Adaptive Neighborhoods in Contrastive Regression Learning for Brain Age Prediction

Jakob Träuble
University of Cambridge

Lucy V Hiscox
Cardiff University

Curtis Johnson
University of Delaware

Carola-Bibiane Schönlieb
University of Cambridge

Gabriele S Kaminski Schierle
University of Cambridge

Angelica I Aviles-Rivero
University of Cambridge

Abstract

In neuroimaging, accurate brain age prediction is key to understanding brain aging and early neurodegenerative signs. Recent advancements in self-supervised learning, particularly contrastive learning, have shown robustness with complex datasets but struggle with non-uniformly distributed data common in medical imaging. We introduce a novel contrastive loss that dynamically adapts during training, focusing on localized sample neighborhoods. Additionally, we incorporate brain stiffness, a mechanical property sensitive to aging. Our approach outperforms state-of-the-art methods and opens new directions for brain aging research.

1 Introduction

Brain age prediction uses neuroimaging data to establish a baseline aging trajectory from healthy samples, casting it as a regression problem. This approach is particularly promising for identifying deviations from normal aging processes that might indicate neurological conditions. Inspired by advancements in computer vision, self-supervised learning techniques, particularly contrastive learning methods, have been effectively adapted for predicting brain age from structural MRI scans [1, 2].

Limitations. Despite their potential, current methods often struggle with generalization, particularly across datasets characterized by non-uniform distributions. To address this limitation, we introduce a novel contrastive loss that focuses on localized sample neighborhoods and adapts dynamically during training, enhancing performance where traditional approaches falter. Given the greater age sensitivity of mechanical over structural properties [3–10] *this is the first application of contrastive learning to brain stiffness maps*, opening new directions in neuroimaging research.

Contrastive Learning. To construct semantically rich and structured representations, contrastive learning has become a widely adopted method for self-supervised representation learning. Contrastive learning adjusts distances in the embedding space to bring similar samples closer and push dissimilar ones apart [11, 12]. Early methods like SimCLR [11] introduced a simple yet effective framework that utilized a contrastive loss function to maximize agreement betweeen differently augmented views of the same data sample (positive pairs) while pushing apart representations of different samples (negative pairs). Subsequently, NNCLR [13] pairs each sample with its nearest neighbor in the feature space, focusing on informative similarities between samples.

Shifting from classification to regression problems, the distinction between positive and negative pairs transitions to a continuous spectrum. This shift necessitates the model's ability to discern varying degrees of similarity, represented as $s_{i,k} = \text{sim}(f(x_i), f(x_k))$, beyond mere categorical differentiation. In response to the challenge of integrating continuous labels such as age, recent advancements propose strategies such as the Y-Aware loss [1], which softens the boundary between positive and negative samples. Similarly, [2] proposed the Threshold and Exponential losses, which

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

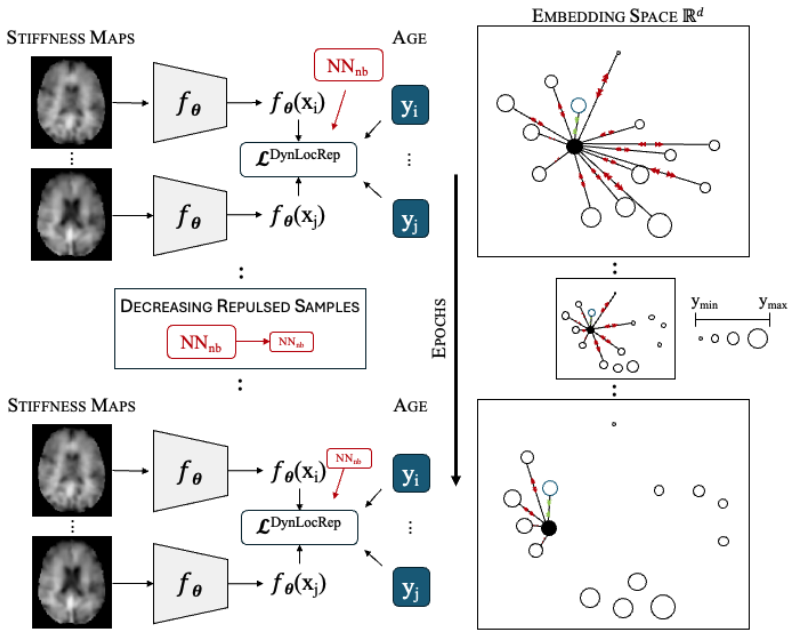


Figure 1: Graphical Illustration of Our Proposed Method. Throughout training (top to bottom), the repulsion is progressively localized as the number of samples, selected as nearest neighbors, are gradually decreased.

adjust the strength of alignment and repulsion based on the similarity between continuous labels. Another work introduced the Rank-N-Contrast loss [14], which employs a comparative ranking strategy among samples. This method ranks samples based on their similarity to a given anchor, creating a ranking-based continuous spectrum of positive and negative pairs. See Appendix A.1 for detailed loss descriptions.

