Cortically-Resolved Recurrent Architecture for fMRI

Da-Woon Heo
Department of Artificial Intelligence
Korea University
Seoul, South Korea
daheo@korea.ac.kr

Kwanseok Oh
Department of Artificial Intelligence
Korea University
Seoul, South Korea
ksohh@korea.ac.kr

Adrián Ayuso-Muñoz
Department of Artificial Intelligence
Korea University
Seoul, South Korea
amadrian@korea.ac.kr

Heung-Il Suk*
Department of Artificial Intelligence
Korea University
Seoul, South Korea
hisuk@korea.ac.kr

Abstract

Decoding cognitive functions and brain disorders from functional magnetic resonance imaging (fMRI) has benefited from deep learning, including fMRI-specific decoders and neuroscience-inspired sequence models. Cortical regions are functionally specialized, integrate information over heterogeneous timescales, and interact through time-varying coupling. We propose CoRTeX, a cortically-resolved recurrent architecture that reflects these properties in its learnable structure. Each region maintains its own recurrent state and learns a temporal integration window, while a memory with region-specific learning and forgetting rates accumulates co-activation among regional trajectories and biases identity-aware attention. We evaluate CoRTeX on task-fMRI stimulus classification (NSD) and resting-state fMRI disease diagnosis (ABIDE and COBRE). CoRTeX achieves competitive performance against strong baselines, including the highest area under the curve on NSD and COBRE. Further analyses show that the regional assignments of learned temporal integration windows influence predictive performance, while class-specific attributions on NSD are consistent with known category-selective regions. Code is available at https://github.com/ku-milab/CoRTeX.

1 Introduction

Decoding cognitive functions and brain diseases from cortical activity is a foundational question in neuroscience, and functional magnetic resonance imaging (fMRI) has played a central role in this effort. Recent advances in deep learning have substantially improved fMRI decoding for tasks such as image reconstruction [Scotti et al., 2024, Takagi and Nishimoto, 2023, Scotti et al., 2023] and brain disease diagnosis [Kawahara et al., 2017, Kim et al., 2021, Bedel et al., 2023].

Deep learning has long drawn on abstractions of brain computation, with sequence models instantiating different aspects of neural processing: RNNs [Elman, 1990] and LSTMs [Hochreiter and Schmidhuber, 1997] abstract working memory and gating; state-space models such as Mamba [Gu and Dao, 2023] capture long-range temporal dependencies via selective state updates; memory-augmented architectures such as Titans [Behrouz et al., 2025] introduce dynamic parameter updates inspired by synaptic plasticity; and the Continuous Thought Machine (CTM) [Darlow et al., 2025] treats neural synchrony as a core computational principle.

*Corresponding author

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

While these approaches abstract general principles of brain computation, comparatively little work has tailored such abstractions to the specific properties of fMRI signals. fMRI signals reflect activity from anatomically distinct cortical regions, each with characteristically different functional roles and temporal integration windows that span seconds to minutes along the cortical hierarchy [Hasson et al., 2008, Honey et al., 2012, Raut et al., 2020]. The hemodynamic response itself also varies across regions and individuals [Handwerker et al., 2004], so regional temporal differences in blood-oxygen-level-dependent (BOLD) signals reflect both neural and hemodynamic sources. Yet neither general-purpose sequence models nor existing fMRI-specific decoders capture this heterogeneity: the former apply the same temporal operator uniformly across all regions [Gu and Dao, 2023, Behrouz et al., 2025], while the latter either collapse the temporal axis into static connectivity before the predictive model [Kawahara et al., 2017] or retain it with a single temporal granularity shared across regions [Kim et al., 2021, Bedel et al., 2023]. As a result, these models do not explicitly expose *which* regions integrate information over *what* timescales, or how inter-regional coupling evolves together with these regional dynamics. This motivates a region-resolved recurrent architecture in which regional identity, timescale, and connectivity are jointly modeled rather than imposed or post-processed.

We address this gap by proposing CoRTeX (Cortical Recurrent Trajectory eXchange), a region-resolved recurrent architecture that processes each cortical region according to its own learned characteristics. Each design choice reflects a property of cortical organization documented in neuroscience, with these properties jointly modeled within a single recurrent computation: (i) *anatomical identity preservation*—each region carries its own per-region parameters and cross-region information is routed through a single identity-aware attention (IAA) channel, reflecting the functional specialization of cortical regions [Kanwisher, 2010]; (ii) *regional temporal heterogeneity*—each region learns its own effective integration window, motivated by the documented hierarchy of intrinsic timescales across cortex [Hasson et al., 2008, Honey et al., 2012, Raut et al., 2020]; and (iii) *adaptive inter-regional coupling*—a recurrent functional connectivity (RFC) computed from the evolving regional trajectories is accumulated within the recurrent loop by a co-activation-driven memory [Ba et al., 2016, Hebb, 2005] indexed by anatomical identity. Given a region-of-interest (ROI)-level BOLD time series, CoRTeX jointly evolves region-specific trajectories and the RFC at every time step.

We evaluate CoRTeX on three fMRI datasets spanning two task families: task-fMRI classification on the Natural Scenes Dataset (NSD) [Allen et al., 2022], and resting-state fMRI (rs-fMRI) disease diagnosis on Autism Brain Imaging Data Exchange (ABIDE) [Di Martino et al., 2014] and Center for Biomedical Research Excellence (COBRE) [Aine et al., 2017]. CoRTeX performs comparably to strong baselines while directly exposing learned ROI-specific integration windows and enabling functional evaluation of their regional assignments.

2 Related Work

Classification models for fMRI. fMRI-based classification has primarily been built on three input representations: single-trial β estimates for task-fMRI [Mahmoudi et al., 2012], Pearson functional connectivity (FC) matrices for rs-fMRI [van den Heuvel and Hulshoff Pol, 2010, Friston, 2011], and the raw BOLD time series. BrainNetCNN [Kawahara et al., 2017] introduces edge-to-edge, edge-to-node, and node-to-graph convolutions tailored to the connectivity matrix structure. However, both β- and FC-based features are static descriptors that collapse the temporal axis of the underlying BOLD signal. To address this, several recent methods model the BOLD signal more directly. STAGIN [Kim et al., 2021] processes sliding-window dynamic FC with a graph neural network, and BolT [Bedel et al., 2023] applies fused window attention to the multivariate BOLD signal. Crucially, none of these methods learns region-specific timescales or carries inter-regional coupling forward as a learned memory within the model. CoRTeX explicitly addresses both.

Recurrent and neuroscience-motivated sequence models. Standard backbones for temporal modeling include RNNs [Elman, 1990], LSTMs [Hochreiter and Schmidhuber, 1997], and selective state-space models such as Mamba [Gu and Dao, 2023]. More recent work incorporates neuroscience-motivated mechanisms. Titans [Behrouz et al., 2025] augments attention with a memory module that updates online based on input "surprise," and CTM [Darlow et al., 2025] assigns each artificial neuron its own private temporal processing unit, using population-level synchronization as its downstream representation. These architectures are designed for general sequence modeling, so their units—

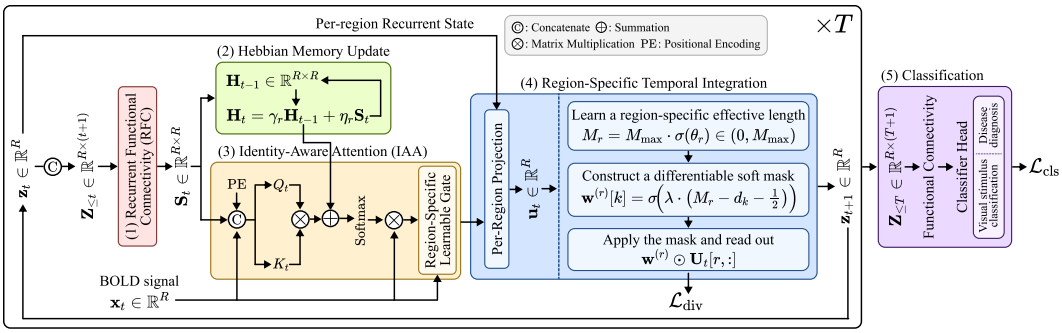


Figure 1: **Overview of CoRTeX.** Over T recurrent steps, CoRTeX evolves per-region states $\mathbf{z}_t$ and a Hebbian memory $\mathbf{H}_t$ from the BOLD input $\mathbf{x}_t$. **(1)** *RFC* computes inter-regional coupling $\mathbf{S}_t$ from the regional trajectories. **(2)** The *Hebbian memory* accumulates $\mathbf{S}_t$ with region-wise forgetting and learning rates (γ_r, η_r) and biases **(3)** *IAA*, which uses NeRF coordinates and one-hot identity as positional encoding. The gated cross-region input is combined with $\mathbf{z}_t$ by a per-region projection, and **(4)** *region-specific temporal integration* reads it out over a learned window M_r. After T steps, the final RFC $\mathbf{S}_T$ is classified by **(5)** the classifier head.

sequence tokens, artificial neurons, or memory slots—are exchangeable by design and are not tied to brain regions. Specifically, Titans adapts its learning and forgetting rates to the input rather than to individual regions, and CTM's per-neuron temporal units are not defined with respect to brain anatomy. CoRTeX, in contrast, anchors its learnable components to anatomically defined cortical regions, learning region-specific dynamics from data (see Appendix A for a detailed comparison).

3 Method

We model inter-regional coupling as an RFC matrix $\mathbf{S}_t$ that summarizes the temporal co-activation of regional trajectories up to time t, extending dynamic FC [Hutchison et al., 2013, Calhoun et al., 2014] with an expanding rather than sliding window.

3.1 Overview

As illustrated in Fig. 1, CoRTeX performs recurrent processing at the level of anatomically defined cortical regions. Unlike standard pipelines that treat FC as a static post-hoc descriptor of observed BOLD, CoRTeX introduces an RFC computed from the model's evolving regional states inside the recurrent loop.

Given a BOLD time series $\mathbf{X} \in \mathbb{R}^{R\times T}$, where R is the number of ROIs and T is the number of time points (volumes), CoRTeX unfolds T recurrent steps, one per time point. At step $t \in \{0, \dots, T-1\}$, the loop receives the current BOLD input $\mathbf{x}_t := \mathbf{X}[:, t] \in \mathbb{R}^R$, the regional state $\mathbf{z}_t \in \mathbb{R}^R$, and the previous Hebbian memory $\mathbf{H}_{t-1} \in \mathbb{R}^{R\times R}$. From these, it computes a recurrent FC matrix $\mathbf{S}_t \in \mathbb{R}^{R\times R}$ (§3.2) and updates the Hebbian memory to $\mathbf{H}_t$ (§3.3). Cross-region information is exchanged through a single IAA channel (§3.4), and per-region representations are updated via region-specific projection and temporal integration (§3.5), emitting $\mathbf{z}_{t+1}$. After T steps, the final FC matrix $\mathbf{S}_T$ is passed to a downstream classifier (§3.6).

3.2 Recurrent Functional Connectivity

Let $\mathbf{z}_0 \in \mathbb{R}^R$ be a learnable initial state, initialized from $\mathcal{N}(0, \sigma^2_{\text{init}})$. The subsequent states $\mathbf{z}_1, \dots, \mathbf{z}_T$ are produced by the components introduced in §3.5, with step $t \in \{0, \dots, T-1\}$ taking $\mathbf{z}_t$ as input and emitting $\mathbf{z}_{t+1}$. At the start of step t, only the partial trajectory $\mathbf{Z}_{\le t} := [\mathbf{z}_0, \dots, \mathbf{z}_t] \in \mathbb{R}^{R\times(t+1)}$ is available. We standardize each row of $\mathbf{Z}_{\le t}$ along the temporal axis to zero mean and approximately unit variance, yielding $\widetilde{\mathbf{Z}}_{\le t}$. The RFC matrix is then proportional to the empirical Pearson correlation

between regional trajectories:

$$\mathbf{S}_t[r,s] = \frac{1}{t+1}\sum_{\tau=0}^{t} \widetilde{\mathbf{Z}}_{\leq t}[r,\tau]\cdot\widetilde{\mathbf{Z}}_{\leq t}[s,\tau], \quad r,s\in\{1,\ldots,R\}. \tag{1}$$

Since standardization requires non-zero temporal variance, Eq. (1) is not well-defined at $t=0$, where only $\mathbf{z}_0$ is available. We therefore use the raw outer product $\mathbf{S}_0[r,s]=\mathbf{z}_0[r]\cdot\mathbf{z}_0[s]$ at this step, with the standardized form taking over from $t=1$ onward. The resulting RFC matrix $\mathbf{S}_t$ feeds into both the Hebbian memory update (§3.3) and the IAA module (§3.4) within the recurrent loop.

3.3 Hebbian-Inspired Plasticity Module

The Hebbian module provides a learnable representation of within-sequence changes in inter-regional coupling that static FC cannot resolve [Liegeois et al., 2017, Vidaurre et al., 2017], drawing inspiration from the Hebbian principle [Hebb, 2005] that co-activated units strengthen their coupling. We instantiate this in the broad sense established by the fast-weights literature [Ba et al., 2016], as a co-activation-driven memory update modulated by a forgetting term rather than as a mechanistic claim about synaptic-level dynamics. Since $\mathbf{S}_t$ summarizes pairwise co-activation between regional trajectories (Eq. (1)), our update accumulates co-activation history into a memory matrix indexed by anatomical identity. The key departure from prior fast-weights formulations [Ba et al., 2016] and memory-augmented architectures [Behrouz et al., 2025] lies in our learning and forgetting rates $\{\eta_r,\gamma_r\}$. They are *region-specific*, rather than shared scalars as in fast weights or input-adaptive rates without anatomical assignment as in Titans, allowing each cortical region to encode its co-activation history under its own update profile.

The Hebbian memory in CoRTeX, $\mathbf{H}_t\in\mathbb{R}^{R\times R}$, accumulates the RFC matrices $\mathbf{S}_t$ across recurrent steps via region-wise learnable forgetting rate $\gamma_r\in(0,1)$ and learning rate $\eta_r>0$ (sigmoid- and softplus-parameterized, respectively). The Hebbian memory is reset to zero at the start of each sequence. At each step $t\geq 0$, it is updated from its previous state and the current $\mathbf{S}_t$:

$$\mathbf{H}_t[r,s] = \mathrm{clip}_{[-c,c]}\Big(\gamma_r\cdot\mathbf{H}_{t-1}[r,s] + \eta_r\cdot\mathbf{S}_t[r,s]\Big), \tag{2}$$

where c is a fixed clipping threshold. This decay-plus-accumulation form follows the fast-weights formulation [Ba et al., 2016]. As the rates η_r and γ_r depend only on the row index r, each row $\mathbf{H}_t[r,:]$ is region r's private record of co-activation with every other region, evolving under its own plasticity parameters without mixing information across rows. A single learnable global scalar α then controls the overall strength with which $\mathbf{H}_t$ enters the attention logits as a bias (§3.4).

3.4 Identity-Aware Attention

The IAA module is the sole mechanism through which information is exchanged between cortical regions in CoRTeX, while keeping each region's output anchored to its own anatomical identity. It is conceptually analogous to *effective connectivity* [Friston, 2011], the directed influence between brain regions. For each region r, we form a feature vector by concatenating the RFC row, position information, and the current BOLD input $\mathbf{x}_t[r]$:

$$\mathbf{f}_t^{(r)} = \big[\,\widehat{\mathbf{S}}_t[r,:]\,;\,\mathbf{e}_r\,;\,\boldsymbol{\gamma}(\mathbf{p}_r)\,;\,\mathbf{x}_t[r]\,\big]\in\mathbb{R}^{d_{\mathrm{feat}}},$$

where $\widehat{\mathbf{S}}_t=\mathrm{LayerNorm}(\mathbf{S}_t)$. Position information combines a one-hot indicator $\mathbf{e}_r\in\{0,1\}^R$, which preserves discrete regional identity, and a neural radiance fields (NeRF)-style sinusoidal encoding $\boldsymbol{\gamma}(\mathbf{p}_r)\in\mathbb{R}^{6L}$ [Mildenhall et al., 2021] of the anatomical centroid $\mathbf{p}_r\in\mathbb{R}^3$ of region r, which encodes its anatomical position (details in Appendix D.1), giving $d_{\mathrm{feat}}=2R+6L+1$.

Stacking the per-region feature vectors into $\mathbf{F}_t\in\mathbb{R}^{R\times d_{\mathrm{feat}}}$, we project to queries $\mathbf{Q}_t=\mathbf{F}_t\mathbf{W}_Q$ and keys $\mathbf{K}_t=\mathbf{F}_t\mathbf{W}_K$ with $\mathbf{W}_Q,\mathbf{W}_K\in\mathbb{R}^{d_{\mathrm{feat}}\times d_{\mathrm{attn}}}$. The attention logits combine the scaled query-key score with the Hebbian bias from §3.3, and a row-wise softmax yields the attention weights:

$$\mathbf{A}_t[r,:] = \mathrm{Softmax}\left(\frac{\mathbf{Q}_t[r]\,\mathbf{K}_t^\top}{\sqrt{d_{\mathrm{attn}}}} + \alpha\cdot\mathbf{H}_t[r,:]\right), \qquad r\in\{1,\ldots,R\}, \tag{3}$$

where $\alpha>0$ is a learnable scalar controlling the Hebbian memory strength. The cross-regional context for region r is the attention-weighted sum of raw BOLD values, $\mathbf{c}_t[r]=\sum_s\mathbf{A}_t[r,s]\,\mathbf{x}_t[s]$. A

region-specific learnable gate $g_r \in (0, 1)$ then balances local input against this context as $\widetilde{\mathbf{x}}_t[r] = g_r \cdot \mathbf{x}_t[r] + (1 - g_r) \cdot \mathbf{c}_t[r]$. Values of g_r close to 1 reflect regions dominated by local input, while values close to 0 reflect regions integrating broadly across the cortex. The gated output $\widetilde{\mathbf{x}}_t \in \mathbb{R}^R$ is passed to the region-specific activation trajectory (§3.5).

3.5 Region-Specific Activation Trajectory

Cortical regions integrate information over different timescales, with systematic variation along the cortical hierarchy [Murray et al., 2014, Hasson et al., 2008, Honey et al., 2012]. Motivated by this, CoRTeX learns two region-specific operations on top of the gated output $\widetilde{\mathbf{x}}_t$ from §3.4, namely a per-region projection that combines the current input with the previous regional state, and a per-region temporal integration window over a first-in-first-out (FIFO) buffer.

