Multi-scale Dual-domain Fusion-Based Spatio-Temporal Graph Learning for fNIRS Signal Classification

Mengxiang Chu[1,†], Xingxing Cheng[1,†], Xichen Wang[1], Jingjing Yu[2], Xiangyu Li[1], Siyi Du[1], Xiaowei He[1], Xiao Li[1], and Hongbo Guo[1,*]

[1] School of Electronic Information (Artificial Intelligence), Northwest University, Xi'an, 710127, China.
[2] School of Physics and Information Technology, Shaanxi Normal University, Xi'an, 710062, China.
[†] Co-first Author; * Corresponding Author, Guohb@nwu.edu.cn.

Abstract. Functional near-infrared spectroscopy (fNIRS) is a promising modality for brain–computer interfaces (BCIs), yet multi-class neural decoding remains limited by physiological noise, hemodynamic response function (HRF) sluggishness, and scale-dependent spatio-temporal heterogeneity. To overcome these challenges, we propose Multi-scale Dual-domain Fusion-Based Spatio-Temporal Graph Learning (MsDFSTGL), a biologically informed hierarchical framework for fNIRS decoding. MsDFSTGL first builds a noise-resilient topological prior by modeling dual-domain HbO2-HbR neurovascular coupling via a Time-Frequency Global Correlation (TFGC) module and Graph Convolutional Networks (GCN). To address HRF sluggishness and scale heterogeneity, parallel multi-scale branches utilize scale-adaptive TFGC and Multi-Stage Differential Encoding (MSDE) to disentangle transient neural triggers from slow hemodynamics. Finally, a Multi-Scale Multi-Head Cross-Attention (MSMHCA) mechanism captures inter-scale dependencies, which are adaptively aggregated by a Cross-Scale Mixture-of-Experts (CS-MoE) to minimize inter-subject variability. Extensive evaluations on three heterogeneous fNIRS datasets demonstrate superior multi-class decoding performance, achieving accuracies of 95.51% on Mental Arithmetic, 76.85% on Unilateral Finger and Foot Tapping, and 34.82% on a challenging Action Recognition Motor Imagery dataset, with strong cross-subject generalization. By explicitly disentangling non-stationary neurovascular dynamics, MsDFSTGL establishes a biologically grounded and robust framework for complex neural decoding in motor rehabilitation. Code is available at https://github.com/astro1109/MsDFSTGL.

Keywords: fNIRS; Time-Frequency Graph Correlation; Differential Encoding; Cross-Scale Mixture-of-Experts; Brain-Computer Interfaces.

1 Introduction

The Functional near-infrared spectroscopy (fNIRS) shows great promise for Brain–Computer Interfaces (BCIs) and stroke rehabilitation due to its porta-

bility and motion tolerance [2, 8, 13]. However, robust multi-class neural decoding remains challenging, with accuracy often plateauing around 70–77.5% for complex motor tasks [12–14, 19, 21, 22, 24, 25]. This limitation arises because such tasks elicit weakly activated, spatially overlapping, and temporally similar hemodynamic responses, making subtle non-stationary dynamics difficult to disentangle [11, 17]. These performance bottlenecks are primarily driven by three fundamental limitations in fNIRS signal modeling: (1) Systemic physiological artifacts severely corrupt neural activation patterns [3, 4]; (2) The intrinsic sluggishness of the hemodynamic response function (HRF) obscures transient neural triggers [20]; and (3) Scale-dependent spatio-temporal heterogeneity and pronounced inter-subject variability severely limit model generalization [6, 18].

Recent deep learning advances have substantially improved connectivity and temporal modeling in fNIRS decoding, yet they still suffer from critical limitations. On the spatial front, methods like ID-CapsuleNet [14], ST prediction with Transformers and foundation models [16] and hybrid GCNs [26] preserve hierarchical patterns and spatial topologies. However, these approaches predominantly construct graphs from single-hemoglobin time-domain correlations, largely overlooking the intrinsic phase-oppositional coupling between oxyhemoglobin (HbO2) and deoxyhemoglobin (HbR), leaving them vulnerable to systemic physiological artifacts [3, 4]. On the temporal front, sequence models such as fNIRS-T [25], MedGNN [7], and fNIRSNet [24] explore multi-resolution spatio-temporal architectures. Nevertheless, they often struggled to effectively overcome the intrinsic sluggishness of the hemodynamic response function (HRF) [20]. Static single-scale [9, 10, 23] and isolated-domain frameworks inadequately model [3, 15] fine-grained transient triggers and macroscopic sustained patterns simultaneously, rendering them highly sensitive to pronounced inter-subject variability [6, 18] and limiting cross-subject clinical generalization.

