ICHOR: A Robust Representation Learning Approach for ASL CBF Maps with Self-Supervised Masked Autoencoders

Xavier Beltran-Urbano[1], Yiran Li[2], Xinglin Zeng[2], Katie R. Jobson[1], Manuel Taso[1], Christopher A. Brown[1], David A. Wolk[1], Corey T. McMillan[1], Ilya M. Nasrallah[1], Paul A. Yushkevich[1], Ze Wang[2], John A. Detre[1], and Sudipto Dolui[1]

[1] University of Pennsylvania, Philadelphia, USA
[2] University of Maryland School of Medicine, Baltimore, USA
sudiptod@pennmedicine.upenn.edu

Abstract. Arterial spin labeling (ASL) perfusion MRI allows direct quantification of regional cerebral blood flow (CBF) without exogenous contrast, enabling noninvasive measurements that can be repeated without constraints imposed by contrast injection. ASL is increasingly acquired in research studies and clinical MRI protocols. Building on successes in structural imaging, recent efforts have implemented deep learning-based methods to improve image quality, enable automated quality control, and derive robust quantitative and predictive biomarkers with ASL-derived CBF. However, progress has been limited by variable image quality, substantial inter-site, vendor and protocol differences, and limited availability of labeled datasets needed to train models that generalize across cohorts. To address these challenges, we introduce ICHOR, a self-supervised pre-training approach for ASL CBF maps that learns transferable representations using 3D masked autoencoders. ICHOR is pre-trained via masked image modeling using a Vision Transformer backbone and can be used as a general-purpose encoder for downstream ASL tasks. For pre-training, we curated one of the largest ASL datasets to date, comprising 11,405 ASL CBF scans from 14 studies spanning multiple sites and acquisition protocols. We evaluated the pre-trained ICHOR encoder on three downstream diagnostic classification tasks and one ASL CBF map quality prediction regression task. Across all evaluations, ICHOR outperformed existing neuroimaging self-supervised pre-training methods adapted to ASL. Pre-trained weights and code will be made publicly available.

Keywords: Arterial Spin Labeling · Cerebral Blood Flow · Self-Supervised Learning · Masked Autoencoders · Vision Transformers · Transfer Learning

1 Introduction

Regional brain perfusion can be quantified via cerebral blood flow (CBF), defined as the volume of blood flowing through a specific region of brain tissue

per unit time. CBF is a fundamental physiological marker of brain function that reflects cerebrovascular integrity and metabolic demand [13], and is sensitive to both normal aging and a wide range of neurological disorders [9]. Conventional perfusion imaging typically relies on exogenous and often radioactive contrast agents, which can limit repeated use and may be contraindicated in vulnerable populations. In contrast, arterial spin labeling (ASL) perfusion MRI [3] provides noninvasive quantification of regional CBF using magnetically labeled arterial blood water as an endogenous tracer. ASL has demonstrated clinical utility in the evaluation and monitoring of Alzheimer's disease (AD), cerebrovascular disorders, brain tumors, epilepsy, and other neurological conditions [17], where regional CBF provides a biomarker of pathology and disease progression [18]. ASL is also included in the imaging protocols of several large-scale neuroimaging studies, including ADNI [11], UK Biobank [12], and CLARiTI [10]. Despite this growing adoption, large-scale ASL analysis can be challenging due to intrinsically low signal and methodological variability [7], which contribute to variable image quality and substantial inter-site and inter-protocol heterogeneity. This heterogeneity, together with the limited availability of labeled datasets, has hindered the development of deep learning (DL) methods in ASL imaging.

Self-supervised learning (SSL) offers a natural way to mitigate limited labeled data by pre-training models on large collections of unlabeled scans and then adapting them to specific downstream tasks with minimal annotation [19]. In SSL, models learn transferable representations by solving surrogate objectives that exploit intrinsic structure in the data [5]. Among recent SSL strategies, masked autoencoders (MAEs) [6] have shown strong performance for volumetric imaging by randomly masking a large fraction of input patches and training the model to reconstruct the missing content from the visible context. This reconstruction objective encourages spatially coherent and semantically meaningful representations that can transfer well to tasks such as prediction, classification, and quality control. Vision Transformers (ViTs) are particularly well-suited to MAE-based pre-training, as their self-attention mechanism captures long-range dependencies and global context in 3D medical images. However, ViT-based models typically require large and diverse pre-training datasets to generalize well, an assumption that is often difficult to meet in medical imaging.

