# HEEGNet: Hyperbolic Embeddings for EEG

**Anonymous authors**
Paper under double-blind review

## 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. Yet, most existing decoding methods rely on Euclidean embeddings, which are not well-suited for capturing hierarchical structures. In contrast, hyperbolic spaces, regarded as the continuous analogue of tree structures, provide a natural geometry for representing hierarchical data. In this study, we first 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-
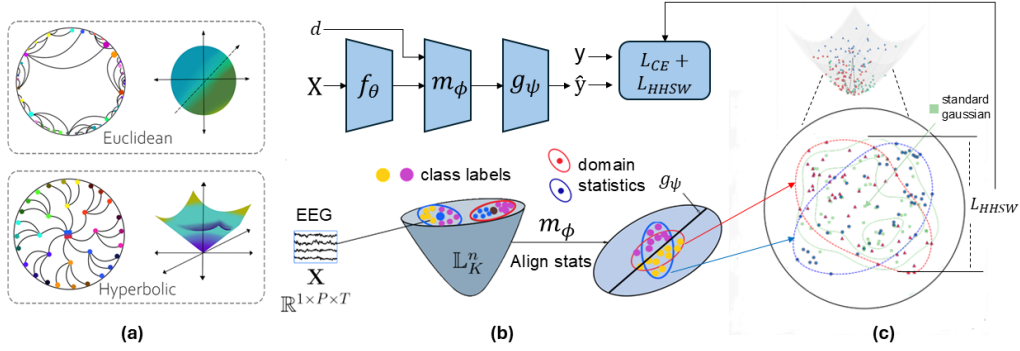
Figure 1: **Framework Overview.** (a) A comparison illustrating that hyperbolic space, with its negative curvature, is better suited for embedding hierarchical data than flat Euclidean space. (b) The HEEGNet architecture, which employs a coarse-to-fine domain adaptation strategy, DSMDBN (proposed). The first stage of DSMDBN aligns domain-specific moment statistics. (c) 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 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

**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.

**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.

**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

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

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:

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

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$_\eta$ on

Riemannian manifolds is defined as the minimizer of the squared distances weighted by $\eta_i$:

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

If the weights $\eta_i$ are uniform, the weighted Fréchet mean reduces to the standard Fréchet mean, and the Fréchet variance $\nu^2$ 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 $\exp_p^K : T_p\mathbb{L}_K^n \to \mathbb{L}_K^n$ and the logarithmic map $\log_p^K : \mathbb{L}_K^n \to T_p\mathbb{L}_K^n$ 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 $\mathrm{PT}_{p \to q}(v) \in \mathbb{L}_K^n$ 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

$$p \oplus_K^{\mathbb{L}} q = \mathrm{Exp}_{\bar{0}}\left(\mathrm{PT}_{\bar{0} \to p}\left(\mathrm{Log}_{\bar{0}}(q)\right)\right), \quad \forall p, q \in \mathbb{L}_K^n, \tag{3}$$

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

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

where $\bar{0} = \left[\sqrt{-1/K}, 0, \ldots, 0\right]^\top$ 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. B.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.5) 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.2) 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.6) also utilizes the distance from instances to hyperplanes to describe the class regions.

**$\delta$-hyperbolicity.** Khrulkov et al. (2020) introduced $\delta$-hyperbolicity as a measure of the degree of tree-like structure inherent in embeddings. The idea is to find the smallest value of $\delta$ for which the triangle inequality holds via the Gromov product (mathematical formulation in App. C.7). 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 $\delta$. 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:

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

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 $p$-Wasserstein distance, $B_\#^v\mu$ denotes the horospherical projection of $\mu$ along direction $v$, and $d\lambda(v)$ is the uniform measure over these directions.

## 3 METHODS

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

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

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

where $\mu$ and $\nu^2$ are Fréchet mean and Fréchet variance, $\gamma \in \mathbb{R}$ is the scaling parameter, and $\epsilon$ is a small value for numerical stability.

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

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

Formally, in our setting we assume that minibatches $\mathcal{B}_k$, that form the union of $N_{\mathcal{B}_k} \leq |\mathcal{D}|$ domain-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 $\frac{M}{N_{\mathcal{B}_k}}$ i.i.d. observations $x_i$, $i = 1, \ldots, M/N_{\mathcal{B}_k}$. In the first stage, $\text{DSMDBN}_{(1)}$ (algorithm 1) explicitly aligns the domain-specific running first- and second-order moments by centering and scaling them with $\nu_\phi^2$, which can be expressed as:

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

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

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

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

## 3.2 HEEGNet

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

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

In a nutshell, the feature extractor $f_\theta$ consists of three Euclidean convolutional layers along with standard components (e.g., BN and pooling), a projection layer ProjX, and a hyperbolic convolutional layer (Fig. 3). We sequentially adopt the three convolutional layers from EEGNet (Lawhern et al., 2018): temporal convolution to learn frequency-specific filters, depthwise spatial convolution to capture electrode-wise patterns, and a second depthwise temporal convolution to summarize information across time. These well-established operations provide spectral-spatial-temporal feature maps with meaningful neurophysiologically properties. The ProjX layer projects the Euclidean feature maps into hyperbolic space $\mathbb{L}_K^n$, after which a hyperbolic convolutional layer performs pointwise convolution to optimally combine these feature maps. The alignment module $m_\phi$ applies the first stage of the proposed DSMDBN (Eq. (8)) to explicitly align domain-specific moments in hyperbolic space, followed by hyperbolic ELU and pooling (Eq. (21)) for dimension reduction. Finally, the classifier $g_\psi$ is parametrized as a hyperbolic MLR (HMLR) layer).

## 4 EXPERIMENTS

In this study, we consider three EEG modalities that have been reported to encode hierarchical structures: visual- and emotion-stimuli scalp EEG, and intracranial EEG (Turner et al., 2023; Sun et al., 2023; Sonkusare et al., 2020). Corresponding EEG-based BCI applications include steady-state visually evoked potentials (SSVEP), code-modulated visually evoked potentials (CVEP), emotion recognition, and intracranial EEG. While all applications hold significant potential for rehabilitation and healthcare (Al-Nafjan et al., 2017; Guo et al., 2022; Elsner et al., 2018), their practical utility remains limited due to poor generalization across domains. We conduct a pilot study to evaluate hyperbolic embeddings and perform comprehensive experiments to assess the proposed HEEGNet.

