BRAINNORM: LEARNING THE HEALTHY BRAIN REPRESENTATIONS FROM STRUCTURAL MRI

Anonymous authors
Paper under double-blind review

ABSTRACT

We introduce **BrainNorm**, a normative foundation model trained and evaluated on ∼66,000 T1-weighted structural MRI (T1w sMRI) scans. Trained exclusively on healthy subjects, BrainNorm learns a *Semantic Atlas Latent* (SAL) space in which anatomically grounded parcel representations are jointly organized by anatomical identity and age. Through contrastive alignment with parcel-specific, age-conditioned healthy reference templates, SAL provides an explicit geometry of healthy anatomical variation that supports age-consistent template matching and represents *localized deviations* directly in the latent space. Across multiple downstream cohorts, BrainNorm is evaluated on various tasks spanning age estimation, brain-age gap estimation, parcel identification, and single- & multi-disease classification. The resulting deviation patterns in SAL space enable zero-shot disease prediction from parcel-wise abnormalities and exhibit regional patterns that closely align with findings across neurodegenerative diseases reported in the clinical literature. Fine-tuning on healthy-only subjects from downstream cohorts further improves performance across various tasks. Across all classification tasks, linear probing on BrainNorm's frozen embeddings outperforms nine baselines fine-tuned with end-to-end downstream supervision. Together, these results demonstrate that learning an anatomically structured, age-conditioned geometry of healthy brain variation yields transferable representations for detecting and characterizing departures from expected brain structure.

1 INTRODUCTION

The pathology of abnormality can be better understood in relation to the normal, and so one must first map the boundaries of normal to discover what lies outside. Normative modeling makes this principle operational for neuroimaging. Given a large healthy reference cohort, it learns the distribution of typical brain structure and asks a clinically sharper question: *where does this individual fall relative to what is expected?* This framework has long been used in medicine to model "normal" variation, most familiarly through pediatric growth charts and centile curves (Bethlehem et al. (2022); Schabdach et al. (2023); Cole (2012); Marquand et al. (2016)). Its relevance is particularly strong in neurodegeneration, where large-scale healthy cohorts are more readily available than well-annotated disease cohorts. Many neurological diseases begin years, sometimes decades, before clinical symptoms appear, emerging as subtle and spatially localized departures from otherwise normal aging trajectories (Young et al. (2018); Bethlehem et al. (2022)). Constructing a reliable normative reference is therefore fundamentally linked to the problem of detecting neurodegenerative anomalies early and at the individual level. Classical approaches relied on probabilistic regression over morphometric features (e.g., Gaussian process and Bayesian normative models), later extended to large multi-site cohorts and lifespan-scale brain charts (Wolfers et al. (2018; 2020); Fraza et al. (2021); Rutherford et al. (2022); Kia et al. (2022); Bethlehem et al. (2022); Ge et al. (2024)). However, they remain constrained by hand-crafted regional summaries and limited representation capacity.

Existing learning-based approaches to normative inference largely fall into two paradigms, each with structural limitations. The first is *brain-age prediction*, where a model is trained on healthy subjects to estimate chronological age from brain morphology, and the brain-age gap (BAG), difference between predicted and actual age, is interpreted as an anomaly score (Cole et al. (2017); Peng et al. (2021); Lee et al. (2022)). While effective as a global biomarker, this formulation collapses complex spatial variation into a single scalar. The BAG can indicate that a brain appears atypical for its age,

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.

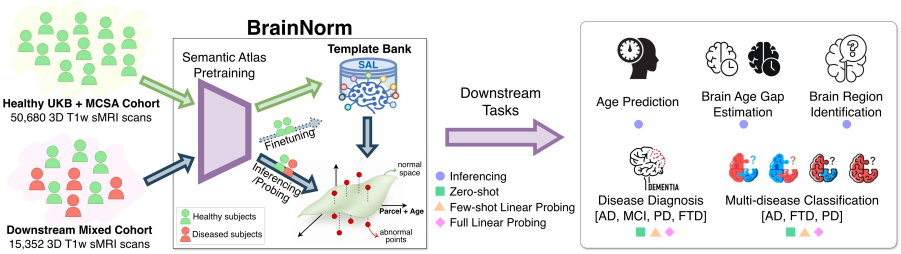


Figure 1: **BrainNorm Framework:** Trained exclusively on healthy subjects, BrainNorm learns (i) age- and anatomy-structured normative scan representations in the *Semantic Atlas Latent* (SAL) space and (ii) parcel-specific, age-conditioned healthy reference templates. By matching the frozen embeddings to corresponding templates, BrainNorm enables zero-shot, few-shot, and full linear probing for downstream prediction and classification tasks on mixed clinical cohorts.

but provides little insight into *where* deviations occur or *how* they manifest across regions. Distinct diseases with different regional atrophy patterns may yield similar brain-age gaps, making the score non-specific and insensitive to spatial heterogeneity.

The second paradigm employs generative models, such as Variational Autoencoders (VAEs) (Pinaya et al. (2019); Kumar et al. (2023); Lawry Aguila et al. (2022); Pinaya et al. (2022b); Zimmerer et al. (2019)) and Diffusion architectures (Pinaya et al. (2022a); Frotscher et al. (2025); Puglisi et al. (2024)), to learn latent distributions of healthy anatomy. These models identify anomalies by detecting where voxel-space reconstruction fails. While intuitive, this approach is computationally expensive and brittle. Voxel-space residuals can reflect not only pathology but also scanner effects, protocols, or registration variability. Similarly, *conditional/ deformable template matching* techniques (Lorenzi et al. (2015); Fu et al. (2025); Dalca et al. (2019); Sivera et al. (2019); Wang et al. (2025)) replace generative reconstruction with age-conditioned anatomical templates, but the underlying operation remains unchanged, i.e., voxel-wise deformation estimation in image space. Consequently, both approaches struggle to disentangle true structural atrophy/ deviation from imaging artifacts, often requiring additional modeling to filter out inherited noise.

Taken together, these observations point to a common bottleneck: the challenge lies not merely in detecting deviations, but in the space where the comparison is performed. Normative modeling does not require reproducing voxels; it can instead compare representations designed to preserve anatomically meaningful structure while reducing dependence on voxel-level correspondence. Recent advances in representation learning have moved towards prediction in representation space rather than pixel reconstruction (Assran et al. (2023); Bardes et al. (2024)). This motivates reframing normative inference from image space to an *anatomically structured latent space*, where meaningful regions are explicitly represented, and deviations can be evaluated locally.

We introduce **BrainNorm**, a normative foundation model, which addresses these limitations through representation learning via *Semantic Atlas Pretraining* (Fig 1). To the best of our knowledge, this is the first normative foundation model trained and tested on a large-scale cohort of $66,032$ 3D T1w sMRI scans from $49{,}579$ subjects. Departing from traditional voxel-level objectives, our framework learns a *Semantic Atlas Latent Space (SAL)*, an anatomically factorized normative latent space, that is jointly organized by anatomy and age, together with parcel-specific, age-conditioned healthy reference templates. Within this space, we evaluate: *how similar is a scan to healthy brains of the same age, and which anatomical parcels deviate from this expectation?* Without disease supervision, SAL space directly supports: (i) parcel-wise normative matching against age-consistent healthy templates, (ii) retrieval-based subject age and brain age-gap estimation, and (iii) zero-shot deviation analysis that localizes neurodegenerative anomalies across anatomical regions, enabling single & multi-disease classifications, and parcel-specific interpretability. Further, we demonstrate that BrainNorm achieves superior performance with simple linear probing on frozen/ healthy-only finetuned representations, compared to baselines finetuned with end-to-end downstream supervision.

2 Related Works

Age prediction: The idea that brain structure encodes biological aging predates deep learning: BrainAGE (Franke et al. (2010)) trained kernel models on morphometric features to predict chrono-

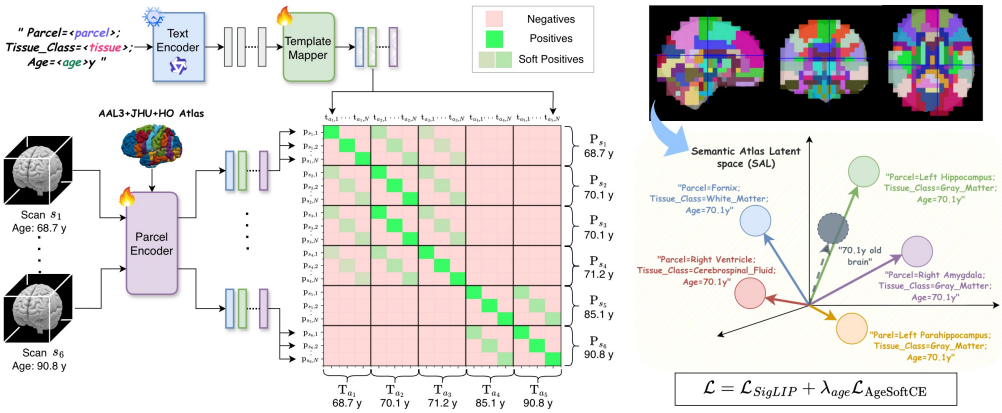


Figure 2: **BrainNorm pretraining framework** for a sample batch of scans. BrainNorm encodes a scan as a set of parcel embeddings in SAL space, rather than compressing it into a single global representation. These embeddings are aligned to an age-indexed, parcel-specific healthy template bank initialized from structured textual priors and refined through a learnable Template Mapper.

logical age, using the Brain-Age Gap as a deviation index (Cole et al. (2018)). Deep learning then shifted the field toward supervised age regression directly from 3D T1w volumes, first with 3D CNNs (Cole et al. (2017); Peng et al. (2021); Lee et al. (2022)) and more recently with 3D ViTs (Zhang et al. (2025a)). However, optimizing strictly for age prediction can reduce sensitivity to disease-related variation (Schulz et al. (2025)), motivating approaches that model age-conditioned normality while preserving deviations from it.

Contrastive Pretraining: Contrastive self-supervision has proved effective for learning representations without explicit labels, motivating its adoption in medical imaging (Ouyang et al. (2022b;a); Zhao et al. (2021)). In neuroimaging, recent works such as y-Aware (Dufumier et al. (2021)), AnatCL (Barbano et al. (2025)) and BrainIAC (Tak et al. (2026)) show that contrastive learning can capture biologically meaningful representations from brain MRI. In parallel, medical vision-language pretraining aligns images with reports or captions (e.g., MedCLIP (Wang et al. (2022)), MedSigLIP (Sellergren et al. (2025)), BioViL-T (Bannur et al. (2023)), BiomedCLIP (Zhang et al. (2025b))), enabling retrieval and zero-shot transfer but typically relying on image-report supervision. Finally, CLIP-style objectives have evolved for better scalability. Among them, SigLIP (Zhai et al. (2023)) replaces softmax-normalized InfoNCE with a sigmoid-based pairwise alignment, naturally accommodating multiple positives and reducing large-batch dependence. We build on this objective to learn age- and anatomy-structured normative representations while retaining the core benefit of contrastive alignment in a shared semantic space.

3 BrainNorm

3.1 Overview

BrainNorm learns a *Semantic Atlas Latent (SAL)* space from healthy cohorts through contrastive pretraining, which is explicitly structured at the anatomical parcel level and ordered by age (Fig 2). Its central objective is to construct an age-indexed normative template bank: $\{\mathbf{t}_{a,p}\}_{a=A_{min},\ldots,A_{max};\ p=1,\ldots,N}$, where each parcel p is associated with an age-indexed trajectory of reference embeddings. These trajectories characterize the expected healthy representation of each parcel across age a. BrainNorm enables: (i) age-consistent matching by alignment with the corresponding healthy template, and (ii) localized deviation modeling by representing parcel-wise departures from age-conditioned healthy references directly in the latent space. By explicitly modeling healthy reference trajectories, BrainNorm transforms normative modeling into a structured representation-learning problem: instead of directly predicting age or disease labels, it learns the geometry of healthy anatomical variation.

BrainNorm operates at parcel-level representations, where each scan i of age a_i, is represented by a set of anatomical parcel embeddings $\{\mathbf{p}_{i,p}\}_{p=1}^{N}$ obtained using a dedicated transformer-based Parcel Encoder (adapted from Venkatesh et al. (2026)), rather than a single global embedding. The

Parcel Encoder uses the atlas to identify patch tokens that overlap each anatomical parcel, forming a parcel-specific token set that defines the parcel's context. Each set is processed independently by a shared Transformer encoder followed by token pooling to produce a parcel-specific representation (details in Appendix Sec B.2). Given these parcel embeddings, the central question becomes: *what geometric structure should a normative latent space exhibit?* BrainNorm's normative reasoning imposes two fundamental constraints in the SAL space:

(i) Parcel specificity: For a parcel embedding $\mathbf{p}_{i,p}$, alignment must be strongest with templates of the same anatomical parcel ($p' = p$), while remaining well separated from other parcels. This ensures anatomical identifiability.

(ii) Age continuity: Healthy aging is continuous. Within a parcel, representations should vary smoothly with age, tolerating small age differences while progressively diverging as the age gap increases. Templates should therefore form ordered trajectories rather than isolated clusters.

To seed this structure, we introduce textual representations as *priors* over the template bank. For each parcel and age, we construct structured texts encoding age and anatomical identity from the brain atlas. We employ Qwen3 (Yang et al. (2025)) embeddings for their strongly structured latent geometry, which exhibit clustering by anatomical semantics while encoding ordinal age signals (see Appendix Sec B.5.4). However, text geometry alone does not guarantee consistency with imaging-derived biological aging trajectories, since brain aging is region-specific, i.e., different parcels follow distinct structural trajectories across the lifespan (Resnick et al. (2003); Marjańska et al. (2017); Lee et al. (2019)). We therefore introduce a lightweight learnable *template mapper* that "bends" the text priors into an age-indexed template bank aligned with parcel embeddings. This provides a principled inductive bias: initialization from a semantically organized language space, followed by alignment using large-scale imaging evidence.

The SAL space is therefore trained to exhibit two properties: (i) parcel embeddings are organized by parcel identity, and (ii) within each parcel, representations follow smooth age-ordered trajectories. To enforce this geometry, we adopt a parcel-specific and age-aware pairwise alignment objective inspired by SigLIP (Zhai et al. (2023)). Cross-parcel pairs serve as negatives to preserve anatomical separation, while within-parcel nearby-age neighborhoods form *soft* positives, in which alignment strength decays smoothly with age difference, encouraging continuous progression across age.

3.2 Semantic Atlas Pretraining

3.2.1 Setup and representations

Parcel embeddings from 3D scans: For each subject i and anatomical parcel $p \in \{1, \ldots, N\}$, a Parcel Encoder produces a representation $\mathbf{z}_{i,p} \in \mathbb{R}^{d'}$, which is projected into a shared SAL space of dimension D and ℓ_2-normalized:

$$\mathbf{p}_{i,p} = \frac{\mathrm{Proj}(\mathbf{z}_{i,p})}{\|\mathrm{Proj}(\mathbf{z}_{i,p})\|_2} \in \mathbb{R}^D, \qquad \|\mathbf{p}_{i,p}\|_2 = 1 \tag{1}$$

The projection layer $\mathrm{Proj}(\cdot)$ enables alignment between image-derived parcel embeddings and text-derived templates within a common semantic space.

Text templates as structured priors: We discretize age as $a \in \{A_{min}, \ldots, A_{max}\}$ (with a minimum difference of 0.1 year). For each age a and parcel p, we construct age-conditioned text of the form:

"`Parcel=<parcel_name>; Tissue_class=<tissue>; Age=<`a`>y`"

These texts are embedded using a frozen Qwen3 text encoder to obtain $\mathbf{e}^{\mathrm{raw}}_{a,p} \in \mathbb{R}^{d''}$. A lightweight Template Mapper, designed as a three-layer MLP with GeLU non-linearity (denoted $\mathrm{Map}(\cdot)$ and detailed in Appendix B.3), projects these prior embeddings into the shared SAL space:

$$\tilde{\mathbf{t}}_{a,p} = \mathrm{Map}(\mathbf{e}^{\mathrm{raw}}_{a,p}) \in \mathbb{R}^D, \qquad \mathbf{t}_{a,p} = \frac{\tilde{\mathbf{t}}_{a,p}}{\|\tilde{\mathbf{t}}_{a,p}\|_2} \in \mathbb{R}^D \tag{2}$$

The resulting set $\{\mathbf{t}_{a,p}\}_{a,p=1}^{A,N}$ forms an age-indexed healthy template bank in the SAL space. Here, $\mathbf{p}_{i,p}$ represents how parcel p *appears* in scan i, while $\mathbf{t}_{a,p}$ represents how parcel p is *expected to appear in a healthy brain* at age a. This assigns each parcel its own age-conditioned reference trajectory, rather than sharing a global prototype.

3.2.2 TRAINING OBJECTIVE

Training enforces two structural biases introduced earlier: (i) cross-parcel separation (parcel specificity) and (ii) smooth age ordering within each parcel (age continuity). We implement these using a pairwise SigLIP alignment objective complemented by a within-parcel age-shaping regularizer.

Age-aware SigLIP alignment: Let $\text{age}_i \in \mathbb{R}$ denote the chronological age of subject i. For each observed parcel embedding $\mathbf{p}_{i,p}$ and candidate parcel template $\mathbf{t}_{a,p'}$, we define a soft target $y_{i,p}^{a,p'} \in [0,1]$:

$$y_{i,p}^{a,p'} = \mathbb{I}[p' = p]\ \mathbb{I}(|\text{age}_i - a| \le w)\ \exp\left(-\frac{(\text{age}_i - a)^2}{2\sigma'^2}\right) \tag{3}$$

where w is an age window (in years) restricting positive candidates, and σ' controls how rapidly the positive weight decays with age mismatch. All cross-parcel pairs ($p' \neq p$) act as negatives by construction, enforcing parcel specificity.

Scaled SigLIP logits: With a learnable logit scale α and bias β, compatibility between an observed parcel $\mathbf{p}_{i,p}$ and template $\mathbf{t}_{a,p'}$ is computed using a scaled cosine logit: $s_{i,p}^{a,p'} = \alpha\, \mathbf{p}_{i,p}^{\top}\mathbf{t}_{a,p'} + \beta$. Since both $\mathbf{p}_{i,p}$ and $\mathbf{t}_{a,p'}$ are unit-normalized, $s_{i,p}^{a,p'}$ supports efficient retrieval over the template bank.

SigLIP loss: Using binary cross-entropy on logits, the age-aware SigLIP objective becomes:

$$\ell_{\text{bce}}(s, y) = -y \log \sigma(s) - (1 - y) \log(1 - \sigma(s)), \tag{4}$$

$$\mathcal{L}_{\text{SigLIP}} = \frac{1}{N_b} \sum_{(i,p)\in\mathcal{B}} \frac{1}{A_{\mathcal{B}} N} \sum_{a\in\mathcal{A}_{\mathcal{B}}} \sum_{p'=1}^{N} \ell_{\text{bce}}\left(s_{i,p}^{a,p'}, y_{i,p}^{a,p'}\right) \tag{5}$$

where $\mathcal{B}$ denotes the set of parcel instances in the batch, $\mathcal{A}_{\mathcal{B}}$ is the set of unique scan ages in the batch, $A_{\mathcal{B}} = |\mathcal{A}_{\mathcal{B}}|$, and $N_b = |\mathcal{B}|$. This enforces two behaviors simultaneously: (i) *parcel discrimination* and (ii) *age-sensitive alignment*, by treating all cross-parcels pairs ($p' \neq p$) and farther ages of within parcel ($p' = p$) as negatives.

Within-parcel age shaping via soft cross-entropy. While SigLIP provides local pairwise supervision, it does not explicitly enforce a smooth unimodal age profile within each parcel. To encourage such structure, we introduce a within-parcel soft age classification regularizer. Let $\mathcal{A} = \{A_{\min}, A_{\min} + 0.1, \ldots, A_{\max}\}$ denote the set of discretized ages, with $A = |\mathcal{A}|$. For each sample i and parcel p, we form the within-parcel age-logit vector $\mathbf{s}_{i,p}^{\text{age}} = \left[s_{i,p}^{a,p}\right]_{a\in\mathcal{A}} \in \mathbb{R}^A$. We define a soft target distribution over ages as

$$q_{i,p}(a) = \frac{\exp\left(-\frac{(\text{age}_i - a)^2}{2\sigma_p^2}\right)}{\sum_{a'\in\mathcal{A}} \exp\left(-\frac{(\text{age}_i - a')^2}{2\sigma_p^2}\right)}, \qquad a \in \mathcal{A}, \tag{6}$$

where $\sigma_p = \exp(\theta_p)$ is a learnable parcel-specific standard deviation. The predicted within-parcel age distribution is: $\hat{q}_{i,p}(a) = \frac{\exp\left(s_{i,p}^{a,p}\right)}{\sum_{a'\in\mathcal{A}} \exp\left(s_{i,p}^{a',p}\right)}$, where $a \in \mathcal{A}$. The resulting soft age cross-entropy loss over a batch $\mathcal{B}$ is:

$$\mathcal{L}_{\text{AgeSoftCE}} = -\frac{1}{N_b} \sum_{(i,p)\in\mathcal{B}} \sum_{a\in\mathcal{A}} q_{i,p}(a) \log \hat{q}_{i,p}(a). \tag{7}$$

This regularizer encourages the within-parcel age logits to form a smooth unimodal profile centered at the subject's age, while allowing the degree of age smoothness to vary across parcels through σ_p.

Total training objective: The final objective combines global alignment and within-parcel age shaping:

$$\mathcal{L} = \mathcal{L}_{\text{SigLIP}} + \lambda_{\text{age}}\, \mathcal{L}_{\text{AgeSoftCE}} \tag{8}$$

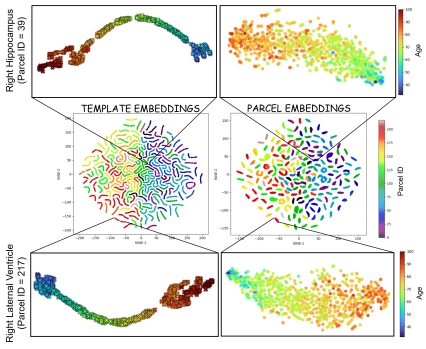


Figure 3: **t-SNE plots of BrainNorm's template ($\mathbf{t}_{a,p}$) and parcel ($\mathbf{p}_{i,p}$) embeddings** of healthy samples in the SAL space show clustering by parcel and age-ordered organization.

where λ_{age} is a hyperparameter (set to 0.1). BrainNorm's pretraining objective yields a structured SAL space. The parcel and template embeddings in the SAL space, as visualized in Fig. 3 with two sample parcels, are ordered by age and clustered by parcel identity. The same objective is used for fine-tuning on downstream cohorts by restricting exclusively to healthy subjects.

4 Experiments and Results

4.1 Datasets & Atlas

We leveraged healthy participants from the *UK Biobank (UKB)* (Sudlow et al. (2015)), a large-scale, publicly available imaging study comprising 48,018 T1-weighted structural MRI scans, for pretraining. Participants with documented neurological disorders were excluded to ensure a clean, healthy cohort. An external healthy cohort of 2,662 scans from the *Mayo Clinic Study of Aging (MCSA)* (Roberts et al. (2008)) (publicly available) was used as an additional held-out evaluation dataset. For downstream evaluations and tasks, we utilized publicly available mixed clinical cohorts with 15,352 scans spanning multiple neurodegenerative disorders, including *Alzheimer's disease* (AD), *Mild Cognitive Impairment* with stable subjects (sMCI) and progressors (pMCI), *Frontotemporal dementia* (FTD), and *Parkinson's disease* (PD). These datasets include the *ADNI* (phases 1,GO,2,3 & 4) (Mueller et al. (2005)), *AIBL* (Fowler et al. (2021)), *MIRIAD* (Malone et al. (2013)), *NIFD* (Rosen et al. (2010)), and *PPMI* (Marek et al. (2011)). BrainNorm was pretrained using 80% of the UKB cohort and evaluated on the remaining 20% held-out test set, along with the full held-out MCSA (Roberts et al. (2008)) dataset. For mixed disease cohorts, all downstream tasks were performed using diagnosis-stratified 5-fold subject-wise cross-validation, with performance reported on the 20% held-out test split within each fold. Each training set further included a 10% internal validation subset. All splits were performed subject-wise to prevent data leakage across training, validation, and test sets. Cohort details and statistics are provided in Appendix Sec A.1. All scans are preprocessed and registered to MNI-152 1mm isotropic space (details in Appendix Sec A.2).

