

DRDCAE-STGNN: An End-to-End Discriminative Autoencoder with Spatio-Temporal Graph Learning for Motor Imagery Classification

Yi Wang, Haodong Zhang and Hongqi Li*, Member, IEEE

Abstract— Motor imagery (MI) based brain-computer interfaces (BCIs) hold significant potential for assistive technologies and neurorehabilitation. However, the precise and efficient decoding of MI remains challenging due to their non-stationary nature and low signal-to-noise ratio. This paper introduces a novel end-to-end deep learning framework of Discriminative Residual Dense Convolutional Autoencoder with Spatio-Temporal Graph Neural Network (DRDCAE-STGNN) to enhance the MI feature learning and classification. Specifically, the DRDCAE module leverages residual-dense connections to learn discriminative latent representations through joint reconstruction and classification, while the STGNN module captures dynamic spatial dependencies via a learnable graph adjacency matrix and models temporal dynamics using bidirectional long short-term memory (LSTM). Extensive evaluations on BCI Competition IV 2a, 2b, and PhysioNet datasets demonstrate state-of-the-art performance, with average accuracies of 95.42%, 97.51%, and 90.15%, respectively. Ablation studies confirm the contribution of each component, and interpretability analysis reveals neurophysiologically meaningful connectivity patterns. Moreover, despite its complexity, the model maintains a feasible parameter count and an inference time of 0.32 ms per sample. These results indicate that our method offers a robust, accurate, and interpretable solution for MI-EEG decoding, with strong generalizability across subjects and tasks and meeting the requirements for potential real-time BCI applications.

Index Terms— Motor imagery classification, graph neural network (GNN), spatio-temporal modeling, interpretable learning, feature extraction.

I. INTRODUCTION

LECTROENCEPHALOGRAPH has emerged as the most widely adopted non-invasive neural signal acquisition method for brain-computer interface (BCI) systems, owing to its portability, high temporal resolution, and relatively low cost [1], [2]. Among various BCI paradigms, motor imagery (MI) has garnered significant research interest due to its independence from external stimuli, relying solely on the user's internal intention to imagine limb movements. This makes MI-BCI a promising tool for neurorehabilitation, prosthetic control, and assistive technologies [3].

Manuscript received Sep 3, 2025. This work was supported by the Natural Science Basic Research Program of Shaanxi Province under Grant 2024JC-YBQN-0659 (Corresponding author: Hongqi Li.)

Y. Wang, H. Zhang, and H. Li are with the School of Software, Northwestern Polytechnical University, Xi'an 710729, China. (e-mail: 3220251460@bit.edu.cn, zhang_haodong@mail.nwpu.edu.cn, li-hongqi@nwpu.edu.cn)

Generally, MI-BCI operates by decoding EEG signals generated when a user imagines movements of specific body parts, such as the left hand, right hand, or feet. These mental tasks induce characteristic neural oscillatory patterns known as event-related desynchronization (ERD) and event-related synchronization (ERS), primarily observed over the sensorimotor cortex [4], [5]. Despite their physiological basis, MI-EEG signals are inherently non-stationary, exhibit low signal-to-noise ratio, and show pronounced variability across subjects and sessions [6]. These characteristics render conventional machine learning approaches, which often depend on hand-crafted features (e.g., power spectral density, common spatial patterns [7], Hjorth parameters [8]) and shallow classifiers like support vector machine (SVM), or linear discriminant analysis (LDA), insufficient for capturing the complex spatio-temporal dynamics essential for robust decoding.

Recent advances in deep learning have shown promise in addressing these challenges. Convolutional neural networks (CNNs) excel at extracting spatially localized features [9], while recurrent architectures such as long short-term memory (LSTM) networks are effective for modeling temporal dynamics [10]. Hybrid models, including CNN-Transformer architectures, have further improved decoding performance by integrating spectral, spatial, and temporal learning [11]. Concurrently, graph neural networks (GNNs) have also been introduced to explicitly model the functional connectivity between EEG electrodes, treating them as nodes in a graph structure [12], [13]. Notwithstanding these developments, most existing methods still face limitations. Fixed or heuristic graph structures may not adequately capture subject-specific neural interactions, while purely reconstruction-based auto-encoders often fail to learn features optimal for classification. There remains a need for an integrated framework that jointly learns discriminative representations and dynamically adapts to the spatio-temporal characteristics of MI-EEG signals.

To bridge this gap, in this work, we propose a novel end-to-end architecture termed DRDCAE-STGNN, which combines a Discriminative Residual Dense Convolutional Autoencoder with a Spatio-Temporal Graph Neural Network. Overall, the DRDCAE module enhances feature learning by deep residual and dense connections, trained with a combined reconstruction and classification loss to ensure both representational fidelity and discriminative power. The STGNN module incorporates a learnable adjacency matrix to adaptively model inter-channel dependencies and utilizes graph convolution and bidirectional

LSTM to capture temporal evolution. Extensive experiments conducted on BCI Competition IV 2a, 2b and PhysioNet Motor Movement datasets demonstrate that our proposed model achieves state-of-the-art performance, superior generalization, and meaningful interpretability of learned brain connectivity patterns. The main contributions of this work are threefold: 1) we propose a novel discriminative autoencoder enriched with residual and dense connections to learn compact and class-relevant feature representations from MI-EEG signals; 2) a dynamic graph-based STGNN architecture is ingeniously adopted to effectively model spatial and temporal dependencies; and 3) comprehensive experiments and ablation studies confirm the superiority and interpretability of our method for MI-EEG classification tasks.

The remainder of this paper is structured as follows. Section II reviews recent related work. Section III details the proposed methodology. Section IV describes the datasets and experimental setup. Comparative results are presented in Section V, and Section VI concludes the paper with discussions on findings and future research directions.

II. RELATED WORKS

The classification of EEG signals has traditionally followed a pipeline involving signal preprocessing, feature extraction, and classification [14]. Each stage has seen extensive research and methodological developments. In the context of MI-EEG classification, spatial and temporal representations are of primary importance, as they directly capture brain-region activations and task-related dynamics. By contrast, frequency-domain features, while widely applied, reflect properties common to many physiological signals and are thus regarded as auxiliary rather than central in this setting. Recent advances have shifted toward deep learning and graph-based models, which can learn representations directly from raw or minimally processed data. This section reviews related work in four key areas: spatial feature learning, temporal modeling, spatio-temporal fusion, and deep feature learning via residual or dense connections.

A. Spatial Feature Learning

Spatial modeling plays a central role in decoding EEG signals, especially for the motor imagery signals, as it aims to capture inter-channel interactions that reflect cortical connectivity. Graph-based approaches have gained prominence for their ability to explicitly model relationships between EEG electrodes. Kipf and Welling's Graph Convolutional Network (GCN) laid the foundation for spectral graph learning [15], and was quickly adopted in EEG applications such as EEG-GCN [12], where predefined sensor adjacency was used to model spatial dependencies. However, such fixed graphs may not generalize across subjects or tasks.

To overcome this limitation, attention mechanisms were introduced to dynamically adjust graph connectivity. Graph Attention Network (GAT) [16] learns edge weights via self-attention, allowing to dynamically adjust the graph topology. More recent approaches treat the adjacency matrix as entirely learnable, enabling task-adaptive graph construction to better

capture dynamic relationships between nodes [17]. Studies [18] and [19] further advanced this idea by introducing cross-frequency and multi-scale graph learning, demonstrating that adaptive graph structures significantly improve decoding accuracy and generalizability compared to static counterparts.

