CAN DIFFERENTIABLE STRUCTURE LEARNING ENABLE LOW-CHANNEL BRAINPRINT AUTHENTICATION ON RESOURCE-CONSTRAINED IOT DEVICES?

Anonymous authors
Paper under double-blind review

ABSTRACT

High-precision Brain-Computer Interface (BCI) systems typically require dense multi-channel EEG acquisition, which conflicts with the low-power and low-channel constraints of IoT edge devices. Existing channel selection methods also lack end-to-end optimization and structural interpretability. To address these issues, we propose the Dynamic Channel Learning Transformer (DCLT), which formulates EEG channel selection as Differentiable Structure Learning for IoT-based brainprint recognition. DCLT introduces an STE-driven Top-K discrete gating mechanism to jointly optimize physical channel constraints and gradient-based learning, followed by Transformer-based global sparse representation modeling. Furthermore, Dynamic-Static Consistency Analysis is proposed to evaluate structural stability and deployment feasibility. Experiments on three EEG datasets show that DCLT maintains over 90% recognition accuracy with only 25% channels retained, while real-world edge evaluation on 23 subjects demonstrates significant reductions in computational cost and inference latency. The anonymous code repository is available at `https://anonymous.4open.science/r/DCLT-DD3C/`

1 INTRODUCTION

With the rapid development of IoT, EEG-based neural interfaces have been widely applied in wearable healthcare, human–computer interaction, and biometric authentication (Laport et al., 2020; Abdellatif et al., 2018). Brainprint recognition is particularly attractive due to its liveness, anti-spoofing capability, and physiological uniqueness (Jin et al., 2020; Changrong et al., 2025).

However, EEG-based recognition for IoT deployment remains challenging due to high dimensionality, redundancy, low SNR, and inter-subject variability. Existing methods often rely on dense 32–64 channel systems (Bozkurt, 2025), leading to increased hardware and computational costs (Alqudah & Moussavi, 2025). Thus, accurate recognition under strict channel constraints is critical for wearable EEG systems (Alotaiby et al., 2015; Yong et al., 2008; Birjandtalab et al., 2017; Arvaneh et al., 2011; Artoni et al., 2018).

Traditional approaches based on handcrafted features and spatial filtering, such as LRMD (Kong et al., 2018), Probabilistic CSP (Wu et al., 2015), and Sparse CSP (Fu et al., 2020), as well as independent channel reduction methods (PCA (Pearson, 1901), ReliefF (Aggarwal et al., 2023)), lack joint optimization. Deep models including EEGNet (Lawhern et al., 2018), TSception (Ding et al., 2023), BrainprintNet (Tu et al., 2026), and disentangled learning (Zheng et al., 2025) improve representation learning but remain limited under sparse channels. Transformer-based models (Dosovitskiy et al., 2020; Song et al., 2022; Wang et al., 2024; Arjun et al., 2021; Zhang et al., 2025) enhance global modeling yet still rely on dense inputs. Recent channel optimization methods based on evolutionary algorithms (Moctezuma & Molinas, 2020; Alyasseri et al., 2022) and attention mechanisms (Bai et al., 2025) suffer from high computation or inconsistency between continuous optimization and discrete deployment (Touvron et al., 2019; Nakanishi et al., 2022). Moreover, channel stability is rarely considered (Yazdani et al., 2021), leaving the gap between physical constraints and gradient-based optimization unresolved (Li et al., 2026a; Dai et al., 2020; Baig et al., 2020).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

To address these limitations, we propose the Dynamic Channel Learning Transformer (DCLT), an IoT-oriented EEG decoding framework that unifies channel selection and representation learning. Unlike conventional channel pruning methods (Faul & Marnane, 2012; Janani et al., 2018; Hejrati et al., 2017), DCLT formulates channel selection as a differentiable structure learning problem (Ming et al., 2020; Li et al., 2026b). An STE-based Top-K gating mechanism (Yin et al., 2019) enables end-to-end discrete channel optimization, while a Vision Transformer (Vaswani et al., 2017; Yang & Modesitt, 2023) captures global dependencies under ultra-low channel settings. Dynamic-Static Consistency Analysis further evaluates channel stability for wearable EEG deployment.

Extensive experiments on multiple EEG benchmarks, including motor imagery, emotion recognition, and brainprint authentication, demonstrate that DCLT achieves robust recognition performance under severe channel compression, validating its effectiveness for resource-constrained IoT neural interfaces.

The main contributions are summarized as follows:

1. To the best of our knowledge, we are the first to introduce differentiable channel selection into EEG brainprint recognition by formulating channel selection as a structure learning problem, enabling joint optimization of electrode configuration and EEG representation learning.
2. We propose an STE-driven Top-K adaptive gating mechanism to bridge discrete hardware constraints and continuous neural optimization.
3. We develop a Transformer-based global EEG modeling framework for robust brainprint recognition under ultra-low channel conditions.

2 METHODOLOGY

This paper proposes the Dynamic Channel Learning Transformer (DCLT), an end-to-end Differentiable Structure Learning framework for high-accuracy EEG representation learning under strict low-channel constraints. DCLT integrates three key components: the Adaptive Channel Selection Module (ACSM), STE-TopK discrete gating mechanism, and Transformer-based global representation backbone (EEG-ViT Backbone), as shown in Fig. 1. The proposed framework unifies adaptive channel optimization and global EEG representation learning, and is extensively validated on three diverse datasets, including PhysioNet MI, DEAP, and AMIGOS.

2.1 Adaptive Graph-Structured Channel Selection Mechanism (ACSM)

2.1.1 Channel Importance and Graph-aware Modeling

Given EEG input $X \in \mathbb{R}^{C\times T}$, DCLT learns channel importance through temporal aggregation and an MLP-based scoring network:

$$s = \sigma(W_2\delta(W_1 z)), \quad z_c = \frac{1}{T}\sum_{t=1}^{T} X_{(c,t)} \tag{1}$$

where s denotes continuous channel importance weights. The learned weights are further formulated as a functional graph:

$$G = (V, E, s) \tag{2}$$

enabling data-driven channel optimization without predefined topology priors. A discrete selection mechanism is then applied to satisfy low-channel IoT deployment constraints while preserving differentiable learning.

