Riemannian Batch Normalization on Correlation Manifolds for EEG Decoding

Jiarui Yang[1], Chen Hu[1], Rui Wang[1](✉), Tianyang Xu[1], Cong Hu[1], Tao Zhou[1], and Xiao-Jun Wu[1]

Jiangnan University, Wuxi, 214122, China
cs_wr@jiangnan.edu.cn

Abstract. Electroencephalography (EEG) plays an important role in brain–computer interfaces (BCIs) and clinical neuroscience by enabling non-invasive measurement of brain dynamics. Covariance descriptors are widely used in EEG decoding for their noise robustness and spatiotemporal statistical representation. Full-rank correlation matrices have emerged as a scale-invariant alternative, but current normalization layers do not account for their intrinsic geometry. In this paper, we propose Correlation Batch Normalization (CorBN), a normalization layer tailored to full-rank correlation matrices in EEG decoding. CorBN leverages the Lie group structure of the correlation manifold to perform geometrically consistent normalization while preserving manifold structures. Our approach ensures more stabilized training of deep neural networks, while respecting the intrinsic geometry of correlation matrices. We conduct extensive experiments on three EEG decoding tasks, and further validate CorBN on a large-scale clinical EEG benchmark. The results show that integrating CorBN into correlation-based networks consistently improves accuracy and accelerates convergence compared with strong baselines. The code can be found at https://github.com/JiaruiYang-AI/CorBN.

Keywords: Correlation matrices · Riemannian manifolds · Batch normalization · Riemannian networks · EEG decoding.

1 Introduction

Electroencephalography (EEG) plays a fundamental role in brain–computer interfaces (BCIs) and clinical neurological assessment. Early deep learning methods applied convolutional architectures directly to raw EEG signals [18,28]. In parallel, representing multichannel EEG segments as covariance matrices and leveraging their Riemannian geometry has shown competitive decoding performance [3,26,2,21]. However, covariance representations are sensitive to signal amplitude variability.

Full-rank correlation matrices alleviate this issue through inherent scale invariance. By normalizing each channel to unit variance, they eliminate amplitude confounds while preserving inter-channel dependencies that are more stable [16]. Nevertheless, the correlation manifold differs geometrically from the general SPD manifold [8]. Recent work has introduced permutation-invariant Riemannian metrics for this space, notably the Off-Log Metric (OLM), which admit closed-form geodesics and Fréchet means [30].

Deep learning on the Riemannian manifolds has progressed rapidly [4,6]. SPDNet [13] introduced end-to-end learning on SPD matrices, followed by extensions to Grassmann manifolds [14,31] and hyperbolic spaces [9]. In EEG decoding, manifold-based attention mechanisms further improved performance by exploiting geometric structure [25]. Building upon the OLM geometry, correlation attention networks [12] incorporated Lie group homomorphism layers, highlighting the importance of respecting the Lie group structure of the correlation manifold.

Despite these advances, a fundamental component for stable optimization is still lacking: batch normalization on the correlation manifold. Batch normalization is critical for accelerating convergence and improving training stability in deep networks [15]. Although Riemannian batch normalization has been developed for SPD matrices [5], and Lie group normalization has been proposed [7], these methods cannot be applied to correlation matrices. The unit diagonal constraint defines a strict submanifold of SPD, and naive normalization operations may violate this constraint, pushing matrices off the manifold.

In this work, we introduce Correlation Batch Normalization (CorBN), the first normalization layer specifically designed for full-rank correlation matrices. Overall, our main contributions are summarized as follows:

1. **Correlation Batch Normalization.** The first batch normalization layer tailored to full-rank correlation matrices under the Off-Log geometry.
2. **Closed-form Implementation.** We derive explicit closed-form expressions for all CorBN components, enabling more efficient computation.
3. **Experimental Validity.** Experiments on three EEG decoding benchmarks, together with a large-scale clinical dataset, demonstrate consistent improvements over state-of-the-art methods.

