Exploring the Enigma of Neural Dynamics Through A Scattering-Transform Mixer Landscape for Riemannian Manifold

Tingting Dan [1] **Ziquan Wei** [1 2] **Won Hwa Kim** [3] **Guorong Wu** [1 2 4 5 6]

Abstract

The human brain is a complex inter-wired system that emerges spontaneous functional fluctuations. In spite of tremendous success in the experimental neuroscience field, a system-level understanding of how brain anatomy supports various neural activities remains elusive. Capitalizing on the unprecedented amount of neuroimaging data, we present a physics-informed deep model to uncover the coupling mechanism between brain structure and function through the lens of data geometry that is rooted in the widespread wiring topology of connections between distant brain regions. Since deciphering the puzzle of self-organized patterns in functional fluctuations is the gateway to understanding the emergence of cognition and behavior, we devise a geometric deep model to uncover manifold mapping functions that characterize the intrinsic feature representations of evolving functional fluctuations on the Riemannian manifold. In lieu of learning unconstrained mapping functions, we introduce a set of graph-harmonic scattering transforms to impose the brain-wide geometry on top of manifold mapping functions, which allows us to cast the manifold-based deep learning into a reminiscent of *MLP-Mixer* architecture (in computer vision) for Riemannian manifold. As a proof-of-concept approach, we explore a neural-manifold perspective to understand the relationship between (static) brain structure and (dynamic) function, challenging the prevailing notion in cognitive neuroscience by proposing that neural activities are essentially excited by brain-wide oscillation waves living on the geometry of human connectomes, instead of being confined to focal areas.

[1]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA. [2]Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA. [3]Computer Science and Engineering / Graduate School of AI, POSTECH, Pohang, 37673, South Korea. [4]Department of Statistics and Operations Research (STOR), University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA. [5]UNC NeuroScience Center, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA. [6]Carolina Institute for Developmental Disabilities, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599, USA. Correspondence to: Guorong Wu <grwu@med.unc.edu>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

1. Introduction

Human brain is a complex system physically wired by massive bundles of nerve fibers (Bassett & Sporns, 2017). On top of intertwined structural connectomes, ubiquitous neural oscillations emerge remarkable functional fluctuations, synchronizing across large-scale neural circuits, that support myriad high-level cognitive functions necessary for everyday living (Bressler & Menon, 2010). With the prevalence of structural and functional neuroimaging technology (aka. Magnetic Resonance Imaging, MRI) in many neuroscience studies, the idea of understanding the mind forms an important concept that the dynamic nature of human brain cannot be understood by thinking of the system as comprised of independent components (Terras, 2012). In this regard, there is a critical need to establish a system-level understanding of how the brain function emerges from anatomical structures and how the self-organized system behavior of functional fluctuations supports the cognitive states.

Like many dynamic systems in the universe, the evolving functional fluctuations manifest remarkable geometric patterns to the extent of self-organized spontaneous co-activation of neural activities. As shown in Fig. 1 (left), functional connectivity (FC), formed by the pairwise correlation between two time courses of BOLD (blood-oxygen-level-dependent) signals (Bullmore & Sporns, 2009), exhibits *small world* properties (Watts & Strogatz, 1998). From a data science perspective, the FC matrix lives on the high-dimensional Riemannian manifold of SPD (symmetric and positive-definite) matrices. Due to many well-studied mathematical properties of SPD matrix, tremendous efforts have been made to statistical modeling (You & Park, 2021), clustering (Qiu et al., 2015), and characterization of temporal dynamics (Dan et al., 2022a) from the manifold instances

of SPD matrices.

Although manifold-based deep learning often presents greater challenges compared to the counterpart machine learning backbones operating in Euclidean space, prioritizing the preservation of data geometry substantially improves both the reliability of model explanations and overall accuracy (Tiwari & Konidaris, 2022). In this regard, various manifold-based deep models have been proposed for learning feature representation of SPD matrices using Riemannian manifold algebras (Huang & Van Gool, 2017; Dan et al., 2022b). As shown in Fig. 1, the driving factor of most deep models for SPD matrices is to find a cascade of mapping functions that progressively project the input SPD matrices into a latent subspace of Riemannian manifold which is in line with the downstream learning tasks. Since there is no constraint on the mapping functions, every element in the mapping function is *independently* updated through the gradients in back-propagation. Thus, it is evident that (1) the learned mapping functions lack explainability, (2) such deep model may demand a relatively large volume of training data, and (3) adapting a pre-trained model to a new dataset may pose challenges.

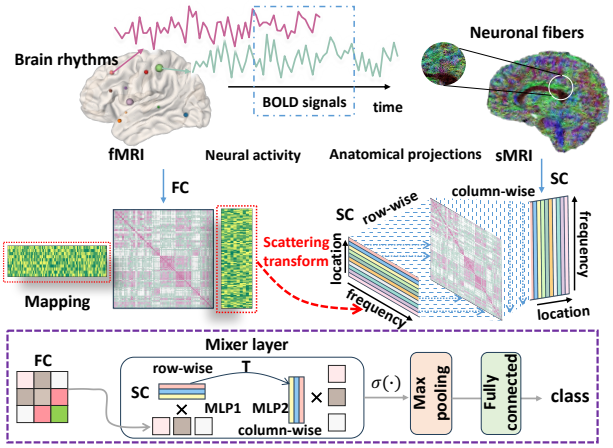

Figure 1. Overview of answering open neuroscience questions using machine learning techniques. *Top*: The motivation of our work is to understand how brain structure supports ubiquitous functional fluctuations. *Middle*: Compared to conventional mapping functions (left) in manifold-based deep learning, we introduce scattering transform to form the basis of mapping functions on the Riemannian manifold (right), which allows us to constrain the mapping of functional connectivities underlying the geometric of structural region-to-region connectome. *Bottom*: We devise a Mixer architecture on top of the row-wise and column-wise scattering transform which not only yields a novel geometric deep model for Riemannian manifold but also provides a new neuroscience insight to understand the coupling mechanism between brain structure and function.

To address these limitations, we introduce the notion of scattering transform to impose a geometric constraint on top of the mapping functions in the Riemannian manifold. Following the notion of recent pioneering work on brain structural and functional coupling (Pang et al., 2023), each scattering transform operator is derived from the graph spectrum of structural connections, essentially operating as harmonic wavelets at various oscillation frequencies (Hammond et al., 2018). To link manifold-based deep learning with scattering transforms, we present a novel mapping mechanism on the Riemannian manifold that allows us to preserve the data geometry in machine learning. Furthermore, we dissect the mapping process on Riemannian manifold into a row-wise and another column-wise scattering transforms (Fig. 1, middle-right) on each SPD matrix, which sets the stage for the learning scenario of Mixer architecture, a reminiscent of *MLP-Mixer* for computer vision (Tolstikhin et al., 2021). Due to the simple network architecture, our scattering-transform Mixer on Riemennian manifold of SPD matrices can be trained efficiently compared to existing manifold-based deep models with a significantly reduced number of parameters. Meanwhile, graph-harmonic scattering transforms offer a new window to interpret the brain-wide contribution of mapping functions on the coupling mechanism between brain structure and function.

Our work makes a significant contribution to the forefront of computational neuroscience, enriching traditional experimental neuroscience through the novel insight into brain structure-function coupling mechanism from the deep model. Specifically, we put the spotlight on the physics principle of scattering-transform in the Mixer, which acts as a stepping stone to understanding the mechanistic role of brain structure on functional fluctuations. We use the geometric pattern of graph harmonics to show that self-organized neural dynamics can be parsimoniously understood as resulting from the wave-to-wave interference formed by superimposing the fundamental resonance modes of the brain's geometry (i.e., topology of region-to-region connectivities) on the subject-specific neural activities.

Furthermore, we connect the interference phenomenon discovered in our deep model to the holography technique (Gabor, 1948) in computer vision, a stereo-imaging technique that generates a hologram by superimposing a reference beam on the wavefront of interest. Following this notion, we present a proof-of-concept theory, called *DeepHoloBrain*, which computationally "records" the cross-frequency couplings (CFC) of time-evolving interference patterns where spontaneous functional fluctuations are constrained by the graph harmonic waves (with predefined oscillation frequencies).

In practice, we evaluate the clinical value of our manifold mixer model in (1) disease early diagnosis and (2) generality of adapting the pre-trained model to a new dataset. Compared with current state-of-the-art deep models, our method not only achieves the best performance in terms of accuracy and consistency but also shows potential in addressing

real-world challenges in clinical practice, particularly in situations where disease data cohorts often have very limited sample sizes.

Together, the contribution of our work has four folds: (1) a new mapping mechanism on the Riemannian manifold constrained by scattering transforms, (2) a novel geometric deep model of scattering-transform Mixer for Riemannian manifold of SPD matrices, (3) a new gateway to understanding the human brain using deep learning technique, and (4) a comprehensive evaluation of the clinical impact on large-scale neuroimages.

