LEARNING SUBJECT STABLE CORTICAL DYNAMICS FROM EEG WITH EVENT SYNCHRONIZED INVARIANT MODES

Anonymous authors
Paper under double-blind review

ABSTRACT

EEG based motor decoding is usually built around a separation that is convenient computationally but weak scientifically. Source imaging is used to estimate where cortical activity occurs, while classification is applied afterward to decide what movement was performed. For fine upper limb actions, this separation is problematic because the most anatomically complete reconstruction is not necessarily the representation that transfers best across people. We introduce Event-Synchronized Invariant Cortical Modes (ESICM), which learns a compact cortical coordinate system jointly shaped by the EEG forward model, movement timing, action discrimination, inter-subject invariance, and cortical geometry. ESICM replaces dense source reconstruction with a small set of leadfield-constrained modes obtained from the training subjects and decodes their multiband temporal dynamics with ESICM-Tiny. No sensor branch, post-hoc source fusion, or subject-specific adaptation is required at inference. Under strict nested leave-one-subject-out evaluation, ESICM achieves the highest mean accuracy across all eight motor-execution and motor-imagery settings, including fine-grained six- and seven-class decoding. On NVIDIA Jetson Nano, the learned cortical representation reduces source preprocessing to 2.425,ms and achieves 4.178,ms median end-to-end latency at 235.7,Hz. These results show that cortical localization, cross-subject invariance, and embedded decoding can be treated as a single representation-learning problem rather than sequential stages. Code and reproduction instructions are available at `https://anonymous.4open.science/r/ESICM_Jetson-DD18/README.md`.

1 INTRODUCTION

Electroencephalography remains one of the most practical sensing modalities for noninvasive brain computer interfaces because it provides millisecond temporal resolution with wearable scalp sensors (Wolpaw et al., 2002). Motor BCIs have progressed from coarse left versus right decisions toward decoding individual upper limb actions and continuous movement variables. Three dimensional hand trajectories can be reconstructed from scalp EEG (Bradberry et al., 2010), while executed and imagined elbow, forearm, and hand movements can be distinguished from low frequency activity (Ofner et al., 2017). Motor information is distributed across complementary temporal and spectral processes, including slow cortical potentials and mu and beta synchronization changes (Pfurtscheller & Da Silva, 1999). Reliable transfer to an unseen subject, however, remains difficult because EEG varies with anatomy, electrode placement, recording conditions, and individual physiology (Lotte et al., 2018; Wu et al., 2020).

Most decoding methods address this problem directly in sensor space. Spatial filters improve task separation (Blankertz et al., 2008), covariance based methods exploit the geometry of trial level spatial structure (Barachant et al., 2011), and convolutional networks learn temporal and spatial filters from EEG (Schirrmeister et al., 2017). EEGNet provides a compact depthwise architecture (Lawhern et al., 2018), while BENDR and EEG Conformer use learned representations and attention to capture longer temporal dependencies (Kostas et al., 2021; Song et al., 2022). Despite these advances, larger models do not remove the effects of limited data, nonstationarity, and subject dependence (Craik et al., 2019; Roy et al., 2019). Subject-independent evaluation is essential be-

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.

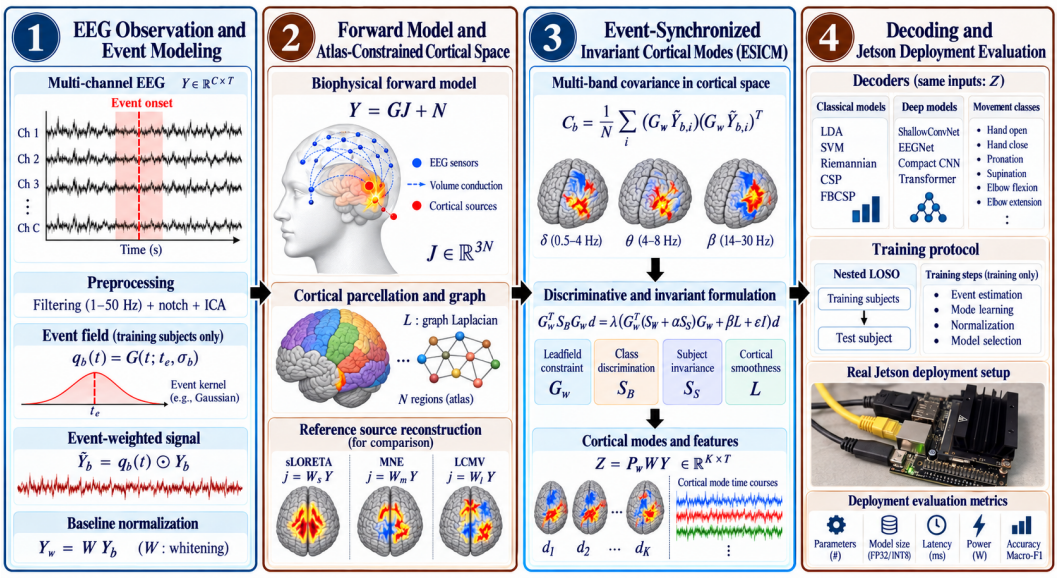

Figure 1: Overview of the proposed ESICM framework for subject-independent upper-limb EEG decoding. The pipeline comprises four stages: (1) event-aware EEG preprocessing and synchronization, and (2) biophysical forward modeling and atlas-constrained cortical representation. (3) Event-Synchronized Invariant Cortical Mode (ESICM) learning, jointly enforcing leadfield consistency, class discrimination, subject invariance, and cortical smoothness, and (4) classical/deep decoding with nested LOSO evaluation and resource-aware Jetson deployment.

cause random trial splits can inflate generalization through subject-specific information (Jayaram & Barachant, 2018), making representation quality as critical as classifier choice.

EEG source imaging provides an anatomical representation by relating scalp measurements to latent cortical activity through a biophysical forward model. Minimum norm estimation provides a distributed inverse solution (Hämäläinen & Ilmoniemi, 1994), dynamic statistical parametric mapping improves spatial normalization (Dale et al., 2000), sLORETA standardizes current density estimates (Pascual-Marqui et al., 2002), and LCMV beamforming constructs spatial filters for candidate cortical generators (Van Veen et al., 1997). These methods move the representation from electrode coordinates toward cortical anatomy, but the inverse problem remains ill posed and is sensitive to head geometry, conductivity assumptions, source orientation, covariance estimation, and inverse regularization (Michel & Brunet, 2019; Mahjoory et al., 2017). Anatomical reconstruction alone therefore does not guarantee that the resulting representation is optimal for distinguishing closely related movements across subjects.

Structured inverse methods reduce some of these limitations. Mixed norm estimation encourages compact spatial support (Gramfort et al., 2012), time frequency formulations model nonstationary source dynamics (Gramfort et al., 2013b), and iterative reweighting can sharpen estimated support (Strohmeier et al., 2016). Sparse Bayesian methods estimate relevance through adaptive variance components (Tipping, 2001; Wipf & Nagarajan, 2009), while learned source imaging incorporates biophysical structure into neural estimators (Sun et al., 2022). These methods improve reconstruction through sparsity, temporal structure, adaptive priors, or learned constraints, yet reconstruction and classification are still commonly optimized as separate stages. This distinction matters because a physiologically plausible source estimate is not necessarily the most discriminative representation, while predictive model weights should not automatically be interpreted as neural generators (Haufe et al., 2014).

Fine-grained upper-limb decoding is challenging because nearby movements recruit overlapping sensorimotor regions, while the most informative activity may lie in only a few coordinated cortical patterns. ESICM addresses this by learning a compact cortical basis directly from the training subjects, combining event-synchronized weighting, baseline noise whitening, the leadfield, cross-

subject stability, and cortical smoothness in a single generalized eigenproblem. Each trial is then decoded from low-dimensional cortical-mode time courses rather than from full source maps or a parallel sensor pathway. Figure 1 summarizes the complete pipeline.