2 Proposed Technique

Problem Setup. The primary challenge in brain age modeling lies in accurately mapping high-dimensional brain imaging data to a continuous age variable. Traditional contrastive learning methods are limited in their capacity to handle the subtle variations in brain stiffness associated with aging due to their global approach. Our technique introduces a dynamic, localized strategy to progressively capture age-related features effectively.

Formally, we aim to train a neural network mapping brain images $x \in \mathcal{X}$ to target ages $y \in \mathbb{R}$. The model comprises two key components: a feature encoder $f : \mathcal{X} \to \mathcal{Z}$, which transforms brain images into an embedding space $\mathcal{Z} \subseteq \mathbb{R}^d$, and an age predictor $g : \mathcal{Z} \to \mathbb{R}$, tasked with estimating the age from these features.

Dynamic Localized Repulsion. The dynamic localized repulsion technique is a cornerstone of our methodology, designed to optimize the contrastive learning framework specifically for the regression tasks inherent in brain age modeling.

To enhance the precision of contrastive regression learning, we introduce a dynamic localized repulsion approach that progressively explores varied scales within the embedding space, as depicted in Figure 1. This methodology systematically adjusts the selection of repulsion candidates, taking into account both their proximity and the evolutionary stage of training. The selection process for repulsed samples is defined by:

$$NN(x_i; \text{epoch}) = \{x_k \mid f_{epoch}(x_k) \text{ is among the } NN_{nb}(\text{epoch}) \text{ nearest neighbors of } f_{epoch}(x_i) \text{ based on } d(f_{epoch}(x_k), f_{epoch}(x_i))\} \quad (1)$$

This defines the samples subject to repulsion by distance d. Our approach narrows the scope of nearest neighbors in repulsion, focusing learning on increasingly localized neighborhoods.

This adaptive mechanism is governed by two critical hyperparameters: the final count of nearest neighbors,$NN_{nb,final}$, representing the ultimate scope of repulsion at the end of training, and the decrement frequency, $NN_{\text{step size}}$, which specifies the interval of epochs for adjustments in the neighbor count, as detailed in Appendix A.3.

In datasets that show non-uniform distributions, especially those with multi-modal characteristics, it is common to find some target areas oversampled and others undersampled. This scenario is typical in neuroimaging datasets (see Appendix A.2.2). Our dynamic localized strategy is designed to address this issue. It starts by segregating distinct groups and then the training objective evolves to focus exclusively on those groups. This process is illustrated in Fig. 1. Our approach aims to reveal more coherent representations throughout the dataset.

Following the methodology proposed by [2], we utilize kernel functions to determine the degrees of positiveness, $w_{i,k} = K(y_i - y_k)$, where $0 \leq w_{i,k} \leq 1$. Larger values of $w_{i,k}$ indicate closer age similarity, bringing their representations closer together. The set of nearest neighbors, $NN_{nb}(\text{epoch})$, is dynamically adjusted during training, and repulsion is applied only to the nearest neighbors, progressively focusing on more localized interactions as training advances. The final dynamic localized repulsion loss is:

$$\mathcal{L}_{DynLocRep} = -\sum_i \sum_{k \neq i} \frac{w_{i,k}}{\sum_t w_{i,t}} \log \left(\frac{\exp(s_{i,k})}{\sum_{x_t \in NN(x_i;\text{epoch})} \exp(s_{i,t}(1 - w_{i,t}))} \right) \tag{2}$$

This overall loss calculates the aggregated contribution of each sample pair within a batch. It normalizes these contributions by the sum of positiveness weights, reflecting age similarity, to adjust the influence of each pair. The softmax function is then applied to these normalized and adjusted similarity scores $s_{i,k}$, which are recalculated for each dynamically defined nearest neighbor set.

What is the Intuition Behind Our Dynamic Localized Repulsion? We address the challenge of non-uniform data distributions in neuroimaging datasets. Traditional models often fail to distinguish between age groups when data is unevenly represented. Our approach refines this by adjusting embeddings dynamically. This is done via (2) which aims to dynamically adjust the embeddings based on age-related similarities. The essence of this equation lies in its ability to modulate the degree of repulsion or attraction between samples within the same batch based on their age proximity $w_{i,k}$. *This formulation allows for adaptive learning where the focus is progressively shifted toward more challenging or informative pairs, potentially those that are not well-aligned in age, thus encouraging the model to learn finer distinctions as training progresses.*

3 Experimental Results

We assembled a dataset of 311 3D brain stiffness maps from healthy control subjects, sourced from multiple clinical studies [15–20]. For detailed information on the data and pre-processing, see Appendix A.2.