2. Background and Related Works

2.1. Open Questions in Network Neuroscience

In the realm of neuroscience, a plethora of studies investigate connections between N predefined regions in the brain (Bassett & Sporns, 2017). Through diffusion-weighted imaging (DWI), we are able to measure the structural connections (SC) between two brain regions Ω_i and Ω_j $(i, j = 1, ..., N)$, *in vivo*, by tracking bundles of nerve fibers. On the other hand, functional magnetic resonance imaging (fMRI) technique enables to capture spontaneous functional connections (FC) which indicate the synchronization between two time courses of neural activities at different brain regions Ω_i and Ω_j. In general, we can form a network of SC or FC, where each element denotes the region-to-region SC/FC connection. A brief summary of an analytic pipeline for constructing SC and FC networks is shown in Appendix A.1. In contrast to the diverse FC patterns associated with cognition and behavior, SC remains relatively static. In this context, several interesting scientific questions arise: *How does the structural foundation of the brain shape its dynamic functional activities? And, how does the coupling between SC and FC contribute to the emergence of cognition and behavior?*

Since each FC matrix is symmetry and positive definite (SPD), striking efforts have been made to understand the self-organized pattern of FC through the lens of Riemannian manifold of SPD matrices (You & Park, 2021; Dan et al., 2022b). Following this spirit, we seek to explore the SC-FC coupling mechanism using geometric deep models on the Riemannian manifold, where the mathematical insight might open a new window to answer neuroscience questions through machine learning.

2.2. Canonical Deep Model on SPD Matrices

Suppose we have a $N \times N$ SPD matrix $\mathbf{X}$ residing on Riemannian manifold, i.e., $\mathbf{X} \in Sym_d^+ (d = N)$, current deep models such as SPDNet (Huang & Van Gool, 2017) are trained to find a set of non-linear mapping functions to obtain a new feature representation $\mathbf{X}_l$ at l^{th} layer by:

$$\mathbf{X}_l = \mathbf{\Psi}_l \mathbf{X}_{l-1} \mathbf{\Psi}_l^\intercal, \tag{1}$$

where the learnable mapping function $\mathbf{\Psi}_l \in \mathbb{R}^{d_l \times d_{l-1}}$ is a row full-rank matrix (usually $d_l \ll d_{l-1}$). $\mathbf{\Psi}_l$ is called positive mapping in Riemannian manifold of SPD matrices (Bhatia, 2009), which ensure the output $\mathbf{X}_l$ maintains the geometric property of SPD matrix, i.e, $\mathbf{X}_l \in Sym_{d_l}^+$. Although it is a common practice to impose column-wise orthogonality on $\mathbf{\Psi}_l$, the optimization of each element in $\mathbf{\Psi}_l$ is completely independent in back-propagation. As shown in the middle-left of Fig. 1, the learned mapping function $\mathbf{\Psi}_l$ lacks interpretability in the context of data geometry. To address this limitation, we introduce the scattering transform technique to decompose each $\mathbf{\Psi}_l$ into two dimensions: frequency and location (the middle-right of Fig. 1).

2.3. Scattering Transforms on Graph

The scattering transform (such as wavelets) (Mollai, 2010; Mallat, 2012; Gama et al., 2019) offers a stable and multi-scale representation for signals in the Euclidean space $\mathbb{R}^d$. A collection of pre-defined scattering transforms can decompose any signals into scale and orientation components separately, which allows us to capitalize on geometric intuition while maintaining useful properties for machine learning such as translational invariance. Following this spirit, the scattering transform technique has been extended to geometric data living on the irregular domains (such as graph and manifold) (Perlmutter et al., 2018; 2020; Gao et al., 2019).

Consider a smooth, compact, and connected N-dimensional Riemannian manifold $\mathcal{M}$ and a square-integrable function $f \in \mathbf{L}^2(\mathcal{M})$. The negative Laplace-Beltrami operator $-\Delta$ on $\mathcal{M}$ possesses a countable set of eigenvalues $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_{K-1}$. There exists a corresponding sequence of eigenfunctions $u_k (k = 0, ..., K-1)$ satisfying $-\Delta u_k = \lambda_k u_k$, where $\{u_k\}_{k \geq 0}$ forms an orthonormal basis for the manifold $\mathcal{M}$. Following the notion in (Perlmutter et al., 2020; 2018), the eigenfunctions $\{u_k\}_{k \geq 0}$ serve as generalized Fourier modes on the manifold $\mathcal{M}$ that is spanned by the inner product with these eigenfunctions. In this context, any function f can be represented as a sum over these eigenfunctions as $f(x) = \sum_{k \geq 0} \widehat{f}(k) u_k(x)$, where $\widehat{f}(k) = \langle f, u_k \rangle_{\mathbf{L}^2(\mathcal{M})} = \int_{\mathcal{M}} f(y) \overline{u_k(y)} dy$ is known as Fourier coefficient corresponding to the eigenfunction u_k.

Furthermore, we can expand Fourier's bases to the wavelet scattering transform on manifold $\mathcal{M}$ by localizing each eigenfunction u_k to particular region $\Omega_i (i = 1, ..., N)$, yielding a frequency-specific and location-specific harmonic wavelet function ψ_i^k. In this work, we construct harmonic wavelets on top of the graph topology (Hammond et al., 2011; Shuman et al., 2013). Suppose the underlying graph has N nodes and the node-to-node connection is encoded

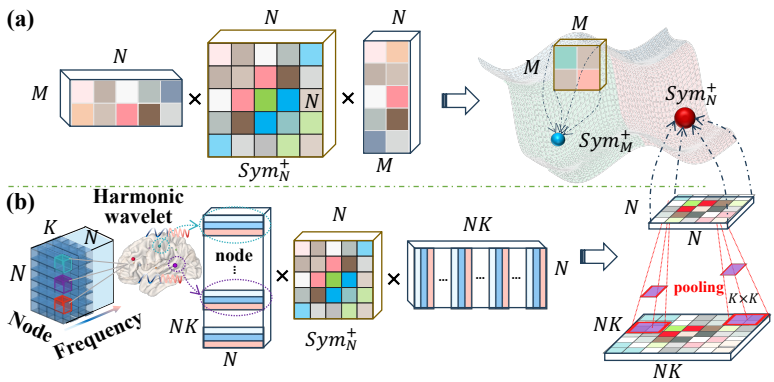

Figure 2. (a) Conventional mapping function (Eq. 1) on Riemannian manifold of SPD matrices, where each element in the mapping matrix is estimated independently. (b) Our new mapping function which consists of (1) constructing the block matrix $\mathbb{P}$ by stacking harmonic wavelets over frequency and location, (2) yielding a *Supra-FC* matrix $\mathbb{X} = \mathbb{P}^\intercal \boldsymbol{X} \mathbb{P}$ by Eq. 1, and (3) applying pooling operation on $\mathbb{X}$ across frequencies at each location. By doing so, the output is a $N \times N$ SPD matrix.

in a $N \times N$ adjacency matrix $\mathbf{A}$. Thus, the graph Laplacian matrix $\mathcal{L}$ can be constructed by $\mathcal{L} = \mathbf{D} - \mathbf{A}$, where $\mathbf{D} \in \mathbb{R}^{N \times N}$ is a diagonal matrix of node-wise total connectivity degree. Next, we apply eigen-decomposition on $\mathcal{L}$, yielding $\mathcal{L} = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^\intercal$, where $\boldsymbol{\Lambda} = diag[\lambda_k]_{k=1}^N$ is a diagonal matrix of eigenvalues and $\boldsymbol{U} = [u_k]_{k=1}^N$ are eigenvectors, each corresponding to the eigenvalue λ_k. For each eigenvector u_k, we can derive in total N harmonic wavelets by localizing u_k to the underlying region Ω_i by (Shuman, 2020):

$$\psi_i^k = \boldsymbol{U} g_k(\gamma \boldsymbol{\Lambda}) \boldsymbol{U}^\intercal \delta_i \tag{2}$$

where δ_i is a graph signal with a value of one at region Ω_i and zero elsewhere. $g_k(\cdot) = e^{-\gamma\lambda_k}$ is a spectral filter that defines the spectral pattern that is localized to each region Ω_i, where γ is a learnable hyper-parameter that controls the scale of spectral pattern.

3. Methods

3.1. New Mapping Function on Riemannian Manifold Constrained by Graph Scattering Transforms

Suppose both SC and FC networks consist of N nodes. We construct $K(K \leq N)$ harmonic wavelets at each node Ω_i of SC network, denoted by $\boldsymbol{\Psi}_i = [\psi_i^k]_{k=1}^K$, where k indicates harmonic frequency. For convenience, we use $\mathbb{P} = [[\boldsymbol{\Psi}_1], [\boldsymbol{\Psi}_2], ..., [\boldsymbol{\Psi}_N]] \in \mathbb{R}^{N \times NK}$ to denote the block matrix by stacking N region-specific matrices of harmonic wavelets $\boldsymbol{\Psi}_i$ one after another, as shown in Fig. 2 (b). Next, we present a new mapping function, i.e., applying scattering transforms $\mathbb{P}$ on FC matrix $\boldsymbol{X}$, yielding a new SPD matrix on the Riemannian manifold $\mathcal{M}$.

