

HEEGNET: HYPERBOLIC EMBEDDINGS FOR EEG

000
 001
 002
 003 **Anonymous authors**
 004 Paper under double-blind review
 005
 006
 007
 008
 009
 010
 011
 012
 013
 014
 015
 016
 017
 018
 019
 020
 021
 022
 023
 024
 025
 026
 027
 028

ABSTRACT

Electroencephalography (EEG)-based brain-computer interfaces facilitate direct communication with a computer, enabling promising applications in human-computer interactions. However, their utility is currently limited because EEG decoding often suffers from poor generalization due to distribution shifts across domains (e.g., subjects). Learning robust representations that capture underlying task-relevant information would mitigate these shifts and improve generalization. One promising approach is to exploit the underlying hierarchical structure in EEG, as recent studies suggest that hierarchical cognitive processes, such as visual processing, can be encoded in EEG. While many decoding methods still rely on Euclidean embeddings, recent work has begun exploring hyperbolic geometry for EEG. Hyperbolic spaces, regarded as the continuous analogue of tree structures, provide a natural geometry for representing hierarchical data. In this study, we first empirically demonstrate that EEG data exhibit hyperbolicity and show that hyperbolic embeddings improve generalization. Motivated by these findings, we propose HEEGNet, a hybrid hyperbolic network architecture to capture the hierarchical structure in EEG and learn domain-invariant hyperbolic embeddings. To this end, HEEGNet combines both Euclidean and hyperbolic encoders and employs a novel coarse-to-fine domain adaptation strategy. Extensive experiments on multiple public EEG datasets, covering visual evoked potentials, emotion recognition, and intracranial EEG, demonstrate that HEEGNet achieves state-of-the-art performance.

1 INTRODUCTION

Electroencephalography (EEG) measures multi-channel electric brain activity (Niedermeyer & da Silva, 2005) and can reveal cognitive processes (Bell & Cuevas, 2012) and emotion states (Suhaimi et al., 2020). EEG-based brain-computer interfaces (BCI) aim to extract meaningful patterns for applications such as attention monitoring (Lee et al., 2015) and emotion recognition (Suhaimi et al., 2020). However, they currently suffer from poor generalization due to distribution shifts across sessions and subjects (Fairclough & Lotte, 2020).

In EEG-based neurotechnology, distribution shifts have traditionally been mitigated by collecting labeled calibration data and training domain-specific models (Lotte et al., 2018), which limits its utility and scalability (Wei et al., 2022). As an alternative, domain adaptation (DA) methods aim to learn a model from source domains that performs well on different (but related) target domains (Ben-David et al., 2010). In EEG, DA primarily addresses cross-session and cross-subject transfer learning problems (Wu et al., 2020). Since target domain data is typically unavailable during training and source domain data is not always available for privacy reasons, model adaptation is often treated in the context of multi-source multi-target source-free unsupervised domain adaptation (SFUDA) (Li et al., 2023). However, existing DA methods do not always work reliably, especially when the distribution shift between domains is large.

Learning robust representations that capture underlying task-relevant information better could reduce distribution shifts and thereby improve generalization in domain adaptation (Bengio et al., 2013; Zhao et al., 2019). One promising approach to achieve such robustness is to exploit the underlying hierarchical structure in EEG data. Indeed, recent studies suggest that the brain’s hierarchical cognitive processes, such as visual processing and emotion regulation, can be encoded in both intracranial EEG (Vlcek et al., 2020) and stimulus-evoked scalp EEG (Turner et al., 2023; Sun et al., 2023). While Euclidean embeddings currently dominate EEG decoding approaches, such embed-

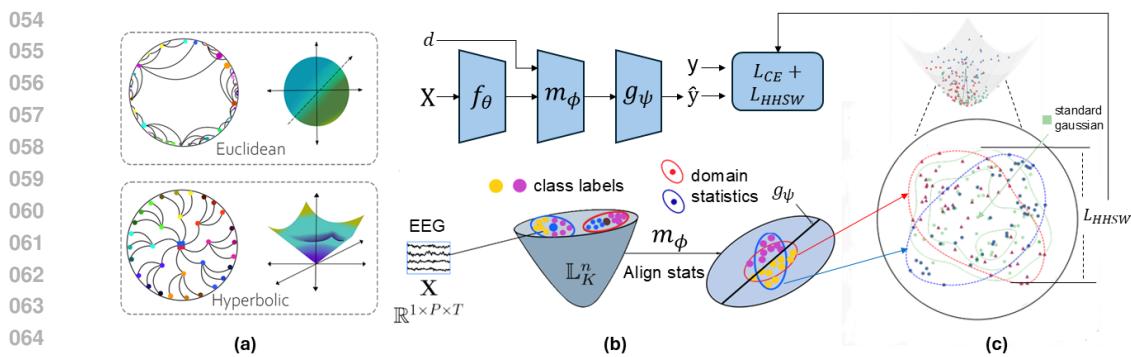


Figure 1: **Framework Overview.** (a) A comparison illustrating that hyperbolic space, with its negative curvature, is better suited for embedding hierarchical data than flat Euclidean space. (b) The HEEGNet architecture, which employs a coarse-to-fine domain adaptation strategy, DSMDBN (proposed). The first stage of DSMDBN aligns domain-specific moment statistics. (c) Top-down view in hyperbolic space: the second stage of DSMDBN aligns each source domain distribution to a standard hyperbolic Gaussian distribution by minimizing the Hyperbolic Horospherical Sliced-Wasserstein (HHSW) discrepancy.

dings are not well-suited to capture the exponential expansion of possible system states described by a hierarchical tree-like process (Peng et al., 2021). Intuitively, this is because, since the space is flat, the circumference and area of a circle grow only linearly and quadratically with the radius, respectively, leading to a mismatch with the underlying data geometry (Fig. 1a). In contrast, in hyperbolic spaces (Fig. 1a) with negative curvature, such quantities grow exponentially with the radius, thereby naturally approximating the exponential expansion of hierarchical processes (Krioukov et al., 2010). Leveraging this representational advantage, hyperbolic embeddings have outperformed Euclidean approaches across various tasks with hierarchical data in computer vision (CV) and natural language processing (NLP) (Ganea et al., 2018; Mettes et al., 2024).

In this study, we argue that hyperbolic spaces are often more appropriate for learning EEG embeddings. To this end, we first conduct a pilot study using the well-established EEGNet architecture (Lawhern et al., 2018) to generate EEG embeddings. We quantify the degree of hierarchical structure in these embeddings and confirm their hyperbolicity. We then modify EEGNet by replacing its multinomial logistic regression (MLR) with a hyperbolic variant, demonstrating that hyperbolic embeddings improve generalization over Euclidean ones. Motivated by the potential of the hyperbolic embeddings, we propose HEEGNet, a hybrid hyperbolic network architecture designed to capture the hierarchical structure of EEG data and learn domain-invariant hyperbolic EEG embeddings. HEEGNet (Fig. 3) integrates both Euclidean and hyperbolic encoders: the Euclidean encoders extract meaningful spectral–spatial–temporal EEG features and project them into hyperbolic space, with the hyperbolic encoders further refining these representations to capture hierarchical relationships more effectively. The hybrid nature of HEEGNet enables Euclidean encoders to leverage well-established signal processing principles and extract meaningful neurophysiological features from EEG signals, while the hyperbolic encoding helps to preserve the hierarchical structure in hyperbolic space. To address distribution shifts across domains, HEEGNet employs a novel *coarse-to-fine* domain alignment strategy. This strategy extends moment alignment, which is the current state-of-the-art (SotA) approach in EEG cross-domain generalization (Roy et al., 2022; Bakas et al., 2025) but whose performance often degrades under large distribution shifts (Li et al., 2025; Rodrigues et al., 2018). Specifically, we propose domain-specific moment-then-distribution batch normalization (DSMDBN, Sec. 3.1), which first explicitly aligns the domain-specific first- and second-order moments (Fig. 1b) using recently developed Riemannian batch normalization methods (Chen et al., 2025a;b). In the second stage (Fig. 1c), the aligned EEG embeddings are further matched at the distribution level by minimizing the Hyperbolic Horospherical Sliced-Wasserstein discrepancy (Bonet et al., 2023) between each source domain and a standard hyperbolic Gaussian. DSMDBN transforms domain-specific inputs into domain-invariant outputs, enabling the extension to multi-source multi-target SFUDA scenarios. We demonstrate that HEEGNet obtains SotA performance on public EEG datasets encoding hierarchy, including visual evoked potentials, emotion

108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
recognition, and intracranial EEG. Additionally, we further validate its performance on public EEG
motor imagery datasets, which are not reported to encode hierarchical information.

2 PRELIMINARIES

2.1 RELATED WORK

115
116
117
118
119
120
121
122
123
124
Brain hierarchy. The human brain is a complex network that supports various hierarchical cognitive
processes. For example, in visual processing, lower cortical areas detect basic features, which higher cortical areas progressively refine into global representations (Hochstein & Ahissar, 2002). Similarly, in emotion regulation, subcortical areas generate primal urges, which are refined by the limbic system with experience, and the neocortex finally regulates them into complex thoughts and feelings (Panksepp, 2011). To study hierarchical brain dynamics, researchers have widely used intracranial EEG, which captures neuronal population activity with high spatial and temporal resolution (Lachaux et al., 2003). As for scalp EEG, Collins et al. (2018) observed distinct EEG responses to varied stimulus frequencies reflecting different visual processing levels. Sun et al. (2023) showed the hierarchical emotion ambiguity processing with distinct EEG patterns.

125
126
127
128
129
130
Hyperbolic neural networks. Hyperbolic neural networks, which perform neural network operations in hyperbolic space, have been widely explored in NLP and CV (Peng et al., 2021; Ganea et al., 2018). In the EEG literature, Chang et al. (2025) performed contrastive pretraining for emotion recognition in hyperbolic space. In contrast, our work introduces a hybrid hyperbolic network architecture and a domain adaptation strategy (DSMDBN), leveraging hyperbolic geometry to advance both representation learning and cross-domain generalization.

131
132
133
134
135
136
137
138
139
140
Domain Adaptation in EEG. Among the DA techniques applied to EEG, moments alignment that align the first and second-order moments either in input (He & Wu, 2019; Gnassounou et al., 2024) or in latent space (Kobler et al., 2022; Bakas et al., 2025) are considered as SotA (Roy et al., 2022; Bakas et al., 2025). Such heuristic alignments are not guaranteed to improve generalization (positive transfer), but tend to improve accuracy under mild distribution shifts (Yair et al., 2019). Theoretically, DA was studied as an upper bound analysis of target risk (Ben-David et al., 2010), using discrepancy terms (such as integral probability metrics (Redko et al., 2019) or f-divergences (Acuna et al., 2021)) between the source and target distributions, suggesting alignment of the feature distributions as a potential solution (Ganin et al., 2016). Inspired by this, the proposed DSMDBN (Fig. 1c) extends moment alignment further with feature distribution alignment.

2.2 MULTI-SOURCE MULTI-TARGET SFUDA

141
142
143
144
145
146
147
148
149
150
Let \mathcal{X} denote the input space, \mathcal{Y} the label space, and \mathcal{D} the set of domain identifiers, where random variables x , y , and d represent features, labels, and domains, respectively. In the multi-source multi-target SFUDA setting, given M labeled source domain datasets $\mathcal{S} = \{(x_i, y_i, d_i) \mid x_i \in \mathcal{X}, y_i \in \mathcal{Y}, d_i \in \mathcal{D}_s\}_{i=1}^{L_s}$ and N unlabeled target domain datasets $\mathcal{T} = \{(x_i, d_i) \mid x_i \in \mathcal{X}, d_i \in \mathcal{D}_t\}_{i=1}^{L_t}$, L_s and L_t denote the sizes of source domains and target domains, and y_i and d_i indicate the associated label and domain of sample i . We assume that all source and target domains share the same feature and label spaces. The goal is to learn a model h from \mathcal{S} that generalizes to unseen, unlabeled \mathcal{T} via SFUDA, where only the trained model is available.

2.3 HYPERBOLIC GEOMETRY

151
152
153
154
155
156
157
158
159
The hyperbolic space is a Riemannian manifold of constant negative curvature $K < 0$. Among its five equivalent models (Cannon et al., 1997), we focus on the Lorentz model due to its numerical stability (Mishne et al., 2023). The n -dimensional Lorentz model is $\mathbb{L}_K^n := \{p \in \mathbb{R}^{n+1} \mid \langle p, p \rangle_{\mathcal{L}} = \frac{1}{K}, p_t > 0\}$ with the Lorentz inner product $\langle p, p \rangle_{\mathcal{L}} = \langle p_s, p_s \rangle - p_t^2$. Following Ratcliffe (2006), we denote the first dimension t as the time component and the remaining dimensions s as the space component (i.e., $p \in \mathbb{L}_K^n = [p_t, p_s^\top]^\top$). The geodesic distance, defined as the shortest distance between any two points, is defined as:

$$d_{\mathcal{L}}(p, q) = \frac{1}{\sqrt{-K}} \cosh^{-1} \left(K \langle p, q \rangle_{\mathcal{L}} \right) \quad (1)$$

Under this distance, it is possible to introduce mean and variance operations into the Lorentz manifold. Specifically, for a set of points $P = \{p_i \in \mathbb{L}_K^n\}_{i \leq M}$, the weighted Fréchet mean wFM_η on Riemannian manifolds is defined as the minimizer of the squared distances weighted by η_i :

$$\mu = \text{wFM}_\eta(\{p_i \in \mathbb{L}_K^n\}_{i=1}^M) = \arg \min_{q \in \mathbb{L}_K^n} \sum_{i=1}^M \eta_i d_{\mathcal{L}}^2(q, p_i), \quad (2)$$

If the weights η_i are uniform, the weighted Fréchet mean reduces to the standard Fréchet mean, and the Fréchet variance ν^2 is defined as the attained value at the Fréchet mean.

For each point $p \in \mathbb{L}_K^n$, there corresponds a tangent space $T_p \mathbb{L}_K^n = \{v \in \mathbb{R}^{n+1} | \langle p, v \rangle_{\mathcal{L}} = 0\}$. To project points between the manifold $p_i \in \mathbb{L}_K^n$ and the tangent space $v_i \in T_p \mathbb{L}_K^n$ at $p \in \mathbb{L}_K^n$, the exponential map $\exp_p^K : T_p \mathbb{L}_K^n \rightarrow \mathbb{L}_K^n$ and the logarithmic map $\log_p^K : \mathbb{L}_K^n \rightarrow T_p \mathbb{L}_K^n$ can be used to transport points $v_i \in T_p \mathbb{L}_K^n$ from the tangent space at p to the tangent space at q , parallel transport $\text{PT}_{p \rightarrow q}(v) \in \mathbb{L}_K^n$ operation can be performed (see App. C.1 for their closed-form expressions).