In neuroimaging, several large-scale pre-trained models have been proposed for brain MRI analysis, including BrainIAC [14] and BrainSegFounder [2], and have demonstrated strong transfer to a range of downstream tasks. However, these efforts have largely focused on structural or multi-contrast anatomical MRI and have not incorporated physiological imaging modalities such as ASL. Consequently, ASL-derived CBF lacks a dedicated pre-trained backbone, despite its clinical relevance and increasing availability in large population studies. This gap motivates the development of pre-trained models tailored to ASL CBF maps that can support a broad range of downstream analyses.

Here, we introduce ICHOR, a self-supervised pre-training framework for ASL CBF maps based on 3D MAEs. To support large-scale pre-training, we curated one of the largest ASL datasets to date, comprising 11,405 CBF scans from 14

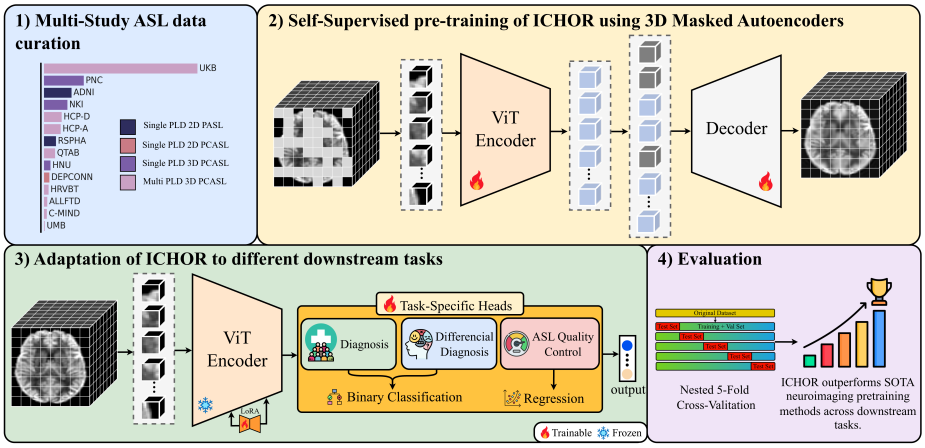

Fig. 1. Overview of the presented study.

studies spanning diverse sites and acquisition protocols. We evaluate the pre-trained ICHOR encoder on three diagnostic classification tasks and one ASL CBF map quality prediction task. Our contributions are: (1) a dedicated SSL pre-training framework for ASL CBF representation learning, (2) a large curated multi-site ASL dataset supporting pre-training across heterogeneous acquisition settings, and (3) an extensive downstream evaluation demonstrating improved performance over existing neuroimaging SSL baselines when adapted to ASL. An overview of this study is illustrated in Fig. 1.

2 Proposed Approach

Our framework adopts the pre-train-and-finetune paradigm for SSL approaches. First, we pre-train an asymmetric encoder-decoder architecture in a self-supervised manner by reconstructing randomly masked patches from the visible context. Next, we adapt the pre-trained encoder for downstream tasks using supervised learning. These two stages are described in the following sections.

2.1 Stage 1: Self-Supervised Model Pre-Training

As illustrated in Fig. 2, the core objective of this stage is to learn robust representations by reconstructing randomly masked portions of the input. This stage can be divided into the following sections:

Model Input. Let $x \in \mathbb{R}^{1\times H\times W\times D}$ denote a preprocessed ASL CBF volume registered to the Montreal Neurological Institute (MNI) space and resampled to a fixed size of $(H, W, D) = (96, 96, 96)$. We split x into $N = (H/P)(W/P)(D/P) = 512$ non-overlapping 3D patches of size $P \times P \times P$ with $P = 12$. We then randomly mask a fraction $\rho = 0.5$ of the patches by selecting masked indices

