INTER-REGIONAL STRUCTURAL SELF-SUPERVISED LEARNING FOR BRAIN MRI: INDIVIDUAL AND POPULATION FORMULATIONS

Anonymous authors
Paper under double-blind review

ABSTRACT

Brain structure is characterized not only by the properties of individual anatomical regions but also by the relationships among them. These inter-regional relationships constitute a fundamental aspect of brain organization. However, existing self-supervised learning (SSL) methods for structural brain MRI predominantly rely on voxel reconstruction or instance-level consistency, without explicitly modeling such inter-regional structural organization.

We introduce Inter-Regional Structural Self-Supervised Learning (IRS-SSL), a framework that establishes *inter-regional structural organization* as its primary self-supervisory principle and inductive bias. Rather than incorporating inter-regional relationships as an auxiliary constraint on a conventional SSL objective, IRS-SSL derives its primary supervisory signal directly from these relationships. We instantiate this principle through two formulations. The individual formulation, IRS-SSL-Ind, characterizes pairwise similarities between regional structural profiles *within each subject*. The population formulation, IRS-SSL-Pop, captures coordinated variation between brain regions *across subjects*. The two formulations encode distinct relational information: IRS-SSL-Ind describes how regions relate within a particular brain, whereas IRS-SSL-Pop captures structural regularities shared across individuals.

We systematically evaluate both formulations across diverse downstream neuroimaging tasks and compare them with state-of-the-art SSL frameworks. Both formulations demonstrate the value of inter-regional structural organization as a source of self-supervision. Notably, IRS-SSL-Ind yields strong and consistent downstream performance. These findings establish inter-regional structural organization as an effective inductive bias for SSL and provide new insights into the relative value of individual- and population-derived relationships for learning representations from structural brain MRI.

1 INTRODUCTION

Self-supervised learning (SSL) has emerged as a powerful paradigm for learning transferable representations from large-scale unlabeled data (Chen et al., 2020; He et al., 2022; Tong et al., 2022; Assran et al., 2023; Bardes et al., 2024). By deriving supervision from the data themselves, SSL reduces dependence on manual annotation and enables pretrained models to be adapted to diverse downstream tasks with limited labeled data.

SSL approaches have been adopted in biomedical imaging, where expert annotation is costly and often constrained by specialized clinical expertise (Zhou et al., 2023; Tak et al., 2026; Liu et al., 2026; Mazher et al., 2025; Caro et al.; Shakouri et al., 2023; Dong et al., 2024; Xu et al., 2025; Jiang et al., 2023; Qiu et al., 2024). Applications span neuroimaging (Tak et al., 2026; Xu et al., 2025; Yang et al., 2025), retinal imaging (Liu et al., 2026; Zhou et al., 2023; Yu et al., 2024), and computed tomography (Zhuang et al., 2025; Li et al., 2025; Valanarasu et al., 2024). For structural brain MRI, most existing methods adapt general-purpose SSL objectives, such as contrastive learning, masked reconstruction, and self-distillation, to volumetric images (Tak et al., 2026; Kaczmarek et al., 2025; Cox et al., 2024; Xu et al., 2025). Although these methods can implicitly learn anatomical patterns from image appearance, neuroanatomical organization is not explicitly incorporated into their supervisory objectives.

A limited body of work has begun to introduce neuroanatomical information into pretraining. AnatCL (Barbano et al., 2026) uses regional anatomical measurements, including cortical thickness, gray-matter volume, and surface area, to modulate the degree of positiveness between subjects in a weakly supervised contrastive objective. Anatomical information therefore guides subject-level alignment, while contrastive learning remains the primary supervisory principle. DAMT (Kim et al., 2024) adopts a multi-task framework built around conventional SSL pretext tasks and supplements them with anatomy-related prediction targets, including brain parcellation, cortical morphology, and regional radiomic features. These studies primarily use neuroanatomy to guide or supplement conventional SSL objectives (See Supp. for related work discussion).

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.

Consequently, existing neuroanatomy-aware methods remain centered on sample-level alignment or voxel- and region-wise anatomical properties. They do not explicitly model how different anatomical regions are structurally related, nor do they investigate such relationships as the primary basis of self-supervision. This leaves a fundamental aspect of brain organization underexplored in representation learning: moving from region-wise anatomical supervision to inter-regional structural self-supervision, and from anatomy as guidance or an auxiliary target to anatomy-derived relationships as the principal inductive bias.

This gap is important because relationships among anatomical regions constitute a foundational aspect of brain organization. Within an individual brain, similarities among regional structural profiles describe subject-specific organization; across a population, coordinated regional variation captures structural regularities shared across individuals. These relationships have been associated with brain development, ageing, cognition, individual variability, and neurodegenerative disease (Alexander-Bloch et al., 2013; Zhao et al., 2020; Liu et al., 2024; Chao et al., 2024; Mongay-Ochoa et al., 2025).

In this paper, we investigate whether inter-regional structural relationships can serve as the principal inductive bias for SSL (Fig. 1). We propose *Inter-Regional Structural Self-Supervised Learning* (IRS-SSL), which derives its primary supervisory signal directly from relationships among anatomical regions, rather than using anatomical information to guide or supplement a conventional SSL objective. We instantiate this principle through two formulations. IRS-SSL-Ind learns from pairwise similarities among regional structural profiles within each subject, emphasizing individualized brain organization. IRS-SSL-Pop learns from structural covariance between regions across subjects, emphasizing anatomical regularities shared at the population level. We investigate the two formulations separately to examine their relative value for structural MRI representation learning.

We evaluate IRS-SSL across diverse downstream neuroimaging tasks under linear-probing, finetuning and data-limited settings. Both formulations demonstrate the value of inter-regional structural self-supervision, while IRS-SSL-Ind achieves particularly strong and consistent performance. Clinical review further supports the clinical relevance of the learned representations. Our main contributions are:

- We introduce inter-regional structural organization as a primary self-supervisory principle for structural brain MRI, moving from region-wise anatomical supervision to relational supervision.
- We propose two neuroscientifically grounded formulations: IRS-SSL-Ind, based on individualized structural similarity, and IRS-SSL-Pop, based on population-level structural covariance.
- We systematically evaluate both formulations across diverse downstream tasks, demonstrating strong performance and revealing the particular effectiveness of individualized inter-regional structure.

2 METHOD

2.1 NEUROSCIENTIFIC BASIS OF INTER-REGIONAL STRUCTURAL ORGANIZATION

Inter-regional structural organization is a foundational property of the brain and can be characterized from both individual and population perspectives (Zhao et al., 2020; Liu et al., 2024; Chao et al., 2024; Mongay-Ochoa et al., 2025). These perspectives provide biologically grounded inter-regional information that is not captured by considering regional properties independently. We translate these two perspectives into self-supervised pretext tasks: **IRS-SSL-Ind** learns from individualized structural relationships within each brain, whereas **IRS-SSL-Pop** learns from population-level structural relationships across brains.

Individual-level inter-regional structural similarity. Within a single brain, anatomical regions can be represented by multivariate profiles of regional structural characteristics. Pairwise similarities between these profiles define an individualized structural network that characterizes how distributed regions are organized within a particular brain (Liu et al., 2024). Importantly, prior studies have shown that these networks are reproducible and sufficiently individualized to support subject identification, while also capturing variation associated with fluid intelligence, mild cognitive impairment subtypes, and frontotemporal dementia (Liu et al., 2024; Nigro et al., 2025). Their network properties have further been associated with cognitive performance and biological markers of neurodegeneration (Nigro et al., 2025). Motivated by this evidence, IRS-SSL-Ind uses individualized inter-regional structural similarity as its self-supervisory principle, encouraging the learned representation to preserve subject-specific patterns of distributed brain organization.

Population-level inter-regional structural covariance. Across a population, anatomical regions exhibit coordinated variation in their structural properties. This structural covariance is commonly characterized through pairwise associations between regional measurements across individuals (Zhao et al., 2020; Carmon et al., 2020; Chao et al., 2024; Mongay-Ochoa et al., 2025). Importantly, prior studies have shown that structural covariance changes across the lifespan and is altered in neurodegenerative conditions, including Alzheimer's disease and frontotemporal dementia (Chao et al., 2024; Mongay-Ochoa et al., 2025; Brown et al., 2025; Mollah et al., 2024; Ghoshal &

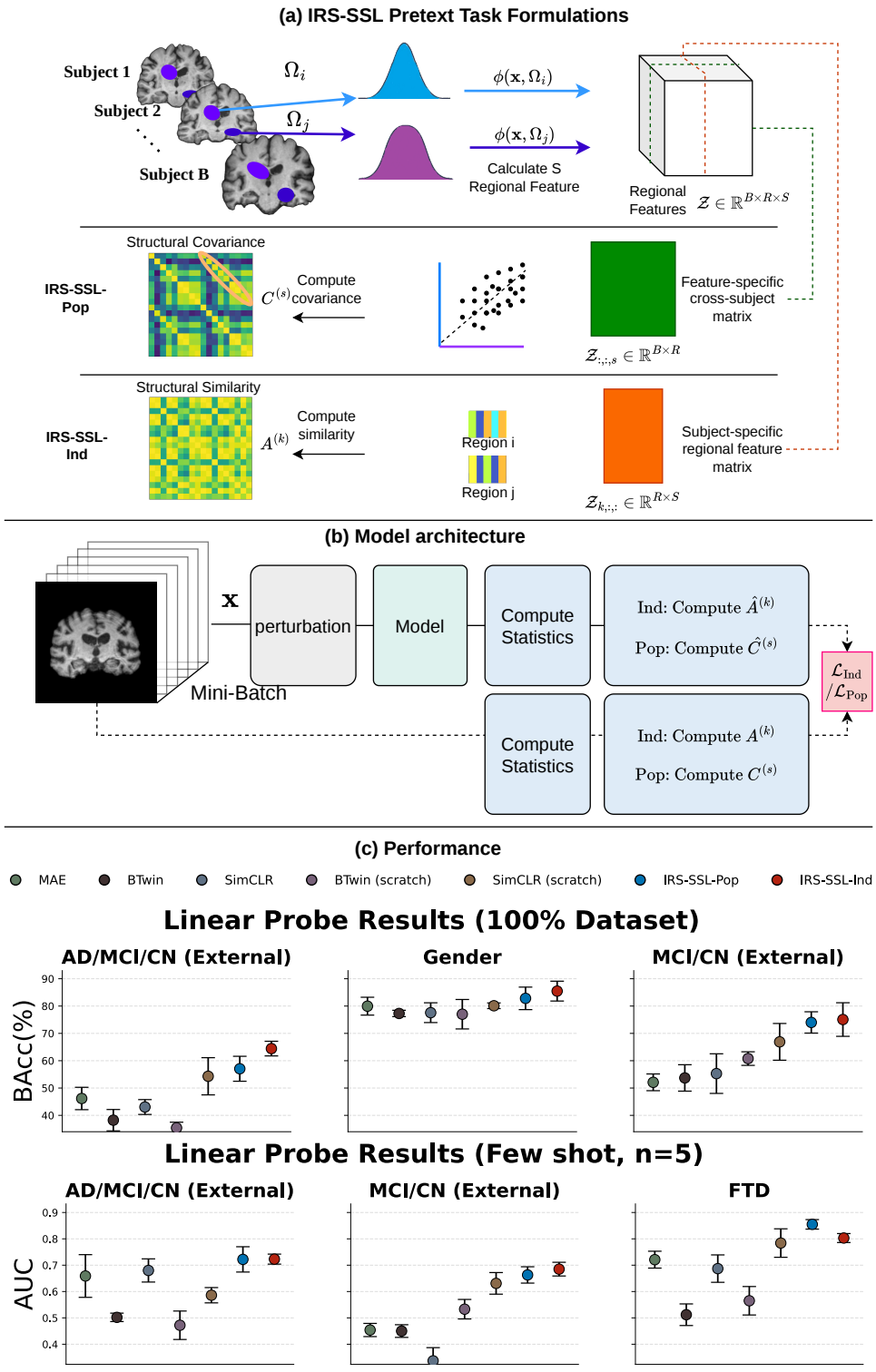


Figure 1: **Overview of Inter-Regional Structural Self-Supervised Learning (IRS-SSL).** IRS-SSL establishes inter-regional structural organization as the primary self-supervisory principle and inductive bias for structural brain MRI. (Sec. 2.1) **(a)** This principle is instantiated through two pretext task formulations: IRS-SSL-Ind derives supervision from pairwise similarities between regional structural profiles within each subject, whereas IRS-SSL-Pop derives supervision from coordinated regional variation across subjects (Sec. 2.2). The two formulations capture subject-specific organization and population-level structural regularities, respectively. **(b)** The proposed principle is not tied to a particular architecture. Perturbed inputs are encoded, and regional features can be obtained through either an image-space or embedding-space realized architecture (Sec. 2.3). The resulting inter-regional relationships are matched to those derived from the original inputs using the corresponding IRS-SSL objective. **(c)** Representative results, together with review by six certified radiologists, support the clinical relevance of the representations learned by IRS-SSL (Sec. 3). **Best viewed in color and with zoom.**

