Surface-based Multi-Axis Longitudinal Disentanglement Using Contrastive Learning for Alzheimer's Disease

Jianwei Zhang[1,2] and Yonggang Shi[1,2]

[1] Stevens Neuroimaging and Informatics Institute, Keck School of Medicine, University of Southern California (USC), Los Angeles, CA 90033, USA
[2] Ming Hsieh Department of Electrical and Computer Engineering, Viterbi School of Engineering, University of Southern California (USC), Los Angeles, CA 90089, USA
yshi@loni.usc.edu

Abstract. Accurate modeling of disease progression is essential for comprehending the heterogeneous neuropathologies such as Alzheimer's Disease (AD). Traditional neuroimaging analysis often confound disease effects with normal aging, complicating the differential diagnosis. Recent advancements in deep learning have catalyzed the development of disentanglement techniques in Autoencoder networks, aiming to segregate longitudinal changes attributable to aging from those due to disease-specific alterations within the latent space. However, existing longitudinal disentanglement methods usually model disease as a single axis factor which ignores the complexity and heterogeneity of Alzheimer's Disease. In response to this issue, we propose a novel Surface-based Multi-axis Disentanglement framework.This framework posits multiple disease axes within the latent space, enhancing the model's capacity to encapsulate the multifaceted nature of AD, which includes various disease trajectories. To assign axes to data trajectories without explicit ground truth labels, we implement a longitudinal contrastive loss leveraging self-supervision, thereby refining the separation of disease trajectories. Evaluated on the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset (N=1321), our model demonstrates superior performance in delineating between cognitively normal (CN), mild cognitive impairment (MCI), and AD subjects,classification of stable MCI vs converting MCI and Amyloid status, compared to the single-axis model. This is further substantiated through an ablation study on the contrastive loss, underscoring the utility of our multi-axis approach in capturing the complex progression patterns of AD. The code is available at: https://github.com/jianweizhang17/MultiAxisDisentanglement.git

Keywords: Alzheimer's Disease · Disentanglement · Contrastive Learning.

1 Introduction

Alzheimer's Disease (AD) presents a significant challenge in neurodegenerative disease research due to its complex pathology and the intertwined effects of

normal aging and disease progression [1–3]. Traditional neuroimaging analysis methods primarily rely on statistical modeling to identify disease-related patterns [4]. However, these methods often assume specific data distributions that may not accurately reflect the underlying biological variability. In contrast, deep learning-based approaches can learn directly from data without strong distributional assumptions, demonstrating superior performance in tasks such as classification and regression. Despite their success, deep learning models suffer from interpretability issues, making it difficult to understand the learned feature representations and decision-making processes [5, 6].

To explicitly model the combined effects of aging and disease progression while leveraging deep learning's capacity to capture nonlinear relationships, various disentanglement-based methods have been proposed [7–10]. These approaches typically encode neuroimaging data into a low-dimensional latent space while enforcing human-interpretable structures to disentangle the effects of disease from other confounding factors. Recent studies have employed variational autoencoders (VAEs) and generative adversarial networks (GANs) to structure the latent space in meaningful ways for human interpretation, ensuring that variations due to aging and pathological changes are independently represented. For instance, the method proposed in [8] utilizes a clustering-classifier alternating strategy to encode atrophy patterns into discrete latent space clusters, yet it does not explicitly model the longitudinal progression of atrophy. Similarly, the model introduced in [7] represents aging and disease progression as orthogonal vectors in latent space, offering a clear framework for measuring disease progression speed. However, given the significant heterogeneity in AD progression patterns [11], assuming a single disease axis fails to capture the full spectrum of disease variability.

To address this limitation, we propose a surface-based multi-axis model that incorporates multiple disease progression axes to more accurately capture the heterogeneity of AD. Additionally, to mitigate the challenge posed by the lack of ground truth labels for different disease trajectories, we introduce a novel longitudinal contrastive loss, enabling self-supervised learning to effectively assign data trajectories to their respective axes. By implementing a multi-axis representation, our approach provides a more flexible and interpretable framework for modeling disease progression in Alzheimer's Disease. Through experiments, our model exhibit better detection power in differentiating CN, MCI and AD data and better accuracy in identifying stable MCI vs converting MCI and amyloid status.

