A MULTI-SCALE SUBMANIFOLD-AWARE SPD ATTENTION NETWORK FOR EEG DECODING

Anonymous authors
Paper under double-blind review

ABSTRACT

Riemannian attention has become an effective approach for electroencephalography (EEG) decoding by modeling temporal dependencies among covariance representations on the Symmetric Positive Definite (SPD) manifold. However, existing methods typically perform attention only in a single fixed, full-dimensional SPD space, where redundant covariance directions may dilute discriminative temporal relations. To address this limitation, we propose a multi-scale Submanifold-aware SPD Attention network (SSAtt) that learns compact SPD representations for explicitly capturing temporal interactions. SSAtt projects each temporal covariance descriptor into one or more lower-dimensional SPD spaces through sample-adaptive congruence transformations. Each transformation combines a learnable base transformation with a hypernetwork-generated dynamic component, while Newton-Schulz orthogonalization promotes semi-orthogonal projection matrices and thereby preserves the positive definiteness of the projected representations. Riemannian attention is then performed in parallel on the original SPD manifold and the learned compact SPD spaces, allowing full-dimensional covariance information and compact discriminative structures to be jointly exploited. Extensive experiments on four EEG benchmarks demonstrate the effectiveness of SSAtt.

1 INTRODUCTION

Electroencephalography (EEG) has been widely used in brain-computer interface (BCI) systems (Teplan, 2002; Iturrate et al., 2009; Subha et al., 2010), particularly for motor imagery (MI) (Ang et al., 2015; Alonso-Valerdi et al., 2015), steady-state visual evoked potential (SSVEP) (Al-Qaysi et al., 2018; Wang et al., 2016; Chen et al., 2015), and error-related potential (ERN) decoding (Lotte et al., 2007). However, prevalent EEG decoding methods typically rely on frequency or spatial filtering and spatiotemporal feature extraction for covariance-based modeling and classification (Ang et al., 2008; Barachant et al., 2010; 2011; Schirrmeister et al., 2017; Lawhern et al., 2018; Wei et al., 2019; Congedo et al., 2017; Lotte et al., 2018; Yger et al., 2016). Although covariance representation encodes second-order statistics of EEG signals, directly using it for classification limits the capability to learn geometry-aware spatiotemporal representations.

Recognizing this limitation, recent research (Gulcehre et al., 2018; Pan et al., 2022; Wang et al., 2024; Hu et al., 2025; Wang et al., 2026) has introduced attention mechanisms into manifold-valued EEG representations to capture long-range dependencies and spatiotemporal fluctuations inherent in EEG signals. Pan et al. (2022) perform self-attention directly on temporally segmented EEG covariance matrices under the Log-Euclidean metric. Wang et al. (2024) extends self-attention to the Grassmannian manifold by representing sequential features as linear subspaces. Hu et al. (2025) instead employs full-rank correlation matrices and develops attention under Off-Log and Log-Scaled geometries. More recently, Wang et al. (2026) formulates a general attention framework on gyrovector spaces and accommodates several matrix manifolds.

However, as illustrated in Fig. 1, existing manifold attention methods generally operate in the original, full-dimensional SPD space, where all covariance directions jointly contribute to query-key affinity. Yet, not all covariance directions are equally informative for EEG decoding: task-relevant channel interactions may coexist with redundant, noisy, or subject-dependent variations. Such heterogeneous directions can perturb geometric affinity estimation and affect the resulting attention weights. Therefore, the key challenge is to construct compact geometric representations that fo-

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.

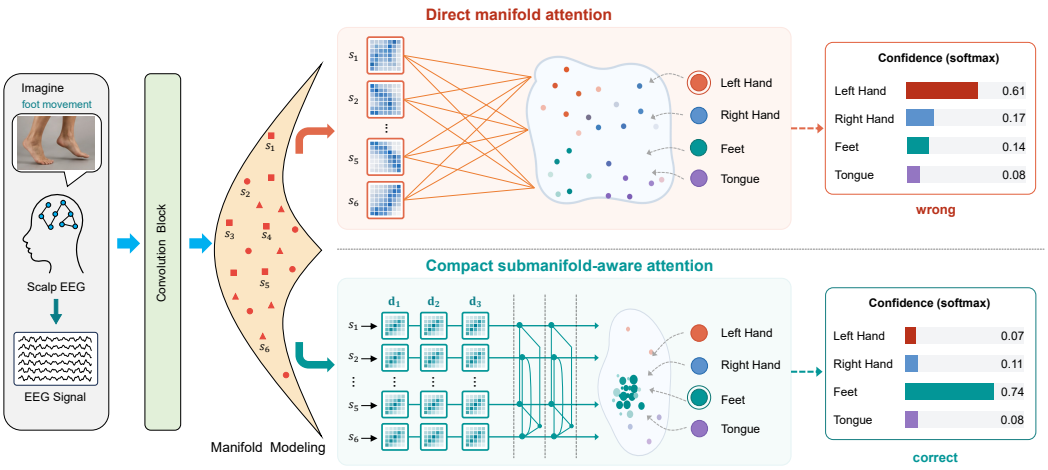


Figure 1: **Upper:** Existing manifold attention operates in a single full-dimensional SPD space. **Lower:** SSAtt additionally constructs multiple compact SPD spaces and performs Riemannian attention independently in each space.

cus on task-relevant and subject-relevant structures while retaining the intrinsic geometry of SPD data. This motivates compact, lower-dimensional SPD representations as complementary spaces for manifold attention (Harandi et al., 2014).

To address this issue, we propose **SSAtt**, a multi-scale submanifold-aware SPD attention network that jointly performs Riemannian attention on the original and multiple compact SPD representations. Rather than simply reducing the dimension before attention, SSAtt learns sample-adaptive transformations to construct compact SPD spaces at multiple scales and performs attention independently within each space under the Log-Euclidean metric. The sample-adaptive transformation combines a learnable base transformation with an input-conditioned dynamic component and is constrained through Newton-Schulz orthogonalization to maintain SPD representations. Our main contributions are summarized as follows:

- **Multi-scale submanifold-aware SPD attention.** We propose a framework that performs Riemannian attention on both the original and multiple learned compact SPD spaces, combining full-dimensional covariance information with compact discriminative structures.
- **Sample-adaptive transformation mechanism.** We introduce a mechanism that combines a learnable base projection with a trial-conditioned dynamic component to construct compact SPD representations adapted to each input covariance structure.
- **Comprehensive empirical validation.** Extensive experiments and ablation studies across four EEG benchmarks validate the effectiveness of SSAtt, while visualization and interpretability analyses reveal its robustness and physiologically meaningful patterns.

2 Preliminary

In this section, we give a brief introduction to the Riemannian geometry on SPD manifold and the basic operation layers used in SPD neural networks. For more in-depth discussions, please refer to Lee (2006); Arsigny et al. (2007); Huang & Van Gool (2017); Gallier & Quaintance (2020).

2.1 Riemannian Geometry on SPD Manifold

SPD Manifold. Let $\mathcal{S}^n_{++} \triangleq \{\mathbf{X} \in \mathbb{R}^{n\times n} \mid \mathbf{X} = \mathbf{X}^T,\ \boldsymbol{v}^T\mathbf{X}\boldsymbol{v} > 0,\ \forall\, \boldsymbol{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}\}$ denote the set of $n \times n$ real symmetric positive definite matrices. Endowed with a Riemannian metric, $\mathcal{S}^n_{++}$ forms the SPD manifold, whose tangent space at any $\mathbf{X} \in \mathcal{S}^n_{++}$ can be identified with $T_{\mathbf{X}}\mathcal{S}^n_{++} \simeq \mathcal{S}^n$, where $\mathcal{S}^n \triangleq \{\mathbf{V} \in \mathbb{R}^{n\times n} \mid \mathbf{V} = \mathbf{V}^T\}$.