Cairns, 2013). These findings establish population-level structural covariance as a biologically and clinically relevant representation. Motivated by this evidence, IRS-SSL-Pop uses inter-regional structural covariance as its self-supervisory principle, encouraging learned representation to preserve regularities shared across individuals.

In what follows, we discuss two self-supervised task formulations to operationalize these complementary inter-regional structural perspectives. IRS-SSL-Ind derives its supervisory signal from individualized relationships among regional structural profiles *within each subject*, thereby emphasizing subject-specific brain organization. In contrast, IRS-SSL-Pop derives its supervisory signal from coordinated regional variation *across subjects*, thereby emphasizing structural regularities shared at the population level.

2.2 Self-Supervised Pretext Task Formulations

Region-level representation. Let $\mathbf{x}^{(k)} \in \mathbb{R}^{H\times W\times D}$ denote the structural MRI of subject k. Using a neuroanatomical atlas, we partition the brain into R anatomically defined regions of interest (ROIs), $\{\Omega_1, \ldots, \Omega_R\}$ (Fig. 1). For each ROI Ω_i, we extract S regional features, $\mathbf{z}_i^{(k)} = \phi(\mathbf{x}^{(k)}, \Omega_i) \in \mathbb{R}^S$, where $\phi(\cdot)$ denotes the regional feature extraction operator. Stacking the ROI-wise feature vectors yields $\mathbf{Z}^{(k)} = [(\mathbf{z}_1^{(k)})^\top, \ldots, (\mathbf{z}_R^{(k)})^\top]^\top \in \mathbb{R}^{R\times S}$. For a mini-batch of size B, we further stack the feature matrices to obtain $\mathcal{Z} \in \mathbb{R}^{B\times R\times S}$, where $\mathcal{Z}_{k,i,:} = \mathbf{z}_i^{(k)}$. The specific choice of regional features is discussed in Sec. 2.3.

IRS-SSL-Ind: Individual-level structural similarity. IRS-SSL-Ind considers *subject-specific regional feature matrix* $\mathcal{Z}_{k,:,:} \in \mathbb{R}^{R\times S}$. Following Liu et al. (2024); Zhao et al. (2020), $\mathbf{z}_i^{(k)}$ is centered and normalized. Let $\widetilde{\mathbf{z}}_i^{(k)}$ denote the normalized feature vector of ROI Ω_i. We define the similarity between regions Ω_i and Ω_j as $A_{i,j}^{(k)} = \frac{1}{S-1}\left(\widetilde{\mathbf{z}}_i^{(k)}\right)^\top \widetilde{\mathbf{z}}_j^{(k)}$, yielding a subject-specific similarity matrix $\mathbf{A}^{(k)} \in \mathbb{R}^{R\times R}$.

The input is also perturbed and passed through the encoder (Fig. 1). The resulting embeddings are either decoded into an image volume, from which regional features $\hat{\mathbf{Z}}^{(k)}$ are extracted in the image-space realized architecture, or mapped directly to $\hat{\mathbf{Z}}^{(k)}$ by an MLP in the embedding-space realized architecture (See architecture realization in Sec. 2.3). Applying the same normalization and pairwise-similarity operation to $\hat{\mathbf{Z}}^{(k)}$ yields the predicted similarity matrix $\hat{\mathbf{A}}^{(k)} \in \mathbb{R}^{R\times R}$.

Since $\mathbf{A}^{(k)}$ is symmetric, we retain only the upper-triangular entries, corresponding to the unique pairs of distinct ROIs. IRS-SSL-Ind preserves the subject-specific inter-regional structure by minimizing

$$\mathcal{L}_{\text{Ind}} = \frac{2}{BR(R-1)} \sum_{k=1}^{B} \sum_{i=1}^{R-1} \sum_{j=i+1}^{R} \left| A_{i,j}^{(k)} - \hat{A}_{i,j}^{(k)} \right|. \tag{1}$$

IRS-SSL-Pop: Population-level structural covariance. IRS-SSL-Pop estimates inter-regional relationships across subjects. For each regional feature $s \in \{1, \ldots, S\}$, we form *feature-specific cross-subject matrix* $\mathbf{M}^{(s)} = \mathcal{Z}_{:,:,s} \in \mathbb{R}^{B\times R}$, where rows correspond to subjects and columns correspond to ROIs. For each ROI, the regional measurements are centered and normalized across the minibatch to obtain $\widetilde{\mathbf{M}}^{(s)}$. The resulting inter-regional covariance matrix is

$$\mathbf{C}^{(s)} = \frac{1}{B-1}\left(\widetilde{\mathbf{M}}^{(s)}\right)^\top \widetilde{\mathbf{M}}^{(s)} \in \mathbb{R}^{R\times R}. \tag{2}$$

As in IRS-SSL-Ind, perturbed inputs are passed through the encoder and either decoded into image volumes or mapped directly to regional features by an MLP, following the two architectural realizations discussed in Sec. 2.3. For each feature s, applying the same normalization and covariance operation to the resulting regional features yields the predicted covariance matrix $\hat{\mathbf{C}}^{(s)} \in \mathbb{R}^{R\times R}$. IRS-SSL-Pop preserves these population-level relationships by minimizing

$$\mathcal{L}_{\text{Pop}} = \frac{1}{S}\sum_{s=1}^{S} \left\| \mathbf{C}^{(s)} - \hat{\mathbf{C}}^{(s)} \right\|_1. \tag{3}$$

The specific choice of retained covariance relationships is discussed in Sec. 2.3.

2.3 Self-Supervised Pretext Task Design Choices

Architectural realization. The central contribution of IRS-SSL lies in its self-supervisory principle rather than in a particular architecture. This principle can be realized through different architectures that enable inter-regional structural relationships to be derived. We investigate two realizations.

Given the input MRI of subject k, $\mathbf{x}^{(k)}$, perturbation is applied. In the *image-space realization*, the perturbed MRI is processed by an encoder–decoder to produce a decoded volume $\hat{\mathbf{x}}^{(k)}$. Regional features are extracted

from $\hat{\mathbf{x}}^{(k)}$ and used to compute individualized structural similarity or population-level structural covariance. In the *embedding-space realization*, an MLP directly predicts the regional features from the encoder embeddings, bypassing volume decoding. The same inter-regional objectives are then applied to the resulting regional features.

Our ablation study shows that the image-space realization yields better downstream performance. We therefore adopt it as the default architecture for subsequent experiments, while emphasizing that the proposed inter-regional self-supervisory principle is not restricted to this particular realization.

Regional feature design. In neuroimaging, regional *morphometric features*, such as cortical thickness and gray-matter volume (Alexander-Bloch et al., 2013; Xu et al., 2018), are commonly used. These features are derived from anatomical segmentation and provide meaningful descriptors of regional morphology. However, directly adopting such features in our self-supervised task is impractical because their extraction requires accurate anatomical segmentation for each sample. Although segmentation can be performed offline for each input volume, obtaining corresponding features from the model output during pretraining would require additional segmentation-related processing at every iteration. The resulting computational overhead would substantially limit the scalability of self-supervised pretraining.

To address this limitation, we instead explore *radiomics-based features* derived from voxel intensities within each ROI. Recent neuroimaging studies have increasingly adopted such features, including mean intensity and statistical descriptors of the regional intensity distribution, to characterize structural variation (Zhao et al., 2020; Nigro et al., 2025). In this work, we consider five established radiomics-based features: mean, variance, entropy, skewness, and kurtosis. As these features can be computed directly using the ROI definitions obtained from the input data, *this design avoids repeated anatomical segmentation during large-scale self-supervised pretraining*. This enables efficient and scalable construction of the inter-regional self-supervisory signals. For IRS-SSL-Pop, our ablation shows that using only mean, variance, and entropy achieves comparable downstream performance. We therefore adopt these features to further reduce pretraining computation.

Selection of population-level structural covariance. Estimating cross-subject covariance for all possible pairs of brain regions provides comprehensive coverage of population-level inter-regional relationships. However, it may include many region pairs without a strong *a priori* basis for coordinated anatomical variation across individuals. Covariance estimates for these weakly related or unrelated pairs may introduce noise and encourage the learning of spurious inter-regional associations.

We instead focus IRS-SSL-Pop on cross-subject covariance between *homologous left–right regions*, a prominent and well-established population-level relationship grounded in the bilateral organization of brain anatomy (Alexander-Bloch et al., 2013). This interhemispheric structural covariance provides a more focused population-level self-supervisory signal. Our ablation study further validates this design choice.

Pretraining schedule. Inter-regional structural organization is a higher-level relational property whose learning can benefit from an initialization that captures basic voxel-level structure. We therefore initialize the model with 50 epochs of masked autoencoding to establish basic voxel representations, followed by 200 epochs of pretraining using either IRS-SSL-Ind or IRS-SSL-Pop alone. Importantly, the MAE objective is removed after initialization and is not jointly optimized with the inter-regional objective. Thus, inter-regional structural organization remains the sole supervisory principle throughout the main pretraining stage.

Our ablation compares this schedule with both the 50-epoch MAE initialization and continued MAE pretraining for the same total of 250 epochs. IRS-SSL formulations outperform these settings. Comparisons with other SSL methods under identical experimental settings further indicate that the gains arise from inter-regional structural self-supervision.

3 Experimental Evaluation

3.1 Experimental Settings

Pretraining datasets. We pretrain IRS-SSL on a collection of structural brain MRI scans aggregated from multiple public datasets, totaling 12,541 MRIs. All scans are skull-stripped, bias-field corrected, and registered to Montreal Neurological Institute (MNI) (Collins, 2009) space. These unlabeled data are used exclusively for self-supervised pretraining. We use the Harvard–Oxford Atlas with 21 predefined ROIs (Rushmore et al., 2022).

Downstream task adaptation. For downstream evaluation, we closely follow the BrainIAC (Tak et al., 2026) evaluation protocol. Representation quality is evaluated using the standard SSL linear probing protocol as well as layer decay fine-tuning across downstream evaluation settings, spanning clinical classification, demographic prediction, and external benchmark tasks, including: MCI vs. CN classification, AD vs. MCI vs. CN classification, FTD vs. CN classification, brain age prediction (regression), gender classification, and tumor grading. *Importantly,*

the pretraining dataset does not contain samples from any downstream evaluation tasks, ensuring strict separation between pretraining and downstream evaluation.