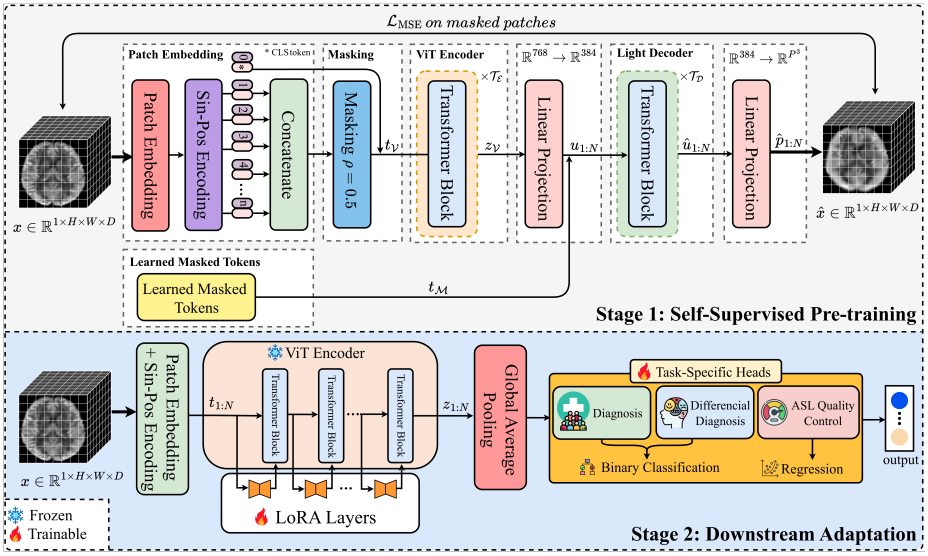

Fig. 2. Schematic of the ICHOR framework. Stage 1 (top) represents the proposed self-supervised 3D MAE pre-training pipeline. Stage 2 (bottom) illustrates the downstream task adaptation.

$\mathcal{M} \subset \{1, \ldots, N\}$ with $|\mathcal{M}| = \rho N$, and denote the visible patches (unmasked patches) by $\mathcal{V} = \{1, \ldots, N\} \setminus \mathcal{M}$.

Vision Encoder. The visible patches are embedded into tokens $t_{\mathcal{V}} \in \mathbb{R}^{|\mathcal{V}|\times 768}$ through a patch embedding layer. Then, sinusoidal positional embeddings are added to encode spatial information. The resulting tokens are processed through a ViT-Base architecture comprising 12 transformer blocks ($T_{\mathcal{E}}$) with 12 attention heads and an MLP dimension of 3072, producing latent representations $z_{\mathcal{V}} \in \mathbb{R}^{|\mathcal{V}|\times 768}$.

Light Decoder. To reconstruct the masked content, we first project the encoder outputs to the decoder dimensions ($\mathbb{R}^{768} \to \mathbb{R}^{384}$). We then construct the full decoder input sequence $u_{1:N} \in \mathbb{R}^{N\times 384}$ by placing the projected visible tokens at their original positions and inserting learnable mask tokens at the masked positions $\mathcal{M}$. The resulting sequence is processed through a light-weight decoder comprising 4 transformer blocks ($T_{\mathcal{D}}$) with 12 attention heads and an MLP dimension of 1536, producing decoder representations $\hat{u}_{1:N} \in \mathbb{R}^{N\times 384}$. Finally, a linear projection layer ($\mathbb{R}^{384} \to \mathbb{R}^{P^3}$) maps the decoder tokens to predict the corresponding patch content $\hat{p}_{1:N} \in \mathbb{R}^{N\times P^3}$.

Model optimization. The model is optimized by minimizing a mean squared error (MSE) loss function computed only over the masked patches:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{|\mathcal{M}|} \sum_{i\in\mathcal{M}} \|\hat{p}_i - p_i\|_2^2 \tag{1}$$

where $\hat{p}_i \in \mathbb{R}^{P^3}$ denotes the predicted patch and $p_i \in \mathbb{R}^{P^3}$ denotes the ground-truth patch.