Per-region projection. At each time t, each region r combines its previous state $\mathbf{z}_t[r]$ with the gated attention output $\widetilde{\mathbf{x}}_t[r]$ through a region-specific two-layer MLP f_r, yielding a per-region update

$$\mathbf{u}_t[r] = f_r(\mathbf{z}_t[r], \widetilde{\mathbf{x}}_t[r]), \tag{4}$$

where f_r has parameters private to region r and shared across time (refer to Appendix D.2).

Region-specific temporal integration. Each region maintains a fixed-length buffer $\mathbf{U}_t[r, :] \in \mathbb{R}^{M_{\max}}$, initialized to a learnable state and updated by a FIFO shift that appends $\mathbf{u}_t[r]$ at the newest position. Each region also learns its own *effective* window length:

$$M_r = M_{\max} \cdot \mathrm{Sigmoid}(\theta_r) \in (0, M_{\max}), \qquad \theta_r \in \mathbb{R}, \tag{5}$$

where θ_r is a learnable parameter and $M_{\max}$ is the buffer length. A small M_r means that a region relies mainly on its most recent updates (short timescale), whereas a large M_r means that it integrates over a longer history (long timescale).

Since M_r is learned from BOLD signals under the classification objective, we interpret it as a task-conditioned integration window in BOLD space, which may also reflect regional differences in hemodynamic response [Handwerker et al., 2004], rather than an intrinsic neural timescale. From M_r, we construct a soft mask over buffer positions:

$$\mathbf{w}^{(r)}[k] = \mathrm{Sigmoid}\Big(\lambda \cdot \big(M_r - d_k - \tfrac{1}{2}\big)\Big), \qquad d_k = M_{\max} - 1 - k, \tag{6}$$

where d_k is the distance of position k from the newest entry and $\lambda > 0$ is a fixed sharpness hyperparameter. Concretely, $\mathbf{w}^{(r)}[k] \approx 1$ for the $\lceil M_r - \frac{1}{2} \rceil$ most recent positions and ≈ 0 elsewhere, with a narrow soft transition at the boundary whose width shrinks as λ grows. For example, with $M_{\max} = 5$ and $M_r = 2$, the mask over buffer positions (oldest to newest) is approximately $[0.00, 0.00, 0.01, 0.99, 1.00]$ at $\lambda = 10$, so only the two most recent entries effectively contribute. Although the mask approaches a hard window as $\lambda \to \infty$, we keep λ finite so that the soft boundary makes M_r differentiable and learnable by gradient descent. We then apply this mask element-wise to the buffer, yielding the masked buffer $\mathbf{w}^{(r)} \odot \mathbf{U}_t[r, :] \in \mathbb{R}^{M_{\max}}$, which is passed through a region-specific readout to produce the next regional state $\mathbf{z}_{t+1}[r]$ (refer to Appendix D.3).

Diversity regularization. To prevent the learned windows from collapsing to a single common value, we add an auxiliary regularization term that promotes heterogeneity across regions by maximizing the variance of the learned windows:

$$\mathcal{L}_{\mathrm{div}} = -\alpha_{\mathrm{div}} \cdot \mathrm{Var}_r(M_r), \tag{7}$$

where $\alpha_{\mathrm{div}} \geq 0$ is the regularization weight, and the negative sign reflects that the term enters the total loss as a penalty whose minimization corresponds to maximizing variance.

3.6 Classification and Training Objective

After T recurrent steps, CoRTeX has produced the full trajectory $\mathbf{Z}_{\leq T} = [\mathbf{z}_0, \ldots, \mathbf{z}_T] \in \mathbb{R}^{R \times (T+1)}$. We summarize this trajectory into a final RFC matrix $\mathbf{S}_T \in \mathbb{R}^{R \times \bar{R}}$ by applying Eq. (1) to the full trajectory rather than its partial prefix, capturing how each pair of cortical regions co-activates over

the full course of recurrent inference. $\mathbf{S}_T$ is subsequently passed to a downstream classifier that produces a class probability $\bar{\mathbf{y}} \in \mathbb{R}^C$ over C categories specified per dataset (§4.1; Appendix D.4).

Given the ground-truth label $\mathbf{y}$, we train CoRTeX end-to-end by minimizing the classification loss with the diversity regularizer of Eq. (7):

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{cls}}(\bar{\mathbf{y}}, \mathbf{y}) + \mathcal{L}_{\text{div}}. \tag{8}$$

$\mathcal{L}_{\text{cls}}$ is a cross-entropy loss, taking the categorical form for single-label classification (ABIDE, COBRE) and a class-balanced binary form for multi-label classification (NSD; 12 non-mutually-exclusive COCO supercategories), with full definitions in Appendix D.5. The training setup and hyperparameters are reported in §4.2 and Appendix C.

4 Experiments

4.1 Datasets

Task-fMRI. We use the NSD [Allen et al., 2022], which provides fMRI scans from 8 subjects viewing natural images from MS-COCO [Lin et al., 2014], parcellated into 360 ROIs using the subject-native HCP multi-modal parcellation (HCP-MMP1; Glasser et al., 2016). We use the 12 MS-COCO supercategories (*e.g.,* animal, food, person, vehicle) as labels for multi-label classification. The BOLD window length is set to $T = 6$ volumes (≈ 8.0 s after stimulus onset), chosen to cover the canonical hemodynamic response function (HRF) peak and a brief post-peak interval [Glover, 1999, Friston et al., 1998]. Further details on this choice, along with data descriptions, preprocessing, and splits, are provided in Appendix B.1.

rs-fMRI. We use two public datasets: ABIDE [Di Martino et al., 2014] (1,072 subjects; 511 autism spectrum disorder (ASD) / 561 typically developing (TD)) and COBRE [Aine et al., 2017] (145 subjects; 71 schizophrenia (SZ) / 74 healthy control (HC)), both parcellated into 200 ROIs using the Schaefer atlas [Schaefer et al., 2018], with its 17-network labels grouped into seven networks. We perform binary classification (ASD vs. TD; SZ vs. HC). Since ABIDE aggregates scans across 17 sites with heterogeneous repetition times (TRs), we resample all time series to a common TR of 2 s via Fourier interpolation and set $T = 98$, the shortest resulting length across subjects. For COBRE, all subjects share 150 volumes, so we use $T = 150$. Data descriptions, preprocessing, and splits are provided in Appendix B.

4.2 Experimental Setup

Implementation details and training configurations. We use the same architectural settings across datasets except for the number of ROIs (R), sequence length (T), maximum integration window ($M_{\max}$), anatomical coordinate space, and classifier head. Models are selected by validation area under the curve (AUC) on the splits described in Appendix B. Experiments were run on PyTorch 2.6.0 / CUDA 12.4 with a single NVIDIA RTX 2080 Ti GPU (11 GB). Full hyperparameters and training protocols are listed in Appendix C.

Baselines. We compare CoRTeX with three categories of baselines. *(1) Classical methods*: Ridge classifier, logistic regression (LogReg), and kernel logistic regression (Kernel-LR) are applied to regional BOLD signals on all datasets, and additionally to dataset-specific features, namely β estimates on NSD and static FC on ABIDE and COBRE. *(2) fMRI-specific deep learning models*: BrainNetCNN, STAGIN, and BolT, applied to ABIDE and COBRE. *(3) Sequence models*: RNN, LSTM, Mamba, Titans, and CTM, applied to all datasets. On NSD, we additionally include MLP baselines on flattened BOLD or FC inputs, where the FC-based MLP serves as a reference despite potentially unreliable Pearson correlation estimates from only $T = 6$ time points [Hutchison et al., 2013, Leonardi and Van De Ville, 2015]. We do not apply BrainNetCNN, STAGIN, and BolT to NSD, as the latter two rely on sliding-window or windowed-attention mechanisms that require additional adaptation to this short temporal setting, and β baselines are not applied to rs-fMRI, as β estimates require task events. All baselines are trained, validated, and tested on identical data splits, with hyperparameters either selected on the validation set or set to the official defaults (Appendix C).

Evaluation metrics. For multi-label classification on NSD, we report macro- and micro-AUC, mean average precision (mAP), macro-F1, and Hamming loss, capturing ranking- and threshold-based performance [Tsoumakas et al., 2010]. For binary classification on ABIDE and COBRE, we report

Table 1: **Multi-label classification on NSD** (mean±std over 8 subjects × 5 seeds).

Type	Method	Macro-F1 (↑)	Macro-AUC (↑)	Micro-AUC (↑)	mAP (↑)	Hamming Loss (↓)	PFC
FC	MLP	0.207±0.041**	0.596±0.039**	0.687±0.032**	0.251±0.034**	0.205±0.013**	8.9×10^{-10}
β	RidgeClassifier	0.215±0.066**	0.847±0.027	0.867±0.022*	**0.560±0.051**	0.135±0.011†	0.322
	LogReg	0.459±0.046**	0.836±0.028**	0.848±0.023**	0.535±0.052	**0.134±0.012**†	0.004
	Kernel-LR	**0.527±0.036**	0.835±0.029**	0.868±0.021**	0.536±0.047	0.190±0.025**	0.003
BOLD	RidgeClassifier	0.497±0.035**	0.820±0.026**	0.853±0.019**	0.505±0.044**	0.194±0.022**	8.1×10^{-5}
	LogReg	0.496±0.038**	0.825±0.025**	0.860±0.019**	0.507±0.044**	0.188±0.026**	8.7×10^{-5}
	Kernel-LR	0.384±0.035**	0.720±0.033**	0.783±0.023**	0.355±0.040**	0.398±0.077**	1.1×10^{-6}
	MLP	0.470±0.040**	0.827±0.025**	0.859±0.020**	0.506±0.039**	0.299±0.056**	8.0×10^{-4}
	RNN	0.313±0.050**	0.809±0.027**	0.845±0.020**	0.475±0.044**	0.137±0.009†	4.0×10^{-7}
	LSTM	0.293±0.060**	0.799±0.032**	0.835±0.023**	0.460±0.051**	0.141±0.010	2.8×10^{-6}
	Mamba	0.301±0.065**	0.801±0.033**	0.838±0.024**	0.463±0.050**	0.140±0.009	3.9×10^{-6}
	Titans	0.301±0.056**	0.792±0.030**	0.831±0.021**	0.452±0.047**	0.141±0.009	8.2×10^{-7}
	CTM	0.387±0.087**	0.796±0.036**	0.840±0.025**	0.437±0.065**	0.170±0.016**	6.6×10^{-5}
	CoRTeX (Ours)	0.522±0.039	**0.849±0.023**	**0.877±0.018**	0.546±0.041	0.142±0.010	–

Best in **bold**, second-best underlined. Paired Wilcoxon test (over 8 subjects, seed-averaged) against CoRTeX: $^*p < 0.05$, $^{**}p < 0.01$. †: significantly better baseline performance. PFC: probability of false claims based on macro-AUC; lower values indicate stronger evidence for CoRTeX outperforming the baseline.

Table 2: **Binary classification on ABIDE and COBRE** (mean±std across 5 folds).

Type	Method	ABIDE (ASD vs. TD)					COBRE (SZ vs. HC)				
		AUC	ACC	SEN	SPEC	PFC	AUC	ACC	SEN	SPEC	PFC
FC	RidgeClassifier	0.722±0.031	0.662±0.031	0.647±0.068	0.677±0.044	0.300	0.749±0.096*	0.668±0.072	0.632±0.158	0.704±0.098	0.115
	LogReg	0.729±0.031	0.669±0.028	0.648±0.041	0.689±0.070	0.464	0.752±0.097*	0.660±0.125	0.631±0.225	0.690±0.165	0.089
	Kernel-LR	**0.740±0.024**	**0.673±0.013**	0.632±0.045	0.710±0.047	0.161	0.757±0.096**	0.599±0.103	0.275±0.303	0.904±0.105	0.139
	BrainNetCNN	0.713±0.029*	0.646±0.041	0.517±0.156	**0.765±0.106**	0.079	0.721±0.099**	0.641±0.106	0.622±0.351	0.666±0.193	0.059
	STAGIN	0.634±0.028***	0.571±0.060	0.544±0.324	0.597±0.364	0.001	0.596±0.051***	0.537±0.045	0.365±0.293	0.706±0.213	0.018
BOLD	RidgeClassifier	0.498±0.021***	0.504±0.031	0.454±0.049	0.549±0.021	1.6×10^{-5}	0.458±0.185***	0.532±0.080	0.481±0.182	0.580±0.093	0.022
	LogReg	0.497±0.026***	0.499±0.027	0.482±0.049	0.514±0.034	4.5×10^{-5}	0.459±0.196***	0.505±0.122	0.480±0.181	0.528±0.122	0.023
	Kernel-LR	0.486±0.033***	0.492±0.034	0.370±0.211	0.603±0.236	1.3×10^{-4}	0.460±0.192***	0.503±0.014	0.071±0.160	**0.920±0.179**	0.022
	RNN	0.523±0.039***	0.520±0.020	0.547±0.039	0.495±0.037	5.7×10^{-4}	0.555±0.111***	0.530±0.080	0.334±0.136	0.715±0.144	0.008
	LSTM	0.669±0.027***	0.613±0.043	0.605±0.099	0.620±0.129	0.016	0.582±0.166***	0.607±0.120	0.714±0.286	0.503±0.285	0.037
	Mamba	0.717±0.025*	0.649±0.022	**0.726±0.196**	0.580±0.204	0.200	0.696±0.118**	0.635±0.038	0.582±0.312	0.690±0.225	0.014
	Titans	0.681±0.017***	0.623±0.029	0.593±0.131	0.651±0.163	0.001	0.642±0.124***	0.614±0.118	0.497±0.213	0.729±0.110	0.055
	CTM	0.698±0.031**	0.647±0.026	0.548±0.154	0.738±0.131	0.027	0.565±0.097***	0.552±0.084	0.343±0.421	0.759±0.428	0.016
	BolT	0.729±0.023	0.656±0.030	0.602±0.259	0.705±0.192	0.464	0.727±0.096**	0.723±0.074	0.705±0.221	0.745±0.157	0.101
	CoRTeX (Ours)	0.730±0.015	0.672±0.023	0.672±0.137	0.673±0.087	–	**0.797±0.132**	**0.786±0.063**	**0.802±0.147**	0.772±0.145	–

Best in **bold**, second-best underlined. Paired DeLong's test against CoRTeX (AUC): $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. PFC: probability of false claims based on AUC; lower values indicate stronger evidence for CoRTeX outperforming the baseline.

AUC, accuracy (ACC), sensitivity (SEN), and specificity (SPEC) [Kawahara et al., 2017, Kim et al., 2021, Bedel et al., 2023]. We assess differences against CoRTeX using paired Wilcoxon signed-rank tests for all metrics on NSD and paired DeLong tests [DeLong et al., 1988] for AUC on ABIDE and COBRE. We further report the probability of false claims (PFC) [Christodoulou et al., 2025] to assess uncertainty in AUC-based outperformance claims.

4.3 Classification Performance

CoRTeX demonstrates competitive performance across all three datasets. On NSD, CoRTeX outperforms all BOLD-based baselines in macro-F1, macro-/micro-AUC, and mAP, achieving the highest AUC scores among all methods, although the margin over β-RidgeClassifier is within statistical parity (PFC $= 0.322$). It also performs comparably to general linear model (GLM)-derived β-value baselines, which summarize each trial into static regional activation estimates, limiting the characterization of region-specific temporal dynamics. For ABIDE, CoRTeX achieves the highest AUC among BOLD-based methods (0.730 ± 0.015), approaching the best FC-based baseline, Kernel-LR (0.740 ± 0.024). Paired DeLong tests are significant against all baselines except the FC-based RidgeClassifier, LogReg, and Kernel-LR, as well as BolT, and PFC is below 0.05 against STAGIN and all BOLD-based baselines except Mamba and BolT. For COBRE, CoRTeX achieves the highest AUC (0.797 ± 0.132) and ACC (0.786 ± 0.063) with balanced sensitivity and specificity. Paired DeLong tests are significant against all baselines, and PFC is below 0.05 against STAGIN and all BOLD-based baselines except Titans and BolT.

Table 3: **Ablation study on COBRE** (AUC, mean±std across 5 folds). Each row indicates whether the corresponding component is included (✓) or removed (×). ΔAUC is the change relative to the full model (top row).

Number	Variant	Hebbian	M_r	PE	AUC	Δ AUC
	Full (CoRTeX)	✓	✓	NeRF & one-hot	$\mathbf{0.797 \pm 0.132}$	–
1	w/o Hebbian	×	✓	NeRF & one-hot	0.735 ± 0.161	-0.062
2	w/o M_r	✓	×	NeRF & one-hot	0.610 ± 0.087	-0.187
3	w/o one-hot only	✓	✓	NeRF	0.724 ± 0.130	-0.073
4	w/o NeRF only	✓	✓	one-hot	0.660 ± 0.094	-0.137
5	w/o NeRF & one-hot	✓	✓	none	0.597 ± 0.148	-0.200
6	w/o Hebbian & M_r	×	×	NeRF & one-hot	0.647 ± 0.139	-0.150
7	w/o Hebbian & NeRF & one-hot	×	✓	none	0.639 ± 0.179	-0.158
8	w/o M_r & NeRF & one-hot	✓	×	none	0.658 ± 0.093	-0.139
9	w/o all three	×	×	none	0.611 ± 0.130	-0.186

4.4 Ablation Study

We ablate the three core components of CoRTeX—positional encoding (PE; NeRF-based coordinates and one-hot ROI identity), region-specific temporal integration (M_r; Eq. (5)), and Hebbian update—on COBRE. Results are reported in Table 3. As these components operate on jointly evolving regional trajectories, their individual ablation effects need not be additive.

Positional encoding has the largest effect within the full architecture. Removing both positional components yields the largest drop ($\Delta\text{AUC} = -0.200$), with NeRF coordinates contributing substantially more than one-hot identity (-0.137 vs. -0.073 individually). Since CoRTeX uses region-specific parameters, ROI identity is partly absorbed by these parameters, whereas NeRF coordinates provide spatial proximity information unavailable elsewhere. This is consistent with prior findings that anatomically adjacent regions share functional properties [Salvador et al., 2005, Margulies et al., 2016]. To test whether this reflects positional information alone, we provide an LSTM baseline with the same encoding. It yields a small gain on NSD and lower AUC on ABIDE and COBRE (Appendix F), suggesting that positional encoding is effective in combination with ROI-indexed states and identity-aware communication rather than on its own.