Atlas Construction: To achieve whole-brain coverage, we selected complementary volumetric atlases and constructed a unified brain atlas combining: AAL3v1 (Tzourio-Mazoyer et al. (2002)) atlas for cortical and subcortical gray matter regions, the JHU white-matter atlas (Hua et al. (2008)) for white matter tracts, and the Harvard–Oxford atlas (FMRIB Software Library) for ventricular and cerebrospinal fluid regions (details in Appendix Sec A.3). This integration yielded 218 anatomical parcels covering the whole brain volume.

4.2 Model and Training Details

BrainNorm employs a Parcel Encoder with a transformer backbone that operates on $8 \times 8 \times 8$ voxel patches. The encoder consists of 8 layers with 8 attention heads, an embedding dimension of 768, and a feed-forward dimension of 2048. Parcel embeddings are projected into a 1024-dimensional SAL space. The Template Mapper projects frozen Qwen3 text embeddings of dimension 4096 into the same 1024-dimensional SAL space (details in Appendix Sec B.3). In total, BrainNorm contains approximately 67M trainable parameters. To construct the age-indexed template bank, we generated $\sim$153K structured texts covering 218 parcels across ages 30–100 years at 0.1-year resolution. Their embeddings were extracted using Qwen3-8B (Yang et al. (2025)) model and stored offline. With $\sim$38K training scans of UKB and the SigLIP objective, pretraining processes $\sim$8M parcel instances per epoch. The batch size was set to 32 scans, which corresponds to $\sim$7k parcel instances per batch. Optimization was performed using AdamW with learning rate 1×10^{-4}, weight decay 1×10^{-4}, and cosine scheduling with 3k warmup steps. Pretraining was performed over 100 hours of wall-clock time on 4 NVIDIA RTX A6000 GPUs for 40K optimization steps.

4.3 Downstream Evaluation Protocol

BrainNorm is pretrained on a large healthy UKB cohort. However, downstream datasets exhibit distributional shifts arising from scanner manufacturer, field strength, acquisition protocol, and demographic variation. To assess robustness under such shifts, we evaluate under two settings:

(i) Direct Transfer (No Finetuning): The pretrained UKB checkpoint is directly deployed to downstream datasets for inference and feature extraction.

Table 1: Parcel Identification Performance (Text(region)→Parcel & Parcel→Text(region)) using CN subjects across pretraining datasets and downstream mixed datasets ($mean_{\pm std}$ over 5-folds). Performance improves with finetuning. [FT: Finetuned on CN-only subjects of mixed cohorts]

Type	Metric	Mode	UKB	MCSA	ADNI	AIBL	MIRIAD	NIFD	PPMI
T→P	Acc@1	No FT	90.87	67.33	$69.50_{\pm 0.35}$	$68.71_{\pm 0.65}$	$73.93_{\pm 1.18}$	$70.62_{\pm 0.59}$	$67.18_{\pm 0.85}$
		FT	90.32	76.06	$79.88_{\pm 1.11}$	$80.01_{\pm 1.25}$	$83.49_{\pm 2.88}$	$82.06_{\pm 0.64}$	$79.47_{\pm 1.09}$
	Acc@5	No FT	99.02	85.19	$87.45_{\pm 0.24}$	$86.20_{\pm 0.48}$	$90.33_{\pm 0.69}$	$86.78_{\pm 0.50}$	$84.93_{\pm 0.58}$
		FT	99.04	94.54	$96.32_{\pm 0.39}$	$96.01_{\pm 0.50}$	$97.74_{\pm 0.83}$	$96.71_{\pm 0.29}$	$95.59_{\pm 0.56}$
P→T	Acc@1	No FT	97.17	75.02	$77.54_{\pm 0.31}$	$76.64_{\pm 0.83}$	$81.30_{\pm 1.01}$	$63.84_{\pm 3.11}$	$74.84_{\pm 0.60}$
		FT	96.72	86.46	$89.64_{\pm 0.78}$	$89.89_{\pm 1.18}$	$93.20_{\pm 1.76}$	$91.35_{\pm 0.37}$	$88.71_{\pm 0.54}$
	Acc@5	No FT	99.80	89.54	$91.17_{\pm 0.23}$	$90.27_{\pm 0.43}$	$93.32_{\pm 0.89}$	$80.40_{\pm 2.62}$	$88.72_{\pm 0.34}$
		FT	99.76	97.81	$98.65_{\pm 0.20}$	$98.64_{\pm 0.30}$	$99.35_{\pm 0.59}$	$98.91_{\pm 0.18}$	$98.13_{\pm 0.27}$

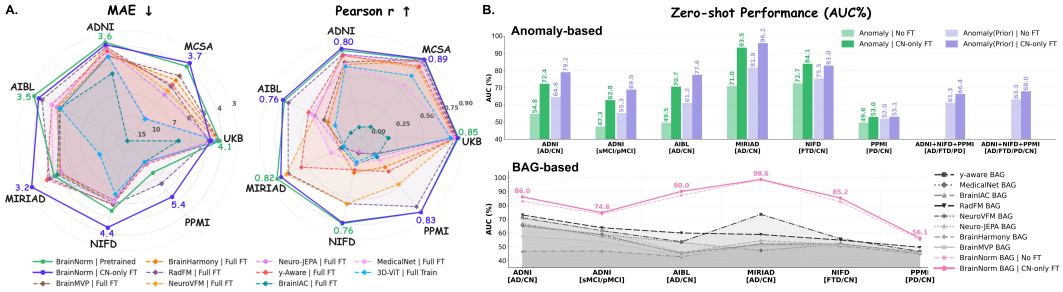


Figure 4: **A. BrainNorm's age estimation** outperforms other supervised baselines' age prediction performance. **B. BrainNorm's zero-shot performance** comparison using anomaly scores and BAG across all classification tasks. Plotted values are mean across 5-folds for mixed cohorts

(ii) CN-only Finetuning (FT): BrainNorm is finetuned using Eq 8 exclusively on cognitively normal (CN) subjects from the downstream training split of ADNI, AIBL, MIRIAD, NIFD, PPMI for each cross-validation split. Then both the Parcel encoder and the Template Mapper are frozen and further downstream tasks are then performed using these frozen embeddings of both CN and disease subjects. This procedure is repeated independently for each fold, preserving the normative objective while allowing limited cohort-specific calibration. (See Appendix Sec B.6 for finetuning details)

Baselines: We evaluated BrainNorm's performance against nine baselines using the same preprocessed scan inputs with end-to-end task-specific supervision for *(i)* age prediction and *(ii)* disease classification. We finetuned foundation models including NeuroJEPA (Huang et al. (2026)), NeuroVFM (Kondepudi et al. (2026)), RadFM (Wu et al. (2025)), BrainHarmony (Dong et al. (2025)), & BrainIAC (Tak et al. (2026)), and pretrained models including BrainMVP (Rui et al. (2025)), MedicalNet (Chen et al. (2019)) & y-Aware (Dufumier et al. (2021)). We additionally trained a ViT (Dosovitskiy et al. (2021)) model from scratch (ViT-B with forward dimension as 2048, having ∼67M params, i.e., nearly the same capacity as BrainNorm). Across all tasks and cohorts, BrainNorm consistently outperforms these baselines (Fig. 4,5). See Appendix Sec B.4 for more details.

4.4 Anatomical Correspondence and Normative Age Inference

a. Parcel Identification: To assess whether BrainNorm encodes *anatomical identity* in a shared semantic space, we evaluate bidirectional parcel-region retrieval. We ask two questions: (i) given a parcel region, can the model recover its anatomical label? and (ii) given an anatomical label, can the model identify the corresponding parcel in the scan? Let $\mathbf{p}_p \in \mathbb{R}^D$ denote the embedding of parcel p extracted from a scan, and let $\mathbf{t}_{a,q}$ denote the template embedding corresponding to anatomical region q and subject's chronological age a. Similarity is computed as $\mathbf{p}_p^\top \mathbf{t}_{a,q}$. For parcel → region retrieval, we rank all templates $\{\mathbf{t}_{a,q}\}_{q=1}^N$ given a parcel embedding $\mathbf{p}_p$. For region → parcel retrieval, we rank all parcel embeddings $\{\mathbf{p}_p\}_{p=1}^N$ given a template query $\mathbf{t}_{a,q}$. A prediction is considered correct for a ground-truth match within the top-k candidates. We report Acc@1 and Acc@5 for both in Table 1. Retrieval accuracy improves consistently after CN-only finetuning on downstream cohorts, indicating stronger anatomical grounding of the learned representations.

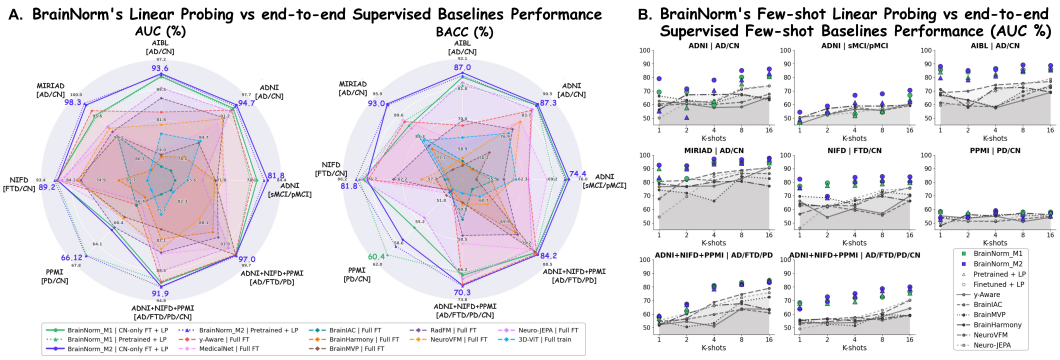


Figure 5: **A. Full data linear probing (LP).** BrainNorm's results of *M1* & *M2* techniques outperform end-to-end supervised baselines across the evaluated classification tasks. **B. Few-shot linear probing** of BrainNorm shows superior performance with both *M1* and *M2* representations compared to end-to-end finetuning of pretrained baselines, indicating stronger transferability under small-label budgets. Plotted values are mean across 5-folds for mixed cohorts.

b. Retrieval-based Age Estimation: Age is inferred via template retrieval across discrete age bins. Let ℓ_a denote the aggregated similarity between a scan and templates at age bin a, computed by averaging parcel-wise similarities. We construct a temperature-scaled softmax distribution and compute age as: $y_{\text{exp}} = \sum_{a \in \mathcal{A}} w_a a, \quad \text{where,} \quad w_a = \frac{\exp(\ell_a/\tau)}{\sum_{a'} \exp(\ell_{a'}/\tau)}$. We report results using $\tau = 0.5$, which sharpens the age-bin distribution and emphasizes high-similarity bins. Performance is evaluated using MAE and Pearson correlation. More metrics are in Appendix Table A10. BrainNorm achieves superior age estimation compared to the end-to-end supervised baseline predictions across datasets (Fig. 4 A). Also, BrainNorm demonstrates better generalizability on the held-out MCSA dataset.

4.5 Neurodegenerative Disease Classification

We evaluate disease sensitivity using zero-shot normative deviation scoring and linear probing (LP).

a. Zero-shot Normative Deviation Scoring: Zero-shot anomaly detection is derived directly from normative deviation measures without training a classifier.

(1) Global Anomaly Score: Let s_p denote similarity between scan parcel p and its age-matched template $\mathbf{t}_{a,p}$. We define parcel deviation score as: $d_p = (1 - s_p)$ and scan-level top-K anomaly score: $D_{global} = \frac{1}{K} \sum_{p \in \mathcal{T}_K} d_p$, where $\mathcal{T}_K$ contains K parcels with largest deviations (lowest similarity).

(2) Disease-Prior Restricted Anomaly: For disease d, top-K deviation aggregation is restricted to pre-defined anatomical parcel sets $\mathcal{P}_d$ derived from prior clinical knowledge of the disease. This produces disease-specific anomaly scores and enables prior-guided zero-shot disease classification.

(3) BAG-based Deviation: As a complementary estimator, we compute parcel-wise age argmax to get subject's age prediction: $y_{\text{mean-argmax}} = \frac{1}{|\mathcal{P}|} \sum_p a_p^*, \quad \text{where,} \quad a_p^* = \arg\max_a s_{i,p}^{a,p}$. Then, Brain Age Gap is defined as, BAG $= y_{\text{mean-argmax}} - y$, where y is chronological age. Visualization of BAG distributions across diagnostic groups is presented in Appendix Fig. A3. Subjects with larger positive BAG may reflect accelerated brain aging associated with neurodegenerative disease.

The zero-shot results before and after finetuning on healthy subjects are presented in Fig 4B. Notably, the disease-prior-restricted anomaly enables *prior-guided zero-shot multi-disease classification* using BrainNorm. (More details in Appendix Sec B.10)

b. Linear Probing (few-shot LP and all LP): To evaluate the disease separability of BrainNorm's representations under supervision, we perform linear probing on frozen embeddings. In particular, we test whether disease-relevant signals can be extracted along: (i) representations from parcel embeddings, and (ii) parcel-wise deviations from age-matched normative templates.

(M1) Parcel Embedding. To test whether anatomical parcel representations alone are discriminative, we concatenate parcel embeddings: $\mathbf{x}_{M1} = concat(\mathbf{p}_1, \ldots, \mathbf{p}_N) \in \mathbb{R}^{ND}$.

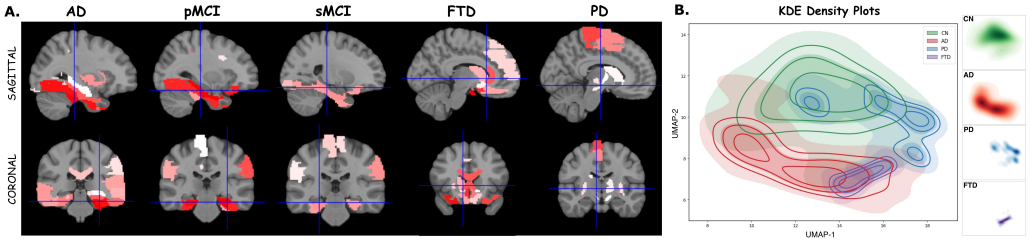


Figure 6: **A. Zero-shot Anomaly Localization** of the top-k anomalous brain parcels across disease cohorts in sagittal & coronal planes. Higher deviation scores from normative reference are represented by darker shades. **B. KDE density plot** of UMAP using parcel deviations ($\mathbf{x}_{M2}$).

(M2) Parcel Deviation. To capture normative deviation patterns, we compute parcel-wise deviations between scan embeddings and age-matched templates, forming: $\mathbf{x}_{M2} = concat(\mathbf{p}_1 - \mathbf{t}_{a,1}, \ldots, \mathbf{p}_N - \mathbf{t}_{a,N}) \in \mathbb{R}^{ND}$. KDE plots of UMAP using $\mathbf{x}_{M2}$ show the clusters across the groups in Fig 6B.

Each feature representation is fed to a linear classifier and evaluated under two setups: **full LP** using all available labeled training scans, and **few-shot LP** using k labeled training scans per class ($k \in \{1, 2, 4, 8, 16\}$). Results for full-data linear probing are reported in Fig. 5A, and few-shot linear probing are reported in Fig. 5B, with mean values across 5-folds. Linear probing with full data on BrainNorm representations consistently outperforms the end-to-end supervised baselines across all evaluated datasets and tasks, demonstrating strong disease-relevant separability in BrainNorm's frozen representations after pretraining and CN-only finetuning.

Further experimental results and details with $mean_{\pm std}$ values are provided in Appendix Sec B. Refer to Appendix Sec B.5.1 for Template Mapper & loss ablations, Sec B.5.2 for text-prior initialization ablations, and Sec B.5.3 for prompt-rephrasing demonstrating BrainNorm's flexibility in querying the SAL space. More details of BrainNorm's embeddings and confounder effects are discussed in Appendix Sec B.13.

5 Interpretability

We visualized top-K anomalous brain regions in Fig. 6 using zero-shot parcel-wise anomaly scores (d_p) relative to age-conditioned healthy templates. For each disease cohort, individual deviation scores are averaged across the test set to generate cohort-level anomaly profile. These top-K parcels are then mapped back into volumetric space to visualize spatial distributions of disease-related deviations. Resulting maps align with findings in clinical neurodegeneration literature, including involvement of the hippocampus and parahippocampal gyrus in AD and MCI(Planche et al. (2022); Braak et al. (2006)); the frontal/prefrontal cortex and temporal lobes in FTD(Rascovsky et al. (2011); Gorno-Tempini et al. (2011)); and the motor cortex and subcortical structures in PD(Hughes et al. (1992); Postuma et al. (2015)). Futher scan-level interpretations are discussed in Appendix Sec B.14.

6 Discussion and Conclusion

We presented **BrainNorm**, a normative foundation model that learns anatomically structured, age-conditioned representations of healthy brain variation within a *Semantic Atlas Latent* (SAL) space. Through Semantic Atlas Pretraining, BrainNorm constructs parcel-specific healthy references against which individual representations can be compared, enabling age estimation, brain-age-gap analysis, anatomical region identification, zero-shot deviation analysis, and disease classification (e.g., AD, PD, FTD, MCI) through linear probing of frozen representations. Across multiple cohorts and tasks, BrainNorm demonstrates strong transferability, with linear probes outperforming end-to-end supervised baselines. BrainNorm currently relies on MNI152 registration to ensure spatial consistency for atlas-guided parcel tokenization. It is developed using T1w sMRI and uses adult brain atlases. Extension to pediatric and neurodevelopmental populations may therefore require age-appropriate anatomical priors. Future work will explore adaptive parcellation to reduce dependence on registration and atlases, and extend the framework to additional modalities such as T2w, FLAIR, and fMRI.

AI USE STATEMENT

In this work, we used generative AI tools solely to assist with grammatical correction, language refinement, and improving the readability of the manuscript. We did not use generative AI tools to develop the research methodology, design experiments, implement methods, process or analyze data, interpret results, formulate scientific claims, generate figures, or conduct literature review. All AI-assisted edits were reviewed and verified by the authors. We take full responsibility for the final content of this work, including all text, claims, results, and conclusions.

REPRODUCIBILITY STATEMENT

To support reproducibility, comprehensive descriptions of model architectures, training configurations, and evaluation protocols are provided in the main text and appendix. The preprocessing pipeline is discussed in Appendix Sec A.2. The model architecture details are provided in Sec 4.2 and Appendix Sec B.2,B.3. Training and finetuning details are provided in Sec 4.2 and Appendix Sec B.6,B.7.

REFERENCES

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 *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15619–15629. IEEE, 2023.

Abhinand Balachandran. Medembed: Medical-focused embedding models, 2024. URL `https://github.com/abhinand5/MedEmbed`.

Shruthi Bannur, Stephanie Hyland, Qianchu Liu, Fernando Perez-Garcia, Maximilian Ilse, et al. Learning to exploit temporal structure for biomedical vision-language processing. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15016–15027. IEEE, 2023.

Carlo Alberto Barbano, Matteo Brunello, Benoit Dufumier, Marco Grangetto, Alzheimer's Disease Neuroimaging Initiative, et al. Anatomical foundation models for brain mris. *Pattern Recognition Letters*, 199:178–184, 2025.

Adrien Bardes, Quentin Garrido, Jean Ponce, Xinlei Chen, Michael Rabbat, Yann LeCun, Mahmoud Assran, and Nicolas Ballas. Revisiting feature prediction for learning visual representations from video. *arXiv preprint arXiv:2404.08471*, 2024.

Richard AI Bethlehem, Jakob Seidlitz, Simon R White, Jacob W Vogel, et al. Brain charts for the human lifespan. *Nature*, 604(7906):525–533, 2022.

Heiko Braak, Irina Alafuzoff, Thomas Arzberger, Hans Kretzschmar, and Kelly Del Tredici. Staging of alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry. *Acta neuropathologica*, 112(4):389–404, 2006.

Sihong Chen, Kai Ma, and Yefeng Zheng. Med3d: Transfer learning for 3d medical image analysis. *arXiv preprint arXiv:1904.00625*, 2019.

James H Cole, Rudra PK Poudel, Dimosthenis Tsagkrasoulis, Matthan WA Caan, Claire Steves, Tim D Spector, and Giovanni Montana. Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker. *NeuroImage*, 163:115–124, 2017.

James H Cole, Stuart J Ritchie, Mark E Bastin, MC Valdés Hernández, et al. Brain age predicts mortality. *Molecular psychiatry*, 23(5):1385–1392, 2018.

Tim J Cole. The development of growth references and growth charts. *Annals of human biology*, 39 (5):382–394, 2012.

Adrian Dalca, Marianne Rakic, John Guttag, and Mert Sabuncu. Learning conditional deformable templates with convolutional networks. *Advances in neural information processing systems*, 32, 2019.

Zijian Dong, Ruilin Li, Joanna Chong, Niousha Dehestani, Yinghui Teng, Yi Lin, Zhizhou Li, Yichi Zhang, Yapei Xie, Leon Ooi, et al. Brain harmony: a multimodal foundation model unifying morphology and function into 1d tokens. *Advances in Neural Information Processing Systems*, 38:122100–122127, 2025.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *ICLR*, 2021.

Benoit Dufumier, Pietro Gori, Julie Victor, Antoine Grigis, Michele Wessa, et al. Contrastive learning with continuous proxy meta-data for 3d mri classification. In *International conference on medical image computing and computer-assisted intervention*, pp. 58–68. Springer, 2021.

FMRIB Software Library. Harvard-oxford cortical structural atlas. `https://fsl.fmrib.ox.ac.uk/fsl/docs/other/datasets.html`. RRID:SCR_001476.

Christopher Fowler, Stephanie R Rainey-Smith, Sabine Bird, Julia Bomke, et al. Fifteen years of the australian imaging, biomarkers and lifestyle (aibl) study: progress and observations from 2,359 older adults spanning the spectrum from cognitive normality to alzheimer's disease. *Journal of Alzheimer's disease reports*, 5(1):443–468, 2021.

Katja Franke, Gabriel Ziegler, Stefan Klöppel, Christian Gaser, Alzheimer's Disease Neuroimaging Initiative, et al. Estimating the age of healthy subjects from t1-weighted mri scans using kernel methods: exploring the influence of various parameters. *Neuroimage*, 50(3):883–892, 2010.

Charlotte J Fraza, Richard Dinga, Christian F Beckmann, and Andre F Marquand. Warped bayesian linear regression for normative modelling of big data. *NeuroImage*, 245:118715, 2021.

Alexander Frotscher, Jaivardhan Kapoor, Thomas Wolfers, and Christian F Baumgartner. Unsupervised anomaly detection in medical imaging using aggregated normative diffusion. *Medical Image Analysis*, 2025.

Jingru Fu, Daniel Ferreira, Örjan Smedby, and Rodrigo Moreno. Decomposing the effect of normal aging and alzheimer's disease in brain morphological changes via learned aging templates. *Scientific Reports*, 15(1), 2025.

Christian Gaser, Katja Franke, Stefan Klöppel, Nikolaos Koutsouleris, Heinrich Sauer, and Alzheimer's Disease Neuroimaging Initiative. Brainage in mild cognitive impaired patients: predicting the conversion to alzheimer's disease. *PloS one*, 8(6), 2013.

Ruiyang Ge, Yuetong Yu, Yi Xuan Qi, Yu-nan Fan, Shiyu Chen, et al. Normative modelling of brain morphometry across the lifespan with centilebrain: algorithm benchmarking and model optimisation. *The Lancet Digital Health*, 6(3), 2024.

Maria Luisa Gorno-Tempini, Argye E Hillis, Sandra Weintraub, Andrew Kertesz, et al. Classification of primary progressive aphasia and its variants. *Neurology*, 76(11):1006–1014, 2011.

Andrew Hoopes, Jocelyn S Mora, Adrian V Dalca, Bruce Fischl, and Malte Hoffmann. Synthstrip: skull-stripping for any brain image. *NeuroImage*, 260, 2022.

Kegang Hua, Jiangyang Zhang, Setsu Wakana, Hangyi Jiang, Xin Li, Daniel S Reich, Peter A Calabresi, James J Pekar, Peter CM van Zijl, and Susumu Mori. Tract probability maps in stereotaxic spaces: analyses of white matter anatomy and tract-specific quantification. *Neuroimage*, 39(1): 336–347, 2008.

