Bayesian Test-Time Adaptation via Dirichlet Feature Projection and GMM-Driven Inference for Motor Imagery EEG Decoding

Huan Luo, Na Lu*, Xiaopeng Wang, Xu Niu

Xi’an Jiaotong University

Introduction

Test-Time Adaptation:

During model inference, adapt the model online using only sequentially arriving unlabeled test samples, enabling real-time calibrated prediction without access to source data or target labels.

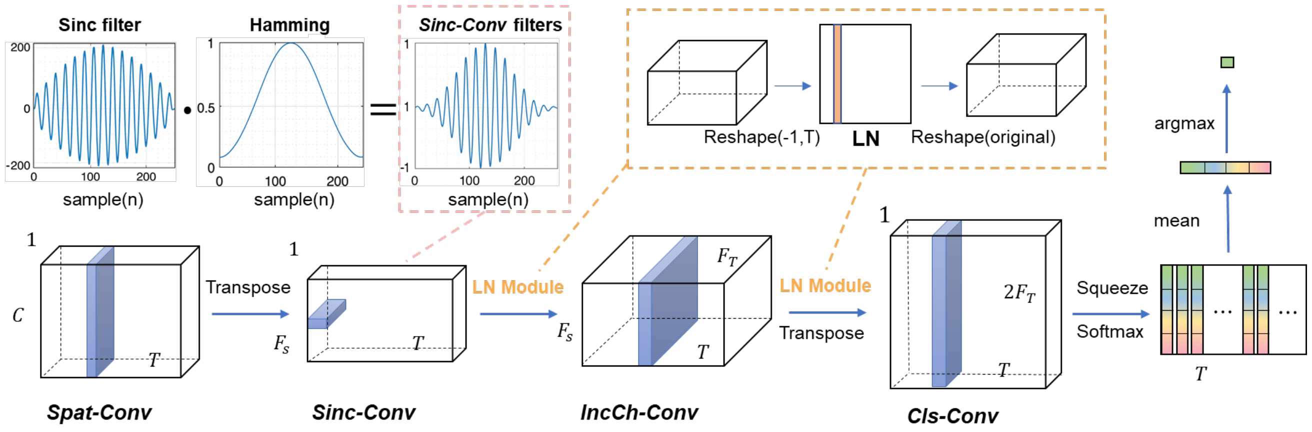
Challenges:

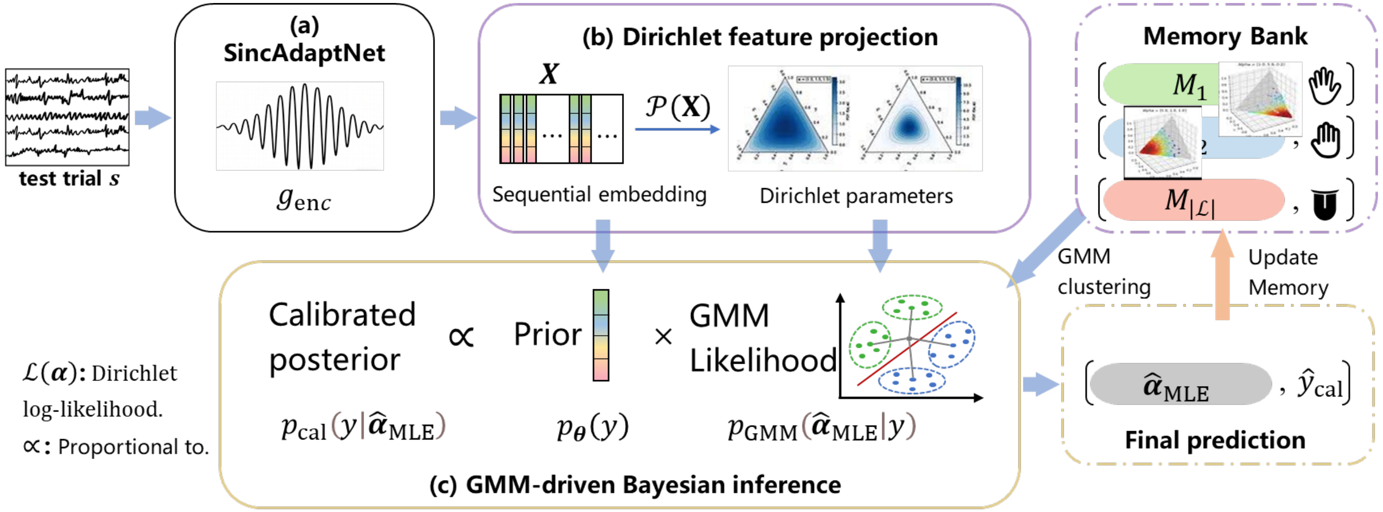
1. No source data;
2. No target labels;
3. Target samples arrive sequentially.