2.2 Stage 2: Downstream Adaptation

After pre-training, we adapt the encoder to downstream tasks using Low-Rank Adaptation (LoRA) [8], which enables efficient fine-tuning while mitigating catastrophic forgetting. As illustrated in Fig. 2, trainable LoRA layers are inserted into the pre-trained ViT encoder while the original weights remain fixed.

During downstream adaptation, no masking is applied, resulting in $N = 512$ visible tokens. The pre-trained encoder processes the input tokens $t_{\mathcal{V}} \in \mathbb{R}^{|\mathcal{V}| \times 768}$ to produce latent representations $z_{1:N} \in \mathbb{R}^{N \times 768}$. Then, these representations are aggregated via global average pooling across spatial dimensions, and the resulting fixed-size feature vector is fed to a task-specific head consisting of one normalization layer followed by a fully connected layer.

3 Experimental Setup

3.1 Pre-training Configuration

To train our model, we created a large multi-site ASL CBF map collection by aggregating data from publicly available and institutional studies (see Fig. 1). In total, after undergoing quality control using an automated tool [15], our dataset included 11,405 scans from 14 studies spanning diverse acquisition protocols and heterogeneous populations. Before model training, all scans were preprocessed following well-established standardized pipelines [16, 4] consisting of: (i) computation of CBF maps from raw data, (ii) spatial normalization to MNI space at $2\,mm^3$ resolution, (iii) cropping to a fixed brain bounding box to reduce background, and (iv) resampling to a fixed grid of $96 \times 96 \times 96$. Finally, (v) CBF intensities were normalized by clipping to $[0, 100]$ and dividing by 100, resulting in values within $[0, 1]$.

For model training, an NVIDIA GeForce RTX 4090 with 24GB memory was used for 400 epochs using AdamW optimizer with a base learning rate of 1.5×10^{-4}, weight decay of 0.05, and batch size of 48. We also applied cosine learning rate decay with 40 epochs of linear warmup. To mitigate dataset imbalance, we used weighted random sampling with soft balancing ($\alpha_{\text{bal}} = 0.5$).

3.2 Downstream Tasks and Evaluation

For all downstream task adaptations, we inserted LoRA modules into the query, key, value, and output projection layers of each transformer block, using rank $r = 8$, scaling factor $\alpha_{\text{LoRA}} = 16$, and dropout $p = 0.2$. Models were trained on an NVIDIA A100 40 GB GPU for 100 epochs with a batch size of 8, using the AdamW optimizer with a base learning rate of 5×10^{-4}, weight decay of 0.05, and cosine learning rate decay with 10 epochs of linear warmup. A weighted binary

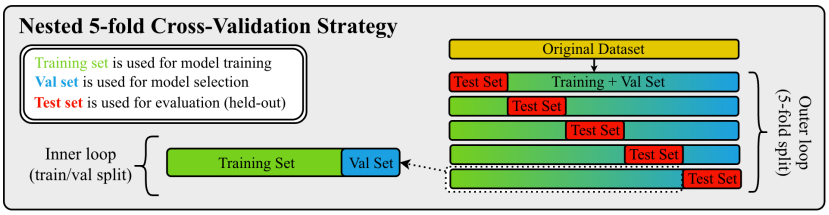

Fig. 3. Nested 5-fold cross-validation strategy used for evaluation.

cross-entropy (BCE) loss was used for classification tasks, with class weights inversely proportional to class frequencies, and MSE loss for the regression task.

Using this experimental setup, we evaluate ICHOR against state-of-the-art (SOTA) baselines, BrainIAC [14], BrainSegFounder [2], and MedicalNet [1], each pre-trained with structural MRI scans. Since MedicalNet is a ResNet-based architecture rather than a transformer, it was fine-tuned by updating all parameters rather than via LoRA modules, with all other hyperparameters kept identical across models. We evaluate across four downstream tasks: (1) binary classification of cognitively unimpaired amyloid-negative (CU A$\beta-$; $n = 103$) versus cognitively impaired amyloid-positive participants (CI A$\beta+$; $n = 44$), where amyloid positivity reflects Alzheimer's disease neuropathology; (2) ASL quality prediction, a regression task targeting a continuous score in $[0, 1]$ ($n = 383$), with a higher score reflecting a better quality CBF map; (3) binary classification of healthy older adults (HOA; $n = 20$) versus small vessel disease patients (SVD; $n = 15$); and (4) differential diagnosis between AD ($n = 27$) and behavioral variant frontotemporal dementia (bvFTD; $n = 36$).