Haoxu Huang, Long Chen, Jingyun Chen, Jinu Hyun, James Ryan Loftus, Kara Melmed, Daniel Orringer, Jennifer Frontera, Seena Dehkharghani, Arjun Masurkar, et al. Learning sparse latent predictive foundation model for multimodal neuroimaging. *arXiv preprint arXiv:2606.14957*, 2026.

Andrew J Hughes, Susan E Daniel, Linda Kilford, and Andrew J Lees. Accuracy of clinical diagnosis of idiopathic parkinson's disease: a clinico-pathological study of 100 cases. *Journal of neurology, neurosurgery & psychiatry*, 55(3):181–184, 1992.

Seyed Mostafa Kia, Hester Huijsdens, Saige Rutherford, Augustijn de Boer, et al. Closing the life-cycle of normative modeling using federated hierarchical bayesian regression. *Plos one*, 17(12), 2022.

Akhil Kondepudi, Akshay Rao, Chenhui Zhao, Yiwei Lyu, Samir Harake, Soumyanil Banerjee, Jacob Ogle, Rushikesh Joshi, Anna-Katharina Meissner, Xinhai Hou, et al. Health system learning enables generalist neuroimaging models. *Nature Medicine*, pp. 1–7, 2026.

Sayantan Kumar, Philip RO Payne, and Aristeidis Sotiras. Normative modeling using multimodal variational autoencoders to identify abnormal brain volume deviations in alzheimer's disease. In *Proceedings of SPIE–the International Society for Optical Engineering*, volume 12465, pp. 1246503, 2023.

Christoph H Lampert, Hannes Nickisch, and Stefan Harmeling. Attribute-based classification for zero-shot visual object categorization. *IEEE transactions on pattern analysis and machine intelligence*, 36(3):453–465, 2013.

Ana Lawry Aguila, James Chapman, Mohammed Janahi, and Andre Altmann. Conditional vaes for confound removal and normative modelling of neurodegenerative diseases. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 430–440. Springer, 2022.

Jeyeon Lee, Brian J Burkett, Hoon-Ki Min, Matthew L Senjem, Emily S Lundt, et al. Deep learning-based brain age prediction in normal aging and dementia. *Nature aging*, 2(5):412–424, 2022.

Jin San Lee, Yu Hyun Park, Seongbeom Park, Uicheul Yoon, et al. Distinct brain regions in physiological and pathological brain aging. *Frontiers in Aging Neuroscience*, 11:147, 2019.

Marco Lorenzi, Xavier Pennec, Giovanni B Frisoni, Nicholas Ayache, Alzheimer's Disease Neuroimaging Initiative, et al. Disentangling normal aging from alzheimer's disease in structural magnetic resonance images. *Neurobiology of aging*, 36:S42–S52, 2015.

Ian B Malone, David Cash, Gerard R Ridgway, David G MacManus, Sebastien Ourselin, Nick C Fox, and Jonathan M Schott. Miriad—public release of a multiple time point alzheimer's mr imaging dataset. *NeuroImage*, 70:33–36, 2013.

Kenneth Marek, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, et al. The parkinson progression marker initiative (ppmi). *Progress in neurobiology*, 95(4):629–635, 2011.

Małgorzata Marjańska, J Riley McCarten, James Hodges, Laura S Hemmy, Andrea Grant, Dinesh K Deelchand, and Melissa Terpstra. Region-specific aging of the human brain as evidenced by neurochemical profiles measured noninvasively in the posterior cingulate cortex and the occipital lobe using 1h magnetic resonance spectroscopy at 7 t. *Neuroscience*, 354:168–177, 2017.

André F. Marquand, Iead Rezek, Jan Buitelaar, and Christian F. Beckmann. Understanding heterogeneity in clinical cohorts using normative models: Beyond case-control studies. *Biological Psychiatry*, 80(7):552–561, 2016. doi: 10.1016/j.biopsych.2015.12.023.

John C Mazziotta, Arthur W Toga, Alan Evans, Peter Fox, and Jack Lancaster. A probabilistic atlas of the human brain: Theory and rationale for its development: The international consortium for brain mapping (icbm). *Neuroimage*, 2(2):89–101, 1995.

Susanne G Mueller, Michael W Weiner, Leon J Thal, Ronald C Petersen, Clifford Jack, et al. The alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics of North America*, 15 (4), 2005.

Jiahong Ouyang, Qingyu Zhao, Ehsan Adeli, Greg Zaharchuk, and Kilian M Pohl. Disentangling normal aging from severity of disease via weak supervision on longitudinal mri. *IEEE Transactions on Medical Imaging*, 41(10):2558–2569, 2022a.

Jiahong Ouyang, Qingyu Zhao, Ehsan Adeli, Greg Zaharchuk, and Kilian M Pohl. Self-supervised learning of neighborhood embedding for longitudinal mri. *Medical image analysis*, 82, 2022b.

Han Peng, Weikang Gong, Christian F Beckmann, Andrea Vedaldi, and Stephen M Smith. Accurate brain age prediction with lightweight deep neural networks. *Medical image analysis*, 68:101871, 2021.

Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Walter HL Pinaya, Andrea Mechelli, and João R Sato. Using deep autoencoders to identify abnormal brain structural patterns in neuropsychiatric disorders: A large-scale multi-sample study. *Human brain mapping*, 40(3):944–954, 2019.

Walter HL Pinaya, Mark S Graham, Robert Gray, Pedro F Da Costa, et al. Fast unsupervised brain anomaly detection and segmentation with diffusion models. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 705–714. Springer, 2022a.

Walter HL Pinaya, Petru-Daniel Tudosiu, Robert Gray, Geraint Rees, Parashkev Nachev, Sébastien Ourselin, and M Jorge Cardoso. Unsupervised brain imaging 3d anomaly detection and segmentation with transformers. *Medical Image Analysis*, 79:102475, 2022b.

Vincent Planche, José V Manjon, Boris Mansencal, Enrique Lanuza, Thomas Tourdias, Gwenaëlle Catheline, and Pierrick Coupé. Structural progression of alzheimer's disease over decades: the mri staging scheme. *Brain communications*, 4(3), 2022.

Ronald B Postuma, Daniela Berg, Matthew Stern, Werner Poewe, C Warren Olanow, Wolfgang Oertel, José Obeso, Kenneth Marek, Irene Litvan, Anthony E Lang, et al. Mds clinical diagnostic criteria for parkinson's disease. *Movement disorders*, 30(12):1591–1601, 2015.

Lemuel Puglisi, Daniel C Alexander, and Daniele Ravì. Enhancing spatiotemporal disease progression models via latent diffusion and prior knowledge. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 173–183. Springer, 2024.

Katya Rascovsky, John R Hodges, David Knopman, Mario F Mendez, et al. Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia. *Brain*, 134(9):2456–2477, 2011.

Susan M Resnick, Dzung L Pham, Michael A Kraut, Alan B Zonderman, and Christos Davatzikos. Longitudinal magnetic resonance imaging studies of older adults: a shrinking brain. *The Journal of Neuroscience*, 23(8):3295–3301, 2003.

Rosebud O Roberts, Yonas E Geda, David S Knopman, Ruth H Cha, V Shane Pankratz, et al. The mayo clinic study of aging: design and sampling, participation, baseline measures and sample characteristics. *Neuroepidemiology*, 30(1):58–69, 2008.

Howard J Rosen, AL Boxer, ML Gorno-Tempini, and MW Weiner. Frontotemporal lobar degeneration neuroimaging initiative (ftldni). *Funded by the National Institute on Aging*, 2010.

Shaohao Rui, Lingzhi Chen, Zhenyu Tang, Lilong Wang, Mianxin Liu, Shaoting Zhang, and Xiaosong Wang. Multi-modal vision pre-training for medical image analysis. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5164–5174. IEEE, 2025.

Saige Rutherford, Seyed Mostafa Kia, Thomas Wolfers, Charlotte Fraza, Mariam Zabihi, Richard Dinga, Pierre Berthet, Amanda Worker, Serena Verdi, Henricus G Ruhe, et al. The normative modeling framework for computational psychiatry. *Nature protocols*, 17(7):1711–1734, 2022.

Jenna M Schabdach, J Eric Schmitt, Susan Sotardi, Arastoo Vossough, et al. Brain growth charts for quantitative analysis of pediatric clinical brain mri scans with limited imaging pathology. *Radiology*, 309(1), 2023.

Marc-Andre Schulz, Nys Tjade Siegel, and Kerstin Ritter. Brain-age models with lower age prediction accuracy have higher sensitivity for disease detection. *PLoS biology*, 23(10), 2025.

Andrew Sellergren, Sahar Kazemzadeh, Tiam Jaroensri, Atilla Kiraly, et al. Medgemma technical report. *arXiv preprint arXiv:2507.05201*, 2025.

Raphaël Sivera, Hervé Delingette, Marco Lorenzi, Xavier Pennec, Nicholas Ayache, Alzheimer's Disease Neuroimaging Initiative, et al. A model of brain morphological changes related to aging and alzheimer's disease from cross-sectional assessments. *NeuroImage*, 198:255–270, 2019.

Jie Song, Chengchao Shen, Jie Lei, An-Xiang Zeng, Kairi Ou, Dacheng Tao, and Mingli Song. Selective zero-shot classification with augmented attributes. In *European Conference on Computer Vision*, pp. 474–490. Springer, 2018.

Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, et al. Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS medicine*, 12(3), 2015.

Divyanshu Tak, Biniam A Garomsa, Anna Zapaishchykova, Tafadzwa L Chaunzwa, et al. A generalizable foundation model for analysis of human brain mri. *Nature Neuroscience*, 29(4):945–956, 2026.

Nathalie Tzourio-Mazoyer, Brigitte Landeau, Dimitri Papathanassiou, Fabrice Crivello, Octave 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.

Madhumitha Venkatesh, Shanawaj S Madarkar, and Konda Reddy Mopuri. Parcel2progression: An anatomy-aware longitudinal framework for alzheimer's disease diagnosis. *arXiv preprint arXiv:2608.08753*, 2026.

Alan Q Wang, Fangrui Huang, Bailey Trang, Wei Peng, Mohammad Abbasi, Kilian Pohl, Mert R Sabuncu, and Ehsan Adeli. Generating novel brain morphology by deforming learned templates. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 207–217. Springer, 2025.

Zifeng Wang, Zhenbang Wu, Dinesh Agarwal, and Jimeng Sun. Medclip: Contrastive learning from unpaired medical images and text. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pp. 3876–3887, 2022.

Thomas Wolfers, Nhat Trung Doan, Tobias Kaufmann, Dag Alnæs, Torgeir Moberget, et al. Mapping the heterogeneous phenotype of schizophrenia and bipolar disorder using normative models. *JAMA psychiatry*, 75(11):1146–1155, 2018.

Thomas Wolfers, Christian F Beckmann, Martine Hoogman, Jan K Buitelaar, Barbara Franke, and Andre F Marquand. Individual differences v. the average patient: mapping the heterogeneity in adhd using normative models. *Psychological medicine*, 50(2):314–323, 2020.

Chaoyi Wu, Xiaoman Zhang, Ya Zhang, Hui Hui, Yanfeng Wang, and Weidi Xie. Towards generalist foundation model for radiology by leveraging web-scale 2d&3d medical data. *Nature Communications*, 16(1):7866, 2025.

An Yang, Anfeng Li, Baosong Yang, Beichen Zhang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Gao, Chengen Huang, Chenxu Lv, et al. Qwen3 technical report. *arXiv preprint arXiv:2505.09388*, 2025.

Alexandra L Young, Razvan V Marinescu, Neil P Oxtoby, Martina Bocchetta, Keir Yong, Nicholas C Firth, David M Cash, David L Thomas, Katrina M Dick, Jorge Cardoso, et al. Uncovering the heterogeneity and temporal complexity of neurodegenerative diseases with subtype and stage inference. *Nature communications*, 9(1):4273, 2018.

Xiaohua Zhai, Basil Mustafa, Alexander Kolesnikov, and Lucas Beyer. Sigmoid loss for language image pre-training. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 11975–11986. IEEE, 2023.

Ruixia Zhang, Fan Yi, Hongjing Mao, Zhengxing Huang, Kai Wang, and Junhang Zhang. Brain age gap as a predictive biomarker that links aging, lifestyle, and neuropsychiatric health. *Communications medicine*, 5(1):441, 2025a.

Sheng Zhang, Yanbo Xu, Naoto Usuyama, Hanwen Xu, Jaspreet Bagga, Robert Tinn, Sam Preston, Rajesh Rao, Mu Wei, Naveen Valluri, et al. A multimodal biomedical foundation model trained from fifteen million image–text pairs. *Nejm Ai*, 2(1), 2025b.

Qingyu Zhao, Zixuan Liu, Ehsan Adeli, and Kilian M Pohl. Longitudinal self-supervised learning. *Medical image analysis*, 71, 2021.

David Zimmerer, Fabian Isensee, Jens Petersen, Simon Kohl, and Klaus Maier-Hein. Unsupervised anomaly localization using variational auto-encoders. In *International conference on medical image computing and computer-assisted intervention*, pp. 289–297. Springer, 2019.

Appendix

CONTENTS

A Datasets and Preprocessing Details **16**

A.1 Dataset Overview 16

A.2 Preprocessing Pipeline 18

A.3 Brain Atlas Construction 18

B Experiment details and More Results **18**

B.1 Code and checkpoint 18

B.2 Parcel Encoder details 19

B.3 Template Mapper details 19

B.4 Baselines 19

B.5 Ablations 20

B.5.1 Template Mapper & Loss Ablations 20

B.5.2 Text-Prior Initialization Ablations 20

B.5.3 Text Query Ablations 21

B.5.4 Choice of Text Encoder 22

B.6 Fine-tuning Details 23

B.7 Implementation details of SigLIP loss 23

B.8 Details of supervised age prediction tasks for prior works 23

B.9 Parcel Identification 24

B.10 Zero-shot Prediction Details 24

B.11 Traditional Normative Model baseline comparison 26

B.12 Formulation of Linear Probing 27

B.13 BrainNorm's embeddings visualizations 28

B.14 Scan-level Interpretability 28

C Limitations, Scope and Future Directions **31**

A DATASETS AND PREPROCESSING DETAILS

A.1 DATASET OVERVIEW

Downstream Clinical Cohorts. For downstream evaluations and tasks, we utilized publicly available mixed clinical cohorts totalling 15,352 scans, spanning multiple neurodegenerative disorders: *Alzheimer's disease* (AD), *Mild Cognitive Impairment* with stable subjects (sMCI) and progressors (pMCI), *Frontotemporal dementia* (FTD), and *Parkinson's disease* (PD). The constituent datasets are described below.

- **ADNI:** The *Alzheimer's Disease Neuroimaging Initiative* (phases 1, GO, 2, 3, and 4) (Mueller et al. (2005)) is a large multi-center study spanning dozens of North American sites (U.S./Canada), with contributions across Siemens, GE, and Philips scanners at

Table A1: Cohort statistics of all datasets used in our experiments.

Dataset	#scans	#subjects	Age (y) (min,max,mean$_{\pm \text{std}}$)	Gender (M / F)	Field strength (1.5T / 3.0T)	Manufacturer (Siemens/GE/Philips)
UKB (CN)	48,018	43,631	$(44.0, 83.0, 64.23_{\pm 7.70})$	20,592 / 23,039	0 / 43,631	43,631 / 0 / 0
MCSA (CN)	2,662	1,611	$(49.0, 89.9, 71.83_{\pm 9.83})$	853 / 758	0 / 1,611	0 / 1,611 / 0
ADNI	5,627	1,703	$(52.4, 93.1, 72.93_{\pm 7.87})$	734 / 969	429 / 1,411	1,063 / 476 / 289
ADNI (MCI)	5,308	687	$(55.6, 91.6, 74.18_{\pm 7.44})$	408 / 279	275 / 512	389 / 258 / 134
AIBL	1,084	599	$(56.3, 93.0, 73.36_{\pm 6.42})$	255 / 343	107 / 541	599 / 0 / 0
MIRIAD	708	69	$(56.0, 86.3, 69.89_{\pm 6.98})$	31 / 38	69 / 0	0 / 69 / 0
NIFD	962	332	$(30.6, 87.1, 64.18_{\pm 8.16})$	176 / 156	0 / 332	317 / 16 / 0
PPMI	1,663	947	$(31.0, 85.0, 62.97_{\pm 10.01})$	576 / 371	214 / 737	548 / 218 / 171

1.5T and 3T, and protocol families such as MPRAGE. It offers broad diagnostic coverage (CN/MCI/AD) with multi-year follow-ups, yielding substantial inter-site and inter-scanner variability in acquisition parameters and subject demographics. The ADNI cohort comprises 493 AD and 1210 CN subjects (corresponding to 1485 AD scans and 4142 CN scans). In the ADNI MCI cohort, both stable MCI (sMCI) and progressive MCI (pMCI) subjects are initially diagnosed with MCI at baseline, but the cognitive condition of the sMCI group remained stable and did not convert to AD within 36 months after their first visit. In contrast, the pMCI group progressed to a clinical diagnosis of AD during the same 36-month follow-up period (Gaser et al. (2013)). All scans from baseline up to, but *not including* the first AD diagnosis, are labeled `pMCI`. The MCI cohort comprises 396 subjects (3915 scans) categorized as sMCI, and 291 subjects (1393 scans) as pMCI.

- **AIBL:** The *Australian Imaging, Biomarkers and Lifestyle* study (Fowler et al. (2021)) is a prospective Australian cohort with imaging conducted primarily at multiple sites in Melbourne and Perth, predominantly on Siemens platforms (1.5T/ 3T). It provides serial T1-weighted scans with rich clinical and biomarker metadata. AIBL cohort includes 104 AD and 500 CN individuals (154 AD scans and 930 CN scans).

- **MIRIAD:** The *Minimal Interval Resonance Imaging in Alzheimer's Disease* dataset (Malone et al. (2013)) is a single-center cohort from London (UCL), acquired on a fixed scanner and protocol with dense revisit intervals (weeks to months). The dataset includes 46 AD and 23 CN participants (corresponding to 465 AD scans and 243 CN scans).

- **NIFD:** The *Neuroimaging in Frontotemporal Dementia* dataset (also known as FTLDNI) (Rosen et al. (2010)) is a multi-site NIH-funded study designed to characterise structural and functional neuroimaging signatures across the FTD spectrum. Data were acquired across three academic medical centres: the University of California San Francisco (UCSF), Mayo Clinic (Rochester), and Massachusetts General Hospital (MGH). T1-weighted MRI scans were acquired at 3T across sites, with each site contributing scanner and protocol-specific variability. The cohort includes 214 FTD patients and 118 cognitively normal controls (605 FTD scans and 357 CN scans).

- **PPMI:** The *Parkinson's Progression Markers Initiative* (Marek et al. (2011)) is a large-scale, multi-center, international longitudinal study launched by the *Michael J. Fox Foundation for Parkinson's Research*, with sites spanning North America and Europe. The study was designed to identify progression biomarkers for Parkinson's disease across clinical, imaging, and biological modalities. Neuroimaging was conducted at both 1.5T and 3T field strengths across multiple scanner platforms (Siemens, GE, Philips), introducing meaningful inter-site and inter-scanner variability. The cohort includes newly diagnosed, untreated PD patients at enrollment alongside healthy controls, with longitudinal follow-up visits. The subset used in this study consists of 734 PD patients and 213 controls (1360 PD scans and 303 CN scans). The broad geographic and institutional diversity of PPMI provides a rigorous testbed for evaluating the model's generalizability to PD-related neurodegeneration.

Detailed per-cohort subject and scan counts, stratified by diagnosis, scanner, and field strength, are provided in Table A1.

A.2 Preprocessing Pipeline

The UKB data considered was already partially preprocessed and registered to MNI space. Thus, only z-score normalization within the brain mask was additionally performed. All other T1-weighted structural MRI scans of downstream cohorts underwent an identical preprocessing pipeline, applied uniformly across all cohorts. Standardizing preprocessing in this way is essential to place all volumes within a common anatomical reference frame so that the parcel-level atlas maps consistently to the same structures across subjects, sites, and scanners.

The pipeline begins with the reorientation of each scan to a canonical via `fslreorient2std` (FSL), bias field correction using `N4BiasFieldCorrection`. The resulting intensity-homogenized volume is then skull-stripped using SynthStrip (Hoopes et al. (2022)). Further, each scan is registered to the MNI152 (Mazziotta et al. (1995)) 1 mm isotropic standard template through affine+nonlinear registration using FSL (`FLIRT`→`FNIRT`). Voxel intensities are then z-score normalized within the brain mask. Finally, the normalized volumes are padded symmetrically from their native MNI grid of $182 \times 218 \times 182$ voxels to $184 \times 224 \times 184$ voxels, to be divisible by the 3D patch size (8^3) used in the model architecture.

A.3 Brain Atlas Construction

We constructed a unified whole-brain atlas in MNI152 space by integrating three complementary, publicly available volumetric atlases, each providing optimal coverage for a distinct tissue compartment to enable anatomically grounded, parcel-level modeling where each input token or feature corresponds to a well-defined neuroanatomical structure. The Atlases utilized are:-

- **AAL3v1** (Tzourio-Mazoyer et al. (2002)) was used to define **cortical and subcortical gray matter** regions. AAL3v1 extends the original AAL atlas with improved subcortical delineations, providing comprehensive bilateral parcellation of cortical gyri and subcortical nuclei (including thalamus, caudate, putamen, pallidum, hippocampus, amygdala, and accumbens), all of which are structures of direct relevance in neurodegenerative disease progression.
- **JHU White-Matter Atlas** (Hua et al. (2008)) was used to define **white matter tract** regions. The JHU atlas provides labeled regions for major white matter fasciculi including the corticospinal tracts, corpus callosum, superior and inferior longitudinal fasciculi, and uncinate fasciculus, among others. Including white matter as explicit parcels allows the model to capture tract-level structural changes, which are prominent in several disorders represented in cohorts (e.g., FTD, PD).
- **Harvard–Oxford Atlas** (FMRIB Software Library) was used to define ventricular regions with cerebrospinal fluid (CSF), including the lateral ventricles, third and fourth ventricles, and surrounding CSF spaces. Ventricular enlargement is a reliable, high-sensitivity biomarker of global atrophy in neurodegenerative conditions, and its explicit inclusion as a modeled compartment ensures this signal is captured at the parcel level rather than treated as background.

Atlas integration. The final integrated atlas yields 218 anatomical parcels providing whole-brain volumetric coverage of the brain, spanning gray matter cortex, subcortical regions, white matter tracts, and ventricular/CSF spaces. The list of parcels considered from each atlas, along with their label IDs, is reported in Table A15.

B Experiment details and More Results

B.1 Code and checkpoint

The complete model code and training scripts will be uploaded to GitHub. The checkpoints will be released post-acceptance.

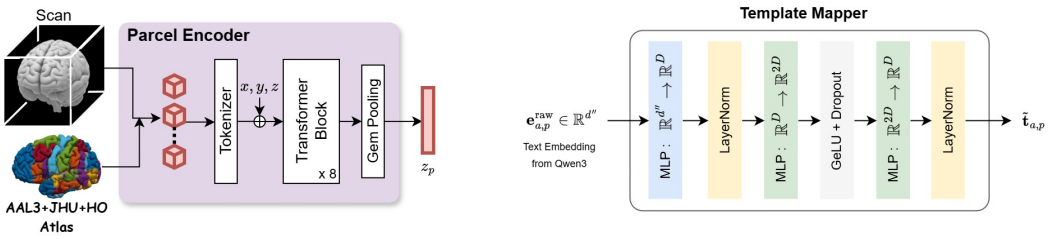


Figure A1: Details of: A. Parcel Encoder and B. the Template Mapper module.

B.2 Parcel Encoder details

The Parcel Encoder is adapted from P2P (Venkatesh et al. (2026)), which encodes a scan as a set of parcel embeddings using the Harvard Oxford atlas prior. For BrainNorm, we extend this technique to encode the parcels of the multi-atlas covering the whole brain. For each parcel, the framework first identifies and groups image patches that contain any information about the parcel. These patches are tokenized, and positional embeddings derived from their spatial coordinates are added to the corresponding tokens. Each parcel-specific token set is then independently processed by a shared Transformer encoder. Finally, the self-attended tokens are aggregated using GeM pooling (with $r = 2$) to obtain a single representation for each parcel. This is illustrated in Fig A1 A.