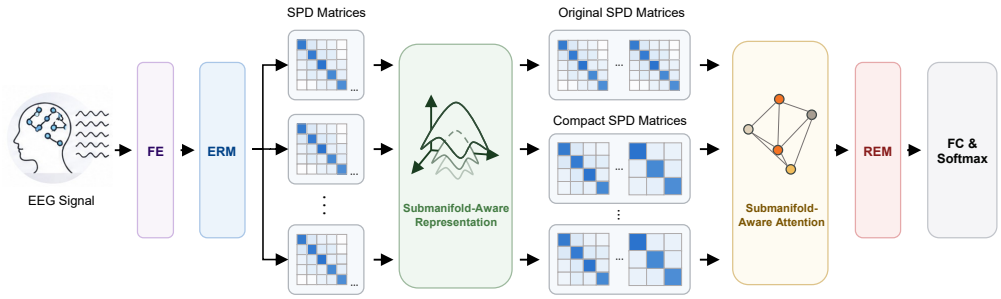


Figure 2: Overview of SSAtt. FE module extracts EEG features, and ERM module forms the original SPD sequence. Sample-adaptive transformations create compact SPD sequences. Attention operates independently in each space. REM module maps the outputs to Euclidean features for classification.

Log-Euclidean Geometry. For $\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T \in \mathcal{S}_{++}^n$, a matrix function is defined as $f(\mathbf{X}) \triangleq \mathbf{U}f(\mathbf{\Lambda})\mathbf{U}^T$, where $f(\mathbf{\Lambda}) = \mathrm{diag}(f(\lambda_1), \ldots, f(\lambda_n))$. The Log-Euclidean geometry maps SPD matrices to the vector space of symmetric matrices through the matrix logarithm. For $\mathbf{X}, \mathbf{Y} \in \mathcal{S}_{++}^n$ and $\mathbf{V} \in T_{\mathbf{X}}\mathcal{S}_{++}^n$, the logarithmic and exponential maps are respectively defined as:

$$\mathrm{Log}_{\mathbf{X}}(\mathbf{Y}) \triangleq D_{\log(\mathbf{X})} \exp \cdot \big(\log(\mathbf{Y}) - \log(\mathbf{X})\big), \tag{1}$$

$$\mathrm{Exp}_{\mathbf{X}}(\mathbf{V}) \triangleq \exp\big(\log(\mathbf{X}) + D_{\mathbf{X}} \log \cdot \mathbf{V}\big), \tag{2}$$

where $D_{\log(\mathbf{X})} \exp \cdot$ and $D_{\mathbf{X}} \log \cdot$ denote the differentials of the matrix exponential and logarithm, respectively. The corresponding Log-Euclidean distance is given below (Huang et al., 2015):

$$d_{\mathrm{LE}}(\mathbf{X}, \mathbf{Y}) \triangleq \|\mathrm{Log}_{\mathbf{I}}(\mathbf{X}) - \mathrm{Log}_{\mathbf{I}}(\mathbf{Y})\|_{\mathrm{F}}\,. \tag{3}$$

where $\mathbf{I}$ is the identity matrix and $\|\cdot\|_{\mathrm{F}}$ is the Frobenius norm.

Weighted Fréchet Mean. Given a set of SPD matrices $\{\mathbf{X}_i\}_{i=1}^m$ with weights $\{\alpha_i\}_{i=1}^m$ satisfying $\alpha_i > 0$ and $\sum_{i=1}^m \alpha_i = 1$, their weighted Fréchet mean under the LEM has the closed-form solution:

$$\mathbf{G}_{\mathrm{LE}} = \exp\left(\sum_{i=1}^{m} \alpha_i \log(\mathbf{X}_i)\right). \tag{4}$$

where $\log(\cdot)$ and $\exp(\cdot)$ denote the matrix logarithm and matrix exponential, respectively.

2.2 Basic Layers of SPD Networks

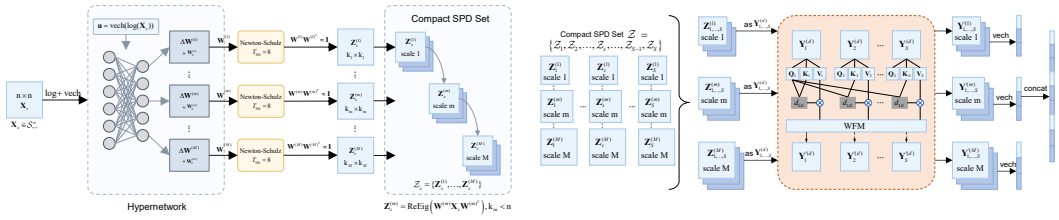
BiMap. Given $\mathbf{X}_{k-1} \in \mathcal{S}_{++}^{n_{k-1}}$, the bilinear mapping $\mathbf{X}_k = \mathbf{W}_k\mathbf{X}_{k-1}\mathbf{W}_k^T$ projects it to a lower-dimensional SPD space, where $\mathbf{W}_k \in \mathbb{R}^{n_k \times n_{k-1}}$ is full-row-rank satisfying semi-orthogonality. **ReEig.** Given the eigendecomposition $\mathbf{X}_{k-1} = \mathbf{U}_{k-1}\mathbf{\Sigma}_{k-1}\mathbf{U}_{k-1}^T$, ReEig rectifies small eigenvalues as $\mathbf{X}_k = \mathbf{U}_{k-1}\max(\mathbf{\Sigma}_{k-1}, \epsilon\mathbf{I})\mathbf{U}_{k-1}^T$. **LogEig.** LogEig maps an SPD matrix to the tangent space at the identity as $\mathbf{X}_k = \mathbf{U}_{k-1}\log(\mathbf{\Sigma}_{k-1})\mathbf{U}_{k-1}^T$ for subsequent Euclidean processing..

3 Methodology

In this section, we provide the technical details of the proposed SSAtt. As illustrated in Fig. 2, SSAtt augments the original SPD representation with compact SPD spaces for multi-scale attention computation. Specifically, Sec. 3.1 describes the submanifold-aware representation module. Sec. 3.2 presents the submanifold-aware SPD attention mechanism, and Sec. 3.3 details the complete network architecture and training objective.

3.1 Submanifold-Aware Representation

Let $\mathcal{X} = \{\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_S\}, \mathbf{X}_s \in \mathcal{S}_{++}^n$ denote the SPD matrix sequence of an EEG trial, where S is the number of temporal segments and n is the dimension of the original SPD space. As illustrated in Fig. 3a, the submanifold-aware representation module constructs one or more compact SPD



(a) (b)

Figure 3: Detailed illustration of the key modules in SSAtt. (a) Submanifold-aware representation module, where a trial-level log-domain descriptor conditions the dynamic projection components. These components are combined with a learnable base projection and orthogonalized by Newton-Schulz iteration to construct multi-scale compact SPD representations. (b) Submanifold-aware attention module, where attention is independently performed on the original and compact SPD spaces under the LEM, followed by Euclidean mapping and concatenation.

spaces through scale-specific sample-adaptive transformations. Formally, the resulting multi-scale compact SPD set is defined as:

$$\mathcal{Z}_s = \left\{ \mathbf{Z}_s^{(m)} \mid \mathbf{Z}_s^{(m)} = \mathrm{ReEig}\left(\mathbf{W}^{(m)} \mathbf{X}_s \mathbf{W}^{(m)^T} \right),\ \mathbf{Z}_s^{(m)} \in \mathcal{S}_{++}^{k_m} \right\}_{m=1}^{M}, \tag{5}$$

where $k_m < n$. Here, $\mathbf{W}^{(m)} \in \mathbb{R}^{k_m \times n}$ denotes the transformation matrix for the m-th scale. The same $\mathbf{W}^{(m)}$ is shared across all temporal segments within the EEG trial, while different scales employ distinct transformation matrices and output dimensions. The ReEig operation rectifies small eigenvalues after the congruence transformation to improve the numerical stability.