2 Preliminary

Lie Groups. A smooth manifold $\mathcal{M}$ is a Lie group if it is endowed with a group operation $\odot$ such that both mappings $m(x, y) \mapsto x \odot y$ and $i(x) \mapsto x^{\odot^{-1}}$ are smooth, where $x^{\odot^{-1}}$ denotes the group inverse.

Full-rank Correlation Matrix Manifold. The manifold of full-rank correlation matrices is defined as $\mathcal{C}^n_{++} = \{C \in \mathcal{S}^n_{++} : \mathrm{Diag}(C) = \mathbf{1}\}$, where $\mathcal{S}^n_{++}$ denotes the space of $n \times n$ symmetric positive definite matrices. Recent work [30] equipped $\mathcal{C}^n_{++}$ with the OLM, which endows the manifold with a Lie group structure. Under OLM, the logarithmic map $\mathrm{Log}^o(\cdot) : \mathcal{C}^n_{++} \to \mathrm{Hol}(n)$ and exponential map $\mathrm{Exp}^o(\cdot) : \mathrm{Hol}(n) \to \mathcal{C}^n_{++}$ are given by

$$\mathrm{Log}^o(C) = \mathrm{Off}(\mathrm{mlog}(C)), \qquad \mathrm{Exp}^o(S) = \mathrm{mexp}(S + \mathcal{D}^o(S)), \tag{1}$$

where $\mathrm{Off}(X) = X - \mathrm{Diag}(\mathrm{Diag}(X))$ extracts the off-diagonal part, $\mathrm{Hol}(n) = \{X \in \mathbb{R}^{n\times n} : X = X^\top, \mathrm{Diag}(X) = \mathbf{0}\}$ is the space of hollow symmetric matrices, and $\mathcal{D}^o(S)$ is the unique diagonal matrix such that $\mathrm{Diag}(\mathrm{mexp}(S+\mathcal{D}^o(S))) = \mathbf{1}$.

The weighted Fréchet Mean (wFM) of $\{C_i\}_{i=1}^N \subset \mathcal{C}_{++}^n$ with weights $\{w_i\}$ is $G = \arg\min_{C\in\mathcal{C}_{++}^n} \sum_{i=1}^N w_i\, d^2(C, C_i)$, which under OLM admits the closed form

$$G = \mathrm{Exp}^o\left(\sum_{i=1}^N w_i \,\mathrm{Log}^o(C_i)\right). \tag{2}$$

3 Proposed Method

3.1 Correlation Batch Normalization (CorBN)

Based on the general framework of BN under the Lie group structure [7], we propose CorBN for geometrically consistent normalization. Recall that Euclidean Batch Normalization (EucBN) [15] normalizes a batch $\{x_i\}_{i=1}^N$ of data points by centering and scaling with first- and second-order statistics:

$$x_i \leftarrow \gamma \cdot \frac{x_i - \mu}{\sqrt{v^2 + \epsilon}} + \beta, \tag{3}$$

where μ and v^2 denote the batch mean and variance. Analogously, CorBN controls the first- and second-order statistics on the correlation manifold.

For $\{C_i\}_{i=1}^N \subset \mathcal{C}_{++}^n$, we first compute the Fréchet mean under the OLM (Eq. 2) as G_B. Each matrix is then centered via left translation to the identity:

$$\bar{C}_i = G_B^{\odot^{-1}} \odot C_i = \mathrm{Exp}^o(\mathrm{Log}^o(C_i) - \mathrm{Log}^o(G_B)). \tag{4}$$

To quantify dispersion, we define the batch variance through squared OLM geodesic distances to the mean.

$$v_B^2 = \frac{1}{N}\sum_{i=1}^N \|\mathrm{Log}^o(C_i) - \mathrm{Log}^o(G_B)\|_F^2. \tag{5}$$

