Ph.D. Dissertation Defense

Novel Deep Learning methods for Early Detection of Neurological Disorders

Jay Shah

Advisors: Prof. Baoxin Li, Prof. Teresa Wu

June 18th, 2025

Acknowledgments

Teresa Wu, Ph.D.

President's Professor
Arizona State University

Baoxin Li, Ph.D.

Professor
Arizona State University

Yi Su, Ph.D.

Director, Comp. Imaging Analysis
Banner Alzheimer's Institute

Yezhou Yang, Ph.D.

Associate Professor
Arizona State University

Yalin Wang, Ph.D.

Professor
Arizona State University

Acknowledgments

Financial support

- National Institutes of Health (NIH)
  Award Number: K23NS070891

- National Institute on Aging of NIH
  Award Numbers: R01AG069453, P30AG072980,
  RF1AG073424, R01AG057708

- National Institute of Neurological Disorders and Stroke (NIH)
  Award Number: 1R61NS11331501

- United States Department of Defense
  Award Numbers: W81XWH1510286, W81XWH1910534

National Institute of
Neurological Disorders
and Stroke

Outline

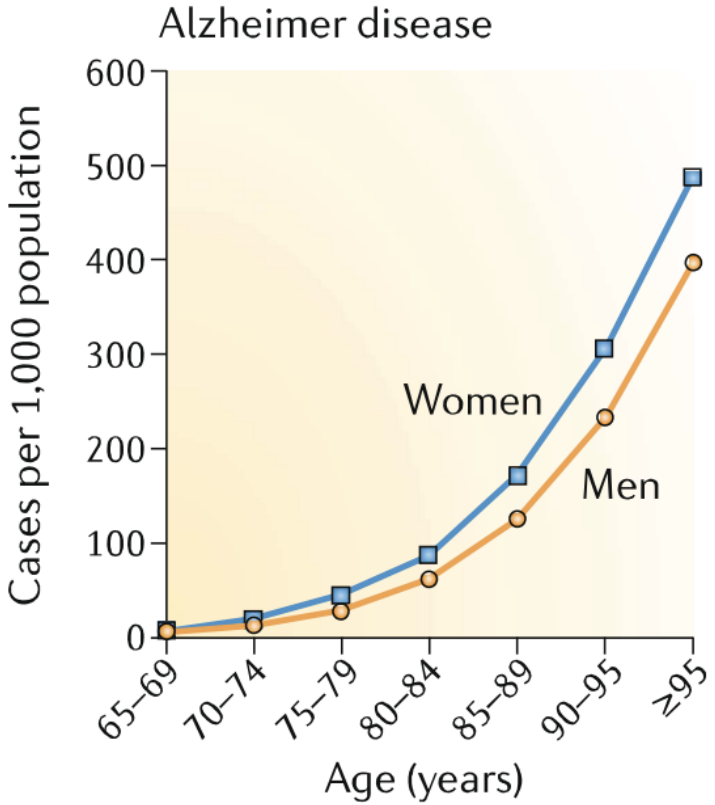
1. Motivation
2. Phase I: Brain Age prediction
3. Phase II: Medical image compression
4. Phase III: PET Imaging Super-Resolution

Outline

1. Motivation
2. Phase I: Brain Age prediction
3. Phase II: Medical image compression
4. Phase III: PET Imaging Super-Resolution

Motivation

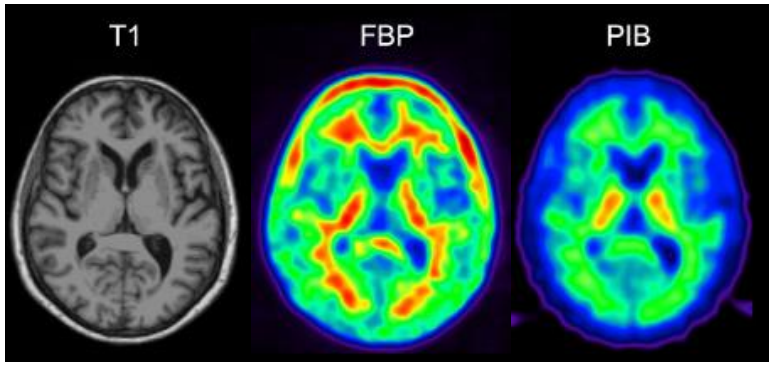
- Age is the biggest known risk factor for most neurodegenerative disorders
  - Alzheimer's disease, Parkinson's disease, and others
- Causes irreversible structural damage
- Early detection
  - Effective interventions
  - Preventing brain damage



Hou, Yujun, et al. "Ageing as a risk factor for neurodegenerative disease." *Nature Reviews Neurology* (2019)

Motivation

Alzheimer's disease (AD) is a multi-factorial disorder



1. Brain structure damage can be quantified using T1w-MRI

2. AD pathogenesis starts with amyloid-beta (Aβ) deposition
   Brain damage ↔ Aβ plaques
   PET tracers (FBP, PiB, others) help quantify early AD onset

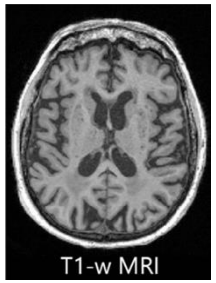
pet = positron emission tomography; FBP = florbetapir; PiB = pittsburgh compound B

Motivation

Deep learning based Brain Age prediction

Models lack precision, show age-related systematic bias

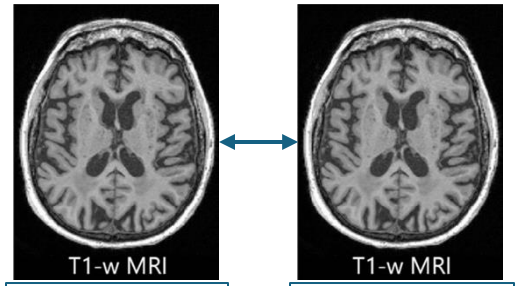
→ **Phase I**



Medical Image compression

Codebook collapse, reconstruction fidelity, practical utility

→ **Phase II**

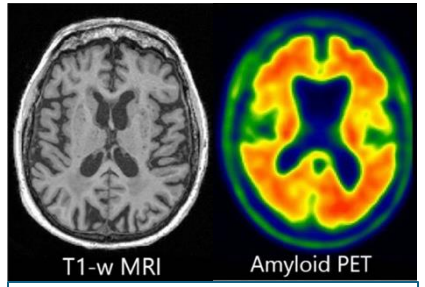


Original | Reconstructed

PET Imaging Super-Resolution

Quantifying Aβ deposition

→ **Phase III**



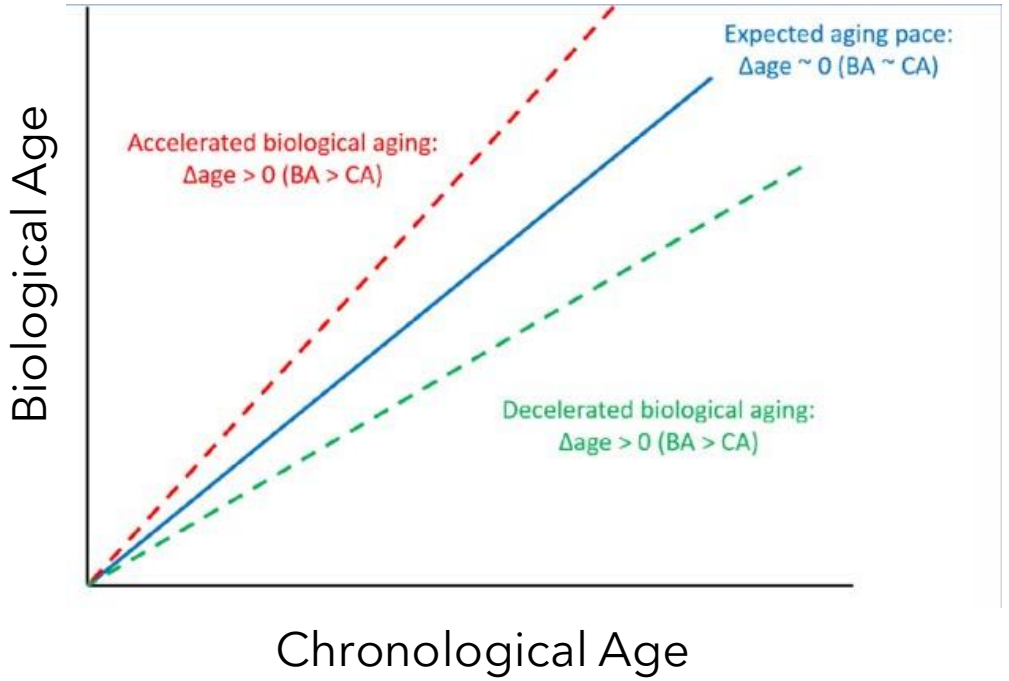
Multimodal

(more details in later slides)

Phase I: Brain Age prediction

Aging in humans is complex

- Biological aging ≠ chronological aging
  brain can age faster or slower
- Variations in individuals
  due to genetic, environmental, neurological predispositions



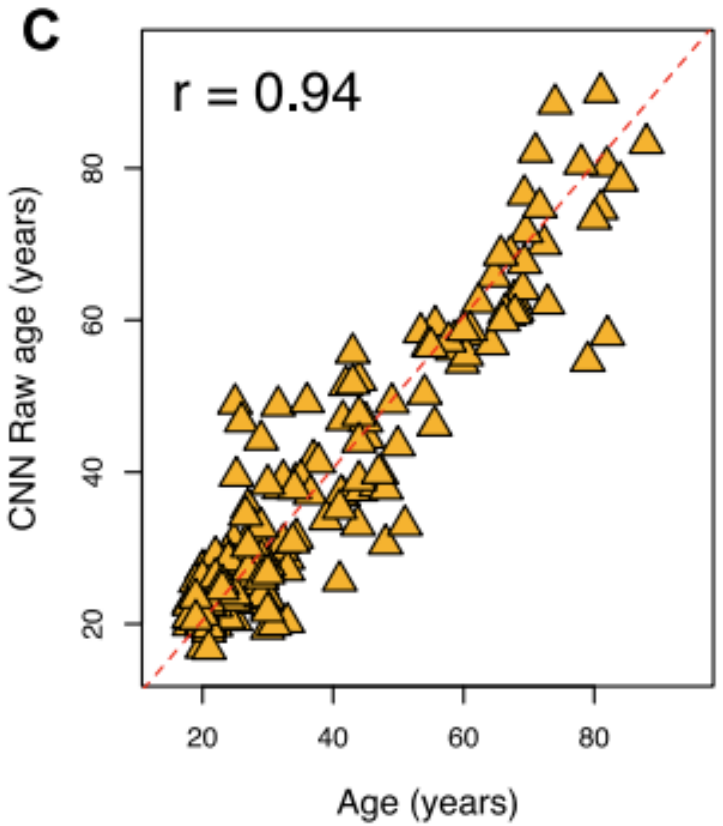
Contents lists available at ScienceDirect

NeuroImage

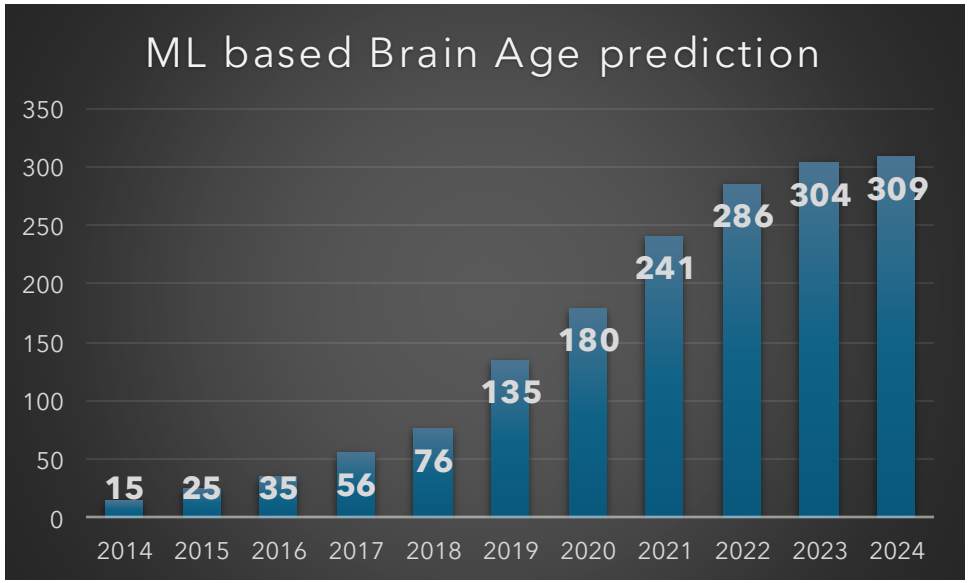
journal homepage: www.elsevier.com/locate/neuroimage

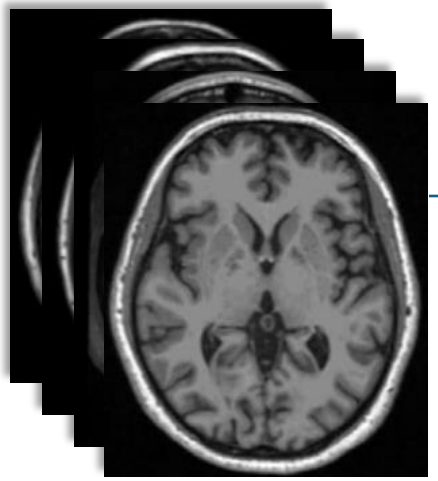
Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker

James H. Cole [a], Rudra P.K. Poudel [b], Dimosthenis Tsagkrasoulis [c], Matthan W.A. Caan [d], Claire Steves [e], Tim D. Spector [e], Giovanni Montana [b,c,*]



Number of articles

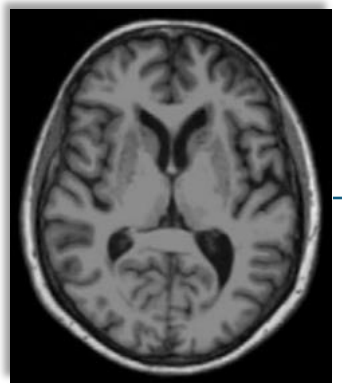




3D MRI scans
(Healthy subjects)

Deep Learning Model

Chronological Age



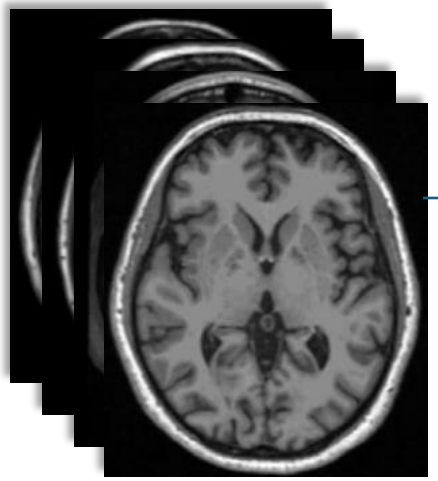
3D MRI scan
(Healthy/AD subject)

Trained Model

Predicted Age

Δage = (predicted – chronological) age

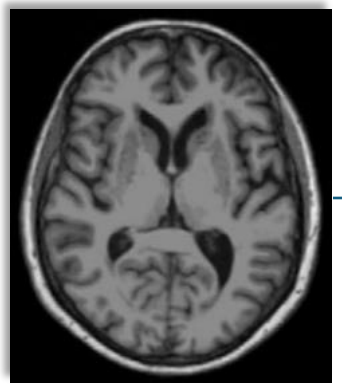
Regression



3D MRI scans
(Healthy subjects)

Deep Learning Model

Chronological Age

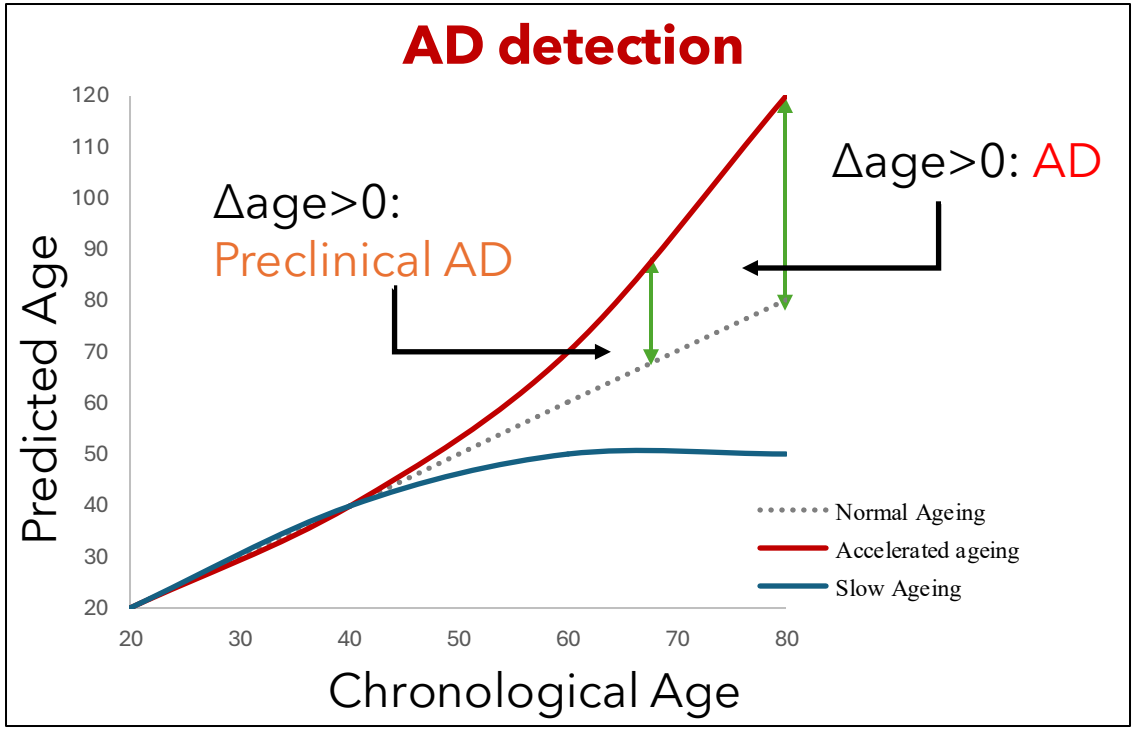


3D MRI scan
(Healthy/AD subject)

Trained Model

Predicted Age

Δage = (predicted – chronological) age



Dataset

- Lifespan cohort (7,377) 3D MRIs Healthy
  - IXI, ABIDE, ICBM, NACC and OASIS (public)
  - age [8-95]

- Discovery cohort (1,584) 3D MRIs Healthy/MCI/AD
  - ADNI database
  - age [55-98]

- Train: Val: Test = 80: 10: 10
  (stratified on age groups 8-12, 12-16, …)

Dataset	Count	Age Range (yrs)	Mean ± STD
NACC	4,132	18 - 95	67.5 ± 10.8
OASIS	1,432	8 - 94	27.9 ± 20.7
ICBM	1,101	18 - 80	37.6 ± 15.4
IXI	536	20 - 86	48.8 ± 16.5
ABIDE	176	18 - 56	26.1 ± 7.0
ADNI	1,584	55 - 98	73.3 ± 7.3

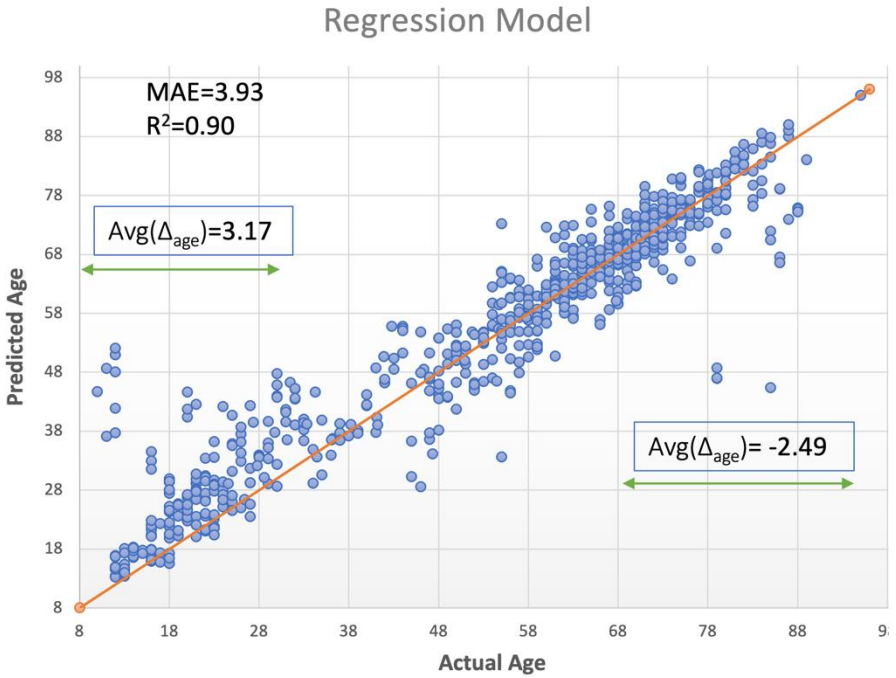
Existing Gaps

Models are not accurate!