To address these limitations, we propose Multi-scale Dual-domain Fusion-Based Spatio-Temporal Graph Learning (MsDFSTGL) for fNIRS multi-class decoding. The design of MsDFSTGL is grounded in neurovascular coupling, temporal phase synchronization, and frequency-energy similarity of functional connectivity. Our contributions are threefold: (1) We introduce an MsDFSTGL to unify noise-resilient spatial topology modeling, HRF-aware temporal disentanglement, and adaptive cross-scale fusion for robust fNIRS classification. (2) We design a Time-Frequency Graph Correlation (TFGC) module to construct noise-resilient spatial topology via dual-domain HbO2/HbR modeling, alleviating connectivity degradation caused by hemodynamic delays. (3) We develop a cascaded multi-scale architecture in which scale-specific TFGC and Multi-Stage Differential Encoding (MSDE) refine multi-resolution features, Multi-Scale Multi-Head Cross-Attention (MSMHCA) models inter-scale dependencies, and Cross-Scale Mixture-of-Experts (CS-MoE) adaptively aggregates features to enhance cross-subject generalization. Superior performance on noisy real-world fNIRS datasets and consistent cross-dataset gains demonstrate that the learned topology and attention patterns capture meaningful brain activity.

2 Method

Robust fNIRS neural decoding requires addressing systemic physiological noise, intrinsic sluggishness of the hemodynamic response function (HRF), and scale-dependent spatiotemporal heterogeneity. To this end, we propose the MsDF-STGL framework whose overall architecture is shown in Fig. 1(a). Given preprocessed paired fNIRS signals $X \in \mathbb{R}^{N\times T\times 2}$, where N, T, and 2 denote the number of recording channels, time points, and hemoglobin modalities (HbO_2/HbR), respectively. We first concatenate HbO2 and HbR along the channel dimension to obtain a $2N$-channel joint representation. The framework workflow is as follows: (1) An initial TFGC constructs a noise-robust global topological prior, followed by a 4-layer ($4L$) GCN for initial spatial aggregation. (2) The aggregated features are fed into parallel multi-scale 1D temporal convolutions (MSConvs) to generate m resolution-specific branches $R_1, R_2, ..., R_m$, where m is the number of multi-scale branches. Each branch is sequentially processed by a scale-specific TFGC, a Multi-Stage Differential Encoding (MSDE), and a 4-layer GCN to produce scale-specific feature queries $Q_1, Q_2, ..., Q_m$. (3) A Multi-Scale Multi-Head Cross-Attention (MSMHCA) module models inter-scale dependencies. (4) A Cross-Scale Mixture-of-Experts (CS-MoE) module adaptively aggregates cross-scale features for final classification via Softmax.

2.1 Time-Frequency Graph Correlation

Conventional single-domain, single-modality graph construction fails to build noise-resilient functional connectivity, leaving GCN aggregation vulnerable to fNIRS physiological artifacts. The TFGC module (Fig. 1(b)) jointly models time-frequency dual-domain connectivity of the $2N$-channel representation to construct robust, scale-adaptive topological graphs. For input features, we compute the temporal Pearson correlation matrix $A_{time} \in \mathbb{R}^{2N\times 2N}$ to capture intra- and inter-modality coupling, and the frequency-domain correlation matrix $A_{freq} \in \mathbb{R}^{2N\times 2N}$ via Fast Fourier Transform (FFT) to filter baseline drifts. A data-driven gating unit, driven by the disparity map $\Delta A = |A_{time} - A_{freq}|$, generates weight matrix W for adaptive fusion symmetric A_m:

$$A_m = W \odot \max(A_{time}, A_{freq}) + (1 - W) \odot \mathrm{Avg}(A_{time}, A_{freq}) \tag{1}$$

where $\odot$ denotes the Hadamard product. A hard thresholding operation enforces topological sparsity on A_m, with the refined adjacency matrix guiding subsequent GCN aggregation.

2.2 Multi-Stage Differential Encoding

The intrinsic sluggishness of HRF obscures transient neural triggers in slow metabolic trends, limiting fNIRS feature discriminability. The MSDE module (Fig. 1(c)) works in parallel with the scale-specific TFGC module, and decouples transient neural gradients from sustained metabolic states via a residual architecture to counteract HRF sluggishness. For scale-specific input features, we