Recently, Chen et al. (2025b, Sec. 5.5) show that Lorentz model admits a gyrovector structure (Chen et al., 2025b, Def. 4), which extends the vector addition and scalar multiplication into manifolds. Specifically, the Lorentz gyroaddition, gyromultiplication, and gyroinverse are

$$p \oplus_K^\mathbb{L} q = \text{Exp}_{\bar{0}}(\text{PT}_{\bar{0} \rightarrow p}(\text{Log}_{\bar{0}}(q))), \quad \forall p, q \in \mathbb{L}_K^n, \quad (3)$$

$$t \odot_K^\mathbb{L} p = \text{Exp}_{\bar{0}}(t \text{Log}_{\bar{0}}(p)), \quad \forall t \in \mathbb{R}, \forall p \in \mathbb{L}_K^n, \quad (4)$$

$$\ominus_K^\mathbb{L} p = -1 \odot_K^\mathbb{L} p = \begin{bmatrix} p_t \\ -p_s \end{bmatrix}, \forall p \in \mathbb{L}_K^n, \quad (5)$$

where $\bar{0} = [\sqrt{-1/K}, 0, \dots, 0]^\top$ is the origin over the Lorentz model. It is also the identity element: $\bar{0} \oplus_K^\mathbb{L} q = q, \forall q \in \mathbb{L}_K^n$. Eqs. (3) and (4) admit closed-form expressions, as shown in App. C.2.

2.3.1 HYPERBOLIC OPERATIONS.

Hyperbolic neural networks. We define the hyperbolic neural networks used in this work within the Lorentz model following Bdeir et al. (2024). App. B.1 provides a brief review of gyrovector spaces, which generalize vector structures to manifolds, serving as foundations to build neural networks in hyperbolic space. The Lorentz convolutional layer (App. C.6) is defined as a matrix multiplication between a linearized kernel and a concatenation of the values in its receptive field. The Lorentz ELU (App. C.3) activation applies the activation function to the space components and concatenates them with the time component. The average pooling layer is implemented by computing the Lorentzian weighted mean of all hyperbolic features within the receptive field. Analogous to the Euclidean MLR classifier, the Lorentz MLR (App. C.7) also utilizes the distance from instances to hyperplanes to describe the class regions.

δ -hyperbolicity. Khrulkov et al. (2020) introduced δ -hyperbolicity as a measure of the degree of tree-like structure inherent in embeddings. The idea is to find the smallest value of δ for which the triangle inequality holds via the Gromov product (mathematical formulation in App. C.8). In this formulation, the definition of a hyperbolic space in terms of the Gromov product can be interpreted as stating that the metric relations between any four points are the same as they would be in a tree, up to an additive constant δ . The lower $\delta \geq 0$ is, the closer the embedding is to hyperbolic space.

Hyperbolic horospherical sliced-Wasserstein discrepancy. The sliced-Wasserstein distance (SWD) is a popular proxy for the Wasserstein distance for comparing probability distributions and has been extensively applied in optimal transport (Lee et al., 2019). Analogous to the Euclidean SWD, (Bonet et al., 2023) defined the hyperbolic sliced-Wasserstein distance by projecting distributions onto horospheres, denoted as Hyperbolic horospherical sliced-Wasserstein discrepancy (HHSW), where distances between the projections of two points belonging to a geodesic with the same direction are conserved. For probability measures $\mu, \nu \in \mathcal{P}_p(\mathbb{L}^d)$ in the Lorentz model, with $p \geq 1$, the p -th power of the HHSW is defined as:

$$\text{HHSW}_p^p(\mu, \nu) = \int_{T_{\bar{0}} \mathbb{L}^d \cap S^d} W_p^p(B_\#^v \mu, B_\#^v \nu) d\lambda(v) \quad (6)$$

216 where $T_{\bar{0}}\mathbb{L}^d \cap S^d$ is the set of unit tangent vectors at the origin $\bar{0}$, W_p^p is the p -th power of the 1D
 217 p -Wasserstein distance, $B_\#^v\mu$ denotes the horospherical projection of μ along direction v , and $d\lambda(v)$
 218 is the uniform measure over these directions.
 219

220 3 METHODS

222 3.1 DOMAIN-SPECIFIC MOMENT-THEN-DISTRIBUTION BATCH NORMALIZATION

224 **Hyperbolic batch normalization.** Batch normalization (BN) (Ioffe & Szegedy, 2015) is a widely
 225 used training technique in deep learning as BN layers speed up convergence and improve general-
 226 ization. Chen et al. (2025a;b) extended the Euclidean BN into the hyperbolic space by gyrovector
 227 structure. The centering and scaling in the Euclidean BN correspond to Lorentz gyroaddition, gy-
 228 roinverse, and gyromultiplication. Given a batch of activations $P = \{p_i \in \mathbb{L}_K^n\}_{i \leq M}$, the core
 229 operations of hyperbolic batch normalization (HBN) are
 230

$$231 \quad \text{HBN}(p_i) = \underbrace{\frac{\gamma}{\sqrt{\nu^2 + \epsilon}}}_{\text{Scaling}} \odot \underbrace{\left(\ominus \mu \oplus p_i \right)}_{\text{Centering}} \quad \forall i \leq M, \quad (7)$$

234 where μ and ν^2 are Fréchet mean and Fréchet variance, $\gamma \in \mathbb{R}$ is the scaling parameter, and ϵ is a
 235 small value for numerical stability.
 236

237 **Domain-specific momentum batch normalization for EEG.** Chang et al. (2019) introduced
 238 domain-specific batch normalization, which employs multiple parallel BN layers that process obser-
 239 vations based on their corresponding domains to mitigate domain shift. However, in EEG scenarios,
 240 where small dataset sizes can make batch statistics unreliable for normalization (Yong et al., 2020).
 241 To address this, Kobler et al. (2022) proposed Domain-Specific Momentum Batch Normalization to
 242 track domain-specific momentum-based running estimates of the first- and second-order moments.
 243 It keeps two separate sets of running estimates, the training estimates are updated using a momen-
 244 tum parameter $\eta_{\text{train}(k)}$ that follows a clamped exponential decay schedule at training step k , while
 245 a fixed momentum parameter η_{test} is used during testing.
 246

247 **Domain-specific moment-then-distribution batch normalization.** While Kobler et al. (2022)
 248 achieved SotA performance, such moment alignment is not guaranteed to improve generalization
 249 (achieve positive transfer) and often struggles under large distribution shifts (Li et al., 2025; Ro-
 250 drigues et al., 2018). On the other hand, feature distribution alignment offers theoretical guarantees,
 251 but can be challenging to achieve in practice, as the feature distributions may be too distant for effec-
 252 tive learning (Chang et al., 2019). Thus, we propose a two-stage DSMDNB strategy by extending
 253 moment alignment in (Kobler et al., 2022) to incorporate the alignment of domain-specific feature
 254 distributions in hyperbolic space.
 255

256 Formally, in our setting we assume that minibatches \mathcal{B}_k , that form the union of $N_{\mathcal{B}_k} \leq |\mathcal{D}|$ domain-
 257 specific minibatches \mathcal{B}_k^d , are drawn from distinct domains $d \in \mathcal{D}_{\mathcal{B}_k} \subseteq \mathcal{D}$. Each \mathcal{B}_k^d contains $\frac{M}{N_{\mathcal{B}_k}}$
 258 i.i.d. observations x_i , $i = 1, \dots, M/N_{\mathcal{B}_k}$. In the first stage, DSMDNB₍₁₎ (algorithm 1) explicitly
 259 aligns the domain-specific running first- and second-order moments by centering and scaling them
 260 with ν_ϕ^2 , which can be expressed as:
 261

$$262 \quad \tilde{p}_i = \text{DSMDNB}_{(1)}(p_i) = \text{HBN}^{d(i)}(p_i; \eta_{\text{test}}, \eta_{\text{train}(k)}). \quad \forall p_i \in \mathcal{B}_k^d, \quad \forall d \in \mathcal{D}_{\mathcal{B}_k} \quad (8)$$

263 DSMDNB₍₂₎ (algorithm 2), moment-aligned domain-specific EEG embeddings are then further (im-
 264 plicitly) aligned by matching them to samples from a standard hyperbolic Gaussian $\mathcal{N}(\bar{0}, 1)$. This
 265 matching is achieved by minimizing the HHSW (Eq. (6)) as a loss term
 266

$$267 \quad \text{DSMDNB}_{(2)}(p_i) = \text{HHSW}^{d(i)}(\tilde{p}_i), \quad \forall p_i \in \mathcal{B}_k^d, \quad \forall d \in \mathcal{D}_{\mathcal{B}_k}. \quad (9)$$

268 This implicit alignment guides the feature extractor toward learning robust, domain-invariant repre-
 269 sentations by matching source distributions to a standard Gaussian, thereby addressing distribution
 270 shifts that moment alignment alone often fails to mitigate. At test-time, HEEGNet applies only
 271 moment alignment, as source data are typically sufficiently diverse to capture the task-relevant vari-
 272 ability in EEG (Rodrigues et al., 2018; Mellot et al., 2023). Learning a Gaussian-aligned feature
 273 space from the sources is sufficient for the extractor to produce normalized target-domain features
 274 after moment alignment, enabling the application of domain matching in the SFUDA scenarios.
 275

270 3.2 HEEGNET
 271

272 Following Kobler et al. (2022), we constrain the hypothesis class \mathcal{H} to functions $h : \mathcal{X} \times \mathcal{D} \rightarrow \mathcal{Y}$
 273 that can be decomposed into a composition of a shared feature extractor $f_\theta : \mathcal{X} \rightarrow \mathbb{L}_K^n$, a domain-
 274 specific alignment module $m_\phi : \mathbb{L}_K^n \times \mathcal{D} \rightarrow \mathbb{L}_K^n$, and a shared classifier $g_\psi : \mathbb{L}_K^n \rightarrow \mathcal{Y}$ with
 275 parameters $\Theta = \{\theta, \phi, \psi\}$. We parametrize $h = g_\psi \circ m_\phi \circ f_\theta$ as a neural network and learn the
 276 entire model in an end-to-end fashion, which we denote as HEEGNet (details in App. D.2).

277 The HEEGNet is designed as a hybrid model, combining Euclidean encoders with hyperbolic neural
 278 networks. Most existing hyperbolic models adopt a hybrid approach, first generating hierarchical
 279 embeddings in Euclidean space and then mapping them to hyperbolic space, leveraging the strong
 280 feature extraction capabilities of well-established Euclidean encoders (Peng et al., 2021). This hy-
 281 brid design is particularly well-suited for EEG data. Studies suggest that EEG signals contain hier-
 282 archical information across temporal (Damera et al., 2020), spectral (Sun et al., 2023), and spatial
 283 (Tonoyan et al., 2017) dimensions. Euclidean encoders, such as convolutional networks, naturally
 284 align with this organization, making them effective for extracting hierarchical and neurophysiolog-
 285 ically meaningful features directly from raw EEG signals (Lawhern et al., 2018). However, such
 286 operations do not preserve physical properties in hyperbolic space, where dimensions are intrinsi-
 287 cally coupled (see App. F for fully hyperbolic experiments).

288 In a nutshell, the feature extractor f_θ consists of three Euclidean convolutional layers along with
 289 standard components (e.g., BN and pooling), a projection layer ProjX, and a hyperbolic convolu-
 290 tional layer (Fig. 3). We sequentially adopt the three convolutional layers from EEGNet (Lawhern
 291 et al., 2018): temporal convolution to learn frequency-specific filters, depthwise spatial convolu-
 292 tion to capture electrode-wise patterns, and a second depthwise temporal convolution to summarize
 293 information across time. These well-established operations provide spectral-spatial-temporal fea-
 294 ture maps with meaningful neurophysiologically properties. The ProjX layer projects the Euclidean
 295 feature maps into hyperbolic space \mathbb{L}_K^n , after which a hyperbolic convolutional layer performs point-
 296 wise convolution to optimally combine these feature maps. The alignment module m_ϕ applies the
 297 first stage of the proposed DSMDNB (Eq. (8)) to explicitly align domain-specific moments in hyper-
 298 bolic space, followed by hyperbolic ELU and pooling (Eq. (21)) for dimension reduction. Finally,
 299 the classifier g_ψ is parametrized as a hyperbolic MLR (HMLR) layer.

300
 301 4 EXPERIMENTS
 302

303 In this study, we consider three EEG modalities that have been reported to encode hierarchical struc-
 304 tures: visual- and emotion-stimuli scalp EEG, and intracranial EEG (Turner et al., 2023; Sun et al.,
 305 2023; Sonkusare et al., 2020). Corresponding EEG-based BCI applications include steady-state
 306 visually evoked potentials (SSVEP), code-modulated visually evoked potentials (CVEP), emotion
 307 recognition, and intracranial EEG. While all applications hold significant potential for rehabilitation
 308 and healthcare (Al-Nafjan et al., 2017; Guo et al., 2022; Elsner et al., 2018), their practical utility
 309 remains limited due to poor generalization across domains. We conduct a pilot study to evaluate hy-
 310 perbolic embeddings and perform comprehensive experiments to assess the proposed HEEGNet. We
 311 further evaluate HEEGNet on public motor imagery datasets that are not being reported to encode
 312 hierarchical information.

313 **Visually evoked potentials (VEP):** Nakanishi (9 subjects/1 session/12 classes) (Nakanishi et al.,
 314 2015), Wang (34/1/40) (Wang et al., 2016), CBVEP40 (12/1/4) (Castillos et al., 2023), and CB-
 315 VEP100 (12/1/4) (Castillos et al., 2023). We used MOABB (Chevallier et al., 2024) to pre-process
 316 the data and extract labeled epochs. Following Pan et al. (2022b), EEG signals were resampled to
 317 256 Hz, bandpass filtered between 1-50 Hz, and segmented each trial into 1- or 2-second segments.

318 **Emotion recognition:** Seed (15/3/3) (Duan et al., 2013) and Faced (123/1/9) (Chen et al., 2023b).
 319 We used the public available pre-processed data. For Seed, EEG signals were resampled to 200 Hz
 320 and bandpass filtered between 0–75 Hz; for Faced, a 0.05–47 Hz bandpass filter was applied.