- Age-related systematic bias
  Young subjects are over-estimated; under-estimation in Old
  Inherent to regression[a]

- Well observed[b], not due to
  Model selection, imbalance, heterogeneity[b]
  Current approaches → post-hoc correction

- Most Alzheimer's patients are age > 50

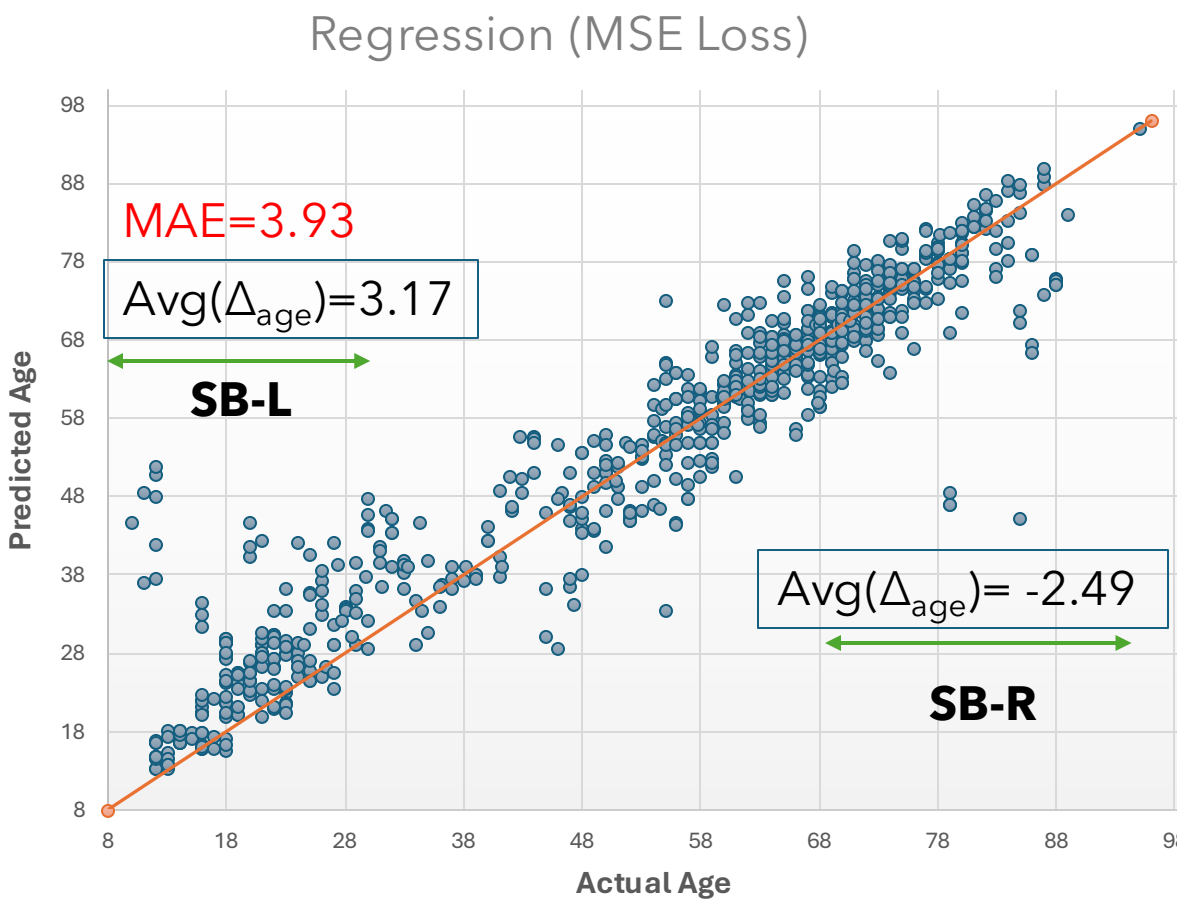


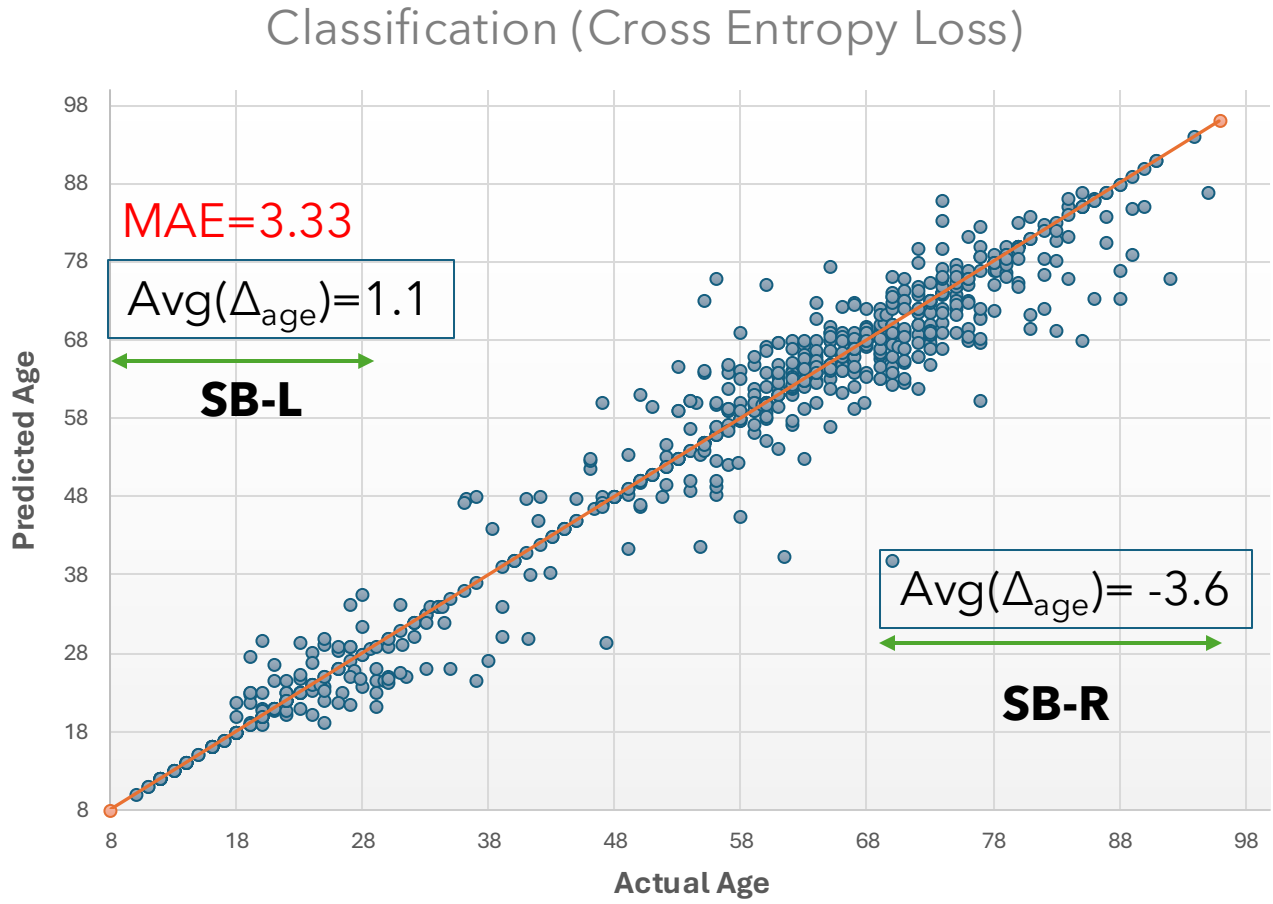
ResNet-18 (regression) model trained on 6617 HC, tested on 760 HC

Hypothesis: Age prediction as regression causes regression-to-mean (RTM)
→ Leading to systematic bias

[a]Gardner, M. J., and J. A. Heady. "Some effects of within-person variability in epidemiological studies." Journal of Chronic Diseases (1973)
[b]Liang, Hualou et al. "Investigating systematic bias in brain age estimation with application to post-traumatic stress disorders". Human Brain Mapping (2019)

Regression as Classification

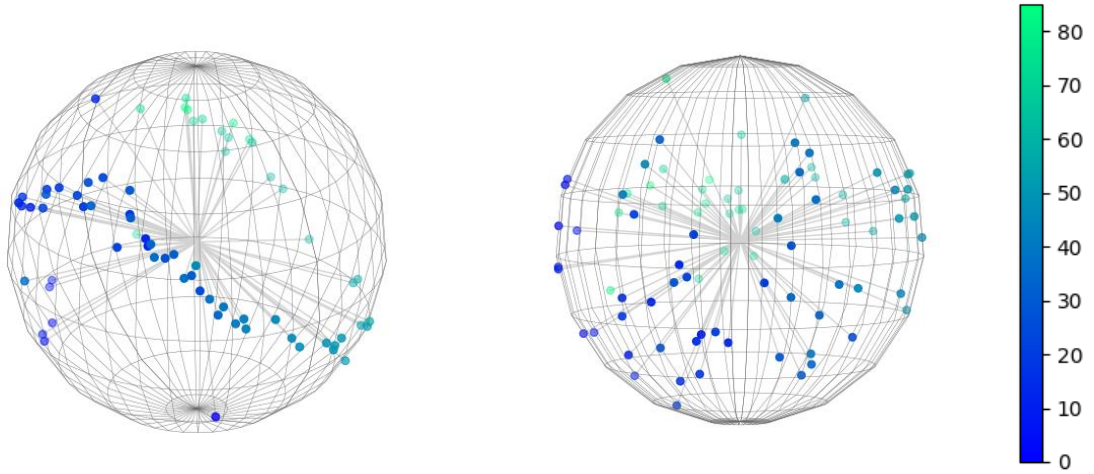




<u>Measuring Systematic Bias</u>:

One standard deviation from mean: systematic bias-left $(\mu - \sigma)$ and right , $(\mu + \sigma)$ [SB-L, SB-R]

Ordinality



MSE
Ordinality score: 0.99

Cross Entropy
Ordinality score: 0.31

- C = {1, 2, … (c-1)} *where c is #classes*
- X = {x_1, x_2, …, x_n} – penultimate layer features
- F_c = {f_1, f_2, …, f_c} – feature centroids

Manhattan distances between f_1 and other feature centroids

- D = {d_{12}, d_{13}, …, d_{1c}}

$$\text{Ordinality score} = \text{Pearson}\,(D, C)$$

<u>Classification beats Regression</u>
Due to ability to learn high entropy discriminative feature representation[a]

But lacks ordinality!
Cross entropy treats each class independent from each other

Ex: *Patient of Age 52 misclassified as 51 vs.14 hampers clinical decision making*

[a]Zhang, Shihao, et al. "Improving Deep Regression with Ordinal Entropy." *ICLR (2023)*

Phase I

How to preserve Ordinality in Classification?

- While reducing RTM bias
- And improving age prediction

ORDER loss

Aim: Ordinal information from target space (age) into learned feature space (***z***)

To guarantee:

$$z_c > z_{c+1} > z_{c+2} > \dots > z_C$$
$$z_c > z_{c-1} > z_{c12} > \dots > z_1$$

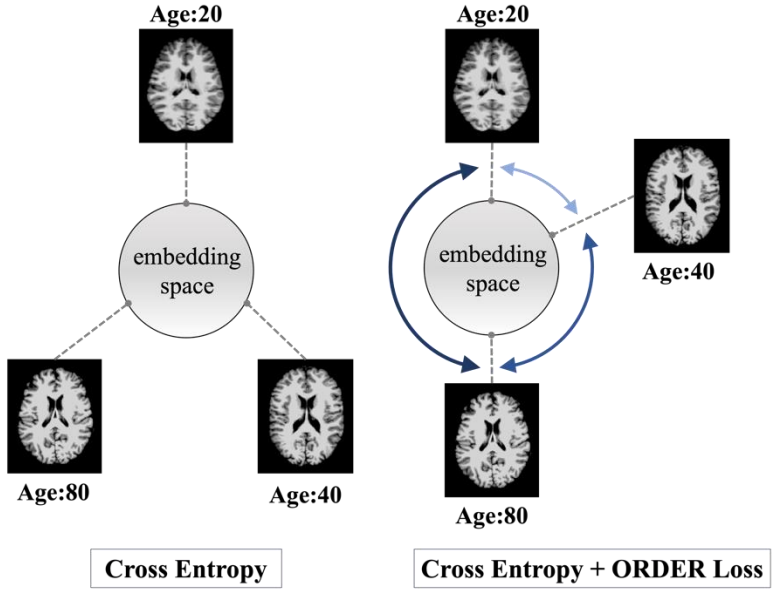
Cross Entropy:

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{z_i}}{\sum_{j=1}^{C}e^{z_j}}$$

Regularization:

$$z_i' = W_{y_i}^T x_i + \varphi(x_i)$$

$$\varphi(x_i) = \frac{1}{N-1}\sum_{j=1,i\neq j}^{N}|i-j||\bar{x}_i - \bar{x}_j|_{manh}$$



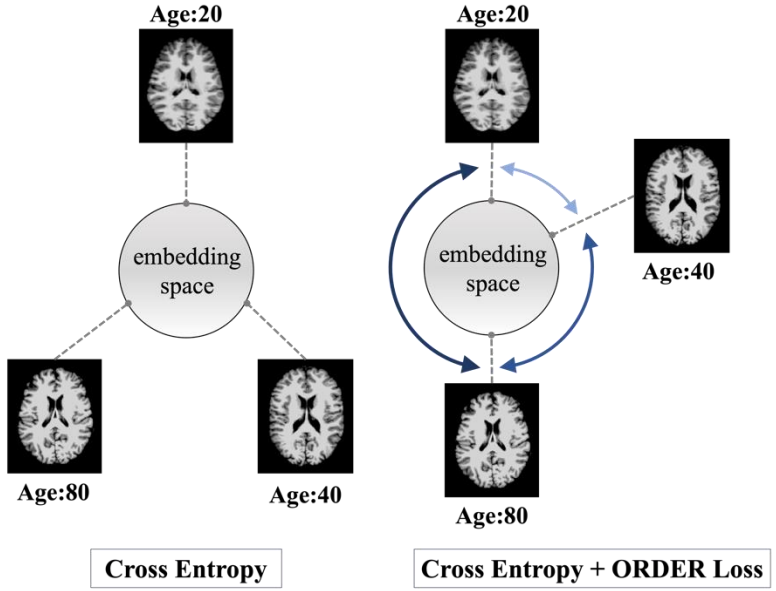
ORDER – ORdinal Distance Encoded Regularization

Manhattan distance (L1 norm) is consistently preferable than the Euclidean distance (L2 norm) for high-dimensional data

Aggarwal, Charu C., et al. "On the surprising behavior of distance metrics in high dimensional space." Database theory—ICDT (2001)

ORDER loss

$$
\begin{aligned}
L_T &= -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T x_i+\varphi(x_i)}}{\sum_{j=1}^{C} e^{W_{y_j}^T x_i}} \\
&= -\frac{1}{N}[\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T x_i}}{\sum_{j=1}^{C} e^{W_{y_j}^T x_i}} + \sum_{i=1}^{N}\varphi(x_i)] \\
&= -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T x_i}}{\sum_{j=1}^{C} e^{W_{y_j}^T x_i}} \\
&- \frac{1}{N(N-1)}\sum_{j=1,i\neq j}^{N}|i-j||\bar{x}_i-\bar{x}_j|_{manh} \\
&= L_{CE}+L_{ORDER}
\end{aligned}
$$



Shah, Jay, et al. "Ordinal classification with distance regularization for robust brain age prediction." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024.