Remark 3.1. A real and symmetric matrix $\boldsymbol{X} \in Sym_N^+$ is said to be SPD matrix if $v\boldsymbol{X}v^\intercal = \sum_{i,j=1}^N v_i x_{ij} v_j > 0$ for any non-zero vector in $v \in \mathbb{R}^N$.

Lemma 3.2. *Given a full-rank matrix $\boldsymbol{W} \in \mathbb{R}^{N \times M}(M < N)$ and a vector $v \in \mathbb{R}^M$, since the columns of $\boldsymbol{W}$ are linearly independent and span $\mathbb{R}^M$, there exists, by the fundamental theorem of linear algebra, a vector $u \in \mathbb{R}^N$ such that $\boldsymbol{W}u = v$ (Proof shown in Proposition A.1).*

Remark 3.3. Given an SPD matrix $\boldsymbol{X} \in Sym_N^+$ and a full-rank matrix $\boldsymbol{W} \in \mathbb{R}^{N \times M}(M < N)$, the function $f : Sym_N^+ \times \mathbb{R}^{N \times M} \to Sym_M^+$ defined as $f(\boldsymbol{X}, \boldsymbol{W}) = \boldsymbol{W}^\intercal \boldsymbol{X} \boldsymbol{W}$ (Huang & Van Gool, 2017).

Proposition 3.4. *The output of $f(\boldsymbol{X}, \boldsymbol{W})$ is a $N \times N$ SPD matrix.*

Proof. According to Lemma 3.2, there exists a vector $u \in \mathbb{R}^N$ such that $\boldsymbol{W}u = v$. Note that u is not the zero vector because v is not zero and $\boldsymbol{W}$ has full rank. Substitute $v = \boldsymbol{W}u$ into the quadratic form: $v^\intercal(\boldsymbol{W}^\intercal \boldsymbol{X} \boldsymbol{W})v = (\boldsymbol{W}u)^\intercal(\boldsymbol{W}^\intercal \boldsymbol{X} \boldsymbol{W})(\boldsymbol{W}u) = u^\intercal \boldsymbol{W}\boldsymbol{W}^\intercal \boldsymbol{X} \boldsymbol{W}\boldsymbol{W}^\intercal u$. Since $\boldsymbol{W}$ has full rank and $M < N$, $\boldsymbol{W}\boldsymbol{W}^\intercal$ is positive definite. As the product of positive definite matrices $\boldsymbol{W}\boldsymbol{W}^\intercal$, $\boldsymbol{X}$, and $\boldsymbol{W}\boldsymbol{W}^\intercal$ yields another positive definite matrix, we have $u^\intercal(\boldsymbol{W}\boldsymbol{W}^\intercal \boldsymbol{X} \boldsymbol{W}\boldsymbol{W}^\intercal)u > 0$. Because u is not the zero vector, the quadratic form is strictly positive. This proves that the output of $f(\boldsymbol{X}, \boldsymbol{W}) = \boldsymbol{W}^\intercal \boldsymbol{X} \boldsymbol{W}$ is indeed an SPD matrix (as shown in Fig. 2 (a)). □

Proposition 3.5. *If the number of rows in $\boldsymbol{W}$ is less than the number of columns, i.e, $N < M$, the mapping $f(\boldsymbol{X}, \boldsymbol{W}) = \boldsymbol{W}^\intercal \boldsymbol{X} \boldsymbol{W}$ only yields a semi-positive definiteness matrix.*

Proof. Similar to the proof for Proposition 3.4, we examine the quadratic form given by $v^\intercal(\boldsymbol{W}^\intercal X \boldsymbol{W})v = (\boldsymbol{W}v)^\intercal X(\boldsymbol{W}v)$ for any vector v. Since $N < M$, there exist a vector v such that $\boldsymbol{W}v = 0$, resulting in $f = (\boldsymbol{X}, \boldsymbol{W}) = 0$. Since $\boldsymbol{X}$ is SPD, $(\boldsymbol{W}v)^\intercal \boldsymbol{X}(\boldsymbol{W}v) > 0$ as long as $\boldsymbol{W}v$ is non-zero vector. Together, we have proven Proposition 3.5. □

Recall that $\mathbb{P} \in \mathbb{R}^{N \times NK}$ is a block matrix (primarily indexed by location Ω_i), where each block element is a $N \times K$ matrix. Since $N < NK$ (row number less than column number), $\tilde{\mathbb{P}} = f(\boldsymbol{X}, \mathbb{P})$ does not guarantee that the output matrix resides on the Riemannian manifold of SPD matrices. However, $\mathbb{X} = \mathbb{P}^\intercal \boldsymbol{X} \mathbb{P} \in \mathbb{R}^{NK \times NK}$ essentially expand each (scalar) element in $\boldsymbol{X}$ to a $K \times K$ matrix. In this context, we coin $\mathbb{X}$ as *Supra-FC* matrix, where each element $\mathbb{X}_{ij}$ is a $K \times K$ block matrix, i.e., $\mathbb{X}_{ij} = [\psi_i^s \boldsymbol{X} \psi_j^t]_{s,t=1}^K$. In light of this, we seek to use the following pooling operation to regain the SPD property while reducing the dimensionality.

Proposition 3.6. *Given a Supra-FC matrix $\mathbb{X}$, we apply max-pooling operation to each block matrix $\mathbb{X}_{ij}$ (over frequencies s and t). Thus, the resulting $N \times N$ matrix*

$$\tilde{\boldsymbol{X}} = \begin{bmatrix} max(\mathbb{X}_{11}) & max(\mathbb{X}_{12}) & \cdots & max(\mathbb{X}_{1N}) \\ max(\mathbb{X}_{21}) & max(\mathbb{X}_{22}) & \cdots & max(\mathbb{X}_{2N}) \\ \vdots & \vdots & \ddots & \vdots \\ max(\mathbb{X}_{N1}) & max(\mathbb{X}_{N2}) & \cdots & max(\mathbb{X}_{NN}) \end{bmatrix}$$

is positive-definite.

Proof. We sketch the proof as follows. By substituting the formulation of wavelets into each $\psi_i^s \boldsymbol{X} \psi_j^t$, we have $\mathbb{X}_{ij}[s,t] = g(\lambda_s)g(\lambda_t)u_s(i)u_s(j)\boldsymbol{X}u_t(i)u_t(j)$ (for convenience, we skip the scaling effect γ and index value δ in Eq. 2). Next, we prove that the lower bound of $max(u_s(i)u_s(j)\boldsymbol{X}u_t(i)u_t(j))$ is $\frac{1}{N}$ in Proposition A.2 of Appendix. Recall that $g(\cdot)$ is an exponential function, we show that $\tilde{X}$ is positive-definite in Proposition A.3 of Appendix since $\tilde{X} = e^B$ where B is a $N \times N$ matrix. □

Remarks. In Fig. 2(b), we demonstrate the workflow of new mapping function on the Riemannian manifold which includes (1) constructing *Supra-FC* matrix $\mathbb{X} = \mathbb{P}^\intercal \boldsymbol{X} \mathbb{P}$, (2) applying block-wise max-pooling operation $\boldsymbol{X} = \text{max-pooling}(\mathbb{X})$ via Proposition 3.6, and (3) make $\tilde{X}$ symmetric by $\tilde{X} \leftarrow \frac{1}{2}(\tilde{X} + \tilde{X}^\intercal)$ (refer to Proposition A.4 in Appendix). Recall that there is a learnable parameter γ in Eq. 2. In this regard, we present the following deep model to optimize scattering transforms in $\mathbb{P}$ through the lens of scaling effect on harmonic wavelets.

3.2. *MLP-Mixer* for Riemannian Manifold

Since each row in $\mathbb{P}$ is associated with a harmonic wavelet, the operation of $\mathbb{P}^\intercal \boldsymbol{X} \mathbb{P}$ can be boiled down to first apply scattering transform to each column of $\boldsymbol{X}$ followed by same operation on its rows. Inspired by the efficient *MLP-Mixer* architecture in computer vision (Tolstikhin et al., 2021), we propose to deploy a set of MLPs to learn the optimal scaling effect γ in the column-wise scattering transforms $\mathbb{P}X$ and then another set of MLPs for the optimal scaling effect γ in the counterpart row-wise scattering transforms. Suppose σ_{row} and σ_{column} denote the MLPs (followed by the ReLU layer on manifold (Huang & Van Gool, 2017)) for row-wise and column-wise scattering transforms, respectively. The *MLP-Mixer* architecture of positive mapping $\mathbb{P}^\intercal X \mathbb{P}$ on Riemannian manifold can be formulated as:

$$\begin{cases} \mathbb{X}_{l+1} = \sigma_{row}\{\text{LN}[\sigma_{column}(\text{LN}(\mathbb{P}^\intercal \tilde{X}_l))\mathbb{P}]\} \\ \tilde{X}_{l+1} = \text{max-pooling}(\mathbb{X}_{l+1}) \\ \tilde{X}_{l+1} = \dfrac{1}{2}(\tilde{X}_{l+1} + (\tilde{X}_{l+1})^\intercal) \end{cases} \tag{3}$$

where LN$(\cdot)$ denotes LayerNorm.