Evaluation metrics. Classification performance is measured using balanced accuracy, AUC, and F1 score, while regression performance is evaluated using mean absolute error (shown as L1) between predicted and chronological age. Results are reported on 20% of held-out test sets.

Pretraining configuration. We compare the proposed IRS-SSL against state-of-the-art SSL approaches spanning major SSL paradigms, including contrastive learning and reconstruction-based SSL. For contrastive SSL, we adopt SimCLR, which has been used in SOTA BrainIAC (Tak et al., 2026) for structural brain MRI. For reconstruction-based SSL, we follow SOTA MiM (Zhuang et al., 2025) and adopt MAE. All the SSL methods are pretrained for 200 epochs with identical initialization (50 epochs MAE) and identical implementation; see Supp. C–D.

3.2 IRS-SSL OUTPERFORMS STANDARD SSL OBJECTIVES AND YIELDS THE STRONGEST OVERALL DOWNSTREAM PERFORMANCE

Table 1: Comparison of SSL pretext strategies evaluated on downstream and external test tasks under linear probing. Results are averaged over three runs; 95% confidence intervals are reported in Supp. E Tab. 12. The best result per column is **bolded** and the second-best is underlined. BAcc = balanced accuracy (%).

	Age	**Gender**			**MCI/CN**			**FTD/CN**		
Pretext	L1↓	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑
MAE	9.556	79.957	0.895	0.804	72.927	0.813	0.735	73.043	0.789	0.722
BarlowTwins	14.911	77.290	0.834	0.775	62.003	0.713	0.610	60.017	0.663	0.600
SimCLR	10.203	77.577	0.846	0.779	69.810	0.762	0.675	71.623	0.804	0.718
BarlowTwins (scratch)	14.623	77.017	0.864	0.772	64.423	0.814	0.639	69.237	0.749	0.679
SimCLR (scratch)	9.437	80.107	0.884	0.803	72.323	0.794	0.729	76.623	0.843	0.767
IRS-SSL-Pop (Ours)	10.322	82.817	0.885	0.829	72.923	0.815	0.741	76.327	**0.861**	0.757
IRS-SSL-Ind (Ours)	**9.238**	**85.470**	**0.918**	**0.856**	**75.290**	**0.826**	**0.757**	**80.593**	0.856	**0.803**

	Tumor Grade			**AD/MCI/CN (ext)**			**Age (ext)**	**MCI/CN (ext)**			**Tumor Segmentation**
Pretext	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	L1↓	BAcc↑	AUC↑	F1↑	DCS↑
MAE	61.100	0.687	0.616	46.173	0.762	0.458	10.614	52.090	0.763	0.507	65.08
BarlowTwins	52.597	0.596	0.489	38.230	0.612	0.354	14.070	53.697	0.643	0.509	57.34
SimCLR	57.613	0.686	0.583	43.050	0.759	0.376	12.059	55.280	0.793	0.606	62.12
BarlowTwins (scratch)	51.297	0.595	0.475	35.413	0.749	0.342	13.565	60.780	0.719	0.537	59.91
SimCLR (scratch)	**64.863**	**0.749**	**0.668**	54.300	0.818	0.501	**10.130**	66.900	0.811	0.597	64.40
IRS-SSL-Pop (Ours)	63.040	0.677	0.624	57.050	0.801	0.495	11.674	74.007	0.858	0.663	66.02
IRS-SSL-Ind (Ours)	61.763	0.692	0.625	**64.433**	**0.834**	**0.605**	10.605	**75.067**	**0.893**	**0.718**	**66.60**

We first investigate whether grounding self-supervision in inter-regional structural covariance improves representation quality. To evaluate this, we compare IRS-SSL-Ind and IRS-SSL-Pop separately against representative SSL pretext strategies (Tab. 1 and Tab. 11), as well as SOTA biomedical SSL models (Tab. 2 and Tab. 3) under the same downstream adaptation setting. Against representative SSL objectives, at least one IRS-SSL formulation ranks in the top two on every reported metric, under both linear probing (Tab. 1) and fine-tuning (11), with IRS-SSL-Ind delivers the strongest overall performance profile. Compared to MAE, Barlow Twins, and SimCLR, IRS-SSL improves on all tasks.

For SOTA SSL biomedical models, we note that the evaluation benchmarks overlap with the pretraining datasets of AMAES (Munk et al., 2024), BrainIAC (Tak et al., 2026), DAMT (Kim et al., 2024), and 3DINO-ViT (Xu et al., 2025); therefore, we exclude these benchmarks from the corresponding tasks whenever dataset overlap occurs. As shown in Tab. 2, under linear probing, IRS-SSL-Ind achieves best or second-best performance on 16 of the 20 evaluation metrics, outperforming all evaluated prior models on 12. IRS-SSL-Pop achieves best or second-best performance on 9 metrics. Furthermore, under layer-decay fine-tuning in Tab. 3, IRS-SSL-Ind achieves best or second-best performance on 16 of the 20 evaluation metrics, outperforming all evaluated prior models on 15. IRS-SSL-Pop achieves best or second-best performance on 14 metrics, ranking first on 9, when each variant is compared independently against the prior models. These comparisons evaluate each variant independently against the prior models, including BrainIAC (Tak et al., 2026), and other strong SSL models. Overall, these results show that *grounding self-supervision in neuroanatomical structure consistently improves over standard SSL objectives.*

Finally, we examine whether the observed gains can be explained simply by the MAE warm-up rather than by the proposed inter-regional objectives (See Supp. E, Tab. 14). IRS-SSL-Ind achieves the best result on 18 of the 20 reported metrics.In particular, both IRS-SSL formulations substantially outperform the MAE controls on the external neurodegenerative classification tasks. These results indicate that continued MAE pretraining alone does not reproduce the gains obtained from inter-regional structural self-supervision.

3.3 IRS-SSL PRODUCES MORE BIOLOGICALLY MEANINGFUL ATTENTION PATTERNS

A pairwise forced-choice preference study was conducted with six certified radiologists. The presentation order of the attention maps was randomized, and radiologists were blinded to model identity. Ninety unique attention

Table 2: Comparison of SOTAs evaluated on downstream and external test tasks under linear probing. The best result per column is **bolded** and the second-best is underlined. BAcc = balanced accuracy (%).

	Age	Gender			MCI/CN			FTD/CN		
Model	L1↓	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑
AnatCL (Barbano et al., 2026)	13.134	80.460	0.889	0.807	**77.360**	**0.857**	**0.792**	74.950	0.845	0.727
DAMT (Kim et al., 2024)	11.348	79.140	0.899	0.794	-	-	-	**80.680**	**0.893**	0.794
BrainIAC (Tak et al., 2026)	10.701	51.822	0.597	0.424	-	-	-	79.110	0.880	0.778
BrainSegFounder (Cox et al., 2024)	15.270	77.520	0.858	0.775	62.540	0.695	0.619	73.720	0.793	0.732
MedicalNet (Chen et al., 2019)	15.370	78.890	0.888	0.793	62.020	0.660	0.623	62.780	0.692	0.626
3DINO-ViT (Xu et al., 2025)	-	-	-	-	-	-	-	79.620	0.865	0.793
AMAES (Munk et al., 2024)	15.506	78.890	0.884	0.793	-	-	-	77.020	0.827	0.761
3D-Neuro-SimCLR (Kaczmarek et al., 2025)	13.892	57.453	0.801	0.457	67.334	0.761	0.682	-	-	-
IRS-SSL-Pop (Ours)	10.322	82.817	0.885	0.829	72.923	0.815	0.741	76.327	0.861	0.757
IRS-SSL-Ind (Ours)	**9.238**	**85.471**	**0.918**	**0.856**	73.893	0.821	0.746	80.593	0.856	**0.803**

	Tumor Grade			AD/MCI/CN (ext)			Age (ext)	MCI/CN (ext)		
Model	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	L1↓	BAcc↑	AUC↑	F1↑
AnatCL (Barbano et al., 2026)	62.460	0.675	0.643	38.320	0.686	0.381	12.132	60.930	0.648	0.540
DAMT (Kim et al., 2024)	**65.120**	0.810	**0.688**	-	-	-	10.836	-	-	-
BrainIAC (Tak et al., 2026)	61.782	0.657	0.633	-	-	-	11.161	-	-	-
BrainSegFounder (Cox et al., 2024)	53.231	0.571	0.531	33.021	0.652	0.307	13.868	52.710	0.741	0.524
MedicalNet (Chen et al., 2019)	51.304	0.511	0.485	41.752	0.618	0.422	13.961	55.273	0.792	0.565
3DINO-ViT (Xu et al., 2025)	60.491	**0.819**	0.622	-	-	-	-	-	-	-
AMAES (Munk et al., 2024)	53.860	0.523	0.541	-	-	-	14.045	-	-	-
3D-Neuro-SimCLR (Kaczmarek et al., 2025)	53.890	0.537	0.538	-	-	-	-	58.600	0.796	0.604
IRS-SSL-Pop (Ours)	63.041	0.677	0.624	57.052	0.801	0.495	11.674	74.007	0.858	0.663
IRS-SSL-Ind (Ours)	61.763	0.692	0.625	**64.433**	**0.834**	**0.605**	**10.605**	**75.067**	**0.893**	**0.718**

Table 3: Comparison of SOTAs evaluated on downstream and external test tasks under finetuning with layer decay. The best result per column is **bolded** and the second-best is underlined. BAcc = balanced accuracy (%).

	Age	Gender			MCI/CN			FTD/CN		
Model	L1↓	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑
AnatCL (Barbano et al., 2026)	-	85.323	**0.960**	0.858	68.160	0.742	0.682	82.287	**0.921**	0.810
DAMT (Kim et al., 2024)	15.515	83.633	0.902	0.839	-	-	-	82.813	0.920	0.822
BrainIAC (Tak et al., 2026)	9.544	82.591	0.869	0.830	-	-	-	74.692	0.856	0.748
BrainSegFounder (Cox et al., 2024)	15.411	81.774	0.929	0.820	69.843	0.786	0.693	79.722	0.888	0.793
MedicalNet (Chen et al., 2019)	7.256	84.658	0.895	0.846	71.242	0.776	0.722	84.744	0.855	0.843
3DINO-ViT (Xu et al., 2025)	-	-	-	-	-	-	-	79.011	0.848	0.778
AMAES (Munk et al., 2024)	12.935	84.651	0.949	0.846	-	-	-	83.412	0.915	0.819
3D-Neuro-SimCLR (Kaczmarek et al., 2025)	11.192	78.332	0.840	0.785	67.541	0.744	0.681	-	-	-
IRS-SSL-Pop (Ours)	7.412	84.423	0.933	0.847	**77.980**	**0.853**	**0.789**	83.843	0.898	0.832
IRS-SSL-Ind (Ours)	**7.096**	**89.253**	0.955	**0.893**	75.890	0.840	0.769	**85.683**	0.897	**0.845**

	Tumor Grade			AD/MCI/CN (ext)			Age (ext)	MCI/CN (ext)		
Pretext	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	L1↓	BAcc↑	AUC↑	F1↑
AnatCL (Barbano et al., 2026)	72.337	0.821	0.753	50.433	0.814	0.495	-	58.217	0.789	0.561
DAMT (Kim et al., 2024)	74.527	**0.831**	**0.769**	-	-	-	14.319	-	-	-
BrainIAC (Tak et al., 2026)	61.160	0.743	0.632	-	-	-	10.160	-	-	-
BrainSegFounder (Cox et al., 2024)	**74.910**	0.795	0.735	50.510	0.730	0.454	14.096	54.490	0.555	0.443
MedicalNet (Chen et al., 2019)	64.386	0.792	0.647	41.388	0.657	0.348	9.908	43.953	0.646	0.451
3DINO-ViT (Xu et al., 2025)	69.020	0.779	0.702	-	-	-	-	-	-	-
AMAES (Munk et al., 2024)	74.210	0.802	0.702	-	-	-	12.606	-	-	-
3D-Neuro-SimCLR (Kaczmarek et al., 2025)	63.720	0.757	0.637	-	-	-	-	59.220	0.669	0.545
IRS-SSL-Pop (Ours)	69.927	0.796	0.731	53.670	0.837	0.493	9.864	**72.067**	0.791	0.675
IRS-SSL-Ind (Ours)	71.243	0.808	0.740	**67.310**	**0.848**	**0.614**	**9.453**	65.970	**0.847**	**0.703**

maps from the MCI/CN task were generated using a ViT-Based GradCAM from three model pairs IRS-SSL-Ind/BarlowTwins, IRS-SSL-Ind/SimCLR, and IRS-SSL-Ind/MAE. Responses from six certified radiologists were recorded and analysed (Tab. 4). Collectively, these findings demonstrate that IRS-SSL-Ind yields superior attention localization performance on the MCI versus CN task. Fig. 2 presents side-by-side Grad-CAM activation maps generated by models pretrained under different SSL objectives (MAE, SimCLR, Barlow Twins, and the proposed IRS-SSL-Ind). These maps were shown pairwise to clinicians for forced-choice preference assessment. Further details of the blinded clinician evaluation protocol, including the questionnaire design and evaluation criteria, are provided in Supp. F.

