CoFlu-Brain: An fMRI Connectivity Foundation Model with Cofluctuation-Based Pretraining

Anonymous authors
Paper under double-blind review

Abstract

Functional magnetic resonance imaging (fMRI) commonly characterizes large-scale brain organization through functional connectivity (FC). However, an observed FC matrix is not a fixed brain graph but a statistical estimate determined by the finite BOLD frames used to compute it. Conventional FC-based approaches typically represent each scan with a single Pearson-correlation matrix, leaving frame-resolved variations in regional interactions unexploited. To address this limitation, we introduce CoFlu-Brain, which organizes framewise cofluctuation observations into multiple retained and dropped views, each summarized as an FC matrix. The model learns connectivity structure reproducible across retained views and predicts the representations of the corresponding dropped views, while adopting SIGReg from LeJEPA to prevent representation collapse. We pretrained CoFlu-Brain on 13,444 subjects from eight dataset groups and evaluated its representations via frozen-encoder linear probing on seven classification and two age-prediction tasks across seven datasets. CoFlu-Brain showed strong transfer across demographic, developmental, and clinical prediction tasks, outperforming the evaluated released models on most tasks. Random frame partitioning improved aggregate transfer, while the complete objective achieved the strongest overall performance. Code is available.

1 Introduction

Functional magnetic resonance imaging (fMRI) provides a non-invasive view of large-scale functional brain organization and its relationships with cognition, behavior, and disease. As large multi-site fMRI collections continue to grow, self-supervised pretraining offers a path toward transferable brain representations without requiring downstream task labels during pretraining (Ortega Caro et al., 2024; Dong et al., 2024; Gijsen et al., 2026). A common route from fMRI measurements to brain-network representations is to parcellate the brain into regions of interest (ROIs), extract a regional BOLD time series from each ROI, and estimate the statistical dependence between regional signals. Pearson correlation organizes these pairwise dependencies into a compact and interpretable functional connectivity (FC) matrix. By explicitly representing interregional functional coupling, FC has become a widely used input for brain-network modeling and connectome-based pretraining (Kan et al., 2022; Yang et al., 2024; Wei et al., 2026a).

An observed Pearson-FC matrix, however, is not a sampling-independent brain graph but a scan-level statistical estimate formed from a finite BOLD sequence. Its value can vary with the frames included in the estimation, such that different frame compositions produce non-identical but statistically related FC estimates.

Existing FC representation models typically operate on a single Pearson-FC matrix or graph for each scan (Kan et al., 2022; Wei et al., 2026a), leaving frame-resolved variation unused during representation learning. Framewise cofluctuation provides a finer-grained description of regional interactions by characterizing the joint standardized activity of regional pairs at individual frames (Zamani Esfahlani et al., 2020; Faskowitz et al., 2020). This perspec-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

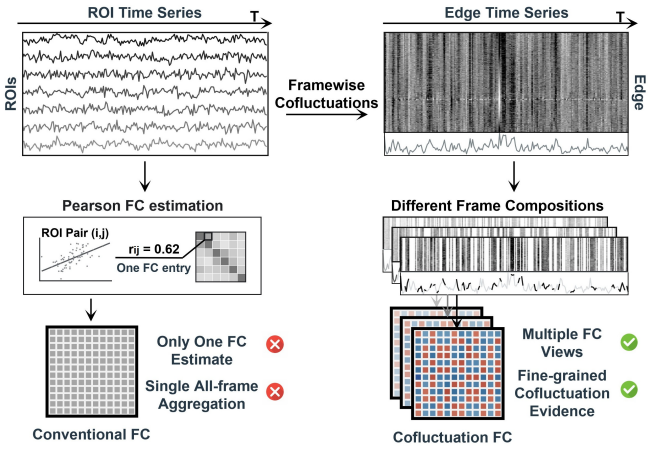

Figure 1: **From a single FC estimate to cofluctuation based multi-view learning.** Conventional FC aggregates all frames into a single connectivity estimate, whereas CoFlu-Brain uses different compositions of framewise cofluctuation observations to construct multiple FC views.

tive makes it possible to construct multiple training views from different compositions of framewise observations.

This gap raises a central question: can different compositions of framewise cofluctuation observations provide self-supervision for learning stable and transferable connectivity representations? We hypothesize that retained frame sets sampled from the same segment provide different views of subject-specific cofluctuation structure. Although the resulting FC views may differ numerically, their shared connectivity structure should remain reproducibly inferable. Moreover, a retained view should contain sufficient information to predict the representation of the corresponding non-overlapping dropped view. The former encourages consistency across retained views, whereas the latter captures predictive relationships between retained and dropped observations.

Motivated by this gap, we propose **CoFlu-Brain**, a cofluctuation-based approach to self-supervised FC pretraining. For each fMRI segment, CoFlu-Brain repeatedly partitions the frames into retained and dropped sets, organizes their framewise cofluctuation observations into separate FC views, and uses the dropped-view representations as prediction targets. Its joint-embedding objective promotes consistent representations across retained views while requiring each retained representation to predict its corresponding dropped-view representation in latent space. We evaluate the resulting representations through multi-dataset pretraining, frozen transfer, and systematic ablations.

Our contributions are:

1. We propose a **cofluctuation-based FC pretraining approach** that constructs retained and dropped views from framewise observations and couples consistency across retained views with retained-to-dropped latent prediction.
2. We develop **CoFlu-Brain**, an FC foundation model pretrained on multi-cohort fMRI data, and demonstrate strong frozen-transfer performance across demographic, developmental, and clinical downstream tasks.
3. We conduct a **systematic empirical analysis** across nine downstream tasks, examining the contributions of the learning objectives and the effects of data coverage, branch count, and temporal sampling, while assessing seed robustness for the primary and temporal-sampling comparisons.

2 Related Work

fMRI representation models differ first in what they model. Time-series methods reconstruct masked BOLD signals, predict masked latent targets, or align representations across temporal views (Ortega Caro et al., 2024; Dong et al., 2024; Gijsen et al., 2026). They preserve temporal dynamics, but interregional connectivity remains an implicit property of the learned activity representation.

Voxel-level models instead operate directly on 4D fMRI volumes (Wang et al., 2026; 2025; Xia et al., 2026). They retain fine-grained spatial information without relying on a predefined atlas, but require high-dimensional inputs and do not use connectivity as the primary modeling unit.

FC models represent each scan as a connectivity matrix or graph, directly encoding interregional coupling (Kawahara et al., 2017; Li et al., 2021; Kan et al., 2022; Yang et al., 2024; Wei et al., 2026a;b). Most treat the resulting Pearson-FC matrix as the object to encode, reconstruct, augment, or align. This overlooks that FC is a finite-sample estimate whose values depend on its supporting frames (Noble et al., 2019; Zamani Esfahlani et al., 2020; Faskowitz et al., 2020). Dynamic-FC methods instead model temporal variation across windows or graph sequences (Hutchison et al., 2013; Kim et al., 2021), but address evolving connectivity states rather than reproducibility across alternative estimates.

CoFlu-Brain retains the interpretability of FC modeling while extending its learning unit from one scan-level estimate to multiple related estimates. These views arise from complementary frame observations and preserve a direct relationship to FC estimation, promoting representations stable across frame compositions and predictive of held-out observations.

Self-supervised methods also differ in how they define their learning signal. Masked autoencoding reconstructs omitted input content, whereas JEPA predicts masked targets in latent space (He et al., 2022; Ortega Caro et al., 2024; Yang et al., 2024; Wei et al., 2026a; Dong et al., 2024; Assran et al., 2023). Joint-embedding methods such as BYOL and DINO instead align representations across views, while VICReg and SIGReg regularize the feature distribution against collapse (Grill et al., 2020; Caron et al., 2021; Bardes et al., 2022; Balestriero & LeCun, 2025).

CoFlu-Brain integrates these principles through FC-specific view relations. Consistency across retained estimates promotes stability under frame resampling; prediction from retained to complementary dropped estimates captures information predictive of held-out observations; and SIGReg maintains representation diversity. Together, these components use the finite-sample structure of FC to define which views should agree and which complementary evidence should be predicted.

3 Method

3.1 Overview

CoFlu-Brain constructs multiple retained–dropped FC pairs from each fMRI segment. As illustrated in Figure 2, a shared encoder maps each estimate to a scan-level representation. Training combines consistency across retained-frame representations, prediction of the paired dropped-frame representation, and distributional regularization. After pretraining, the projector and predictor are discarded, and the shared FC encoder provides representations for downstream tasks.