The main contribution of ESICM is therefore not simply a new decoder, but a learned cortical coordinate system optimized jointly for discrimination, cross-subject stability, and neuroanatomical structure. We make four contributions: (1) an event-synchronized, leadfield-constrained objective that jointly models class separation, subject invariance, and cortical geometry; (2) compact discriminative cortical modes that replace high-dimensional source-amplitude reconstruction; (3) a leakage-controlled nested leave-one-subject-out protocol in which event estimation, mode learning, normalization, model selection, and decoding use training subjects only, with MNE, sLORETA, and LCMV retained as conventional source-space references; and (4) a lightweight temporal decoder evaluated jointly for predictive performance and edge efficiency using accuracy, macro-F1, parameter count, memory, and inference latency (Banbury et al., 2021; Tragoudaras et al., 2023). This combination of physiologically constrained representation learning and compact deployment is the central strength of the proposed framework.

2 METHODOLOGY

2.1 ESICM: LEARNING A CORTICAL COORDINATE SYSTEM

Let $X_i \in \mathbb{R}^{C\times T}$ denote EEG trial i with movement label y_i and subject identity s_i. Conventional inverse pipelines first estimate a dense cortical signal and then learn a decoder on the reconstructed amplitudes. ESICM instead learns the cortical coordinate system itself:

$$X_i \to \{Z_i^b\}_{b\in\mathcal{B}} \to e_i \to \hat{y}_i, \qquad \mathcal{B} = \{\text{Broad}, \text{MRCP}, \mu, \beta\}. \tag{1}$$

The four views span 0.3–30, 0.3–3, 8–13, and 13–30 Hz and capture complementary slow and oscillatory motor activity (Pfurtscheller & Da Silva, 1999). We use the upper-limb ME and MI dataset of Ofner et al. (2017), retaining the fixed 61-channel montage and cue-locked interval $[-2, 3]$ s.

EEG generation is modeled as

$$Y(t) = GJ(t) + \varepsilon(t), \qquad G_w = WG, \tag{2}$$

where $G \in \mathbb{R}^{C\times R}$ is the cortical leadfield. Prestimulus covariance is estimated from the training subjects and regularized before whitening, $W = \Sigma_n^{-1/2}$, following standard covariance stabilization principles (Ledoit & Wolf, 2004). The same forward model is used for MNE, sLORETA, and LCMV reference reconstructions (Hämäläinen & Ilmoniemi, 1994; Pascual-Marqui et al., 2002; Van Veen et al., 1997). Cortical geometry is represented by graph Laplacian $L \in \mathbb{R}^{R\times R}$ (Shuman et al., 2013).

Algorithm 1 summarizes ESICM training. For each frequency view, ESICM learns a training-only event field, estimates class and subject covariance structure, and solves the cortical eigenproblem to obtain invariant modes. These modes are converted into fixed sensor-to-cortical projections and used to train the compact prototype decoder. At inference, only multiband filtering, cortical projection, normalization, and decoding are required, enabling efficient subject-independent deployment.

2.2 TRAINING-ONLY EVENT SYNCHRONIZATION

Rather than prescribing a movement window, ESICM estimates where discriminative activity occurs from the training population. For subject s, sensor j, band b, and time t, a Fisher-style score is

$$F_{s,j}^b(t) = \frac{\mathrm{Var}_c\left[\mu_{s,c,j}^b(t)\right]}{\frac{1}{N_c}\sum_c \sigma_{s,c,j}^{2,b}(t) + \epsilon}, \tag{3}$$

following the classical between-versus-within variability principle (Sugiyama, 2007). Strong sensor responses are aggregated within subject, followed by a median across training subjects. After temporal smoothing, the resulting population curve $f_b(t)$ is converted to a bounded event field

$$q_b(t) = q_{\min} + (1 - q_{\min}) \operatorname{clip}\left(\frac{f_b(t) - P_5(f_b)}{P_{95}(f_b) - P_5(f_b) + \epsilon}, 0, 1 \right), \tag{4}$$

with $q_{\min} = 0.1$. Hence no test-subject activity and no fixed class-specific temporal crop determine the event representation.

For whitened EEG $Y^b_{w,i} = W X^b_i$, temporal evidence enters the spatial statistics through

$$C^b_i = \frac{Y^b_{w,i} \operatorname{diag}(q_b)(Y^b_{w,i})^\top}{\sum_t q_b(t) + \epsilon}, \qquad C^b_i \leftarrow \frac{C^b_i}{\operatorname{tr}(C^b_i) + \epsilon}. \tag{5}$$

Covariance-based discrimination is closely related to the principle underlying CSP and its regularized variants (Ramoser et al., 2000; Blankertz et al., 2008; Lu et al., 2010); ESICM differs by transferring this structure through the cortical forward model and explicitly modeling subject instability.

2.3 Invariant Cortical Mode Objective

Let C^b_0, C^b_c, and $C^b_{s,c}$ denote the global, class, and subject-class covariance means. We define the movement contrast

$$S^b_B = \frac{1}{N_c} \sum_c (C^b_c - C^b_0)(C^b_c - C^b_0)^\top, \tag{6}$$

and the subject-instability operator

$$S^b_S = \frac{1}{|\mathcal{S}| N_c} \sum_{s,c} (C^b_{s,c} - C^b_c)(C^b_{s,c} - C^b_c)^\top. \tag{7}$$

S^b_B rewards movement-dependent structure, whereas S^b_S measures how much the same movement changes across subjects. The cortical mode d is obtained from

$$\mathcal{R}_b(d) = \frac{d^\top G_w^\top S^b_B G_w d}{d^\top \left[G_w^\top (C^b_0 + \alpha S^b_S) G_w + \beta L + \epsilon I \right] d}. \tag{8}$$

Maximization yields

$$G_w^\top S^b_B G_w d = \lambda \left[G_w^\top (C^b_0 + \alpha S^b_S) G_w + \beta L + \epsilon I \right] d. \tag{9}$$

The generalized eigenvalue form is related to discriminative covariance filtering (Ramoser et al., 2000; Blankertz et al., 2008), while $d^\top L d$ follows graph-smoothness regularization (Belkin et al., 2006; Shuman et al., 2013). The contribution is their coupling through G_w with a training-derived event field and the explicit subject-instability term S^b_S. Thus, ESICM does not seek the source configuration that best reconstructs every cortical generator; it seeks physically realizable cortical directions that preserve movement structure while rejecting subject-dependent variation.

The leading K eigenvectors form $D_b = [d_{b,1}, \ldots, d_{b,K}]$, with scalp realization

$$A_b = G_w D_b. \tag{10}$$

A ridge-stabilized dual projection (Hansen, 2010) gives

Algorithm 1 Event Synchronized Invariant Cortical Modes

Require: Training trials $\mathcal{D} = \{X_i, y_i, s_i\}$, leadfield G, cortical Laplacian L, number of modes K
Ensure: Bandwise projections $\{P_b\}$, event fields $\{q_b\}$, decoder Θ

```
Construct EEG views {X_i^b}_{b∈B}, B = {Broad, MRCP, μ, β}
Estimate training baseline whitener W and set G_w ← WG
for each band b ∈ B do
    Estimate subject robust event field q_b(t)
    Compute event weighted covariances {C_i^b}
    Form class contrast S_B^b and subject instability S_S^b
    Solve
        G_w^⊤ S_B^b G_w D_b = Λ_b [G_w^⊤ (C_0^b + α S_S^b) G_w + βL + εI] D_b
    Retain the leading K cortical modes in D_b
    A_b ← G_w D_b
    P_b ← (A_b^⊤ A_b + ρI)^{-1} A_b^⊤ W
    Z_i^b ← P_b X_i^b
end for
Stack cortical trajectories Z_i ← Stack_b(Z_i^b)
Train event conditioned prototype decoder Θ on {Z_i, y_i}
return {P_b, q_b}_{b∈B}, Θ
```

$$P_b = (A_b^\top A_b + \rho I)^{-1} A_b^\top W, \qquad Z_i^b = P_b X_i^b. \tag{11}$$

Consequently, deployment no longer requires a dense inverse solution. Each band is reduced by a fixed $K \times C$ sensor-to-mode operator. The modes are therefore interpreted as discriminative cortical coordinates rather than unrestricted estimates of neural generators (Haufe et al., 2014).