3.4 Architectural and biologically motivated design choices matter

We investigate the key design choices underlying IRS-SSL, including architectural realization, batch size, neuroanatomical parcellation, regional feature design, and the selection of population-level structural relationships. We first compare the image-space and embedding-space realizations introduced in Sec. 2.3. Using IRS-SSL-Ind, the

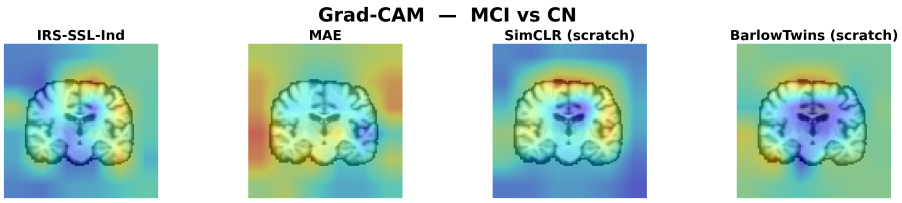


Figure 2: **Model attention analysis.** Grad-CAM attention maps from models pretrained with different SSL objectives. The maps were presented pairwise to clinicians for forced-choice preference assessment.

Table 4: Attention-map preferences (two-sided exact binomial test against $p = 0.5$).

Baseline	Prefer IRS-SSL-Ind	p-value
MAE	65/90	2.97×10^{-5}
SimCLR	77/90	3.18×10^{-12}
Barlow Twins	73/90	1.95×10^{-9}

image-space realization outperforms the embedding-space alternative across all six evaluated downstream tasks and all 16 reported metrics (Tab. 5). These results motivate our use of the image-space realization as the default architecture in all subsequent experiments.

We then ablate the remaining design choices. Specifically, for IRS-SSL-Ind, we ablate minibatch size and neuroanatomical atlas selection (Tabs. 6). For IRS-SSL-Pop, we additionally ablate the region-pairing strategy used to define structural covariance and the choice of radiomic feature statistics (Tab. 7). Across both formulations, performance is stable across batch sizes, with bs = 64 giving the best overall trade-off between statistical stability and efficiency. Both the Harvard-Oxford (Rushmore et al., 2022) and Desikan-Killiany (Desikan et al., 2006) atlases support strong performance, confirming IRS-SSL is not tied to a specific parcellation; Harvard-Oxford is generally stronger for IRS-SSL-Ind, while the pattern is more mixed for IRS-SSL-Pop, indicating that anatomically meaningful parcellation matters more than any single atlas choice.

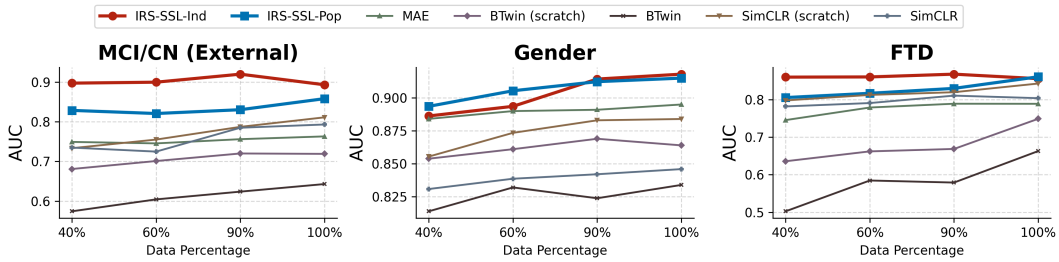


Figure 3: **Effect of downstream dataset size on representation transfer performance.** We evaluate SimCLR, MAE, and IRS-SSL under varying amounts of labeled downstream training data. Across neurodegenerative classification tasks, IRS-SSL maintains stronger or more stable performance than standard SSL objectives across data regimes.

For IRS-SSL-Pop specifically, restricting covariance to homologous left–right region pairs outperforms both full pairwise covariance and random pairing on most tasks, confirming that biologically motivated pairing yields a more effective population-level signal than modeling all or arbitrary region relationships. Similarly, the default low-order radiomic descriptor set (mean, variance, entropy) matches or outperforms the extended set with added skewness and kurtosis on most tasks, while reducing pretraining computation—showing that simple intensity statistics are sufficient to capture informative structural covariance. *Overall, moderate batch sizes, anatomically meaningful parcellation, homologous region pairing, and low-order radiomic features each contribute to IRS-SSL's effectiveness, though their individual impact is task-dependent.*

Table 5: Comparison of IRS-SSL-Ind and its embedding-space realization. The best result per column is **bolded**. BAcc = balanced accuracy (%).

	AD/MCI/CN (ext)			Age	Gender			MCI/CN			FTD/CN			Tumor Grade		
Model	BAcc↑	AUC↑	F1↑	L1↓	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑
IRS-SSL-Ind	**64.433**	**0.834**	**0.605**	**9.238**	**85.470**	**0.918**	**0.856**	**75.290**	**0.826**	**0.757**	**80.593**	**0.856**	**0.803**	**59.130**	**0.692**	**0.599**
IRS-SSL-Ind (embedding-space realization)	51.231	0.787	0.489	10.228	80.460	0.882	0.807	70.111	0.789	0.711	76.141	0.803	0.749	57.823	0.628	0.586

Table 6: Ablation study of IRS-SSL-Ind design choices across downstream tasks. † denotes the default IRS-SSL-Ind configuration; the best result within each section is shown in **bold**.

	AD/MCI/CN (ext)			Age	Gender			MCI/CN			FTD			Tumor Grade		
Config	BAcc↑	AUC↑	F1↑	L1↓	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑
(a) Batch size																
bs = 32	45.450	0.781	0.452	**8.963**	83.370	**0.920**	0.837	73.625	0.822	0.752	74.370	0.828	0.738	55.525	0.705	0.554
bs = 64†	**64.433**	0.834	**0.605**	9.238	**85.470**	0.918	**0.856**	**75.290**	**0.826**	**0.757**	**80.593**	0.856	**0.803**	59.130	0.692	0.599
bs = 128	56.915	**0.836**	0.552	9.153	79.550	0.899	0.799	73.130	0.742	0.735	78.510	**0.861**	0.775	**59.805**	**0.728**	**0.609**
(b) Brain atlas																
Desikan-Killiany	56.930	0.805	0.532	**8.765**	82.670	**0.923**	0.830	73.483	**0.837**	0.748	76.860	**0.866**	0.760	**61.123**	0.680	**0.618**
Harvard-Oxford†	**64.433**	**0.834**	**0.605**	9.238	**85.470**	0.918	**0.856**	**75.290**	0.826	**0.757**	**80.593**	0.856	**0.803**	59.130	**0.692**	0.599

Table 7: Ablation study of IRS-SSL-Pop design choices across downstream tasks. Radiomic statistics: μ (mean), σ^2 (variance), Ent (entropy), Skew (skewness), and Kurt (kurtosis). † denotes the default IRS-SSL-Pop configuration; the best result within each section is shown in **bold**.

	AD/MCI/CN (ext)			Age	Gender			MCI/CN			FTD			Tumor Grade		
Config	BAcc↑	AUC↑	F1↑	L1↓	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑
(a) Batch size																
bs = 32	45.763	0.728	0.433	10.771	77.623	0.866	0.776	**73.380**	0.794	0.739	75.883	0.846	0.755	**64.620**	**0.768**	**0.655**
bs = 64†	**57.050**	**0.801**	**0.495**	**10.322**	**82.817**	**0.885**	**0.829**	72.923	**0.815**	**0.741**	**76.327**	**0.861**	**0.757**	63.040	0.677	0.624
bs = 128	41.960	0.712	0.410	10.511	77.957	0.864	0.779	70.173	0.771	0.705	75.860	0.820	0.752	57.777	0.613	0.577
(b) Brain atlas																
Desikan-Killiany	54.830	**0.850**	**0.562**	**9.629**	78.580	0.863	0.786	67.390	0.813	0.674	74.040	0.844	0.732	**66.390**	**0.774**	**0.677**
Harvard-Oxford†	**57.050**	0.801	0.495	10.322	**82.817**	**0.885**	**0.829**	**72.923**	**0.815**	**0.741**	**76.327**	**0.861**	**0.757**	63.040	0.677	0.624
(c) Pairing strategy																
IRS-SSL-Pop†	**57.050**	**0.801**	**0.495**	10.322	**82.817**	**0.885**	**0.829**	72.923	**0.815**	0.741	**76.327**	**0.861**	**0.757**	**63.040**	0.677	**0.624**
IRS-SSL-Pop (full cov.)	51.230	0.787	0.489	10.228	74.640	0.880	0.748	66.390	0.774	0.671	76.140	0.803	0.749	57.820	0.628	0.586
IRS-SSL-Pop (random pair)	44.240	0.792	0.374	**9.780**	72.820	0.846	0.730	**75.030**	0.793	**0.743**	66.160	0.689	0.646	51.930	**0.679**	0.509
(d) Radiomic statistics																
μ, σ^2, Ent†	57.050	**0.801**	0.495	10.322	**82.817**	**0.885**	**0.829**	**72.923**	**0.815**	**0.741**	**76.327**	**0.861**	**0.757**	**63.040**	0.677	**0.624**
μ, σ^2, Ent, Skew, Kurt	**57.700**	0.797	**0.540**	**9.875**	73.570	0.875	0.737	70.140	0.791	0.705	74.550	0.813	0.740	53.230	**0.723**	0.531

3.5 IRS-SSL IMPROVES ROBUSTNESS AND DATA EFFICIENCY

We further evaluate the robustness and data efficiency of IRS-SSL by varying the amount of labeled downstream training data from 40% to 100% and comparing against representative reconstruction- and contrastive-based SSL baselines (Fig. 3). Across these data regimes, IRS-SSL maintains stronger or more stable downstream performance than MAE, Barlow Twins, and SimCLR. We additionally evaluate an extreme few-shot setting with only $n = 5$ labeled samples (Appendix E, Tab. 15), where one of the two IRS-SSL formulations achieves the highest balanced accuracy on all six evaluated classification tasks.