2 Method

2.1 Model Structure

Following [7], the main idea of longitudinal disentanglement is to leverage orthogonality where the directions of aging and disease are represented as orthogonal axes in the latent space constructed by an encoder-decoder pair. The reasoning is that normal subjects should have a longitudinal trajectory in the latent space

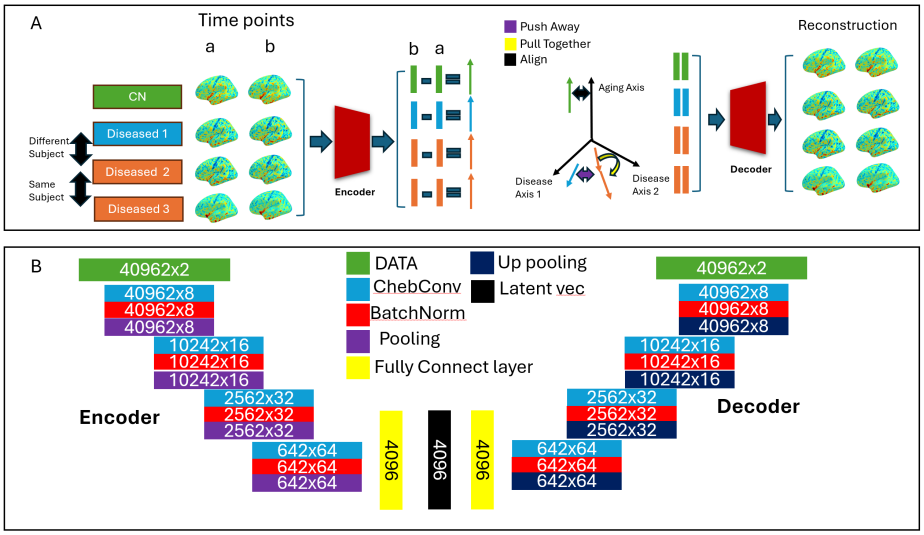

Fig. 1. An illustration of the proposed model. (A) The general frame work of the model. The 4 pairs of data, 1 pair is CN data, 3 are nonCN data, two of which are from the same subject, are encoded into trajectory in latent space and disentangled by alignment loss and contrastive loss. (B) The structure of the encoder-decoder built out of graph convolution, batch norm and icosahedron pooling and up pooling.

that is parallel to aging axis since normal data should be free of disease effect and disease subjects(Mild Cognitive impaired(MCI), Alzheimer's Disease(AD)) should have only disease component after subtracting the aging component. To adapt [7] to surface data, a graph convolution based encoder-decoder pair is designed. Given a series of N longitudinal cortical thickness data $x_0...x_N$, the model first encode the data into a latent vector space and then decode it to reconstruct for data fidelity. The network consists of an encoder, a decoder, an aging axis and N disease axes. The encoder and decoder each have 3 blocks of convolution layer, batch normalization layer, leaky ReLU layer and pooling or unpooling layer. To adapt the convolution from 3d volume to our cortical surface data, the ChebShev convolution [12] is chosen due to its robust and efficient properties. For pooling layer, we take advantage of the hierarchical structure of the icosahedron. To down sample from the ith order ico-sphere to the (i-1)th order, the max of the 1-ring neighborhood of every (i-1)th vertices is computed. For up pooling, we take the transposed convolution strategy. To up sample from the ith to (i+1)th ico-sphere, the additional vertices from the (i+1)th order is zero padded and followed by a convolution layer. There are two fully connected layers at the end of the encoder and start of the decoder for converting data into and out of the latent space. The latent space is formulated as a C-dimensional vector space.

2.2 Loss and Representation

Given an encoder and a decoder, E and D, an input data x can be encoded into a latent representation z and then reconstructed as $\hat{x}$. Our model takes, as input, four pairs of longitudinal data, (x_{1A}, x_{1B}), (x_{2A}, x_{2B}), (x_{3A}, x_{3B}), (x_{4A}, x_{4B}), where A corresponds to the earlier time point and B to the later time point. Each pair also has an associated time interval $(\Delta t)_i$, $i = 1, 2, 3, 4$. Pair 1 consists of cognitively normal (CN) data from the same subject. Pairs 2, 3, and 4 contain non-CN (MCI, AD) data, where pairs 2 and 3 come from the same subject, while pair 4 belongs to a different subject. All data are encoded into $(z_{1A}, z_{1B}), (z_{2A}, z_{2B}), (z_{3A}, z_{3B}), (z_{4A}, z_{4B})$. The difference value is defined as $\Delta z_i = z_{iB} - z_{iA}$ for $i = 1, 2, 3, 4$. The aging axis $\mathbf{a}$ and the set of N disease axes $\{\mathbf{d}_i\}_{i=1}^N$ are unit vectors in the latent space, with each $\mathbf{d}_i$ made orthogonal to $\mathbf{a}$ by subtracting any aging component. We do not require the $\mathbf{d}_i$ vectors to be mutually orthogonal, as we do not assume complete independence among different disease trajectories. We employ a combination of existing losses from prior work and our new contrastive loss. The first 5 losses are adapted from [7].

