FUSION OF SPATIO-TEMPORAL AND MULTI-SCALE FREQUENCY FEATURES FOR DRY ELECTRODES MI-EEG DECODING

Tianyi Gong[1†], *Can Han*[2†], *Junxi Wu*[3], *Dahong Qian*[2]

[1] School of Data Science, The Chinese University of Hong Kong, Shenzhen, China
[2] School of Biomedical Engineering, Shanghai Jiao Tong University, Shanghai, China
[3] OYMotion Technology Co., Ltd
122090863@link.cuhk.edu.cn, {hancan, dahong.qian}@sjtu.edu.cn, JCjunxi@gmail.com

ABSTRACT

Dry-electrode Motor Imagery (MI) Electroencephalography (EEG) enables fast, comfortable, real-world Brain Computer Interface by eliminating gels and shortening setup for at-home and wearable use. However, dry recordings pose three main issues: lower Signal-to-Noise Ratio with more baseline drift and sudden transients; weaker and noisier data with poor phase alignment across trials; and bigger variances between sessions. These drawbacks lead to larger data distribution shift, making features less stable for MI-EEG tasks. To address these problems, we introduce STGMFM, a tri-branch framework tailored for dry-electrode MI-EEG, which models complementary spatio-temporal dependencies via dual graph orders, and captures robust envelope dynamics with a multi-scale frequency mixing branch, motivated by the observation that amplitude envelopes are less sensitive to contact variability than instantaneous waveforms. Physiologically meaningful connectivity priors guide learning, and decision-level fusion consolidates a noise-tolerant consensus. On our collected dry-electrode MI-EEG, STGMFM consistently surpasses competitive CNN/Transformer/graph baselines. Codes are available at `https://github.com/Tianyi-325/STGMFM`

***Index Terms*—** EEG, motor imagery, dry electrodes, graph neural network, frequency mixing

1. INTRODUCTION

Non-invasive Electroencephalography (EEG) is the most accessible neural sensing modality for daily-life Brain Computer Interface (BCI) applications. Among non-invasive options, dry-electrode systems avoid conductive gel, shorten setup/cleanup, and better fit out-of-lab use [1, 2]. This practicality directly makes it suitable for rehabilitation at home, longitudinal self-training and wearable assistive interaction, where user comfort, compliance, and scalability are critical. However, dry recordings differ systematically from wet data: the unstable skin–electrode coupling often elevates and fluctuates impedance, lowering Signal-to-Noise (SNR) and amplifying motion/contact artifacts [1, 2]. Besides, repositioning across days leads to cross-session drift. What is more, inter-subject variability is exacerbated due to hair and pressure differences. In practice, dry-cap protocols also limit the number and duration of trials to reduce discomfort and time burden, which increases the premium on data-efficient and noise-robust decoding and makes frequent recalibration impractical. Currently, most existing Motor Imagery (MI) decoders were designed/validated under wet-electrode assumptions, and most of them transfer poorly to dry conditions due to the noise characteristics, distribution shift, and trial scarcity [3]. As a result, explicit algorithmic tailoring for dry-electrode MI remains under-explored.

Non-invasive BCIs with dry electrodes ease setup, but they typically suffer lower SNR and contact variability, which depresses MI accuracy and calls for noise-robust, data-efficient models [1, 2, 4]. Classical MI pipelines relied on Common Spatial Pattern/bandpower and linear classifiers or Riemannian geometry on covariances [3, 5–7], while deep CNNs improved end-to-end decoding yet still assume relatively clean, well-calibrated signals [8–13]. Graph formulations encode inter-channel relations and have shown benefits for spatio–temporal fusion and cross-session robustness [14]—most notably ST-GF by Wang *et al.* [15]; multi-branch/domain-generalization schemes further diversify representations via decision fusion [16]. In time-series modeling, recent "mixers" learn multi-period structure with compact, multi-resolution tokenization, including TimeMixer and its successor TimeMixer++ by Wang *et al.* [17, 18], as well as related sequence learners [19, 20]. Connectivity priors such as PLV provide phase-synchrony graphs that are amplitude-invariant and physiologically meaningful [21, 22].