4 CONCLUSION

In this work, we introduced IRS-SSL, a framework that grounds self-supervised task design in neuroanatomical principles for structural brain MRI. Unlike conventional SSL objectives that primarily operate at the level of individual samples or instance-wise representations, IRS-SSL models biologically meaningful cross-sample structural relationships through inter-regional covariance consistency. Extensive experiments across diverse downstream tasks demonstrate that IRS-SSL consistently improves representation quality over representative SSL paradigms and achieves state-of-the-art performance among structural brain MRI representation learning methods. Qualitative analyses further show improved biological grounding in the learned representations. Overall, this work provides evidence that grounding SSL task design in domain-specific scientific structure can improve representation learning in scientifically structured domains beyond general-purpose objectives.

Limitation. As inter-regional structural organization is a higher-level relational property, IRS-SSL currently benefits from a brief warm-up to initialize basic voxel-level structural knowledge. Our study is limited to structural MRI and a specific set of anatomically defined inter-regional relationships; extending IRS-SSL to other forms of brain organization, imaging modalities, and more heterogeneous populations remains an important direction. From an ethical perspective, performance may vary across acquisition sites and demographic groups, and the learned representations should not be used for clinical decision-making without appropriate external and prospective validation.

AI USE STATEMENT

Generative AI tools were used solely to edit and improve the clarity, grammar, and readability of author-written text. They were not used to generate research ideas, design the methodology or experiments, implement the proposed methods, analyze data, or interpret experimental results. All AI-assisted edits were reviewed by the authors to ensure technical accuracy and fidelity to the intended meaning. We take full responsibility for the final content of this work, including any text produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The full codebase for IRS-SSL, including training scripts and evaluation protocols, is available at the following anonymous repository: `https://anonymous.4open.science/r/IRS-SSN-108E`.

REFERENCES

Aaron Alexander-Bloch, Jay N Giedd, and Ed Bullmore. Imaging structural co-variance between human brain regions. *Nature reviews neuroscience*, 14(5):322–336, 2013.

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

Carlo Alberto Barbano, Matteo Brunello, Benoit Dufumier, and Marco Grangetto. Anatomical foundation models for brain mris. *Pattern Recognition Letters*, 199:178–184, 2026. ISSN 0167-8655. doi: https://doi.org/10.1016/j.patrec.2025.11.028. URL `https://www.sciencedirect.com/science/article/pii/S0167865525003848`.

Adrien Bardes, Quentin Garrido, Jean Ponce, Xinlei Chen, Michael Rabbat, Yann LeCun, Mido Assran, and Nicolas Ballas. Revisiting feature prediction for learning visual representations from video. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL `https://openreview.net/forum?id=QaCCuDfBk2`. Featured Certification.

Duane L Beekly, Erin M Ramos, William W Lee, Woodrow D Deitrich, Mary E Jacka, Joylee Wu, Janene L Hubbard, Thomas D Koepsell, John C Morris, Walter A Kukull, et al. The national alzheimer's coordinating center (nacc) database: the uniform data set. *Alzheimer Disease & Associated Disorders*, 21(3):249–258, 2007.

Jesse A Brown, Alex J Lee, Kristen Fernhoff, Taylor Pistone, Lorenzo Pasquini, Amy B Wise, Adam M Staffaroni, Maria Luisa Mandelli, Suzee E Lee, Adam L Boxer, et al. Functional network collapse in neurodegenerative disease. *Nature communications*, 16(1):10273, 2025.

Evan Calabrese, Javier Villanueva-Meyer, Jeffrey Rudie, Andreas Rauschecker, Ujjwal Baid, Spyridon Bakas, Soonmee Cha, John Mongan, and Christopher Hess. The university of california san francisco preoperative diffuse glioma MRI (UCSF-PDGM) (version 5), 2022. URL `https://doi.org/10.7937/tcia.bdgf-8v37`.

Jona Carmon, Jil Heege, Joe H. Necus, Thomas W. Owen, Gordon Pipa, Marcus Kaiser, Peter N. Taylor, and Yujiang Wang. Reliability and comparability of human brain structural covariance networks. *NeuroImage*, 220:117104, 2020. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2020.117104. URL `https://www.sciencedirect.com/science/article/pii/S1053811920305905`.

Josue Ortega Caro, Antonio Henrique de Oliveira Fonseca, Syed A Rizvi, Matteo Rosati, Christopher Averill, James L Cross, Prateek Mittal, Emanuele Zappala, Rahul Madhav Dhodapkar, Chadi Abdallah, et al. Brainlm: A foundation model for brain activity recordings. In *The Twelfth International Conference on Learning Representations*.

Dong Chao, Thalamuthu Anbupalam, Jiang Jiyang, Karen A. Mather, Perminder S. Sachdev, and Wei Wen. Brain structural covariances in the ageing brain in the uk biobank. *Brain structure & function*, 229(5):1165–1177, 2024. doi: 10.1007/s00429-024-02794-4.

Sihong Chen, Kai Ma, and Yefeng Zheng. Med3d: Transfer learning for 3d medical image analysis, 2019. URL `https://arxiv.org/abs/1904.00625`.

Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 1597–1607. PMLR, 13–18 Jul 2020. URL `https://proceedings.mlr.press/v119/chen20j.html`.

DL Collins. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, 47: S102, 2009.

Joseph Cox, Peng Liu, Skylar E. Stolte, Yunchao Yang, Kang Liu, Kyle B. See, Huiwen Ju, and Ruogu Fang. Brainsegfounder: Towards 3d foundation models for neuroimage segmentation. *Medical Image Analysis*, 97: 103301, 2024. ISSN 1361-8415. doi: https://doi.org/10.1016/j.media.2024.103301. URL https://www.sciencedirect.com/science/article/pii/S1361841524002263.

Maria Correia de Verdier, Rachit Saluja, Louis Gagnon, Dominic LaBella, Ujjwall Baid, Nourel Hoda Tahon, Martha Foltyn-Dumitru, Jikai Zhang, Maram Alafif, Saif Baig, Ken Chang, Gennaro D'Anna, Lisa Deptula, Diviya Gupta, Muhammad Ammar Haider, Ali Hussain, Michael Iv, Marinos Kontzialis, Paul Manning, Farzan Moodi, Teresa Nunes, Aaron Simon, Nico Sollmann, David Vu, Maruf Adewole, Jake Albrecht, Udunna Anazodo, Rongrong Chai, Verena Chung, Shahriar Faghani, Keyvan Farahani, Anahita Fathi Kazerooni, Eugenio Iglesias, Florian Kofler, Hongwei Li, Marius George Linguraru, Bjoern Menze, Ahmed W. Moawad, Yury Velichko, Benedikt Wiestler, Talissa Altes, Patil Basavasagar, Martin Bendszus, Gianluca Brugnara, Jaeyoung Cho, Yaseen Dhemesh, Brandon K. K. Fields, Filip Garrett, Jaime Gass, Lubomir Hadjiiski, Jona Hattangadi-Gluth, Christopher Hess, Jessica L. Houk, Edvin Isufi, Lester J. Layfield, George Mastorakos, John Mongan, Pierre Nedelec, Uyen Nguyen, Sebastian Oliva, Matthew W. Pease, Aditya Rastogi, Jason Sinclair, Robert X. Smith, Leo P. Sugrue, Jonathan Thacker, Igor Vidic, Javier Villanueva-Meyer, Nathan S. White, Mariam Aboian, Gian Marco Conte, Anders Dale, Mert R. Sabuncu, Tyler M. Seibert, Brent Weinberg, Aly Abayazeed, Raymond Huang, Sevcan Turk, Andreas M. Rauschecker, Nikdokht Farid, Philipp Vollmuth, Ayman Nada, Spyridon Bakas, Evan Calabrese, and Jeffrey D. Rudie. The 2024 brain tumor segmentation (brats) challenge: Glioma segmentation on post-treatment mri, 2024. URL https://arxiv.org/abs/2405.18368.

Rahul S Desikan, Florent Ségonne, Bruce Fischl, Brian T Quinn, Bradford C Dickerson, Deborah Blacker, Randy L Buckner, Anders M Dale, R Paul Maguire, Bradley T Hyman, et al. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. *Neuroimage*, 31(3):968–980, 2006.

Zijian Dong, Ruilin Li, Yilei Wu, Thuan Tinh Nguyen, Joanna Chong, Fang Ji, Nathanael Tong, Christopher Chen, and Juan Helen Zhou. Brain-jepa: Brain dynamics foundation model with gradient positioning and spatiotemporal masking. *Advances in Neural Information Processing Systems*, 37:86048–86073, 2024.

Nupur Ghoshal and Nigel J Cairns. Unravelling the mysteries of frontotemporal dementia. *Missouri Medicine*, 110 (5):411, 2013.

Ellen Grober, Richard B Lipton, Reisa A Sperling, Kathryn V Papp, Keith A Johnson, Dorene M Rentz, Amy E Veroff, Paul S Aisen, and Ali Ezzati. Associations of stages of objective memory impairment with amyloid pet and structural mri: The a4 study. *Neurology*, 98(13):e1327–e1336, 2022.

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

Fabian Isensee, Marianne Schell, Irada Pflueger, Gianluca Brugnara, David Bonekamp, Ulf Neuberger, Antje Wick, Heinz-Peter Schlemmer, Sabine Heiland, Wolfgang Wick, Martin Bendszus, Klaus H. Maier-Hein, and Philipp Kickingereder. Automated brain extraction of multisequence mri using artificial neural networks. *Human Brain Mapping*, 40(17):4952–4964, 2019. doi: https://doi.org/10.1002/hbm.24750. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hbm.24750.

IXI Consortium. Ixi dataset. https://brain-development.org/ixi-dataset/, 2004. Brain Development Dataset, Imperial College London and collaborators.

Yankai Jiang, Mingze Sun, Heng Guo, Xiaoyu Bai, Ke Yan, Le Lu, and Minfeng Xu. Anatomical invariance modeling and semantic alignment for self-supervised learning in 3d medical image analysis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 15859–15869, October 2023.

Emily Kaczmarek, Justin Szeto, Brennan Nichyporuk, and Tal Arbel. Building a general simclr self-supervised foundation model across neurological diseases to advance 3d brain mri diagnoses. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 1310–1319, 2025.

Jonghun Kim, Mansu Kim, and Hyunjin Park. Domain aware multi-task pre-training of 3d swin transformer for brain mri. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, pp. 2124–2144, December 2024.

Pamela J LaMontagne, Tammie LS Benzinger, John C Morris, Sarah Keefe, Russ Hornbeck, Chengjie Xiong, Elizabeth Grant, Jason Hassenstab, Krista Moulder, Andrei G Vlassenko, et al. Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease. *medrxiv*, pp. 2019–12, 2019.

Thomas Z Li, Aravind R Krishnan, Lianrui Zuo, John M Still, Kim L Sandler, Fabien Maldonado, Thomas A Lasko, and Bennett A Landman. Self-supervised learning of imaging and clinical signatures using a multimodal joint-embedding predictive architecture. *arXiv preprint arXiv:2509.15470*, 2025.

Han Liu, Zhe Ma, Lijiang Wei, Zhenpeng Chen, Yun Peng, Zhicheng Jiao, Harrison Bai, and Bin Jing. A radiomics-based brain network in t1 images: construction, attributes, and applications. *Cerebral Cortex*, 34(2):bhae016, 2024.

