A APPENDIX: RELATED WORKS

Past literature has explored various approaches to generalize Euclidean classifiers to non-Euclidean domains, particularly focusing on SPD matrix learning on Riemannian manifolds.

Huang et al. (Huang & Gool, 2017) introduced a Riemannian network that maintains SPD structure across layers, addressing geometric preservation in deep learning, while methods like Log-Euclidean Metric Learning (LEML) often flatten SPD manifolds, which compromises geometric fidelity.

Chen et al. (Chen et al., 2024) proposed a Riemannian Multinomial Logistic Regression (RMLR) framework for SPD networks, leveraging Pullback Euclidean Metrics (PEMs) like LEM and LCM for classification. However, their approach primarily emphasizes PEMs and encounters computational challenges when scaling to high-dimensional SPD data.

Brooks et al. (Brooks et al., 2019) introduced Riemannian Batch Normalization (RBN) to normalize data on the SPD manifold while preserving geometric properties. Although effective for small datasets, RBN becomes computationally inefficient on large-scale or high-dimensional SPD matrices.

Wang et al. (Wang et al., 2023) developed DreamNet, a deep Riemannian network designed for SPD matrix learning, utilizing stacked autoencoders and residual blocks. While DreamNet shows promising performance, its increasing depth leads to slower convergence and higher training time, particularly on smaller datasets.

Katsman et al. (Katsman et al., 2023) proposed Riemannian Residual Neural Networks (RResNet), extending residual connections to Riemannian manifolds using geodesic-based updates. Although this method theoretically preserves manifold consistency, explicit geodesic computations remain expensive, especially on complex or high-dimensional manifolds.

Recent state-of-the-art EEG models such as Hybrid BiLSTMs (Mouazen et al., 2025), EmoSTT transformers (Li et al., 2025), and LSTM-based imagined speech classifiers (Abdulghani et al., 2023) have shown strong results on DEAP, SEED, and Inner Speech, respectively. These methods leverage sequential or attention-based modeling but lack intrinsic geometry preservation; we include them to address recent critiques on task-specific baselines and to ensure fair contextualization of EGN's performance.

In contrast to these model-specific methods, our proposed EGN inherently respects the manifold architecture throughout the learning process, maintaining geometric consistency across layers. This design allows for effective classification while preserving the intrinsic structure of SPD data.

B APPENDIX: THEORETICAL PROOFS

This appendix provides proofs for the theorems stated in the main paper.

B.1 PROOF OF THEOREM 1 (EQUIVARIANCE OF BILINEAR MAPPING)

Let $\mathbf{\Sigma} \in \mathcal{S}^d_{++}$ be an SPD matrix, and let $\mathbf{W} \in \mathbb{R}^{d\times m}$ satisfy $\mathbf{W}^\top\mathbf{W} = \mathbf{I}_m$. Define the bilinear mapping

$$\mathcal{B}(\mathbf{\Sigma}) = \mathbf{W}^\top\mathbf{\Sigma}\mathbf{W}.$$

Let $\mathbf{g} \in \mathcal{O}(d)$, i.e., $\mathbf{g}^\top\mathbf{g} = \mathbf{I}_d$. Consider the orthogonal group action $\mathbf{g}\cdot\mathbf{\Sigma} := \mathbf{g}\mathbf{\Sigma}\mathbf{g}^\top$. Then:

$$\mathcal{B}(\mathbf{g}\cdot\mathbf{\Sigma}) = \mathbf{W}^\top\mathbf{g}\mathbf{\Sigma}\mathbf{g}^\top\mathbf{W}.$$

Using associativity:

$$\mathcal{B}(\mathbf{g}\mathbf{\Sigma}\mathbf{g}^\top) = (\mathbf{W}^\top\mathbf{g})\mathbf{\Sigma}(\mathbf{W}^\top\mathbf{g})^\top.$$

Define the induced transformation $\tilde{\mathbf{g}} := \mathbf{W}^\top\mathbf{g}\mathbf{W} \in \mathbb{R}^{m\times m}$. Then:

$$\mathcal{B}(\mathbf{g}\cdot\mathbf{\Sigma}) = \tilde{\mathbf{g}}\,\mathcal{B}(\mathbf{\Sigma})\,\tilde{\mathbf{g}}^\top.$$

Hence, the transformed bilinear output is conjugated by $\tilde{\mathbf{g}} = \mathbf{W}^\top\mathbf{g}\mathbf{W}$, preserving the SPD structure under the induced transformation.