Scaling is performed by rescaling the centered tangent vectors before mapping them back onto the manifold:

$$\hat{C}_i = \mathrm{Exp}^o\left(\frac{s}{\sqrt{v_B^2+\epsilon}}\,\mathrm{Log}^o(\bar{C}_i)\right), \tag{6}$$

where $s \in \mathbb{R}$ is a learnable scale parameter. Finally, we introduce learnable bias parameter $B \in \mathcal{C}_{++}^n$, applied via group action:

$$\tilde{C}_i = B \odot \hat{C}_i = \mathrm{Exp}^o\Big(\mathrm{Log}^o(B) + \mathrm{Log}^o(\hat{C}_i)\Big). \tag{7}$$

During training, the running mean and variance are updated as follows:

$$G_S^{(t+1)} = \mathrm{Exp}^o\Big((1-\eta)\mathrm{Log}^o(G_S^{(t)}) + \eta\mathrm{Log}^o(G_B)\Big), \tag{8}$$

$$v_S^{(t+1)} = (1-\eta)v_S^{(t)} + \eta v_B^2, \tag{9}$$

where $\eta \in (0, 1)$ is the momentum parameter.

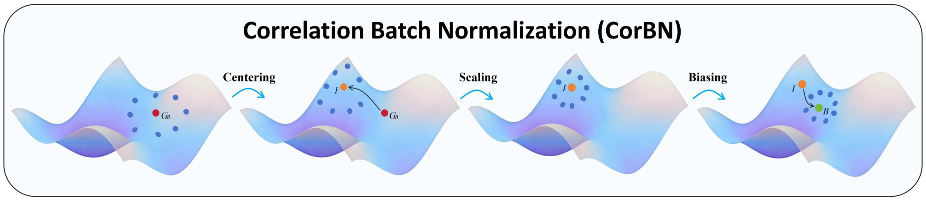


Fig. 1. An overview of CorBN. Given a batch of correlation matrices on $\mathcal{C}^n_{++}$, the Fréchet mean is computed and used for centering via left translation to the identity I_n. The centered matrices are scaled in the tangent space to control dispersion, followed by a learnable bias applied through group composition to shift the mean on the manifold.

Algorithm 1 Correlation Batch Normalization (CorBN)

Input: Batch $\{C_1, \ldots, C_N\} \subset \mathcal{C}^n_{++}$, running mean G_S, running variance v_S, learnable bias B, learnable scale s, momentum η
Output: Normalized batch $\{\tilde{C}_1, \ldots, \tilde{C}_N\}$

if training **then**
  $G_B \leftarrow \mathrm{Exp}^o\left(\frac{1}{N}\sum_{i=1}^N \mathrm{Log}^o(C_i)\right)$, $v_B^2 \leftarrow \frac{1}{N}\sum_{i=1}^N \|\mathrm{Log}^o(C_i) - \mathrm{Log}^o(G_B)\|_F^2$
  $G_S \leftarrow \mathrm{Exp}^o((1-\eta)\,\mathrm{Log}^o(G_S) + \eta\,\mathrm{Log}^o(G_B))$, $v_S \leftarrow (1-\eta)\,v_S + \eta\,v_B^2$
end if
$(\bar{G}, \bar{v}^2) \leftarrow (G_B, v_B^2)$ if training, else $(\bar{G}, \bar{v}^2) \leftarrow (G_S, v_S^2)$
for $i \leftarrow 1$ **to** N **do**
  $\bar{C}_i \leftarrow \bar{G}^{\odot^{-1}} \odot C_i$
  $\hat{C}_i \leftarrow \mathrm{Exp}^o\left(\frac{s}{\sqrt{\bar{v}^2 + \epsilon}} \cdot \mathrm{Log}^o(\bar{C}_i)\right)$
  $\tilde{C}_i \leftarrow B \odot \hat{C}_i$
end for
return $\{\tilde{C}_1, \ldots, \tilde{C}_N\}$