B. Temporal Dependency Modeling

EEG signals are inherently temporal, and capturing their dynamics is essential for decoding motor intentions. Early approaches combined spatial encoding with recurrent modules. EEGNet, originally proposed for compact CNN-based EEG decoding, was extended with LSTM layers to form EEGNet-LSTM, enabling sequence modeling of MI signals [20]. Similarly, DeepConvNet-LSTM integrated convolutional and recurrent layers to capture temporal dependencies in P300 and epilepsy detection [21]. Despite their effectiveness, these recurrent neural networks (RNNs) are often computationally intensive and prone to vanishing gradients. Temporal convolutional networks (TCNs) such as TS-TCN offer an alternative with larger receptive fields and more stable training [22]. More recently, Transformer-based models have been applied to EEG decoding, leveraging self-attention to capture long-range temporal dependencies. Specifically, EEG-Transformer [23] and multi-branch fusion Transformers [24] have shown remarkable performance by attending to globally relevant time steps, marking a shift from local recurrent modeling to global temporal attention.

C. Fusion of Spatial-Temporal Features

Joint modeling of spatial and temporal features has proven highly effective across various EEG paradigms, where the spatio-temporal graph neural networks (STGNNs) have been particularly successful, combining graph convolutions for spatial modeling with sequential modules (e.g., LSTMs or TCNs) for temporal processing [25], [26]. For example, ST-SCGNN [27] dynamically constructs graph structures for cross-subject emotion recognition and consciousness detection, demonstrating the value of adaptive spatio-temporal modeling. Similar approaches have been applied to MI-EEG, where models such as EEG-STGNN and BrainGNN [28] use gated recurrent units and attention mechanisms to model evolving spatial interactions over time. In addition, Yan et al. [29] proposed a temporal-spatial embedding and dynamic aggregation network for EEG motion intention recognition, while Zheng et al. [30] introduced a deep multi-dilation model for 3D EEG. These studies consistently show that integrated spatio-temporal models outperform methods that treat spatial and temporal features independently.

D. Residual and Dense Connection-based Research

To improve robustness under the noisy and low-resource conditions, autoencoders have been widely employed for unsupervised EEG feature extraction. However, standard autoencoders often lack discriminative power. To address this, hybrid architectures are proposed that combine reconstruction loss with supervised learning loss. More specifically, deep residual

connections, as popularized by ResNet, help mitigate the vanishing gradient problem and enable training of deeper networks. Discriminative convolutional autoencoders (DCAEs) [31] further enhance latent feature quality through auxiliary classification loss or conditional guidance. Variational auto-encoders (VAEs) and generative adversarial networks (GANs) have also been used for EEG data augmentation and domain adaptation [32], which learn subject-invariant features by introducing probabilistic priors or adversarial training. These strategies, combined with dense connectivity, enable efficient and hierarchical feature compression, forming a strong foundation for downstream decoding tasks.

E. Research Gap and Contribution

Despite these advancements, most existing methods exhibit limitations in one or more aspects. Spatial models often rely on static graphs, temporal models may overlook spatial interactions, and autoencoders may not optimize features for classification. Moreover, few frameworks seamlessly integrate discriminative feature learning with dynamic spatio-temporal modeling, and most existing approaches often degrade under subject variability and noise, leaving room for improvement in both discriminability and generalization. Concluding the current advances, discriminative residual dense convolutional autoencoder (DRDCAE) excels at robust feature compression but lacks explicit modeling of complex spatio-temporal dependencies. STGNN is able to effectively capture dynamic relations while remaining vulnerable in terms of discriminative power and noise resistance. This motivates the integration of the two into a unified framework that combines discriminative feature learning with adaptive spatio-temporal modeling for MI-EEG classification. Our proposed framework follows this principle by unifying residual-dense encoding with classification-aware learning objectives.

III. APPROACH OF DRDCAE-STGNN NETWORK

The overall structure of the proposed DRDCAE-STGNN framework is illustrated in Fig. 1, where the model accepts the preprocessed three-dimensional EEG data (samples \times channels \times features) containing both temporal-spectral and spatial features. Following standardization (i.e., Step 1 - 4) depicted in the middle part of Fig. 1, the data is fed into the model. The backbone network mainly consists of two parts: discriminative residual dense convolutional autoencoder (DRDCAE) and the spatio-temporal graph neural network (STGNN), which are used for high-dimensional extraction of local features and global modeling across channels and time.

A. DRDCAE Module

The DRDCAE is designed to learn compact and discriminative feature representations through a combination of reconstruction and classification objectives, and consists of an encoder, decoder and classifier (see in Fig. 1). Unlike conventional autoencoders, it incorporates residual learning and dense connectivity to enhance gradient flow and feature reuse across layers.

Specifically, in the DRDCAE encoder module, the local features of each channel are first extracted through the densely connected DenseBlock. This module employs a series of 1D convolution operations, connecting all historical feature maps to enhance the feature expression capability and the stability of gradient propagation. The computation in each residual block can be formulated as:

$$y_l = \text{ReLU}(K_l(y_{l-1} + y_{l-2} + b_l)) \quad (1)$$

where K_l and b_l represents the convolutional kernel and bias term at the l -th layer, respectively. y_{l-1} and y_{l-2} denote the feature maps from the previous two layers. The residual structure allows the network to better capture the difference between input and output by adding past features before convolution.

Then, a 1×1 convolution is used to compress the channel dimension and map the data to a fixed-dimensional hidden feature space, resulting in the encoded representation. To maintain the original feature information and strengthen the model's reconstruction ability, DRDCAE sets up a symmetric decoder structure. The hidden representation is restored to its original dimensions using transposed convolution, and a Sigmoid activation function is introduced to generate the reconstructed output. Specifically, the decoder reconstructs the original image $\hat{x} \in \mathbb{R}^{H \times W \times C}$ from the latent representation z using transposed convolution operations, described as:

$$\hat{x} = \text{Sigmoid}\left(\left(\sum_{k=1}^K K_k^T z_k + b\right)\right) \quad (2)$$

where K_k^T denotes the transposed convolution kernel and z_k represents the k -th partition of the latent feature map.

A critical innovation of DRDCAE is its joint optimization via a composite loss function:

$$\mathcal{L}_{DRDCAE} = \mathcal{L}_{rec} + \lambda \mathcal{L}_{cls} \quad (3)$$

$$\mathcal{L}_{rec} = \frac{1}{N} \sum_{i=1}^N \|x_i - \hat{x}_i\|^2 \quad (4)$$

where \mathcal{L}_{rec} is the mean squared reconstruction error, which evaluates the discrepancy between the original input x_i and its reconstruction \hat{x}_i , N denotes the number of samples. \mathcal{L}_{cls} is the cross-entropy classification loss, the hyperparameter λ balances the two objectives, ensuring the learned features are both representative and discriminative.

B. Spatio-Temporal Graph Neural Network

After the initial feature extraction, the encoded features from the encoder are fed into the STGNN module to further model the spatial connectivity between EEG channels and the dynamic evolution of features in the time series. The STGNN is based on a trainable adjacency matrix and consists of three key components: adaptive graph convolution, bidirectional temporal modeling, and attention-based pooling.

1) Data Input of Learnable Graph Construction: To initialize the graph topology required by the model, Pearson correlation coefficients were calculated between all EEG channels