B.3 Template Mapper details

The template mapper is a 3-layer MLP as presented in Fig A1 B. The first layer projects the high dimensional text embedding to the SAL's dimension. The later part of the module is inspired from the transformer's Feed Forward network module with GeLU non-linearity. This allows the template embeddings to adapt according to the parcel representations.

B.4 Baselines

We evaluated BrainNorm's performance against nine baselines using the same preprocessed scan inputs, fine-tuned or trained under end-to-end task-specific supervision. These baselines are discussed in detail below:

- **BrainIAC** (Tak et al. (2026)): A foundation model, with 88.3M trainable parameters, trained and validated on 48,965 multiparametric brain MRIs using contrastive self-supervised pretraining.
- **BrainHarmony** (Dong et al. (2025)): A foundation model, with 88.4M trainable parameters, pretrained on 64,594 T1w sMRI brain scans.
- **NeuroJEPA** (Huang et al. (2026)): A foundation model, with ∼122M parameters, pretrained on 1,551,862 multiparametric brain MRI scans using a sparse latent predictive objective.
- **NeuroVFM** (Kondepudi et al. (2026)): A foundation model, with 86M parameters, pretrained on 5.24 million clinical MRI and CT volumes using Volumetric Joint-Embedding Predictive Architecture (Vol-JEPA).
- **RadFM** (Wu et al. (2025)): A generalist foundation model, with 104M parameters, trained on 13M 2D images and 615K 3D scans spanning a wide range of radiological modalities, covering 17 medical systems.
- **y-Aware** (Dufumier et al. (2021)): A 3D CNN model, with ∼11M params, pretrained on 10^4 multi-site healthy brain MRI scans using contrastive learning and conditioned on age.
- **BrainMVP** (Rui et al. (2025)): A multi-modal vision pre-training framework, with ∼22M trainable params, for brain image analysis using multi-parametric MRI scans. It was pretrained on 16,022 brain MRI scans.
- **MedicalNet** (Chen et al. (2019)): A 3D-ResNet152 model, with 117M parameters, pretrained on 23 medical datasets.

Table A2: **Ablation of the Template Mapper and AgeSoftCE objective.** (1) SigLIP + Template Mapper + AgeSoftCE (ours); (2) SigLIP + Template Mapper, without AgeSoftCE; (3) SigLIP without Template Mapper and AgeSoftCE

Dataset	Ablation	Text→Parcel		Parcel→Text		Age-MAE ↓	Pearson-r ↑
		Acc@1 ↑	Acc@5 ↑	Acc@1 ↑	Acc@5 ↑		
UKB	1	90.87	99.02	97.17	99.80	**4.14**	**0.85**
UKB	2	**90.99**	**99.09**	**97.99**	**99.99**	4.98	0.77
UKB	3	44.21	49.26	47.23	49.89	10.21	0.43
MCSA	1	67.33	85.19	**75.02**	**89.54**	**3.94**	**0.88**
MCSA	2	**70.53**	**86.60**	74.19	88.12	6.78	0.84
MCSA	3	30.23	35.89	38.85	41.56	12.51	0.46

- **3D ViT** (Dosovitskiy et al. (2021)): A ViT-B model trained from scratch with forward dimension as 2048 and ∼67M parameters, providing a capacity-matched comparison to BrainNorm.

BrainNorm was compared with a wide range of baselines that operate directly on volumetric MRI inputs, including models pretrained on near-equal or larger dataset sizes and capacities. This is to ensure fair comparison with respect to the scale of pretraining. We also compare against both domain-specific and general-purpose volumetric foundation models.

B.5 Ablations

B.5.1 Template Mapper & Loss Ablations

While the frozen text encoder provides a semantically structured representation with parcel clustering and age ordering, the observed separability and continuity in the visualization alone do not establish that these representations are sufficiently aligned with the parcels' biological aging trajectories. The Template Mapper is introduced to adapt these representations to learn image-derived, parcel-specific aging patterns using the AgeSoftCE loss. We conducted targeted ablations of the Template Mapper and the soft age objective to isolate their respective contributions. These variants are evaluated on parcel identification and age prediction tasks on the UKB test set and the external MCSA cohort.

The results in Table A2 clearly separate the roles of the two components. Removing the Template Mapper causes retrieval accuracy to collapse to near-chance levels and substantially degrades age prediction on both datasets, confirming that the frozen text embeddings alone are not aligned with the image-derived parcel representations. Removing only AgeSoftCE largely preserves parcel retrieval performance but worsens age MAE and Pearson correlation, particularly on the external MCSA cohort. This indicates that the soft age objective is specifically responsible for imposing an age-sensitive structure on the learned parcel representations. Thus, the Template Mapper establishes cross-modal parcel alignment, while AgeSoftCE promotes the continuous age structure required for normative modelling.

B.5.2 Text-Prior Initialization Ablations

We directly evaluate the role of the text prior by replacing Qwen3 templates with:

1. randomly initialized embeddings,
2. fixed categorical embeddings, and
3. 218 learnable parcel embeddings modulated by age through FiLM conditioning (Perez et al. (2018)).

Results on the UKB test set and the external MCSA cohort are reported in Table A3. These results show that categorical and learnable embeddings can preserve parcel identity, but are substantially weaker for age modelling. In particular, categorical initialization achieves near-perfect parcel retrieval on UKB and remains strong on MCSA, yet exhibits considerably higher age-prediction error. In contrast, Qwen3 achieves the lowest age MAE and highest Pearson correlation on both UKB

Table A3: **Ablation of text-prior initialization strategies.** We compare Qwen3-based template initialization against randomly initialized embeddings, fixed categorical embeddings, and learnable parcel embeddings with FiLM-based age conditioning.

Dataset	Initialization	Text→Parcel		Parcel→Text		Age-MAE ↓	Pearson-r ↑
		Acc@1 ↑	Acc@5 ↑	Acc@1 ↑	Acc@5 ↑		
UKB	Random	55.93	58.34	56.02	58.19	10.32	0.68
UKB	Categorical	**99.99**	**99.99**	**99.98**	**99.99**	10.05	0.73
UKB	Learnable	91.20	99.21	96.91	97.71	7.29	0.79
UKB	Qwen3	90.87	99.02	97.17	99.80	**4.14**	**0.85**
MCSA	Random	15.37	17.80	14.97	17.65	17.40	0.67
MCSA	Categorical	**94.99**	**98.17**	**92.83**	**97.67**	17.13	0.70
MCSA	Learnable	68.52	86.33	73.24	86.53	8.90	0.78
MCSA	Qwen3	67.33	85.19	75.02	89.54	**3.94**	**0.88**

Table A4: **Robustness to semantic rephrasing of text queries.** We evaluate four semantically equivalent prompt formulations that vary in syntax and word order.

Dataset	Template	Text→Parcel		Parcel→Text		Age-MAE ↓	Pearson-r ↑
		Acc@1 ↑	Acc@5 ↑	Acc@1 ↑	Acc@5 ↑		
UKB	1	**90.87**	**99.02**	**97.17**	**99.80**	4.14	0.85
UKB	2	86.10	97.87	94.49	99.66	4.47	0.85
UKB	3	85.96	97.89	93.43	99.58	3.99	0.85
UKB	4	80.39	96.57	93.48	99.45	**3.91**	0.85
MCSA	1	**67.33**	**85.19**	**75.02**	**89.54**	3.94	0.88
MCSA	2	62.32	83.30	71.67	88.76	3.88	**0.89**
MCSA	3	61.80	83.03	70.27	88.22	**3.71**	0.88
MCSA	4	60.23	81.78	69.60	87.64	3.80	0.88

and MCSA. This suggests that the semantic structure provided by the pretrained text representations yields more effective and generalizable age-conditioned parcel representations than parcel identity alone.

B.5.3 Text Query Ablations

BrainNorm's textual inputs provide flexibility in querying the SAL space, rather than constraining the model to a single fixed template. We evaluate four semantically equivalent text-query formulations with different syntax and word order:

- **Template 1 (ours):** Parcel=<parcel_name>; Tissue_class=<tissue>; Age=<a>y
- **Template 2:** The parcel is <parcel_name>, its tissue class is <tissue>, and the subject is <a> years old.
- **Template 3:** <parcel_name> of a <a>-year-old subject with <tissue> tissue.
- **Template 4:** A <a>-year-old subject showing the <parcel_name>, which has <tissue> tissue.

The results are presented in Table A4. Despite changes in syntax and word order, all prompt formulations retain strong parcel retrieval and comparable age-estimation performance across both the UKB test set and the external MCSA cohort. While retrieval performance decreases moderately for more substantially rephrased prompts, age MAE and Pearson correlation remain highly stable, indicating that the learned SAL representations preserve the underlying age-related structure across semantically equivalent formulations. These results demonstrate that BrainNorm is not restricted to a single fixed textual template and that the SAL space can be queried through multiple equivalent natural-language descriptions without retraining. This flexibility is inherently unavailable to categorical or learnable parcel-ID embeddings, which do not support semantic rephrasing at inference time.

Table A5: **Choice of Text Encoder.** Comparison of text encoders for the desired embedding geometry. The representation should (1) cluster samples by **parcel identity** and (2) exhibit a smooth **age trajectory within each parcel**. Metrics on the left measure parcel clustering, while metrics on the right measure age ordering within parcels. Best values are highlighted in bold.

	Parcel Clustering			Age Ordering Within Parcel		
Encoder	Silhouette ↑	Hit@1 ↑	NMI ↑	Spearman ↑	Pair Order Acc ↑	kNN Age MAE ↓
Qwen3-Embedding-8B (Yang et al. (2025))	0.4637	0.9999	0.9476	**0.9682**	**0.9311**	**0.5114**
MedEmbed-small (Balachandran (2024))	**0.6667**	**1.0000**	**0.9962**	0.5483	0.6922	1.9061
MedSigLIP (Sellergren et al. (2025))	0.1224	0.7284	0.8790	0.2906	0.5971	6.0694
BiomedCLIP (Zhang et al. (2025b))	0.6133	0.9999	0.9863	0.3816	0.6323	2.4691
Hashing+SVD	0.4440	0.9998	0.9939	0.0024	0.5010	3.4503
TF-IDF	0.2225	0.0412	0.9928	0.0002	0.4996	4.1861

B.5.4 Choice of Text Encoder

Why semantic atlas pretraining instead of direct trajectory regression? Although the retrieval-based age estimator resembles non-parametric regression at inference time, BrainNorm differs in the representation over which retrieval is performed. Classical trajectory estimators such as k-NN, kernel regression, Gaussian processes, or brain-chart models typically operate on predefined morphometric summaries or low-dimensional regional statistics. In contrast, BrainNorm learns high-dimensional parcel-level representations directly from structural MRI and aligns them with age-conditioned semantic atlas descriptors. The text encoder is therefore not used to inject unconstrained natural-language knowledge, but to provide a structured semantic coordinate system over parcel identity, tissue class, and age. This encourages prompts from the same anatomical parcel to remain coherent while allowing age-conditioned prompts to form smooth within-parcel trajectories. Consequently, age retrieval, brain age gap estimation, parcel identification, and zero-shot anomaly scoring are all performed in a shared normative representation space rather than through independent per-task regressors. This is broader than a non-parametric age trajectory estimator. The retrieval formulation is therefore used as an interpretable inference mechanism over the learned normative representation.

Selecting a semantic embedding space with parcel-aware age geometry: BrainNorm's representation requires a specific geometric structure in the SAL space: clustering by parcel and ordering by age. This structure allows the model to interpret age-conditioned prompts as continuous trajectories within each anatomical region. To identify an embedding model that satisfies these requirements, we evaluate several candidate text encoders using two complementary groups of metrics.

Parcel clustering metrics: These metrics evaluate whether prompts corresponding to the same anatomical parcel remain grouped together in the embedding space. We first compute the *silhouette score*, which measures how well each sample is separated from samples of other parcels relative to samples from its own parcel. Higher values indicate that embeddings belonging to the same parcel form compact clusters that are clearly separated from other parcels. To evaluate local parcel consistency, we compute *nearest-neighbor parcel retrieval (Hit@1)*, defined as the fraction of samples whose nearest neighbor in the embedding space belongs to the same parcel. Finally, we measure *normalized mutual information (NMI)* between the ground-truth parcel assignments and clusters obtained by applying k-means to the embeddings (with k equal to the number of parcels). NMI evaluates how well the clustering structure induced by the embeddings aligns with the true anatomical parcel labels.

Age trajectory metrics: These metrics assess whether embeddings within a parcel exhibit a consistent progression with age. For each parcel, we first compute the *Spearman rank correlation* between age and the dominant direction of variation in the embeddings, obtained via principal component analysis. A high correlation indicates that age progression aligns with the principal trajectory of the parcel embeddings. We additionally evaluate *pairwise age-order accuracy*, which measures the fraction of sample pairs within a parcel for which the relative ordering of their embedding coordinates matches the chronological ordering of their ages. This metric quantifies whether the embedding space preserves the temporal ordering of samples. Finally, we compute the *local nearest-neighbor age error*, defined as the mean absolute difference in age between each sample

and its nearest neighbors within the same parcel. Low error values indicate that nearby embeddings correspond to samples with similar ages, reflecting a smooth age trajectory within each parcel.

Table A5 reports the results across several biomedical and SOTA text encoders as well as classical baselines (TF-IDF and Hashing+SVD). These classical methods embeddings do not guarantee smooth age geometry and semantic organization simultaneously unless explicitly regularized. While several encoders achieve strong parcel discrimination, **Qwen3-Embedding-8B** produces the most consistent age-structured geometry. In particular, it yields substantially higher parcel-wise age monotonicity and pairwise age-order accuracy while simultaneously achieving the lowest local age prediction error. These results indicate that Qwen3 embeddings naturally organize prompts into parcel-specific manifolds where age varies smoothly along the dominant direction. Consequently, we adopt Qwen3 as the text encoder for constructing *parcel-specific age-conditioned representations* for BrainNorm.

B.6 Fine-tuning Details

BrainNorm was fine-tuned in a 5-fold setting using only CN subjects from the mixed downstream cohort, with the same training objective as during pretraining (Eq. 8). The learning rates for the Parcel Encoder and Template Mapper were set to 5×10^{-6} and 1×10^{-6}, respectively, and finetuning was performed for 5–10 epochs. For linear probing, the learning rate is set to 1×10^{-3} for 15–20 epochs. All other training settings were kept identical to those used during pretraining.

Learnable parameters: The learned parcel-specific σ_p lie in the range 0.60–0.76, with the majority concenterated around 0.6.

B.7 Implementation details of SigLIP loss

In our implementation, SigLIP alignment is performed between parcel embeddings extracted from MRI scans and a bank of age-conditioned textual templates representing anatomical parcels. We perform distributed training across 4 GPUs.

Distributed pair construction: Across all GPUs, the union of these ages is gathered using distributed communication. Template coordinates for the corresponding (p, a) unique combinations are then retrieved from the template bank. To efficiently compute similarities across a large number of parcel templates, the template bank is sharded across GPUs along the parcel dimension. Each GPU retrieves and processes templates for a subset of parcels, computes the mapped template embeddings locally, and then gathers these embeddings across GPUs using distributed communication to obtain all similarities of the template bank in that batch. This strategy avoids replicating the entire template bank on every GPU while still allowing each batch to compute similarities against the full set of parcel–age templates.

Positive pair selection: Positive supervision is defined using parcel identity and an age window of ±1 year around the subject age. For a given parcel embedding, all templates corresponding to the same parcel and whose ages fall within this window are treated as positive pairs. When multiple valid templates exist within the age window, their contributions are handled using soft targets rather than selecting a single positive pair. This allows nearby age templates to contribute to the learning signal while preserving the strongest supervision at the exact age when available.

Similarity computation: After gathering the template bank across GPUs, cosine similarities between parcel embeddings and all candidate templates are computed in the shared embedding space. These similarities are used to compute SigLIP binary cross-entropy loss as described in main paper.

B.8 Details of supervised age prediction tasks for prior works

For age prediction, pretrained checkpoints from all baseline models are used to initialize their respective backbone architectures. The original task-specific heads are replaced with regression layers designed to output continuous age estimates, after which the models are fine-tuned on the target dataset (CN-only) for chronological age prediction. For 3D-ViT, the model is trained from scratch. Prior to training, target age values are normalized using the mean and standard deviation of the

Table A6: Parcel identification on held-out test sets and downstream datasets using only CN subjects (5-fold mean$_{\pm \mathrm{std}}$ where available). PT: pretrained encoder used directly for inference. FT: finetuned only on the CN subset of the corresponding dataset or finetuned on pooled CN data from ADNI+AIBL+MIRIAD+NIFD+PPMI (All).

Dataset	Mode	CN data	Text→Parcel		Parcel→Text	
			Acc@1 ↑	Acc@5 ↑	Acc@1 ↑	Acc@5 ↑
UKB	No FT	UKB	90.87	99.02	97.17	99.80
	FT	UKB	90.32	99.04	96.72	99.76
MCSA	No FT	UKB	67.33	85.19	75.02	89.54
	FT	All	76.06	94.54	86.46	97.81
ADNI	No FT	UKB	$69.50_{\pm 0.35}$	$87.45_{\pm 0.24}$	$77.54_{\pm 0.31}$	$91.17_{\pm 0.23}$
	FT	ADNI	$76.64_{\pm 3.01}$	$94.75_{\pm 1.48}$	$86.95_{\pm 2.18}$	$97.76_{\pm 0.87}$
	FT	All	$79.88_{\pm 1.11}$	$96.32_{\pm 0.39}$	$89.64_{\pm 0.78}$	$98.65_{\pm 0.20}$
AIBL	No FT	UKB	$68.71_{\pm 0.65}$	$86.20_{\pm 0.48}$	$76.64_{\pm 0.83}$	$90.27_{\pm 0.43}$
	FT	AIBL	$77.57_{\pm 1.54}$	$94.29_{\pm 0.69}$	$85.53_{\pm 0.91}$	$97.04_{\pm 0.35}$
	FT	All	$80.01_{\pm 1.25}$	$96.01_{\pm 0.50}$	$89.89_{\pm 1.18}$	$98.64_{\pm 0.30}$
MIRIAD	No FT	UKB	$73.93_{\pm 1.18}$	$90.33_{\pm 0.69}$	$81.30_{\pm 1.01}$	$93.32_{\pm 0.89}$
	FT	MIRIAD	$76.35_{\pm 2.43}$	$92.19_{\pm 1.22}$	$83.56_{\pm 1.83}$	$94.92_{\pm 1.10}$
	FT	All	$83.49_{\pm 2.88}$	$97.74_{\pm 0.83}$	$93.20_{\pm 1.76}$	$99.35_{\pm 0.59}$
NIFD	No FT	UKB	$70.62_{\pm 0.59}$	$86.78_{\pm 0.50}$	$63.84_{\pm 3.11}$	$80.40_{\pm 2.62}$
	FT	NIFD	$75.71_{\pm 0.39}$	$92.39_{\pm 0.26}$	$83.47_{\pm 0.41}$	$95.10_{\pm 0.27}$
	FT	All	$82.06_{\pm 0.64}$	$96.71_{\pm 0.29}$	$91.35_{\pm 0.37}$	$98.91_{\pm 0.18}$
PPMI	No FT	UKB	$67.18_{\pm 0.85}$	$84.93_{\pm 0.58}$	$74.84_{\pm 0.60}$	$88.72_{\pm 0.34}$
	FT	PPMI	$71.56_{\pm 0.82}$	$90.01_{\pm 0.53}$	$78.69_{\pm 0.65}$	$92.85_{\pm 0.57}$
	FT	All	$79.47_{\pm 1.09}$	$95.59_{\pm 0.56}$	$88.71_{\pm 0.54}$	$98.13_{\pm 0.27}$

Table A7: Disease-specific prior parcel IDs used for region-restricted analysis.

Class	Parcel IDs
AD/MCI	38, 39, 40, 41, 36, 37, 68, 69, 66, 67, 62, 63, 84, 85, 18, 19, 216, 217
PD	156, 157, 158, 159, 74, 75, 72, 73, 76, 77, 154, 155, 160, 161, 162, 163, 164, 165, 0, 1, 14, 15, 58, 59, 172, 173
FTD	2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 82, 83, 86, 87, 80, 81, 84, 85, 32, 33, 146, 147, 148, 149, 150, 151, 24, 25, 26, 27, 28, 29, 30, 31, 20, 21, 152, 153

training set to stabilize optimization. The models then predict normalized age values, and Smooth L1 loss is applied between predicted and normalized targets to provide robustness against outliers while maintaining sensitivity to smaller errors. During inference, predicted normalized ages are denormalized back to the original age scale for evaluation using metrics such as MAE, RMSE, Pearson correlation, and R^2. This approach allows all three pretrained architectures to adapt their learned feature spaces while preserving the benefits of prior pretraining.

Table A10 provides the results of age prediction for BrainNorm and baselines across all datasets. The results are reported as $mean_{\pm std}$ across all folds of downstream mixed cohorts.

B.9 Parcel Identification

The detailed results of both Text→Parcel and Parcel→Text is presented in Table A6 with $mean_{\pm std}$ values across all 5-folds. Across all datasets, the performance improves with finetuning on more healthy scans.

B.10 Zero-shot Prediction Details

Note: No diseased samples or disease labels are used to construct the normative references, define the scoring functions, or fit task-specific classifiers. Disease labels are used only for evaluation. For Disease-Prior Restricted Anomaly [Sec. 4.5(a)(2)], disease-relevant parcels are defined solely from established clinical neurodegeneration literature. We therefore refer to this setting as *prior-guided zero-shot classification*, conceptually related to attribute-based zero-shot learning, where

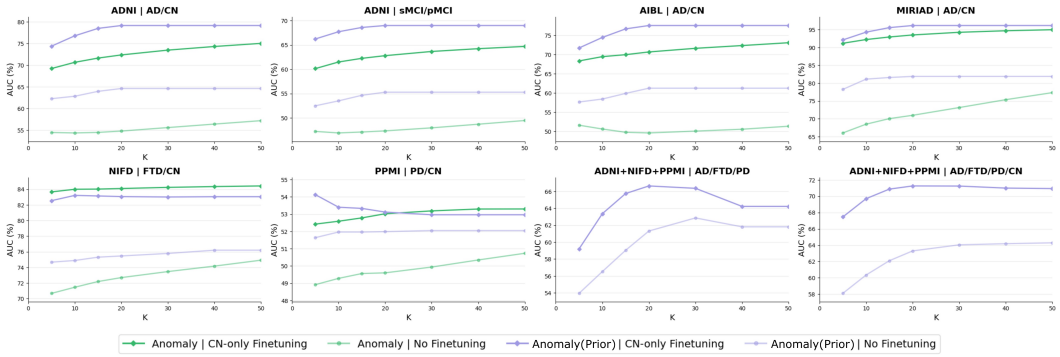


Figure A2: Zero-shot performance with increasing top-K parcels considered for both with and without fine-tuning on the full dataset.

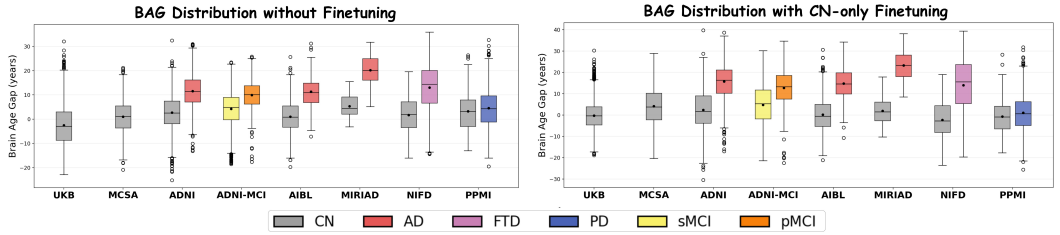


Figure A3: **BAG distribution** improves with finetuning across all datasets and groups.

unseen classes are identified through externally defined attributes rather than labeled class examples (Lampert et al. (2013); Song et al. (2018)).

Anomaly-based: The zero-shot experiments that utilized priors for anomaly prediction incorporated only those anatomical regions known to be affected according to the neurodegeneration literature (Planche et al. (2022); Braak et al. (2006); Rascovsky et al. (2011); Gorno-Tempini et al. (2011); Hughes et al. (1992); Postuma et al. (2015)). We identified the most relevant parcels associated with each diagnosis, as listed in Table A7. These parcel IDs can be mapped to their corresponding anatomical regions using Table A15.