Region-specific temporal integration is the second key driver. Replacing M_r with a fixed window yields the second-largest single-component effect ($\Delta\text{AUC} = -0.187$), indicating that letting each ROI attend over its own temporal extent, rather than a uniform window, contributes substantially to performance. On NSD, relocating the learned M_r across ROIs at test time, which preserves their marginal distribution, lowers macro-AUC from 0.849 to 0.819, below even a uniform window at the mean (0.829). The learned region-to-window assignment outperforms the permuted configurations in all 40 checkpoints (8 subjects × 5 seeds), with an average macro-AUC improvement of 0.03 (Appendix G). The spatial distribution of the learned M_r is examined in §4.5 and Appendix K. An $M_{\max}$ sensitivity analysis supports our choice (Appendix E).

Hebbian update plays a distinct role. The co-activation memory shows a smaller standalone effect ($\Delta\text{AUC} = -0.062$), but serves a different function from M_r. Whereas M_r explicitly controls the recent history accessed by each region's readout, the Hebbian memory introduces a separate history-dependent mechanism for inter-regional communication through region-specific learning and forgetting rates. Sharing these rates does not recover the full model's performance (Appendix L).

4.5 Regional Temporal Heterogeneity Analysis

We treat the learned M_r as a task-conditioned effective integration window rather than an estimate of an intrinsic neural constant. Unlike post-hoc attribution on an LSTM, which also preserves ROI identity at the input, M_r can be examined directly at the ROI level. Under SmoothGrad [Smilkov et al., 2017] on NSD, the LSTM yields temporal profiles with limited variation across ROIs, with a normalized between-ROI standard deviation of 4.07% compared with 16.31% for CoRTeX when averaged over 8 subjects and 5 seeds (Appendix I). To test whether the regional structure of M_r is induced by the diversity regularizer (Eq. (7)), we sweep α_{div} over $\{0, 10^{-3}, 10^{-2}, 1\}$ on NSD (Appendix H). Dispersion across ROIs more than doubles while macro-AUC varies by less than 0.005, and the per-ROI ranking remains largely preserved relative to $\alpha_{\text{div}} = 0$, with Spearman $\rho \approx 0.95$ at the operating point and $\rho \approx 0.87$ at $\alpha_{\text{div}} = 1$. Since substantial dispersion remains without the

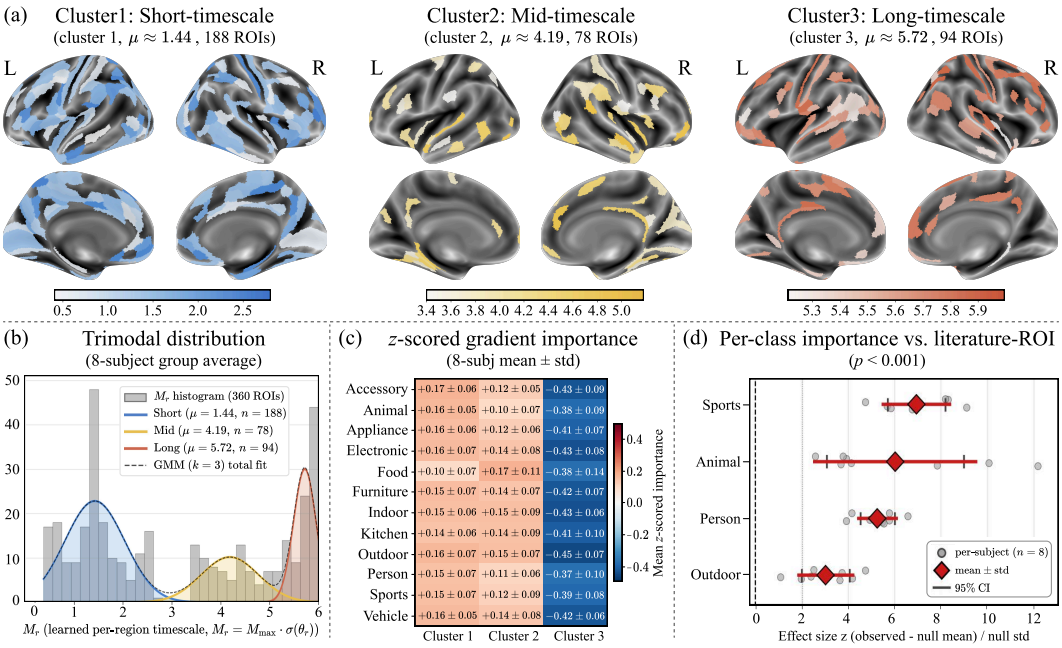


Figure 2: **Trimodal regional timescale organization on NSD task-fMRI within the 6-volume (≈ 8.0 s) trial window.** (a) Cortical projection of learned per-region timescales M_r (360 ROIs) by GMM cluster. (b) Distribution of M_r (8-subject group average) with three-component GMM fit ($k = 3$ selected by BIC). (c) z-scored gradient importance per supercategory and cluster (mean±std over 8 subjects). (d) Per-class importance vs. literature-defined class-specific ROIs (effect size against permutation null; all $p < 0.001$).

regularizer, it scales regional separation rather than determining which region receives which window. Related rs-fMRI analyses are given in Appendix K.

Discretized timescale regimes within the trial window. On NSD, the learned M_r exhibit a trimodal structure within the 6-volume (≈ 8.0 s) trial window (Fig. 2b). A Bayesian information criterion (BIC)-selected Gaussian mixture model (GMM) over $k \in \{1, \ldots, 6\}$ ($k^\star = 3$) partitions the 360 cortical ROIs into short- ($\mu \approx 1.44$, $n = 188$), mid- ($\mu \approx 4.19$, $n = 78$), and long-timescale ($\mu \approx 5.72$, $n = 94$) regimes, and Hartigan's dip test rejects unimodality ($p < 10^{-4}$) in all 40 evaluations (8 subjects × 5 seeds). These within-trial regimes differ from the seconds-to-minutes hierarchy of intrinsic timescales [Hasson et al., 2008, Murray et al., 2014, Raut et al., 2020], although their anatomical assignment is qualitatively in line with its sensory end. Most visual-hierarchy ROIs (V1–V4, V7, FFC, MT, LO) fall in the short regime (Fig. 2a), default-mode and frontoparietal cortex span both short and long regimes, and the short regime spans all Yeo-7 networks (per-ROI values in Appendix J).

Cluster- and class-level importance. Gradient importance $|\partial f_c / \partial \mathbf{x}|$, z-scored across the 360 ROIs within each subject and class and averaged within each GMM cluster (Fig. 2c), is consistently highest in short-timescale regions and lowest in long-timescale regions across classes, with *Food* as a partial exception. This relation also holds modestly at the ROI level, where regions with shorter windows tend to have higher class-averaged importance (Spearman $\rho = -0.136$, spin test $p = 0.007$ [Alexander-Bloch et al., 2018]). As a consistency check, we compared per-class importance with literature-defined, category-selective ROIs, namely face/body/motion regions for *Person* (L_LO2, L_V4t, R_FFC, R_FST, R_MST) [Kanwisher et al., 1997, Downing et al., 2001], motion areas for *Sports* (R_FST, R_MST) [Tootell et al., 1995], ventral-stream regions for *Animal* (V4t, LO2, VVC) [Grill-Spector and Weiner, 2014], and parahippocampal scene-selective cortex and adjacent ventral temporal regions for *Outdoor* (PHA3, PeEc, FFC) [Epstein and Kanwisher, 1998]. For all four classes, importance in these ROIs exceeded a permutation null (Fig. 2d, all $p < 0.001$), indicating that class-specific attribution is consistent with category-relevant cortex.

5 Conclusion

We introduced CoRTeX, a region-resolved recurrent architecture that jointly models regional BOLD dynamics, ROI-specific temporal integration, and adaptive inter-regional coupling. Each region maintains its own recurrent state and learns an explicit integration window, while co-activation among evolving regional trajectories is accumulated with region-specific learning and forgetting rates to guide identity-aware attention. Across NSD, ABIDE, and COBRE, CoRTeX achieves competitive decoding performance while enabling direct examination of learned ROI-to-window assignments and class-specific regional attribution.

On NSD, the learned M_r exhibit a trimodal pattern within the 6-volume trial window, whereas ABIDE and COBRE show unimodal distributions. These differences may reflect task structure, although $M_{\max}$ and sequence lengths also differ. The learned windows should not be interpreted as intrinsic neural timescales. On NSD, class-specific attribution is consistent with category-selective cortex. On ABIDE, multi-network attribution differences survive false discovery rate (FDR) and Bonferroni correction in the correctly classified subset but not in the full cohort. The default mode network (DMN)-leaning COBRE pattern remains exploratory (Appendix K). These correspondences serve as consistency checks rather than new neuroscientific evidence.

Several limitations remain. CoRTeX does not consistently outperform the strongest β- and FC-based classifiers, which may benefit from stimulus-aligned GLM estimates or full-scan connectivity summaries. In contrast, CoRTeX models raw BOLD trajectories and evolving inter-regional relationships, supporting prediction alongside explicit ROI-level temporal analysis. Incorporating complementary information from such summary features is a promising direction. Performance may depend on dataset size, particularly for the small COBRE cohort with higher variability, and $M_{\max}$ remains a selected hyperparameter (Appendix E). Future work includes relaxing this bound, controlling classifier-head differences across datasets, modeling regional HRF variability to distinguish neural from hemodynamic contributions to M_r, and comparing attribution with β-maps.

More broadly, by tying learnable components to anatomically defined cortical regions, CoRTeX provides a basis for fMRI decoders that support region-level temporal analysis, and its per-region analyses may help generate hypotheses about cortical organization for further neuroscientific investigation. The learned windows and attributions are task-conditioned model properties rather than direct estimates of neural mechanisms, and the disease diagnosis results are intended for research only.

Acknowledgments and Disclosure of Funding

This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2019-II190079, Artificial Intelligence Graduate School Program (Korea University)), the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2026-25518987, K-NeuroMind NEXUS: Neurocognitive-Affective Next-Generation Cross-modal Foundation Platform), and Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2024-00457882, National AI Research Lab Project). The authors declare no competing financial interests.

References

CJ Aine, Henry Jeremy Bockholt, Juan R Bustillo, José M Cañive, Arvind Caprihan, Charles Gasparovic, Faith M Hanlon, Jon M Houck, Rex E Jung, John Lauriello, et al. Multimodal neuroimaging in schizophrenia: Description and dissemination. *Neuroinformatics*, 15(4):343–364, 2017.

Aaron F Alexander-Bloch, Haochang Shou, Siyuan Liu, Theodore D Satterthwaite, David C Glahn, Russell T Shinohara, Simon N Vandekar, and Armin Raznahan. On testing for spatial correspondence between maps of human brain structure and function. *NeuroImage*, 178:540–551, 2018.

Emily J Allen, Ghislain St-Yves, Yihan Wu, Jesse L Breedlove, Jacob S Prince, Logan T Dowdle, Matthias Nau, Brad Caron, Franco Pestilli, Ian Charest, et al. A massive 7T fMRI dataset to bridge cognitive neuroscience and artificial intelligence. *Nature Neuroscience*, 25(1):116–126, 2022.

Jimmy Ba, Geoffrey E Hinton, Volodymyr Mnih, Joel Z Leibo, and Catalin Ionescu. Using fast weights to attend to the recent past. In *Advances in Neural Information Processing Systems*, volume 29, 2016.

Hasan A Bedel, Irmak Sivgin, Onat Dalmaz, Salman UH Dar, and Tolga Çukur. BolT: Fused window transformers for fMRI time series analysis. *Medical Image Analysis*, 88:102841, 2023.

Ali Behrouz, Peilin Zhong, and Vahab Mirrokni. Titans: Learning to memorize at test time. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Yashar Behzadi, Khaled Restom, Joy Liau, and Thomas T Liu. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage*, 37(1):90–101, 2007.

Robyn L Bluhm, Jodi Miller, Ruth A Lanius, Elizabeth A Osuch, Kristine Boksman, RWJ Neufeld, Jean Théberge, Betsy Schaefer, and Peter Williamson. Spontaneous low-frequency fluctuations in the BOLD signal in schizophrenic patients: Anomalies in the default network. *Schizophrenia Bulletin*, 33(4):1004–1012, 2007.

Vince D Calhoun, Robyn Miller, Godfrey Pearlson, and Tulay Adalı. The chronnectome: Time-varying connectivity networks as the next frontier in fMRI data discovery. *Neuron*, 84(2):262–274, 2014.

Evangelia Christodoulou, Annika Reinke, Pascaline Andrè, Patrick Godau, Piotr Kalinowski, Rola Houhou, Selen Erkan, Carole H Sudre, Ninon Burgos, Sofiène Boutaj, et al. False promises in medical imaging AI? Assessing validity of outperformance claims. *arXiv preprint arXiv:2505.04720*, 2025.

Luke Nicholas Darlow, Ciaran Regan, Sebastian Risi, Jeffrey Seely, and Llion Jones. Continuous thought machines. In *Advances in Neural Information Processing Systems*, volume 38, 2025.

Elizabeth R DeLong, David M DeLong, and Daniel L Clarke-Pearson. Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach. *Biometrics*, 44(3):837–845, 1988.

Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, Francisco X Castellanos, Kaat Alaerts, Jeffrey S Anderson, Michal Assaf, Susan Y Bookheimer, Mirella Dapretto, et al. The Autism Brain Imaging Data Exchange: Towards a large-scale evaluation of the intrinsic brain architecture in autism. *Molecular Psychiatry*, 19(6):659–667, 2014.

Paul E Downing, Yuhong Jiang, Miles Shuman, and Nancy Kanwisher. A cortical area selective for visual processing of the human body. *Science*, 293(5539):2470–2473, 2001.

Jeffrey L Elman. Finding structure in time. *Cognitive Science*, 14(2):179–211, 1990.

Russell Epstein and Nancy Kanwisher. A cortical representation of the local visual environment. *Nature*, 392(6676):598–601, 1998.

Karl J Friston. Functional and effective connectivity: A review. *Brain Connectivity*, 1(1):13–36, 2011.

Karl J Friston, Paul Fletcher, Oliver Josephs, Andrew Holmes, Michael D Rugg, and Robert Turner. Event-related fMRI: Characterizing differential responses. *NeuroImage*, 7(1):30–40, 1998.

Abigail G Garrity, Godfrey D Pearlson, Kristen McKiernan, Dan Lloyd, Kent A Kiehl, and Vince D Calhoun. Aberrant "default mode" functional connectivity in schizophrenia. *American Journal of Psychiatry*, 164(3):450–457, 2007.

Matthew F Glasser, Timothy S Coalson, Emma C Robinson, Carl D Hacker, John Harwell, Essa Yacoub, Kamil Ugurbil, Jesper Andersson, Christian F Beckmann, Mark Jenkinson, et al. A multi-modal parcellation of human cerebral cortex. *Nature*, 536(7615):171–178, 2016.

Gary H Glover. Deconvolution of impulse response in event-related BOLD fMRI. *NeuroImage*, 9(4): 416–429, 1999.

Kalanit Grill-Spector and Kevin S Weiner. The functional architecture of the ventral temporal cortex and its role in categorization. *Nature Reviews Neuroscience*, 15(8):536–548, 2014.

Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

Daniel A Handwerker, John M Ollinger, and Mark D'Esposito. Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. *NeuroImage*, 21(4):1639–1651, 2004.

Uri Hasson, Eunice Yang, Ignacio Vallines, David J Heeger, and Nava Rubin. A hierarchy of temporal receptive windows in human cortex. *Journal of Neuroscience*, 28(10):2539–2550, 2008.

Donald Olding Hebb. *The organization of behavior: A neuropsychological theory*. Psychology Press, 2005.

Dan Hendrycks and Kevin Gimpel. Gaussian error linear units (GELUs). *arXiv preprint arXiv:1606.08415*, 2016.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8): 1735–1780, 1997.

Christopher J Honey, Thomas Thesen, Tobias H Donner, Lauren J Silbert, Chad E Carlson, Orrin Devinsky, Werner K Doyle, Nava Rubin, David J Heeger, and Uri Hasson. Slow cortical dynamics and the accumulation of information over long timescales. *Neuron*, 76(2):423–434, 2012.

R Matthew Hutchison, Thilo Womelsdorf, Elena A Allen, Peter A Bandettini, Vince D Calhoun, Maurizio Corbetta, Stefania Della Penna, Jeff H Duyn, Gary H Glover, Javier Gonzalez-Castillo, et al. Dynamic functional connectivity: Promise, issues, and interpretations. *NeuroImage*, 80: 360–378, 2013.

Nancy Kanwisher. Functional specificity in the human brain: A window into the functional architecture of the mind. *Proceedings of the National Academy of Sciences*, 107(25):11163–11170, 2010.

Nancy Kanwisher, Josh McDermott, and Marvin M Chun. The fusiform face area: A module in human extrastriate cortex specialized for face perception. *Journal of Neuroscience*, 17(11): 4302–4311, 1997.

Jeremy Kawahara, Colin J Brown, Steven P Miller, Brian G Booth, Vann Chau, Ruth E Grunau, Jill G Zwicker, and Ghassan Hamarneh. BrainNetCNN: Convolutional neural networks for brain networks; towards predicting neurodevelopment. *NeuroImage*, 146:1038–1049, 2017.

Byung-Hoon Kim, Jong Chul Ye, and Jae-Jin Kim. Learning dynamic graph representation of brain connectome with spatio-temporal attention. In *Advances in Neural Information Processing Systems*, volume 34, 2021.

Nora Leonardi and Dimitri Van De Ville. On spurious and real fluctuations of dynamic functional connectivity during rest. *NeuroImage*, 104:430–436, 2015.

Raphael Liegeois, Timothy O Laumann, Abraham Z Snyder, Juan Zhou, and BT Thomas Yeo. Interpreting temporal fluctuations in resting-state functional connectivity MRI. *NeuroImage*, 163: 437–455, 2017.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft COCO: Common objects in context. In *European Conference on Computer Vision*, pages 740–755. Springer, 2014.

Abdelhak Mahmoudi, Sylvain Takerkart, Fakhita Regragui, Driss Boussaoud, and Andrea Brovelli. Multivoxel pattern analysis for fMRI data: A review. *Computational and Mathematical Methods in Medicine*, 2012(1):961257, 2012.

Daniel S Margulies, Satrajit S Ghosh, Alexandros Goulas, Marcel Falkiewicz, Julia M Huntenburg, Georg Langs, Gleb Bezgin, Simon B Eickhoff, F Xavier Castellanos, Michael Petrides, et al. Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proceedings of the National Academy of Sciences*, 113(44):12574–12579, 2016.

Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. NeRF: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

John D Murray, Alberto Bernacchia, David J Freedman, Ranulfo Romo, Jonathan D Wallis, Xinying Cai, Camillo Padoa-Schioppa, Tatiana Pasternak, Hyojung Seo, Daeyeol Lee, et al. A hierarchy of intrinsic timescales across primate cortex. *Nature Neuroscience*, 17(12):1661–1663, 2014.

Ryan V Raut, Abraham Z Snyder, and Marcus E Raichle. Hierarchical dynamics as a macroscopic organizing principle of the human brain. *Proceedings of the National Academy of Sciences*, 117 (34):20890–20897, 2020.

Raymond Salvador, John Suckling, Martin R Coleman, John D Pickard, David Menon, and ED Bullmore. Neurophysiological architecture of functional magnetic resonance images of human brain. *Cerebral Cortex*, 15(9):1332–1342, 2005.

Alexander Schaefer, Ru Kong, Evan M Gordon, Timothy O Laumann, Xi-Nian Zuo, Avram J Holmes, Simon B Eickhoff, and BT Thomas Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity MRI. *Cerebral Cortex*, 28(9):3095–3114, 2018.

Paul Scotti, Atmadeep Banerjee, Jimmie Goode, Stepan Shabalin, Alex Nguyen, Aidan Dempster, Nathalie Verlinde, Elad Yundler, David Weisberg, Kenneth Norman, et al. Reconstructing the mind's eye: fMRI-to-image with contrastive learning and diffusion priors. In *Advances in Neural Information Processing Systems*, volume 36, pages 24705–24728, 2023.

Paul Steven Scotti, Mihir Tripathy, Cesar Torrico, Reese Kneeland, Tong Chen, Ashutosh Narang, Charan Santhirasegaran, Jonathan Xu, Thomas Naselaris, Kenneth A. Norman, and Tanishq Mathew Abraham. MindEye2: Shared-subject models enable fMRI-to-image with 1 hour of data. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 44038–44059. PMLR, 21–27 Jul 2024.

Daniel Smilkov, Nikhil Thorat, Been Kim, Fernanda Viégas, and Martin Wattenberg. SmoothGrad: Removing noise by adding noise. *arXiv preprint arXiv:1706.03825*, 2017.

Yu Takagi and Shinji Nishimoto. High-resolution image reconstruction with latent diffusion models from human brain activity. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14453–14463, 2023.

Matthew Tancik, Pratul Srinivasan, Ben Mildenhall, Sara Fridovich-Keil, Nithin Raghavan, Utkarsh Singhal, Ravi Ramamoorthi, Jonathan Barron, and Ren Ng. Fourier features let networks learn high frequency functions in low dimensional domains. In *Advances in Neural Information Processing Systems*, volume 33, pages 7537–7547, 2020.

Roger B Tootell, John B Reppas, Kenneth K Kwong, Rafael Malach, Richard T Born, Thomas J Brady, Bruce R Rosen, and John W Belliveau. Functional analysis of human MT and related visual cortical areas using magnetic resonance imaging. *Journal of Neuroscience*, 15(4):3215–3230, 1995.

Grigorios Tsoumakas, Ioannis Katakis, and Ioannis Vlahavas. Mining multi-label data. *Data Mining and Knowledge Discovery Handbook*, pages 667–685, 2010.

Martijn P van den Heuvel and Hilleke E Hulshoff Pol. Exploring the brain network: A review on resting-state fMRI functional connectivity. *European Neuropsychopharmacology*, 20(8):519–534, 2010.

Diego Vidaurre, Stephen M Smith, and Mark W Woolrich. Brain network dynamics are hierarchically organized in time. *Proceedings of the National Academy of Sciences*, 114(48):12827–12832, 2017.

Susan Whitfield-Gabrieli, Heidi W Thermenos, Snezana Milanovic, Ming T Tsuang, Stephen V Faraone, Robert W McCarley, Martha E Shenton, Alan I Green, Alfonso Nieto-Castanon, Peter LaViolette, et al. Hyperactivity and hyperconnectivity of the default network in schizophrenia and in first-degree relatives of persons with schizophrenia. *Proceedings of the National Academy of Sciences*, 106(4):1279–1284, 2009.

BT Thomas Yeo, Fenna M Krienen, Jorge Sepulcre, Mert R Sabuncu, Danial Lashkari, Marisa Hollinshead, Joshua L Roffman, Jordan W Smoller, Lilla Zöllei, Jonathan R Polimeni, et al. The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, 106(3):1125–1165, 2011.

A Architectural Comparison with Sequence-Modeling Baselines

We compare CoRTeX with recurrent and sequence-modeling baselines along key architectural dimensions (Table A1). While these models span diverse paradigms—recurrent, attention-based, and state-space—CoRTeX is distinguished by two design choices: (1) its computational primitive is anchored to anatomically defined cortical regions rather than exchangeable abstractions, and (2) inter-regional communication is shaped by a within-sequence Hebbian-like memory update rather than static parameters.

Distinction from memory-augmented models. CoRTeX shares with Titans [Behrouz et al., 2025] the use of inference-time updates, but the two differ fundamentally in what is updated and how: Titans applies surprise-based gradient updates to a shared memory module that operates over exchangeable tokens, while CoRTeX applies Hebbian-like updates ($\mathbf{H}_t = \gamma_r \mathbf{H}_{t-1} + \eta_r \mathbf{S}_t$) to a region-pair-specific memory matrix indexed by anatomical identity, with each region carrying its own learning and forgetting rates.

Beyond these architectural dimensions, CoRTeX additionally supports region-level analyses, such as the spatial organization of learned timescales and class-specific regional attribution, that are not directly available in models with exchangeable units. Spiking neural networks are omitted because their spike-timing representation is mismatched with the blood-oxygen-level-dependent (BOLD)-rate signal at typical fMRI temporal resolution.

Table A1: **Architectural comparison between CoRTeX and sequence-modeling baselines along key dimensions.** ✓: explicitly built into the architecture; ✗: not natively supported.

Dimension	LSTM	Transformer	Mamba	Titans	CTM	CoRTeX (Ours)
Computational primitive	Hidden state vector	Token	State-space vector	Token & memory slots	Artificial neuron	**Cortical region (anatomically defined)**
Identity of primitives	Exchangeable	Exchangeable	Exchangeable	Exchangeable	Exchangeable	**Fixed by parcellation**
Recurrent processing	✓ Gated recurrence	✗	✓ Selective SSM	✓ Attention with memory	✓ Internal ticks	✓ Recurrent over time points
Inter-unit communication	Fully-connected gates	Self-attention	Channel-wise selective SSM	Cross-attention + memory	Synapse MLP + synchronization	**Identity-aware attention**
Per-unit timescale	Shared (via gates)	N/A	Input-dependent (shared Δ)	Multi-scale memory (short/long)	Per-neuron private temporal unit	**Per-region learnable timescale**
Inference-time memory update	✗	✗	✗	✓ Surprise-based update on shared memory	✗	✓ **Hebbian-like update on region-pair memory**

B Dataset Details

The datasets used in this work are publicly available at the following links:

- NSD: `https://naturalscenesdataset.org/`
- ABIDE: `https://fcon_1000.projects.nitrc.org/indi/abide/abide_I.html`
- COBRE: `http://fcon_1000.projects.nitrc.org/indi/retro/cobre.html`

B.1 Task-fMRI Dataset

Natural Scenes Dataset. The Natural Scenes Dataset (NSD; Allen et al., 2022) is a large-scale fMRI dataset comprising 8 subjects, each viewing up to 10,000 unique natural images drawn from the Microsoft COCO corpus [Lin et al., 2014], with each image presented up to three times for a maximum of 30,000 trials per subject. Because some participants did not complete the full acquisition schedule, the number of scanning sessions varies across subjects. Subjects 1, 2, 5, and 7 completed

all 40 sessions (30,000 trials each), subjects 3 and 6 completed 32 sessions (24,000 trials each), and subjects 4 and 8 completed 30 sessions (22,500 trials each). We explicitly accounted for this imbalance in both preprocessing and the splits described below.

NSD supercategory. For multi-label classification, we group the 80 fine-grained MS-COCO categories into 12 supercategories using the official COCO metadata, as listed in Table A2.

Table A2: **MS-COCO supercategory hierarchy used for NSD multi-label classification.** The 80 fine-grained object categories are grouped into 12 supercategories using the official COCO metadata (`supercategory` field in `instances_*.json`).

Supercategory	#	Fine categories
Person	1	person
Vehicle	8	bicycle, car, motorcycle, airplane, bus, train, truck, boat
Outdoor	5	traffic light, fire hydrant, stop sign, parking meter, bench
Animal	10	bird, cat, dog, horse, sheep, cow, elephant, bear, zebra, giraffe
Accessory	5	backpack, umbrella, handbag, tie, suitcase
Sports	10	frisbee, skis, snowboard, sports ball, kite, baseball bat, baseball glove, skateboard, surfboard, tennis racket
Kitchen	7	bottle, wine glass, cup, fork, knife, spoon, bowl
Food	10	banana, apple, sandwich, orange, broccoli, carrot, hot dog, pizza, donut, cake
Furniture	6	chair, couch, potted plant, bed, dining table, toilet
Electronic	6	TV, laptop, mouse, remote, keyboard, cell phone
Appliance	5	microwave, oven, toaster, sink, refrigerator
Indoor	7	book, clock, vase, scissors, teddy bear, hair drier, toothbrush
Total	**80**	

Train/validation/test split. For each subject, the held-out test set consisted of the "shared1000" images, a fixed 1,000-image subset presented to all subjects, yielding between 2,188 and 3,000 test trials per subject depending on session coverage. The remaining trials, corresponding to the COCO `train2017` and `val2017` partitions, were pooled, and in each run approximately 4% of them, matching the COCO validation proportion, were randomly sampled as the validation set for model selection, with the rest used for training. All experiments were repeated with five random seeds.

Neural signals. We used two complementary signals from NSD, both derived from the 1.8 mm isotropic functional preprocessing (`func1pt8mm`; acquired at a repetition time (TR) of 1.6 s and temporally resampled by NSD to 1.333 s): (i) the pre-denoised BOLD time series, and (ii) the single-trial response amplitudes (β weights) obtained via the `betas_fithrf_GLMdenoise_RR` (b3) pipeline. Following the NSD data manual, the `int16` β values were divided by 300 to recover percent-signal-change units.

ROI parcellation. Voxel-level signals were reduced to region-of-interest (ROI) features using the subject-native HCP multi-modal parcellation (HCP-MMP1; Glasser et al., 2016) provided by NSD, which partitions the cortex into 180 parcels per hemisphere (left: labels 1–180; right: labels 181–360 after concatenation), for a total of 360 cortical ROIs. For each ROI, BOLD and β signals were obtained by averaging across all voxels belonging to that parcel.

BOLD trial extraction. Using the stimulus onset times from NSD's design information, BOLD signals were z-scored per ROI at the run level and then sliced into per-trial windows. To capture the canonical hemodynamic response, each window spanned 6 volumes (8.0 s) starting from the stimulus onset, covering the typical hemodynamic response function (HRF) peak at 4–6 s post-onset and a brief post-peak interval [Glover, 1999, Friston et al., 1998].

B.2 Resting-State fMRI Datasets

B.2.1 Autism Brain Imaging Data Exchange

The Autism Brain Imaging Data Exchange (ABIDE; Di Martino et al., 2014) aggregates resting-state fMRI scans from multiple international sites for the study of autism spectrum disorder (ASD). From the original release of $1{,}102$ subjects, we excluded 30 subjects whose preprocessed time series contained one or more zero-valued ROIs, yielding $1{,}072$ subjects used in our experiments (511 with ASD and 561 typically developing (TD) controls).

Demographics. ASD subjects ($n = 511$) had a mean age of 16.75 ± 8.16 years (range 7.0 to 64.0), with 450 males and 61 females. TD subjects ($n = 561$) had a mean age of 17.03 ± 7.68 years (range 6.5 to 56.2), with 462 males and 99 females. Across the full cohort, the mean age was 16.90 ± 7.91 years, with 912 males and 160 females.

B.2.2 Center for Biomedical Research Excellence

The Center for Biomedical Research Excellence (COBRE; Aine et al., 2017) dataset provides resting-state fMRI for the study of schizophrenia. COBRE recordings are typically 150 volumes per subject. From the original 148 subjects, we excluded 3 subjects, namely 2 subjects with unknown diagnostic labels (marked as "Disenrolled") and 1 subject whose time series was shorter than the required length ($T = 67$ vs. $T = 150$). This yielded 145 subjects used in our experiments (71 with schizophrenia (SZ) and 74 healthy controls (HC)).

Demographics. SZ subjects ($n = 71$) had a mean age of 38.14 ± 13.99 years (range 18.0 to 65.0), with 57 males and 14 females. HC subjects ($n = 74$) had a mean age of 35.82 ± 11.58 years (range 18.0 to 65.0), with 51 males and 23 females. Across the full cohort, the mean age was 36.96 ± 12.83 years, with 108 males and 37 females.

B.2.3 Preprocessing

For ABIDE, we used the C-PAC preprocessed release from the Preprocessed Connectomes Project's ABIDE Preprocessed Initiative (`http://preprocessed-connectomes-project.org/abide/`), which includes band-pass filtering (0.01–0.1 Hz) and global signal regression. The COBRE dataset was preprocessed using C-PAC with the following pipeline. The standard preprocessing steps included slice-timing correction, motion correction, distortion correction, registration to a standard template, and Gaussian spatial smoothing (4 mm FWHM). Nuisance signals were then regressed out using 24 motion parameters, the CompCor approach with five principal components [Behzadi et al., 2007], and both linear and quadratic detrending to remove low-frequency drifts. A band-pass filter (0.01–0.1 Hz) was applied as the final step.

ROI parcellation. For both ABIDE and COBRE, voxel-level BOLD signals were reduced to ROI features using the 200-region Schaefer parcellation [Schaefer et al., 2018] by averaging across the constituent voxels, and ROI-wise BOLD signals were z-scored across volumes.

B.2.4 Temporal Harmonization via Fourier Resampling

ABIDE aggregates scans from 17 acquisition sites with heterogeneous TRs (ranging from $1.5\,\mathrm{s}$ to $3.0\,\mathrm{s}$), while COBRE was acquired at a single TR of $2.0\,\mathrm{s}$. To enable cross-site training under a unified temporal grid, we resampled all ABIDE BOLD time series to a common target TR of $2.0\,\mathrm{s}$ using Fourier-domain resampling. COBRE required no resampling because all subjects share the same TR.

Fourier-domain resampling. For each ROI time series $x \in \mathbb{R}^{T_{\mathrm{orig}}}$, we compute the real Fast Fourier Transform (FFT), truncate or zero-pad the frequency spectrum to match the target length, and apply the inverse real FFT, with an amplitude correction factor of $T_{\mathrm{new}}/T_{\mathrm{orig}}$. This procedure is equivalent to `scipy.signal.resample` and is applied independently along the temporal axis of each ROI,

$$T_{\mathrm{new}} = \mathrm{round}\left(T_{\mathrm{orig}} \cdot \frac{\mathrm{TR}_{\mathrm{orig}}}{\mathrm{TR}_{\mathrm{target}}}\right), \qquad \mathrm{TR}_{\mathrm{target}} = 2.0\,\mathrm{s}. \tag{9}$$

We chose Fourier resampling over linear interpolation for two reasons. First, unlike linear interpolation, it does not attenuate the amplitude of low-frequency components. Second, the low-frequency band of interest for BOLD (0.01–0.1 Hz) is preserved without distortion.

Per-site TRs. Table A3 summarizes the native TR, resampling direction, and resulting time-series length for each of the 17 ABIDE sites. The minimum post-resampling length across all subjects was 98 volumes (from OHSU, with $T_{\text{orig}} = 78$ at TR of 2.5 s), and we therefore set $T_{\min} = 98$ as the unified segment length.

Table A3: **ABIDE site-specific TRs and Fourier resampling to** $\text{TR}_{\text{target}} = 2.0\,\text{s}$.

Site	TR (s)	Direction	T_{orig}	$T_{\text{resampled}}$
CALTECH	2.000	no change	146	146
CMU	2.000	no change	202 to 316	202 to 316
KKI	2.500	upsample	124 to 152	155 to 190
LEUVEN	1.667	downsample	246	205
MAX_MUN	3.000	upsample	116 to 196	174 to 294
NYU	2.000	no change	176	176
OHSU	2.500	upsample	78	98 ($T_{\min}$)
OLIN	1.500	downsample	206	154
PITT	1.500	downsample	196	147
SBL	2.200	upsample	196	216
SDSU	2.000	no change	176	176
STANFORD	2.000	no change	176 to 236	176 to 236
TRINITY	2.000	no change	146	146
UCLA	3.000	upsample	116	174
UM	2.000	no change	296	296
USM	2.000	no change	232 to 236	232 to 236
YALE	2.000	no change	196	196

Segment splitting. For subjects whose resampled length exceeded $T_{\min} = 98$, we partitioned the time series into non-overlapping segments of length $T_{\min}$. For example, a UM_1 subject with $T = 296$ yields $\lfloor 296/98 \rfloor = 3$ segments. Subject-level splits were preserved, so all segments from a given subject were assigned to the same train, validation, or test partition to prevent leakage. After segmentation, ABIDE yielded 1,702 segments from 1,072 subjects (approximately 1.59 segments per subject), while COBRE yielded 145 segments from 145 subjects, each with $T = 150$ and therefore a single segment per subject.

B.3 Train/Validation/Test Split

For ABIDE, we performed 5-fold cross-validation stratified by both acquisition site and diagnostic label (ASD/TD) to ensure that each fold preserves the joint distribution of sites and labels. Within each fold, 10% of the training subjects were held out as the validation set. For COBRE, we performed 5-fold cross-validation stratified by diagnostic label (SZ/HC), with the same 10% validation split within each training fold. At test time, predictions from all segments of a subject were averaged to obtain a subject-level prediction, and all metrics are reported at the subject level.