Although these advances are compelling, most were devised and tuned with wet-electrode data rather than dry EEG. Therefore, they fail to remediate the low-SNR, contact-motion-artifact instability. They also emphasize localized patterns and lack multi-view learning, limiting the learning

†These authors contributed to the work equally.

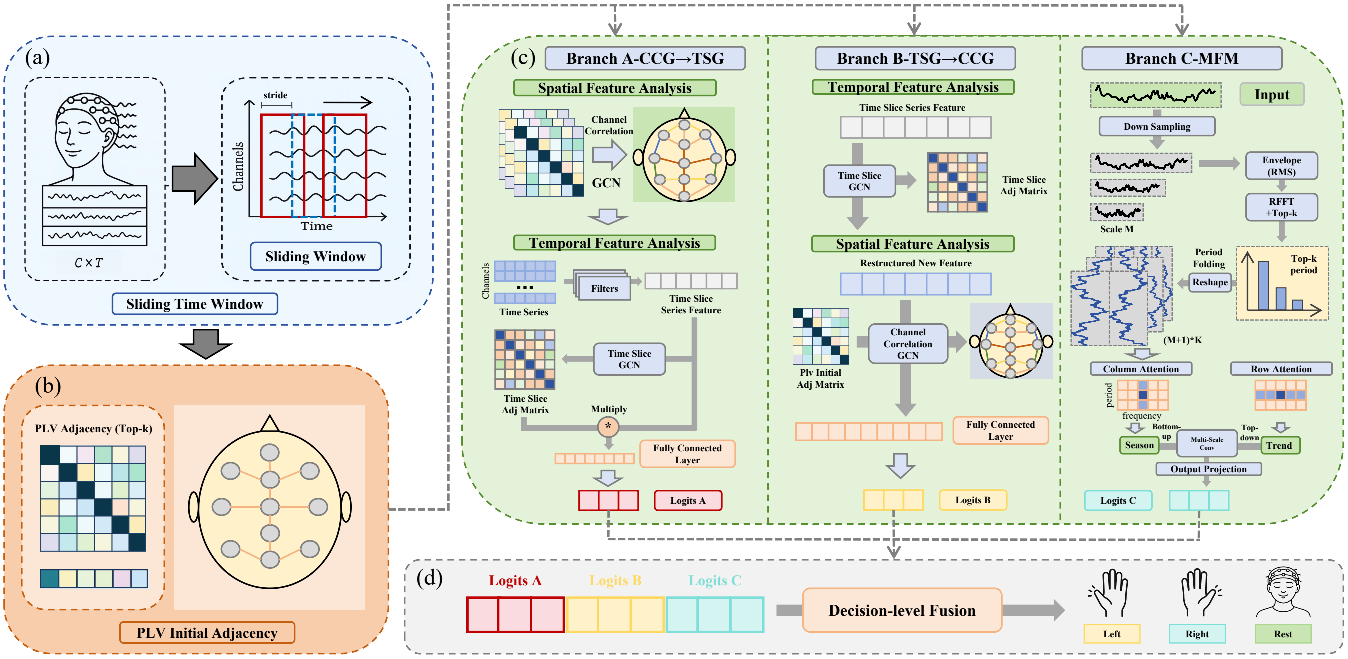

Fig. 1. Pipeline of the proposed model. Raw EEG is windowed into overlapping segments and processed by three branches. **A** (CCG→TSG) aligns functional connectivity via a PLV-initialized channel graph before aggregating on a time-slice graph. **B** (TSG→CCG) first reconstructs intra-trial continuity and then fuses channels on the CCG. **C** (MFM) operates on the amplitude envelope (RMS) with multi-resolution imaging and cross-scale mixing. Linear head fuses branch logits at the decision level.