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

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

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

**Evaluation.** We treat sessions as domains, and use either a leave-one-group-out (source domain number $\leq 10$) 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 setting 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 training schedule described above. We begin by training and evaluating EEGNet on cross-domain tasks. Using the trained EEGNet, we then generate embeddings of the target domain classification space and quantify their degree of inherent tree-likeness using $\delta$-hyperbolicity (Eq. (27)) following Krioukov et al. (2010). Following Khrulkov et al. (2020), we report the scale-invariant metric $\delta_{rel} \in [0, 1]$, where the lower $\delta_{rel}$ 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 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 are shown in Fig. 2a, b). These findings strongly motivate the use of hyperbolic embeddings in cross-domain generalization.

| Dataset | Visual | | Emotion | | Intracranial |
| --- | --- | --- | --- | --- | --- |
| Metric | Nakanishi (n=9) | Wang (n=34) | Seed (n=45) | Faced (n=123) | Boran (n=37) |
| $\delta_{rel}$ | $0.306 \pm 0.027$ | $0.333 \pm 0.039$ | $0.072 \pm 0.025$ | $0.132 \pm 0.047$ | $0.017 \pm 0.048$ |

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

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

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

## 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), SchirrmeisterNet (Schirrmeister 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), TSception (Ding et al., 2022). We use the implementation provided in the public available repositories for all architectures, stick to all hyperparameters as provided, and use the standard cross-entropy loss as training objective.

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

**HEEGNet.** We parametrize $h = g_\psi \circ m_\phi \circ f_\theta$ as a neural network and learn the entire model in an end-to-end fashion, and use the standard cross-entropy loss with HHSW (Eq. (9)) as the training objective. The HHSW loss weight is a hyperparameter to be tuned, we set it 0.01 for emotion and 0.5 for other dataset in our experiments. During source-free target domain adaptation, HEEGNet keeps the fitted source feature extractor $f_\theta$ and linear classifier $g_\psi$ fixed and estimates domain-specific first- and second-order statistics by solving Eq. (2) for moments alignment $m_\phi$.

Tab. 3 summarizes the results across all three EEG modalities, presenting the grand average scores with general EEG decoding baseline methods. Extended results for VEP-specific and emotion recognition-specific architectures, as well as per-dataset results, are provided in Supplementary Tab. 6 and Tab. 7. At the overall grand average level, HEEGNet (DSMDBN) outperforms all baseline methods. While input space alignment methods like EA and STMA improved cross-domain generalization, a gap still remains compared to our proposed two-stage alignment DSMDBN. A visualization of the two stages of DSMDBN is shown in Fig. 2c, d. Interestingly, HEEGNet (DSMDBN+EA), the integration of DSMDBN with the input space alignment method EA, achieves superior performance. We conduct an ablation study to systematically investigate the effects of different alignment strategies within the HEEGNet architecture. We further evaluate HEEGNet on public motor imagery datasets that not being reported to encode hierarchical information (details in App. F).
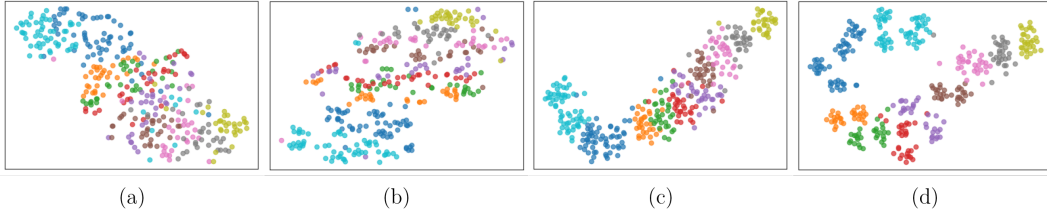


(a)　　　　　　　(b)　　　　　　　(c)　　　　　　　(d)

Figure 2: **t-SNE visualizations of classification space for Nakanishi (subject 4).** Each color denotes a distinct class. (a) EEGNet, (b) EEGNet with HMLR, (c) DSMDBN stage 1, and (d) DSMDBN stage 2. The plots illustrate how hyperbolic embedding and our proposed two-stage alignment progressively enhance class separability.

**Ablation study.** Tab. 4 summarizes the effectiveness of three different alignment methods, input alignment, moment alignment, and distribution alignment, within the HEEGNet architecture. We highlight three important observations. First, moment alignment in hyperbolic space is the primary driver of performance improvement. The absence of moment alignment consistently results in a significant performance drop of at least 12.3% compared to the best-performing configuration. Second, distribution alignment proves effective only when paired with moment alignment, validating our proposed DSMDBN approach, which first aligns moments to facilitate subsequent distribution alignment. Third, moment alignment in the latent space outperforms input space alignment, supporting the findings of Bakas et al. (2025) that alignment benefits from enhanced class discrimination in the latent representation. Fourth, DSMDBN augmented with input alignment yields the best performance, indicating that multistage alignment a promising strategy in cross-domain generalization.

## 5 CONCLUSIONS

In this work, we introduced HEEGNet, a hybrid hyperbolic network architecture designed to capture the hierarchical structure of EEG data and learn domain-invariant hyperbolic embeddings. Our