Methods

- Baseline loss functions

- 3D ResNet-18
- Stratified oversampling [8-12, 12-16, …]
- 100 epochs, *AdamW* opt, batch size=4
- *LR*=1e-3, *weight decay*=1e-2

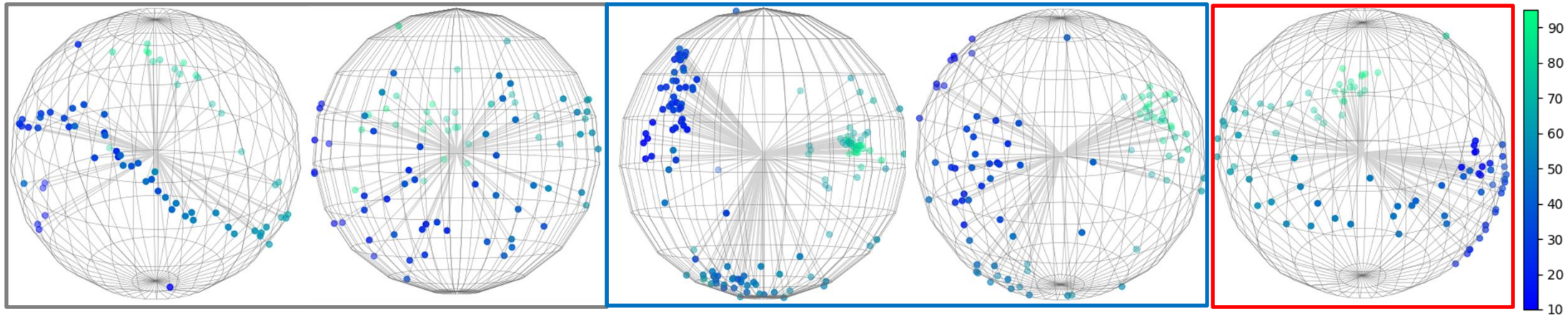
	Method (Loss)
Regression	MSE
	MSE + Euclidean norm[a]
Classification	CE
	CE + mean-variance[b]
Ours	CE + ORDER

CE=cross entropy
MSE=mean squared error

[a]Zhang, Shihao, et al. "Improving Deep Regression with Ordinal Entropy." ICLR (2023).
[b]Pan, Hongyu, et al. "Mean-variance loss for deep age estimation from a face." CVPR (2018).

Results

On Lifespan (healthy) cohort



MSE | Cross Entropy | MSE + Euclidean Norm | Mean-Variance Loss | Cross Entropy + ORDER

Embedding space analysis (512-dim)

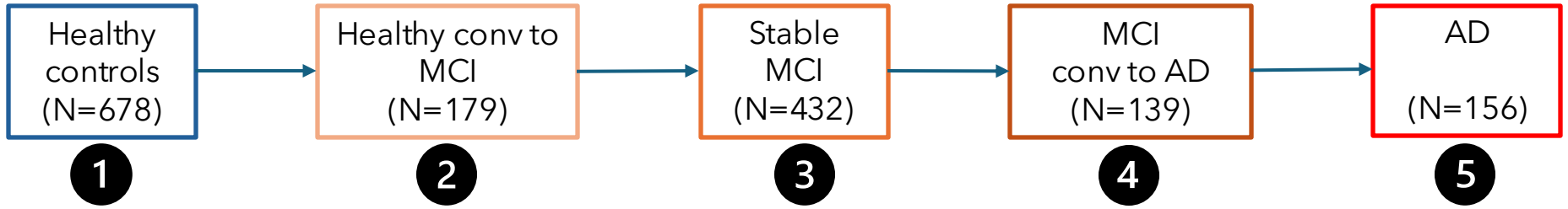
Results

On Lifespan (healthy) cohort

	Method (Loss)	**MAE**	**Ordinality**	**Systematic Bias**	
				SB-L	**SB-R**
Regression	MSE	3.93	**0.99**	3.4	-4.2
	MSE + Euclidean norm	4.57	0.95	4.8	-4.1
Classification	CE	3.33	0.31	1.1	-3.6
	CE + mean-variance	2.65	0.58	0.4	-4.2
Ours	CE + ORDER	**2.56**	0.98	**0.1**	**-2.5**

Results

5 clinical groups with increasing order of severity



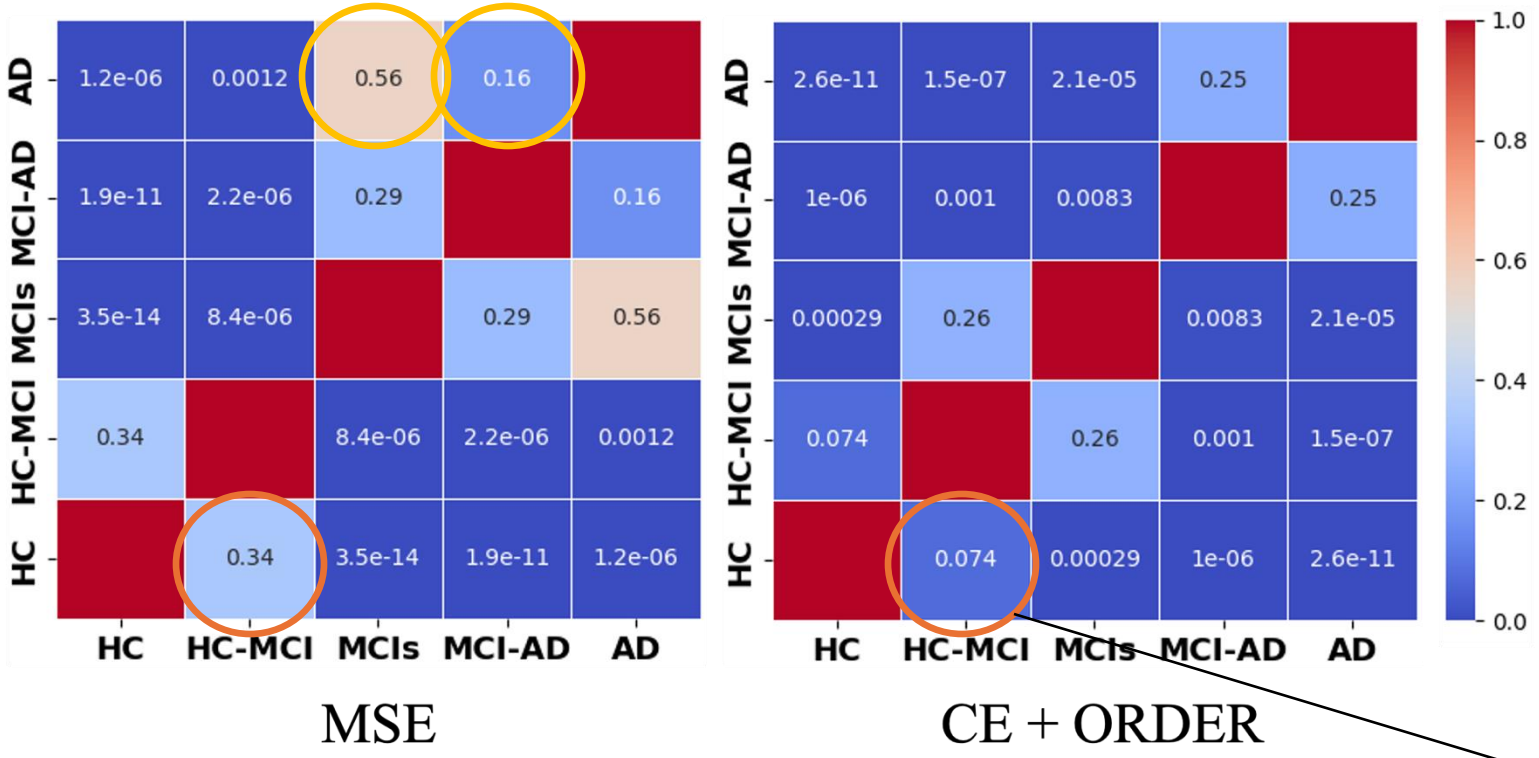
On Discovery (mixed) cohort

	Method (Loss)	**Healthy** (1)	**HC conv MCI** (2)	**MCI-stable** (3)	**MCI conv AD** (4)	**AD** (5)	**Pearson Correlation**
Reg	MSE	-1.2	-0.8	-0.3	0.8	1.5	**0.98**
	MSE + Euclidean norm	-2.7	-1.9	-1.7	-0.9	0.9	0.94
CLS	CE	-1.9	-1.5	-3.4	-2.3	-4.1	-0.75
	CE + mean-variance	-1.6	-0.3	-0.5	0.8	2.8	0.94
Ours	CE + ORDER	-1.5	-0.7	-0.3	1.2	2.0	**0.98**

Correlation with disease severity

MSE vs. ORDER

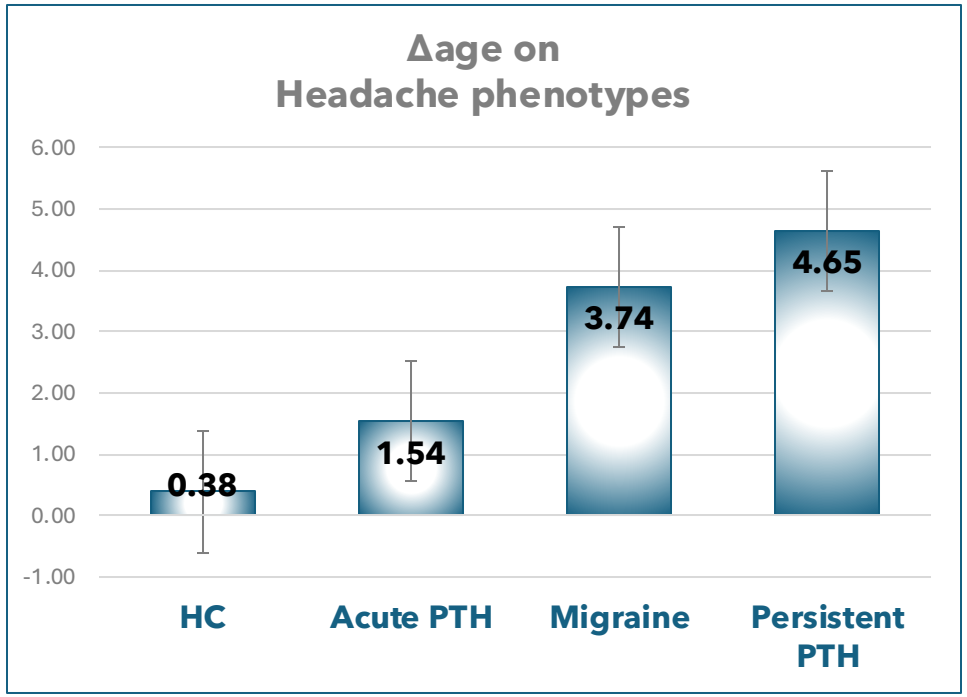
Statistical significances between clinical groups as *p-values* on predicted BrainAGE



More accurate for early detection

- MSE – disruptive trend
- CE + ORDER – consistent trend

Headache detection



HC = Healthy Controls
PTH = Post Traumatic Headache

*in-house data collected from Mayo Clinic, Arizona

Findings:

- Δage(P-PTH) < Δage(A-PTH)
  suggesting more structural decline related to PTH persistence over time
- Headache frequency associated with structural damage
  Δage(P-PTH) > Δage(Mig) > Δage(A-PTH)
- Early detection potential
  structural decline acutely following TBI at risk for developing persistent PTH

Shah, Jay, et al. "Capturing MRI Signatures of Brain Age as a Potential Biomarker to Predict Persistence of Post-traumatic Headache (S20. 006)." *Neurology*. Vol. 102. No. 17_supplement_1. Hagerstown, MD: Lippincott Williams & Wilkins, 2024.

Findings

1. Cross-entropy learn high-entropy (discriminative) feature representation
   To reduce RTM bias from regression

2. ORDER loss can preserve ordinality in feature space
   To improve overall prediction accuracy

3. Model achieved MAE=2.56 on Healthy
   Compared to 3.93 (MSE), 35% improvement⬆
   Biomarker reliability

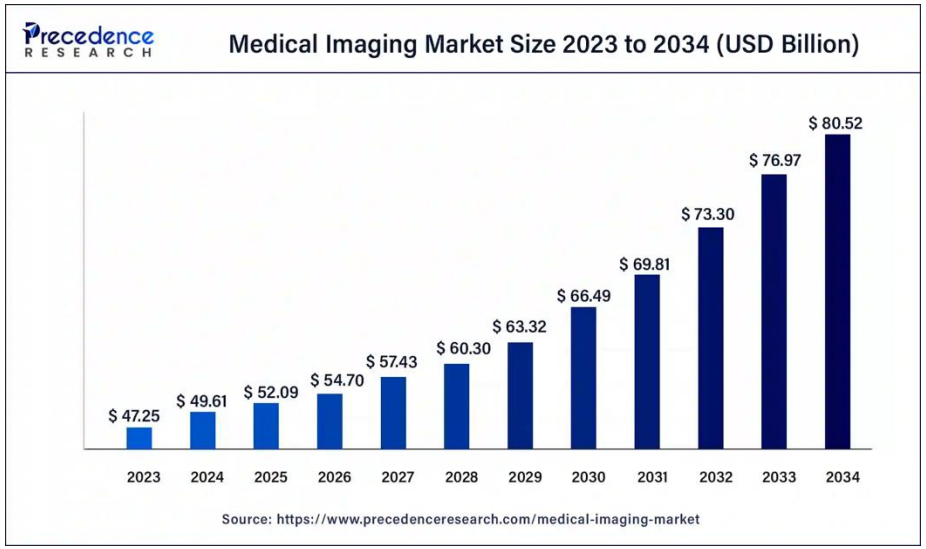
4. Can detect subtle difference in clinical groups
   Crucial for early detection (Alzheimer's & Headache)

Locality Constrained Vector Quantization

Motivation

- Rapid growth of medical imaging in modern medicine

  Petabytes (PB) of MRI data generated annually
  Radiology data at Stanford grew by ~450 TB per year[a]



Requires: Network bandwidth & Storage

- <u>Efficient compression matters</u>
  1. Storage burden
     FreeSurfer processed image: 16-60 MB
     Entire folder: 300-370 MB

  2. Impractical for telemedicine[b]
     Limited bandwidth (rural or mobile)

[a]Mesterhazy, Joseph et al. "High performance on-demand de-identification of a petabyte-scale medical imaging data lake." arXiv preprint (2020).

[b]Elhadad, Ahmed, et al. "Reduction of NIFTI files storage and compression to facilitate telemedicine services based on quantization hiding of downsampling approach." Scientific Reports (2024)

Image Compression

Lossless

- Huffman coding
  - ~3.7:1 on DICOM
  - Cannot exploit spatial correlations[a]
- JPEG-LS
  - 2-3x on MRI[b]
- gzip
  - ~30–40% on NIfTI
  - nontrivial CPU overhead
  - Not ideal for real-time telemedicine

Lossy

- DCT based JPG
  - Scalar quantization introduces artifacts[c]
  - Loss of anatomical info (edges)
- JPEG2000
  - Limited real utility
  - Info loss at higher rates[d]
- 3D wavelet + DWT-VQ
  - Volumetrics wavelets improve distortion
  - Lacks end-to-end optimization
  - Heavy compute cost[e]

[a]Rahmat, Romi Fadillah et al. "Analysis of DICOM Image Compression Alternative Using Huffman Coding." Journal of healthcare engineering 17 Jun. 2019
[b]https://dicom.nema.org/medical/dicom/current/output/chtml/part05/sect_8.2.3.html

[c]Luo, Ying et al. "Removing the blocking artifacts of block-based DCT compressed images." IEEE transactions on Image Processing (2003)
[d]Dennison, Don et al.. "Informatics challenges–lossy compression in medical imaging." Journal of Digital Imaging (2014)
[e]Bruylants, Tim et al. "Wavelet based volumetric medical image compression." Signal processing: Image communication (2015)

Neural Image compression

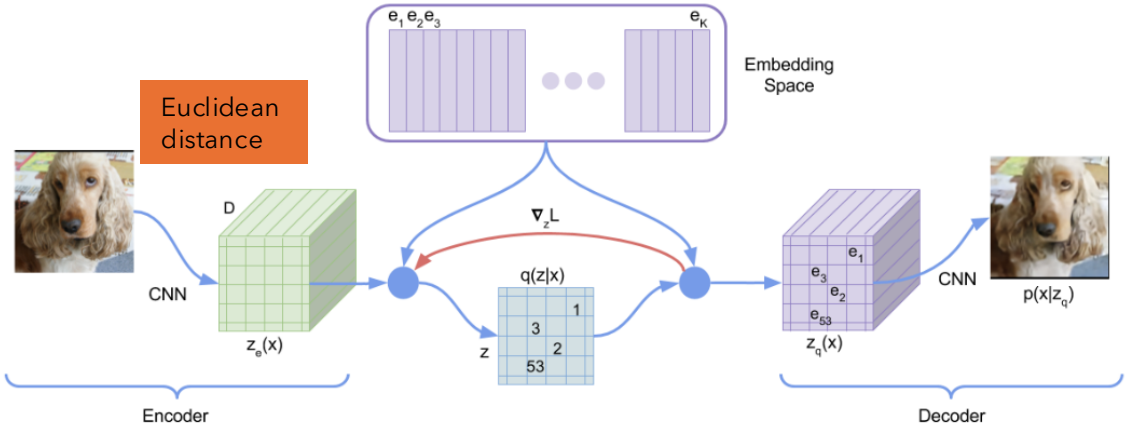
- Auto-Encoders (AE)
  - Latent maps via MSE
  - Blurry reconstructions and no entropy coding control
- Variational AE (VAE)
  - KL regularization for smoothness
  - Suffers from blurriness
- Vector-Quantizaed VAE
  - Discrete codebooks reduce blur

- VQVAE consists:
  1. Analysis transform: $y = g_a(x)$
  2. Quantization: $\hat{y} = Q(y)$
  3. Entropy coding
  4. Synthesis transform: $\hat{x} = g_s(\hat{y})$
- Traditional methods focus on:
  - entropy coding (3),
  - ignoring quantization step (2)

  → Euclidean Nearest Neighbor

Revisiting Quantization

- Encoder → continuous latent vector e_z,
  quantized to nearest codebook entry e^k via Euclidean
- Commitment loss term $\| sg[z_e] - e^{k*} \|$
  encourages encoder outputs close to their assigned embeddings



Van Den Oord, Aaron, and Oriol Vinyals. "Neural discrete representation learning." *Advances in NeurIPS* (2017).