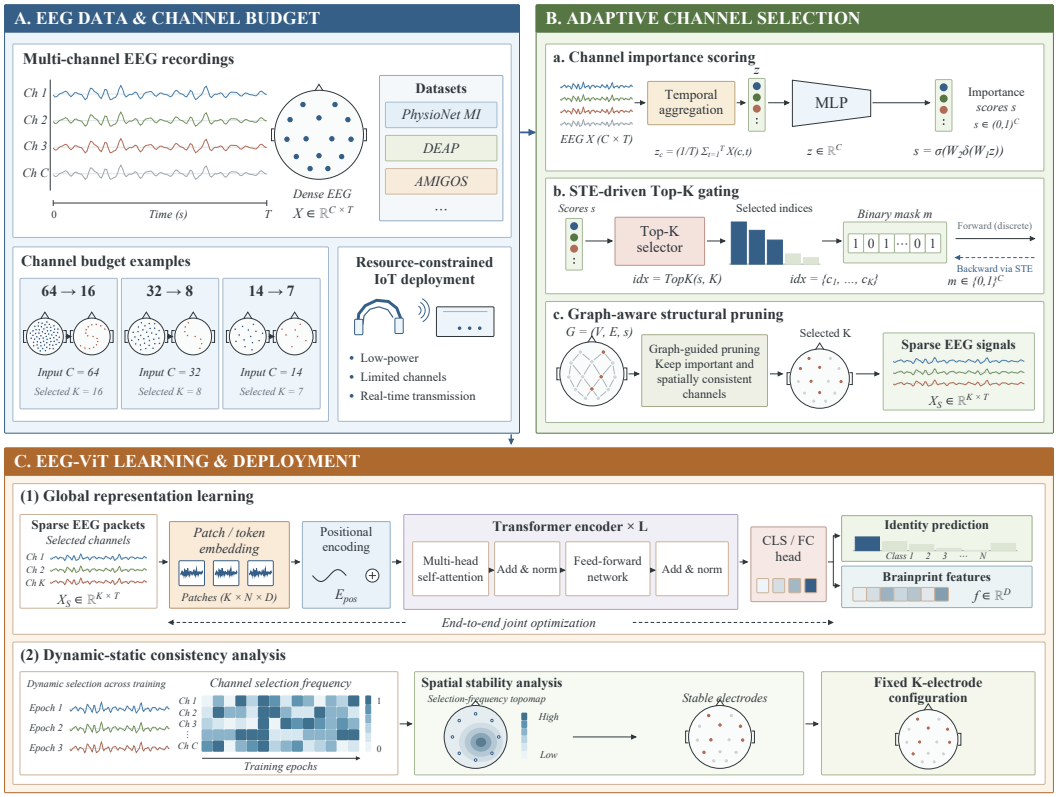


Figure 1: Overall framework of the proposed Dynamic Channel Learning Transformer (DCLT) for differentiable EEG channel selection and brainprint recognition. (A) Dense multi-channel EEG signals are acquired under predefined channel budgets for resource-constrained IoT deployment. (B) The adaptive channel selection module learns channel importance scores, performs STE-driven Top-K gating, and applies graph-aware structural pruning to obtain sparse EEG signals. (C) The selected EEG channels are encoded by EEG-ViT for global representation learning and identity prediction through end-to-end joint optimization. Dynamic–static consistency analysis further aggregates sample-adaptive channel selections into a stable fixed K-electrode configuration for efficient IoT-oriented deployment.

2.2 Top-K Discrete Channel Selection and STE Optimization Mechanism

As shown in Appendix Fig. 3, DCLT converts continuous channel importance scores s into discrete channel selections through an STE-based Top-K gating mechanism to satisfy ultra-low-channel deployment constraints. The selected indices are obtained as:

$$idx = \arg \mathrm{TopK}(s, K) \tag{3}$$

and the binary mask is defined as:

$$M_{hard}^{(c)} = \begin{cases} 1, & c \in idx \\ 0, & otherwise \end{cases} \tag{4}$$

Since Top-K selection is non-differentiable, STE is introduced:

$$M_{STE} = M_{hard} + (s - \mathrm{sg}[s]) \tag{5}$$

where $\mathrm{sg}(\cdot)$ denotes stop-gradient. The selected EEG representation is:

$$\tilde{X} = X \odot M_{STE} \tag{6}$$

During forward propagation, M_{STE} behaves as a hard mask to preserve exactly K channels, while gradients are passed through s during backpropagation, enabling end-to-end optimization.

To avoid premature pruning, DCLT further adopts progressive structural pruning, gradually reducing the channel budget from full-channel soft attention to the target K_{tgt}:

$$K_{cur}^{(e)} : C \to 3K_{tgt} \to 2K_{tgt} \to K_{tgt} \tag{7}$$

This strategy enables smooth channel compression while preserving discriminative EEG representations.

2.3 EEG-ViT Feature Modeling Network

DCLT employs a Vision Transformer (ViT) as the global feature extractor by converting selected EEG signals into image-like representations:

$$X' \in \mathbb{R}^{C \times T} \to \text{Image-like Tensor} \tag{8}$$

Dataset-specific patch strategies are adopted to preserve temporal and spatial information under different channel densities. The ViT encoder then captures long-range dependencies through self-attention:

$$Z^{(l)} = \mathrm{MSA}(\mathrm{LN}(Z^{(l-1)})) + Z^{(l-1)} \tag{9}$$

Unlike CNNs relying on local spatial priors, ViT enables adaptive global dependency modeling, providing robust representation learning under sparse EEG configurations.

2.4 Multi-objective Optimization Function

The proposed DCLT is optimized by jointly considering recognition accuracy and channel-selection stability:

$$\mathcal{L} = \mathcal{L}_{\mathrm{cls}} + \lambda \mathcal{L}_{\mathrm{reg}}, \tag{10}$$

where $\mathcal{L}_{\mathrm{cls}}$ denotes the cross-entropy classification loss and λ is the regularization coefficient, set to 0.01.

To encourage the continuous channel-importance scores to become well polarized before discrete Top-K selection, we define

$$\mathcal{L}_{\mathrm{reg}} = \frac{1}{C} \sum_{c=1}^{C} s_c(1 - s_c), \tag{11}$$

where $s_c \in (0, 1)$ denotes the continuous importance score of channel c. This regularization is minimized when the importance scores approach binary states, thereby reducing ambiguous channel rankings and promoting stable discrete channel selection. The hard Top-K operator subsequently enforces the exact channel budget.

2.5 Dynamic-Static Consistency Evaluation

Traditional EEG channel selection methods typically report only classification accuracy, failing to answer a crucial question:

Has the model genuinely learned task-relevant channels, or merely overfitted dataset-specific biases?

To address this limitation, this work proposes a novel evaluation paradigm termed **Dynamic-Static Consistency Analysis**.

2.5.1 Dynamic Channel Selection

During standard inference, the model dynamically generates channel masks for each input sample:

$$X' = X \odot \tilde{m}(X) \tag{12}$$

where $\tilde{m}$ is generated by the selector network, resulting in sample-specific channel combinations.

Dynamic channel selection provides strong adaptability and effectively captures individual differences, typically achieving the highest predictive performance. The corresponding performance is denoted as:

$$Acc_{dynamic}$$

2.5.2 Static Top-K Channel Selection

To obtain a deployable fixed electrode configuration, DCLT derives a global static channel subset based on training-set selection frequency:

$$F(c) = \sum_{i=1}^{N_{train}} \mathbb{I}(c \in idx_i), \tag{13}$$

where idx_i denotes the Top-K channels selected for each sample. The final subset is obtained by $S_{static} = \arg \mathrm{TopK}(F, K)$. During inference, the selected channels are converted into a fixed spatial mask, and dynamic selection is removed, simulating hardware-constrained EEG systems. The resulting performance is denoted as Acc_{static}.

