Brain-Specialized Self-Supervised Learning: Neuroanatomy-Aware Spatial Priors for 3D Brain MRI

Kyeong Ho Kim [1, †], Minkyeong Kwon [1, †], Yu-Mi Kim [2], Mi Kyung Kim [2], Min-Ho Shin [3], Insung Chung [4], Sang Baek Koh [5], Hyeon Chang Kim [6] and Jong-Min Lee [1, 7, *]

[1] Department of Artificial Intelligence, Hanyang University, Seoul, Republic of Korea
[2] Department of Preventive Medicine, Hanyang University College of Medicine, Seoul, Republic of Korea
[3] Department of Preventive Medicine, Chonnam National University Medical School, Gwangju, Republic of Korea
[4] Department of Occupational and Environmental Medicine, Keimyung University School of Medicine, Daegu, Republic of Korea
[5] Department of Preventive Medicine and Institute of Occupational Medicine, Yonsei Wonju College of Medicine, Wonju, Republic of Korea
[6] Department of Preventive Medicine, Yonsei University College of Medicine, Seoul, Republic of Korea
[7] Department of Biomedical Engineering, Hanyang University, Seoul, Republic of Korea
ljm@hanyang.ac.kr
[*] Corresponding author, [†] These authors contributed equally. Shared first authorship

Abstract. Self-supervised learning enables pretraining on large unlabeled 3D brain MRI cohorts. However, many objectives are adapted from natural images and underuse neuroanatomical spatial structure. We build on an iBOT-style teacher–student 3D vision transformer (ViT) with an exponential moving average (EMA) teacher. We align masked token representations in latent space. We also learn subject-level discrimination from classification tokens (CLS). To inject spatial priors, we propose distance-modulated patch decorrelation. It suppresses spurious similarity more for far-apart patches. It remains tolerant to nearby regions. We pretrain on 6,811 unlabeled T1-weighted scans from five independent cohorts. These cohorts do not overlap with the downstream datasets. We evaluate transfer on AD-related classification using the Alzheimer's Disease Neuroimaging Initiative (ADNI) and Open Access Series of Imaging Studies 3 (OASIS-3), with AD vs. NC evaluated on both datasets and MCI vs. NC evaluated on ADNI only. We also test tumor grading on Brain Tumor Segmentation (BraTS). Beyond downstream metrics, we analyze distance-dependent patch similarity. We also explain why region of interest (ROI)-based priors can be unreliable at our patch granularity.

Keywords: Self-supervised learning, 3D brain MRI, Masked image modeling, Contrastive learning, Distance-modulated patch decorrelation.

1 Introduction

Deep neural networks are now routinely used in neuroimaging pipelines, but performance in many clinical and research settings is still constrained by the limited scale of reliably labeled 3D MRI cohorts [1, 2]. Manual annotation for volumetric scans is resource-intensive and difficult to standardize across sites, whereas unlabeled acquisitions can often be aggregated at much larger scale [3]. This imbalance has accelerated interest in self-supervised learning (SSL) as a label-efficient route to transferable MRI representations [4].

Most SSL recipes in vision fall into two complementary objectives: masked prediction, which learns from partial observation, and instance-level discrimination, which promotes separability across samples [5-7]. Masked image modeling (MIM) scales well with transformer encoders by predicting masked targets from visible context [5, 6]. In parallel, contrastive or bootstrapping-based methods learn invariances by aligning representations across views while avoiding collapse, using explicit negatives (e.g., InfoNCE) or momentum-teacher self-distillation [7-10]. Hybrid teacher–student formulations extend these ideas to token-level objectives, unifying masked token learning and global discrimination [11].