Reconstruction Loss $\mathcal{L}_{recon}$ To ensure that the latent representation retains essential information about the input, we apply a mean squared error (MSE) reconstruction loss to all four pairs. The loss is computed for both time points A and B:

$$\mathcal{L}_{\text{recon}} = \frac{1}{8} \sum_{i=1}^{4} \big(\|x_{iA} - \hat{x}_{iA}\|^2 + \|x_{iB} - \hat{x}_{iB}\|^2 \big). \tag{1}$$

Aging Direction Loss $\mathcal{L}_{da}$ To enforce a consistent aging trajectory, we apply the aging direction loss to the cognitively normal pair (pair 1). We encourage alignment with the aging vector $\mathbf{a}$ via cosine similarity:

$$\mathcal{L}_{\text{da}} = 1 - \frac{\Delta \mathbf{z}_1}{||\Delta \mathbf{z}_1||_2} * \mathbf{a}. \tag{2}$$

Disease Direction Loss $\mathcal{L}_{dd}$ To separate disease effects from aging, we first subtract the aging components off the diseased pair 2 latent difference Δz_2, which should then be parallel to one of the disease axis. In this case, we choose the disease axis $\{\mathbf{d}_i\}_{i=1}^N$ with the largest projection to be the target.

$$\Delta \mathbf{z}_{2,\text{disease}} \;=\; \Delta \mathbf{z}_2 \;-\; \left(\frac{\Delta \mathbf{z}_2 \cdot \mathbf{a}}{\|\mathbf{z}_2\|^2} \right) \mathbf{a}. \tag{3}$$

$$\mathcal{L}_{\text{dd}} = 1 \;-\; \max_i \; \frac{\Delta \mathbf{z}_{2,\text{disease}} \cdot \mathbf{d}_i}{||\Delta \mathbf{z}_{2,\text{disease}}||_2}, i = 1...N \tag{4}$$

KL Loss $\mathcal{L}_{kl}$ The kl loss is applied pair 1 and pair 2 under the assumption is that all subject should have a similar aging component distribution regardless of disease or normal. This loss is to ensure that disease subjects and normal

subjects have similar aging components. For $\Delta\mathbf{z}_1$ and $\Delta\mathbf{z}_2$, let $\alpha_i = \frac{\Delta\mathbf{z}_i \cdot \mathbf{a}}{\|\mathbf{a}\|}$ be its projection onto $\mathbf{a}$. Their KL divergence is:

$$\mathcal{L}_{\text{kl}} \;=\; \log\!\Big(\tfrac{\sigma_2}{\sigma_1}\Big) \;+\; \frac{\sigma_1^2 \;+\; \big(\mu_1 - \mu_2\big)^2}{2\,\sigma_2^2} \;-\; 0.5, \tag{5}$$

where $\{\mu_1, \sigma_1\}$ and $\{\mu_2, \sigma_2\}$ are the mean and standard deviation of α_1(CN) and α_2(nonCN), respectively.