3 Experiments

3.1 Datasets

We evaluate DCLT on three public EEG datasets, including PhysioNet MI (Schalk et al., 2009), DEAP (Koelstra et al., 2011), and AMIGOS (Miranda-Correa et al., 2018), covering diverse channel densities (64/32/14 channels) and identity scales (32–109 subjects). Each subject is treated as an identity class for brainprint recognition, with evaluations conducted under channel compression settings of 64-to-16, 32-to-8, and 14-to-7. Furthermore, real-world EEG acquisition experiments are performed on 23 participants using the NeuroScan system and Curry 9 software, where 60-channel EEG signals are recorded at 400 Hz. The collected data are used to evaluate DCLT's deployment feasibility in terms of compact electrode selection, computational cost, and inference latency.

3.1.1 Evaluation Protocol

Following the evaluation protocol in prior EEG biometric studies (Zheng et al., 2025), continuous EEG signals were segmented into non-overlapping 1-second windows, with 90% of the samples used for training and the remaining 10% for testing.

3.2 Brainprint Recognition under Ultra-low Channel Constraints

3.2.1 High-density Array Compression: PhysioNet Evaluation (64 $\rightarrow$ 16 Channels)

Although high-density EEG arrays provide rich spatial information, they also introduce redundancy and inter-channel correlations. We evaluate PhysioNet MI under an extreme channel budget of $K = 16$, retaining only 25% electrodes. As shown in Table 1, the STE-driven Top-K gating mechanism enables DCLT to preserve discriminative identity information and achieve 91.17% accuracy under severe channel compression. The selected channels mainly concentrate in sensorimotor and parieto-occipital regions, indicating that DCLT identifies meaningful functional areas rather than redundant patterns.

Table 1: Brainprint recognition performance comparison on the PhysioNet MI dataset under 64-channel and 16-channel settings (109-class classification).

Methods	Channel Selection	64-Channel Setting (Accuracy %)				16-Channel Setting (Accuracy %)			
		MM-H	MM-HF	MI-H	MI-HF	MM-H	MM-HF	MI-H	MI-HF
LRMD (Kong et al., 2018)	ML	92.25±0.62	91.02±0.71	91.18±0.68	91.95±0.64	80.30±1.02	76.24±1.15	79.46±1.08	77.12±1.12
CSP+rLDA (Wu et al., 2015; Fu et al., 2020)	ML	81.42±0.83	81.15±0.79	81.39±0.81	81.08±0.85	76.84±1.10	75.42±1.21	76.75±1.14	76.26±1.18
FBCNet (Mane et al., 2020)	CNN	98.13±0.24	97.86±0.28	98.05±0.22	97.72±0.31	87.47±0.74	86.54±0.81	88.80±0.69	87.08±0.77
EEGNet (Lawhern et al., 2018)	CNN	93.20±0.55	91.94±0.63	92.31±0.58	91.85±0.61	80.21±1.01	78.66±1.05	81.79±0.96	78.95±1.08
EEGViT (Dosovitskiy et al., 2020)	Transformer (ViT)	96.45±0.38	96.10±0.41	96.54±0.36	96.01±0.43	86.51±0.82	83.67±0.91	85.96±0.85	84.89±0.88
EEGConformer (Song et al., 2022)	CNN+Transformer	93.62±0.52	92.28±0.60	92.70±0.57	92.16±0.63	80.28±1.03	78.42±1.10	78.85±1.06	77.94±1.12
EEGPT (Wang et al., 2024)	Transformer	94.58±0.47	94.05±0.51	95.61±0.43	94.98±0.49	78.26±1.14	83.90±0.92	83.29±0.96	83.92±0.90
TSception (Ding et al., 2023)	CNN	96.36±0.39	96.01±0.42	95.42±0.45	96.95±0.35	86.64±0.83	85.15±0.88	87.58±0.79	85.40±0.86
ArjunViT (Arjun et al., 2021)	Transformer (ViT)	95.51±0.44	95.40±0.46	94.63±0.52	95.18±0.47	84.99±0.91	85.86±0.87	85.17±0.90	84.52±0.93
DMAE-EEG (Zhang et al., 2025)	Transformer	97.74±0.30	97.45±0.33	97.86±0.28	97.32±0.35	87.40±0.76	86.85±0.80	88.34±0.72	86.80±0.81
BrainprintNet (Tu et al., 2026)	CNN	98.68±0.21	98.30±0.24	98.29±0.23	98.54±0.22	88.12±0.70	88.48±0.68	88.61±0.66	88.55±0.67
AD-CCL (Zheng et al., 2025)	CNN+Attention	98.81±0.20	98.80±0.19	98.95±0.18	98.39±0.23	88.24±0.69	88.62±0.65	88.85±0.63	88.81±0.64
DCLT	**Transformer (ViT)**	**99.01±0.16**	**98.85±0.18**	**99.34±0.12**	**98.62±0.17**	**91.17±0.52**	**90.03±0.56**	**90.28±0.54**	**90.09±0.55**

To validate the effectiveness of the learned channel subset, the DCLT-selected golden 16-channel configuration is fixed for all baselines. As detailed in Appendix Table 2, Monte Carlo sampling with 10 random subsets is used to reduce selection bias. DCLT consistently outperforms CNN- and Transformer-based baselines as well as random channel selections, demonstrating that jointly learning channel selection and global dependency modeling enables robust EEG brainprint recognition with limited electrodes.

Table 2: Performance comparison between random 16-channel selection and DCLT-selected golden 16-channel subsets on the PhysioNet MI dataset (109-class classification).

Methods	Channel Selection	16-Channel Random (Accuracy %)				Δ (Opt − Rand) (Accuracy %)			
		MM-H	MM-HF	MI-H	MI-HF	MM-H	MM-HF	MI-H	MI-HF
LRMD (Kong et al., 2018)	ML	61.75±1.12	58.90±1.25	63.46±1.08	59.20±1.21	+18.55±0.72	+17.34±0.75	+16.00±0.69	+17.92±0.73
CSP+rLDA (Wu et al., 2015; Fu et al., 2020)	ML	59.31±1.18	58.31±1.27	58.12±1.20	59.73±1.16	+17.53±0.78	+17.11±0.81	+18.63±0.76	+16.53±0.80
FBCNet (Mane et al., 2020)	CNN	75.20±0.86	74.42±0.91	75.04±0.83	74.94±0.88	+12.27±0.55	+12.12±0.58	+13.76±0.52	+12.14±0.57
EEGNet (Lawhern et al., 2018)	CNN	65.66±1.04	64.42±1.10	65.36±1.02	65.37±1.06	+14.55±0.66	+14.24±0.70	+16.43±0.63	+13.58±0.68
EEGViT (Dosovitskiy et al., 2020)	Transformer (ViT)	72.61±0.92	70.05±0.98	72.86±0.89	70.58±0.95	+13.90±0.61	+13.62±0.64	+13.10±0.60	+14.31±0.62
EEGConformer (Song et al., 2022)	CNN+Transformer	62.22±1.09	61.28±1.15	63.58±1.05	60.73±1.12	+18.06±0.74	+17.14±0.77	+15.27±0.71	+17.21±0.76
EEGPT (Wang et al., 2024)	Transformer	72.91±0.90	71.57±0.96	72.40±0.93	71.03±0.94	+10.35±0.59	+12.33±0.62	+10.89±0.58	+12.89±0.60
TSception (Ding et al., 2023)	CNN	74.18±0.88	73.09±0.92	74.22±0.86	73.34±0.90	+12.46±0.57	+12.06±0.60	+13.36±0.55	+12.06±0.58
ArjunViT (Arjun et al., 2021)	Transformer (ViT)	72.13±0.94	72.90±0.91	72.93±0.89	73.62±0.87	+12.86±0.63	+12.96±0.61	+12.24±0.59	+10.90±0.60
DMAE-EEG (Zhang et al., 2025)	Transformer	75.61±0.84	74.20±0.89	74.37±0.86	74.88±0.83	+11.79±0.54	+12.65±0.57	+13.97±0.53	+11.92±0.55
BrainprintNet (Tu et al., 2026)	CNN	77.24±0.81	76.99±0.84	75.07±0.86	76.82±0.82	+10.88±0.51	+11.49±0.54	+13.54±0.52	+11.73±0.53
AD-CCL (Zheng et al., 2025)	CNN+Attention	77.52±0.80	76.33±0.83	75.62±0.85	77.04±0.81	+10.72±0.50	+12.29±0.53	+13.23±0.51	+11.77±0.52
DCLT	**Transformer (ViT)**	**77.09±0.62**	**77.37±0.60**	**78.94±0.58**	**77.55±0.61**	**+14.08±0.42**	**+12.66±0.40**	**+11.34±0.39**	**+12.54±0.41**