Remarks. It is clear that the input to each *MLP-Mixer* is a SPD matrix $\tilde{X}_l$. The backbone of positive mapping on the Riemannian manifold is nothing but a set of MLPs and block-wise max-polling operations, yielding a new SPD matrix $\tilde{X}_{l+1}$ in the same dimensions. In our implementation, we employ the classic positive mapping function in Eq. 1 to learn the low-dimensional feature representations. After that, we project the learned SPD matrix to the tangent space (Huang et al., 2015) and then connect to the fully-connected layer for various machine learning applications such as classification.

3.3. *DeepHoloBrain*: A Proof-of-Concept Approach to Explore the Enigma of Neural Dynamics Through the Insight of Deep Model

Scattering Transforms: Stepping Stone between Deep Learning and SC-FC Couplings. In functional MRI studies, the FC matrix $\boldsymbol{X}$ encodes pairwise correlations between two time course of neural activities $h_i = [h_i(1), h_i(2), ..., h_i(T)]$ at region Ω_i and $h_j(t)$ at region Ω_j, where T is the number of time points. For convenience, we stack each h_i row-by-row and form a whole-brain time course matrix $\boldsymbol{H} \in \mathbb{R}^{N \times T}$ (purple lines in Fig. 3). In this context, the FC matrix can be computed by $\boldsymbol{X} = \boldsymbol{H}\boldsymbol{H}^\intercal$ (brown box), where each element $x_{ij} = h_i h_j^\intercal$.

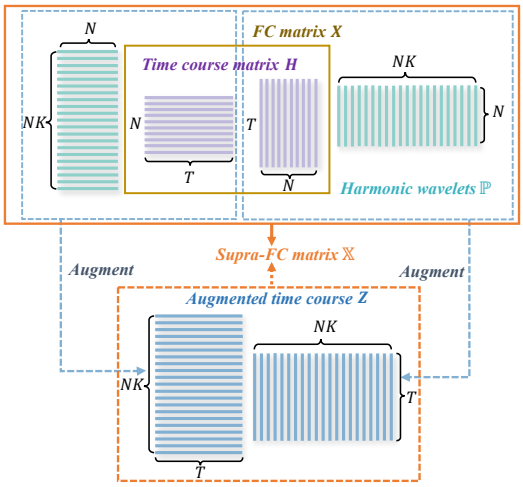

Figure 3. Two ways to compute *Supra-FC* matrix $\mathbb{X}$. Top: We apply scattering transforms $\mathbb{P}$ on FC matrix $\boldsymbol{X}$ (brown box) and obtain *Supra-FC* matrix by $\mathbb{X} = \mathbb{P}^\intercal \boldsymbol{X} \mathbb{P}$. Bottom: An alternative way is to break the FC matrix into the covariance of time course matrix $\boldsymbol{H}$ (purple lines in the brown box). Then the inner product between each scattering transform and the snapshot of time course (columns in $\boldsymbol{H}$ is essentially the result of modulating the whole-brain signal of neural activity using harmonic wavelets from SC, which is constrained by the geometry of brain anatomy. Eventually, the correlation between augmented neural signals $\boldsymbol{Z}$ (blue lines) results in the same *Supra-FC* matrix $\mathbb{X}$, accompanied by an in-depth neuroscience underpinning for SC-FC coupling mechanisms.

In Proposition 3.6, we apply harmonic wavelets on FC matrix $\boldsymbol{X}$. Here, we break down FC matrix $\boldsymbol{X}$ into $\boldsymbol{H}\boldsymbol{H}^\intercal$. Furthermore, we study each column in $\boldsymbol{H}$, which is the whole-brain snapshot of neural activities $h^t \in \mathbb{R}^N$ at time t.

As the blue dashed box in Fig. 3, we first apply harmonic wavelets $\mathbb{P}$ to each snapshot h^t, yielding an augmented time course matrix $\boldsymbol{Z} = \mathbb{P}\boldsymbol{H} \in \mathbb{R}^{NK\times T}$ (blue lines, compared to original signals $\boldsymbol{H}$). After that, the inner project of $\boldsymbol{Z}$ and $\boldsymbol{Z}^\intercal$ results in the *Supra-FC* matrix $\mathbb{X}$ (orange dashed box).

Neuroscience Insights. The oscillation patterns of each harmonic wavelet ψ_i^k, constrained by the local topology of the structural connectome, characterize the frequency-specific neural activities supported by the underlying neural circuit. The inner project $\langle \psi_i^k, h^t \rangle$ over time essentially allows us to modulate the observed neural activity signals with the predefine bandpass filters, which gives rise to coupled neural oscillations at distinct frequencies. In analogy to the holography technique in computer vision, the physics insight behind is that $\mathbb{X} = \boldsymbol{Z}\boldsymbol{Z}^\intercal$ (orange dashed box in Fig. 3) records interference patterns generated by two SC-modulated neural activity signals.

Since the geometry of harmonic wavelets is governed by the topology from SC, we have initialed the effort of building an integrated framework to analyze the SC-FC coupling mechanism using well-studied interference principles, which yields a new research paradigm using deep learning technology, called *DeepHoloBrain*.

Potential Applications. *First*, our manifold-based deep model can be used to investigate SC-FC coupling mechanisms in response to specific tasks or stimuli. Specifically, the input to our model is time course from task-fMRI images. The driving factor of machine learning is to predict the underlying cognitive tasks based on the learned intrinsic feature representation of functional fluctuations that describe the latent brain states. In addition, we integrate node-specific and frequency-specific attention components on top of the scattering transforms $\mathbb{P}$, which allows us to uncover how task-dependent spontaneous FC is supported by a collection of anatomical neural circuits in SC.

Second, our approach offers an integrated approach to predict disease risks using both SC and FC information. Current machine learning methods primarily focus on feature fusion (Sarwar et al., 2021). We model the SC-FC relationship through scattering transforms with great mathematical insight. Thus, our model offers in-depth network neuroscience underpinning to uncover the synergistic effect of brain structure/function and elucidate their mechanistic role in modifying cognitive performance in the context of neurological diseases.

4. Experiments

4.1. Data Description and Experimental Setting

Data Description. To thoroughly validate the effectiveness of our proposed model, we utilize three existing public datasets for our experiments.

The Lifespan Human Connectome Project Aging (HCP-A) dataset (Bookheimer et al., 2019). HCP-A is instrumental in task recognition research, offering a comprehensive view of the aging process. It includes data from 717 subjects, encompassing both fMRI (4,846 time series) and DWI (717) scans. This rich collection facilitates in-depth analyses of both functional and structural connectivity. HCP-A dataset includes data from four brain tasks associated with memory: VISMOTOR, CARIT, FACENAME, and Resting State. In the related experiments, these tasks are treated as distinct categories in a four-class classification problem.

Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (Weiner et al., 2015). The ADNI dataset serves as an invaluable resource, featuring a collection of 250 fMRI time series and 1,012 meticulously processed structural connectomes. In our experiment, we selectively focused on 250 subjects who had undergone both DWI and fMRI scans. This careful selection ensured that we had access to detailed data on both functional connectivity and structural connectomes for each subject. Additionally, the ADNI data includes clinical diagnostic labels, encompassing a spectrum of cognitive states: Cognitive Normal (CN), Subjective Memory Complaints (SMC), Early-Stage Mild Cognitive Impairment (EMCI), Late-Stage Mild Cognitive Impairment (LMCI), and Alzheimer's Disease (AD). Considering the data balance issue, we simplified these categories into two broad groups based on disease severity. Specifically, we combined CN, SMC, and EMCI into a single 'CN' group, representing less severe conditions, while LMCI and AD were grouped together as the 'AD' group. This categorization is intended to facilitate a binary classification framework for our analysis.

Open Access Series of Imaging Studies (OASIS) dataset (LaMontagne et al., 2019). The OASIS dataset presents a substantial collection of data from 924 subjects, comprising 3,322 fMRI sessions in total. Each subject has BOLD signals and structural connectome data, forming a comprehensive foundation for our analysis. In our experiment, we focused on binary classification: subjects in preclinical stages of Alzheimer's disease (categorized as preclinical-AD) or those manifesting dementia-related conditions (categorized as AD), while healthy individuals were classified under the 'CN' group. This classification strategy aids in diagnosing the preclinical conditions at the early stage.