3.2 Cofluctuation-based retained and dropped views

Let $X_b \in \mathbb{R}^{T\times N}$ denote the preprocessed fMRI segment of sample b. For notational simplicity, we describe the construction for a fixed sample and suppress the sample index b below. For a selected frame set A, we independently standardize each regional signal within A, yielding z_t^A at frame t. We define the framewise cofluctuation matrix as

$$E_t^A = z_t^A (z_t^A)^\top.$$

CoFlu-Brain summarizes the cofluctuation observations associated with A as an FC view:

$$R_{ij}^A = \frac{1}{|A|-1}\sum_{t\in A} E_{t,ij}^A, \qquad i \neq j.$$

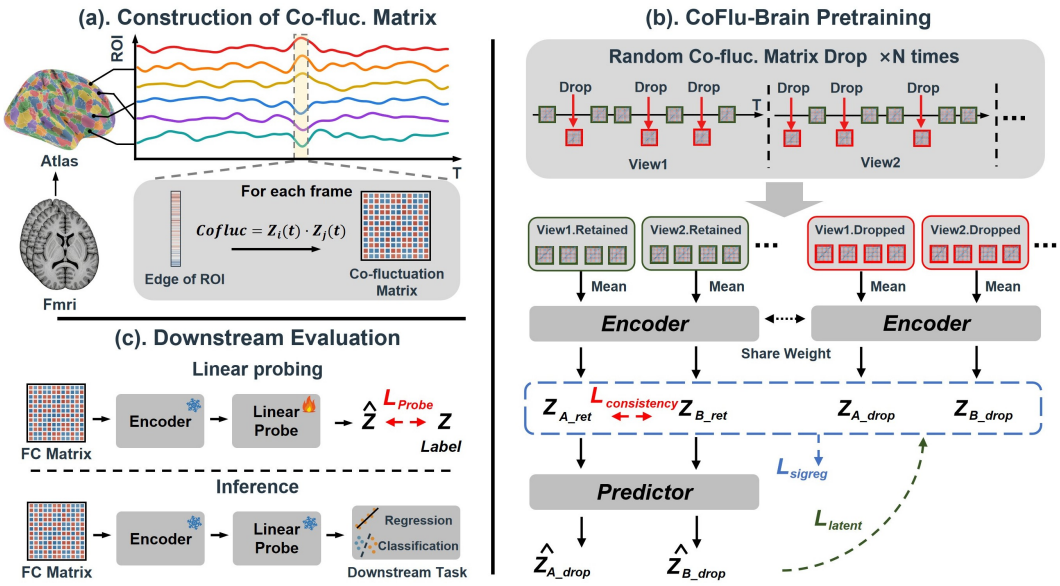

Figure 2: **CoFlu-Brain pretraining and downstream evaluation.** **(a)** Framewise cofluctuations from standardized regional BOLD signals. **(b)** Paired retained–dropped FC views train a shared encoder through consistency, latent prediction, and SIGReg. **(c)** Frozen-encoder linear probing and downstream inference.

This decomposition makes the frame support of each finite-sample FC estimate explicit and enables the complementary view relations used below.

The resulting matrix is Fisher transformed on its off-diagonal entries before being passed to the encoder. For each partition v, we sample a retained proportion r_v. We draw the retained frame set S_v without replacement and assign the remaining frames to the dropped set D_v. The two sets do not overlap within a pair, $S_v \cap D_v = \emptyset$, while independently sampled partitions may reuse frames across different pairs. We independently standardize the retained and dropped sets and construct their cofluctuation-based FC views:

$$C^{S_v} = \mathrm{FisherZ}(R^{S_v}), \qquad C^{D_v} = \mathrm{FisherZ}(R^{D_v}).$$

The Fisher transformation is applied only to off-diagonal entries, and all diagonal entries are set to zero. For a minibatch, we restore the sample index b, writing S_{vb}, D_{vb}, $C_b^{S_{vb}}$, and $C_b^{D_{vb}}$.

The retained and dropped views summarize disjoint sets of framewise cofluctuation observations from the same scan. Repeating this construction for V independently sampled partitions exposes the model to different cofluctuation compositions. The learning objective encourages connectivity representations that remain consistent across retained views and predictive of their paired dropped views.

3.3 Model architecture

The encoder represents each sampled frame set through its cofluctuation-based FC estimate $C \in \mathbb{R}^{N\times N}$. With diagonal entries set to zero, we treat each row $c_i \in \mathbb{R}^N$ as the signed connectivity profile of ROI i. A shared row-wise encoder $\phi : \mathbb{R}^N \to \mathbb{R}^d$ maps the N connectivity profiles to N ROI tokens. A Vision Transformer (ViT) encoder F_θ (Dosovitskiy et al., 2021) then contextualizes these tokens through interactions among ROIs:

$$H(C) = F_\theta([\phi(c_1), \ldots, \phi(c_N)]) \in \mathbb{R}^{N\times d}.$$

Mean pooling over the contextualized ROI tokens produces the scan-level FC representation

$$f_\theta(C) = \frac{1}{N}\sum_{i=1}^{N} H_i(C) \in \mathbb{R}^d.$$

For each of the V retained–dropped pairs and each sample b in a minibatch of size B, the same encoder and projector $g : \mathbb{R}^d \to \mathbb{R}^{d_p}$ are applied to both FC estimates:

$$p_{vb} = g\Big(f_\theta(C_b^{S_{vb}})\Big), \qquad q_{vb} = g\Big(f_\theta(C_b^{D_{vb}})\Big), \quad v = 1, \dots, V, \quad b = 1, \dots, B.$$

A predictor $h : \mathbb{R}^{d_p} \to \mathbb{R}^{d_p}$ maps each retained-frame projection to $\hat{q}_{vb} = h(p_{vb})$, which is trained to predict the corresponding dropped-frame projection q_{vb}. The retained and dropped branches share all encoder and projector parameters, and both receive gradients during pretraining. After pretraining, the projector and predictor are discarded; only the encoder representation $f_\theta(C)$ is used for frozen downstream probing. For downstream evaluation, C was computed from all 200 frames of each input segment. Architectural dimensions and optimization settings are reported in Appendix A.3.

3.4 Learning across FC estimates

Reproducibility across retained estimates. Retained-frame FC matrices from the same scan differ numerically because each is supported by a different collection of cofluctuations. We encourage their representations to preserve the structure that can be inferred reproducibly across estimates:

$$\bar{p}_b = \frac{1}{V}\sum_{v=1}^{V} p_{vb}, \qquad \mathcal{L}_{\text{con}} = \frac{1}{VBd_p}\sum_{v=1}^{V}\sum_{b=1}^{B} \|p_{vb} - \bar{p}_b\|_2^2 .$$

Here, V, B, and d_p denote the number of temporal partitions, the minibatch size, and the projection dimension, respectively.

Prediction of dropped-frame evidence. Consistency captures structure shared across retained estimates. To also learn across paired non-overlapping frame subsets, a predictor h infers the representation supported by the dropped frames:

$$\mathcal{L}_{\text{pred}} = \frac{1}{VB}\sum_{v=1}^{V}\sum_{b=1}^{B} \left[1 - \cos(h(p_{vb}), q_{vb})\right].$$

Both branches share parameters and receive gradients without requiring a stop-gradient target or momentum encoder. This term trains each retained FC representation to predict its paired dropped-frame representation.

Non-collapse regularization. Consistency and prediction specify relationships among paired FC estimates. We additionally apply SIGReg to the collection of retained and dropped projections following LeJEPA (Balestriero & LeCun, 2025), regularizing their distribution against representational collapse. The complete objective is

$$\mathcal{L} = \mathcal{L}_{\text{pred}} + \lambda_{\text{con}}\gamma(e)\mathcal{L}_{\text{con}} + \lambda_{\text{reg}}\mathcal{L}_{\text{SIGReg}},$$

where $\gamma(e)$ controls the contribution of consistency through an initial full-weight phase followed by linear decay:

$$\gamma(e) = \begin{cases} 1, & e \le E_w, \\ \max\left(1 - \dfrac{e - E_w}{E_d}, 0\right), & e > E_w. \end{cases}$$

Here, e is the training epoch, E_w is the duration of the initial full-weight phase, and E_d is the subsequent decay duration. In the main configuration, $E_w = 1$ and $E_d = 24$, so the consistency weight reaches zero at epoch 25. The schedule emphasizes alignment early in training and shifts the later objective toward dropped-frame prediction. The three loss terms respectively promote latent prediction, reproducibility across finite FC estimates, and a non-degenerate representation distribution.

4 Results

4.1 Experimental setup

CoFlu-Brain was pretrained on 13,444 subjects from eight dataset groups: CHCP, HCP, ABCD, ABCC, HBN, PNC, MDD, and AOMIC PIOP1/2 (Ge et al., 2023; Van Essen et al., 2013; Casey et al., 2018; Feczko et al., 2021; Alexander et al., 2017; Satterthwaite et al., 2014; Yan et al., 2019; Snoek et al., 2021). All inputs were represented as Schaefer-400 cortical time series (Schaefer et al., 2018), and 200-frame segments were used for pretraining. Downstream evaluation comprised two in-cohort tasks from HCP and ABCD, whose unlabeled scans contributed to pretraining, and seven out-of-cohort tasks from SALD, PPMI, ADNI, ABIDE, and ADHD200, which were excluded from pretraining (Van Essen et al., 2013; Casey et al., 2018; Wei et al., 2018; Marek et al., 2011; Weiner et al., 2010; Di Martino et al., 2014; ADHD-200 Consortium, 2012). No downstream labels were used during pretraining.

Frozen representations were evaluated with logistic regression for classification and ridge regression for age prediction. For age prediction, nMAE was computed after standardizing the target using the training-fold mean and population standard deviation. Cross-dataset ranks used weighted F1 for classification and nMAE for age prediction. Endpoint ranks were averaged within each cohort, and the dispersion of method mean ranks was evaluated using two million method-label permutations within cohort blocks. Full details are provided in Appendix A.