| Model | SFUDA | VEP (n=67) | Emotion (n=168) | Intracranial (n=37) | Overall (n=272) |
|---|---|---|---|---|---|
| EEGNet | w/o | 35.7±15.5 ● | 38.1±27.0 ● | 55.4±9.1 ● | 39.9±23.6 ● |
| | EA | 55.8±20.8 ● | 73.8±22.3 ● | 59.1±11.2 ● | 67.4±22.3 ● |
| | STMA | 30.3±8.9 ● | 70.7±20.5 ● | 51.2±5.9 ● | 58.1±24.0 ● |
| EEGConformer | w/o | 29.1±9.1 ● | 83.2±16.9 · | 55.5±7.0 ● | 66.1±27.2 ● |
| | EA | 46.1±23.1● | 82.8±16.8 ● | 61.7±13.1 · | 70.9±24.1 ● |
| | STMA | 27.8±7.0 ● | 78.2±15.9 ● | 53.1±5.2 ● | 62.4±25.2 ● |
| ATCNet | w/o | 33.3±15.4 ● | 17.5±11.7 ● | 57.5±9.0 ● | 26.9±18.5 ● |
| | EA | 52.2±21.4● | 54.1±15.5 ● | 57.1±9.0 ● | 54.0±16.5 ● |
| | STMA | 39.4±15.6 ● | 56.4±16.3 ● | 54.2±5.7 ● | 51.9±16.7 ● |
| TSLANet | w/o | 24.4±14.1 ● | 35.5±12.9 ● | 55.1±9.3 ● | 35.5±15.7 ● |
| | EA | 37.3±28.1 ● | 44.8±18.1 ● | 54.9±10.4 ● | 44.3±20.8 ● |
| | STMA | 20.5±5.4 ● | 49.6±15.1 ● | 54.1±6.2 ● | 43.0±17.9 ● |
| SchirrmeisterNet | w/o | 40.2±27.5 ● | 51.9±22.4 ● | 56.9±7.7 ● | 49.7±23.1 ● |
| | EA | 35.7±24.7 ● | 79.0±16.4 ● | 65.0±10.2 | 66.5±25.7 ● |
| | STMA | 33.1±16.8 ● | 64.2±16.7 ● | 53.9±5.5 ● | 55.2±20.4 ● |
| FBCNet | w/o | 22.8±8.0 ● | 28.5±19.5 ● | 61.1±10.6 · | 31.6±20.2 ● |
| | EA | 38.8±28.4 ● | 51.3±20.7 ● | 61.3±11.0 · | 49.6±22.9 ● |
| | STMA | 21.6±5.4 ● | 33.0±20.9 ● | 53.0±6.8 ● | 32.9±19.2 ● |
| HEEGNet (*proposed*) | DSMDBN | **79.6±19.8** | 82.8±16.6 | 58.4±9.6 | 78.7±18.6 |
| | DSMDBN+EA | 77.4±12.6 | **86.7±12.5** | **66.7±12.5** | **81.7±14.4** |

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

| Alignment | | | 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) |

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

pilot study and empirical analyses indicate that EEG data exhibits hyperbolicity and that hyperbolic embeddings improve generalization compared to Euclidean ones. By integrating both Euclidean and hyperbolic encoders and employing a novel two-stage domain adaptation strategy (DSMDBN), HEEGNet effectively aligns domain-specific moments and distributions in hyperbolic space. Experiments on multiple EEG datasets, including visual evoked potentials, emotion recognition, and intracranial EEG, demonstrate state-of-the-art performance. A limitation of the HEEGNet is that moment alignment methods are difficult to extend to online settings, as moment estimation becomes unreliable with sequentially arriving data. Future work will explore the online setting and more effective encoders to capture the hierarchical structure of EEG signals.

## REPRODUCIBILITY STATEMENT

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

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

`https://anonymous.4open.science/r/HEEGNet-F655`

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

**Data Availability and Processing Steps.** All datasets used in our experiments are publicly available VEP, intracranialand EEG and emotion recognition datasets. Our preprocessing and evaluation schemes are detailed in Section 4.

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

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

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

SEED: `https://bcmi.sjtu.edu.cn/home/seed/`

Faced: `https://www.synapse.org/Synapse:syn50614194/files/`

The public intracranialand EEG datasets we used can be found in links below.

Boran: `https://doi.gin.g-node.org/10.12751/g-node.d76994/`

## ETHICS STATEMENT

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

## REFERENCES

David Acuna, Guojun Zhang, Marc T Law, and Sanja Fidler. f-domain adversarial learning: Theory and algorithms. In *International Conference on Machine Learning*, pp. 66–75. PMLR, 2021.

Abeer Al-Nafjan, Manar Hosny, Yousef Al-Ohali, and Areej Al-Wabil. Review and classification of emotion recognition based on eeg brain-computer interface system research: a systematic review. *Applied Sciences*, 7(12):1239, 2017. doi: 10.3390/app7121239.

Hamdi Altaheri, Ghulam Muhammad, and Mansour Alsulaiman. Physics-informed attention temporal convolutional network for eeg-based motor imagery classification. *IEEE transactions on industrial informatics*, 19(2):2249–2258, 2022. doi: 10.1109/TII.2022.3197419.

Stylianos Bakas, Siegfried Ludwig, Dimitrios A Adamos, Nikolaos Laskaris, Yannis Panagakis, and Stefanos Zafeiriou. Latent alignment in deep learning models for eeg decoding. *Journal of Neural Engineering*, 22(1):016047, 2025. doi: 10.1088/1741-2552/adb336.

Ahmad Bdeir, Kristian Schwethelm, and Niels Landwehr. Fully hyperbolic convolutional neural networks for computer vision. In *ICLR*, 2024.

Gary Bécigneul and Octavian-Eugen Ganea. Riemannian adaptive optimization methods. *arXiv preprint arXiv:1810.00760*, 2018.

Martha Ann Bell and Kimberly Cuevas. Using eeg to study cognitive development: Issues and practices. *Journal of cognition and development*, 13(3):281–294, 2012. doi: 10.1080/15248372.2012.691143.

Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79:151–175, 2010. doi: 10.1007/s10994-009-5152-4.

Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013. doi: 10.1109/TPAMI.2013.50.

Clément Bonet, Laetitia Chapel, Lucas Drumetz, and Nicolas Courty. Hyperbolic sliced-wasserstein via geodesic and horospherical projections. In *Topological, Algebraic and Geometric Learning Workshops 2023*, pp. 334–370. PMLR, 2023.

Ece Boran, Tommaso Fedele, Adrian Steiner, Peter Hilfiker, Lennart Stieglitz, Thomas Grunwald, and Johannes Sarnthein. Dataset of human medial temporal lobe neurons, scalp and intracranial eeg during a verbal working memory task. *Scientific data*, 7(1):30, 2020. doi: 10.6084/m9.figshare.11500485.

James W Cannon, William J Floyd, Richard Kenyon, Walter R Parry, et al. Hyperbolic geometry. *Flavors of geometry*, 31(59-115):2, 1997.

Kalou Cabrera Castillos, Simon Ladouce, Ludovic Darmet, and Frédéric Dehais. Burst c-vep based bci: Optimizing stimulus design for enhanced classification with minimal calibration data and improved user experience. *NeuroImage*, 284:120446, 2023. doi: 10.1016/j.neuroimage.2023.120446.