Penalty Loss $\mathcal{L}_{pen}$ The penalty loss is to ensure that pair 1 (CN) subjects have as little disease components as possible, which is defined as ratio of sum of disease components and aging component. Ideally, CN data should have no disease components where this loss converges to 0.The loss is shown as below:

$$\mathcal{L}_{\text{pen}} = \frac{\sum_{i \in 1 \ldots N} ||\Delta\mathbf{z}_1 \cdot \mathbf{d}_i||}{||\Delta\mathbf{z}_1 \cdot \mathbf{a}||_2} \tag{6}$$

Contrastive Loss $\mathcal{L}_{contrast}$ We introduce a contrastive loss, inspired by SimCLR [13], to structure disease progression in a self-supervised manner. It enforces alignment between disease trajectories from the same subject while separating them from those of different subjects. Specifically, we operate on three shifts: $\Delta\mathbf{z}_{\text{base}}$ (pair 2), $\Delta\mathbf{z}_{\text{pos}}$ (pair 3), and $\Delta\mathbf{z}_{\text{neg}}$ (pair 4), where pair 2 and 3 from the same subject should share the same disease axis and pair 4 from a different subject should be pushed away from that same disease axis. We first subtract their aging components along $\mathbf{a}$ and normalize them in unit vectors, then project them onto $\{\mathbf{d}_i\}_{i=1}^N$ to obtain $\mathbf{q}_{\text{base}}, \mathbf{q}_{\text{pos}}, \mathbf{q}_{\text{neg}}$, where $\mathbf{q}$ is a length N vector of cosine similarity to each of the N disease axis. Finally, we define:

$$\mathcal{L}_{\text{contrast}} \;=\; -\log \frac{\exp\big(S_{\text{base,pos}}\big)}{\sum_j \exp\big(S_{\text{base},j}\big)}, \quad S_{ij} \;=\; \frac{\mathbf{q}_i \;\cdot\; \mathbf{q}_j}{\tau}, \tag{7}$$

where τ is the temperature parameter, and $S_{\text{base,pos}}$ measures the cosine similarity between $\mathbf{q}_{\text{base}}$ and $\mathbf{q}_{\text{pos}}$ which pull data from the same subject towards the same axis and push data from different subjects away from the same axis. The index j iterates through (base, pos, neg).

Total Loss Combining the above objectives, our final loss is a weighted sum:

$$\begin{aligned}\mathcal{L}_{\text{total}} \;=\; (\lambda_{\text{recon}}\,\mathcal{L}_{\text{recon}} \;+\; \lambda_{\text{da}}\,\mathcal{L}_{\text{da}} \;+\; \lambda_{\text{dd}}\,\mathcal{L}_{\text{dd}} \;+\; \lambda_{\text{kl}}\,\mathcal{L}_{\text{kl}} \\ +\; \lambda_{\text{pen}}\,\mathcal{L}_{\text{pen}} \;+\; \lambda_{\text{contrast}}\,\mathcal{L}_{\text{contrast}})\end{aligned} \tag{8}$$

In order to quantize disease severity for each group, we employ the same representation as [7]. For each pair of longitudinal data Δz, we compute the aging and disease speed as equation below:

$$V_{aging} = \frac{\Delta z \cdot \mathbf{a}}{\Delta t} \tag{9}$$

$$V_{disease} = \max_i \frac{\Delta z \cdot \mathbf{d_i}}{\Delta t} \tag{10}$$

Δt is the time interval between a pair of data. The aging and disease speed quantize the longitudinal changes at a time point, however, the exact value of the speed is not interpretable as it is a result of network optimization. The relative comparison among different groups, on the other hand, can provide us interpretability on the model's behavior.

2.3 Implementation

The network is implemented through Pytorch 2.3 [14] and torch geometric [15]. The detail dimension sizes are shown in Figure 1 B. The network is trained at learning rate of 0.0002 using Adam optimizer [16] for 90 epoches on a NVIDIA A5000 GPU. The weights for reconstruction loss, da loss, dd loss, kl loss, penalty loss and contrast loss are 4,1,1,1,1,1.

3 Experiments and Results

3.1 Dataset Processing

All experiments are conducted on the Alzheimer's Disease Neuroimaging Initiative(ADNI) dataset [17]. In total 5749 T1 weighted MRI scans from 1321 subjects are selected. Each scan is assigned a diagnosis label of CN, MCI and AD based on closest timestamp. Subjects with inconsistent longitudinal diagnosis, such as AD subjects turning into MCI, are removed to reduce noise. For CN group, only subjects that remain CN on all scans are kept. In MCI group, subjects are divided into stable MCI(sMCI) and converting MCI(cMCI) where sMCI remains MCI on all scans and cMCI has transition in diagnosis from MCI to AD. After the screening, there are 495 CN, 446 sMCI, 148 cMCI and 232 AD. All the T1 MRI images are processed through FreeSurfer 6.0 to extract cortical thickness of left and right hemisphere. [18] The thickness maps are registered by FreeSurfer in the spherical domain and resampled to a standard 6th order icosahedron of 40962 vertices using FreeSurfer *mris_surf2surf* command. We use data from both hemispheres which will be represented as a 40962 by 2 tensor as input to the model.