C Implementation Details

The fMRI-specific baselines (BrainNetCNN, STAGIN, and BolT) are trained with their official codebases under the same split as CoRTeX, using their publicly released default hyperparameters. However, to ensure that baseline performance was not constrained by model size, we swept MLP, RNN, LSTM, Mamba, and Titans over hidden dimensions ($\{64, 128, 256, 384, 512, 1024\}$) and layer counts ($\{2, 4, 6\}$), with chunk sizes ($\{4, 16, 32, 64\}$) additionally considered for Titans. For these models, we also searched batch size over $\{4, 8, 16, 32\}$ and learning rate over $\{1\times10^{-3}, 5\times10^{-4}, 2\times10^{-4}, 1.5\times10^{-4}, 1\times10^{-4}\}$. Configurations were selected by validation performance. The sweep spans 0.143 M to 47.673 M parameters for the LSTM on NSD and 0.102 M to 47.008 M on ABIDE and COBRE, covering sizes larger than CoRTeX in both settings. The selected sizes vary across

datasets, and the best-performing configuration is not consistently the largest. Tables A6 and A7 report the parameter counts of the selected configurations.

Hyperparameter selection. For CoRTeX, hyperparameters were selected based on validation performance on each dataset. We searched the learning rate over values in $[10^{-6}, 10^{-3}]$ (*e.g.,* $1 \times 10^{-3}, 5 \times 10^{-4}, 2 \times 10^{-4}, 1.5 \times 10^{-4}, 1 \times 10^{-4}$), batch size over $\{4, 8, 16, 32\}$, and the number of training epochs over $\{30, 50, 60, 100\}$. Other hyperparameters (weight decay, warmup ratio, dropout, etc.) were initialized to common defaults and lightly adjusted when validation performance plateaued. The final values reported in Tables A4 and A5 are those that achieved the best validation performance on each dataset.

Compute resources. Experiments were run on PyTorch 2.6.0 / CUDA 12.4 with a single NVIDIA RTX 2080 Ti GPU (11 GB). The main runs reported in this paper total ~ 70 GPU-hours (~ 4 h ABIDE, ~ 1 h COBRE, ~ 66 h NSD). Additional compute was spent on hyperparameter search and exploratory experiments not reported here.

Table A4: **CoRTeX training and data hyperparameters across the three datasets.** R denotes the number of ROIs and T the input temporal length.

Hyperparameter	**ABIDE**	**COBRE**	**NSD**
Data			
Atlas / parcels (R)	Schaefer-200	Schaefer-200	Glasser-360
Input length T	98	150	6
n classes	2 (binary)	2 (binary)	12 (multi-label)
Optimization			
Optimizer		AdamW	
Learning rate	1×10^{-5}	1×10^{-5}	1.5×10^{-4}
Weight decay	1×10^{-6}	1×10^{-6}	1×10^{-5}
Batch size	16	16	4
Epochs	60	60	60
Warmup (epochs)	10	10	10
Cosine LR floor (`min_lr_ratio`)	0.05	0.05	0.05
Gradient clipping (max-norm)	1.0	1.0	1.0
Dropout (readout)	0.1	0.1	0.1
Early stopping	none	none	patience 10

Table A5: **CoRTeX architectural hyperparameters across the three datasets.**

Hyperparameter	**ABIDE**	**COBRE**	**NSD**
Region-Specific Activation Trajectory (§3.5)			
The number of volumes as input	T=98	T=150	T=6
Region-specific readout hidden width d_{rsr} (Eq. (12))	1024	1024	1024
$M_{\max}$	20	20	6
A fixed sharpness λ (Eq. (6))	10.0	10.0	10.0
θ_r init std (Eq. (5))	1.0	1.0	1.0
Diversity weight α_{div} (Eq. (7))	0.01	0.01	0.01
Per-region projection hidden d_{prp} (§D.2)	16	16	16
Initial $\mathbf{z}_0$, $\mathbf{U}_0$ scale	0.1	0.1	0.1
Identity-Aware Attention (§3.4)			
Attention dim d_{attn}	64	64	64
NeRF frequencies	6	6	6
Centroid space	MNI152	MNI152	native voxel
Hebbian α init (Eq. (3))	0.5	0.5	0.5
Hebbian rate (η_r, γ_r) init std (Eq. (2))	0.5	0.5	0.5
Hebbian $\mathbf{H}_t$ clip (Eq. (2))	5.0	5.0	5.0
Classification Head			
Head type	MLP	MLP	BrainNetCNN
MLP - hidden	$(1024, 256)$	$(1024, 256)$	–
MLP - activation	ReLU	ReLU	–
MLP input	$\mathbf{S}_T$ upper-tri $(R(R-1)/2)$	$\mathbf{S}_T$ upper-tri	–
BrainNetCNN - E2E filters	–	–	8
BrainNetCNN - E2N filters	–	–	16
BrainNetCNN - FC hidden	–	–	64
Total parameters			
# parameters	25.22 M	25.22 M	3.64 M

Table A6: **Parameter counts of deep models** (in millions).

Method	Feature	NSD	ABIDE	COBRE
MLP	FC	66.437	–	–
MLP	BOLD	2.744	–	–
BrainNetCNN	FC	–	0.933	0.933
STAGIN	FC	–	1.004	1.004
BolT	BOLD	–	2.672	2.672
RNN	BOLD	0.293	3.357	0.142
LSTM	BOLD	0.385	3.565	0.301
Mamba	BOLD	6.971	6.884	3.494
Titans	BOLD	1.343	2.883	0.338
CTM	BOLD	15.237	1.348	1.348
CoRTeX	BOLD	3.639	25.222	25.222

Table A7: **Parameter counts of linear and kernel baselines.** D is the input dimension.

Dataset	Feature	D	RidgeClassifier / LogReg	Kernel-LR
NSD	β	360	4,332	6,012
NSD	BOLD	2,160	25,932	6,012
ABIDE	FC	19,900	19,901	1,001
ABIDE	BOLD	12,800	12,801	1,001
COBRE	FC	19,900	19,901	104–106
COBRE	BOLD	12,800	12,801	104–106

Note: For ABIDE and COBRE, classical BOLD baselines use 64 PCA-derived features per ROI (12,800 in total). PCA was fitted on the training set of each fold and applied to the corresponding validation and test sets.

D Architecture Details

This appendix provides the full mathematical specifications and implementation details of CoRTeX's components, summarized in §3.

D.1 NeRF-Based Positional Encoding

Each cortical region r is associated with a 3D centroid $\mathbf{p}_r \in [-1, 1]^3$, computed as the volume-weighted center of mass of the region's voxels and linearly rescaled to $[-1, 1]$ along each axis. The atlas and reference space depend on the dataset. For NSD, we use the HCP-MMP1 atlas [Glasser et al., 2016] in each subject's native volume space as provided by the NSD release ($R = 360$), and for ABIDE and COBRE, we use the Schaefer 200-region atlas [Schaefer et al., 2018] in MNI152 space ($R = 200$). We then apply a NeRF-style sinusoidal encoding [Mildenhall et al., 2021] with L frequency bands:

$$\boldsymbol{\gamma}(\mathbf{p}) = \Big[\big(\sin(\pi\mathbf{p}), \cos(\pi\mathbf{p})\big), \ldots, \big(\sin(2^{L-1}\pi\mathbf{p}), \cos(2^{L-1}\pi\mathbf{p})\big) \Big] \in \mathbb{R}^{6L}, \tag{10}$$

where $\sin$ and $\cos$ are applied element-wise to $\mathbf{p}$ and the per-frequency 6-vectors are concatenated across L frequencies. This encoding provides high-frequency components that help the attention module discriminate anatomically nearby regions, addressing the spectral bias of MLPs toward low-frequency functions [Tancik et al., 2020]. We use $L = 6$ in our experiments.

D.2 Per-Region Projection

For each region r, the per-region projection f_r is a region-specific two-layer feed-forward network with hidden dimension d_{prp} applied to the concatenation of the previous regional state $\mathbf{z}_t[r]$ and the gated attention output $\widetilde{\mathbf{x}}_t[r]$:

$$\mathbf{u}_t[r] = \mathbf{w}_2^{(r)\top} \phi\big(\mathbf{W}_1^{(r)}[\mathbf{z}_t[r]; \widetilde{\mathbf{x}}_t[r]] + \mathbf{b}_1^{(r)}\big) + \mathbf{b}_2^{(r)}, \tag{11}$$

where ϕ is the GELU non-linearity [Hendrycks and Gimpel, 2016] and $\mathbf{W}_1^{(r)} \in \mathbb{R}^{d_{\text{prp}} \times 2}$, $\mathbf{b}_1^{(r)} \in \mathbb{R}^{d_{\text{prp}}}$, $\mathbf{w}_2^{(r)} \in \mathbb{R}^{d_{\text{prp}}}$, and $\mathbf{b}_2^{(r)} \in \mathbb{R}$ are specific to region r and are not shared across regions. Dropout with rate $p = 0.1$ is applied to the hidden activation during training. The same region-specific MLPs are applied at every time point, so any trajectory-dependent behavior must arise from the evolving state variables ($\mathbf{z}_t$, $\mathbf{U}_t$, $\mathbf{S}_t$, $\mathbf{H}_t$) rather than from time-varying parameters.

D.3 Region-Specific Readout

For each region r, the readout h_r is a region-specific two-layer feed-forward network with hidden dimension d_{rsr} that maps the masked buffer (Eq. (6)) to the next regional state:

$$\mathbf{z}_{t+1}[r] = \mathbf{w}_2^{(r)\top} \phi\big(\mathbf{W}_1^{(r)}(\mathbf{w}^{(r)} \odot \mathbf{U}_t[r, :]) + \mathbf{b}_1^{(r)}\big) + \mathbf{b}_2^{(r)}, \tag{12}$$

where ϕ is the GELU non-linearity [Hendrycks and Gimpel, 2016], and $\mathbf{W}_1^{(r)} \in \mathbb{R}^{d_{\text{rsr}} \times M_{\max}}$, $\mathbf{b}_1^{(r)} \in \mathbb{R}^{d_{\text{rsr}}}$, $\mathbf{w}_2^{(r)} \in \mathbb{R}^{d_{\text{rsr}}}$, and $\mathbf{b}_2^{(r)} \in \mathbb{R}$ are private to region r. Dropout with rate $p = 0.1$ is applied to the hidden activation during training.

D.4 Classifier Heads

We use one of two classifier heads on top of $\mathbf{S}_T$, depending on the dataset and task (Appendix C).

BrainNetCNN-style head. Following BrainNetCNN [Kawahara et al., 2017], we use its two structured operations for connectivity matrices, namely edge-to-edge (E2E) and edge-to-node (E2N) convolutions, both of which respect the row/column structure of the input. In our implementation, $\mathbf{S}_T \in \mathbb{R}^{R\times R}$ is processed by an E2E layer with 8 filters, followed by an E2N layer with 16 filters, a fully-connected layer with hidden dimension 64, and a final linear projection to C class logits, in place of the original node-to-graph layer. Dropout with probability $p = 0.1$ is applied after the fully-connected hidden layer. We use Leaky ReLU activations with negative slope 0.1.

Multi-layer perceptron. We extract the strict upper triangle of $\mathbf{S}_T$ (omitting the diagonal) to obtain a vector of dimension $R(R-1)/2$, exploiting the symmetry of the functional connectivity (FC) matrix to halve the input size relative to a full flatten. This vector is processed by a 3-layer MLP with hidden widths $(1024, 256)$ and ReLU activations, producing C class logits, with dropout ($p = 0.1$) applied after each hidden layer.

D.5 Classification Loss

Single-label classification (ABIDE, COBRE). $\bar{\mathbf{y}}$ is the softmax-activated probability over C classes and $\mathcal{L}_{\text{cls}}$ is the standard categorical cross-entropy:

$$\mathcal{L}_{\text{cls}} = -\sum_{c=1}^{C} \mathbf{y}_c \log \bar{\mathbf{y}}_c. \tag{13}$$

Multi-label classification (NSD). Each $\bar{\mathbf{y}}_c$ is the independent sigmoid-activated probability for class c, and $\mathcal{L}_{\text{cls}}$ is a class-balanced binary cross-entropy:

$$\mathcal{L}_{\text{cls}} = -\frac{1}{C}\sum_{c=1}^{C} \big[\, w_c\, \mathbf{y}_c \log \bar{\mathbf{y}}_c + (1-\mathbf{y}_c)\log(1-\bar{\mathbf{y}}_c)\,\big], \qquad w_c = \frac{1}{\sqrt{f_c+\epsilon}}, \tag{14}$$

where $f_c \in (0, 1]$ is the positive frequency of class c in the training set and ϵ is a small constant for numerical stability. The weight w_c is applied only to the positive term, so that positive samples of rarer classes receive larger weights, addressing the long-tailed distribution of COCO supercategories in NSD.

E Sensitivity Analysis: $M_{\max}$ Choice

We examine the sensitivity of CoRTeX to the maximum integration window $M_{\max}$ on COBRE by sweeping $M_{\max} \in \{5, 10, 15, 20, 25, 30\}$. Performance peaks at $M_{\max} = 20$ (used in the main experiments), with a secondary peak at $M_{\max} = 5$ and lower values at intermediate settings (Table A8).

Table A8: **Sensitivity to $M_{\max}$ on COBRE. Values are mean±std across five cross-validation folds.** The sweep uses a different random seed from the main experiments, so the result at $M_{\max} = 20$ differs slightly from Table 2.

$M_{\max}$	5	10	15	20†	25	30
ACC	0.585 ± 0.062	0.619 ± 0.123	0.524 ± 0.047	0.696 ± 0.102	0.614 ± 0.093	0.558 ± 0.094
AUC	0.707 ± 0.118	0.687 ± 0.166	0.604 ± 0.079	0.775 ± 0.135	0.644 ± 0.166	0.608 ± 0.118
SEN	0.688 ± 0.252	0.714 ± 0.290	0.266 ± 0.255	0.718 ± 0.134	0.492 ± 0.345	0.517 ± 0.372
SPEC	0.489 ± 0.209	0.520 ± 0.296	0.773 ± 0.229	0.677 ± 0.157	0.731 ± 0.209	0.600 ± 0.386

† Setting used in the main experiments.

Interpretation of the non-monotonic pattern. Given the fold-level standard deviations of AUC (0.08 to 0.17), differences among these settings should be interpreted with caution.

Practical choice and limitation. $M_{\max}$ is a user-set hyperparameter rather than a learned quantity. For task-fMRI with short trial windows ($T = 6$ on NSD), $M_{\max} = T$ is a natural choice, fixed by the trial structure rather than tuned for accuracy. For longer time series, $M_{\max}$ needs to be selected empirically. In our experiments, $M_{\max} = 20$ was used for ABIDE and COBRE, consistent with the sensitivity sweep reported in Table A8. Removing the need to specify a fixed $M_{\max}$—for example, by parameterizing M_r directly without an explicit upper bound, or by allowing the buffer length to grow adaptively during training—is a natural direction for future work.

F Positional Encoding in the LSTM Baseline

We equip the LSTM with CoRTeX's positional encoding (PE; one-hot ROI identity and NeRF-style coordinates). Without ROI-wise attention, we test three PE injection strategies, each using a shared nonlinear projection with normalization: (i) concatenating BOLD and PE before projection to one scalar per ROI, (ii) adding projected PE to BOLD, and (iii) concatenating BOLD with scalar-compressed PE before projection. We report the best variant per dataset alongside a parameter-matched LSTM. Table A9 shows a small gain on NSD but lower AUC on ABIDE and COBRE.

Table A9: **Effect of positional encoding on the LSTM baseline.** NSD reports macro-AUC; ABIDE and COBRE report AUC. The LSTM column comes from a separate run, so its values differ slightly from Tables 1 and 2.

Dataset	LSTM	LSTM + PE (best)	Δ	Parameter-matched LSTM
NSD	0.787±0.034	0.792±0.031	$+0.005$	0.788±0.032
ABIDE	0.654±0.044	0.620±0.029	-0.033	0.645±0.031
COBRE	0.578±0.161	0.487±0.082	-0.092	0.523±0.107

G Relocating Learned Temporal Windows (NSD)

We test whether CoRTeX relies on which ROI receives which window, rather than on the marginal distribution of $\{M_r\}$ alone. For each of the 40 trained NSD checkpoints (8 subjects $\times$ 5 seeds), we evaluate test macro-AUC without retraining under three configurations: the learned M_r, a *permuted* configuration that reassigns the learned values across ROIs, and a *uniform* configuration that sets every M_r to its mean. Relocating the windows reduces performance below even the uniform configuration (Tables A10 and A11), indicating that the trained model depends on the assignment of learned timescales to specific regions. Because the windows are relocated without retraining, this tests the trained model's reliance on its assignment rather than whether a different assignment could be learned equally well.

Table A10: **Macro-AUC under each M_r configuration** (mean±std over 40 checkpoints).

M_r configuration	Macro-AUC
Learned	0.849±0.022
Uniform at the mean	0.829±0.024
Permuted across ROIs	0.819±0.023

Table A11: **Paired comparisons across 40 checkpoints.** "Checkpoints" counts those satisfying each comparison.

Comparison	Checkpoints	ΔMacro-AUC	p (Wilcoxon)
Learned > permuted	40/40	$+0.030 \pm 0.009$	1.8×10^{-12}
Uniform > permuted	39/40	$+0.009 \pm 0.006$	6.0×10^{-11}

H Sensitivity to Diversity Regularization (NSD)

We sweep α_{div} from 0 to 1 on NSD. The standard deviation of M_r across ROIs (ROI-SD) more than doubles across the range, while macro-AUC varies by less than 0.005, suggesting that the regularizer neither degrades nor substantially improves predictive performance. Since $\text{Var}_r(M_r)$ is invariant under permutation of the ROI index, the regularizer encourages windows to differ across regions without specifying which regions should have shorter or longer windows. Consistent with this, the per-ROI ranking is largely preserved relative to $\alpha_{\text{div}} = 0$ (Table A12). Given the limited differences in predictive performance, we use $\alpha_{\text{div}} = 0.01$ across all three datasets.

Table A12: **Diversity regularization sweep on NSD.** ROI-SD denotes the standard deviation of M_r across 360 ROIs, and ρ denotes the Spearman rank correlation between the per-ROI M_r maps at each α_{div} and at $\alpha_{\text{div}} = 0$. All values are averaged over 8 subjects for each seed. The main experiments use $\alpha_{\text{div}} = 0.01$.