A key limitation is that these formulations were largely optimized for natural-image statistics and are often applied to brain MRI without objectives that explicitly reflect neuroanatomical structure [4, 12]. Brain MRI has weaker texture cues and a highly regular anatomical layout. As a result, patch tokens from distinct locations can be visually and statistically similar, reducing patch distinctiveness and hindering patch-centric SSL training [12, 13]. This motivates SSL objectives that incorporate spatial organization as an inductive bias rather than treating patch tokens as exchangeable.

We propose a brain-tailored hybrid SSL framework built on an iBOT-style teacher–student transformer with an EMA teacher [9-11]. Our method integrates latent-space masked token alignment and subject-level global discrimination and introduces a spatially grounded patch objective. Instead of pixel-space reconstruction with a heavy decoder, we align masked token representations directly in latent space [5, 11]. We apply an InfoNCE-style loss to CLS tokens to improve subject-level separability [7]. To address low patch distinctiveness in brain MRI, we further introduce a distance-modulated patch decorrelation term that penalizes excessive similarity more strongly for spatially distant patches [14-17].

In summary, we propose a brain-specific hybrid teacher–student SSL framework that combines latent masked token alignment, global discrimination, and a spatially grounded patch objective. Our key contribution is a distance-modulated patch decorrelation prior that exploits 3D spatial organization to encourage distance-aware patch representations and support downstream transfer in task-dependent settings.

2 Related Work

2.1 SSL for vision transformers and volumetric imaging

Masked image modeling (e.g., MAE, SimMIM) is a strong baseline for transformer pretraining and transfers well to downstream tasks [5, 6]. Self-distillation methods (e.g., BYOL, DINO) learn invariances via momentum teachers without explicit negatives, while InfoNCE-based contrastive learning remains influential for instance discrimination [7, 9, 10]. Hybrid teacher–student approaches such as iBOT unify masked token learning with global alignment and have been adapted to volumetric inputs for medical imaging SSL [4, 11, 12]. Despite these advances, objectives optimized for natural images may be insufficient to capture neuroimaging properties [4, 12, 13].

2.2 Spatial priors and anatomical constraints

Neuroimaging studies often exploit spatial organization, and distance-dependent structure has been reported in brain connectivity and organization [16, 17]. ROI atlases (e.g., AAL, Desikan–Killiany) provide standardized partitions, but patch-level ROI supervision can be ambiguous due to tissue mixing and partial-volume effects at common patch granularities [18-20]. Accordingly, we study two forms of anatomical priors. We propose a distance-modulated patch decorrelation prior as the main approach, while ROI-based priors serve as an alternative evaluated for comparison.

3 Method

3.1 Overview

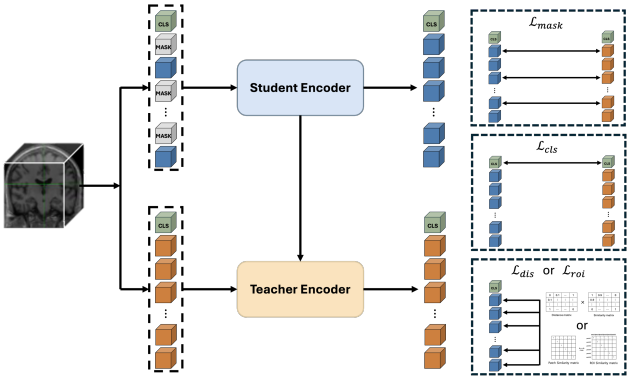

Fig. 1. Overview of our teacher–student SSL pretraining for unlabeled 3D T1 MRI. The student encoder receives a masked token sequence, while an EMA teacher processes an unmasked view. Training combines latent masked token alignment (L_{mask}), CLS token contrastive discrimination (L_{cls}), and a spatial objective (L_{dis} or L_{roi}).

As illustrated in Fig. 1, given an unlabeled 3D T1-weighted MRI volume $x \in \mathbb{R}^{H\times W\times D}$, we pretrain a vision transformer in a teacher–student framework [9-11]. The teacher processes an unmasked view, while the student receives a masked view. Pretraining optimizes (i) latent masked token alignment in latent space (without pixel reconstruction) [5, 6, 11], (ii) InfoNCE-based subject-level discrimination on CLS tokens [7], and

(iii) a spatial objective via distance-aware spatial prior or ROI-based similarity regression.

