Novel Deep Learning Techniques for Early Detection of Neurological Disorders

by

Jay Shah

A Dissertation Presented in Partial Fulfillment
of the Requirement for the Degree
Doctor of Philosophy

Approved June 2025 by the
Graduate Supervisory Committee:

Baoxin Li, Co-Chair
Teresa Wu, Co-Chair
Yi Su
Yalin Wang
Yezhou Yang

ARIZONA STATE UNIVERSITY

August 2025

ABSTRACT

Early detection is critical in neurological disorders such as Alzheimer's disease and headaches, offering a window for interventions that may delay progression and improve outcomes. However, traditional diagnostic approaches often fail to provide timely and accurate assessments due to the multifactorial nature of these conditions. Multimodal neuroimaging, particularly magnetic resonance imaging (MRI) and positron emission tomography (PET), provides complementary structural and functional information that can enhance diagnostic accuracy if leveraged effectively. This dissertation presents a three-step deep learning framework that addresses key challenges in medical imaging analysis. First, a robust brain age prediction model is introduced that reformulates age estimation as an ordinal classification task. This mitigates systematic bias in regression-based methods and better captures deviations from normative aging, critical for identifying early signs of neurological disorders. Building on this, the second chapter proposes a novel quantization technique to improve latent representation learning in medical imaging. It leverages Mahalanobis-distance-based soft assignment to multiple codebook vectors, addressing limitations of traditional vector quantization such as codebook collapse and poor reconstruction diversity. When integrated into downstream tasks such as brain age prediction, the proposed quantization method maintains performance in low-resource settings. The final chapter focuses on enhancing early prognosis through a multimodal generative approach. Specifically, an MRI-guided latent diffusion model is developed to super-resolve low-resolution PET scans. By conditioning the diffusion process on high-resolution anatomical MRI, the model improves spatial resolution and quantification of pathological biomarkers such as beta-amyloid. Collectively, these contributions form a clinically motivated, resource-efficient, and multimodal deep learning pipeline for advancing the early detection of neurological disorders.

DEDICATION

Dedicated to everyone who ever took a chance on me.

ACKNOWLEDGEMENTS

This thesis marks the culmination of a five-year journey, and there are many individuals to whom I am deeply grateful for their pivotal roles in shaping this experience.

First and foremost, I would like to express my sincere gratitude to Dr. Teresa Wu for taking a chance on me. I joined her research group amidst the COVID pandemic as a computer science engineer with a budding interest in deep learning for medical research. Her unwavering support and the freedom she provided to explore research questions I was passionate about have been invaluable. Through our many discussions, I learned to refine research questions that address the fundamental "so what?"—a challenge that I believe is among the most difficult yet rewarding aspects of scientific inquiry. I am also profoundly thankful to Dr. Baoxin Li for his steadfast support and mentorship throughout my Ph.D. journey.

I am immensely grateful to my other Ph.D. committee members, Dr. Yi Su, Dr. Yalin Wang, and Dr. Yezhou Yang for their insightful feedback and guidance. Dr. Su's mentorship and our thought-provoking conversations over the years have deepened my understanding of the challenges and effectiveness of deep neural networks in real-world clinical applications. Dr. Wang's expertise in machine learning for medical imaging and Dr. Yang's knowledge on fundamentals of computer vision have been instrumental in strengthening this dissertation.

I am especially appreciative of my research collaborators at the Mayo Clinic, Dr. Todd Schwedt and Dr. Catherine Chong. Their clinical expertise, passion for interdisciplinary research, and exceptional scientific writing have been a constant source of inspiration. I am also particularly grateful to Dr. Gina Dumkrieger and Dr. Simona Nikolova from Mayo Clinic and Dr. Javad Sohankar from Banner Health for their support and guidance in my research projects.

Additionally, I would like to extend my gratitude to Dr. Yonas Geda, Dr. Janina

Krell-Roesch, Jeremy Syrjanen, Dr. Bettina Barisch-Fritz, and Dr. Qi Wang for their invaluable contributions to my exploration of early detection in neurological disorders using non-imaging data. The insights I gained from these collaborations played a crucial role in shaping the direction of my Ph.D. research and my long-term career goals.

A special thank you goes to my colleagues and co-authors at the AMCII lab at ASU, whose collaboration and enriching brainstorming sessions made this journey intellectually stimulating and enjoyable: Dr. Md Mahfuzur Rahman Siddiquee, Dr. Abhidnya Patharkar, Dr. Fulin Cai, Maitry Trivedi, Dr. Jiajing Huang, Dr. Yaoxin Zhuo, Dr. Nathan Gaw, Dr. Fei Gao, Dr. Yiming Che, Dr. Josh Xu, Dr. Firas Al-Hindawi, Teng Li, Fazle Rafsani, Devam Seth, Amogh Joshi, You Zhou, Asra Aghaei, Yutong Wen, Dominic Baker, and Hari Varshan.

Finally, my deepest gratitude goes to my parents, my sister, and my family members for their unwavering support, sacrifices, and belief in me—without which this journey would not have been possible. I am also indebted to my girlfriend, whose constant encouragement has meant the world to me. Her unwavering belief in my potential and her presence through every high and low made this journey possible. A particularly heartfelt thank you goes to my close friends in Arizona, who became my family away from home. The countless road trips, hikes, dinners, and shared laughter provided emotional support and joy throughout this sometimes-stressful Ph.D. journey.

TABLE OF CONTENTS

Page

LIST OF TABLES ix

LIST OF FIGURES xii

CHAPTER

1 INTRODUCTION 1

1.1 Overview of Contributions 4

2 BACKGROUND 9

2.1 Machine Learning and non-imaging AD Biomarkers 10

2.2 Early detection in Alzheimer's Disease 18

2.2.1 Deep Learning and MRI data 19

2.2.2 Deep Learning and PET imaging data 21

2.3 Early detection in Headache 23

3 ROBUST BRAIN AGE PREDICTION 27

3.1 Motivation 27

3.2 Related Works 31

3.2.1 Neuroimaging-based Brain Age prediction 31

3.2.2 Systematic Bias in Predicted Brain Age 32

3.2.3 Regression as Ordinal Classification 34

3.3 Proposed Framework 35

3.3.1 ORDER Loss 36

3.3.2 Evaluation Metrics 38

3.4 Experiment Design 40

3.4.1 Datasets and Preprocessing 40

3.4.2 Implementation Details 45

3.5 Results and Analyses 45

CHAPTER Page

3.5.1 BrainAGE prediction 45
3.5.2 Ordinality and Systematic Bias 46
3.5.3 Alzheimer's Disease detection 48
3.5.4 Headache Phenotypes 49
3.5.5 Ablation studies 53
3.6 Conclusion 55
4 LOCALITY CONSTRAINED VECTOR QUANTIZATION 56
4.1 Motivation 56
4.1.1 Role of distance metric in Quantization 59
4.1.2 Locality in Quantization 60
4.2 Related Works 62
4.2.1 Addressing Codebook collapse in VQ 63
4.2.2 Locality and Soft Quantization 64
4.3 Proposed Framework 66
4.3.1 Locality Constrained Vector Quantization 66
4.4 Experiment Design 69
4.4.1 Datasets 70
4.4.2 Implementation Details 71
4.4.3 Evaluation 73
4.4.4 Downstream Task: Brain Age Prediction 74
4.5 Results and Analyses 75
4.5.1 Discussion 78
4.5.2 Future work 81
4.6 Conclusion 84

CHAPTER Page

5 MRI-GUIDED PET SUPER-RESOLUTION 86
5.1 Motivation 86
5.2 Related Works 87
5.2.1 Partial Volume Correction 87
5.2.2 Image Super-Resolution 88
5.3 Proposed Framework 91
5.3.1 LDM-RR: MRI-guided PET Resolution Recovery 91
5.3.2 Compression Models 92
5.3.3 Diffusion Model 92
5.3.4 Implementation Details 96
5.4 Experiment Design 98
5.4.1 Datasets and Simulation Procedure 98
5.4.2 Image Analysis and Simulation 101
5.4.3 Statistical Analysis 102
5.4.4 Simulated Data Analysis 102
5.4.5 Longitudinal Analysis 103
5.4.6 Cross-tracer Analysis 103
5.5 Results and Analyses 104
5.5.1 Qualitative Assessments 104
5.5.2 Evaluation on Simulated Data 104
5.5.3 Evaluation on Real Longitudinal Amyloid PET Data 105
5.5.4 Evaluation on Real Cross-tracer Amyloid PET Data 106
5.5.5 Discussion 107
5.6 Conclusion 113

CHAPTER Page

6 FUTURE WORK & CONCLUSION 114
6.1 Future Work 114
6.2 Conclusion 116
REFERENCES 118
APPENDIX
A STATEMENT OF CO-AUTHOR PERMISSIONS 138
B LOSS FUNCTION FOR LDM-RR: PET RESOLUTION RECOVERY FRAMEWORK 140
C FIGURES for RIED-NET 142
D FIGURES for PHYSICAL ACTIVITY DATA 145
E RESULTS for LCVQ 148

LIST OF TABLES

Table Page

2.1 Summary of Literature on Machine Learning Methods for Brain Aging Analysis, Spanning Neuropsychiatric Symptoms (NPS), MRI and PET Biomarkers, and Multi-modal Combinations Thereof. 11

3.1 Age Range with Distribution and Number of Samples for Each Cohort. The Lifespan Cohort Comprises NACC, OASIS, ICBM, IXI, and ABIDE, Whereas the Discovery Cohort Consists of Samples from the ADNI Cohort. 41

3.2 Summary of Age Range with Distribution and Number of Samples for Each Cohort. 44

3.3 Brain Age Prediction Results on Lifespan Cohort. Here, MAE measures the difference between predicted and actual chronological age on the same test set. Here, **bold** numbers represent the best results, while underlined numbers represent second-best results. 47

3.4 Performance Evaluation of All Methods in Preserving Ordinality and Addressing Systematic Bias in Brain Age Prediction Using Metrics Defined In Sec.3.3.2 and Sec.3.3.2. Here, systematic bias is measured one standard deviation from mean: systematic bias-left ($\mu - \sigma$) and right ($\mu + \sigma$) [SB-L, SB-R]. 48

3.5 Average Brainage Values Across the Five Clinical Groups of AD. The last column is the Pearson correlation between average BrainAGE values and disease severity of clinical groups in increasing order from HC to AD (1-5). MV: Mean-Variance. 50

3.6 Ablation Studies on the Proposed Framework Components Evaluated by MAE, Ordinality, and Systematic Bias Scores. Here, k denotes different L_k-norm distance metrics defined in Eq.3.7 54

4.1 Age Range with Distribution and Number of Samples for Each Cohort. The Lifespan Cohort Comprises NACC, OASIS, ICBM, IXI, and ABIDE. Whereas the Discovery Cohort Consists of Samples from the ADNI Cohort. 72

4.2 Comparison Between Various Quantization Approaches on Image Reconstruction Task. All methods use the same base architecture and codebooks $(1024, 32)$ were initialized with K-means across all implementations. AR: affine reparameterization, LCVQ: locality constrained vector quantization. 76

4.3 Comparison of Brain Age Prediction Performance on a Test of the Lifespan Cohort, with and Without Quantization. Here, training time measures the time taken solely to train the brain age prediction model. 78

4.4 Comparison of Brain Age Prediction Performance on an Independent Discovery Cohort (ADNI), with and Without Quantization. ResNet-18 models were trained to predict brain age using MSE loss on raw imaging and quantized data. Brain Age delta measures the (predicted age - actual age), and MAE measures its absolute value on AD, MCI, and healthy control subjects. AD:alzheimer's disease, MCI:mild cognitive impairment, HC:healthy controls 79

Table Page

4.5 Ablation Study of the Top-k Parameter in the Lcvq Framework along with an Adaptive k Selection Method. All methods use the same base architecture and codebooks $(1024, 32)$ were initialized with K-means across all implementations. 82

5.1 Summary of Demographic Information of the Three Cohorts Included in This Study. 100

5.2 Statistical Power in Detecting Longitudinal Changes Measured by Mean, Standard Deviation, and P-value of an Annualized Rate of Amyloid Accumulation and Sample Size (SS) per Arm Estimates to Detect a 25% Reduction in Amyloid Accumulation Rate Due to Treatment (80% Power and a Two-tailed Type-I Error of P=0.05) 107

5.3 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. 108

E.1 Comparison of Baseline Vqvae, Affine Reparameterization (AR), and Proposed LCVQ Quantization Methods on CIFAR-10 and MNIST Datasets, Evaluated by Mean Squared Error (MSE) and Codebook Utilization Using Perplexity (PER%). 149

LIST OF FIGURES

Figure Page

2.1 Building Three Different Sets of ML Models Using a Stepwise Feature Addition Approach. (1) Including Demographic Features, Then Adding Plasma-derived Biomarkers, Then Adding NPS; (2) Including Demographic Features, Then Adding Neuroimaging-derived Biomarkers, Then Adding NPS; And (3) Including Demographic Features, Then Adding CSF-derived Biomarkers, Then Adding NPS. 16

3.1 Standard Cross-entropy Vs. Cross-entropy with ORDER Loss: Here, cross-entropy loss (left) encourages the model to learn high-entropy feature representations where embeddings are spread out. However, it fails to capture ordinal information from labels. Our proposed ORDER loss with cross entropy (right, Eq.3.5) preserves ordinality by spreading the features proportional to the Manhattan distance between normalized features weighted by absolute age difference. The illustrated example (right) shows an embedding space where learned representations of MRI scans with ages 20, 40, and 80 are distributed apart from one another, with distances proportional to absolute age differences. . . . 30

Figure Page

3.2 Overview of Proposed Brain Age Prediction Framework. (a) A 3D ResNet-18 model is trained using the lifespan cohort with cross-entropy and ORDER as loss functions. Age is calculated as the weighted average of class probabilities from the softmax classifier. (b) During inference, the Brain Age Gap Estimate (BrainAGE) is calculated as the difference between predicted biological age and actual chronological age. (c) The trajectory plot offers a visual interpretation of predicted BrainAGE and its associations with aging patterns. The preclinical AD stage is when the patient behaves cognitively normal, but underlying changes in the brain due to accelerated aging happening at a subtle rate can be captured using BrainAGE. 36

3.3 A T-SNE Visualization of Embeddings from Models' Penultimate Layer: (a) When using MSE loss, embeddings maintain ordinal relationships but are tightly packed, resulting in a low-entropy feature space (b) MSE with Euclidean distance loss spreads out embeddings but struggles to preserve ordinal relationships accurately (c) Cross-entropy (CE) further spreads embeddings, creating a high-entropy space, but at the cost of losing ordinal information (d) Mean-variance loss combined with cross-entropy creates a high-entropy feature space and slightly improves ordinality (Tab.3.4). (e) ORDER loss combined with cross-entropy achieves the best balance: it accurately preserves ordinality, maintains a high-entropy space, and improves overall performance. Embeddings are color-coded based on their ground truth age values $[10 - 95]$. 46

Figure Page

3.4 Heatmap of Statistical Significances Between the Five Clinical Groups of AD Calculated as p Values from a T-test on Predicted BrainAGE from Respective Groups, for MSE and Cross-entropy with ORDER Loss Models. 50

3.5 Brain Age Delta (Predicted–chronological Age) Across Different Headache Phenotypes on the Mayo Clinic Dataset 51

3.6 Visualization of Headache Phenotype Clustering in a 3D LDA-reduced Feature Space. The plot demonstrates three discernible clusters, effectively separating Migraine, persistent PTH, acute PTH, and healthy control groups based on model-extracted neuroimaging features. Each dot represents an individual subject. 52

4.1 Visual Comparison of Original T1w MRI Scans and Reconstructed MRI Scans after Quantization Using the Proposed LCVQ Approach. . . 76