Evaluation Protocol. We used a 3D ResNet-18 model (33.5M parameters), pre-trained on over 5000 T_1 3D MRI brain images from the openBHB dataset [21], using the best reported method from the OpenBHB challenge [21]. We fully fine-tune (i.e. updating all weights) the pre-trained ResNet-18 on our brain stiffness dataset and evaluated the learned representations using a Ridge Regression estimator [2] to predict age. As an evaluation metric, we calculated the mean absolute error (MAE) on the test set, averaging the results across five random seeds. Further details can be found in Appendix A.4.

Results and Discussion. We begin by evaluating the representations of stiffness maps learned using our dynamic localized repulsion loss against those using current state-of-the-art classification and contrastive regression losses. Table 1 demonstrates the effectiveness of our approach, evidenced by the Mean Absolute Error (MAE) metric. Notably, our method significantly outperforms contrastive classification losses such as SimCLR [11] and NNCLR [13], which achieve higher MAEs of 9.600 ± 1.701 and 8.526 ± 1.442, respectively.

When comparing our method to existing state-of-the-art contrastive regression losses, our approach demonstrates superior accuracy in predicting brain age. Upon careful examination, we can observe that Rank-N-Contrast [14] shows the highest MAE, suggesting it may be less adept at capturing the nuanced patterns within the data necessary for precise age prediction. Y-Aware and Exponential [1, 2]

Table 1: Representation comparison demonstrates the superior performance of our method over state-of-the-art contrastive classification and regression losses.

Contrastive Classification Loss	**MAE [years]**
SimCLR [22]	9.600 ± 1.701
NNCLR [13]	8.526 ± 1.442
Contrastive Regression Loss	
Rank-N-Contrast [14]	5.266 ± 0.587
Y-Aware [1, 2]	3.852 ± 0.212
Threshold [2]	4.420 ± 0.503
Exponential [1, 2]	3.824 ± 0.215
Dynamical Localized Repulsion (Ours)	$\mathbf{3.724 \pm 0.220}$

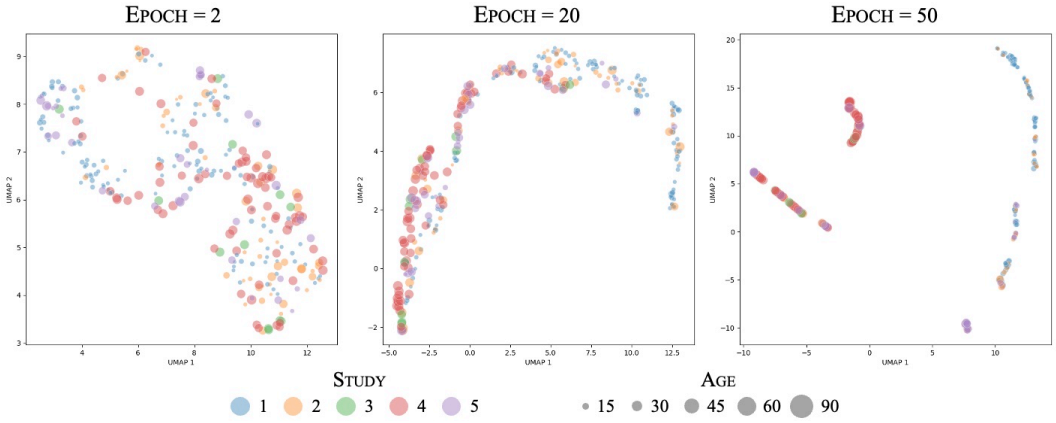


Figure 2: UMAP visualizations of representations show model improvements throughout various learning stages. As epochs increase, the clusters become more distinct and separate, indicating a more defined representation of the underlying data features.

losses show improvements over Rank-N-Contrast. These methods appear to better align with the underlying age-related changes in brain stiffness but still fall short compared to our approach. Threshold [2] loss offers a competitive performance, yet it does not achieve the same level of accuracy as our method. This indicates that while these methods handle some data variability effectively, they might not fully capture localized age-related changes as our method does, which achieves the lowest MAE. This shows that focusing on localized neighborhoods and dynamic adaptation significantly contribute to its improved performance. Refer to Appendix A.5 for ablation studies on various settings.