Zixuan Liu, Hanwen Xu, Addie Woicik, Linda G. Shapiro, Marian Blazes, Yue Wu, Verena Steffen, Catherine A. Cukras, Cecilia S. Lee, Miao Zhang, Aaron Y. Lee, and Sheng Wang. A three-dimensional multi-modal foundation model for optical coherence tomography. *Nature Biomedical Engineering*, 2026. doi: 10.1038/s41551-026-01662-2.

Kenneth Marek, Danna Jennings, Shirley Lasch, Andrew Siderowf, Caroline Tanner, Tanya Simuni, Chris Coffey, Karl Kieburtz, Emily Flagg, Sohini Chowdhury, Werner Poewe, Brit Mollenhauer, Paracelsus-Elena Klinik, Todd Sherer, Mark Frasier, Claire Meunier, Alice Rudolph, Cindy Casaceli, John Seibyl, Susan Mendick, Norbert Schuff, Ying Zhang, Arthur Toga, Karen Crawford, Alison Ansbach, Pasquale De Blasio, Michele Piovella, John Trojanowski, Les Shaw, Andrew Singleton, Keith Hawkins, Jamie Eberling, Deborah Brooks, David Russell, Laura Leary, Stewart Factor, Barbara Sommerfeld, Penelope Hogarth, Emily Pighetti, Karen Williams, David Standaert, Stephanie Guthrie, Robert Hauser, Holly Delgado, Joseph Jankovic, Christine Hunter, Matthew Stern, Baochan Tran, Jim Leverenz, Marne Baca, Sam Frank, Cathi-Ann Thomas, Irene Richard, Cheryl Deeley, Linda Rees, Fabienne Sprenger, Elisabeth Lang, Holly Shill, Sanja Obradov, Hubert Fernandez, Adrienna Winters, Daniela Berg, Katharina Gauss, Douglas Galasko, Deborah Fontaine, Zoltan Mari, Melissa Gerstenhaber, David Brooks, Sophie Malloy, Paolo Barone, Katia Longo, Tom Comery, Bernard Ravina, Igor Grachev, Kim Gallagher, Michelle Collins, Katherine L. Widnell, Suzanne Ostrowizki, Paulo Fontoura, Tony Ho, Johan Luthman, Marcel van der Brug, Alastair D. Reith, and Peggy Taylor. The parkinson progression marker initiative (ppmi). *Progress in Neurobiology*, 95(4):629–635, 2011. ISSN 0301-0082. doi: https://doi.org/10.1016/j.pneurobio.2011.09.005. URL https://www.sciencedirect.com/science/article/pii/S0301008211001651. Biological Markers for Neurodegenerative Diseases.

Moona Mazher, Geoff J. M. Parker, and Daniel C. Alexander. Towards generalisable foundation models for brain mri, 2025. URL https://arxiv.org/abs/2510.23415.

Sha Abbas Mollah, Aditya Nayak, Swapna Barhai, and Uddip Maity. A comprehensive review on frontotemporal dementia: its impact on language, speech and behavior. *Dementia & Neuropsychologia*, 18:e20230072, 2024.

Neus Mongay-Ochoa, Gabriel Gonzalez-Escamilla, Vinzenz Fleischer, Deborah Pareto, Àlex Rovira, Jaume Sastre-Garriga, and Sergiu Groppa. Structural covariance analysis for neurodegenerative and neuroinflammatory brain disorders. *Brain*, pp. awaf151, 2025.

Asbjørn Munk, Jakob Ambsdorf, Sebastian Llambias, and Mads Nielsen. Amaes: Augmented masked autoencoder pretraining on public brain mri data for 3d-native segmentation. *arXiv preprint arXiv:2408.00640*, 2024.

Salvatore Nigro, Marco Filardi, Benedetta Tafuri, Roberto De Blasi, Maria Teresa Dell'Abate, Alessia Giugno, Valentina Gnoni, Giammarco Milella, Daniele Urso, Chiara Zecca, Stefano Zoccolella, and Giancarlo Logroscino. Radiomics feature similarity: A novel approach for characterizing brain network changes in patients with behavioral variant frontotemporal dementia. *NeuroImage: Clinical*, 46:103780, 2025. doi: 10.1016/j.nicl.2025.103780.

Ronald Carl Petersen, Paul S Aisen, Laurel A Beckett, Michael C Donohue, Anthony Collins Gamst, Danielle J Harvey, CR Jack Jr, William J Jagust, Leslie M Shaw, Arthur W Toga, et al. Alzheimer's disease neuroimaging initiative (adni) clinical characterization. *Neurology*, 74(3):201–209, 2010.

Jianing Qiu, Jian Wu, Hao Wei, Peilun Shi, Minqing Zhang, Yunyun Sun, Lin Li, Hanruo Liu, Hongyi Liu, Simeng Hou, Yuyang Zhao, Xuehui Shi, Junfang Xian, Xiaoxia Qu, Sirui Zhu, Lijie Pan, Xiaoniao Chen, Xiaojia Zhang, Shuai Jiang, Kebing Wang, Chenlong Yang, Mingqiang Chen, Sujie Fan, Jianhua Hu, Aiguo Lv, Hui Miao, Li Guo, Shujun Zhang, Cheng Pei, Xiaojuan Fan, Jianqin Lei, Ting Wei, Junguo Duan, Chun Liu, Xiaobo Xia, Siqi Xiong, Junhong Li, Kyle Lam, Benny Lo, Yih Chung Tham, Tien Yin Wong, Ningli Wang, and Wu Yuan. Development and validation of a multimodal multitask vision foundation model for generalist ophthalmic artificial intelligence. *NEJM AI*, 1(12):AIoa2300221, 2024. doi: 10.1056/AIoa2300221. URL https://ai.nejm.org/doi/full/10.1056/AIoa2300221.

Rosebud O. Roberts, Yonas E. Geda, David S. Knopman, Ruth H. Cha, V. Shane Pankratz, Bradley F. Boeve, Robert J. Ivnik, Eric G. Tangalos, Ronald C. Petersen, and Walter A. Rocca. The mayo clinic study of aging: Design and sampling, participation, baseline measures and sample characteristics. *Neuroepidemiology*, 30(1):58–69, 02 2008. ISSN 0251-5350. doi: 10.1159/000115751. URL https://doi.org/10.1159/000115751.

Jonathan D Rohrer and Howard J Rosen. Neuroimaging in frontotemporal dementia. *International Review of Psychiatry*, 25(2):221–229, 2013.

Chris Rorden and John Absher. Stroke outcome optimization project (soop), 2023. URL https://doi.org/10.18112/openneuro.ds004889.v1.0.0. BIDS Version 1.6.0, CC0 license.

R Jarrett Rushmore, Kyle Sunderland, Holly Carrington, Justine Chen, Michael Halle, Andras Lasso, Elizabeth Rizzoni, Brynn Vessey, et al. Anatomically curated segmentation of human subcortical structures in high resolution magnetic resonance imaging: An open science approach. *Frontiers in Neuroanatomy*, 16:894606, 2022.

Mohammadreza Shakouri, Fatemeh Iranmanesh, and Mahdi Eftekhari. Dino-cxr: A self supervised method based on vision transformer for chest x-ray classification. In George Bebis, Golnaz Ghiasi, Yi Fang, Andrei Sharf, Yue Dong, Chris Weaver, Zhicheng Leo, Joseph J LaViola Jr., and Luv Kohli (eds.), *Advances in Visual Computing*, pp. 320–331. Springer Nature Switzerland, 2023. ISBN 978-3-031-47966-3. doi: 10.1007/978-3-031-47966-3_25.

Divyanshu Tak, Biniam A. Garomsa, Anna Zapaishchykova, Tafadzwa L. Chaunzwa, Juan Carlos Climent Pardo, Zezhong Ye, John Zielke, Yashwanth Ravipati, Suraj Pai, Sri Vajapeyam, Maryam Mahootiha, Mitchell Parker, Luke R. G. Pike, Ceilidh Smith, Ariana M. Familiar, Kevin X. Liu, Sanjay Prabhu, Omar Arnaout, Pratiti Bandopadhayay, Ali Nabavizadeh, Sabine Mueller, Hugo JWL Aerts, Raymond Y. Huang, Tina Y. Poussaint, and Benjamin H. Kann. A generalizable foundation model for analysis of human brain MRI. *Nature Neuroscience*, 29(4):945–956, 2026. ISSN 1546-1726. doi: 10.1038/s41593-026-02202-6. URL https://doi.org/10.1038/s41593-026-02202-6.

Zhan Tong, Yibing Song, Jue Wang, and Limin Wang. Videomae: Masked autoencoders are data-efficient learners for self-supervised video pre-training. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 10078–10093. Curran Associates, Inc., 2022.

Jeya Maria Jose Valanarasu, Yucheng Tang, Dong Yang, Ziyue Xu, Can Zhao, Wenqi Li, Vishal M. Patel, Bennett Allan Landman, Daguang Xu, Yufan He, and Vishwesh Nath. Disruptive autoencoders: Leveraging low-level features for 3d medical image pre-training. In Ninon Burgos, Caroline Petitjean, Maria Vakalopoulou, Stergios Christodoulidis, Pierrick Coupe, Hervé Delingette, Carole Lartizien, and Diana Mateus (eds.), *Proceedings of The 7nd International Conference on Medical Imaging with Deep Learning*, volume 250 of *Proceedings of Machine Learning Research*, pp. 1553–1570. PMLR, 03–05 Jul 2024. URL https://proceedings.mlr.press/v250/valanarasu24a.html.

David C. Van Essen, Stephen M. Smith, Deanna M. Barch, Timothy E.J. Behrens, Essa Yacoub, and Kamil Ugurbil. The wu-minn human connectome project: An overview. *NeuroImage*, 80:62–79, 2013. ISSN 1053-8119. doi: https://doi.org/10.1016/j.neuroimage.2013.05.041. URL https://www.sciencedirect.com/science/article/pii/S1053811913005351. Mapping the Connectome.

Dongtao Wei, Kaixiang Zhuang, Lei Ai, Qunlin Chen, Wenjing Yang, Wei Liu, Kangcheng Wang, Jiangzhou Sun, and Jiang Qiu. Structural and functional brain scans from the cross-sectional southwest university adult lifespan dataset. *Scientific data*, 5(1):180134, 2018.

Tony Xu, Sepehr Hosseini, Chris Anderson, Anthony Rinaldi, Rahul G Krishnan, Anne L Martel, and Maged Goubran. A generalizable 3d framework and model for self-supervised learning in medical imaging. *npj Digital Medicine*, 8(1):639, 2025.

Yueyang Xu, Ashish Raj, Jonathan Victor, and Alzheimer's Disease Neuroimaging Initiative. Systematic differences between visually-relevant global and local image statistics of brain mri and natural scenes. *bioRxiv*, pp. 271338, 2018.

Guangqian Yang, Kangrui Du, Zhihan Yang, Ye Du, Eva Yi Wah Cheung, Yongping Zheng, Mo Yang, Zoe Kourtzi, Carola-Bibiane Schönlieb, Shujun Wang, and Alzheimer's Disease Neuroimaging Initiative. Adfound: A foundation model for diagnosis and prognosis of alzheimer's disease. *IEEE Journal of Biomedical and Health Informatics*, 29(11):8395–8408, 2025. doi: 10.1109/JBHI.2025.3576436.

Kai Yu, Yang Zhou, Yang Bai, Zhi Da Soh, Xinxing Xu, Rick Siow Mong Goh, Ching-Yu Cheng, and Yong Liu. UrFound: Towards Universal Retinal Foundation Models via Knowledge-Guided Masked Modeling . In *proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024*, volume LNCS 15012. Springer Nature Switzerland, October 2024.