	Seed 20260504			Seed 42		
α_{div}	ROI-SD	ρ	Macro-AUC	ROI-SD	ρ	Macro-AUC
0	1.213±0.005	1.000±0.000	0.846	1.202±0.004	1.000±0.000	0.848
0.001	1.438±0.032	0.977±0.007	0.846	1.411±0.015	0.983±0.004	0.848
0.01	1.974±0.112	0.946±0.014	0.847	1.942±0.096	0.948±0.012	0.851
1.0	2.990±0.005	0.872±0.013	0.848	2.986±0.005	0.871±0.011	0.850

I Post-Hoc Temporal Attribution of the LSTM Baseline (NSD)

Because the LSTM receives ROI-indexed inputs, ROI-by-time attribution can in principle be extracted post hoc. We applied SmoothGrad [Smilkov et al., 2017] to the LSTM baseline and to CoRTeX on NSD (8 subjects $\times$ 5 seeds each) with identical settings. Each test input $\mathbf{x} \in \mathbb{R}^{R\times T}$ ($R = 360, T = 6$) is a repetition-averaged trial, in which the BOLD responses to repeated presentations of the same image are averaged to reduce trial-level noise. For each such trial and each of its positive classes c, the attribution is $A_c[r,t] = \frac{1}{K}\sum_{k=1}^{K}\left|\partial f_c(\mathbf{x}+\boldsymbol{\epsilon}_k)/\partial x_{r,t}\right|$, where f_c is the class-c logit, $K = 32$, and $\boldsymbol{\epsilon}_k \sim \mathcal{N}(0, \sigma^2\mathbf{I})$ with $\sigma = 0.15\,(\max \mathbf{x} - \min \mathbf{x})$ set per trial, within the 10–20% range recommended by Smilkov et al. [2017]. Attributions are averaged over the trial's positive classes, normalized over the T volumes for each ROI, and averaged over test trials, yielding a temporal profile $p_r(t)$ with $\sum_t p_r(t) = 1$. We summarize each profile by its effective span $s_r = 1/\sum_t p_r(t)^2 \in [1, T]$ and quantify regional differentiation as the standard deviation of s_r across ROIs, normalized by T, the maximum possible span. Averaged over 8 subjects and 5 seeds, the normalized between-ROI standard deviation is 4.07% for the LSTM and 16.31% for CoRTeX (higher for CoRTeX in every subject), indicating that the LSTM's post-hoc temporal profiles are largely shared across ROIs. This comparison concerns how separable the regional temporal profiles are, not their biological fidelity.

J Per-ROI Learned Timescales (NSD)

The following tables list the learned per-region timescales M_r for all 360 HCP-MMP1 ROIs [Glasser et al., 2016] on NSD, together with their Yeo-7 network [Yeo et al., 2011], assigned at the atlas level by majority vote over fsaverage surface vertices (mean purity 0.83). M_{mean} and M_{std} are computed across 8 NSD subjects (5 random seeds each, then averaged within subject). The *Regime* column gives the Gaussian mixture model (GMM) cluster assignment (short/mid/long) reported in §4.5. Table A13 summarizes M_r by Yeo-7 network, and the per-ROI values are split by hemisphere: Table A14 (left, ROI 0–179) and Table A15 (right, ROI 180–359).

Table A13: **Learned M_r on NSD summarized by Yeo-7 network** (mean and variance across ROIs). Per-ROI values are listed in Tables A14 and A15.

Yeo-7 network	n	Mean M_r	Variance M_r
Visual	54	2.88	3.05
Somatomotor	54	3.42	4.46
DorsalAttn	46	2.99	3.59
VentralAttn	48	2.86	3.35
Limbic	29	3.24	3.76
Frontoparietal	44	2.88	3.50
Default	85	3.51	3.81
All ROIs	360	3.15	3.74

Table A14: **Per-ROI learned timescales M_r on NSD (left hemisphere).** See the introductory paragraph above for definitions. `L_` prefix denotes the left hemisphere.

Idx	ROI	Yeo-7 network	M_{mean}	M_{std}	Regime	Idx	ROI	Yeo-7 network	M_{mean}	M_{std}	Regime
0	L_V1	Visual	1.483	0.070	short	90	L_11l	Frontoparietal	0.446	0.032	short
1	L_MST	DorsalAttn	0.672	0.074	short	91	L_13l	Limbic	5.998	0.003	long
2	L_V6	Visual	4.254	0.918	mid	92	L_OFC	Limbic	5.348	0.480	long
3	L_V2	Visual	0.800	0.209	short	93	L_47s	Default	5.938	0.115	long
4	L_V3	Visual	5.733	0.452	long	94	L_LIPd	DorsalAttn	2.584	0.067	short
5	L_V4	Visual	1.366	0.255	short	95	L_6a	DorsalAttn	1.447	0.292	short
6	L_V8	Visual	4.994	0.603	mid	96	L_i6-8	Default	1.443	0.491	short
7	L_4	Somatomotor	1.341	0.302	short	97	L_s6-8	Default	5.733	0.208	long
8	L_3b	Somatomotor	1.077	0.375	short	98	L_43	VentralAttn	1.345	0.363	short
9	L_FEF	DorsalAttn	0.655	0.087	short	99	L_OP4	Somatomotor	5.998	0.003	long
10	L_PEF	DorsalAttn	5.887	0.180	long	100	L_OP1	Somatomotor	5.574	0.541	long
11	L_55b	Somatomotor	0.431	0.150	short	101	L_OP2-3	Somatomotor	5.864	0.184	long
12	L_V3A	Visual	1.251	0.338	short	102	L_52	Somatomotor	1.327	0.346	short
13	L_RSC	Default	5.877	0.162	long	103	L_RI	Somatomotor	0.881	0.268	short
14	L_POS2	Default	1.531	0.089	short	104	L_PFcm	VentralAttn	2.620	0.118	short
15	L_V7	Visual	1.348	0.241	short	105	L_PoI2	VentralAttn	3.668	0.435	mid
16	L_IPS1	DorsalAttn	1.526	0.020	short	106	L_TA2	Somatomotor	4.025	0.970	mid
17	L_FFC	Visual	0.601	0.047	short	107	L_FOP4	VentralAttn	1.496	0.211	short
18	L_V3B	Visual	0.836	0.337	short	108	L_MI	VentralAttn	0.591	0.136	short
19	L_LO1	Visual	4.907	0.976	mid	109	L_Pir	Limbic	1.428	0.057	short
20	L_LO2	Visual	0.599	0.026	short	110	L_AVI	Frontoparietal	4.110	0.799	mid
21	L_PIT	Visual	5.103	0.544	mid	111	L_AAIC	VentralAttn	4.262	1.038	mid
22	L_MT	Visual	0.612	0.062	short	112	L_FOP1	VentralAttn	0.938	0.316	short
23	L_A1	Somatomotor	5.758	0.336	long	113	L_FOP3	VentralAttn	1.540	0.040	short
24	L_PSL	VentralAttn	5.448	0.727	long	114	L_FOP2	Somatomotor	0.596	0.029	short
25	L_SFL	Default	5.763	0.202	long	115	L_PFt	DorsalAttn	1.187	0.287	short
26	L_PCV	Default	1.544	0.074	short	116	L_AIP	DorsalAttn	5.734	0.175	long
27	L_STV	Default	5.325	0.660	long	117	L_EC	Limbic	5.431	0.528	long
28	L_7Pm	Frontoparietal	5.673	0.385	long	118	L_PreS	Default	4.993	0.435	mid
29	L_7m	Default	2.302	0.392	short	119	L_H	Limbic	0.519	0.210	short
30	L_POS1	Default	1.152	0.387	short	120	L_ProS	Visual	5.396	0.637	long
31	L_23d	Default	5.993	0.008	long	121	L_PeEc	Limbic	1.400	0.321	short
32	L_v23ab	Default	3.800	0.565	mid	122	L_STGa	Default	4.947	0.522	mid
33	L_d23ab	Default	1.343	0.385	short	123	L_PBelt	Somatomotor	0.637	0.042	short
34	L_31pv	Default	0.948	0.349	short	124	L_A5	Default	5.855	0.189	long
35	L_5m	Somatomotor	1.337	0.999	short	125	L_PHA1	Default	4.511	0.089	mid
36	L_5mv	VentralAttn	1.247	0.295	short	126	L_PHA3	Visual	3.665	0.584	mid
37	L_23c	VentralAttn	1.437	0.041	short	127	L_STSda	Default	0.555	0.032	short
38	L_5L	Somatomotor	1.417	0.252	short	128	L_STSdp	Default	5.499	0.466	long
39	L_24dd	Somatomotor	5.819	0.196	long	129	L_STSvp	Default	0.391	0.032	short
40	L_24dv	Somatomotor	1.111	0.456	short	130	L_TGd	Limbic	1.401	0.314	short
41	L_7AL	DorsalAttn	2.421	0.333	short	131	L_TE1a	Default	2.473	0.121	short
42	L_SCEF	VentralAttn	5.923	0.111	long	132	L_TE1p	Default	5.866	0.178	long
43	L_6ma	VentralAttn	5.714	0.224	long	133	L_TE2a	Limbic	1.945	0.586	short
44	L_7Am	DorsalAttn	2.628	0.063	short	134	L_TF	Limbic	5.464	0.655	long
45	L_7PL	DorsalAttn	5.390	0.529	long	135	L_TE2p	DorsalAttn	1.417	0.226	short
46	L_7PC	DorsalAttn	4.030	0.768	mid	136	L_PHT	DorsalAttn	4.792	0.702	mid
47	L_LIPv	DorsalAttn	2.131	0.370	short	137	L_PH	DorsalAttn	1.389	0.322	short
48	L_VIP	DorsalAttn	5.833	0.174	long	138	L_TPOJ1	Somatomotor	5.710	0.159	long

(continued on next page)

(continued from previous page)

Idx	ROI	Yeo-7 network	M_{mean}	M_{std}	Regime	Idx	ROI	Yeo-7 network	M_{mean}	M_{std}	Regime
49	L_MIP	DorsalAttn	3.464	0.046	mid	139	L_TPOJ2	VentralAttn	5.263	0.371	long
50	L_1	Somatomotor	5.651	0.172	long	140	L_TPOJ3	DorsalAttn	5.634	0.218	long
51	L_2	Somatomotor	0.572	0.071	short	141	L_DVT	Visual	1.899	0.357	short
52	L_3a	Somatomotor	5.993	0.008	long	142	L_PGp	DorsalAttn	5.739	0.227	long
53	L_6d	Somatomotor	3.570	0.097	mid	143	L_IP2	Frontoparietal	5.960	0.081	long
54	L_6mp	Somatomotor	3.790	0.494	mid	144	L_IP1	Frontoparietal	1.761	0.407	short
55	L_6v	Somatomotor	2.473	0.206	short	145	L_IP0	DorsalAttn	1.333	0.267	short
56	L_p24pr	VentralAttn	2.561	0.112	short	146	L_PFop	VentralAttn	3.419	0.047	mid
57	L_33pr	Frontoparietal	1.311	0.402	short	147	L_PF	VentralAttn	1.713	0.501	short
58	L_a24pr	VentralAttn	1.511	0.097	short	148	L_PFm	Frontoparietal	2.222	0.376	short
59	L_p32pr	VentralAttn	1.560	0.072	short	149	L_PGi	Default	5.327	0.418	long
60	L_a24	Default	0.595	0.055	short	150	L_PGs	Default	1.389	0.229	short
61	L_d32	Default	0.955	0.332	short	151	L_V6A	Visual	5.644	0.421	long
62	L_8BM	Frontoparietal	1.409	0.283	short	152	L_VMV1	Visual	3.569	0.305	mid
63	L_p32	Default	2.111	0.367	short	153	L_VMV3	Visual	4.640	0.519	mid
64	L_10r	Default	2.544	0.117	short	154	L_PHA2	Default	4.845	0.582	mid
65	L_47m	Default	1.609	0.048	short	155	L_V4t	Visual	1.487	0.102	short
66	L_8Av	Default	1.406	0.210	short	156	L_FST	DorsalAttn	2.046	0.650	short
67	L_8Ad	Default	5.953	0.067	long	157	L_V3CD	Visual	5.777	0.173	long
68	L_9m	Default	5.391	0.457	long	158	L_LO3	Visual	2.293	0.241	short
69	L_8BL	Default	2.565	0.095	short	159	L_VMV2	Visual	4.420	0.055	mid
70	L_9p	Default	1.238	0.300	short	160	L_31pd	Default	4.325	0.896	mid
71	L_10d	Default	4.607	0.853	mid	161	L_31a	Default	1.818	1.091	short
72	L_8C	Frontoparietal	0.768	0.251	short	162	L_VVC	Visual	1.482	0.057	short
73	L_44	Default	5.513	0.397	long	163	L_25	Limbic	2.542	0.109	short
74	L_45	Default	1.427	0.294	short	164	L_s32	Limbic	5.395	0.531	long
75	L_47l	Default	5.867	0.234	long	165	L_pOFC	Limbic	5.575	0.402	long
76	L_a47r	Default	5.755	0.199	long	166	L_PoI1	VentralAttn	1.645	0.168	short
77	L_6r	VentralAttn	1.753	0.871	short	167	L_Ig	Somatomotor	1.445	0.106	short
78	L_IFJa	Frontoparietal	3.833	0.823	mid	168	L_FOP5	VentralAttn	1.833	0.409	short
79	L_IFJp	Frontoparietal	5.616	0.248	long	169	L_p10p	Default	4.416	0.918	mid
80	L_IFSp	Frontoparietal	3.539	0.093	mid	170	L_p47r	Frontoparietal	1.149	0.485	short
81	L_IFSa	Frontoparietal	2.360	0.354	short	171	L_TGv	Limbic	5.845	0.197	long
82	L_p9-46v	Frontoparietal	1.295	0.324	short	172	L_MBelt	Somatomotor	5.811	0.140	long
83	L_46	Frontoparietal	4.109	0.955	mid	173	L_LBelt	Somatomotor	1.200	0.443	short
84	L_a9-46v	Frontoparietal	5.256	0.589	long	174	L_A4	Somatomotor	5.024	0.515	mid
85	L_9-46d	Frontoparietal	0.587	0.017	short	175	L_STSva	Default	3.459	0.042	mid
86	L_9a	Default	5.945	0.128	long	176	L_TE1m	Default	4.344	0.935	mid
87	L_10v	Limbic	1.515	0.054	short	177	L_PI	VentralAttn	1.530	0.037	short
88	L_a10p	Default	5.841	0.176	long	178	L_a32pr	Frontoparietal	0.590	0.047	short
89	L_10pp	Limbic	5.841	0.202	long	179	L_p24	Default	4.337	0.787	mid

Table A15: **Per-ROI learned timescales M_r on NSD (right hemisphere).** See the introductory paragraph above for definitions. `R_` prefix denotes the right hemisphere.

Idx	ROI	Yeo-7 network	M_{mean}	M_{std}	Regime	Idx	ROI	Yeo-7 network	M_{mean}	M_{std}	Regime
180	R_V1	Visual	0.571	0.041	short	270	R_11l	Frontoparietal	0.535	0.100	short
181	R_MST	Visual	4.402	0.030	mid	271	R_13l	Limbic	3.435	0.054	mid
182	R_V6	Visual	3.741	0.693	mid	272	R_OFC	Limbic	1.401	0.320	short
183	R_V2	Visual	3.750	0.775	mid	273	R_47s	Default	2.253	0.436	short
184	R_V3	Visual	1.431	0.169	short	274	R_LIPd	DorsalAttn	1.441	0.330	short
185	R_V4	Visual	1.350	0.356	short	275	R_6a	DorsalAttn	1.315	0.385	short
186	R_V8	Visual	4.637	0.517	mid	276	R_i6-8	Frontoparietal	5.979	0.035	long
187	R_4	Somatomotor	1.457	0.098	short	277	R_s6-8	Default	5.221	0.518	long
188	R_3b	Somatomotor	2.579	0.032	short	278	R_43	Somatomotor	1.326	0.356	short
189	R_FEF	DorsalAttn	5.890	0.165	long	279	R_OP4	Somatomotor	3.425	0.102	mid
190	R_PEF	DorsalAttn	3.563	0.073	mid	280	R_OP1	Somatomotor	3.876	0.803	mid
191	R_55b	VentralAttn	1.341	0.280	short	281	R_OP2-3	Somatomotor	3.975	0.821	mid
192	R_V3A	Visual	1.084	0.379	short	282	R_52	Somatomotor	5.942	0.111	long
193	R_RSC	Default	4.696	0.981	mid	283	R_RI	Somatomotor	5.813	0.378	long
194	R_POS2	Frontoparietal	1.511	0.171	short	284	R_PFcm	VentralAttn	1.430	0.130	short
195	R_V7	Visual	1.516	0.060	short	285	R_PoI2	VentralAttn	4.785	0.284	mid
196	R_IPS1	DorsalAttn	1.537	0.050	short	286	R_TA2	Somatomotor	3.602	0.336	mid
197	R_FFC	Visual	1.083	0.424	short	287	R_FOP4	VentralAttn	4.352	0.764	mid
198	R_V3B	Visual	5.858	0.173	long	288	R_MI	VentralAttn	1.868	0.655	short

(continued on next page)

(continued from previous page)