Table 3: Comparison of brainprint recognition performance on the DEAP dataset under 32-channel and 8-channel settings (32 classes), and on the AMIGOS dataset under 14-channel and 7-channel settings.

Methods	Backbone	DEAP 32-Channel Setting	DEAP 8-Channel Setting (Accuracy %)		AMIGOS 14-Channel Setting	AMIGOS 7-Channel Setting (Accuracy %)	
			8-Channel Random (%)	8-Channel Optimal (%)		7-Channel Random (%)	7-Channel Optimal (%)
LRMD (Kong et al., 2018)	ML	91.29±0.42	69.87±1.12	79.45±0.86	78.01±0.48	69.37±1.15	74.98±0.91
CSP+rLDA (Wu et al., 2015; Fu et al., 2020)	ML	91.84±0.39	68.23±1.18	77.10±0.93	83.87±0.41	72.01±1.11	79.48±0.88
FBCNet (Mane et al., 2020)	CNN	98.37±0.21	81.66±0.84	89.66±0.63	87.29±0.35	77.63±0.92	83.72±0.70
EEGNet (Lawhern et al., 2018)	CNN	97.95±0.26	78.56±0.97	87.37±0.71	86.37±0.38	76.59±0.95	82.58±0.74
EEGViT (Dosovitskiy et al., 2020)	Transformer (ViT)	97.73±0.28	80.10±0.90	89.21±0.68	86.01±0.40	77.20±0.91	83.05±0.72
EEGConformer (Song et al., 2022)	CNN+Transformer	95.60±0.45	70.52±1.09	79.47±0.85	78.63±0.53	69.32±1.14	75.20±0.90
EEGPT (Wang et al., 2024)	Transformer	97.16±0.31	79.33±0.92	87.71±0.73	85.59±0.42	77.75±0.90	82.57±0.75
TSception (Ding et al., 2023)	CNN	98.16±0.24	83.68±0.80	89.69±0.62	89.20±0.34	78.57±0.88	85.38±0.67
ArjunViT (Arjun et al., 2021)	Transformer (ViT)	98.06±0.25	81.80±0.86	88.02±0.70	88.32±0.36	78.21±0.90	85.02±0.69
DMAE-EEG (Zhang et al., 2025)	Transformer	98.87±0.18	82.93±0.82	90.79±0.59	89.75±0.30	78.88±0.86	85.37±0.65
BrainprintNet (Tu et al., 2026)	CNN	99.47±0.14	83.11±0.78	93.64±0.51	91.25±0.27	82.48±0.79	87.89±0.55
AD-CCL (Zheng et al., 2025)	CNN + Attention	99.34±0.15	82.79±0.81	94.87±0.48	91.34±0.28	80.90±0.82	87.76±0.57
DCLT	**Transformer (ViT)**	**99.82±0.10**	**84.33±0.63**	**95.21±0.39**	**92.59±0.22**	**81.58±0.58**	**88.05±0.42**

3.3 Standard-Density Array Compression: Evaluation on the DEAP Dataset ($32 \rightarrow 8$ Channels)

A 32-channel EEG configuration is widely used in clinical studies, while 8 channels represent a practical wearable EEG setting. We evaluate DCLT on DEAP under $K = 8$ to assess IoT deployment capability. As shown in Table 3, DCLT achieves 95.21% recognition accuracy with only eight

selected channels, preserving identity information under severe compression. Using the same selected channels, DCLT outperforms CNN- and Transformer-based baselines by jointly optimizing channel selection and global dependency learning. The selected electrodes ([16, 21, 22, 19, 20, 23, 25, 27], corresponding to Fp2, FC6, FC2, F4, F8, Cz, T8, and CP2) exhibit neurophysiological relevance, demonstrating its ability to discover compact and meaningful configurations for low-channel EEG recognition.

3.4 Extreme Sparse Array Compression: Evaluation on the AMIGOS Dataset ($14 \rightarrow 7$ Channels)

As shown in Table 3, AMIGOS represents an extreme wearable EEG setting with only 14 channels. Under $K = 7$, DCLT achieves 88.05% accuracy while preserving discriminative brainprint features. The selected channels ([4,3,1,6,5,0,2], corresponding to T7, FC5, F7, O1, P7, AF3, and F3) span multiple functional regions. Compared with CNN-based methods using the same channels, DCLT benefits from dynamic selection and global modeling, demonstrating effective discovery of compact electrode configurations for ultra-low-channel EEG recognition.

3.5 Comparative Analysis with State-of-the-Art Channel Selection Methods

To decouple the contribution of the dynamic channel selection mechanism from the EEG-ViT backbone, we conduct a controlled comparison to evaluate the effectiveness of the STE-driven differentiable structure learning paradigm. Under identical backbones and channel budgets (PhysioNet MI: $K = 16$, DEAP: $K = 8$, AMIGOS: $K = 7$), DCLT is compared with representative channel selection strategies, including full-channel, random sampling, statistical reduction (PCA (Pearson, 1901), ReliefF (Aggarwal et al., 2023)), evolutionary optimization (BGWO (Alyasseri et al., 2022), NSGA-III (Moctezuma & Molinas, 2020)), and attention-based truncation (GCMTE-Net (Bai et al., 2025)). The results are summarized in Table 4.

Table 4: Comparison of channel selection methods under different channel budgets.