321 **Intracranial EEG:** Boran (9/2-7/2) (Boran et al., 2020). We use MNE-python (Gramfort et al.,
 322 2014) to pre-process data and extract labeled epochs. Following Frauscher et al. (2018), EEG signals
 323 were resampled to 1000 Hz, bandpass filtered between 0.3-70 Hz.

Evaluation. We consider cross-domain adaptation within each dataset. We treat sessions as domains, and use either a leave-one-group-out (source domain number ≤ 10) or a 10-fold leave-groups-out cross-validation scheme to fit and evaluate models. For the intracranial EEG dataset, due to the different number of electrodes across subjects, we consider the cross-session adaptation setting. We fit and evaluate models independently for each subject, treating the session as the grouping variable. For other datasets, we consider the cross-subject adaptation and treat the subject as the grouping variable. We use either the standard Adam optimizer (Kingma & Ba, 2014) for Euclidean frameworks or the Riemannian Adam optimizer (Bécigneul & Ganea, 2018) for geometric frameworks, both with default hyperparameters in PyTorch. We split the source domain data into training and validation sets (80% / 20% splits, randomized and stratified by domain and label) and iterated through the training set for 100 epochs. Early stopping was fit with a single stratified (domain and labels) inner train/validation split.

4.1 PILOT STUDY.

We select five datasets as representatives to motivate the use of hyperbolic embeddings. All experiments are conducted under the cross-domain setting and follow the evaluation scheme described above. We begin by training and evaluating EEGNet on cross-domain tasks. Following Krioukov et al. (2010), we then use the target domain raw EEG data and the trained EEGNet to generate embeddings of the intermediate layer (after first two convolutional layers), and the classification space to quantify their degree of inherent tree-likeness using δ -hyperbolicity respectively (Eq. (27)). Following Khrulkov et al. (2020), we report the scale-invariant metric $\delta_{rel} \in [0, 1]$, where the lower δ_{rel} is the higher the hyperbolicity of the embeddings. We then modify EEGNet by replacing its MLR with its hyperbolic variant, HMLR, and repeat the training and evaluation procedure.

Tab. 1 indicates that all dataset both raw EEG data and EEGNet generated embeddings exhibit hierarchical structures, confirming the suitability of hyperbolic geometry for capturing the underlying information. Furthermore, replacing the Euclidean MLR layer in EEGNet with its hyperbolic counterpart consistently enhances performance across all datasets (Tab. 2), suggesting that hyperbolic geometry captures more robust representations across domains and improves generalization. t-SNE visualizations Fig. 2a, b for the identical subject and session show that under the same encoder structure, hyperbolic representations enhance class separability. These findings strongly motivate the use of hyperbolic embeddings in cross-domain generalization.

Table 1: **Pilot Study: δ_{rel} of datasets.** The lower the δ_{rel} , the closer the dataset to hyperbolic space. The number of domains in each dataset is indicated by n .

Dataset	Visual		Emotion		Intracranial
	Nakanishi (n=9)	Wang (n=34)	Seed (n=45)	Faced (n=123)	Boran (n=37)
Raw δ_{rel}	0.244 ± 0.064	0.219 ± 0.053	0.052 ± 0.026	0.103 ± 0.077	0.157 ± 0.045
Intermediate δ_{rel}	0.263 ± 0.036	0.240 ± 0.045	0.061 ± 0.026	0.111 ± 0.050	0.141 ± 0.041
Classification space δ_{rel}	0.306 ± 0.027	0.333 ± 0.039	0.072 ± 0.025	0.132 ± 0.047	0.017 ± 0.048

4.2 MAIN EXPERIMENTS

Baseline models. We included six deep learning architectures EEGNet (Lawhern et al., 2018), EEGConformer (Song et al., 2022), ATCNet (Altaheri et al., 2022), TSLANet (Eldele et al., 2024), SchirrmmeisterNet (Schirrmmeister et al., 2017), and FBCNet (Mane et al., 2021) that proposed or extensively used for general EEG decoding. We consider four deep learning architectures specifically designed for VEP: EEGInception (Santamaria-Vazquez et al., 2020), DDGCNN (Zhang et al., 2024), SSVEPNet (Pan et al., 2022a), SSVEPFormer (Chen et al., 2023a); and two deep learning architectures for emotion recognition: EmT (Ding et al., 2025), TScception (Ding et al., 2022). We further considered two foundation models: LaBraM (Jiang et al., 2024) and CBraMod (Wang et al., 2024a) in evaluation. We use the implementation provided in the public available repositories for

378
 379
 380
 381
 382
Table 2: Pilot Study: comparison of Euclidean / Hyperbolic MLR. The averages of test-set scores
 383 are shown above using balanced accuracy (the score and the standard deviation are shown for each
 384 dataset). The number of domains in each dataset is indicated by n . Permutation-paired t-tests were
 385 used to identify significant differences between EEGNet+HMLR (*Hyperbolic*) and EEGNet (1e4
 386 permutations). Significant differences are highlighted using dots: • $p \leq 0.05$, • $p \leq 0.01$, • $p \leq 0.001$.
 387
 388
 389
 390

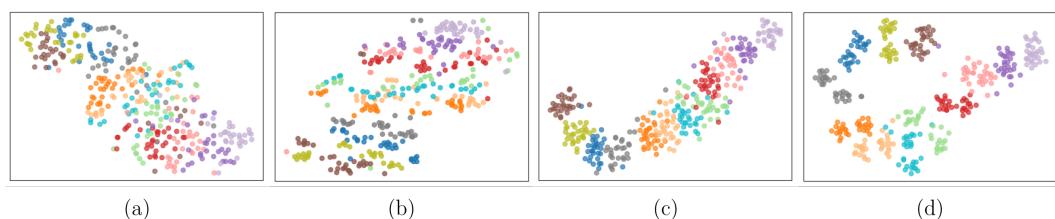
Dataset Model	Visual		Emotion		Intracranial
	Nakanishi (n = 9)	Wang (n = 34)	Seed (n = 45)	Faced (n = 123)	Boran (n = 37)
EEGNet	57.2 ± 19.8 •	37.4 ± 12.9 •	74.5 ± 19.8	24.8 ± 13.7 •	55.4 ± 9.1
EEGNet + HMLR	60.8 ± 20.7	39.2 ± 13.8	75.1 ± 20.0	38.8 ± 12.4	57.4 ± 8.7

391 all architectures, stick to all hyperparameters as provided, and use the standard cross-entropy loss as
 392 training objective.
 393

394 **SFUDA baselines.** We consider two SFUDA baselines: Euclidean alignment (EA) (He & Wu,
 395 2019), and spatio-temporal Monge alignment (STMA) (Gnassounou et al., 2024). These alignment
 396 methods are model-agnostic techniques that are applied to the EEG data in the input space before a
 397 model is fitted. For example, EA aligns the EEG trials covariance matrix directly in the input space.
 398 Therefore, we combine them with different models (e.g., EEGNet + EA) in our evaluation.
 399

400 **HEEGNet.** We parametrize $h = g_\psi \circ m_\phi \circ f_\theta$ as a neural network and learn the entire model in
 401 an end-to-end fashion, and use the standard cross-entropy loss with HHSW (Eq. (9)) as the training
 402 objective. The HHSW loss weight is a hyperparameter to be tuned, we set it to 0.01 for emotion and
 403 0.5 for other datasets in our experiments. **Following Bdeir et al. (2024), we set the curvature to -1
 404 by default and implemented the norm normalization to assure numerical stabilization for training.**
 405 **Computational cost for hyperbolic operation is briefly discussed in App. D.4.** During source-free
 406 target domain adaptation, HEEGNet keeps the fitted source feature extractor f_θ and linear classi-
 407 fier g_ψ fixed and estimates domain-specific first- and second-order statistics by solving Eq. (2) for
 408 moments alignment m_ϕ .
 409

410 Tab. 3 summarizes the results across all three EEG modalities, presenting the grand average
 411 scores with general EEG decoding baseline methods. Extended results for VEP-specific, emotion
 412 recognition-specific and foundation models, as well as per-dataset results, are provided in Sup-
 413 plementary Tab. 8 and Tab. 9. At the overall grand average level, HEEGNet (DSMDBN) outper-
 414 forms all baseline methods. While input space alignment methods like EA and STMA improved
 415 cross-domain generalization, a gap still remains compared to our proposed two-stage alignment
 416 DSMDBN. A visualization of the two stages of DSMDBN is shown in Fig. 2c, d. Interestingly,
 417 HEEGNet (DSMDBN+EA), the integration of DSMDBN with the input space alignment method
 418 EA, achieves superior performance. We conduct an ablation study to systematically investigate the
 419 effects of different alignment strategies within the HEEGNet architecture.
 420



421
 422
 423
 424
 425
Figure 2: t-SNE visualizations of classification space for Nakanishi (subject 4, session1). Each
 426 color denotes a distinct class. (a) EEGNet, (b) EEGNet with HMLR, (c) DS MDBN stage 1, and
 427 (d) DS MDBN stage 2. The plots illustrate how hyperbolic embedding and our proposed two-stage
 428 alignment progressively enhance class separability.
 429

430 **Ablation study.** Tab. 4 summarizes the effectiveness of three different alignment methods, input
 431 alignment (EA), moment alignment, and distribution alignment, within the HEEGNet architecture.
 432

432
 433 **Table 3: Main experiment results.** Grand average of test-set scores across three EEG modalities
 434 (balanced accuracy (%); higher is better; mean \pm std) Permutation-paired t-tests were used to
 435 identify significant differences between HEEGNet (DSMDBN+EA) and baseline methods (1e4 per-
 436 mutations, 18 tests, t-max correction). Significance markers: • $p \leq 0.05$, • $p \leq 0.01$, • $p \leq 0.001$.

437 Model	438 SFUDA	439 VEP (n=67)	440 Emotion (n=168)	441 Intracranial (n=37)	442 Overall (n=272)
443 EEGNet	w/o	35.7 \pm 15.5 •	38.1 \pm 27.0 •	55.4 \pm 9.1 •	39.9 \pm 23.6 •
	EA	55.8 \pm 20.8 •	73.8 \pm 22.3 •	59.1 \pm 11.2 •	67.4 \pm 22.3 •
	STMA	30.3 \pm 8.9 •	70.7 \pm 20.5 •	51.2 \pm 5.9 •	58.1 \pm 24.0 •
444 EEGConformer	w/o	29.1 \pm 9.1 •	83.2 \pm 16.9 •	55.5 \pm 7.0 •	66.1 \pm 27.2 •
	EA	46.1 \pm 23.1•	82.8 \pm 16.8 •	61.7 \pm 13.1•	70.9 \pm 24.1•
	STMA	27.8 \pm 7.0 •	78.2 \pm 15.9 •	53.1 \pm 5.2 •	62.4 \pm 25.2 •
445 ATCNet	w/o	33.3 \pm 15.4 •	17.5 \pm 11.7 •	57.5 \pm 9.0 •	26.9 \pm 18.5 •
	EA	52.2 \pm 21.4•	54.1 \pm 15.5 •	57.1 \pm 9.0 •	54.0 \pm 16.5 •
	STMA	39.4 \pm 15.6 •	56.4 \pm 16.3 •	54.2 \pm 5.7 •	51.9 \pm 16.7 •
446 TSLANet	w/o	24.4 \pm 14.1 •	35.5 \pm 12.9 •	55.1 \pm 9.3 •	35.5 \pm 15.7 •
	EA	37.3 \pm 28.1 •	44.8 \pm 18.1 •	54.9 \pm 10.4 •	44.3 \pm 20.8 •
	STMA	20.5 \pm 5.4 •	49.6 \pm 15.1 •	54.1 \pm 6.2 •	43.0 \pm 17.9 •
447 SchirrmesteinerNet	w/o	40.2 \pm 27.5 •	51.9 \pm 22.4 •	56.9 \pm 7.7 •	49.7 \pm 23.1 •
	EA	35.7 \pm 24.7 •	79.0 \pm 16.4 •	65.0 \pm 10.2	66.5 \pm 25.7 •
	STMA	33.1 \pm 16.8 •	64.2 \pm 16.7 •	53.9 \pm 5.5 •	55.2 \pm 20.4 •
448 FBCNet	w/o	22.8 \pm 8.0 •	28.5 \pm 19.5 •	61.1 \pm 10.6 •	31.6 \pm 20.2 •
	EA	38.8 \pm 28.4 •	51.3 \pm 20.7 •	61.3 \pm 11.0 •	49.6 \pm 22.9 •
	STMA	21.6 \pm 5.4 •	33.0 \pm 20.9 •	53.0 \pm 6.8 •	32.9 \pm 19.2 •
449 HEEGNet (<i>proposed</i>)	DSMDBN	79.6\pm19.8	82.8 \pm 16.6	58.4 \pm 9.6	78.7 \pm 18.6
	DSMDBN+EA	77.4 \pm 12.6	86.7\pm12.5	66.7\pm12.5	81.7\pm14.4

463 We highlight three important observations. First, moment alignment in hyperbolic space is the pri-
 464 mary driver of performance improvement. The absence of moment alignment consistently results
 465 in a significant performance drop of at least 12.3% compared to the best-performing configuration.
 466 Second, distribution alignment proves effective only when paired with moment alignment, validating
 467 our proposed DSMDBN approach, which first aligns moments to facilitate subsequent distribution
 468 alignment. Third, moment alignment in the latent space outperforms input space alignment, support-
 469 ing the findings of Bakas et al. (2025) that alignment benefits from enhanced class discrimination in
 470 the latent representation. Fourth, DSMDBN augmented with input alignment yields the best perfor-
 471 mance, indicating that multistage alignment a promising strategy in cross-domain generalization.
 472

473 **Table 4: Ablation results.** Grand average of all test-sets scores (balanced accuracy (%), higher is
 474 better) relative to the combination of alignment in HEEGNet. Permutation-paired t-test values and
 475 adjusted p-values indicate the effect strength (1e4 permutations, 7 tests, t-max correction).