5.1 Training Three Modality-specific Autoencoder Models to Compress High-dimensional Simulated DP, Simulated FBP, and MRI Data into a Lower-dimensional Latent Representation. 93

5.2 Proposed LDM-RR Framework's Training Process for PET Super-resolution. LDM is conditioned on latent representations of T1-MRI and simFBP and uses a combination of image and noise scale losses to generate corresponding high-resolution simDP. 97

5.3 Visualization of a Simulated Digital Phantom (SimDP) and Simulated FBP (SimFBP) from the Data Simulation Pipeline Using T1-MRI and the LDM-RR Generated Synthetic Super-resolution FBP. 102

Figure Page

5.4 Visual Comparison of Generated FBP Scans Using RL-RR and Our LDM-RR to Real FBP and T1-MRI for a Sample from the OASIS-3 Cohort. 105

5.5 Comparison of Mean Recovery Coefficient (RC) Using Different Methods on a Held-out Test of 338 Samples Randomly Selected from the Simulated Dataset. A Value Closer to 1 Indicates High Performance. . . 106

5.6 Comparisons of the Results from LDM Trained Using Different Loss Functions. 109

C.1 Overall Architecture of RIED-Net. It uses a U-Net-like structure with a residual inception shortcut to enhance training efficiency. It includes 5 encoding and 4 decoding blocks. Input is a 2D florbetapir image slice (256×256×1). Each block processes data via 3×3 or 1×1 convolutions (ReLU), with skip connections between corresponding encoder-decoder layers. The final layer outputs a synthetic 2D PiB slice using a 1×1 convolution. Convolutional, deconvolutional, and copy operations are represented with distinct arrows. 143

C.2 Visual Comparison of Synthetic Images Generated Using RIED-net to Real Pittsburgh Compound-b (PIB) Data for the OASIS (Panels a and B) and GAAIN (Centiloid Project; Panels C and D) Datasets Used in Our Study. Panels A and C Show Representative Images from Amyloid Negative Participants, and Panels B & D Are Examples from Amyloid Positive Participants. Abbreviations: FBP, Florbetapir; SynPIB, Synthetic PIB. 144

Figure Page

D.1 Classification of Decline Vs. Non-decline in the Cognitive Outcome Variables, Presented by SHAP Plots for CG and IG. 146
D.2 (Same as the above Plot) Classification of Decline Vs. Non-decline in the Cognitive Outcome Variables, Presented by SHAP Plots for CG and IG. 147

Chapter 1

INTRODUCTION

Neurological disorders pose a massive burden on the global healthcare industry. Over 3 billion people worldwide were living with a neurological condition in 2021, making these disorders the leading cause of ill health and disability globally (Organization *et al.*, 2024). Traditional clinical frameworks often identify pathology only after irreversible damage has occurred. This delayed detection limits the efficacy of therapeutic interventions, as it is too late by the time symptoms manifest; ergo, early detection is key to the ultimate success in combating neurodegenerative disorders and for effective intervention.

But, early detection of neurological disorders is challenging, especially for Alzheimer's disease (AD). Modern deep learning techniques applied to complex neuroimaging data have helped capture new biomarker signatures for early detection previously unknown with acceptable accuracy (Jónsson *et al.*, 2019; Bethlehem *et al.*, 2022), primarily using algorithms that can analyze high-dimensional imaging datasets and uncover patterns imperceptible to medical experts. Age being the most significant risk factor, deep learning based **brain-age prediction** using magnetic resonance imaging (MRI) data can capture the multivariate pattern of age-related changes in the brain associated with healthy aging (Cole *et al.*, 2018; Beheshti *et al.*, 2020; Baecker *et al.*, 2021a). Although these models have achieved mean absolute errors (MAE) as low as 3.39 years (Jónsson *et al.*, 2019), they are known to overestimate younger subjects and underestimate older ones due to regression-to-the-mean effects, observed across methods like ridge regression, support vector machines, and deep learning models (Liang *et al.*, 2019). This bias persists even after post-hoc corrections, as age-adjusted resid-

uals remain correlated with chronological age (Zhang *et al.*, 2023a). The predicted outcomes from a machine learning model for continuous outcomes are often systematically biased (Belitz and Stackelberg, 2021), which is problematic for the applications of regression models, leading to inaccuracies in any downstream analysis.

Although brain age prediction using MRI has shown strong performance, it remains constrained by the high dimensionality and storage demands of 3D neuroimaging. MRI volumes consist of millions of voxels per scan, making them computationally intensive and inefficient for large-scale deployment or privacy-sensitive scenarios like federated learning. Transmitting full-resolution images is impractical; instead, compact, informative representations are needed to reduce bandwidth and enable secure collaboration. Recent advances in deep generative modeling have demonstrated the power of learning priors over high-dimensional data (Goodfellow *et al.*, 2014; Ho *et al.*, 2020), with conditional generation increasingly applied to healthcare (Xia *et al.*, 2021; Tudosiu *et al.*, 2024). Among these, vector quantized variational autoencoders (VQ-VAEs) offer a compelling way to encode images into discrete latent tokens from a learned codebook (Van Den Oord *et al.*, 2017). These discrete representations support both compression and downstream tasks such as classification, regression, segmentation and generation. Similar discretization mechanisms have been central to large language models, where tokenization and codebook design directly affect model capacity and performance (Sun *et al.*, 2024; Gao *et al.*, 2024). For example, Llama 3's expanded vocabulary enabled more efficient encoding and improved generative performance, underscoring the importance of codebook quality[1] .

A major limitation of VQ-VAEs, however, is *codebook collapse*, where only a small subset of codewords are utilized during training, limiting representational diversity (Huh *et al.*, 2023). In neuroimaging, this can obscure subtle structural variations

[1]https://ai.meta.com/blog/meta-llama-3/

critical for diagnosis. Tudosiu *et al.* (2024) observed this in FLAIR-based brain MRI models, which exhibited lower codebook perplexity and less anatomical detail in reconstructions compared to T1-weighted counterparts. To address this, we propose a **Mahalanobis distance-based quantization** approach that replaces the standard nearest-neighbor assignment with a k-nearest neighbor strategy using a covariance-aware distance metric. This allows each latent vector to activate multiple statistically relevant codewords, improving codebook utilization, reducing quantization error, and preserving fine-grained anatomical structure. Our experiments demonstrate that this approach retains actual brain age prediction accuracy while offering advantages for privacy-preserving and resource-efficient learning, making it suitable for clinical and federated settings.

While structural changes captured by MRI are important, they may not be sufficient for detecting preclinical AD pathology. Molecular biomarkers such as amyloid-β—detectable via positron emission tomography (PET)—are critical for identifying the biological onset of AD years before symptoms emerge (Chapleau *et al.*, 2022). However, PET imaging faces a resolution-accuracy trade-off: clinically feasible low-resolution PET scans obscure subtle amyloid accumulations, while higher resolution comes at the cost of increased scan time and radiation. Although deep learning-based PET super-resolution approaches have been proposed (Song *et al.*, 2020b; Yoshimura *et al.*, 2022), they often fail to accurately resolve amyloid-specific distributions and are prone to smoothing artifacts. To address this, the final part of this thesis explores a **multimodal diffusion-based framework** for PET super-resolution, where high-resolution anatomical guidance from MRI is used to enhance low-resolution PET. By leveraging the complementary strengths of MRI (structural detail) and PET (molecular specificity), this approach enables more accurate reconstruction of amyloid distributions and enhances quantification reliability. We build on recent advances in latent

diffusion models and demonstrate that incorporating MRI guidance during both encoding and generation yields super-resolved PET images that are not only sharper but also more clinically informative, thus supporting more precise early detection of AD.

Motivated by these aforementioned limitations of existing deep learning techniques applied to real-world medical imaging data, this thesis introduces three innovative deep learning methodologies that address the following questions, broadly applicable to deep learning research and not solely limited to computational medical imaging analysis:

1. How to address systematic bias of regression-to-mean (RTM) in image-based regression models when using mean squared error (MSE) loss?

2. How to address the well-known problem of codebook collapse in vector quantization which is critical for capturing subtle features or anatomical distortions directly impacting diagnostic decisions in medical image reconstruction?

3. How can diffusion models better leverage image-guidance for medical image super-resolution while preserving the accuracy of anatomical variations?

1.1 Overview of Contributions

Chapter 3 presents a novel framework for robust brain age prediction by reformulating the task as an ordinal classification problem, augmented by a novel distance regularization technique. In this approach, continuous brain age values are discretized into ordered categories, thereby capitalizing on the inherent progression of aging. We

enforce the model to recognize that misclassifications between adjacent age groups are less severe than those spanning distant categories. Central to this framework is the integration of distance regularization via *ORdinal Distance Encoded Regularization* (ORDER) loss, which imposes a penalty that is proportional to the ordinal distance between the predicted and actual age groups. This ensures that the network learns a smooth transition between adjacent age categories, effectively capturing subtle, age-related structural changes in the brain. The deep convolutional architecture employed here extracts high-level features from T1-weighted brain MRI data, thereby enhancing the sensitivity to anatomical nuances indicative of aging. During the training phase, the model optimizes a composite loss function that combines the traditional cross-entropy loss for ordinal classification with the bespoke distance regularization term (Shah *et al.*, 2024c). This dual-loss formulation not only guides the network to correctly assign images to their respective age bins but also minimizes the error magnitude when predictions deviate from the true age, resulting in nuanced and robust brain age estimation. At inference, similar to Pan *et al.* (2018), the probability distribution across the ordered categories is leveraged to derive a continuous age estimate, blending the results of an ordinal classification model with the precision of regression.

Extensive experiments on both public and institutional datasets demonstrate that this framework significantly outperforms standard regression and conventional classification approaches. Notably, the incorporation of distance regularization contributes to improved stability and reliability, particularly in scenarios with imbalanced age distributions or limited sample sizes. The empirical results underscore the potential of this approach for clinical applications, where accurate brain age prediction can serve as an early indicator of neurodegenerative conditions (Shah *et al.*, 2022b, 2023b, 2024c) and other potentially age-associated brain disorders (Shah *et al.*, 2024b).

Chapter 4 presents a novel framework for quantization-aware compression of

high-dimensional brain MRI using a Mahalanobis distance-based vector quantization strategy. This chapter introduces *Locality-Constrained Vector Quantization* (LCVQ), a replacement for the conventional quantization step in VQ-VAE, which incorporates statistical structure into the codeword selection process. We took inspiration from findings of much earlier quantization research (Yang *et al.*, 2009; Wang *et al.*, 2010), where authors showed that by relaxing the quantization step to aggregate more than one codebook vector via locality or sparsity constraints can lead to better feature representation learning. Instead of relying on a single nearest neighbor in Euclidean space, our method computes Mahalanobis distances, taking into account the global covariance of the codebook, and selects the top-K most relevant vectors for each latent. The selected neighbors are then averaged to yield a smooth, locality-preserving quantized representation. This formulation not only mitigates codebook collapse by promoting broader codebook usage but also preserves the geometric continuity of the latent space.

Integrated into a 3D VQ-VAE pipeline, the proposed LCVQ framework retains anatomical reconstruction fidelity and downstream predictive performance in brain age estimation, all while achieving compression rates conducive to real-world deployment. Furthermore, we argue that the compact-discrete latent codes produced by this approach offer practical advantages for federated and privacy-preserving learning frameworks, where data efficiency is paramount. Experiments on large-scale neuroimaging datasets demonstrate that LCVQ outperforms state-of-the-art VQ-based baselines—including morphology-preserving (Tudosiu *et al.*, 2024) and affine-regularized variants (Huh *et al.*, 2023) in terms of codebook perplexity, perceptual similarity, and predictive accuracy. To our knowledge, this is the first use of Mahalanobis-informed quantization in medical image compression, highlighting its potential to enhance both interpretability and utility in clinical AI pipelines.

Chapter 5 introduces an innovative framework that leverages multimodal imaging and synthetic data to render high-resolution PET, given that the ground-truth for those images doesn't exist in reality. More specifically, it is aimed at enhancing the quantification of amyloid PET scans through an MRI-guided super-resolution approach by leveraging latent diffusion models (Shah *et al.*, 2024a). It harnesses the complementary strengths of high-resolution MRI and the generative prowess of latent diffusion models to reconstruct PET images with higher resolution. By conditioning the latent diffusion process on detailed anatomical information derived from MRI, the framework is designed to generate super-resolved PET images that not only exhibit enhanced spatial resolution but also retain critical biochemical information necessary for accurate amyloid quantification. At the core of this approach lies a dual-path architecture: one branch processes the lower-resolution PET input, while the other extracts rich anatomical features from corresponding MRI scans. The latent diffusion model then iteratively refines the PET image by integrating these features, effectively learning a mapping that respects both the inherent distribution of PET signal intensities and the structural guidance provided by MRI. This fusion of modalities enables a more faithful reconstruction of fine-grained details, thereby reducing partial volume effects and improving the reliability of subsequent amyloid burden assessments.

Model training was done using a combined loss objective that balances reconstruction fidelity with anatomical consistency, ensuring that the generated high-resolution PET images are both quantitatively accurate and visually coherent. Empirical evaluations on different datasets demonstrate that this MRI-guided super-resolution strategy significantly outperforms conventional interpolation methods (Tohka and Reilhac, 2008) and deep learning-based super-resolution techniques (Ho *et al.*, 2020; Pinaya *et al.*, 2022a), providing improvements in both visual quality and quantitative metrics. We believe such enhancements are critical for the early diagnosis and tracking

the progression of neurodegenerative diseases like Alzheimer's, where precise amyloid quantification plays a pivotal role. Moreover, the modularity of this framework allows for future integration with other imaging modalities and clinical parameters, a potential way for broader applications in multimodal neuroimaging diagnostics and data synthesis.

In summary, this dissertation contributes to the field of deep learning-based early detection of brain disorders using diverse medical imaging data. The proposed *ORDER* loss in an ordinal classification can better structural proxies of aging from MRI data, aiding early detection, whereas the *LDM-RR* framework allows better quantification of amyloid deposits in the brain using PET data. At the same time, *LCVQ* addresses the common challenge of codebook collapse in VQ-based image reconstruction models. Through comprehensive evaluations on real-world datasets, we demonstrate these methods' applicability and effectiveness.

Chapter 2

BACKGROUND

Understanding the progression of Alzheimer's disease requires synthesizing insights from a wide range of clinical, behavioral, and biological signals. While imaging-based biomarkers like MRI and PET have dominated much of the machine learning (ML) research in AD, there is growing interest in leveraging non-imaging features, including neuropsychiatric symptoms (NPS) and fluid biomarkers for early risk prediction. These modalities not only complement structural and functional imaging but may also capture upstream changes that precede overt cognitive decline. This chapter reviews the landscape of ML applications in AD research, with a focus on integrating underutilized non-imaging biomarkers into predictive frameworks.

Deep learning breakthroughs in neuroimaging now dominate AD research. Yet front-line clinical practice still leans on low-cost, non-invasive measures: neuropsychiatric-symptom inventories, activities-of-daily-living scales, routine blood tests that can be collected at scale and frequently register pathophysiologic change earlier than MRI or amyloid-PET (Geda *et al.*, 2014; Gill *et al.*, 2020). Meta-analytic evidence shows that ML models built on such inexpensive signals already boost diagnostic sensitivity and specificity (Chang *et al.*, 2021). Beginning the review with these traditional ML approaches, therefore grounds the thesis in real-world clinical workflows and highlights how behavioural or fluid markers complement the computationally intensive imaging pipelines developed later. Recent multimodal studies that weave neuropsychiatric data, fluid biomarkers, and neuroimaging into a single predictive framework (Qiu *et al.*, 2020) illustrate this synergy and directly motivate the deep-learning methods introduced in subsequent chapters.