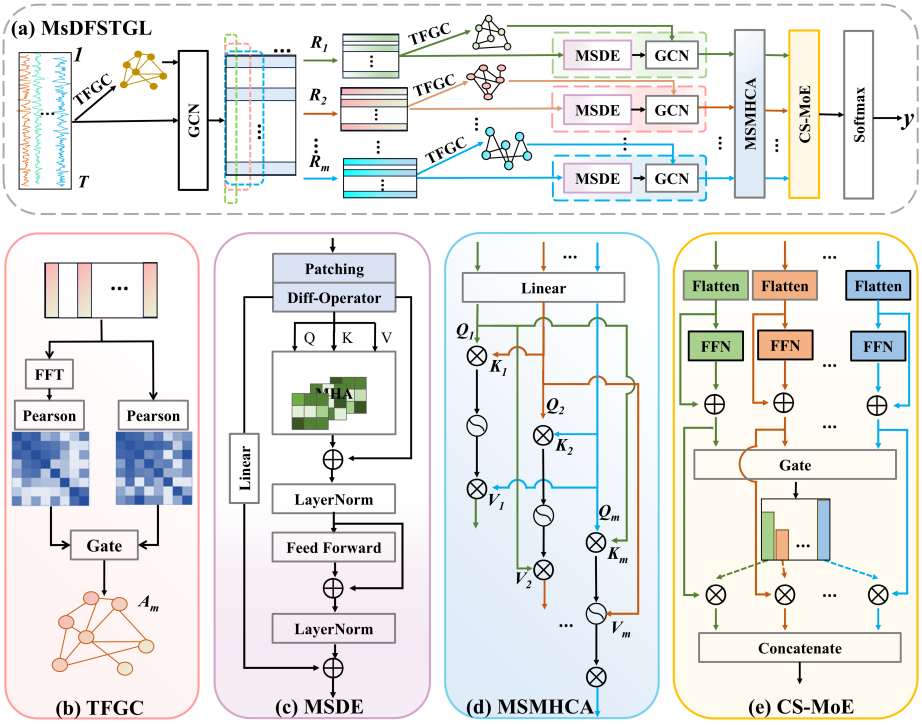

Fig. 1. MsDFSTGL framework and core modules: (a) Overall architecture; (b) TFGC for noise-robust global topological prior; (c) MSDE for HRF sluggishness mitigation; (d) MSMHCA for inter-scale dependency modeling; (e) CS-MoE for adaptive cross-scale aggregation.

first standardize sequence length via scale-specific patching, where the patching length is aligned with the temporal resolution of its corresponding multi-scale branch. We then extract first-order temporal gradients within sliding windows $\Delta F_{SW} = F_{SW+1} - F_{SW}$ to amplify instantaneous neural variations and filter slow baseline drifts. Differential features are fed into a Multi-Head Attention (MHA) block to model transient activation dynamics, followed by layer normalization, feed-forward network and residual connections to generate enhanced transient representations. The output of MSDE is fused with the topological prior from the parallel TFGC module, forming enriched node features for the subsequent 4-layer GCN.

2.3 Multi-Scale Multi-Head Cross-Attention

Existing multi-scale fusion strategies ignore cross-scale complementary dependencies, leading to insufficient utilization of heterogeneous scale-specific information. The MSMHCA module (Fig. 1(d)) models inter-scale dependencies via cross-attention for adaptive multi-scale information aggregation. For each target

scale m, we take the scale-specific GCN output Q_m as the Query matrix, and concatenate features from all other scales to form Key and Value matrices:

$$K_m = V_m = \text{Concat}(Q_1, ..., Q_{m-1}, Q_{m+1}, ..., Q_n) \tag{2}$$

Cross-scale attention is computed via scaled dot-product attention, producing multi-scale outputs:

$$\text{CrossAttn}(Q_m, K_m, V_m) = \text{Softmax}\left(\frac{Q_m K_m^\top}{\sqrt{d_k}}\right) V_m \tag{3}$$

where d_k is the Key vector dimension. This enables each scale to adaptively aggregate complementary discriminative information from other scales, adapting to individual differences in biomarker scale distribution.

2.4 Cross-Scale Mixture-of-Experts

Pronounced inter-subject variability in fNIRS signals leads to heterogeneous scale distribution of discriminative information, limiting cross-subject model generalization. The proposed Cross-Scale Mixture-of-Experts (CS-MoE) module (Fig. 1(e)) achieves adaptive multi-scale feature fusion via a gating mechanism to alleviate this issue. For the m cross-attended scale-specific features, each is routed to an independent expert network composed of a flattening operation, feed-forward network (FFN) and residual connection to preserve scale-specific discriminative information. A centralized gating network takes all expert outputs as input, generates a softmax-normalized gating weight w_m for each corresponding expert, and adaptively weights each expert output with its matched w_m. All weighted expert features are concatenated and fed into the final Softmax layer for classification. This dynamic strategy enables the model to focus on the most discriminative scales for each input, effectively enhancing cross-subject generalization.