of deep, invariant features to bridge cross-session/subject shifts. In addition, multi-scale spatio and temporal structure is under-explored, leaving deep cross-scale cues untapped. Motivated by these gaps, we propose **STGMFM**, which unifies dual ordered spatio-temporal graph branches (CCG→TSG and TSG→CCG) so that spatial-first and temporal-first propagation can hedge against different noise pathways. And we augments them with a multi-cale frequency mixer that learns multi-resolution envelope dynamics, aligned with ERD/ERS to offer a robust temporal cue when instantaneous waveform detail is unreliable. Physiological connectivity priors (PLV) provide a relatively detailed and precise starting point that is then adapted end-to-end to accommodate subject/session variability, while decision-level fusion forms a simple consensus that resists over-reliance on any single branch. In experiments on a dry-electrode MI-EEG dataset, STGMFM outperforms CNN/transformer/graph baselines, proving it to be a practical path toward dry-electrode BCIs.

Main contributions: (1) We introduce a PLV-initialized EEG graph that creates a more detailed and precise prior adjacency, suitable for low-SNR dry recordings. (2) We design a dual-order spatio-temporal graph to yield complementary evidence. (3) We introduce a lightweight Multi-Scale Frequency Mixer that extracts phase-invariant, multi-scale temporal cues aligned with ERD/ERS. (4) We employ a stable training recipe (decision-level fusion, L1/L2 regularization, cosine annealing) that reduces overfitting and over-reliance.

2. METHODS

2.1. Overview

We present a triple-branch framework for cross-subject MI EEG decoding. Raw trials are segmented with overlapping windows. Branches A/B pair a channel-correlation graph (CCG) and a time-slice graph (TSG) in opposite orders (CCG→TSG vs. TSG→CCG), yielding complementary biases—spatial alignment before temporal aggregation vs. temporal stabilization before spatial fusion. Branch C consumes amplitude envelopes to build a lightweight multi-resolution representation of rhythmic/trend dynamics. Each branch outputs logits, and a shallow head fuses them at the decision level, delivering robust generalization without fragile feature-level alignment. Figure 1 illustrates the overall pipeline.

2.2. Notation and Sliding Windowing

Let a trial be $\mathbf{X} \in \mathbf{R}^{C\times T}$ with C channels and T time steps. Using window length W_n and stride Str, the number of windows is

$$W = \left\lfloor \frac{T - W_n}{\mathrm{Str}} \right\rfloor + 1. \tag{1}$$

After segmentation, we obtain a 4-D tensor $\mathbf{X} \in \mathbf{R}^{N\times W\times C\times T_w}$ for batch size N and per-window length $T_w = W_n$. Windowing provides natural nodes for the TSG and a consistent statistical unit for the graph modules and the envelope branch, avoiding drift on long unsegmented sequences.

2.3. PLV-driven Initial Channel Graph

To encode functional connectivity as a prior, we build an initial adjacency from the phase-locking value (PLV). For channels i and j with analytic phases $\phi_i(t)$ and $\phi_j(t)$,

$$\mathrm{PLV}(i,j) = \left| \frac{1}{T} \sum_{t=1}^{T} e^{\mathrm{i}(\phi_i(t) - \phi_j(t))} \right|. \tag{2}$$

We remove self-loops, sparsify by per-row Top-k (or a threshold), symmetrize, and degree-normalize to obtain $\tilde{\mathbf{A}} = \mathbf{D}^{-1/2}\mathbf{A}\mathbf{D}^{-1/2}$. The same $\tilde{\mathbf{A}}$ initializes the CCG in both branches. During training, a small learnable increment adapts $\tilde{\mathbf{A}}$ to subject-specific variability while retaining interpretability. At the same time, it also provides more detailed and precise adjacency. Fig. 2 shows the before/after learning comparison of adjacency under PLV-initial and the basic spatial-neighborhood prior.