2.1 Machine Learning and non-imaging AD Biomarkers

Reviews and original studies have examined ML in the context of AD biomarkers derived from neuroimaging, cerebrospinal fluid (CSF), or plasma. For example, a meta-analysis reported that biomarker-based ML techniques might increase the sensitivity and specificity of the AD diagnosis (Chang *et al.*, 2021), and NPS are known to be associated with AD biomarkers in brain aging (Gill *et al.*, 2020). However, to date, less is known as to whether ML approaches are also useful in investigating the associations between NPS and AD biomarkers in brain aging. Since this thesis aims to develop AI-driven methods for early detection of neurological disorders, including AD, we also did a literature review with an emphasis on the potential of incorporating NPS, alongside traditional imaging biomarkers, to achieve this goal (Shah *et al.*, 2023c). Tab.2.1 provides a comprehensive overview of all reviewed ML studies (summarized in 2023) categorized by their focus on NPS, AD biomarkers (MRI and PET), and cognitive outcomes. The table highlights the diversity of data modalities and analytical methods in AD research, with a particular emphasis on how multimodal integration and advanced ML techniques (e.g., deep learning, ensemble models) enhance diagnostic accuracy, prognostic value, and our understanding of the relationships between NPS, other AD biomarkers, and cognition. This literature review was aimed at facilitating comparison and identifying gaps, particularly the underrepresentation of studies jointly analyzing NPS and AD biomarkers.

Further, the brain age prediction framework developed in the thesis (Chapter 3) could potentially benefit from the inclusion of NPS data. Since accelerated brain aging, as captured by brain age prediction models, is also associated with cognitive decline and AD risk, incorporating NPS as additional features in the brain age prediction framework might enhance its accuracy and provide a more holistic assess-

ment of an individual's risk profile. For instance, individuals exhibiting a brain age significantly older than their chronological age and reporting specific NPS could be identified as being at a particularly high risk of developing AD.

Table 2.1: Summary of Literature on Machine Learning Methods for Brain Aging Analysis, Spanning Neuropsychiatric Symptoms (NPS), MRI and PET Biomarkers, and Multi-modal Combinations Thereof.

Study	**Data**	**ML Model**	**Key findings**
Gill *et al.* (2020)	NPS, AD biomarkers	Logistic Model Tree	Predicted future cognitive status using baseline clinical, neuropsychiatric, and MRI data. MBI scores showed higher diagnostic value than clinical or volumetric features. ML model identified 2–7 optimal classification features.
Mallo *et al.* (2020)	NPS	Random Forest	Predicted MCI-to-dementia conversion using socio-demographics, health status, and NPS proxies with 67% F1-score, 88% accuracy. Random forest outperformed 8 other ML models. Top predictors: NPI-Q severity, stress, and GDS-15 scores.
Mar *et al.* (2020)	NPS	Random Forest	ML models detected psychotic and depressive symptoms in dementia. The psychotic cluster model had higher discriminatory power.
Palermo *et al.* (2021)	NPS	LSTM	Sensor-based data used to detect agitation in dementia. LSTM models captured temporal dependencies and predicted agitation with 75% accuracy.
Kang *et al.* (2019)	NPS	3-layer ANN, Logistic Regression	Neural net outperformed logistic regression in 2-class and 3-class tasks. Achieved 97% accuracy predicting NC, MCI, ADD using neuropsychological data.

Continued on next page

Table 2.1 – Continued from previous page

Study	**Data**	**ML Model**	**Key findings**
Klöppel *et al.* (2018)	MRI	SVM	SVM model distinguished depression vs early AD in challenging cases. Not useful for diagnosis, but the authors recommend it for individual-level analysis.
Salvatore *et al.* (2015)	MRI	PCA, FDR, SVM	ML model found similar atrophy in AD and MCI converters, suggesting early MRI-detectable structural biomarkers.
Sørensen *et al.* (2018)	ADNI	Ensemble SVM	Ensemble SVMs were more robust and accurate than single classifiers, useful for multi-class AD classification.
Cao *et al.* (2017a)	MRI	Multiple kernel learning, k-NN, manifold learning	Manifold learning improved oversampling and dimensionality reduction. k-NN outperformed SVM. Kernel learning better represents nonlinear structures.
Pang *et al.* (2019)	MRI	Semi-supervised Deep Autoencoder	Deep learning enabled effective hippocampus segmentation without registration—useful for AD diagnosis.
Chen *et al.* (2021)	MRI	CNN	CNNs outperformed SVM/FCNN in segmentation and speed. AD indicators like reduced gray matter and cortex thickness were better captured.
Qiao *et al.* (2021)	MRI	CNN	Contrastive loss on CNN grouped similar subjects and improved MMSE ranking. Helped distinguish AD, NC, and MCI. Used CNN with MRI input.
Ambastha *et al.* (2017)	MRI	Ensemble CNNs, AdaBoost	Ensemble CNN with dual-region MRI inputs improved AD prediction. Captured degenerative patterns tied to behavior.

Continued on next page

Table 2.1 – Continued from previous page

Study	**Data**	**ML Model**	**Key findings**
Bhagwat *et al.* (2019)	MRI	Anatomically partitioned ANN	ANN predicted cognitive scores at baseline and 1 year using structural MRI data, including hippocampal segmentations and cortical parcellations.
Qiao *et al.* (2022)	MRI	3D CNN	Converted MMSE regression into multi-class classification using ordinal bins. Better predicted MMSE over time, highlighting subtle cognitive changes.
Cui *et al.* (2019)	MRI	CNN, RNN	Combined CNNs (for spatial) and RNNs (for longitudinal) features. Outperformed baseline models and handled missing data. Focused on longitudinal AD analysis.
Son *et al.* (2020)	PET	2D & 3D CNNs	Deep models matched or outperformed expert PET readers, especially for ambiguous cases. 2D CNNs performed better than 3D CNNs.
Choi *et al.* (2020)	PET	3D CNN	3D CNN captured cognitive patterns across AD, MCI, and PD. Transfer learning aided performance on small datasets.
Whittington *et al.* (2021)	PET	Linear regression	Voxel-wise regression enabled a new tau quantification method accounting for complex deposition patterns beyond amyloid.
Qiu *et al.* (2020)	multiple biomarkers	FCN + MLP	Adding age, gender, and MMSE improved performance over MRI alone (acc: 0.968 vs 0.834).
Donnelly-Kehoe *et al.* (2018)	multiple biomarkers	Decision trees (RF, SVM, AdaBoost)	Random forest performed best. Morphological features alone are insufficient for good accuracy—MMSE needed for cognitive profiling.

Continued on next page

Table 2.1 – Continued from previous page

Study	**Data**	**ML Model**	**Key findings**
Liu *et al.* (2020)	multiple biomarkers	Group Guided Fused Laplacian Sparse Group Lasso	Multimodal, multitask framework using graph Laplacian showed significance in longitudinal modeling.
Bhagwat *et al.* (2018)	multiple biomarkers	Siamese Neural Network	Follow-up info improved multimodal performance. The two-timepoint input offered the best performance. Helped identify stable/declining patterns.
Beltran *et al.* (2020)	multiple biomarkers	Random forests, gradient boosting	Cost-effective screening using ML models vs expensive biomarkers. Focused on trajectories without hard thresholds.
Delmotte *et al.* (2021)	multiple biomarkers	Linear mixed effects model	Explored MMSE trajectories across ATN classes. A–/T–/N+ class showed distinct MMSE decline over 3 years.
Frölich *et al.* (2017)	multiple biomarkers	SVM	SVM with linear kernel and 9 predictors. Hippocampal volume and tau had the best predictive power. Multi-modality helped at fixed sensitivity.

Abbreviations: AD, Alzheimer's disease; ADD, Alzheimer's disease dementia; ADNI, Alzheimer's Disease Neuroimaging Initiative; ANN, artificial neural network; ATN, amyloid/tau/neurodegeneration; CNN, convolutional neural network; CU, cognitively unimpaired; EMCI, early MCI; FCN, fully convolutional network; FDR, Fisher Discriminant Ratio; GDS-15, Geriatric Depression Scale-15 items; HC, healthy controls; k-NN, k-nearest neighbor; LMCI, late MCI; LSTM, Long Short-Term Memory network; MBI, mild behavioral impairment; MCI, mild cognitive impairment; MMSE, Mini-Mental Status Exam; MRI, magnetic resonance imaging; NC, normal cognition;

NPI-Q, Neuropsychiatric Inventory-Questionnaire; NPS, neuropsychiatric symptoms; PET, positron emission tomography; PCA, principal component analysis; RNN, recurrent neural network; SVM, support vector machine.

We found that while a substantial number of ML studies have focused on utilizing traditional AD biomarkers, such as neuroimaging data (Franke *et al.*, 2010; Franke and Gaser, 2012), cognitive scores, and various omics data (Trivedi *et al.*, 2023), to construct predictive diagnostic models for AD, there is a noticeably limited number of studies that have specifically investigated NPS, either in isolation or in combination with these established biomarkers. This disparity underscores a potential gap in the current research landscape, suggesting that the role of NPS in ML-driven AD research might be underexplored (Shah *et al.*, 2023c). One key takeaway from our analysis is the potential predictive power of NPS, especially when considered in conjunction with established AD biomarkers, for identifying individuals who are at an elevated risk of developing mild cognitive impairment (MCI) or dementia (Gill *et al.*, 2020; Shah *et al.*, 2023c). Since the evidence suggests that NPS can manifest even in the early stages of the disease and may even precede cognitive decline (Geda *et al.*, 2014). The combination of these clinical symptoms with biological markers offers a more comprehensive picture of an individual's risk profile.

We further tested this hypothesis and found that, indeed, incorporating NPS into machine learning models, with a simple step-wise feature addition approach as shown in Fig.2.1, generally led to improved performance in predicting cognitive decline (Shah *et al.*, 2025). Specifically, the ML models demonstrated better performance when NPS were included alongside neuroimaging biomarkers for predicting decline in global cognition, as well as in the cognitive domains of language and visuospatial skills, suggesting that NPS provide valuable additional insights when combined with structural or functional brain imaging data in the context of predicting cognitive

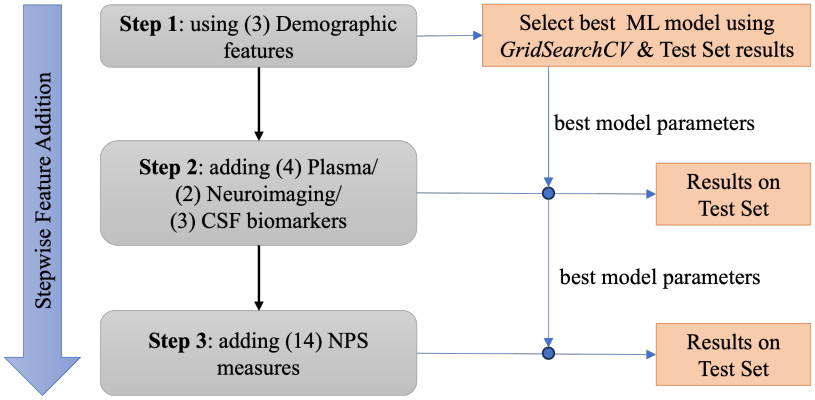

Figure 2.1: Building Three Different Sets of ML Models Using a Stepwise Feature Addition Approach. (1) Including Demographic Features, Then Adding Plasma-derived Biomarkers, Then Adding NPS; (2) Including Demographic Features, Then Adding Neuroimaging-derived Biomarkers, Then Adding NPS; And (3) Including Demographic Features, Then Adding CSF-derived Biomarkers, Then Adding NPS.

decline. Furthermore, we found that the inclusion of NPS also improved the prediction of decline in visuospatial skills when using plasma-derived biomarkers, and the models showed enhanced performance in predicting decline in attention/executive function, language, and memory when NPS were added to models that included CSF-derived AD biomarkers. We do note that the effect sizes of this improvement were generally small, indicating that while NPS contribute valuable information, they might not be the sole determinant of cognitive decline, and their impact might be subtle.

Extending this multimodal approach to modifiable lifestyle factors, our recent work demonstrates how machine learning can identify intervention-specific predictors of cognitive trajectory in dementia patients undergoing physical activity (PA) programs (Barisch-Fritz *et al.*, 2025). SVM models achieved moderate classification accuracy (40.6-75.6%) for cognitive decline using PA adherence metrics and functional performance measures like activities of daily living (ADL). ADL performance

emerged as a stronger predictor than traditional PA metrics, suggesting functional capacity captures neurodegeneration aspects complementary to biological markers (see Appendix D). Whereas, adherence patterns in the intervention group correlated with cognitive outcomes, implying ML could detect early signs of intervention resistance. To better understand individual feature contributions to model predictions, we used SHapley Additive exPlanations (SHAP) plots. These plots (Fig.D.1 and Fig.D.2) revealed that ADL-related variables consistently ranked among the most influential features in both intervention and control groups, underscoring their predictive utility. Biological markers (e.g., brain age delta), symptom profiles (NPS), and modifiable factors (PA/adherence) could create composite risk signatures, and future implementations of our brain age prediction system could incorporate such lifestyle/functional metrics.

In addition to the use of ML for traditional imaging and fluid biomarkers, recent advances have demonstrated the power of interpretable deep learning frameworks in elucidating the molecular changes underlying AD using bulk RNA-sequencing data. We developed and applied an interpretable deep learning approach to postmortem brain transcriptomic data from multiple brain regions in large clinical cohorts (Trivedi *et al.*, 2023). This framework not only achieved high accuracy in classifying AD versus control samples across internal and external datasets but also enabled the identification of key gene modules implicated in disease pathogenesis. Notably, our analyses revealed common gene signatures associated with microglial activation across brain regions, as well as sex-specific transcriptional modules in neurons, providing new insights into the molecular basis of sex differences in AD vulnerability and progression. These findings underscore the potential of explainable AI approaches to bridge the gap between high-dimensional molecular data and clinically meaningful biomarkers and to advance our understanding of the complex, multi-factorial etiology of AD.

Overall, these studies show how machine learning can bring together different types of data, like brain scans, fluid biomarkers, symptom profiles, lifestyle factors, and even gene expression, to better understand and predict AD. Each type of data offers a different piece of information, and combining them leads to more accurate models and deeper biological insights.

2.2 Early detection in Alzheimer's Disease

AD is a devastating neurodegenerative disorder that is currently affecting over 6.9 million people in the United States[1] and 55 million people worldwide, and its associated economic burden is anticipated to increase exponentially in the coming decades. AD is characterized by an extended preclinical phase that can span decades before the onset of clinical symptoms and is neuropathologically defined by the presence of neuritic amyloid plaques and tau neurofibrillary tangles (Long and Holtzman, 2019). While genetic predispositions, such as carrying the APOE4 allele, contribute to disease risk, *aging* remains the most significant risk factor (Niccoli and Partridge, 2012). Consequently, gaining deeper insights into the aging process and its interaction with AD-specific pathophysiology is essential for advancing effective treatment and prevention strategies.

However, biological aging is a highly complex process that lacks a universally accepted definition (Niccoli and Partridge, 2012). As a result, chronological age is frequently used as an imperfect proxy, as it reflects elapsed time rather than the underlying biological mechanisms (Horvath and Raj, 2018). In recent years, the availability of large-scale datasets and advances in deep learning (LeCun *et al.*, 2015; Cole *et al.*, 2015) have led to the development of various biological "clocks" designed to characterize aging at different levels using macroscopic structural imaging data

[1] 2024 Alzheimer's disease facts and figures; https://doi.org/10.1002/alz.13809

(Cole *et al.*, 2017) notably with a mean absolute error of approximately 4 years. Continued advancements in computational methods are expected to further enhance model accuracy (Jónsson *et al.*, 2019; Shah *et al.*, 2024c). However, existing biological clocks face limitations, including biases arising from constrained training datasets, the misalignment of linear modeling approaches with the inherently nonlinear nature of biological aging, survival biases, and regression modeling. These factors frequently result in systematic errors, leading to age overestimation in younger individuals and underestimation in older populations (Liang *et al.*, 2019).