3.2 Statistical Completeness of CorBN

Euclidean BN can be interpreted statistically as normalizing a Gaussian distribution by controlling its first- and second-order moments. To extend this principle to the correlation manifold $\mathcal{C}^n_{++}$, we adopt the Riemannian Gaussian distribution under OLM geometry. Following [7,27], we consider the Riemannian Gaussian distribution on $\mathcal{C}^n_{++}$ with mean $M \in \mathcal{C}^n_{++}$ and variance σ^2, denoted $\mathcal{N}(M, \sigma^2)$, whose density is

$$p(C \mid M, \sigma^2) = k(\sigma)\exp\left(-\frac{(d^o(C,M))^2}{2\sigma^2}\right), \tag{10}$$

where $k(\sigma)$ is the normalizing constant and $d^o(\cdot,\cdot)$ is the geodesic distance under OLM. On $\mathcal{C}^n_{++}$, left group translation generalizes addition and subtraction, providing natural definitions of centering and biasing, while scaling is performed in the tangent space at the Fréchet mean. The following propositions characterize how CorBN controls the Riemannian mean and variance at both population and sample levels.

Proposition 1 (Population Properties). *Let $C_1, \dots, C_N \sim \mathcal{N}(M, \sigma^2)$ be i.i.d. samples on $\mathcal{C}^n_{++}$. Then, we have*

1. *The maximum likelihood estimator of M coincides with the Fréchet mean:*

$$\hat{M}_{\text{MLE}} = \underset{C \in \mathcal{C}^n_{++}}{\arg\min} \sum_{i=1}^{N} d^o(C, C_i)^2 = Exp^o\left(\frac{1}{N}\sum_{i=1}^{N} Log^o(C_i)\right).$$

2. *For any $G \in \mathcal{C}^n_{++}$, the distribution is equivariant under left group action:*

$$C \sim \mathcal{N}(M, \sigma^2) \quad \Longrightarrow \quad G \odot C \sim \mathcal{N}(G \odot M, \sigma^2).$$

Proposition 2 (Sample Properties). *Let $\{C_i\}_{i=1}^N \subset \mathcal{C}^n_{++}$ and define the scaling map $\phi_s(C) = Exp^o(s\, Log^o(C))$. Then, we have:*

1. *The sample Fréchet mean is equivariant:*

$$\text{FM}\{G \odot C_i\} = G \odot \text{FM}\{C_i\}, \quad \forall G \in \mathcal{C}^n_{++}.$$

2. *The weighted dispersion with respect to the identity scales quadratically:*

$$\sum_{i=1}^{N} w_i (d^o(\phi_s(C_i), I_n))^2 = s^2 \sum_{i=1}^{N} w_i (d^o(C_i, I_n))^2.$$

Overall, these results show that CorBN controls both the Riemannian mean and variance. After centering, the mean is mapped to the identity I_n. The scaling map ϕ_s then transforms $\mathcal{N}(I_n, \sigma^2)$ into $\mathcal{N}(I_n, s^2\sigma^2)$, mirroring the role of γ in EucBN.

4 Experiments

We evaluate CorBN on four EEG datasets: BCIC-IV-2a [29], which contains motor imagery signals from 9 subjects performing four-class tasks with 22 channels at 250 Hz; MAMEM-SSVEP-II [23], which includes data from 11 subjects responding to five-frequency visual stimuli using 8 occipital channels; BCI-ERN [20], which consists of recordings from 16 subjects performing a P300-based spelling task with 56 channels; and TUAB [24], which contains over 400K samples from a clinical abnormal-EEG screening benchmark. We compare CorBN against several state-of-the-art methods, including CNN-based approaches such as EEGNet [18], ShallowCNet [28], SCCNet [32], FBCNet [19], TCNet-Fusion [22], and MBEEGSE [1], as well as Riemannian geometry-based methods including SPDNet [13], SPDNetBN [5], MAtt [25], GDLNet [31], and our correlation attention baseline CorAtt [12].

Table 1. Performance comparison on the three involved EEG datasets. Wherein, MI and SSVEP report accuracy (%), while ERN reports AUC (%).