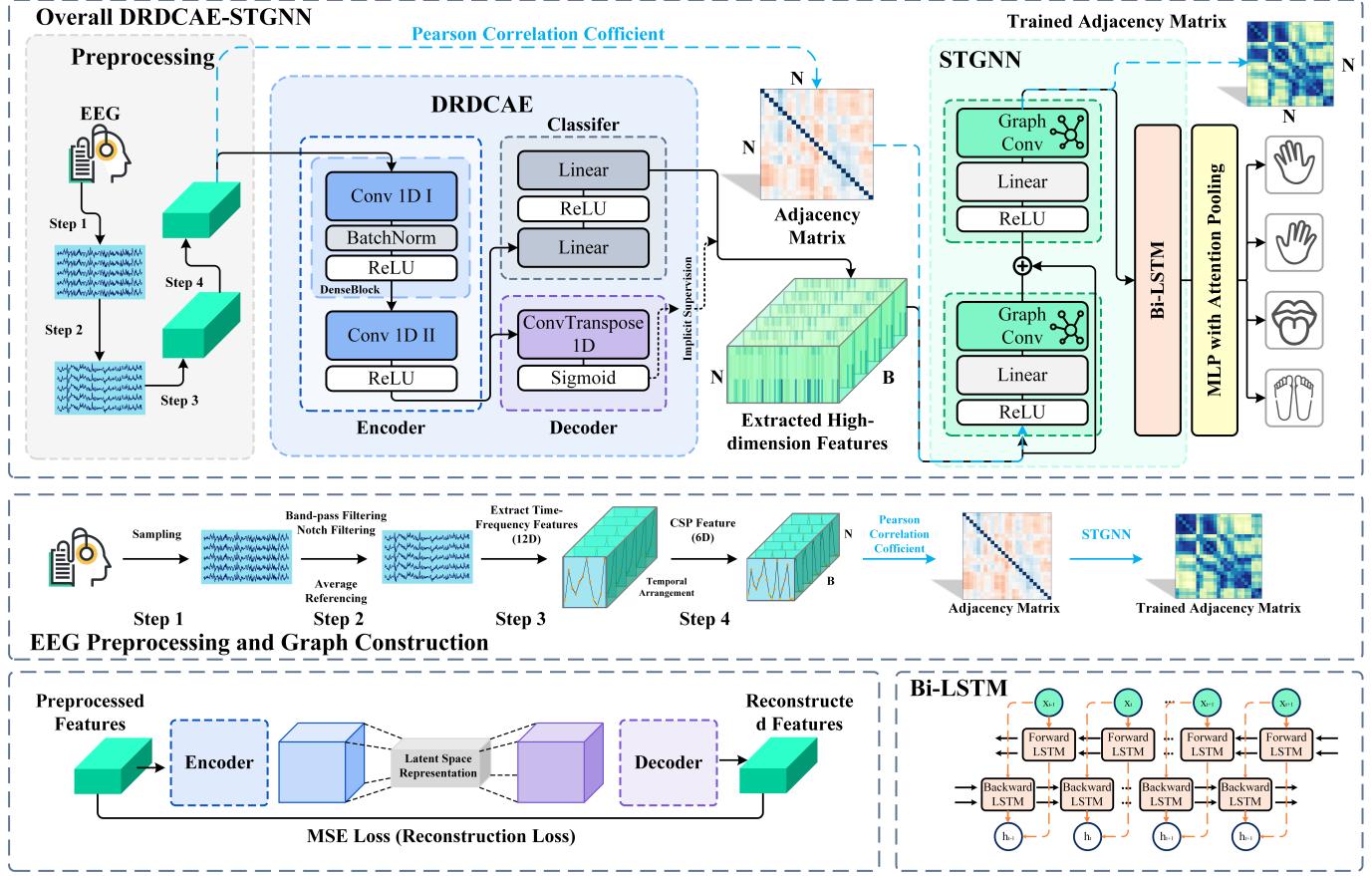


Fig. 1: The proposed DRDCAE-STGNN model architecture, where the overall structure is at the top, the signal processing flow diagram is in the middle, and the partial implementation details of DRDCAE and STGNN are at the bottom, respectively.

based on their averaged time series, resulting in a symmetric adjacency matrix $\mathbf{A} \in \mathbb{R}^{C \times C}$. During training, this matrix is made learnable to adapt to subject-specific connectivity patterns. To prevent overfitting and ensure numerical stability, self-loops are masked, and the matrix is normalized row-wise via softmax:

$$\tilde{\mathbf{A}} = \text{softmax}(\mathbf{A} \odot \mathbf{M}) \quad (5)$$

where $\mathbf{M}=\mathbf{I}-\mathbf{I}$ is a mask matrix that removes self-connections, \odot denotes the element-wise (Hadamard) product. As training progresses, the model automatically adjusts the connection strengths between nodes in $\tilde{\mathbf{A}}$ based on their semantic contribution to the classification task. This mechanism enables the transition of the graph structure from a static prior to a task-driven, learnable topology.

2) *Graph Convolutional Layers*: As shown in Fig. 1, we employ two graph convolutional layers to propagate spatial information across the node dimension. Each GraphConv integrates residual connections and batch normalization to mitigate the vanishing gradient problem and improve training stability. A mask is introduced in the graph convolution operation to prevent self-connections between channels. The graph convolution operation is defined as:

$$H^{(l+1)} = \text{ReLU}(\text{BN}(\tilde{\mathbf{A}} \cdot H^{(l)} + H^{(l)})) \quad (6)$$

where W^l and H^l denote the learnable weight matrix and the node feature matrix at the l -th layer, $\text{BN}(\bullet)$ represents batch normalization, and ReLU is the non-linear activation function.

3) *Bidirectional Temporal Modeling*: The output of the graph convolution layers is reshaped into a sequence format and processed by a bidirectional LSTM (Bi-LSTM) to capture temporal dependencies:

$$\overrightarrow{\mathbf{h}}, \overleftarrow{\mathbf{h}} = \text{Bi-LSTM}(\mathbf{x}_t, \overrightarrow{\mathbf{h}}, \overleftarrow{\mathbf{h}}) \quad (7)$$

where $x \in \mathbb{R}^{B \times N \times H}$ is the input feature vector at time step t , B is the batch size, N is the number of nodes (EEG channels), H is the hidden dimension. The forward and backward hidden states are concatenated to form a comprehensive temporal representation.

4) *Attention-Based Pooling*: To extract task-relevant global representations from all temporal node embeddings, an attention-weighted pooling mechanism is introduced along the node dimension. Attention weights α_i are computed through a two-layer fully connected network as:

$$\alpha_i = \frac{\exp(w_i^T \text{ReLU}(w_i h_i))}{\sum_{j=1}^N \exp(w_j^T \text{ReLU}(w_j h_j)))} \quad (8)$$

$$h_{\text{goa}} = \sum_{i=1}^N \alpha_i \cdot h_i \quad (9)$$

Here, w_i and w_j are learnable weight vectors used to transform the node features h_i and h_j , respectively. In the context of graph neural networks, h_i typically represents the multi-dimensional feature vector of node i (e.g., an EEG channel). The attention weight α_i measures the contribution

of node i , and the resulting global representation is passed to a softmax classifier for final prediction.

C. End-to-End Training

The entire DRDCAE-STGNN model is trained in an end-to-end manner. The overall loss function combines the DRDCAE reconstruction and classification losses with the STGNN classification loss:

$$\mathcal{L}_{total} = \mathcal{L}_{REC} + \gamma \mathcal{L}_{cls} \quad (10)$$

where γ controls the contribution of the STGNN module.

In summary, the integrated architecture with adaptive graph learning strategy retains neurophysiological prior knowledge while introducing structural flexibility. It allows the model to emphasize task-relevant inter-channel dependencies and suppress irrelevant ones, thereby improving the spatial representation capacity of the graph-based feature extractor.

IV. DATASET AND EXPERIMENTAL SETUP

To rigorously evaluate the performance and generalizability of the proposed DRDCAE-STGNN model, experiments were conducted on three publicly available motor imagery EEG datasets: BCI Competition IV Dataset 2a (BCI 2a), BCI Competition IV Dataset 2b (BCI 2b), and the PhysioNet EEG Motor Movement/Imagery Dataset (PhysioNet) [33]. These datasets vary in complexity, number of channels, and subject pools, providing a comprehensive benchmark.

A. Datasets

Dataset I: BCI 2a-This dataset contains EEG recordings from 9 subjects performing four-class MI tasks (left hand, right hand, feet, and tongue). Data was recorded using 22 EEG electrodes and 3 EOG channels at a sampling rate of 250 Hz. Each subject completed two sessions on separate days, each comprising 288 trials (72 per class), resulting in 576 trials per subject. In this study, all EEG channels signal was used, with the time window set for each trial between 3-6s.

Dataset II: BCI 2b-This dataset features binary-class MI (left hand vs. right hand) from 9 subjects. Signals were recorded from 3 bipolar channels (C3, Cz, C4) at 250 Hz. Each subject participated in five sessions, with the first two without feedback and the latter three with feedback. For consistency with common practice, we utilized data from all sessions, totaling 400 trials per subject. Its lower channel count presents a distinct challenge for spatial modeling.