- Codebook update: minimizing the average Euclidean distance to the batch of assigned encoder outputs,
  effectively K-means–style centroid updates

Observation: Reliance on plain Euclidean distance treats all latent dimensions equally and ignores their covariance[a]

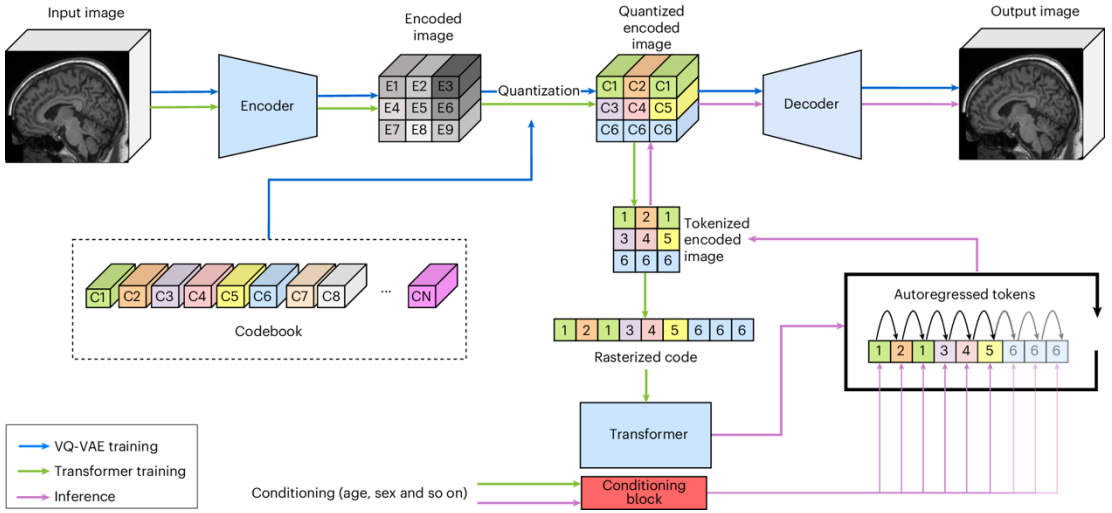
[a]Mimmack, Gillian M., Simon J. Mason, and Jacqueline S. Galpin. "Choice of distance matrices in cluster analysis: Defining regions." Journal of climate (2001)

Brain Imaging Gen (existing work)

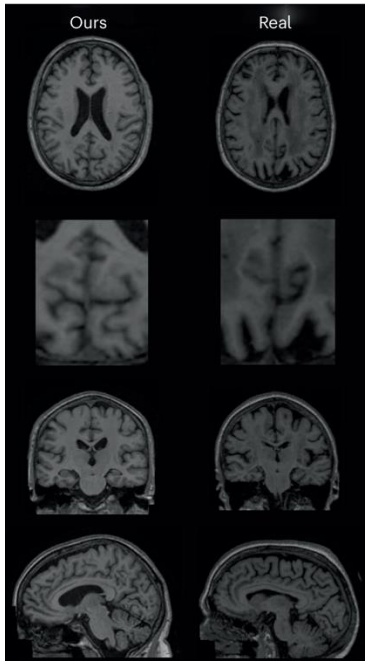
- Generates morphology preserving synthetic MRIs
  - Stage-1: VQVAE to compress
  - Stage-2: Autoregressive transformer for conditional generation (age, sex, etc.)

Novelties:

1. Freq domain sharpness (Anatomy):
   $MSE\left(X,\hat{X}\right) + MSE\left(FFT(X), FFT(\hat{X})\right)$
2. Perceptual loss (Stability):
   2D AlexNet-based LPIPS
3. PatchGAN adversarial term (Realism)
   Discriminator (LSGAN)



Data	FID	MS-SSIM
UKBB	0.0026	0.67 ± 0.05
ADNI	0.0075	0.69 ± 0.07



We use this VQVAE as baseline

Tudosiu, Petru-Daniel, et al. "Realistic morphology-preserving generative modelling of the brain." *Nature Machine Intelligence* (2024)

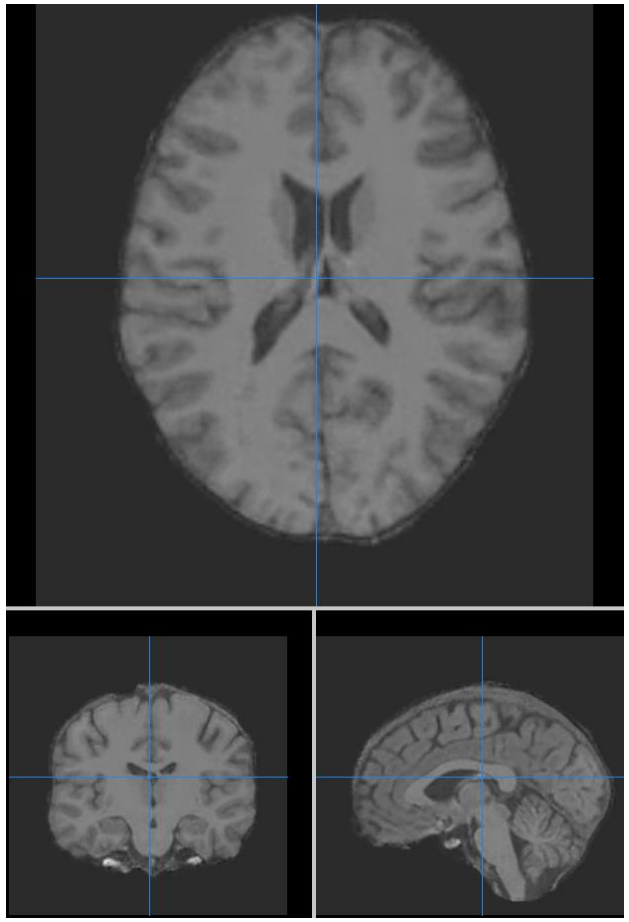
Dataset

- Lifespan cohort (7,932) 3D MRIs Healthy
  - IXI, ABIDE, ICBM, NACC and OASIS (public)
  - age [18-93]

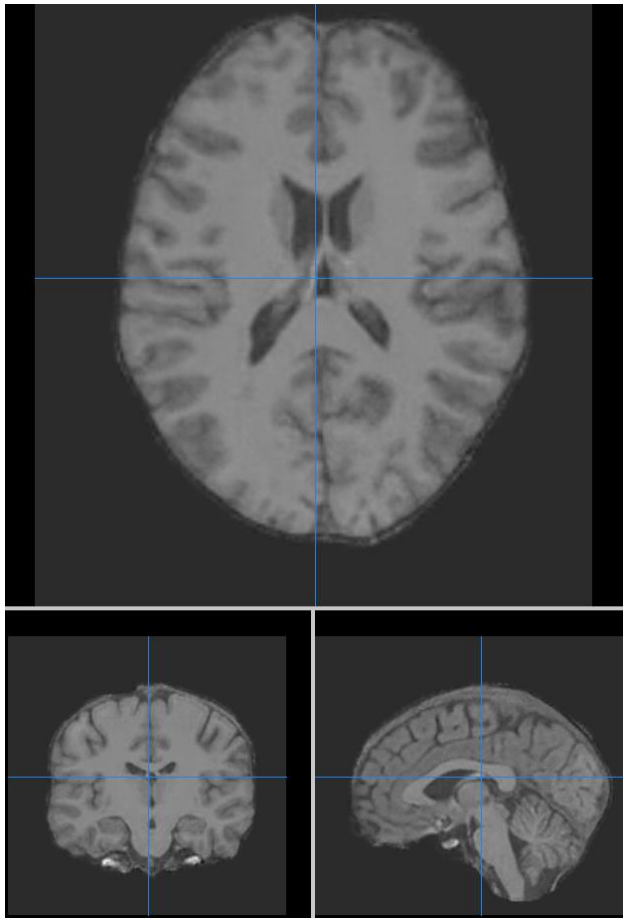
- Discovery cohort (9,913) 3D MRIs Healthy/MCI/AD
  - ADNI database
  - age [49-98]

Dataset	Count	Age Range (yrs)	Mean ± STD (yrs)
NACC	4,649	18 - 93	67.8 ± 11.1
OASIS	1,839	18 - 93	55.4 ± 25.1
ICBM	814	19 - 80	41.6 ± 15.2
IXI	529	20 - 86	48.5 ± 16.5
ABIDE	101	18 - 56	25.9 ± 7.6
ADNI	9,913	49 - 98	75.2 ± 7.5