For each scan, let $\mathbf{s} = [s_1, \ldots, s_P]^\top \in \mathbb{R}^P$, denote the parcel-wise similarity vector between the scan embeddings and the age-matched normative template, where higher similarity indicates stronger agreement with the normative reference and lower similarity indicates greater deviation. The global anomaly score and disease-prior restricted anomaly scores are then computed (Sec 4.5a) from this similarity profile using $K = 20$. The results are reported in Table A11. A sweep on increasing K values are also recorded and plotted in Fig A2

For binary zero-shot diagnosis, these anomaly scores are used directly as continuous diagnostic variables. In particular, we do not train any task-specific classifier on top of the frozen representations. Instead, the scalar anomaly score itself is treated as the ranking variable, and performance is evaluated using the area under the ROC curve (AUC).

For multi-disease zero-shot evaluation, we compute one class-specific anomaly score for each disease using its corresponding anatomical prior subset. These scores are not summed across diseases. Rather, each disease hypothesis produces its own score, which is interpreted as class-specific diagnostic evidence. The cognitively normal class is represented using the negative global anomaly score,

$$\ell_{\text{CN}} = -D_{\text{global}},$$

while each disease class d is represented by its own prior-restricted anomaly score, $\ell_d = D_d$. These class-specific scores are converted into probabilities through a softmax over class logits,

$$P(y = d \mid x) = \frac{\exp(\ell_d)}{\sum_c \exp(\ell_c)}.$$

Thus, multi-disease zero-shot prediction is based on relative comparison across disease-specific anomaly scores rather than on any learned classifier.

Overall, this protocol evaluates whether BrainNorm's frozen age-conditioned normative representation, together with anatomical disease priors, is sufficient to separate controls from disease and to distinguish among multiple disease classes without explicit supervised training. Detailed results of zero-shot prediction for all methods across all tasks and datasets are provided in Table A11.

BAG-based: BrainNorm also enables zero-shot disease prediction through Brain Age Gap (BAG). For each scan x_i with chronological age y_i, we first obtain the retrieval-based predicted age $\hat{y}_i$ from the age-conditioned normative template space and compute:

$$\mathrm{BAG}_i = \hat{y}_i - y_i. \tag{9}$$

We use BAG_i directly as a disease-agnostic anomaly score, where larger positive values indicate that the brain appears older than expected under the learned normative aging trajectory.

We report threshold-free AUROC performance for all tasks. The values are reported as $mean_{\pm std}$ across all 5-folds in Table A11. Similarly for all baselines works, we used the same finetuned checkpoint for age prediction to calculate BAG for comparison.

Visualization of BAG distributions across diagnostic groups before and after finetuning are presented in Fig. A3. CN-only finetuning yields improved group separation across cohorts.

B.11 Traditional Normative Model baseline comparison

BrainNorm characterizes abnormality as deviation from parcel-specific, age-conditioned healthy reference templates. We assess the abnormalities captured by BrainNorm in two complementary settings:

1. **subject-level anomaly scores**, evaluated through zero-shot disease classification (Sec. 4.5a); and
2. **parcel-level deviation representations**, evaluated by linearly probing the frozen representations (Sec. 4.5b, BrainNorm_M2).

We compare both settings against Hierarchical Bayesian Regression (HBR), using conventional morphometric features extracted with FreeSurfer, including regional volume, cortical surface area, and cortical thickness, following the traditional normative-modeling formulation of Rutherford et al. (Rutherford et al. (2022)). HBR is pretrained on UK Biobank and evaluated across AD/CN and sMCI/pMCI from ADNI, AD/CN from AIBL & MIRIAD, PD/CN from PPMI, FTD/CN from NIFD.

For zero-shot disease classification, we consider: (i) Direct transfer of the UKB-trained normative model to the target cohort; and (ii) CN-only target-cohort adaptation using only healthy control subjects, without using disease labels. For linear probing without fine-tuning, we probe BrainNorm's frozen parcel-deviation representations (BrainNorm_M2) and compare them against a linear probe trained on HBR's concatenated regional morphometric deviation scores. All results are reported as the mean over 5-fold cross-validation.

Inference. These results reveal important distinctions between the two normative-modelling pipelines.

- HBR provides a good classical normative baseline and achieves higher subject-level anomaly-score AUC than unadapted BrainNorm on several direct-transfer tasks. Following healthy-only target-cohort fine-tuning, BrainNorm improves substantially and achieves stronger anomaly-score performance on four of the six evaluated tasks, demonstrating that its learned healthy references can be effectively adapted to cohort-specific variation without disease supervision. Nevertheless, BrainNorm enables zero-shot tasks on direct sMRI scan inputs unlike other existing foundation models.
- Healthy-only fine-tuning particularly improves BrainNorm on AD/CN, sMCI/pMCI, and FTD/CN, suggesting that cohort-specific calibration better aligns the learned parcel-wise

Table A8: **Zero-shot disease classification using subject-level anomaly scores.** Comparison between HBR and BrainNorm under direct transfer from UKB and healthy-only adaptation. Results are reported as AUC (%). Best and second-best results for each task may be highlighted in bold and italics, respectively.

Task	HBR Direct Transfer	HBR CN-only Adaptation	BrainNorm Direct Transfer	BrainNorm CN-only Fine-tuning
AD/CN [ADNI]	**75.20**	72.01	54.78	*72.37*
AD/CN [MIRIAD]	*90.81*	88.97	71.00	**93.51**
AD/CN [AIBL]	**71.04**	*72.51*	49.55	70.70
sMCI/pMCI [ADNI]	58.81	*59.89*	47.33	**62.79**
PD/CN [PPMI]	*51.86*	49.96	49.60	**53.02**
FTD/CN [NIFD]	67.85	*75.21*	72.69	**84.09**

Table A9: **Linear probing of frozen normative-deviation representations.** No CN-only adaptation or downstream representation fine-tuning is performed. BrainNorm_M2 uses frozen parcel-deviation representations, while HBR uses concatenated regional morphometric deviation scores. Δ denotes BrainNorm minus HBR. Results are reported in %.

Task	HBR BAcc	HBR AUC	BrainNorm BAcc	BrainNorm AUC	Δ BAcc	Δ AUC
AD/CN [ADNI]	82.29	90.25	**86.85**	**94.27**	+4.56	+4.02
AD/CN [AIBL]	83.53	91.06	**85.91**	**93.39**	+2.38	+2.33
AD/CN [MIRIAD]	88.47	94.57	**92.72**	**98.26**	+4.25	+3.69
sMCI/pMCI [ADNI]	67.47	74.55	**71.96**	**80.75**	+4.49	+6.20
PD/CN [PPMI]	55.87	60.42	**59.08**	**66.12**	+3.21	+5.70
FTD/CN [NIFD]	70.14	75.42	**81.82**	**89.15**	+11.68	+13.73
AD/PD/FTD	80.85	92.86	**83.98**	**97.01**	+3.13	+4.15
AD/PD/FTD/CN	68.69	86.64	**70.32**	**91.91**	+1.63	+5.27

healthy references with the target population. Importantly, BrainNorm operates directly on sMRI and derives normative deviations in the learned representation space without requiring predefined morphometric measurements.

- Beyond subject-level anomaly scores, BrainNorm represents abnormality through high-dimensional parcel-deviation vectors. Linear probing of these frozen representations consistently outperforms probing HBR's concatenated morphometric deviation scores, indicating that the learned deviation representations retain disease-relevant information beyond what is captured by scalar regional abnormalities alone.
- The strong performance of BrainNorm's frozen parcel-deviation representations even without CN-only fine-tuning further indicates that disease-relevant deviation structure is already encoded during healthy-only pretraining, although simple global anomaly aggregation may not fully capture this information.

BrainNorm's subject-level anomaly scoring and its parcel-level deviation representations captures complementary aspects of the learned normative space.

B.12 Formulation of Linear Probing

To test whether disease-related signals are linearly extractable from the frozen representation, we train lightweight supervised heads on top of the feature constructions (BrainNorm_M1, BrainNorm_M2). Additionally, we also consider:

(M3) Parcel similarity vector. We directly use the age-matched parcel-template cosine similarities $\mathbf{x}_{M3} = [s_1, \ldots, s_P]^\top \in \mathbb{R}^P$, where $s = \mathbf{p}_{a,p} \cdot \mathbf{t}_{a,p}$.

Classifier architecture. Each feature vector $\mathbf{x}$ is fed to a single-layer MLP. For binary tasks, the MLP outputs logits for two classes; for multiclass tasks, it outputs C logits followed by softmax at evaluation time. Detailed results of linear-probing prediction and few-shot prediction for all three methods across all tasks and datasets are provided in Table A12, and Table A13, A14 respectively.

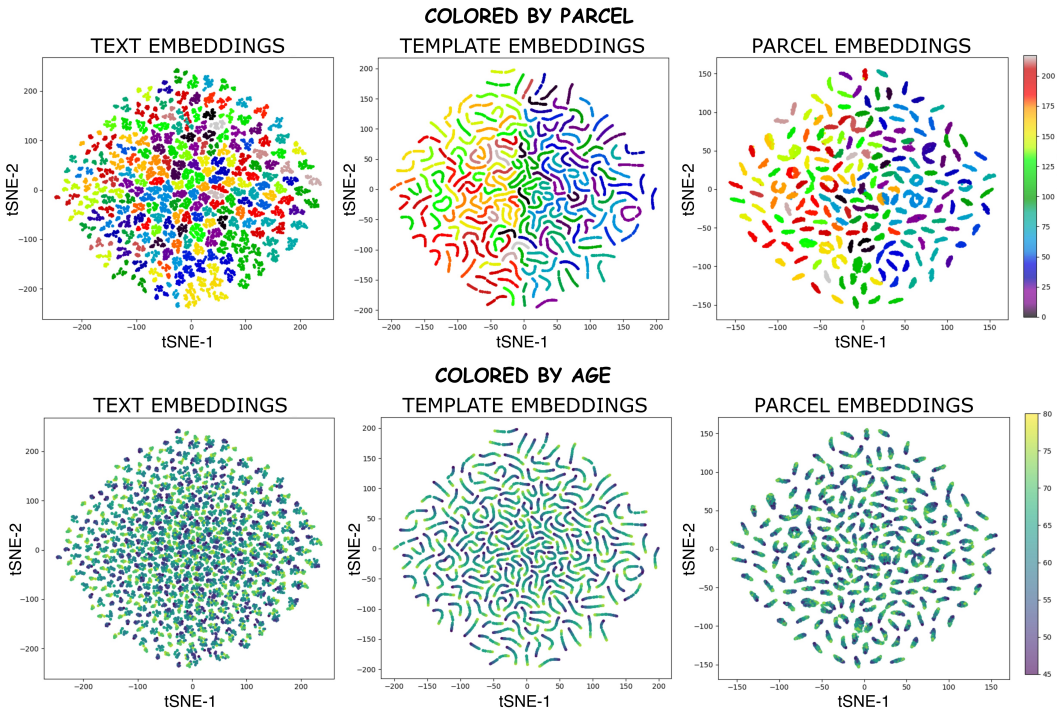


Figure A4: t-SNE visualizations of text priors ($\mathbf{e}_{a,p}^{\text{raw}}$), learned templates ($\mathbf{t}_{a,p}$), and parcel embeddings ($\mathbf{p}_{i,p}$) on held-out UKB (Sudlow et al. (2015)) test data. The learned SAL space exhibits visible anatomical clustering and ribbon-like derived structure of age-ordering within each parcel.

B.13 BrainNorm's embeddings visualizations

Fig. A4 visualizes the learned Semantic Atlas Latent (SAL) space, where BrainNorm's template and parcel embeddings organize according to both anatomical identity and age. Embeddings from the same parcel form coherent clusters, while within each cluster, age-conditioned embeddings follow a smooth ordering. This suggests that the learned space preserves parcel-specific identity while capturing normative age-related variation.

The Fig A5 visualizes the two-dimensional UMAP projection of residual whole-brain embeddings computed after subtracting age-matched normative template representations. The scatter plot shows individual subjects, with colours indicating diagnostic groups and age gradients within each group. The inset density panels provide group-specific density structure, while the combined density map shows the relative spatial organization and overlap between diagnostic groups. The residual embedding visualization shows that age-matched normative deviations organize subjects into clusters while also preserving an apparent age-related ordering within the embedding space. The plots suggests that CN subjects largely occupy the upper residual space, whereas AD shifts toward the lower-left/lower-central region, PD extends toward the upper-right with broader dispersion, and FTD forms a more compact lower-right cluster. The density plots further summarize these class-wise regions, showing high-density zones with some overlap between disease groups. This indicates that the learned residual embeddings capture both diagnostic deviation patterns and age-dependent variation, suggesting that the model retains meaningful information about normal aging while also separating disease-associated departures from the normative trajectory. More importantly, the confounder visualizations show no obvious clustering by scanner manufacturer, sex, or field strength, providing qualitative evidence that these factors do not dominate the observed low-dimensional organization.

B.14 Scan-level Interpretability

To examine whether the learned normative representation provides interpretable evidence at the individual-subject level, we analyze representative correct and incorrect predictions for AD, FTD,

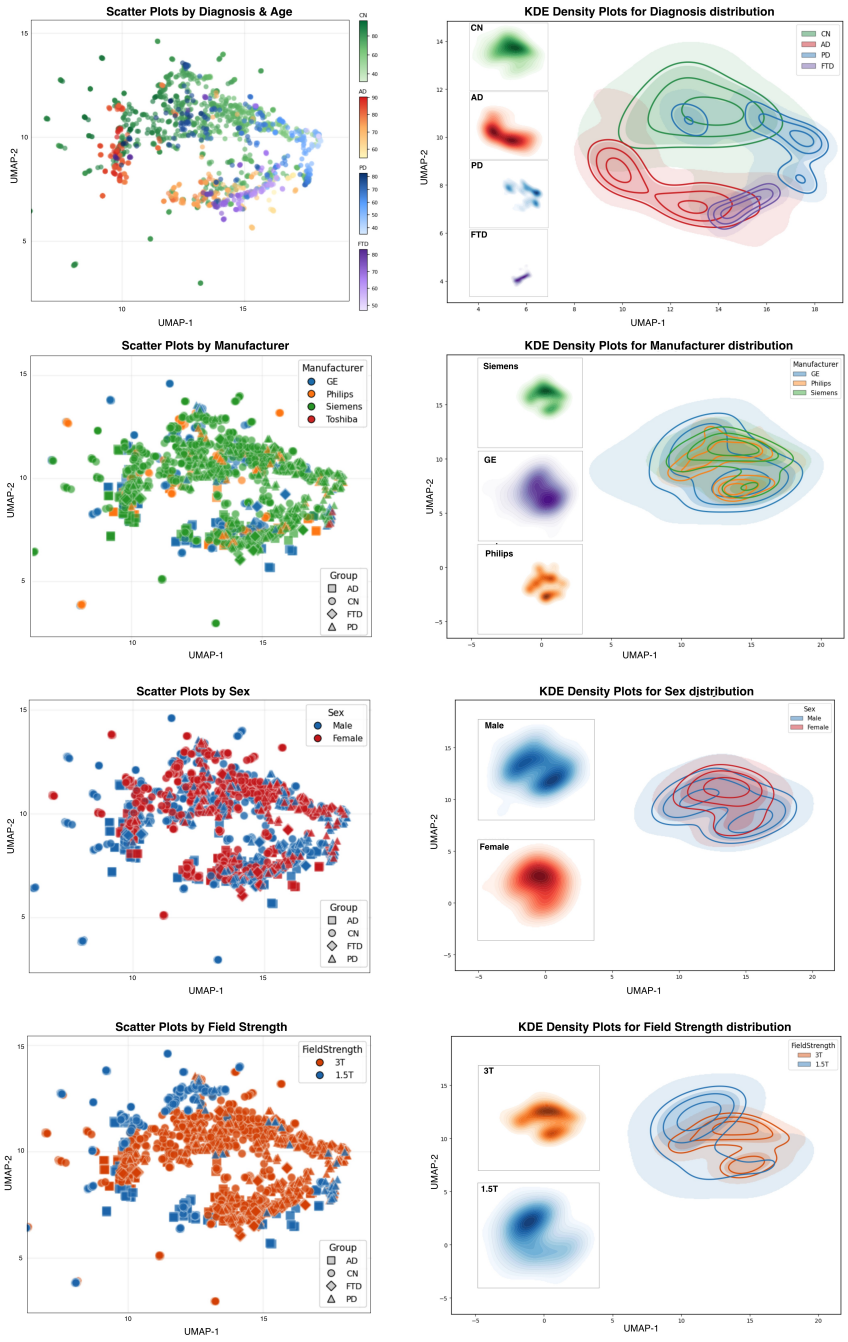


Figure A5: UMAP scatter plots and KDE density maps visualize the distribution of scan-level residual embeddings across neurodegenerative cohorts. The diagnosis-based scatter plot is colored by clinical group with an age gradient, revealing an apparent age-related organization in the embedding space, where relatively younger scans tend to lie toward the right and older scans toward the left. The diagnosis-specific KDE maps further show structured group-wise density patterns, suggesting that the residual embedding space captures disease-associated deviations from the normative representation. In contrast, the manufacturer-, sex-, and field-strength-based scatter plots and KDE maps show that scans from different vendors, sex categories, and field strengths are broadly dispersed across the embedding space rather than forming distinct confounder-specific clusters. This suggests that the observed diagnostic organization is unlikely to be primarily driven by scanner manufacturer, sex, or field-strength effects, supporting the interpretation that the learned SAL space captures biologically meaningful anomaly deviations.

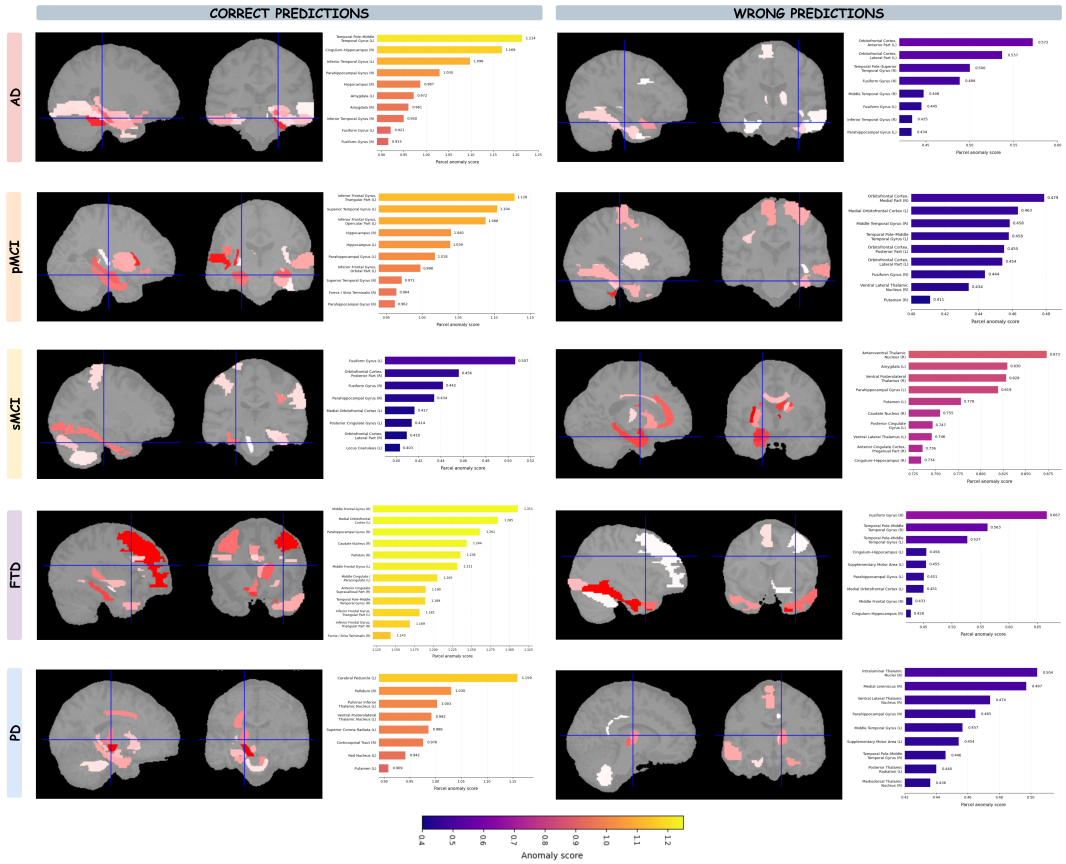


Figure A6: **Scan-level interpretability across neurodegenerative diseases.** Representative correct and incorrect predictions are shown for AD, pMCI, sMCI, FTD, and PD. Correct predictions correspond to AD, pMCI, sMCI, FTD, and PD subjects assigned to their respective ground-truth classes, whereas wrong predictions correspond to AD, FTD, and PD subjects misclassified as CN, pMCI subjects misclassified as sMCI, and sMCI subjects misclassified as pMCI. For each subject, the spatial map shows the highest-ranking parcel-wise deviations from the age-matched normative representation, with the image colors normalized independently within that subject to emphasize the spatial distribution of its regional abnormalities. Higher deviation scores from normative reference are represented by darker shades. The adjacent bar plot reports the corresponding raw parcel anomaly scores, with bar colors mapped using a common anomaly-score scale across all scans to enable direct comparison of deviation magnitude between subjects. Across correct predictions, the highlighted parcels show strong disease-relevant deviations. In contrast, wrong predictions generally retain anatomically relevant regional patterns but exhibit anomaly magnitudes that are shifted in the direction of the predicted class: AD, FTD, PD, and pMCI false negatives show comparatively weaker deviations, whereas sMCI subjects misclassified as pMCI show comparatively stronger deviations.

PD, sMCI, and pMCI. For each scan, parcel-wise anomaly is quantified as the deviation from its age-matched normative representation, and the highest-ranking disease-relevant parcels are projected back onto the image. The corresponding parcel anomaly scores are shown using a shared color scale across subjects, such that both the spatial localization and absolute magnitude of normative deviation can be compared directly (Fig. A6).

Correct predictions exhibit distinct and anatomically coherent deviation patterns across diseases. Correctly identified AD demonstrates prominent temporal and limbic abnormalities, including the temporal pole, hippocampus, parahippocampal gyrus, amygdala, inferior temporal cortex, and fusiform gyri. Similarly, the correctly classified pMCI subject exhibits strong hippocampal, parahippocampal, and temporal-related deviations, consistent with a more pronounced progression-like pat-

tern. In contrast, the correctly identified sMCI subject shows substantially weaker regional deviations. FTD is characterized by strong frontal, orbitofrontal, cingulate, temporal, and striatal deviations, while the correctly identified PD subject shows a more motor-system-dominant pattern involving the cerebral peduncle, pallidum, putamen, thalamic nuclei, corona radiata, corticospinal tract, and red nucleus. Thus, the normative anomaly representation does not collapse different disorders onto a common abnormality score, but preserves heterogeneous regional patterns at the individual-scan level.

The failure cases further clarify the model's decision mechanism. AD and FTD subjects predicted as CN retain partially disease-consistent temporal or frontotemporal abnormalities, respectively, but their anomaly magnitudes are markedly reduced relative to the correctly classified counterparts. A similar behavior is observed for PD, where the false-negative scan contains thalamic and motor-related abnormalities but lacks the stronger and more complete basal-ganglia–motor pathway pattern observed in the correct prediction. The MCI errors show the complementary behavior: a pMCI subject classified as sMCI exhibits only weak frontal-temporal deviations, whereas an sMCI subject classified as pMCI shows substantially elevated limbic, cingulate, and thalamic abnormalities. These examples suggest that errors arise not because the model attends to anatomically unrelated regions, but because the strength and completeness of disease-relevant normative deviations place the subject on the opposite side of the decision boundary.

Together, these results demonstrate that the learned parcel-wise normative space supports subject-specific interpretation at two complementary levels: it localizes where an individual deviates from the expected healthy representation and quantifies how strongly those regions deviate. This provides an interpretable explanation for both successful and failed predictions. Future work can explore reducing these wrong predictions by improving the alignment of the learned normative representation, enabling more accurate capture of disease-relevant deviations at the scan and parcel levels.