2.4. Branch A: CCG→TSG

Branch A first performs graph propagation on the channel graph to align truly cooperating electrodes and suppress mismatched ones. With node features $\mathbf{H}^{(l)} \in \mathbf{R}^{C\times d}$, one CCG layer reads

$$\mathbf{H}^{(l+1)} = \sigma\Big(\tilde{\mathbf{A}}\,\mathbf{H}^{(l)}\,\mathbf{W}_s^{(l)}\Big), \quad l = 0, \dots, K_s - 1. \tag{3}$$

We visualized the changes in spatial connectivity before and after learning, see Fig. 2. The resulting spatially denoised representation then passes through a shared temporal convolutional block (depthwise–pointwise 1D convolution with GELU/normalization) and is aggregated over the *time-slice graph*. Let $\mathbf{U}^{(l)} \in \mathbf{R}^{W\times d}$ be slice-level features and $\tilde{\mathbf{S}} \in \mathbf{R}^{W\times W}$ the learnable slice adjacency; one TSG layer is

$$\mathbf{U}^{(l+1)} = \sigma\Big(\tilde{\mathbf{S}}\,\mathbf{U}^{(l)}\,\mathbf{W}_t^{(l)}\Big), \quad l = 0, \dots, K_t - 1. \tag{4}$$

Placing temporal aggregation on top of a connectivity-aligned representation improves the SNR of time reconstruction by reducing sensitivity to phase misalignment across channels. A global average pooling (GAP) and a linear head yield $\mathbf{z}_A$. We insert a lightweight shared temporal block between CCG and TSG to align feature spaces and add local temporal expressiveness without disturbing the graph inductive bias.

2.5. Branch B: TSG→CCG

Branch B starts by reconstructing intra-trial continuity on the slice graph. Continuous MI segments become tightly clustered while non-task fragments are attenuated. The graph-filtered features are then rearranged to the channel dimension

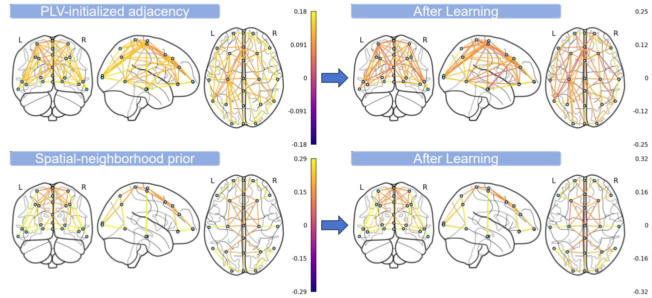

Fig. 2. Before/after training comparison of adjacency under two priors: PLV initialization and the spatial-neighborhood prior

and passed through the CCG propagation with $\tilde{\mathbf{A}}$. By stabilizing time first, B prevents short-term temporal noise from being prematurely diffused spatially. A shared temporal block and classifier head produce $\mathbf{z}_B$. Shared temporal block is also placed between TSG and CCG.

2.6. Branch C: Multi-Scale Frequency Mixer

The third branch operates on the amplitude envelope (RMS) (e.g., μ/β band), which makes ERD/ERS modulation explicit and phase-invariant. We first form multi-resolution "time images" by selecting dominant periods/frequencies (rfft) and mapping the 1D signal to a compact time–frequency lattice. We then decouple periodic (seasonal) patterns from slower trend evolution using dual-axis processing, and finally carry out multi-scale and multi-resolution mixing to aggregate evidence. This imaging—decoupling—mixing pipeline turns long dependencies into local ones on the image plane and yields a compact cross-scale representation. A lightweight configuration (small channel width and a single block, with one down-sampling and Top-1 dominant period) suits the low-sample EEG regime. The branch outputs $\mathbf{z}_C$.

2.7. Decision-level Fusion

Let $[\cdot;\cdot]$ denote concatenation. We fuse logits with a shallow linear head,

$$\mathbf{z} = \mathbf{W}\,[\mathbf{z}_A;\mathbf{z}_B;\mathbf{z}_C] + \mathbf{b}, \qquad \hat{y} = \arg\max_k \mathbf{z}_A. \tag{5}$$

We avoid gating/temperature mechanisms: in small EEG datasets they introduce extra degrees of freedom that are hard to calibrate across subjects and can overfit. A fixed linear combiner preserves class-wise alignment ability while remaining stable.

2.8. Objective and Optimization

We train with cross-entropy and combine ℓ_1 sparsity on graph-increment parameters with ℓ_2 weight decay (AdamW).

Table 1. Avg±Std results on dry-EEG under three evaluation protocols. We report Accuracy (ACC), Cohen's kappa, and F1.