4.2 Main Results

We evaluated CoFlu-Brain against four released fMRI representation models: BrainGFM, BrainMASS, BrainLM, and the Large Connectome Model (Wei et al., 2026a; Yang et al., 2024; Ortega Caro et al., 2024; Wei et al., 2026b).

Table 1. **Linear probing of fMRI models for classification and age prediction.** Values are mean ± sample SD across five participant-grouped folds. Classification tasks report accuracy and weighted F1 (↑); age-prediction tasks report nMAE (↓) and Pearson's r (↑). Red boldface marks the best mean for each metric. CoFlu-Brain averages three pretrained checkpoints per fold; baselines use one released checkpoint. Overall rank test: cohort-blocked permutation $p < 0.001$.

Dataset	Model	CoFlu-Brain	BrainGFM	BrainMASS	BrainLM	LCM
HCP / Sex	Accuracy ↑	**0.7749 ± 0.0111**	0.6428 ± 0.0123	0.7207 ± 0.0053	0.6258 ± 0.0138	0.7292 ± 0.0163
	Weighted F1 ↑	**0.7745 ± 0.0103**	0.6432 ± 0.0119	0.7203 ± 0.0048	0.6258 ± 0.0129	0.7284 ± 0.0157
ABCD / Sex	Accuracy ↑	0.5953 ± 0.0195	0.5336 ± 0.0482	**0.6216 ± 0.0516**	0.4965 ± 0.0352	0.5436 ± 0.0308
	Weighted F1 ↑	0.5961 ± 0.0190	0.5317 ± 0.0528	**0.6215 ± 0.0515**	0.4967 ± 0.0350	0.5420 ± 0.0331
PPMI / Group	Accuracy ↑	**0.5918 ± 0.0063**	0.5871 ± 0.0335	0.5813 ± 0.0280	0.5269 ± 0.0192	0.5501 ± 0.0262
	Weighted F1 ↑	**0.5775 ± 0.0173**	0.5502 ± 0.0468	0.5751 ± 0.0301	0.5237 ± 0.0300	0.5354 ± 0.0373
ADNI / AD	Accuracy ↑	0.6841 ± 0.0306	**0.6870 ± 0.0194**	0.5826 ± 0.0418	0.6217 ± 0.1049	0.5957 ± 0.1136
	Weighted F1 ↑	**0.6835 ± 0.0307**	0.6832 ± 0.0236	0.5811 ± 0.0393	0.6183 ± 0.1078	0.5934 ± 0.1179
ADNI / MCI	Accuracy ↑	**0.5570 ± 0.0415**	0.5515 ± 0.0351	0.5070 ± 0.0339	0.5412 ± 0.0661	0.5513 ± 0.0233
	Weighted F1 ↑	**0.5538 ± 0.0441**	0.5480 ± 0.0367	0.5045 ± 0.0346	0.5367 ± 0.0673	0.5506 ± 0.0232
ABIDE / ASD	Accuracy ↑	**0.5602 ± 0.0474**	0.4911 ± 0.0734	0.5487 ± 0.0520	0.5087 ± 0.1294	0.4971 ± 0.0618
	Weighted F1 ↑	**0.5567 ± 0.0476**	0.4902 ± 0.0737	0.5472 ± 0.0527	0.5057 ± 0.1333	0.4913 ± 0.0630
ADHD200 / Diagnosis	Accuracy ↑	**0.5899 ± 0.0155**	0.5297 ± 0.0313	0.5559 ± 0.0257	0.5241 ± 0.0105	0.5797 ± 0.0397
	Weighted F1 ↑	**0.5891 ± 0.0151**	0.5286 ± 0.0285	0.5549 ± 0.0254	0.5235 ± 0.0089	0.5806 ± 0.0381
SALD / Age	nMAE ↓	**0.6571 ± 0.0431**	0.8548 ± 0.0605	0.9064 ± 0.0877	0.9824 ± 0.0473	0.8759 ± 0.0641
	Pearson r ↑	**0.6422 ± 0.0418**	0.4262 ± 0.0600	0.4415 ± 0.1125	0.3180 ± 0.1073	0.4335 ± 0.0793
ADHD200 / Age	nMAE ↓	**0.6557 ± 0.0370**	0.8406 ± 0.0797	0.9328 ± 0.0377	1.1863 ± 0.0455	0.9135 ± 0.0545
	Pearson r ↑	**0.5967 ± 0.0249**	0.2696 ± 0.0593	0.3525 ± 0.0526	0.2374 ± 0.0783	0.4433 ± 0.0476
Cohort-averaged rank		**1.14**	3.36	2.93	4.50	3.07

CoFlu-Brain led eight of nine tasks and achieved the best rank across the seven evaluation cohorts (1.14; Table 1).

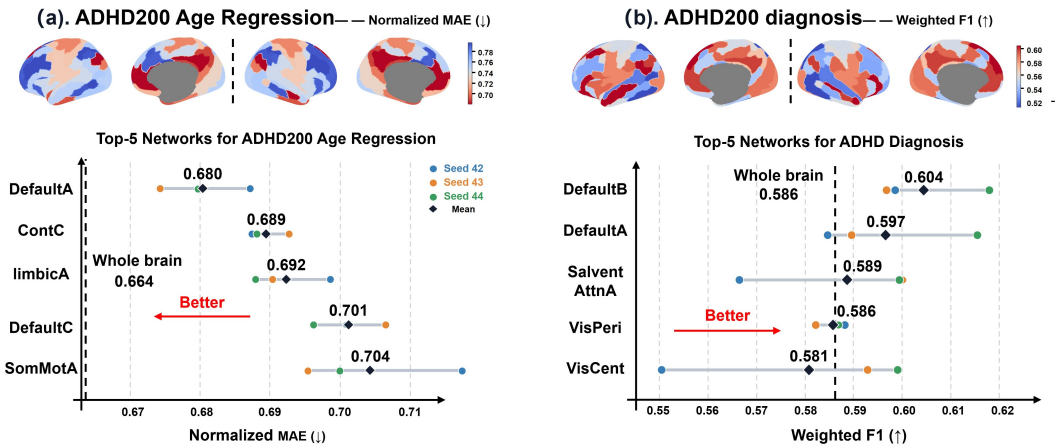

Figure 3: **Network-restricted probing on ADHD200.** **(a)** Age nMAE (↓). **(b)** Diagnosis weighted F1 (↑). Cortical maps summarize mean performance across all 17 networks, and the corresponding point-range plots detail the five best-performing networks for each task. Points indicate individual pretraining seeds, diamonds indicate seed means, horizontal segments indicate seed ranges, and dashed vertical lines indicate whole-brain performance.

4.3 Functional-network analysis on ADHD200

We examined how predictive information for age and ADHD diagnosis was distributed across the Yeo-17 functional systems (Yeo et al., 2011). For each network, we retained its subject-specific within-network FC, replaced the remaining entries with their probe-training means, and evaluated the resulting representation using the frozen encoder and linear probe (Figure 3).

Network rankings were consistent across the three pretrained encoders, with mean pairwise rank correlations of $\rho = 0.906$ for age and $\rho = 0.857$ for diagnosis. The prominence of default-network components in both tasks is consistent with their progressive integration during brain maturation (Dosenbach et al., 2010; Fair et al., 2008). Their joint contribution with the salience/ventral-attention system to diagnosis further reflects the distributed network organization associated with ADHD (Sripada et al., 2014b;a; Norman et al., 2023) (see Appendix B.2 for complete results).

4.4 Ablation and sensitivity analyses

We evaluated temporal sampling (Appendix C.1) and pretraining objectives (Appendix C.2) through ablations, and examined sensitivity to data coverage (Appendix C.3), temporal branch count (Appendix C.4), and consistency-weight scheduling (Appendix C.5). For compact presentation, all four tabular analyses use the same task subset; the appendices report all nine tasks.

4.4.1 Random versus ordered frame partitioning

To compare temporal sampling strategies, we evaluated random and ordered partitions at matched retained-frame ratios of 0.4–0.7. Random partitioning sampled retained frames throughout the sequence and used the complement as the dropped view; ordered partitioning used the first k and remaining $T - k$ frames. Both used identical standardization, FC construction, architecture, and objective, differing only in frame indices. This comparison assesses the empirical effect of the two frame-partitioning strategies on downstream transfer.

On seed-averaged performance, random partitioning outperformed ordered partitioning across all seven evaluation cohorts (exact sign test, $p = 0.016$; Figure 4).

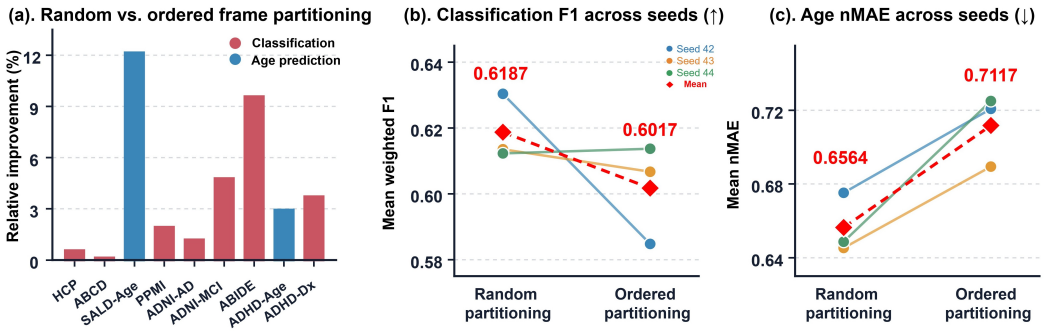

Figure 4: **Random versus ordered frame partitioning.** **(a)** Relative weighted-F1 gain for classification and relative nMAE reduction for age prediction; positive values favor random partitioning. **(b)** Mean weighted F1 across seven classification tasks. **(c)** Mean nMAE across two age-prediction tasks. Lines connect paired pretraining seeds; red diamonds denote seed means.