Due to the limited sample size of some datasets, we employed a nested 5-fold cross-validation strategy to evaluate the models, as illustrated in Fig. 3. Each dataset was randomly partitioned into five stratified folds to preserve class distribution across splits, and in each iteration, one fold was held out as an independent test set while the remaining folds were used for training, with an inner 80%/20% stratified train/validation split for model training and selection. Performance is reported as the mean across the five held-out test folds, using the area under the curve (AUC), accuracy, precision, recall, and F1-score for classification tasks, and MSE, root MSE (RMSE), mean absolute error (MAE), R^2 and Pearson correlation (PC) for regression tasks. All datasets used for downstream adaptation were obtained from institutional cohorts independent of the pre-training data.

4 Results and Discussion

4.1 Comparison with State-of-the-Art

Results across all downstream tasks are presented in Table 1. Compared to the pre-trained baselines trained on structural imaging, ICHOR improves overall

Table 1. Results obtained by adapting the pre-trained encoder to downstream tasks using LoRA. Rand denotes randomly initialized model parameters. Arrows indicate desired direction. Metrics are reported as mean across the five outer folds. (Unit: %)

Method	CU A$\beta-$ vs CI A$\beta+$					ASL-Quality Prediction				
	AUC ↑	Accuracy ↑	Precision ↑	Recall ↑	F1-score ↑	MSE ↓	MAE ↓	RMSE ↓	R^2 ↑	PC ↑
Rand	68.24	57.12	38.5	**74.44**	50.34	1.21	8.14	10.96	83.99	92.03
MedicalNet [1]	70.31	68.80	51.07	56.94	51.80	1.31	8.29	11.32	82.54	91.21
BrainSegFounder [2]	65.90	59.90	40.54	72.5	51.28	1.21	8.25	10.94	83.86	91.98
BrainIAC [14]	55.97	57.93	37.86	54.72	41.71	1.44	8.72	11.97	80.94	90.30
Ours	**78.93**	**73.54**	**57.89**	63.61	**59.37**	**1.04**	**7.64**	**10.15**	**86.29**	**93.11**
Method	HOA vs SVD					AD vs bvFTD				
	AUC ↑	Accuracy ↑	Precision ↑	Recall ↑	F1-score ↑	AUC ↑	Accuracy ↑	Precision ↑	Recall ↑	F1-score ↑
Rand	55.83	42.85	40.95	**86.67**	**54.46**	84.52	81.28	86.17	83.93	83.99
MedicalNet [1]	48.33	48.57	38.57	53.33	41.33	98.02	89.10	94.16	86.78	89.61
BrainSegFounder [2]	48.33	51.43	28.57	40.00	31.43	85.80	82.94	85.27	86.67	85.45
BrainIAC [14]	50.00	54.29	41.91	60.00	48.19	86.11	85.89	93.05	83.92	86.67
Ours	**73.33**	**71.43**	**66.67**	46.67	52.67	**100.00**	**98.33**	**100.00**	**97.14**	**98.46**

performance across all four tasks, with the strongest gains on the diagnostic classification problems. A likely contributor to these gains is modality mismatch during pre-training. Structural MRI encodes anatomical contrast, whereas ASL CBF maps capture a physiological perfusion signal. As a result, representations learned from structural MRI, including features related to tissue boundaries, morphological patterns, and intensity gradients, may transfer less effectively to perfusion-derived CBF distributions, which can be particularly limiting for diagnostic tasks with subtle inter-class differences (for example, HOA vs SVD).

This effect is less pronounced for the ASL quality prediction task, where structural MRI pre-trained baselines perform more competitively. Scan quality is influenced by low-level signal properties such as signal-to-noise ratio, smoothness, and intensity contrast, which are more transferable across MRI modalities than the disease-specific perfusion patterns required for diagnostic classification.