Fig. 2 shows UMAP embeddings across epochs 2, 20 and 50, illustrating the progressive refinement of the feature space. Initially scattered, the representations form more defined clusters as training progresses, reflecting the model's increasing ability to capture age-related variations. By epoch 50, distinct groupings indicate a better understanding of underlying age-related features. These visualizations confirm our approach's effectiveness and offer intuitive insights into how contrastive learning can be harnessed for regression tasks in medical imaging.

4 Conclusion

We introduced a dynamic localized repulsion approach for contrastive regression learning, addressing generalization challenges in medical imaging with non-uniform data distributions. Applied to brain stiffness maps, our method enhances robustness and performance in brain age prediction. *Our research marks the first application of self-supervised learning to explore mechanical brain properties*, opening avenues for understanding structural changes related to aging and neurological conditions.

Future work includes expanding to neurological disease cohorts and integrating multimodal imaging data for better prediction accuracy.

Acknowledgments

JT acknowledges support from the Gates Cambridge Trust via the Gates Cambridge Scholarship. CJ acknowledges partial support from the National Institutes of Health grant R01-AG058853. GSKS acknowledges funding from the Wellcome Trust (065807/Z/01/Z) (203249/Z/16/Z), the UK Medical Research Council (MRC) (MR/K02292X/1), ARUK (ARUK-PG013-14), Michael J Fox Foundation (16238; 022159), and Infinitus China Ltd. LVH is supported by the Wellcome Trust (grant number: 226420/Z/22/Z).

References

[1] Benoit Dufumier, Pietro Gori, Julie Victor, Antoine Grigis, Michele Wessa, Paolo Brambilla, Pauline Favre, Mircea Polosan, Colm McDonald, Camille Marie Piguet, et al. Contrastive learning with continuous proxy meta-data for 3d mri classification. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part II 24*, pages 58–68. Springer, 2021.

[2] Carlo Alberto Barbano, Benoit Dufumier, Edouard Duchesnay, Marco Grangetto, and Pietro Gori. Contrastive learning for regression in multi-site brain age prediction. In *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, pages 1–4. IEEE, 2023.

[3] Lucy V Hiscox, Hillary Schwarb, Matthew DJ McGarry, and Curtis L Johnson. Aging brain mechanics: Progress and promise of magnetic resonance elastography. *Neuroimage*, 232: 117889, 2021.

[4] Ingolf Sack, Kaspar-Josche Streitberger, Dagmar Krefting, Friedemann Paul, and Jürgen Braun. The influence of physiological aging and atrophy on brain viscoelastic properties in humans. *PloS one*, 6(9):e23451, 2011.

[5] Matthew C Murphy, John Huston III, Clifford R Jack Jr, Kevin J Glaser, Matthew L Senjem, Jun Chen, Armando Manduca, Joel P Felmlee, and Richard L Ehman. Measuring the characteristic topography of brain stiffness with magnetic resonance elastography. *PloS one*, 8(12):e81668, 2013.

[6] Arvin Arani, Matthew C Murphy, Kevin J Glaser, Armando Manduca, David S Lake, Scott A Kruse, Clifford R Jack Jr, Richard L Ehman, and John Huston 3rd. Measuring the effects of aging and sex on regional brain stiffness with mr elastography in healthy older adults. *Neuroimage*, 111:59–64, 2015.

[7] Tomohiro Takamura, Utaroh Motosugi, Yu Sasaki, Takashi Kakegawa, Kazuyuki Sato, Kevin J Glaser, Richard L Ehman, and Hiroshi Onishi. Influence of age on global and regional brain stiffness in young and middle-aged adults. *Journal of Magnetic Resonance Imaging*, 51(3): 727–733, 2020.

[8] Lucy V Hiscox, Curtis L Johnson, Matthew DJ McGarry, Michael Perrins, Aimee Littlejohn, Edwin JR van Beek, Neil Roberts, and John M Starr. High-resolution magnetic resonance elastography reveals differences in subcortical gray matter viscoelasticity between young and healthy older adults. *Neurobiology of aging*, 65:158–167, 2018.

[9] Peyton L Delgorio, Lucy V Hiscox, Ana M Daugherty, Faria Sanjana, Ryan T Pohlig, James M Ellison, Christopher R Martens, Hillary Schwarb, Matthew DJ McGarry, and Curtis L Johnson. Effect of aging on the viscoelastic properties of hippocampal subfields assessed with high-resolution mr elastography. *Cerebral Cortex*, 31(6):2799–2811, 2021.

[10] Rebecca G Clements, Claudio Cesar Claros-Olivares, Grace McIlvain, Austin J Brockmeier, and Curtis L Johnson. Mechanical property based brain age prediction using convolutional neural networks. *bioRxiv*, pages 2023–02, 2023.