Kun Zhao, Qiang Zheng, Tongtong Che, Dyrba Martin, Qiongling Li, Yanhui Ding, Yuanjie Zheng, Yong Liu, and Shuyu Li. Regional radiomics similarity networks (r2sn) in the human brain: reproducibility, small-world and biological basis. *bioRxiv*, 2020. doi: 10.1101/2020.12.09.418509.

Yukun Zhou, Mark A. Chia, Siegfried K. Wagner, Murat S. Ayhan, Dominic J. Williamson, Robbert R. Struyven, Timing Liu, Moucheng Xu, Mateo G. Lozano, Peter Woodward-Court, Yuka Kihara, Naomi Allen, John E.J. Gallacher, Thomas Littlejohns, Tariq Aslam, Paul Bishop, Graeme Black, Panagiotis Sergouniotis, Denize Atan, Andrew D. Dick, Cathy Williams, Sarah Barman, Jenny H. Barrett, Sarah Mackie, Tasanee Braithwaite, Roxana O. Carare, Sarah Ennis, Jane Gibson, Andrew J. Lotery, Jay Self, Usha Chakravarthy, Ruth E. Hogg, Euan Paterson, Jayne Woodside, Tunde Peto, Gareth Mckay, Bernadette Mcguinness, Paul J. Foster, Konstantinos Balaskas, Anthony P. Khawaja, Nikolas Pontikos, Jugnoo S. Rahi, Gerassimos Lascaratos, Praveen J. Patel, Michelle Chan, Sharon Y.L. Chua, Alexander Day, Parul Desai, Cathy Egan, Marcus Fruttiger, David F. Garway-Heath, Alison Hardcastle, Sir Peng T. Khaw, Tony Moore, Sobha Sivaprasad, Nicholas Strouthidis, Dhanes Thomas, Adnan Tufail, Ananth C. Viswanathan, Bal Dhillon, Tom Macgillivray, Cathie Sudlow, Veronique Vitart, Alexander Doney, Emanuele Trucco, Jeremy A. Guggeinheim, James E. Morgan, Chris J. Hammond, Katie Williams, Pirro Hysi, Simon P. Harding, Yalin Zheng, Robert Luben, Phil Luthert, Zihan Sun, Martin McKibbin, Eoin O'Sullivan, Richard Oram, Mike Weedon, Chris G. Owen, Alicja R. Rudnicka, Naveed Sattar, David Steel, Irene Stratton, Robyn Tapp, Max M. Yates, Axel Petzold, Savita Madhusudhan, Andre Altmann, Aaron Y. Lee, Eric J. Topol, Alastair K. Denniston, Daniel C. Alexander, and Pearse A. Keane. A foundation model for generalizable disease detection from retinal images. *Nature 2023 622:7981*, 622:156–163, 9 2023. ISSN 1476-4687. doi: 10.1038/s41586-023-06555-x. URL https://www.nature.com/articles/s41586-023-06555-x.

Jiaxin Zhuang, Linshan Wu, Qiong Wang, Peng Fei, Varut Vardhanabhuti, Lin Luo, and Hao Chen. Mim: Mask in mask self-supervised pre-training for 3d medical image analysis. *IEEE Transactions on Medical Imaging*, 44(9): 3727–3740, 2025. doi: 10.1109/TMI.2025.3564382.

A Code and Implementation Details

The full codebase for IRS-SSL, including training scripts and evaluation protocols, is available at the following anonymous repository: `https://anonymous.4open.science/r/IRS-SSN-108E`.

B Related Work

Anatomy-agnostic SSL for 3D brain MRI. A major line of work adapts general-purpose SSL to volumetric brain MRI without explicitly using neuroanatomical information in the pretraining objective. Contrastive approaches include BrainIAC Tak et al. (2026) and 3D-Neuro-SimCLR Kaczmarek et al. (2025), both of which extend SimCLR-style learning to large collections of 3D brain MRI. BrainSegFounder Cox et al. (2024) uses a two-stage pretraining scheme built on standard volumetric pretext tasks, including masked volume inpainting, rotation prediction, and contrastive coding. Reconstruction-based methods such as AMAES Munk et al. (2024) use masked autoencoding together with augmentation reversal for large-scale brain MRI pretraining. Related 3D foundation models include 3DINO-ViT Xu et al. (2025), which extends DINO-style self-distillation to volumetric medical imaging. Although these methods differ in architecture and pretext task, neuroanatomical structure is learned implicitly from the images rather than introduced as an explicit supervisory signal.

Anatomy-aware SSL for 3D brain MRI. Only limited work has explicitly incorporated neuroanatomical information into pretraining. AnatCL Barbano et al. (2026) extends weakly supervised contrastive learning by using age and region-wise anatomical measurements, including cortical thickness, gray-matter volume, and surface area, to modulate the degree of positiveness between subjects. Thus, anatomical information guides subject-level alignment, while contrastive learning remains its primary supervisory principle. DAMT Kim et al. (2024) adopts a multi-task framework built around conventional SSL pretext tasks, supplemented with anatomy-related prediction targets such as brain parcellation, cortical morphology, and regional radiomic features. Its supervision therefore remains substantially grounded in standard SSL objectives, with anatomical knowledge incorporated through additional prediction tasks.

These methods demonstrate the value of incorporating anatomical knowledge into representation learning. However, they primarily use anatomical properties at the voxel or region level to guide or supplement conventional SSL objectives; they do not explicitly model how different anatomical regions are structurally related. Despite its foundational role in brain organization and its demonstrated relevance to cognition, ageing, and neurodegenerative disease (Alexander-Bloch et al., 2013; Zhao et al., 2020; Liu et al., 2024; Chao et al., 2024; Mongay-Ochoa et al., 2025), inter-regional structural organization has not been explored as a primary self-supervisory principle for structural brain MRI. In this work, we address this gap by investigating whether this biologically and clinically meaningful principle can serve as the principal inductive bias for SSL.

C Dataset Description

Table 8: Pre-training dataset count

Data set	Count
A4 Grober et al. (2022)	1819
MCSA Roberts et al. (2008)	3095
NACC Beekly et al. (2007)	1626
OASIS LaMontagne et al. (2019)	424
PPMI Marek et al. (2011)	2401
SOOP Rorden & Absher (2023)	1393
WU1200 Van Essen et al. (2013)	1783
Total	12,541

C.1 Data Pre-processing

All datasets were processed through standardised preprocessing steps to ensure data consistency and quality control across all structural magnetic resonance imaging (MRI) scans. Raw DICOM images were first converted to NIfTI format using the `dcm2nii` package. To correct for low-frequency intensity inhomogeneities inherent to MRI acquisition, N4 bias field correction was applied using `SimpleITK`. All images were then skull stripped using the `HD-BET` framework Isensee et al. (2019) and finally rigidly registered to the MNI Collins (2009) standard brain space.

Table 9: Fine-tuning and evaluation dataset statistics

	Dataset	Task	Total
	ADNI Petersen et al. (2010)	AD/MCI/CN Classification	2549
		MCI/CN Classification	1969
	IXI IXI Consortium (2004)	Age Prediction and Gender Classification	562
	NIFD Rohrer & Rosen (2013)	FTD Classification	1466
	UCSF-PDGM Calabrese et al. (2022)	Tumour Grade	466
	BRATS-GLI de Verdier et al. (2024)	Tumour Segmentation	2700
External	OASIS-3 LaMontagne et al. (2019)	AD/MCI/CN Classification	251
		MCI/CN Classification	230
	SALD Wei et al. (2018)	Age Prediction	434
		Gender Classification	494

D Pretraining Details

In this section, we report our IRS-SSN pretraining details. Tab. 10 summarizes the main hyperparameters used during pretraining.

Transformers in medical MRI are increasingly used because their self-attention mechanism can capture both local tissue details and global anatomical context across entire 3D volumes. As such, following Zhuang et al. (2025), all methods use the same parameter initialization obtained from MAE pretraining for 50 epochs on the unlabeled pretraining dataset. All methods are pretrained on a single GPU, [NVIDIA RTX 6000 Ada Generation]

Table 10: NAISSL pretraining hyperparameters.

Category	NAISSL
Data	
Dataset	Multi-site T1-weighted MRI
Input resolution	96x96x96
Patch size	16x16x16
Sampling rate	1
Batch size	64
Model	
Backbone	ViT-B/16
Decoder depth	8
Decoder embed dim	576
Mask tokens	Enabled
Masking	
Mask type	random patch masking
Spatial scale	0.6
Temporal scale	1.0
Optimization	
Epochs	200
Optimizer	AdamW
Base learning rate	1.0×10^{-4}
Warm-up epochs	50
Weight decay	0.05

E ADDITIONAL RESULTS

Table 11: Comparison of SSL pretext strategies evaluated on downstream and external task performance of SSL pretext strategies evaluated under layer decay fine-tuning. The best result per column is **bolded** and the second-best is underlined. BAcc = balanced accuracy (%).

	Age	Gender			MCI/CN			FTD/CN		
Pretext	L1↓	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑
MAE	7.207	79.767	0.906	0.801	72.437	0.8	0.731	82.1	0.86	0.807
BarlowTwins	14.792	80.26	0.903	0.808	68.98	0.817	0.696	78.5	0.854	0.775
SimCLR	11.222	81.523	0.906	0.819	71.437	0.791	0.727	81.867	0.869	0.807
BarlowTwins (scratch)	13.762	80.273	0.908	0.804	71.31	0.81	0.72	78.797	0.852	0.777
SimCLR (scratch)	**6.646**	83.36	0.91	0.836	70.253	0.767	0.71	78.127	0.879	0.773
IRS-SSL-Ind (Ours)	7.096	**89.253**	**0.955**	**0.893**	75.89	0.84	0.769	**85.683**	0.897	**0.845**
IRS-SSL-Pop (Ours)	7.412	84.423	0.933	0.847	**77.98**	**0.853**	**0.789**	83.843	**0.898**	0.832

	Tumor Grade			AD/MCI/CN (ext)			Age (ext)	MCI/CN (ext)			Tumor Segmentation
Pretext	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	L1↓	BAcc↑	AUC↑	F1↑	DCS↑
MAE	64.643	0.742	0.664	53.65	0.823	0.475	9.799	60.153	0.765	0.585	**72.67**
BarlowTwins	63.707	0.717	0.645	41.34	0.747	0.396	13.832	55.27	0.81	0.564	63.25
SimCLR	61.367	0.692	0.633	45.237	0.734	0.441	11.395	50.337	0.722	0.497	66.02
BarlowTwins (scratch)	64.928	0.804	0.664	46.65	0.784	0.421	13.063	50.67	0.75	0.504	62.13
SimCLR (scratch)	**74.097**	**0.838**	**0.77**	49.313	0.771	0.459	**9.056**	64.057	0.826	0.627	71.08
IRS-SSL-Ind (Ours)	71.243	0.808	0.74	**67.31**	**0.848**	**0.614**	9.453	65.97	**0.847**	**0.703**	69.5
IRS-SSL-Pop (Ours)	69.927	0.796	0.731	53.67	0.837	0.493	9.864	**72.067**	0.791	0.675	71.13

Table 12: 95% confidence intervals computed across three runs for the linear probing results reported in Table 1.