Method	Cross Session			Cross Subject			Cross-Subject + Single-Session Fine-tuning		
	ACC(%)	kappa	F1(%)	ACC(%)	kappa	F1(%)	ACC(%)	kappa	F1(%)
ShallowNet [9]	47.30 ± 7.68	0.2095 ± 0.1154	45.39 ± 7.85	50.38 ± 10.92	0.2621 ± 0.1643	48.55 ± 12.13	52.06 ± 11.30	0.2809 ± 0.1695	50.18 ± 12.03
EEGNet [8]	47.67 ± 11.78	0.2152 ± 0.1767	45.96 ± 12.41	52.21 ± 10.81	0.2936 ± 0.1625	50.17 ± 11.85	54.94 ± 10.23	0.3241 ± 0.1535	52.71 ± 11.62
EEG-TCNet [11]	44.52 ± 11.86	0.1678 ± 0.1779	40.53 ± 13.40	46.25 ± 10.71	0.1936 ± 0.1415	41.65 ± 11.26	52.66 ± 9.12	0.2899 ± 0.1368	48.15 ± 11.36
EEGConformer [12]	48.55 ± 8.62	0.2282 ± 0.1295	46.07 ± 9.62	50.23 ± 9.61	0.2534 ± 0.1441	48.01 ± 10.81	55.83 ± 11.64	0.3374 ± 0.1746	53.82 ± 12.81
BaseNet [23]	46.07 ± 11.16	0.1911 ± 0.1674	41.80 ± 13.52	52.05 ± 10.21	0.2960 ± 0.1535	48.59 ± 12.55	53.97 ± 10.22	0.3095 ± 0.1534	50.14 ± 12.05
LMDANet [13]	46.13 ± 10.70	0.1920 ± 0.1604	43.58 ± 11.58	47.89 ± 8.87	0.2184 ± 0.1333	45.78 ± 10.46	51.33 ± 11.26	0.2701 ± 0.1689	49.03 ± 12.17
STGENet [15]	47.08 ± 10.58	0.2034 ± 0.1602	43.01 ± 12.60	47.66 ± 8.70	0.2155 ± 0.1253	44.96 ± 9.63	50.83 ± 10.28	0.2604 ± 0.1552	48.75 ± 11.21
STGMFM	**49.25 ± 4.16**	**0.2368 ± 0.0643**	**47.50 ± 5.16**	**57.26 ± 8.34**	**0.3592 ± 0.1247**	**56.52 ± 8.08**	**59.81 ± 6.71**	**0.3972 ± 0.1006**	**59.22 ± 6.72**

The total loss is

$$\mathcal{L} = \mathcal{L}_{\mathrm{CE}} + \lambda_s \|\Delta\tilde{\mathbf{A}}\|_1 + \lambda_t \|\Delta\tilde{\mathbf{S}}\|_1 + \beta \sum_{\theta} \|\theta\|_2^2. \qquad (6)$$

Cosine-annealed learning rates promote smooth convergence and better generalization.

3. EXPERIMENTS AND RESULTS

3.1. Dataset and Protocols

We collected a dry-electrode MI-EEG dataset with a 23-channel cap at 250 Hz. 19 subjects each completed two sessions on different days; within each session, three repeated runs were recorded for a three-class MI task. We evaluate three realistic protocols: (i) **Cross-Session**: train on one session and test on the other session from the same subject; (ii) **Cross-Subject**: leave one subject out for testing, train on the remaining subjects; (iii) **Cross-Subject + Single-Session Fine-tuning**: pre-train on other subjects and adapt using one session of the target subject. Trials are segmented using windows of 125 samples (0.5 s) with a stride of 125, yielding nine time slices per trial. We train 1000 epochs with AdamW (initial learning rate 2×10^{-3}), batch size 32, dropout 0.2, cosine-annealing schedule, and L1/L2 regularization on graph weights and classifier. All experiments run on a NVIDIA RTX 4090.