Recently, multi-modal deep learning approaches integrating MRI and PET imaging have been proposed for early AD detection, as these modalities provide complementary insights into structural and molecular pathologies (Song *et al.*, 2021; Venugopalan *et al.*, 2021; Castellano *et al.*, 2024). MRI can capture detailed anatomical changes, such as hippocampal atrophy and cortical thinning, which correlate with neurodegeneration. Meanwhile, amyloid PET directly visualizes β-amyloid plaques—a neuropathological hallmark of AD through radiotracers like 18F-Florbetaben, offering functional evidence of disease progression even before structural changes manifest.

More broadly, deep learning has become a pivotal tool in leveraging T1-weighted MRI data for the early diagnosis of AD. By automating the detection of structural and functional biomarkers, such as brain atrophy and amyloid burden, deep learning models provide a non-invasive, cost-effective alternative to traditional diagnostic methods like PET scans or cerebrospinal fluid (CSF) analysis.

2.2.1 Deep Learning and MRI data

Deep learning-based techniques have attracted considerable research attention in the field of medical image analysis (Ker *et al.*, 2017). Their ability to learn directly from raw data makes them an ideal framework for exploring how such methods can

enhance clinical workflows. More fundamentally, it has transformed structural MRI analysis through its capacity to extract latent neuroanatomical patterns directly from raw T1-weighted images, with brain age prediction emerging as a pivotal biomarker for neurodegenerative disease detection (Cole *et al.*, 2018; Jónsson *et al.*, 2019). By training convolutional neural networks (CNNs) on large-scale datasets of healthy individuals, researchers can quantify the discrepancy between predicted brain age and chronological age (brain age gap), which correlates with AD progression and cognitive decline (Cole *et al.*, 2017; Jónsson *et al.*, 2019; Dartora *et al.*, 2024). Early approaches achieved mean absolute errors (MAE) of 3.39 years using residual CNNs with transfer learning across multi-site datasets (Jónsson *et al.*, 2019), while recent innovations have reduced MAE to 2.67 years through streamlined preprocessing pipelines requiring only rigid registration to MNI space (Dartora *et al.*, 2024). This technical advance enables broader clinical adoption by minimizing computational barriers traditionally associated with feature extraction and nonlinear registration.

Current brain age estimation methodologies face three interrelated limitations that hinder clinical translation. First, precision constraints persist despite technical advancements; state-of-the-art models achieve mean absolute errors (MAEs) of 4.73-8.38 years in healthy adults (More *et al.*, 2023; Yu *et al.*, 2024), exceeding the subtle annual brain age acceleration (0.5-1 year) characteristic of preclinical neurodegeneration (Moguilner *et al.*, 2024). This resolution gap fundamentally limits early detection capabilities, as MAEs often surpass the biological signal of interest (Yu *et al.*, 2024). Second, standardization challenges permeate the field. Variability in preprocessing pipelines (e.g., registration methods, intensity normalization) and feature selection (voxel-wise vs. region-of-interest approaches) creates divergent outcomes. Whereas, scanner-related variance introduces additional bias—models trained on 3T MRI data exhibit systematic overestimations (around 5.6 years) when applied to 1.5T scans

from equivalent populations (Moguilner *et al.*, 2024). Even advanced harmonization techniques only reduce cross-protocol MAE variance from 5.63 to 2.25 years (Yu *et al.*, 2024).

Third, systematic bias in brain age prediction models, characterized by overestimation of younger individuals' brain age and underestimation in older cohorts, arises fundamentally from regression to the mean (RTM)—a statistical phenomenon where extreme initial measurements gravitate toward the population average in subsequent assessments (Liang *et al.*, 2019; Lee and Chen, 2025). This bias persists across diverse methodologies (ridge regression, support vector machines, Gaussian processes, and deep neural networks) and datasets, indicating it is inherent to the statistical properties of age prediction rather than specific algorithmic flaws. RTM-induced bias creates artificially inflated brain age gaps (BAGs) in youth and diminished gaps in older adults, complicating the detection of subtle, clinically relevant deviations (e.g., 0.5–1 year acceleration in preclinical AD). While post-hoc corrections (e.g., general linear models) can mitigate this bias (Liang *et al.*, 2019), residual errors (mean absolute errors $\geq$ 2.38–7.14 years) remain orders of magnitude larger than pathological aging rates, limiting sensitivity for early disease biomarkers.

2.2.2 Deep Learning and PET imaging data

PET is a molecular imaging technique that uses radioactive tracers to visualize and measure biochemical changes in the brain. Once injected, these radiotracers bind to specific targets, such as amyloid or tau proteins, and emit positrons as they decay. When a positron collides with an electron, they annihilate, producing two gamma photons that travel in opposite directions. These photons are detected by the PET scanner, allowing the reconstruction of high-resolution images that reflect underlying molecular activity.

Deep learning models have recently been applied to these PET scans to detect subtle pathological changes at the earliest stages of AD, enabling improved prediction, monitoring well before the onset of clinical symptoms, improving quantitative accuracy, and addressing technical limitations inherent to multimodal data. Convolutional neural networks (CNNs) and other architectures have demonstrated exceptional performance in classifying AD stages, with studies achieving AUC values up to 0.98 for distinguishing AD from normal cognition using FDG-PET data (Duan *et al.*, 2023). Multimodal PET/MRI integration has the potential to further enhance diagnostic precision by combining metabolic, molecular, and anatomical data, while overcoming challenges like tracer variability through innovative deep learning solutions for cross-tracer harmonization.

We developed a deep residual inception encoder-decoder network (RIED-Net) that addressed a critical barrier in amyloid PET quantification: variability in the amyloid burden measurements across tracers like Pittsburgh Compound-B and florbetapir (Shah *et al.*, 2022a; Gao *et al.*, 2025). By training the model on 92 subjects with 10-fold cross-validation, the model harmonized global amyloid burden and voxel-wise measurements, achieving significantly stronger between-tracer correlations ($p < 0.001$). This encoder-decoder architecture learned spatial mappings to align tracer-specific uptake patterns while preserving disease-relevant features, enabling more reliable multisite and longitudinal studies (see Appendix C). The harmonization framework's success (validated on an external cohort of 46 subjects) underscores deep learning's capacity to standardize biomarker measurements, which is a prerequisite for accurate disease monitoring.

Moreover, technical limitations hinder optimal integration. PET's low spatial resolution (typically 4-6mm) causes partial volume effects (PVE) (Tohka and Reilhac, 2008), blurring the amyloid signal across tissue boundaries, and complicating quan-

tification. MRI's superior resolution ($\leq$1mm) helps mitigate this through anatomical guidance, but conventional fusion methods struggle with misalignment artifacts and information loss during registration (Song *et al.*, 2021; Shah *et al.*, 2024a). These challenges are particularly acute in early-stage AD, where subtle amyloid deposition and minimal structural changes demand millimeter-level precision for accurate diagnosis.

2.3 Early detection in Headache

There is currently no recognized way of accurately predicting who will recover from post-traumatic headache (PTH) during the acute phase following concussion and who will go on to develop persistent PTH (PPTH), a condition that is difficult to treat effectively. It is of particular interest to identify brain imaging and clinical feature biomarkers using machine learning algorithms that distinguish individuals at high risk for developing PPTH from patients who are likely to acutely recover from PTH prior to three months. Data-driven biomarker discovery can determine important clinical factors and neuropathological mechanisms underlying PPTH and interrogate the contribution of clinical, demographic data, and speech signatures on predicting PTH persistence.

Drawing upon the success of deep learning in the computer vision field, neuroimaging research has also benefited from DL for the study of neurodegenerative diseases, potentially contributing to earlier diagnosis, disease staging, prognosis, and therapeutic development (Myszczynska *et al.*, 2020; Ramzan *et al.*, 2020). However, the development of brain imaging–based DL models for migraine and other headache types has been limited. Our efforts have been among the few research studies exploring DL-based headache detection using structural MRI scans. This mainly includes a supervised classification method for automated biomarker extraction for different headache subtypes (Rahman Siddiquee *et al.*, 2023) and unsupervised anomaly de-

tection methods (Rahman Siddiquee *et al.*, 2022; Siddiquee *et al.*, 2024). The former leveraged a supervised 3D ResNet architecture to classify headache subtypes and automatically localize predictive brain regions, identifying distinct structural patterns for migraine, acute, and persistent post-traumatic headache. In contrast, the latter methods, HealthyGAN and Brainomaly, introduced unsupervised anomaly detection frameworks using generative image translation to learn deviations from healthy brain structure without requiring disease labels. HealthyGAN employs a one-directional image-to-image translation approach that learns to map unannotated mixed datasets to healthy images, relaxing the cycle-consistency constraint common in traditional GANs. This method generates difference maps by subtracting the translated healthy image from the input, highlighting potential anomalies. Building upon this, Brainomaly introduces an additive map–based translation mechanism tailored for neuroimaging, where the model predicts voxel-wise changes needed to transform an input MRI into a healthy counterpart. This approach not only enhances the detection of subtle structural abnormalities but also incorporates a pseudo-AUC metric for model selection in the absence of annotated data. Notably, Brainomaly achieved high AUCs in detecting headache-related abnormalities, demonstrating clinical potential for early detection even in unlabeled MRI datasets. Complementing these efforts, we also explored the prediction of headache recovery using multimodal machine learning algorithms to build predictive models that predict the likelihood of symptom resolution in the APTH cohort (Joshi *et al.*, 2024). The multimodal SVM model combining T2* MRI features and speech biomarkers achieved an AUC of 0.87 for predicting headache improvement at three months in patients with acute post-traumatic headache, outperforming single-modality models (AUC=0.66 for T2* alone and AUC=0.75 for speech alone). This underscores how integrating diverse data modalities can guide personalized prognostication for headache recovery. Furthermore, in our recent work (Rafsani

et al., 2025), we leveraged a language-image foundation model pre-trained on 15 million biomedical image-text pairs to fine-tune on a relatively small MRI dataset comprising various headache subtypes. This approach achieved high classification accuracies (up to 91.67%) and enabled the identification of discriminative brain regions, underscoring the potential of large-scale pretraining for enhancing headache classification and biomarker discovery from limited neuroimaging data.

However, much similar to AD, the relationship between neurodegeneration and headaches is complex and multifaceted. A meta-analysis found that headache disorders are associated with a 1.35-fold increased risk of all-cause dementia, a 1.49-fold increased risk of AD, and a 1.72-fold increased risk of vascular dementia (Qu *et al.*, 2022). Similar to that, another study reported that patients with migraine developed AD with a 1.32 hazard ratio compared to those without migraine history (Kim *et al.*, 2023). Chong *et al.* (2014) found unique age-related patterns of cortical thinning and thickening in migraine patients, suggesting an interaction between migraine and the aging process. Whereas, a recent study found no significant structural brain differences between elderly migraine sufferers and non-sufferers, challenging the notion of long-term structural changes due to migraine (Acarsoy *et al.*, 2024). These contrasting findings underscore the complexity of headaches' effects on the brain and necessitate the need to further investigate the role of headaches in developing accelerated aging patterns and better predict headache persistence (Shah *et al.*, 2024b). Advancements in imaging techniques, particularly MRI, have played a pivotal role in understanding the human brain, allowing direct quantification of such alterations in vivo and at a large scale. This has contributed significantly to headache research and the search for data-driven imaging biomarkers (Rahman Siddiquee *et al.*, 2023). However, identifying and understanding the alterations in brain structure and function linked to the onset, progression, and episodic occurrence of headache disorders

remains an active area of research interest (Schwedt *et al.*, 2024). We hypothesize that MR imaging completed between 7-28 days since the onset of PTH will show greater magnitude of alterations and greater distribution of changes in brain structure and function in patients that go on to have persistence of PTH compared to concussed patients that recover from PTH during the acute phase and compared to healthy controls. In Sec.3.5.4 we demonstrate that the bio-signatures of brain age captured from MRI in patients with acute post-traumatic headache (aPTH) attributed to mild traumatic brain injury (mTBI) may serve as a surrogate neuroimaging biomarker for predicting PTH persistence.

Chapter 3

ROBUST BRAIN AGE PREDICTION

3.1 Motivation

Normal aging causes structural changes in the human brain across the adult lifespan, which is a major risk factor for the decline in physical health and cognitive ability (Cole *et al.*, 2018). Aging also exposes an individual to an increased risk of cancer (López-Otín *et al.*, 2013) and various neurological disorders such as Parkinson's Disease (Beheshti *et al.*, 2020), vascular dementia (Haan and Wallace, 2004), headache (Chong *et al.*, 2014; Shah *et al.*, 2024b), mild cognitive impairment (MCI), and Alzheimer's Disease (AD) (Franke and Gaser, 2012). However, aging in humans is a complex and heterogeneous phenomenon. Even though each individual ages at the same rate chronologically, their biological age does not follow the same trajectory due to genetic factors, environmental influences, underlying neurological conditions, and other unknown factors (López-Otín *et al.*, 2013). Measuring this deviation from normal aging can allow a better understanding of associations between cognitive impairment and aging (Draper and Ponsford, 2008; Gaser *et al.*, 2013) and identify patients at risk for clinical trials (Cole *et al.*, 2018). Hence, there is a growing interest in predicting biological age, most commonly derived from an individual's structural MRI data. The difference between predicted biological age and chronological age, also known as the Brain Age Gap Estimate (BrainAGE) (Franke *et al.*, 2010), can be used to monitor accelerated or decelerated brain aging.

Measuring deviation from normal aging relies heavily on the base model's performance to predict normal aging, i.e., accurately predicting the biological age of healthy

subjects. A model's performance on a healthy cohort is often assessed using mean absolute error (MAE), which calculates the mean of absolute BrainAGE. Existing deep learning-based regression approaches (Jónsson *et al.*, 2019; Cole *et al.*, 2017; Ito *et al.*, 2018) have limited clinical applications because the models have reported MAE $4-5$ years in healthy cohorts, suggesting the lack of discriminative power to interrogate BrainAGEs of different clinical groups (Gaser *et al.*, 2013). Moreover, a common challenge in brain prediction models is the issue of systematic bias (Beheshti *et al.*, 2019; Liang *et al.*, 2019; Zhang *et al.*, 2023a), where there is an overestimation in the predicted age of young subjects and an underestimation of old subjects. If BrainAGE were to be used as a reliable imaging biomarker for measuring brain health, the effect of systematic bias is of concern. For example, AD patients are often aged 50+, and age underestimation will impact early detection. Studies have investigated whether this bias is induced due to model or data selection used for training (Liang *et al.*, 2019). We argue that systematic bias is inherent to brain age prediction due to its formulation as a regression analysis. This study has two primary objectives:

1. Addressing the systematic bias of regression-to-mean to improve the robustness of brain age prediction.

2. Enhancing the model's performance in predicting normal aging in healthy cohorts, thereby facilitating more accurate disease detection in downstream tasks.