	Age	Gender			MCI/CN			FTD/CN		
Pretext	L1↓	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑
MAE	0.233	3.267	0.014	0.033	1.098	0.031	0.038	8.443	0.01	0.067
BarlowTwins	1.236	1.162	0.017	0.014	4.503	0.101	0.080	4.448	0.052	0.042
SimCLR	0.588	3.618	0.027	0.037	3.13	0.034	0.108	4.687	0.028	0.048
BarlowTwins (scratch)	0.138	5.389	0.010	0.056	1.025	0.011	0.040	6.552	0.039	0.078
SimCLR (scratch)	0.232	1.030	0.025	0.013	4.083	0.062	0.168	2.02	0.013	0.022
IRS-SSL-Pop (Ours)	0.566	4.119	0.022	0.044	3.231	0.037	0.036	1.522	0.035	0.013
IRS-SSL-Ind (Ours)	0.462	3.630	0.008	0.032	2.625	0.002	0.015	3.213	0.017	0.039

	Tumor Grade			AD/MCI/CN (ext)			Age (ext)	MCI/CN (ext)			Tumor Segmentation
Pretext	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	L1↓	BAcc↑	AUC↑	F1↑	DCS↑
MAE	3.267	0.032	0.034	4.094	0.041	0.074	0.492	3.063	0.018	0.077	0.994
BarlowTwins	8.53	0.096	0.135	3.911	0.041	0.106	2.058	4.832	0.041	0.039	1.627
SimCLR	1.061	0.014	0.026	2.717	0.007	0.046	0.621	7.250	0.02	0.016	1.317
BarlowTwins (scratch)	5.579	0.222	0.136	2.080	0.023	0.02	0.216	2.484	0.013	0.127	2.474
SimCLR (scratch)	1.908	0.156	0.031	6.805	0.021	0.039	0.238	8.723	0.037	0.061	0.627
IRS-SSL-Pop (Ours)	9.973	0.034	0.097	4.566	0.028	0.088	0.569	3.899	0.034	0.079	1.255
IRS-SSL-Ind (Ours)	6.460	0.043	0.041	2.682	0.052	0.036	0.372	6.145	0.017	0.083	1.301

Table 13: 95% confidence intervals computed across three runs for the fine-tuning layer decay results reported in 11.

	Age	Gender			MCI/CN			FTD/CN		
Pretext	L1↓	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑
MAE	0.316	5.219	0.039	0.054	3.076	0.045	0.03	1.013	0.006	0.010
BarlowTwins	0.735	4.613	0.010	0.046	6.244	0.019	0.058	2.397	0.046	0.016
SimCLR	0.164	4.066	0.029	0.051	1.505	0.016	0.022	1.547	0.015	0.011
BarlowTwins (scratch)	1.357	10.758	0.033	0.102	5.800	0.031	0.076	7.235	0.033	0.087
SimCLR (scratch)	0.541	3.267	0.035	0.032	3.779	0.023	0.039	2.709	0.018	0.051
IRS-SSL-Pop (Ours)	0.294	3.699	0.051	0.042	4.979	0.039	0.055	1.664	0.013	0.013
IRS-SSL-Ind (Ours)	0.33	0.359	0.003	0.001	1.006	0.08	0.015	1.190	0.035	0.012

	Tumor Grade			AD/MCI/CN (ext)			Age (ext)	MCI/CN (ext)			Tumor Segmentation
Pretext	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	L1↓	BAcc↑	AUC↑	F1↑	DCS↑
MAE	2.445	0.189	0.007	3.619	0.036	0.032	0.445	23.112	0.080	0.181	1.077
BarlowTwins	4.957	0.043	0.058	7.172	0.022	0.093	0.934	9.406	0.154	0.122	0.323
SimCLR	0.961	0.038	0.019	6.223	0.084	0.093	0.996	3.653	0.036	0.101	2.131
BarlowTwins (scratch)	7.655	0.030	0.101	8.193	0.08	0.014	0.242	8.002	0.023	0.098	2.384
SimCLR (scratch)	1.886	0.125	0.026	4.341	0.038	0.049	0.324	5.029	0.124	0.078	2.399
IRS-SSL-Pop (Ours)	3.772	0.041	0.038	1.936	0.034	0.065	0.551	6.215	0.068	0.089	2.582
IRS-SSL-Ind (Ours)	2.445	0.037	0.022	5.118	0.024	0.048	0.442	0.990	0.066	0.048	1.312

F CLINICIAN EVALUATION PROTOCOL

To further assess whether the learned representations produce clinically meaningful attention patterns, we conducted a blinded pairwise evaluation with six certified radiologists. The evaluation focused on the MCI versus cognitively

Table 14: Comparison of pretraining warmup evaluated on downstream and external test tasks under linear probing. The best result per column is **bolded** and the second-best is underlined. BAcc = balanced accuracy (%).

	Age	Gender			MCI/CN			FTD/CN		
Model	L1↓	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑
Warm up (MAE)	10.107	84.153	0.915	0.844	70.947	0.809	0.721	75.947	0.803	0.753
MAE250	9.556	79.957	0.895	0.804	72.927	0.813	0.735	73.043	0.789	0.722
IRS-SSL-Pop (Ours)	10.322	82.817	0.885	0.829	72.923	0.815	0.741	76.327	**0.861**	0.757
IRS-SSL-Ind (Ours)	**9.238**	**85.47**	**0.918**	**0.856**	**73.893**	**0.821**	**0.746**	**80.593**	0.856	**0.803**

	Tumor Grade			AD/MCI/CN (ext)			Age (ext)	MCI/CN (ext)		
Pretext	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	L1↓	BAcc↑	AUC↑	F1↑
Warm up (MAE)	60.257	0.679	0.615	46.168	0.735	0.473	11.224	56.613	0.776	0.583
MAE250	61.1	0.687	0.616	46.173	0.762	0.458	10.614	52.09	0.763	0.507
IRS-SSL-Pop (Ours)	**63.04**	0.677	0.624	57.05	0.801	0.495	11.674	74.007	0.858	0.663
IRS-SSL-Ind (Ours)	61.763	**0.692**	**0.625**	**64.433**	**0.834**	**0.605**	**10.605**	**75.067**	**0.893**	**0.718**

Table 15: Few-shot ($n = 5$) results in linear probing settings. Best results are bold; second-best results are underlined.

	Gender			MCI/CN			FTD/CN		
Pretext	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑
MAE	72.273	0.803	0.723	51.400	0.575	0.458	64.090	0.721	0.636
BarlowTwins	72.533	0.788	0.711	60.460	0.670	0.553	50.287	0.512	0.442
SimCLR	71.813	0.783	0.719	46.227	0.386	0.353	65.223	0.687	0.652
BarlowTwins (scratch)	70.593	0.806	0.694	57.073	0.644	0.516	53.600	0.565	0.522
SimCLR (scratch)	70.340	0.761	0.696	54.337	0.601	0.519	70.713	0.784	0.709
IRS-SSL-pop (ours)	73.313	0.811	0.721	55.013	0.588	0.489	**78.903**	**0.855**	**0.786**
IRS-SSL-Ind (ours)	**75.390**	**0.814**	**0.750**	**64.397**	**0.717**	**0.622**	73.843	0.803	0.730

	Tumor Grade			AD/MCI/CN (ext)			MCI/CN (ext)		
Pretext	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑	BAcc↑	AUC↑	F1↑
MAE	61.650	0.628	0.582	49.030	0.659	0.360	50.620	0.454	0.418
BarlowTwins	54.350	0.569	0.493	31.873	0.502	0.242	50.207	0.450	0.442
SimCLR	59.097	0.662	0.575	49.423	0.680	0.350	45.247	0.337	0.345
BarlowTwins (scratch)	56.601	0.628	0.534	33.933	0.472	0.324	55.733	0.533	0.465
SimCLR (scratch)	65.217	**0.670**	0.626	38.240	0.586	0.286	59.150	0.631	0.451
IRS-SSL-pop (ours)	60.507	0.666	0.574	**52.360**	0.722	0.424	60.827	0.663	0.591
IRS-SSL-Ind (ours)	**66.363**	0.650	**0.652**	49.860	**0.723**	**0.447**	**72.017**	**0.685**	**0.655**

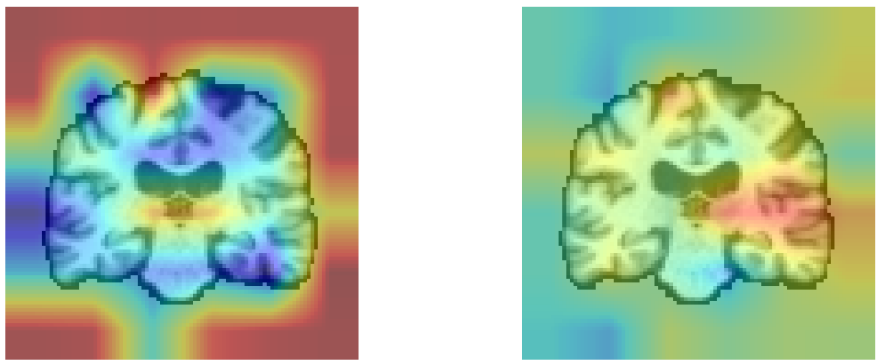
normal (CN) classification task and compared Grad-CAM attention maps obtained from IRS-SSL-Ind against those obtained from models pretrained with MAE, SimCLR, and Barlow Twins.

Each radiologist completed a 45-case questionnaire. For each case, two Grad-CAM attention maps corresponding to the same coronal MRI slice were displayed side-by-side. One attention map was generated by IRS-SSL-Ind and the other by one of the three comparison models. Model identities were not disclosed to the radiologists, and the presentation order of the two attention maps was randomized to reduce positional and model-identity bias.

Radiologists were asked to select the attention map that was more clinically appropriate for distinguishing MCI from (see Fig. 4). Specifically, they were instructed to consider: (1) the anatomical locations receiving high attention, indicated by warmer colors; (2) whether these regions were consistent with structures known to be relevant to MCI, such as the medial temporal lobe, hippocampus, and posterior cingulate; and (3) the overall anatomical plausibility and specificity of the attention pattern. An example was provided before the evaluation to familiarize the radiologists with the assessment procedure.

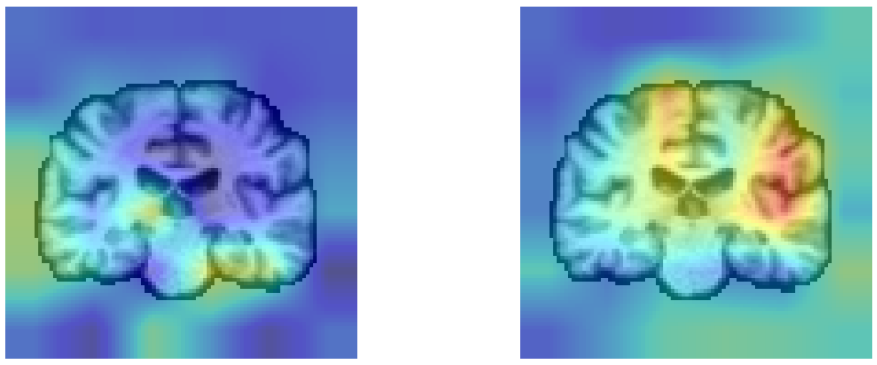
The questionnaire contained 15 pairwise comparisons for each of the three pretraining baselines, resulting in 45 comparisons per radiologist. Across the six radiologists, this yielded 90 independent forced-choice responses for each model pair and 270 responses in total.

1.



Rank the two columns (1 better) based on most appropriate attention	
Column 1	Column 2

2.



Rank the two columns (1 better) based on most appropriate attention	
Column 1	Column 2

Figure 4: **Example of the blinded clinician evaluation.** For each case, radiologists were presented with two Grad-CAM attention maps generated from the same coronal MRI slice and were asked to select the map with the more clinically appropriate attention pattern for distinguishing MCI from CN. Model identities were hidden and the left–right presentation order was randomized. Evaluators were instructed to consider localization to MCI-relevant anatomical regions as well as the overall plausibility and specificity of the attention pattern.