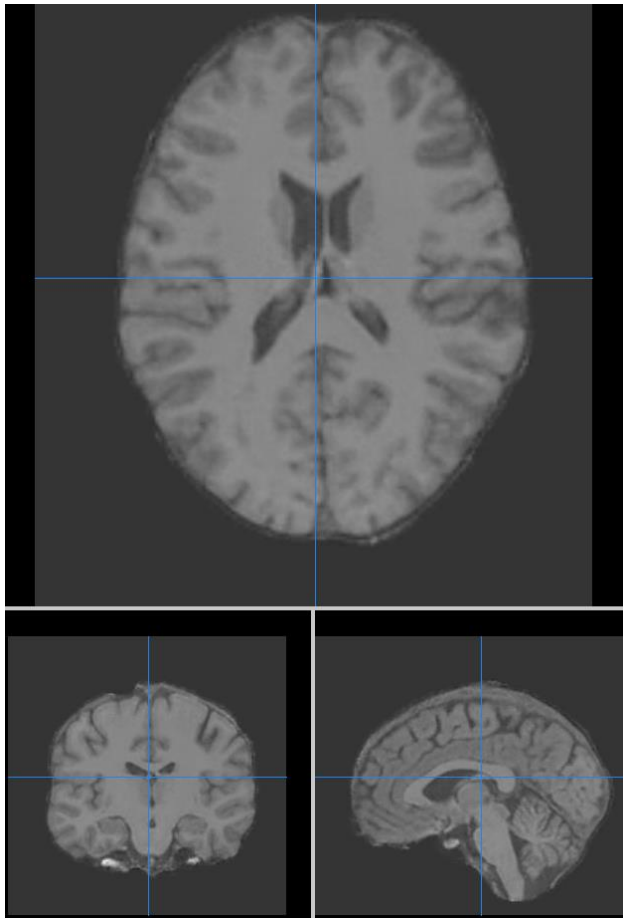
Reconstructed MRIs



K=1



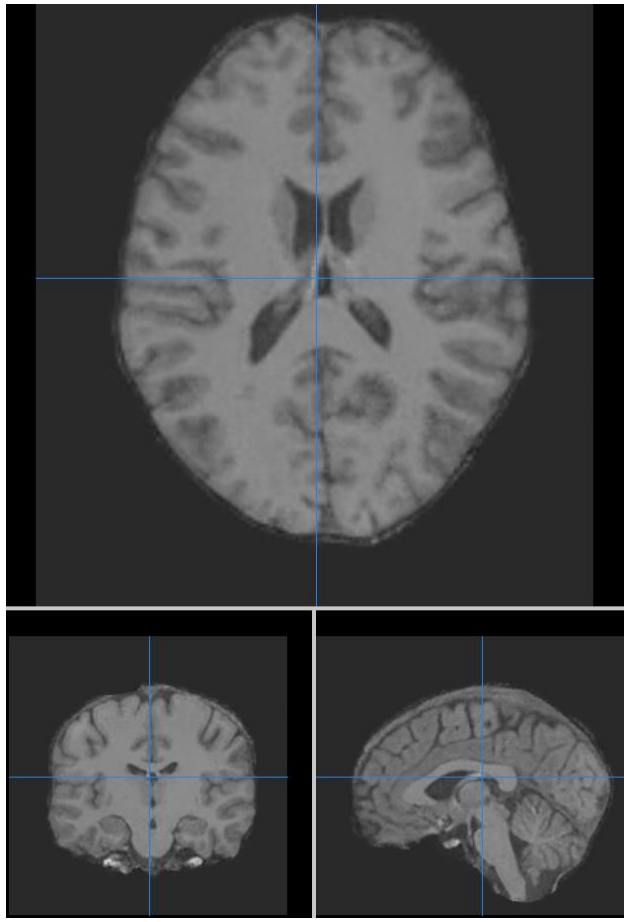
K=2



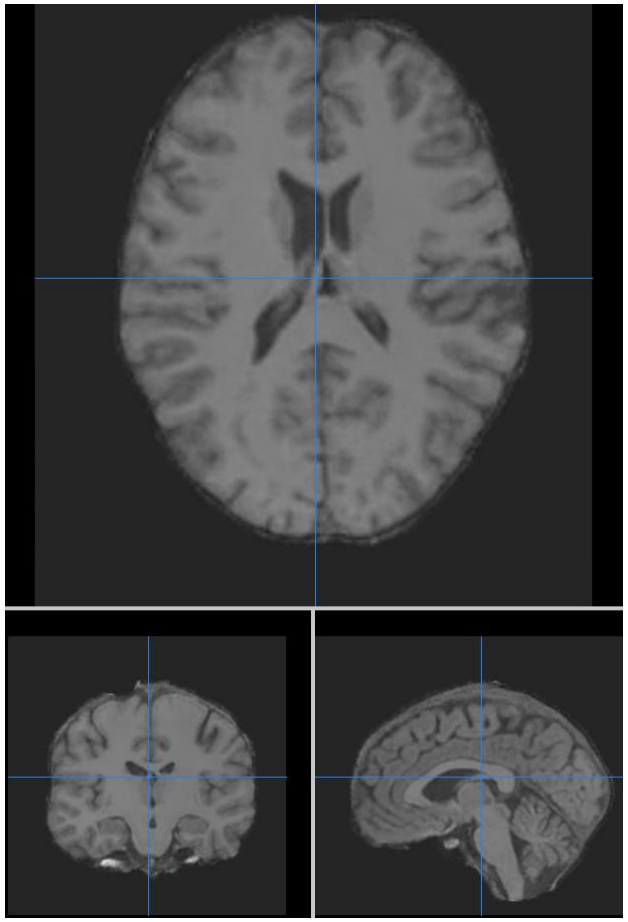
K=3

K → kth nearest codebook element using Euclidean distance

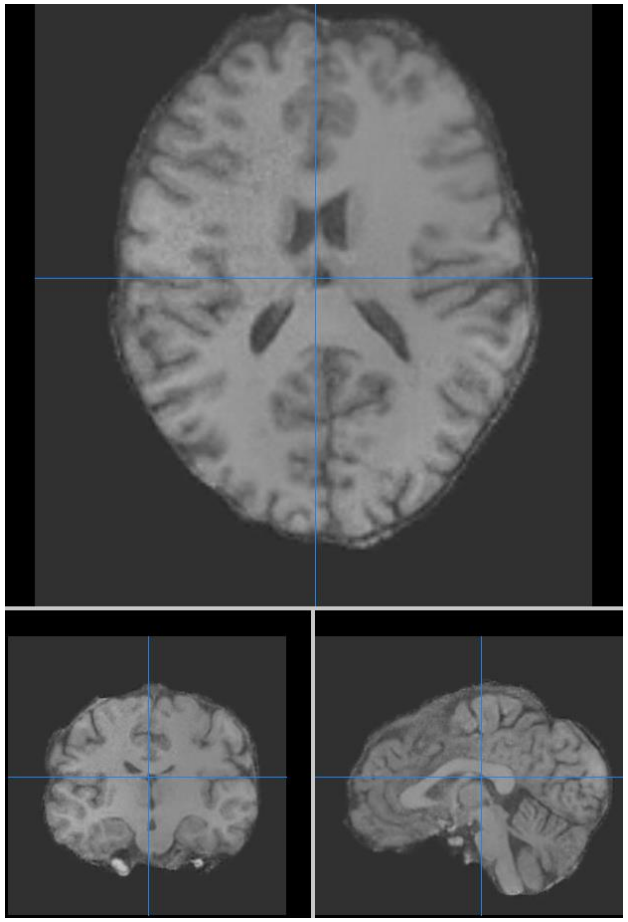
Reconstructed MRIs



K=4



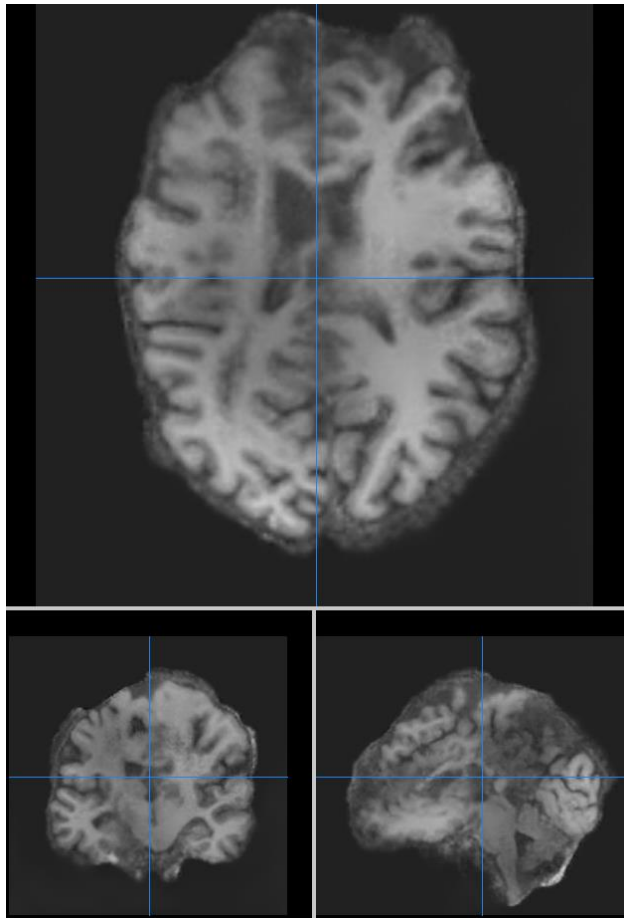
K=5



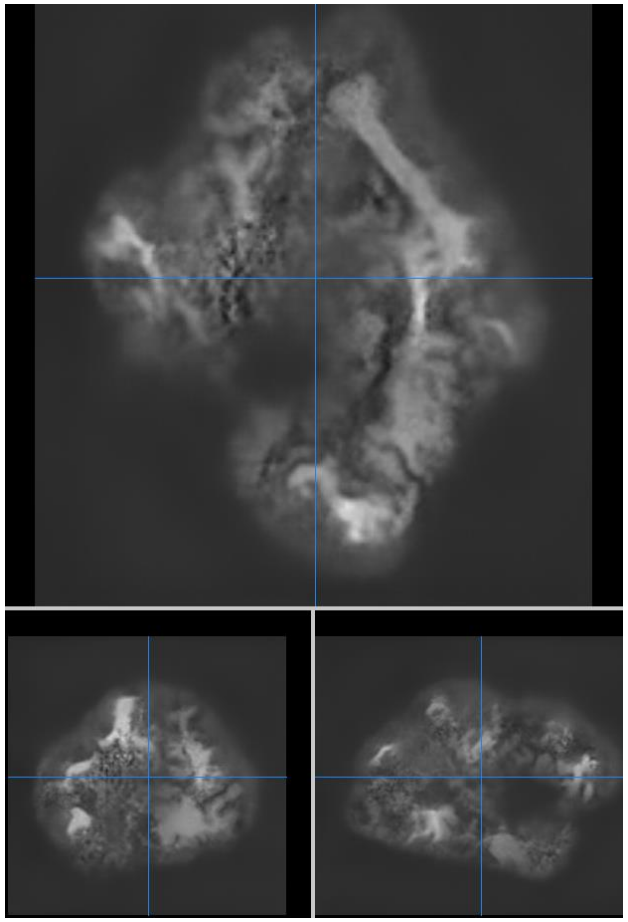
K=15

K → kth nearest codebook element using Euclidean distance

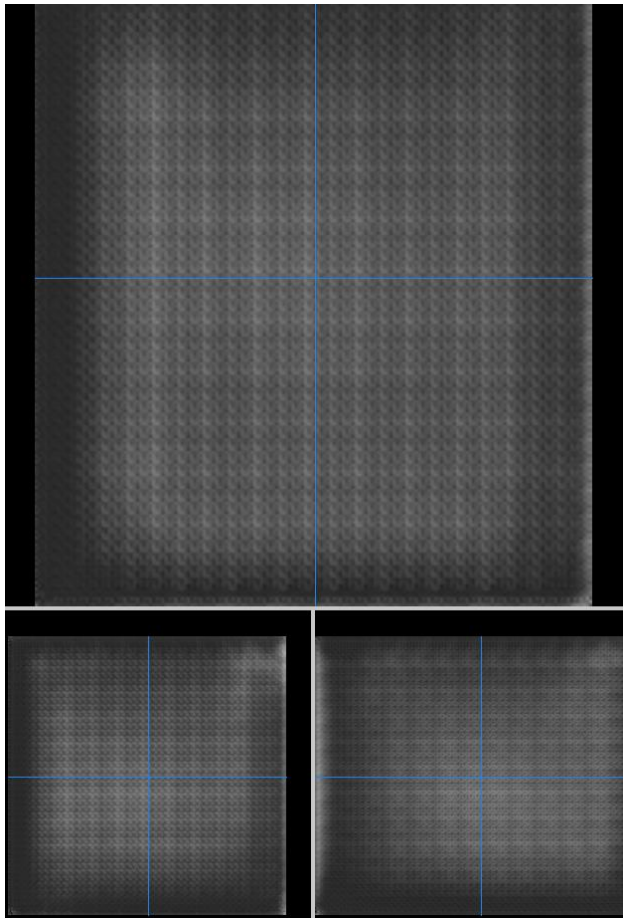
Reconstructed MRIs



K=25



K=35



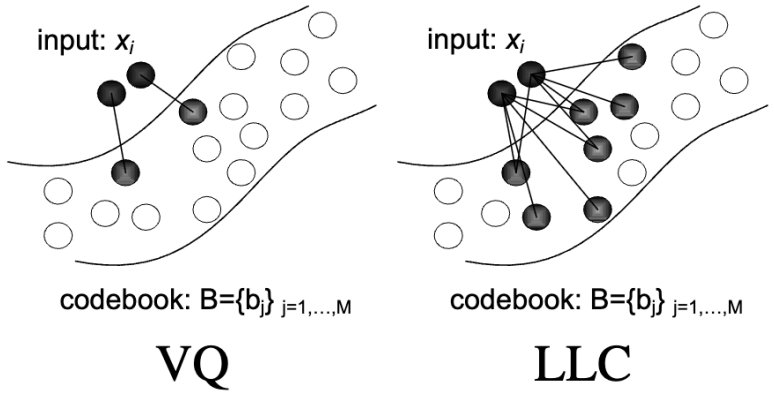
K=45

K → kth nearest codebook element using Euclidean distance

Empirical observations

- Codebook neighborhoods encode coherent semantic information
- Codebook is underutilized[a] (79/2048 ~ 4%)

1. Locality constrained Linear coding (Wang et al., 2010)
   replaces hard VQ in Bag-of-Features
   each descriptor → into its local coordinate system (K nearest bases)
   - How to extend LLC to deep neural nets?



Wang, Jinjun, et al. "Locality-constrained linear coding for image classification." *IEEE computer society conference on CVPR*, 2010.

2. Improving Codebook Utilization

Category	Cons	Names
Regularization & Reset	Agnostic to the local geometry of latent space	HVQ-VAE, Jukebox
Soft & Stochastic Quantization	Random/poor selection of codes	SQVAE, Affine Reparam, CVQ-VAE, Gumbel

- How local-structure/feature-covariance improve codebook usage?

[a]Huh, Minyoung, et al. "Straightening out the straight-through estimator: Overcoming optimization challenges in vector quantized networks." ICML, 2023.

Locality Constrained Vector Quantization

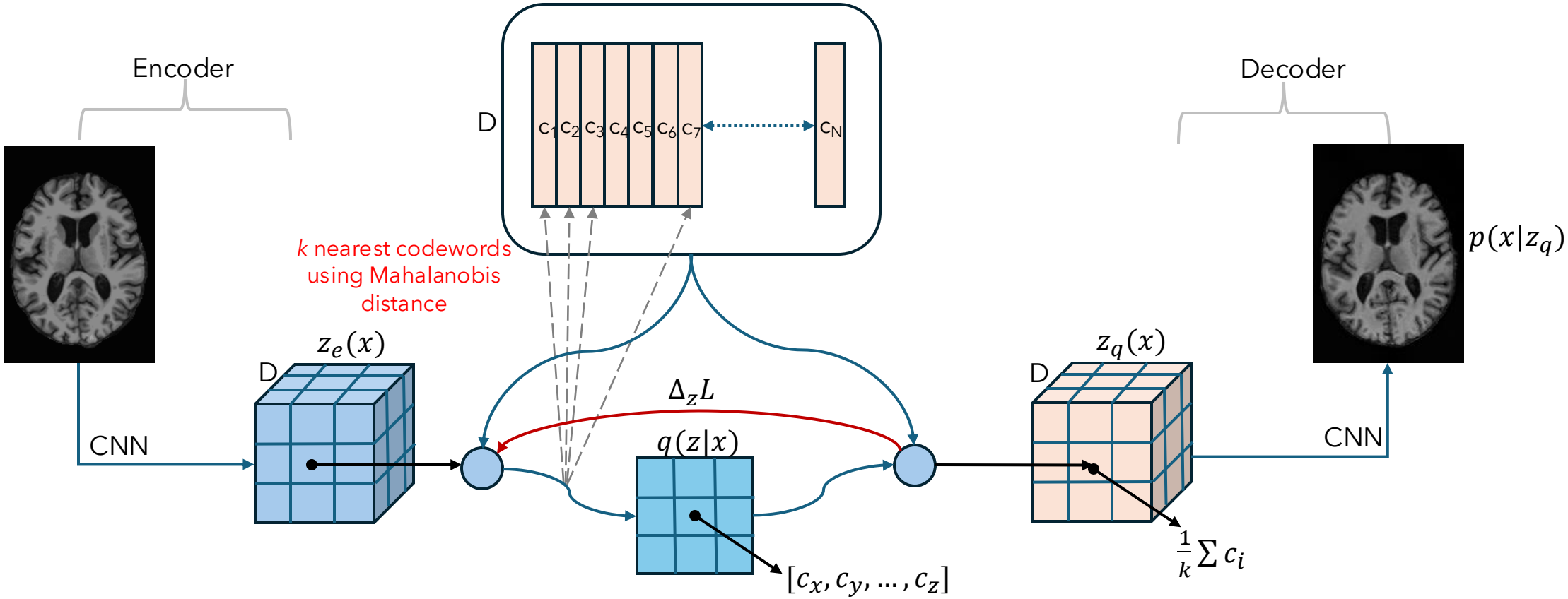
1. Locality informed soft-quantization
2. Optimal codebook usage

Proposed method

Locality constrained VQ (LCVQ)

Problem: Euclidean nearest-neighbor ignores latent covariance

Goal: Leverage local latent geometry via *Mahalanobis distance* to improve codebook utilization and reconstruction fidelity



LCVQ

1. Center codebook & calculate covariance
   - How latent dimensions co-vary across embeddings
2. Mahalanobis distance to each c_i
   - Prioritizes codewords that lie along high-variance axes
3. Top-K selection & aggregation
   - Instead of a single "hard" nearest neighbor, average the K most "informative" neighbors under the true geometry

Algorithm 1: Mahalanobis-based Top-K Quantization for a single input x

Input: $x \in \mathbb{R}^D$, $C \in \mathbb{R}^{N \times D}$ (codebook), K

$\bar{C} = C - \text{mean}(C)$ # Center codebook

$\Sigma = \frac{1}{N-1} \bar{C}^\top \bar{C}$ # Compute covariance

$\Sigma^{-1} = \text{pinv}(\Sigma)$ # Pseudo-inverse for stability

$d_i = \sqrt{(x - C_i)^\top \Sigma^{-1} (x - C_i)} \quad \forall i = 1, \ldots, N$ # Mahalanobis distances

$I = \text{argsort}(d)[1:K]$ # Indices of top-K neighbors

$q = \frac{1}{K} \sum_{i \in I} C_i$ # Average embeddings

Return q

Richer Representations → Leverages local neighborhoods in covariance-adjusted space.

Higher Codebook Utilization → Spreads assignments across more embeddings

Smoother Reconstructions → Soft aggregation reduces quantization artifacts, anatomical errors.

Results (Reconstruction)

Methods	MSE (10^{-3}) ↓	MS-SSIM ↑	Perplexity ↑
VQ-VAE[a]	2.01	0.9421	66.5
w/ Affine[b]	2.12	0.9419	63.1
w/ LCVQ (k=15)	**1.20**	**0.9684**	**368.9**
w/ Affine[b] + LCVQ (k=15)	1.52	0.9602	354.9

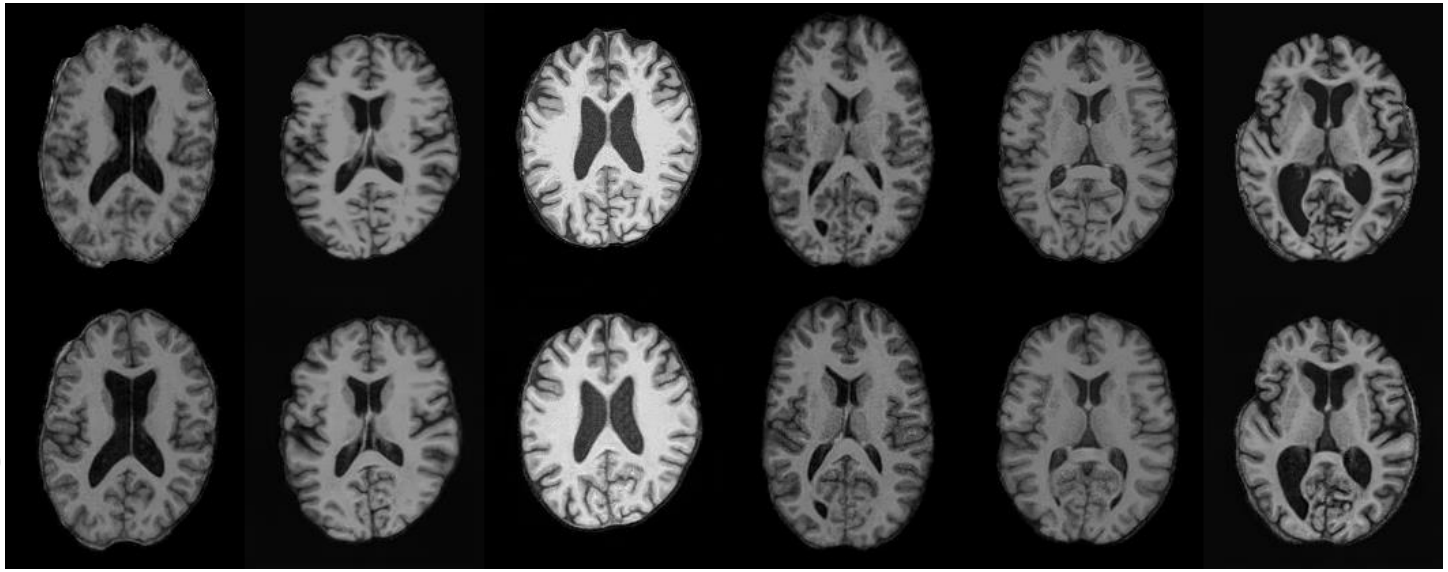
Perplexity → codebook usage

Given a distribution q over codebook entries, the perplexity is defined as:

$$\text{codebook perplexity} = \exp\left(-\sum_i q_i \log q_i\right)$$

Original

Predictions (LCVQ)



[a]Tudosiu, Petru-Daniel, et al. "Realistic morphology-preserving generative modelling of the brain." *Nature Machine Intelligence* (2024)

[b]Huh, Minyoung, et al. "Straightening out the straight-through estimator: Overcoming optimization challenges in vector quantized networks." ICML, 2023.

Results (Downstream)

Brain Age prediction

On Lifespan (healthy)

Data	MAE (years)	Training time
Raw Imaging (w/o quantization)	5.10	~ 4 days
Quantized Imaging (LCVQ)	5.32	~ 15 minutes

Raw Imaging → (176, 208, 176)
Quantized Imaging → (11, 13, 11)

Compression ratio → 4096: 1

On Discovery (Unhealthy)

Data	MAE (years)				Brain Age delta (years)			
	AD ↑	MCI	HC ↓	All	AD ↑	MCI	HC ↓	All
Raw Imaging (w/o quantization)	5.93	4.93	4.58	5.00	3.01	-0.14	-2.15	-0.21
Quantized Imaging (LCVQ)	6.01	4.53	4.29	4.73	3.43	0.50	-1.56	0.36

AD=Alzheimer's disease, MCI=mild cognitive impairment, HC=healthy controls

Future work

Ablation studies (k)

- k=5 → perplexity ↑ (tight neighborhood)
- k=100 → MSE ↑ SSIM ↓ (diffuse neighborhood)

k in VQVAE + LCVQ	MSE (10^{-3}) ↓	MS-SSIM ↑	Perplexity ↑
5	1.40	0.9652	**391.9**
15	**1.20**	**0.9684**	368.9
25	1.21	0.9678	330.5
50	1.32	0.9679	329.2
75	1.32	0.9672	258.7
100	1.44	0.9627	295.3
Adaptive (k=15, α=10)	1.81	0.9544	316.94

- Outperforms very large k
- Underperforms fixed-15 on some metrics
- α and k need refinement.

Adaptive k via χ^2 Thresholding

Squared Mahalanobis distance follows a chi-squared dist with D degrees of freedom[a]:

$$d_M^2(x, c_i) = (x - c_i)^T \sum^{-1} (x - c_i) \sim \chi_D^2$$

Thresholding:

$$\tau^2 = \chi_D^2 (1 - \alpha),\ \mathcal{N}_\tau(x) = \{c_i | d_M^2(x, c_i) \leq \tau^2\}$$

If $|\mathcal{N}_\tau| = 0$, revert to top-k

[a]https://en.wikipedia.org/wiki/Mahalanobis_distance

Findings

1. Locality constrained quantization
   Co-variance aware distance for high entropy codeword selection
   Addresses codebook collapse

2. Top-k codeword aggregation
   Reduce quantization artifacts, capture data diversity

3. Retains downstream performance
   Lesser computation cost

4. Use cases
   Low resource settings, federated learning, storage optimization

Improving PET quantification using Deep Learning

Motivation

Amyloid PET

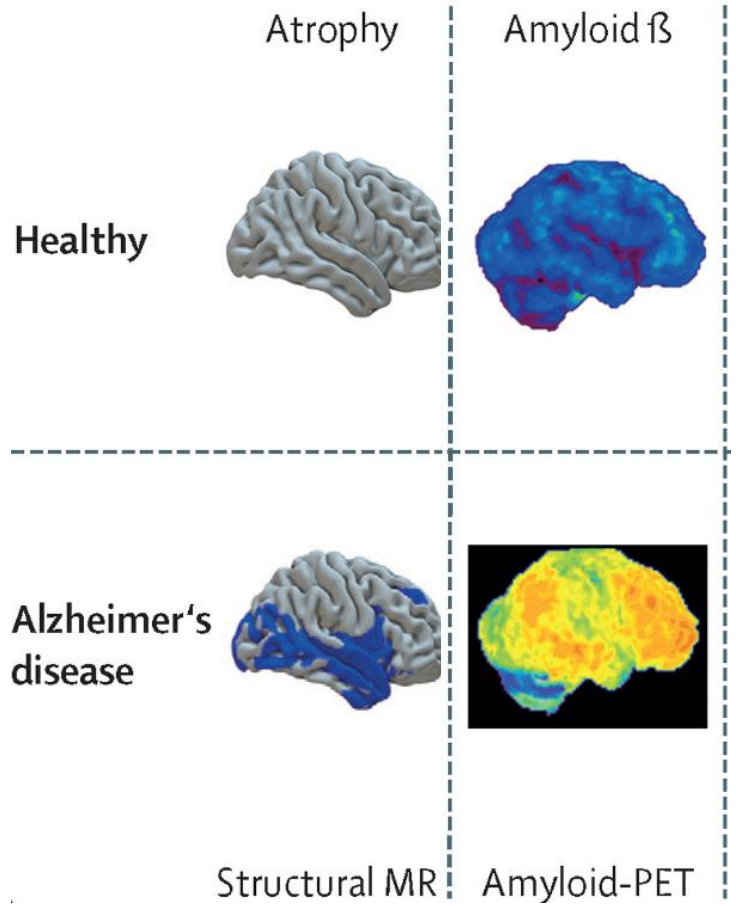
1. Measures amyloid beta (Aβ) protein deposits in brain

2. Detect pathological changes <u>earlier</u> than clinical symptoms (~15 yrs)

Comparison

- MRI shows general neurodegeneration
  Amyloid PET is more specific to AD pathology

- Amyloid PET can detect earlier pathological changes than MRI



Chételat, Gaël, et al. "Amyloid-PET and 18F-FDG-PET in the diagnostic investigation of AD and other dementias." *The Lancet Neurology* (2020)

Motivation

(1) Low Spatial Resolution

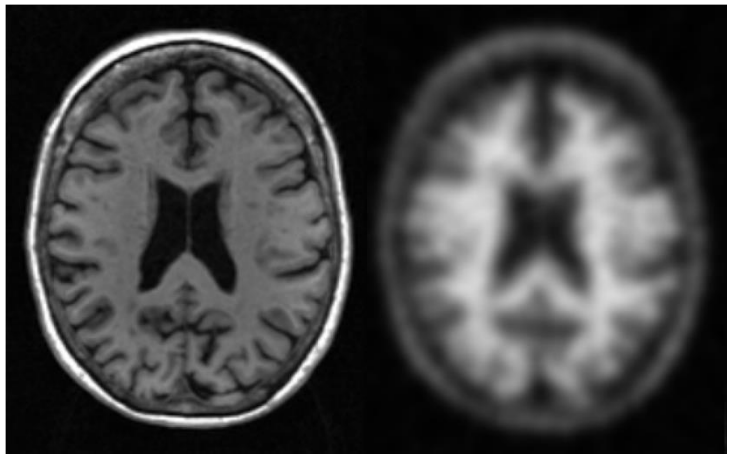
Scanners at 4-6 mm FWHM – partial volume effect (PVE)