4.4.2 Pretraining objectives

We next assessed the contribution of the three objective components by removing prediction, consistency, or SIGReg while preserving the remaining pretraining configuration. Table 2 presents four downstream tasks, while Appendix C.2 reports the complete nine-task results and representation diagnostics.

Table 2. **Downstream performance under objective-component ablation.** Each objective variant was pretrained independently with seed 42 under matched configurations and is reported as mean ± SD across five participant-grouped folds. Metrics are weighted F1 (↑) and age nMAE (↓). Red boldface marks the best mean for each task. Overall rank test: cohort-blocked permutation $p = 0.0014$.

Objective \ Dataset	HCP / Sex	ADNI / AD	ADHD200 / Diagnosis	ADHD200 / Age
	Weighted F1 ↑	Weighted F1 ↑	Weighted F1 ↑	nMAE ↓
Full	**0.7874 ± 0.0115**	**0.7219 ± 0.0657**	**0.5872 ± 0.0215**	**0.6523 ± 0.0315**
No prediction	0.6565 ± 0.0187	0.5023 ± 0.0850	0.5348 ± 0.0309	0.7416 ± 0.0567
No consistency	0.7789 ± 0.0110	0.6866 ± 0.1175	0.5725 ± 0.0384	0.6611 ± 0.0539
No SIGReg	0.6942 ± 0.0153	0.6065 ± 0.0667	0.5643 ± 0.0288	0.7132 ± 0.0482
Cohort-averaged rank: Full 1.50; No prediction 3.57; No consistency 1.79; No SIGReg 3.14				

The complete objective ranked best overall (1.50; Table 2). Removing prediction or SIGReg generally reduced transfer, while consistency produced task-dependent gains.

4.4.3 Pretraining data coverage

To examine how the breadth of pretraining data affected transfer, we increased the pretraining manifest from 12.5% to 100% while keeping the architecture and number of optimization updates fixed. Table 3 shows four downstream tasks, and Appendix C.3 reports all nine.

Table 3. **Pretraining data coverage and downstream performance.** Values are mean ± SD across probe folds. Metrics are weighted F1 (↑) and age nMAE (↓). Red boldface marks the best mean for each task.

Coverage \ Dataset	HCP / Sex	ADNI / AD	ADHD200 / Diagnosis	ADHD200 / Age
	Weighted F1 ↑	Weighted F1 ↑	Weighted F1 ↑	nMAE ↓
12.5%	0.7265 ± 0.0126	0.6627 ± 0.0137	0.5678 ± 0.0134	0.7553 ± 0.0318
25%	0.7376 ± 0.0117	**0.7280 ± 0.0348**	0.5609 ± 0.0097	0.7453 ± 0.0341
50%	0.7326 ± 0.0240	0.6786 ± 0.0301	0.5314 ± 0.0333	0.7152 ± 0.0174
100%	**0.7737 ± 0.0092**	0.7235 ± 0.0336	**0.5756 ± 0.0298**	**0.6785 ± 0.0224**

Using the complete manifest improved all four tasks relative to 12.5% coverage and produced the best result on three of them (Table 3).

4.4.4 Number of temporal branches

Each temporal branch provides a distinct retained-frame FC estimate during pretraining. We varied the number of branches from two to eight to determine how many retained–dropped estimate pairs were most effective, while retaining the same update budget and normalizing the consistency objective across branch counts. Table 4 shows four downstream tasks; Appendix C.4 reports all nine together with training-time details.

Table 4. **Downstream performance by temporal branch count.** Values are mean ± SD across probe folds. Metrics are weighted F1 (↑) and age nMAE (↓). Red boldface marks the best mean for each task.

Branches \ Dataset	HCP / Sex	ADNI / AD	ADHD200 / Diagnosis	ADHD200 / Age
	Weighted F1 ↑	Weighted F1 ↑	Weighted F1 ↑	nMAE ↓
B2	0.7557 ± 0.0205	0.6518 ± 0.0536	**0.5762 ± 0.0387**	0.7088 ± 0.0091
B4	**0.7737 ± 0.0092**	**0.7235 ± 0.0336**	0.5756 ± 0.0298	**0.6785 ± 0.0224**
B6	0.7601 ± 0.0296	0.6429 ± 0.0818	0.5637 ± 0.0183	0.6872 ± 0.0195
B8	0.7643 ± 0.0122	0.6686 ± 0.0470	0.5713 ± 0.0269	0.6894 ± 0.0338

Four branches achieved the best performance on three of the four tasks, whereas increasing from four to eight branches raised the pretraining cost from 10.54 to 26.72 GPU hours without consistent performance gains (Table 4).

4.4.5 Consistency-weight schedule

Consistency aligns retained-frame compositions, while prediction captures relationships between retained- and dropped-frame FC estimates. We hypothesized that emphasizing consistency early and prediction later could balance these signals. We therefore compared Decay-25, Decay-50, and Constant consistency schedules. Table 5 presents the three schedules, while Appendix C.5 reports their performance on all nine downstream tasks.

Table 5. **Downstream performance under consistency-weight schedules.** Values are mean ± SD across probe folds. Metrics are weighted F1 (↑) and age nMAE (↓). Red boldface marks the best mean for each task.

Schedule \ Dataset	HCP / Sex	ADNI / AD	ADHD200 / Diagnosis	ADHD200 / Age
	Weighted F1 ↑	Weighted F1 ↑	Weighted F1 ↑	nMAE ↓
Decay-25	**0.7737 ± 0.0092**	**0.7235 ± 0.0336**	**0.5756 ± 0.0298**	0.6785 ± 0.0224
Decay-50	0.7588 ± 0.0204	0.6736 ± 0.0236	0.5433 ± 0.0282	0.6839 ± 0.0239
Constant	0.7687 ± 0.0164	0.6627 ± 0.0350	0.5688 ± 0.0300	**0.6568 ± 0.0270**

Decay-25 performed best on all three classification tasks, whereas the constant schedule performed best for age prediction (Table 5).

5 Discussion

CoFlu-Brain shows that relationships among FC estimates supported by different frame compositions provide useful self-supervision. Its representations transferred across demographic, developmental, and clinical tasks, while random partitioning improved aggregate transfer. The complete objective combines retained-view consistency, dropped-view prediction, and distributional regularization as complementary signals.

Limitations and future directions. CoFlu-Brain uses the framewise cofluctuation decomposition of Pearson connectivity to construct multiple FC estimates with explicitly controlled frame support. Because each view aggregates the selected framewise products into a subset-specific FC matrix, the current model does not preserve the timing, order, or amplitude of individual cofluctuation events. Its scope is instead to learn from structured relationships among finite-sample FC estimates: reproducibility across retained-frame compositions and prediction of complementary held-out estimates. Future work could combine this estimate-level objective with order-aware or edge-time-series encoders to capture transient dynamics. Biological validation, causal analysis, and integration with voxel-level features remain important directions beyond the predictive evaluations considered here.

AI use statement

Generative AI tools were used to assist manuscript organization and language editing, table formatting, and consistency auditing between reported results and experiment records. They were not used to generate experimental data. The authors reviewed all AI-assisted text, independently verified the reported numerical values against the underlying result files, and take responsibility for the final content of the submission.

Ethics statement

This work performs secondary analysis of de-identified neuroimaging data from established research cohorts and does not recruit new participants or attempt participant re-identification. Data access and use are governed by the terms of the respective source datasets. Cohort composition, preprocessing provenance, and pretraining-data exposure are documented to support appropriately scoped interpretation of the transfer results.

Reproducibility statement

Section 3 specifies the model and objective, and Section 4.1 defines the pretraining and downstream evaluation protocol. Supplementary Appendix A documents the datasets, preprocessing, baselines, architecture, and training and probing configurations. Appendix B provides seed-wise and complete network-level results, while Appendix C reports the full ablation and sensitivity analyses. The implementation, configuration files, verification scripts, and pretrained checkpoints are available in the anonymous repository: https://anonymous.4open.science/r/review-artifact-2027/.

References

ADHD-200 Consortium. The ADHD-200 consortium: A model to advance the translational potential of neuroimaging in clinical neuroscience. *Frontiers in Systems Neuroscience*, 6: 62, 2012. doi: 10.3389/fnsys.2012.00062.

Lindsay M. Alexander, Jasmine Escalera, Lei Ai, Charissa Andreotti, Karina Febre, Alexander Mangone, Natan Vega-Potler, Nicolas Langer, Alexis Alexander, Meagan Kovacs, et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific Data*, 4:170181, 2017. doi: 10.1038/sdata.2017.181.

Mahmoud Assran, Quentin Duval, Ishan Misra, Piotr Bojanowski, Pascal Vincent, Michael Rabbat, Yann LeCun, and Nicolas Ballas. Self-supervised learning from images with a joint-embedding predictive architecture. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15619–15629, 2023. doi: 10.1109/CVPR 52729.2023.01499.