To obtain sample-adaptive transformations, each scale-specific raw projection matrix is decomposed into a learnable base component and an input-conditioned dynamic component:

$$\widetilde{\mathbf{W}}^{(m)} = \mathbf{W}_b^{(m)} + \Delta\mathbf{W}^{(m)}, \qquad \mathbf{W}_b^{(m)}, \Delta\mathbf{W}^{(m)} \in \mathbb{R}^{k_m \times n}. \tag{6}$$

The base component $\mathbf{W}_b^{(m)}$ is initialized as a random semi-orthogonal matrix and learned as a shared projection across samples. The dynamic component $\Delta\mathbf{W}^{(m)}$ is generated by a lightweight hypernetwork conditioned on the log-domain descriptor $\mathbf{u} = \mathrm{vech}(\log(\mathbf{X})) \in \mathbb{R}^{\frac{n(n+1)}{2}}$:

$$\Delta\mathbf{W}^{(m)} = \mathrm{Hypernetwork}(\mathbf{u}) = \mathrm{reshape}(f(\mathbf{u})), \tag{7}$$

where $f(\cdot)$ denotes a neural network. Thereby, $\mathbf{W}_b^{(m)}$ captures shared projection parameters, while $\Delta\mathbf{W}^{(m)}$ models sample-specific variation. For multiple scales, the same construction is applied with different output dimensions $k_1, k_2, \ldots, k_M$.

Since $\mathbf{W}^{(m)}$ is required to satisfy the semi-orthogonality constraint in Eq. (5), while $\widetilde{\mathbf{W}}^{(m)}$ does not necessarily satisfy this constraint, we obtain $\mathbf{W}^{(m)}$ from $\widetilde{\mathbf{W}}^{(m)}$ using the differentiable Newton-Schulz iteration:

$$\mathbf{W}_{t+1} = \frac{1}{2}\left(3\mathbf{I} - \mathbf{W}_t \mathbf{W}_t^T\right)\mathbf{W}_t. \tag{8}$$

After T_{NS} iterations, the resulting projection matrix approximately satisfies $\mathbf{W}^{(m)}\mathbf{W}^{(m)T} \approx \mathbf{I}$. In our implementation, we set $T_{\mathrm{NS}} = 8$, and the complete procedure is summarized in Algorithm 1.

3.2 Submanifold-Aware SPD Attention

Given the original SPD sequence and the learned compact SPD sequences, attention is performed independently within each manifold space, as illustrated in Fig. 3b. For a manifold of dimension d, we denote its input sequence as:

$$\mathcal{Y}^{(d)} = \left\{ \mathbf{Y}_1^{(d)}, \mathbf{Y}_2^{(d)}, \ldots, \mathbf{Y}_S^{(d)} \right\}, \qquad \mathbf{Y}_s^{(d)} \in \mathcal{S}_{++}^{d}, \tag{9}$$

where $\mathcal{Y}^{(d)}$ corresponds to either the original SPD sequence or one of the compact SPD sequences.

Algorithm 1 Newton-Schulz Orthogonalization

Require: Raw projection matrix $\widetilde{\mathbf{W}} \in \mathbb{R}^{k\times n}$, iteration number T_{NS}
Ensure: Orthogonalized projection matrix $\mathbf{W}$
1: $\mathbf{W}_0 \leftarrow \widetilde{\mathbf{W}} / \left(\|\widetilde{\mathbf{W}}\|_F + \epsilon\right)$
2: $\mathbf{I} \leftarrow$ identity matrix in $\mathbb{R}^{k\times k}$
3: **for** $t = 0$ to $T_{\mathrm{NS}} - 1$ **do**
4: $\quad \mathbf{W}_{t+1} \leftarrow \frac{1}{2}\left(3\mathbf{I} - \mathbf{W}_t\mathbf{W}_t^T\right)\mathbf{W}_t$
5: **end for**
6: **return** $\mathbf{W}_{T_{\mathrm{NS}}}$

For each $\mathbf{Y}_i^{(d)}$, query, key, and value matrices are generated through BiMap function:

$$\mathbf{Q}_i^{(d)} = \mathbf{W}_Q^{(d)}\mathbf{Y}_i^{(d)}\mathbf{W}_Q^{(d)T}, \quad \mathbf{K}_i^{(d)} = \mathbf{W}_K^{(d)}\mathbf{Y}_i^{(d)}\mathbf{W}_K^{(d)T}, \quad \mathbf{V}_i^{(d)} = \mathbf{W}_V^{(d)}\mathbf{Y}_i^{(d)}\mathbf{W}_V^{(d)T}. \tag{10}$$

where $\mathbf{W}_Q^{(d)}, \mathbf{W}_K^{(d)}, \mathbf{W}_V^{(d)} \in \mathbb{R}^{d\times d}$ are to-be-learned transformation matrices. The query-key affinity is measured by the Log-Euclidean distance:

$$D_{ij}^{(d)} = \left\|\mathrm{Log}_{\mathbf{I}}\left(\mathbf{Q}_i^{(d)}\right) - \mathrm{Log}_{\mathbf{I}}\left(\mathbf{K}_j^{(d)}\right)\right\|_{\mathrm{F}}. \tag{11}$$

The normalized attention weight is defined as:

$$a_{ij}^{(d)} = \frac{\exp\left(e_{ij}^{(d)}\right)}{\sum_{l=1}^{S}\exp\left(e_{il}^{(d)}\right)}, \quad e_{ij}^{(d)} = \frac{1}{1+\log\left(1+D_{ij}^{(d)}\right)}. \tag{12}$$

where $e_{ij}^{(d)}$ is the attention score. The attention output is obtained by the weighted Fréchet mean under the LEM:

$$\mathbf{R}_i^{(d)} = \mathrm{Exp}_{\mathbf{I}}\left(\sum_{j=1}^{S} a_{ij}^{(d)}\,\mathrm{Log}_{\mathbf{I}}\left(\mathbf{V}_j^{(d)}\right)\right). \tag{13}$$

The resulting output sequence is denoted as:

$$\mathcal{R}^{(d)} = \left\{\mathbf{R}_1^{(d)}, \mathbf{R}_2^{(d)}, \ldots, \mathbf{R}_S^{(d)}\right\}. \tag{14}$$

The same attention operation is independently applied to the original SPD space and each compact SPD space, yielding:

$$\mathcal{R} = \left\{\mathcal{R}^{(n)}, \mathcal{R}^{(k_1)}, \ldots, \mathcal{R}^{(k_M)}\right\}. \tag{15}$$

The resulting attention outputs remain separate across SPD spaces and are concatenated only after Riemannian-to-Euclidean mapping.

3.3 End-to-End EEG Decoding Network

As illustrated in Fig. 2, the end-to-end EEG decoding network further includes the following components in addition to the two modules introduced above.

Feature extraction (FE). Given an EEG trial $\mathbf{E} \in \mathbb{R}^{C\times T}$, a feature extraction module $f_{\mathrm{FE}}(\cdot)$ is first employed to obtain the Euclidean feature representation $\mathbf{F} = f_{\mathrm{FE}}(\mathbf{E}) \in \mathbb{R}^{N\times L}$. The specific feature extraction configurations vary across datasets and are detailed in App. A.2.

Euclidean-to-Riemannian mapping (ERM). The feature matrix $\mathbf{F}$ is partitioned into S segments, denoted as $\{\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_S\}$, where $\mathbf{F}_s \in \mathbb{R}^{N\times L_s}$. For each segment, its feature mean is computed as $\bar{\mathbf{F}}_s = \frac{1}{L_s}\sum_{l=1}^{L_s}\mathbf{F}_s(:,l)$, and the covariance matrix is obtained as $\mathbf{C}_s = \frac{(\mathbf{F}_s - \bar{\mathbf{F}}_s)(\mathbf{F}_s - \bar{\mathbf{F}}_s)^T}{L_s - 1}$. We then apply trace normalization and diagonal loading to obtain $\mathbf{X}_s = \frac{\mathbf{C}_s}{\mathrm{tr}(\mathbf{C}_s)} + \epsilon\mathbf{I}$, where $\epsilon > 0$ is a small constant. The resulting matrices form the SPD sequence $\mathcal{X} = \{\mathbf{X}_1, \ldots, \mathbf{X}_S\}$, which is then processed by the submanifold-aware representation and attention modules.