3.2 SSL for vision transformers and volumetric imaging

We partition the input volume into non-overlapping 3D patches and embed them into a token sequence:

$$\mathrm{z}_0 = [\mathrm{z}_{\mathrm{cls}};\ \mathrm{z}_1; \ldots ; \mathrm{z}_N] + \mathrm{p},$$

where $\mathbf{z}_{\mathrm{cls}}$ is a learnable global token, $\mathbf{p}$ is a positional embedding, and N is the number of patches. We denote the student encoder as $f_s(\cdot)$ and the teacher encoder as $f_t(\cdot)$. The teacher parameters are updated by EMA:

$$\theta_t \leftarrow \tau\theta_t + (1-\tau)\theta_s,$$

where $\tau \in (0,1)$ is the momentum coefficient [9–11]. For the student view, a subset of patch tokens is masked at ratio $r$and replaced with a learnable [MASK] token.

3.3 Projection heads

Let $\mathrm{h}^s \in \mathbb{R}^{(N+1)\times C}$ and h^t be student/teacher outputs. We apply a projector $g(\cdot)$ and a student-side predictor $q(\cdot)$, followed by normalization:

$$\mathrm{u}^t = \mathrm{norm}(g(\mathrm{h}^t)), \mathrm{u}^s = \mathrm{norm}(q(g(\mathrm{h}^s))).$$

Teacher gradients are stopped, as in momentum-teacher self-distillation [9-11].

3.4 Objectives

Latent masked token alignment. Let $\mathcal{M}$ be masked patch indices. We match student and teacher representations for masked tokens in latent space:

$$\mathcal{L}_{\mathrm{mask}} = \frac{1}{|\ \mathcal{M}\ |}\sum_{i\in\mathcal{M}} \|\ \mathrm{u}_i^s - \mathrm{sg}(\mathrm{u}_i^t)\ \|_2^2,$$

where $\mathrm{sg}(\cdot)$ is stop-gradient. This avoids pixel reconstruction while retaining masked context learning [5, 6, 11].

CLS token contrastive discrimination. We apply an InfoNCE-style objective to CLS embeddings. For batch size B, with normalized student/teacher CLS tokens $\mathbf{c}_b^s$ and $\mathbf{c}_b^t$:

$$\mathcal{L}_{\mathrm{cls}} = -\frac{1}{B}\sum_{b=1}^{B} \log \frac{\exp\left(\langle \mathrm{c}_b^s, \mathrm{sg}(\mathrm{c}_b^t)\rangle/T\right)}{\sum_{k=1}^{B}\exp\left(\langle\ \mathrm{c}_b^s, \mathrm{sg}(\mathrm{c}_k^t)\rangle/T\right)},$$

where T is temperature [7].

Spatially weighted decorrelation prior. To mitigate excessive similarity among patch embeddings while respecting spatial structure, we penalize pairwise similarity with distance-dependent weights. Let $\tilde{\mathbf{v}}_i^s$ be normalized student patch embeddings (excluding CLS), and $S_{ij} = \langle \tilde{\mathrm{v}}_i^s, \tilde{\mathrm{v}}_j^s \rangle$. With normalized inter-patch distances $d_{ij} \in [0,1]$ we use a monotone weight $w(d_{ij})$defined as $w(d_{ij}) = d_{ij}$. We compute d_{ij} as the Euclidean distance between 3D patch-center coordinates and normalize it to $[0,1]$ by dividing by the maximum pairwise distance among patch centers within each volume. The distance-modulated term is computed over all unique token pairs within each volume:

$$\mathcal{L}_{\text{dis}} = \frac{1}{N(N-1)} \sum_{i \neq j} w(\, d_{ij})\, S_{ij}^2.$$

This objective is related in spirit to redundancy-reduction regularizers (decorrelation) but introduces explicit spatial modulation for brain MRI [14-17].