Ines Chami, Zhitao Ying, Christopher Ré, and Jure Leskovec. Hyperbolic graph convolutional neural networks. In *Advances in Neural Information Processing Systems*, pp. 4869–4880, 2019.

Jiang Chang, Zhixin Zhang, Yuhua Qian, and Pan Lin. Multi-scale hyperbolic contrastive learning for cross-subject eeg emotion recognition. *IEEE Transactions on Affective Computing*, 2025. doi: 10.1109/TAFFC.2025.3535542.

Woong-Gi Chang, Tackgeun You, Seonguk Seo, Suha Kwak, and Bohyung Han. Domain-specific batch normalization for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, pp. 7354–7362, 2019.

Jianbo Chen, Yangsong Zhang, Yudong Pan, Peng Xu, and Cuntai Guan. A transformer-based deep neural network model for ssvep classification. *Neural Networks*, 164:521–534, 2023a. doi: 10.1016/j.neunet.2023.04.045.

Jingjing Chen, Xiaobin Wang, Chen Huang, Xin Hu, Xinke Shen, and Dan Zhang. A large finer-grained affective computing eeg dataset. *Scientific Data*, 10(1):740, 2023b. doi: 10.7303/syn50614194.

Ziheng Chen, Yue Song, Xiaojun Wu, and Nicu Sebe. Gyrogroup batch normalization. In *ICLR*, 2025a.

Ziheng Chen, Xiao-Jun Wu, Bernhard Schölkopf, and Nicu Sebe. Riemannian batch normalization: A gyro approach. *arXiv preprint arXiv:2509.07115*, 2025b.

Sylvain Chevallier, Igor Carrara, Bruno Aristimunha, Pierre Guetschel, Sara Sedlar, Bruna Lopes, Sébastien Velut, Salim Khazem, and Thomas Moreau. The largest eeg-based bci reproducibility study for open science: the moabb benchmark. *arXiv preprint arXiv:2404.15319*, 2024.

Elliot Collins, Amanda K Robinson, and Marlene Behrmann. Distinct neural processes for the perception of familiar versus unfamiliar faces along the visual hierarchy revealed by eeg. *NeuroImage*, 181:120–131, 2018. doi: 10.1016/j.neuroimage.2018.06.080.

Srikanth R Damera, Jacob G Martin, Clara Scholl, Judy S Kim, Laurie Glezer, Patrick S Malone, and Maximilian Riesenhuber. From shape to meaning: Evidence for multiple fast feedforward hierarchies of concept processing in the human brain. *Neuroimage*, 221:117148, 2020. doi: j.neuroimage.2020.117148.

Yi Ding, Neethu Robinson, Su Zhang, Qiuhao Zeng, and Cuntai Guan. Tsception: Capturing temporal dynamics and spatial asymmetry from eeg for emotion recognition. *IEEE Transactions on Affective Computing*, 14(3):2238–2250, 2022. doi: 10.1109/TAFFC.2022.3169001.

Yi Ding, Chengxuan Tong, Shuailei Zhang, Muyun Jiang, Yong Li, Kevin JunLiang Lim, and Cuntai Guan. Emt: A novel transformer for generalized cross-subject eeg emotion recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 2025. doi: 10.1109/TNNLS.2025.3552603.

Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. Differential entropy feature for eeg-based emotion classification. In *2013 6th international IEEE/EMBS conference on neural engineering (NER)*, pp. 81–84. IEEE, 2013.

Emadeldeen Eldele, Mohamed Ragab, Zhenghua Chen, Min Wu, and Xiaoli Li. Tslanet: rethinking transformers for time series representation learning. In *Proceedings of the 41st International Conference on Machine Learning*, pp. 12409–12428, 2024.

Bernhard Elsner, Joachim Kugler, and Jan Mehrholz. Transcranial direct current stimulation (tdcs) for upper limb rehabilitation after stroke: future directions. *Journal of neuroengineering and rehabilitation*, 15(1):106, 2018. doi: 10.1186/s12984-018-0459-7.

Stephen H Fairclough and Fabien Lotte. Grand challenges in neurotechnology and system neuroergonomics. *Frontiers in Neuroergonomics*, 1:602504, 2020. doi: 10.3389/fnrgo.2020.602504.

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

Birgit Frauscher, Nicolas Von Ellenrieder, Rina Zelmann, Irena Doležalová, Lorella Minotti, André Olivier, Jeffery Hall, Dominique Hoffmann, Dang Khoa Nguyen, Philippe Kahane, et al. Atlas of the normal intracranial electroencephalogram: neurophysiological awake activity in different cortical areas. *Brain*, 141(4):1130–1144, 2018. doi: 10.1093/brain/awy035.

Octavian Ganea, Gary Bécigneul, and Thomas Hofmann. Hyperbolic neural networks. *Advances in neural information processing systems*, 31, 2018.

Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario March, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of machine learning research*, 17(59):1–35, 2016.

Théo Gnassounou, Antoine Collas, Rémi Flamary, Karim Lounici, and Alexandre Gramfort. Multi-source and test-time domain adaptation on multivariate signals using spatio-temporal monge alignment. *arXiv preprint arXiv:2407.14303*, 2024.

Alexandre Gramfort, Martin Luessi, Eric Larson, Denis A Engemann, Daniel Strohmeier, Christian Brodbeck, Lauri Parkkonen, and Matti S Hämäläinen. Mne software for processing meg and eeg data. *neuroimage*, 86:446–460, 2014. doi: 10.1016/j.neuroimage.2013.10.027.

Ning Guo, Xiaojun Wang, Dehao Duanmu, Xin Huang, Xiaodong Li, Yunli Fan, Hailan Li, Yongquan Liu, Eric Hiu Kwong Yeung, Michael Kai Tsun To, et al. Ssvep-based brain computer interface controlled soft robotic glove for post-stroke hand function rehabilitation. *IEEE transactions on neural systems and rehabilitation engineering*, 30:1737–1744, 2022. doi: 10.1109/TNSRE.2022.3185262.

He He and Dongrui Wu. Transfer learning for brain–computer interfaces: A euclidean space data alignment approach. *IEEE Transactions on Biomedical Engineering*, 67(2):399–410, 2019. doi: 10.1109/TBME.2019.2913914.