Idx	ROI	Yeo-7 network	M_{mean}	M_{std}	Regime	Idx	ROI	Yeo-7 network	M_{mean}	M_{std}	Regime
199	R_LO1	Visual	0.611	0.026	short	289	R_Pir	Limbic	5.844	0.172	long
200	R_LO2	Visual	2.522	0.118	short	290	R_AVI	Frontoparietal	1.922	0.345	short
201	R_PIT	Visual	2.408	0.301	short	291	R_AAIC	VentralAttn	5.354	0.610	long
202	R_MT	Visual	4.452	0.058	mid	292	R_FOP1	VentralAttn	3.553	0.045	mid
203	R_A1	Somatomotor	1.319	0.427	short	293	R_FOP3	VentralAttn	2.614	0.101	short
204	R_PSL	VentralAttn	1.493	0.087	short	294	R_FOP2	Somatomotor	5.633	0.763	long
205	R_SFL	Default	5.931	0.132	long	295	R_PFt	DorsalAttn	0.551	0.035	short
206	R_PCV	DorsalAttn	0.977	0.306	short	296	R_AIP	DorsalAttn	2.005	0.415	short
207	R_STV	VentralAttn	5.148	0.668	mid	297	R_EC	Limbic	2.038	0.655	short
208	R_7Pm	Frontoparietal	3.825	0.822	mid	298	R_PreS	Visual	5.222	0.548	long
209	R_7m	Default	2.450	0.104	short	299	R_H	Limbic	2.482	0.234	short
210	R_POS1	Default	1.852	0.344	short	300	R_ProS	Visual	0.568	0.052	short
211	R_23d	Default	1.093	0.278	short	301	R_PeEc	Limbic	2.058	0.371	short
212	R_v23ab	Default	3.542	0.421	mid	302	R_STGa	Default	1.265	0.370	short
213	R_d23ab	Default	1.278	0.326	short	303	R_PBelt	Somatomotor	5.840	0.235	long
214	R_31pv	Default	0.933	0.428	short	304	R_A5	Somatomotor	5.658	0.435	long
215	R_5m	Somatomotor	1.366	0.271	short	305	R_PHA1	Visual	3.734	0.280	mid
216	R_5mv	VentralAttn	2.587	0.090	short	306	R_PHA3	Visual	3.523	0.051	mid
217	R_23c	VentralAttn	5.880	0.183	long	307	R_STSda	Default	5.194	0.501	mid
218	R_5L	Somatomotor	5.766	0.457	long	308	R_STSdp	Default	5.675	0.185	long
219	R_24dd	Somatomotor	1.527	0.060	short	309	R_STSvp	Default	1.313	0.334	short
220	R_24dv	Somatomotor	1.329	0.435	short	310	R_TGd	Limbic	4.147	0.385	mid
221	R_7AL	DorsalAttn	5.889	0.157	long	311	R_TE1a	Default	5.294	0.531	long
222	R_SCEF	VentralAttn	4.616	0.964	mid	312	R_TE1p	Frontoparietal	1.544	0.085	short
223	R_6ma	VentralAttn	5.883	0.172	long	313	R_TE2a	Limbic	0.485	0.066	short
224	R_7Am	DorsalAttn	3.856	0.673	mid	314	R_TF	Limbic	3.745	0.627	mid
225	R_7PL	DorsalAttn	1.151	0.420	short	315	R_TE2p	DorsalAttn	2.413	0.250	short
226	R_7PC	DorsalAttn	0.794	0.233	short	316	R_PHT	DorsalAttn	1.483	0.054	short
227	R_LIPv	DorsalAttn	5.999	0.002	long	317	R_PH	DorsalAttn	4.679	0.361	mid
228	R_VIP	DorsalAttn	2.357	0.789	short	318	R_TPOJ1	VentralAttn	0.616	0.038	short
229	R_MIP	DorsalAttn	5.360	0.443	long	319	R_TPOJ2	DorsalAttn	1.823	0.277	short
230	R_1	Somatomotor	5.956	0.095	long	320	R_TPOJ3	DorsalAttn	5.654	0.169	long
231	R_2	Somatomotor	3.738	0.215	mid	321	R_DVT	Visual	3.798	0.810	mid
232	R_3a	Somatomotor	5.720	0.169	long	322	R_PGp	DorsalAttn	2.310	0.325	short
233	R_6d	Somatomotor	5.759	0.410	long	323	R_IP2	Frontoparietal	2.202	0.450	short
234	R_6mp	Somatomotor	5.631	0.408	long	324	R_IP1	Frontoparietal	5.995	0.013	long
235	R_6v	Somatomotor	0.564	0.039	short	325	R_IP0	DorsalAttn	1.511	0.080	short
236	R_p24pr	VentralAttn	0.636	0.088	short	326	R_PFop	VentralAttn	5.792	0.157	long
237	R_33pr	VentralAttn	0.563	0.068	short	327	R_PF	VentralAttn	4.318	0.832	mid
238	R_a24pr	VentralAttn	1.129	0.346	short	328	R_PFm	Frontoparietal	3.767	0.845	mid
239	R_p32pr	VentralAttn	5.601	0.245	long	329	R_PGi	Default	5.821	0.184	long
240	R_a24	Default	1.500	0.097	short	330	R_PGs	Default	0.796	0.522	short
241	R_d32	Default	5.110	0.465	mid	331	R_V6A	Visual	2.387	0.322	short
242	R_8BM	Frontoparietal	1.366	0.303	short	332	R_VMV1	Visual	1.741	0.278	short
243	R_p32	Default	1.883	0.682	short	333	R_VMV3	Visual	1.558	0.072	short
244	R_10r	Default	4.388	1.009	mid	334	R_PHA2	Visual	4.322	0.998	mid
245	R_47m	Default	5.398	0.548	long	335	R_V4t	Visual	4.926	0.659	mid
246	R_8Av	Frontoparietal	0.526	0.046	short	336	R_FST	DorsalAttn	1.121	0.252	short
247	R_8Ad	Default	5.885	0.135	long	337	R_V3CD	Visual	0.582	0.043	short
248	R_9m	Default	5.921	0.131	long	338	R_LO3	Visual	3.399	0.069	mid
249	R_8BL	Default	3.962	0.838	mid	339	R_VMV2	Visual	4.436	0.070	mid
250	R_9p	Default	0.537	0.044	short	340	R_31pd	Default	5.945	0.070	long
251	R_10d	Default	0.528	0.116	short	341	R_31a	Default	4.312	1.012	mid
252	R_8C	Frontoparietal	4.405	1.126	mid	342	R_VVC	Visual	1.555	0.135	short
253	R_44	Frontoparietal	0.656	0.045	short	343	R_25	Limbic	1.409	0.260	short
254	R_45	Default	1.549	0.069	short	344	R_s32	Default	2.551	0.112	short
255	R_47l	Default	1.500	0.091	short	345	R_pOFC	Limbic	0.602	0.041	short
256	R_a47r	Frontoparietal	5.040	0.564	mid	346	R_PoI1	VentralAttn	5.209	0.467	long
257	R_6r	VentralAttn	1.554	0.405	short	347	R_Ig	Somatomotor	0.552	0.080	short
258	R_IFJa	Frontoparietal	3.480	0.050	mid	348	R_FOP5	VentralAttn	1.151	0.382	short
259	R_IFJp	DorsalAttn	5.849	0.128	long	349	R_p10p	Frontoparietal	2.344	0.338	short
260	R_IFSp	Frontoparietal	5.870	0.122	long	350	R_p47r	Frontoparietal	5.881	0.150	long
261	R_IFSa	Frontoparietal	1.434	0.243	short	351	R_TGv	Limbic	3.501	0.068	mid
262	R_p9-46v	Frontoparietal	2.005	0.501	short	352	R_MBelt	Somatomotor	5.952	0.086	long
263	R_46	Frontoparietal	5.953	0.093	long	353	R_LBelt	Somatomotor	4.518	0.961	mid
264	R_a9-46v	Frontoparietal	2.816	0.310	short	354	R_A4	Somatomotor	1.362	0.388	short
265	R_9-46d	Frontoparietal	2.435	0.176	short	355	R_STSva	Default	4.144	1.002	mid
266	R_9a	Default	4.070	0.744	mid	356	R_TE1m	Default	1.566	0.039	short

(continued on next page)

(continued from previous page)

Idx	ROI	Yeo-7 network	M_{mean}	M_{std}	Regime	Idx	ROI	Yeo-7 network	M_{mean}	M_{std}	Regime
267	R_10v	Limbic	5.532	0.415	long	357	R_PI	VentralAttn	0.617	0.100	short
268	R_a10p	Frontoparietal	1.139	0.363	short	358	R_a32pr	Frontoparietal	1.902	0.522	short
269	R_10pp	Limbic	1.497	0.050	short	359	R_p24	Default	5.839	0.196	long

K Regional Temporal Heterogeneity on rs-fMRI

Unlike NSD, the M_r distributions on both rs-fMRI datasets (ABIDE and COBRE) are unimodal. This difference may reflect the absence of event-locked trial structure in rs-fMRI, although $M_{\max}$ and sequence length also differ between settings. To examine network-level structure while maintaining comparability with the NSD analysis (§4.5), we partition the 200 Schaefer ROIs into three timescale groups using quantile splits of the per-ROI mean M_r (5-fold mean). These partitions are analytic conveniences rather than data-driven clusters.

K.1 ABIDE

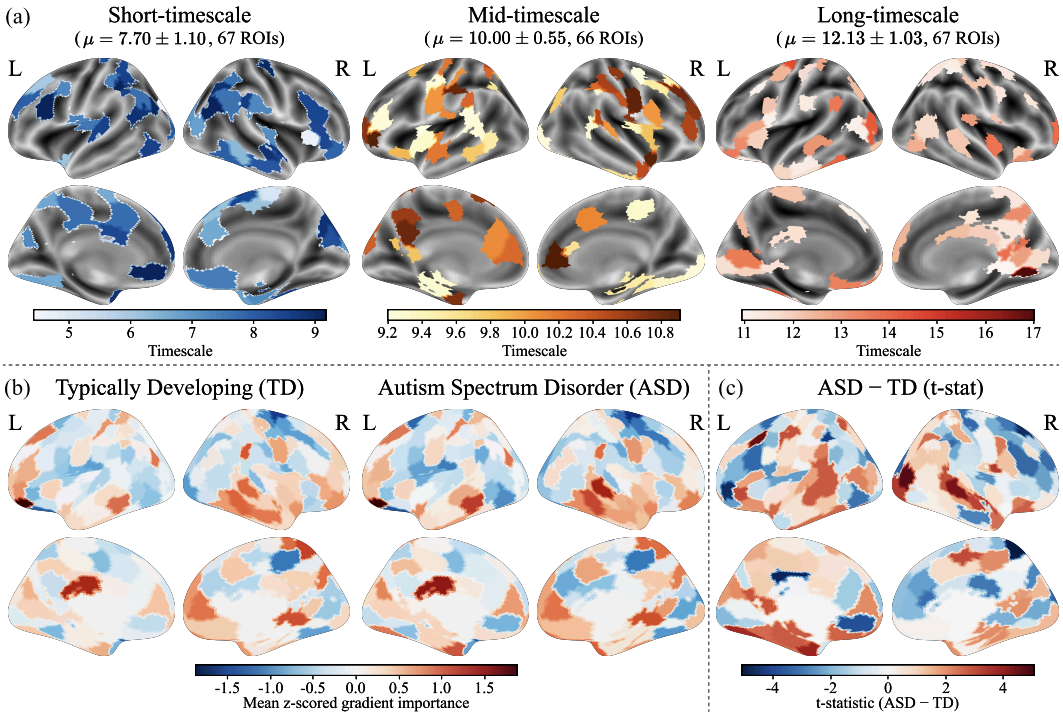


Figure A1: **Timescale organization and group-discriminative regions in ABIDE.** (a) Per-ROI M_r timescale partitions cortical regions into short ($n = 67$, $\mu = 7.70 \pm 1.10$), mid ($n = 66$, $\mu = 10.00 \pm 0.55$), and long ($n = 67$, $\mu = 12.13 \pm 1.03$) timescale groups (quantile-based; representative ROIs in Table A17). The long-timescale tier includes transmodal (Default 21%, Cont 15%) and sensory (SomMot 19%, Vis 16%) ROIs in roughly equal shares, close to their atlas-wide proportions. (b) Subject-level gradient importance (z-scored within each subject) for correctly classified subjects ($n = 720$; 342 ASD, 378 TD), averaged within group. (c) Between-group t-statistic (ASD−TD). 73 ROIs reach uncorrected $p < 0.05$ ($|t| > 1.96$) and 43 reach $p < 0.01$. Of the 73, 45 survive BH–FDR $q < 0.05$ and 9 survive Bonferroni $p < 0.05/200$ (Tables A20, A16). Among the BH–FDR survivors, the discriminative signal spans multiple networks with prominent Cont (13) and Default (11) involvement, followed by SomMot (8) and Vis (5), with smaller contributions from DorsAttn, Limbic, and VentAttn (≤ 3 each). The strongest effect appears at right dorsal-attention postcentral region (RH_DorsAttnB_PostC_3, $t = -5.09$, $p < 10^{-6}$, $q_{\text{BH}} < 10^{-3}$).

Table A16: All 45 ROIs surviving BH–FDR $q < 0.05$ on ABIDE (correctly classified subjects only, $n = 720$; ASD $n = 342$, TD $n = 378$). t is the two-sample t-statistic for z-scored gradient saliency, ASD$-$TD; positive t indicates higher importance for ASD. All listed ROIs satisfy BH–FDR $q < 0.05$ (†); †† marks the 9 ROIs that additionally survive Bonferroni $p < 0.05/200$.

#	ROI	Network	Hemi	t	p	q_{BH}	p_{Bonf}	sig
1	RH_DorsAttnB_PostC_3	DorsAttn	R	-5.09	$< 10^{-4}$	0.0001	0.0001	††
2	RH_VisCent_ExStr_4	Vis	R	$+4.74$	$< 10^{-4}$	0.0002	0.0005	††
3	RH_ContB_PFCld_3	Cont	R	-4.65	$< 10^{-4}$	0.0002	0.0008	††
4	LH_DefaultA_pCunPCC_2	Default	L	-4.62	$< 10^{-4}$	0.0002	0.0009	††
5	LH_ContB_PFClv_1	Cont	L	-4.56	$< 10^{-4}$	0.0002	0.0012	††
6	LH_ContB_PFCl_1	Cont	L	$+4.50$	$< 10^{-4}$	0.0003	0.0016	††
7	RH_TempPar_2	Default	R	$+4.37$	$< 10^{-4}$	0.0004	0.0028	††
8	LH_ContA_IPS_2	Cont	L	-4.19	$< 10^{-4}$	0.0008	0.0064	††
9	LH_LimbicA_TempPole_1	Limbic	L	$+3.76$	0.0002	0.0042	0.0374	††
10	LH_DefaultA_PFCm_1	Default	L	-3.66	0.0003	0.0054	0.0536	†
11	LH_VisCent_ExStr_1	Vis	L	$+3.61$	0.0003	0.0060	0.0657	†
12	RH_VisCent_ExStr_5	Vis	R	-3.38	0.0008	0.0117	0.1522	†
13	RH_SomMotB_S2_3	SomMot	R	$+3.36$	0.0008	0.0117	0.1658	†
14	RH_ContA_PFCl_2	Cont	R	-3.33	0.0009	0.0117	0.1798	†
15	RH_ContB_IPL_2	Cont	R	-3.33	0.0009	0.0117	0.1839	†
16	RH_VisCent_ExStr_3	Vis	R	$+3.32$	0.0009	0.0117	0.1880	†
17	RH_SomMotA_4	SomMot	R	-3.25	0.0012	0.0142	0.2418	†
18	RH_SomMotA_11	SomMot	R	-3.21	0.0014	0.0152	0.2742	†
19	RH_TempPar_1	Default	R	$+3.19$	0.0015	0.0155	0.2973	†
20	LH_ContA_PFCl_1	Cont	L	-3.18	0.0016	0.0155	0.3107	†
21	LH_ContA_PFCd_1	Cont	L	-3.16	0.0016	0.0156	0.3267	†
22	RH_DefaultA_PFCm_2	Default	R	-3.14	0.0018	0.0162	0.3569	†
23	RH_SalVentAttnB_PFCl_1	VentAttn	R	-3.12	0.0019	0.0162	0.3735	†
24	RH_SomMotA_3	SomMot	R	$+3.06$	0.0023	0.0192	0.4617	†
25	RH_TempPar_3	Default	R	$+2.99$	0.0029	0.0224	0.5859	†
26	RH_ContA_PFCd_1	Cont	R	-2.97	0.0031	0.0224	0.6105	†
27	LH_DefaultB_PFCv_3	Default	L	$+2.96$	0.0032	0.0224	0.6359	†
28	RH_ContC_Cingp_1	Cont	R	-2.96	0.0032	0.0224	0.6375	†
29	LH_DorsAttnA_TempOcc_2	DorsAttn	L	-2.95	0.0032	0.0224	0.6495	†
30	LH_VisCent_ExStr_5	Vis	L	-2.94	0.0034	0.0226	0.6777	†
31	LH_SomMotB_Cent_2	SomMot	L	$+2.88$	0.0041	0.0265	0.8230	†
32	RH_ContB_Temp_1	Cont	R	-2.84	0.0047	0.0292	0.9351	†
33	LH_DorsAttnA_SPL_3	DorsAttn	L	$+2.81$	0.0051	0.0306	1.0000	†
34	LH_DefaultC_Rsp_1	Default	L	$+2.80$	0.0052	0.0306	1.0000	†
35	RH_ContC_pCun_1	Cont	R	-2.78	0.0055	0.0316	1.0000	†
36	LH_SomMotB_Aud_3	SomMot	L	$+2.74$	0.0063	0.0345	1.0000	†
37	LH_TempPar_1	Default	L	$+2.74$	0.0064	0.0345	1.0000	†
38	LH_DefaultB_Temp_4	Default	L	$+2.71$	0.0069	0.0359	1.0000	†
39	RH_LimbicB_OFC_2	Limbic	R	$+2.70$	0.0070	0.0359	1.0000	†
40	LH_DefaultC_PHC_1	Default	L	$+2.66$	0.0079	0.0395	1.0000	†
41	RH_ContA_IPS_2	Cont	R	-2.65	0.0083	0.0405	1.0000	†
42	LH_SalVentAttnA_FrOper_2	VentAttn	L	-2.61	0.0091	0.0435	1.0000	†
43	RH_SomMotB_S2_1	SomMot	R	-2.60	0.0094	0.0438	1.0000	†
44	LH_SomMotA_8	SomMot	L	-2.57	0.0104	0.0471	1.0000	†
45	LH_LimbicB_OFC_2	Limbic	L	$+2.55$	0.0110	0.0489	1.0000	†

We extend the rs-fMRI timescale analysis to ABIDE by linking the learned timescale organization to the autism-discriminative signal at the subject level, and contrast the resulting spatial pattern with the schizophrenia-discriminative signal on COBRE (§K.2, presented after ABIDE for narrative clarity since ABIDE provides the primary false discovery rate (FDR)-controlled evidence).

Quantile-based timescales. The M_r distribution on ABIDE is also unimodal (Hartigan's dip test $p = 0.92$, BIC favors $k = 1$, $\Delta\text{BIC}_{2-1} = +16.2$). Quantile splits yield short ($M_r \in [4.41, 9.19]$, $n = 67$), mid ($M_r \in [9.19, 10.91]$, $n = 66$), and long ($M_r \in [10.92, 17.02]$, $n = 67$) groups.

Timescale groups span sensory and transmodal networks. Mapping each timescale group to Yeo-7 networks shows that the long-timescale group contains transmodal (Default 21%, Cont 15%) and sensory (SomMot 19%, Vis 16%) ROIs in roughly equal shares, close to their atlas-wide proportions (Default 21.5%, Cont 18.5%, SomMot 17%, Vis 12%). Unlike COBRE, where control-network ROIs are over-represented in the long-timescale group, the ABIDE organization therefore shows no clear alignment with the canonical sensorimotor-to-transmodal timescale gradient [Murray et al., 2014, Margulies et al., 2016, Hasson et al., 2008]. The short-timescale group is likewise heterogeneous (Cont 22%, Default 19%, Vis 13%, VentAttn 13%) rather than purely sensory, indicating that the model assigns fast dynamics to a varied set of regions across networks (Fig. A1a). This lack of

clear separation may reflect the heterogeneity of the multisite ABIDE cohort and the presence of long-timescale visual ROIs (*e.g.,* RH_VisPeri_ExStrInf_1, $M_r = 17.02$).