Channel Selection Methods	Detection Model	16-Channel (Accuracy %)				8-Ch / 7-Ch
		MM-H	MM-HF	MI-H	MI-HF	DEAP / AMIGOS
Random channel training	Transformer (ViT)	77.09±0.62	77.37±0.60	78.94±0.58	77.55±0.61	69.87±1.05 / 69.37±1.10
PCA (Pearson, 1901)	Transformer (ViT)	83.30±0.74	82.64±0.79	83.36±0.72	81.79±0.81	75.07±0.91 / 73.39±0.98
ReliefF (Aggarwal et al., 2023)	Transformer (ViT)	84.59±0.70	82.15±0.82	83.07±0.76	82.89±0.78	76.77±0.88 / 75.11±0.94
BGWO (Alyasseri et al., 2022)	SVM	75.33±0.86	75.45±0.83	76.79±0.81	78.43±0.79	70.70±1.02 / 68.40±1.15
BGWO (Alyasseri et al., 2022)	Transformer (ViT)	85.69±0.66	84.47±0.71	84.06±0.73	82.82±0.76	79.58±0.82 / 77.08±0.89
NSGA-III (Moctezuma & Molinas, 2020)	Transformer (ViT)	86.65±0.63	84.31±0.70	84.19±0.74	83.09±0.72	79.13±0.85 / 79.65±0.83
NSGA-III (Moctezuma & Molinas, 2020)	SVM	78.34±0.81	77.61±0.84	78.15±0.80	77.26±0.86	72.62±0.96 / 71.68±1.01
GCMTE-Net (Bai et al., 2025)	Transformer	87.05±0.59	86.52±0.62	88.50±0.57	85.78±0.65	83.99±0.76 / 83.65±0.79
GCMTE-Net (Bai et al., 2025)	Transformer (ViT)	88.68±0.55	85.25±0.63	86.91±0.60	86.43±0.61	85.53±0.72 / 84.37±0.75
DCLT	**Transformer (ViT)**	**91.17±0.52**	**90.03±0.56**	**90.28±0.54**	**90.09±0.55**	**95.21±0.39 / 88.05±0.42**

3.5.1 Overcoming Feature Fragmentation in Unsupervised Dimensionality Reduction

As shown in Table 4, PCA and ReliefF achieve limited performance even with a strong ViT backbone, since channel selection is decoupled from recognition optimization. This two-stage strategy ignores task-relevant neurophysiological structures, while DCLT jointly optimizes channel selection and representation learning through differentiable structure learning.

Table 5: Ablation study on core components of the proposed DCLT framework

Model Variant	**MM-H**	**MM-HF**	**MI-H**	**MI-HF**	**DEAP**	**AMIGOS**
CNN Backbone	89.88 ± 0.82	87.21 ± 0.91	88.34 ± 0.76	86.77 ± 0.95	92.84 ± 0.54	86.96 ± 0.88
Only ViT	86.51 ± 1.05	83.67 ± 1.12	85.96 ± 0.97	84.89 ± 1.08	89.21 ± 0.83	83.05 ± 1.15
w/o Regularization	87.95 ± 0.93	86.61 ± 0.86	87.38 ± 0.89	86.77 ± 0.92	91.53 ± 0.67	85.24 ± 0.96
w/o Top-K Constraint	99.01 ± 0.21	98.85 ± 0.24	99.34 ± 0.18	98.62 ± 0.26	99.82 ± 0.10	92.59 ± 0.51
w/o STE (Soft Mask)	88.69 ± 0.87	86.47 ± 0.94	87.17 ± 0.83	86.89 ± 0.90	92.80 ± 0.58	85.72 ± 0.91
DCLT (Ours)	**91.17 ± 0.63**	**90.03 ± 0.71**	**90.28 ± 0.58**	**90.09 ± 0.66**	**95.21 ± 0.42**	**88.05 ± 0.73**

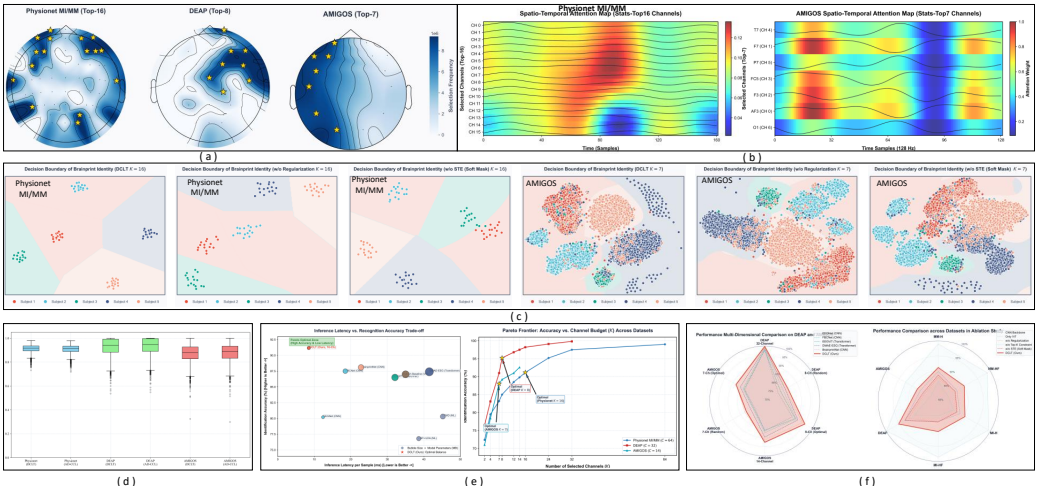


Figure 2: **Comprehensive Evaluation of DCLT.** (a) Selected core channels on scalp topoplots; (b) Spatiotemporal attention dynamics; (c) t-SNE visualization of decision boundaries; (d) Metric distributions under missing electrodes; (e) Accuracy–latency trade-offs and optimal channel budgets; (f) Overall performance comparison across evaluation metrics.

3.5.2 End-to-End Optimization Beyond Heuristic Search

Evolutionary methods such as BGWO and NSGA-III improve channel selection but rely on costly heuristic search and separate channel sampling from model optimization. Soft-attention approaches (GCMTE-Net) further suffer from train–test inconsistency between continuous weighting and discrete deployment. In contrast, DCLT adopts STE-driven differentiable optimization, enabling unified topology learning and robust EEG recognition under sparse channel conditions.

3.6 Dynamic-Static Consistency Analysis and Structural Stability Verification

To evaluate channel stability, we compare dynamic Top-K selection with a fixed training-derived subset. The consistency gap is defined as

$$\Delta = Acc_{\text{dynamic}} - Acc_{\text{static}} \tag{14}$$

where a smaller Δ indicates better stability. As shown in Appendix Fig. 4, DCLT maintains a negligible gap under high-channel settings, with only a modest increase under stronger compression ($\Delta = 0.93\%, 2.32\%, 3.15\%, 5.30\%$ for $K = 32, 16, 8, 4$), demonstrating robust channel selection even under severe sparsity.

3.7 Interpretability of Channel Selection and Neurophysiological Mapping

Additional channel-selection frequency statistics are provided in Appendix C.1 and Fig. 5.