Method	MI	SSVEP	ERN
EEGNet [18]	61.84±6.39	53.72±7.23	74.28±2.47
ShallowCNet [28]	57.43±6.25	56.93±6.97	71.86±2.64
SCCNet [32]	71.95±5.05	62.11±7.70	70.93±2.31
FBCNet [19]	56.52±3.07	53.09±5.67	60.47±3.06
TCNet-Fusion [22]	71.45±4.45	45.00±6.45	70.46±2.94
MBEEGSE [1]	64.58±6.07	56.45±7.27	75.46±2.34
SPDNet [13]	72.93±4.33	62.30±3.12	72.05±4.43
SPDNetBN [5]	73.02±3.67	62.76±3.01	72.34±3.46
MAtt [25]	74.71±5.01	65.19±3.14	75.68±2.23
GDLNet [31]	69.32±2.89	65.52±2.86	78.23±2.52
CorAtt [12]	75.01±2.78	67.39±3.22	78.78±3.40
CorAtt-BN	**75.49±1.98**	**68.38±2.82**	**79.59±3.27**

Table 2. Performance comparison on the large-scale clinical TUAB (AUROC, %).

Method	LaBraM-base [17]	LaBraM-base + SPDNetBN	**LaBraM-base + CorBN**
AUROC	90.22±0.09	90.34±0.21	**90.98±0.11**

4.1 Main Results

Table 1 reports the performance comparison. On the MI, SSVEP, and ERN tasks, Riemannian geometry-based deep learning methods consistently outperform their Euclidean counterparts, underscoring the importance of modeling the intrinsic submanifold structure of EEG features. Among manifold-based models, CorAtt-BN achieves the best performance on all three decoding tasks. Specifically, compared with SPDNetBN, which performs Riemannian batch normalization on the broader SPD manifold, CorAtt-BN achieves accuracy gains of +2.47%, +5.62%, and +7.25%. This suggests that the scale-invariant property inherent to full-rank correlation matrices captures inter-channel relationships more faithfully than covariance descriptors. More importantly, relative to the baseline CorAtt, CorAtt-BN yields consistent improvements of +0.48% on MI, +0.99% on SSVEP, and +0.81% AUC on ERN. Notably, the standard deviations are reduced to ±1.98, ±2.82, and ±3.27, indicating improved training stability.

This stems from the two-stage design: CorBN before and after attention jointly regulates first- and second-order statistics throughout the network.

Compatibility with EEG Foundation Models. To examine whether CorBN scales to large-scale clinical data, we further validate it on TUAB [24]. We integrate CorBN into LaBraM-base [17] through an auxiliary geometric branch, where SPDNetBN or CorBN normalizes the branch features before fusion. As shown in Table 2, LaBraM-base attains 90.22% AUROC; the SPDNetBN branch

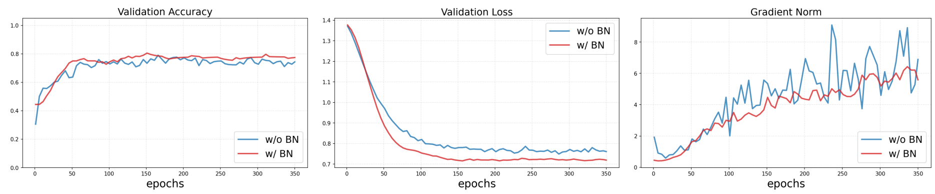


Fig. 2. Training dynamics on subject 9 of BCIC-IV-2a with and without CorBN. From left to right: validation accuracy, validation loss, and gradient norm. CorBN accelerates convergence, reduces loss, and stabilizes gradient magnitudes during training.

Table 3. Component-wise ablation of CorBN (%).

Components	MI	SSVEP	ERN
Centering only	75.01±1.68	68.09±2.73	75.57±3.63
Centering + Scaling	75.12±2.49	68.24±3.00	75.70±3.54
Full (Cent.+Scal.+Bias)	**75.49±1.98**	**68.38±2.82**	**79.59±3.27**

yields only a marginal change (90.34%), whereas CorBN improves it to 90.98%, confirming its effectiveness on large-scale clinical data and compatibility with EEG foundation models.