Finally, the randomly initialized baseline underperforms ICHOR on most metrics across the four tasks, though in some settings achieving higher recall and F1, but performs competitively with the structural MRI pre-trained baselines. One likely reason is that the pre-trained models are adapted to ASL using LoRA, which may limit adaptation under strong modality mismatch, whereas the randomly initialized model is trained entirely on ASL data. Additionally, several downstream cohorts are small, and training from random initialization is prone to overfitting, which can inflate recall and F1. Metrics such as AUC, which are robust to such effects, confirm that ICHOR consistently outperforms random initialization, underscoring the value of modality-specific pre-training.

4.2 Effect of Masking Ratio

We evaluated the effect of the masking ratio ρ on both reconstruction quality and downstream task performance. As shown in Fig. 4, increasing ρ yields visually sharper reconstructions across acquisition protocols and scanner vendors. However, downstream performance does not monotonically improve with

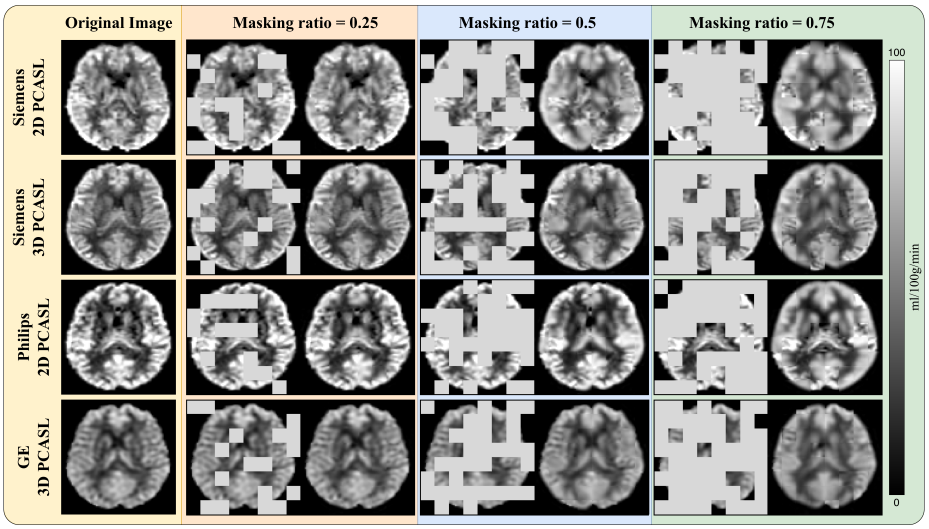

Fig. 4. Reconstruction results for pre-trained models trained with different masking ratios across imaging protocols and scanner vendors.

Table 2. Performance on CU A$\beta-$ vs CI A$\beta+$ task for different ρ values. (Unit: %)

Metrics	AUC ↑	Accuracy ↑	Precision ↑	Recall ↑	F1-score ↑
$\rho = 0.25$	76.74	69.54	55.42	**65.55**	56.94
$\rho = 0.5$	**78.93**	**73.54**	**57.90**	63.61	**59.37**
$\rho = 0.75$	71.90	69.44	50.32	61.66	54.67

reconstruction quality. To quantify this effect, we pre-trained ICHOR using $\rho \in \{0.25, 0.5, 0.75\}$ and adapted each pre-trained encoder to the CU A$\beta-$ vs CI A$\beta+$ classification task using the same downstream training protocol. As reported in Table 2, $\rho = 0.5$ achieves the best overall downstream performance across most metrics. This suggests a tradeoff in which low masking ratios make the reconstruction objective less informative, whereas high masking ratios provide insufficient visible context and make reconstruction overly challenging.

5 Conclusion

We presented ICHOR, a self-supervised pre-training framework for ASL-derived CBF maps. Across four downstream tasks, ICHOR achieves the best overall performance relative to pre-trained neuroimaging baselines trained on structural MRI and adapted to ASL. By leveraging a large and diverse multi-study, multi-site ASL dataset for pre-training, ICHOR addresses the limited availability of labeled ASL data and provides a transferable encoder for downstream ASL analyses. Future directions include extending ICHOR to quality enhancement tasks such as denoising and artifact correction, leveraging ICHOR representations for

longitudinal disease progression and treatment response modeling, and expanding the pre-training corpus to further strengthen generalization across diverse populations, protocols, and scanner platforms.