2.4 Mode-Only Cortical Decoder

The bandwise trajectories are stacked as $Z_i \in \mathbb{R}^{B \times K \times T}$. The event field remains active during decoding through

$$\widetilde{Z}_{i,b,k,t} = Z_{i,b,k,t} \left[1 + g q_b(t)\right], \qquad g = \mathrm{softplus}(\theta_g). \tag{12}$$

Two depthwise-separable temporal blocks model the resulting cortical dynamics, following the parameter-efficient principle of depthwise and pointwise factorization (Chollet, 2017; Lawhern et al., 2018). If $H_i(t)$ denotes the final temporal feature map, the trial embedding combines event-weighted, global, and dispersion information,

$$e_i = \psi\left(\left[\sum_t \bar{q}(t) H_i(t),\ \mathrm{mean}_t\, H_i(t),\ \mathrm{std}_t\, H_i(t)\right]\right). \tag{13}$$

Movement classes are represented by learned prototypes p_c, following metric-space prototype learning (Snell et al., 2017). Prediction uses normalized similarity,

$$\ell_{i,c} = \tau \frac{e_i^\top p_c}{\|e_i\|_2 \|p_c\|_2}, \qquad \hat{y}_i = \arg\max_c \ell_{i,c}, \tag{14}$$

with learned temperature τ. The proposed decoder receives only ESICM modes; no sensor logits, teacher network, distillation, or sensor-source fusion enters the prediction.

2.5 Leakage-Free Subject Generalization

Evaluation uses nested leave-one-subject-out validation. For held-out subject $s^\star$,

Table 1: Cross-subject motor execution decoding across the complete configuration space of sensor, conventional source, and learned cortical-mode representations. Mean accuracy ± standard deviation (%) is reported over 15 held-out subjects under strict LOSO evaluation; bold marks the highest accuracy for each task.

Configuration	Four+Rest	HO/HC/Rest	Six	Six+Rest
Sensor + LDA	30.03 ± 4.10	39.43 ± 4.03	18.81 ± 2.09	17.48 ± 1.94
Sensor + MDM	28.29 ± 3.72	35.81 ± 6.54	18.01 ± 1.64	16.36 ± 1.95
Sensor + CNN	55.74 ± 9.65	63.39 ± 12.96	31.12 ± 6.91	36.42 ± 8.83
Sensor + EEGNet	50.83 ± 9.46	59.37 ± 13.51	29.66 ± 11.18	32.61 ± 11.07
Sensor + Transformer	41.90 ± 6.46	49.17 ± 12.47	25.09 ± 6.54	29.58 ± 7.59
MNE + LDA	31.27 ± 5.24	42.18 ± 8.56	19.25 ± 3.17	19.32 ± 4.57
MNE + MDM	31.61 ± 4.03	37.25 ± 4.31	17.99 ± 1.97	17.39 ± 2.55
MNE + CNN	49.41 ± 8.23	58.43 ± 11.83	25.70 ± 5.44	30.13 ± 7.01
MNE + EEGNet	49.61 ± 7.20	55.13 ± 9.91	29.92 ± 9.61	31.85 ± 7.89
MNE + Transformer	37.87 ± 4.96	45.15 ± 9.41	20.03 ± 3.67	21.90 ± 4.75
sLORETA + LDA	31.61 ± 4.78	38.69 ± 6.09	19.23 ± 3.43	19.05 ± 4.75
sLORETA + MDM	27.61 ± 3.76	35.99 ± 4.74	18.36 ± 2.86	16.92 ± 3.27
sLORETA + CNN	49.30 ± 7.15	57.97 ± 10.68	26.10 ± 6.49	30.78 ± 7.00
sLORETA + EEGNet	49.72 ± 7.23	55.23 ± 10.41	29.71 ± 9.53	31.63 ± 7.96
sLORETA + Transformer	38.09 ± 4.75	44.62 ± 8.28	20.27 ± 3.67	21.83 ± 4.77
LCMV + LDA	35.28 ± 7.53	45.50 ± 11.90	21.23 ± 5.31	21.91 ± 6.54
LCMV + MDM	28.58 ± 3.98	39.11 ± 5.73	17.54 ± 1.57	16.04 ± 2.07
LCMV + CNN	52.84 ± 9.52	57.76 ± 7.68	28.20 ± 6.21	32.36 ± 7.30
LCMV + EEGNet	51.53 ± 8.72	59.81 ± 13.22	29.46 ± 10.18	32.79 ± 9.28
LCMV + Transformer	40.10 ± 6.54	48.04 ± 12.16	22.54 ± 5.48	26.19 ± 6.63
ESICM + ESICM-Tiny	$\mathbf{58.05 \pm 9.16}$	$\mathbf{66.82 \pm 11.00}$	$\mathbf{31.17 \pm 8.48}$	$\mathbf{38.90 \pm 8.95}$

$$\mathcal{D}_{\text{test}} = \{i : s_i = s^\star\}, \qquad \mathcal{D}_{\text{train}} = \{i : s_i \neq s^\star\}. \tag{15}$$

Every supervised quantity, including q_b, W, S_B^b, S_S^b, D_b, P_b, normalization, hyperparameters, and decoder parameters, is estimated without the test subject. Inner folds are also grouped by subject. Comparisons include MNE, sLORETA, LCMV, Riemannian covariance models (Barachant et al., 2011), ShallowConvNet (Schirrmeister et al., 2017), EEGNet (Lawhern et al., 2018), Transformer, and compact depthwise decoders.

2.6 System Implementation

Offline workstation: Model construction and nested subject-independent training were implemented in Python 3.11 with PyTorch 2.3 and MNE-Python 1.8 (Gramfort et al., 2014; 2013a) on an Alienware Aurora R16 equipped with a 13th Gen Intel Core i9-13900F, 32 GB RAM, and an NVIDIA GeForce RTX 2080 Ti with 11 GB memory. The workstation estimates the expensive subject-population quantities q_b, W, D_b, and P_b. These quantities are frozen after training; deployment requires only the learned band filters, sensor-to-mode matrices, normalization statistics, and compact decoder.

NVIDIA Jetson P3450: Edge deployment was evaluated independently on an NVIDIA Jetson Nano P3450. ESICM-Tiny was exported to ONNX and compiled with TensorRT FP16, while the complete preprocessing and cortical-mode construction stages were retained in the deployment path to measure true end-to-end cost rather than decoder latency alone. We report latency, throughput, memory footprint, device utilization, temperature, power, and energy per inference following TinyML benchmarking practice (Banbury et al., 2021; Tragoudaras et al., 2023). Predictive metrics are taken exclusively from strict LOSO evaluation, whereas all-subject models are used only for hardware characterization.

3 Results

We evaluate the proposed framework across two complementary EEG benchmarks. First, using the upper-limb dataset of Ofner et al. (2017), we assess subject-independent decoding over eight set-

Table 2: Cross-subject motor imagery decoding across the complete configuration space of sensor, conventional source, and learned cortical-mode representations. Mean accuracy ± standard deviation (%) is reported over 15 held-out subjects under strict LOSO evaluation; bold marks the highest accuracy for each task.