BTTA-DG:

- ➢ **Sinc-Based Adaptive Bandpass Filtering Network**
- ➢ **Dirichlet Feature Projection**
- ➢ **GMM-Driven Bayesian Inference**

Sinc-Based Adaptive Bandpass Filtering Network





- $\mathcal{L}(\alpha)$: Dirichlet log-likelihood.
- $\propto$: Proportional to.

Core Formula

Ⅰ Dirichlet Feature Projection

Time-Varying Predictive Embeddings

$\boldsymbol{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_T] \in \mathbb{R}^{|\mathcal{L}| \times T}$

$\boldsymbol{x}_j \in \Delta^{|\mathcal{L}|-1}$

$\boldsymbol{x}_{ij} \geq 0,\ \sum_i \boldsymbol{x}_{ij} = 1$

Dirichlet Assumption

$\boldsymbol{X} \sim Dir(\boldsymbol{\alpha}),\ \ \boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_{|\mathcal{L}|}) \in \mathbb{R}_+^{|\mathcal{L}|}$

supp $Dir(\boldsymbol{\alpha}) \subseteq \Delta^{|\mathcal{L}|-1}$

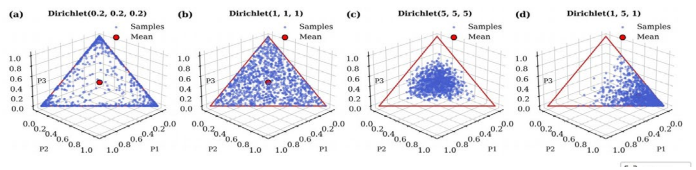
MLE Projection $\mathcal{P}$

$\widehat{\boldsymbol{\alpha}}_{\text{MLE}} = \underset{\boldsymbol{\alpha}}{\operatorname{argmax}} \sum_j^T \log D(\boldsymbol{x}_j; \boldsymbol{\alpha})$

$D(\boldsymbol{x}_j; \boldsymbol{\alpha}) = \frac{\Gamma(\alpha_0)}{\prod_i^{|\mathcal{L}|} \Gamma(\alpha_i)} \prod_i^{|\mathcal{L}|} \boldsymbol{x}_{ij}^{\alpha_i - 1}$

$\alpha_0 = \sum_i \alpha_i$: scale (↑ → low uncertainty)

α_i: concentration of class i (↑ → prior to class i)



(a) vs (b) vs (c): $\alpha_0 \uparrow$ | (b) vs (d): $\alpha_i \uparrow$

Ⅱ Accelerated Dirichlet Estimation

Moment-Based Initialization

$\bar{x}_i = (1/\mathrm{T}) \sum_j x_{ij} \qquad \sigma_i^2 = (1/\mathrm{T}) \sum_j (x_{ij} - \bar{x}_i)^2$

$\alpha_0^{\text{init}} \approx (1/|\mathcal{L}|) \sum_i [\bar{x}_i (1 - \bar{x}_i)/\sigma_i^2 - 1]$

$\alpha_i^{\text{init}} = \alpha_0^{\text{init}} \cdot \bar{x}_i$

Fixed-point MLE

$\alpha_i^{\text{new}} = \psi^{-1}\left(\psi(\alpha_0^{\text{old}}) + (1/\mathrm{T}) \sum_j^T \log x_{ij}\right)$