Size of the object is smaller than twice the FWHM of scanner

1. Underestimation of radiotracer uptake
   Especially in small brain structures

2. Spillover between GM, WM, and CSF
   Within tissue variability complicates correction

3. Longitudinal tracking
   Hampers monitoring disease progression



T1w MRI PET (FBP)

Motivation

(2) Lack of Standardization

- Multiple PET tracers exist[a]
  florbetapir (FBP), florbetaben (FBB), flutemetamol, and NAV4694

- Cross-tracer variability
  Tracer-dependent characteristics leads to inconsistencies
  Lack of consensus in multi-center studies[a]

Super-Resolution

- *To address PVE (due to low-res)*
- *To reduce inter-tracer variability*

[a]Shah, Jay, et al. "Deep residual inception encoder-decoder network for amyloid PET harmonization." *Alzheimer's & Dementia* (2022)

Existing Gaps

Low Spatial Resolution of PET Imaging

Partial Volume Correction (PVC)

1. Iterative deconvolution methods

   - Unblur image by estimating & removing PSF of imaging
   - Iteratively deconvolve using the estimated PSF

2. Challenges

   - Noise amplification
   - Low resolution recovery
   - These are region-based approaches

     Ideal is voxel-level resolution recovery

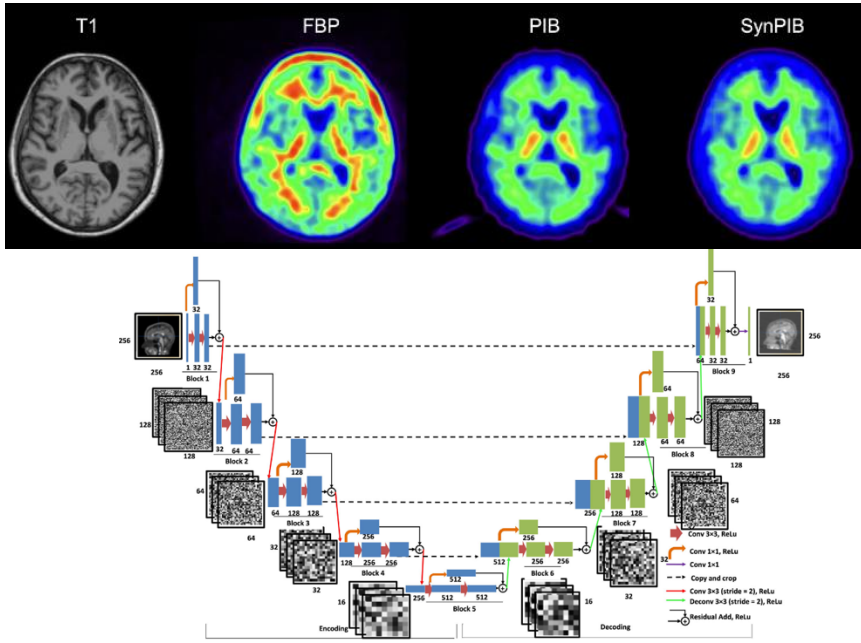
Tohka, Jussi, and Anthonin Reilhac. "Deconvolution-based partial volume correction in Raclopride-PET and Monte Carlo comparison to MR-based method." *Neuroimage* (2008)

Existing Gaps

Cross-tracer Harmonization

Lack of Standardization

1. Paired image-to-image translation
   - Cross-tracer translation
2. Challenges
   - Can’t generalize to other tracers
   - Loss of tracer-specific information
   - Requires paired datasets
   - Bias from imperfect translation



Shah, Jay, et al. "Deep residual inception encoder-decoder network for amyloid PET harmonization." Alzheimer's & Dementia 18.12 (2022): 2448-2457.

Phase III

Improving PET quantification using Diffusion model based Super-resolution

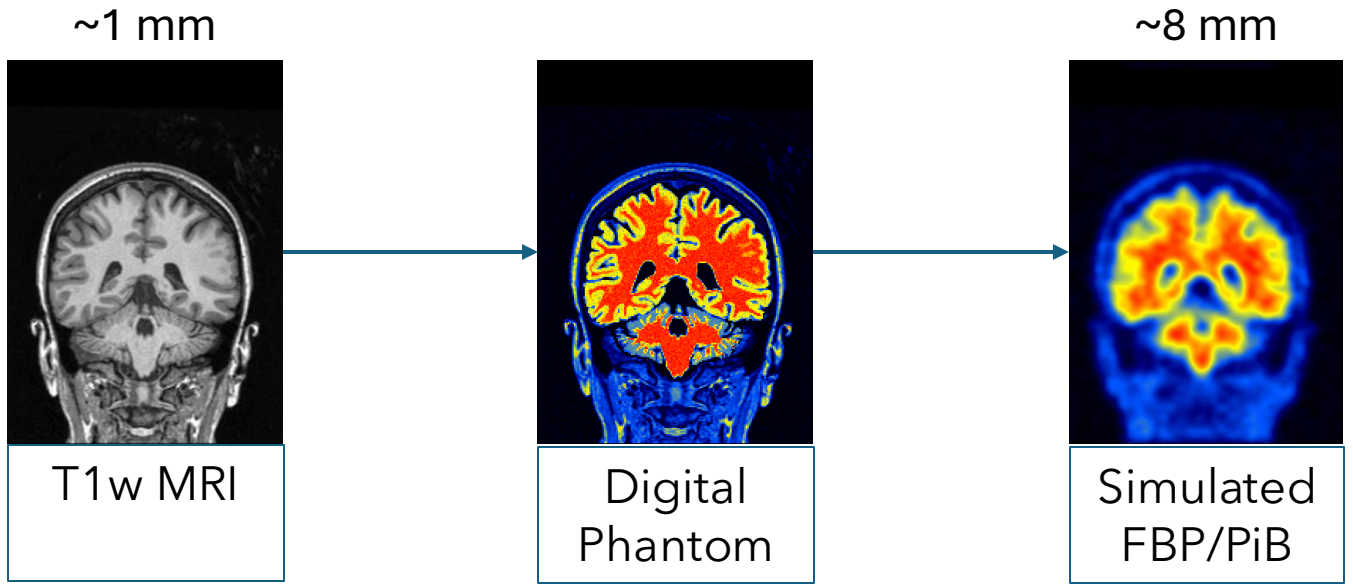
Objectives

1. Improve absolute quantification
2. Detect longitudinal changes (progression)
3. Improve cross-tracer Harmonization

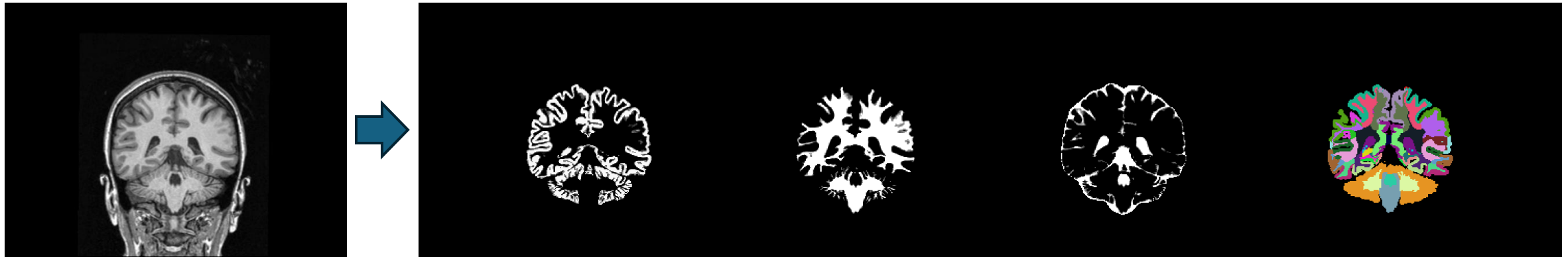
Data Simulation

Ground truth high resolution PET does not exist!

Data Simulation – based on PET imaging physics



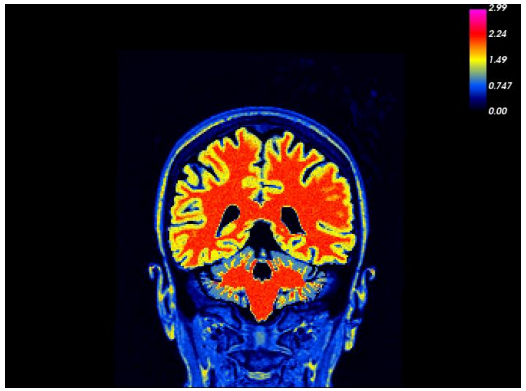
Digital Phantoms



V=f_gm*gmv + f_wm*wmv + f_csf*csfv + f_bg*bgv + f_gm*abetav

Tracer 1 (PIB)
gmv ~ (1, 0.04)
wmv ~ (2.2, 0.066)
csfmv ~ (0.05, 0.001)
bgv = nt1v*.7
abetav ~ (0.5, 0.1)

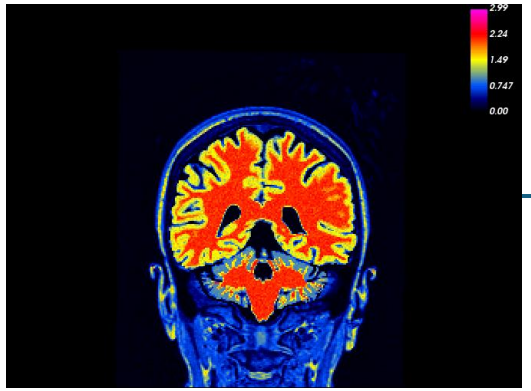
f_gm+f_wm+f_csf+f_bg = 1
Cerebellum-Cortex abeta = 0
Brain-Stem abeta = 0
Normalize to Cerebellum-Cortex
MCSUVR=TargetROI/Cerebellum-Cortex



Tracer 2 (FBP)
gm2=gm1
wm2=1.2*wm1
csf2=csf1
bg2 = 1.2*bg1
abeta2 =0.75*abeta1

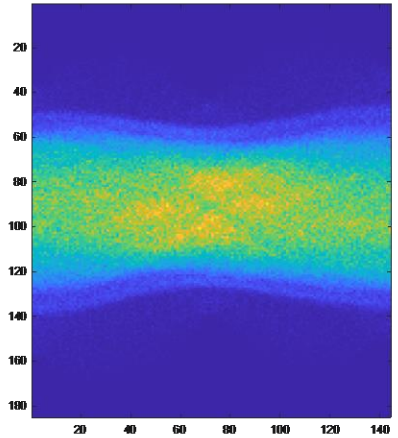
Su, Yi, et al. "Partial volume correction in quantitative amyloid imaging." *Neuroimage* (2015)

Simulated PET

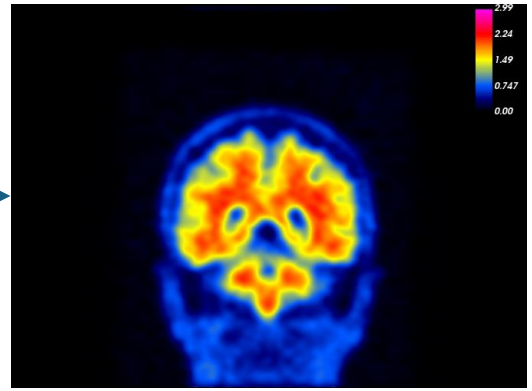


Digital Phantom
(dpPIB)

smoothing
forward projection
add noise



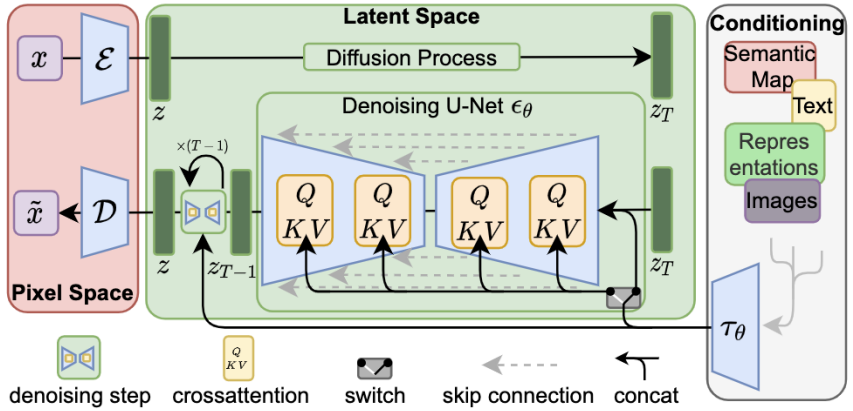
reconstruction
smoothing



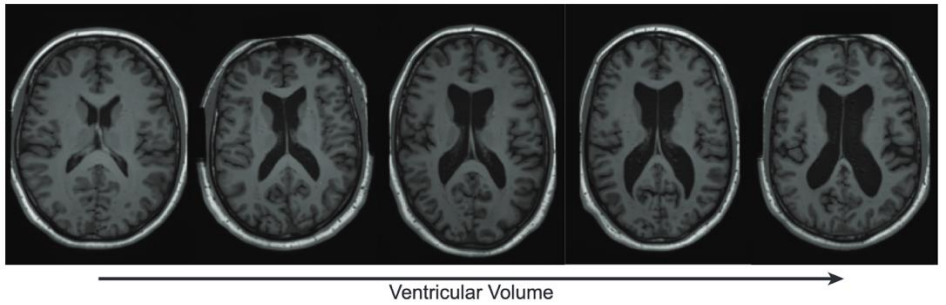
Simulated PET
(spPIB)

Diffusion Models

- Diffusion Models outperform GANs in medical imaging synthesis[a]
  - More diversity, stable training, conditioning strategies
  - Limited by computation complexity
- Latent diffusion models
  - Denoising in latent space
  - Ideal for medical imaging (256^3 dim data)
- Ideal for Super-Resolution



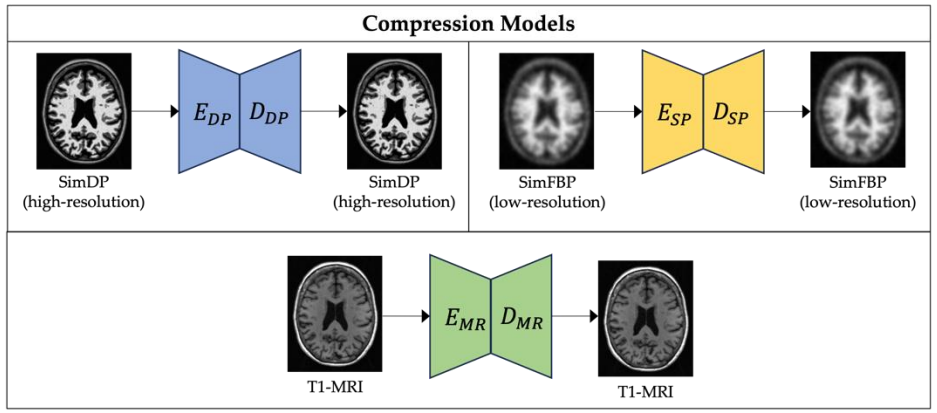
Rombach, Robin, et al. "High-resolution image synthesis with latent diffusion models." *CVPR* (2022)

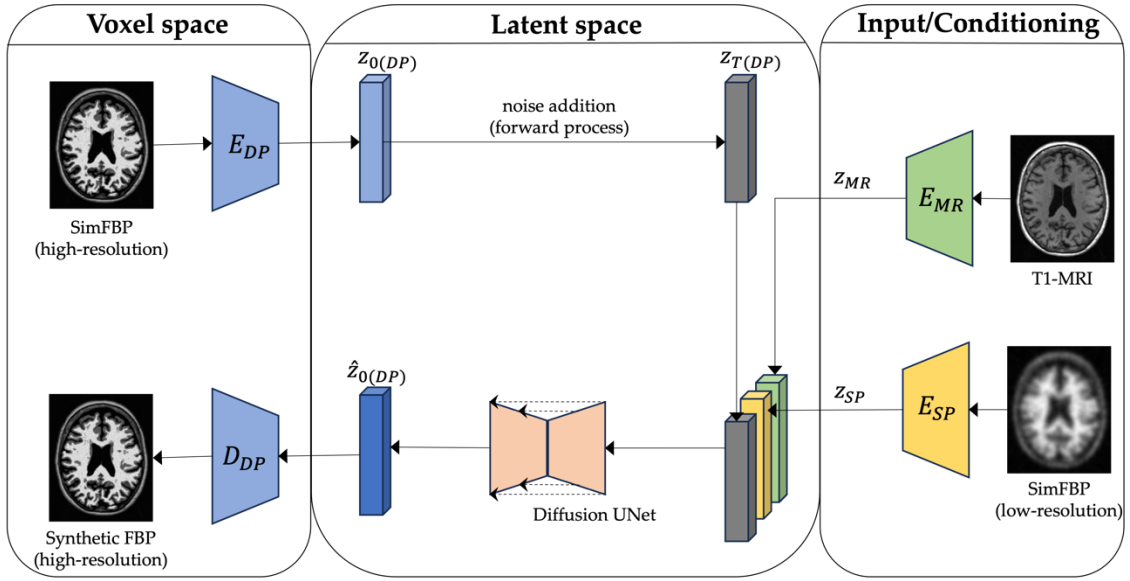


Pinaya, Walter HL, et al. "Brain imaging generation with latent diffusion models." *MICCAI Workshop on Deep Gen Models* (2022)

[a]Khader, Firas, et al. "Denoising diffusion probabilistic models for 3D medical image generation." *Scientific Reports* (2023)

Method (baseline)





AutoencoderKL (3D)[a]

- Attention layers only at last level
- 32 base channels, with multiplier of [1,2,2]
- one residual block per level
- latent space [16×16×16], 3 latent channels.
- 80 training epochs, minibatch of 60
- Adam optimizer, base lr=0.0001.
- patch-based discriminator in our adversarial loss with 32 base channels, lr=0.0001.