Dataset III: PhysioNet-This large-scale dataset includes data from 109 subjects performing both motor execution and imagery of hand and foot movements. We focused on the two-class (left/right hand) and four-class (both hands, both feet) imagination tasks. Signals were recorded from 64 electrodes at a sampling rate of 160 Hz.

B. Data Preprocessing

A uniform preprocessing pipeline was applied to all datasets using MNE-Python. Continuous EEG data were bandpass-filtered between 0.5-100 Hz and a 50 Hz notch filter was applied to suppress line noise. The data was then re-referenced using a common average reference (CAR). Epochs were extracted from 0 to 3 seconds relative to the cue onset. A sliding window method is then used to segment each signal for subsequent feature extraction and modeling analysis. For all datasets, trials were split into an 80/20 training-testing ratio per subject.

C. Model Implementation and Training

The proposed DRDCAE-STGNN was implemented in PyTorch. Model parameters were initialized using Kaiming initialization (for convolutional layers) and Xavier (for graph and linear layers) strategies to ensure stable training. The model was trained end-to-end using the Adam optimizer, where the learning rate and batch size are being $1e^{-3}$ and 32 for DRDCAE and $2e^{-4}$ and 16 for STGNN, respectively. Dropout ($p = 0.3$) was applied to mitigate overfitting. The loss balance parameters λ and γ were empirically set to 0.3 and 1.0, respectively. DRDCAE was trained using a weighted sum of reconstruction (MSE) and classification (cross-entropy) losses, while STGNN used cross-entropy loss. Training proceeded for a maximum of 100 epochs with early stopping based on validation loss. Detailed layer configuration and parameter counts for each component of the proposed DRDCAE-STGNN model are provided in TableI.

TABLE I:
LAYER CONFIGURATION AND PARAMETER COUNTS OF PROPOSED
MODEL

Module	Layer*	Input	Output	Params
DARCAE	DB	(B, F, N)	($B, 66, N$)	$\approx 4.8K$
	Conv1D	($B, 66, N$)	($B, 64, N$)	4.3K
	CT1D	($B, 64, N$)	(B, F, N)	1.2K
	Linear	($B, 64$)	($B, 64$)	4.2K
	Linear	($B, 64$)	($B, \text{num, classes}$)	260
STGNN	GC1	(B, N, F)	(B, N, H)	0.6K
	GC2	(B, N, H)	(B, N, H)	1.0K
	Bi-L.	(B, N, H)	(B, N, H)	16.6K
	Attn.	($B, N, 2H$)	($B, N, 2H$)	4.2K
	Linear	($B, 2H$)	($B, 64$)	4.2K
	Linear	($B, 64$)	($B, \text{num, classes}$)	260

* DB: Dense Block, CT1D: ConvTranspose 1D, Attn.: Attention, GC: Graph Convolution, Bi-L.: Bi-LSTM

V. RESULTS

A. Optimization of Sliding Window Parameters

The selection of temporal segmentation parameters critically influences the model's ability to capture discriminative features. We systematically evaluated the impact of window length (ω) and step size (s) with their different combinations. The window length is set to 125, 250, and 500 sampling points, while the step size parameters are set to 62, 125, and 250 sampling points, resulting in a total of 9 experimental groups.

By comparing model performance, the proper selection of time resolution and feature redundancy is explored. As an

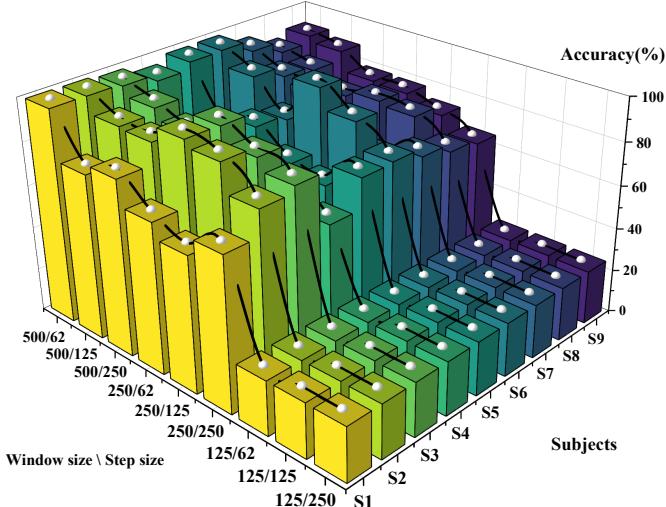


Fig. 2: Classification accuracy for nine subjects on the BCI-2a dataset under different window length ω and step size s .

example, the classification accuracy across different sliding window parameters of BCI-2a dataset is given. As illustrated in Fig.2, the configuration with a window length of 500 points (2 seconds) and a step size of 62 points (0.248 seconds) yields the highest and most stable performance across all nine subjects. Hence, this parameter set provides an optimal trade-off, where the longer window captures sufficient spectral information while the highly overlapping step ensures robustness to temporal misalignments and augments the effective training sample size. Such the configuration was consequently adopted for all subsequent experiments.

B. Comparative Performance Evaluation

The proposed DRDCAE-STGNN framework was evaluated against state-of-the-art methods on three public benchmarks. Results, detailed in TABLE II, TABLE III, and TABLE IV, demonstrate its superior and robust performance. Specifically, on BCI IV-2a dataset (4-class), the proposed model achieved an average accuracy of 95.42% ($\kappa = 0.92$, more than half of the subjects had an accuracy rate higher than 97%), signifi-

cantly outperforming comparable methods of EEGNET [34], MB3DC [36], DRDA [38], GMM-JCSFE [41] ($p < 0.01$), and SS-MEMDBF [35], JSTFD-LDA [37], FBRTS [39], ATCnet [40], TriNet [43], EEG-DG [45] ($p < 0.05$). It also surpassed recent benchmarks such as EEGNET-tabnet (92.47%) [42] and ACNN-STADP (90.54%) [44], while exhibiting the lower standard deviation (4.22), indicating exceptional consistency across subjects (S1-S9).

For the BCI IV-2b dataset (2-class), the model attained a mean accuracy of 97.51% ($\kappa = 0.96$), maintaining above 95% for all subjects and exceeding the performance of other advanced models like Trinet [43] (92.64%), GMM-JCSFE [41] (89.63%), and EEG-DG [45] (87.12%). The remarkably low standard deviation (0.90) further confirms its robustness and adaptability to low-channel configurations. On the PhysioNet dataset, the framework demonstrated strong generalization, achieving accuracies of 95.9% ($\kappa = 0.986$) for the binary task and 90.2% ($\kappa = 0.868$) for the more challenging four-class task. The results of the F1-score and AUC also exhibit a substantial improvement over existing methods, such as MBEEGNet [51], highlighting the model's capability to handle larger subject pools and higher channel counts.

Overall, the proposed DRDCAE-STGNN outperforms compared baselines in both accuracy and robustness, validating the effectiveness of its residual dense encoding and spatio-temporal graph modeling.

C. Ablation Study

Ablation experiments were conducted with four model variants to quantify the contribution of each architectural component. The results on one representative dataset of BCI-2a are listed in TABLE V, which unequivocally validate the design choices. Specifically, the full model of DRDCAE-STGNN (Model A) achieved the highest mean accuracy (95.42%). Replacing the STGNN with a two-layer fully connected network (Model B) resulted in a performance drop (94.36%), underscoring the critical role of adaptive spatial modeling via graph convolutions. Removing the DRDCAE encoder and using raw time-frequency and CSP features (Model C) caused a significant decline in accuracy (85.34%), emphasizing the

TABLE II:
CLASSIFICATION ACCURACY AND KAPPA VALUES OF DIFFERENT METHODS ON DATASET I [AVG ACC: THE AVERAGE ACCURACY (%); S1-S9 MEANS SUBJECT 1-9 INVOLVED IN THE DATASET I; ‘-’ INDICATES THAT THE EXACT VALUES WERE NOT REPORTED BY THE ORIGINAL PAPER; *: $P < 0.05$, **: $P < 0.01$.]