3.7.1 Neurophysiological Interpretation of Selected Channels

Across three datasets, DCLT consistently identifies compact and neurophysiologically meaningful channel subsets under different channel budgets. PhysioNet MI emphasizes frontal and motor regions, DEAP highlights frontal–motor representations with reduced occipital activation, and AMIGOS preserves frontal dominance with complementary temporal–parietal–occipital activity. These results demonstrate that DCLT adaptively captures stable physiological signatures while suppressing redundant channels.

3.8 Ablation Study

To validate the contribution of each component in DCLT, ablation experiments are conducted by removing key modules, including the Transformer backbone, STE-based gating, sparsity regularization, and Top-K constraint. As shown in Table 5, each component contributes to robust brainprint recognition under channel constraints.

Replacing the Transformer with CNN or vanilla ViT leads to performance degradation, demonstrating the necessity of global dependency modeling under sparse EEG topology. Removing STE-based hard selection prevents effective electrode reduction, while removing sparsity regularization causes unstable channel convergence. Furthermore, relaxing the Top-K constraint verifies that DCLT can effectively purify redundant information and preserve discriminative identity features under extreme channel compression.

Table 6: Inference performance comparison between full-channel and DCLT compact configurations on a real-world dataset (23 subjects)

Metric	**Base** ($K = 60$)	**DCLT** ($K = 16$)	Δ
Accuracy (%)	99.10	90.72	-8.32
Single Latency (ms)	7.36	5.49	-1.87
Throughput (ms)	0.0317	0.0224	-0.0093

3.9 System Overhead and Inference Performance for Edge Deployment

To evaluate IoT-oriented deployment efficiency, we conduct edge-constrained simulations on EEG data from 23 subjects with batch size 1. As shown in Table 6, reducing the channel configuration from 60 to 16 enables DCLT to maintain 90.72% recognition accuracy while reducing single-sample latency from 7.36 ms to 5.49 ms (25.4%) and throughput latency from 0.0317 ms to 0.0224 ms (29.3%). These results demonstrate that DCLT achieves an effective accuracy-efficiency trade-off and is suitable for resource-constrained IoT-BCI deployment.

3.10 Visualization and Interpretability Analysis

Fig. 2 visualizes the channel selection behavior, attention patterns, decision boundaries, robustness, deployment efficiency, and overall performance of DCLT.

The visualization shows compact spatial channel sets, focused spatiotemporal attention, clearer class separation, stable performance under missing electrodes, and favorable accuracy–latency trade-offs. Panel-by-panel discussion is provided in Appendix C.2.

4 Conclusion

This paper proposes DCLT, a differentiable structure learning framework integrating STE-driven Top-K channel selection and Dynamic-Static Consistency Analysis for IoT-oriented EEG recognition. DCLT bridges hardware constraints and neural optimization, achieving over 90% accuracy with only 25% channels while reducing computational complexity and inference latency through optimized configurations.

AI Use Statement

Generative AI tools were used to assist with English translation and language editing, mathematical typesetting, and manuscript organization. They were not used to formulate research hypotheses or mathematical claims, develop the proposed methodology, design or conduct experiments, generate or process experimental data, implement the proposed method, interpret experimental results, or determine the scientific conclusions of this work. All AI-assisted content was reviewed and verified by the authors, who take full responsibility for the final manuscript.

REFERENCES

A. A. Abdellatif, M. G. Khafagy, A. Mohamed, et al. Eeg-based transceiver design with data decomposition for healthcare iot applications. *IEEE Internet of Things Journal*, 5(5):3569–3579, 2018.

N. Aggarwal, U. Shukla, G. J. Saxena, et al. Mean based relief: An improved feature selection method based on relieff. *Applied Intelligence*, 53(19):23004–23028, 2023.

T. Alotaiby, F. E. A. El-Samie, S. A. Alshebeili, et al. A review of channel selection algorithms for eeg signal processing. *EURASIP Journal on Advances in Signal Processing*, 2015(1):66, 2015.

A. M. Alqudah and Z. Moussavi. A review of deep learning for biomedical signals: Current applications, advancements, future prospects, interpretation, and challenges. *Computers, Materials & Continua*, 83(3), 2025.

Z. A. A. Alyasseri et al. Eeg channel selection for person identification using binary grey wolf optimizer. *IEEE Access*, 10:10500–10513, 2022. doi: 10.1109/ACCESS.2021.3135805.

A. Arjun, A. S. Rajpoot, and M. R. Panicker. Introducing attention mechanism for eeg signals: Emotion recognition with vision transformers. In *Proc. 43rd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, pp. 5723–5726, 2021.

F. Artoni, A. Delorme, and S. Makeig. Applying dimension reduction to eeg data by principal component analysis reduces the quality of its subsequent independent component decomposition. *NeuroImage*, 175:176–187, 2018.

M. Arvaneh, C. Guan, K. K. Ang, et al. Optimizing the channel selection and classification accuracy in eeg-based bci. *IEEE transactions on biomedical engineering*, 58(6):1865–1873, 2011.

Y. Bai, L. Yang, R. Gao, Q. Wang, J. Song, and K. Yao. Eeg channel selection algorithm based on attention mechanism for emotion recognition. In *2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pp. 6950–6957, 2025. doi: 10.1109/BIBM66473.2025.11356915.

M. Z. Baig, N. Aslam, and H. P. H. Shum. Filtering techniques for channel selection in motor imagery eeg applications: a survey. *Artificial intelligence review*, 53(2):1207–1232, 2020.

J. Birjandtalab, M. B. Pouyan, D. Cogan, et al. Automated seizure detection using limited-channel eeg and non-linear dimension reduction. *Computers in biology and medicine*, 82:49–58, 2017.

E. C. E. Bozkurt. Lightweight design for eeg acquisition: a trade-off between reconstruction accuracy and data efficiency. *arXiv preprint*, 2025.

Y. Changrong, L. Mingxu, and L. Yongjin. Research on brainprint recognition system based on bp neural network. In *2025 10th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA)*, pp. 80–84. IEEE, 2025.

C. Dai, D. Pi, and S. I. Becker. Shapelet-transformed multi-channel eeg channel selection. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(5):1–27, 2020.

Y. Ding, N. Robinson, S. Zhang, Q. Zeng, and C. Guan. Tsception: Capturing temporal dynamics and spatial asymmetry from eeg for emotion recognition. *IEEE Trans. Affect. Comput.*, 14(3): 2238–2250, 2023.

A. Dosovitskiy, L. Beyer, A. Kolesnikov, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

S. Faul and W. Marnane. Dynamic, location-based channel selection for power consumption reduction in eeg analysis. *Computer methods and programs in biomedicine*, 108(3):1206–1215, 2012.

R. Fu, M. Han, Y. Tian, and P. Shi. Improvement motor imagery eeg classification based on sparse common spatial pattern and regularized discriminant analysis. *J. Neurosci. Methods*, 343:108833, 2020.

B. Hejrati, A. Fathi, and F. Abdali-Mohammadi. Efficient lossless multi-channel eeg compression based on channel clustering. *Biomedical Signal Processing and Control*, 31:295–300, 2017.