Traditionally, brain age estimation is formulated as a regression task since the problem of interest is understanding which bio-signatures from imaging data have a statistically significant effect on age. And more importantly, it is clinically relevant to study how these signatures change across different age groups and track their progres-

sion. To accomplish this, capturing ordinal information from the ground-truth age is critical; hence, regression is preferable. However, it is known that regression models suffer from systematic bias. To address this issue, we propose reformulating the task of brain age prediction as multi-class classification. However, in classification, each class is treated independently of the other and hence cannot capture the ordinality of target labels (Zhang *et al.*, 2023b). To counter this, we propose a novel *ORdinal Distance Encoded Regularization* (**ORDER**) loss in conjunction with cross-entropy loss for multi-class ordinal classification. ORDER loss (Shah *et al.*, 2024c) is calculated based on the Manhattan distance between samples in the training mini-batch within both feature space and target space. As depicted in Fig.3.1, it scales the distance between learned features in high-dimensional space by a weighted magnitude of the chronological age difference (see Sec. 3.3.1). We propose a new ordinality metric to quantify the relative ordering of feature representations compared to their actual target label ordering. Results show that our proposed framework preserves ordinality in feature space and improves brain age prediction by a statistically significant amount compared to existing deep learning approaches (Jónsson *et al.*, 2019; Cole *et al.*, 2017; Ito *et al.*, 2018).

One challenge in medical imaging is heterogeneity in the quality of MRI scans due to different scanners and acquisition protocols. Several studies have confined themselves to a single cohort to train and evaluate model performance (Jónsson *et al.*, 2019; Peng *et al.*, 2021), which could affect multi-site studies or generalization performance. Contrary to that, it is shown that deep learning (Mårtensson *et al.*, 2020) and machine learning models (Franke *et al.*, 2010; Franke and Gaser, 2012; Kaufmann *et al.*, 2019) are not only robust to scanner differences, but diversity in data due to heterogeneous sources can improve model generalization. In this study, we decide to combine cohorts from 5 public data sources to train and validate our model collected

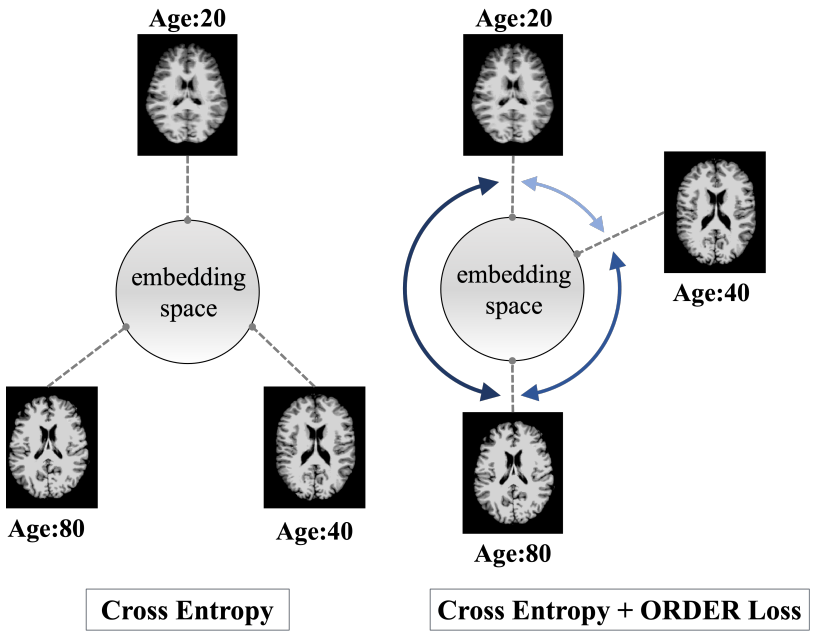

Figure 3.1: Standard Cross-entropy Vs. Cross-entropy with ORDER Loss: Here, cross-entropy loss (left) encourages the model to learn high-entropy feature representations where embeddings are spread out. However, it fails to capture ordinal information from labels. Our proposed ORDER loss with cross entropy (right, Eq.3.5) preserves ordinality by spreading the features proportional to the Manhattan distance between normalized features weighted by absolute age difference. The illustrated example (right) shows an embedding space where learned representations of MRI scans with ages 20, 40, and 80 are distributed apart from one another, with distances proportional to absolute age differences.

from (1) National Alzheimer's Coordinating Center's (NACC), (2) Open Access Series of Imaging Studies (OASIS) (Marcus *et al.*, 2007, 2010), (3) International Consortium for Brain Mapping (ICBM), (4) Information eXtraction from Images (IXI), and (5) Autism Brain Imaging Data Exchange (ABIDE) (Di Martino *et al.*, 2014). Additionally, we evaluate its performance for disease detection on an independent dataset. In summary, the main contributions of this research are the following:

1. We formulate Brain Age prediction as an ordinal classification task that outperforms existing regression-based methods by a significant margin.

2. We introduce a novel *ORDER loss* for classification that preserves the ordinality in the learned feature space from target labels, which here is Age.

3. Our proposed framework addresses the well-observed issue of systematic bias in predicted biological age from neuroimaging data.

4. The developed model detected subtle differences between clinical groups of AD, which were not accurately captured by the regression model or other approaches.

3.2 Related Works

3.2.1 Neuroimaging-based Brain Age prediction

Prior studies on brain age prediction from neuroimaging data (Franke *et al.*, 2010; Cole *et al.*, 2018; Gaser *et al.*, 2013; Valizadeh *et al.*, 2017; Baecker *et al.*, 2021b; Liem *et al.*, 2017; Cole *et al.*, 2015; Franke and Gaser, 2012; Beheshti *et al.*, 2020) use regression techniques such as Gaussian process regression, support vector regression, and relevance vector regression. These approaches involve extensive pre-processing of raw structural MRI data and extracting imaging features such as cortical thickness,

regional volumes, or surface area using tools such as FreeSurfer or Statistical Parametric Mapping (SPM). Input to the machine learning models are these pre-processed brain morphological features, and chronological age is the target variable.

More recent studies have also explored deep neural networks to predict brain age using raw neuroimaging data (Cole *et al.*, 2017; Jónsson *et al.*, 2019; Peng *et al.*, 2021; Jiang *et al.*, 2020; Shah *et al.*, 2022b, 2023b; Bashyam *et al.*, 2020) and results demonstrate that deep neural networks outperform traditional machine learning approaches given sufficient training data (Bashyam *et al.*, 2020; Cole *et al.*, 2017; Ito *et al.*, 2018). Since deep learning methods perform automatic feature extraction from raw structural MRI data, it allows capturing previously unseen imaging signatures related to aging in the brain and makes the model less prone to any biases from pre-processing steps, making it more generalizable.

3.2.2 Systematic Bias in Predicted Brain Age

In brain age prediction, predicted biological age is often observed to be systematically biased towards the cohort's mean age (Liang *et al.*, 2019; Le *et al.*, 2018; Smith *et al.*, 2019; Treder *et al.*, 2021; Beheshti *et al.*, 2019) affected by regression to the mean (RTM) effect, limiting its potential clinical utility. This causes an unexpected overestimation of predicted brain age in young subjects and underestimation among old subjects. Historically, the RTM effect has been attributed to within-subject and between-subject variability (Gardner and Heady, 1973). This systematic bias in predicted brain age is not specific to the choice of learning algorithm, data sample imbalance across age groups, or imaging data heterogeneity due to different scanners (Liang *et al.*, 2019). Since brain age prediction is traditionally formulated as a regression problem, RTM is a characteristic phenomenon of regression analysis.

Recently, Lee and Chen (2025) showed how systematic bias emerges naturally in

machine learning regression models optimized for mean squared error (MSE) loss. For tasks like brain age prediction, models achieve lower MSE by systematically underestimating high values and overestimating low values (linear central tendency warp). This mainly occurs because: squared error penalizes large deviations asymmetrically and predictions regress toward the mean to minimize extreme errors.

Studies that aim to mitigate this systematic bias propose post-hoc correction methods where predicted age is scaled by slope and intercept derived from regression of predicted age or BrainAGE (Cole *et al.*, 2018; de Lange *et al.*, 2019; Beheshti *et al.*, 2019) on chronological age. Le *et al.* (2018) used chronological age as a covariate when analyzing group-level differences in BrainAGE, whereas Cole *et al.* (2018) did not include chronological age in the final adjustment scheme. However, it increased the variance in predicted BrainAGE (Beheshti *et al.*, 2019). Other studies (Beheshti *et al.*, 2019; de Lange *et al.*, 2019) included chronological age in the final age adjustment, but these methods are likely to be inaccurate when the age range of the independent testing dataset differs from the age range of the model's training data. Recently, Zhang *et al.* (2023a) found that these correction methods do not properly address the systematic bias in predictions. Experiments from that study also show that even though linear (Beheshti *et al.*, 2019; Cole *et al.*, 2018) and quadratic (Smith *et al.*, 2019) correction methods push average BrainAGE close to zero, bias in BrainAGE for same-age subjects gets worse.

More fundamentally, correcting the predicted BrainAGE in a two-step process by explicitly controlling for age would make downstream analysis questionable. This highlights the need to develop an end-to-end method that addresses systematic bias in brain age prediction and is more accurate in predicting normal aging.

3.2.3 Regression as Ordinal Classification

Predicting brain age from imaging data is an ordinal classification task (also known as ordinal regression) since the labels exhibit a natural order. Gutiérrez *et al.* (2015) conducted a comprehensive exploration of ordinal classification methodologies, categorizing them into three main groups: naive approaches using regression or nominal classification methods, ordinal binary decomposition, and threshold models. However, the efficacy of ordinal decomposition approaches relies heavily on task-specific decomposition strategies, while threshold models demand meticulous calibration of hyperparameters to achieve optimal convergence (Rosati *et al.*, 2022). In this study, we compare our approach with the nominal classification and regression techniques previously documented in the literature.

In computer vision, it is shown that classification can outperform regression in many tasks, such as age estimation from face images (Pan *et al.*, 2018; Lanitis *et al.*, 2004; Rothe *et al.*, 2018), object counting (Liu *et al.*, 2019), and depth estimation (Cao *et al.*, 2017b). The target space is discretized into the same-size intervals, and surprisingly, models are more accurate in predicting a range of values rather than estimating actual values on a continuous scale. The exact reason for classification outperforming regression has been less explored before. Zhang *et al.* (2023b) suggests that classification benefits from its ability to learn high-entropy feature representations compared to regression, which accounts for the performance gap. Inspired by these insights, we transform the task of brain age prediction from regression to multi-class classification. In brain age prediction, the target output follows a continuous scale consisting of the human life age span. Despite the performance improvement, classification models treat each class label independently from each other, where each wrong prediction is penalized equally. For instance, given a sample with a true age

of 53, cross-entropy (CE) penalizes the model by the same magnitude if the wrong prediction was 21 or 52. Hence, the ordinal relationship between target labels is not accurately captured in learned representations of brain age using CE or other loss functions proposed in previous studies (Pan *et al.*, 2018; Zhang *et al.*, 2023b) (Sec. 3.5.2).

One of the initial works that proposed deep learning-based classification for age estimation from facial images was by Rothe *et al.* (2018), where they used the expected mean of softmax weights as the estimated age. Pan *et al.* (2018) also used softmax expected value for age estimation with an additional mean-variance loss used in training. Mean loss minimizes the difference between the mean of the estimated distribution and the ground truth, while the variance loss minimizes the variance of the estimated distribution, resulting in a concentrated distribution. Different from these approaches, Zhang *et al.* (2023b) observed that classification allows learning high-entropy feature representation with a more diverse feature set compared to regression. They introduce an Euclidean distance-based loss with mean squared error (MSE) loss for regression to increase the marginal entropy such that learned features are spread out while preserving target ordinality. The latter two studies (Pan *et al.*, 2018; Zhang *et al.*, 2023b) also highlight preserved ordinality in learned feature space from their proposed approaches. However, for brain age prediction, we show that this is not the case when compared to a regression model (Sec. 3.5.2).

3.3 Proposed Framework

Fig.3.2 gives an overview of our framework for robust brain age prediction. In this section, we first describe our proposed ORDER loss that encodes ordinal information within target labels into learned feature space. Then, in addition to MAE, we define two metrics to measure our model's performance in preserving ordinality

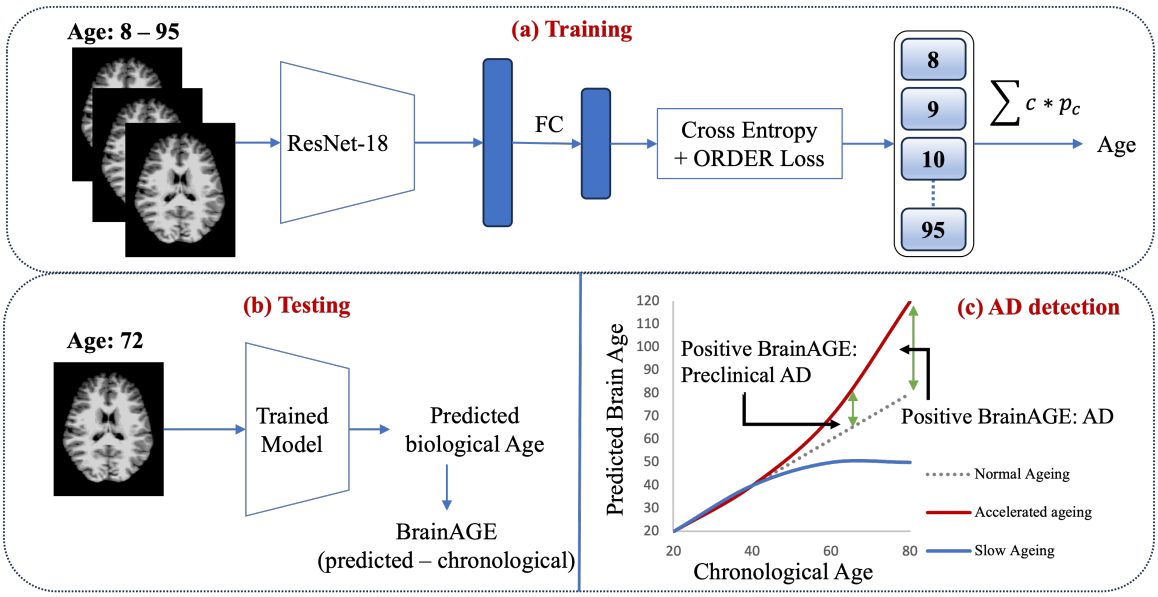

Figure 3.2: Overview of Proposed Brain Age Prediction Framework. (a) A 3D ResNet-18 model is trained using the lifespan cohort with cross-entropy and ORDER as loss functions. Age is calculated as the weighted average of class probabilities from the softmax classifier. (b) During inference, the Brain Age Gap Estimate (BrainAGE) is calculated as the difference between predicted biological age and actual chronological age. (c) The trajectory plot offers a visual interpretation of predicted BrainAGE and its associations with aging patterns. The preclinical AD stage is when the patient behaves cognitively normal, but underlying changes in the brain due to accelerated aging happening at a subtle rate can be captured using BrainAGE.

and minimizing systematic bias compared to established methods.