Configuration	Four+Rest	HO/HC/Rest	Six	Six+Rest
Sensor + LDA	30.04 ± 3.81	38.76 ± 3.58	17.93 ± 1.79	17.34 ± 1.94
Sensor + MDM	26.82 ± 2.95	34.84 ± 4.45	17.20 ± 1.87	15.29 ± 1.72
Sensor + CNN	35.96 ± 5.16	43.51 ± 6.97	20.40 ± 2.14	21.65 ± 3.83
Sensor + EEGNet	32.86 ± 3.59	43.26 ± 3.78	19.57 ± 2.50	18.68 ± 2.81
Sensor + Transformer	29.54 ± 3.40	36.47 ± 3.00	18.38 ± 2.27	16.81 ± 2.29
MNE + LDA	29.93 ± 3.00	38.95 ± 4.34	17.34 ± 1.88	16.16 ± 1.90
MNE + MDM	27.25 ± 3.04	35.66 ± 2.87	17.48 ± 1.33	15.18 ± 1.49
MNE + CNN	36.25 ± 5.95	44.11 ± 5.10	20.08 ± 2.94	21.18 ± 2.88
MNE + EEGNet	33.83 ± 4.15	43.05 ± 5.06	18.51 ± 1.99	19.53 ± 2.79
MNE + Transformer	29.64 ± 3.20	38.26 ± 4.15	17.49 ± 1.92	15.67 ± 2.23
sLORETA + LDA	30.17 ± 3.41	37.88 ± 3.56	17.40 ± 1.59	16.06 ± 1.59
sLORETA + MDM	27.56 ± 2.98	36.24 ± 2.76	17.21 ± 1.18	15.40 ± 1.57
sLORETA + CNN	34.58 ± 6.18	45.45 ± 5.91	19.83 ± 2.69	21.25 ± 3.12
sLORETA + EEGNet	33.19 ± 3.73	43.70 ± 5.58	18.35 ± 2.42	19.63 ± 2.65
sLORETA + Transformer	29.08 ± 2.58	37.37 ± 3.21	17.78 ± 2.10	15.67 ± 2.33
LCMV + LDA	27.56 ± 2.76	35.64 ± 3.62	16.95 ± 1.79	15.75 ± 1.87
LCMV + MDM	25.75 ± 3.21	34.93 ± 4.97	16.35 ± 1.70	14.77 ± 2.76
LCMV + CNN	31.33 ± 6.55	40.72 ± 6.62	17.18 ± 2.50	16.73 ± 3.10
LCMV + EEGNet	30.45 ± 4.53	38.36 ± 4.58	17.89 ± 1.92	17.24 ± 3.18
LCMV + Transformer	26.42 ± 2.95	36.11 ± 4.92	17.26 ± 2.16	14.77 ± 2.32
ESICM + ESICM-Tiny	$\mathbf{39.40 \pm 6.06}$	$\mathbf{47.60 \pm 7.94}$	$\mathbf{21.03 \pm 3.30}$	$\mathbf{24.04 \pm 4.48}$

tings spanning four class configurations under both motor execution (ME) and motor imagery (MI), enabling analysis across different movement granularities and task conditions. Second, we evaluate the source-localization and decoding behavior on the BCI2000 motor-imagery dataset Schalk et al. (2004), using four-class discrimination of left fist, right fist, both fists, and both feet imagery. Together, these experiments examine predictive performance, cross-subject robustness, and the consistency of the learned cortical representation across distinct EEG datasets and motor paradigms.

3.1 Subject-Independent Decoding Across Eight Settings

We evaluate ESICM against 20 sensor- and conventional source-space configurations under strict 15-subject LOSO evaluation across four ME and four MI settings (Tables 1 and 2). ESICM obtains the highest mean accuracy in all eight settings, with the largest gains appearing in Four+Rest, HO/HC/Rest, and Six+Rest decoding. The consistent ranking across executed and imagined movements, and across different class granularities, supports the transferability of the learned cortical-mode representation.

For ME, ESICM attains $58.05 \pm 9.16\%$, $66.82 \pm 11.00\%$, $31.17 \pm 8.48\%$, and $38.90 \pm 8.95\%$ for Four+Rest, HO/HC/Rest, Six, and Six+Rest, respectively. The corresponding MI accuracies are $39.40 \pm 6.06\%$, $47.60 \pm 7.94\%$, $21.03 \pm 3.30\%$, and $24.04 \pm 4.48\%$. Importantly, the advantage persists as the decoding problem becomes increasingly fine-grained, including the challenging six- and seven-class configurations, where inter-class cortical differences are substantially more subtle.

Although previous studies have investigated the same upper-limb dataset, direct numerical comparison is generally not methodologically meaningful because evaluation protocols and classification objectives differ substantially. For example, Mammone et al. (2020) perform source-level decoding using binary classification, whereas our study evaluates strict cross-subject generalization under substantially harder multi-class configurations with up to seven classes. Rather than optimizing a separate inverse solution and classifier for each task, ESICM uses a single formulation that jointly learns a compact cortical coordinate system and its decoder. Its consistent ranking across all eight LOSO settings therefore provides evidence that the learned modes capture transferable motor structure rather than task- or subject-specific discriminative patterns.

Table 3: Mechanism ablation of ESICM under identical 15-subject LOSO evaluation. Accuracy is reported as mean ± standard deviation (%) across held-out subjects. Bold denotes the highest accuracy for each decoding setting.

Configuration	Four+Rest	Hand Open/Close/Rest	Six Movements	Six+Rest
Motor Execution (ME)				
Full ESICM	$\mathbf{58.05 \pm 9.16}$	$\mathbf{66.82 \pm 11.00}$	$\mathbf{31.17 \pm 8.48}$	$\mathbf{38.90 \pm 8.95}$
Without Event Weighting in Cortical Modes	52.21 ± 7.88	62.03 ± 10.92	30.98 ± 8.47	33.62 ± 7.04
Without Event Conditioning in Decoder	48.45 ± 9.37	57.47 ± 14.19	29.59 ± 9.23	31.08 ± 9.67
Without Event Synchronization	46.40 ± 8.06	55.84 ± 12.74	28.33 ± 8.23	30.83 ± 7.46
Without Subject-Invariance Regularization	51.38 ± 6.75	62.81 ± 12.45	30.95 ± 7.45	34.58 ± 7.99
Without Cortical-Graph Regularization	52.05 ± 8.51	61.90 ± 12.12	28.94 ± 7.51	33.65 ± 9.12
Motor Imagery (MI)				
Full ESICM	$\mathbf{39.40 \pm 6.06}$	$\mathbf{47.60 \pm 7.94}$	$\mathbf{21.03 \pm 3.30}$	$\mathbf{24.04 \pm 4.48}$
Without Event Weighting in Cortical Modes	34.25 ± 6.41	43.80 ± 6.43	18.57 ± 2.69	19.71 ± 4.89
Without Event Conditioning in Decoder	30.92 ± 5.22	40.95 ± 5.67	18.12 ± 2.55	17.81 ± 3.05
Without Event Synchronization	31.29 ± 5.23	41.15 ± 5.73	18.16 ± 2.22	17.64 ± 2.66
Without Subject-Invariance Regularization	33.82 ± 6.13	42.93 ± 6.97	19.28 ± 2.63	19.67 ± 4.53
Without Cortical-Graph Regularization	33.34 ± 6.17	43.01 ± 6.07	19.69 ± 2.05	19.06 ± 3.90

3.2 Mechanism Ablation

We ablate ESICM by mechanism while fixing LOSO splits, training budget, cortical dimensionality, and decoder capacity. The full model remains best across all eight ME/MI settings (Table 3), showing that its gains are not attributable to a single component. Event synchronization is dominant: removing it lowers mean accuracy from 40.88% to 33.71% (-7.17 points), while removing event-conditioned decoding causes a comparable 6.58-point loss. Event weighting, subject-invariance, and cortical-graph regularization contribute additional gains of 3.98, 3.95, and 4.42 points, respectively. The consistent degradation across both ME and MI reveals a complementary mechanism: temporal synchronization supplies discriminative structure, while subject and cortical constraints stabilize how that structure is represented across individuals. ESICM therefore benefits from coordinated representation learning rather than any isolated architectural component.