■

B.2 Proof of Theorem 2 (Convex Combination of SPD Matrices).

Let $\mathbf{X}_1, \mathbf{X}_2 \in \mathcal{S}^+$ be symmetric positive definite (SPD) matrices. Then any convex combination

$$\mathbf{X} = \alpha \mathbf{X}_1 + (1-\alpha)\mathbf{X}_2, \quad \text{for } \alpha \in [0,1],$$

is also an SPD matrix.

Proof. We show that $\mathbf{X} \in \mathcal{S}^+$, i.e., that $\mathbf{X}$ is symmetric and positive definite.

(1) Symmetry: Since $\mathbf{X}_1$ and $\mathbf{X}_2$ are symmetric, we have

$$\mathbf{X}^T = (\alpha \mathbf{X}_1 + (1-\alpha)\mathbf{X}_2)^T =$$

$$\alpha \mathbf{X}_1^T + (1-\alpha)\mathbf{X}_2^T = \alpha \mathbf{X}_1 + (1-\alpha)\mathbf{X}_2 = \mathbf{X}.$$

Thus, $\mathbf{X}$ is symmetric.

(2) Positive definiteness: Let $\mathbf{v} \in \mathbb{R}^n$, $\mathbf{v} \neq \mathbf{0}$. Then

$$\mathbf{v}^T \mathbf{X} \mathbf{v} = \alpha \mathbf{v}^T \mathbf{X}_1 \mathbf{v} + (1-\alpha)\mathbf{v}^T \mathbf{X}_2 \mathbf{v}.$$

Since $\mathbf{X}_1 \succ 0$ and $\mathbf{X}_2 \succ 0$, it follows that $\mathbf{v}^T \mathbf{X}_1 \mathbf{v} > 0$ and $\mathbf{v}^T \mathbf{X}_2 \mathbf{v} > 0$. Therefore,

$$\mathbf{v}^T \mathbf{X} \mathbf{v} > 0.$$

Conclusion: $\mathbf{X}$ is symmetric and positive definite, hence $\mathbf{X} \in \mathcal{S}^+$. ■

B.3 Proof of Theorem 3 (Geometric Consistency of Geodesic Attention Weights).

Let $\Sigma \in \mathcal{S}_m^+$ be an SPD matrix and $\{\mathbf{P}_c\}_{c=1}^C \subset \mathcal{S}_m^+$ be class prototypes. Define the geodesic attention weights via temperature-scaled softmax:

$$\alpha_c(\Sigma) = \frac{\exp\left(-\frac{1}{\tau}\, d^2(\Sigma, \mathbf{P}_c)\right)}{\sum_{j=1}^C \exp\left(-\frac{1}{\tau}\, d^2(\Sigma, \mathbf{P}_j)\right)},$$

where $d(\cdot,\cdot)$ is the affine-invariant geodesic distance on $\mathcal{S}_m^+$:

$$d(\Sigma, \mathbf{P}_c) = \left\| \log\left(\Sigma^{-1/2} \mathbf{P}_c \Sigma^{-1/2}\right) \right\|_F.$$

Then for any invertible matrix $\mathbf{A} \in \mathbb{R}^{m \times m}$, it holds that

$$\alpha_c(\mathbf{A}\Sigma\mathbf{A}^T) = \alpha_c(\Sigma).$$

Proof. Let $\Sigma' = \mathbf{A}\Sigma\mathbf{A}^T$ and $\mathbf{P}'_c = \mathbf{A}\mathbf{P}_c\mathbf{A}^T$ for each class c. By Theorem 4 (Affine Invariance of Geodesic Distance), we know:

$$d(\Sigma', \mathbf{P}'_c) = d(\Sigma, \mathbf{P}_c), \quad \forall c.$$

Substituting into the attention formula:

$$\alpha_c(\Sigma') = \frac{\exp\left(-\frac{1}{\tau}\, d^2(\Sigma', \mathbf{P}'_c)\right)}{\sum_{j=1}^C \exp\left(-\frac{1}{\tau}\, d^2(\Sigma', \mathbf{P}'_j)\right)}$$

$$= \frac{\exp\left(-\frac{1}{\tau}\, d^2(\Sigma, \mathbf{P}_c)\right)}{\sum_{j=1}^C \exp\left(-\frac{1}{\tau}\, d^2(\Sigma, \mathbf{P}_j)\right)} = \alpha_c(\Sigma).$$