3.2 Disentanglement Evaluation

In this experiment, we want to evaluate the impact of the number of disease axis on aging and disease disentanglement, as well as the effect of the contrastive loss. We follow the training scheme described in the method section and only vary the number of disease axis. During training, the model is only provided CN vs nonCN labels, therefore, we can rate the model's performance based on their ability to differentiate CN vs MCI and CN vs AD. After determining the optimal axis number, we train the same model without the contrastive loss for ablation

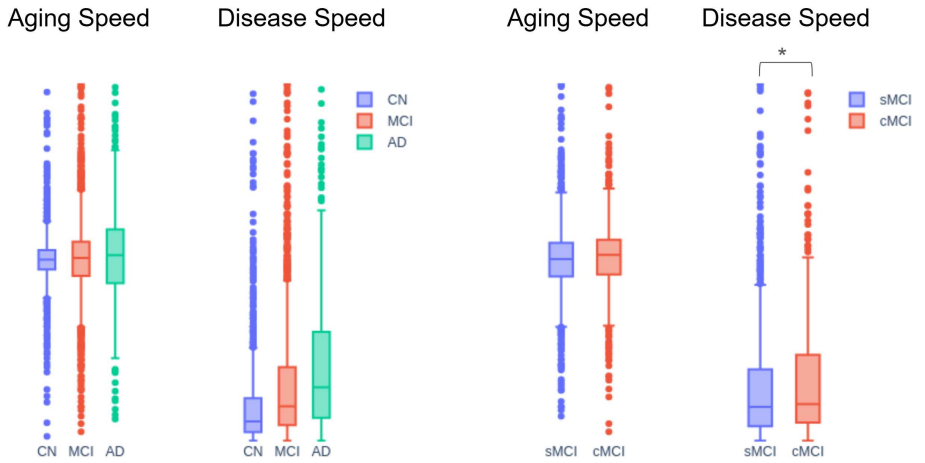

Fig. 2. Distribution of aging and disease speed among CN, MCI, AD, as well as sMCI and cMCI subjects. Between each group, the aging speed is similar while disease varies across different disease groups. The y axis is not shown and is different for each graph since there is no intrinsic meaning in magnitude of aging or disease speed, only in relative term.

study. The results are shown in Table 1. We compute the disease speed from each model using equation 9 and 10 and compute the Cohen'd and p-value from t-test as evaluation metric. The results show that the optimal number of disease axis is 2, which has a higher cohen's d and lower p-value in both CN vs MCI and CN vs AD. The 2-axis model also shows superior performance than single axis which demonstrate the necessity of multi-axis system for disentanglement. Table 1 also shows that the 2 axis model with contrast loss performs better than that without contrast loss which confirms that the contrast loss is an essential addition for increased performance. Figure 2 shows the distribution of aging and disease speed among differnt groups from the optimal multi axis model.

Model	CN vs MCI Disease Speed		CN vs AD Disease Speed	
	p-value ↓	Cohen's d ↑	p-value ↓	Cohen's d ↑
1 Axis	1.5551e-15	0.3079	1.3179e-32	0.5405
2 Axis	**1.1517e-19**	**0.3431**	**1.7353e-36**	**0.6271**
3 Axis	4.3966e-15	0.2965	2.2377e-27	0.5262
4 Axis	4.0447e-18	0.3273	1.0465e-34	0.6106
5 Axis	1.4278e-12	0.2590	5.6348e-14	0.4096
2 Axis w/o Contrast	3.6444e-13	0.2682	1.5783e-20	0.4971

Table 1. Disentanglement evaluation: This table shows the comparison of multi-axis with different axis number and single axis in terms of the respective disease speed's ability to distinguish CN vs MCI and CN vs AD.