ROI similarity regression. Using a parcellation with K regions, we compute soft ROI compositions $\mathrm{r}_i \in \mathbb{R}^K$ for each patch i from voxel-wise ROI fractions. We define ROI affinity as $A_{ij} = \langle \mathrm{r}_i, \mathrm{r}_j \rangle$ and regress feature similarity toward ROI similarity:

$$\mathcal{L}_{\text{roi}} = \frac{1}{N(N-1)} \sum_{i \neq j} (\, S_{ij} - A_{ij})^2.$$

We treat this ROI-based term as an alternative anatomical prior for comparison with our main distance-modulated decorrelation prior [18-20].

3.5 Total loss

We use a short stabilization period where $\mathcal{L}_{\text{mask}}$ dominates, then enable the global and spatial terms. The final objective is:

$$\mathcal{L} = \mathcal{L}_{\text{mask}} + \lambda_{\text{cls}} \mathcal{L}_{\text{cls}} + \lambda_{\text{sp}} \mathcal{L}_{\text{sp}}$$

where $\mathcal{L}_{sp} \in \mathcal{L}_{dis}, \mathcal{L}_{roi}$.

4 Experiments

4.1 Data

We pretrain on an unlabeled T1-weighted MRI corpus aggregated from five cohorts (OASIS-1/2 [21], IXI, and two private cohorts; 6,811 volumes), with subject-level deduplication and no overlap with downstream evaluation sets. Transferability is assessed by fine-tuning on ADNI (NC/AD, NC/MCI) [22], OASIS-3 (NC/AD) [23], and BraTS 2018 (high-grade glioma (HGG) vs. low-grade glioma (LGG)) [24]. All datasets share a unified preprocessing pipeline and are resampled before patch tokenization.

4.2 Baselines and evaluation protocol

We compare against supervised-from-scratch and backbone-matched SSL baselines (MAE [5], SimMIM [6], iBOT [11]) using a common 3D ViT-style encoder, and include M3T [25] as a supervised baseline for AD classification. For all downstream tasks in Sec. 4.1, we use subject-wise 5-fold cross-validation and report mean ± std.

4.3 Implementation details

We pretrain a MONAI 3D ViT-B in a teacher–student setup with EMA momentum of 0.996 and a masking ratio of 0.75 for 600 epochs. After a 50 epoch MIM only warmup, we add CLS contrastive learning and an additional spatial term ($\lambda\_cls = 1$ and $\lambda\_sp = 1$). Fine-tuning runs for 200 epochs with subject-wise 5-fold cross validation at 126^3 input resolution, 14×14×14 patches, and a cosine annealing learning-rate schedule with linear warmup. These settings were selected to follow standard MAE/iBOT style training practice and to balance spatial resolution, token length and computational cost. Although we did not conduct an exhaustive sensitivity analysis over λ_cls, λ_sp and patch size, a systematic investigation of these design choices would be valuable future work.

5 Results & Discussion

5.1 Overall performance across benchmarks

Table 1. Downstream performance (accuracy (ACC)/ area under the ROC curve (AUC), mean ± std) on ADNI-AD/NC, ADNI-MCI/NC, OASIS-3, and BraTS.