477	478 Alignment			479 Metrics	
	Moments	Distribution	Input	mean (std)	t-val (p-val)
✓	✓	✓	-	-	-
✓	✓	✗	-3.0 (18.6)	3.6 (0.0003)	
✓	✗	✓	-5.0 (15.6)	8.8 (0.0001)	
✓	✗	✗	-5.0 (19.3)	5.5 (0.0001)	
✗	✓	✗	-12.3 (23.1)	10.4 (0.0001)	
✗	✗	✗	-12.9 (22.4)	11.7 (0.0001)	
✗	✓	✓	-15.5 (21.4)	14.9 (0.0001)	
✗	✗	✓	-15.5 (22.2)	16.0 (0.0001)	

486 **4.3 MOTOR IMAGERY**
 487

488 We consider two publicly available motor imagery datasets. Pre-processing included resampling
 489 EEG signals to 250 or 256 Hz, applying temporal filters to capture frequencies between 4 and 36 Hz,
 490 and extracting 3-second epochs linked to specific class labels. In addition to Euclidean models, we
 491 evaluate several manifold-based decoding approaches for motor imagery. These include Grassmann
 492 manifold methods: GDLNet (Wang et al., 2024b) and GyroAtt-Gr (Wang et al., 2025); Symmetric
 493 Positive Definite (SPD) manifold methods: TSMNet (Kobler et al., 2022), MAtt (Pan et al., 2022b),
 494 and CSPNet (Ju & Guan, 2022); as well as Symmetric Positive Semi-Definite (SPSD) manifold
 495 methods: GyroAtt-SPSD (Wang et al., 2025). The results, summarized in Tab. 5, indicate that our
 496 proposed HEEGNet delivers competitive performance.
 497

497 **Table 5: Motor imagery per dataset results.** Average of test-set scores (balanced accuracy (%);
 498 higher is better; mean \pm std).

Manifold	Model	SFUDA	Dataset	
			BNCI2014001	BNCI2015001
Euclidean	ATCNet	w/o	42.7 \pm 16.4	60.2 \pm 8.4
	EEGConformer	w/o	42.6 \pm 16.7	60.1 \pm 10.7
	EEGNet	w/o	43.6 \pm 16.7	61.3 \pm 8.8
		EA	49.9 \pm 16.9	72.5 \pm 14.2
	STMA	STMA	49.7 \pm 16.9	69.9 \pm 14.6
		EEGIceptionMI	39.7 \pm 12.7	59.5 \pm 9.2
		ShallowNet	42.2 \pm 16.2	58.7 \pm 5.8
		LaBraM	33.3 \pm 16.6	70.8 \pm 31.4
Grassmann	CBraMod	w/o	30.6 \pm 4.0	59.1 \pm 8.1
	GDLNet	w/o	46.3 \pm 5.1	63.3 \pm 14.2
SPSD	GyroAtt-Gr	w/o	52.1 \pm 14.2	75.3 \pm 13.7
	GyroAtt-SPSD	w/o	51.7 \pm 13.1	74.9 \pm 12.6
SPD	TSMNet	MAtt	45.3 \pm 11.3	63.1 \pm 10.1
		CSPNet	45.2 \pm 9.3	64.2 \pm 13.4
		w/o	43.0 \pm 13.3	61.7 \pm 11.4
	SPDDSBN	EA	51.2 \pm 15.1	72.5 \pm 13.6
		STMA	52.5 \pm 16.4	70.1 \pm 14.2
		SPDDSBN	54.6 \pm 16.1	74.3 \pm 14.7
Hyperbolic	HEEGNet	DSMDBN+EA	54.1 \pm 15.9	75.8 \pm 13.0

524 **5 DISCUSSIONS**
 525

526 In this work, we introduced HEEGNet, a hybrid hyperbolic network architecture designed to cap-
 527 ture the hierarchical structure of EEG data and learn domain-invariant hyperbolic embeddings. Our
 528 pilot study and empirical analyses indicate that EEG data exhibits hyperbolicity and that hyperbolic
 529 embeddings improve generalization compared to Euclidean ones. By integrating both Euclidean
 530 and hyperbolic encoders and employing a novel two-stage domain adaptation strategy (DSMDBN),
 531 HEEGNet effectively aligns domain-specific moments and distributions in hyperbolic space. Ex-
 532 periments on multiple EEG datasets, including visual evoked potentials, emotion recognition, and
 533 intracranial EEG, demonstrate state-of-the-art performance.
 534

535 Despite these strengths, HEEGNet relies on stable estimates of domain-specific statistics, which
 536 become unreliable in online settings where data arrive sequentially. Like all hyperbolic neural net-
 537 works, hyperbolic operations add additional computational cost. Future work will explore online
 538 extensions of hyperbolic normalization and the development of more expressive encoders capable
 539 of capturing hierarchical structure in EEG signals.

540 **REPRODUCIBILITY STATEMENT**
 541

542 We have made great efforts to ensure the reproducibility of our results. The primary components
 543 necessary for replication are detailed throughout the main paper and supplementary materials.
 544

545 **Code Availability.** Our proposed novel hybrid hyperbolic neural network architecture HEEG is
 546 described in Section 3.2, the novel domain adaptation strategy is described in Section 3.1 . To
 547 ensure the reproducibility of HEEGNet, we provide an anonymous link to the downloadable source
 548 code below.

549 <https://anonymous.4open.science/r/HEEGNet-F655>
 550

551 The provided code is designed to allow other researchers to replicate the methods and main results
 552 described in the paper with minimal effort. In particular, it includes implementations of all the key
 553 algorithms, model architectures, and evaluation procedures used in our experiments. To ensure ease
 554 of use, the code is well-documented and structured to facilitate understanding and modification.
 555 Instructions for setting up the environment, installing necessary dependencies, and running the code
 556 are included in a detailed README file.

557 **Data Availability and Processing Steps.** All datasets used in our experiments are publicly avail-
 558 able VEP, intracranial EEG and emotion recognition datasets. Our preprocessing and evaluation
 559 schemes are detailed in Section 4.

560 All the public VEP datasets we used can be found in the link below.

562 https://moabb.neurotechx.com/docs/dataset_summary.html

563 All the public emotion recognition datasets we used can be found in links below.

565 SEED: <https://bcmi.sjtu.edu.cn/home/seed/>

566 Faced: <https://www.synapse.org/Synapse:syn50614194/files/>

568 The public intracranial and EEG datasets we used can be found in links below.

569 Boran: <https://doi.gin.g-node.org/10.12751/g-node.d76994/>

571 **ETHICS STATEMENT**
 572

573 This research employs computational approaches exclusively on publicly available datasets, with no
 574 involvement of human subjects or handling of sensitive data. We follow ICLR’s ethical guidelines
 575 and declare no competing interests, prioritizing the responsible use of our findings and transparent
 576 reporting to facilitate reproducibility.

577

578

579

580

581

582

583

584

585

586

587

588

589

590

591

592

593

594 REFERENCES
595

- 596 David Acuna, Guojun Zhang, Marc T Law, and Sanja Fidler. f-domain adversarial learning: Theory
597 and algorithms. In *International Conference on Machine Learning*, pp. 66–75. PMLR, 2021.
- 598 Abeer Al-Nafjan, Manar Hosny, Yousef Al-Ouali, and Areej Al-Wabil. Review and classification of
599 emotion recognition based on eeg brain-computer interface system research: a systematic review.
600 *Applied Sciences*, 7(12):1239, 2017. doi: 10.3390/app7121239.
- 601 Hamdi Altaheri, Ghulam Muhammad, and Mansour Alsulaiman. Physics-informed attention tem-
602 poral convolutional network for eeg-based motor imagery classification. *IEEE transactions on*
603 *industrial informatics*, 19(2):2249–2258, 2022. doi: 10.1109/TII.2022.3197419.
- 604 Stylianos Bakas, Siegfried Ludwig, Dimitrios A Adamos, Nikolaos Laskaris, Yannis Panagakis, and
605 Stefanos Zafeiriou. Latent alignment in deep learning models for eeg decoding. *Journal of Neural*
606 *Engineering*, 22(1):016047, 2025. doi: 10.1088/1741-2552/adb336.
- 607 Ahmad Bdeir, Kristian Schwethelm, and Niels Landwehr. Fully hyperbolic convolutional neural
608 networks for computer vision. In *ICLR*, 2024.
- 609 Gary Bécigneul and Octavian-Eugen Ganea. Riemannian adaptive optimization methods. *arXiv*
610 *preprint arXiv:1810.00760*, 2018.
- 611 Martha Ann Bell and Kimberly Cuevas. Using eeg to study cognitive development: Issues and
612 practices. *Journal of cognition and development*, 13(3):281–294, 2012. doi: 10.1080/15248372.
613 2012.691143.
- 614 Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wort-
615 man Vaughan. A theory of learning from different domains. *Machine learning*, 79:151–175,
616 2010. doi: 10.1007/s10994-009-5152-4.
- 617 Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new
618 perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828,
619 2013. doi: 10.1109/TPAMI.2013.50.
- 620 Clément Bonet, Laetitia Chapel, Lucas Drumetz, and Nicolas Courty. Hyperbolic sliced-wasserstein
621 via geodesic and horospherical projections. In *Topological, Algebraic and Geometric Learning*
622 *Workshops 2023*, pp. 334–370. PMLR, 2023.
- 623 Ece Boran, Tommaso Fedele, Adrian Steiner, Peter Hilfiker, Lennart Stieglitz, Thomas Grunwald,
624 and Johannes Sarnthein. Dataset of human medial temporal lobe neurons, scalp and intracranial
625 eeg during a verbal working memory task. *Scientific data*, 7(1):30, 2020. doi: 10.6084/m9.
626 figshare.11500485.
- 627 James W Cannon, William J Floyd, Richard Kenyon, Walter R Parry, et al. Hyperbolic geometry.
628 *Flavors of geometry*, 31(59–115):2, 1997.
- 629 Kalou Cabrera Castillos, Simon Ladouce, Ludovic Darmet, and Frédéric Dehais. Burst c-vep based
630 bci: Optimizing stimulus design for enhanced classification with minimal calibration data and
631 improved user experience. *NeuroImage*, 284:120446, 2023. doi: 10.1016/j.neuroimage.2023.
632 120446.
- 633 Ines Chami, Zhitao Ying, Christopher Ré, and Jure Leskovec. Hyperbolic graph convolutional neural
634 networks. In *Advances in Neural Information Processing Systems*, pp. 4869–4880, 2019.
- 635 Jiang Chang, Zhixin Zhang, Yuhua Qian, and Pan Lin. Multi-scale hyperbolic contrastive learning
636 for cross-subject eeg emotion recognition. *IEEE Transactions on Affective Computing*, 2025. doi:
637 10.1109/TAFFC.2025.3535542.
- 638 Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han. Domain-specific
639 batch normalization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF con-
640 ference on Computer Vision and Pattern Recognition*, pp. 7354–7362, 2019.

- 648 Jianbo Chen, Yangsong Zhang, Yudong Pan, Peng Xu, and Cuntai Guan. A transformer-based
 649 deep neural network model for ssvep classification. *Neural Networks*, 164:521–534, 2023a. doi:
 650 10.1016/j.neunet.2023.04.045.
- 651 Jingjing Chen, Xiaobin Wang, Chen Huang, Xin Hu, Xinkle Shen, and Dan Zhang. A large finer-
 652 grained affective computing eeg dataset. *Scientific Data*, 10(1):740, 2023b. doi: 10.7303/
 653 syn50614194.
- 654 Ziheng Chen, Yue Song, Xiaojun Wu, and Nicu Sebe. Gyrogroup batch normalization. In *ICLR*,
 655 2025a.
- 656 Ziheng Chen, Xiao-Jun Wu, Bernhard Schölkopf, and Nicu Sebe. Riemannian batch normalization:
 657 A gyro approach. *arXiv preprint arXiv:2509.07115*, 2025b.
- 658 Sylvain Chevallier, Igor Carrara, Bruno Aristimunha, Pierre Guetschel, Sara Sedlar, Bruna Lopes,
 659 Sébastien Velut, Salim Khazem, and Thomas Moreau. The largest eeg-based bci reproducibility
 660 study for open science: the moabb benchmark. *arXiv preprint arXiv:2404.15319*, 2024.
- 661 Elliot Collins, Amanda K Robinson, and Marlene Behrmann. Distinct neural processes for the
 662 perception of familiar versus unfamiliar faces along the visual hierarchy revealed by eeg. *NeuroImage*,
 663 181:120–131, 2018. doi: 10.1016/j.neuroimage.2018.06.080.
- 664 Srikanth R Damera, Jacob G Martin, Clara Scholl, Judy S Kim, Laurie Glezer, Patrick S Malone,
 665 and Maximilian Riesenhuber. From shape to meaning: Evidence for multiple fast feedforward
 666 hierarchies of concept processing in the human brain. *Neuroimage*, 221:117148, 2020. doi:
 667 j.neuroimage.2020.117148.
- 668 Yi Ding, Neethu Robinson, Su Zhang, Qiuhan Zeng, and Cuntai Guan. Tsception: Capturing tem-
 669 poral dynamics and spatial asymmetry from eeg for emotion recognition. *IEEE Transactions on*
 670 *Affective Computing*, 14(3):2238–2250, 2022. doi: 10.1109/TAFFC.2022.3169001.
- 671 Yi Ding, Chengxuan Tong, Shuailei Zhang, Muyun Jiang, Yong Li, Kevin JunLiang Lim, and Cuntai
 672 Guan. Emtn: A novel transformer for generalized cross-subject eeg emotion recognition. *IEEE*
 673 *Transactions on Neural Networks and Learning Systems*, 2025. doi: 10.1109/TNNLS.2025.
 674 3552603.
- 675 Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. Differential entropy feature for eeg-based emotion
 676 classification. In *2013 6th international IEEE/EMBS conference on neural engineering (NER)*,
 677 pp. 81–84. IEEE, 2013.
- 678 Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, and Xiaoli Li. Tslanet: rethinking
 679 transformers for time series representation learning. In *Proceedings of the 41st International*
 680 *Conference on Machine Learning*, pp. 12409–12428, 2024.
- 681 Bernhard Elsner, Joachim Kugler, and Jan Mehrholz. Transcranial direct current stimulation (tdcs)
 682 for upper limb rehabilitation after stroke: future directions. *Journal of neuroengineering and*
 683 *rehabilitation*, 15(1):106, 2018. doi: 10.1186/s12984-018-0459-7.
- 684 Stephen H Fairclough and Fabien Lotte. Grand challenges in neurotechnology and system neuroer-
 685 gonomics. *Frontiers in Neuroergonomics*, 1:602504, 2020. doi: 10.3389/fnrgo.2020.602504.
- 686 Josef Faller, Carmen Vidaurre, Teodoro Solis-Escalante, Christa Neuper, and Reinhold Scherer. Au-
 687 tocalibration and recurrent adaptation: Towards a plug and play online erd-bci. *IEEE Transactions*
 688 *on Neural Systems and Rehabilitation Engineering*, 20(3):313–319, 2012.
- 689 Birgit Frauscher, Nicolas Von Ellenrieder, Rina Zelmann, Irena Doležalová, Lorella Minotti, André
 690 Olivier, Jeffery Hall, Dominique Hoffmann, Dang Khoa Nguyen, Philippe Kahane, et al. Atlas
 691 of the normal intracranial electroencephalogram: neurophysiological awake activity in different
 692 cortical areas. *Brain*, 141(4):1130–1144, 2018. doi: 10.1093/brain/awy035.
- 693 Octavian Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. *Advances in*
 694 *neural information processing systems*, 31, 2018.