Shaul Hochstein and Merav Ahissar. View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, 36(5):791–804, 2002.

Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pp. 448–456. pmlr, 2015.

Valentin Khrulkov, Leyla Mirvakhabova, Evgeniya Ustinova, Ivan Oseledets, and Victor Lempitsky. Hyperbolic image embeddings. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6418–6428, 2020.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

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

Max Kochurov, Rasul Karimov, and Serge Kozlukov. Geoopt: Riemannian optimization in pytorch. *arXiv preprint arXiv:2005.02819*, 2020.

Dmitri Krioukov, Fragkiskos Papadopoulos, Maksim Kitsak, Amin Vahdat, and Marián Boguná. Hyperbolic geometry of complex networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics*, 82(3):036106, 2010. doi: 10.1103/PhysRevE.82.036106.

J Ph Lachaux, D Rudrauf, and P Kahane. Intracranial eeg and human brain mapping. *Journal of Physiology-Paris*, 97(4-6):613–628, 2003. doi: 10.1016/j.jphysparis.2004.01.018.

Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. Eegnet: a compact convolutional neural network for eeg-based brain–computer interfaces. *Journal of neural engineering*, 15(5):056013, 2018. doi: 10.1088/1741-2552/aace8c.

Chen-Yu Lee, Tanmay Batra, Mohammad Haris Baig, and Daniel Ulbricht. Sliced wasserstein discrepancy for unsupervised domain adaptation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10285–10295, 2019.

Yi-Chieh Lee, Wen-Chieh Lin, Fu-Yin Cherng, and Li-Wei Ko. A visual attention monitor based on steady-state visual evoked potential. *IEEE Transactions on neural systems and rehabilitation engineering*, 24(3):399–408, 2015. doi: TNSRE.2015.2501378.

Shanglin Li, Motoaki Kawanabe, and Reinmar J Kobler. Spdim: Source-free unsupervised conditional and label shift adaptation in eeg. In *The Thirteenth International Conference on Learning Representations*, 2025.

Siyang Li, Ziwei Wang, Hanbin Luo, Lieyun Ding, and Dongrui Wu. T-time: Test-time information maximization ensemble for plug-and-play bcis. *IEEE Transactions on Biomedical Engineering*, 71(2):423–432, 2023. doi: 10.1109/TBME.2023.3303289.

Fabien Lotte, Laurent Bougrain, Andrzej Cichocki, Maureen Clerc, Marco Congedo, Alain Rako-tomamonjy, and Florian Yger. A review of classification algorithms for eeg-based brain–computer interfaces: a 10 year update. *Journal of neural engineering*, 15(3):031005, 2018. doi: 10.1088/1741-2552/aab2f2.

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

Apolline Mellot, Antoine Collas, Pedro LC Rodrigues, Denis Engemann, and Alexandre Gramfort. Harmonizing and aligning m/eeg datasets with covariance-based techniques to enhance predictive regression modeling. *Imaging Neuroscience*, 1:1–23, 2023. doi: 10.1162/imag_a_00040.

Pascal Mettes, Mina Ghadimi Atigh, Martin Keller-Ressel, Jeffrey Gu, and Serena Yeung. Hyperbolic deep learning in computer vision: A survey. *International Journal of Computer Vision*, 132 (9):3484–3508, 2024. doi: 10.1007/s11263-024-02043-5.

Gal Mishne, Zhengchao Wan, Yusu Wang, and Sheng Yang. The numerical stability of hyperbolic representation learning. In *International Conference on Machine Learning*, pp. 24925–24949. PMLR, 2023.

Masaki Nakanishi, Yijun Wang, Yu-Te Wang, and Tzyy-Ping Jung. A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials. *PloS one*, 10(10):e0140703, 2015. doi: 10.1371/journal.pone.0140703.

Xuan Son Nguyen. The Gyro-structure of some matrix manifolds. In *NeurIPS*, 2022.

Ernst Niedermeyer and FH Lopes da Silva. *Electroencephalography: basic principles, clinical applications, and related fields*. Lippincott Williams & Wilkins, 2005. ISBN 978-0190228484.

Yudong Pan, Jianbo Chen, Yangsong Zhang, and Yu Zhang. An efficient cnn-lstm network with spectral normalization and label smoothing technologies for ssvep frequency recognition. *Journal of neural engineering*, 19(5):056014, 2022a. doi: 10.1088/1741-2552/ac8dc5.

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

Jaak Panksepp. Cross-species affective neuroscience decoding of the primal affective experiences of humans and related animals. *PloS one*, 6(9):e21236, 2011.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, et al. Scikit-learn: Machine learning in python. *the Journal of machine Learning research*, 12:2825–2830, 2011. 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.

Kamil Vlcek, Iveta Fajnerova, Tereza Nekovarova, Lukas Hejtmanek, Radek Janca, Petr Jezdik, Adam Kalina, Martin Tomasek, Pavel Krsek, Jiri Hammer, et al. Mapping the scene and object processing networks by intracranial eeg. *Frontiers in human neuroscience*, 14:561399, 2020. doi: 10.3389/fnhum.2020.561399.

Yijun Wang, Xiaogang Chen, Xiaorong Gao, and Shangkai Gao. A benchmark dataset for ssvep-based brain–computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 25(10):1746–1752, 2016. doi: 10.1109/TNSRE.2016.2627556.

Xiaoxi Wei, A. Aldo Faisal, Moritz Grosse-Wentrup, Alexandre Gramfort, Sylvain Chevallier, Vinay Jayaram, Camille Jeunet, Stylianos Bakas, Siegfried Ludwig, Konstantinos Barmpas, Mehdi Bahri, Yannis Panagakis, Nikolaos Laskaris, Dimitrios A. Adamos, Stefanos Zafeiriou, William C. Duong, Stephen M. Gordon, Vernon J. Lawhern, Maciej Śliwowski, Vincent Rouanne, and Piotr Tempczyk. 2021 BEETL Competition: Advancing Transfer Learning for Subject Independence and Heterogenous EEG Data Sets. In Douwe Kiela, Marco Ciccone, and Barbara Caputo (eds.), *Proceedings of the NeurIPS 2021 Competitions and Demonstrations Track*, volume 176 of *Proceedings of Machine Learning Research*, pp. 205–219. PMLR, 2022.