3 Experiments and Results

We utilized three fNIRS datasets to evaluate MsDFSTGL: (1) Mental Arithmetic (MA) comprised data from 8 subjects performing MA tasks and rest, totaling 174 trials. (2) Unilateral Finger and Foot Tapping (UFFT) involved 30 subjects performing left/right hand tapping (LHT, RHT) and foot tapping (FT), totaling 2,250 trials. (3) Action Recognition Motor Imagery (AR-MI) was a self-collected dataset from 9 subjects performing six imagery tasks, including left/right arm swing (LA, RA), left/right leg kick (LL,RL), walk (WA), run (RU), totaling 36 trials per subject. HbO2 and HbR were obtained via standardized preprocessing [5]. Continuous recordings were segmented into task trials, and each trial was treated as an independent sample. Thus, MA contains 174 trials from 8 subjects, each represented as $52 \times 2 \times 140$ (channels $\times$ hemoglobin $\times$ time points). The same construction was applied to UFFT ($20 \times 2 \times 256$) and AR-MI (36

$\times$ 2 $\times$ 192). Data were partitioned into training and testing sets 8:2. Empirically, MsDFSTGL was optimized with 4 GCN layers and a 10% adjacency matrix sparsity threshold, as retaining denser connections degraded performance. Multi-scale temporal projections utilized kernel sizes of 2 and 16, with patching lengths set to 2 and 16 in corresponding MSDE models. The model was trained using Adam with learning rate 1×10^{-3}, and early stopping used a patience of 10 epochs. We compared MsDFSTGL with nine methods, including an artificial neural network (ANN) [25], classical deep learning models (CNN [21], LSTM [1], GCN), CapsuleNet variants (CapsuleNet [22], ID-CapsuleNet [14]), and specialized fNIRS transformers (fNIRSNet [24], fNIRS-PreT [25], fNIRS-T [25]). All baseline models were implemented with the same training configuration, data preprocessing pipeline, and dataset split strategy as the proposed method to ensure fair comparison. All models underwent 50 rounds of 5-fold cross-validation. Evaluation metrics included Accuracy (ACC), Precision (PRE), Recall (REC), F1-score (F1), and Kappa coefficient.

3.1 Results

Comparative experiment: As shown in Fig. 2 and Table 1, MsDFSTGL achieved state-of-the-art performance on the UFFT dataset (ACC: 76.85%, F1: 76.72%), outperforming fNIRS-T. On the MA dataset, it yielded a highly competitive ACC of 95.51%, closely matching the top-performing fNIRSNet. Furthermore, MsDFSTGL outperformed all baselines on the challenging AR-MI dataset (ACC: 34.82%, F1: 33.70%). Paired two-sample t-tests with FDR correction confirmed the superior performance of MsDFSTGL over most baseline methods across three datasets ($p < 0.05$). The notably lower classification accuracy on the AR-MI dataset compared with MA and UFFT datasets is attributed to the higher cognitive and motor complexity of motor imagery tasks and the spatially overlapping cortical activation patterns they elicit [17, 12, 3, 8, 24]. These results demonstrate the model's robust decoding capabilities for both cognitive and fine-grained motor tasks.

Ablation experiment: We conducted comprehensive ablation studies on all three datasets (Table 2) to quantify the contribution of each component. Replacing the multi-scale kernels with fixed-2 ones (w/o MS-Convs) resulted in the most severe degradation, with ACC decreasing by 9.80% on MA, 8.67% on UFFT, and 11.37% on AR-MI, indicating the necessity of modeling heterogeneous temporal dynamics. Removing the differential branch (w/o MSDE) further reduced performance, particularly on MA (-7.51%) and UFFT (-3.29%), demonstrating the importance of hemodynamic gradients for transient activation detection. Bypassing the TFGC and global GCN stage (w/o Global TFGC) led to ACC drops of 6.37%, 4.45%, and 6.18% on MA, UFFT, and AR-MI, respectively. Eliminating either the temporal (w/o TFGC-Time) or spectral (w/o TFGC-Freq) branch consistently impaired performance, confirming that dual-domain global initialization provides a robust spatial prior. Finally, replacing MSMHCA and CS-MoE with simple feature concatenation reduced ACC by 4.81%–7.55% , highlighting validating the value of explicit inter-scale dependency modeling the effectiveness