Table 1: The results of different methods on BNCI2014001 and BNCI2015001.

Manifold	Method	BNCI2014001		BNCI2015001	
		Inter-session	Inter-subject	Inter-session	Inter-subject
Euclidean	FBCSP+SVM (Ang et al., 2008)	60.6 ± 4.9	32.3 ± 7.3	81.5 ± 4.4	58.6 ± 13.4
	EEGNet (Lawhern et al., 2018)	41.8 ± 5.8	43.3 ± 17.0	72.4 ± 8.4	59.2 ± 9.5
	ShConvNet (Schirrmeister et al., 2017)	51.3 ± 2.3	42.2 ± 16.2	74.1 ± 4.2	58.7 ± 5.8
	FBCSP+DSS+LDA (Hehenberger et al., 2021)	71.3 ± 1.8	48.3 ± 14.3	84.6 ± 4.8	67.7 ± 14.3
SPD	TSM+SVM (Barachant et al., 2011)	61.8 ± 4.1	34.7 ± 8.6	75.7 ± 5.1	56.0 ± 6.0
	FB+TSM+LR (Kobler et al., 2021)	69.8 ± 4.8	36.5 ± 8.2	80.9 ± 6.0	60.6 ± 10.9
	URPA+MDM (Rodrigues et al., 2018)	59.5 ± 2.7	46.8 ± 14.6	79.2 ± 4.6	70.3 ± 16.1
	SPDOT+TSM+SVM (Yair et al., 2019)	66.8 ± 3.8	38.6 ± 8.6	77.5 ± 2.9	63.3 ± 8.1
	TSMNet (Kobler et al., 2022)	69.0 ± 3.6	51.6 ± 16.5	85.8 ± 4.3	77.0 ± 13.7
	Graph-CSPNet (Ju & Guan, 2023)	71.9 ± 13.3	45.2 ± 9.3	79.8 ± 14.6	64.2 ± 13.4
Gr Att	GDLNet (Wang et al., 2024)	58.1 ± 8.9	46.3 ± 5.1	76.9 ± 13.6	63.3 ± 14.2
	GyroAtt-Gr (Wang et al., 2026)	72.5 ± 7.3	52.1 ± 14.2	85.0 ± 7.7	75.3 ± 13.7
SPD Att	MAtt (Pan et al., 2022)	66.5 ± 8.9	45.3 ± 11.3	80.8 ± 14.8	63.1 ± 10.1
	GyroAtt-SPD (Wang et al., 2026)	75.3 ± 6.5	52.3 ± 14.1	85.7 ± 5.5	76.6 ± 13.7
	SSAtt	$\mathbf{76.4 \pm 3.4}$	$\mathbf{53.5 \pm 15.7}$	$\mathbf{86.8 \pm 4.1}$	$\mathbf{77.8 \pm 13.4}$

Riemannian-to-Euclidean mapping (REM). After submanifold-aware attention, the resulting SPD output sequences are mapped to the tangent space at the identity using the logarithmic map. For each output $\mathbf{R}_i^{(d)}$, its Euclidean representation is given by $\mathbf{r}_i^{(d)} = \mathrm{vech}\Big(\mathrm{Log}_{\mathbf{I}}(\mathbf{R}_i^{(d)})\Big)$. The representations from the original and compact SPD spaces are then concatenated as $\mathbf{g} = \mathrm{Concat}_{d,i}\Big(\mathbf{r}_i^{(d)}\Big)$.

Classification objective. The final representation $\mathbf{g}$ is mapped to class logits by $\mathbf{z} = \mathbf{W}_c\mathbf{g} + \mathbf{b}_c$, followed by $\mathbf{p} = \mathrm{softmax}(\mathbf{z})$. The network is optimized end-to-end with the cross-entropy loss $\mathcal{L} = -\sum_{c=1}^{C_{\mathrm{cls}}} y_c \log p_c$, where C_{cls} denotes the number of classes, y_c is the ground-truth label indicator, and p_c is the predicted probability for class c.

4 Experiments

Experimental setup. Following prior works (Kobler et al., 2022; Pan et al., 2022; Wang et al., 2026), we evaluate the proposed SSAtt on four widely used EEG decoding benchmarks covering three representative BCI paradigms: Motor Imagery (MI), Steady-State Visual Evoked Potential (SSVEP), and Error-Related Negativity (ERN). Specifically, BNCI2014001 (Tangermann et al., 2012) and BNCI2015001 (Faller et al., 2012) are used for MI decoding, MAMEM-SSVEP-II (Nikolopoulos, 2021) is used for SSVEP decoding, and BCI-ERN (Margaux et al., 2012) is used for ERN decoding. For BNCI2014001 and BNCI2015001, we conduct both inter-session and inter-subject evaluations and use balanced accuracy as the primary metric. For MAMEM-SSVEP-II, we report classification accuracy, while AUC is adopted for BCI-ERN due to its class imbalance. For more details on datasets and the corresponding preprocessing, please refer to App. A.1, and detailed implementation configurations are provided in App. A.2.

Baselines. To provide a comprehensive comparison, we evaluate SSAtt against both Euclidean deep learning methods and geometry-aware deep learning methods. The Euclidean baselines include FBCSP+SVM (Ang et al., 2008), FBCSP+DSS+LDA (Hehenberger et al., 2021), ShallowConvNet (Schirrmeister et al., 2017), EEGNet (Lawhern et al., 2018), SCCNet (Wei et al., 2019), EEG-TCNet (Ingolfsson et al., 2020), TCNet-Fusion (Musallam et al., 2021), FBCNet (Mane et al., 2021), and MBEEGSE (Altuwaijri et al., 2022). The geometry-aware baselines include TSM+SVM (Barachant et al., 2011), FB+TSM+LR (Kobler et al., 2021), URPA+MDM (Rodrigues et al., 2018), SPDOT+TSM+SVM (Yair et al., 2019), TSMNet (Kobler et al., 2022), Graph-CSPNet (Ju & Guan, 2023), SPDNet (Huang & Van Gool, 2017), GrNet (Huang et al., 2018), SPDNetBN (Brooks et al., 2019), MAtt (Pan et al., 2022), GDLNet (Wang et al., 2024), CorAtt (Hu et al., 2025), and GyroAtt (Wang et al., 2026). These baselines cover Euclidean CNN-based EEG modeling, SPD manifold learning, Grassmannian learning, Correlation manifold attention, and Gyrovector space attention.

Table 2: The results of different methods on MAMEM-SSVEP-II and BCI-ERN.

Manifold	Method	MAMEM-SSVEP-II	BCI-ERN
Euclidean	ShallowConvNet (Schirrmeister et al., 2017)	56.9 ± 7.0	71.9 ± 2.6
	EEGNet (Lawhern et al., 2018)	53.7 ± 7.2	74.3 ± 2.5
	SCCNet (Wei et al., 2019)	62.1 ± 7.7	70.9 ± 2.3
	EEG-TCNet (Ingolfsson et al., 2020)	55.5 ± 7.7	77.1 ± 2.5
	TCNet-Fusion (Musallam et al., 2021)	45.0 ± 6.5	70.5 ± 2.9
	FBCNet (Mane et al., 2021)	53.1 ± 5.7	60.5 ± 3.1
	MBEEGSE (Altuwaijri et al., 2022)	56.5 ± 7.3	75.5 ± 2.3
Grassmann	GrNet (Huang et al., 2018)	61.2 ± 3.6	72.2 ± 4.6
	GDLNet (Wang et al., 2024)	65.5 ± 2.9	78.2 ± 2.5
	GyroAtt-Gr (Wang et al., 2026)	67.1 ± 1.6	78.4 ± 1.4
Correlation	CorAtt (Hu et al., 2025)	67.4 ± 3.2	78.8 ± 3.4
SPD	SPDNet (Huang & Van Gool, 2017)	62.3 ± 3.1	72.1 ± 4.4
	SPDNetBN (Brooks et al., 2019)	62.8 ± 3.0	72.3 ± 3.5
	MAtt (Pan et al., 2022)	65.2 ± 3.1	75.7 ± 2.2
	GyroAtt-SPD (Wang et al., 2026)	66.2 ± 2.5	76.1 ± 4.2
	SSAtt	**68.8 ± 1.6**	**79.4 ± 2.5**

Table 3: Running time (s/epoch) comparison on four datasets.