3.3 Deployment Efficiency

Table 4 shows that the dominant deployment cost of conventional source-space pipelines lies in constructing the cortical representation itself, requiring approximately 3.43 ms before decoding begins. ESICM removes this bottleneck by learning a compact cortical basis directly for prediction, reducing source preprocessing to 2.425 ms, about 29% below conventional source methods. The resulting pipeline reaches 4.178 ms median end-to-end latency, 4.887 ms p99 latency, and 235.7 Hz throughput, corresponding to 3.0% lower median latency and 1.9% higher throughput than the fastest conventional source-space configuration. The key distinction is that ESICM does not compress a dense source estimate after reconstruction; it learns the low-dimensional cortical coordinates that are needed for decoding in the first place. The expensive representation learning is therefore absorbed during training, while deployment uses fixed sensor-to-mode projections and a compact temporal decoder. This makes cortical representation design part of the systems optimization itself and shifts source-space inference from reconstruction-first processing to representation-native deployment.

3.4 Subject-Stable Cortical Representation Analysis

We examine whether ESICM generalizes beyond aggregate LOSO accuracy from three complementary views. Appendix Table 5 compares the learned cortical representation with the full sensor/source decoder set, including EEGConformer and a DANN-style subject-adversarial baseline (Song et al., 2022; Tang & Zhang, 2020). For EEG-DANN, subject identity is adversarially suppressed using only the training subjects; no held-out-subject data enter optimization. Appendix Tables 6 and 7 further test subject-wise consistency and cross-fold cortical reproducibility. Together, these analyses distinguish representation-level invariance from gains produced by decoder capacity or adversarial training alone.

Table 4: Edge deployment efficiency of sensor-space, conventional source-space, and learned cortical-mode pipelines on NVIDIA Jetson Nano. All models use TensorRT FP16 with batch size 1. Preprocessing measures representation construction prior to network inference.

Configuration	Params ↓	Preproc. (ms) ↓	E2E p50 (ms) ↓	E2E p99 (ms) ↓	Throughput (Hz) ↑
Sensor Space					
Sensor + CNN	48,540	0.402	1.742	1.973	567.6
Sensor + EEGNet	2,316	0.401	1.176	1.488	829.9
Sensor + TinyDepthwiseCNN	2,377	0.401	1.205	1.520	810.8
Sensor + Transformer	70,804	0.401	1.702	1.967	581.5
Conventional Source Space					
MNE + CNN	72,348	3.427	4.843	5.354	204.0
MNE + EEGNet	**2,812**	3.426	4.579	4.802	217.9
MNE + TinyDepthwiseCNN	3,152	3.424	4.307	**4.531**	231.3
MNE + Transformer	84,196	3.425	4.801	5.226	206.2
sLORETA + CNN	72,348	3.426	4.870	5.287	203.5
sLORETA + EEGNet	**2,812**	3.429	4.573	4.841	218.1
sLORETA + TinyDepthwiseCNN	3,152	3.424	4.313	4.736	229.1
sLORETA + Transformer	84,196	3.425	4.842	5.250	205.6
LCMV + CNN	72,348	3.427	4.867	5.438	200.6
LCMV + EEGNet	**2,812**	3.424	4.578	4.768	217.9
LCMV + TinyDepthwiseCNN	3,152	3.426	4.319	4.715	229.2
LCMV + Transformer	84,196	3.423	4.805	5.237	205.9
Learned Cortical Modes					
ESICM + ESICM-Tiny	15,618	**2.425**	**4.178**	4.887	**235.7**

3.5 Cross-Subject Generalization on BCI2000

On the PhysioNet EEG Motor Movement/Imagery dataset (Schalk et al., 2004), ESICM–Tiny reaches $66.88 \pm 14.45\%$ accuracy, $65.78 \pm 14.97\%$ macro-F1, and $\kappa = 0.525$ under strict subject-disjoint evaluation. It exceeds the strongest conventional source pipeline, sLORETA+CNN, by 6.82 accuracy points while using 78.4% fewer trainable parameters. Because prior results on the same four motor-imagery classes use different evaluation protocols (Lun et al., 2022; Zhou et al., 2024; Lim et al., 2025), we avoid direct leaderboard claims and provide the full controlled comparison in Appendix Section A.1. The result supports the central ESICM hypothesis: cross-subject transfer improves when invariance is learned in the cortical coordinate system itself rather than deferred to the decoder.

4 Conclusion

We introduced ESICM, a source-space EEG representation that learns *which cortical coordinates should be preserved for decoding* rather than first reconstructing a dense source distribution. By coupling training-derived event synchronization, leadfield consistency, subject-invariance regularization, and cortical smoothness, ESICM transforms high-dimensional EEG into compact cortical trajectories optimized jointly for discrimination and cross-subject stability. Across eight subject-independent ME/MI settings, the same representation remains competitive across movement granularities and achieves the strongest performance in the fine-grained six-class ME task, while the ablation analysis identifies event synchronization as the dominant contributor to predictive performance. Cross-fold cortical analysis further shows that the learned modes remain geometrically reproducible across held-out subjects, supporting the interpretation of ESICM as a stable cortical coordinate system rather than an unconstrained latent embedding. Evaluation on BCI2000 further demonstrates that this representation transfers to a distinct motor-imagery benchmark while retaining a compact decoder. Finally, Jetson deployment shows that moving dimensionality reduction into the learned cortical representation reduces source-space preprocessing and provides a favorable accuracy–latency operating point for embedded inference. Together, these results suggest that subject-independent EEG decoding can be formulated as a unified problem of learning *when*, *where*, and *which* cortical dynamics should be preserved for prediction, rather than treating source reconstruction, feature compression, and classification as independent stages.

REFERENCES

Colby Banbury, Vijay Janapa Reddi, Peter Torelli, Jeremy Holleman, Nat Jeffries, Csaba Kiraly, Pietro Montino, David Kanter, Sebastian Ahmed, Danilo Pau, et al. Mlperf tiny benchmark. *arXiv preprint arXiv:2106.07597*, 2021.

Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Multiclass brain–computer interface classification by riemannian geometry. *IEEE Transactions on Biomedical Engineering*, 59(4):920–928, 2011.

Mikhail Belkin, Partha Niyogi, and Vikas Sindhwani. Manifold regularization: A geometric framework for learning from labeled and unlabeled examples. *Journal of machine learning research*, 7 (11), 2006.

Benjamin Blankertz, Ryota Tomioka, Steven Lemm, Motoaki Kawanabe, and Klaus-Robert Muller. Optimizing spatial filters for robust eeg single-trial analysis. *IEEE Signal processing magazine*, 25(1):41–56, 2008.

Trent J Bradberry, Rodolphe J Gentili, and José L Contreras-Vidal. Reconstructing three-dimensional hand movements from noninvasive electroencephalographic signals. *The journal of Neuroscience*, 30(9):3432–3437, 2010.

François Chollet. Xception: Deep learning with depthwise separable convolutions. In *2017 IEEE conference on computer vision and pattern recognition (CVPR)*, pp. 1800–1807. Ieee, 2017.

Alexander Craik, Yongtian He, and Jose L Contreras-Vidal. Deep learning for electroencephalogram (eeg) classification tasks: a review. *Journal of neural engineering*, 16(3):031001, 2019.

Anders M Dale, Arthur K Liu, Bruce R Fischl, Randy L Buckner, John W Belliveau, Jeffrey D Lewine, and Eric Halgren. Dynamic statistical parametric mapping: combining fmri and meg for high-resolution imaging of cortical activity. *neuron*, 26(1):55–67, 2000.

Alexandre Gramfort, Matthieu Kowalski, and Matti Hämäläinen. Mixed-norm estimates for the m/eeg inverse problem using accelerated gradient methods. *Physics in medicine and biology*, 57 (7):1937–1961, 2012.

Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Roman Goj, Mainak Jas, Teon Brooks, Lauri Parkkonen, et al. Meg and eeg data analysis with mne-python. *Frontiers in Neuroinformatics*, 7:267, 2013a.

Alexandre Gramfort, Daniel Strohmeier, Jens Haueisen, Matti S Hämäläinen, and Matthieu Kowalski. Time-frequency mixed-norm estimates: Sparse m/eeg imaging with non-stationary source activations. *NeuroImage*, 70:410–422, 2013b.

Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Lauri Parkkonen, and Matti S Hämäläinen. Mne software for processing meg and eeg data. *neuroimage*, 86:446–460, 2014.