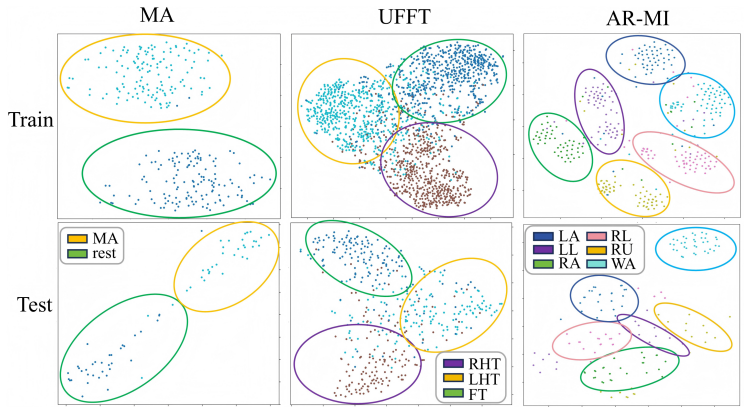

Fig. 2. t-SNE visualization on three datasets.

Table 1. Performance comparison with baselines on MA, UFFT, and AR-MI datasets. * $p < 0.05$, ** $p < 0.01$ compared with MsDFSTGL (paired two-sample t-test with FDR correction).

Dataset	Methods	ACC (%)	PRE (%)	REC (%)	F1 (%)	Kappa (%)
MA	ANN	75.49±4.32**	76.12±4.45**	74.85±4.51**	75.48±4.48**	68.25±5.60**
	GCN	86.68±3.23**	87.15±3.18**	86.22±3.45**	86.68±3.30**	82.40±4.25**
	CNN	87.00±3.32**	86.50±3.50**	86.50±3.50**	87.00±3.37**	82.85±4.30**
	LSTM	88.45±3.41**	88.92±3.35**	87.95±3.58**	88.43±3.46**	84.75±4.45**
	CapsuleNet	60.81±5.54**	61.55±5.80**	59.88±6.10**	60.70±5.95**	48.65±7.20**
	fNIRS-PreT	94.84±2.71	95.12±2.65	94.58±2.90	94.85±2.77	93.10±3.55
	fNIRS-T	95.29±3.17	95.58±3.08	95.02±3.30	95.30±3.19	93.72±3.95
	ID-CapsuleNet	95.01±2.41	95.28±2.35	94.75±2.58	95.01±2.46	93.35±3.15
	fNIRSNet	**95.59±2.44**	**95.85±2.38**	95.35±2.60	**95.60±2.49**	**94.12±3.20**
	MsDFSTGL	95.51±1.46	95.68±1.52	**95.42±1.44**	95.54±1.48	94.01±1.95
UFFT	ANN	47.54±2.07**	48.40±2.12**	47.01±2.12**	47.70±1.98**	30.05±2.96**
	GCN	69.82±2.29**	70.19±2.58**	69.44±2.36**	69.81±2.22**	59.76±2.86**
	CNN	72.89±1.70**	73.38±1.86**	72.13±1.74**	72.75±1.66**	63.85±2.21**
	LSTM	61.94±1.95**	62.35±2.12**	61.25±1.96**	61.79±1.97**	49.25±2.43**
	CapsuleNet	36.92±1.94**	37.70±2.03**	36.91±1.97**	37.30±1.98**	15.89±2.56**
	fNIRS-PreT	70.66±1.45**	69.82±1.51**	71.44±1.59**	70.62±1.42**	60.88±1.94**
	fNIRS-T	75.49±2.07*	75.16±2.36*	76.56±2.35	75.86±2.09	67.32±2.99*
	ID-CapsuleNet	74.03±2.26*	73.24±2.39*	74.19±2.35*	73.71±2.33*	65.37±2.93*
	fNIRSNet	73.16±2.15*	73.33±2.17*	72.93±2.47*	73.13±2.21*	64.21±2.69*
	MsDFSTGL	**76.85±1.72**	**76.78±1.46**	**76.85±1.72**	**76.72±1.73**	**64.53±2.58**
AR-MI	ANN	22.40±3.10**	21.85±3.25**	22.75±3.15**	22.05±3.05**	6.88±3.72**
	GCN	27.50±2.60**	26.95±2.70**	27.80±2.55**	27.15±2.65**	13.00±3.12**
	CNN	29.80±2.50**	29.15±2.68**	30.20±2.45**	29.35±2.52**	15.76±3.00**
	LSTM	28.15±2.75**	27.60±2.82**	28.45±2.90**	27.80±2.78**	13.78±3.30**
	CapsuleNet	19.50±3.20**	19.10±3.35**	19.85±3.25**	19.25±3.18**	3.40±3.84**
	fNIRS-PreT	30.50±2.45*	29.85±2.55*	30.95±2.62*	30.10±2.40*	16.60±2.94*
	fNIRS-T	32.45±2.85*	31.90±2.92*	32.85±2.80*	32.05±2.75*	18.94±3.42*
	ID-CapsuleNet	33.10±2.58*	32.45±2.65*	33.56±2.72*	32.68±2.60*	19.72±3.10*
	fNIRSNet	33.95±2.45	32.88±2.51	34.12±2.60*	33.15±2.48	20.68±2.95*
	MsDFSTGL	**34.82±2.80**	**33.25±2.63**	**35.05±2.79**	**33.70±2.60**	**21.75±3.33**

of adaptive expert fusion in handling inter-subject variability. All core components of the proposed framework contribute significantly to the final decoding performance, verifying the rationality of our module design.