Method	BNCI2014001		BNCI2015001		MAMEM-SSVEP-II	BCI-ERN
	Inter-session	Inter-subject	Inter-session	Inter-subject		
MAtt	8.08 ± 0.07	180.87 ± 1.27	6.85 ± 0.03	164.00 ± 0.68	4.21 ± 0.02	1.05 ± 0.03
GyroAtt-SPD (LEM)	10.12 ± 0.11	211.27 ± 1.97	8.70 ± 0.05	182.97 ± 0.97	5.56 ± 0.03	1.64 ± 0.05
SSAtt	11.06 ± 0.13	221.99 ± 0.73	10.07 ± 0.13	197.16 ± 1.24	6.41 ± 0.04	3.01 ± 0.03

Main results. Tabs. 1 and 2 summarize the results on four EEG benchmarks. SSAtt achieves the best performance in all six evaluation settings, with balanced accuracies of 76.4% and 53.5% on BNCI2014001 and 86.8% and 77.8% on BNCI2015001 under inter-session and inter-subject evaluation, respectively. Notably, SSAtt improves over GyroAtt-SPD by 1.2% on both cross-subject settings. On MAMEM-SSVEP-II and BCI-ERN, SSAtt further achieves 68.8% accuracy and 79.4% AUC, surpassing CorAtt by 1.4% and 0.6%, respectively. Beyond MAtt, SSAtt consistently improves over geometry-aware attention models based on different manifold formulations, including GDLNet, GyroAtt-SPD, and CorAtt. These gains across both inter-session and inter-subject evaluations suggest that the benefit of SSAtt stems from its multi-scale compact SPD representations rather than from a specific attention geometry.

Computational cost. Tab. 3 reports the average training time per epoch of the manifold-attention methods, including forward propagation, backpropagation, and parameter updates. SSAtt incurs additional computation due to the use of multiple compact SPD spaces, sample-adaptive transformation generation, and Newton–Schulz orthogonalization. Despite this overhead, SSAtt consistently improves decoding performance over the competing methods, while the runtime remains at the scale of several seconds in the inter-session and single-task settings. The larger runtime observed in the inter-subject settings is mainly due to their substantially larger training sets. Overall, SSAtt achieves a favorable trade-off between decoding performance and computational cost.

Ablation on sample-adaptive transformation. Tab. 4 illustrates the contribution of sample-adaptive transformation to compact SPD representation learning. Generally, the complete configuration consistently outperforms the fixed transformation across all six evaluation settings, with

Table 4: Ablation results of sample-adaptive transformation components on four datasets.

Transformation	Orthogonalization	W		BNCI2014001		BNCI2015001		MAMEM	BCI-ERN
		$\mathbf{W}_b$	$\Delta\mathbf{W}$	Inter-session	Inter-subject	Inter-session	Inter-subject		
Fixed	w/o	w/o	w/o	74.9 ± 3.6	49.8 ± 15.5	86.1 ± 4.2	77.0 ± 13.0	66.9 ± 2.8	78.3 ± 3.6
Sample-Adaptive	Lie Algebra	w/	w/	74.7 ± 4.4	50.8 ± 17.1	85.8 ± 4.4	75.5 ± 13.5	67.3 ± 3.2	78.4 ± 3.8
Sample-Adaptive	Newton–Schulz	w/	w/o	74.7 ± 4.6	51.0 ± 16.5	85.7 ± 3.5	76.0 ± 13.9	67.6 ± 2.5	78.1 ± 3.2
Sample-Adaptive	Newton–Schulz	w/o	w/	70.0 ± 3.4	51.4 ± 18.5	82.5 ± 5.8	75.9 ± 14.0	67.6 ± 2.6	77.6 ± 4.0
Sample-Adaptive	Newton–Schulz	w/	w/	**76.4 ± 3.4**	**53.5 ± 15.7**	**86.8 ± 4.1**	**77.8 ± 13.4**	**68.8 ± 1.6**	**79.4 ± 2.5**

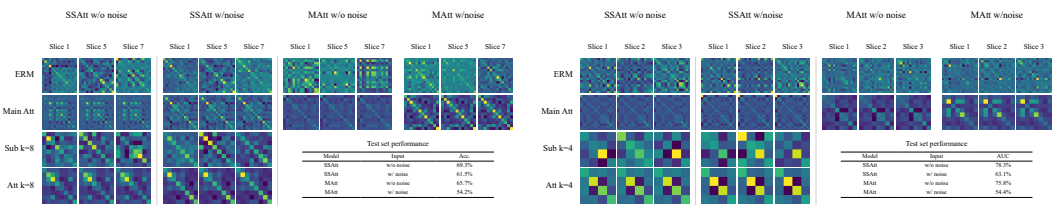
Table 5: Ablation results on compact SPD dimensions and multi-scale configurations.

Configuration	Dims.	BNCI2014001		BNCI2015001		MAMEM-SSVEP-II		BCI-ERN	
		Inter-session	Inter-subject	Inter-session	Inter-subject	Dims.	Acc. (%)	Dims.	AUC (%)
Low (1 scale)	[11]	75.0 ± 3.9	53.1 ± 16.4	$\mathbf{86.8 \pm 4.1}$	75.8 ± 13.7	[4]	66.9 ± 2.7	[4]	$\mathbf{79.4 \pm 2.5}$
Low (2 scales)	[9,11]	74.9 ± 3.7	$\mathbf{53.5 \pm 15.7}$	85.4 ± 3.9	77.0 ± 12.5	[3,4]	66.8 ± 2.7	[3,4]	78.1 ± 2.8
Low (3 scales)	[9,11,13]	74.5 ± 4.9	52.2 ± 15.9	85.4 ± 4.2	76.3 ± 13.9	[3,4,5]	67.0 ± 2.5	[3,4,5]	77.7 ± 3.6
Mid (1 scale)	[22]	$\mathbf{76.4 \pm 3.4}$	52.2 ± 15.7	85.8 ± 3.7	76.3 ± 13.7	[8]	$\mathbf{68.8 \pm 1.6}$	[9]	78.0 ± 3.3
Mid (2 scales)	[18,22]	75.0 ± 4.0	52.8 ± 17.2	85.1 ± 4.7	77.1 ± 12.9	[8,9]	66.0 ± 2.9	[8,9]	77.0 ± 3.7
Mid (3 scales)	[18,22,26]	75.2 ± 4.2	52.0 ± 17.3	85.0 ± 4.0	76.0 ± 13.1	[7,8,9]	67.5 ± 2.6	[8,9,10]	77.2 ± 3.7
High (1 scale)	[34]	75.1 ± 3.9	51.8 ± 17.7	85.6 ± 4.2	76.3 ± 13.6	[12]	66.6 ± 3.1	[13]	77.6 ± 3.4
High (2 scales)	[30,34]	75.2 ± 3.6	52.6 ± 15.7	85.0 ± 4.0	76.5 ± 13.6	[12,13]	66.8 ± 2.5	[13,14]	77.2 ± 3.2
High (3 scales)	[30,34,38]	75.0 ± 3.9	51.5 ± 17.0	85.3 ± 4.0	77.1 ± 13.1	[11,12,13]	66.6 ± 2.8	[12,13,14]	78.1 ± 3.3
Low+Mid	[11,22]	74.6 ± 4.5	51.8 ± 16.8	85.7 ± 4.0	76.2 ± 13.5	[4,8]	67.7 ± 3.0	[4,9]	77.6 ± 3.7
Mid+High	[22,34]	75.4 ± 3.6	52.5 ± 17.1	85.0 ± 3.8	76.9 ± 13.4	[8,12]	67.3 ± 2.7	[9,13]	76.4 ± 3.6
Low+High	[11,34]	75.0 ± 3.2	52.1 ± 16.5	85.6 ± 3.8	76.4 ± 13.2	[4,12]	67.1 ± 3.6	[4,13]	78.3 ± 3.6
Low+Mid+High	[11,22,34]	75.3 ± 3.9	51.5 ± 17.3	85.3 ± 4.0	$\mathbf{77.8 \pm 13.4}$	[4,8,12]	66.7 ± 2.8	[4,9,13]	77.2 ± 3.7