6 Disclosure of Interests

DAW reports receiving consulting fees from Eisai and Beckman Coulter and serving on a DSMB for GSK. The remaining authors declare no competing interests.

7 Acknowledgments

This work was supported by the National Institutes of Health under grant numbers P01-AG066597, R01-AG090414, R01-AG080734, P30-AG072979, R01-AG081693, R01-EB031080, R01-AG056014, R33-AG-080518, R01-EB031080. Additional support was provided by the DeCrane Family Primary Progressive Aphasia Fund, the Fred and Barbara Erb Foundation and the Alzheimer's Association (AACSF-23-1152241).

References

1. Chen, S., Ma, K., Zheng, Y.: Med3d: Transfer learning for 3d medical image analysis (2019), https://arxiv.org/abs/1904.00625
2. Cox, J., Liu, P., Stolte, S.E., Yang, Y., Liu, K., See, K.B., Ju, H., Fang, R.: Brainsegfounder: Towards 3d foundation models for neuroimage segmentation (2024), https://arxiv.org/abs/2406.10395
3. Detre, J.A., Leigh, J.S., Williams, D.S., Koretsky, A.P.: Perfusion imaging. Magnetic resonance in medicine **23**(1), 37–45 (1992)
4. Dolui, S., Tisdall, D., Vidorreta, M., Jacobs, D.R., Nasrallah, I.M., Bryan, R.N., Wolk, D.A., Detre, J.A.: Characterizing a perfusion-based periventricular small vessel region of interest. NeuroImage: Clinical **23**, 101897 (2019). https://doi.org/https://doi.org/10.1016/j.nicl.2019.101897, https://www.sciencedirect.com/science/article/pii/S2213158219302475
5. Gui, J., Chen, T., Zhang, J., Cao, Q., Sun, Z., Luo, H., Tao, D.: A survey on self-supervised learning: Algorithms, applications, and future trends (2024), https://arxiv.org/abs/2301.05712
6. He, K., Chen, X., Xie, S., Li, Y., Dollár, P., Girshick, R.: Masked autoencoders are scalable vision learners (2021), https://arxiv.org/abs/2111.06377
7. Hernandez-Garcia, L., Aramendía-Vidaurreta, V., Bolar, D.S., Dai, W., Fernández-Seara, M.A., Guo, J., Madhuranthakam, A.J., Mutsaerts, H., Petr, J., Qin, Q., Schollenberger, J., Suzuki, Y., Taso, M., Thomas, D.L., van Osch, M.J.P., Woods, J., Zhao, M.Y., Yan, L., Wang, Z., Zhao, L., Okell, T.W.: Recent technical developments in asl: A review of the state of the art. Magnetic Resonance in Medicine **88**(5), 2021–2042 (2022). https://doi.org/https://doi.org/10.1002/mrm.29381, https://onlinelibrary.wiley.com/doi/abs/10.1002/mrm.29381
8. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: Lora: Low-rank adaptation of large language models (2021), https://arxiv.org/abs/2106.09685
9. Mokhber, N., Shariatzadeh, A., Avan, A., Saber, H., Babaei, G.S., Chaimowitz, G., Azarpazhooh, M.R.: Cerebral blood flow changes during aging process and in cognitive disorders: A review. The Neuroradiology Journal **34**(4), 300–307 (2021). https://doi.org/10.1177/19714009211002778, https://doi.org/10.1177/19714009211002778, pMID: 33749402
10. Mormino, E.C., Biber, S.A., Rahman-Filipiak, A., Arfanakis, K., Clark, L., Dage, J.L., Detre, J.A., Dickerson, B.C., Donohue, M.C., Kecskemeti, S., Hohman, T.J., Jagust, W.J., Keene, D.C., Kukull, W., Levendovszky, S.R., Rosen, H., Thompson, P.M., Villemagne, V.L., Wolk, D.A., Okonkwo, O.C., Rabinvovici, G.D., Rivera-Mindt, M., Foroud, T., Johnson, S.C.: The consortium for clarity in adrd research through imaging (clariti). Alzheimer's & Dementia **21**(1), e14383 (2025). https://doi.org/https://doi.org/10.1002/alz.14383, https://alz-journals.onlinelibrary.wiley.com/doi/abs/10.1002/alz.14383
11. Petersen, R.C., Aisen, P.S., Beckett, L.A., Donohue, M.C., Gamst, A.C., Harvey, D.J., Jack, C.R., Jagust, W.J., Shaw, L.M., Toga, A.W., Trojanowski, J.Q., Weiner, M.W.: Alzheimer's disease neuroimaging initiative (adni). Neurology **74**(3), 201–209 (2010). https://doi.org/10.1212/WNL.0b013e3181cb3e25, https://www.neurology.org/doi/abs/10.1212/WNL.0b013e3181cb3e25
12. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., Liu, B., Matthews, P., Ong, G., Pell, J., Silman, A., Young, A., Sprosen, T., Peakman, T.,