Table 2. Ablation study results on MA, UFFT, and AR-MI datasets.

Methods	MA		UFFT		AR-MI	
	ACC (%)	F1 (%)	ACC (%)	F1 (%)	ACC (%)	F1 (%)
MsDFSTGL	95.51 ± 1.46	95.54 ± 1.48	76.85 ± 1.72	76.72 ± 1.73	34.82 ± 2.80	33.70 ± 2.60
w/o MS-Convs	85.71 ± 6.19	85.41 ± 6.47	68.18 ± 3.72	68.01 ± 3.83	23.45 ± 3.64	22.46 ± 2.62
w/o MSDE	88.00 ± 10.52	87.47 ± 11.56	73.56 ± 2.81	73.42 ± 3.02	26.45 ± 6.39	25.90 ± 6.96
w/o Global TFGC	89.14 ± 6.41	88.99 ± 6.71	72.40 ± 1.06	72.39 ± 1.14	28.64 ± 2.94	27.22 ± 2.37
w/o TFGC-Freq	91.37 ± 2.91	91.37 ± 2.91	73.11 ± 2.39	73.12 ± 2.44	29.09 ± 7.04	27.17 ± 7.15
w/o TFGC-Time	90.29 ± 6.02	90.18 ± 6.23	74.40 ± 1.20	74.30 ± 1.40	30.22 ± 3.38	26.04 ± 2.26
w/o MSMHCA	90.05 ± 5.87	89.92 ± 6.03	72.68 ± 2.15	72.59 ± 2.08	27.86 ± 3.21	26.95 ± 3.10
w/o CS-MoE	88.57 ± 7.23	88.40 ± 7.53	72.04 ± 2.01	72.14 ± 1.98	27.27 ± 5.50	26.59 ± 5.33

Leave-one-subject-out classification experiment: We conducted a Leave-One-Subject-Out (LOSO) cross-validation (Table 3) to evaluate cross-subject generalization, where data from one subject was iteratively held out for testing while the remaining subjects constituted the training set. MsDFSTGL demonstrated robust generalization performance against inter-subject variability with the highest ACC of 94.23% and PRE of 94.68% on the MA dataset, and maintained a competitive ACC of 77.24% on the UFFT dataset. On the challenging AR-MI dataset, MsDFSTGL achieved the highest ACC of 41.55%, significantly surpassing both random chance (16.67%) and Transformer models. This verified the model's robustness against inter-subject variability and highlighted its potential for calibration-free BCI applications.

Table 3. Performance comparison of LOSO cross-validation.

Methods	MA		UFFT		AR-MI	
	ACC (%)	PRE (%)	ACC (%)	PRE (%)	ACC (%)	PRE (%)
ANN	68.84 ± 9.92	70.18 ± 12.05	45.78 ± 8.76	46.88 ± 9.05	24.50 ± 8.50	25.10 ± 9.20
GCN	82.38 ± 10.23	83.15 ± 12.18	72.94 ± 15.62	73.91 ± 16.37	33.50 ± 10.50	32.80 ± 11.20
CNN	82.47 ± 11.64	83.96 ± 13.59	73.47 ± 17.11	75.34 ± 16.07	34.20 ± 11.20	34.85 ± 12.05
LSTM	83.68 ± 12.42	84.88 ± 12.62	66.89 ± 14.29	67.92 ± 14.60	29.80 ± 12.00	28.95 ± 12.80
CapsuleNet	85.42 ± 10.96	84.34 ± 12.93	33.38 ± 4.18	23.80 ± 9.89	21.50 ± 9.00	20.80 ± 10.50
fNIRS-PreT	92.97 ± 6.70	93.55 ± 7.39	76.98 ± 16.35	77.66 ± 16.09	39.10 ± 9.20	38.50 ± 9.80
fNIRS-T	93.84 ± 5.99	96.30 ± 4.22	**78.22 ± 16.69**	**79.36 ± 16.67**	40.25 ± 9.50	40.60 ± 10.10
ID-CapsuleNet	88.72 ± 10.47	88.20 ± 12.41	75.78 ± 16.73	77.52 ± 15.83	38.50 ± 10.10	37.90 ± 11.20
fNIRSNet	87.85 ± 11.30	86.87 ± 13.00	75.72 ± 7.45	76.67 ± 9.49	37.80 ± 9.80	37.20 ± 10.50
MsDFSTGL	**94.23 ± 6.24**	**94.68 ± 1.46**	77.24 ± 18.20	78.99 ± 17.49	**41.55 ± 8.61**	**41.49 ± 10.36**