Dongrui Wu, Yifan Xu, and Bao-Liang Lu. Transfer learning for eeg-based brain–computer interfaces: A review of progress made since 2016. *IEEE Transactions on Cognitive and Developmental Systems*, 14(1):4–19, 2020. doi: 10.1109/TCDS.2020.3007453.

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

Hongwei Yong, Jianqiang Huang, Deyu Meng, Xiansheng Hua, and Lei Zhang. Momentum batch normalization for deep learning with small batch size. In *European Conference on Computer Vision*, pp. 224–240. Springer, 2020.

Shubin Zhang, Dong An, Jincun Liu, Jiannan Chen, Yaoguang Wei, and Fuchun Sun. Dynamic decomposition graph convolutional neural network for ssvep-based brain–computer interface. *Neural Networks*, 172:106075, 2024. doi: 10.1016/j.neunet.2023.12.029.

Han Zhao, Remi Tachet Des Combes, Kun Zhang, and Geoffrey Gordon. On learning invariant representations for domain adaptation. In *International conference on machine learning*, pp. 7523–7532. PMLR, 2019.

APPENDIX CONTENTS

## A LARGE LANGUAGE MODEL USAGE STATEMENT

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

## B GYROVECTOR SPACES

### B.1 GENERAL DEFINITION

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

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

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

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

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

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

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

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

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

**Definition B.2** (Gyrocommutative Gyrogroups (Ungar, 2022)). A gyrogroup $(G, \oplus)$ is gyrocommutative if it satisfies

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

Similarly, the gyrovector space generalizes the vector space,

**Definition B.3** (Gyrovector Spaces (Chen et al., 2025b)). A gyrocommutative gyrogroup $(G, \oplus)$ equipped with a scalar gyromultiplication $\otimes : \mathbb{R} \times G \to G$ is called a gyrovector space if it satisfies the following axioms for $s, t \in \mathbb{R}$ and $p, q, z \in G$:
(V1) Identity Scalar Multiplication: $1 \otimes p = p$.
(V2) Scalar Distributive Law: $(s + t) \otimes p = s \otimes p \oplus t \otimes p$.
(V3) Scalar Associative Law: $(st) \otimes p = s \otimes (t \otimes p)$.
(V4) Gyroautomorphism: $\mathrm{gyr}[p, q](t \otimes z) = t \otimes \mathrm{gyr}[p, q]z$.
(V5) Identity Gyroautomorphism: $\mathrm{gyr}[s \otimes p, t \otimes p] = \mathbb{I}$, where $\mathbb{I}$ is the identity map.

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

### B.2 GEOMETRIC INTUITION

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

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

$$T_p\mathbb{R}^n = T_q\mathbb{R}^n = \mathbb{R}^n, \tag{10}$$

$$\mathrm{Log}_p(q) = q - p, \tag{11}$$

$$\mathrm{Exp}_p(v) = v + p \tag{12}$$

$$\mathrm{PT}_{p \to q}(v) = v, \tag{13}$$

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

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

$$x + y = \text{Exp}_p\big(\text{PT}_{\mathbf{0} \to p}(\text{Log}_{\mathbf{0}}(y))\big). \tag{14}$$

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

$$t \odot p = \text{Exp}_{\mathbf{0}}(t \, \text{Log}_{\mathbf{0}}(p)) = tp. \tag{15}$$

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

**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 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 reviews the closed-form expression of the Lorentz gyro operators, which are more efficient than the Riemannian definitions Eqs. (3) and (4) (Chen et al., 2025b, Sec. 6.1). The Lorentz gyroaddition and gyromultiplication have the closed-form solution:

$$\text{Gyroaddition:} \quad p \oplus_K^{\mathcal{M}} q = \begin{cases} p, & q = \overline{0}, \\ q, & p = \overline{0}, \\ \begin{bmatrix} \dfrac{1 - \frac{D - KN}{|K|}}{D + KN} \\ \dfrac{2(A_s p_s + A_q q_s)}{D + KN} \end{bmatrix}, & \text{Others.} \end{cases} \tag{16}$$

$$\text{Gyromultiplication:} \quad t \otimes_K^{\mathcal{M}} p = \begin{cases} \overline{0}, & t = 0 \lor p = \overline{0}, \\ \dfrac{1}{\sqrt{|K|}} \begin{bmatrix} \cosh\big(t \cosh^{-1}(\sqrt{|K|}p_t)\big) \\ \dfrac{\sinh\big(t \cosh^{-1}(\sqrt{|K|}p_t)\big)}{\|p_s\|} p_s \end{bmatrix}, & t \neq 0, \end{cases} \tag{17}$$

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

$$a = 1 + \sqrt{|K|}p_t, \quad b = 1 + \sqrt{|K|}q_t, \quad n_p = \|p_s\|^2, \quad n_q = \|q_s\|^2, \quad s_{pq} = \langle p_s, q_s \rangle,$$

$$D = a^2 b^2 - 2Kabs_{pq} + K^2 n_p n_q, \quad N = a^2 n_q + 2abs_{pq} + b^2 n_p.$$

## C LORENTZ OPERATIONS

### C.1 RIEMANNIAN OPERATORS

The exponential map $\text{Exp}_p^K : T_p\mathbb{L}_K^n \to \mathbb{L}_K^n$ and logarithmic map $\text{Log}_p^K : T_p\mathbb{L}_K^n \to \mathbb{L}_K^n$ 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 point $p \in \mathbb{L}_K^n$.

$$\text{Exp}_p^K(v) = \cosh(\alpha)p + \sinh(\alpha)\frac{v}{\alpha}, \quad \text{with} \quad \alpha = \sqrt{-K}\,\|v\|_{\mathcal{L}}, \quad \|v\|_{\mathcal{L}} = \sqrt{\langle v, v \rangle_{\mathcal{L}}} \tag{18}$$

$$\text{Log}_p^K(q) = \frac{\cosh^{-1}(\beta)}{\sqrt{\beta^2 - 1}} \cdot (q - \beta p), \quad \text{with} \quad \beta = K\langle p, q \rangle_{\mathcal{L}} \tag{19}$$