3.3.1 ORDER Loss

To better understand the intuition behind the proposed ORDER loss, we first review the original cross-entropy loss (L_{CE}), which is formulated as:

$$
\begin{aligned}
L_{CE} &= -\frac{1}{N}\sum_{i=1}^{N} y_i \log(\hat{y}_i) = -\frac{1}{N}\sum_{i=1}^{N} \log(\hat{y}_i) \\
&= -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{W_{y_i}^T x_i}}{\sum_{j=1}^{C} e^{W_{y_j}^T x_i}}
\end{aligned}
\tag{3.1}
$$

where x_i is input to the last fully connected layer corresponding to i-th sample from training data N, y_i is the hot encoding of the true label, $\hat{y}_i$ is the predicted probability, and $W_{y_j}^T$ is j-th column of last fully connected layer ($j \in [1, C], C$ is number of classes). $W_{y_i}^T x_i$ often denoted as z_i, is the target logit of i-th sample (Pereyra *et al.*, 2017).

$$
L_{CE} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \tag{3.2}
$$

In brain age prediction, our main aim is to understand how the dependent variable (age) changes with variations in the independent variables (imaging features). Given a sample from class c, cross-entropy loss forces $z_c > z_j (\forall j \neq c)$. However, when the class labels are ordered, it does not guarantee that learned feature representation follows the same order, i.e., $z_c > z_{c+1} > z_{c+2} > ... > z_C$ and $z_c > z_{c-1} > z_{c12} > ... > z_1$. Even though L_{CE} increases the marginal entropy of feature space, resulting in a diverse feature set, the marginal ordering between class labels is not correctly captured. Keeping the diversity of features from L_{CE} intact, we adjust the target logit z_i with the corresponding feature vector x_i's distance to other features $x_j (\forall j \neq i)$ in a batch of samples, weighted by the distance between class labels.

$$
z_i' = W_{y_i}^T x_i + \varphi(x_i) \tag{3.3}
$$

where,

$$
\varphi(x_i) = \frac{1}{N-1}\sum_{j=1, i\neq j}^{N} |i-j||\bar{x}_i - \bar{x}_j|_{manh} \tag{3.4}
$$

and $\bar{x}$ is L_2 normalized vector $\bar{x} = x/max(||x||_2)$. Substituting Eq.3.3 in Eq.3.2 we get new loss L_T, which can be decomposed into L_{CE} and ORDER loss (L_{ORDER})

$$\begin{aligned} L_T &= -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T x_i+\varphi(x_i)}}{\sum_{j=1}^{C} e^{W_{y_j}^T x_i}} \\ &= -\frac{1}{N}[\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T x_i}}{\sum_{j=1}^{C} e^{W_{y_j}^T x_i}} + \sum_{i=1}^{N}\varphi(x_i)] \\ &= -\frac{1}{N}\sum_{i=1}^{N}\log\frac{e^{W_{y_i}^T x_i}}{\sum_{j=1}^{C} e^{W_{y_j}^T x_i}} \\ &\quad - \frac{1}{N(N-1)}\sum_{j=1,i\neq j}^{N}|i-j||\bar{x}_i - \bar{x}_j|_{manh} \\ &= L_{CE} + L_{ORDER} \end{aligned} \tag{3.5}$$

L_{ORDER} uses the Manhattan distance to calculate the distance between two features x_i and x_j in high-dimensional space. Euclidean distance is the most common metric to measure similarity or distances between two data points. However, Aggarwal *et al.* (2001) found that, due to the curse of dimensionality in high-dimensional space, the sparsity of features is significantly high, making them almost equidistant from each other. The ratio between the closest and farthest points from a reference sample approaches 1 in high-dimensional space (Domingos, 2012). This further explains the inability of a classification model to capture ordinal information. We explored different orders of distance metrics for ORDER loss, but Manhattan distance performed best (see Sec.3.5.5).

3.3.2 Evaluation Metrics

Measuring Ordinality

To the best of our knowledge, we found no defined metrics in the literature that measure the ordinality of feature representations from a deep learning model with

reference to the order of ground truth. Given n images and c ordered classes, we first obtain n features of 512 dimensions from the penultimate layer of a trained model $\{x_1, x_2, ..., x_n\}$. From those features, we calculate c feature centroids $\{f_1, f_2, ..., f_c\}$ using ground-truth labels corresponding to each class. After that, we compute the Manhattan distances between f_1 and other feature centroid as $D = \{d_{12}, d_{13}, ...d_{1c}\}$ where,

$$d_{ij} = |d_i - d_j|_{manh} \tag{3.6}$$

Since class labels here are age values in a chronologically increasing order, we get $C = \{1, 2, ..., (c-1)\}$ as the distance of the first class to others. We define the ordinality metric as the Pearson correlation coefficient between D and C. For a model that perfectly captures ordinal relationships in feature representations, the ordinality score is close to 1. Pearson's correlation between two continuous variables measures how much change in one variable is associated with a proportional change in the other variable. Using this metric, we evaluate our model's performance to capture age-related order information from labels compared to other approaches (see Tab.3.4). An ordinality score close to +1 indicates that the learned features have a similar ranking order as their corresponding ground-truth labels, and a lower value indicates otherwise.

Quantifying Systematic Bias

Previous approaches discussed in Sec.3.3.2 that propose post-hoc correction methods use correlation of predicted BrainAGE and chronological age as a measure of underlying systematic bias (Le *et al.*, 2018; Liang *et al.*, 2019). Using chronological age to adjust BrainAGE would reduce age dependence on BrainAGE, i.e., $r = 0$. However,

it does not address the inherent systematic bias effect caused due to regression. Additionally, this correction method would be questionable when the test dataset does not have the same age range as the training dataset.

To objectively quantify systematic bias caused by regression to the mean effect, we compare the predicted BrainAGE at one standard deviation away from mean (Gardner and Heady, 1973), i.e., for values less than $(\mu - \sigma)$ and greater than $(\mu + \sigma)$, where μ and σ are mean and standard deviation of target age values of the test set. We refer to these two groups as Systematic Bias - Left and Right (SB-L, SB-R). Since there is an overestimation of predicted biological age in young subjects and an underestimation in old subjects, the bias causes higher BrainAGE and lower BrainAGE values for those respective sub-groups. We compare these scores across different methods, and a value closer to 0 indicates better performance in addressing systematic bias (see Tab.3.4).

3.4 Experiment Design

3.4.1 Datasets and Preprocessing

Since most medical imaging datasets are part of multi-center studies, differences in scanners, imaging protocols, variations in vendors, and their hardware account for heterogeneity in data. Deep learning models are known to be robust against heterogeneity in data. In fact, including more heterogeneous data in model training improves its generalization on out-of-distribution data (Mårtensson *et al.*, 2020). With that consideration, we used a combined lifespan cohort of $7,377$ T1-weighted MRI scans of healthy participants from five different public sources.

Table 3.1: Age Range with Distribution and Number of Samples for Each Cohort. The Lifespan Cohort Comprises NACC, OASIS, ICBM, IXI, and ABIDE, Whereas the Discovery Cohort Consists of Samples from the ADNI Cohort.

Dataset	**Count**	**Age Range** (yrs)	**Mean ± STD** (yrs)
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

Lifespan Cohort

All the age prediction models were trained, validated, and tested on a healthy cohort (age: 8-95 years) collected from (1) NACC Uniform Data Set (UDS) from 1999 to March 2021 (2) OASIS (3) ICBM (4) IXI and (5) ABIDE. These cohorts included both 1.5T and 3T scans with predominantly Caucasian participants but also included other race/ethnic groups. The number of samples and age range per cohort are summarized in Tab.3.1.

All five cohorts were preprocessed using an in-house data preprocessing pipeline. T1-weighted MR images were first aligned to the MNI template with rigid transformation, and then intensity normalized and conformed using FreeSurfer v7 to generate preprocessed images at 1 mm isotropic voxels with a 256 x 256 x 256 matrix.

Alzheimer's Disease dataset

Additionally, we collected 1,584 MRI scans from the AD Neuroimaging Initiative (ADNI) database containing a mix of healthy, cognitively impaired, and AD patients. We used this cohort as an independent testing and discovery dataset to evaluate model performance in predicting age and its ability to differentiate clinical groups in AD. Priority was given to scans with matching PET data and participants who had longitudinal follow-ups. For healthy controls (HCs), a random subset was selected from the overall ADNI set and included in this analysis. The diagnostic status was determined based on ADNI clinical data. In this analysis, HC (N=678) participants had normal cognition and did not convert to MCI or AD in follow-up visits. HC to MCI converters (HC-MCI, N=179) are participants who had normal cognition at baseline but converted to MCI during follow-up. MCI-stable (MCIs, N=432) participants had a baseline diagnosis of MCI and stayed unchanged in follow-ups. MCI to AD converters (MCI-AD, N=139) are those participants with an MCI diagnosis at baseline and subsequently converted to AD. AD (N=156) patients are those who were diagnosed with AD at baseline.

Headache dataset

Our institutional dataset of headache patients consisted of T1-weighted MRI data from 93 individuals diagnosed with migraine, 99 with APTH, and 49 with PPTH, in accordance with the diagnostic criteria of the International Classification of Headache Disorders (ICHD) available at the time the participant was enrolled (ICHD-3 beta or ICHD-3)(of the International Headache Society , IHS). Additionally, we included 104 healthy controls from the Mayo Clinic as the reference group. For each data cohort and headache disorder, the information of participants is summarized in Tab.3.2.

All participants were imaged at the Mayo Clinic Arizona. This study was approved by the Mayo Clinic Institutional Review Board (IRB), and all participants provided written informed consent for their participation. At the time of enrollment, migraine participants were diagnosed with episodic or chronic migraine, with or without aura, based on the most recent edition of the International Classification of Headache Disorders (ICHD-3 beta or ICHD-3). Participants with APTH or PPTH had PTH attributed to mild traumatic brain injury (mTBI) according to the latest ICHD criteria (ICHD-3 beta or ICHD-3). Individuals with a history of moderate or severe traumatic brain injury were excluded from the study. Participants with APTH were enrolled between 0 to 59 days following mTBI, while those with PPTH were enrolled at any point after they had PTH for longer than three months. Healthy controls (HC) were excluded if they had a history of any headache type other than infrequent tension-type headache.

Image acquisition was performed using 3 Tesla Siemens scanners (Siemens Magnetom Skyra, Erlangen, Germany) equipped with a 20-channel head and neck coil. Anatomical T1-weighted images were captured using magnetization-prepared rapid gradient echo (MPRAGE) sequences. The imaging parameters for acquiring T1-weighted images were as follows: repetition time (TR) = 2400 ms, echo time (TE) = 3.03 ms, flip angle (FA) = 8°, and voxel size = $1 \times 1 \times 1.25$ mm^3.

The participants with Mig. had a mean age of 39.9 years ($\pm$11.6), and 74.7% were female. They had a mean headache frequency of 15.3 days per month; 37 had episodic migraine and 59 had chronic migraine. Additionally, 49 patients reported experiencing an aura with at least some of their migraine attacks. The PTH group included 48 participants with APTH, with a mean age of 41.6 years ($\pm$12.7), and 60.4% were female. Individuals with APTH experienced their first PTH symptoms an average of 24.5 days ($\pm$14.5) prior to imaging. They reported headaches on approximately

76.3% (±29.6%) of days following the onset of APTH. Their MTBI was due to motor vehicle accidents (n = 20), falls (n = 21), and direct hits to the head (n = 7). The dataset also included 49 patients with PPTH. The mean age of the PPTH patients was 38.1 years (±10.5 years), and 34.7% were female. PPTH participants experienced headaches on an average of 15.3 days (±7.4 days) per month. The mTBIs leading to PPTH were attributed to various causes, including sports-related injuries (n = 8), falls (n = 12), motor vehicle accidents (n = 7), and blast injuries (n = 22).

Details regarding participant enrollment, demographic characteristics, and other relevant information of public datasets used here are provided in the original publications or documentation of the respective datasets. Readers are encouraged to refer to these sources for comprehensive descriptions of the study populations and methodologies.

Table 3.2: Summary of Age Range with Distribution and Number of Samples for Each Cohort.

Source	**Number of subjects**	**Diagnostic status**	**Age Range (years)**
Mayo Clinic, Arizona	99	Acute PTH	18–70 (42.2 ± 15.3)
	49	Persistent PTH	19–63 (38.1 ± 10.6)
	93	Migraine	22–66 (39.6 ± 11.7)
	137	HC	18–66 (39.5 ± 12.0)
ABIDE	176	HC	18–56 (26.1 ± 7.0)
ICBM	1,101	HC	18–80 (37.6 ± 15.4)
IXI	536	HC	20–86 (48.8 ± 16.5)
NACC	4,132	HC	18–95 (67.5 ± 10.8)
OASIS	1,432	HC	8–94 (27.9 ± 20.7)

PTH = post traumatic headache; **HC** = healthy controls.

3.4.2 Implementation Details

A 3D ResNet-18 was adopted as the base deep-learning model, and the input to the model is 3-dimensional MRI scans with a batch size of 4. Stratified oversampling was employed in classification models, and for regression, we stratified samples based on age groups $(8-12, 12-16, ..., 92-96)$ to perform oversampling. We used AdamW optimizer with a $1e^{-3}$ learning rate and weight decay of $1e^{-2}$. Each model was trained for 100 epochs with early stopping to avoid over-fitting. All experiments were performed on NVIDIA's $A100$ $80GB$ GPUs to train, validate, and test the models. The implementation code is made public for reproducibility and further research[1] .

We evaluate our proposed brain age prediction framework and other baseline methods on a combined healthy cohort using three different metrics specific to this task. Evaluation metrics include MAE, Ordinality, and Systematic Bias scores.

3.5 Results and Analyses

3.5.1 BrainAGE prediction

We compare our proposed method's performance in predicting the brain age of healthy individuals from a lifespan cohort to four baseline methods, including two regression and two classification models (Tab.3.3). For classification models, age values were rounded off to the closest integer and assigned respective class labels. Only 535(7.3%) samples from the lifespan cohort had non-integer age values.

Our model performed best on the healthy test set with MAE 2.56, outperforming standard MSE and cross-entropy loss models. Among other competing methods, the classification model with mean-variance loss performed best. Model with cross-entropy loss outperforms MSE model due to its ability to learn high entropy

[1]https://github.com/jaygshah/Robust-Brain-Age-Prediction

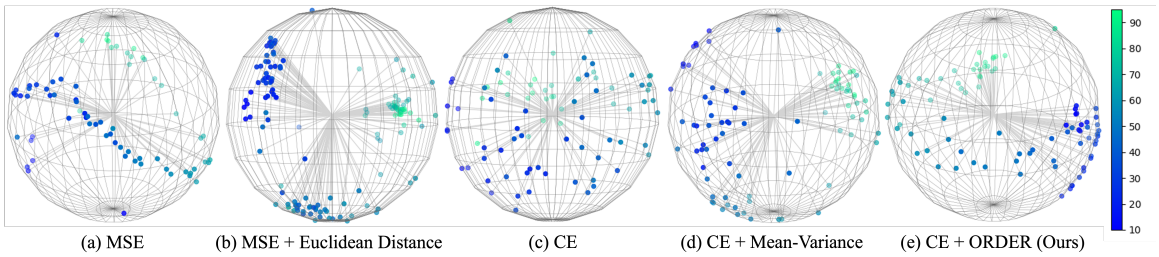

Figure 3.3: A T-SNE Visualization of Embeddings from Models' Penultimate Layer: (a) When using MSE loss, embeddings maintain ordinal relationships but are tightly packed, resulting in a low-entropy feature space (b) MSE with Euclidean distance loss spreads out embeddings but struggles to preserve ordinal relationships accurately (c) Cross-entropy (CE) further spreads embeddings, creating a high-entropy space, but at the cost of losing ordinal information (d) Mean-variance loss combined with cross-entropy creates a high-entropy feature space and slightly improves ordinality (Tab.3.4). (e) ORDER loss combined with cross-entropy achieves the best balance: it accurately preserves ordinality, maintains a high-entropy space, and improves overall performance. Embeddings are color-coded based on their ground truth age values $[10 - 95]$.

features (Fig.3.3), where inter-class features are spread out, and intra-class features are compact (Boudiaf *et al.*, 2020). Surprisingly, adding an Euclidean distance-based regularizer to the MSE loss did not improve the regression model's performance. Our method's performance is also significantly better than MAE reported by prior studies using regression analysis (Ito *et al.*, 2018; Jónsson *et al.*, 2019; Cole *et al.*, 2017), however, on different cohorts.

3.5.2 Ordinality and Systematic Bias

We further evaluate our model's ability to preserve ordinality and address systematic bias in predicted BrainAGE using metrics defined in Sec.3.3.2 and Sec.3.3.2.

Table 3.3: Brain Age Prediction Results on Lifespan Cohort. Here, MAE measures the difference between predicted and actual chronological age on the same test set. Here, **bold** numbers represent the best results, while underlined numbers represent second-best results.