4 Discussion and Conclusion

The fNIRS-based multi-class neural decoding is severely limited by systemic physiological noise, HRF sluggishness and scale-dependent spatiotemporal het-

erogeneity, hindering the clinical translation of fNIRS-BCI systems, we propose a multi-scale dual-domain spatio-temporal graph learning framework MsDFSTGL. The MsDFSTGL constructs noise-resilient topological priors via the dual-domain TFGC module, decouples transient neural triggers via the parallel MSDE module, models cross-scale dependencies via MSMHCA, and adaptively fuses features via CS-MoE. Comprehensive evaluations on three diverse fNIRS datasets demonstrate that MsDFSTGL achieves excellent multi-class decoding performance, outperforming both classical and advanced deep learning baselines. LOSO performance (94.23% ACC on MA and 77.24% ACC on UFFT) further validates the cross-subject generalizability of MsDFSTGL. Our survey of fNIRS-based motor imagery studies published between 2012 and 2026 shows that multi-class research remains limited, with most reported accuracies below 40% for four-class tasks. On the six-class AR-MI dataset, MsDFSTGL achieved 34.82% ACC and 41.55% ACC under the random split and LOSO settings, respectively, substantially exceeding the 16.67% random baseline and demonstrating promising scalability to more challenging multi-class motor imagery classification. Stable performance across 20-channel 3-class, 36-channel 6-class, and 52-channel 2-class tasks further demonstrates strong scalability. This work provides a robust, calibration-efficient technical paradigm for fNIRS-based neural decoding, laying a critical foundation for the practical clinical application of fNIRS-BCI systems in motor rehabilitation and cognitive neuroscience research.

Acknowledgments. This study was funded by: National Natural Science Foundation of China (12571457, 62506298); General Project of the Key Research and Development Plan of Shaanxi Province (2025SF-YBXM-240).

Disclosure of Interests. The authors have no conflicts to disclose.

References

1. Asgher, U., Khalil, K., Khan, M.J., Ahmad, R., Butt, S.I., Ayaz, Y., Naseer, N., Nazir, S.: Enhanced accuracy for multiclass mental workload detection using long short-term memory for brain–computer interface. Frontiers in neuroscience **14**, 584 (2020)
2. Chen, J., Xia, Y., Zhou, X., Vidal Rosas, E., Thomas, A., Loureiro, R., Cooper, R.J., Carlson, T., Zhao, H.: fnirs-eeg bcis for motor rehabilitation: a review. Bioengineering **10**(12), 1393 (2023)
3. Chu, M., Ma, Y., He, X., Li, X., Ren, J., Zhong, Z., Yu, J., Guo, H.: T&f-dfc fusionnet: Time&frequency-dynamic functional connectivity fusion network for adhd diagnosis in children based on fnirs. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 627–637. Springer (2025)
4. Chu, M., Ma, Y., Tian, R., Liu, S., Yu, J., He, X., Guo, H.: Advances in functional near-infrared spectroscopy: Physical principles and expanding applications in neuroscience. Journal of Biophotonics **18**(11), e202500147 (2025)
5. Chu, M., Ren, J., Yang, L., Ma, Y., Zhong, Z., He, X., Guo, H., Li, X., Yu, J., Wang, Z.: Exploring the neural mechanisms of adhd in children: a multifeature cross-task fnirs analysis. Cerebral Cortex **35**(6) (2025)