Method	ADNI AD/NC		ADNI MCI/NC		OASIS-3		BraTS 2018	
	ACC	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**
ViT-B	79.3 (±2.6)	0.81 (±0.07)	57.4 (±3.1)	0.59 (±0.04)	80.4 (±2.6)	0.61 (±0.08)	73.3 (±0.7)	0.58 (±0.03)
M3T	**88.0** (±1.1)	**0.92** (±0.01)	60.6 (±1.8)	0.65 (±0.02)	82.3 (±1.8)	0.78 (±0.01)	68.7 (±4.5)	0.65 (±0.13)
MAE	83.9 (±0.67)	0.89 (±0.08)	62.6 (±2.6)	0.67 (±0.02)	79.4 (±2.3)	0.71 (±0.03)	71.5 (±4.3)	0.65 (±0.09)
Sim-MIM	81.1 (±3.2)	0.86 (±0.02)	57.3 (±2.9)	0.60 (±0.04)	81.4 (±1.5)	0.71 (±0.06)	70.1 (±9.0)	0.59 (±0.12)
iBOT	86.4 (±2.1)	**0.92** (±0.01)	64.2 (±2.6)	0.69 (±0.03)	81.3 (±0.6)	**0.79** (±0.05)	73.3 (±0.02)	0.59 (±0.13)
Ours	86.0 (± 0.9)	0.91 (±0.01)	**64.6** (±2.7)	**0.71** (±0.03)	**83.9** (±0.7)	0.76 (±0.07)	**74.0** (±4.8)	**0.73** (±0.06)

Table 1 summarizes downstream performance across four benchmarks. DIS does not uniformly outperform iBOT across all tasks. In particular, on ADNI-AD/NC, performance is comparable but not improved over iBOT, likely because AD/NC separation is relatively clearer and strong baselines already perform well. In contrast, DIS shows gains on ADNI-MCI/NC and BraTS 2018, where disease cues may be more subtle, spatially distributed, or heterogeneous. Overall, these results suggest that DIS provides

task-dependent benefits rather than uniform improvements across benchmarks, with greater advantages in tasks that require modeling complex spatial disease patterns.

5.2 Ablation study: contribution of each objective

Table 2. Ablation of the pretraining objectives ($\mathcal{L}_{\text{mask}}$, $\mathcal{L}_{\text{cls}}$, $\mathcal{L}_{dis}$, and $\mathcal{L}_{roi}$) evaluated by downstream ACC/AUC (mean ± std) on ADNI-AD/NC, ADNI-MCI/NC, and OASIS-3.

$\mathcal{L}_{\text{mask}}$	$\mathcal{L}_{\text{cls}}$	$\mathcal{L}_{dis}$	$\mathcal{L}_{roi}$	ADNI AD/NC		ADNI MCI/NC		OASIS-3	
				ACC	**AUC**	**ACC**	**AUC**	**ACC**	**AUC**
X	X	X	X	79.3 (±2.6)	0.81 (±0.07)	57.4 (±3.1)	0.59 (±0.04)	80.4 (±2.6)	0.61 (±0.08)
O	X	X	X	84.1 (±1.3)	0.90 (±0.09)	57.8 (±2.6)	0.59 (±0.04)	81.5 (±3.2)	0.74 (±0.05)
O	O	X	X	**86.3** (±1.1)	**0.91** (±0.01)	63.1 (±2.1)	0.69 (±0.02)	81.0 (±1.9)	**0.76** (±0.04)
O	O	O	X	86.0 (± 0.9)	**0.91** (±0.01)	**64.6** (±2.7)	**0.71** (±0.03)	**83.9** (±0.7)	**0.76** (±0.07)
O	O	X	O	85.3 (±1.6)	0.89 (±0.00)	61.1 (±2.0)	0.64 (±0.01)	81.2 (±1.2)	0.69 (±0.08)

Table 2 reports a progressive ablation study. We use the masked token objective as the base pretext task, then add the CLS objective and either the proposed distance-modulated spatial prior or the ROI-based prior. Adding CLS improves transfer, while adding DIS further improves ADNI-MCI/NC and OASIS-3 accuracy compared with the CLS-only variant, although gains are not uniform across all metrics. In contrast, the ROI-based prior degrades performance, likely because patch-level ROI targets are noisy at our token granularity.