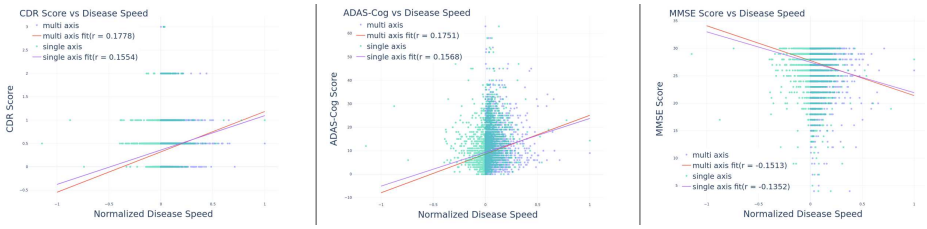

Fig. 3. Correlation between the disease speed and 3 Cognitive Scores: ADAS, CDR and MMSE. The correlation values are shown in the legend. The disease speeds from our model and single axis model are used for comparison. All three scores show higher correlation with our model than the single axis model.

3.3 Correlation with Cognitive Scores

We also performed correlation test of the disease speed with the 3 types of cognitive score,Alzheimer's Disease Assessment Scale(ADAS), Clinical Dementia Rating(CDR) and Mini-Mental State Examination(MMSE). Going through all the longitudinal pairs of data, we select the earlier data from all pair to ensure that each scan only has 1 disease speed for consistency. Figure 3 shows the plot of disease speed against 3 cognitive scores for both single axis and multi axis model, which is the optimal model picked from the previous section. The correlation(r) between disease speed and cognitive scores are shown on the graph. The results show that multi-axis model consistently yields higher correlation than single axis model, which indicates that the multi-axis has higher clinical relevance.

3.4 Classification on sMCI vs cMCI and Amyloid status

We also performed classification of sMCI vs cMCI and amyloid status defined by [19]. As baseline comparison, we also trained a pure Autoencoder [20] model

Model	sMCI vs cMCI balanced Accuracy ↑	Amyloid Status balanced Accuracy↑
AE [20]	0.5251	0.5183
Single axis	0.5690	0.5205
Ours	**0.5948**	**0.5365**

Table 2. Classification of sMCI vs cMCI and Amyloid status. The table shows the classification accuracy on sMCI vs cMCI and Amyloid Status using latent code from vanilla autoencoder, Single Axis and Multi-Axis(ours).

without regularization on the latent space. All the data are transformed into 4096 latent vector using the respective encoder. The classifier is a simple 3 layer multi-layer perception with RELU as activation function and cross entropy as loss. The classifiers are all trained for 20 epoches. To balance the dataset, we randomly picked equal amount of sMCI data to match that of the cMCI data which has less

number of data. For cMCI, we only pick time points before their transition to AD to ensure no data leakage from post-conversion stages.For amyloid, we use the entire dataset. Table 2 shows that resulting accuracy of all three models. Multi axis model achieved the highest classification accuracy in both sMCI vs cMCI and amyloid status, which indicates that multi-axis model can better represent biological status in diseased subject.

4 Conclusion

In conclusion, this Surface-based Multi-axis Disentanglement framework successfully differentiates normal aging from AD-specific alterations by learning multiple disease axes in the latent space. Leveraging a longitudinal contrastive loss for self-supervision allows the model to capture the multifaceted nature of AD, including distinct subtypes and trajectories. Evaluation on the ADNI dataset confirms that our model more accurately separates CN, MCI, and AD subjects than single-axis models and robustly handles the complexity of AD progression. These findings highlight the value of multi-axis disentanglement for characterizing heterogeneous neuropathologies and underscore the promise of this approach for deeper insights into disease evolution and improved diagnostic stratification.

Acknowledgments. This work was supported by the National Institute of Health (NIH) under grants R01EB022744, RF1AG077578, RF1AG064584, U19AG078109, and P30AG066530.

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

References

1. Yana Blinkouskaya and Johannes Weickenmeier. Brain shape changes associated with cerebral atrophy in healthy aging and alzheimer's disease. *Frontiers in Mechanical Engineering*, 7:705653, 2021.
2. Takashi Ohnishi, Hiroshi Matsuda, Takeshi Tabira, Takashi Asada, and Masatake Uno. Changes in brain morphology in alzheimer disease and normal aging: is alzheimer disease an exaggerated aging process? *American Journal of Neuroradiology*, 22(9):1680–1685, 2001.
3. Ilsa Rovira, Arya Biragyn, LaVerne Brown, Zorina S Galis, Malgorzata Klauzinska, Svetlana Kotliarova, Janine M Simmons, Anil Wali, Ronit Yarden, Dan Xi, et al. Health and aging trajectories: Shared and competing risks and resiliencies for chronic diseases associated with aging. a trans-nih workshop. *Frontiers in Public Health*, 13:1462217.
4. Yuria Utsumil, Ognjen Oggi Rudovicl, Kelly Petersonl, Ricardo Guerrero, and Rosalind W. Picardl. Personalized gaussian processes for forecasting of alzheimer's disease assessment scale-cognition sub-scale (adas-cog13). In *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 4007–4011, 2018.