- 702 Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François
 703 Laviotte, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks.
 704 *Journal of machine learning research*, 17(59):1–35, 2016.
- 705 Théo Gnassounou, Antoine Collas, Rémi Flamary, Karim Lounici, and Alexandre Gramfort. Multi-
 706 source and test-time domain adaptation on multivariate signals using spatio-temporal monge
 707 alignment. *arXiv preprint arXiv:2407.14303*, 2024.
- 708 Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian
 709 Brodbeck, Lauri Parkkonen, and Matti S Hämäläinen. Mne software for processing meg and eeg
 710 data. *neuroimage*, 86:446–460, 2014. doi: 10.1016/j.neuroimage.2013.10.027.
- 711 Ning Guo, Xiaojun Wang, Dehao Duanmu, Xin Huang, Xiaodong Li, Yunli Fan, Hailan Li,
 712 Yongquan Liu, Eric Hiu Kwong Yeung, Michael Kai Tsun To, et al. Ssvep-based brain
 713 computer interface controlled soft robotic glove for post-stroke hand function rehabilitation. *IEEE
 714 transactions on neural systems and rehabilitation engineering*, 30:1737–1744, 2022. doi:
 715 10.1109/TNSRE.2022.3185262.
- 716 He He and Dongrui Wu. Transfer learning for brain–computer interfaces: A euclidean space data
 717 alignment approach. *IEEE Transactions on Biomedical Engineering*, 67(2):399–410, 2019. doi:
 718 10.1109/TBME.2019.2913914.
- 719 Shaul Hochstein and Merav Ahissar. View from the top: Hierarchies and reverse hierarchies in the
 720 visual system. *Neuron*, 36(5):791–804, 2002.
- 721 Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by
 722 reducing internal covariate shift. In *International conference on machine learning*, pp. 448–456.
 723 pmlr, 2015.
- 724 Wei-Bang Jiang, Li-Ming Zhao, and Bao-Liang Lu. Large brain model for learning generic repre-
 725 sentations with tremendous eeg data in bci. *arXiv preprint arXiv:2405.18765*, 2024.
- 726 Ce Ju and Cuntai Guan. Tensor-cspnet: A novel geometric deep learning framework for motor
 727 imagery classification. *IEEE Transactions on Neural Networks and Learning Systems*, 34(12):
 728 10955–10969, 2022.
- 729 Valentin Khrulkov, Leyla Mirvakhabova, Evgeniya Ustinova, Ivan Oseledets, and Victor Lempitsky.
 730 Hyperbolic image embeddings. In *Proceedings of the IEEE/CVF conference on computer vision
 731 and pattern recognition*, pp. 6418–6428, 2020.
- 732 Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint
 733 arXiv:1412.6980*, 2014.
- 734 Reinmar Kobler, Jun-ichiro Hirayama, Qibin Zhao, and Motoaki Kawanabe. Spd domain-specific
 735 batch normalization to crack interpretable unsupervised domain adaptation in eeg. *Advances in
 736 Neural Information Processing Systems*, 35:6219–6235, 2022.
- 737 Max Kochurov, Rasul Karimov, and Serge Kozlukov. Geoopt: Riemannian optimization in pytorch.
 738 *arXiv preprint arXiv:2005.02819*, 2020.
- 739 Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguná.
 740 Hyperbolic geometry of complex networks. *Physical Review E—Statistical, Nonlinear, and Soft
 741 Matter Physics*, 82(3):036106, 2010. doi: 10.1103/PhysRevE.82.036106.
- 742 J Ph Lachaux, D Rudrauf, and P Kahane. Intracranial eeg and human brain mapping. *Journal of
 743 Physiology-Paris*, 97(4-6):613–628, 2003. doi: 10.1016/j.jphysparis.2004.01.018.
- 744 Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and
 745 Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer
 746 interfaces. *Journal of neural engineering*, 15(5):056013, 2018. doi: 10.1088/1741-2552/aace8c.
- 747 Chen-Yu Lee, Tanmay Batra, Mohammad Haris Baig, and Daniel Ulbricht. Sliced wasserstein
 748 discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on
 749 computer vision and pattern recognition*, pp. 10285–10295, 2019.

- 756 Yi-Chieh Lee, Wen-Chieh Lin, Fu-Yin Cherng, and Li-Wei Ko. A visual attention monitor based
 757 on steady-state visual evoked potential. *IEEE Transactions on neural systems and rehabilitation*
 758 *engineering*, 24(3):399–408, 2015. doi: TNSRE.2015.2501378.
- 759
- 760 Shanglin Li, Motoaki Kawanabe, and Reinmar J Kobler. Spdim: Source-free unsupervised condi-
 761 tional and label shift adaptation in eeg. In *The Thirteenth International Conference on Learning*
 762 *Representations*, 2025.
- 763
- 764 Siyang Li, Ziwei Wang, Hanbin Luo, Lieyun Ding, and Dongrui Wu. T-time: Test-time information
 765 maximization ensemble for plug-and-play bcis. *IEEE Transactions on Biomedical Engineering*,
 766 71(2):423–432, 2023. doi: 10.1109/TBME.2023.3303289.
- 767
- 768 Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rak-
 769 tomamonyjy, and Florian Yger. A review of classification algorithms for eeg-based brain-
 770 computer interfaces: a 10 year update. *Journal of neural engineering*, 15(3):031005, 2018. doi:
 771 10.1088/1741-2552/aab2f2.
- 772
- 773 Ravikiran Mane, Effie Chew, Karen Chua, Kai Keng Ang, Neethu Robinson, A Prasad Vinod,
 774 Seong-Whan Lee, and Cuntai Guan. Fbcnet: A multi-view convolutional neural network for
 775 brain-computer interface. *arXiv preprint arXiv:2104.01233*, 2021.
- 776
- 777 Apolline Mellot, Antoine Collas, Pedro LC Rodrigues, Denis Engemann, and Alexandre Gramfort.
 778 Harmonizing and aligning m/eeg datasets with covariance-based techniques to enhance predictive
 779 regression modeling. *Imaging Neuroscience*, 1:1–23, 2023. doi: 10.1162/imag_a_00040.
- 780
- 781 Pascal Mettes, Mina Ghadimi Atigh, Martin Keller-Ressel, Jeffrey Gu, and Serena Yeung. Hyper-
 782 bolic deep learning in computer vision: A survey. *International Journal of Computer Vision*, 132
 783 (9):3484–3508, 2024. doi: 10.1007/s11263-024-02043-5.
- 784
- 785 Gal Mishne, Zhengchao Wan, Yusu Wang, and Sheng Yang. The numerical stability of hyperbolic
 786 representation learning. In *International Conference on Machine Learning*, pp. 24925–24949.
 787 PMLR, 2023.
- 788
- 789 Masaki Nakanishi, Yijun Wang, Yu-Te Wang, and Tzzy-Ping Jung. A comparison study of canonical
 790 correlation analysis based methods for detecting steady-state visual evoked potentials. *PloS one*,
 791 10(10):e0140703, 2015. doi: 10.1371/journal.pone.0140703.
- 792
- 793 Xuan Son Nguyen. The Gyro-structure of some matrix manifolds. In *NeurIPS*, 2022.
- 794
- 795 Ernst Niedermeyer and FH Lopes da Silva. *Electroencephalography: basic principles, clinical*
 796 *applications, and related fields*. Lippincott Williams & Wilkins, 2005. ISBN 978-0190228484.
- 797
- 798 Yudong Pan, Jianbo Chen, Yangsong Zhang, and Yu Zhang. An efficient cnn-lstm network with
 799 spectral normalization and label smoothing technologies for ssvep frequency recognition. *Journal*
 800 *of neural engineering*, 19(5):056014, 2022a. doi: 10.1088/1741-2552/ac8dc5.
- 801
- 802 Yue-Ting Pan, Jing-Lun Chou, and Chun-Shu Wei. Matt: A manifold attention network for eeg
 803 decoding. In *Advances in Neural Information Processing Systems*, volume 35, pp. 31116–31129,
 804 2022b.
- 805
- Jaak Panksepp. Cross-species affective neuroscience decoding of the primal affective experiences
 806 of humans and related animals. *PloS one*, 6(9):e21236, 2011.
- 807
- Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor
 808 Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-
 809 performance deep learning library. *Advances in neural information processing systems*, 32, 2019.
- 810
- Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier
 811 Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn:
 812 Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011.
 813 doi: 10.5555/1953048.2078195.

- Wei Peng, Tuomas Varanka, Abdelrahman Mostafa, Henglin Shi, and Guoying Zhao. Hyperbolic deep neural networks: A survey. *IEEE Transactions on pattern analysis and machine intelligence*, 44(12):10023–10044, 2021. doi: 10.1109/TPAMI.2021.3136921.
- John G Ratcliffe. *Foundations of Hyperbolic Manifolds*. Springer, 2006.
- Ievgen Redko, Emilie Morvant, Amaury Habrard, Marc Sebban, and Younes Bennani. *Advances in domain adaptation theory*. Elsevier, 2019.
- 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. doi: 10.1109/TBME.2018.2889705.
- Raphaëlle N. Roy, Marcel F. Hinss, Ludovic Darmet, Simon Ladouce, Emilie S. Jahanpour, Bertille Somon, Xiaoqi Xu, Nicolas Drougard, Frédéric Dehais, and Fabien Lotte. Retrospective on the First Passive Brain-Computer Interface Competition on Cross-Session Workload Estimation. *Frontiers in Neuroergonomics*, 3:838342, 2022. doi: 10.3389/fnrgo.2022.838342.
- Eduardo Santamaria-Vazquez, Victor Martinez-Cagigal, Fernando Vaquerizo-Villar, and Roberto Hornero. Eeg-inception: a novel deep convolutional neural network for assistive erp-based brain–computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(12):2773–2782, 2020. doi: 10.1109/TNSRE.2020.3048106.
- 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. doi: doi.org/10.1002/hbm.23730.
- Ryohei Shimizu, YUSUKE Mukuta, and Tatsuya Harada. Hyperbolic neural networks++. In *International Conference on Learning Representations*, 2021.
- Yonghao Song, Qingqing Zheng, Bingchuan Liu, and Xiaorong Gao. Eeg conformer: Convolutional transformer for eeg decoding and visualization. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 31:710–719, 2022. doi: 10.1109/TNSRE.2022.3230250.
- Saurabh Sonkusare, Vinh T Nguyen, Rosalyn Moran, Johan van der Meer, Yudan Ren, Nikitas Koussis, Sasha Dionisio, Michael Breakspear, and Christine Guo. Intracranial-eeg evidence for medial temporal pole driving amygdala activity induced by multi-modal emotional stimuli. *Cortex*, 130:32–48, 2020. doi: 10.1016/j.cortex.2020.05.018.
- Nazmi Sofian Suhaimi, James Mountstephens, and Jason Teo. Eeg-based emotion recognition: a state-of-the-art review of current trends and opportunities. *Computational intelligence and neuroscience*, 2020(1):8875426, 2020. doi: 10.1155/2020/8875426.
- Sai Sun, Hongbo Yu, Rongjun Yu, and Shuo Wang. Functional connectivity between the amygdala and prefrontal cortex underlies processing of emotion ambiguity. *Translational psychiatry*, 13(1):334, 2023. doi: 10.1038/541398-023-02625-W.
- Michael Tangermann, Klaus-Robert Müller, Ad Aertsen, Niels Birbaumer, Christoph Braun, Clemens Brunner, Robert Leeb, Carsten Mehring, Kai J Miller, Gernot Mueller-Putz, Guido Nolte, et al. Review of the BCI competition IV. *Frontiers in Neuroscience*, 6:55, 2012.
- Yelena Tonoyan, Theerasak Chanwimalueang, Danilo P Mandic, and Marc M Van Hulle. Discrimination of emotional states from scalp-and intracranial eeg using multiscale rényi entropy. *PLoS One*, 12(11):e0186916, 2017. doi: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0186916>.
- William Turner, Tessel Blom, and Hinze Hogendoorn. Visual information is predictively encoded in occipital alpha/low-beta oscillations. *Journal of Neuroscience*, 43(30):5537–5545, 2023. doi: 10.1523/JNEUROSCI.0135-23.2023.
- Abraham Albert Ungar. *Analytic Hyperbolic Geometry and Albert Einstein’s Special Theory of Relativity (Second Edition)*. World Scientific, 2022.