C Limitations, Scope and Future Directions

BrainNorm currently relies on spatial normalization to MNI152 space to ensure anatomical priors are consistently transferred across subjects. Reducing dependence on spatial normalization remains an important direction for future work. BrainNorm uses a fixed multi-atlas parcellation designed to provide complementary whole-brain coverage across gray matter, white matter, and ventricular regions. Replacing this unified atlas with a single coarser or finer alternative would therefore change not only parcel granularity, but also anatomical and tissue coverage, making it difficult to attribute performance differences solely to atlas resolution. A controlled multi-atlas granularity analysis is similarly non-trivial, since the granularity of gray-matter, white-matter, and ventricular regions can be varied independently, resulting in numerous atlas combinations with different spatial coverage and parcel definitions. Systematically characterizing these combinations and their influence on the learned representations, therefore, warrants a dedicated study, which we identify as an important direction for future work.

Finally, BrainNorm is currently trained and evaluated on adult and aging cohorts, consistent with the anatomical priors, healthy reference populations, and neurological conditions considered in this study. The framework is extensible to pediatric and neurodevelopmental settings, but such extension would require age-appropriate normative cohorts, anatomical references, and validation across developmental trajectories that differ substantially from adult aging. Nevertheless, adult and aging neuroimaging itself spans considerable anatomical variability, multiple neurological disorders, and diverse clinical settings, providing a broad and challenging scope for the present evaluation. Future work will also investigate multimodal extensions incorporating T2w, FLAIR, and fMRI to capture complementary aspects of normative brain variation. We consider both age-range expansion and multimodal modeling important directions for future work.

Table A10: BrainNorm's Age estimation on held-out test sets and downstream datasets using only CN subjects vs Prior works age prediction with full supervision. FT denotes finetuning on the CN subset of the corresponding dataset or pooled CN data from ADNI+AIBL+MIRIAD+NIFD+PPMI (All). Values are reported as mean$_{\pm \mathrm{std}}$ over 5 folds where available.

Dataset	Method	Mode	CN Train/FT data	MAE ↓	RMSE ↓	Pearson r ↑	R^2 ↑
UKB	3D-ViT	Full Train	UKB	4.23	5.65	0.77	0.58
	BrainIAC	Full FT	UKB	6.76	8.30	0.16	0.16
	BrainMVP	Full FT	UKB	4.15	5.21	0.83	0.69
	BrainHarmony	Full FT	UKB	4.15	**5.20**	0.83	0.66
	RadFM	Full FT	UKB	4.83	5.62	0.82	0.67
	NeuroVFM	Full FT	UKB	4.25	5.11	0.81	**0.72**
	Neuro-JEPA	Full FT	UKB	4.64	5.58	0.81	0.62
	MedicalNet	Full FT	UKB	4.16	5.27	0.82	0.53
	y-Aware	Full FT	UKB	4.17	5.23	0.82	0.54
	BrainNorm	No FT	UKB	**4.14**	5.22	**0.85**	0.54
	BrainNorm	CN-only FT	All	4.69	5.77	0.84	0.44
MCSA	3D-ViT	Infer	UKB	9.06	10.68	0.68	0.02
	BrainIAC	Infer	UKB	18.49	19.83	0.04	0.01
	BrainMVP	Infer	UKB	4.70	5.77	0.88	0.66
	BrainHarmony	Infer	UKB	5.05	6.16	0.86	0.61
	RadFM	Infer	UKB	5.55	6.81	0.86	0.52
	NeuroVFM	Infer	UKB	5.93	7.21	0.78	0.46
	Neuro-JEPA	Infer	UKB	6.24	7.46	0.85	0.43
	MedicalNet	Infer	UKB	8.75	10.42	0.56	0.10
	y-Aware	Infer	UKB	5.43	6.78	0.80	0.53
	BrainNorm	No FT	UKB	3.94	4.90	0.88	0.75
	BrainNorm	CN-only FT	All	**3.66**	**4.61**	**0.89**	**0.78**
ADNI	3D-ViT	Full Train	ADNI	$4.48_{\pm 0.16}$	$5.65_{\pm 0.23}$	$0.62_{\pm 0.03}$	$0.38_{\pm 0.03}$
	BrainIAC	Full FT	ADNI	$5.60_{\pm 0.06}$	$7.11_{\pm 0.12}$	$0.01_{\pm 0.13}$	$0.00_{\pm 0.01}$
	BrainMVP	Full FT	ADNI	$4.46_{\pm 0.13}$	$5.70_{\pm 0.16}$	$0.66_{\pm 0.04}$	$0.37_{\pm 0.04}$
	BrainHarmony	Full FT	ADNI	$4.38_{\pm 0.17}$	$5.52_{\pm 0.17}$	$0.64_{\pm 0.04}$	$0.40_{\pm 0.05}$
	RadFM	Full FT	ADNI	$3.87_{\pm 0.24}$	$4.71_{\pm 0.22}$	$0.74_{\pm 0.01}$	$0.60_{\pm 0.02}$
	NeuroVFM	Full FT	ADNI	$3.85_{\pm 0.03}$	$4.94_{\pm 0.03}$	$0.73_{\pm 0.00}$	$0.52_{\pm 0.01}$
	Neuro-JEPA	Full FT	ADNI	$3.98_{\pm 0.06}$	$4.98_{\pm 0.09}$	$0.74_{\pm 0.01}$	$0.60_{\pm 0.01}$
	MedicalNet	Full FT	ADNI	$4.38_{\pm 0.09}$	$5.54_{\pm 0.17}$	$0.64_{\pm 0.01}$	$0.40_{\pm 0.02}$
	y-Aware	Full FT	ADNI	$4.14_{\pm 0.13}$	$5.25_{\pm 0.16}$	$0.73_{\pm 0.02}$	$0.46_{\pm 0.02}$
	BrainNorm	No FT	UKB	$\mathbf{3.56_{\pm 0.13}}$	$\mathbf{4.49_{\pm 0.19}}$	$0.78_{\pm 0.03}$	$\mathbf{0.61_{\pm 0.05}}$
	BrainNorm	CN-only FT	ADNI	$3.88_{\pm 0.29}$	$4.81_{\pm 0.41}$	$\mathbf{0.80_{\pm 0.03}}$	$0.54_{\pm 0.10}$
	BrainNorm	CN-only FT	All	$3.72_{\pm 0.20}$	$4.66_{\pm 0.30}$	$0.80_{\pm 0.03}$	$0.57_{\pm 0.07}$
AIBL	3D-ViT	Full Train	AIBL	$5.24_{\pm 0.37}$	$6.44_{\pm 0.54}$	$0.19_{\pm 0.13}$	$0.01_{\pm 0.01}$
	BrainIAC	Full FT	AIBL	$5.23_{\pm 0.38}$	$6.43_{\pm 0.54}$	$0.02_{\pm 0.06}$	$0.01_{\pm 0.01}$
	BrainMVP	Full FT	AIBL	$5.23_{\pm 0.07}$	$6.44_{\pm 0.08}$	$0.31_{\pm 0.01}$	$0.01_{\pm 0.03}$
	BrainHarmony	Full FT	AIBL	$5.10_{\pm 0.38}$	$6.26_{\pm 0.56}$	$0.27_{\pm 0.10}$	$0.04_{\pm 0.02}$
	RadFM	Full FT	AIBL	$3.96_{\pm 0.44}$	$4.94_{\pm 0.48}$	$0.70_{\pm 0.08}$	$0.50_{\pm 0.11}$
	NeuroVFM	Full FT	AIBL	$5.04_{\pm 0.06}$	$6.20_{\pm 0.06}$	$0.41_{\pm 0.01}$	$0.07_{\pm 0.02}$
	Neuro-JEPA	Full FT	AIBL	$4.99_{\pm 0.13}$	$6.11_{\pm 0.16}$	$0.41_{\pm 0.07}$	$0.10_{\pm 0.05}$
	MedicalNet	Full FT	AIBL	$4.62_{\pm 0.33}$	$5.76_{\pm 0.41}$	$0.45_{\pm 0.07}$	$0.19_{\pm 0.07}$
	y-Aware	Full FT	AIBL	$4.66_{\pm 0.15}$	$5.68_{\pm 0.21}$	$0.48_{\pm 0.08}$	$0.21_{\pm 0.09}$
	BrainNorm	No FT	UKB	$3.50_{\pm 0.26}$	$4.42_{\pm 0.45}$	$0.74_{\pm 0.02}$	$0.52_{\pm 0.04}$
	BrainNorm	CN-only FT	AIBL	$\mathbf{3.39_{\pm 0.26}}$	$\mathbf{4.31_{\pm 0.45}}$	$\mathbf{0.76_{\pm 0.03}}$	$\mathbf{0.55_{\pm 0.05}}$
	BrainNorm	CN-only FT	All	$3.72_{\pm 0.20}$	$4.67_{\pm 0.35}$	$0.76_{\pm 0.02}$	$0.47_{\pm 0.05}$
MIRIAD	3D-ViT	Full Train	MIRIAD	$8.15_{\pm 4.43}$	$9.03_{\pm 4.04}$	$0.09_{\pm 0.75}$	$0.01_{\pm 4.18}$
	BrainIAC	Full FT	MIRIAD	$5.03_{\pm 1.28}$	$6.48_{\pm 1.83}$	$0.12_{\pm 0.59}$	$0.08_{\pm 0.33}$
	BrainMVP	Full FT	MIRIAD	$5.24_{\pm 0.04}$	$7.05_{\pm 0.10}$	$0.10_{\pm 0.09}$	$0.01_{\pm 0.03}$
	BrainHarmony	Full FT	MIRIAD	$5.20_{\pm 1.32}$	$6.67_{\pm 1.73}$	$0.14_{\pm 0.23}$	$0.01_{\pm 0.24}$
	RadFM	Full FT	MIRIAD	$4.49_{\pm 1.07}$	$5.71_{\pm 1.37}$	$0.70_{\pm 0.08}$	$0.40_{\pm 0.29}$
	NeuroVFM	Full FT	MIRIAD	$5.18_{\pm 0.17}$	$6.87_{\pm 0.23}$	$0.72_{\pm 0.10}$	$0.01_{\pm 0.07}$
	Neuro-JEPA	Full FT	MIRIAD	$5.32_{\pm 0.05}$	$6.92_{\pm 0.07}$	$0.13_{\pm 0.14}$	$0.02_{\pm 0.02}$
	MedicalNet	Full FT	MIRIAD	$4.74_{\pm 1.36}$	$5.84_{\pm 1.84}$	$0.23_{\pm 0.57}$	$0.13_{\pm 0.29}$
	y-Aware	Full FT	MIRIAD	$4.31_{\pm 1.06}$	$5.39_{\pm 1.15}$	$0.56_{\pm 0.33}$	$0.20_{\pm 0.43}$
	BrainNorm	No FT	UKB	$4.83_{\pm 1.30}$	$5.31_{\pm 1.42}$	$\mathbf{0.82_{\pm 0.17}}$	$0.29_{\pm 0.40}$
	BrainNorm	CN-only FT	MIRIAD	$3.46_{\pm 1.56}$	$3.89_{\pm 1.61}$	$0.80_{\pm 0.19}$	$0.52_{\pm 0.30}$
	BrainNorm	CN-only FT	All	$\mathbf{3.23_{\pm 1.18}}$	$\mathbf{3.78_{\pm 1.27}}$	$0.78_{\pm 0.24}$	$\mathbf{0.52_{\pm 0.28}}$
NIFD	3D-ViT	Full Train	NIFD	$6.53_{\pm 0.60}$	$8.08_{\pm 0.98}$	$0.14_{\pm 0.13}$	$0.01_{\pm 0.04}$
	BrainIAC	Full FT	NIFD	$5.97_{\pm 0.72}$	$7.71_{\pm 0.87}$	$0.16_{\pm 0.13}$	$0.01_{\pm 0.02}$
	BrainMVP	Full FT	NIFD	$6.12_{\pm 0.05}$	$7.77_{\pm 0.01}$	$0.05_{\pm 0.03}$	$0.01_{\pm 0.02}$
	BrainHarmony	Full FT	NIFD	$5.97_{\pm 0.71}$	$7.72_{\pm 0.89}$	$0.25_{\pm 0.15}$	$0.01_{\pm 0.01}$
	RadFM	Full FT	NIFD	$5.83_{\pm 0.77}$	$6.95_{\pm 0.83}$	$0.56_{\pm 0.18}$	$0.18_{\pm 0.18}$
	NeuroVFM	Full FT	NIFD	$5.98_{\pm 0.02}$	$7.71_{\pm 0.07}$	$0.56_{\pm 0.02}$	$0.01_{\pm 0.02}$
	Neuro-JEPA	Full FT	NIFD	$6.11_{\pm 0.06}$	$7.82_{\pm 0.07}$	$0.03_{\pm 0.07}$	$0.02_{\pm 0.02}$
	MedicalNet	Full FT	NIFD	$6.01_{\pm 0.64}$	$7.76_{\pm 0.87}$	$0.15_{\pm 0.09}$	$0.01_{\pm 0.01}$
	y-Aware	Full FT	NIFD	$5.88_{\pm 0.75}$	$7.67_{\pm 0.91}$	$0.22_{\pm 0.09}$	$0.02_{\pm 0.02}$
	BrainNorm	No FT	UKB	$5.44_{\pm 0.42}$	$6.75_{\pm 0.62}$	$0.76_{\pm 0.08}$	$0.20_{\pm 0.25}$
	BrainNorm	CN-only FT	NIFD	$\mathbf{4.33_{\pm 0.40}}$	$5.51_{\pm 0.56}$	$\mathbf{0.77_{\pm 0.08}}$	$0.47_{\pm 0.15}$
	BrainNorm	CN-only FT	All	$4.35_{\pm 0.62}$	$\mathbf{5.44_{\pm 0.67}}$	$0.75_{\pm 0.08}$	$\mathbf{0.49_{\pm 0.14}}$
PPMI	3D-ViT	Full Train	PPMI	$9.33_{\pm 0.91}$	$11.93_{\pm 1.23}$	$0.11_{\pm 0.24}$	$0.01_{\pm 0.04}$
	BrainIAC	Full FT	PPMI	$9.39_{\pm 0.68}$	$11.92_{\pm 0.81}$	$0.14_{\pm 0.07}$	$0.00_{\pm 0.01}$
	BrainMVP	Full FT	PPMI	$9.34_{\pm 0.06}$	$11.63_{\pm 0.04}$	$0.10_{\pm 0.01}$	$0.01_{\pm 0.04}$
	BrainHarmony	Full FT	PPMI	$9.18_{\pm 0.75}$	$11.63_{\pm 0.99}$	$0.26_{\pm 0.20}$	$0.03_{\pm 0.01}$
	RadFM	Full FT	PPMI	$6.53_{\pm 0.59}$	$9.94_{\pm 0.87}$	$0.76_{\pm 0.03}$	$0.43_{\pm 0.11}$
	NeuroVFM	Full FT	PPMI	$9.23_{\pm 0.24}$	$11.72_{\pm 0.46}$	$0.48_{\pm 0.07}$	$0.02_{\pm 0.08}$
	Neuro-JEPA	Full FT	PPMI	$9.26_{\pm 0.13}$	$11.60_{\pm 0.20}$	$0.13_{\pm 0.13}$	$0.02_{\pm 0.03}$
	MedicalNet	Full FT	PPMI	$9.05_{\pm 0.84}$	$11.55_{\pm 1.09}$	$0.18_{\pm 0.08}$	$0.01_{\pm 0.02}$
	y-Aware	Full FT	PPMI	$8.54_{\pm 0.88}$	$11.00_{\pm 1.33}$	$0.32_{\pm 0.14}$	$0.13_{\pm 0.13}$
	BrainNorm	No FT	UKB	$7.42_{\pm 0.91}$	$9.68_{\pm 1.07}$	$0.83_{\pm 0.03}$	$0.29_{\pm 0.07}$
	BrainNorm	CN-only FT	PPMI	$6.33_{\pm 0.77}$	$8.64_{\pm 1.07}$	$\mathbf{0.84_{\pm 0.03}}$	$0.44_{\pm 0.06}$
	BrainNorm	CN-only FT	All	$\mathbf{5.44_{\pm 0.55}}$	$\mathbf{7.56_{\pm 0.94}}$	$0.83_{\pm 0.03}$	$\mathbf{0.57_{\pm 0.05}}$

Table A11: Zero-shot disease classification performance (AUC in %) across binary and multi-class tasks using top-k@20. Results are reported as mean$_{\pm \mathrm{std}}$. FT (own): finetuning on the corresponding dataset CN only. FT (all CN): finetuning on pooled CN data from ADNI+AIBL+MIRIAD+NIFD+PPMI. A: Anomaly. P+A: Disease Prior restricted Anomaly. BAG: Brain Age Gap

Dataset/Task	Method	Full FT	No FT	FT (own)	FT (all CN)
ADNI [AD/CN]	**A**	-	$54.78_{\pm 4.04}$	$72.80_{\pm 4.39}$	$72.37_{\pm 5.07}$
	P+A	-	$64.61_{\pm 3.00}$	$79.27_{\pm 3.94}$	$79.16_{\pm 3.79}$
	BAG	-	$82.80_{\pm 1.95}$	$86.19_{\pm 2.62}$	$86.02_{\pm 2.12}$
	BrainIAC+BAG	$46.68_{\pm 3.97}$	-	-	-
	BrainMVP+BAG	$57.58_{\pm 2.20}$	-	-	-
	BrainHarmony+BAG	$66.65_{\pm 3.78}$	-	-	-
	RadFM+BAG	$72.93_{\pm 1.50}$	-	-	-
	NeuroVFM+BAG	$65.00_{\pm 1.00}$	-	-	-
	Neuro-JEPA+BAG	$70.42_{\pm 0.75}$	-	-	-
	MedicalNet+BAG	$65.75_{\pm 4.39}$	-	-	-
	y-aware+BAG	$70.83_{\pm 3.69}$	-	-	-
ADNI [sMCI/pMCI]	**A**	-	$47.33_{\pm 4.11}$	$62.49_{\pm 6.14}$	$62.79_{\pm 6.55}$
	P+A	-	$55.28_{\pm 3.04}$	$69.44_{\pm 2.62}$	$69.01_{\pm 3.37}$
	BAG	-	$73.59_{\pm 5.01}$	$74.98_{\pm 4.64}$	$74.58_{\pm 5.03}$
	BrainIAC+BAG	$46.85_{\pm 1.63}$	-	-	-
	BrainMVP+BAG	$53.62_{\pm 1.50}$	-	-	-
	BrainHarmony+BAG	$57.18_{\pm 8.79}$	-	-	-
	RadFM+BAG	$63.71_{\pm 0.77}$	-	-	-
	NeuroVFM+BAG	$58.61_{\pm 0.53}$	-	-	-
	Neuro-JEPA+BAG	$61.60_{\pm 0.36}$	-	-	-
	MedicalNet+BAG	$58.86_{\pm 3.99}$	-	-	-
	y-aware+BAG	$61.60_{\pm 5.63}$	-	-	-
AIBL [AD/CN]	**A**	-	$49.55_{\pm 5.98}$	$68.09_{\pm 3.42}$	$70.70_{\pm 2.23}$
	P+A	-	$61.24_{\pm 6.10}$	$79.83_{\pm 1.97}$	$77.60_{\pm 3.60}$
	BAG	-	$87.14_{\pm 1.19}$	$90.56_{\pm 1.89}$	$89.97_{\pm 2.54}$
	BrainIAC+BAG	$42.72_{\pm 5.34}$	-	-	-
	BrainMVP+BAG	$43.38_{\pm 0.17}$	-	-	-
	BrainHarmony+BAG	$45.74_{\pm 6.25}$	-	-	-
	RadFM+BAG	$59.91_{\pm 4.23}$	-	-	-
	NeuroVFM+BAG	$45.00_{\pm 1.00}$	-	-	-
	Neuro-JEPA+BAG	$46.03_{\pm 1.31}$	-	-	-
	MedicalNet+BAG	$53.10_{\pm 8.24}$	-	-	-
	y-aware+BAG	$53.50_{\pm 6.05}$	-	-	-
MIRIAD [AD/CN]	**A**	-	$71.00_{\pm 8.74}$	$80.58_{\pm 8.03}$	$93.51_{\pm 7.58}$
	P+A	-	$81.89_{\pm 7.23}$	$85.91_{\pm 7.87}$	$96.16_{\pm 3.59}$
	BAG	-	$98.52_{\pm 2.47}$	$98.89_{\pm 2.32}$	$98.56_{\pm 3.22}$
	BrainIAC+BAG	$52.12_{\pm 8.61}$	-	-	-
	BrainMVP+BAG	$50.74_{\pm 0.03}$	-	-	-
	BrainHarmony+BAG	$54.32_{\pm 6.76}$	-	-	-
	RadFM+BAG	$58.81_{\pm 5.63}$	-	-	-
	NeuroVFM+BAG	$52.00_{\pm 1.00}$	-	-	-
	Neuro-JEPA+BAG	$51.36_{\pm 0.45}$	-	-	-
	MedicalNet+BAG	$47.22_{\pm 8.27}$	-	-	-
	y-aware+BAG	$73.37_{\pm 8.64}$	-	-	-
NIFD [FTD/CN]	**A**	-	$72.69_{\pm 5.95}$	$79.89_{\pm 7.02}$	$84.09_{\pm 8.06}$
	P+A	-	$75.46_{\pm 6.65}$	$78.91_{\pm 6.13}$	$83.05_{\pm 7.68}$
	BAG	-	$82.38_{\pm 4.13}$	$82.80_{\pm 4.61}$	$85.18_{\pm 3.07}$
	BrainIAC+BAG	$50.52_{\pm 2.15}$	-	-	-
	BrainMVP+BAG	$51.08_{\pm 0.03}$	-	-	-
	BrainHarmony+BAG	$51.68_{\pm 2.54}$	-	-	-
	RadFM+BAG	$54.78_{\pm 3.31}$	-	-	-
	NeuroVFM+BAG	$52.00_{\pm 0.00}$	-	-	-
	Neuro-JEPA+BAG	$51.34_{\pm 0.05}$	-	-	-
	MedicalNet+BAG	$52.10_{\pm 3.38}$	-	-	-
	y-aware+BAG	$55.73_{\pm 3.31}$	-	-	-
PPMI [PD/CN]	**A**	-	$49.60_{\pm 2.48}$	$50.26_{\pm 6.23}$	$53.02_{\pm 7.96}$
	P+A	-	$51.99_{\pm 4.53}$	$51.65_{\pm 7.30}$	$53.12_{\pm 5.18}$
	BAG	-	$55.03_{\pm 4.57}$	$55.59_{\pm 5.33}$	$56.10_{\pm 6.14}$
	BrainIAC+BAG	$44.90_{\pm 4.30}$	-	-	-
	BrainMVP+BAG	$45.42_{\pm 0.01}$	-	-	-
	BrainHarmony+BAG	$45.69_{\pm 3.95}$	-	-	-
	RadFM+BAG	$49.53_{\pm 1.78}$	-	-	-
	NeuroVFM+BAG	$46.00_{\pm 0.00}$	-	-	-
	Neuro-JEPA+BAG	$45.49_{\pm 0.25}$	-	-	-
	MedicalNet+BAG	$44.96_{\pm 3.66}$	-	-	-
	y-aware+BAG	$46.54_{\pm 3.80}$	-	-	-
ADNI+NIFD+PPMI [AD/FTD/PD]	**P+A**	-	$61.31_{\pm 1.50}$	$67.20_{\pm 4.12}$	$66.44_{\pm 3.59}$
ADNI+NIFD+PPMI [AD/FTD/PD/CN]	**P+A**	-	$63.30_{\pm 1.01}$	$68.12_{\pm 2.14}$	$68.02_{\pm 1.83}$

Table A12: Full-training-set performance (%) across eight classification tasks. Baselines use end-to-end full finetuning (Full FT). BrainNorm uses linear probing (LP) without finetuning or after CN-only finetuning on the target cohort (own) or all cohorts (all). Values are $\mathrm{mean}_{\pm\mathrm{std}}$. For each task and metric, the best result across all models and methods is in **bold**, and the second-best is underlined.