Year	Methods	S1	S2	S3	S4	S5	S6	S7	S8	S9	Avg Acc	Std	Kappa
2018	EEGNet [34]	85.71	78.57	91.15	95.67	89.20	85.12	79.23	81.28	80.67	85.59**	6.02	0.77
2019	SS-MEMDBF [35]	91.49	60.56	94.16	76.72	58.52	68.52	78.57	97.01	93.85	79.93**	14.99	0.54
2019	MB3DC [36]	73.97	60.14	82.93	72.29	75.84	68.99	76.00	76.86	84.67	75.02**	7.29	0.64
2020	JSTFD-LDA [37]	86.40	55.90	96.30	73.10	89.50	58.20	76.10	93.80	86.60	79.60**	14.78	-
2021	DRDA [38]	83.19	55.14	87.43	75.28	62.29	57.15	86.18	83.61	82.00	74.75**	12.96	0.66
2022	FBRTS [39]	86.10	65.20	90.00	63.80	75.60	52.40	91.10	89.00	86.50	77.70*	14.14	0.71
2023	ATCNet [40]	88.50	70.50	97.60	81.00	83.00	73.60	93.10	90.30	91.00	85.40*	9.08	0.81
2024	GMM-JCSFE [41]	89.56	86.28	87.38	84.70	85.96	86.75	84.07	86.44	85.49	78.96**	1.02	-
2024	EEGNet-tabnet [42]	94.60	79.30	96.70	86.10	84.20	78.80	96.20	90.20	90.40	92.47	6.83	0.88
2025	TriNet [43]	90.10	75.30	95.40	84.00	86.00	79.00	94.80	89.60	91.80	88.50*	6.90	0.86
2025	ACNN-STADP [44]	90.01	77.00	97.10	88.70	88.90	84.70	98.60	95.40	94.50	90.54	6.81	0.87
2025	EEG-DG [45]	89.24	64.93	94.79	85.76	68.75	61.46	95.14	88.89	87.15	81.79*	13.06	0.75
2025	DRDCAE-STGNN	99.13	98.96	97.92	94.79	98.26	98.57	93.06	87.50	90.62	95.42	4.22	0.92

TABLE III:

COMPARISON RESULTS OF DIFFERENT METHODS ON DATASET II [AVG ACC: THE AVERAGE ACCURACY(%) S1 S9 MEANS SUBJECT 1 9 INVOLVED IN THE DATASET II; ‘-’ INDICATES THAT THE EXACT VALUES WERE NOT REPORTED BY THE ORIGINAL PAPER; *: P<0.05, **: P<0.01.]

Year	Methods	S1	S2	S3	S4	S5	S6	S7	S8	S9	Avg Acc	Std	Kappa
2018	EEGNet [34]	75.94	57.64	58.43	98.13	81.25	88.75	84.06	93.44	89.69	80.48**	14.40	0.61
2021	FBCaps Net [46]	79.47	58.34	59.59	96.40	84.06	88.09	82.12	90.47	89.12	80.85**	13.36	-
2021	DRDA [38]	81.37	62.86	63.63	95.94	93.56	88.19	85.00	95.25	90.00	83.98*	12.67	0.68
2022	FBRTS [39]	82.40	75.20	86.90	95.20	89.70	80.20	90.50	91.20	91.10	86.90**	6.40	-
2023	Conformer [47]	82.50	65.71	63.75	98.44	86.56	90.31	87.81	94.38	92.19	84.63	11.49	0.6926
2024	GMM-JCSFE [41]	87.53	91.55	86.08	94.81	90.58	90.31	88.76	87.21	89.90	89.63**	2.64	-
2025	Trinet [43]	93.65	90.43	91.70	95.43	96.09	92.26	90.48	88.67	95.08	92.64**	2.57	-
2025	EEG-DG [17]	82.5	67.50	72.19	98.44	96.56	90.94	89.38	95.00	91.56	87.12*	10.89	0.74
2025	DRDCAE-STGNN	98.5	97.5	98.25	95.95	96.43	97.00	98.25	97.50	98.25	97.51	0.90	0.96

TABLE IV:

CLASSIFICATION ACCURACY AND KAPPA VALUES OF DIFFERENT METHODS ON DATASET III [AVG ACC: THE AVERAGE ACCURACY (%); AUC: AREA UNDER CURVE]

Year	Methods	Avg Acc	F1 score	AUC	Kappa	Year	Methods	Avg Acc	F1 score	AUC	Kappa
2018	EEGnet [48]	76.1	0.758	0.521	2019	2019	4-layer LSTM [52]	54.6	0.526	0.569	0.090
2019	LSTM [49]	49.9	0.494	0.507	-0.003	2020	EEGNet fusion [48]	76.2	0.784	0.849	0.523
2020	C-LSTM [50]	74.7	0.736	0.797	0.495	2021	TS-SEFFNet [53]	67.1	0.667	0.759	0.342
2022	MBEEGNet [51]	76.0	0.755	0.853	0.520	2022	MBEEGNet [51]	76.5	0.757	0.846	0.530
2025	DRDCAE-STGNN	95.9	0.958	0.917	0.986	2025	DRDCAE-STGNN	90.2	0.901	0.936	0.868

TABLE V:

ABLATION STUDY RESULTS (CLASSIFICATION ACCURACIES, %) OF FOUR MODEL VARIANTS ON THE BCI COMPETITION IV-2A DATASET

Subjects	Model A	Model B	Model C	Model D
S1	99.13	94.35	92.01	85.07
S2	98.96	96.88	93.75	81.94
S3	97.92	95.14	93.75	78.82
S4	94.79	91.67	83.33	57.64
S5	98.26	95.49	84.38	72.92
S6	98.57	95.14	82.29	67.01
S7	93.06	96.18	93.06	71.53
S8	87.50	91.32	73.61	48.96
S9	90.62	93.06	71.88	44.10
Avg Acc	95.42	94.36	85.34	67.55

importance of its discriminative feature learning and noise reduction capabilities. In addition, eliminating residual connections within the graph layers (Model D) led to the most severe performance degradation (67.55%), confirming their necessity for stabilizing the training of deep graph networks and mitigating gradient issues.

By comparing model performance, the proper selection of time resolution and feature redundancy is explored. As an example, the classification accuracy across different sliding window parameters of BCI-2a dataset is given. As illustrated in Fig. 2, the configuration with a window length of 500 points (2 seconds) and a step size of 62 points (0.248 seconds) yields the highest and most stable performance across all nine subjects. Hence, this parameter set provides an optimal trade-off, where the longer window captures sufficient spectral information while the highly overlapping step ensures robustness to temporal misalignments and augments the effective training sample size. Such the configuration was consequently adopted for all subsequent experiments.

D. Analysis of Computational Efficiency

We performed a complexity analysis with empirical evaluations to quantify the proposed model complexity, where the

forward computational paths and tensor operation costs of each architecture were considered based on consistent input configurations. The running hardware configuration is AMD Ryzen 7 4800H with Radeon Graphics CPU and NVIDIA GeForce GTX 1650 Ti GPU. The results were measured on the BCI-2a dataset and were benchmarked against several classical models, as reflected in TABLE VI. As we can observe, while the integrated architecture of the DRDCAE-STGNN (with the coupled residual and dense connections, the graph convolution, bidirectional temporal modeling, and attention pooling) results in a longer training time (46.82 s), it maintains a moderate parameter count (41.92K). Most critically, the average inference time per sample is 0.3174 ms, well below the 100 ms threshold for real-time BCI operation. This demonstrates a favorable balance between high decoding accuracy and practical computational efficiency for potential deployment in the real world.