Subject-level gradient importance spans multiple networks. Among correctly classified subjects ($n = 720$), two-sample t-tests (ASD vs. TD) per ROI identified 73 ROIs at uncorrected $p < 0.05$, of which 45 survive Benjamini–Hochberg FDR correction at $q < 0.05$ and 9 survive Bonferroni $p < 0.05/200$ (Tables A20, A16). Among the FDR survivors, the discriminative signal spans multiple networks with prominent Cont (13, $\sim 29\%$) and Default (11, $\sim 24\%$) involvement, followed by SomMot (8) and Vis (5), with smaller contributions from DorsAttn, Limbic, and VentAttn (≤ 3 each). This contrasts with the exploratory default mode network (DMN)-leaning pattern on COBRE (4 of 8 nominally significant ROIs in DMN; §K.2). The strongest effect appears at right dorsal-attention postcentral region (RH_DorsAttnB_PostC_3, $t = -5.09$, $p < 10^{-6}$, $q_{\text{BH}} < 10^{-3}$, reduced importance in ASD), with additional strong effects in visual cortex (RH_VisCent_ExStr_4, $t = +4.74$, increased), prefrontal Cont (RH_ContB_PFCld_3, $t = -4.65$), and posterior cingulate DMN (LH_DefaultA_pCunPCC_2, $t = -4.62$). This distributed pattern, with the largest shares in control and default mode networks, is consistent with reports of distributed ASD-related connectivity alterations, dominated by cortico-cortical and interhemispheric hypoconnectivity [Di Martino et al., 2014].

Table A17: **Top-3 representative ROIs per timescale on ABIDE.**

Timescale	ROI	Network	M_r
Short	LH_VisCent_ExStr_5	Vis	4.41
	RH_SalVentAttnB_Ins_2	VentAttn	4.63
	RH_SomMotA_10	SomMot	5.10
Mid	LH_SomMotA_2	SomMot	10.04
	RH_ContA_PFCl_1	Cont	10.04
	LH_SomMotB_Cent_1	SomMot	10.03
Long	RH_VisPeri_ExStrInf_1	Vis	17.02
	LH_SomMotA_6	SomMot	14.22
	LH_VisCent_ExStr_4	Vis	13.89

The qualitative dissociation from COBRE aligns with prior literature, in which schizophrenia is often reported with prominent DMN involvement [Whitfield-Gabrieli et al., 2009, Garrity et al., 2007], whereas autism has been associated with distributed connectivity alterations [Di Martino et al., 2014]. We note that the COBRE ($n = 145$) and ABIDE ($n = 1072$) cohorts differ in size, site composition, and age range, so the difference in absolute sensitivity (8 vs. 73 nominally significant ROIs at $p < 0.05$, 0 vs. 45 at FDR $q < 0.05$) partly reflects statistical power, whereas the spatial pattern (DMN-leaning vs. multi-network) is the more interpretable contrast.

K.2 COBRE

Table A18: **Top-10 ROIs by raw p on COBRE** (correctly classified subjects only, $n = 114$, SZ $n = 57$, HC $n = 57$). No ROI survived BH–FDR $q < 0.05$, likely reflecting the limited sample size, so this table reports the uncorrected ranking. t is for SZ$-$HC. $*$: uncorrected $p < 0.05$.

#	ROI	Network	Hemi	t	p	q_{BH}	p_{Bonf}	sig
1	RH_DefaultB_Temp_1	Default	R	$+2.61$	0.0103	0.7833	1.0000	$*$
2	LH_SomMotB_Cent_2	SomMot	L	$+2.59$	0.0108	0.7833	1.0000	$*$
3	RH_SalVentAttnA_ParMed_1	VentAttn	R	$+2.56$	0.0117	0.7833	1.0000	$*$
4	RH_VisCent_ExStr_1	Vis	R	-2.35	0.0204	0.9095	1.0000	$*$
5	RH_DefaultB_AntTemp_1	Default	R	$+2.17$	0.0319	0.9095	1.0000	$*$
6	LH_SomMotB_Aud_2	SomMot	L	-2.15	0.0333	0.9095	1.0000	$*$
7	RH_DefaultA_PFCm_3	Default	R	-2.09	0.0392	0.9095	1.0000	$*$
8	LH_DefaultB_Temp_3	Default	L	-2.01	0.0470	0.9095	1.0000	$*$
9	RH_LimbicB_OFC_4	Limbic	R	$+1.97$	0.0508	0.9095	1.0000	
10	LH_DefaultB_PFCd_1	Default	L	-1.97	0.0518	0.9095	1.0000	

We complement the ABIDE analysis with a parallel—but exploratory—analysis on COBRE, linking the learned timescale organization to the schizophrenia-discriminative signal at the subject level.

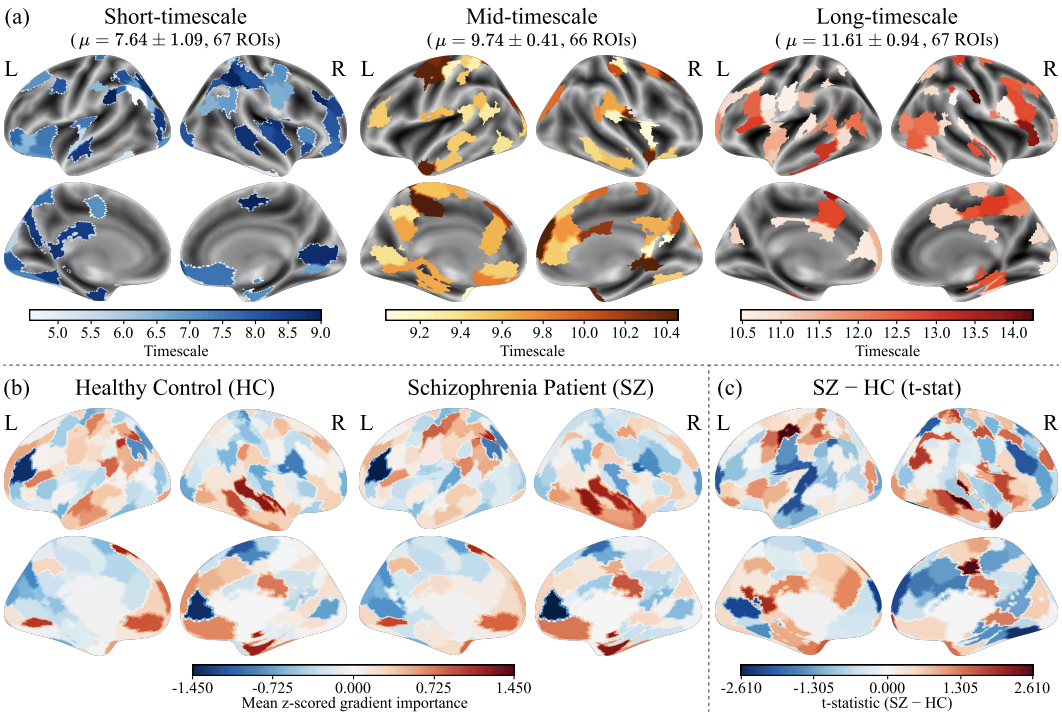


Figure A2: **Timescale organization and group-discriminative regions in COBRE.** (a) Quantile splits of per-ROI M_r partition cortical regions into short ($n = 67$, $\mu = 7.64 \pm 1.09$), mid ($n = 66$, $\mu = 9.74 \pm 0.41$), and long ($n = 67$, $\mu = 11.61 \pm 0.94$) timescale groups. In the long-timescale group, control-network ROIs are over-represented relative to the atlas (25% vs. 18.5%), whereas default-mode ROIs appear at their atlas-wide proportion (21%). (b) Subject-level gradient importance (z-scored within each subject) for correctly classified subjects ($n = 114$, 57 SZ, 57 HC), averaged within group. (c) Between-group t-statistic (SZ−HC). 8 ROIs (Table A18) reach $p < 0.05$ (uncorrected, $|t| > 1.98$, none surviving BH–FDR $q < 0.05$), 4 of which reside in the default mode network (strongest effect at right lateral temporal DMN, RH_DefaultB_Temp_1, $t = +2.61$, $p = 0.010$). Results are exploratory (see the Statistical caveats paragraph).

Because of COBRE's modest sample size, this analysis is hypothesis-generating rather than confirmatory (see Statistical caveats below); we report it primarily to contrast the spatial signature with that of ASD on ABIDE.

Quantile-based timescales. The M_r distribution on COBRE is unimodal (Hartigan's dip test $p = 0.28$, BIC favors $k = 1$ over multimodal GMM). Following the quantile-split procedure described above, we obtain short ($M_r \in [4.55, 9.02]$, $n = 67$), mid ($M_r \in [9.03, 10.46]$, $n = 66$), and long ($M_r \in [10.48, 14.27]$, $n = 67$) groups.

Hierarchical organization partially aligns with the canonical gradient. Mapping each timescale group to Yeo-7 networks shows that control-network ROIs are over-represented in the long-timescale group relative to the atlas (25% vs. 18.5%), whereas default-mode ROIs appear at their atlas-wide proportion (21%), partially consistent with the canonical sensorimotor-to-transmodal timescale gradient [Murray et al., 2014, Margulies et al., 2016, Hasson et al., 2008]. The short-timescale group, however, shows a more mixed distribution (Cont 18%, Default 18%, SomMot 16%) rather than a purely sensory composition, indicating that the model assigns fast dynamics to a heterogeneous set of regions across multiple networks (Fig. A2a).

Subject-level gradient importance (exploratory). For each test subject, we computed gradient saliency $|\partial(\text{class-1 logit})/\partial(\text{BOLD input})|$ summed over the time axis, averaged across the subject's segments, and z-scored across the 200 ROIs (Fig. A2b). Restricting to correctly classified subjects

($n = 114$, 78.6%), two-sample t-tests (SZ vs. HC) per ROI identified 8 ROIs with uncorrected $p < 0.05$ (Fig. A2c), 4 of which lie in the DMN, with the strongest effect at right lateral temporal DMN (RH_DefaultB_Temp_1, $t = +2.61$, $p = 0.010$). This nominal pattern is broadly compatible with prior reports of DMN involvement in schizophrenia [Garrity et al., 2007, Whitfield-Gabrieli et al., 2009], subject to the statistical caveats discussed below.

Statistical caveats. Two limitations temper the COBRE ROI-level findings. **(i) Power.** COBRE's modest sample ($n = 145$ overall, $n = 114$ correct) limits power for 200-ROI comparisons, and no ROI survives BH–FDR $q < 0.05$ at either n (Table A20). **(ii) Circularity.** Restricting to correctly classified subjects raised nominally significant counts from 0 (full cohort) to 8 (correct-only), partly reflecting non-independence, since the saliency maps are computed from the same model whose correct predictions define the subset. We therefore treat the COBRE pattern as suggestive and hypothesis-generating rather than confirmatory. The well-powered ABIDE analysis (§K.1), which applies the same correct-only restriction yet yields 45 FDR-surviving ROIs (none on COBRE), provides the primary statistically controlled evidence.

Table A19: **Representative ROIs per timescale on COBRE** (top-3 for short and long; top-2 for mid).

Timescale	ROI	Network	M_r
Short	LH_DefaultA_IPL_1	Default	4.55
	RH_LimbicA_TempPole_2	Limbic	4.97
	LH_DorsAttnA_TempOcc_1	DorsAttn	5.21
Mid	RH_SalVentAttnB_PFCmp_1	VentAttn	9.73
	RH_TempPar_4	Default	9.76
Long	RH_DorsAttnB_PostC_1	DorsAttn	14.27
	RH_DefaultB_PFCv_1	Default	13.79
	LH_DefaultB_PFCd_4	Default	13.71

Together, the over-representation of control-network ROIs in the long-timescale group provides partial alignment with known cortical timescale gradients [Murray et al., 2014, Margulies et al., 2016], while the exploratory subject-level analysis suggests a DMN-leaning pattern of group differences (4/8 nominally significant ROIs in DMN, uncorrected) that is broadly compatible with prior reports of DMN dysfunction in schizophrenia [Whitfield-Gabrieli et al., 2009, Bluhm et al., 2007] but does not survive multiple-comparison correction at the available sample size.

Read together with the FDR-controlled ABIDE analysis (§K.1), the two datasets yield qualitatively distinct spatial signatures, namely a DMN-leaning pattern in schizophrenia versus a multi-network pattern (Cont and Default jointly $\sim 53\%$ of FDR survivors, with substantial SomMot/Vis involvement) in autism. While COBRE's small sample precludes firm claims, this qualitative dissociation aligns with prior literature reporting prominent DMN involvement in schizophrenia [Whitfield-Gabrieli et al., 2009, Garrity et al., 2007] and distributed connectivity alterations in autism [Di Martino et al., 2014], and motivates future work with larger SZ cohorts to test the COBRE pattern in a confirmatory setting.

Table A20: **Subject-level ROI t-test corrected for multiple comparisons across 200 Schaefer ROIs** (group means of z-scored gradient saliency, ASD vs. TD on ABIDE and SZ vs. HC on COBRE). BH: Benjamini–Hochberg; Bonf.: Bonferroni ($p < 0.05/200$).

			Uncorrected		BH FDR		Bonf.
Dataset	Variant	n_{subj}	$p < 0.05$	$p < 0.01$	$q < 0.05$	$q < 0.10$	$p < 0.05$
ABIDE	all	1072	7	4	0	0	0
ABIDE	correct only	720	73	43	45	63	9
COBRE	all	145	0	0	0	0	0
COBRE	correct only	114	8	0	0	0	0

L Memory Variants on COBRE

To isolate the contribution of region-specific plasticity, we compare four variants of the co-activation pathway on COBRE. The variants differ only in the Hebbian bias added to the attention logits (§3.4), which uses region-specific rates η_r, γ_r (CoRTeX), a single rate pair shared by all regions as in standard fast weights, the current recurrent functional connectivity (RFC) $\mathbf{S}_t$ without accumulation, or no bias at all. All four variants, including the full model, are retrained once per fold with the same 5-fold splits and hyperparameters as Table 3, so the rows are paired. The retrained full model reproduces the AUC of Table 3 within 0.002 (0.795 vs. 0.797), well below the gaps between variants (0.059 to 0.082). As shown in Table A21, the instantaneous-RFC and no-memory variants perform nearly identically, and sharing the rates across regions does not recover the performance of region-specific rates, which is also the only variant with balanced sensitivity and specificity (SEN 0.745, SPEC 0.719).

Table A21: **Co-activation memory variants on COBRE** (mean±std across 5 folds).

Variant	ACC	AUC	SEN	SPEC
Region-specific η_r, γ_r (CoRTeX)	0.730 ± 0.068	0.795 ± 0.127	0.745 ± 0.181	0.719 ± 0.201
Rates shared across regions	0.605 ± 0.112	0.713 ± 0.143	0.718 ± 0.290	0.493 ± 0.463
Instantaneous RFC bias (no accumulation)	0.620 ± 0.139	0.736 ± 0.125	0.517 ± 0.273	0.720 ± 0.280
No memory	0.619 ± 0.147	0.729 ± 0.152	0.473 ± 0.300	0.760 ± 0.203

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Refer to Section 1 for our contributions and scope.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Limitations and directions for future work are discussed in §5, with $M_{\max}$ sensitivity in Appendix E and the exploratory status of the COBRE analysis discussed in Appendix K.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [N/A]

Justification: The paper does not include theoretical results.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Detailed dataset descriptions are provided in §4 and Appendix B; implementation and training details are in §4.2 and Appendix C; full architectural specifications are in Appendix D. Code is available at the link in the abstract.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: All datasets used (NSD, ABIDE, COBRE) are publicly available; access links and preprocessing procedures are provided in Appendix B. The source code is released as the repository linked in the abstract.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

Answer: [Yes]

Justification: We describe the datasets, preprocessing, and train/validation/test splits in Section 4.1 and Appendix B. The experimental setup, including the optimizer (AdamW with cosine schedule), hyperparameter search procedure, and final values, is provided in Section 4.2 and Appendix C (Tables A4 and A5).

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

Answer: [Yes]

Justification: We report statistical significance in Tables 1 and 2. For NSD, we train one model per subject with 5 random seeds (40 runs total: 8 subjects $\times$ 5 seeds) on the fixed dataset split, and compute significance via the Wilcoxon signed-rank test against CoRTeX. For ABIDE and COBRE, we use 5-fold cross-validation and compute significance on AUC via DeLong's paired test against CoRTeX on the concatenated test predictions across folds. Both tests are non-parametric and do not assume normality. Significance levels: $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. Further details are in Appendices B and C.

Guidelines:

- The answer [N/A] means that the paper does not include experiments.
- The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Hardware (a single NVIDIA RTX 2080 Ti GPU, 11 GB) is reported in Section 4.2. The total compute for the main reported runs ($\sim$ 70 GPU-hours: $\sim$ 66 h NSD, $\sim$4 h ABIDE, $\sim$1 h COBRE), along with disclosure of additional compute spent on hyperparameter search and exploratory experiments not reported in the paper, is provided in Appendix C.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conforms with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: We discuss broader impacts at the end of the conclusion (Section 5). CoRTeX may contribute positively to neuroscientific investigation of cortical organization through its interpretable per-region analyses. A potential negative impact is unvalidated clinical use of the diagnostic models, which could lead to misdiagnosis. We therefore frame the ABIDE and COBRE results as research-only. Additionally, performance in groups underrepresented in the training cohorts (e.g., female participants in ABIDE) has not been separately evaluated.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not pose a high risk for misuse.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: All datasets and code are credited by citation. NSD is used under the NSD Terms and Conditions (via the NSD Data Access Agreement); ABIDE I is distributed under CC BY-NC-SA 3.0 and COBRE under CC BY-NC through INDI, and both are used for non-commercial research. Baseline implementations were obtained from publicly available repositories and used under their respective licenses.

Guidelines:

- The answer [N/A] means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The details of the model and code have been provided in Section 4.2 and Appendix C, with the source code released under the MIT License in the repository linked in the abstract.

Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [N/A]

Justification: The paper does not involve crowdsourcing or new human-subject research.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve human-subject research requiring IRB approval.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: LLMs were used only for writing assistance (grammar and editing) and are not part of the core methodology.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.