[11] 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.

[12] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020.

[13] Debidatta Dwibedi, Yusuf Aytar, Jonathan Tompson, Pierre Sermanet, and Andrew Zisserman. With a little help from my friends: Nearest-neighbor contrastive learning of visual representations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9588–9597, 2021.

[14] Kaiwen Zha, Peng Cao, Jeany Son, Yuzhe Yang, and Dina Katabi. Rank-n-contrast: Learning continuous representations for regression. *Advances in Neural Information Processing Systems*, 36, 2024.

[15] Lucy V Hiscox, Matthew DJ McGarry, Hillary Schwarb, Elijah EW Van Houten, Ryan T Pohlig, Neil Roberts, Graham R Huesmann, Agnieszka Z Burzynska, Bradley P Sutton, Charles H Hillman, et al. Standard-space atlas of the viscoelastic properties of the human brain. *Human brain mapping*, 41(18):5282–5300, 2020.

[16] Philip V Bayly, Ahmed Alshareef, Andrew K Knutsen, Kshitiz Upadhyay, Ruth J Okamoto, Aaron Carass, John A Butman, Dzung L Pham, Jerry L Prince, KT Ramesh, et al. Mr imaging of human brain mechanics in vivo: new measurements to facilitate the development of computational models of brain injury. *Annals of biomedical engineering*, 49:2677–2692, 2021.

[17] Faria Sanjana, Peyton L Delgorio, Lucy V Hiscox, Theodore M DeConne, Joshua C Hobson, Matthew L Cohen, Curtis L Johnson, and Christopher R Martens. Blood lipid markers are associated with hippocampal viscoelastic properties and memory in humans. *Journal of Cerebral Blood Flow & Metabolism*, 41(6):1417–1427, 2021.

[18] Peyton L Delgorio, Lucy V Hiscox, Grace McIlvain, Mary K Kramer, Alexa M Diano, Kyra E Twohy, Alexis A Merritt, Matthew DJ McGarry, Hillary Schwarb, Ana M Daugherty, et al. Hippocampal subfield viscoelasticity in amnestic mild cognitive impairment evaluated with mr elastography. *NeuroImage: Clinical*, 37:103327, 2023.

[19] Lucy V Hiscox, Curtis L Johnson, Matthew DJ McGarry, Helen Marshall, Craig W Ritchie, Edwin JR Van Beek, Neil Roberts, and John M Starr. Mechanical property alterations across the cerebral cortex due to alzheimer's disease. *Brain Communications*, 2(1):fcz049, 2020.

[20] Peyton L Delgorio, Lucy V Hiscox, Ana M Daugherty, Faria Sanjana, Grace McIlvain, Ryan T Pohlig, Matthew DJ McGarry, Christopher R Martens, Hillary Schwarb, and Curtis L Johnson. Structure–function dissociations of human hippocampal subfield stiffness and memory performance. *Journal of Neuroscience*, 42(42):7957–7968, 2022.

[21] Benoit Dufumier, Antoine Grigis, Julie Victor, Corentin Ambroise, Vincent Frouin, and Edouard Duchesnay. Openbhb: a large-scale multi-site brain mri data-set for age prediction and debiasing. *NeuroImage*, 263:119637, 2022.

[22] 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.

[23] Bruce Fischl. Freesurfer. *Neuroimage*, 62(2):774–781, 2012.

[24] Brian B Avants, Nick Tustison, Gang Song, et al. Advanced normalization tools (ants). *Insight j*, 2(365):1–35, 2009.

[25] Rasmus Rothe, Radu Timofte, and Luc Van Gool. Dex: Deep expectation of apparent age from a single image. In *Proceedings of the IEEE international conference on computer vision workshops*, pages 10–15, 2015.

A Appendix

A.1 State-of-the-Art Contrastive Regression Losses

In this Appendix, we provide detailed descriptions of the state-of-the-art contrastive regression losses referenced in the main text. These methods adapt contrastive learning for regression tasks by modifying the loss functions to handle continuous labels.