$\alpha_0^{\text{new}} = \sum_i \alpha_i^{\text{new}} \qquad \widehat{\boldsymbol{\alpha}}_{\text{MLE}} = \boldsymbol{\alpha}^{\text{new}};$

Newton Iteration for $\psi^{-1}(v)$

$u^{\text{new}} = u^{\text{old}} - [\psi(u^{\text{old}}) - v]/\psi'(u^{\text{old}})$

$u^{\text{init}} = \exp(v + 0.5) \quad [v \geq -2.22]$

$u^{\text{init}} = -1/(v + \gamma) \quad [v < -2.22]$

ψ: Digamma $\qquad \psi'$: Trigamma

Convergence

Until: $\|\Delta\alpha\|/\|\alpha^{old}\| < 10^{-4}$ 5~10 iters/trial

Memory Buffer Update

If $\boldsymbol{conf}(p_{cal}) \geq \tau_{\text{conf}}$ AND $\boldsymbol{H}(p_{cal}) \leq \tau_{\text{ent}}$:

$\boldsymbol{M}_{\hat{y}_{\text{cal}}} \leftarrow \boldsymbol{M}_{\hat{y}_{\text{cal}}} \cup \{\widehat{\boldsymbol{\alpha}}_{MLE}\}$ (**FIFO**)

Ⅲ GMM-Driven Bayesian Inference

GMM Likelihood Modeling

$p_{\text{GMM}}(\boldsymbol{\alpha}|y) = \sum_1^K \pi_{yk} \mathcal{N}(\boldsymbol{\alpha}; \boldsymbol{\mu}_{yk}, \boldsymbol{\Sigma}_{yk})$

$\sum_1^K \pi_{yk} = 1,\ \pi_{yk} > 0$

K : components $\quad \boldsymbol{\mu}, \boldsymbol{\Sigma}$: mean/std

Bayesian Posterior Calibration

$p_\theta(y) = f_\theta(s))$

$p_{\text{cal}}(y|\widehat{\boldsymbol{\alpha}}_{\text{MLE}}) \propto p_{\text{GMM}}(\widehat{\boldsymbol{\alpha}}_{\text{MLE}}|y) \cdot p_{\boldsymbol{\theta}}(y)$

Final Calibrated Prediction

$\hat{y}_{\text{cal}} = \operatorname{argmax}_y p_{\text{cal}}(y|\widehat{\boldsymbol{\alpha}}_{\text{MLE}})$

Complete Pipeline

$\boldsymbol{X} = g_{\text{enc}}(\boldsymbol{s})$ (Feature Extraction)

$\widehat{\boldsymbol{\alpha}}_{\text{MLE}} = \mathcal{P}(\boldsymbol{X})$ (Dirichlet Feature Projection)

$p_{\text{GMM}}(\widehat{\boldsymbol{\alpha}}_{\text{MLE}}|y)$ (GMM Likelihood)

$p_{\text{cal}} \rightarrow \hat{y}_{cal}$ (Bayesian Calibration)

Idea

- $\hat{\alpha}$ depends on current trial
- GMM models historical distribution
- gradient-free

Principles:
1. Temporal embeddings -> compact parameter $\boldsymbol{\alpha}$;
2. Accelerated MLE (fixed-point + Newton);
3. GMM models the distribution of historical $\boldsymbol{\alpha}$;
4. Bayesian inference fuses prior with historical evidence

Experiments

Cross-Subject Adaptation Accuracy (BNCI2014001)

Table 1: Cross-subject adaptation accuracy (%) on BNCI2014001, with an asterisk(*) denoting the significance level (*: p<0.05).