3.2. Baselines and Overall Performance

We compare against some representative models (ShallowNet [9], EEGNet [8], EEG-TCNet [11], EEGConformer [12], BaseNet [23], LMDANet [13], STGENet [15].) We report 3 averaged metrics over 19 subjects: Accuracy, Cohen's kappa and F1. Table 1 summarize the results. STGMFM attains the best accuracy across all three protocols.

3.3. Ablation and Component Analysis (on Cross-Subject)

We quantify the contribution of each component using the Cross-Subject protocol. Table 2 reports the results.

Table 2. Ablation study on dry-EEG under Cross Subject. We report Avg±Std results for Accuracy (ACC), Cohen's kappa, and F1. Best results are **bold**.

Variant	ACC(%)	kappa	F1(%)
Only A&B	54.19 ± 8.87	0.3132 ± 0.1327	53.53 ± 8.78
Only C	51.14 ± 9.16	0.2669 ± 0.1376	50.11 ± 9.45
No PLV initialization	56.76 ± 8.71	0.3517 ± 0.1278	56.11 ± 9.03
Spatial adjacency fixed	52.20 ± 9.48	0.2834 ± 0.1420	51.64 ± 9.50
No L1/L2 regularization	54.32 ± 9.46	0.3153 ± 0.1416	53.56 ± 9.38
Gated fusion	52.89 ± 9.43	0.2936 ± 0.1412	52.22 ± 9.35
STGMFM (full)	**57.26 ± 8.34**	**0.3592 ± 0.1247**	**56.52 ± 8.08**

Modules Analysis: *(i) Dual graph orders (A: CCG→ TSG; B: TSG→CCG)*: The two orders inject complementary inductive biases that hedge against distinct error pathways in dry-EEG: A first stabilizes spatial connectivity then consolidates temporal relations while B emphasizes slice-wise temporal consistency. *(ii) Multi-Scale Frequency-Mixer (C)*: MFM focuses on multi-scale envelopes and long-range modulations with reduced phase sensitivity, recovering ERD/ERS-like patterns that are blurred by dry-electrode noise. *(iii) PLV-initialized, learnable graphs*: PLV provides physiologically plausible priors that narrow the search space and improve early-epoch stability, yet learnability preserves subject/session adaptability. At the same time, it also provides more detailed and precise adjacency. *(iv) Regularization and simple fusion*: L1 prevents spurious edges/feature dominance, while L2 and cosine annealing smooth optimization.

4. CONCLUSION

We introduced STGMFM, a tri-branch fusion network for dry-electrode MI-EEG. Two complementary graph orders (CCG→TSG and TSG→CCG) and a Multi-Scale Frequency-Mixer jointly model spatio-temporal dependencies and envelope regularities. With PLV-initialized learnable graphs, cosine-annealed training, and L1/L2 regularization, STGMFM surpasses strong CNN/Transformer/GCN baselines across cross-session, cross-subject, and pre-train & adapt protocols on a 23-channel dry-EEG dataset. Ablations confirm the role of each module under low-SNR, contact variability: dual graph orders curb noise propagation, MFM captures robust ERD/ERS-like envelopes, and simple decision-level fusion generalizes better than gated schemes. Future work will explore adaptive PLV estimation, subject-aware fusion, and channel pruning for on-device inference.

5. REFERENCES

[1] Miguel A. Lopez-Gordo, Daniel Sanchez-Morillo, and Francisco Valle Pineda, "Dry eeg electrodes," *Sensors*, vol. 14, no. 7, pp. 12847–12870, 2014.

[2] Tony S. Grummett, Roy E. Leibbrandt, Timothy W. Lewis, Daniel DeLosAngeles, David M. Powers, John O. Willoughby, and Kenneth J. Pope, "Validity of artifact-limited eeg data from dry electrodes," *Clinical Neurophysiology*, vol. 126, no. 3, pp. 556–563, 2015.

[3] Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, et al., "A review of classification algorithms for eeg-based brain–computer interfaces: a 10-year update," *Journal of Neural Engineering*, vol. 15, no. 3, pp. 031005, 2018.

[4] Keun-Tae Kim, Jaehyung Lee, and Song Joo Lee, "Convolutional neural network approach for mi and ssep-based hybrid bci using dry electrodes," *Biomedical Signal Processing and Control*, vol. 110, pp. 108343, 2025.