We summarize the processing steps for SC and FC data in the Appendix A.1. In all of the following experiments, we partition each into 90 regions using AAL atlas (Tzourio-Mazoyer et al., 2002). Thus, SC is a 90×90 matrix where each element is quantified by the number of fibers linking two brain regions. We further normalize the fiber count by the total fiber count of the underlying subject. Regarding

FC, we first calculate the mean time course of BOLD signals in each region from fMRI. Then FC is a 90×90 matrix by Pearson's correlation between mean time courses in different regions, which indicates temporal synchronization of neuronal activity in the brain (as shown in Fig. 6).

Experimental Setup. In our evaluations, we have conducted a thorough evaluation of the proposed *DeepHoloBrain* model, focusing on two key classification challenges in neuroscience: brain task recognition and disease diagnosis. Since the geometric patterns in FC matrix have been widely investigated in the neuroscience field (You & Park, 2021), our comparative analysis includes a range of state-of-the-art SPD matrix learning methods: Covariance Discriminative Learning (CDL) (Wang et al., 2012), Log-Euclidean Metric Learning (LEML) (Huang et al., 2015), SPD Manifold Learning (SPDML) (Harandi et al., 2014), Affine-Invariant Metric (AIM) (Pennec et al., 2006), Riemannian Sparse Representation (RSR) (Harandi et al., 2012), DeepO2P network (Ionescu et al., 2015), and SPDNet (Huang & Van Gool, 2017). For each method, we utilize the source codes and meticulously fine-tuned parameters according to the specifications provided by their respective authors to ensure the integrity of our evaluation. The effectiveness of these comparative methods has been rigorously gauged using a comprehensive set of metrics, including accuracy, recall, F1-score, and precision. For all classification tasks, we use a cross-entropy loss function that compares the ground truth $G_{m,c}$ for the cth class of the mth sample against the corresponding prediction $\hat{G}_{m,c}$. Regarding our deep model, a regularization term is added to the scale parameter γ in harmonic wavelets (Eq. 2) to prevent it from taking negative values:

$$\ell(\gamma) = -\frac{1}{M}\sum_{m=1}^{M}\sum_{c\in C} G_{m,c}\log \hat{G}_{m,c}(\gamma) + \beta|\gamma|[\gamma < 0] \tag{4}$$

Here, β is a hyperparameter, M represents the sample size, and $|\gamma|[\gamma < 0]$ denotes the absolute value of γ when it is less than zero. We update the scale parameter using the rule $\gamma \leftarrow \gamma - \alpha_\gamma \frac{\partial \ell}{\partial \gamma}$, where α_γ is the learning rate for γ. This update is performed via gradient-based methods along with other learnable parameters. For gradient back-propagation, we employ the Riemannian manifold optimizer (Huang & Van Gool, 2017), with a learning rate of 0.001 and over 1000 epochs, the batch size is set to 16, weight decay is set to 1e-5, momentum is set as 0.9, the transformation matrix of SPDNet is initialized to a semi-positive definite matrix ($N \times 64$). Results from each experimental setup are reported using five-fold cross-validation, with both the mean and standard deviation documented for thorough analysis. The permutation t-test (the number of permutations is set to 10000) is used for statistical tests. All experiments were conducted on an Intel(R) Xeon(R) Gold 6448Y, paired with an NVIDIA RTX 6000 GPU.

4.2. Performance on Task-specific Recognition

In this series of experiments, we demonstrate task-specific recognition using several methods including SPDNet, CDL, LEML, SPDML, AIM, RSR, DeepO2P, and our *DeepHoloBrain* (denoted as 'OURS') on the HCP-A dataset. The numerical results, presented in Table 1, clearly indicate that *DeepHoloBrain* significantly outperforms (indicated by '*') the other seven methods, at a significance level $p < 0.05$. Notably, all methods exhibit remarkable performance in this task. This can be attributed to the distinctively designed tasks (VISMOTOR, CARIT, FACENAME, and Resting State) within the dataset. However, the transformation matrix $\mathbf{\Psi}$ in Eq. 1, a crucial component in these methods, is learned through a 'black-box' deep learning approach, leaving the exact nature of the parameters learned somewhat obscure. To gain a deeper understanding of brain dynamics, we have developed two attention mappings: one for node-wise analysis and the other for graph wavelet construction (frequency). After retraining the model to focus solely on the VISMOTOR and FACENAME tasks, the visual results of the node-wise attention map, as shown in Fig. 4 (left), reveal a distinct spatial pattern in the sensorimotor and visual regions (color bar denotes the degree). This indicates that anatomical regions within the areas detected by our model are highly relevant to the VISMOTOR and FACENAME tasks, respectively. Such findings align with current research (Bedel et al., 2023; Glasser et al., 2016), suggesting that our model is capable of capturing intrinsic features that differentiate various brain tasks. Furthermore, Fig. 4 (right) displays the oscillation patterns underlying the learned most relevant frequencies, indicating a distinct preference for high frequencies in the sensorimotor regions and lower frequencies within the visual region. This informative location-frequency attention map suggests that different cognitive tasks engage distinct neural circuits, manifesting as varied frequency responses across brain regions.

Table 1. Results on brain task recognition for HCP-Aging dataset.

METHODS	ACCURACY	RECALL	F1-SCORE
SPDNET	0.984± 0.003	0.975± 0.004	0.978± 0.004
CDL	0.976± 0.003	0.962± 0.005	0.966± 0.005
LEML	0.961± 0.022	0.903± 0.039	0.929± 0.036
SPDML	0.944± 0.015	0.908± 0.027	0.920±0.019
AIM	0.952± 0.014	0.9114± 0.016	0.929±0.015
RSR	0.966± 0.005	0.944± 0.010	0.951 ±0.008
DEEPO2P	0.977± 0.004	0.963± 0.006	0.969± 0.005
OURS	**0.995± 0.003***	**0.989± 0.003***	**0.993± 0.003***

4.3. Performance on Disease Diagnosis

Disease Early Diagnosis. We evaluate the prediction accuracy on ADNI and OASIS data, respectively, in forecasting the risk of developing AD by SPDNet, CDL, LEML, SPDML, AIM, RSR, DeepO2P and *DeepHoloBrain*

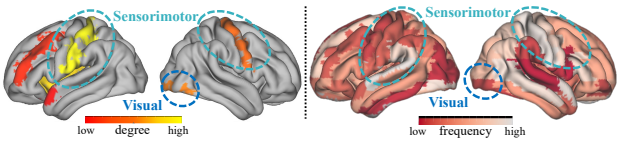

Figure 4. Left: Uncovered anatomical regions (using node-wise attention) for task VISMOTOR and task FACENAME, respectively. Right: Oscillation patterns underlying the more relevant harmonic frequencies for VISMOTOR and FACENAME tasks, revealing task-specific wavelet dynamics.

(marked 'OURS') in Table 2 (top and middle). At the significance level of 0.001, our method outperforms all other counterpart methods in terms of classification accuracy (indicated by '*').

Altered SC-FC Coupling in AD. Here, we seek to identify the alterations of SC-FC coupling that underline the progression of AD. In this regard, we apply the node-size attention module to characterize the regional contribution of SC-FC coupling in recognizing AD from CN, which implies the most vulnerable brain regions in AD progression. As illustrated in Fig. 5, the analysis on both OASIS and ADNI datasets reveals that the regions predominantly affected are within the default mode regions (highlighted by red circles). This finding aligns with numerous studies that have established a strong correlation between these regions and human memory (Greicius et al., 2004; Koch et al., 2012).

Performance on Learning Consistency across Data Cohorts. Herein, we focus on consistency in terms of learned attention maps from ADNI and OASIS. Although both are AD-specific studies, variations exist in the choice of scanner, study design, and demographic data of the subjects between the two studies. Particularly, we skip the data harmonization routine (Richter et al., 2022) and train our deep model on ADNI and OASIS separately. The spatial maps of region-specific and frequency-specific attentions are shown in Fig. 5 side by side. It is evident that the oscillation patterns in frequency-specific attention are quite consistent between ADNI and OASIS. Although we only find one region (circles in Fig. 5 bottom) consistently selected in two data cohorts, the top-ranked regions in OASIS and ADNI (circles in Fig. 5 top) are both located in the default-mode network (Greicius et al., 2004; Lee et al., 2020; Mevel et al., 2011), a sub-network frequently affected by AD.

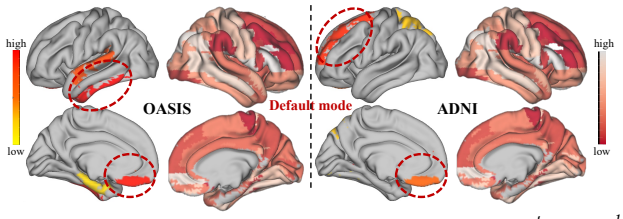

Figure 5. Consistency evaluation for region-specific (1st and 3rd columns) and frequency-specific (2nd and 4th columns) attentions learned from OASIS (left) and ADNI (right).