A. S. Janani, T. S. Grummett, H. Bakhshayesh, et al. How many channels are enough? evaluation of tonic cranial muscle artefact reduction using ica with different numbers of eeg channels. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 101–105. IEEE, 2018.

X. Jin, J. Tang, X. Kong, et al. Ctnn: A convolutional tensor-train neural network for multi-task brainprint recognition. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 29:103–112, 2020.

S. Koelstra, C. Muhl, M. Soleymani, et al. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.

X. Kong, W. Kong, Q. Fan, Q. Zhao, and A. Cichocki. Taskindependent eeg identification via low-rank matrix decomposition. In *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, pp. 412–419, 2018.

F. Laport, A. Dapena, P. M. Castro, et al. A prototype of eeg system for iot. *International journal of neural systems*, 30(07):2050018, 2020.

V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance. Eegnet: A compact convolutional neural network for eeg-based brain–computer interfaces. *J. Neural Eng.*, 15(5):056013, 2018.

Q. Li, L. Hu, X. Wang, et al. A novel approach with channel reduction scheme for depression symptom identification based on resting-state eeg. *IEEE Sensors Journal*, 2026a.

Z. Li, S. He, J. Huang, et al. Decoding multilingual imagined speech from scalp eeg via dynamic differentiable graph hierarchical fusion network. *Information Fusion*, 120:104262, 2026b.

R. Mane, N. Robinson, A. P. Vinod, S.-W. Lee, and C. Guan. A multiview cnn with novel variance layer for motor imagery brain computer interface. In *Proc. 42nd Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, pp. 2950–2953, 2020.

Y. Ming, D. Pelusi, C. N. Fang, et al. Eeg data analysis with stacked differentiable neural computers. *Neural Computing and Applications*, 32(12):7611–7621, 2020.

J. A. Miranda-Correa, M. K. Abadi, N. Sebe, et al. Amigos: A dataset for affect, personality and mood research on individuals and groups. *IEEE transactions on affective computing*, 12(2):479–493, 2018.

L. A. Moctezuma and M. Molinas. Towards a minimal eeg channel array for a biometric system using resting-state and a genetic algorithm for channel selection. *Scientific Reports*, 10(1):14917, 2020.

H. Nakanishi, M. Suzuki, and Y. Matsuo. Fixing the train-test objective discrepancy: Iterative image inpainting for unsupervised anomaly detection. *Journal of Information Processing*, 30:495–504, 2022.

K. Pearson. Liii. on lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin philosophical magazine and journal of science*, 2(11):559–572, 1901.

G. Schalk et al. Eeg motor movement/imagery dataset (version 1.0.0). PhysioNet, 2009. RRID:SCR_007345.

Y. Song, Q. Zheng, B. Liu, and X. Gao. Eeg conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 31:710–719, 2022.

H. Touvron, A. Vedaldi, M. Douze, et al. Fixing the train-test resolution discrepancy. In *Advances in neural information processing systems*, volume 32, 2019.

Y. Tu, S. Li, X. Chen, and D. Wu. Brainprintnet: A multiscale cross-band fusion network for eeg-based brainprint recognition. *IEEE Transactions on Information Forensics and Security*, 21: 2757–2768, 2026. doi: 10.1109/TIFS.2026.3672000.

A. Vaswani, N. Shazeer, N. Parmar, et al. Attention is all you need. In *Advances in neural information processing systems*, volume 30, 2017.

G. Wang, W. Liu, Y. He, C. Xu, L. Ma, and H. Li. Eegpt: Pretrained transformer for universal and reliable representation of eeg signals. In *Proc. Adv. Neural Inf. Process. Syst.*, pp. 39249–39280, 2024.

W. Wu, Z. Chen, X. Gao, Y. Li, E. N. Brown, and S. Gao. Probabilistic common spatial patterns for multichannel eeg analysis. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(3):639–653, 2015.

Richard Yang and Emma Modesitt. Vit2eeg: leveraging hybrid pretrained vision transformers for eeg data. *arXiv preprint arXiv:2308.00454*, 2023.

D. Yazdani, R. Cheng, D. Yazdani, et al. A survey of evolutionary continuous dynamic optimization over two decades—part b. *IEEE Transactions on Evolutionary Computation*, 25(4):630–650, 2021.

P. Yin, J. Lyu, S. Zhang, et al. Understanding straight-through estimator in training activation quantized neural nets. *arXiv preprint arXiv:1903.05662*, 2019.

X. Yong, R. K. Ward, and G. E. Birch. Sparse spatial filter optimization for eeg channel reduction in brain-computer interface. In *2008 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 417–420. IEEE, 2008.

Y. Zhang, Y. Yu, H. Li, et al. Dmae-eeg: A pretraining framework for eeg spatiotemporal representation learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2025.

C. Zheng et al. Disentangled representation learning for robust brainprint recognition. *IEEE Transactions on Information Forensics and Security*, 20:9263–9274, 2025. doi: 10.1109/TIFS.2025.3602266.

A Additional Benchmark Results

The following table reports the complete random-channel comparison supporting the PhysioNet analysis in the main text.

B Supplementary Method and Consistency Figures

The gating schematic and channel-budget consistency plot support the corresponding analyses in the main text.

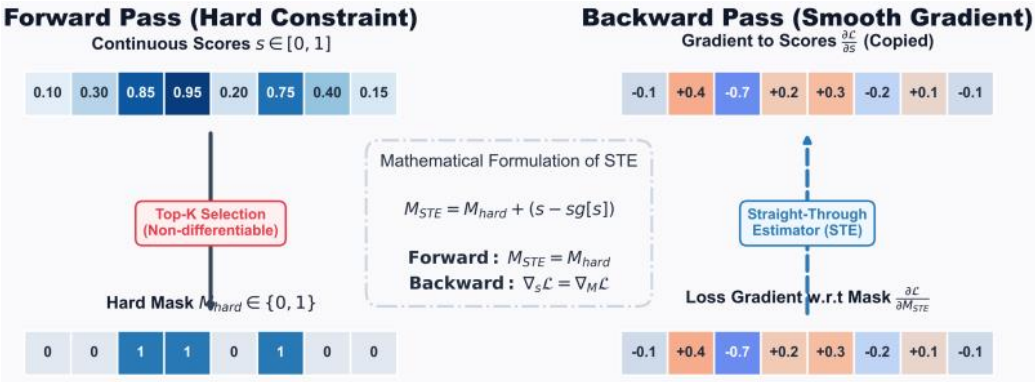


Figure 3: Schematic illustration of the STE-driven Top-K discrete gating mechanism.