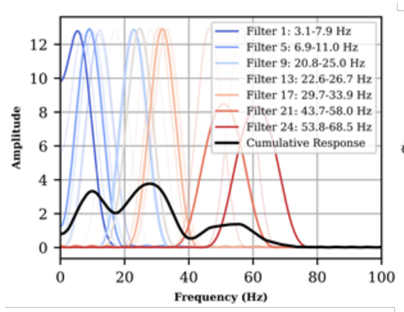
Setting	Method	S0	S1	S2	S3	S4	S5	S6	S7	S8	Avg.
Source	CSP [5]	83.33	52.08	97.92	75.00	56.25	67.36	72.22	88.19	71.53	73.77
	EEGNet [20]	83.19	60.28	92.08	67.92	57.22	72.50	64.86	86.11	79.44	$73.73_{\pm 1.11}$
	SincAdaptNet	84.97	63.93	97.68	77.13	56.22	72.68	67.26	93.86	79.56	$77.03_{\pm 1.31}$
Online TTA	BN-adapt [39]	84.97	63.93	97.68	77.13	56.22	72.68	67.26	93.86	79.56	$77.03_{\pm 1.31}$
	Tent [42]	75.97	57.92	94.51	68.54	52.22	65.21	59.38	90.14	68.19	$70.23_{\pm 3.28}$
	PL [21]	76.46	56.67	97.92	70.34	52.29	66.32	60.42	93.89	72.15	$71.83_{\pm 3.21}$
	CoTTA [44]	85.00	63.68	98.05	76.32	57.22	72.08	67.64	94.63	80.48	$77.24_{\pm 1.51}$
	SAR [31]	84.24	63.40	97.36	76.25	54.72	69.10	67.50	93.54	80.28	$76.27_{\pm 1.92}$
	T-TIME [23]	84.44	61.94	97.43	76.11	56.60	69.38	63.13	94.65	79.38	$75.90_{\pm 1.95}$
	OTTA [46]	84.43	63.60	97.14	**77.63**	57.63	73.04	66.44	95.26	**83.14**	$77.58_{\pm 1.33}$
	BTTA-DG	**87.51***	**66.67***	**98.61***	77.08	**57.64**	**73.61**	**68.75***	**95.83***	82.64	$\mathbf{78.70}^{*}_{\pm 1.32}$

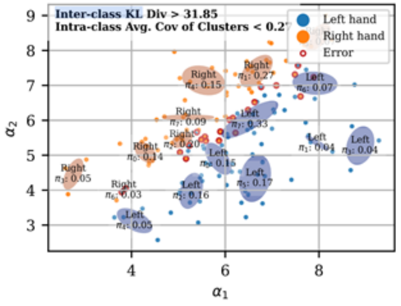
Per-Trial Inference Time

Table 2: Average inference time (ms) on BNCI2014001.

Method	BN-adapt [39]	Tent [42]	PL [21]	CoTTA [44]	SAR [31]	T-TIME [23]	OTTA [46]	BTTA-DG
Time	5.1	18.4	17.8	23.0	32.5	18.5	19.5	**15.7**

Sinc Filter Responses & Dirichlet–GMM Visualization





Ablation Study (Across 4 Datasets)

Table 8: Ablation Study across settings and MI datasets. Mean ± s.d. accuracy (%).

Method	BNCI2014001		BNCI2014002	BNCI2015001	SHU MI
	cross-session	cross-subject	cross-subject	cross-subject	cross-subject
SincAdaptNet (Source Only)	80.62 ± 2.70	75.30 ± 1.82	76.40 ± 1.62	73.92 ± 1.95	61.02 ± 1.70
BTTA-DG w/o EA	81.88 ± 2.58	76.85 ± 1.65	77.55 ± 1.59	75.06 ± 1.88	61.90 ± 1.68
SincAdaptNet + EA	82.33 ± 2.62	77.03 ± 1.31	78.05 ± 2.48	75.48 ± 1.84	62.42 ± 1.72
SincAdaptNet + EA + GMM	82.47 ± 2.59	77.55 ± 1.35	78.25 ± 2.36	75.62 ± 1.82	62.58 ± 1.70
SincAdaptNet + EA + Dirichlet	84.04 ± 2.55	77.61 ± 1.43	78.88 ± 1.47	76.36 ± 1.78	63.24 ± 1.78
BTTA-DG (Full Model)	**86.50 ± 2.49**	**78.70 ± 1.32**	**80.29 ± 1.07**	**77.92 ± 1.76**	**64.06 ± 1.92**

Sinc Filter Rhythm Alignment Analysis

Table 6: Quantitative alignment of learned Sinc filter passbands with known MI-EEG rhythms on BNCI2014001. Values denote, for each subject, the percentage (%) of filters whose passbands overlap each band.