Method	**MAE**
MSE	3.93
MSE + Distance (Zhang *et al.*, 2023b)	4.57
CE (Rothe *et al.*, 2018)	3.33
CE + Mean-Variance (Pan *et al.*, 2018)	2.65
CE + ORDER (Ours, Shah *et al.* (2024c))	**2.56**

As expected, the model with MSE loss had the highest ordinality score (Tab.3.4). Our classification model with ORDER loss had an ordinality score much closer to standard MSE loss than other methods, demonstrating its effectiveness in learning ordinal information. Fig.3.3 offers a visual comparison of learned feature space using different loss functions to confirm this further.

Furthermore, the model with ORDER loss also performed best in reducing systematic bias measured by average BrainAGE values at one standard deviation away from the mean. The mean of the test set was 53.4 with a standard deviation of 22.2. Hence, the bias scores reported in Tab.3.4 are BrainAGE values for age < 31.2 (SB-L) and age > 75.6 (SB-R). Values closer to zero reflect a better reduction in systematic bias. Both MSE-based models had a higher systematic bias due to the inherent RTM effect. Due to its ability to learn class-specific and diverse feature sets, cross-entropy loss reduces bias effects for SB-L and SB-R groups. Incorporating order information allows the model to learn the relative ranking of labels, further improving ordinal

Table 3.4: Performance Evaluation of All Methods in Preserving Ordinality and Addressing Systematic Bias in Brain Age Prediction Using Metrics Defined In Sec.3.3.2 and Sec.3.3.2. Here, systematic bias is measured one standard deviation from mean: systematic bias-left ($\mu - \sigma$) and right ($\mu + \sigma$) [SB-L, SB-R].

Method	**Ordinality**	**Systematic Bias**	
		SB-L	**SB-R**
MSE	**0.99**	3.4	-4.2
MSE + Distance	0.95	4.8	-4.1
CE	0.31	1.1	-3.6
CE + Mean-Variance	0.58	<u>0.4</u>	<u>-4.2</u>
CE + ORDER	<u>0.98</u>	**0.1**	**-2.5**

classification performance.

3.5.3 Alzheimer's Disease detection

AD has a prolonged preclinical phase where brain changes manifest subtly as accelerated aging (Long and Holtzman, 2019). Fig.3.2 illustrates this phase, showing accelerated aging diverging slightly from normal aging. MCI, a pre-dementia stage, involves greater cognitive decline than typical aging (Selkoe, 1997). BrainAGE can help detect and monitor this stage early.

We use our discovery cohort (Sec.3.4.1) obtained from ADNI with five clinical groups to test BrainAGE prediction using different methods. Trained models were applied to this cohort using the abovementioned methods to calculate BrainAGE. These five groups were ranked $[1-5]$ in an increasing order of disease severity as HC $<$ HC-MCI $<$ MCI-stable $<$ MCI-AD $<$ AD. Since disease severity is proportional

to accelerated aging, we expect the average predicted BrainAGE to follow the same order. Pearson correlation was calculated between the model's predicted BrainAGE and the rank of disease severity. A high correlation would indicate the model's ability to accurately characterize aging signatures along the AD continuum via estimated BrainAGE. From Tab.3.5, we see that only the model with MSE and our proposed loss has a high correlation.

We further compare the ability of MSE and ORDER loss models to detect subtle differences between these clinical groups accurately. From Fig.3.4, we see that the MSE model had a more disruptive trend in predicted BrainAGEs between groups, i.e., there was a higher difference between AD and MCI-AD ($p = 0.16$) compared to AD and MCI-stable ($p = 0.56$). Whereas the ORDER loss model had an overall consistent trend in statistical significance between groups associated with actual disease severity, highlighting its better discriminative power. It also better detected differences between HC and HC-MCI subjects ($p = 0.07$) compared to MSE ($p = 0.34$), which is crucial for early AD detection. Although our model's performance wasn't as strong as MSE in distinguishing between HC-MCI and MCI-stable, we posit that this could be attributed to the definitions of clinical groups used here. The absence of clinical tools to definitively differentiate HC-MCI from MCI-stable groups, given that subjects exhibit normal cognitive behavior and no discernible symptoms despite age-related brain changes, might contribute to this outcome. We plan to work with a clinician to further investigate these observations from both groups.

3.5.4 Headache Phenotypes

Fig.3.5 summarizes the trend of the neuroimaging-derived aging biomarker and its association with headache severity and different phenotypes. The model predicted Δage using baseline imaging was significantly higher for patients with migraine (3.74$\pm$

Table 3.5: Average Brainage Values Across the Five Clinical Groups of AD. The last column is the Pearson correlation between average BrainAGE values and disease severity of clinical groups in increasing order from HC to AD (1-5). MV: Mean-Variance.

Method	HC	HC-MCI	MCIs	MCI-AD	AD	Correlation
MSE	-1.2	-0.8	-0.3	0.8	1.5	0.98
MSE + Distance	-2.7	-1.9	-1.7	-0.9	0.9	0.94
CE	-1.9	-1.5	-3.4	-2.3	-4.1	-0.75
CE + MV	-1.6	-0.3	-0.5	0.8	2.8	0.94
CE + ORDER	-1.5	-0.8	-0.3	1.2	2.0	0.98

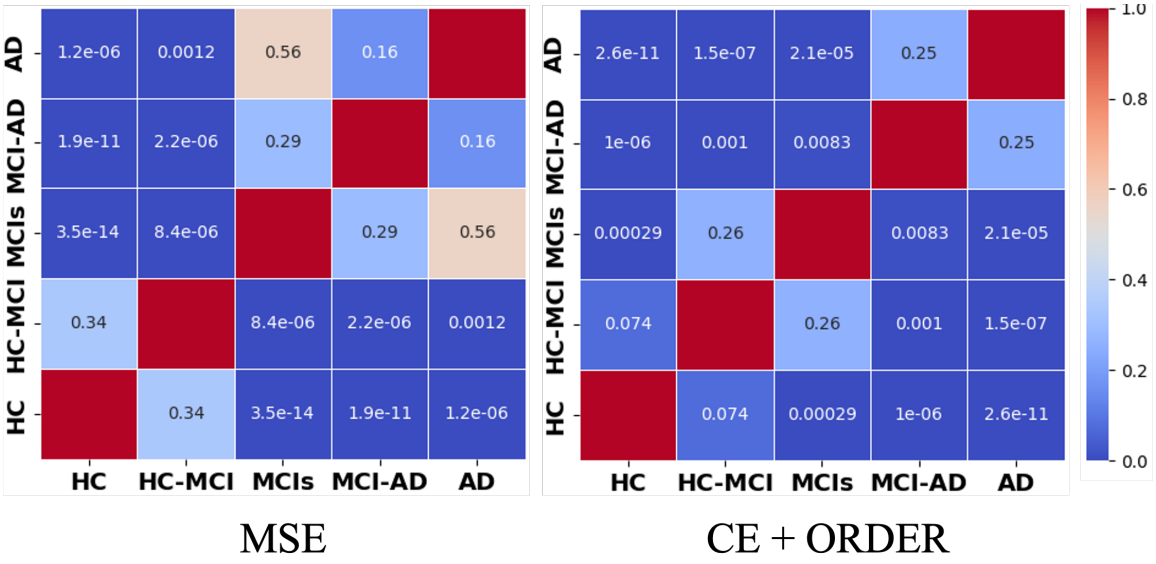

Figure 3.4: Heatmap of Statistical Significances Between the Five Clinical Groups of AD Calculated as p Values from a T-test on Predicted BrainAGE from Respective Groups, for MSE and Cross-entropy with ORDER Loss Models.

1.03 years, $p = 0.03$) and PPTH (4.65 ± 1.41 years, $p = 0.01$) compared to healthy controls (0.46 ± 0.87 years). For APTH patients, Δage was higher compared to healthy

controls (1.25 ± 0.87 years) with statistically significant differences compared to the PPTH ($p = 0.04$) and migraine ($p = 0.08$) groups. One-way ANOVA analysis revealed a statistically significant effect of group on Δage ($F = 3.53$, $p < 0.05$), indicating that the model-predicted Δage varied significantly across the groups. Although the differences between APTH and healthy control ($p = 0.55$) and PPTH and migraine ($p = 0.67$) groups were not significant with pair-wise t-tests.

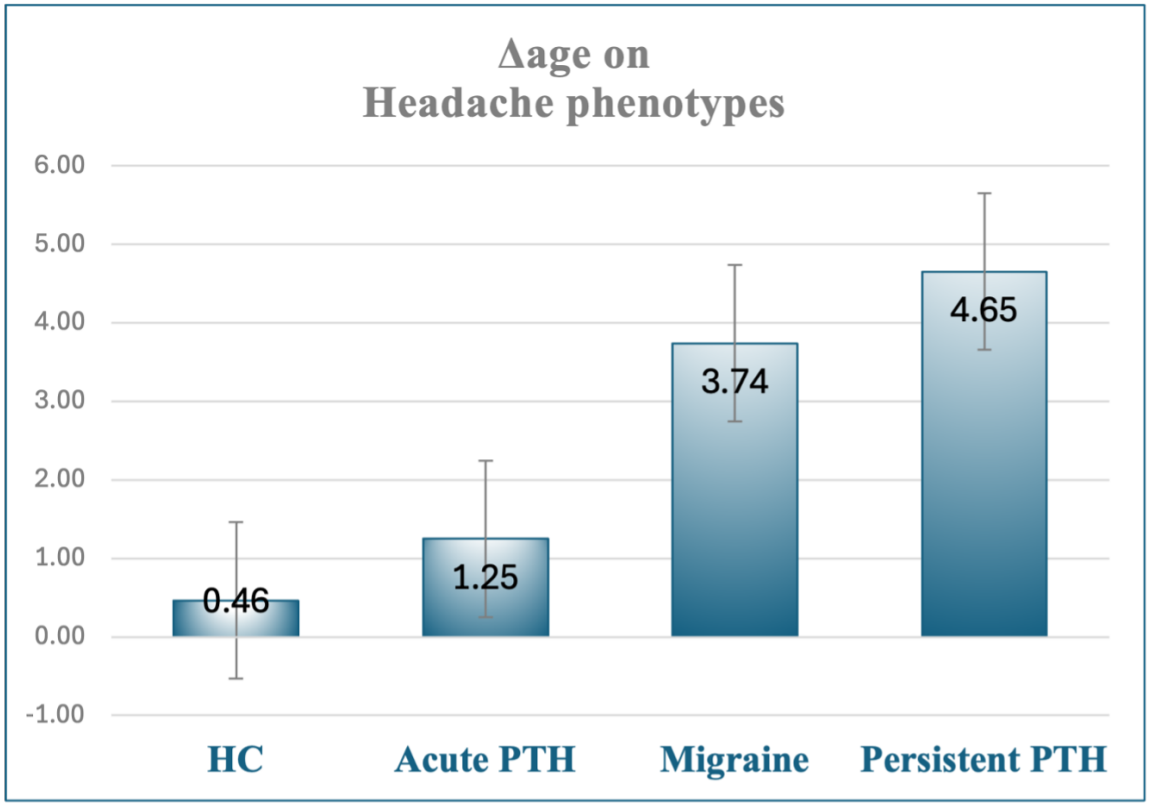

Figure 3.5: Brain Age Delta (Predicted–chronological Age) Across Different Headache Phenotypes on the Mayo Clinic Dataset

To further investigate whether the neuroimaging features learned by our brain age prediction model could differentiate between headache phenotypes, we extracted the 512-dimensional latent representations from the final layer of the ResNet-18 model for each participant. These latent features were then subjected to linear discrimi-

nant analysis (LDA) using the participants' headache phenotype labels (Migraine, Persistent PTH, Acute PTH, and Healthy Controls) as the basis for supervised dimensionality reduction and clustering.

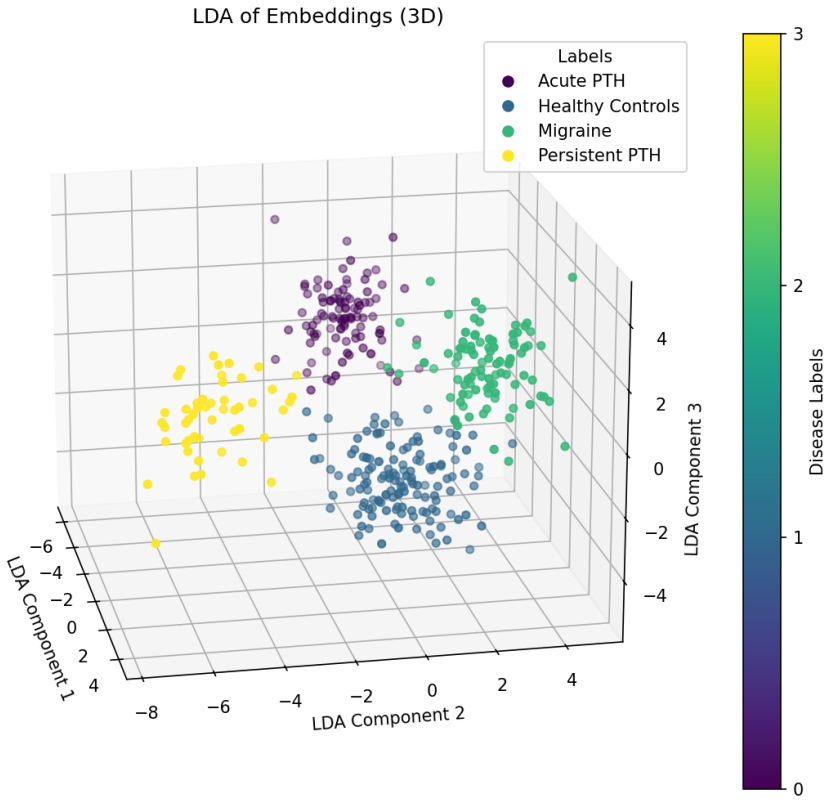

Figure 3.6: Visualization of Headache Phenotype Clustering in a 3D LDA-reduced Feature Space. The plot demonstrates three discernible clusters, effectively separating Migraine, persistent PTH, acute PTH, and healthy control groups based on model-extracted neuroimaging features. Each dot represents an individual subject.

The LDA revealed a clear separation of the headache phenotypes in the reduced dimensional space, indicating that the latent features encode distinct neurobiological signatures associated with different headache conditions. Visual inspection of the

LDA plot (see Fig.3.6) demonstrated the formation of four discernible clusters, suggesting that the model captures patterns that group subjects based on their headache classification. Quantitative evaluation of this clustering further substantiated the robustness of these groupings. A Silhouette Score of 0.602 indicated that clusters were well-separated and internally dense, with data points showing greater similarity to their assigned cluster than to neighboring clusters. This was complemented by a Calinski-Harabasz Index of 645.986, where a higher value signifies well-defined and distinct clusters. Furthermore, a Davies-Bouldin Index of 0.542, for which lower values are preferable, confirmed that the clusters were compact and effectively separated from one another. Collectively, these metrics provide strong quantitative support for the observation of distinct, meaningful groupings within the latent space that correspond to the predefined headache labels. Similar to our earlier research (Rahman Siddiquee *et al.*, 2023), we will investigate which brain regions were identified as most significant by the brain age prediction model within each headache phenotype.