[5] Herbert Ramoser, Johannes Müller-Gerking, and Gert 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.

[6] Kai Keng Ang, Zhuoyi Chin, Haihong Zhang, and Cuntai Guan, "Filter bank common spatial pattern (FBCSP) in brain–computer interface," in *IEEE IJCNN*, 2008, pp. 2390–2397.

[7] Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten, "Classification of covariance matrices using a riemannian-based kernel for bci applications," in *Int'l Conf. on Neural Engineering*, 2013, pp. 204–207.

[8] Vernon J. Lawhern, Amelia J. Solon, Nicholas R. Waytowich, Stephen M. Gordon, Chou P. Hung, and Brent J. Lance, "Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, pp. 056013, 2018.

[9] Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominik Fiederer, et al., "Deep learning with convolutional neural networks for eeg decoding and visualization," in *Human Brain Mapping*, 2017.

[10] Can Han, Chen Liu, Yaqi Wang, Crystal Cai, Jun Wang, and Dahong Qian, "A spatial-spectral and temporal dual prototype network for motor imagery brain-computer interface," *arXiv preprint arXiv:2407.03177*, 2024.

[11] Trausti Magnus Ingolfsson, Michael Hersche, Luca Benini, and Abbas Rahimi, "Eeg-tcnet: An accurate temporal convolutional network for embedded motor-imagery brain-machine interfaces," *arXiv preprint arXiv:2006.00622*, 2020.

[12] Y. Song, Q. Zheng, B. Liu, et al., "Eeg conformer: Convolutional transformer for eeg decoding and visualization," *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 2022.

[13] Zhaojie Miao, Ziyi Wang, Peiying Xiong, et al., "Lmda-net: A lightweight multi-dimensional attention network for eeg decoding," *NeuroImage*, 2023.

[14] T. Zhang, X. Chen, and Y. Wang, "Spatio-temporal graph convolutional networks for eeg decoding," in *ICASSP*, 2022.

[15] Xuhui Wang, Sigang Yu, Kui Zhao, Geng Chen, Enze Shi, and Shu Zhang, "Graph-based fusion of spatial and temporal features for eeg motor imagery decoding," in *IEEE BIBM*, 2024.

[16] Y. Du et al., "Multi-view ensemble learning for eeg motor imagery classification," in *ICASSP*, 2023.

[17] Shiyu Wang, Haixu Wu, Xiaoming Shi, Tengge Hu, Huakun Luo, Lintao Ma, James Y. Zhang, and Jun Zhou, "TimeMixer: Decomposable multiscale mixing for time series forecasting," in *Int'l Conf. on Learning Representations (ICLR)*, 2024.

[18] Yue Wang, Zhenyi He, Chang Zhou, Mohammed A. AlZain, Ferrari Dacrema, Jiao Zhang, Zuozhu Liu, Hao Chen, Qiang Zhang, Jian Tang, Li Erran Li, and Hao Wang, "Timemixer++: A general time series pattern machine for universal predictive analysis," in *Int'l Conf. on Learning Representations (ICLR)*, 2025.

[19] Y. Nie et al., "Patchtst: A patch-mixing transformer for time series," in *ICLR*, 2023.

[20] Haixu Wu et al., "Timesnet: Temporal 2d-variation modeling for general time series analysis," in *ICLR*, 2023.

[21] Jean-Philippe Lachaux, Eugenio Rodriguez, Jacques Martinerie, and Francisco J. Varela, "Measuring phase synchrony in brain signals," *Human Brain Mapping*, vol. 8, no. 4, pp. 194–208, 1999.

[22] H. Kang, C. Dong, Y. Ye, et al., "Plv-guided graph convolutional network with lstm for mi-eeg classification," *Biomedical Signal Processing and Control*, 2026.

[23] Martin Wimpff, Leonardo Gizzi, Jan Zerfowski, and Bin Yang, "Eeg motor imagery decoding: A framework for comparative analysis with channel attention mechanisms," *arXiv preprint arXiv:2310.11198*, 2023.