	57.2
✓	✓	58.9

^{*}EEGNet accuracy reproduced using hyperparameters from the SHU dataset source code [13].

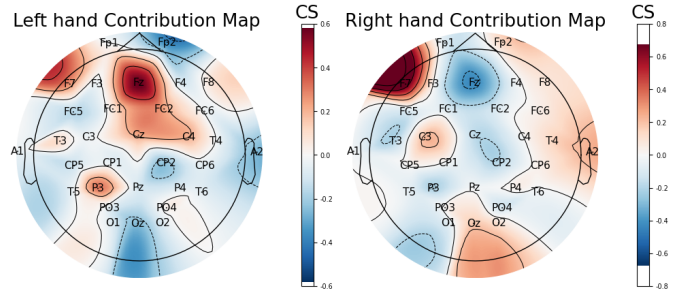


Fig. 3: Topographic contribution maps (Gradient $\times$ Input) averaged over all correctly classified trials belonging to a specific class (left or right motor representation) for Subject 6, Session 4. Warm colors indicate greater channel importance.

to obtain a single attribution score per channel. Channels were ranked by the average absolute attribution values, highlighting those most influential to the model’s decisions.

Scalp topographies were generated using MNE-Python to visualize spatial patterns of channel importance, aggregating attribution maps across trials per class. Figure 3 shows that left- and right-hand imagery produce distinct activation patterns. Notably, right-hand motor imagery is associated with decreased activation in central frontoparietal regions (Fz, F4, FC2, Cz, and FC1) and increased activity in posterior areas (PO3, O1, Oz, and O2), whereas left-hand imagery exhibits the opposite pattern. These observations suggest the model distinguishes classes based on both the presence and suppression of activity in specific regions.

To assess the reliability of the interpretation method, electrode deletion tests were conducted. Electrodes were ranked according to their attribution importance, and increasing proportions of either the most or the least important channels were set to zero. At each step, the model’s prediction confidence was recorded.

For left-hand imagery (Figure 4(a)), the confidence dropped from about 0.90 to just below 0.40 after deleting 60% of the most influential electrodes, while removing the least important channels had only a minor effect. For right-hand imagery (Figure 4(b)), deleting the most important electrodes reduced confidence from nearly 0.95 to around 0.75 at 20% deletion and to about 0.70 at 40%, with a slight recovery to approximately 0.80 at 60%. In contrast, removing the least important channels caused negligible changes.

Overall, the clear separation between the deletion curves obtained by removing the most and the least important electrodes supports the validity of the proposed interpretation method. However, the moderately high area under the curves suggests that, while the method captures the dominant discriminative patterns, it may not fully account for all fine-grained features contributing to the model’s decisions.

VII. CONCLUSION AND FUTURE WORK

This work introduced EEG-MFTNet, a deep learning model for motor imagery EEG classification. Extending EEGNet, it integrates a multi-scale temporal convolutional block and

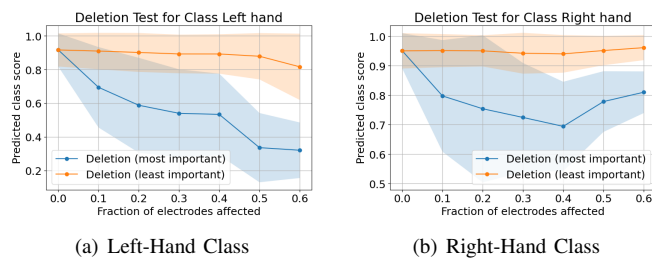


Fig. 4: EEG-MFTNet's average prediction confidence during Class-specific electrode deletion tests. Blue: most important electrodes deleted; Orange: least important deleted; shaded areas show standard deviation.

a parallel Transformer encoder to capture frequency-specific patterns and long-range dependencies, addressing cross-session variability. Evaluated on the SHU dataset under a subject-dependent cross-session protocol, EEG-MFTNet achieved 58.9% accuracy, outperforming EEGNet, AA-EEGNet, and EEG-GENet. Despite slightly higher complexity, it maintained low latency and few parameters. Ablation study confirmed the value of both modules, while interpretability analyses revealed attention to class-specific electrodes. Future work will explore cross-session adaptation, broader datasets, and online learning.