TABLE VI:
COMPARISON OF EFFICIENCY AND COMPLEXITY FOR VARIOUS MODELS

Models	Parameter	Training Time (s)	Inference Time (10 ⁻² ms)
SVM	0	1.2	-
CNN-LSTM	97.46 K	18.86	5.75
CNN-GNN	16.99 K	13.12	4.92
DCGAN-D [54]	432.97 K	9.13	3.15
DRDCAE-STGNN	41.92 K	46.82	31.74

E. Interpretability of Learned Graph Structures

The learnable adjacency matrix, initialized from raw EEG features, is dynamically updated during training to capture task-relevant inter-channel dependencies and thus providing a window into the model’s spatial reasoning. Fig. 3 visualizes the graph connectivity before and after training across all three datasets. As we can see, the initial adjacency matrices after

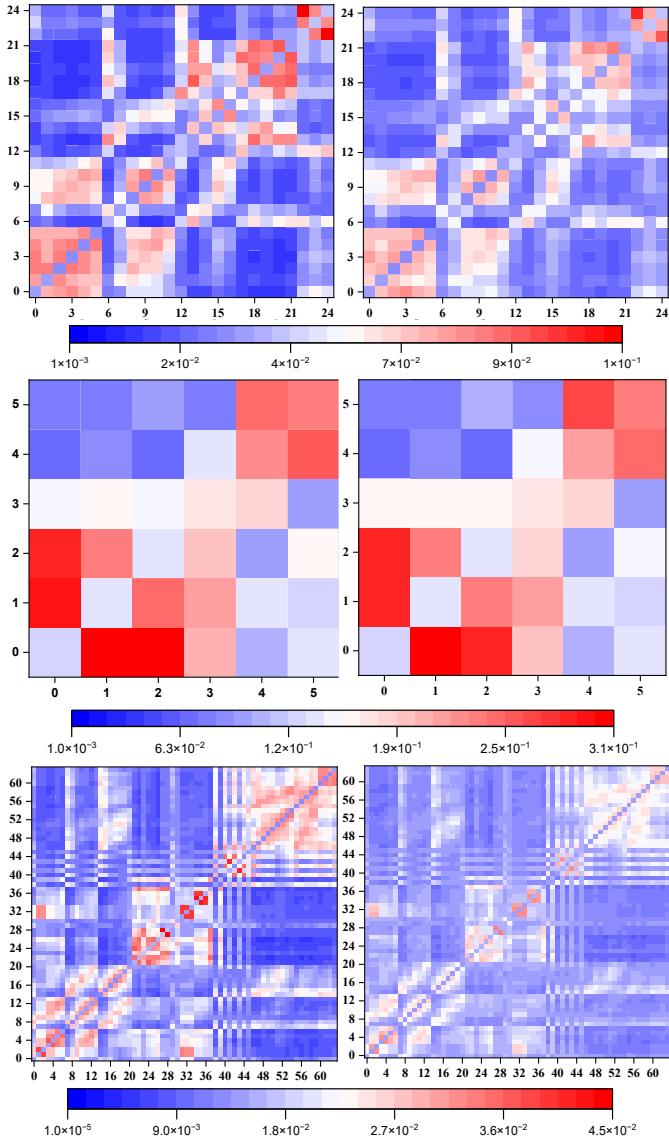


Fig. 3: Channel connectivity adjacency matrices before (left) and after (right) training across three datasets of BCI2a (top), BCI2b (middle), and PhysioNet (bottom). The coordinate numbers correspond to the EEG electrode under the International 10-20 system, and the color maps represent the connection strength between EEG channels.

Softmax normalization exhibit sparse and low-weight connections (more deep blue parts that indicating weaker relation), with only moderate symmetric correlations between adjacent channels, reflecting the high noise and non-stationarity of EEG signals. In contrast, post-training matrices reveal that the model consistently enhanced physiologically meaningful connections. Especially, considering the exact EEG electrode location, it formed clusters between frontal-parietal and sensorimotor regions, while suppressing spurious or noisy links. This indicates an adaptive learning process that prioritizes motor imagery task-relevant neural pathways.

Further analysis via the chord diagram (Fig. 4) identifies the top 10 connections that underwent the largest strengthening during training on dataset I of BCI-2a. The prominence of links involving central (e.g., C3), parietal (e.g., CP1, P1), and frontal (e.g., Fz, FC4) electrodes aligns perfectly with the

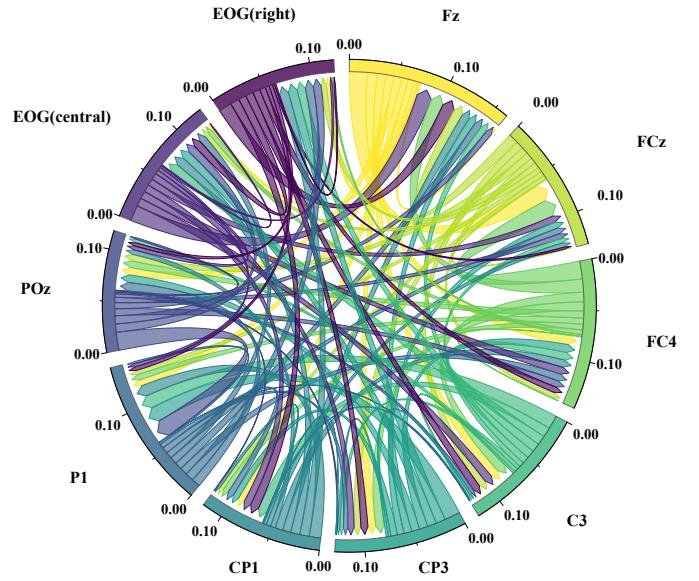


Fig. 4: Chord diagram of the top 10 EEG channel pairs with the largest changes in connection strength.

neuroanatomical correlates of motor imagery, providing strong evidence for the neurophysiological validity and interpretability of the model’s learned representations.

VI. DISCUSSION AND CONCLUSION

A. Discussion of the Results and Research Contributions

To further improve the decoding effect while maintain the interpretability, a novel DRDCAE-STGNN framework has been proposed in this study. The consistently superior performance of the framework across three distinct datasets underscores its efficacy as a powerful tool for MI-EEG decoding. The results affirm that the integration of discriminative feature learning (via DRDCAE) with adaptive spatio-temporal modeling (via STGNN) addresses core challenges in the field: the non-stationarity of EEG signals and the high variability between subjects.

The key innovation of this work is the end-to-end learning of a neuro-physiologically plausible graph structure. Unlike methods relying on static, anatomically predefined graphs, our learnable adjacency matrix allowed the model to discover and emphasize task-specific functional connectivity. The post-hoc analysis (Fig. 3, Fig. 4) revealed that the model autonomously strengthened connections within well-established sensorimotor networks (e.g., between frontal, central, and parietal regions) while suppressing irrelevant noise. This capability transforms the model from a black-box classifier into a more transparent tool that can offer insights into the underlying neural mechanisms of MI. Importantly, without explicit anatomical priors, the model automatically enhances connections between motor-related regions (e.g., Fz-FCz, FC4-FCz, Fz-FC4), consistent with classical motor imagery studies [55]. Indeed, STGNN employs a soft-attention masked graph learning mechanism, enabling spatial sparsity and dynamic inter-channel dependencies, thus offering greater adaptability compared with static graphs.

The implications of this work are threefold: 1) It presents a highly accurate and generalizable model for MI-BCIs, with direct potential applications in neurorehabilitation and assistive control; 2) It advances the interpretability of deep learning models in neuroscience by demonstrating how learnable graph structures can reveal biologically meaningful connectivity; 3) It provides a versatile framework that can be adapted for other EEG-based tasks that rely on spatio-temporal feature learning.

B. Limitations and Future Work

In conclusion, we have proposed and validated the DRDCAE-STGNN, a novel deep learning framework that sets a new state-of-the-art for MI-EEG classification. Its success stems from a synergistic design: the DRDCAE module learns robust and discriminative latent representations by jointly optimizing for reconstruction and classification, while the STGNN module dynamically captures the evolving spatial relationships between EEG channels and their temporal context. Extensive experimental results on public benchmarks confirm that our method outperforms existing approaches in both accuracy and robustness. The adaptive graph learning component further provides neurophysiologically interpretable insights into functional brain connectivity patterns during motor imagery.