Randall Balestriero and Yann LeCun. LeJEPA: Provable and scalable self-supervised learning without the heuristics. *arXiv preprint arXiv:2511.08544*, 2025.

Adrien Bardes, Jean Ponce, and Yann LeCun. VICReg: Variance-invariance-covariance regularization for self-supervised learning. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=IXexLXymbZ9.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9630–9640, 2021. doi: 10.1109/ICCV48922.2021.00951.

B. J. Casey, Tariq Cannonier, May I. Conley, Alexandra O. Cohen, Deanna M. Barch, Mary M. Heitzeg, Mary E. Soules, Theresa Teslovich, Danielle V. Dellarco, Hugh Garavan, et al. The adolescent brain cognitive development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, 32:43–54, 2018. doi: 10.1016/j.dcn.2018.03.001.

Adriana Di Martino, Chao-Gan Yan, Qingyang Li, Erin Denio, F. Xavier 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. doi: 10.1038/mp.2013.78.

Zijian Dong, Ruilin Li, Yilei Wu, Thuan Tinh Nguyen, Joanna Su Xian Chong, Fang Ji, Nathanael Ren Jie Tong, Christopher Li Hsian Chen, and Juan Helen Zhou. Brain-JEPA: Brain dynamics foundation model with gradient positioning and spatiotemporal masking. In *Advances in Neural Information Processing Systems*, volume 37, pp. 86048–86073, 2024. doi: 10.52202/079017-2732. URL https://proceedings.neurips.cc/paper_files/paper/2024/hash/9c3828adf1500f5de3c56f6550dfe43c-Abstract-Conference.html.

Nico U. F. Dosenbach, Binyam Nardos, Alexander L. Cohen, Damien A. Fair, Jonathan D. Power, Jessica A. Church, Steven M. Nelson, Gagan S. Wig, Alecia C. Vogel, Christina N. Lessov-Schlaggar, et al. Prediction of individual brain maturity using fMRI. *Science*, 329 (5997):1358–1361, 2010. doi: 10.1126/science.1194144.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

Damien A. Fair, Alexander L. Cohen, Nico U. F. Dosenbach, Jessica A. Church, Francis M. Miezin, Deanna M. Barch, Marcus E. Raichle, Steven E. Petersen, and Bradley L. Schlaggar. The maturing architecture of the brain's default network. *Proceedings of the National Academy of Sciences*, 105(10):4028–4032, 2008. doi: 10.1073/pnas.0800376105.

Joshua Faskowitz, Farnaz Zamani Esfahlani, Youngheun Jo, Olaf Sporns, and Richard F. Betzel. Edge-centric functional network representations of human cerebral cortex reveal overlapping system-level architecture. *Nature Neuroscience*, 23:1644–1654, 2020. doi: 10.1038/s41593-020-00719-y.

Eric Feczko, Greg Conan, Scott Marek, Brenden Tervo-Clemmens, Michaela Cordova, Olivia Doyle, et al. Adolescent brain cognitive development (ABCD) community MRI collection and utilities. *bioRxiv*, 2021. doi: 10.1101/2021.07.09.451638.

Jianqiao Ge, Guoyuan Yang, Meizhen Han, Sizhong Zhou, Weiwei Men, Lang Qin, Bingjiang Lyu, Hai Li, Haobo Wang, Hengyi Rao, et al. Increasing diversity in connectomics with the chinese human connectome project. *Nature Neuroscience*, 26(1):163–172, 2023. doi: 10.1038/s41593-022-01215-1.

Sam Gijsen, Marc-Andre Schulz, and Kerstin Ritter. Brain-Semantoks: Learning semantic tokens of brain dynamics with a self-distilled foundation model. In *The Fourteenth International Conference on Learning Representations*, pp. 13186–13208, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/162aa614b2684b3724375241195f55b4-Abstract-Conference.html.

Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, Bilal Piot, Koray Kavukcuoglu, Rémi Munos, and Michal Valko. Bootstrap your own latent: A new approach to self-supervised learning. In *Advances in Neural Information Processing Systems*, volume 33, pp. 21271–21284, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/f3ada80d5c4ee70142b17b8192b2958e-Abstract.html.

Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick. Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16000–16009, 2022.

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. doi: 10.1016/j.neuroimage.2013.05.079.

Xuan Kan, Wei Dai, Hejie Cui, Zilong Zhang, Ying Guo, and Carl Yang. Brain network transformer. In *Advances in Neural Information Processing Systems*, volume 35, pp. 25586–25599, 2022. doi: 10.52202/068431-1855. URL https://proceedings.neurips.cc/paper_files/paper/2022/hash/a408234a9b80604a9cf6ca518e474550-Abstract-Conference.html.

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. doi: 10.1016/j.neuroimage.2016.09.046.

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, pp. 4314–4327, 2021.

Xiaoxiao Li, Yuan Zhou, Nicha Dvornek, Muhan Zhang, Siyuan Gao, Juntang Zhuang, Dustin Scheinost, Lawrence H. Staib, Pamela Ventola, and James S. Duncan. BrainGNN: Interpretable brain graph neural network for fMRI analysis. *Medical Image Analysis*, 74: 102233, 2021. doi: 10.1016/j.media.2021.102233.

Kenneth Marek, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, Tanya Simuni, Chris Coffey, Karl Kieburtz, Emily Flagg, Sohini Chowdhury, et al. The parkinson progression marker initiative (PPMI). *Progress in Neurobiology*, 95(4):629–635, 2011. doi: 10.1016/j.pneurobio.2011.09.005.

Stephanie Noble, Dustin Scheinost, and R. Todd Constable. A decade of test-retest reliability of functional connectivity: A systematic review and meta-analysis. *NeuroImage*, 203: 116157, 2019. doi: 10.1016/j.neuroimage.2019.116157.

Luke J. Norman, Gustavo Sudre, Jolie Price, Gauri G. Shastri, and Philip Shaw. Evidence from "big data" for the default-mode hypothesis of ADHD: a mega-analysis of multiple large samples. *Neuropsychopharmacology*, 48(2):281–289, 2023. doi: 10.1038/s41386-022-01408-z.

Josue Ortega Caro, Antonio Henrique de Oliveira Fonseca, Syed Rizvi, Matteo Rosati, Christopher Averill, James Cross, Prateek Mittal, Emanuele Zappala, Rahul Dhodapkar, Chadi Abdallah, and David van Dijk. BrainLM: A foundation model for brain activity recordings. In *The Twelfth International Conference on Learning Representations*, pp. 565–576, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/hash/029ce70401321de3808b3ac39e1ab167-Abstract-Conference.html.

Theodore D. Satterthwaite, Mark A. Elliott, Kosha Ruparel, James Loughead, Karthik Prabhakaran, Monica E. Calkins, Ryan Hopson, Chad Jackson, Jack Keefe, Marisa Riley, et al. Neuroimaging of the philadelphia neurodevelopmental cohort. *NeuroImage*, 86: 544–553, 2014. doi: 10.1016/j.neuroimage.2013.07.064.

Alexander Schaefer, Ru Kong, Evan M. Gordon, Timothy O. Laumann, Xi-Nian Zuo, Avram J. Holmes, Simon B. Eickhoff, and B. T. Thomas Yeo. Local-global parcellation of the human cerebral cortex from intrinsic functional connectivity mri. *Cerebral Cortex*, 28(9):3095–3114, 2018. doi: 10.1093/cercor/bhx179.

Lukas Snoek, Maite M. van der Miesen, Tinka Beemsterboer, Andries van der Leij, Annemarie Eigenhuis, and H. Steven Scholte. The amsterdam open MRI collection, a set of multimodal MRI datasets for individual difference analyses. *Scientific Data*, 8:85, 2021. doi: 10.1038/s41597-021-00870-6.

Chandra Sripada, Daniel Kessler, Yu Fang, Robert C. Welsh, Krishan Prem Kumar, and Michael Angstadt. Disrupted network architecture of the resting brain in attention-deficit/hyperactivity disorder. *Human Brain Mapping*, 35(9):4693–4705, 2014a. doi: 10.1002/hbm.22504.

Chandra S. Sripada, Daniel Kessler, and Mike Angstadt. Lag in maturation of the brain's intrinsic functional architecture in attention-deficit/hyperactivity disorder. *Proceedings of the National Academy of Sciences*, 111(39):14259–14264, 2014b. doi: 10.1073/pnas.1407787111.

Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Olivier Etard, Nicolas Delcroix, Bernard Mazoyer, and Marc Joliot. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1):273–289, 2002. doi: 10.1006/nimg.2001.0978.

David C. Van Essen, Stephen M. Smith, Deanna M. Barch, Timothy E. J. Behrens, Essa Yacoub, and Kamil Ugurbil. The WU-Minn human connectome project: An overview. *NeuroImage*, 80:62–79, 2013. doi: 10.1016/j.neuroimage.2013.05.041.

Mo Wang, Junfeng Xia, Wenhao Ye, Enyu Liu, Kaining Peng, Jianfeng Feng, Quanying Liu, and Hongkai Wen. SLIM-Brain: A data- and training-efficient foundation model for fMRI data analysis. *arXiv preprint arXiv:2512.21881*, 2025.