- 864 Kamil Vlcek, Iveta Fajnerova, Tereza Nekovarova, Lukas Hejtmanek, Radek Janca, Petr Jezdik,
 865 Adam Kalina, Martin Tomasek, Pavel Krsek, Jiri Hammer, et al. Mapping the scene and object
 866 processing networks by intracranial eeg. *Frontiers in human neuroscience*, 14:561399, 2020. doi:
 867 10.3389/fnhum.2020.561399.
- 868 Jiquan Wang, Sha Zhao, Zhiling Luo, Yangxuan Zhou, Haiteng Jiang, Shijian Li, Tao Li, and
 869 Gang Pan. Cbramod: A criss-cross brain foundation model for eeg decoding. *arXiv preprint*
 870 *arXiv:2412.07236*, 2024a.
- 871 Rui Wang, Chen Hu, Ziheng Chen, Xiao-Jun Wu, and Xiaoning Song. A grassmannian manifold
 872 self-attention network for signal classification. In *Proceedings of the Thirty-Third International*
 873 *Joint Conference on Artificial Intelligence*, pp. 5099–5107, 2024b.
- 874 Rui Wang, Chen Hu, Xiaoning Song, Xiaojun Wu, Nicu Sebe, and Ziheng Chen. Towards a gen-
 875 eral attention framework on gyrovector spaces for matrix manifolds. In *The Thirty-ninth Annual*
 876 *Conference on Neural Information Processing Systems*, 2025.
- 877 Yijun Wang, Xiaogang Chen, Xiaorong Gao, and Shangkai Gao. A benchmark dataset for ssvep-
 878 based brain-computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation En-*
 879 *gineering*, 25(10):1746–1752, 2016. doi: 10.1109/TNSRE.2016.2627556.
- 880 Xiaoxi Wei, A. Aldo Faisal, Moritz Grosse-Wentrup, Alexandre Gramfort, Sylvain Chevallier,
 881 Vinay Jayaram, Camille Jeunet, Stylianos Bakas, Siegfried Ludwig, Konstantinos Barmpas,
 882 Mehdi Bahri, Yannis Panagakis, Nikolaos Laskaris, Dimitrios A. Adamos, Stefanos Zafeiriou,
 883 William C. Duong, Stephen M. Gordon, Vernon J. Lawhern, Maciej Śliwowski, Vincent Rouanne,
 884 and Piotr Tempczyk. 2021 BEETL Competition: Advancing Transfer Learning for Subject In-
 885 dependence and Heterogenous EEG Data Sets. In Douwe Kiela, Marco Ciccone, and Barbara
 886 Caputo (eds.), *Proceedings of the NeurIPS 2021 Competitions and Demonstrations Track*, vol-
 887 ume 176 of *Proceedings of Machine Learning Research*, pp. 205–219. PMLR, 2022.
- 888 Dongrui Wu, Yifan Xu, and Bao-Liang Lu. Transfer learning for eeg-based brain-computer inter-
 889 faces: A review of progress made since 2016. *IEEE Transactions on Cognitive and Developmental*
 890 *Systems*, 14(1):4–19, 2020. doi: 10.1109/TCDS.2020.3007453.
- 891 Or Yair, Felix Dietrich, Ronen Talmon, and Ioannis G Kevrekidis. Domain adaptation with optimal
 892 transport on the manifold of spd matrices. *arXiv preprint arXiv:1906.00616*, 2019.
- 893 Hongwei Yong, Jianqiang Huang, Deyu Meng, Xiansheng Hua, and Lei Zhang. Momentum batch
 894 normalization for deep learning with small batch size. In *European Conference on Computer*
 895 *Vision*, pp. 224–240. Springer, 2020.
- 896 Shubin Zhang, Dong An, Jincun Liu, Jiannan Chen, Yaoguang Wei, and Fuchun Sun. Dynamic de-
 897 composition graph convolutional neural network for ssvep-based brain-computer interface. *Neu-*
 898 *ral Networks*, 172:106075, 2024. doi: 10.1016/j.neunet.2023.12.029.
- 899 Han Zhao, Remi Tachet Des Combes, Kun Zhang, and Geoffrey Gordon. On learning invariant
 900 representations for domain adaptation. In *International conference on machine learning*, pp.
 901 7523–7532. PMLR, 2019.
- 902
 903
 904
 905
 906
 907
 908
 909
 910
 911
 912
 913
 914
 915
 916
 917

918	APPENDIX CONTENTS	
919		
920	A Large language model usage statement	19
921		
922	B Gyrovector spaces	19
923	B.1 General definition	19
924		
925	C Lorentz operations	20
926	C.1 Riemannian operators	20
927	C.2 Gyrovector operators	20
928	C.3 Lorentz non-linear activation	21
929	C.4 Lorentz concatenation	21
930	C.5 Lorentz fully-connected layer	21
931	C.6 Lorentz convolutional layer	21
932	C.7 Lorentz multinomial logistic regression	22
933	C.8 δ -hyperbolicity	22
934		
935	D HEEGNet details	23
936	D.1 Algorithm	23
937	D.2 Model architecture	24
938	D.3 Software and hardware	24
939	D.4 Computational Cost	25
940		
941	E Full experiments	26
942	E.1 Dataset details	26
943	E.2 Visually evoked potentials	27
944	E.3 Emotion recognition	28
945	E.4 Intracranial EEG results	29
946		
947	F Fully hyperbolic neural network experiment	30
948		
949		
950		
951		
952		
953		
954		
955		
956		
957		
958		
959		
960		
961		
962		
963		
964		
965		
966		
967		
968		
969		
970		
971		

972 **A LARGE LANGUAGE MODEL USAGE STATEMENT**
973

974 Large language models were partially used in this article to refine the contents.
975

976 **B GYROVECTOR SPACES**
977

978 **B.1 GENERAL DEFINITION**
979

980 This subsection briefly reviews the gyrovector space (Ungar, 2022), which generalizes the vector
981 structure into manifold. It has shown great success in building neural networks over different Rie-
982 mannian spaces, such as hyperbolic (Ganea et al., 2018; Chami et al., 2019; Shimizu et al., 2021),
983 symmetric positive definite (SPD) (Nguyen, 2022), and Grassmannian (Nguyen, 2022) manifold.
984

985 We start from the gyrogroup. Intuitively, gyrogroups are natural generalizations of groups. Unlike
986 groups, gyrogroups are non-associative but have gyroassociativity characterized by gyrations.
987

988 **Definition B.1** (Gyrogroups (Ungar, 2022)). Given a nonempty set G with a binary operation $\oplus : G \times G \rightarrow G$, (G, \oplus) forms a gyrogroup if its binary operation satisfies the following axioms for any $p, q, z \in G$:
989

990 (G1) There is at least one element $e \in G$ called a left identity (or neutral element) such that $e \oplus p = p$.
991

992 (G2) There is an element $\ominus p \in G$ called a left inverse of p such that $\ominus p \oplus p = e$.
993

994 (G3) There is an automorphism $\text{gyr}[p, q] : G \rightarrow G$ for each $p, q \in G$ such that
995

$$p \oplus (q \oplus z) = (p \oplus q) \oplus \text{gyr}[p, q]z \quad (\text{Left Gyroassociative Law}).$$
996

997 The automorphism $\text{gyr}[p, q]$ is called the gyroautomorphism, or the gyration of G generated by p, q .
998

999 (G4) Left reduction law: $\text{gyr}[p, q] = \text{gyr}[p \oplus q, q]$.
1000

1000 **Definition B.2** (Gyrocommutative Gyrogroups (Ungar, 2022)). A gyrogroup (G, \oplus) is gyrocom-
1001 mutative if it satisfies
1002

$$p \oplus q = \text{gyr}[p, q](q \oplus p) \quad (\text{Gyrocommutative Law}).$$
1003

1003 Similarly, the gyrovector space generalizes the vector space,
1004

1004 **Definition B.3** (Gyrovector Spaces (Chen et al., 2025b)). A gyrocommutative gyrogroup (G, \oplus)
1005 equipped with a scalar gyromultiplication $\otimes : \mathbb{R} \times G \rightarrow G$ is called a gyrovector space if it satisfies
1006 the following axioms for $s, t \in \mathbb{R}$ and $p, q, z \in G$:
1007

1008 (V1) Identity Scalar Multiplication: $1 \otimes p = p$.
1009

1010 (V2) Scalar Distributive Law: $(s + t) \otimes p = s \otimes p \oplus t \otimes p$.
1011

1011 (V3) Scalar Associative Law: $(st) \otimes p = s \otimes (t \otimes p)$.
1012

1012 (V4) Gyroautomorphism: $\text{gyr}[p, q](t \otimes z) = t \otimes \text{gyr}[p, q]z$.
1013

1013 (V5) Identity Gyroautomorphism: $\text{gyr}[s \otimes p, t \otimes p] = \mathbb{I}$, where \mathbb{I} is the identity map.
1014

1015 The vector space, equipped with addition and scalar multiplication, forms the foundation of Eu-
1016 clidean deep learning. Similarly, the gyrovector space, endowed with gyroaddition and scalar gyro-
1017 multiplication, offers a powerful tool for designing neural networks over non-Euclidean manifolds.
1018

1019
1020
1021
1022
1023
1024
1025

1026 **C LORENTZ OPERATIONS**

1028 **C.1 RIEMANNIAN OPERATORS**

1030 The exponential map $\text{Exp}_p^K : T_p \mathbb{L}_K^n \rightarrow \mathbb{L}_K^n$ and logarithmic map $\text{Log}_p^K : T_p \mathbb{L}_K^n \rightarrow \mathbb{L}_K^n$ project
 1031 points between the manifold $p_i \in \mathbb{L}_K^n$ and the tangent space $v_i \in T_p \mathbb{L}_K^n$ at point $p \in \mathbb{L}_K^n$.

1034
$$\text{Exp}_p^K(v) = \cosh(\alpha)p + \sinh(\alpha)\frac{v}{\alpha}, \quad \text{with } \alpha = \sqrt{-K} \|v\|_{\mathcal{L}}, \quad \|v\|_{\mathcal{L}} = \sqrt{\langle v, v \rangle_{\mathcal{L}}} \quad (10)$$

1037
$$\text{Log}_p^K(q) = \frac{\cosh^{-1}(\beta)}{\sqrt{\beta^2 - 1}} \cdot (q - \beta p), \quad \text{with } \beta = K \langle p, q \rangle_{\mathcal{L}} \quad (11)$$

1041 To transport points $v_i \in T_p \mathbb{L}_K^n$ from the tangent space at p to the tangent space at q , parallel transport
 1042 $\text{PT}_{p \rightarrow q}(v) \in \mathbb{L}_K^n$ can be used:

1044
$$\text{PT}_{p \rightarrow q}(v) = v - \frac{K(q, v)_K}{1 + K(p, q)_K}(p + q) \quad (12)$$

1048 **C.2 GYROVECTOR OPERATORS**

1050 The definitions in Eqs. (3) and (4) are direct generalizations of Euclidean vector operations. In
 1051 Euclidean space, vector addition and scalar multiplication can be understood geometrically as oper-
 1052 ations on rays emanating from the origin.

1053 Let us first review Euclidean geometry. For \mathbb{R}^n , we have

1055 $T_p \mathbb{R}^n = T_q \mathbb{R}^n = \mathbb{R}^n, \quad (13)$

1057 $\text{Log}_p(q) = q - p, \quad (14)$

1058 $\text{Exp}_p(v) = v + p \quad (15)$

1059 $\text{PT}_{p \rightarrow q}(v) = v, \quad (16)$

1061 where $p, q \in \mathbb{R}^n$ and $v \in T_x \mathbb{R}^n$

1063 To compute the Euclidean vector addition $p + q$, one may regard q as the ray from $\mathbf{0}$ to q , parallel
 1064 translate this ray to the base point x , and then shoot it out from p . The above process can be
 1065 expressed as

1066
$$x + y = \text{Exp}_p(\text{PT}_{\mathbf{0} \rightarrow p}(\text{Log}_{\mathbf{0}}(y))). \quad (17)$$

1068 Similarly, Euclidean scalar multiplication corresponds to taking the ray from $\mathbf{0}$ to x , scaling its
 1069 length by t , and then shooting it out from the origin again:

1071
$$t \odot p = \text{Exp}_{\mathbf{0}}(t \text{Log}_{\mathbf{0}}(p)) = tp. \quad (18)$$

1073 Hence, the Lorentz gyroaddition and gyromultiplication extend this geometric intuition of Euclidean
 1074 linear operations to curved manifolds.

1076 **Expressions.** Let $p = [p_t, p_s]^\top$ and $q = [q_t, q_s]^\top$ be points in \mathbb{L}_K^n , where $p_t, q_t \in \mathbb{R}$ are the
 1077 time scalars, and $p_s, q_s \in \mathbb{R}^n$ are the spatial parts. Let $t \in \mathbb{R}$ be a real scalar. The following
 1078 reviews the closed-form expression of the Lorentz gyro operators, which are more efficient than the
 1079 Riemannian definitions Eqs. (3) and (4) (Chen et al., 2025b, Sec. 6.1). The Lorentz gyroaddition

and gyromultiplication have the closed-form solution:

$$\text{Gyroaddition: } p \oplus_K^M q = \begin{cases} p, & q = \bar{0}, \\ q, & p = \bar{0}, \\ \left[\begin{array}{c} 1 - \frac{D-KN}{|K|} \\ \hline D+KN \\ \frac{2(A_s p_s + A_q q_s)}{D+KN} \end{array} \right], & \text{Others.} \end{cases} \quad (19)$$

$$\text{Gyromultiplication: } t \otimes_K^M p = \begin{cases} \bar{0}, & t = 0 \vee p = \bar{0}, \\ \frac{1}{\sqrt{|K|}} \left[\begin{array}{c} \cosh(t \cosh^{-1}(\sqrt{|K|} p_t)) \\ \hline \sinh(t \cosh^{-1}(\sqrt{|K|} p_t)) \\ \|p_s\| p_s \end{array} \right], & t \neq 0, \end{cases} \quad (20)$$

where $A_s = ab^2 - 2Kbs_{pq} - Kan_q$ and $A_q = b(a^2 + Kn_p)$ with the following:

$$\begin{aligned} a &= 1 + \sqrt{|K|} p_t, & b &= 1 + \sqrt{|K|} q_t, & n_p &= \|p_s\|^2, & n_q &= \|q_s\|^2, & s_{pq} &= \langle p_s, q_s \rangle, \\ D &= a^2 b^2 - 2Kabs_{pq} + K^2 n_p n_q, & N &= a^2 n_q + 2abs_{pq} + b^2 n_p. \end{aligned}$$

C.3 LORENTZ NON-LINEAR ACTIVATION

The Lorentz ELU activation applies the activation function to the space components and concatenates them with the time component:

$$C_{\text{activated}} = \left[\begin{array}{c} \sqrt{\|\text{ELU}(\mathbf{p}_s)\|^2 - 1/K} \\ \text{ELU}(\mathbf{p}_s) \end{array} \right]. \quad (21)$$

C.4 LORENTZ CONCATENATION

Given a set of hyperbolic points $\{p_i \in \mathbb{L}_K^n\}_{i=1}^N$, the Lorentz direct concatenation is defined as:

$$\mathbf{y} = \text{HCat}(\{p_i\}_{i=1}^N) = \left[\sqrt{\sum_{i=1}^N p_{it}^2 + \frac{N-1}{K}}, p_{1s}^T, \dots, p_{Ns}^T \right]^T, \quad (22)$$

where $\mathbf{y} \in \mathbb{L}_K^{nN} \subset \mathbb{R}^{nN+1}$.