6. Edelman, B.J., Zhang, S., Schalk, G., Brunner, P., Müller-Putz, G., Guan, C., He, B.: Non-invasive brain-computer interfaces: state of the art and trends. IEEE reviews in biomedical engineering **18**, 26–49 (2024)
7. Fan, W., Fei, J., Guo, D., Yi, K., Song, X., Xiang, H., Ye, H., Li, M.: Towards multi-resolution spatiotemporal graph learning for medical time series classification. In: Proceedings of the ACM on Web Conference 2025. pp. 5054–5064 (2025)
8. Guan, S., Li, Y., Luo, Y., Niu, H., Gao, Y., Yang, D., Li, R.: Disentangling the impact of motion artifact correction algorithms on functional near-infrared spectroscopy–based brain network analysis. Neurophotonics **11**(4), 045006–045006 (2024)
9. Ismail Fawaz, H., Forestier, G., Weber, J., Idoumghar, L., Muller, P.A.: Deep learning for time series classification: a review. Data mining and knowledge discovery **33**(4), 917–963 (2019)
10. Jin, M., Koh, H.Y., Wen, Q., Zambon, D., Alippi, C., Webb, G.I., King, I., Pan, S.: A survey on graph neural networks for time series: Forecasting, classification, imputation, and anomaly detection. arXiv preprint arXiv:2307.03759 (2023)
11. Khanam, F.: Four-class motor imagery eeg signal classification using pca, wavelet and two-stage neural network. International Journal of Advanced Computer Science and Applications (2019)
12. Li, L., Zhang, M., Chen, Y., Wang, K., Zhou, G., Huang, Q.: Tagl: temporal-guided adaptive graph learning network for coordinated movement classification. IEEE Transactions on Industrial Informatics **20**(11), 12554–12564 (2024)
13. Li, Y., Li, S., Yuan, Z., Zhao, S., Wan, F., Xu, T., Zhang, H., Wang, H.: Ivcan: An improved visual curve attention network for fnirs-based motor imagery/execution classification. Biomedical Signal Processing and Control **104**, 107679 (2025)
14. Li, Y., Xu, T., Li, J., Wan, F., Wang, H.: Improved dilation capsulenet for motor imagery and mental arithmetic classification based on fnirs. Brain-Apparatus Communication: A Journal of Bacomics **3**(1), 2335886 (2024)
15. Liu, H., Yang, D., Liu, X., Chen, X., Liang, Z., Wang, H., Cui, Y., Gu, J.: Todynet: temporal dynamic graph neural network for multivariate time series classification. Information Sciences **677**, 120914 (2024)
16. Mao, Y., Zhou, H., Chen, L., Qi, R., Sun, Z., Rong, Y., He, X., Chen, M., Mumtaz, S., Frascolla, V., et al.: A survey on spatio-temporal prediction: From transformers to foundation models. ACM Computing Surveys **58**(4), 1–36 (2025)
17. Milu, M.M.H., Rahman, M.A., Rashid, M.A., Kuwana, A., Kobayashi, H.: Improvement of classification accuracy of four-class voluntary-imagery fnirs signals using convolutional neural networks. Engineering, Technology & Applied Science Research **13**(2), 10425–10431 (2023)
18. Sarafraz, G., Behnamnia, A., Hosseinzadeh, M., Balapour, A., Meghrazi, A., Rabiee, H.R.: Domain adaptation and generalization of functional medical data: A systematic survey of brain data. ACM Computing Surveys **56**(10), 1–39 (2024)
19. Shamsi, F., Najafizadeh, L.: Multi-class classification of motor execution tasks using fnirs. In: 2019 IEEE Signal Processing in Medicine and Biology Symposium (SPMB). pp. 1–5. IEEE (2019)
20. Taylor, A.J., Kim, J.H., Ress, D.: Temporal stability of the hemodynamic response function across the majority of human cerebral cortex. Human brain mapping **43**(16), 4924–4942 (2022)
21. Trakoolwilaiwan, T., Behboodi, B., Lee, J., Kim, K., Choi, J.W.: Convolutional neural network for high-accuracy functional near-infrared spectroscopy in a brain–computer interface: three-class classification of rest, right-, and left-hand motor execution. Neurophotonics **5**(1), 011008–011008 (2018)

22. Van Quang, N., Chun, J., Tokuyama, T.: Capsulenet for micro-expression recognition. In: 2019 14th IEEE international conference on automatic face & gesture recognition (FG 2019). pp. 1–7. IEEE (2019)
23. Wang, Y., Huang, N., Li, T., Yan, Y., Zhang, X.: Medformer: A multi-granularity patching transformer for medical time-series classification. arXiv preprint arXiv:2405.19363 (2024)
24. Wang, Z., Fang, J., Zhang, J.: Rethinking delayed hemodynamic responses for fnirs classification. IEEE Transactions on Neural Systems and Rehabilitation Engineering **31**, 4528–4538 (2023)
25. Wang, Z., Zhang, J., Zhang, X., Chen, P., Wang, B.: Transformer model for functional near-infrared spectroscopy classification. IEEE Journal of Biomedical and Health Informatics **26**(6), 2559–2569 (2022)
26. Zafar, A., Dad Kallu, K., Atif Yaqub, M., Ali, M.U., Hyuk Byun, J., Yoon, M., Su Kim, K.: A hybrid gcn and filter-based framework for channel and feature selection: An fnirs-bci study. International Journal of Intelligent Systems **2023**(1), 8812844 (2023)