Conclusion: The geodesic attention weights are invariant under affine transformations of the input and prototypes. ■

B.4 Proof of the Lemma (Riemannian Mean Pooling Preservation)

Let $\{\Sigma_i\}_{i=1}^k \subset \mathcal{S}_m^+$ be a set of SPD matrices. The Riemannian (Karcher) mean

$$\mu = \arg\min_{Y \in \mathcal{S}_m^+} \sum_{i=1}^{k} \delta_g^2(\Sigma_i, Y),$$

where δ_g denotes the affine-invariant geodesic distance, satisfies $\mu \in \mathcal{S}_m^+$.

The manifold $\mathcal{S}_m^+$ equipped with the affine-invariant Riemannian metric is a Hadamard manifold: it is geodesically complete, simply connected, and has non-positive sectional curvature. In such spaces, the Karcher mean of any finite set $\{\Sigma_i\}$ exists and is unique.

Moreover, the Riemannian mean μ lies along the geodesics that connect elements in $\mathcal{S}_m^+$, and therefore remains within the manifold. Specifically, for any pair $\Sigma_i, \Sigma_j \in \mathcal{S}_m^+$, the geodesic $\gamma(t)$ joining them satisfies $\gamma(t) \in \mathcal{S}_m^+$ for all $t \in [0, 1]$. Since μ minimizes the Fréchet functional over such geodesics, it inherits the SPD property.

Thus, the Riemannian mean μ preserves both positive definiteness and the intrinsic manifold structure, i.e., $\mu \in \mathcal{S}_m^+$. ■

B.5 Statement and Proof of Affine Invariance of Geodesic Distance

Statement. Let $\mathbf{X}, \mathbf{Y} \in \mathcal{S}_{++}^n$ be two SPD matrices, and let $\mathbf{A} \in \mathbb{R}^{n \times n}$ be any invertible matrix. The affine-invariant Riemannian distance satisfies:

$$d(\mathbf{X}, \mathbf{Y}) = d(\mathbf{A}\mathbf{X}\mathbf{A}^\top, \mathbf{A}\mathbf{Y}\mathbf{A}^\top).$$

Proof Sketch. This result follows from the affine invariance of the AIRM, as shown in (Arsigny et al., 2007). Let $\mathbf{X}, \mathbf{Y} \in \mathcal{S}_{++}^n$. Then:

$$\begin{aligned}
& d(\mathbf{A}\mathbf{X}\mathbf{A}^\top, \mathbf{A}\mathbf{Y}\mathbf{A}^\top) \\
&= \left\| \log\left((\mathbf{A}\mathbf{X}\mathbf{A}^\top)^{-1/2} (\mathbf{A}\mathbf{Y}\mathbf{A}^\top) (\mathbf{A}\mathbf{X}\mathbf{A}^\top)^{-1/2} \right) \right\|_F \\
&= \left\| \log\left(\mathbf{A}^{-\top} \mathbf{X}^{-1/2} \mathbf{Y} \mathbf{X}^{-1/2} \mathbf{A}^\top \right) \right\|_F \\
&= \left\| \mathbf{A}^{-\top} \log\left(\mathbf{X}^{-1/2} \mathbf{Y} \mathbf{X}^{-1/2} \right) \mathbf{A}^\top \right\|_F \\
&= \left\| \log\left(\mathbf{X}^{-1/2} \mathbf{Y} \mathbf{X}^{-1/2} \right) \right\|_F = d(\mathbf{X}, \mathbf{Y}),
\end{aligned}$$

where the third step uses the similarity-invariance of the matrix logarithm and the unitary invariance of the Frobenius norm.