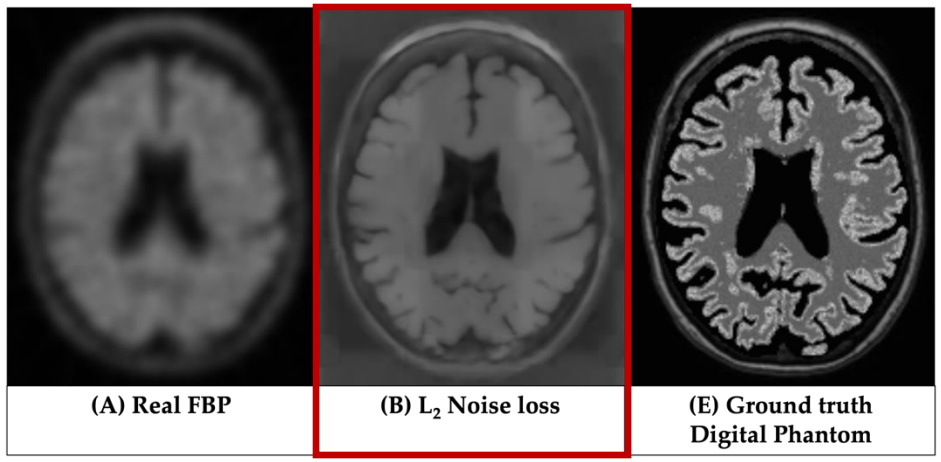
LDM (3D)

- U-net architecture, 32 base channels, multiplier of [1,2,2]
- one residual block per level
- 9 input channels (3 each for simFBP, simDP, MRI latents).
- Adam optimizer with a base lr=0.0001.
- DDPM scheduler with 1000 timesteps (training), with a linear variance schedule (0.0015, 0.0195)
- DDIM scheduler with 250 timesteps (inference)

[a]Pinaya, Walter HL, et al. "Brain imaging generation with latent diffusion models." *MICCAI Workshop on Deep Gen Models* (2022)

Preliminary results

Inputs:	low-res PET, T1 MRI
Diffusion-Unet:	L_2 loss on noise scale $(\epsilon, \hat{\epsilon})$



Observations

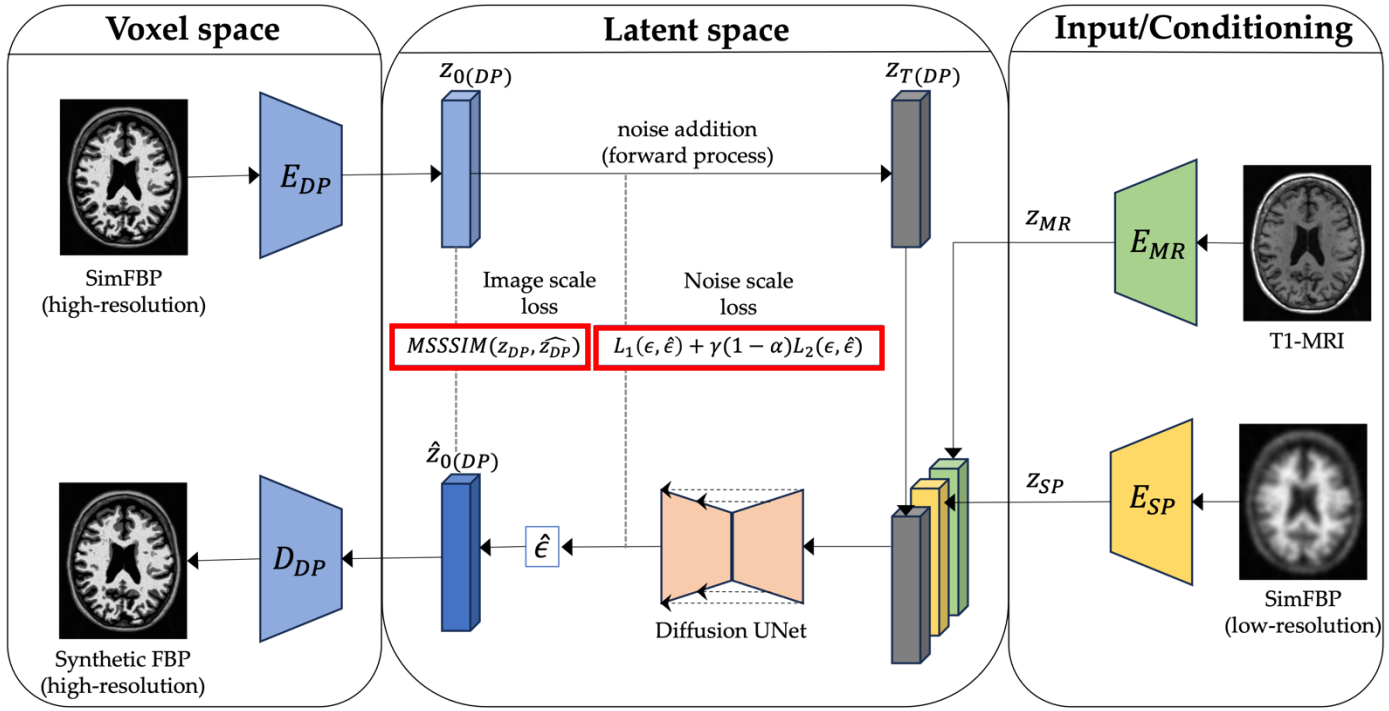
1. Minimizing loss on noise-scale does not guarantee accurate image-scale reconstruction
2. Cannot retain (brain) structure information
3. Combination of L_1/L_2 and MS-SSIM[b] loss is more suitable for image restoration/super-resolution[a]
    - L_2 can be sensitive to outliers
    - L_1 suffers non-differentiability at zero

Preserving structure & voxel level intensity is key to PET Quantification accuracy!

[a]Zhao, Hang, et al. "Loss functions for image restoration with neural networks." *IEEE Transactions on computational imaging* (2016)
[b]MS-SSIM=multi scale structural similarity index

Method

Latent diffusion model for resolution recovery (LDM-RR)



$$\hat{z_0} = \frac{z_t - \sqrt{1-\alpha_t}\hat{\epsilon}}{\sqrt{\alpha_t}}$$

$$loss_1 = (1-\alpha)L_2(z_0, \hat{z_0}) + \alpha MSSSIM(z_0, \hat{z_0})$$

$$loss_2 = L_1(\epsilon, \hat{\epsilon}) = |\epsilon - \hat{\epsilon}|$$

$$loss_{combined} = \underbrace{L_1(\epsilon, \hat{\epsilon}) + \gamma(1-\alpha)L_2(\boldsymbol{\epsilon}, \hat{\boldsymbol{\epsilon}})}_{\text{noise-scale}} + \underbrace{\alpha MSSSIM(\boldsymbol{z_0}, \hat{\boldsymbol{z_0}})}_{\text{image-scale}}$$

Shah, Jay, et al. "Enhancing Amyloid PET Quantification: MRI-Guided Super-Resolution Using Latent Diffusion Models." Life 14.12 (2024): 1580.

Datasets

Training cohort

- 3,376 MRI scans from ADNI
- Simulated paired FBP/PiB scans

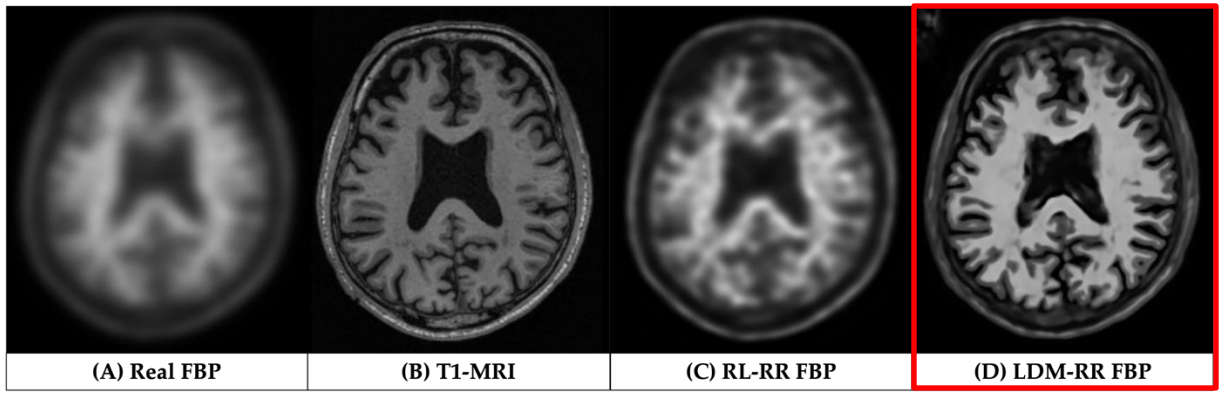
Longitudinal cohort

Cross-tracer cohort

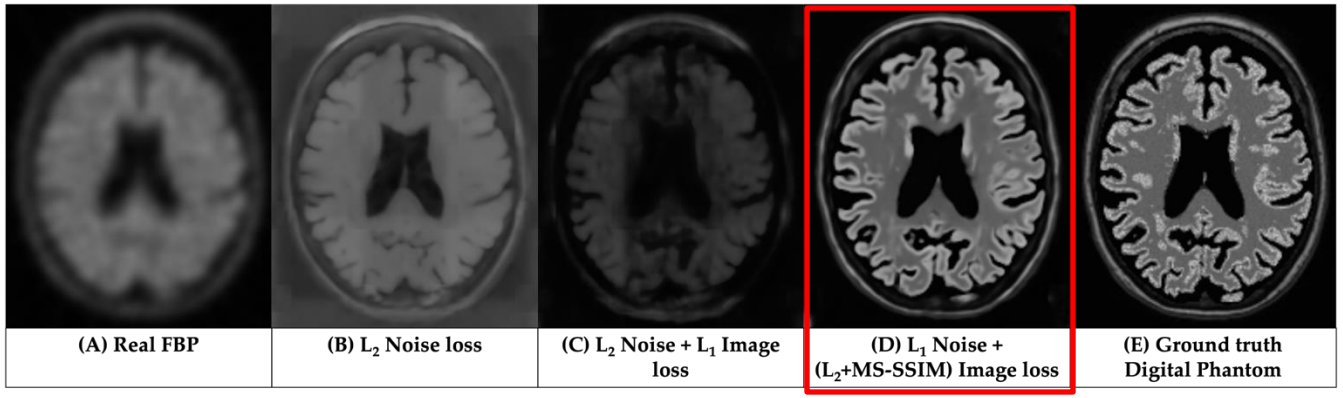
Cohort	ADNI	OASIS-3	Centiloid
Sample count	334 FBPs 167 baseline-followup	113 (FBP-PIB pairs)	46 (FBP-PIB pairs)
Age (SD) yrs	75.1 (6.9)	68.1 (8.7)	58.4 (21.0)
Education (SD) yrs	16.1 (2.7)	15.8 (2.6)	NA
Male (%)	182 (54.5%)	48 (42.5%)	27 (58.7%)
Cognitive impairment (%)	236 (70.6%)	5 (4.4%)	24 (52.2%)
APOE4+ (%)	218 (65.3%)	38 (33.6%)	15 (46.9*%) [*14/46 unknown]
PET interval (SD) yrs	2.0 (0.06)	NA	NA

Results

Qualitative analysis



Compared to traditional Iterative deconvolution-based correction method[a]



Compared to traditional LDMs with L_2 loss minimization on noise scale

[a]Tohka, Jussi, and Anthonin Reilhac. "Deconvolution-based partial volume correction in Raclopride-PET and Monte Carlo comparison to MR-based method." *Neuroimage* (2008)

Results

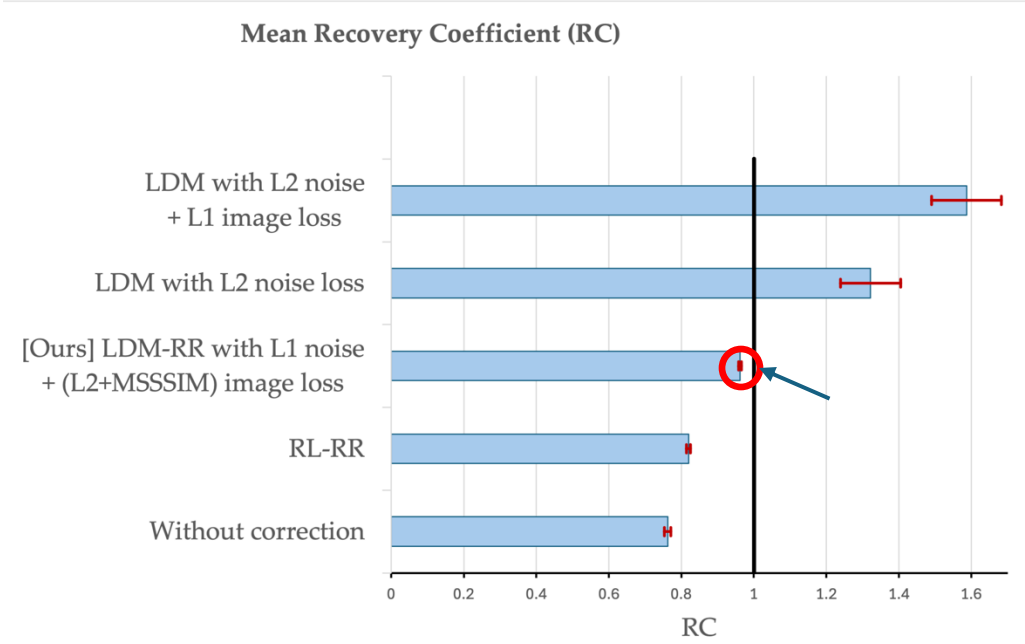
On Simulated dataset

MCSUVR: Mean Cortical Standardized Uptake Value Ratio
measures amyloid plaque accumulation in brain PET
normalized measure of radiotracer uptake.

1. RC: Recovery Coefficient

   - $RC = \frac{MCSUVR\ (Synthetic\ DP)}{MCSUVR\ (Simulated\ DP)}$

   - Synthetic DP = synthetic high-resolution PET
     Simulated DP = simulated digital phantom
   - ~1 is ideal



Comparison of mean recovery coefficient (RC) using different methods on a held-out test of 338 samples randomly selected from the simulated dataset.

Results

On Longitudinal cohort

1. Annualized Rate

   - $rate = \frac{\Delta MCSUVR\ (followup - baseline)}{time\ interval\ (yrs)}$

   - A higher rate of change = higher statistical power to detect longitudinal changes in amyloid deposition

2. SS: Sample Size

   - # participants per arm needed to detect a 25% reduction in amyloid accumulation rate due to treatment with 80% power and a two-tailed type-I error of p=0.05 in hypothetical anti-amyloid treatment trials.

   - A smaller SS indicates greater statistical power.

Annualized rate	Raw	RL-RR	LDM-RR
Mean	0.0278	0.0377	**0.0459**
SD	0.0664	0.0807	0.0881
p-value	1.0E-07	5.0E-09	1.3E-10
SS	1431	1154	**926**

Statistical power in detecting longitudinal changes measured

Results

On Cross-tracer cohort

- Combined dataset (OASIS + Centiloid)
- Agreement of PET-derived global amyloid burden between FBP and PiB
- Using Pearson correlation & Steiger's t-test p-values

Method	Pearson Correlation	Steiger's p-value
w/o Correction	0.9163	N/A
RL-RR	0.9308	<0.0001 (RL-RR vs. without correction)
LDM-RR	**0.9411**	0.0001 (LDM-RR vs. without correction)
		0.0421 (LDM-RR vs. RL-RR)

Comparison of RL and LDM-RR methods in improving the MCSUVR agreement between FBP and PIB tracers shown by Pearson correlation and Steiger's test.

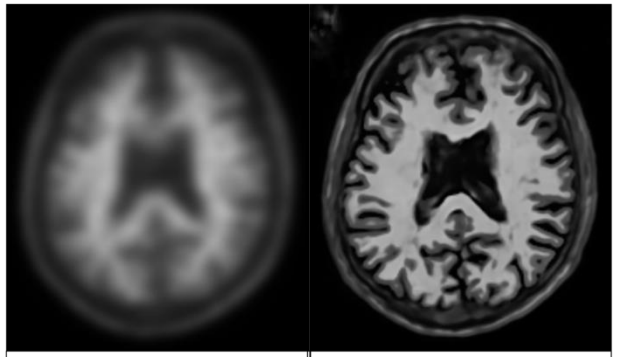
Findings

1. Super-resolution with latent diffusion models

   - Adding image-scale loss penalty can preserve global image structure
   - Noise-scale loss does not guarantee accurate reconstruction

2. Super-resolution for PET Quantification

   - Using simulated dataset from domain knowledge
   - Voxel-level enhancement using MRI
   - Better longitudinal tracking
   - Cross-tracer harmonization



3. Deep Learning can address Partial volume effect in PET

   - Bettering Early detection and disease monitoring

Products

1. Shah, J., Siddiquee, M. M. R., Su, Y., Wu, T., & Li, B. (2024). Ordinal Classification with Distance Regularization for Robust Brain Age Prediction. *In Proceedings of the IEEE/CVF WACV*

2. Shah, J., Gao, F., Li, B., Ghisays, V., Luo, J., Chen, Y., ... & Wu, T. (2022). Deep residual inception encoder-decoder network for amyloid PET harmonization. *Alzheimer's & Dem*entia
   - [**Patent**]: Gao, F., Su, Y., Shah, J. and Wu, T., *Banner Health and Arizona State University*, (2025). Deep residual inception encoder-decoder network for amyloid PET harmonization. U.S. Patent 12,186,114.