C.5 LORENTZ FULLY-CONNECTED LAYER

Let $p \in \mathbb{L}_K^n$ denote the input vector and $\mathbf{W} \in \mathbb{R}^{m \times n+1}$, $v \in \mathbb{R}^{n+1}$ the weight parameters, the Lorentz fully-connected layer (LFC) is defined as:

$$y = \text{LFC}(p) = \left[\begin{array}{c} \sqrt{\|\psi(\mathbf{W}p + \mathbf{b})\|^2 - 1/K} \\ \psi(\mathbf{W}p + \mathbf{b}) \end{array} \right] \quad (23)$$

$$\phi(\mathbf{W}p, v) = \lambda \sigma(v^T p + b') \frac{\mathbf{W}\psi(p) + \mathbf{b}}{\|\mathbf{W}\psi(p) + \mathbf{b}\|} \quad (24)$$

where $\lambda > 0$ denotes a trainable scaling factor, $\mathbf{b} \in \mathbb{R}^n$ is the bias vector, and ψ and σ represent the activation and sigmoid functions, respectively.

C.6 LORENTZ CONVOLUTIONAL LAYER

Given a hyperbolic input feature map $p = \{\mathbf{p}_{h,w} \in \mathbb{L}_K^n\}_{h,w=1}^{H,W}$ as an ordered set of n -dimensional hyperbolic feature vectors, the features within the receptive field of the kernel $\mathbf{K} \in \mathbb{R}^{m \times n \times \tilde{H} \times \tilde{W}}$ are $\{\mathbf{p}_{h'+\delta_{\tilde{h}}, w'+\delta_{\tilde{w}}} \in \mathbb{L}_K^n\}_{\tilde{h}, \tilde{w}=1}^{\tilde{H}, \tilde{W}}$, where (h', w') denotes the starting position and δ is the stride parameter. The Lorentz convolutional layer is defined as $\text{LFC}(\text{HCAT}(\{\mathbf{p}_{h'+\delta_{\tilde{h}}, w'+\delta_{\tilde{w}}} \in \mathbb{L}_K^n\}_{\tilde{h}, \tilde{w}=1}^{\tilde{H}, \tilde{W}}))$, where HCAT and LFC, denote hyperbolic concatenation and a Lorentz fully-connected layer performing the affine transformation and parameterizing the kernel and bias.

1134
1135

C.7 LORENTZ MULTINOMIAL LOGISTIC REGRESSION

1136 Similar to the Euclidean MLR, the Lorentz MLR performs classification by measuring distances to
 1137 decision hyperplanes. The output logit for class c is computed from the hyperbolic distance between
 1138 p and its corresponding hyperplane. For a hyperbolic input point $p \in \mathbb{L}_K^n$ and C possible classes,
 1139 each class $c \in \{1, \dots, C\}$ is associated with a decision hyperplane parameterized by $a_c \in \mathbb{R}$ and
 1140 $z_c \in \mathbb{R}^n$.

$$1141 v_{z_c, a_c}(p) = \frac{1}{\sqrt{-K}} \text{sign}(\alpha_c) \beta_c \left| \sinh^{-1} \left(\sqrt{-K} \frac{\alpha_c}{\beta_c} \right) \right|, \quad (25)$$

1143 where

$$1144 \alpha_c = \cosh(\sqrt{-K} a_c) \langle z_c, p_s \rangle - \sinh(\sqrt{-K} a_c) \|z_c\| p_t,$$

$$1145 \beta_c = \sqrt{\|\cosh(\sqrt{-K} a_c) z_c\|^2 - (\sinh(\sqrt{-K} a_c) \|z_c\|)^2}.$$

1148 C.8 δ -HYPERBOLICITY

1149

1150 Khrulkov et al. (2020) introduced δ -hyperbolicity as a measure of the degree of tree-like structure
 1151 inherent in embeddings. The idea is to find the smallest value of δ for which the triangle inequality
 1152 holds via the Gromov product. In this formulation, the definition of a hyperbolic space in terms of
 1153 the Gromov product can be seen as the metric relations between any four points are the same as they
 1154 would be in a tree, up to an additive constant δ . Formally, given the Lorentz model \mathbb{L}_K^n with distance
 1155 $d_{\mathcal{L}}$, the Gromov product of $z, q \in \mathbb{L}_K^n$ with respect to $p \in \mathbb{L}_K^n$ as:

$$1156 \quad (p, q)_z = \frac{1}{2} (d_{\mathcal{L}}(p, z) + d_{\mathcal{L}}(q, z) - d_{\mathcal{L}}(p, q)). \quad (26)$$

1158 The Lorentz model is said to be δ -hyperbolic for some $\delta \geq 0$ if it satisfies the four-point condition,
 1159 which states that for any $p, q, z, w \in \mathbb{L}_K^n$:

$$1161 \quad (p, q)_w \geq \min\{(p, z)_w, (q, z)_w\} - \delta. \quad (27)$$

1162 The metric relations between any four points are the same as they would be in a tree, up to the
 1163 additive constant δ . The lower $\delta \geq 0$ is, the higher the hyperbolicity of the embedding.
 1164

1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187

1188 D HEEGNET DETAILS
11891190 D.1 ALGORITHM
11911192 **Algorithm 1** Hyperbolic domain-specific momentum batch normalization (HDSMBN)
1193**Input:**

1194 batch $\mathcal{B}_k = \{p_i \in \mathbb{L}_K^n, d(i) \in \mathcal{D}_{\mathcal{B}_k}\}_{i=1}^M$ at training step k , $d(i)$ indicates the associated domain d
 1195 domain-specific running mean $\tilde{\mu}_{k-1}^d$ ($\tilde{\mu}_0^d = \bar{0}$) and variance $\tilde{\nu}_{k-1}^{2d}$ ($\tilde{\nu}_0^{2d} = 1$) for training
 1196 domain-specific running mean $\hat{\mu}_{k-1}^d$ ($\hat{\mu}_0^d = \bar{0}$) and variance $\hat{\nu}_{k-1}^{2d}$ ($\hat{\nu}_0^{2d} = 1$) for testing
 1197 momentum for training and testing $\eta_{train(k)}, \eta_{test} \in [0, 1]$, learnable parameter ν_ϕ^2
 1198

Output: normalized batch $\{\tilde{p}_i\} = \text{HBN}^{d(i)}(p_i)$ **if** training **then** Compute domain-specific batch mean μ_k^d and variance ν_k^{2d} ▷ using Eq. (2) $\tilde{\mu}_k^d = \text{wFM}_{\eta_{train(k)}}(\tilde{\mu}_{k-1}^d, \mu_k^d)$ ▷ update running mean using Eq. (2) $\tilde{\nu}_k^{2d} = (1 - \eta_{train(k)}) \tilde{\nu}_{k-1}^{2d} + \eta_{train(k)} \nu_k^{2d}$ $\hat{\mu}_k^d = \text{wFM}_{\eta_{test}}(\hat{\mu}_{k-1}^d, \mu_k^d)$ ▷ update running mean using Eq. (2) $\hat{\nu}_k^{2d} = (1 - \eta_{test}) \hat{\nu}_{k-1}^{2d} + \eta_{test} \nu_k^{2d}$ **end if** $(\mu_k^d, \nu_k^{2d}) \leftarrow (\tilde{\mu}_k^d, \tilde{\nu}_k^{2d})$ **if** training **else** $(\hat{\mu}_k^d, \hat{\nu}_k^{2d})$ $\tilde{p}_i = \frac{\nu_\phi^2}{\sqrt{\nu_k^{2d} + \epsilon}} (\ominus \mu_k^d \oplus p_i)$ ▷ use Eq. (7) to recentering and rescale each domain1212 **Algorithm 2** Horospherical Hyperbolic Sliced-Wasserstein loss (HHSW)
1213**Input:**

1214 batch $\mathcal{B} = \{\tilde{p}_i \in \mathbb{L}_K^n, d(i) \in \mathcal{D}_{\mathcal{B}}\}_{i=1}^M$, $d(i)$ indicates the associated domain d
 1215 number of slices $S = 1000$, exponent $p = 2$

Output: scalar loss \mathcal{L}_{swd} Initialize $\mathcal{L}_{\text{swd}} \leftarrow 0$ **for** each domain $d \in \mathcal{D}_{\mathcal{B}_k}$ **do** Extract domain-specific samples $\mathcal{P}^d = \{p_i \mid d(i) = d\}$ Sample Gaussian noise $Z^d \sim \mathcal{N}(0, I)$ with shape $\text{shape}(\mathcal{P}^d)$ Normalize: $Z^d \leftarrow \frac{Z^d}{\|Z^d\|_2 + \epsilon}$ Map to hyperbolic manifold: $Q^d \leftarrow \exp_0^K(Z^d)$ Compute domain loss: $\ell^d \leftarrow \text{HHSW}_p^p(\mathcal{P}^d, Q^d)$ ▷ Eq. (6) $\mathcal{L}_{\text{swd}} \leftarrow \mathcal{L}_{\text{swd}} + \ell^d$ **end for****return** \mathcal{L}_{swd}

1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241

D.2 MODEL ARCHITECTURE

Table 6: HEEGNet architecture details. P : electrodes; T : temporal samples; C : classes.

Layer	Output (dim)	Parameter (dim)	Operation	Space
<i>Input: $1 \times P \times T$</i>				
TempConv	$8 \times P \times T$	$8 \times 1 \times 1 \times 64$	convolution	Euclidean
BN	$8 \times P \times T$	8	batch norm	Euclidean
SpatConv	$16 \times 1 \times T$	$16 \times 8 \times P \times 1$	depthwise conv	Euclidean
BN	$16 \times 1 \times T$	16	batch norm	Euclidean
Activation	$16 \times 1 \times T$	–	ELU	Euclidean
AvgPool	$16 \times 1 \times \lfloor T/4 \rfloor$	–	pooling	Euclidean
Dropout	$16 \times 1 \times \lfloor T/4 \rfloor$	–	dropout=0.25	Euclidean
DepthConv	$16 \times 1 \times \lfloor T/4 \rfloor$	$16 \times 1 \times 1 \times 16$	depthwise conv	Euclidean
ProjX	$17 \times \lfloor T/4 \rfloor$	–	projection	Euclidean
PointConv	$17 \times \lfloor T/4 \rfloor$	–	pointwise conv	Hyperbolic
DSMDBN ₍₁₎	$17 \times \lfloor T/4 \rfloor$	–	DSMDBN	Hyperbolic
Activation	$17 \times \lfloor T/4 \rfloor$	–	ELU	Hyperbolic
AvgPool	$17 \times \lfloor T/32 \rfloor$	–	pooling	Hyperbolic
Flatten	$16 \cdot \lfloor T/32 \rfloor + 1$	–	flatten	Hyperbolic
MLR	C	$(16 \cdot \lfloor T/32 \rfloor + 1) \times C$	-	Hyperbolic

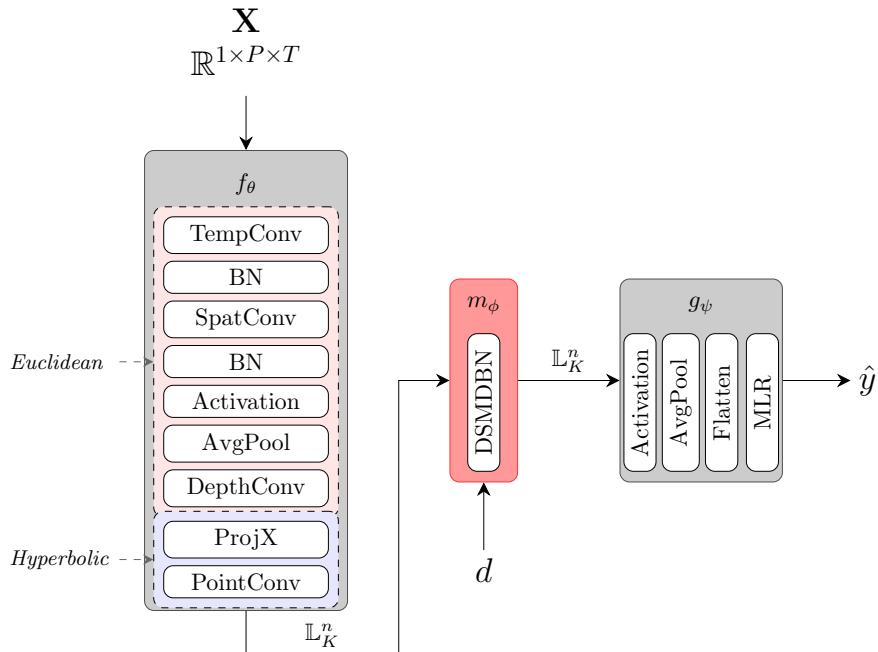


Figure 3: HEEGNet architecture.

D.3 SOFTWARE AND HARDWARE

We used publicly available Python code for baseline methods and implemented custom methods using the packages torch (Paszke et al., 2019), scikit-learn (Pedregosa et al., 2011), geoopt (Kochurov et al., 2020). We conducted all experiments on standard computation PCs with 32-core CPUs, 128 GB of RAM, and a single GPU.

1296 **D.4 COMPUTATIONAL COST**
 1297

1298 Hyperbolic neural networks remain in their early development stage and introduce additional com-
 1299 putational cost, because of the need for exponential and logarithmic mappings. To evaluate the
 1300 practical computational cost and the efficiency of our implementations, we compared the per-epoch
 1301 runtimes of four models: the original EEGNet, HEEGNet without SFUDA, HEEGNet with mo-
 1302 ments alignment (DSMDBN₁), and HEEGNet with DSMDBN.

1303 Table 7: **Hyperbolic computational costs.** Comparisons of the runtime (seconds) per epoch.
 1304

Model	SFUDA	Cost (seconds/epoch)
EEGNet	w/o	0.420
HEEGNet	w/o	1.157
HEEGNet	DSMDBN ₁	1.320
HEEGNet	DSMDBN	1.451

1350 **E FULL EXPERIMENTS**

1351

1352 **E.1 DATASET DETAILS**

1353

1354 **Nakanishi2015** is an SSVEP dataset including EEG from 9 subjects recorded with 8 channels at 256
 1355 Hz. Each subject performed a 12-class joint frequency-phase modulation paradigm with 15 trials
 1356 per class, each trial lasting 4.15 seconds. The dataset consists of a single session per subject and was
 1357 originally designed to evaluate online BCI performance. This benchmark is widely used for 12-class
 1358 SSVEP decoding.