larger improvement observed under the inter-subject evaluation of BNCI2014001. Removing either the base projection $\mathbf{W}_b$ or the dynamic component $\Delta\mathbf{W}$ also degrades performance across all settings. In particular, excluding $\Delta\mathbf{W}$ causes drops of 2.5% and 1.8% on the inter-subject protocols of BNCI2014001 and BNCI2015001, respectively, indicating the benefit of adapting the transformation to subject-dependent covariance variations. Meanwhile, the degradation caused by removing $\mathbf{W}_b$ shows that the dynamic component is not sufficient on its own. These results indicate that $\mathbf{W}_b$ and $\Delta\mathbf{W}$ play complementary roles, with the former providing a base projection reference and the latter adapting it to the covariance structure of each trial.

Ablation on compact SPD representation design. To characterize how compact SPD representations should be configured, we vary both their dimensions and scale combinations, as summarized in Tab. 5. The results show that representation capacity is determined jointly by dimensionality and the number of scales: a single compact scale yields the best performance in four of the six evaluation settings, while additional scales are most beneficial for the inter-subject protocols of the two MI datasets. This task-dependent behavior supports the motivation of SSAtt, where attention is performed in learned compact SPD geometries rather than being restricted to a single fixed-dimensional SPD space. The results further indicate that the benefit of multi-scale representation does not arise from adding compact spaces indiscriminately, but from selecting a small set of dimensions that provides complementary views of the covariance structure underlying each EEG decoding task.

Visualization and robustness. Fig. 4 compares the representations of SSAtt and MAtt under clean EEG inputs and inputs corrupted with additive Gaussian noise. For SSAtt, the "Main Att", "Sub", and "Att" maps preserve similar structures across the two conditions. In contrast, the "Main Att" maps of MAtt change more noticeably under noise. This qualitative difference is reflected in test set performance: on MAMEM-SSVEP-II, the accuracy of SSAtt drops by 7.8 percentage points, compared with 11.5 points for MAtt; on BCI-ERN, the corresponding AUC drops are 15.2 and 21.4 points. Together, these results suggest that attention restricted to a single full-dimensional SPD space is more sensitive to noise-induced covariance variations, whereas compact SPD spaces provide additional geometric views that stabilize the learned temporal dependencies. This behavior is consistent with the motivation of SSAtt. By concentrating attention on compact covariance structures



(a) MAMEM-SSVEP-II (b) BCI-ERN

Figure 4: Robustness of SSAtt and MAtt to noise. Heatmaps show selected slices: ERM inputs, Main Att original-space outputs, and Sub/Att compact-space representations before/after attention ($k = 8$ for SSVEP, $k = 4$ for ERN). Insets report the test set performance.

while retaining the original SPD space, the model can reduce the influence of perturbation-sensitive directions without discarding the full-dimensional representation.

EEG interpretability analysis. Fig. 5 presents the gradient-based responses of SSAtt on the MAMEM-SSVEP-II dataset across five stimulus frequencies. The temporal-channel heatmaps consistently show the strongest responses around the Oz electrode, while the corresponding topomaps exhibit a pronounced concentration over the occipital region across frequencies. This spatial localization agrees with the visual-evoked origin of SSVEP responses (Herrmann, 2001; Han et al., 2018) and suggests that SSAtt emphasizes task-relevant visual activity. Fig. 6 shows the corresponding visualization on the BCI-ERN dataset. For both correct and error trials, the strongest responses are concentrated in the frontocentral region, with a clear peak around the FCz electrode. The temporal-channel heatmaps further show pronounced activity around FCz during the early event-related response period. This localization is consistent with established neurophysiological evidence linking ERN-related activity to medial frontal and anterior cingulate regions. The consistent FCz-centered responses across the two feedback conditions suggest that SSAtt captures physiologically meaningful error-related activity while preserving the task-dependent response patterns needed to distinguish correct and erroneous trials (Hajcak, 2012).

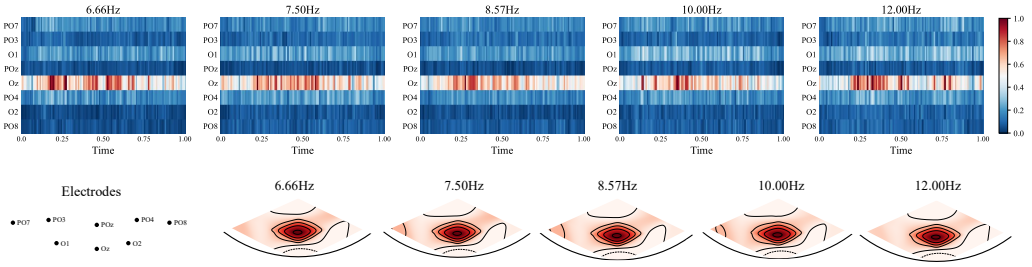


Figure 5: Spatial topomaps and gradient heatmaps of SSAtt on the MAMEM-SSVEP-II (subject S11) dataset. Strong responses appear around the Oz electrode, especially within 0–0.5s. In heatmaps, the x-axis denotes time and the y-axis denotes EEG channels.

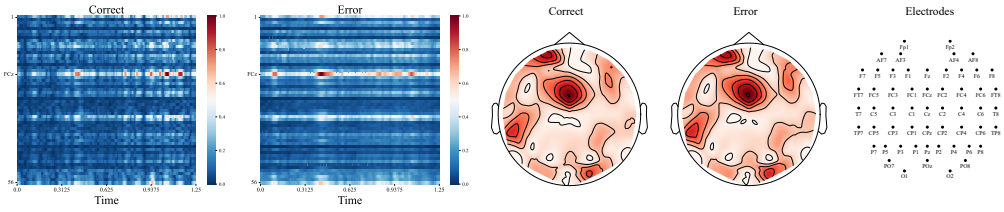


Figure 6: Spatial topomaps and gradient heatmaps of SSAtt for two classes on the BCI-ERN dataset. The x-axis represents time, and the y-axis represents EEG channels.

5 Conclusion

This paper presented SSAtt, a multi-scale submanifold-aware SPD attention network for EEG decoding. SSAtt learns sample-adaptive compact SPD representations through shared and dynamic projection components, and performs attention jointly on the original SPD space and the learned compact SPD spaces. Experiments on four EEG benchmarks show that SSAtt consistently outperforms representative Euclidean and geometry-aware baselines, while ablation studies verify the contributions of sample-adaptive transformation and compact representation design. Visualization analyses further indicate that SSAtt produces stable attention patterns under Gaussian perturbations and captures physiologically meaningful spatiotemporal EEG activity. Future work will explore automatic selection of compact dimensions and scale configurations, as well as efficient implementations for real-time BCI applications and extensions to other manifold-valued biosignals.

AI USE STATEMENT

In this work, we used generative AI tools only to polish the manuscript's language, including its grammar, wording, clarity, and readability. We have reviewed all AI-assisted revisions and take full responsibility for the final content.

REFERENCES

ZT Al-Qaysi, BB Zaidan, AA Zaidan, and MS Suzani. A review of disability eeg based wheelchair control system: Coherent taxonomy, open challenges and recommendations. *Computer methods and programs in biomedicine*, 164:221–237, 2018.

Luz Maria Alonso-Valerdi, Ricardo Antonio Salido-Ruiz, and Ricardo A Ramirez-Mendoza. Motor imagery based brain–computer interfaces: An emerging technology to rehabilitate motor deficits. *Neuropsychologia*, 79:354–363, 2015.