3. Shah, J., Che, Y., Sohankar, J., Luo, J., Li, B., Su, Y., & Wu, T. (2024). Enhancing PET Quantification: MRI-Guided Super-Resolution Using Latent Diffusion Models. *Life Journal*

4. Shah, J., Siddiquee, M. M. R., Krell-Roesch, J., Syrjanen, J. A., Kremers, W. K., Vassilaki, M., Forzani, E., Wu, T. & Geda, Y. E. (2023). Neuropsychiatric Symptoms and Commonly Used Biomarkers of Alzheimer's Disease: A Literature Review from a Machine Learning Perspective. *Journal of Alzheimer's Disease*

5. Shah, J., Krell-Roesch, J., Forzani, E., Knopman, D., Jack, C., Peterson, R., Che, Y., Wu, T. & Geda, Y. E. (2023). Predicting cognitive decline from neuropsychiatric symptoms and Alzheimer's disease biomarkers: A machine learning approach to a population-based data. *Journal of Alzheimer's Disease*

6. Siddiquee, M. M. R., Shah, J., Wu, T., Chong, C., Schwedt, T. J., Dumkrieger, G., ... & Li, B. (2024). Brainomaly: Unsupervised neurologic disease detection utilizing unannotated t1-weighted brain mr images. *In Proceedings of the IEEE/CVF WACV*

Products

7. Siddiquee, M. M. R., Shah, J., Wu, T., Chong, C., Schwedt, T., & Li, B. (2022, September). Healthygan: Learning from unannotated medical images to detect anomalies associated with human disease. *MICCAI SASHIMI*

8. Siddiquee, M. M. R., Shah, J., Chong, C., Nikolova, S., Dumkrieger, G., Li, B., ... & Schwedt, T. J. (2023). Headache classification and automatic biomarker extraction from structural MRIs using deep learning. *Brain Communications*

9. Trivedi, M. R., Shah, J., Readhead, B., Su, Y., Wu, T., & Wang, Q. (2023, December). Interpretable deep learning framework towards understanding molecular changes in human brains with Alzheimer's disease: implication for microglia activation and sex differences in AD. *Nature Publishing Journal, Aging*

10. Barisch-Fritz, B., Shah, J., Krafft, J., Geda, Y., Wu, T., Woll, A., & Krell-Roesch, J. (2025). Physical activity and the outcome of cognitive trajectory: a machine learning approach. *European Review of Aging & Physical Activity*

11. Che, Y., Rafsani, F., Shah, J., Siddiquee, M. M. R., & Wu, T. (2025). AnoFPDM: Anomaly Segmentation with Forward Process of Diffusion Models for Brain MRI. *In Proceedings of the IEEE/CVF WACV*

In progress

1. Shah, J., Dumkreiger, G. Chong, C., Schwedt, T., Wu, T. (2025) Capturing Brain Ageing signatures across different Headache disorders using deep learning. *Brain*

2. Rafsani, F., Shah, J., & Wu, T. (2025) DinoAtten3D: Slice-Level Attention Aggregation of DinoV2 for 3D Brain MRI Anomaly Detection. *ICCV Workshop*

3. Rafsani, F., Sheth, D., Che, Y., Shah, J., Siddiquee, M. M. R., Chong, C., Nikolova, S., Dumkreiger, G., Li, B., Wu, T. & Schwedt, T. (2025). Using Large-scale Contrastive Language- Image Pre-training to Maximize MRI-based Headache Classification. *Nature Scientific Reports*

4. Joshi, A., Che, Y., Shah, J., Siddiquee, M. M. R., Chong, C., Nikolova, S., Dumkreiger, G., Li, B., Wu, T. & Schwedt, T. (2025). A Pilot Study: Traumatic Brain Injury Recovery Prediction with Harmonized Brain MRI and CT. *Brain Communications*

Abstracts

1. [**Oral presentation**] Shah, J., Siddiquee, M.M.R., ... & Wu, T. (2024). Capturing MRI Signatures of Brain Age as a Potential Biomarker to Predict Persistence of Post-traumatic Headache. *American Academy of Neurology, Annual Meeting* ***&*** *NIH Heal Annual Meeting*

2. Siddiquee, M.M.R., Shah, J., ... & Wu, T. (2024). Applying Generative Adversarial Network on Structural Brain MRI for Unsupervised Classification of Headache. *American Academy of Neurology, Annual Meeting* ***&*** *NIH Heal Annual Meeting*

3. Joshi, A., Siddiquee, M.M.R., Shah, J., ... & Wu, T. (2024). Prediction of Headache Improvement Using Multimodal Machine Learning in Patients with Acute Post-traumatic Headache. *American Academy of Neurology, Annual Meeting* ***&*** *NIH Heal Annual Meeting*

4. Shah, J., Luo, J., ... & Wu, T. (2023). A multi-class deep learning model to estimate brain age while addressing systematic bias of regression to the mean. *Alzheimer's Association International Conference*.

5. Shah, J., Sohankar, J., ... & Su, Y. (2023). A 2.5D residual U-Net for improved amyloid harmonization preserving spatial information. *Alzheimer's Association International Conference*.

6. Joshi, A., Shah, J., ... & Wang, Q. (2023). Interpretable deep learning framework towards understanding molecular changes associated with neuropathology in human brains with Alzheimer's disease. *Alzheimer's Association International Conference*.

7. Siddiquee, M.M.R., Shah, J., Schwedt, T., Chong, C.,... & Wu, T. (2022). Classification and Biomarker Discovery of Persistent Post-traumatic Headache (PPTH) Using Deep Learning on Structural Brain MRI Data. *INFORMS Annual Meeting*

Abstracts

8. Shah, J., Syrjanen, Krell-Roesch, J., ... & Geda, Y. (2023). Participant-specific interrogation of population-based data to predict cognitive decline from neuropsychiatric symptoms and neuroimaging biomarkers: A machine learning approach. *American Academy of Neurology Annual Meeting*

9. Shah, J., ... & Su, Y. (2022). MRI signatures of Brain Age in the Alzheimer's Disease continuum. *Alzheimer's Association International Conference*

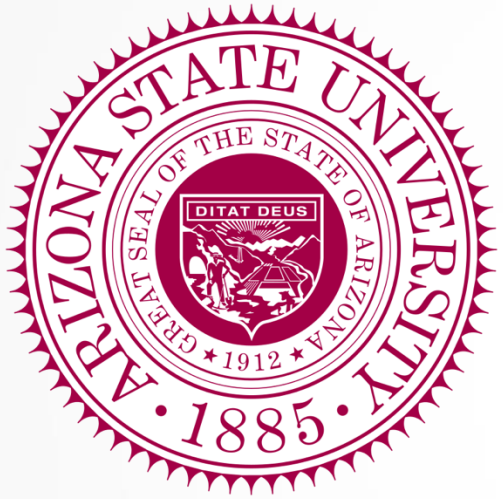
10. Shah, J., Chen, K., Reiman, E., Li, B., Wu, T., Su, Y. (2022). Transfer Learning based Deep Encoder Decoder Network for Amyloid PET Harmonization with Small Datasets. *Alzheimer's Association International Conference*

11. Siddiquee, M.M.R., Shah, J., Chong, C., Schwedt, T., ... & Wu, T. (2022). Classification of Post-Traumatic Headache (PTH) using Deep Learning on Structural Brain MRI data. *American Headache Society 64th Annual Scientific Meeting*

12. Siddiquee, M.M.R., Shah, J., Chong, C., Schwedt, T., ... & Wu, T. (2022). Migraine Classification using Deep Learning on Structural Brain MRI data. *American Headache Society 64th Annual Scientific Meeting*

13. Shah, J., Chong, C., Schwedt, T., ... & Wu, T. (2021). Interpreting Deep Learning Model Predictions using Shapley Values. *INFORMS Annual Meeting*

14. Shah, J., Ghisays, V., Luo, J., Chen, Y., Lee, W., Li, B., ... & Wu, T. (2021). Deep Residual Inception Encoded-Decoder Neural Network for amyloid PET harmonization. *Alzheimer's Association International Conference* ***&*** *Arizona Alzheimer's Consortium*



Thank You

Questions?

Jay Shah, PhD Student

School of Computing and Augmented Intelligence
Arizona State University

Email: jgshah1@asu.edu

Appendix

Ablation

k	**Method (Loss)**	**MAE**	**Ordinality**	**Systematic Bias**	
				SB-L	**SB-R**
1/2	CE	6.05	0.85	5.31	-5.19
2/3	CE	18.51	0.13	30.67	28.27
1	**CE**	**2.56**	**0.98**	**0.11**	**-2.5**
1	MSE	4.66	0.95	2.19	-4.98
2	CE	2.90	0.10	0.93	-3.04
2	MSE	4.57	0.95	4.83	-4.13

We use Manhattan distance in regularization:

$$\varphi(x_i) = \frac{1}{N-1} \sum_{j=1, i \neq j}^{N} |i-j||\bar{x}_i - \bar{x}_j|_{manh}$$

Exploring other L_k norm distances

$$L_k(x, y) = \sum_{i=1}^{d} [||x^i - y^i||^k]^{\frac{1}{k}}$$

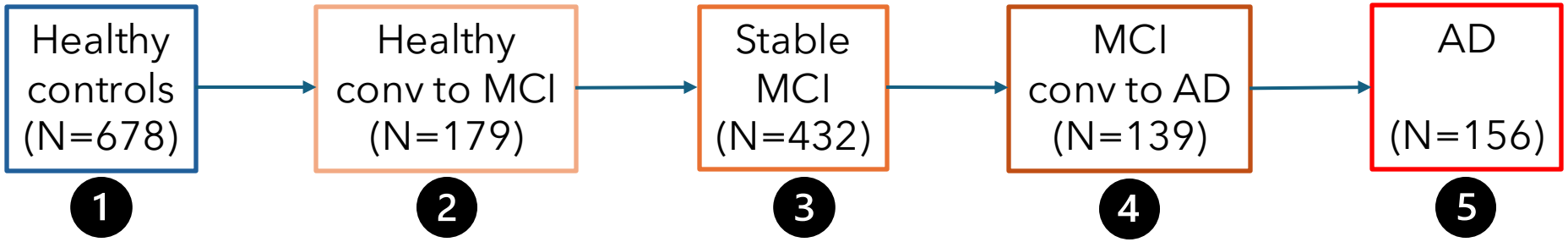
Results align with an established study*, which suggests Manhattan distance (L_1) is more suitable than Euclidean in high dimensional learning

*Aggarwal, Charu C., et al. "On the surprising behavior of distance metrics in high dimensional space." ICDT (2001).

Results

On Discovery (mixed) cohort

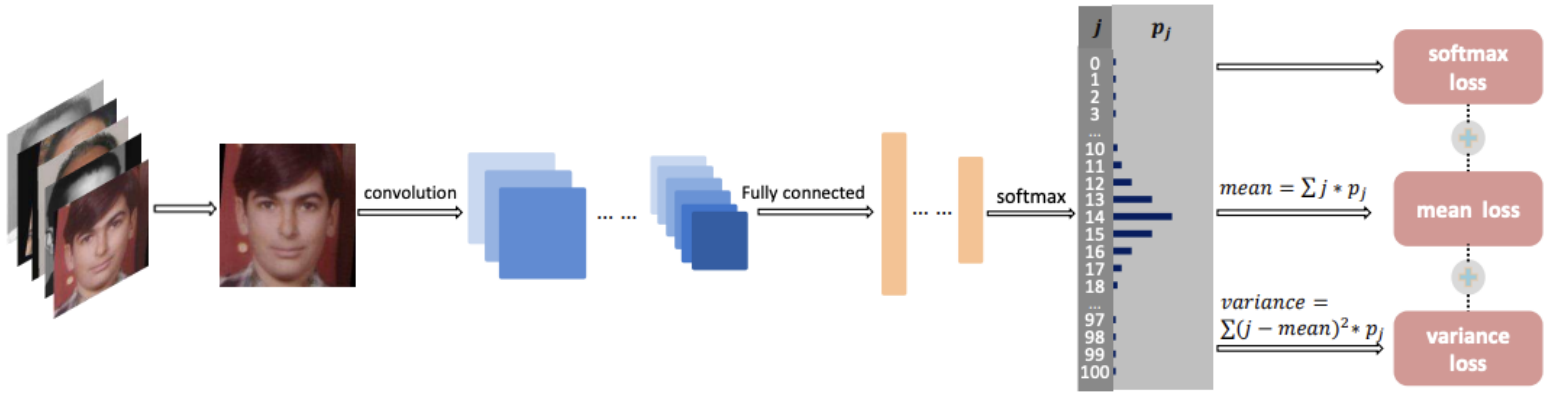
5 clinical groups with increasing order of severity



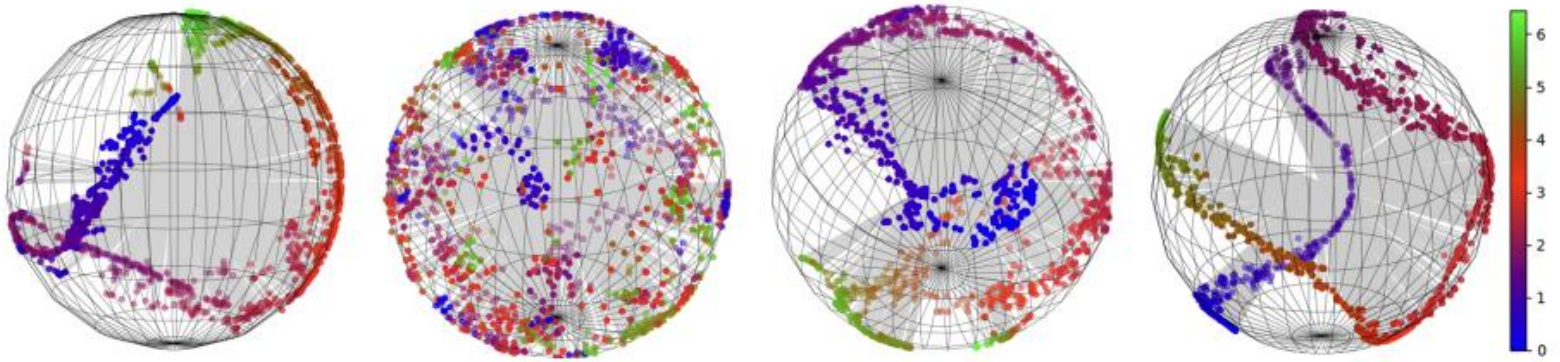
Definitions

1. HC conv to MCI: normal cognition at baseline, converted to MCI during follow-up.
2. MCI-stable: baseline diagnosis of MCI, unchanged in follow-ups.
3. MCI conv to AD: MCI diagnosis at baseline, subsequently converted to AD.
4. AD: diagnosed with AD at baseline.

Methods



Mean-variance Loss

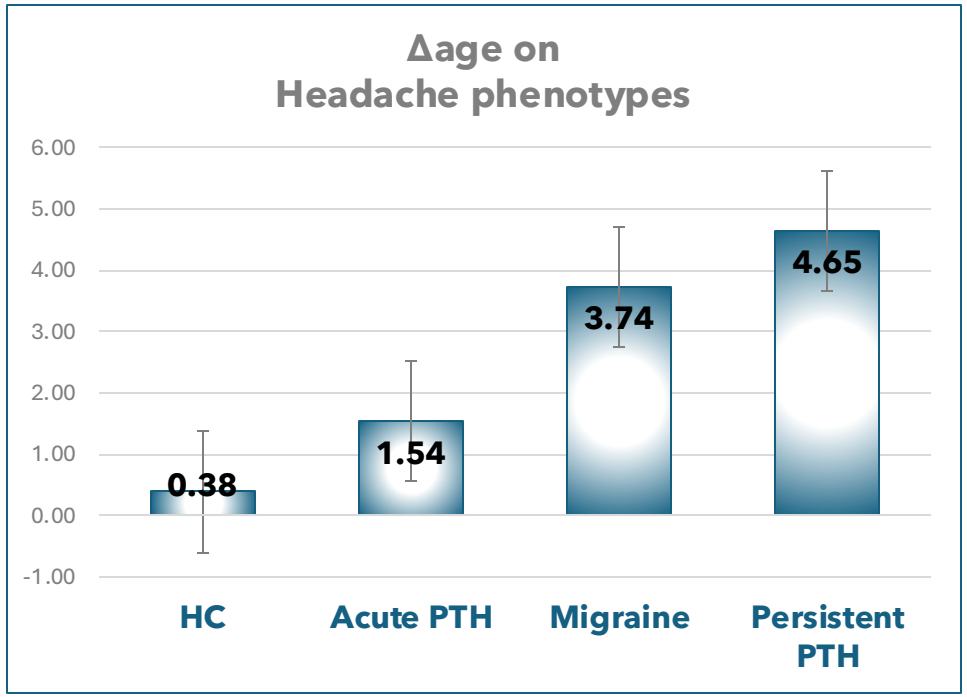


(a) Regression (b) Regression $+\mathcal{L}'_d$ (c) Regression $+\mathcal{L}_d$ (d) Regression $+\mathcal{L}_d + \mathcal{L}_t$

MSE + Euclidean Norm
(Ordinal entropy Loss)

[a]Zhang, Shihao, et al. "Improving Deep Regression with Ordinal Entropy." ICLR (2023).
[b]Pan, Hongyu, et al. "Mean-variance loss for deep age estimation from a face." CVPR (2018).

Headache detection



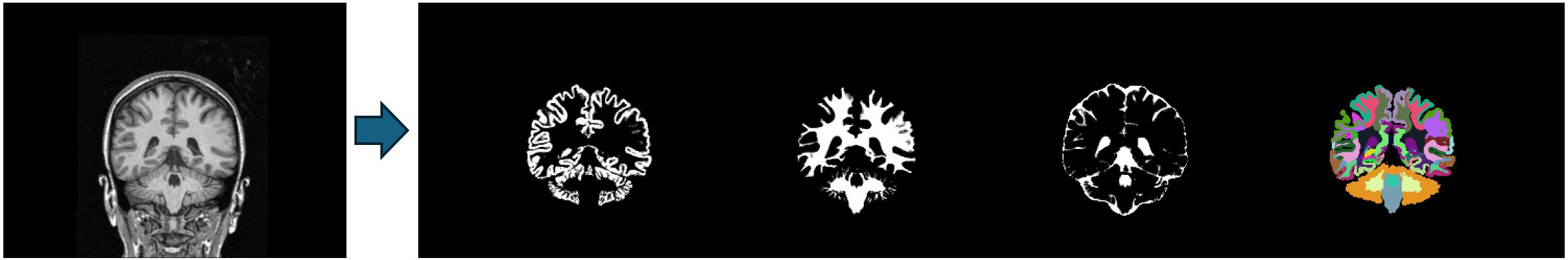
HC = Healthy Controls
PTH = Post Traumatic Headache

*in-house data collected from Mayo Clinic, Arizona

Findings:

- Δage(P-PTH) < Δage(A-PTH)
  suggesting more structural decline related to PTH persistence over time
- Headache frequency associated with structural damage
  Δage(P-PTH) > Δage(Mig) > Δage(A-PTH)
- Early detection potential
  structural decline acutely following TBI at risk for developing persistent PTH

Digital Phantoms



$$V = f_{gm} * gmv + f_{wm} * wmv + f_{csf} * csfv + f_{bg} * bgv + f_{gm} * abetav$$

V:	The observed PET signal in a voxel
f_gm:	The true radiotracer concentration in gray matter
f_wm:	The true radiotracer concentration in white matter
f_csf:	The true radiotracer concentration in cerebrospinal fluid
f_bg:	The true radiotracer concentration in background (non-brain tissue)
gmv:	The fraction of gray matter in the voxel
wmv:	The fraction of white matter in the voxel
csfv:	The fraction of cerebrospinal fluid in the voxel
bgv:	The fraction of background tissue in the voxel
abetav:	The fraction of amyloid-beta in the voxel

Comparisons

Methods	Pros	Cons	For Medical Imaging
LCVQ	• Captures anisotropic latent structure • High, organic codebook utilization (no resets)	• Slight extra compute	Best: sharp reconstructions, reproducible latents, ↓ MSE / ↑ MS-SSIM
Soft / Stochastic Quantization	• End-to-end differentiable; prevents hard collapse • Minimal extra parameters	• Probabilistic sampling → noisy, less deterministic latents • Over-softens → blurred details; can hurt subtle pathology signals	*Adequate*: for generic images; weak for fine neuro-features
Reset / Regularization	• Simple add-on to VQ-VAE • Rarely-used codes are alive without softness	• Tuning reset freq & penalty weights needed • Still Euclidean; no locality or covariance awareness	*Moderate*: avoids collapse, but recon detail & task scores plateau
FSQ (Finite Scalar Quantization)	• No learnable codebook • Guaranteed non-collapse	• Fixed grid can't adapt to data manifold	*Low*: coarse, anisotropic errors degrade PET/MRI precision