However, several limitations of our current work must be acknowledged. First, while the model demonstrates strong overall performance, its accuracy varies across subjects (e.g., S8 and S9 in BCI-2a). This variability suggests that the current architecture, though adaptive, may not fully compensate for the pronounced physiological and cognitive differences between individuals. Second, the computational cost during training, though justified by the performance gains, is non-trivial. This could pose a barrier for deployment in scenarios requiring rapid model personalization. Third, while the graph provides high-level interpretability, the exact contribution of specific time-points or frequency bands to the final decision remains less transparent, indicating a need for more granular explainability techniques.

Looking forward, we plan to address the current limitations through several avenues. First, to enhance the cross-subject generalization ability, we will attempt to strengthen model adaptability through subject-invariant representation learning that combines adversarial domain alignment with covariance correction on the symmetric positive definite (SPD) manifold [56]. Furthermore, we will introduce parameter-efficient test-time adaptation modules to minimize calibration overhead for new users, thereby enhancing the practical utility of our framework in real-world BCI applications. Second, for the multi-granular interpretability, we will develop enhanced explanation methods that integrate gradient-based relevance analysis with perturbation testing to generate channel- and time-frequency-level attributions. These attributions will be quantitatively validated against well-established μ/β motor rhythm patterns [57]. Finally, we will extend the framework to asynchronous ERP detection and emotion recognition within multi-task EEG paradigms, and evaluate it with rigorous leave-one-subject-out and cross-dataset validation protocols to ensure robustness and scalability. These efforts will ultimately facilitate more

reliable and trustworthy BCI systems for clinical and assistive applications.

REFERENCES

- [1] X. Gu, Z. Duan, J. Wang, and D. C. Wunsch, “EEG-based brain-computer interfaces (BCIs): A survey of recent studies on signal sensing technologies and computational intelligence approaches and their applications,” *arXiv: 2001.11337*, Jan. 2020.
- [2] D. Wu, Y. Xu, and B.-L. Lu, “Transfer learning for EEG-based brain-computer interfaces: A review of progress made since 2016,” *arXiv: 2004.06286*, Apr. 2020.
- [3] A. Lebedev and M. A. L. Nicolelis, “Brain-machine interfaces: past, present and future,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 13, no. 2, pp. 231-240, Jun. 2005.
- [4] 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, Art. no. 2798, Mar. 2023.
- [5] J. Fumanal-Idocin, Y.-K. Wang, C.-T. Lin, J. Fernández, J. A. Sanz, and H. Bustince, “Motor-imagery-based brain computer interface using signal derivation and aggregation functions,” *Neurocomputing*, vol. 421, pp. 1-10, Apr. 2021.
- [6] T. Lotey, P. Keserwani, D. P. Dogra, and P. P. Roy, “Feature reweighting for EEG-based motor imagery classification,” *arXiv:2308.02515*, Aug. 2023.
- [7] C. Vidaurre, N. Krämer, B. Blankertz, and A. Schlögl, “Time domain parameters as a feature for EEG-based brain-computer interfaces,” *Neural Netw.*, vol. 22, no. 9, pp. 1313-1319, 2009.
- [8] Y. LeCun, Y. Bengio, and G. Hinton, “Deep learning for computer vision: A brief review,” *Proc. IEEE*, vol. 103, no. 6, pp. 1077-1095, Jun. 2015.
- [9] F. M. Garcia-Moreno, M. Bermudez-Edo, M. J. Rodríguez-Fórtiz, and J. L. Garrido, “A CNN-LSTM deep learning classifier for motor imagery EEG detection using a low-invasive and low-cost BCI headband,” in *Proc. 16th Int. Conf. Intell. Environ. (IE)*, 2020, pp. 1-4.
- [10] Z. Khademi, F. Ebrahimi, and H. M. Kordy, “A transfer learning-based CNN and LSTM hybrid deep learning model to classify motor imagery EEG signals,” *Comput. Biol. Med.*, vol. 143, Art. no. 105288, 2022.
- [11] H. Li, H. Zhang and Y. Chen, “Dual-TSST: A dual-branch temporal-spectral-spatial Transformer model for EEG decoding,” *IEEE J. Biomed. Health Inform.*, doi: 10.1109/JBHI.2025.3577611.
- [12] A. Demir, T. Koike-Akino, Y. Wang, M. Haruna, and D. Erdogmus, “EEG-GNN: Graph neural networks for classification of electroencephalogram (EEG) signals,” in *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, 2021, pp. 1061-1067.
- [13] Y. Hou *et al.*, “GCNs-Net: A graph convolutional neural network approach for decoding time-resolved EEG motor imagery signals,” *arXiv:2006.08924*, Jun. 2020.
- [14] A. K. Singh and S. Krishnan, “Trends in EEG signal feature extraction applications,” *Front. Artif. Intell.*, vol. 5, p. 1072801, 2022.
- [15] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks,” in *Proc. 5th Int. Conf. Learn. Represent. (ICLR)*, 2017, pp. 1-14.
- [16] P. Velickovic, G. Cucurull, A. Casanova, A. Romero, P. Lio, and Y. Bengio, “Graph attention networks,” in *Proc. 6th Int. Conf. Learn. Represent. (ICLR)*, 2018.
- [17] S. Saha and R. Bowden, “Learning adaptive neighborhoods for graph neural networks,” in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, 2023, pp. 22541-22550.
- [18] T. Wei and H. Wang, “Cross-frequency graph neural networks for EEG based motor imagery,” in *Proc. 8th Int. Conf. Adv. Algorithms Control Eng. (ICAACE)*, 2025, pp. 2477-2480.
- [19] H. Shi, Z. Pei, Z. Tang, M. Li, Y. Fan, and J. Zhang, “A dense multi-scale temporal feature fusion and spatial information embedding model for electroencephalogram-based motor imagery classification,” in *Proc. 8th Int. Conf. Adv. Algorithms Control Eng. (ICAACE)*, 2025, pp. 1-6.
- [20] P. Saideepthi, A. Chowdhury, P. Gaur, and R. B. Pachori, “Sliding window along with EEGNet-based prediction of EEG motor imagery,” *IEEE Sens. J.*, vol. 23, no. 15, pp. 17703-17713, Aug. 2023.
- [21] A. Ditthaphron, N. Banluesombatkul, S. Ketrat, E. Chuangsawanich, and T. Wilaiprasitpon, “Universal joint feature extraction for P300 EEG classification using multi-task autoencoder,” *IEEE Access*, vol. 7, pp. 68415-68428, 2019.