5.3 Representation analysis: distance-dependent patch similarity

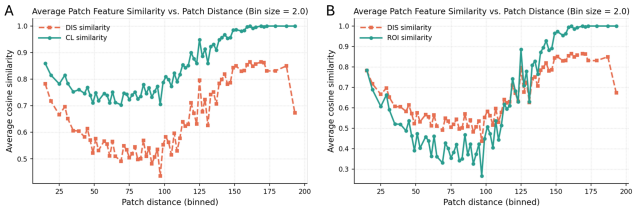

Fig. 2. (A) DIS vs. CLS-only. (B) DIS vs. ROI within typical intra-brain distances. Cosine similarity between patch embeddings as a function of 3D inter-patch distance.

As shown in Fig. 2A, DIS yields a steeper decay of patch similarity with increasing 3D distance than CLS-only. This indicates clearer separation of distant anatomy.

At very large distances, both objectives can exhibit spuriously high similarity. We attribute this behavior to background-dominated corner patches. We therefore focus on typical intra-brain distance ranges to mitigate such artifacts.

Fig. 2B further compares DIS with the ROI-based prior within this regime. ROI appears to induce more sharply separated patch embeddings in the similarity plot. However, this does not translate into improved downstream performance and even degrades it (Table 2). This discrepancy suggests a mismatch between atlas-level ROI-based priors and the requirements of our self-supervised pretraining objective. We analyze this issue in Sec. 5.4.

5.4 Why ROI-based priors fail at the chosen patch granularity

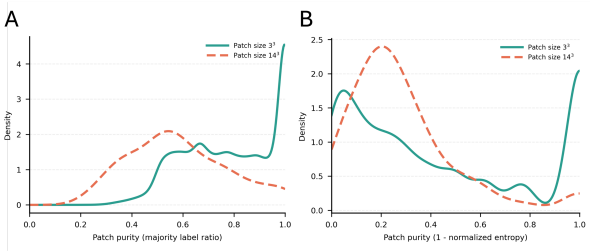

Fig. 3. Patch-level ROI purity for 14×14×14 tokens (majority ROI fraction and entropy), showing many mixed-region patches and noisy ROI-based prior.

To explain why ROI-based priors degrade downstream performance (Table 2), we examine the quality of ROI supervision at our token granularity. With 14×14×14 patches, many tokens span multiple anatomical regions. Consequently, voxel-derived ROI soft targets become ambiguous.

Fig. 3 confirms low ROI purity for a large fraction of patches. Specifically, many patches show a low majority ROI fraction and high entropy, indicating substantial label noise. Moreover, this supervision imposes rigid, atlas-level constraints that are often misaligned with mixed-region tokens. As a result, it interferes with transferable representation learning and leads to performance drops.

6 Conclusion

We present a teacher–student SSL framework for 3D brain MRI that combines masked latent-token alignment and subject-level discrimination with a lightweight distance-modulated decorrelation prior. Across the evaluated benchmarks, DIS shows task-dependent transfer benefits, with gains on ADNI-MCI/NC and BraTS, and yields a more distance-structured organization of patch embeddings. Our findings suggest that inter-patch distance provides an effective annotation-free spatial cue for scalable 3D MRI pretraining, particularly when disease signals are subtle and distributed.

Acknowledgments. This research was supported by an Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korean government (MSIT) (No. RS-2020-II201373, Artificial Intelligence Graduate School Program (Hanyang University)) and by a grant of the Korea Dementia Research Project through the Korea Dementia Research

Center (KDRC), funded by the Ministry of Health & Welfare and Ministry of Science and ICT, Republic of Korea (grant number: RS-2020-KH106773). This study used clinical data provided by the Korea Dementia Research Center's Trial Ready Registry (KDRC TRR), supported by the Ministry of Science and ICT & the Ministry of Health & Welfare, Republic of Korea. TRR data were collected with informed consent under institutional review board–approved protocols.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Data use declaration. Publicly available datasets analyzed in this study include IXI, ADNI, OASIS, and BraTS 2018, accessed under their respective data use agreements/licenses. In addition, we analyze two private deidentified cohorts collected under local policies.