Ghadir Ali Altuwaijri, Ghulam Muhammad, Hamdi Altaheri, and Mansour Alsulaiman. A multi-branch convolutional neural network with squeeze-and-excitation attention blocks for eeg-based motor imagery signals classification. *Diagnostics*, 12(4):995, 2022.

Kai Keng Ang, Zheng Yang Chin, Haihong Zhang, and Cuntai Guan. Filter bank common spatial pattern (fbcsp) in brain-computer interface. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, pp. 2390–2397. IEEE, 2008.

Kai Keng Ang, Karen Sui Geok Chua, Kok Soon Phua, Chuanchu Wang, Zheng Yang Chin, Christopher Wee Keong Kuah, Wilson Low, and Cuntai Guan. A randomized controlled trial of eeg-based motor imagery brain-computer interface robotic rehabilitation for stroke. *Clinical EEG and neuroscience*, 46(4):310–320, 2015.

Vincent Arsigny, Pierre Fillard, Xavier Pennec, and Nicholas Ayache. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM journal on matrix analysis and applications*, 29(1):328–347, 2007.

Alexandre Barachant, Stéphane Bonnet, Marco Congedo, and Christian Jutten. Riemannian geometry applied to bci classification. In *International conference on latent variable analysis and signal separation*, pp. 629–636. Springer, 2010.

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.

Daniel Brooks, Olivier Schwander, Frédéric Barbaresco, Jean-Yves Schneider, and Matthieu Cord. Riemannian batch normalization for spd neural networks. *Advances in neural information processing systems*, 32, 2019.

Xiaogang Chen, Yijun Wang, Masaki Nakanishi, Xiaorong Gao, Tzyy-Ping Jung, and Shangkai Gao. High-speed spelling with a noninvasive brain–computer interface. *Proceedings of the national academy of sciences*, 112(44):E6058–E6067, 2015.

Marco Congedo, Alexandre Barachant, and Rajendra Bhatia. Riemannian geometry for eeg-based brain-computer interfaces; a primer and a review. *Brain-Computer Interfaces*, 4(3):155–174, 2017.

Josef Faller, Carmen Vidaurre, Teodoro Solis-Escalante, Christa Neuper, and Reinhold Scherer. Autocalibration and recurrent adaptation: Towards a plug and play online erd-bci. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 20(3):313–319, 2012.

JEAN QUAINTANCE Gallier and Jocelyn Quaintance. *Differential geometry and lie groups*, volume 12. Springer, 2020.

Caglar Gulcehre, Misha Denil, Mateusz Malinowski, Ali Razavi, Razvan Pascanu, Karl Moritz Hermann, Peter Battaglia, Victor Bapst, David Raposo, Adam Santoro, et al. Hyperbolic attention networks. *arXiv preprint arXiv:1805.09786*, 2018.

Greg Hajcak. What we've learned from mistakes: Insights from error-related brain activity. *Current Directions in Psychological Science*, 21(2):101–106, 2012.

Chengcheng Han, Guanghua Xu, Jun Xie, Chaoyang Chen, and Sicong Zhang. Highly interactive brain–computer interface based on flicker-free steady-state motion visual evoked potential. *Scientific reports*, 8(1):5835, 2018.

Mehrtash T Harandi, Mathieu Salzmann, and Richard Hartley. From manifold to manifold: Geometry-aware dimensionality reduction for spd matrices. In *European conference on computer vision*, pp. 17–32. Springer, 2014.

Lea Hehenberger, Reinmar J Kobler, Catarina Lopes-Dias, Nitikorn Srisrisawang, Peter Tumfart, John B Uroko, Paul R Torke, and Gernot R Müller-Putz. Long-term mutual training for the cybathlon bci race with a tetraplegic pilot: a case study on inter-session transfer and intra-session adaptation. *Frontiers in Human Neuroscience*, 15:635777, 2021.

Christoph S Herrmann. Human eeg responses to 1–100 hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. *Experimental brain research*, 137 (3):346–353, 2001.

Chen Hu, Rui Wang, Xiaoning Song, Tao Zhou, Xiao-Jun Wu, Nicu Sebe, and Ziheng Chen. A correlation manifold self-attention network for eeg decoding. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pp. 5372–5380, 2025.

Zhiwu Huang and Luc Van Gool. A riemannian network for spd matrix learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 31, 2017.

Zhiwu Huang, Ruiping Wang, Shiguang Shan, Xianqiu Li, and Xilin Chen. Log-euclidean metric learning on symmetric positive definite manifold with application to image set classification. In *International conference on machine learning*, pp. 720–729. PMLR, 2015.

Zhiwu Huang, Jiqing Wu, and Luc Van Gool. Building deep networks on grassmann manifolds. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

Thorir Mar Ingolfsson, Michael Hersche, Xiaying Wang, Nobuaki Kobayashi, Lukas Cavigelli, and Luca Benini. Eeg-tcnet: An accurate temporal convolutional network for embedded motor-imagery brain–machine interfaces. In *2020 IEEE international conference on systems, man, and cybernetics (SMC)*, pp. 2958–2965. IEEE, 2020.

Inaki Iturrate, Javier Antelis, and Javier Minguez. Synchronous eeg brain-actuated wheelchair with automated navigation. In *2009 IEEE International Conference on Robotics and Automation*, pp. 2318–2325. IEEE, 2009.

Ce Ju and Cuntai Guan. Graph neural networks on spd manifolds for motor imagery classification: a perspective from the time–frequency analysis. *IEEE Transactions on Neural Networks and Learning Systems*, 35(12):17701–17715, 2023.

Reinmar Kobler, Jun-ichiro Hirayama, Qibin Zhao, and Motoaki Kawanabe. Spd domain-specific batch normalization to crack interpretable unsupervised domain adaptation in eeg. *Advances in Neural Information Processing Systems*, 35:6219–6235, 2022.

Reinmar J Kobler, Jun-Ichiro Hirayama, Lea Hehenberger, Catarina Lopes-Dias, Gernot R Müller-Putz, and Motoaki Kawanabe. On the interpretation of linear riemannian tangent space model parameters in m/eeg. In *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pp. 5909–5913. IEEE, 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.

John M Lee. *Riemannian manifolds: an introduction to curvature*. Springer Science & Business Media, 2006.

Fabien Lotte, Marco Congedo, Anatole Lécuyer, Fabrice Lamarche, and Bruno Arnaldi. A review of classification algorithms for eeg-based brain–computer interfaces. *Journal of neural engineering*, 4(2):R1–R13, 2007.

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.

Ravikiran Mane, Effie Chew, Karen Chua, Kai Keng Ang, Neethu Robinson, A Prasad Vinod, Seong-Whan Lee, and Cuntai Guan. Fbcnet: A multi-view convolutional neural network for brain-computer interface. *arXiv preprint arXiv:2104.01233*, 2021.

Perrin Margaux, Maby Emmanuel, Daligault Sébastien, Bertrand Olivier, and Mattout Jérémie. Objective and subjective evaluation of online error correction during p300-based spelling. *Advances in Human-Computer Interaction*, 2012(1):578295, 2012.

Yazeed K Musallam, Nasser I AlFassam, Ghulam Muhammad, Syed Umar Amin, Mansour Alsulaiman, Wadood Abdul, Hamdi Altaheri, Mohamed A Bencherif, and Mohammed Algabri. Electroencephalography-based motor imagery classification using temporal convolutional network fusion. *Biomedical Signal Processing and Control*, 69:102826, 2021.

Spiros Nikolopoulos. Mamem eeg ssvep dataset ii (256 channels, 11 subjects, 5 frequencies presented simultaneously). *Centre Res. Technol. Hellas, Inf. Technol. Inst., Thessaloniki, Greece, Tech. Rep*, pp. 5, 2021.

Yue-Ting Pan, Jing-Lun Chou, and Chun-Shu Wei. Matt: A manifold attention network for eeg decoding. *Advances in Neural Information Processing Systems*, 35:31116–31129, 2022.