Computational Complexity Analysis. We analyze the per-forward-pass complexity of CorBN for a batch of N correlation matrices of size $n \times n$. Each $\mathrm{Log}^o(\cdot)$ and $\mathrm{Exp}^o(\cdot)$ requires an eigendecomposition with complexity $\mathcal{O}(n^3)$. Computing the batch Fréchet mean involves N $\mathrm{Log}^o(\cdot)$ calls and one $\mathrm{Exp}^o(\cdot)$ call, while centering, scaling, and biasing introduce $\mathcal{O}(N)$ additional $\mathrm{Exp}^o(\cdot)$ operations. The overall complexity is therefore $\mathcal{O}(Nn^3)$. This matches the asymptotic complexity of SPDNetBN. However, under OLM geometry, CorBN admits a closed-form Fréchet mean and avoids the K iterative Karcher updates required by SPDNetBN ($\mathcal{O}(KNn^3)$). Furthermore, previously computed $\{\mathrm{Log}^o(C_i)\}$ are reused, preventing redundant eigendecompositions. Consequently, CorBN achieves a smaller constant factor in practice while maintaining geometric consistency.

Training Dynamics Analysis. To examine the impact of CorBN on optimization, we compare training and validation curves with and without CorBN. Figure 2 shows representative results on BCIC-IV-2a (subject 9), with curves smoothed using an exponential moving average for clarity.

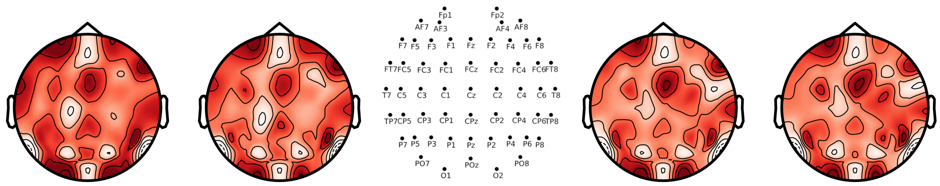
CorBN leads to faster convergence, lower validation loss, and reduced gradient-norm fluctuation in later epochs, indicating improved numerical stability.

Component-wise Ablation. Table 3 isolates each CorBN operation. The full configuration achieves the best performance across all datasets, where the learnable bias contributes the most by restoring a task-adaptive manifold mean. This confirms that centering, scaling, and bias are complementary.

Effect of CorBN Position. We further evaluate different placements of CorBN within the network. As shown in Table 4, applying CorBN either before or after attention improves performance over the baseline. Using CorBN at both posi-

Table 4. Performance comparison (%) under different CorBN positions in CorAtt.

CorBN Position	MI	SSVEP	ERN
No CorBN (baseline)	75.01±2.78	67.39±3.22	78.78±3.40
After E2R only	75.47±1.77	67.74±2.81	78.95±3.15
After Attention only	75.27±1.65	67.85±2.95	78.89±2.73
Both positions	**75.49±1.98**	**68.38±2.82**	**79.59±3.27**



Correct Error Electrodes Correct Error

(a) Without CorBN (b) With CorBN

Fig. 3. Visualization results of CorAtt-OLM on the BCI-ERN dataset S7 model. (a) Spatial topo-maps without CorBN showing activations for Correct and Error trials. (b) Spatial topo-maps with CorBN. Strong gradient activations are centered around the FCz electrode region in both cases, consistent with ERN neurophysiology.

tions yields the best results across all datasets, suggesting that normalization before and after attention provides complementary stabilization effects.

4.2 Model Interpretation

To verify that CorBN preserves neurophysiologically meaningful patterns, we visualize gradient activations on two EEG datasets. For BCI-ERN (Figure 3), gradient responses for 'Correct' versus 'Error' trials center around the FCz electrode in both configurations, consistent with the role of the anterior cingulate cortex in ERN generation [10]. For MAMEM-SSVEP-II (Figure 4), activations across all five stimulus frequencies concentrate around the Oz electrode, consistent with its location in the primary visual cortex [11].