Model	Method	ADNI AD/CN		ADNI sMCI/pMCI		AIBL AD/CN		MIRIAD AD/CN	
		BACC	AUC	BACC	AUC	BACC	AUC	BACC	AUC
3D-ViT	Full FT	$78.06_{\pm 3.56}$	$83.63_{\pm 2.61}$	$60.82_{\pm 2.80}$	$64.70_{\pm 4.93}$	$63.26_{\pm 2.58}$	$78.26_{\pm 3.00}$	$82.29_{\pm 4.84}$	$84.52_{\pm 4.29}$
BrainHarmony	Full FT	$66.51_{\pm 7.98}$	$78.78_{\pm 5.70}$	$55.10_{\pm 3.05}$	$70.42_{\pm 2.22}$	$53.79_{\pm 1.15}$	$72.36_{\pm 4.57}$	$79.82_{\pm 9.19}$	$84.38_{\pm 8.49}$
BrainMVP	Full FT	$71.94_{\pm 2.15}$	$83.11_{\pm 4.65}$	$51.79_{\pm 3.61}$	$70.59_{\pm 3.28}$	$53.00_{\pm 4.91}$	$72.89_{\pm 4.64}$	$73.73_{\pm 5.49}$	$91.49_{\pm 4.98}$
BrainIAC	Full FT	$66.70_{\pm 2.56}$	$74.66_{\pm 3.70}$	$56.56_{\pm 1.87}$	$62.17_{\pm 1.20}$	$54.69_{\pm 3.79}$	$69.89_{\pm 4.94}$	$83.85_{\pm 5.85}$	$91.27_{\pm 5.21}$
NeuroVFM	Full FT	$81.70_{\pm 2.28}$	$90.40_{\pm 1.43}$	$60.25_{\pm 4.40}$	$72.27_{\pm 3.58}$	$55.82_{\pm 8.01}$	$80.56_{\pm 7.19}$	$74.60_{\pm 8.29}$	$93.11_{\pm 6.06}$
Neuro-JEPA	Full FT	$85.66_{\pm 1.48}$	$92.71_{\pm 1.49}$	$63.59_{\pm 5.96}$	$67.91_{\pm 7.16}$	$83.43_{\pm 6.99}$	$90.01_{\pm 3.75}$	$85.87_{\pm 9.85}$	$93.45_{\pm 7.21}$
RadFM	Full FT	$79.33_{\pm 6.58}$	$90.77_{\pm 2.14}$	$58.03_{\pm 5.52}$	$71.00_{\pm 2.57}$	$69.54_{\pm 7.41}$	$87.36_{\pm 4.42}$	$80.71_{\pm 9.82}$	$92.41_{\pm 9.80}$
MedicalNet	Full FT	$81.79_{\pm 2.80}$	$88.79_{\pm 2.76}$	$67.88_{\pm 3.75}$	$72.80_{\pm 4.61}$	$60.62_{\pm 5.55}$	$73.68_{\pm 3.79}$	$88.37_{\pm 5.37}$	$92.22_{\pm 7.57}$
y-Aware	Full FT	$85.27_{\pm 1.74}$	$93.07_{\pm 1.57}$	$63.63_{\pm 7.94}$	$78.41_{\pm 3.10}$	$67.67_{\pm 9.36}$	$83.79_{\pm 8.13}$	$87.81_{\pm 5.65}$	$96.74_{\pm 2.15}$
BrainNorm_M1	No FT+LP	$87.07_{\pm 1.61}$	$93.35_{\pm 2.50}$	$69.88_{\pm 1.92}$	$77.90_{\pm 2.48}$	$82.38_{\pm 7.91}$	$92.82_{\pm 0.91}$	$87.92_{\pm 9.24}$	$95.63_{\pm 5.10}$
	CN-only FT (own)+LP	$86.31_{\pm 1.57}$	$93.42_{\pm 2.18}$	$71.88_{\pm 1.60}$	$78.84_{\pm 2.39}$	$85.10_{\pm 4.30}$	$92.48_{\pm 0.70}$	$87.87_{\pm 8.34}$	$95.63_{\pm 5.22}$
	CN-only FT (all)+LP	$86.97_{\pm 1.93}$	$93.65_{\pm 2.20}$	$72.40_{\pm 2.00}$	$79.29_{\pm 2.15}$	$85.34_{\pm 4.36}$	$92.92_{\pm 0.68}$	$86.58_{\pm 7.73}$	$95.83_{\pm 5.74}$
BrainNorm_M2	No FT+LP	$86.85_{\pm 2.87}$	$94.27_{\pm 2.22}$	$71.96_{\pm 4.10}$	$80.75_{\pm 3.80}$	$85.91_{\pm 5.47}$	$93.39_{\pm 1.71}$	$92.72_{\pm 5.00}$	$\mathbf{98.26_{\pm 3.88}}$
	CN-only FT (own)+LP	$\mathbf{87.67_{\pm 3.13}}$	$94.62_{\pm 2.12}$	$73.59_{\pm 2.82}$	$81.11_{\pm 2.99}$	$85.82_{\pm 3.24}$	$93.62_{\pm 1.66}$	$92.60_{\pm 5.09}$	$98.13_{\pm 4.10}$
	CN-only FT (all)+LP	$87.34_{\pm 2.78}$	$\mathbf{94.66_{\pm 1.95}}$	$\mathbf{74.43_{\pm 3.03}}$	$\mathbf{81.76_{\pm 3.36}}$	$\mathbf{87.01_{\pm 4.80}}$	$\mathbf{93.64_{\pm 1.30}}$	$\mathbf{92.97_{\pm 8.97}}$	$97.92_{\pm 4.63}$
BrainNorm_M3	No FT+LP	$76.48_{\pm 2.61}$	$83.81_{\pm 2.42}$	$64.53_{\pm 2.83}$	$70.41_{\pm 3.42}$	$56.66_{\pm 6.26}$	$58.70_{\pm 6.46}$	$86.49_{\pm 6.65}$	$90.39_{\pm 6.27}$
	CN-only FT (own)+LP	$78.81_{\pm 2.05}$	$86.19_{\pm 1.77}$	$69.61_{\pm 2.76}$	$76.21_{\pm 4.11}$	$58.23_{\pm 4.86}$	$62.61_{\pm 3.59}$	$84.63_{\pm 8.87}$	$92.93_{\pm 8.51}$
	CN-only FT (all)+LP	$79.27_{\pm 2.19}$	$86.71_{\pm 1.31}$	$68.68_{\pm 2.46}$	$75.34_{\pm 4.65}$	$62.39_{\pm 7.90}$	$66.74_{\pm 7.25}$	$90.32_{\pm 8.63}$	$97.15_{\pm 4.73}$

Model	Method	NIFD FTD/CN		PPMI PD/CN		ADNI+NIFD+PPMI AD/FTD/PD		ADNI+NIFD+PPMI AD/FTD/PD/CN	
		BACC	AUC	BACC	AUC	BACC	AUC	BACC	AUC
3D-ViT	Full FT	$52.96_{\pm 1.98}$	$60.68_{\pm 7.12}$	$50.05_{\pm 0.27}$	$54.75_{\pm 3.05}$	$58.86_{\pm 1.65}$	$79.72_{\pm 1.08}$	$46.68_{\pm 5.41}$	$76.54_{\pm 3.04}$
BrainHarmony	Full FT	$52.59_{\pm 3.56}$	$69.42_{\pm 3.36}$	$50.20_{\pm 0.96}$	$55.90_{\pm 2.76}$	$74.49_{\pm 3.83}$	$90.57_{\pm 1.85}$	$49.07_{\pm 3.65}$	$81.73_{\pm 2.24}$
BrainMVP	Full FT	$56.60_{\pm 4.34}$	$82.26_{\pm 6.45}$	$51.00_{\pm 0.91}$	$59.41_{\pm 2.75}$	$82.79_{\pm 2.63}$	$96.78_{\pm 0.92}$	$49.06_{\pm 2.00}$	$79.66_{\pm 1.15}$
BrainIAC	Full FT	$53.86_{\pm 5.68}$	$66.12_{\pm 5.12}$	$50.50_{\pm 0.71}$	$57.38_{\pm 4.45}$	$55.65_{\pm 1.50}$	$79.17_{\pm 1.44}$	$53.36_{\pm 2.00}$	$72.24_{\pm 1.15}$
NeuroVFM	Full FT	$61.22_{\pm 5.53}$	$82.03_{\pm 5.88}$	$50.00_{\pm 0.00}$	$59.64_{\pm 2.99}$	$62.26_{\pm 7.80}$	$86.97_{\pm 3.55}$	$49.68_{\pm 8.34}$	$83.80_{\pm 2.31}$
Neuro-JEPA	Full FT	$80.30_{\pm 2.29}$	$86.88_{\pm 3.83}$	$54.51_{\pm 6.49}$	$56.70_{\pm 7.19}$	$81.83_{\pm 2.63}$	$94.87_{\pm 0.96}$	$68.53_{\pm 3.70}$	$88.56_{\pm 1.45}$
RadFM	Full FT	$70.08_{\pm 6.41}$	$81.09_{\pm 5.85}$	$50.27_{\pm 0.61}$	$61.18_{\pm 4.06}$	$75.39_{\pm 3.20}$	$92.00_{\pm 0.80}$	$57.44_{\pm 3.06}$	$84.61_{\pm 1.44}$
MedicalNet	Full FT	$78.75_{\pm 3.52}$	$87.69_{\pm 2.76}$	$50.67_{\pm 0.60}$	$59.51_{\pm 4.40}$	$82.06_{\pm 2.71}$	$94.64_{\pm 1.42}$	$59.51_{\pm 3.33}$	$84.87_{\pm 2.23}$
y-Aware	Full FT	$77.92_{\pm 6.58}$	$89.14_{\pm 4.01}$	$50.10_{\pm 0.70}$	$57.28_{\pm 2.53}$	$83.75_{\pm 1.52}$	$96.70_{\pm 0.82}$	$69.87_{\pm 1.96}$	$90.81_{\pm 1.44}$
BrainNorm_M1	No FT+LP	$80.87_{\pm 3.79}$	$88.55_{\pm 4.49}$	$\mathbf{60.43_{\pm 2.59}}$	$66.01_{\pm 5.30}$	$83.21_{\pm 4.27}$	$96.91_{\pm 0.88}$	$67.10_{\pm 2.52}$	$90.73_{\pm 0.93}$
	CN-only FT (own)+LP	$81.49_{\pm 5.36}$	$88.71_{\pm 4.60}$	$55.36_{\pm 4.42}$	$61.20_{\pm 4.92}$	$82.90_{\pm 3.75}$	$96.89_{\pm 0.72}$	$67.74_{\pm 1.97}$	$90.82_{\pm 1.32}$
	CN-only FT (all)+LP	$79.45_{\pm 2.44}$	$88.42_{\pm 3.07}$	$56.12_{\pm 5.79}$	$61.27_{\pm 5.74}$	$83.33_{\pm 3.73}$	$96.85_{\pm 0.63}$	$67.55_{\pm 2.12}$	$91.01_{\pm 1.22}$
BrainNorm_M2	No FT+LP	$81.82_{\pm 6.92}$	$89.15_{\pm 4.29}$	$59.08_{\pm 3.38}$	$\mathbf{66.12_{\pm 5.23}}$	$83.98_{\pm 4.48}$	$\mathbf{97.01_{\pm 0.99}}$	$\mathbf{70.32_{\pm 1.77}}$	$\mathbf{91.91_{\pm 1.24}}$
	CN-only FT (own)+LP	$\mathbf{82.62_{\pm 7.61}}$	$\mathbf{89.34_{\pm 4.77}}$	$57.73_{\pm 1.73}$	$62.03_{\pm 5.16}$	$84.22_{\pm 3.94}$	$96.84_{\pm 0.95}$	$70.23_{\pm 2.31}$	$91.78_{\pm 1.63}$
	CN-only FT (all)+LP	$78.45_{\pm 2.86}$	$88.22_{\pm 4.43}$	$58.00_{\pm 4.14}$	$61.65_{\pm 6.53}$	$\mathbf{84.24_{\pm 3.84}}$	$97.00_{\pm 0.83}$	$70.09_{\pm 1.83}$	$91.82_{\pm 1.63}$
BrainNorm_M3	No FT+LP	$77.64_{\pm 2.85}$	$82.97_{\pm 6.26}$	$52.40_{\pm 2.47}$	$48.49_{\pm 4.08}$	$66.85_{\pm 2.82}$	$86.86_{\pm 1.77}$	$51.31_{\pm 3.44}$	$83.11_{\pm 1.11}$
	CN-only FT (own)+LP	$72.04_{\pm 3.47}$	$82.26_{\pm 4.20}$	$50.19_{\pm 2.90}$	$49.62_{\pm 5.17}$	$66.29_{\pm 4.13}$	$84.84_{\pm 2.59}$	$44.93_{\pm 2.10}$	$79.08_{\pm 1.90}$
	CN-only FT (all)+LP	$75.45_{\pm 4.54}$	$81.40_{\pm 5.59}$	$51.59_{\pm 4.72}$	$46.92_{\pm 5.33}$	$64.87_{\pm 3.63}$	$83.87_{\pm 2.94}$	$44.40_{\pm 2.45}$	$79.76_{\pm 1.98}$

Table A13: Few-shot performance across binary tasks and datasets for BrainNorm variants. Each method is shown under three modes: No FT, CN-only FT on the corresponding dataset only, and CN-only FT on pooled CN data from ADNI+AIBL+MIRIAD+PPMI+NIFD. Values are reported as mean $\pm$ standard deviation.

Method	Mode	$k = 1$	$k = 2$	$k = 4$	$k = 8$	$k = 16$
ADNI — [AD/CN]						
BrainNorm_M1	No FT + LP	$55.31_{\pm 5.18}$	$57.69_{\pm 5.01}$	$62.54_{\pm 5.82}$	$75.21_{\pm 7.97}$	$80.60_{\pm 5.06}$
	CN-only FT(own) + LP	$68.86_{\pm 3.78}$	$67.60_{\pm 5.95}$	$59.59_{\pm 7.55}$	$80.23_{\pm 4.77}$	$80.84_{\pm 2.49}$
	CN-only FT (all) + LP	$69.22_{\pm 3.65}$	$69.47_{\pm 5.01}$	$59.60_{\pm 4.49}$	$80.33_{\pm 4.61}$	$81.09_{\pm 2.21}$
BrainNorm_M2	No FT + LP	$55.30_{\pm 3.75}$	$50.70_{\pm 4.70}$	$70.57_{\pm 7.48}$	$78.28_{\pm 6.23}$	$83.16_{\pm 4.02}$
	CN-only FT(own) + LP	$79.51_{\pm 5.37}$	$69.28_{\pm 2.02}$	$77.13_{\pm 3.10}$	$84.95_{\pm 5.88}$	$86.08_{\pm 3.34}$
	CN-only FT (all) + LP	$79.09_{\pm 5.48}$	$71.64_{\pm 6.71}$	$78.14_{\pm 5.73}$	$85.18_{\pm 5.24}$	$86.20_{\pm 3.85}$
BrainNorm_M3	No FT + LP	$58.12_{\pm 5.38}$	$55.21_{\pm 5.64}$	$52.04_{\pm 7.94}$	$54.60_{\pm 6.48}$	$62.73_{\pm 8.69}$
	CN-only FT(own) + LP	$69.24_{\pm 5.52}$	$64.68_{\pm 5.49}$	$64.81_{\pm 6.61}$	$51.97_{\pm 5.69}$	$61.45_{\pm 9.33}$
	CN-only FT (all) + LP	$67.74_{\pm 3.82}$	$61.49_{\pm 5.65}$	$61.68_{\pm 6.62}$	$43.83_{\pm 4.94}$	$64.79_{\pm 3.60}$
AIBL — [AD/CN]						
BrainNorm_M1	No FT + LP	$84.06_{\pm 4.36}$	$80.07_{\pm 8.27}$	$82.50_{\pm 5.89}$	$84.36_{\pm 5.29}$	$85.55_{\pm 4.04}$
	CN-only FT(own) + LP	$85.46_{\pm 3.15}$	$78.31_{\pm 7.09}$	$85.17_{\pm 3.81}$	$85.49_{\pm 3.82}$	$86.21_{\pm 3.09}$
	CN-only FT (all) + LP	$85.83_{\pm 4.26}$	$84.11_{\pm 5.60}$	$86.68_{\pm 4.23}$	$85.61_{\pm 4.43}$	$85.97_{\pm 4.06}$
BrainNorm_M2	No FT + LP	$79.87_{\pm 2.29}$	$78.24_{\pm 8.15}$	$81.14_{\pm 7.51}$	$85.26_{\pm 6.41}$	$85.53_{\pm 5.29}$
	CN-only FT(own) + LP	$85.53_{\pm 3.17}$	$82.40_{\pm 7.13}$	$83.63_{\pm 7.06}$	$87.91_{\pm 5.26}$	$88.19_{\pm 3.81}$
	CN-only FT (all) + LP	$88.23_{\pm 2.36}$	$85.15_{\pm 4.84}$	$85.89_{\pm 7.04}$	$89.24_{\pm 3.46}$	$88.75_{\pm 3.40}$
BrainNorm_M3	No FT + LP	$55.27_{\pm 2.65}$	$56.55_{\pm 7.61}$	$54.55_{\pm 5.61}$	$51.73_{\pm 7.23}$	$58.54_{\pm 9.07}$
	CN-only FT(own) + LP	$67.42_{\pm 4.35}$	$49.57_{\pm 9.01}$	$54.73_{\pm 9.72}$	$47.20_{\pm 7.82}$	$67.12_{\pm 8.45}$
	CN-only FT (all) + LP	$65.50_{\pm 7.56}$	$47.37_{\pm 9.43}$	$52.35_{\pm 9.79}$	$55.75_{\pm 9.26}$	$61.38_{\pm 8.51}$
MIRIAD — [AD/CN]						
BrainNorm_M1	No FT + LP	$90.07_{\pm 8.61}$	$91.11_{\pm 9.15}$	$94.79_{\pm 5.81}$	$94.67_{\pm 5.52}$	$94.41_{\pm 5.98}$
	CN-only FT(own) + LP	$88.08_{\pm 9.44}$	$88.11_{\pm 9.80}$	$94.40_{\pm 5.92}$	$94.62_{\pm 5.77}$	$94.42_{\pm 6.14}$
	CN-only FT (all) + LP	$81.37_{\pm 9.26}$	$82.70_{\pm 9.34}$	$94.77_{\pm 6.48}$	$94.83_{\pm 6.45}$	$94.74_{\pm 6.53}$
BrainNorm_M2	No FT + LP	$83.50_{\pm 7.97}$	$89.89_{\pm 9.89}$	$93.39_{\pm 6.16}$	$93.81_{\pm 4.24}$	$97.19_{\pm 4.64}$
	CN-only FT(own) + LP	$88.97_{\pm 7.89}$	$77.30_{\pm 9.43}$	$95.30_{\pm 3.90}$	$94.24_{\pm 5.05}$	$96.91_{\pm 5.62}$
	CN-only FT (all) + LP	$92.50_{\pm 6.31}$	$92.10_{\pm 9.14}$	$97.17_{\pm 3.87}$	$96.69_{\pm 4.35}$	$97.66_{\pm 4.64}$
BrainNorm_M3	No FT + LP	$73.05_{\pm 9.60}$	$60.20_{\pm 9.30}$	$69.40_{\pm 7.21}$	$51.42_{\pm 9.09}$	$76.14_{\pm 8.85}$
	CN-only FT(own) + LP	$77.72_{\pm 7.91}$	$70.01_{\pm 7.57}$	$74.96_{\pm 8.04}$	$58.39_{\pm 9.93}$	$72.24_{\pm 9.21}$
	CN-only FT (all) + LP	$83.78_{\pm 7.90}$	$64.72_{\pm 9.94}$	$77.91_{\pm 7.68}$	$70.99_{\pm 9.19}$	$84.47_{\pm 9.17}$
NIFD — [FTD/CN]						
BrainNorm_M1	No FT + LP	$75.92_{\pm 7.00}$	$77.55_{\pm 3.57}$	$77.85_{\pm 3.42}$	$78.86_{\pm 3.28}$	$79.76_{\pm 3.13}$
	CN-only FT(own) + LP	$75.77_{\pm 6.77}$	$77.36_{\pm 3.82}$	$77.79_{\pm 3.40}$	$78.79_{\pm 3.08}$	$78.78_{\pm 3.78}$
	CN-only FT (all) + LP	$77.15_{\pm 6.71}$	$79.56_{\pm 3.15}$	$80.12_{\pm 2.94}$	$80.54_{\pm 3.07}$	$80.44_{\pm 3.43}$
BrainNorm_M2	No FT + LP	$76.13_{\pm 7.36}$	$68.57_{\pm 9.56}$	$80.34_{\pm 6.08}$	$81.50_{\pm 6.52}$	$81.25_{\pm 5.35}$
	CN-only FT(own) + LP	$78.08_{\pm 8.74}$	$67.56_{\pm 8.15}$	$81.31_{\pm 6.33}$	$82.17_{\pm 6.67}$	$81.50_{\pm 5.79}$
	CN-only FT (all) + LP	$82.31_{\pm 6.36}$	$69.58_{\pm 9.81}$	$83.51_{\pm 4.47}$	$84.22_{\pm 5.04}$	$84.03_{\pm 4.44}$
BrainNorm_M3	No FT + LP	$57.96_{\pm 7.49}$	$43.29_{\pm 8.14}$	$40.93_{\pm 7.22}$	$67.92_{\pm 4.26}$	$63.60_{\pm 9.48}$
	CN-only FT(own) + LP	$59.37_{\pm 7.62}$	$44.14_{\pm 8.86}$	$53.96_{\pm 8.69}$	$54.86_{\pm 8.47}$	$59.08_{\pm 6.90}$
	CN-only FT (all) + LP	$52.81_{\pm 9.17}$	$39.87_{\pm 9.47}$	$43.67_{\pm 9.06}$	$60.71_{\pm 9.85}$	$62.56_{\pm 9.62}$
PPMI — [PD/CN]						
BrainNorm_M1	No FT + LP	$58.39_{\pm 5.47}$	$57.75_{\pm 5.56}$	$58.55_{\pm 5.24}$	$57.85_{\pm 5.31}$	$57.60_{\pm 4.83}$
	CN-only FT(own) + LP	$57.54_{\pm 4.87}$	$56.23_{\pm 5.61}$	$58.12_{\pm 5.11}$	$57.48_{\pm 5.47}$	$57.47_{\pm 5.15}$
	CN-only FT (all) + LP	$58.51_{\pm 5.04}$	$56.56_{\pm 5.78}$	$58.19_{\pm 5.25}$	$58.09_{\pm 5.41}$	$58.03_{\pm 5.62}$
BrainNorm_M2	No FT + LP	$52.79_{\pm 2.76}$	$54.52_{\pm 4.30}$	$57.71_{\pm 6.21}$	$53.23_{\pm 5.58}$	$55.02_{\pm 2.60}$
	CN-only FT(own) + LP	$52.05_{\pm 2.87}$	$52.71_{\pm 5.56}$	$56.76_{\pm 6.38}$	$53.18_{\pm 5.25}$	$55.33_{\pm 3.59}$
	CN-only FT (all) + LP	$53.85_{\pm 3.35}$	$54.85_{\pm 6.76}$	$59.25_{\pm 6.55}$	$55.57_{\pm 3.42}$	$57.16_{\pm 4.66}$
BrainNorm_M3	No FT + LP	$50.32_{\pm 2.39}$	$49.96_{\pm 3.37}$	$49.39_{\pm 3.19}$	$51.02_{\pm 4.09}$	$50.59_{\pm 4.21}$
	CN-only FT(own) + LP	$51.26_{\pm 2.31}$	$49.83_{\pm 3.60}$	$49.70_{\pm 3.43}$	$49.80_{\pm 3.53}$	$50.98_{\pm 1.78}$
	CN-only FT (all) + LP	$48.73_{\pm 3.39}$	$46.82_{\pm 4.09}$	$47.94_{\pm 4.54}$	$46.90_{\pm 4.77}$	$48.56_{\pm 2.95}$
ADNI — [sMCI/pMCI]						
BrainNorm_M1	No FT + LP	$47.51_{\pm 8.51}$	$54.44_{\pm 8.45}$	$54.86_{\pm 7.34}$	$55.02_{\pm 7.63}$	$63.71_{\pm 5.73}$
	CN-only FT(own) + LP	$45.83_{\pm 8.04}$	$54.48_{\pm 9.32}$	$53.24_{\pm 9.68}$	$55.27_{\pm 7.51}$	$65.14_{\pm 5.47}$
	CN-only FT (all) + LP	$45.81_{\pm 9.68}$	$53.77_{\pm 9.69}$	$53.10_{\pm 9.34}$	$54.49_{\pm 8.52}$	$66.92_{\pm 6.68}$
BrainNorm_M2	No FT + LP	$48.93_{\pm 8.31}$	$55.41_{\pm 4.76}$	$61.50_{\pm 6.58}$	$61.44_{\pm 6.34}$	$62.36_{\pm 6.41}$
	CN-only FT(own) + LP	$48.85_{\pm 8.94}$	$59.82_{\pm 9.02}$	$67.08_{\pm 6.04}$	$65.56_{\pm 7.99}$	$69.30_{\pm 4.74}$
	CN-only FT (all) + LP	$48.59_{\pm 8.33}$	$59.09_{\pm 6.11}$	$67.03_{\pm 7.30}$	$67.97_{\pm 8.38}$	$70.60_{\pm 6.67}$
BrainNorm_M3	No FT + LP	$50.52_{\pm 5.35}$	$46.67_{\pm 6.03}$	$45.99_{\pm 5.67}$	$47.60_{\pm 8.16}$	$47.01_{\pm 8.25}$
	CN-only FT(own) + LP	$59.41_{\pm 3.15}$	$50.33_{\pm 7.92}$	$53.25_{\pm 7.06}$	$53.89_{\pm 9.06}$	$53.85_{\pm 8.44}$
	CN-only FT (all) + LP	$57.61_{\pm 5.56}$	$51.30_{\pm 8.49}$	$55.07_{\pm 7.93}$	$52.41_{\pm 6.33}$	$51.81_{\pm 8.90}$

Table A14: Few-shot performance across multi-class tasks for BrainNorm variants. Each method is shown under three modes: No FT, CN-only FT on the corresponding dataset only, and CN-only FT on pooled CN data from ADNI+AIBL+MIRIAD+PPMI+NIFD. Values are reported as mean ± standard deviation.