5. Ahmed Elnakib, Fahmi Khalifa, Ahmed Soliman, Ahmed Shalaby, and Mostafa Elhosseini. Emerging artificial intelligence technologies for neurological and neuropsychiatric research. *Frontiers in Neuroscience*, 18:1518442, 2024.
6. Qian Zhang, XiaoLi Yang, and ZhongKui Sun. Classification of alzheimer's disease progression based on smri using gray matter volume and lateralization index. *Plos one*, 17(3):e0262722, 2022.
7. Jiahong Ouyang, Qingyu Zhao, Ehsan Adeli, Greg Zaharchuk, and Kilian M. Pohl. Disentangling Normal Aging From Severity of Disease via Weak Supervision on Longitudinal MRI. *IEEE Transactions on Medical Imaging*, 41(10):2558–2569, 2022.
8. Sohyun Kang, Sung-Woo Kim, Joon-Kyung Seong, and Alzheimer's Disease Neuroimaging Initiative. Disentangling brain atrophy heterogeneity in Alzheimer's disease: A deep self-supervised approach with interpretable latent space. *NeuroImage*, 297:120737, 2024.
9. Qingyu Zhao, Zixuan Liu, Ehsan Adeli, and Kilian M Pohl. Longitudinal self-supervised learning. *Medical image analysis*, 71:102051, 2021.
10. Gurur Gamgam, Alkan Kabakcioglu, Demet Yüksel Dal, and Burak Acar. Disentangled attention graph neural network for alzheimer's disease diagnosis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 219–228. Springer, 2024.
11. Seyed Hani H Hojjati, Gloria Chinag, Tracy Butler, Bardiya Ghaderi Yazdi, Sindy Ozoria, Peter Chernek, Jenseric Calimag, Yaakov Stern, Devangere Devanand, Sonja Blum, et al. Tau progression index (tpi): An individualized, clinically applicable, multimodally-derived score to predict ad progression. *medRxiv*, pages 2025–01, 2025.
12. Michaël Defferrard, Xavier Bresson, and Pierre Vandergheynst. Convolutional neural networks on graphs with fast localized spectral filtering. *Advances in neural information processing systems*, 29, 2016.
13. Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PmLR, 2020.
14. Adam Paszke, Sam Gross, Soumith Chintala, Gregory Chanan, Edward Yang, Zachary DeVito, Zeming Lin, Alban Desmaison, Luca Antiga, and Adam Lerer. Automatic differentiation in pytorch. 2017.
15. Matthias Fey and Jan E. Lenssen. Fast graph representation learning with PyTorch Geometric. In *ICLR Workshop on Representation Learning on Graphs and Manifolds*, 2019.
16. Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
17. Susanne G. Mueller, Michael W. Weiner, Leon J. Thal, Ronald C. Petersen, Clifford Jack, William Jagust, John Q. Trojanowski, Arthur W. Toga, and Laurel Beckett. The alzheimer's disease neuroimaging initiative. *Neuroimaging Clinics of North America*, 15(4):869–877, 2005. Alzheimer's Disease: 100 Years of Progress.
18. Anders M. Dale, Bruce Fischl, and Martin I. Sereno. Cortical surface-based analysis: I. segmentation and surface reconstruction. *NeuroImage*, 9(2):179–194, 1999.
19. Susan M Landau, Mark A Mintun, Abhinay D Joshi, Robert A Koeppe, Ronald C Petersen, Paul S Aisen, Michael W Weiner, William J Jagust, and Alzheimer's Disease Neuroimaging Initiative. Amyloid deposition, hypometabolism, and longitudinal cognitive decline. *Annals of neurology*, 72(4):578–586, 2012.
20. G. E. Hinton and R. R. Salakhutdinov. Reducing the dimensionality of data with neural networks. *Science*, 313(5786):504–507, 2006.