Frequency band	S0	S1	S2	S3	S4	S5	S6	S7	S8	Avg.
μ (8–13 Hz)	24.17	25.83	25.83	23.33	24.17	27.50	24.17	24.17	25.00	24.91
β (13–30 Hz)	32.50	35.00	28.33	28.33	43.33	25.83	30.00	28.33	29.17	31.20
γ (30–45 Hz)	35.83	30.83	34.17	40.00	26.67	40.00	40.00	40.00	36.67	36.02
Other	7.50	8.33	11.67	8.33	5.83	6.67	5.83	7.50	9.17	7.87

+4%
Average Accuracy

78.70%
BNCI2014001 Cross-Subject

80.29%
BNCI2014002 Cross-Subject

15.7ms
Per-Trial Inference Time

92.1%
μ+β+γ Frequency Band Coverage

KL>31
Inter-Class Dirichlet Divergence

Experiments

Cross-Session Adaptation Accuracy (BNCI2014001)

Table 5: Cross-session adaptation accuracy (%) on BNCI2014001, with an asterisk(*) denoting the significance level (*: p<0.05).

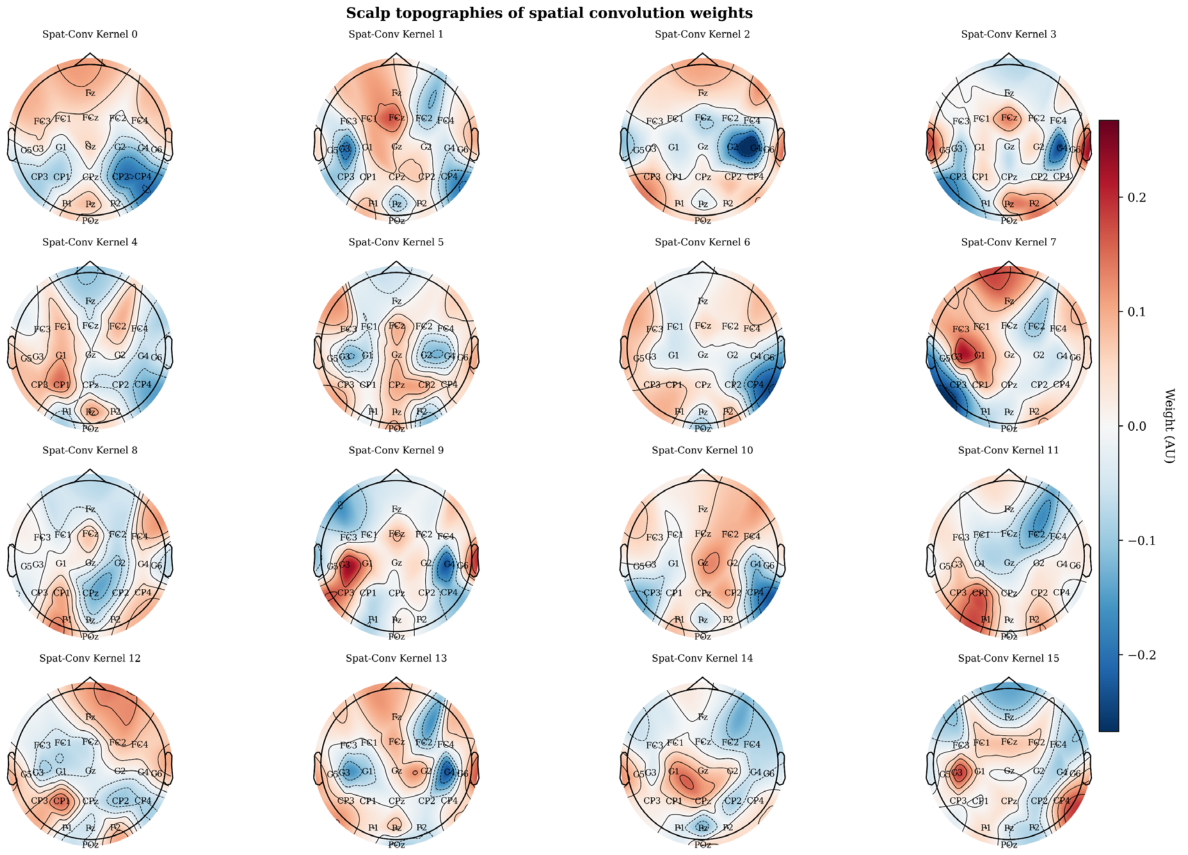
Setting	Method	S0	S1	S2	S3	S4	S5	S6	S7	S8	Avg.
Source	CSP	88.19	54.17	97.22	65.97	48.61	70.14	68.06	94.44	90.97	75.31
	EEGNet	86.81	**63.54**	94.65	70.97	72.92	68.61	73.26	93.47	92.71	$79.66_{\pm 2.52}$
	EEG Conformer	87.92	58.75	97.15	70.35	75.83	68.54	77.77	95.63	89.23	$80.13_{\pm 3.18}$
	SincAdaptNet	84.69	56.64	98.51	70.32	86.22	72.06	82.47	97.54	92.47	$82.33_{\pm 2.62}$
Online TTA	BN-adapt	84.69	56.64	98.51	70.32	86.22	72.06	82.47	97.54	92.47	$82.33_{\pm 2.62}$
	Tent	80.35	51.18	99.03	57.92	64.38	62.22	64.51	93.06	90.83	$73.72_{\pm 4.77}$
	PL	77.71	51.74	98.75	58.13	75.28	64.58	68.75	97.01	91.46	$75.93_{\pm 4.72}$