Mo Wang, Wenhao Ye, Junfeng Xia, Junxiang Zhang, Xuanye Pan, Minghao Xu, Haotian Deng, Hongkai Wen, and Quanying Liu. Omni-fMRI: A universal atlas-free fMRI foundation model. *arXiv preprint arXiv:2601.23090*, 2026.

Dongtao Wei, Kaixiang Zhuang, Lei Ai, Qunlin Chen, Wenjing Yang, Wei Liu, Kangcheng Wang, Jiangzhou Sun, and Jiang Qiu. Structural and functional brain scans from the cross-sectional southwest university adult lifespan dataset. *Scientific Data*, 5:180134, 2018. doi: 10.1038/sdata.2018.134.

Xinxu Wei, Kanhao Zhao, Yong Jiao, Lifang He, and Yu Zhang. A brain graph foundation model: Pre-training and prompt-tuning across broad atlases and disorders. In *The Fourteenth International Conference on Learning Representations*, pp. 156800–156829, 2026a. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/fdf5965c14708e90eba7f8f77d24ff4c-Abstract-Conference.html.

Ziquan Wei, Tingting Dan, and Guorong Wu. Large connectome model: An fMRI foundation model of brain connectomes empowered by brain-environment interaction in multi-task learning landscape. *Proceedings of the AAAI Conference on Artificial Intelligence*, 40(3):2155–2163, 2026b. doi: 10.1609/aaai.v40i3.37198.

Michael W. Weiner, Paul S. Aisen, Clifford R. Jack, William J. Jagust, John Q. Trojanowski, Leslie Shaw, Andrew J. Saykin, John C. Morris, Nigel Cairns, Laurel A. Beckett, et al. The alzheimer's disease neuroimaging initiative: Progress report and future plans. *Alzheimer's & Dementia*, 6(3):202–211.e7, 2010. doi: 10.1016/j.jalz.2010.03.007.

Junfeng Xia, Wenhao Ye, Junxiang Zhang, Xuanye Pan, Mo Wang, and Quanying Liu. BrainWorld: A structural-prior-conditioned generative model for whole-brain 4d fMRI dynamics. *arXiv preprint arXiv:2606.17742*, 2026.

Chao-Gan Yan, Xiao Chen, Le Li, F. Xavier Castellanos, Tong-Jian Bai, Qi-Jing Bo, Jun Cao, Guan-Mao Chen, Ning-Xuan Chen, Wei Chen, et al. Reduced default mode network functional connectivity in patients with recurrent major depressive disorder. *Proceedings of the National Academy of Sciences*, 116(18):9078–9083, 2019. doi: 10.1073/pnas.1900390116.

Yanwu Yang, Chenfei Ye, Guinan Su, Ziyao Zhang, Zhikai Chang, Hairui Chen, Piu Chan, Yue Yu, and Ting Ma. BrainMASS: Advancing brain network analysis for diagnosis with large-scale self-supervised learning. *IEEE Transactions on Medical Imaging*, 43(11):4004–4016, 2024. doi: 10.1109/TMI.2024.3414476.

B. T. 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. doi: 10.1152/jn.00338.2011.

Farnaz Zamani Esfahlani, Youngheun Jo, Joshua Faskowitz, Lisa Byrge, Daniel P. Kennedy, Olaf Sporns, and Richard F. Betzel. High-amplitude cofluctuations in cortical activity drive functional connectivity. *Proceedings of the National Academy of Sciences*, 117(45): 28393–28401, 2020. doi: 10.1073/pnas.2005531117.

A Experimental Details

A.1 Datasets and downstream tasks

Pretraining datasets. The pretraining corpus contained 13,444 usable subjects from eight dataset groups. Only fMRI inputs, without downstream labels, were used during pretraining.

Preprocessing. We started from cohort-specific preprocessed derivatives: the HCP minimal preprocessing pipeline for HCP and CHCP, ABCD-HCP derivatives for ABCD and ABCC, fMRIPrep derivatives for AOMIC PIOP1/2, and available MNI-space preprocessed derivatives for HBN, PNC, and the multisite MDD data. Thus, low-level operations, including motion correction and any nuisance-signal handling, were inherited from the corresponding source pipeline rather than re-estimated with a single cross-cohort pipeline. We then applied a common input harmonization procedure: all fMRI volumes were represented at 2-mm isotropic resolution, and only datasets with a source TR different from 0.72 s were temporally interpolated to 0.72 s using cubic B-spline interpolation. Regional signals were extracted using the Schaefer-400 atlas, and each regional time series was Z-standardized along the temporal dimension. Files with exactly 200 frames were used directly; during pretraining, contiguous 200-frame windows were sampled from longer series, whereas sequences shorter than 200 frames were excluded without zero-padding.

Chinese Human Connectome Project (CHCP). CHCP contributed healthy-adult resting-state scans acquired with multiple phase-encoding directions (Ge et al., 2023).

Human Connectome Project (HCP). HCP contributed young-adult resting-state scans to pretraining (Van Essen et al., 2013). It also provided an internal downstream task for sex classification; the original data partitions remained separated by subject.

Adolescent Brain Cognitive Development (ABCD). ABCD contributed pediatric resting-state scans to pretraining (Casey et al., 2018). A separate internal downstream evaluation predicted sex, with participant identities kept disjoint across the original partitions.

ABCC Data Share. ABCC contributed a large pediatric imaging collection (Feczko et al., 2021). Subjects overlapping with ABCD were removed from ABCC before constructing the pretraining manifest.

Healthy Brain Network (HBN). HBN contributed pediatric and adolescent scans spanning a broad range of mental-health and learning characteristics (Alexander et al., 2017).

Philadelphia Neurodevelopmental Cohort (PNC). PNC contributed neurodevelopmental resting-state scans from youth and young adults (Satterthwaite et al., 2014).

Major Depressive Disorder (MDD). The multisite MDD collection contributed resting-state scans from participants with recurrent depression and controls (Yan et al., 2019).

Amsterdam Open MRI Collection (AOMIC). AOMIC PIOP1 and PIOP2 contributed young-adult resting-state scans (Snoek et al., 2021).

Downstream datasets and tasks. Seven datasets provided nine downstream tasks. HCP and ABCD were treated as in-cohort evaluations because their unlabeled scans contributed to pretraining; no downstream labels were used. The remaining cohorts were excluded from the pretraining mixture and provided out-of-cohort transfer evaluations.

Parkinson's Progression Markers Initiative (PPMI). PPMI was used for three-class diagnostic-group classification (Marek et al., 2011).

Alzheimer's Disease Neuroimaging Initiative (ADNI). ADNI provided two binary classification tasks: Alzheimer's disease and mild cognitive impairment (Weiner et al., 2010).

Autism Brain Imaging Data Exchange (ABIDE). ABIDE was used for autism-spectrum-disorder classification (Di Martino et al., 2014). ABIDE and ABIDE2 were both excluded from pretraining.

ADHD-200. ADHD-200 provided ADHD diagnosis classification and age prediction (ADHD-200 Consortium, 2012). The complete cohort was excluded from pretraining.

Southwest University Adult Lifespan Dataset (SALD). SALD was used for lifespan age prediction and was excluded from pretraining (Wei et al., 2018).

For datasets distributed with predefined partitions, we combined the available samples before constructing participant-grouped cross-validation folds (Table 6).

Table 6. **Downstream evaluation cohorts and prediction targets.** Participants are grouped across all available segment-level examples when constructing cross-validation folds.

Dataset / task	Participants	Examples	Target
HCP sex	1,000	5,997	Binary sex
ABCD sex	500	999	Binary sex
PPMI group	474	1,896	Three diagnostic groups
ADNI AD	230	230	AD vs. control
ADNI MCI	292	292	MCI vs. control
ABIDE ASD	173	173	ASD vs. control
ADHD200 diagnosis	646	1,923	ADHD vs. control
SALD age	493	493	Chronological age
ADHD200 age	646	1,923	Chronological age

A.2 Baseline models

The linear-probing comparison used four ROI-based fMRI representation models. Each encoder was kept frozen, and its scan-level representation was evaluated with the same task-specific linear-probe family used for CoFlu-Brain. Each model used its native input parcellation and feature extraction pathway.

BrainGFM. BrainGFM pretrains a shared graph encoder across multiple parcellations using graph contrastive learning and masked graph autoencoding (Wei et al., 2026a). We used the authors' released pretrained checkpoint with Schaefer-100 inputs (Schaefer et al., 2018). Raw Pearson-FC rows served as node features, and FC values above 0.3 defined the graph adjacency. The encoder's graph-level output was used as the frozen scan representation.

BrainMASS. BrainMASS constructs augmented FC views by dropping BOLD observations and learns with masked-ROI modeling and latent representation alignment (Yang et al., 2024). In CoFlu-Brain, each partition forms a complementary retained–dropped FC pair; multiple retained estimates are aligned, and each paired dropped-view representation is predicted. We used the authors' released Schaefer-100 checkpoint (Schaefer et al., 2018) and released downstream feature path: Transformer-encoded FC rows followed by dimensionality reduction and flattening.

BrainLM. BrainLM is a masked-autoencoding foundation model for regional fMRI time series (Ortega Caro et al., 2024). We used the released 111M-parameter ViT-MAE checkpoint with its native 424-region input. The concatenation of the final class token and mean patch representation was used as the frozen scan feature.