5 Conclusion

We introduced Correlation Batch Normalization (CorBN), a principled normalization framework for neural networks operating on correlation matrices derived from EEG and other neural signals. By exploiting the Lie group structure under the Off-Log Metric, CorBN extends batch normalization to the correlation manifolds, reducing geometric operations to computationally efficient transformations while preserving manifold constraints. We showed that CorBN enables geometrically consistent centering and scaling via exponential and logarithmic

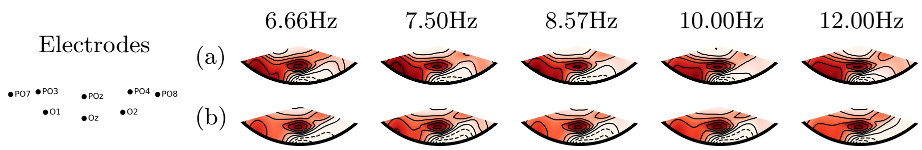


Fig. 4. Spatial topo-maps of CorAtt-OLM on the MAMEM-SSVEP-II dataset S11 model across five stimulus frequencies. (a) W/o CorBN. (b) W/ CorBN. Strong activations are centered around the Oz electrode, consistent with SSVEP neurophysiology.

maps, ensuring isometric behavior and proper batch statistics normalization. Extensive experiments on EEG-based tasks and neurological disorder classification demonstrate consistent improvements over state-of-the-art methods.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (62306127, 62576153, 62332008), the Natural Science Foundation of Jiangsu Province (BK20231040), the Fundamental Research Funds for the Central Universities (JUSRP124015), and the National Key R&D Program of China (2023YFF1105102, 2023YFF1105105).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Altuwaijri, G.A., Muhammad, G., Altaheri, H., Alsulaiman, M.: A multi-branch convolutional neural network with squeeze-and-excitation attention blocks for eeg-based motor imagery signals classification. Diagnostics (2022)
2. Arsigny, V., Fillard, P., Pennec, X., Ayache, N.: Log-euclidean metrics for fast and simple calculus on diffusion tensors. Magn. Reson. Med. (2006)
3. Barachant, A., Bonnet, S., Congedo, M., Jutten, C.: Multiclass brain–computer interface classification by riemannian geometry. IEEE Trans. Biomed. Eng. (2011)
4. Bronstein, M.M., Bruna, J., LeCun, Y., Szlam, A., Vandergheynst, P.: Geometric deep learning: going beyond euclidean data. IEEE Signal Process. Mag. (2017)
5. Brooks, D., Schwander, O., Barbaresco, F., Schneider, J.Y., Cord, M.: Riemannian batch normalization for spd neural networks. In: NeurIPS (2019)
6. Chakraborty, R., Bouza, J., Manton, J.H., Vemuri, B.C.: Manifoldnet: A deep neural network for manifold-valued data with applications. IEEE Trans. Pattern Anal. Mach. Intell. (2020)
7. Chen, Z., Song, Y., Liu, Y., Sebe, N.: A lie group approach to riemannian batch normalization. arXiv preprint arXiv:2403.11261 (2024)
8. David, P., Gu, W.: A riemannian structure for correlation matrices. Oper. Matrices (2019)
9. Ganea, O., Bécigneul, G., Hofmann, T.: Hyperbolic neural networks. In: NeurIPS (2018)
10. Hajcak, G.: What we've learned from mistakes: Insights from error-related brain activity. Curr. Dir. Psychol. Sci. (2012)