Method	Contrastive Regression Loss
Rank-N-Contrast	$\mathcal{L}^{RnC} = -\sum_i \sum_{k \neq i} \log \frac{\exp(s_k)}{\sum_{x_{i,t} \in S_{i,j}} \exp(s_{i,t})}$ with $S_{i,j} := \{x_k \| k \neq i, d(y_i, y_k) \geq d(y_i, y_j)\}$
Y-Aware	$\mathcal{L}^{y-aware} = -\sum_i \sum_{k \neq i} \frac{w_{i,k}}{\sum_t w_{i,t}} \log \left(\frac{\exp(s_{i,k})}{\sum_{t \neq k} \exp(s_{i,t})} \right)$
Threshold	$\mathcal{L}^{threshold} = -\sum_i \sum_{k \neq i} \frac{w_{i,k}}{\sum_t \delta_{w_{i,t} < w_{i,k}} w_{i,t}} \log \left(\frac{\exp(s_{i,k})}{\sum_{t \neq k} \delta_{w_{i,t} < w_{i,k}} \exp(s_{i,t})} \right)$
Exponential	$\mathcal{L}^{exp} = -\sum_i \sum_{k \neq i} \frac{w_{i,k}}{\sum_t w_{i,t}} \log \left(\frac{\exp(s_{i,k})}{\sum_{t \neq k} \exp(s_{i,t}(1 - w_{i,t}))} \right)$

Table 2: Overview of Contrastive Regression Losses. This details existing methods each employing distinct strategies to refine the contrastive learning process for regression tasks.

Descriptions:

- **Rank-N-Contrast loss** [14]: Construct positive and negative pairs based on their ranking of label differences, focusing on labels with smaller label distances.
- **Y-Aware loss** [1]: Uses a weighting function $w_{i,k}$ to adjust the influence of each sample pair according to label similarity, allowing for a smooth transition between positive and negative pairs.
- **Threshold loss** [2]: Introduces a threshold to define positive pairs, considering only those within a certain label distance, and adjusts the loss accordingly.
- **Exponential loss** [2]: Applies an exponential weighting to modulate the repulsion between samples based on label differences, emphasizing pairs with similar labels.

A.2 Detailed Data Description

This section provides comprehensive details about the datasets used in our study and the preprocessing steps applied to ensure data uniformity and quality.

A.2.1 Datasets

We assembled a dataset of 311 3D brain stiffness maps from healthy control subjects, sourced from multiple clinical studies [15, 16, 19, 18, 17, 20]. Table 3 presents a detailed breakdown of these datasets. All datasets were collected in accordance with ethical standards, under protocols approved by the respective local institutional review boards.

Table 3: Compilation of Dataset Information Across Multiple Studies: This table presents a detailed breakdown of the datasets used in our analysis. The aggregated data encompasses a diverse age range and a balanced gender ratio, facilitating a comprehensive evaluation of brain stiffness in healthy controls.

Study	Published In	#Subjects	Age [years]	Sex [F:M]
1	[15]	134	23.4 ± 4.0	78:56
2	[16]	60	37.8 ± 20.9	34:26
3	[19]	12	69.4 ± 2.4	6:6
4	[18]	68	69.3 ± 5.8	49:19
5	[17, 20]	37	49.1 ± 16.6	16:21
Total	**-**	**311**	**41.0±21.9**	**183:128**

A.2.2 Age Distribution

Fig. 3 illustrates the age distribution of participants from the different studies. The distribution is bimodal, indicating two predominant age groups among the subjects.

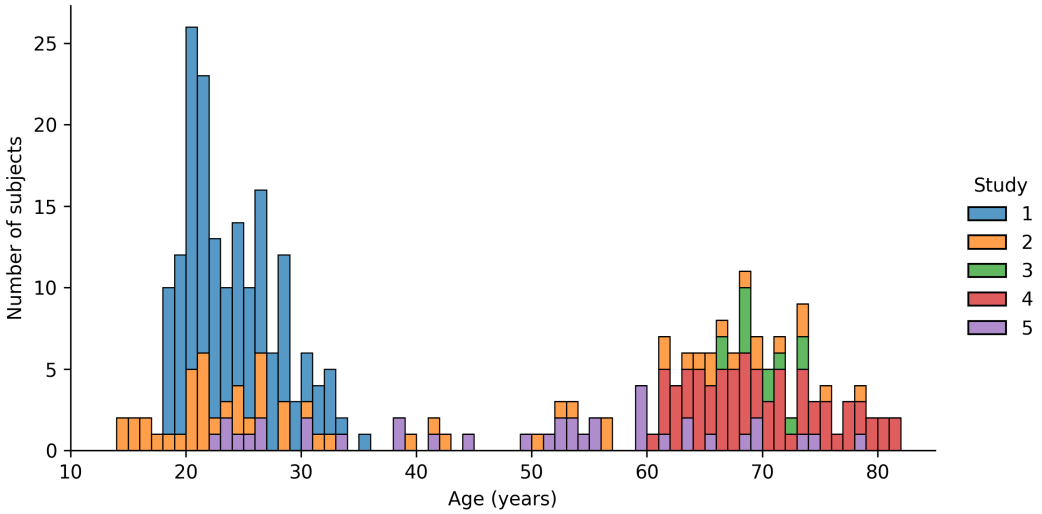


Figure 3: Age Distribution of Participants from Multi-Site MR Elastography Studies. Contribution to the 311 healthy control (HC) stiffness brain maps of different clinical studies is highlighted in color. The distribution is bimodal, indicating two predominant age groups among the subjects.