Matti S Hämäläinen and Risto J Ilmoniemi. Interpreting magnetic fields of the brain: minimum norm estimates. *Medical & biological engineering & computing*, 32(1):35–42, 1994.

Per Christian Hansen. *Discrete inverse problems: insight and algorithms*. SIAM, 2010.

Stefan Haufe, Frank Meinecke, Kai Görgen, Sven Dähne, John-Dylan Haynes, Benjamin Blankertz, and Felix Bießmann. On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, 87:96–110, 2014.

Vinay Jayaram and Alexandre Barachant. Moabb: trustworthy algorithm benchmarking for bcis. *Journal of neural engineering*, 15(6):066011, 2018.

Demetres Kostas, Stephane Aroca-Ouellette, and Frank Rudzicz. Bendr: Using transformers and a contrastive self-supervised learning task to learn from massive amounts of eeg data. *Frontiers in Human Neuroscience*, 15:653659, 2021.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018.

Olivier Ledoit and Michael Wolf. A well-conditioned estimator for large-dimensional covariance matrices. *Journal of multivariate analysis*, 88(2):365–411, 2004.

Galvin Brice S Lim, Brian Godwin S Lim, Argel A Bandala, John Anthony C Jose, Timothy Scott C Chu, and Edwin Sybingco. Agtcnet: A graph-temporal approach for principled motor imagery eeg classification. *IEEE Access*, 13:187383–187409, 2025.

Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rakotomamonjy, and Florian Yger. A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. *Journal of neural engineering*, 15(3):031005, 2018.

Haiping Lu, How-Lung Eng, Cuntai Guan, Konstantinos N Plataniotis, and Anastasios N Venetsanopoulos. Regularized common spatial pattern with aggregation for eeg classification in small-sample setting. *IEEE transactions on Biomedical Engineering*, 57(12):2936–2946, 2010.

Xiangmin Lun, Jianwei Liu, Yifei Zhang, Ziqian Hao, and Yimin Hou. A motor imagery signals classification method via the difference of eeg signals between left and right hemispheric electrodes. *Frontiers in Neuroscience*, 16:865594, 2022.

Keyvan Mahjoory, Vadim V Nikulin, Loïc Botrel, Klaus Linkenkaer-Hansen, Marco M Fato, and Stefan Haufe. Consistency of eeg source localization and connectivity estimates. *Neuroimage*, 152:590–601, 2017.

Nadia Mammone, Cosimo Ieracitano, and Francesco C Morabito. A deep cnn approach to decode motor preparation of upper limbs from time–frequency maps of eeg signals at source level. *Neural networks*, 124:357–372, 2020.

Christoph M Michel and Denis Brunet. Eeg source imaging: a practical review of the analysis steps. *Frontiers in neurology*, 10:325, 2019.

Patrick Ofner, Andreas Schwarz, Joana Pereira, and Gernot R Müller-Putz. Upper limb movements can be decoded from the time-domain of low-frequency eeg. *PloS one*, 12(8):e0182578, 2017.

Roberto Domingo Pascual-Marqui et al. Standardized low-resolution brain electromagnetic tomography (sloreta): technical details. *Methods find exp clin pharmacol*, 24(Suppl D):5–12, 2002.

Gert Pfurtscheller and FH Lopes Da Silva. Event-related eeg/meg synchronization and desynchronization: basic principles. *Clinical neurophysiology*, 110(11):1842–1857, 1999.

Herbert Ramoser, Johannes Muller-Gerking, and Gert Pfurtscheller. Optimal spatial filtering of single trial eeg during imagined hand movement. *IEEE transactions on rehabilitation engineering*, 8(4):441–446, 2000.

Yannick Roy, Hubert Banville, Isabela Albuquerque, Alexandre Gramfort, Tiago H Falk, and Jocelyn Faubert. Deep learning-based electroencephalography analysis: a systematic review. *Journal of neural engineering*, 16(5):051001, 2019.

Gerwin Schalk, Dennis J McFarland, Thilo Hinterberger, Niels Birbaumer, and Jonathan R Wolpaw. Bci2000: a general-purpose brain-computer interface (bci) system. *IEEE Transactions on biomedical engineering*, 51(6):1034–1043, 2004.

Robin Tibor Schirrmeister, Jost Tobias Springenberg, Lukas Dominique Josef Fiederer, Martin Glasstetter, Katharina Eggensperger, Michael Tangermann, Frank Hutter, Wolfram Burgard, and Tonio Ball. Deep learning with convolutional neural networks for eeg decoding and visualization. *Human brain mapping*, 38(11):5391–5420, 2017.

David I Shuman, Sunil K Narang, Pascal Frossard, Antonio Ortega, and Pierre Vandergheynst. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine*, 30(3):83–98, 2013.

Jake Snell, Kevin Swersky, and Richard Zemel. Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.

Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. Eeg conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2022.

Daniel Strohmeier, Yousra Bekhti, Jens Haueisen, and Alexandre Gramfort. The iterative reweighted mixed-norm estimate for spatio-temporal meg/eeg source reconstruction. *IEEE transactions on medical imaging*, 35(10):2218–2228, 2016.

Masashi Sugiyama. Dimensionality reduction of multimodal labeled data by local fisher discriminant analysis. *Journal of machine learning research*, 8(5), 2007.

Rui Sun, Abbas Sohrabpour, Gregory A Worrell, and Bin He. Deep neural networks constrained by neural mass models improve electrophysiological source imaging of spatiotemporal brain dynamics. *Proceedings of the National Academy of Sciences*, 119(31):e2201128119, 2022.

Xingliang Tang and Xianrui Zhang. Conditional adversarial domain adaptation neural network for motor imagery eeg decoding. *Entropy*, 22(1):96, 2020.

Michael E. Tipping. Sparse bayesian learning and the relevance vector machine. *Journal of Machine Learning Research*, 1:211–244, 2001.

Antonios Tragoudaras, Charalampos Antoniadis, and Yehia Massoud. Tinyml for eeg decoding on microcontrollers. In *2023 IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 1–5. IEEE, 2023.

Barry D Van Veen, Wim Van Drongelen, Moshe Yuchtman, and Akifumi Suzuki. Localization of brain electrical activity via linearly constrained minimum variance spatial filtering. *IEEE Transactions on biomedical engineering*, 44(9):867–880, 1997.

David Wipf and Srikantan Nagarajan. A unified bayesian framework for meg/eeg source imaging. *NeuroImage*, 44(3):947–966, 2009.

Jonathan R Wolpaw, Niels Birbaumer, Dennis J McFarland, Gert Pfurtscheller, and Theresa M Vaughan. Brain–computer interfaces for communication and control. *Clinical neurophysiology*, 113(6):767–791, 2002.

Dongrui Wu, Yifan Xu, and Bao-Liang Lu. Transfer learning for eeg-based brain–computer interfaces: A review of progress made since 2016. *IEEE Transactions on Cognitive and Developmental Systems*, 14(1):4–19, 2020.

Kai Zhou, Aierken Haimudula, and Wanying Tang. Dual-branch convolution network with efficient channel attention for eeg-based motor imagery classification. *IEEE Access*, 12:74930–74943, 2024.

AI Use Statement

Generative AI tools were used only for limited writing support, including grammar checking, minor wording and typing assistance, and suggestions for organizing parts of the manuscript. They were not used to develop the methodology, write or debug the experimental code, generate data, produce numerical results, or perform the scientific analysis. All technical content, experiments, results, and conclusions were produced and verified by the authors, who take full responsibility for the final manuscript.

Ethics Statement

This work uses only previously collected, de-identified EEG datasets and involves no new human-subject recruitment or intervention. Experiments were conducted on the upper-limb motor execution and motor imagery dataset of Ofner et al. (2017) and the public BCI2000 motor-imagery dataset (Schalk et al., 2004). All evaluations preserve strict subject separation under LOSO validation, and no held-out-subject information is used during supervised representation learning. ESICM is intended solely as a research framework for noninvasive brain–computer interfaces.