Method	Mode	$\mathbf{k=1}$	$\mathbf{k=2}$	$\mathbf{k=4}$	$\mathbf{k=8}$	$\mathbf{k=16}$
ADNI+NIFD+PPMI — [AD/FTD/PD]						
BrainNorm_M1	No FT + LP	$57.02_{\pm 9.59}$	$64.05_{\pm 9.45}$	$80.38_{\pm 3.50}$	$82.35_{\pm 2.93}$	$85.18_{\pm 3.57}$
	CN-only FT(own) + LP	$58.15_{\pm 9.92}$	$62.82_{\pm 9.94}$	$81.12_{\pm 3.45}$	$83.36_{\pm 3.50}$	$85.13_{\pm 2.73}$
	CN-only FT (all) + LP	$57.74_{\pm 9.61}$	$62.38_{\pm 9.24}$	$81.08_{\pm 3.52}$	$83.39_{\pm 3.95}$	$84.92_{\pm 3.55}$
BrainNorm_M2	No FT + LP	$58.97_{\pm 8.17}$	$66.91_{\pm 8.10}$	$78.65_{\pm 5.27}$	$82.12_{\pm 3.50}$	$84.34_{\pm 3.06}$
	CN-only FT(own) + LP	$58.67_{\pm 8.78}$	$67.13_{\pm 8.68}$	$79.83_{\pm 5.17}$	$81.17_{\pm 4.52}$	$83.08_{\pm 1.11}$
	CN-only FT (all) + LP	$58.46_{\pm 9.95}$	$67.38_{\pm 9.30}$	$80.40_{\pm 4.89}$	$81.71_{\pm 4.10}$	$83.33_{\pm 1.88}$
BrainNorm_M3	No FT + LP	$53.18_{\pm 6.50}$	$50.36_{\pm 7.08}$	$52.60_{\pm 8.68}$	$53.73_{\pm 8.67}$	$63.86_{\pm 8.68}$
	CN-only FT(own) + LP	$51.18_{\pm 8.62}$	$50.39_{\pm 5.39}$	$52.84_{\pm 8.99}$	$53.47_{\pm 5.39}$	$66.13_{\pm 3.71}$
	CN-only FT (all) + LP	$50.95_{\pm 9.40}$	$54.15_{\pm 6.12}$	$53.28_{\pm 7.56}$	$53.04_{\pm 6.34}$	$66.91_{\pm 4.28}$
ADNI+NIFD+PPMI — [AD/FTD/PD/CN]						
BrainNorm_M1	No FT + LP	$68.04_{\pm 3.85}$	$67.86_{\pm 3.92}$	$69.10_{\pm 5.04}$	$73.87_{\pm 2.06}$	$75.60_{\pm 2.76}$
	CN-only FT(own) + LP	$68.27_{\pm 3.33}$	$69.52_{\pm 5.52}$	$71.80_{\pm 3.52}$	$72.81_{\pm 4.60}$	$76.47_{\pm 3.14}$
	CN-only FT (all) + LP	$68.42_{\pm 2.91}$	$69.51_{\pm 5.51}$	$71.95_{\pm 3.21}$	$72.48_{\pm 4.55}$	$76.61_{\pm 2.51}$
BrainNorm_M2	No FT + LP	$65.18_{\pm 7.85}$	$69.72_{\pm 6.86}$	$73.32_{\pm 4.96}$	$77.32_{\pm 2.31}$	$78.88_{\pm 2.49}$
	CN-only FT(own) + LP	$62.96_{\pm 8.59}$	$71.88_{\pm 2.75}$	$74.11_{\pm 4.60}$	$78.12_{\pm 2.59}$	$79.92_{\pm 2.56}$
	CN-only FT (all) + LP	$63.78_{\pm 8.25}$	$72.75_{\pm 2.23}$	$75.15_{\pm 4.24}$	$78.76_{\pm 2.80}$	$80.04_{\pm 2.52}$
BrainNorm_M3	No FT + LP	$60.02_{\pm 6.14}$	$63.33_{\pm 6.43}$	$63.04_{\pm 4.05}$	$67.11_{\pm 2.37}$	$69.83_{\pm 3.60}$
	CN-only FT(own) + LP	$59.40_{\pm 6.11}$	$62.17_{\pm 8.15}$	$61.34_{\pm 7.88}$	$67.60_{\pm 2.32}$	$67.22_{\pm 2.58}$
	CN-only FT (all) + LP	$58.54_{\pm 6.03}$	$59.74_{\pm 6.16}$	$60.22_{\pm 7.41}$	$64.34_{\pm 6.19}$	$64.31_{\pm 6.45}$

Table A15: Atlas parcel index mapping used in BrainNorm.

Parcel ID	Atlas	Tissue_Class	Label ID	Region
0	AAL3	Gray_Matter	1	Left Precentral Gyrus
1	AAL3	Gray_Matter	2	Right Precentral Gyrus
2	AAL3	Gray_Matter	3	Left Superior Frontal Gyrus
3	AAL3	Gray_Matter	4	Right Superior Frontal Gyrus
4	AAL3	Gray_Matter	5	Left Middle Frontal Gyrus
5	AAL3	Gray_Matter	6	Right Middle Frontal Gyrus
6	AAL3	Gray_Matter	7	Left Inferior Frontal Gyrus Opercular Part
7	AAL3	Gray_Matter	8	Right Inferior Frontal Gyrus Opercular Part
8	AAL3	Gray_Matter	9	Left Inferior Frontal Gyrus Triangular Part
9	AAL3	Gray_Matter	10	Right Inferior Frontal Gyrus Triangular Part
10	AAL3	Gray_Matter	11	Left Inferior Frontal Gyrus Orbital Part
11	AAL3	Gray_Matter	12	Right Inferior Frontal Gyrus Orbital Part
12	AAL3	Gray_Matter	13	Left Rolandic Operculum
13	AAL3	Gray_Matter	14	Right Rolandic Operculum
14	AAL3	Gray_Matter	15	Left Supplementary Motor Area
15	AAL3	Gray_Matter	16	Right Supplementary Motor Area
16	AAL3	Gray_Matter	17	Left Olfactory Cortex
17	AAL3	Gray_Matter	18	Right Olfactory Cortex
18	AAL3	Gray_Matter	19	Left Superior Medial Frontal Gyrus
19	AAL3	Gray_Matter	20	Right Superior Medial Frontal Gyrus
20	AAL3	Gray_Matter	21	Left Medial Orbitofrontal Cortex
21	AAL3	Gray_Matter	22	Right Medial Orbitofrontal Cortex
22	AAL3	Gray_Matter	23	Left Gyrus Rectus
23	AAL3	Gray_Matter	24	Right Gyrus Rectus
24	AAL3	Gray_Matter	25	Left Orbitofrontal Cortex Medial Part
25	AAL3	Gray_Matter	26	Right Orbitofrontal Cortex Medial Part
26	AAL3	Gray_Matter	27	Left Orbitofrontal Cortex Anterior Part
27	AAL3	Gray_Matter	28	Right Orbitofrontal Cortex Anterior Part
28	AAL3	Gray_Matter	29	Left Orbitofrontal Cortex Posterior Part
29	AAL3	Gray_Matter	30	Right Orbitofrontal Cortex Posterior Part
30	AAL3	Gray_Matter	31	Left Orbitofrontal Cortex Lateral Part
31	AAL3	Gray_Matter	32	Right Orbitofrontal Cortex Lateral Part
32	AAL3	Gray_Matter	33	Left Insula
33	AAL3	Gray_Matter	34	Right Insula
34	AAL3	Gray_Matter	37	Left Middle Cingulate and Paracingulate Gyri
35	AAL3	Gray_Matter	38	Right Middle Cingulate and Paracingulate Gyri
36	AAL3	Gray_Matter	39	Left Posterior Cingulate Gyrus
37	AAL3	Gray_Matter	40	Right Posterior Cingulate Gyrus
38	AAL3	Gray_Matter	41	Left Hippocampus
39	AAL3	Gray_Matter	42	Right Hippocampus
40	AAL3	Gray_Matter	43	Left Parahippocampal Gyrus
41	AAL3	Gray_Matter	44	Right Parahippocampal Gyrus
42	AAL3	Gray_Matter	45	Left Amygdala
43	AAL3	Gray_Matter	46	Right Amygdala
44	AAL3	Gray_Matter	47	Left Calcarine Fissure and Cortex
45	AAL3	Gray_Matter	48	Right Calcarine Fissure and Cortex
46	AAL3	Gray_Matter	49	Left Cuneus
47	AAL3	Gray_Matter	50	Right Cuneus
48	AAL3	Gray_Matter	51	Left Lingual Gyrus
49	AAL3	Gray_Matter	52	Right Lingual Gyrus
50	AAL3	Gray_Matter	53	Left Superior Occipital Gyrus
51	AAL3	Gray_Matter	54	Right Superior Occipital Gyrus
52	AAL3	Gray_Matter	55	Left Middle Occipital Gyrus
53	AAL3	Gray_Matter	56	Right Middle Occipital Gyrus
54	AAL3	Gray_Matter	57	Left Inferior Occipital Gyrus
55	AAL3	Gray_Matter	58	Right Inferior Occipital Gyrus
56	AAL3	Gray_Matter	59	Left Fusiform Gyrus
57	AAL3	Gray_Matter	60	Right Fusiform Gyrus
58	AAL3	Gray_Matter	61	Left Postcentral Gyrus
59	AAL3	Gray_Matter	62	Right Postcentral Gyrus
60	AAL3	Gray_Matter	63	Left Superior Parietal Gyrus
61	AAL3	Gray_Matter	64	Right Superior Parietal Gyrus

Continued on next page

Parcel ID	Atlas	Tissue_Class	Label ID	Region
62	AAL3	Gray_Matter	65	Left Inferior Parietal Gyrus
63	AAL3	Gray_Matter	66	Right Inferior Parietal Gyrus
64	AAL3	Gray_Matter	67	Left Supramarginal Gyrus
65	AAL3	Gray_Matter	68	Right Supramarginal Gyrus
66	AAL3	Gray_Matter	69	Left Angular Gyrus
67	AAL3	Gray_Matter	70	Right Angular Gyrus
68	AAL3	Gray_Matter	71	Left Precuneus
69	AAL3	Gray_Matter	72	Right Precuneus
70	AAL3	Gray_Matter	73	Left Paracentral Lobule
71	AAL3	Gray_Matter	74	Right Paracentral Lobule
72	AAL3	Gray_Matter	75	Left Caudate Nucleus
73	AAL3	Gray_Matter	76	Right Caudate Nucleus
74	AAL3	Gray_Matter	77	Left Putamen
75	AAL3	Gray_Matter	78	Right Putamen
76	AAL3	Gray_Matter	79	Left Pallidum
77	AAL3	Gray_Matter	80	Right Pallidum
78	AAL3	Gray_Matter	83	Left Heschl's Gyrus
79	AAL3	Gray_Matter	84	Right Heschl's Gyrus
80	AAL3	Gray_Matter	85	Left Superior Temporal Gyrus
81	AAL3	Gray_Matter	86	Right Superior Temporal Gyrus
82	AAL3	Gray_Matter	87	Left Temporal Pole: Superior Temporal Gyrus
83	AAL3	Gray_Matter	88	Right Temporal Pole: Superior Temporal Gyrus
84	AAL3	Gray_Matter	89	Left Middle Temporal Gyrus
85	AAL3	Gray_Matter	90	Right Middle Temporal Gyrus
86	AAL3	Gray_Matter	91	Left Temporal Pole: Middle Temporal Gyrus
87	AAL3	Gray_Matter	92	Right Temporal Pole: Middle Temporal Gyrus
88	AAL3	Gray_Matter	93	Left Inferior Temporal Gyrus
89	AAL3	Gray_Matter	94	Right Inferior Temporal Gyrus
90	AAL3	Gray_Matter	95	Left Crus I of Cerebellar Hemisphere
91	AAL3	Gray_Matter	96	Right Crus I of Cerebellar Hemisphere
92	AAL3	Gray_Matter	97	Left Crus II of Cerebellar Hemisphere
93	AAL3	Gray_Matter	98	Right Crus II of Cerebellar Hemisphere
94	AAL3	Gray_Matter	99	Left Lobule III of Cerebellar Hemisphere
95	AAL3	Gray_Matter	100	Right Lobule III of Cerebellar Hemisphere
96	AAL3	Gray_Matter	101	Left Lobule IV-V of Cerebellar Hemisphere
97	AAL3	Gray_Matter	102	Right Lobule IV-V of Cerebellar Hemisphere
98	AAL3	Gray_Matter	103	Left Lobule VI of Cerebellar Hemisphere
99	AAL3	Gray_Matter	104	Right Lobule VI of Cerebellar Hemisphere
100	AAL3	Gray_Matter	105	Left Lobule VIIB of Cerebellar Hemisphere
101	AAL3	Gray_Matter	106	Right Lobule VIIB of Cerebellar Hemisphere
102	AAL3	Gray_Matter	107	Left Lobule VIII of Cerebellar Hemisphere
103	AAL3	Gray_Matter	108	Right Lobule VIII of Cerebellar Hemisphere
104	AAL3	Gray_Matter	109	Left Lobule IX of Cerebellar Hemisphere
105	AAL3	Gray_Matter	110	Right Lobule IX of Cerebellar Hemisphere
106	AAL3	Gray_Matter	111	Left Lobule X of Cerebellar Hemisphere
107	AAL3	Gray_Matter	112	Right Lobule X of Cerebellar Hemisphere
108	AAL3	Gray_Matter	113	Lobule I-II of Vermis
109	AAL3	Gray_Matter	114	Lobule III of Vermis
110	AAL3	Gray_Matter	115	Lobule IV-V of Vermis
111	AAL3	Gray_Matter	116	Lobule VI of Vermis
112	AAL3	Gray_Matter	117	Lobule VII of Vermis
113	AAL3	Gray_Matter	118	Lobule VIII of Vermis
114	AAL3	Gray_Matter	119	Lobule IX of Vermis
115	AAL3	Gray_Matter	120	Lobule X of Vermis
116	AAL3	Gray_Matter	121	Left Anteroventral Thalamic Nucleus
117	AAL3	Gray_Matter	122	Right Anteroventral Thalamic Nucleus
118	AAL3	Gray_Matter	123	Left Lateral Posterior Thalamic Nucleus
119	AAL3	Gray_Matter	124	Right Lateral Posterior Thalamic Nucleus
120	AAL3	Gray_Matter	125	Left Ventral Anterior Thalamic Nucleus
121	AAL3	Gray_Matter	126	Right Ventral Anterior Thalamic Nucleus
122	AAL3	Gray_Matter	127	Left Ventral Lateral Thalamic Nucleus
123	AAL3	Gray_Matter	128	Right Ventral Lateral Thalamic Nucleus
124	AAL3	Gray_Matter	129	Left Ventral Posterolateral Thalamic Nucleus
125	AAL3	Gray_Matter	130	Right Ventral Posterolateral Thalamic Nucleus

Continued on next page

Parcel ID	Atlas	Tissue_Class	Label ID	Region
126	AAL3	Gray_Matter	131	Left Intralaminar Thalamic Nuclei
127	AAL3	Gray_Matter	132	Right Intralaminar Thalamic Nuclei
128	AAL3	Gray_Matter	133	Left Reuniens Thalamic Nucleus
129	AAL3	Gray_Matter	134	Right Reuniens Thalamic Nucleus
130	AAL3	Gray_Matter	135	Left Mediodorsal Medial Magnocellular Thalamic Nucleus
131	AAL3	Gray_Matter	136	Right Mediodorsal Medial Magnocellular Thalamic Nucleus
132	AAL3	Gray_Matter	137	Left Mediodorsal Lateral Parvocellular Thalamic Nucleus
133	AAL3	Gray_Matter	138	Right Mediodorsal Lateral Parvocellular Thalamic Nucleus
134	AAL3	Gray_Matter	139	Left Lateral Geniculate Thalamic Nucleus
135	AAL3	Gray_Matter	140	Right Lateral Geniculate Thalamic Nucleus
136	AAL3	Gray_Matter	141	Left Medial Geniculate Thalamic Nucleus
137	AAL3	Gray_Matter	142	Right Medial Geniculate Thalamic Nucleus
138	AAL3	Gray_Matter	143	Left Pulvinar Inferior Thalamic Nucleus
139	AAL3	Gray_Matter	144	Right Pulvinar Inferior Thalamic Nucleus
140	AAL3	Gray_Matter	145	Left Pulvinar Medial Thalamic Nucleus
141	AAL3	Gray_Matter	146	Right Pulvinar Medial Thalamic Nucleus
142	AAL3	Gray_Matter	147	Left Pulvinar Anterior Thalamic Nucleus
143	AAL3	Gray_Matter	148	Right Pulvinar Anterior Thalamic Nucleus
144	AAL3	Gray_Matter	149	Left Pulvinar Lateral Thalamic Nucleus
145	AAL3	Gray_Matter	150	Right Pulvinar Lateral Thalamic Nucleus
146	AAL3	Gray_Matter	151	Left Anterior Cingulate Cortex Subgenual Part
147	AAL3	Gray_Matter	152	Right Anterior Cingulate Cortex Subgenual Part
148	AAL3	Gray_Matter	153	Left Anterior Cingulate Cortex Pregenual Part
149	AAL3	Gray_Matter	154	Right Anterior Cingulate Cortex Pregenual Part
150	AAL3	Gray_Matter	155	Left Anterior Cingulate Cortex Supracallosal Part
151	AAL3	Gray_Matter	156	Right Anterior Cingulate Cortex Supracallosal Part
152	AAL3	Gray_Matter	157	Left Nucleus Accumbens
153	AAL3	Gray_Matter	158	Right Nucleus Accumbens
154	AAL3	Gray_Matter	159	Left Ventral Tegmental Area
155	AAL3	Gray_Matter	160	Right Ventral Tegmental Area
156	AAL3	Gray_Matter	161	Left Substantia Nigra Pars Compacta
157	AAL3	Gray_Matter	162	Right Substantia Nigra Pars Compacta
158	AAL3	Gray_Matter	163	Left Substantia Nigra Pars Reticulata
159	AAL3	Gray_Matter	164	Right Substantia Nigra Pars Reticulata
160	AAL3	Gray_Matter	165	Left Red Nucleus
161	AAL3	Gray_Matter	166	Right Red Nucleus
162	AAL3	Gray_Matter	167	Left Locus Coeruleus
163	AAL3	Gray_Matter	168	Right Locus Coeruleus
164	AAL3	Gray_Matter	169	Dorsal Raphe Nucleus
165	AAL3	Gray_Matter	170	Median Raphe Nucleus
166	JHU	White_Matter	1	Middle Cerebellar Peduncle
167	JHU	White_Matter	2	Pontine Crossing Tract
168	JHU	White_Matter	3	Genu Of Corpus Callosum
169	JHU	White_Matter	4	Body Of Corpus Callosum
170	JHU	White_Matter	5	Splenium Of Corpus Callosum
171	JHU	White_Matter	6	Fornix
172	JHU	White_Matter	7	Right Corticospinal Tract
173	JHU	White_Matter	8	Left Corticospinal Tract
174	JHU	White_Matter	9	Right Medial Lemniscus
175	JHU	White_Matter	10	Left Medial Lemniscus
176	JHU	White_Matter	11	Right Inferior Cerebellar Peduncle
177	JHU	White_Matter	12	Left Inferior Cerebellar Peduncle
178	JHU	White_Matter	13	Right Superior Cerebellar Peduncle
179	JHU	White_Matter	14	Left Superior Cerebellar Peduncle
180	JHU	White_Matter	15	Right Cerebral Peduncle
181	JHU	White_Matter	16	Left Cerebral Peduncle
182	JHU	White_Matter	17	Right Anterior Limb Of Internal Capsule
183	JHU	White_Matter	18	Left Anterior Limb Of Internal Capsule
184	JHU	White_Matter	19	Right Posterior Limb Of Internal Capsule
185	JHU	White_Matter	20	Left Posterior Limb Of Internal Capsule
186	JHU	White_Matter	21	Right Retrolenticular Part Of Internal Capsule
187	JHU	White_Matter	22	Left Retrolenticular Part Of Internal Capsule
188	JHU	White_Matter	23	Right Anterior Corona Radiata
189	JHU	White_Matter	24	Left Anterior Corona Radiata

Continued on next page

Parcel ID	Atlas	Tissue_Class	Label ID	Region
190	JHU	White_Matter	25	Right Superior Corona Radiata
191	JHU	White_Matter	26	Left Superior Corona Radiata
192	JHU	White_Matter	27	Right Posterior Corona Radiata
193	JHU	White_Matter	28	Left Posterior Corona Radiata
194	JHU	White_Matter	29	Right Posterior Thalamic Radiation
195	JHU	White_Matter	30	Left Posterior Thalamic Radiation
196	JHU	White_Matter	31	Right Sagittal Stratum
197	JHU	White_Matter	32	Left Sagittal Stratum
198	JHU	White_Matter	33	Right External Capsule
199	JHU	White_Matter	34	Left External Capsule
200	JHU	White_Matter	35	Right Cingulum (cingulate Gyrus)
201	JHU	White_Matter	36	Left Cingulum (cingulate Gyrus)
202	JHU	White_Matter	37	Right Cingulum (hippocampus)
203	JHU	White_Matter	38	Left Cingulum (hippocampus)
204	JHU	White_Matter	39	Right Fornix / Stria Terminalis
205	JHU	White_Matter	40	Left Fornix / Stria Terminalis
206	JHU	White_Matter	41	Right Superior Longitudinal Fasciculus
207	JHU	White_Matter	42	Left Superior Longitudinal Fasciculus
208	JHU	White_Matter	43	Right Superior Fronto-occipital Fasciculus
209	JHU	White_Matter	44	Left Superior Fronto-occipital Fasciculus
210	JHU	White_Matter	45	Right Inferior Fronto-occipital Fasciculus
211	JHU	White_Matter	46	Left Inferior Fronto-occipital Fasciculus
212	JHU	White_Matter	47	Right Uncinate Fasciculus
213	JHU	White_Matter	48	Left Uncinate Fasciculus
214	JHU	White_Matter	49	Right Tapetum
215	JHU	White_Matter	50	Left Tapetum
216	HO	Cerebrospinal_Fluid	2	Left Lateral Ventricle
217	HO	Cerebrospinal_Fluid	13	Right Lateral Ventricle