Thus, AIRM-based geodesic distance remains unchanged under any invertible affine transformation. ■

C Appendix: Manifold Components

C.1 Geometric Bias Block Details

The Geometric Bias Block $\mathcal{G}$ introduces an SPD-preserving transformation to incorporate flexible geometric shifts in the learned representation. Given an input SPD matrix $\mathbf{\Sigma}' \in \mathcal{S}_m^+$, the block computes:

$$\mathbf{\Sigma}'' = \mathcal{G}(\mathbf{\Sigma}') = \mathbf{D}\mathbf{\Sigma}'\mathbf{D}^\top,$$

where $\mathbf{D} \in \mathcal{S}_m^+$ is a symmetric positive definite bias matrix that can be learned or adaptively estimated.

SPD Preservation. The transformation $\mathbf{\Sigma}'' = \mathbf{D}\mathbf{\Sigma}'\mathbf{D}^\top$ preserves positive definiteness: - If $\mathbf{\Sigma}' \in \mathcal{S}_m^+$ and $\mathbf{D} \in \mathcal{S}_m^+$, then $\mathbf{\Sigma}'' \in \mathcal{S}_m^+$ by congruence invariance of SPD matrices.

Bias Matrix Configurations. We support three configurations for the bias matrix $\mathbf{D}$:

- **Global Learnable Bias (Mode: "learnable"):** A fixed parameter $\mathbf{D}_0 \in \mathcal{S}_m^+$ shared across all inputs, initialized near the identity and enforced to remain SPD via symmetric square or regularized parameterization.
- **Adaptive Bias (Mode: "adaptive"):** A fully connected neural network $f : \mathbb{R}^{m(m+1)/2} \to \mathbb{R}^{m\times m}$ takes as input the vectorized upper triangle of $\mathbf{\Sigma}'$ and outputs a matrix $\mathbf{D}$, which is symmetrized and SPD-corrected:

$$\mathbf{D} = f(\mathrm{vec}(\mathbf{\Sigma}')) \Rightarrow \frac{1}{2}(\mathbf{D} + \mathbf{D}^\top) + \epsilon\mathbf{I}.$$

- **Low-Rank Adaptive Bias (Mode: "lowrank_adaptive"):** Two networks generate low-rank factors $\mathbf{U}, \mathbf{V} \in \mathbb{R}^{m\times r}$ such that:

$$\mathbf{D} = \mathbf{U}\mathbf{V}^\top, \quad \text{then symmetrized and SPD-enforced.}$$

  This reduces parameter cost while retaining adaptive capability.

Additionally, an optional learnable scalar α is applied:

$$\mathbf{\Sigma}'' \leftarrow \alpha \cdot \mathbf{\Sigma}''.$$

Summary. The Geometric Bias Block provides a flexible mechanism for introducing learned or input-adaptive geometric variation while maintaining the manifold structure. It serves as a generalization of bias layers for SPD-valued inputs and plays a critical role in preserving Riemannian consistency across layers of the network.

C.2 Gradient Lifting Bridge for SPD Backpropagation

The Riemannian manifold $\mathcal{S}_m^+$ of symmetric positive definite matrices does not support naive Euclidean gradient descent without violating geometric constraints. To enable end-to-end training of our SPD-valued network with class prototypes, we adopt a geometry-aware backpropagation strategy via a Gradient Lifting Bridge.

During classification, let $\mathbf{P}_c \in \mathcal{S}_m^+$ denote the Riemannian prototype of class c, and let $\Sigma_i \in \mathcal{S}_m^+$ be the input matrix for sample i. Let $p(c \mid \Sigma_i)$ be the softmax probability for class c, based on geodesic distances. The classification loss gradient $\nabla_{\Sigma_i}\mathcal{L}$ must be transported in a way that respects the manifold geometry.

To handle this, we introduce the Gradient Lifting Bridge, which computes a lifted tangent-space gradient that aligns with the geometry of $\mathcal{S}_m^+$. In the main text, the lifted gradient was introduced in simplified form:

$$\text{Lifted Gradient} = \sum_{c=1}^{C} \nabla p(c \mid \Sigma_i) \cdot \mathbf{P}_c,$$

where $\mathbf{P}_c \in \mathcal{S}_m^+$ denotes the SPD prototype of class c, and $\nabla p(c \mid \Sigma_i)$ is the scalar gradient of the predicted probability with respect to the input matrix Σ_i.