Pedro Luiz Coelho Rodrigues, Christian Jutten, and Marco Congedo. Riemannian procrustes analysis: transfer learning for brain–computer interfaces. *IEEE Transactions on Biomedical Engineering*, 66(8):2390–2401, 2018.

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.

D Puthankattil Subha, Paul K Joseph, Rajendra Acharya U, and Choo Min Lim. Eeg signal analysis: a survey. *Journal of medical systems*, 34(2):195–212, 2010.

Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J Miller, Gernot R Müller-Putz, et al. Review of the bci competition iv. *Frontiers in neuroscience*, 6:55, 2012.

M Teplan. Fundamental of eeg measurement measurement science review. *Measurement Science Review*, 2, 2002.

Rui Wang, Chen Hu, Ziheng Chen, Xiao-Jun Wu, and Xiaoning Song. A grassmannian manifold self-attention network for signal classification. In *IJCAI*, pp. 5099–5107, 2024.

Rui Wang, Chen Hu, Xiaoning Song, Xiaojun Wu, Nicu Sebe, and Ziheng Chen. Towards a general attention framework on gyrovector spaces for matrix manifolds. *Advances in Neural Information Processing Systems*, 38:112051–112091, 2026.

Yu-Te Wang, Masaki Nakanishi, Yijun Wang, Chun-Shu Wei, Chung-Kuan Cheng, and Tzyy-Ping Jung. An online brain-computer interface based on ssveps measured from non-hair-bearing areas. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(1):14–21, 2016.

Chun-Shu Wei, Toshiaki Koike-Akino, and Ye Wang. Spatial component-wise convolutional network (sccnet) for motor-imagery eeg classification. In *2019 9th International IEEE/EMBS Conference on Neural Engineering (NER)*, pp. 328–331. IEEE, 2019.

Or Yair, Felix Dietrich, Ronen Talmon, and Ioannis G Kevrekidis. Domain adaptation with optimal transport on the manifold of spd matrices. *arXiv preprint arXiv:1906.00616*, 2019.

Florian Yger, Maxime Berar, and Fabien Lotte. Riemannian approaches in brain-computer interfaces: a review. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10): 1753–1762, 2016.

A APPENDIX

A.1 DATASETS AND EEG PREPROCESSING

BNCI2014001. BNCI2014001 is used for MI decoding and contains EEG recordings from 9 subjects performing four motor imagery classes: left hand, right hand, both feet, and tongue. Each subject completed two sessions on different days, with six runs per session and 48 trials per run, resulting in 288 trials per session.

BNCI2015001. BNCI2015001 is another MI dataset recorded from electrodes around the sensorimotor cortex, mainly covering C3, Cz, and C4 according to the international 10–20 system. Signals were acquired at 512 Hz using a g.GAMMAsys active electrode system with a g.USBamp amplifier, with a 0.5–100 Hz band-pass filter and a 50 Hz notch filter applied during recording.

MAMEM-SSVEP-II. MAMEM-SSVEP-II is used for time-synchronous SSVEP decoding and includes EEG recordings from 11 subjects. Participants gazed at visual stimuli flickering at five frequencies: 6.66, 7.50, 8.57, 10.00, and 12.00 Hz. Each subject completed five sessions, and EEG signals were recorded from 256 channels at 250 Hz.

BCI-ERN. BCI-ERN is used for ERN decoding and was collected from a P300-based BCI spelling task, where erroneous feedback elicits ERN responses. It is an imbalanced binary classification dataset, with EEG signals recorded from 56 electrodes at 600 Hz. We use the 16 subjects released in the original competition setting.

EEG preprocessing. For BNCI2014001 and BNCI2015001, we follow the preprocessing protocol of prior work (Kobler et al., 2022). The EEG signals are processed using MOABB and MNE. Specifically, the signals are resampled to 250/256 Hz, temporally filtered in the 4–36 Hz frequency band, and segmented into short label-aligned trials with duration no longer than 3 s.

For MAMEM-SSVEP-II, we follow the preprocessing protocol used in (Pan et al., 2022). The preprocessing steps include: (1) selecting eight occipital channels related to visual responses, including PO7, PO3, POz, PO4, PO8, O1, Oz, and O2; (2) applying a band-pass filter between 1 and 50 Hz; and (3) segmenting each trial into four non-overlapping 1-second samples from 1 s to 5 s after cue onset. Each resulting SSVEP sample contains 8 channels and 125 temporal points.

For BCI-ERN, we follow the preprocessing settings adopted in previous EEG studies (Pan et al., 2022). The preprocessing steps include: (1) downsampling the EEG signals from 600 Hz to 128 Hz; and (2) applying a band-pass filter between 1 and 40 Hz. After preprocessing, each trial is represented as a 1.25-second EEG segment with 56 channels and 160 temporal samples.

A.2 IMPLEMENTATION DETAILS

Network architecture. The architectures used for the four datasets are summarized in Tab. 6. Across all datasets, the feature extraction block produces EEG features, which are partitioned and converted to SPD matrix sequences by covariance pooling. The resulting sequences are processed by SSAtt and subsequently mapped to Euclidean representations by the REM block for classification.

Table 6: SSAtt architectures across four datasets.

Block	MAMEM-SSVEP-II	BCI-ERN	BNCI2014001	BNCI2015001	Operation
Input data	1×8×125	1×56×160	1×22×750	1×13×768	
TempConv	125×1×125	22×1×160	4×22×750	5×13×768	Convolution
SpatConv	15×1×126	48×1×161	43×1×750	44×1×768	Convolution
Split & CovPool	7×15×15	3×16×16	3×43×43	3×44×44	Split + Covariance
SPDDSMBN	*w/o*	*w/o*	3×43×43	3×44×44	Domain Alignment
SSAtt	Full: 7×15×15 Sub: 7×8×8	Full: 3×16×16 Sub: 3×4×4	Full: 3×43×43 Sub: 3×22×22	Full: 3×44×44 Sub: 3×11×11	Attention
REM	7×{15,8}	3×{16,4}	3×{43,22}	3×{44,11}	$\mathrm{Log}(\cdot)$
Flat	1092	438	3597	3168	$\mathrm{vech}(\cdot)$
Classifier	5	2	4	2	FC + Softmax

Feature extraction. For MAMEM-SSVEP-II, following the preprocessing and feature extraction strategy in (Pan et al., 2022), the feature extractor consists of a channel-mixing convolution and a temporal convolution, with batch normalization applied after each block. The resulting feature map is divided into seven non-overlapping temporal segments, and covariance pooling is independently performed on each segment.

For BCI-ERN, the feature extractor first applies a channel-mixing convolution and then employs three parallel temporal convolutions, each followed by batch normalization and ELU activation. Their outputs are concatenated along the channel dimension, split into three channel-wise groups, and transformed into SPD matrices through covariance pooling.

For BNCI2014001 and BNCI2015001, following the configurations in (Wang et al., 2026), the feature extractor consists of a temporal convolution for learning local signal patterns and a subsequent spatial convolution for integrating information across EEG channels, with batch normalization applied after each convolution. The resulting feature maps are partitioned using a temporal pyramid, yielding three segments for BNCI2014001 and BNCI2015001. Covariance pooling is independently performed on each segment to construct the corresponding SPD matrix sequences. These SPD sequences are then processed by SPDDSMBN to alleviate distribution shifts across subjects and recording runs before being fed into SSAtt.

SSAtt configuration. For the SSAtt block, each full-dimensional SPD matrix is represented by the original space together with several learned compact SPD spaces. The sample-adaptive transformations combine a shared base projection with an input-conditioned dynamic component, and Newton–Schulz iteration is applied to obtain approximately semi-orthogonal projection matrices. The number of Newton–Schulz iterations is set to $T_{\mathrm{NS}} = 8$. Attention is independently evaluated on the original SPD sequence and each compact SPD sequence under the Log-Euclidean metric. The outputs from all attention spaces are mapped to Euclidean representations by the REM block, vectorized, concatenated, and passed to the final fully connected classifier.