Large Connectome Model. The Large Connectome Model (LCM) learns connectome representations through multitask brain–environment interactions (Wei et al., 2026b). We used the released AAL-116 backbone and decoder checkpoints (Tzourio-Mazoyer et al., 2002),

computed raw Pearson FC following the released input convention, and mean-pooled the final decoder-query states to obtain a frozen scan representation.

A.3 Configuration

CoFlu-Brain was pretrained from scratch using two-rank distributed training. The model, optimization, and frozen-probe configurations are summarized in Table 7.

Table 7. **CoFlu-Brain model, pretraining, and linear-probe configurations.**

Configuration	Value
Model configuration	
Input representation	
ROI atlas	Schaefer-400
Input segment length	200 frames
FC-row embedder	
MLP dimensions	$400 \to 512 \to 128 \to 512$
ViT encoder	
Encoder depth	6 blocks
Embedding dimension	512
Attention heads	8
Readout	Mean pooling over ROI tokens
Projection and prediction heads	
Projector MLP dimensions	$512 \to 2048 \to 2048 \to 128$
Predictor MLP dimensions	$128 \to 512 \to 128$
Pretraining configuration	
Optimizer	AdamW
Learning rate	3×10^{-4}
Weight decay	0.1
Dropout	0.2
Gradient clipping	5.0
Total pair batch size	64
Training epochs	150
Optimizer updates	9,600
Temporal partitions	4
Retained-frame ratio	$U[0.4, 0.7]$
Consistency weight	0.5
SIGReg weight	0.1
Consistency schedule	Full weight at epoch 1; linear decay to zero at epoch 25
Checkpoint selection	Last checkpoint
Frozen-probe configuration	
Encoder	Frozen
Representation	Final encoder representation
Primary, temporal-sampling, and objective comparison folds	5
Other sensitivity-analysis folds	3
Fold construction	Participant-grouped and task-stratified
Fold seed	42
Evaluation cohort	All available downstream samples
Five-fold probe fitting	Four folds; evaluation on the remaining fold
Three-fold probe fitting	Two folds; evaluation on the remaining fold
Feature scaling	StandardScaler fitted on the training folds only
Classification model	Logistic regression, $C = 1$
Maximum iterations	2,000
Classification metrics	Accuracy and weighted F1
Age-regression model	Ridge regression, $\alpha = 1$
Age-target scaling	Training-fold mean and population standard deviation
Age-regression metric	$\text{nMAE} = \text{MAE}/\sigma_{\text{train}}$
Fold summary	Mean $\pm$ sample SD

B Expanded Main Results

B.1 Seed-wise linear-probing results

Table 8 reports the five-fold performance of each independently pretrained CoFlu-Brain encoder. The final column matches the seed-averaged result in Table 1.

Table 8. **CoFlu-Brain linear probing across pretraining seeds.** Values are mean ± sample SD across five participant-grouped folds. Metrics are weighted F1 (↑) and age nMAE (↓). The final column averages scores across the three pretraining seeds within each fold before computing mean ± SD.

Dataset \ Run	Seed 42	Seed 43	Seed 44	Seed average
HCP / Sex ↑	0.7865 ± 0.0113	0.7648 ± 0.0155	0.7722 ± 0.0142	0.7745 ± 0.0103
ABCD / Sex ↑	0.5899 ± 0.0184	0.5828 ± 0.0497	0.6157 ± 0.0182	0.5961 ± 0.0190
SALD / Age ↓	0.6982 ± 0.0586	0.6378 ± 0.0437	0.6353 ± 0.0373	0.6571 ± 0.0431
PPMI / Group ↑	0.5802 ± 0.0206	0.5785 ± 0.0314	0.5736 ± 0.0187	0.5775 ± 0.0173
ADNI / AD ↑	0.7219 ± 0.0657	0.6643 ± 0.0425	0.6643 ± 0.0492	0.6835 ± 0.0307
ADNI / MCI ↑	0.5461 ± 0.0553	0.5585 ± 0.0573	0.5567 ± 0.0524	0.5538 ± 0.0441
ABIDE / ASD ↑	0.6013 ± 0.1107	0.5623 ± 0.0553	0.5066 ± 0.0621	0.5567 ± 0.0476
ADHD200 / Age ↓	0.6523 ± 0.0315	0.6528 ± 0.0426	0.6620 ± 0.0455	0.6557 ± 0.0370
ADHD200 / Diagnosis ↑	0.5867 ± 0.0207	0.5836 ± 0.0175	0.5969 ± 0.0394	0.5891 ± 0.0151

B.2 Complete network-restricted probing results

Table 9 expands the top-five presentation in Figure 3 to all 17 Yeo networks. Each value is the mean and sample SD across the three independently pretrained CoFlu-Brain encoders. The table reports predictive performance from retaining subject-specific FC within one network while replacing all remaining entries with their probe-training mean.

Table 9. **Functional-network selection on ADHD200.** Gray shading marks the top-five networks identified for each task by network-restricted probing. Values are mean ± SD across three independent pretraining runs. Age prediction reports nMAE (↓); diagnosis reports weighted F1 (↑). ROI counts indicate network size.

Network	ROIs	Age nMAE ↓	Diagnosis F1 ↑
DefaultA	34	**0.6804 ± 0.0065**	**0.5965 ± 0.0165**
ContC	12	**0.6894 ± 0.0029**	0.5619 ± 0.0132
LimbicA	13	**0.6923 ± 0.0056**	0.5442 ± 0.0047
DefaultC	13	**0.7012 ± 0.0052**	0.5288 ± 0.0125
SomMotA	39	**0.7042 ± 0.0116**	0.5435 ± 0.0226
LimbicB	11	0.7196 ± 0.0138	0.5372 ± 0.0092
DorsAttnB	25	0.7245 ± 0.0099	0.5490 ± 0.0101
SomMotB	31	0.7262 ± 0.0187	0.5806 ± 0.0153
VisPeri	23	0.7314 ± 0.0180	**0.5857 ± 0.0032**
TempPar	16	0.7446 ± 0.0040	0.5032 ± 0.0179
SalVentAttnA	34	0.7447 ± 0.0089	**0.5886 ± 0.0192**
DefaultB	32	0.7520 ± 0.0100	**0.6043 ± 0.0117**
VisCent	24	0.7526 ± 0.0070	**0.5807 ± 0.0264**
DorsAttnA	27	0.7562 ± 0.0110	0.4899 ± 0.0170
SalVentAttnB	17	0.7912 ± 0.0284	0.5330 ± 0.0115
ContB	25	0.7934 ± 0.0020	0.5169 ± 0.0131
ContA	24	0.7966 ± 0.0048	0.5157 ± 0.0066

The complete table preserves the patterns summarized in the main text: DefaultA provided the strongest age-prediction result across all three seeds, whereas diagnosis was led by DefaultB and followed by DefaultA and SalVentAttnA.

C Ablation and Sensitivity Analyses

C.1 Random versus ordered frame partitioning

Table 10 reports the task-level results for random and ordered frame partitioning.

C.2 Objective components

Table 11 reports the complete results for all objective variants.

The complete objective led five of the nine tasks and achieved the highest mean classification F1 (0.6306). Removing prediction or SIGReg generally reduced transfer performance,

Table 10. **Downstream performance under random and ordered frame partitioning.** Each seed entry is mean ± sample SD across five participant-grouped folds. The final column averages scores across pretraining seeds within each fold before computing mean ± SD. Metrics are weighted F1 (↑) and age nMAE (↓). Red boldface marks the better seed-averaged score for each task.

(a) Random partitioning

Dataset \ Run	Seed 42	Seed 43	Seed 44	Seed average
HCP sex ↑	0.7865 ± 0.0113	0.7648 ± 0.0155	0.7722 ± 0.0142	**0.7745 ± 0.0103**
ABCD sex ↑	0.5899 ± 0.0184	0.5828 ± 0.0497	0.6157 ± 0.0182	**0.5961 ± 0.0190**
SALD age ↓	0.6982 ± 0.0586	0.6378 ± 0.0437	0.6353 ± 0.0373	**0.6571 ± 0.0431**
PPMI group ↑	0.5802 ± 0.0206	0.5785 ± 0.0314	0.5736 ± 0.0187	**0.5775 ± 0.0173**
ADNI AD ↑	0.7219 ± 0.0657	0.6643 ± 0.0425	0.6643 ± 0.0492	**0.6835 ± 0.0307**
ADNI MCI ↑	0.5461 ± 0.0553	0.5585 ± 0.0573	0.5567 ± 0.0524	**0.5538 ± 0.0441**
ABIDE ASD ↑	0.6013 ± 0.1107	0.5623 ± 0.0553	0.5066 ± 0.0621	**0.5567 ± 0.0476**
ADHD200 age ↓	0.6523 ± 0.0315	0.6528 ± 0.0426	0.6620 ± 0.0455	**0.6557 ± 0.0370**
ADHD200 diagnosis ↑	0.5867 ± 0.0207	0.5836 ± 0.0175	0.5969 ± 0.0394	**0.5891 ± 0.0151**