1359 **Wang2016** is an SSVEP dataset including EEG from 34 subjects recorded with 64 channels at 250
 1360 Hz. Each subject performed a 40-class visual stimulation paradigm, where 40 flickering targets
 1361 were presented. The experiment comprised 6 blocks per subject, with 40 trials per block and 6 trials
 1362 per class in total. Each trial lasted 6 seconds, and subjects were instructed to gaze at the cued target
 1363 while avoiding blinks during stimulation. This dataset provides a large-scale benchmark for 40-class
 1364 SSVEP decoding.

1365 **CBVEP** is a c-VEP and burst-VEP dataset including EEG from 12 subjects recorded with 32 chan-
 1366 nels at 500 Hz. Each subject performed a 4-class visual stimulation paradigm, with 15 trials per
 1367 class and trial duration of 2.2 seconds. EEG was recorded using a BrainProduct LiveAmp 32 sys-
 1368 tem with electrodes placed according to the 10–20 system, referenced to FCz and grounded at FPz.
 1369 Participants focused on cued targets during each stimulation phase, and post-experiment subjective
 1370 ratings of visual comfort, tiredness, and intrusiveness were also collected. This dataset provides a
 1371 benchmark for 4-class c-VEP decoding.

1372 **SEED** is an emotion recognition dataset including EEG and eye movement data from 15 subjects,
 1373 recorded with a 62-channel system. Each subject performed a 3-class emotion elicitation task, where
 1374 they watched 15 different film clips (approximately 4 minutes each) designed to evoke positive,
 1375 neutral, or negative emotions. The experiment consisted of 15 trials, with each trial including a 5 s
 1376 hint before the clip, 45 s for self-assessment, and a 15 s rest period afterward.

1377 **Faced** is a fine-grained emotion recognition dataset including EEG from 123 subjects recorded
 1378 with 32 channels at 250 Hz. Each subject performed a 9-class emotion elicitation task, where they
 1379 watched 28 video clips selected to evoke a range of emotions including amusement, inspiration,
 1380 joy, tenderness, anger, fear, disgust, sadness, and neutral states. This dataset provides a large-scale,
 1381 fine-grained, and balanced benchmark for 9-class emotion recognition from EEG signals.

1382 **Boran** is an intracranial EEG (iEEG) dataset including recordings from 9 patients with drug-resistant
 1383 focal epilepsy. The dataset contains simultaneous recordings from stereotactically implanted depth
 1384 electrodes in the medial temporal lobe and from scalp EEG electrodes placed according to the 10–20
 1385 system. Macroelectrode iEEG was recorded at 4 kHz and microelectrode iEEG at 32 kHz, while
 1386 scalp EEG was recorded at 256 Hz. Recordings were performed using an ATLAS system for iEEG
 1387 and a NicoletOne system for scalp EEG. The dataset provides high-resolution iEEG and single-
 1388 neuron data from the human medial temporal lobe for studying epilepsy.

1389 **BNCI2015001** (Faller et al., 2012) is a motor imagery dataset including EEG recordings from 12
 1390 subjects with 13 channels at 512 Hz. Each subject performed sustained right hand versus both
 1391 feet motor imagery across 200 trials per class, resulting in a total of 14,400 trials. The experiment
 1392 comprised 3 sessions, each with a single run of 5-second trials. EEG was recorded from Laplacian
 1393 derivations centered on C3, Cz, and C4 according to the international 10–20 system.

1394 **BNCI2014001** (Tangermann et al., 2012) is a motor imagery dataset containing EEG from 9 subjects
 1395 recorded with 22 channels at 250 Hz. The paradigm involves four motor imagery tasks: left hand,
 1396 right hand, both feet, and tongue, with 144 trials per class per subject. Each subject completed two
 1397 sessions on different days, each comprising 6 runs of 48 4-second trials.

1398
 1399
 1400
 1401
 1402
 1403

1404
 1405 **E.2 VISUALLY EVOKED POTENTIALS**

1406 **Table 8: VEP per dataset results.** Average of test-set scores (balanced accuracy (%); higher is
 1407 better; mean \pm std).

Model	SFUDA	Dataset			
		CBVEP100	CBVEP40	Nakanishi	Wang
ATCNet	w/o	25.4 \pm 1.8	25.2 \pm 3.7	56.3 \pm 20.6	32.8 \pm 13.1
	EA	74.4 \pm 12.1	74.6 \pm 9.0	54.7 \pm 18.9	35.8 \pm 10.8
	STMA	25.6 \pm 1.5	24.5 \pm 1.0	58.9 \pm 18.7	44.3 \pm 11.1
EEGConformer	w/o	25.0 \pm 0.0	25.1 \pm 0.2	29.4 \pm 10.2	31.9 \pm 11.0
	EA	75.8 \pm 10.3	73.1 \pm 9.6	29.4 \pm 9.0	30.5 \pm 8.5
	STMA	25.0 \pm 0.0	24.8 \pm 0.5	24.3 \pm 7.4	30.7 \pm 8.2
EEGNet	w/o	24.9 \pm 0.9	25.7 \pm 4.9	57.2 \pm 19.8	37.4 \pm 12.9
	EA	79.5 \pm 11.2	77.4 \pm 8.2	56.0 \pm 16.5	39.7 \pm 9.5
	STMA	24.6 \pm 2.9	24.1 \pm 3.3	40.0 \pm 10.7	32.0 \pm 8.4
FBCNet	w/o	29.4 \pm 3.9	30.4 \pm 3.1	28.3 \pm 7.5	16.4 \pm 4.5
	EA	75.8 \pm 10.4	74.3 \pm 10.1	27.7 \pm 7.4	16.2 \pm 3.0
	STMA	25.0 \pm 0.5	24.2 \pm 2.4	27.9 \pm 7.3	17.8 \pm 3.4
ShallowNet	w/o	76.2 \pm 7.2	74.3 \pm 8.4	29.4 \pm 7.7	18.4 \pm 6.7
	EA	66.7 \pm 14.8	64.9 \pm 13.7	27.6 \pm 7.2	16.7 \pm 4.4
	STMA	54.7 \pm 8.3	52.5 \pm 6.4	28.6 \pm 8.3	19.8 \pm 4.3
TSLANet	w/o	30.4 \pm 12.5	41.8 \pm 19.4	16.1 \pm 4.3	18.4 \pm 6.2
	EA	73.5 \pm 9.6	73.5 \pm 9.1	14.0 \pm 1.7	18.0 \pm 5.5
	STMA	25.9 \pm 1.7	26.3 \pm 2.4	15.4 \pm 3.3	17.8 \pm 4.0
DDGCNN	w/o	24.4 \pm 1.8	25.5 \pm 2.2	26.2 \pm 12.9	30.3 \pm 10.7
	EA	72.2 \pm 11.1	70.8 \pm 11.4	54.7 \pm 21.2	46.1 \pm 12.2
EEGInception	w/o	28.9 \pm 5.5	34.2 \pm 20.9	60.7 \pm 22.0	38.7 \pm 12.8
	EA	79.9 \pm 8.8	75.5 \pm 11.1	61.5 \pm 19.0	41.8 \pm 9.6
SSVEPNet	w/o	26.0 \pm 3.8	28.1 \pm 4.4	70.8 \pm 20.5	55.1 \pm 16.0
	EA	72.4 \pm 13.3	68.5 \pm 10.9	68.9 \pm 20.2	53.9 \pm 13.0
SSVEPFormer	w/o	24.0 \pm 3.9	26.7 \pm 6.4	77.2 \pm 21.0	72.3 \pm 15.6
	EA	72.2 \pm 8.4	72.1 \pm 7.0	77.0 \pm 19.8	72.6 \pm 12.6
LaBraM	w/o	22.5 \pm 39.4	17.5 \pm 31.2	22.2 \pm 41.5	72.2 \pm 13.6
CBraMod	w/o	25.1 \pm 5.7	24.6 \pm 17.6	31.7 \pm 17.3	17.5 \pm 0.7
HEEGNet	DSMDBN	95.8 \pm 6.2	92.6 \pm 22.1	79.8 \pm 20.8	69.3 \pm 15.5
	DSMDBN+EA	83.8 \pm 9.5	79.0 \pm 10.2	81.9 \pm 18.7	73.5 \pm 11.4

1458 E.3 EMOTION RECOGNITION
 1459

1460 Table 9: **Emotion recognition per dataset results.** Average of test-set scores (balanced accuracy
 1461 (%); higher is better; mean \pm std).

Model	SFUDA	Dataset	
		Faced	Seed
ATCNet	w/o	11.2 \pm 0.9	34.8 \pm 9.8
	EA	57.5 \pm 15.0	44.6 \pm 13.1
	STMA	59.3 \pm 15.5	48.4 \pm 15.7
EEGConformer	w/o	85.4 \pm 14.3	77.5 \pm 21.7
	EA	89.0 \pm 11.1	65.9 \pm 18.1
	STMA	81.5 \pm 13.9	69.2 \pm 17.7
EEGNet	w/o	24.8 \pm 13.7	74.5 \pm 19.8
	EA	83.9 \pm 14.5	46.2 \pm 15.4
	STMA	78.2 \pm 16.6	50.1 \pm 15.2
FBCNet	w/o	19.0 \pm 8.4	54.5 \pm 17.4
	EA	42.1 \pm 12.7	76.3 \pm 17.4
	STMA	23.1 \pm 8.9	60.0 \pm 20.3
ShallowNet	w/o	43.1 \pm 16.6	76.2 \pm 17.6
	EA	80.3 \pm 16.7	75.6 \pm 15.2
	STMA	64.6 \pm 17.2	63.3 \pm 15.5
TSLANet	w/o	33.3 \pm 11.8	41.8 \pm 13.8
	EA	37.2 \pm 12.4	65.6 \pm 14.4
	STMA	46.2 \pm 13.1	58.7 \pm 16.4
EMT	w/o	32.9 \pm 13.4	48.4 \pm 14.5
	EA	38.6 \pm 11.4	42.8 \pm 16.1
TSception	w/o	14.3 \pm 4.6	42.2 \pm 15.4
	EA	84.3 \pm 13.3	55.0 \pm 18.0
CBraMod	w/o	33.3 \pm 11.6	47.4 \pm 9.6
LaBraM	w/o	14.9 \pm 3.3	45.1 \pm 5.2
HEEGNet	DSMDBN	84.1 \pm 14.7	79.4 \pm 20.8
	DSMDBN+EA	<u>89.7 \pm 11.2</u>	78.7 \pm 12.4

1497
 1498
 1499
 1500
 1501
 1502
 1503
 1504
 1505
 1506
 1507
 1508
 1509
 1510
 1511

1512 **E.4 INTRACRANIAL EEG RESULTS**
 1513

1514 **Table 10: Intracranial EEG result.** Average of test-set scores (balanced accuracy (%); higher is
 1515 better; mean \pm std).

Model	SFUDA	Intracranial
EEGNet	w/o	55.4 ± 9.1
	EA	59.1 ± 11.2
	STMA	51.2 ± 5.9
EEGConformer	w/o	55.5 ± 7.0
	EA	61.7 ± 13.1
	STMA	53.1 ± 5.2
ATCNet	w/o	57.5 ± 9.0
	EA	57.1 ± 9.0
	STMA	54.2 ± 5.7
TSLANet	w/o	55.1 ± 9.3
	EA	54.9 ± 10.4
	STMA	54.1 ± 6.2
SchirrmeisterNet	w/o	56.9 ± 7.7
	EA	65.0 ± 10.2
	STMA	53.9 ± 5.5
FBCNet	w/o	61.1 ± 10.6
	EA	61.3 ± 11.0
	STMA	53.0 ± 6.8
LaBraM	w/o	44.4 ± 49.7
CBraMod	w/o	53.6 ± 6.2
HEEGNet (proposed)	DSMDBN	58.4 ± 9.6
	DSMDBN+EA	66.7 ± 12.5

1516
 1517
 1518
 1519
 1520
 1521
 1522
 1523
 1524
 1525
 1526
 1527
 1528
 1529
 1530
 1531
 1532
 1533
 1534
 1535
 1536
 1537
 1538
 1539
 1540
 1541
 1542
 1543
 1544
 1545
 1546
 1547
 1548
 1549
 1550
 1551
 1552
 1553
 1554
 1555
 1556
 1557
 1558
 1559
 1560
 1561
 1562
 1563
 1564
 1565

1566 F FULLY HYPERBOLIC NEURAL NETWORK EXPERIMENT
 1567

1568 To investigate the necessity of the hybrid design in HEEGNet, we conducted experiments by replac-
 1569 ing Euclidean convolutional operations in EEGNet (Lawhern et al., 2018) with hyperbolic variants
 1570 as follows:
 1571

- 1572 • **None (EEGNet)** : Original EEGNet
- 1573 • **HMLR**: Replaced the final MLR layer with hyperbolic MLR
- 1574 • **HMLR + Block2**: Additionally replaced Block2 in EEGNet (the depth-wise convolution
 1575 and the point-wise convolution)
- 1576 • **Full Hyperbolic**: Fully hyperbolic models
 1577

1578 As shown in Tab. 11, the performance on the 4 representative datasets demonstrates a degradation
 1579 pattern as more Euclidean operations are replaced with hyperbolic variants. These result suggests
 1580 that the Euclidean backbone is essential for effective downstream classification, supporting the ra-
 1581 tionale behind our hybrid architecture, which combines Euclidean encoders with hyperbolic neural
 1582 networks.
 1583

1584 **Table 11: Fully hyperbolic neural network experiments.** The averages of test-set scores are shown
 1585 above using balanced accuracy (the score and the standard deviation are shown for each dataset).
 1586 The number of domains in each dataset is indicated by n .
 1587

1588 1589 1590 1591 1592 1593 1594 1595 1596	1597 1598 1599 1600 1601 1602 1603 1604 1605 1606 1607 1608 1609 1610 1611 1612 1613 1614 1615 1616 1617 1618 1619	Visual		Emotion	Intracranial	
		Dataset	Nakanishi ($n = 9$)	Wang ($n = 34$)	Seed ($n = 45$)	
		Model			Boran ($n = 37$)	
		None (EEGNet)	57.2 \pm 19.8	37.4 \pm 12.9	74.5 \pm 19.8	55.4 \pm 9.1
		HMLR	60.8 \pm 20.7	39.2 \pm 13.8	75.1 \pm 20.0	57.4 \pm 8.7
		HMLR + Block2	53.4 \pm 17.5	32.5 \pm 11.9	76.9 \pm 22.1	55.5 \pm 7.3
		Fully Hyperbolic	27.3 \pm 12.0	22.5 \pm 8.8	70.7 \pm 21.6	52.7 \pm 8.9