To ensure geometric correctness, the full manifold-aware lifting is performed using the affine-invariant Riemannian metric. Specifically, we define the lifted gradient in the tangent space $T_{\Sigma_i}\mathcal{S}_m^+$ as:

$$\nabla_{\Sigma_i} = \sum_{c=1}^{C} \nabla p(c \mid \Sigma_i) \cdot \left(\Sigma_i^{1/2} \log\left(\Sigma_i^{-1/2}\mathbf{P}_c\Sigma_i^{-1/2}\right)\Sigma_i^{1/2}\right).$$

where: - $\nabla p(c \mid \Sigma_i) \in \mathbb{R}$ is the scalar gradient of the class probability with respect to Σ_i, - $\mathbf{P}_c \in \mathcal{S}_m^+$ is the prototype of class c, - the matrix logarithm is computed in the affine-invariant Riemannian metric.

This construction ensures that the gradient flow remains on the tangent space of the manifold at Σ_i, and the update step via the exponential map:

$$\Sigma_i^{(t+1)} = \exp_{\Sigma_i^{(t)}} \left(-\eta \cdot \widetilde{\nabla}_{\Sigma_i^{(t)}} \right)$$

guarantees that the resulting matrix remains in $\mathcal{S}_m^+$. Here, η is the learning rate and $\exp_\Sigma$ is the exponential map from the tangent space at Σ back to the manifold.

Motivation and Role. Standard backpropagation assumes a flat Euclidean geometry, which would distort learning dynamics on SPD manifolds and may lead to outputs that are no longer SPD. The Gradient Lifting Bridge remedies this by lifting scalar classification gradients to matrix-valued gradients consistent with the Riemannian structure. This mechanism is essential to ensure geometry-preserving optimization and enable faithful end-to-end training of all manifold-aware layers.

D Appendix : Component Present

Table 4: Component-wise comparison of geometry-preserving models. EGN uniquely integrates all key Riemannian operations in a unified, end-to-end trainable pipeline

Method	Equiv. Map	Geom. Bias	Riem. Pool	Geo. Attn	Riem. Backprop
SPDNet (Huang & Gool, 2017)	✔	✘	✘	✘	✘
SPDNet-BN (Brooks et al., 2019)	✘	✘	✘	✘	✘
RResNet (Katsman et al., 2023)	✘	✘	✔	✘	✔
DreamNet (Wang et al., 2022)	✘	✘	✔	✘	✘
EGN (Ours)	✔	✔	✔	✔	✔

E Appendix : Extended Ablation Study

We evaluate the same ablations across all five datasets. Results are consistent: geodesic attention contributes most, and the true geodesic variant performs best. However, the fixed architecture offers a good trade-off between accuracy and speed.

Table 5: Extended ablation results (% accuracy) across all tasks.

Model Variant	**DEAP**	**SEED**	**Inner Speech**	**Psychiatric**	**PaSC**
EGN (Full)	**93.1**	**92.8**	**93.6**	**93.3**	**93.7**
- w/o Geodesic Attention	89.4	90.1	91.2	90.6	91.2
- w/o Geometric Bias	88.6	89.3	90.0	89.1	90.4
- w/o Riemannian Pooling	87.9	88.7	88.3	88.0	89.5
EGN (Fixed Architecture)	91.3	90.7	92.4	91.5	92.0
EGN (True Geodesic)	93.1	92.8	93.6	93.3	93.7

Robustness to SPD Perturbations. We tested EGN under additive Gaussian noise and low-rank perturbations on SPD inputs (e.g., $\Sigma' = \Sigma + \epsilon I$, $\epsilon \in [10^{-3}, 10^{-1}]$). The model retained stable performance within small perturbation levels and degraded gracefully at high noise, confirming stability against moderate ill-conditioning.

F Appendix : Code and Reproducibility

[Codes, resources, and pretrained models will be released publicly upon publication to ensure full reproducibility.]