Table 2. Prognosis accuracies on forecasting AD risk for OASIS (top) and ADNI (shaded, middle) datasets. Bold font denotes the best performance. 'OURS' refers to our proposed model, named *DeepHoloBrain*. When we mention 'baseline' (darker shade), they specifically denote the version of the models that have been trained using a particular dataset (i.e., OASIS and ADNI). Additionally, 'baseline+' (lighter shade, bottom) represents an enhanced variant of the models, which has been pre-trained on the HCP-A dataset and subsequently fine-tuned on the ADNI dataset.

METHODS	ACCURACY	RECALL	F1-SCORE
SPDNET	0.871± 0.018	0.593± 0.017	0.613± 0.024
CDL	0.827± 0.623	0.570± 0.062	0.581 ±0.047
LEML	0.796± 0.193	0.624± 0.045	0.632± 0.117
SPDML	0.786± 0.148	0.739± 0.038	0.697±0.091
AIM	0.816± 0.098	0.540± 0.043	0.539±0.072
RSR	0.840± 0.020	0.690± 0.020	0.689 ±0.028
DEEPO2P	0.857 ± 0.019	0.684± 0.02	0.675± 0.021
OURS	**0.885±0.017***	**0.740± 0.045***	**0.6974± 0.041***
SPDNET	0.800±0.085	0.670± 0.067	0.627± 0.090
CDL	0.710±0.095	0.500± 0.018	0.415±0.064
LEML	0.704±0.095	0.523±0.019	0.474±0.065
SPDML	0.672± 0.079	0.543± 0.037	0.529± 0.055
AIM	0.708± 0.089	0.500± 0.000	0.413± 0.032
RSR	0.740± 0.106	0.610± 0.059	0.608± 0.080
DEEPO2P	0.760± 0.089	0.614± 0.068	0.625±0.082
OURS	**0.820± 0.071***	**0.625± 0.049***	**0.647±0.079***

	ACCURACY	RECALL	F1-SCORE	PRECISION
SPDNET+	0.7120	**0.7120**	0.6632	0.7409
CDL+	0.6640	0.5253	0.5168	0.5397
LEML+	0.6080	0.5501	0.5466	0.5462
SPDML+	0.5880	0.5803	0.5581	0.5671
AIM+	0.6160	0.5356	0.5356	0.5356
RSR+	0.6880	0.6880	**0.8690**	0.4716
DEEPO2P+	0.6960	0.6880	0.5971	0.7369
OURS+	**0.7400***	0.6103	0.6081	**0.7892***

Generality as A Pre-trained Model. One of the critical challenges of deploying computer-assisted diagnosis in clinical routine is the limited sample size, especially for disease cohorts. Considering the fact that (1) the pre-defined scattering transforms underscore the coupling mechanism of SC-FC relationship and (2) the *MLP-Mixer* architecture enables us to train the deep model with much fewer parameters than conventional methods, we hypothesize that our model pre-trained on a large dataset of normal healthy brains can produce reasonable classification results for disease diagnosis after fine-tuning on a limited amount disease-specific data. To test this hypothesis, we pre-train a regression model based on Montreal Cognitive Assessment (MoCA) score on HCP-A data and fine-tune a classification mode on ADNI data. The experimental results, as presented in Table 2 (bottom), it is apparent that our method has achieved significant improvement in accuracy and precision, with a marginal decrease in recall and F1-score. Furthermore, we compute the AUC for SPDNet and our *DeepHoloBrain*, where our method achieves 1.76% improvement over SPDNet. Considering early diagnosis of AD is critical in developing

interventional therapeutics, the results in Table 2 (bottom) suggest our method has better clinical value than SPDNet in terms of diagnosis sensitivity.

5. Discussion

In this section, we delve into the performance comparison of task recognition and disease diagnosis using graph-based and transformer-based methods, which have been extensively employed in neuroscience research. Specifically, we evaluate six methods, comprising four GCN-based approaches and two transformer models. The comparison encompasses GCN (Kipf & Welling, 2017), GAT (Veličković et al., 2017), GIN (Xu et al., 2018), BrainNetCNN (Kawahara et al., 2017), Graph Transformer (GraphT) (Shi et al., 2020), and BolT (Bedel et al., 2023), assessed on both the HCP-A dataset (top) and an ADNI dataset (bottom) in Table 3. Notably, for GCN-based methods, we utilize FC as the feature representation underlying the SC topology. For BolT and BrainNetCNN, only FC information is used in the deep model.

Table 3. Performance on task recognition for HCP-A dataset (top) and forecasting AD risk for ADNI datasets (bottom), respectively.

Methods	Accuracy	Precision	F1-score
GCN	0.690± 0.057	0.511± 0.157	0.582± 0.116
GAT	0.719± 0.123	0.540± 0.222	0.611 ±0.182
GIN	0.585± 0.124	0.513± 0.162	0.694± 0.066
GraphT	0.721± 0.127	0.542± 0.227	0.614±0.189
BrainNetCNN	0.725± 0.135	0.546± 0.235	0.617±0.195
BolT	0.863± 0.013	0.865± 0.015	0.864 ±0.014
GCN	0.541±0.313	0.447± 0.369	0.472± 0.366
GAT	0.659±0.287	0.562± 0.319	0.596±0.319
GIN	0.600±0.106	0.728±0.073	0.642±0.093
GraphT	0.541± 0.275	0.595± 0.297	0.497± 0.276
BrainNetCNN	0.800± 0.085	0.638± 0.124	0.643± 0.102
BolT	0.812± 0.075	0.618± 0.142	0.645± 0.105

It is evident that GCN-based methods yield less satisfactory results in recognizing cognitive tasks for normal adults (in HCP-A), largely due to the neuroscience fact that the topological pattern usually does not highly correlate with functional fluctuations exhibited in cognitive tasks. Meanwhile, some GCN-based methods are able to yield reasonable prediction accuracy of AD risk since SC alteration is a hallmark of AD progression in clinical findings. Taken together, these results suggest that effectively employing GCN-based techniques from machine learning in neuroscience requires tailored adaptation of domain-specific expertise. Following this spirit, our *DeepHoloBrain* integrates physics principles, neuroscience insights, and the power of machine learning, which allows us to consistently outperform existing deep models. A good integration of SC and FC information effectively provides a greater advantage than using SC or FC alone. For instance, although BolT uses a more advanced machine learning backbone (Transformer) than our *DeepHoloBrain* (MLPs), BolT (using FC only) shows less accuracy in both cognitive task recognition (0.863 vs. 0.995) and disease diagnosis (0.812 vs. 0.820) than our method (using both SC and FC).

6. Conclusions

In this work, we have ventured into developing a unique deep learning framework, *DeepHoloBrain*, which utilizes a scattering-transform mixer on the Riemannian manifold to delve into the complexities of neural dynamics. At its core, *DeepHoloBrain* is a physics-informed model that aims to unravel the intricate coupling mechanism between brain structure and function, viewed through the prism of data geometry. Our model innovatively combines: (1) A new mapping mechanism on the Riemannian manifold using scattering transforms; (2) A pioneering geometric deep learning structure, tailored for SPD matrices; (3) An insightful approach to understanding the human brain with deep learning; (4) A thorough evaluation against large-scale neuroimaging data, proving its effectiveness and practicality. Our extensive experimental results affirm the model's effectiveness and practicality, marking a promising pilot work in the field of neuroscience.

Software and Data

HCP-A data can be found in https://www.humanconnectome.org/study/hcp-lifespan-aging, ADNI can be downloaded in https://adni.loni.usc.edu/, OASIS can be downloaded in https://sites.wustl.edu/oasisbrains/home/oasis-3/.

Acknowledgements

This work was supported by the National Institutes of Health AG070701, AG073927, AG068399, and Foundation of Hope.

Impact Statement

This paper advances Machine Learning and Neuroscience by introducing novel concepts and methods. Our major contributions to the machine learning field are we propose (1) a new mapping mechanism on the Riemannian manifold, constrained by scattering transforms, and (2) a novel geometric deep model - Scattering-transform Mixer for the Riemannian manifold of SPD matrices. From an application standpoint, it offers (1) a novel approach to understanding the human brain via deep learning techniques, and (2) a comprehensive evaluation of these techniques' clinical impact on large-scale neuroimaging datasets. These advancements promise significant implications in both theoretical and prac-

tical domains of neural science and machine learning. Potential applications can be referred to the last paragraph of Section 3.3.

References

Bassett, D. S. and Sporns, O. Network neuroscience. *Nature Neuroscience*, 20(3):353–364, 2017.

Bedel, H. A., Sivgin, I., Dalmaz, O., Dar, S. U., and Çukur, T. Bolt: Fused window transformers for fmri time series analysis. *Medical Image Analysis*, 88:102841, 2023.

Bhatia, R. *Positive definite matrices*. Princeton University Press, 2009.