A.2.3 Data Preprocessing

To enhance data quality and uniformity, we applied several preprocessing steps to the stiffness maps:

1. **Skull Stripping**: Each map underwent skull stripping using Freesurfer [23] to isolate brain tissue from non-relevant anatomical structures.
2. **Bias Field Correction**: We applied bias field correction to remove intensity gradients that could affect analyses.
3. **Spatial Normalization**: To address data heterogeneity across different studies, we performed affine registration of the images to the MNI152 template at an isotropic resolution of 2mm^3 using ANTs [24], ensuring consistent orientation and scale among all datasets.
4. **Intensity Normalization**: Finally, we normalized the quantitative stiffness images by setting their mean to zero and standard deviation to one across the dataset.

A.2.4 Neuroimaging Modalities Comparison

Figure 4 presents a comparison of neuroimaging modalities. Each row shows three orthogonal views (sagittal, coronal, and axial) of the brain images, highlighting differences in mechanical (stiffness maps) and structural (T1-weighted MRI) properties across different ages.

A.3 Calculation of Number of Nearest Neighbors

In our dynamic localized repulsion approach, the number of nearest neighbors (NN_{nb}) decreases progressively during training. Algorithm 1 outlines the computation of NN_{nb} at each epoch.

A.4 Evaluation Protocol Details

We used an 80:20 train-test split over 50 epochs with a batch size of 32, utilizing the Adam optimizer. The initial learning rate was set to 1×10^{-4} and decreased by 10% every 10 epochs. A weight decay of 5×10^{-5} was applied to prevent overfitting. Hyperparameters $NN_{nb,final}$ and $NN_{\text{step size}}$ were

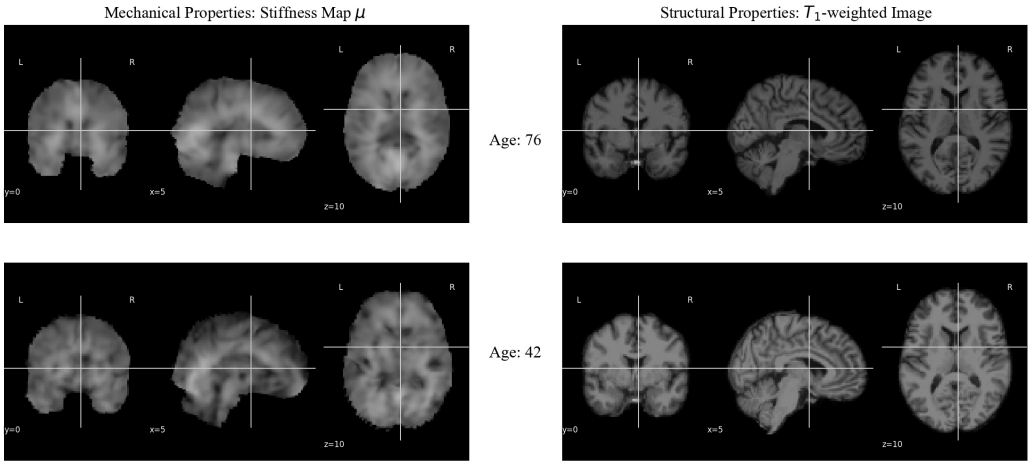


Figure 4: Comparison of Neuroimaging Modalities.

Algorithm 1 Calculation of Number of Nearest Neighbors NN_{nb} (epoch)

Require: $\texttt{NN}_{\texttt{step size}} < \texttt{max epochs} \wedge \texttt{NN}_{\texttt{nb, final}} \geq 0$

$\texttt{steps completed} \leftarrow \left\lfloor \frac{\texttt{current epoch}}{\texttt{NN}_{\texttt{step size}}} \right\rfloor$

$\texttt{total steps} \leftarrow \left\lfloor \frac{\texttt{max epochs}}{\texttt{NN}_{\texttt{step size}}} \right\rfloor$

$\texttt{NN}_{\texttt{nb}}\ \texttt{decrement per step} \leftarrow \frac{\texttt{batch size} - \texttt{NN}_{\texttt{nb, final}}}{\texttt{total steps} - 1}$

$\texttt{NN}_{\texttt{nb}} \leftarrow \texttt{batch size} - (\texttt{NN}_{nb}\ \texttt{decrement per step} \times \texttt{steps completed})$

$\texttt{NN}_{\texttt{nb}} \leftarrow \max(\texttt{NN}_{\texttt{nb}}, \texttt{NN}_{\texttt{nb, final}})$

optimized via random search across 30 iterations, resulting in $NN_{nb,final} = 14$ and $NN_{\text{step size}} = 1$, which were used throughout the benchmarks and ablation studies. Our implementation is based on Barbano et al. [2]. All models were trained using an NVIDIA A100-SXM-80GB GPU.