11. Herrmann, C.S.: Human eeg responses to 1–100 hz flicker: resonance phenomena in visual cortex and their potential correlation to cognitive phenomena. Exp. Brain Res. (2001)
12. Hu, C., Wang, R., Song, X., Zhou, T., Wu, X.J., Sebe, N., Chen, Z.: A correlation manifold self-attention network for eeg decoding. In: IJCAI (2025)
13. Huang, Z., Van Gool, L.: A riemannian network for spd matrix learning. In: AAAI (2017)
14. Huang, Z., Wu, J., Van Gool, L.: Building deep networks on grassmann manifolds. In: AAAI (2018)
15. Ioffe, S.: Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)
16. Jalili, M., Knyazeva, M.G.: Constructing brain functional networks from eeg: partial and unpartial correlations. J. Integr. Neurosci. (2011)
17. Jiang, W.B., Zhao, L., Lu, B.L.: Large brain model for learning generic representations with tremendous eeg data in bci. In: International Conference on Learning Representations (2024)
18. Lawhern, V.J., Solon, A.J., Waytowich, N.R., Gordon, S.M., Hung, C.P., Lance, B.J.: Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. J. Neural Eng. (2018)
19. Mane, R., Chew, E., Chua, K., Ang, K.K., Robinson, N., Vinod, A.P., Lee, S.W., Guan, C.: Fbcnet: A multi-view convolutional neural network for brain-computer interface. arXiv preprint arXiv:2104.01233 (2021)
20. Margaux, P., Emmanuel, M., Sébastien, D., Olivier, B., Jérémie, M.: Objective and subjective evaluation of online error correction during p300-based spelling. Adv. Hum.-Comput. Interact. (2012)
21. Moakher, M.: A differential geometric approach to the geometric mean of symmetric positive-definite matrices. SIAM J. Matrix Anal. Appl. (2005)
22. Musallam, Y.K., AlFassam, N.I., Muhammad, G., Amin, S.U., Alsulaiman, M., Abdul, W., Altaheri, H., Bencherif, M.A., Algabri, M.: Electroencephalography-based motor imagery classification using temporal convolutional network fusion. Biomed. Signal Process. Control (2021)
23. Nikolopoulos, S., Petrantonakis, P.C., Georgiadis, K., Kalaganis, F., Liaros, G., Lazarou, I., Adam, K., Papazoglou-Chalikias, A., Chatzilari, E., Oikonomou, V.P., et al.: A multimodal dataset for authoring and editing multimedia content: the mamem project. Data Brief (2017)
24. Obeid, I., Picone, J.: The temple university hospital eeg data corpus. Frontiers in neuroscience (2016)
25. Pan, Y.T., Chou, J.L., Wei, C.S.: Matt: A manifold attention network for eeg decoding. In: NeurIPS (2022)
26. Pennec, X., Fillard, P., Ayache, N.: A riemannian framework for tensor computing. Int. J. Comput. Vis. (2006)
27. Said, S., Bombrun, L., Berthoumieu, Y., Manton, J.H.: Riemannian gaussian distributions on the space of symmetric positive definite matrices. IEEE Trans. Inf. Theory (2017)
28. Schirrmeister, R.T., Springenberg, J.T., Fiederer, L.D.J., Glasstetter, M., Eggensperger, K., Tangermann, M., Hutter, F., Burgard, W., Ball, T.: Deep learning with convolutional neural networks for eeg decoding and visualization. Hum. Brain Mapp. (2017)
29. Tangermann, M., Müller, K.R., Aertsen, A., Birbaumer, N., Braun, C., Brunner, C., Leeb, R., Mehring, C., Miller, K.J., Müller-Putz, G.R., et al.: Review of the bci competition iv. Front. Neurosci. (2012)

30. Thanwerdas, Y.: Permutation-invariant log-euclidean geometries on full-rank correlation matrices. SIAM J. Matrix Anal. Appl. (2024)
31. Wang, R., Hu, C., Chen, Z., Wu, X.J., Song, X.: A grassmannian manifold self-attention network for signal classification. In: IJCAI (2024)
32. Wei, C.S., Koike-Akino, T., Wang, Y.: Spatial component-wise convolutional network (sccnet) for motor-imagery eeg classification. In: NER (2019)