(b) Ordered partitioning

Dataset \ Run	Seed 42	Seed 43	Seed 44	Seed average
HCP sex ↑	0.7769 ± 0.0151	0.7585 ± 0.0301	0.7752 ± 0.0103	0.7702 ± 0.0138
ABCD sex ↑	0.5817 ± 0.0366	0.6118 ± 0.0402	0.5926 ± 0.0324	0.5954 ± 0.0171
SALD age ↓	0.7936 ± 0.0675	0.7126 ± 0.0689	0.7376 ± 0.0581	0.7480 ± 0.0345
PPMI group ↑	0.5596 ± 0.0236	0.5614 ± 0.0275	0.5786 ± 0.0293	0.5665 ± 0.0247
ADNI AD ↑	0.6467 ± 0.0477	0.6809 ± 0.0815	0.6988 ± 0.0696	0.6755 ± 0.0480
ADNI MCI ↑	0.4627 ± 0.0818	0.5566 ± 0.0443	0.5663 ± 0.0315	0.5285 ± 0.0164
ABIDE ASD ↑	0.4958 ± 0.1181	0.4996 ± 0.1064	0.5288 ± 0.0583	0.5081 ± 0.0636
ADHD200 age ↓	0.6479 ± 0.0467	0.6662 ± 0.0422	0.7124 ± 0.0455	0.6755 ± 0.0410
ADHD200 diagnosis ↑	0.5700 ± 0.0226	0.5782 ± 0.0207	0.5557 ± 0.0161	0.5680 ± 0.0103

Table 11. **Objective-component ablation across nine downstream tasks.** Each objective variant was pretrained independently with seed 42 under matched configurations and is reported as mean ± SD across five participant-grouped folds. Metrics are weighted F1 (↑) and age nMAE (↓). Red boldface marks the best mean for each task.

Dataset \ Objective	Full	No prediction	No consistency	No SIGReg
HCP / Sex ↑	**0.7874 ± 0.0115**	0.6565 ± 0.0187	0.7789 ± 0.0110	0.6942 ± 0.0153
ABCD / Sex ↑	0.5899 ± 0.0184	0.5383 ± 0.0174	**0.6140 ± 0.0260**	0.5511 ± 0.0379
PPMI / Group ↑	**0.5802 ± 0.0206**	0.5399 ± 0.0460	0.5576 ± 0.0285	0.5426 ± 0.0720
ADNI / AD ↑	**0.7219 ± 0.0657**	0.5023 ± 0.0850	0.6866 ± 0.1175	0.6065 ± 0.0667
ADNI / MCI ↑	0.5461 ± 0.0553	0.4872 ± 0.0901	**0.5501 ± 0.0705**	0.5377 ± 0.0410
ABIDE / ASD ↑	0.6013 ± 0.1107	**0.6082 ± 0.0749**	0.5994 ± 0.0835	0.4136 ± 0.0587
ADHD200 / Diagnosis ↑	**0.5872 ± 0.0215**	0.5348 ± 0.0309	0.5725 ± 0.0384	0.5643 ± 0.0288
SALD / Age ↓	0.6982 ± 0.0586	0.7674 ± 0.0551	**0.6778 ± 0.0783**	0.7625 ± 0.0577
ADHD200 / Age ↓	**0.6523 ± 0.0315**	0.7416 ± 0.0567	0.6611 ± 0.0539	0.7132 ± 0.0482

while consistency produced task-dependent gains. Without SIGReg, representations became highly redundant, with a view cosine of 0.9999999 and a covariance penalty of 639.19, compared with 8.81 for the complete objective.

C.3 Pretraining data coverage

Table 12 reports every downstream task for all data scales; aggregate means are omitted because the task-level pattern is more informative.

Table 12. **Pretraining data coverage and downstream performance across nine tasks.** Values are mean ± SD across probe folds. Metrics are weighted F1 (↑) and age nMAE (↓). Red boldface marks the best mean for each task.

Dataset \ Coverage	12.5%	25%	50%	100%
HCP / Sex ↑	0.7265 ± 0.0126	0.7376 ± 0.0117	0.7326 ± 0.0240	**0.7737 ± 0.0092**
ABCD / Sex ↑	0.5947 ± 0.0266	**0.6153 ± 0.0383**	0.5905 ± 0.0332	0.5604 ± 0.0149
PPMI / Group ↑	0.5204 ± 0.0402	**0.5844 ± 0.0347**	0.5625 ± 0.0310	0.5562 ± 0.0582
ADNI / AD ↑	0.6627 ± 0.0137	**0.7280 ± 0.0348**	0.6786 ± 0.0301	0.7235 ± 0.0336
ADNI / MCI ↑	**0.6321 ± 0.0267**	0.5847 ± 0.0322	0.5884 ± 0.0090	0.5967 ± 0.0422
ABIDE / ASD ↑	0.4414 ± 0.0404	0.5256 ± 0.0614	0.5144 ± 0.0275	**0.5427 ± 0.0500**
ADHD200 / Diagnosis ↑	0.5678 ± 0.0134	0.5609 ± 0.0097	0.5314 ± 0.0333	**0.5756 ± 0.0298**
SALD / Age ↓	0.9158 ± 0.0296	0.7858 ± 0.0347	0.7831 ± 0.0281	**0.7610 ± 0.0802**
ADHD200 / Age ↓	0.7553 ± 0.0318	0.7453 ± 0.0341	0.7152 ± 0.0174	**0.6785 ± 0.0224**

Increasing data coverage improves both age-regression tasks and five of seven classification tasks from 12.5% to 100%. The individual classification responses remain nonmonotonic.

C.4 Number of cofluctuation views

Table 13 reports every downstream task for all branch-count variants. The main text reports the four-task summary.

Table 13. **Temporal branch-count comparison across nine downstream tasks.** Values are mean ± SD across probe folds. Metrics are weighted F1 (↑) and age nMAE (↓). Red boldface marks the best mean for each task.

Dataset \ Branches	2 branches	4 branches	6 branches	8 branches
HCP / Sex ↑	0.7557 ± 0.0205	**0.7737 ± 0.0092**	0.7601 ± 0.0296	0.7643 ± 0.0122
ABCD / Sex ↑	0.5713 ± 0.0092	0.5604 ± 0.0149	0.5713 ± 0.0444	**0.5895 ± 0.0308**
PPMI / Group ↑	0.5220 ± 0.0446	0.5562 ± 0.0582	**0.5846 ± 0.0373**	0.5827 ± 0.0331
ADNI / AD ↑	0.6518 ± 0.0536	**0.7235 ± 0.0336**	0.6429 ± 0.0818	0.6686 ± 0.0470
ADNI / MCI ↑	0.5514 ± 0.0207	**0.5967 ± 0.0422**	0.5357 ± 0.0371	0.5617 ± 0.0287
ABIDE / ASD ↑	0.4849 ± 0.0917	**0.5427 ± 0.0500**	0.5274 ± 0.0394	0.4959 ± 0.0208
ADHD200 / Diagnosis ↑	**0.5762 ± 0.0387**	0.5756 ± 0.0298	0.5637 ± 0.0183	0.5713 ± 0.0269
SALD / Age ↓	0.7605 ± 0.0446	0.7610 ± 0.0802	**0.7139 ± 0.0128**	0.7256 ± 0.0274
ADHD200 / Age ↓	0.7088 ± 0.0091	**0.6785 ± 0.0224**	0.6872 ± 0.0195	0.6894 ± 0.0338

Pretraining cost increased from 10.54 GPU-hours for four branches to 26.72 GPU-hours for eight branches. Performance varied nonmonotonically across tasks, with no consistent benefit from increasing the branch count beyond four. All runs completed the same 9,600-update budget.

C.5 Consistency-weight schedule

Table 14 expands the three schedules compared in the main text to all nine downstream tasks.

Table 14. **Consistency-weight schedule comparison across nine downstream tasks.** Values are mean ± SD across probe folds. Metrics are weighted F1 (↑) and age nMAE (↓). Red boldface marks the best mean for each task.

Dataset \ Schedule	Decay-25	Decay-50	Constant
HCP / Sex ↑	**0.7737 ± 0.0092**	0.7588 ± 0.0204	0.7687 ± 0.0164
ABCD / Sex ↑	0.5604 ± 0.0149	**0.5955 ± 0.0349**	0.5916 ± 0.0442
PPMI / Group ↑	**0.5562 ± 0.0582**	0.5179 ± 0.0120	0.5540 ± 0.0048
ADNI / AD ↑	**0.7235 ± 0.0336**	0.6736 ± 0.0236	0.6627 ± 0.0350
ADNI / MCI ↑	**0.5967 ± 0.0422**	0.5636 ± 0.0107	0.5431 ± 0.0253
ABIDE / ASD ↑	0.5427 ± 0.0500	0.4734 ± 0.0243	**0.5669 ± 0.0179**
ADHD200 / Diagnosis ↑	**0.5756 ± 0.0298**	0.5433 ± 0.0282	0.5688 ± 0.0300
SALD / Age ↓	0.7610 ± 0.0802	0.7607 ± 0.0261	**0.7576 ± 0.0617**
ADHD200 / Age ↓	0.6785 ± 0.0224	0.6839 ± 0.0239	**0.6568 ± 0.0270**

Decay-25 led five of the seven classification tasks, Decay-50 led ABCD sex, and Constant led ABIDE ASD and both age-prediction tasks. The complete results preserve the task-dependent schedule pattern summarized in the main text.