Bookheimer, S. Y., Salat, D. H., Terpstra, M., Ances, B. M., Barch, D. M., Buckner, R. L., Burgess, G. C., Curtiss, S. W., Diaz-Santos, M., Elam, J. S., et al. The lifespan human connectome project in aging: an overview. *Neuroimage*, 185:335–348, 2019.

Bressler, S. L. and Menon, V. Large-scale brain networks in cognition: emerging methods and principles. *Trends in Cognitive Sciences*, 14(6):277–290, 2010.

Bullmore, E. and Sporns, O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, 10(3):186–198, 2009.

Dan, T., Huang, Z., Cai, H., Laurienti, P. J., and Wu, G. Learning brain dynamics of evolving manifold functional mri data using geometric-attention neural network. *IEEE Transactions on Medical Imaging*, 41(10):2752–2763, 2022a.

Dan, T., Huang, Z., Cai, H., Lyday, R. G., Laurienti, P. J., and Wu, G. Uncovering shape signatures of resting-state functional connectivity by geometric deep learning on riemannian manifold. *Human Brain Mapping*, 43(13): 3970–3986, 2022b.

Gabor, D. A new microscopic principle. *Nature*, 161:777–778, 1948.

Gama, F., Ribeiro, A., and Bruna, J. Stability of graph scattering transforms. *Advances in Neural Information Processing Systems*, 32, 2019.

Gao, F., Wolf, G., and Hirn, M. Geometric scattering for graph data analysis. In *International Conference on Machine Learning*, pp. 2122–2131. PMLR, 2019.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C. F., Jenkinson, M., et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615): 171–178, 2016.

Greicius, M. D., Srivastava, G., Reiss, A. L., and Menon, V. Default-mode network activity distinguishes alzheimer's disease from healthy aging: evidence from functional mri. *Proceedings of the National Academy of Sciences*, 101 (13):4637–4642, 2004.

Hammond, D. K., Vandergheynst, P., and Gribonval, R. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011.

Hammond, D. K., Vandergheynst, P., and Gribonval, R. The spectral graph wavelet transform: Fundamental theory and fast computation. In *Vertex-Frequency Analysis of Graph Signals*, pp. 141–175. Springer, 2018.

Harandi, M. T., Sanderson, C., Hartley, R., and Lovell, B. C. Sparse coding and dictionary learning for symmetric positive definite matrices: A kernel approach. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part II 12*, pp. 216–229. Springer, 2012.

Harandi, M. T., Salzmann, M., and Hartley, R. From manifold to manifold: Geometry-aware dimensionality reduction for spd matrices. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part II 13*, pp. 17–32. Springer, 2014.

Huang, Z. and Van Gool, L. A riemannian network for spd matrix learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 31, 2017.

Huang, Z., Wang, R., Shan, S., Li, X., and Chen, X. Log-euclidean metric learning on symmetric positive definite manifold with application to image set classification. In *International Conference on Machine Learning*, pp. 720–729. PMLR, 2015.

Ionescu, C., Vantzos, O., and Sminchisescu, C. Matrix backpropagation for deep networks with structured layers. In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2965–2973, 2015.

Kawahara, J., Brown, C. J., Miller, S. P., Booth, B. G., Chau, V., Grunau, R. E., Zwicker, J. G., and Hamarneh, G. Brainnetcnn: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, 2017.

Kipf, T. N. and Welling, M. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.

Koch, W., Teipel, S., Mueller, S., Benninghoff, J., Wagner, M., Bokde, A. L., Hampel, H., Coates, U., Reiser, M., and Meindl, T. Diagnostic power of default mode network resting state fmri in the detection of alzheimer's disease. *Neurobiology of Aging*, 33(3):466–478, 2012.

LaMontagne, P. J., Benzinger, T. L., Morris, J. C., Keefe, S., Hornbeck, R., Xiong, C., Grant, E., Hassenstab, J., Moulder, K., Vlassenko, A. G., et al. Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. *MedRxiv*, pp. 2019–12, 2019.

Lee, J. S., Kim, J. H., and Lee, S.-K. The relationship between neuropsychiatric symptoms and default-mode network connectivity in alzheimer's disease. *Psychiatry Investigation*, 17(7):662, 2020.

Mallat, S. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, 2012.

Mevel, K., Chételat, G., Eustache, F., Desgranges, B., et al. The default mode network in healthy aging and alzheimer's disease. *International Journal of Alzheimer's Disease*, 2011, 2011.

Mollai, S. Recursive interferometric representations. In *2010 18th European Signal Processing Conference*, pp. 716–720. IEEE, 2010.

Pang, J. C., Aquino, K. M., Oldehinkel, M., Robinson, P. A., Fulcher, B. D., Breakspear, M., and Fornito, A. Geometric constraints on human brain function. *Nature*, pp. 1–9, 2023.

Pennec, X., Fillard, P., and Ayache, N. A riemannian framework for tensor computing. *International Journal of Computer Vision*, 66:41–66, 2006.

Perlmutter, M., Wolf, G., and Hirn, M. Geometric scattering on manifolds. *arXiv preprint arXiv:1812.06968*, 2018.

Perlmutter, M., Gao, F., Wolf, G., and Hirn, M. Geometric wavelet scattering networks on compact riemannian manifolds. In *Mathematical and Scientific Machine Learning*, pp. 570–604. PMLR, 2020.

Qiu, A., Lee, A., Tan, M., and Chung, M. K. Manifold learning on brain functional networks in aging. *Medical Image Analysis*, 20(1):52–60, 2015.

Richter, S., Winzeck, S., Correia, M. M., Kornaropoulos, E. N., Manktelow, A., Outtrim, J., Chatfield, D., Posti, J. P., Tenovuo, O., Williams, G. B., et al. Validation of cross-sectional and longitudinal combat harmonization methods for magnetic resonance imaging data on a travelling subject cohort. *Neuroimage: Reports*, 2(4):100136, 2022.

Sarwar, T., Tian, Y., Yeo, B. T., Ramamohanarao, K., and Zalesky, A. Structure-function coupling in the human connectome: A machine learning approach. *Neuroimage*, 226:117609, 2021.

Shi, Y., Huang, Z., Feng, S., Zhong, H., Wang, W., and Sun, Y. Masked label prediction: Unified message passing model for semi-supervised classification. *arXiv preprint arXiv:2009.03509*, 2020.

Shuman, D. I. Localized spectral graph filter frames: A unifying framework, survey of design considerations, and numerical comparison. *IEEE Signal Processing Magazine*, 37(6):43–63, 2020.

Shuman, D. I., Narang, S. K., Frossard, P., Ortega, A., and Vandergheynst, P. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Processing Magazine*, 30(3):83–98, 2013.

Terras, A. *Harmonic analysis on symmetric spaces and applications II*. Springer Science & Business Media, 2012.

Tiwari, S. and Konidaris, G. Effects of data geometry in early deep learning. *Advances in Neural Information Processing Systems*, 35:30099–30113, 2022.

Tolstikhin, I. O., Houlsby, N., Kolesnikov, A., Beyer, L., Zhai, X., Unterthiner, T., Yung, J., Steiner, A., Keysers, D., Uszkoreit, J., et al. Mlp-mixer: An all-mlp architecture for vision. *Advances in Neural Information Processing Systems*, 34:24261–24272, 2021.

Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., and Joliot, M. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*, 15(1): 273–289, 2002.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., and Bengio, Y. Graph attention networks. *arXiv preprint arXiv:1710.10903*, 2017.

Wang, R., Guo, H., Davis, L. S., and Dai, Q. Covariance discriminative learning: A natural and efficient approach to image set classification. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2496–2503. IEEE, 2012.

Watts, D. J. and Strogatz, S. H. Collective dynamics of 'small-world'networks. *Nature*, 393(6684):440–442, 1998.

Weiner, M. W., Veitch, D. P., Aisen, P. S., Beckett, L. A., Cairns, N. J., Cedarbaum, J., Donohue, M. C., Green,

R. C., Harvey, D., Jack Jr, C. R., et al. Impact of the alzheimer's disease neuroimaging initiative, 2004 to 2014. *Alzheimer's & Dementia*, 11(7):865–884, 2015.

Xu, K., Hu, W., Leskovec, J., and Jegelka, S. How powerful are graph neural networks? *arXiv preprint arXiv:1810.00826*, 2018.

You, K. and Park, H.-J. Re-visiting riemannian geometry of symmetric positive definite matrices for the analysis of functional connectivity. *Neuroimage*, 225:117464, 2021.

A. Appendix

A.1. Construction of SC and FC Networks

The data preprocessing involves the following steps to derive FC and SC matrices:

• **fMRI Data Processing**[1]**:**

▷ *Structural MRI (T1-weighted) preprocessing*, including brain extraction, tissue segmentation, spatial normalization, cost function masking, longitudinal processing, and brain mask refinement.