REFERENCES

- [1] L. F. Nicolas-Alonso and J. Gomez-Gil, "Brain computer interfaces, a review," *Sensors*, vol. 12, no. 2, pp. 1211–1279, 2012. [Online]. Available: <https://doi.org/10.3390/s120201211>
- [2] H. Altaheri, G. Muhammad, M. Alsulaiman, S. U. Amin, G. A. Altuwajri, W. Abdul, M. A. Bencherif, and M. Faisal, "Deep learning techniques for classification of electroencephalogram (eeg) motor imagery (mi) signals: A review," *Neural Computing and Applications*, vol. 35, no. 20, pp. 14 681–14 722, 2023.
- [3] J. Meng, Y. Wei, X. Mai, S. Li, X. Wang, R. Luo, M. Ji, and X. Zhu, "Paradigms and methods of noninvasive brain-computer interfaces in motor or communication assistance and rehabilitation: a systematic review," *Medical & Biological Engineering & Computing*, pp. 1–25, 2025.
- [4] R. A. Ramadan and A. V. Vasilakos, "Brain computer interface: control signals review," *Neurocomputing*, vol. 223, pp. 26–44, 2017.
- [5] Y. Sun, X. Chen, B. Liu, L. Liang, Y. Wang, S. Gao, and X. Gao, "Signal acquisition of brain-computer interfaces: A medical-engineering crossover perspective review," *Fundamental Research*, vol. 5, no. 1, pp. 3–16, 2025.
- [6] N. Padfield, J. Zabalza, H. Zhao, V. Masero, and J. Ren, "Eeg-based brain-computer interfaces using motor-imagery: Techniques and challenges," *Sensors (Basel, Switzerland)*, vol. 19, no. 6, p. 1423, 2019. [Online]. Available: <https://doi.org/10.3390/s19061423>
- [7] A. Saibene, M. Caglioni, S. Corchs, and F. Gasparini, "Eeg-based bcis on motor imagery paradigm using wearable technologies: A systematic review," *Sensors*, vol. 23, no. 5, p. 2798, 2023. [Online]. Available: <https://doi.org/10.3390/s23052798>
- [8] B. J. Edelman, S. Zhang, G. Schalk, P. Brunner, G. Müller-Putz, C. Guan, and B. He, "Non-invasive brain-computer interfaces: state of the art and trends," *IEEE reviews in biomedical engineering*, 2024.
- [9] H. Ramoser, J. Müller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial eeg during imagined hand movement," *IEEE transactions on rehabilitation engineering*, vol. 8, no. 4, pp. 441–446, 2000.
- [10] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter bank common spatial pattern (fbcs) in brain-computer interface," in *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*. IEEE, 2008, pp. 2390–2397.
- [11] X. Li, Z. Yang, X. Tu, J. Wang, and J. Huang, "Mfrc-net: Multi-scale feature residual convolutional neural network for motor imagery decoding," *IEEE Journal of Biomedical and Health Informatics*, 2024.
- [12] 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, p. 102738, 2024.
- [13] J. Ma, B. Yang, W. Qiu, Y. Li, S. Gao, and X. Xia, "A large eeg dataset for studying cross-session variability in motor imagery brain-computer interface," *Scientific Data*, vol. 9, no. 1, p. 531, 2022.
- [14] R. T. Schirrmester, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangemann, 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.
- [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, p. 056013, 2018.
- [16] 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 *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2020, pp. 2958–2965.
- [17] M. Riyad, M. Khalil, and A. Adib, "Mi-eegnet: A novel convolutional neural network for motor imagery classification," *Journal of Neuroscience Methods*, vol. 353, p. 109037, 2021.
- [18] A. Salami, J. Andreu-Perez, and H. Gillmeister, "Eeg-itnet: An explainable inception temporal convolutional network for motor imagery classification," *IEEE Access*, vol. 10, pp. 36 672–36 685, 2022.
- [19] T. Liang, X. Yu, X. Liu, H. Wang, X. Liu, and B. Dong, "Eeg-cdilnet: a lightweight and accurate cnn network using circular dilated convolution for motor imagery classification," *Journal of Neural Engineering*, vol. 20, no. 4, p. 046031, 2023.
- [20] H. Wang, H. Yu, and H. Wang, "Eeg_genet: A feature-level graph embedding method for motor imagery classification based on eeg signals," *Biocybernetics and Biomedical Engineering*, vol. 42, no. 3, pp. 1023–1040, 2022.
- [21] Y. El Ouahidi, V. Gripon, B. Pasdeloup, G. Bouallegue, N. Farrugia, and G. Lioi, "A strong and simple deep learning baseline for bci motor imagery decoding," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2024.
- [22] Y. Hong, X. Zeng, F. Wu, and J. Wang, "Mafnet: Multi-domain features attention-based fusion network for cross-subject motor imagery classification," in *2024 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2024, pp. 1–7.
- [23] I. A. Abdellaoui, J. G. Fernández, C. Sahinli, and S. Mehrkanon, "Enhancing brain decoding using attention augmented deep neural networks," in *Proceedings of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning*. ESANN, 2021, pp. 183–188.
- [24] L. Goené and S. Mehrkanon, "Dual stream graph transformer fusion networks for enhanced brain decoding," in *Proceedings of European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN)*. ESANN, 2024, pp. 375–380.
- [25] S. Kazatzidis and S. Mehrkanon, "A novel dual-stream time-frequency contrastive pretext tasks framework for sleep stage classification," in *Proceedings of International Joint Conference on Neural Networks (IJCNN)*. IEEE, 2024, pp. 1–8.
- [26] S. Kuang, J. Shi, K. van der Heijden, and S. Mehrkanon, "BAST-Mamba: Binaural audio spectrogram mamba transformer for binaural sound localization," *Neurocomputing*, p. 130804, 2025.
- [27] C. Fan, B. Yang, X. Li, S. Gao, and P. Zan, "Eeg-based feature classification combining 3d-convolutional neural networks with generative adversarial networks for motor imagery," *Journal of Integrative Neuroscience*, vol. 23, no. 8, p. 153, 2024.
- [28] J. Cui, L. Yuan, Z. Wang, R. Li, and T. Jiang, "Towards best practice of interpreting deep learning models for eeg-based brain computer interfaces," *Frontiers in Computational Neuroscience*, vol. 17, 2023. [Online]. Available: <https://doi.org/10.3389/fncom.2023.1232925>