References

1. Litjens, G., et al., A survey on deep learning in medical image analysis. Medical image analysis, 2017. 42: p. 60-88.
2. Shen, D., G. Wu, and H.-I. Suk, Deep learning in medical image analysis. Annual review of biomedical engineering, 2017. 19(1): p. 221-248.
3. Cheplygina, V., M. De Bruijne, and J.P. Pluim, Not-so-supervised: a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. Medical image analysis, 2019. 54: p. 280-296.
4. Zhou, Z., et al. Models genesis: Generic autodidactic models for 3d medical image analysis. in International conference on medical image computing and computer-assisted intervention. 2019. Springer.
5. He, K., et al. Masked autoencoders are scalable vision learners. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.
6. Xie, Z., et al. Simmim: A simple framework for masked image modeling. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.
7. Oord, A.v.d., Y. Li, and O. Vinyals, Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748, 2018.
8. Chen, X., S. Xie, and K. He. An empirical study of training self-supervised vision transformers. in Proceedings of the IEEE/CVF international conference on computer vision. 2021.
9. Grill, J.-B., et al., Bootstrap your own latent-a new approach to self-supervised learning. Advances in neural information processing systems, 2020. 33: p. 21271-21284.
10. Caron, M., et al. Emerging properties in self-supervised vision transformers. in Proceedings of the IEEE/CVF international conference on computer vision. 2021.
11. Zhou, J., et al., ibot: Image bert pre-training with online tokenizer. arXiv preprint arXiv:2111.07832, 2021.
12. Zhuang, X., et al. Self-supervised feature learning for 3d medical images by playing a rubik's cube. in Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part IV 22. 2019. Springer.
13. Wu, L., J. Zhuang, and H. Chen. Voco: A simple-yet-effective volume contrastive learning framework for 3d medical image analysis. in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024.

14. Zbontar, J., et al. Barlow twins: Self-supervised learning via redundancy reduction. in International conference on machine learning. 2021. PMLR.
15. Bardes, A., J. Ponce, and Y. LeCun, Vicreg: Variance-invariance-covariance regularization for self-supervised learning. arXiv preprint arXiv:2105.04906, 2021.
16. Bullmore, E. and O. Sporns, The economy of brain network organization. Nature reviews neuroscience, 2012. 13(5): p. 336-349.
17. Ercsey-Ravasz, M., et al., A predictive network model of cerebral cortical connectivity based on a distance rule. Neuron, 2013. 80(1): p. 184-197.
18. Tzourio-Mazoyer, N., et al., Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. Neuroimage, 2002. 15(1): p. 273-289.
19. Desikan, R.S., et al., An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage, 2006. 31(3): p. 968-980.
20. Yeo, B.T., et al., The organization of the human cerebral cortex estimated by intrinsic functional connectivity. Journal of neurophysiology, 2011.
21. Marcus, D.S., et al., Open Access Series of Imaging Studies (OASIS): cross-sectional MRI data in young, middle aged, nondemented, and demented older adults. Journal of cognitive neuroscience, 2007. 19(9): p. 1498-1507.
22. Mueller, S.G., et al., The Alzheimer's disease neuroimaging initiative. Neuroimaging Clinics, 2005. 15(4): p. 869-877.
23. LaMontagne, P.J., et al., OASIS-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and Alzheimer disease. medrxiv, 2019: p. 2019.12. 13.19014902.
24. Menze, B.H., et al., The multimodal brain tumor image segmentation benchmark (BRATS). IEEE transactions on medical imaging, 2014. 34(10): p. 1993-2024.
25. Jang, J. and D. Hwang. M3T: three-dimensional Medical image classifier using Multi-plane and Multi-slice Transformer. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022.