A.5 Ablation Studies

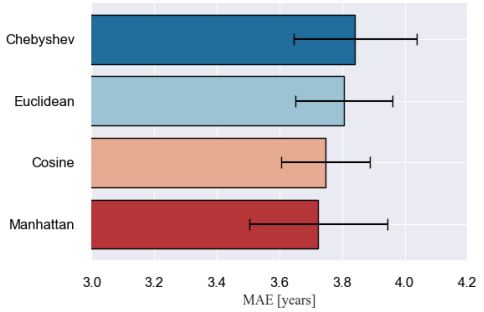
We conducted several ablation studies to examine the impact of different components on model performance.

A.6 Impact of Distance Norms

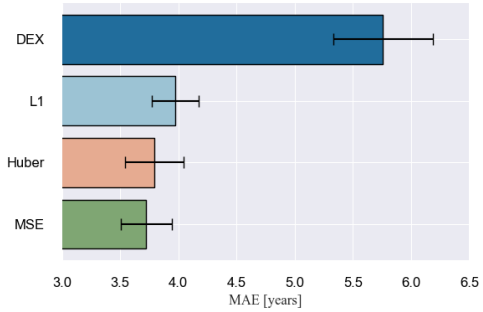
We investigated the role of the distance norm used for nearest neighbor selection in our dynamic localized repulsion approach. As shown in Figure 5a, the method demonstrates robustness regarding the choice of distance norm. The Manhattan norm achieved the lowest MAE of 3.724 ± 0.220 years, outperforming the Cosine (MAE = 3.748 ± 0.142 years), Euclidean (MAE = 3.806 ± 0.154 years), and Chebyshev norms (MAE = 3.842 ± 0.196 years).

A.7 Comparison of Regression Losses

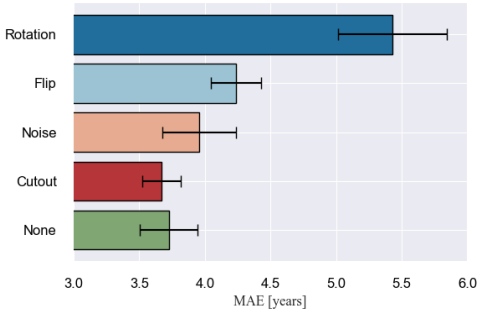
We evaluated different auxiliary loss functions to investigate their impact on performance: Mean Squared Error (MSE), Mean Absolute Error (L1), Huber loss, and DEX loss [25]. The encoder was first trained using $\mathcal{L}_{DynLocRep}$, then frozen while a predictor was trained using the auxiliary loss. As shown in Figure 5b, the MSE loss achieved the lowest MAE of 3.724 ± 0.220 years, suggesting that traditional regression losses are more effective when combined with contrastive learning representations.



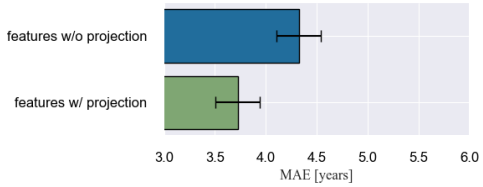
(a) Ablation study for different distance norms when selecting nearest neighbors shows Manhattan norm achieves lowest MAE.



(b) Ablation study for different regression losses shows MSE loss achieves lowest MAE compared to Huber, L1 and DEX.



(c) Ablation study for different augmentations reveals the Cutout method achieves the lowest MAE, outperforming other augmentations.



(d) Ablation study for feature extraction shows feature extraction after non-linear projection mapping achieves lower MAE.

Figure 5: Ablation studies for distance norms, regression losses, augmentations, and feature extraction with/without projection mapping.

A.8 Effect of Data Augmentations

We explored the impact of various augmentations—Noise, Cutout, Rotation, and Flip—on model performance. Figure 5c indicates that Cutout provided the best performance with an MAE of 3.667 ± 0.147 years. Rotation significantly degraded performance due to disruption of consistent brain orientation.

A.9 Feature Extraction with/without Projection

We compared model performance when features were extracted after the projection layer versus before. Figure 5d shows that extracting features after the projection layer resulted in a lower MAE of 3.724 ± 0.220 years compared to 4.323 ± 0.216 years without projection, aligning with practices in frameworks like SimCLR.