Reproducibility Statement

The paper describes the complete ESICM formulation, preprocessing, source-space construction, nested LOSO protocol, ablations, and embedded deployment settings. Experiments on both the Ofner upper-limb dataset (Ofner et al., 2017) and BCI2000 (Schalk et al., 2004) use fixed subject-independent evaluation procedures. The released implementation will include data preparation, ESICM training and evaluation, baseline comparisons, ablations, ONNX export, Jetson benchmarking, configuration files, and random seeds required to reproduce the reported results.

A Appendix

Additional analysis of subject-independent motor decoding. The appendix expands the main results with analyses that are difficult to include in the primary paper but are important for interpreting ESICM under strict cross-subject evaluation. Table 5 reports the complete set of sensor-space, conventional source-space, and learned cortical-mode baselines, allowing the contribution of the representation itself to be separated from the choice of decoder. Table 6 complements these aggregate results with subject-wise paired comparisons over the same 15 LOSO folds, showing how consistently the observed differences persist across unseen individuals rather than only at the level of mean accuracy. Table 7 examines the representation directly by measuring cross-fold subspace stability, matched-mode reproducibility, and the concentration of learned activity within motor and premotor cortex. Finally, Table 8 evaluates the same framework on the independent BCI2000 motor-imagery benchmark, testing whether the learned cortical representation transfers beyond the upper-limb dataset used in the main experiments.

These analyses are especially relevant because fine-grained LOSO motor decoding is fundamentally different from within-subject classification. The model must distinguish closely related movements in a subject whose EEG is absent from both representation learning and decoder training, while coping with substantial inter-subject variation in anatomy, signal amplitude, spectral structure, and movement timing. The difficulty increases further for six- and seven-class settings, where neighboring upper-limb actions recruit overlapping sensorimotor regions, and for motor imagery, where the neural response is weaker and less temporally constrained than during execution. The appendix therefore tests more than whether ESICM improves classification accuracy. It asks whether the same cortical coordinate system remains useful across subjects, reproducible across folds, anatomically meaningful, and transferable to a second dataset. This is central to the novelty of ESICM: the method does not treat source localization, cross-subject alignment, and decoding as separate stages, but learns a compact cortical representation in which all three are coupled.

A.1 Controlled Representation–Decoder Comparison

Table 8 isolates the effect of the representation by pairing the same family of decoders with sensor EEG, MNE, sLORETA, and LCMV features. All configurations use the same data partition and evaluation pipeline. This distinction is important: if the gain came mainly from decoder capacity, replacing the representation while retaining a strong CNN or compact depthwise network should recover similar performance. It does not. The strongest sensor model reaches 58.71% and the strongest fixed-source model reaches 60.06%, compared with 66.88% for ESICM–Tiny.

The result also separates ESICM from a conventional "source space + classifier" pipeline. MNE, sLORETA, and LCMV first reconstruct a high-dimensional cortical field and leave the classifier to discover which components remain useful across subjects. ESICM learns those coordinates directly from the training subjects. Event timing, forward-model structure, cortical geometry, and inter-subject variability therefore shape the space in which classification is performed. The improvement appears before the final decision layer rather than through a larger decoder.

Published results on EEGMMIDB require care because "four-class classification" does not imply the same generalization problem. Lun et al. (Lun et al., 2022) reconstruct cortical activity with BEM/WMNE and classify left fist, right fist, both fists, and both feet using a dual CNN. Their 96.36% group-level result is obtained after splitting the trials of each subject into training and test portions and then pooling those portions across subjects. Subjects are therefore represented on both sides of the split. The experiment demonstrates strong trial discrimination, but it does not measure performance on a person absent from training.

Table 5: Cross-subject representation–decoder comparison on Four+Rest ME. Mean ± standard deviation (%) across the same 15 LOSO subjects. EEGConformer tests a modern attention-based decoder, while EEG-DANN directly tests subject-adversarial invariance. ESICM-Tiny operates on the learned compact cortical coordinates.

Representation	Decoder	N	Accuracy (%)	Macro-F1 (%)
Sensor	LDA	15	30.03 ± 4.10	26.05 ± 5.06
	MDM	15	28.29 ± 3.72	21.43 ± 5.63
	CNN	15	55.74 ± 9.65	54.11 ± 10.15
	EEGNet	15	50.83 ± 9.46	49.03 ± 9.79
	Transformer	15	41.90 ± 6.46	40.15 ± 6.07
	TinyDepthwiseCNN	15	50.45 ± 8.72	48.86 ± 8.83
	EEGConformer	15	47.95 ± 10.34	45.90 ± 10.95
	EEG-DANN	15	34.21 ± 5.35	31.32 ± 5.64
MNE	LDA	15	31.27 ± 5.24	25.91 ± 6.27
	MDM	15	31.61 ± 4.03	25.12 ± 7.25
	CNN	15	49.41 ± 8.23	48.35 ± 8.04
	EEGNet	15	49.61 ± 7.20	48.25 ± 7.15
	Transformer	15	37.87 ± 4.96	36.45 ± 3.81
	TinyDepthwiseCNN	15	46.78 ± 6.65	45.75 ± 6.94
sLORETA	LDA	15	31.61 ± 4.78	28.42 ± 6.01
	MDM	15	27.61 ± 3.76	21.78 ± 6.25
	CNN	15	49.30 ± 7.15	48.23 ± 6.93
	EEGNet	15	49.72 ± 7.23	48.32 ± 7.29
	Transformer	15	38.09 ± 4.75	36.87 ± 3.83
	TinyDepthwiseCNN	15	46.78 ± 6.65	45.75 ± 6.94
LCMV	LDA	15	35.28 ± 7.53	31.25 ± 8.15
	MDM	15	28.58 ± 3.98	20.02 ± 6.86
	CNN	15	52.84 ± 9.52	51.49 ± 9.71
	EEGNet	15	51.53 ± 8.72	49.88 ± 8.88
	Transformer	15	40.10 ± 6.54	38.13 ± 8.01
	TinyDepthwiseCNN	15	49.83 ± 10.20	48.01 ± 10.27
ESICM	**ESICM-Tiny**	15	**58.05 ± 9.16**	**60.95 ± 8.43**

Zhou et al. (Zhou et al., 2024) report 69.58% with DB-ATCNet on the same four PhysioNet motor-imagery classes. Their PhysioNet experiment uses random 10-fold partitioning of the data, with 90% used for training and 10% for testing in each fold. Although the study describes this setting as subject-independent, the reported split is defined over the data rather than explicitly by held-out subject identity. Its number is consequently not directly equivalent to strict LOSO evaluation.

Lim et al. (Lim et al., 2025) address this issue explicitly. AGTCNet evaluates unseen groups of subjects using a subject-new (SN) framework and reports $65.44 \pm 2.81\%$ accuracy and $\kappa = 0.54 \pm 0.04$ for four-class EEGMMIDB decoding. This is the closest published comparison among the three because the training and evaluation subjects are disjoint. ESICM reaches $66.88 \pm 14.45\%$ under strict LOSO, but the difference should still not be interpreted as a direct state-of-the-art margin because the fold construction, preprocessing, and optimization procedures are different.

The comparison is more useful at the level of modeling assumptions. Lun et al. (Lun et al., 2022) show that reconstructed cortical signals can be highly discriminative when subjects are shared across training and test data. Zhou et al. (Zhou et al., 2024) pursue generalization through a substantially more expressive sensor-space architecture combining dual-branch convolution, channel attention, self-attention, and temporal fusion. Lim et al. (Lim et al., 2025) encode scalp topology directly through graph-temporal learning. ESICM takes a different route. It does not ask the decoder to remove inter-subject variability after features have already been constructed. Instead, inter-subject stability is part of the definition of the cortical representation itself.

This distinction explains why the controlled result in Table 8 is more informative than a raw cross-paper accuracy ranking. Increasing decoder capacity does not close the gap: the CNN and Transformer baselines remain below ESICM in sensor space and after each fixed inverse solution. Nor is conventional source localization alone sufficient: the strongest sLORETA configuration trails ESICM by 6.82 points in accuracy and 7.60 points in macro-F1. The gain appears only when the cortical coordinates are learned together with the constraints required for transfer.