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 $\text{PT}_{p \to q}(v) \in \mathbb{L}_K^n$ can be used:

$$\text{PT}_{p \to q}(v) = v - \frac{K(q, v)_K}{1 + K(p, q)_K}(p + q) \tag{20}$$

### C.2 LORENTZ NON-LINEAR ACTIVATION

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

$$C_{\text{activated}} = \begin{bmatrix} \sqrt{\|\text{ELU}(\mathbf{p}_s)\|^2 - 1/K} \\ \text{ELU}(\mathbf{p}_s) \end{bmatrix}. \tag{21}$$

### C.3 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:

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

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

### C.4 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:

$$y = \text{LFC}(p) = \begin{bmatrix} \sqrt{\|\psi(\mathbf{W}p + \mathbf{b})\|^2 - 1/K} \\ \psi(\mathbf{W}p + \mathbf{b}) \end{bmatrix} \tag{23}$$

$$\phi(\mathbf{W}p, v) = \lambda\sigma(\mathbf{v}^Tp + b')\frac{\mathbf{W}\psi(p) + \mathbf{b}}{\|\mathbf{W}\psi(p) + \mathbf{b}\|} \tag{24}$$

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

### C.5 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 $\mathbf{K} \in \mathbb{R}^{m \times n \times \bar{H} \times \bar{W}}$ 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 $\delta$ is the stride parameter. The Lorentz convolutional layer is defined as $\text{LFC}\big(\text{HCAT}(\{\mathbf{p}_{h'+\delta_{\tilde{h}}, w'+\delta_{\tilde{w}}} \in \mathbb{L}_K^n\}_{\tilde{h},\tilde{w}=1}^{\tilde{H},\tilde{W}})\big)$, where HCAT and LFC, denote hyperbolic concatenation and a Lorentz fully-connected layer performing the affine transformation and parameterizing the kernel and bias.

## C.6 LORENTZ MULTINOMIAL LOGISTIC REGRESSION

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

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

where

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

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

## C.7 $\delta$-HYPERBOLICITY

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

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

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

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

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

# D HEEGNET DETAILS

## D.1 ALGORITHM

---

**Algorithm 1** Hyperbolic domain-specific momentum batch normalization (HDSMBN)

---

**Input:**
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

domain-specific running mean $\tilde{\mu}_{k-1}^d (\tilde{\mu}_0^d = \bar{0})$ and variance $\tilde{\nu}^2{}_{k-1}^d (\tilde{\nu}^2{}_0^d = 1)$ for training

domain-specific running mean $\hat{\mu}_{k-1}^d (\hat{\mu}_0^d = \bar{0})$ and variance $\hat{\nu}^2{}_{k-1}^d (\hat{\nu}^2{}_0^d = 1)$ for testing

momentum for training and testing $\eta_{train(k)}, \eta_{test} \in [0,1]$, learnable parameter $\nu_\phi^2$

**Output:** normalized batch $\{\tilde{p}_i\} = \mathrm{HBN}^{d(i)}(p_i)$

    **if** training then **then**

        Compute domain-specific batch mean $\mu_k^d$ and variance $\nu^2{}_k^d$          ▷ using Eq. (2)

        $\tilde{\mu}_k^d = \mathrm{wFM}_{\eta_{train(k)}}(\tilde{\mu}_{k-1}^d, \mu_k^d)$         ▷ update running mean using Eq. (2)

        $\tilde{\nu}^2{}_k^d = (1 - \eta_{train(k)})\, \tilde{\nu}^2{}_{k-1}^d + \eta_{train(k)}\, \nu^2{}_k^d$

        $\hat{\mu}_k^d = \mathrm{wFM}_{\eta_{test}}(\hat{\mu}_{k-1}^d, \mu_k^d)$         ▷ update running mean using Eq. (2)

        $\hat{\nu}^2{}_k^d = (1 - \eta_{test})\, \hat{\nu}^2{}_{k-1}^d + \eta_{test}\, \nu^2{}_k^d$

    **end if**

    $(\mu_k^d, \nu^2{}_k^d) \leftarrow (\tilde{\mu}_k^d, \tilde{\nu}^2{}_k^d)$ **if training else** $(\hat{\mu}_k^d, \hat{\nu}^2{}_k^d)$

    $\tilde{p}_i = \frac{\nu_\phi^2}{\sqrt{\nu^2{}_k^d + \epsilon}}\left( \ominus \mu_k^d \oplus p_i \right)$         ▷ use Eq. (7) to recentering and rescale each domain

---

---

**Algorithm 2** Horospherical Hyperbolic Sliced-Wasserstein loss (HHSW)

---

**Input:**
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 $d$

number of slices $S = 1000$, exponent $p = 2$

**Output:** scalar loss $\mathcal{L}_{\mathrm{swd}}$

    Initialize $\mathcal{L}_{\mathrm{swd}} \leftarrow 0$

    **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 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: $\ell^d \leftarrow \mathrm{HHSW}_p^p(\mathcal{P}^d, Q^d)$         ▷ Eq. (6)

        $\mathcal{L}_{\mathrm{swd}} \leftarrow \mathcal{L}_{\mathrm{swd}} + \ell^d$

    **end for**

    **return** $\mathcal{L}_{\mathrm{swd}}$

---

## D.2 MODEL ARCHITECTURE

Table 5: HEEGNet architecture details. $P$: electrodes; $T$: temporal samples; $C$: classes.