- [22] Y. Zuo, H. Zhu, S. Wang, K. Peng, R. Zhao, and L. Zhang, "TS-TCN: An integrated architecture for time-sensitive networking and train communication network," in *Proc. 16th Int. Conf. Commun. Softw. Netw. (ICCSN)*, Ningbo, China, 2024, pp. 142-148.
- [23] J. Luo, Q. Cheng, H. Wang, Q. Du, Y. Wang, and Y. Li, "MI-MBFT: Superior motor imagery decoding of raw EEG data based on a multibranch and fusion transformer framework," *IEEE Sens. J.*, vol. 24, no. 21, pp. 34879-34891, Nov. 2024.
- [24] W. Chen, Y. Luo, and J. Wang, "Three-branch temporal-spatial convolutional transformer for motor imagery EEG classification," *IEEE Access*, vol. 12, pp. 79754-79764, 2024.
- [25] X. Li, B. Qian, J. Wei, A. Li, X. Liu, and Q. Zheng, "Classify EEG and reveal latent graph structure with spatio-temporal graph convolutional neural network," in *Proc. IEEE Int. Conf. Data Min. (ICDM)*, Beijing, China, 2019, pp. 389-398.
- [26] C. Tang, L. Yang, G. Cao, J. Du, and Q. Zhang, "ST-GCN: EEG emotion recognition via spectral graph and temporal analysis with graph convolutional networks," in *Proc. IEEE Int. Conf. Bioinform. Biomed. (BIBM)*, Lisbon, Portugal, 2024, pp. 3748-3751.
- [27] J. Pan *et al.*, "ST-SCGNN: A spatio-temporal self-constructing graph neural network for cross-subject EEG-based emotion recognition and consciousness detection," *IEEE J. Biomed. Health Inform.*, vol. 28, no. 2, pp. 777-788, Feb. 2024.
- [28] X. Li *et al.*, "BrainGNN: Interpretable brain graph neural network for fMRI analysis," *Med. Image Anal.*, vol. 74, Art. no. 102233, 2021.
- [29] W. Yan, J. Liu, J. Luo, W. Liu, Y. Ma, and Q. Zhang, "A temporal-spatial embedding and dynamic aggregation network with adaptive weighting spectrum for EEG motion intention recognition," *IEEE Trans. Instrum. Meas.*, vol. 74, pp. 1-11, 2025, Art. no. 4007611.
- [30] Y. Zheng, X. Zhao, L. Yao and L. Cao, "Deep multidilation temporal and spatial dependence modeling in stereoscopic 3-D EEG for visual discomfort assessment," *IEEE Trans. Syst. Man Cybern.: Syst.*, vol. 54, no. 4, pp. 2125-2136, April 2024.
- [31] H. Zheng *et al.*, "DCAE: A dual conditional autoencoder framework for the reconstruction from EEG into image," *Biomed. Signal Process. Control*, vol. 81, Art. no. 104440, 2023.
- [32] C. Tian, Y. Ma, J. Cammon, F. Fang, Y. Zhang and M. Meng, "Dual-encoder VAE-GAN with spatiotemporal features for emotional EEG data augmentation," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 2018-2027, 2023.
- [33] X. Wang, V. Liesaputra, Z. Liu, Y. Wang, and Z. Huang, "An in-depth survey on deep learning-based motor imagery electroencephalogram (EEG) classification," *Artif. Intell. Med.*, vol. 147, Art. no. 102738, 2024.
- [34] B. Xu, Y. Xu, H. Liu, Y. Luo, W. Zhao, and D. Wang, "Wavelet transform time-frequency image and convolutional network-based motor imagery EEG classification," *IEEE Access*, vol. 7, pp. 6084-6093, 2019.
- [35] P. Gaur, R. B. Pachori, H. Wang, and G. Prasad, "A multi-class EEG-based BCI classification using multivariate empirical mode decomposition based filtering and Riemannian geometry," *Expert Syst. Appl.*, vol. 95, pp. 201-211, 2018.
- [36] X. Zhao, H. Zhang, G. Zhu, F. You, S. Kuang, and L. Sun, "A multi-branch 3D convolutional neural network for EEG-based motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 10, pp. 2164-2177, Oct. 2019.
- [37] A. Jiang, J. Shang, X. Liu, Y. Tang, H. K. Kwan, and Y. Zhu, "Efficient CSP algorithm with spatio-temporal filtering for motor imagery classification," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 4, pp. 1006-1016, Apr. 2020.
- [38] H. Zhao, Q. Zheng, K. Ma, H. Li, and Y. Zheng, "Deep representation-based domain adaptation for nonstationary EEG classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 2, pp. 535-545, Feb. 2021.
- [39] H. Fang, J. Jin, I. Daly, and X. Wang, "Feature extraction method based on filter banks and Riemannian tangent space in motor-imagery BCI," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 6, pp. 2504-2514, Jun. 2022.
- [40] H. Altaheri, G. Muhammad, and M. Alsulaiman, "Physics-informed attention temporal convolutional network for EEG-based motor imagery classification," *IEEE Trans. Ind. Inform.*, vol. 19, no. 2, pp. 2249-2258, Feb. 2023.
- [41] C. Liao, S. Zhao, and J. Zhang, "Motor imagery recognition based on GMM-JCSFE model," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 3348-3357, 2024.
- [42] P. Chaudhary, N. Dhankhar, A. Singhal, and K. P. S. Rana, "A two-stage transformer based network for motor imagery classification," *Med. Eng. Phys.*, vol. 128, Art. no. 104154, 2024.
- [43] H. F. Abbasi, M. A. Abbasi, S. Jianbo, X. Liping, and X. Yu, "TriNet: A hybrid feature integration approach for motor imagery classification in brain-computer interface," *IEEE Access*, vol. 13, pp. 115406-115418, 2025.
- [44] A. Raza and M. Z. Yusoff, "An adaptive convolutional neural network with spatio-temporal attention and dynamic pathways (ACNN-STADP) for robust EEG-based motor imagery classification," *IEEE Access*, vol. 13, pp. 106387-106405, 2025.
- [45] X.-C. Zhong, Y. Zhou, S. Wang, J. Li, M. Xu, and C. Guan, "EEG-DG: A multi-source domain generalization framework for motor imagery EEG classification," *IEEE J. Biomed. Health Inform.*, vol. 29, no. 4, pp. 2484-2495, Apr. 2025.
- [46] M. Xu, J. Yao, Y. Zheng, and Y. Lin, "Multi-rhythm capsule network recognition structure for motor imagery classification," *J. Web Eng.*, vol. 20, no. 3, pp. 759-774, May 2021.
- [47] Y. Song, Q. Zheng, B. Liu, and X. Gao, "EEG Conformer: Convolutional transformer for EEG decoding and visualization," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 31, pp. 710-719, 2023.
- [48] K. Roots, Y. Muhammad, and N. Muhammad, "Fusion convolutional neural network for cross-subject EEG motor imagery classification," *Computers*, vol. 9, no. 3, p. 72, 2020.
- [49] Z. Tayeb *et al.*, "Validating deep neural networks for online decoding of motor imagery movements from EEG signals," *Sensors*, vol. 19, no. 1, Art. no. 210, 2019.
- [50] D. Freer and G.-Z. Yang, "Data augmentation for self-paced motor imagery classification with C-LSTM," *J. Neural Eng.*, vol. 17, no. 1, Art. no. 016041, 2020.
- [51] G. A. Altuwairji and G. Muhammad, "A multibranch of convolutional neural network models for electroencephalogram-based motor imagery classification," *Biosensors*, vol. 12, no. 1, p. 22, 2022.
- [52] X. Wang, V. Liesaputra, Z. Liu, Y. Wang, and Z. Huang, "An in-depth survey on deep learning-based motor imagery electroencephalogram (EEG) classification," *Artif. Intell. Med.*, vol. 147, Art. no. 102738, 2024.
- [53] Y. Li, L. Guo, Y. Liu, J. Liu, and F. Meng, "A temporal-spectral-based squeeze-and-excitation feature fusion network for motor imagery EEG decoding," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1534-1545, 2021.
- [54] K. Rasheed, J. Qadir, T. J. O'Brien, L. Kuhlmann and A. Razi, "A generative model to synthesize EEG data for epileptic seizure prediction," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 2322-2332, 2021.
- [55] M. Lotze, P. Montoya, M. Erb, E. Hülsmann, H. Flor, U. Klose, N. Birbaumer, and W. Grottkau, "Activation of cortical and cerebellar motor areas during executed and imagined hand movements: an fMRI study," *J. Cog. Neurosci.*, vol. 11, no. 5, pp. 491-501, Sep. 1999.
- [56] T. Jia, L. Meng, S. Li, J. Liu, and D. Wu, "Federated motor imagery classification for privacy-preserving brain-computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 32, pp. 3442-3451, 2024.
- [57] S. Li, Z. Wang, H. Luo, L. Ding, and D. Wu, "T-TIME: Test-time information maximization ensemble for plug-and-play BCIs," *IEEE Trans. Biomed. Eng.*, vol. 71, no. 2, pp. 423-432, 2024.