Table 6: Paired cross-subject analysis for Four+Rest motor execution over the same 15 LOSO folds. ΔAcc. denotes the mean subject-wise accuracy difference (ESICM minus baseline). p_{Holm} is the two-sided Wilcoxon signed-rank p-value after Holm correction across the 16 planned comparisons, and r_{rb} is the paired rank-biserial effect size. W/T/L gives the number of held-out subjects on which ESICM wins, ties, or loses relative to each baseline.

Comparator	Baseline Acc. (%)	ΔAcc. (pp)	p_{Holm}	r_{rb}	W/T/L
Sensor Space					
Sensor + CNN	55.74 ± 9.65	$+2.31$	0.021545	$+0.667$	12/0/3
Sensor + EEGNet	50.83 ± 9.46	$+7.23$	0.000977	$+1.000$	14/1/0
Sensor + TinyDepthwiseCNN	50.45 ± 8.72	$+7.61$	0.001221	$+0.950$	14/0/1
Sensor + Transformer	41.90 ± 6.46	$+16.16$	0.000977	$+1.000$	15/0/0
MNE Source Space					
MNE + CNN	49.41 ± 8.23	$+8.65$	0.001221	$+0.950$	13/0/2
MNE + EEGNet	49.61 ± 7.20	$+8.44$	0.000977	$+1.000$	15/0/0
MNE + TinyDepthwiseCNN	46.78 ± 6.65	$+11.27$	0.000977	$+1.000$	15/0/0
MNE + Transformer	37.87 ± 4.96	$+20.18$	0.000977	$+1.000$	15/0/0
sLORETA Source Space					
sLORETA + CNN	49.30 ± 7.15	$+8.75$	0.001221	$+0.950$	13/0/2
sLORETA + EEGNet	49.72 ± 7.23	$+8.33$	0.000977	$+1.000$	15/0/0
sLORETA + TinyDepthwiseCNN	46.78 ± 6.65	$+11.27$	0.000977	$+1.000$	15/0/0
sLORETA + Transformer	38.09 ± 4.75	$+19.96$	0.000977	$+1.000$	15/0/0
LCMV Source Space					
LCMV + CNN	52.84 ± 9.52	$+5.22$	0.000977	$+1.000$	15/0/0
LCMV + EEGNet	51.53 ± 8.72	$+6.52$	0.000977	$+1.000$	15/0/0
LCMV + TinyDepthwiseCNN	49.83 ± 10.20	$+8.23$	0.000977	$+0.967$	14/0/1
LCMV + Transformer	40.10 ± 6.54	$+17.96$	0.000977	$+1.000$	15/0/0

Table 7: Cross-fold reproducibility of the learned cortical coordinate system over 15 LOSO folds. Subspace agreement is invariant to mode sign and ordering; matched-mode agreement measures spatial consistency after mode matching; motor concentration denotes the fraction of cortical-mode energy within motor and premotor regions.

Band	Subspace stability	Matched-mode agreement	Motor/premotor concentration
Broad	0.935	0.840	0.379
MRCP	0.915	0.826	0.390
Mu	0.968	0.915	0.473
Beta	0.942	0.884	0.457

BCI2000 therefore serves a different purpose from the upper-limb experiments in the main paper. It tests whether the same representation principle survives a change in dataset, subject population, and motor paradigm. The result suggests that ESICM is not tied to one collection of upper-limb movements or to one decoder. Its useful object is the learned coordinate system itself: a small cortical space in which movement structure remains separable when the person changes.

A.2 Limitations

ESICM is evaluated on two independent EEG datasets, but broader validation across larger cohorts, additional montages, and clinical populations would further establish its generality. The learned modes are discriminative cortical coordinates constrained by biophysics and anatomy rather than direct source estimates, and the current experiments focus on event-aligned motor paradigms. Future work will extend the framework to asynchronous decoding and fully closed-loop BCI operation.

Table 8: Controlled comparison on the four-class BCI2000 motor-imagery task (left fist, right fist, both fists, and both feet). All configurations use the same subject-disjoint evaluation, preprocessing, and class definition. Accuracy and macro-F1 are reported as mean ± standard deviation (%); κ is reported on a 0–100 scale.

Representation	Decoder	Accuracy (%) ↑	Macro-F1 (%) ↑	κ (%) ↑	Params. ↓
Sensor	LogCov-LDA	33.39 ± 8.40	28.13 ± 10.81	11.18	0
Sensor	Riemannian-MDM	25.96 ± 3.00	14.41 ± 4.70	1.28	0
Sensor	CNN	58.34 ± 14.82	57.11 ± 15.52	44.45	50,844
Sensor	EEGNet	56.19 ± 13.39	55.47 ± 13.60	41.58	2,364
Sensor	Transformer	55.64 ± 12.48	53.70 ± 13.40	40.85	72,100
Sensor	TinyDepthwiseCNN	58.71 ± 7.44	57.75 ± 7.51	44.95	2,452
MNE	LogCov-LDA	28.89 ± 3.21	24.18 ± 4.27	5.18	0
MNE	Riemannian-MDM	26.62 ± 5.14	15.79 ± 7.54	2.17	0
MNE	CNN	57.44 ± 16.32	55.24 ± 17.13	43.25	72,348
MNE	EEGNet	54.92 ± 9.01	53.46 ± 9.31	39.90	2,812
MNE	Transformer	52.70 ± 11.14	51.40 ± 11.51	36.93	84,196
MNE	TinyDepthwiseCNN	54.14 ± 10.51	52.80 ± 10.39	38.85	3,152
sLORETA	LogCov-LDA	29.50 ± 5.81	24.03 ± 6.21	6.00	0
sLORETA	Riemannian-MDM	25.11 ± 3.68	15.70 ± 5.25	0.15	0
sLORETA	CNN	60.06 ± 14.73	58.18 ± 15.93	46.75	72,348
sLORETA	EEGNet	56.62 ± 9.07	55.22 ± 8.78	42.16	2,812
sLORETA	Transformer	50.76 ± 11.08	48.96 ± 11.54	34.35	84,196
sLORETA	TinyDepthwiseCNN	51.72 ± 9.89	50.50 ± 9.69	35.63	3,152
LCMV	LogCov-LDA	37.33 ± 9.03	33.93 ± 9.74	16.44	0
LCMV	Riemannian-MDM	26.00 ± 3.10	13.18 ± 5.58	1.33	0
LCMV	CNN	51.18 ± 14.03	49.29 ± 15.18	34.91	72,348
LCMV	EEGNet	47.27 ± 10.80	44.77 ± 10.52	29.70	2,812
LCMV	Transformer	48.39 ± 14.96	46.18 ± 15.98	31.19	84,196
LCMV	TinyDepthwiseCNN	49.82 ± 11.82	47.42 ± 11.79	33.09	3,152
Learned cortical modes	ESICM–Tiny	$\mathbf{66.88 \pm 14.45}$	$\mathbf{65.78 \pm 14.97}$	**52.51**	15,618

Table 9: Protocol-aware context for four-class motor-imagery results on the PhysioNet EEGMMIDB dataset. All studies use left fist, right fist, both fists, and both feet. Reported accuracies are shown for context, not as a common leaderboard, because subject separation differs across evaluation protocols.

Study	Representation / model	Subjects	Evaluation protocol	Accuracy (%)
Lun et al. (Lun et al., 2022)	WMNE cortical ROIs + dual CNN	10	Per-subject 90/10 trial split, then pooled	96.36
Zhou et al. (Zhou et al., 2024)	Sensor EEG + DB-ATCNet	103	Random 10-fold data partition	69.58
Lim et al. (Lim et al., 2025)	Sensor graph + AGTCNet	103	Disjoint-subject SN / LMSO evaluation	65.44 ± 2.81
ESICM (ours)	Learned cortical modes + ESICM–Tiny	106	Strict subject-disjoint LOSO	66.88 ± 14.45