▷ *BOLD preprocessing*, encompassing reference image estimation, head-motion estimation, slice time correction, susceptibility distortion correction, EPI to T1w registration, resampling into standard spaces, sampling to Freesurfer surfaces, HCP Grayordinates, and confounds estimation.

▷ *Generation of BOLD time-series data for FC matrices construction.*

• **DWI Data Processing**[2]**:**

▷ *Initial processing*, including image and gradient orientation conformity, distortion grouping, denoising, distortion correction, head motion correction, and B0 template construction.

▷ *Reconstruction steps*, such as ODF/FOD estimation, anisotropy scalar computation, and tractography.

▷ *Estimation of structural connectomes to generate SC matrices.*

Figure 6 illustrates the process of constructing FC and SC networks.

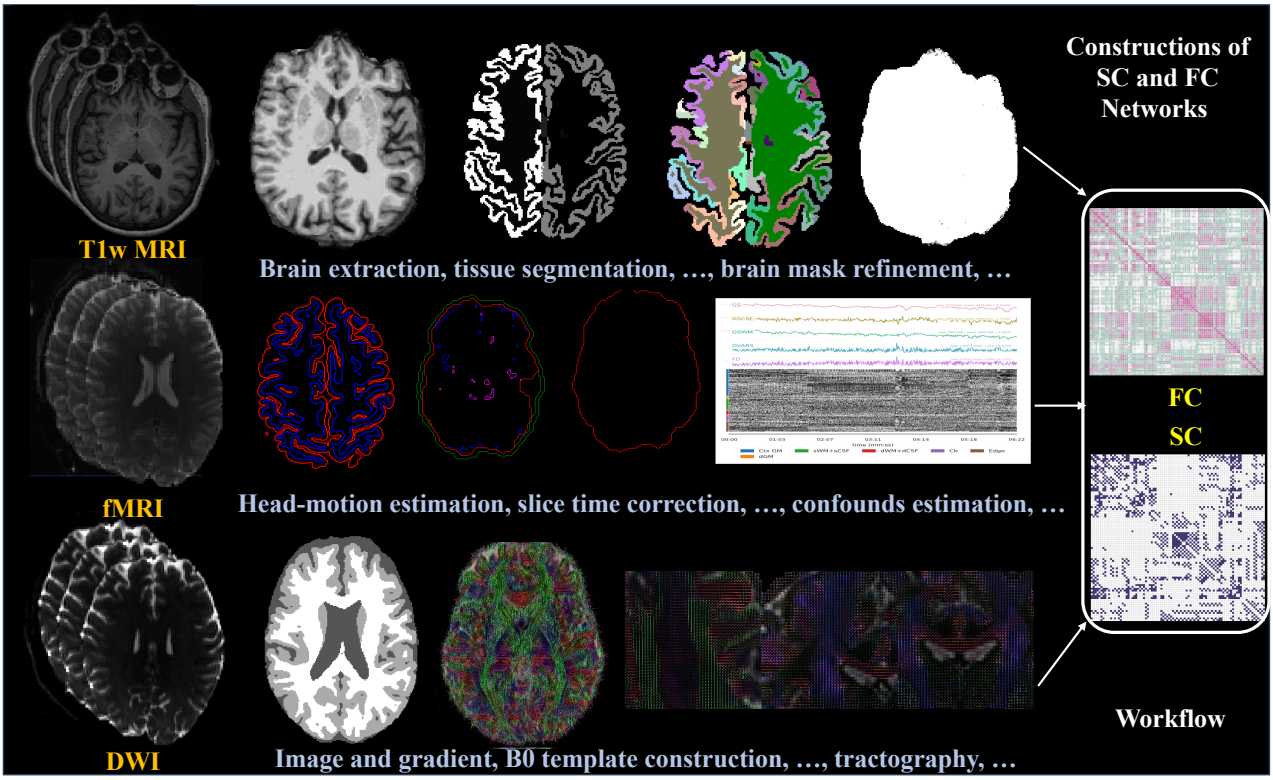

Figure 6. Workflow of constructing FC and SC.

[1] https://fmriprep.org/en/stable/
[2] https://qsiprep.readthedocs.io/en/latest/

A.2. Proof

Proposition A.1. *Given a full-rank matrix $W \in \mathbb{R}^{n\times m}$ and a vector $v \in \mathbb{R}^m$, since the columns of W are linearly independent and span $\mathbb{R}^m$, there exists, by the fundamental theorem of linear algebra, a vector $u \in \mathbb{R}^n$ such that $Wu = v$.*

Proof. Since W is full rank, its column space is a subspace of $\mathbb{R}^m$. If the rank of W is m, it means that its column space is actually the entire $\mathbb{R}^m$. Thus, every vector in $\mathbb{R}^m$ can be linearly represented by the columns of W, that is, for each $v \in \mathbb{R}^m$, there exist coefficients $c_1, c_2, \ldots, c_m$ such that $v = c_1w_1 + c_2w_2 + \ldots + c_mw_m$, where $w_1, w_2, \ldots, w_m$ are the columns of W. We can write this equation to matrix form as

$$v = W \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_m \end{bmatrix},$$

where the vector of coefficients is $u \in \mathbb{R}^n$ which satisfies $Wu = v$, proving the existence of such a vector u for every $v \in \mathbb{R}^m$. □

Proposition A.2. *Remark. Suppose we have a graph Laplacian matrix $\mathcal{L} = D - A$, where D is the degree matrix and A is the adjacency matrix, the first eigenvalue is typically $\lambda_0 = 0$, corresponding to the trivial eigenvector u_0. This is because the graph Laplacian represents the nodal connectivity, and the constant vector represents an equal state of all nodes, naturally belonging to the kernel of $\mathcal{L}$.*

The eigenvector u_0 is usually a vector with all equal elements. For an undirected graph, it is often chosen as $u_0 = \frac{1}{\sqrt{N}}(1, 1, \ldots, 1)^\intercal$, where N is the number of nodes in the graph. Thus, the first vector u_0 has all positive elements and this choice satisfies the definition of an eigenvector:

$$\mathcal{L}u_0 = (D - A)u_0 = 0 \cdot u_0 = 0$$

and the normalization condition of u_0 (i.e., $\|u_0\| = 1$), since each element is $\frac{1}{\sqrt{N}}$, the norm (length) of the vector is 1.

In this context, the value range harmonic wavelet is $[-\infty, \frac{1}{N}, +\infty]$, thus the lower band of $max(\cdot)$ operator is $\frac{1}{N} > 0$.

Proposition A.3. *Given a symmetric matrix A, $B = \exp(A)$ is a SPD matrix.*

Proof. To prove that $\exp(A)$ is positive definite, we use the fact that any symmetric matrix can be diagonalized. This means that there exists an orthogonal matrix Q and a diagonal matrix D such that $A = QDQ^\intercal$.

Then, we have:

$$\exp(A) = \exp(QDQ^\intercal) = Q\exp(D)Q^\intercal$$

where the exponential of a diagonal matrix is obtained by applying the exponential function to its diagonal elements.

Consider any non-zero vector v, we have:

$$v^\intercal \exp(A)v = v^\intercal Q\exp(D)Q^\intercal v$$

Let $y = Q^\intercal v$, then we get:

$$v^\intercal \exp(A)v = y^\intercal \exp(D)y$$

Since $\exp(D)$ is positive definite (its diagonal elements are the results of the exponential function, hence all positive), we have:

$$y^\intercal \exp(D)y > 0$$

unless y is a zero vector. However, since Q is orthogonal, $y = 0$ if and only if $v = 0$. Therefore, for all non-zero v, the expression is positive.

This proves that $B = \exp(A)$ is a positive definite matrix.

□

Proposition A.4. *Given a positive definite matrix A, then $B = (A + A^\intercal)/2$ is a SPD matrix.*

Proof. **1: Symmetry**: The symmetry of B can be shown quite straightforwardly. Given A, a positive definite matrix, let $A^\intercal$ denotes its transpose. The matrix B is defined as $B = \frac{A+A^\intercal}{2}$. Since the transpose of a sum of matrices is the sum of their transposes, we have $B^\intercal = \frac{A^\intercal+(A^\intercal)^\intercal}{2}$. Since $(A^\intercal)^\intercal = A$, it follows that $B^\intercal = B$. Hence, B is symmetric.

2: Positive Definiteness: For any non-zero vector v, consider the quadratic form $v^\intercal Bv$:

$$v^\intercal Bv = v^\intercal \left(\frac{A + A^\intercal}{2}\right) v$$

$$= \frac{1}{2}v^\intercal Av + \frac{1}{2}x^\intercal A^\intercal v$$

Since A is positive definite, $v^\intercal Av > 0$ for all non-zero v. Similarly, $v^\intercal A^\intercal v = (Av)^\intercal v = v^\intercal Av > 0$ as A is positive definite. Obviously, $v^\intercal Bv = (\frac{1}{2}v^\intercal Av + \frac{1}{2}v^\intercal A^\intercal v) > 0$. Hence, B is positive definite.

□