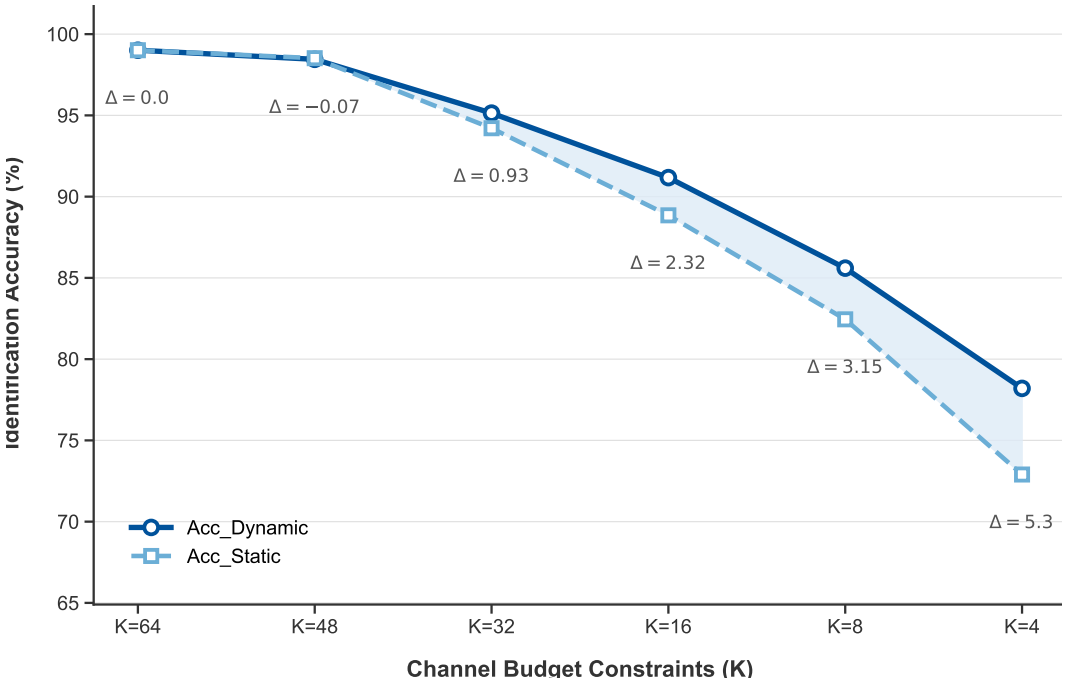


Figure 4: Performance consistency evolution under varying channel budgets.

C Additional Interpretability Analyses

C.1 Frequency Statistics of Dynamically Routed Channel Selection

To investigate the optimization behavior of the Adaptive Channel Selection Module (ACSM), Fig. 5 illustrates the cumulative activation frequency of EEG channels selected by the STE-based hard mask. Across all datasets, DCLT exhibits a consistent long-tail distribution, where only a small subset of channels is selected frequently while redundant channels are progressively suppressed. Despite different compression ratios (64→16, 32→8, and 14→7), the learned channel distributions converge to stable sparse patterns, demonstrating that DCLT reliably identifies discriminative electrodes and provides efficient channel configurations for resource-constrained wearable EEG systems.

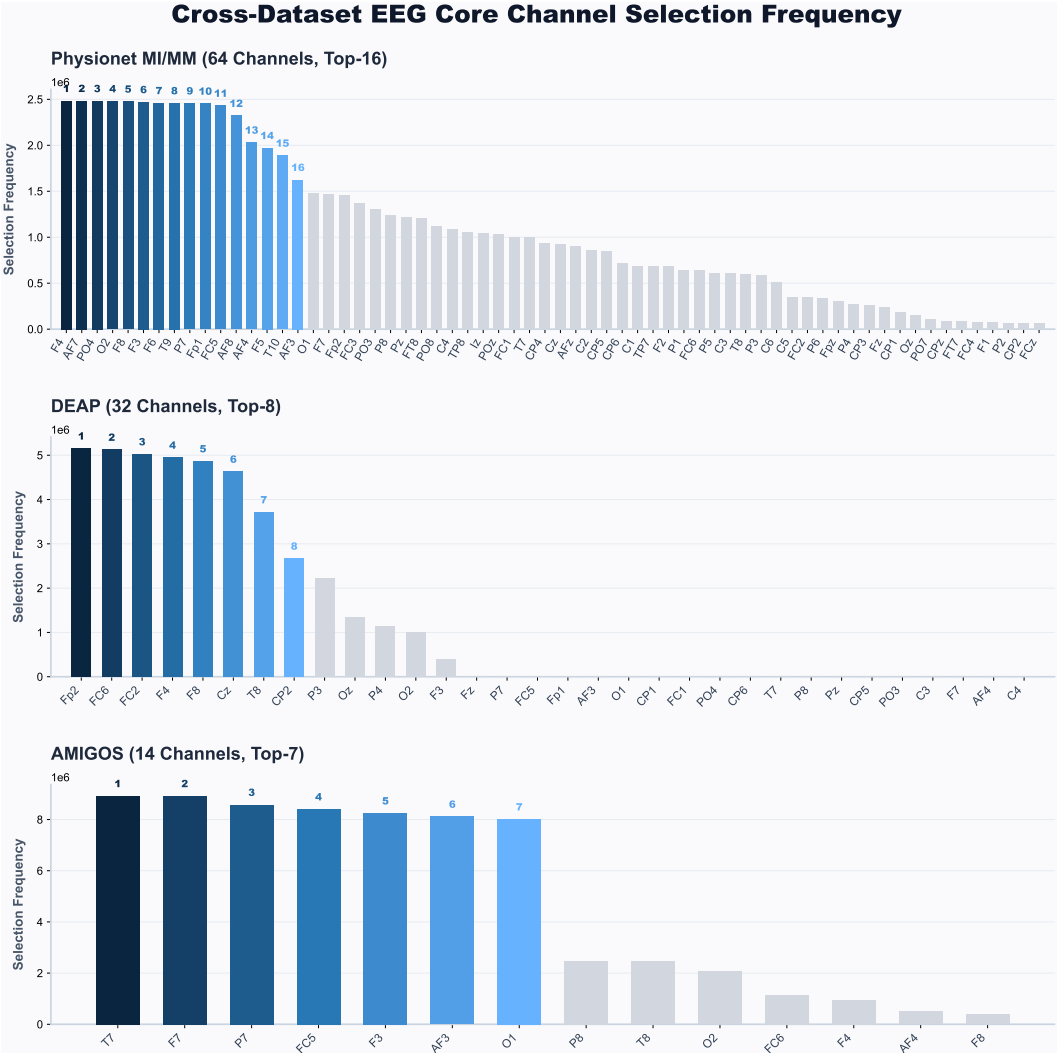


Figure 5: Channel selection frequency distribution across electrodes during training.

C.2 Detailed Analysis of Figure 2

(a) Spatial topological distribution. Fig. 2(a) shows compact channel sets in neurophysiologically relevant regions across datasets.

(b) Spatiotemporal attention. Fig. 2(b) shows attention focused on a few electrodes and time intervals, suppressing redundant features.

(c) Decision boundaries. Fig. 2(c) shows tighter within-class clusters and clearer between-class separation than ablated variants, supporting STE optimization and sparsity regularization.

(d) Robustness. Fig. 2(d) shows higher median accuracy and lower variance across repeated experiments.

(e) Deployment trade-off. Fig. 2(e) shows competitive accuracy with lower inference latency at smaller channel budgets.

(f) Overall performance. Fig. 2(f) summarizes the balance of recognition accuracy, channel efficiency, and deployment feasibility.