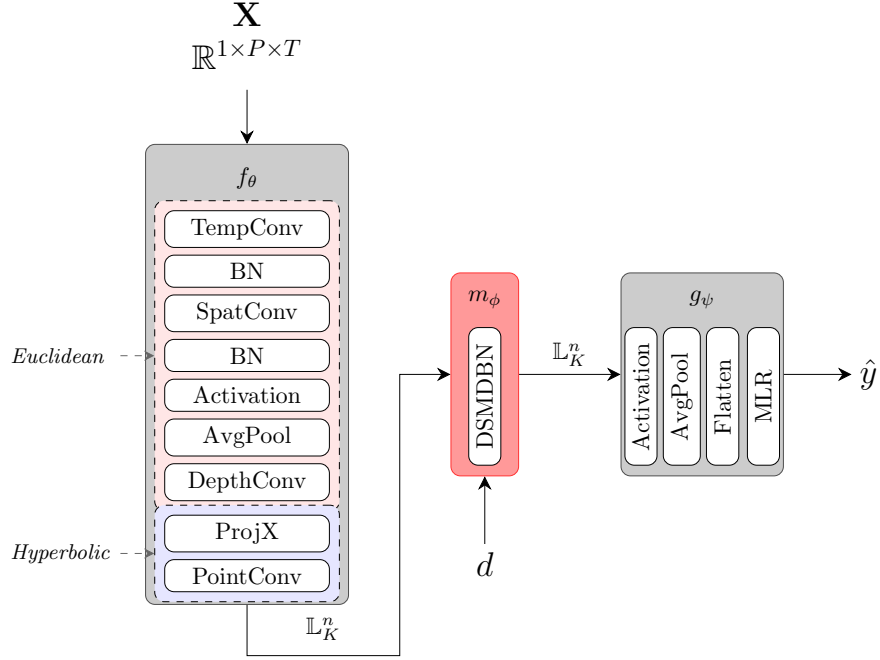| Layer | Output (dim) | Parameter (dim) | Operation | Space |
|---|---|---|---|---|
| | *Input:* $1 \times P \times T$ | | | |
| TempConv | $8 \times P \times T$ | $8 \times 1 \times 1 \times 64$ | convolution | Euclidean |
| BN | $8 \times P \times T$ | $8$ | batch norm | Euclidean |
| SpatConv | $16 \times 1 \times T$ | $16 \times 8 \times P \times 1$ | 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.

# E  FULL EXPERIMENTS

## E.1  DATASET DETAILS

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

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

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

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

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

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

## E.2 VISUALLY EVOKED POTENTIALS

Table 6: **VEP per dataset results.** Average of test-set scores (balanced accuracy (%); higher is better; mean ± 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 | $72.2 \pm 8.4$ | $72.1 \pm 7.0$ | $77.0 \pm 19.8$ | $72.6 \pm 12.6$ |
| HEEGNet | DSMDBN | $\underline{\mathbf{95.8 \pm 6.2}}$ | $\underline{\mathbf{92.6 \pm 22.1}}$ | $79.8 \pm 20.8$ | $69.3 \pm 15.5$ |
| | DSMDBN+EA | $83.8 \pm 9.5$ | $79.0 \pm 10.2$ | $\underline{\mathbf{81.9 \pm 18.7}}$ | $\underline{\mathbf{73.5 \pm 11.4}}$ |

## E.3 EMOTION RECOGNITION

Table 7: **Emotion recognition per dataset results.** Average of test-set scores (balanced accuracy (%); higher is better; mean ± 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 | $84.3 \pm 13.3$ | $55.0 \pm 18.0$ |
| HEEGNet | DSMDBN | $84.1 \pm 14.7$ | $\mathbf{79.4 \pm 20.8}$ |
| | DSMDBN+EA | $\mathbf{89.7 \pm 11.2}$ | $78.7 \pm 12.4$ |

# F MOTOR IMAGERY EXPERIMENTS

## F.1 DATASET DETAILS

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

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

## F.2 MOTOR IMAGERY BASELINES

We consider two public available motor imagery datasets. Pre-processing included resampling EEG signals to 250 or 256 Hz, applying temporal filters to capture frequencies between 4 and 36 Hz, and extracting 3-second epochs linked to specific class labels. We further include the TSMNet (Kobler et al., 2022) as the considered baseline methods, SFUDA method spddsbn is proposed by (Kobler et al., 2022). The results suggest that our proposed HEEGNet achieves competitive performance, even on datasets for which hierarchical information has not been previously reported.

Table 8: **Motor imagery per dataset results.** Average of test-set scores (balanced accuracy (%); higher is better; mean ± std).

| Model | SFUDA | Dataset | |
| --- | --- | --- | --- |
| | | **BNCI2014001** | **BNCI2015001** |
| 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 | $49.7 \pm 16.9$ | $69.9 \pm 14.6$ |
| EEGInceptionMI | w/o | $39.7 \pm 12.7$ | $59.5 \pm 9.2$ |
| TSMNet | w/o | $43.0 \pm 13.3$ | $61.7 \pm 11.4$ |
| | EA | $51.2 \pm 15.1$ | $72.5 \pm 13.6$ |
| | STMA | $52.5 \pm 16.4$ | $70.1 \pm 14.2$ |
| | SPDDSBN | $\mathbf{54.6 \pm 16.1}$ | $74.3 \pm 14.7$ |
| HEEGNet | DSMDBN+EA | $54.1 \pm 15.9$ | $\underline{\mathbf{75.8 \pm 13.0}}$ |

# G   FULLY HYPERBOLIC NEURAL NETWORK EXPERIMENT

To investigate the necessity of the hybrid design in HEEGNet, we conducted ablation experiments by progressively replacing Euclidean convolutional operations in EEGNet (Lawhern et al., 2018) with hyperbolic variants. The original EEGNet architecture contains four convolutional layers, and we systematically substituted them as follows:

- **HEEGNet0**: Replaced the final MLR layer with hyperbolic MLR
- **HEEGNet1**: Additionally replaced the point-wise convolution
- **HEEGNet2**: Further replaced the depthwise convolution layer
- **HEEGNet3**: Further replaced the spatial convolution layer
- **HEEGNet4**: Fully hyperbolic models

As shown in Tab. 9, the performance on the Nakanishi dataset demonstrates a clear degradation pattern as more Euclidean operations are replaced with hyperbolic variants. The baseline EEGNet achieves $57.2 \pm 19.8\%$ accuracy, while HEEGNet0 shows improved performance at $60.8 \pm 20.7\%$. However, further replacements lead to consistent performance drops: HEEGNet1 ($53.9 \pm 19.1\%$), HEEGNet2 ($53.4\pm17.5\%$), HEEGNet3 ($52.7\pm17.1\%$), and the fully hyperbolic HEEGNet4 achieving only $27.3 \pm 12.0\%$. These result motivated us to keep a hybrid design.

Table 9: **Fully hyperbolic experiments**.Average of test-set scores on different (balanced accuracy (%); higher is better; mean ± std).

| Model | Nakanishi |
| --- | --- |
| EEGNet | $57.2 \pm 19.8$ |
| HEEGNet0 | $60.8 \pm 20.7$ |
| HEEGNet1 | $53.9 \pm 19.1$ |
| HEEGNet2 | $53.4 \pm 17.5$ |
| HEEGNet3 | $52.7 \pm 17.1$ |
| HEEGNet4 | $27.3 \pm 12.0$ |