	CoTTA	85.63	54.51	99.44	69.03	86.46	72.36	82.71	98.33	93.06	$82.39_{\pm 2.83}$
	SAR	86.32	55.00	99.24	71.11	86.11	70.49	82.92	96.53	91.88	$82.18_{\pm 2.99}$
	T-TIME	78.40	54.03	98.33	69.79	81.94	70.49	80.69	97.50	91.53	$80.30_{\pm 3.42}$
	OTTA	**89.71**	55.89	96.79	72.42	91.58	73.67	87.49	96.03	91.58	$83.91_{\pm 2.25}$
	BTTA-DG	85.42	62.50	**100.00***	**76.39***	**91.67**	**77.78***	**90.97**	**100.00***	**93.75**	$\mathbf{86.50^{*}_{\pm 2.49}}$

Test-Time Adaptation under Online Class Imbalance

Table 9: Performance of BTTA-DG under varying online class imbalance ratios on the BNCI2014001 dataset. As imbalance increases, the model specializes, improving minority class accuracy.

Class Ratio (0 : 1)	Accuracy Class 0 (%)	Accuracy Class 1 (%)	Overall Accuracy (%)
1 : 1	77.01 ± 1.53	80.40 ± 1.45	$\mathbf{78.70 \pm 1.32}$
1 : 0.75	74.07 ± 1.44	80.45 ± 1.34	76.81 ± 1.25
1 : 0.5	69.75 ± 1.52	83.02 ± 1.42	74.17 ± 1.28
1 : 0.25	64.67 ± 1.18	85.19 ± 1.64	68.77 ± 1.19

2D Scalp Topography of Learned SpatConv Kernels



BTTA-DG achieves the highest average accuracy of 86.50% in cross-session adaptation, maintaining strong robustness under online class imbalance. The learned *SpatConv* kernel topographies exhibit interpretable spatial attention patterns consistent with motor-related cortical regions.