Collins, R.: Uk biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLOS Medicine **12**(3), 1–10 (03 2015). https://doi.org/10.1371/journal.pmed.1001779, https://doi.org/10.1371/journal.pmed.1001779

13. Taghvaei, M., Dolui, S., Sadaghiani, S., Shakibajahromi, B., Brown, C., Khandelwal, P., Xie, S.X., Das, S., Yushkevich, P.A., Wolk, D.A., Detre, J.A.: Regional cerebral blood flow reflects both neurodegeneration and microvascular integrity across the alzheimer's continuum. Alzheimer's & Dementia **21**(1), e14382 (2025). https://doi.org/https://doi.org/10.1002/alz.14382, https://alz-journals.onlinelibrary.wiley.com/doi/abs/10.1002/alz.14382
14. Tak, D., Garomsa, B.A., Chaunzwa, T.L., Zapaishchykova, A., Climent Pardo, J.C., Ye, Z., Zielke, J., Ravipati, Y., Vajapeyam, S., Mahootiha, M., Smith, C., Familiar, A.M., Liu, K.X., Prabhu, S., Bandopadhayay, P., Nabavizadeh, A., Mueller, S., Aerts, H.J., Huang, R.Y., Poussaint, T.Y., Kann, B.H.: A foundation model for generalized brain mri analysis. medRxiv (2024). https://doi.org/10.1101/2024.12.02.24317992, https://www.medrxiv.org/content/early/2024/12/03/2024.12.02.24317992
15. Urbano, X., Taso, M., Nasrallah, I., Detre, J., Wang, Z., Dolui, S.: Qei-net: A deep learning-based automated quality evaluation index for asl cbf maps. https://doi.org/10.58530/2025/0730
16. Wang, Z., Aguirre, G.K., Rao, H., Wang, J., Fernández-Seara, M.A., Childress, A.R., Detre, J.A.: Empirical optimization of asl data analysis using an asl data processing toolbox: Asltbx. Magnetic Resonance Imaging **26**(2), 261–269 (2008). https://doi.org/https://doi.org/10.1016/j.mri.2007.07.003, https://www.sciencedirect.com/science/article/pii/S0730725X07003517
17. Wolf, R.L., Detre, J.A.: Clinical neuroimaging using arterial spin-labeled perfusion magnetic resonance imaging. Neurotherapeutics **4**(3), 346–359 (2007). https://doi.org/https://doi.org/10.1016/j.nurt.2007.04.005, https://www.sciencedirect.com/science/article/pii/S1878747923006360, advances in Neuroimaging/Neuroethics
18. Wolk, D.A., Detre, J.A.: Arterial spin labeling mri: an emerging biomarker for alzheimer's disease and other neurodegenerative conditions. Current opinion in neurology **25**(4), 421–428 (2012)
19. Zeng, X., Abdullah, N., Putra, S.: Self-supervised learning framework application for medical image analysis: a review and summary. BioMedical Engineering OnLine **23** (10 2024). https://doi.org/10.1186/s12938-024-01299-9