3.5.5 Ablation studies

Distance Metric: We explored different L_k norm distance metrics for ORDER loss and found Manhattan distance best performing across all evaluation measures. L_k norm distance between two points x and y in high-dimensional space, given ($x, y \in R^d$), is be defined as:

$$L_k(x, y) = \sum_{i=1}^{d} [||x^i - y^i||^k]^{\frac{1}{k}} \tag{3.7}$$

Aggarwal *et al.* (2001) showed that Manhattan distance ($k = 1$) is a more suitable distance metric than Euclidean ($k = 2$) for high-dimensional data. They recommended using $k \leq 1$ to improve downstream classification performance. Later studies showed that fractional distance metrics, i.e., ($k < 1$), do not systematically address

Table 3.6: Ablation Studies on the Proposed Framework Components Evaluated by MAE, Ordinality, and Systematic Bias Scores. Here, k denotes different L_k-norm distance metrics defined in Eq.3.7

k	**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	<u>0.95</u>	2.19	-4.98
2	CE	<u>2.90</u>	0.10	<u>0.93</u>	<u>-3.04</u>
2	MSE	4.57	<u>0.95</u>	4.83	-4.13

.

the issue of the curse of dimensionality (Mirkes *et al.*, 2020) but should be a choice depending on the training data distribution. For the high-dimensional neuroimaging dataset used here, we found Manhattan distance more accurate at preserving ordinality and improving class separability compared to Euclidean or fractional distance metrics (see Tab.3.6).

ORDER loss with Classification vs. Regression: We experimented with the proposed ORDER loss using both classification and regression frameworks. As discussed in the paragraph above, since Euclidean and Manhattan distances performed significantly better than fractional distances, we explored regression models with our loss for $k = \{1, 2\}$ (Tab.3.6). Results show that distance-based regularization does not work well in regression models. Our model with cross-entropy loss and Manhattan distance-based ordinal regularization performed best across the three metrics.

3.6 Conclusion

This paper proposes a novel ordinal-distance regularization loss for robust brain age prediction using deep learning. We show that this loss in an ordinal classification framework outperforms regression-based brain age prediction methods, reduces systematic bias in predictions, and preserves ordinality in learned feature space. Improved performance is attributed to ordering information encoded in the model using ORDER loss and the ability of cross-entropy loss to learn high-entropy feature representations. The predicted BrainAGE from this model is a more reliable imaging biomarker for diagnosing AD and headache and predicting its early onset. We believe this framework can be generalized to other regression tasks to improve prediction and address the RTM effect, if present, which we aim to investigate further in future work.

Chapter 4

LOCALITY CONSTRAINED VECTOR QUANTIZATION

4.1 Motivation

The previous chapter focused on developing a novel deep learning method to aid the early detection of neurological disorders using high-dimensional imaging data such as MRI. Amid the rapid growth of modern medicine, it is no surprise that medical imaging has become crucial for accurate diagnosis by physicians and computer-aided tools (Hong *et al.*, 2020). However, there is a more practical problem that poses a great challenge to deploying these technological solutions for clinical applications- the sheer volume of this imaging data that places heavy demands on storage capacity (Elhadad *et al.*, 2024). For example, FreeSurfer[1] preprocessing requires $300-370MB$ of storage per image, which would multiply to terrabytes of data for data cohorts of size included in studies similar to ours (Tab.4.1). Additionally, while telemedicine is emerging as a key trend, limited transmission bandwidth poses a significant challenge. To address these issues, extensive research has focused on the compression of medical images (Liu *et al.*, 2017; Hussain *et al.*, 2018; Xin and Fan, 2021; Elhadad *et al.*, 2024) to make it more practical for storage or training deep learning models in low-resource settings.

Lossless image compression techniques such as Huffman coding (Huffman, 1952) and arithmetic coding (Langdon and Rissanen, 1981) can reconstruct identical original images. However, such methods can only obtain a low compression ratio around 1 to 4; a higher compression ratio is hard to obtain (Jiang *et al.*, 2012). Lossy compression

[1]https://surfer.nmr.mgh.harvard.edu/fswiki/DownloadAndInstall5.3

techniques leverage the limitations of human perception to achieve higher compression ratios, where they remove data that is less noticeable to the human eye (Spaulding *et al.*, 2002). Although for medical imaging analysis, physicians and scientists prefer to work with uncorrupted data, and the modest compression offered by lossless coding is often insufficient for either storage or downstream purposes. In these cases, a lossy compression method that preserves the diagnostic information is needed.

Vector Quantization (VQ) is an efficient information source coding method, and quantization is one of the core components in lossy image compression. The principle of it is constructing a vector based on several scalar data groups, so in theory, vector quantization coding is superior to scalar quantization coding. VQ (Gray, 1984; Van Den Oord *et al.*, 2017) is a foundational technique in signal processing for discretizing continuous data spaces through codebook-based representations. While effective for compression, traditional VQ methods lacked integration with deep learning frameworks for end-to-end representation learning. The advent of Variational Autoencoders (VAEs) introduced a probabilistic approach to learning continuous latent spaces via encoder-decoder architectures, enabling powerful data generation and reconstruction. However, standard VAEs suffered from limitations such as posterior collapse (Van Den Oord *et al.*, 2017), where latent variables become uninformative and inefficient latent utilization due to their continuous nature. These challenges motivated the development of vector quantized variational autoencoders (VQ-VAEs), which synergize VQ's discrete codebooks with VAEs' generative capabilities. By quantizing encoder outputs to the nearest codebook vector, VQ-VAE enforces a structured discrete latent space that inherently avoids posterior collapse while maintaining gradient flow via the straight-through estimator (Bengio *et al.*, 2013). This architecture achieves state-of-the-art performance in unsupervised phoneme discovery, speaker conversion, and high-fidelity image generation, while later extensions like VQ-VAE-2 introduced

hierarchical quantization for multi-scale feature learning.

Among many other challenges in training VQVAE architectures, codebook collapse is a significant challenge (Takida *et al.*, 2022; Huh *et al.*, 2023; Baykal *et al.*, 2024), occurring when only a small subset of the learned discrete codebook vectors are consistently selected during training, leaving many vectors underutilized or entirely "dead". This challenge is mainly introduced by the use of the straight-through-estimation (STE) technique, which effectively reduces the model's representational capacity, hindering its ability to capture the diversity of the input data and potentially degrading reconstruction quality. To address this, various techniques have been proposed, like-stochastic sampling, which introduces randomness to encourage exploration of less-frequently chosen codes (Takida *et al.*, 2022), repeated K-means, which aims to initialize a more robust codebook offline (Łańcucki *et al.*, 2020), and replacement policies, which actively monitor usage and reset inactive codes citep zeghidour2021soundstream. Recently, Finite Scalar Quantization (FSQ), proposed by Mentzer *et al.* (2023), circumvents the problem by adopting a different quantization strategy altogether, avoiding the traditional vector codebook structure. However, these approaches can have limitations: stochasticity might introduce noise into the codeword selection process to diversify usage during training. Further, this does not ensure consistent and meaningful selections, even though it encourages a more uniform codebook utilization with external regularization. Replacement policies can be heuristic fixes that don't necessarily improve the relevance of selected codes, whereas FSQ changes the quantization paradigm. Different from these, Baykal *et al.* (2024) recently introduced evidential uncertainty into the quantization process, effectively softening hard assignments and allowing for probabilistic selection of latent codes. Crucially, most of these methods do not explicitly leverage the learned geometric relationships between the codebook vectors themselves during the online lookup pro-

cess.

4.1.1 Role of distance metric in Quantization

More generally, the problem of VQ can be described as follows: Suppose that we have a set $\mathcal{H}$ of n samples $x_1, x_2, \ldots, x_n$ in $\mathbb{R}^d$ that we want to partition into exactly c disjoint subsets $\mathcal{H}_1, \ldots, \mathcal{H}_c$. Each subset is to represent a cluster, with samples in the same cluster being somehow more similar than samples in different clusters. Each member of the ith cluster is assigned to a reproduction vector v_i drawn from c reproduction codewords. Usually, Euclidean distance is used to find the most similar or closest vector in the codebook for this assignment (Van Den Oord *et al.*, 2017; Zeghidour *et al.*, 2021). The most important question in any design of a vector quantizer is how to determine "more similar", where similarity measures can be in the form of distance metrics. A general form of the distance between vectors x and m is

$$D = \|x - m\|_A^2 = (x - m)' A^{-1} (x - m)$$

where A is any positive definite $d \times d$ matrix. The most common idea of measuring the similarity between two patterns x and z is to measure the Euclidean distance $D = \|x - z\|_I^2$, where I is the identity matrix. The smaller the distance, the greater the similarity. However, choosing the Euclidean distance to measure dissimilarity assumes an isotropic feature space weighting. Consequently, the resulting clusters will be invariant to rigid-body motions of the data in the feature space (translation and rotation). Nevertheless, a simple scaling of the coordinate axes can result in a different grouping of data into clusters. In such cases, scaling becomes important, and Euclidean distance clustering is not appropriate. An alternative distance metric that depends on the distribution of the data itself is the Mahalanobis distance where in

equation 1 A becomes the covariance matrix of a pattern population, m is the mean vector, and x represents a variable pattern.

4.1.2 Locality in Quantization

The importance of locality (codebook neighborhood) is largely overlooked in modern deep learning based quantization approaches. In an earlier research work, Wang *et al.* (2010) emphasized the importance of preserving local geometric consistency in feature representation. They proposed locality-constrained linear coding (LLC), where instead of representing an input using a global linear combination over an entire codebook or a single codebook vector, it is more effective to encode inputs using only a local subset of basis vectors—those that are closest in feature space—thereby maintaining the manifold structure of the data. This property is particularly important in deep neural networks where slight distortions in intermediate representations can propagate and amplify through subsequent layers.

The effectiveness of locality constraints stems from their ability to reduce reconstruction error. While traditional vector quantization (VQ) represents each descriptor by a single basis in the codebook (resulting in large quantization errors), locality-constrained methods enable more accurate reconstruction by allowing multiple bases to contribute. Somewhat similar to that, additive quantization's (AQ) core idea is to represent a vector as a sum of multiple codewords, although each from a separate codebook, which allows finer-grained approximation than standard VQ (Babenko and Lempitsky, 2014). Recent research has extended locality-based approaches to address the challenges of neural network quantization (Lee *et al.*, 2023). Their framework combines a transformer encoder with a locality-aware decoder to effectively capture local information in data. This approach selectively aggregates latent tokens via cross-attention mechanisms for coordinate inputs and progressively

decodes with coarse-to-fine modulation through multiple frequency bandwidths.

The core insight of our method is that neural network weights exhibit locality patterns similar to those observed in natural data distributions. Our quantization method extends these locality principles to neural network compression. Similar to how LLC uses local coordinates rather than global combinations, it identifies and preserves neighborhood relationships in the weight space.

Key observations here:

1. Most quantization methods use Euclidean distance, which assumes uniform geometry-neglecting the internal geometry of the codebook—its distribution, correlation, and anisotropy. Mahalanobis distance adjusts for these variations by accounting for the covariance in the data.

2. Enforcing locality in quantization can better preserve semantic continuity using similar or neighboring codewords and discourage over-reliance on a few dominant codewords, potentially mitigating codebook collapse.

We propose a novel quantization method that integrates statistical locality into the codeword assignment process by leveraging the Mahalanobis distance metric. Unlike conventional vector quantization approaches that rely on Euclidean distance—which assumes isotropic feature space and ignores inter-feature correlations—our method computes a global covariance matrix over the codebook embeddings and uses its inverse to measure similarity between latent vectors and codewords. During quantization, each input embedding is compared to all codewords using this Mahalanobis metric, and the top-K closest codewords are selected. Rather than performing a hard assignment to a single nearest neighbor, we average the embeddings of these top-K

codewords to produce a soft, locality-preserving quantized representation. This strategy promotes robust codebook utilization, preserves local geometric structure in the latent space, and reduces the risk of codebook collapse without relying on stochastic sampling or heuristic replacement strategies. Please refer to Sec.4.3 for a detailed workflow of this method.

In summary, the contributions of this chapter include:

1. A novel Mahalanobis distance-based vector quantization scheme for VQ-VAE, which selects multiple nearest codebook vectors ($k = 25$) for each latent and thus incorporates latent distribution information during quantization;

2. An integrated framework applying this quantization to 3D brain MRI, demonstrating improved reconstruction fidelity and morphology preservation as compared to a state-of-the-art VQ-VAE baseline (Tudosiu *et al.*, 2024).

3. Evidence that the proposed compression enhances downstream task performance, illustrated by improved brain age prediction on public datasets relative to baseline compression.

To our knowledge, this is the first work to leverage Mahalanobis distance in the VQ-VAE quantization process for medical images, and to show that doing so can yield not only better compression but also tangible benefits for a clinically relevant prediction task.

4.2 Related Works

Generative deep learning models have been increasingly used in neuroimaging to address data scarcity and augment datasets (Khader *et al.*, 2023). Tudosiu *et al.* (2024) recently leveraged a VQ-VAE with a transformer to generate synthetic 3D

brain MRIs that preserved realistic anatomy. Their morphology-preserving generative model was validated with rigorous comparisons of tissue segmentations and volumetric measures between real and generated brains, confirming that key anatomical structures (e.g. ventricles, hippocampus) remained plausible. This represents a significant advance, as prior GAN-based approaches often produced visually realistic images but did not guarantee that the finer morphometric details were accurate. The success of VQ-VAE in capturing brain anatomy suggests it is a promising tool not only for generation but also for compression and representation learning in neuroimaging.

4.2.1 Addressing Codebook collapse in VQ

Codebook collapse is a critical failure mode in vector quantization systems, characterized by the underutilization of codebook elements where only a small subset actively participates in quantization (Zeghidour *et al.*, 2021). While the precise reason remains an open research question, empirical observations suggest that codebook vectors closer to encoder embeddings receive stronger gradient signals during training, leading to preferential updates causing distributional divergence between frequently used and inactive codes (Huh *et al.*, 2023). This phenomenon, termed *internal codebook covariate shift*, progressively marginalizes underused vectors, effectively reducing the model's representational capacity.

Heuristic approaches, such as codebook reset policies, replace inactive codes with randomly sampled vectors or convex combinations of active neighbors (Zeghidour *et al.*, 2021). While simple to implement, these methods require careful tuning of replacement timing and frequency to avoid destabilizing training dynamics. Takida *et al.* (2022) introduced stochastic quantization (SQ-VAE), where randomness is added to the code selection to encourage diversity and better codebook utilization. Others have employed the Gumbel-Softmax trick (Jang *et al.*, 2016) to create a differ-

entiable relaxation of quantization, allowing smoother gradient updates and reducing the codebook training instability.

Alternative quantization paradigms such as Additive Quantization (AQ) and Product Quantization (PQ) have been proposed to address limitations of hard nearest-neighbor assignment and improve codebook utilization. AQ represents an input vector as the sum of codewords drawn from multiple codebooks, allowing finer-grained approximations in high-dimensional or non-Euclidean latent spaces by minimizing error with respect to the inner product distance (Babenko and Lempitsky, 2014). This is particularly effective in domains like image retrieval and recommendation systems, where inner-product similarities dominate. In contrast, PQ divides the vector space into orthogonal subspaces and quantizes each independently (Jegou *et al.*, 2010), effectively reducing quantization error in Euclidean settings by leveraging the local structure of high-dimensional embeddings. PQ has been widely adopted in approximate nearest neighbor search due to its strong balance between compression efficiency and search accuracy. Both approaches highlight the importance of structured or distribution-aware quantization, aligning with recent efforts to address codebook underutilization in VQ-based models.

4.2.2 Locality and Soft Quantization

Locality-based approaches in vector quantization represent an important advancement that addresses key limitations of traditional VQ methods. While conventional vector quantization focuses primarily on minimizing overall quantization error across the entire feature space, locality-based methods explicitly consider the spatial or feature-space relationships between data points during the quantization process. This consideration of local structure is particularly important when working with high-dimensional data that often lies on a lower-dimensional manifold, where preserving

neighborhood relationships becomes crucial for maintaining the intrinsic geometric structure of the data. Locality-based quantization techniques attempt to preserve these local relationships, resulting in more meaningful and effective representations for tasks such as similarity search, pattern recognition, and data compression.

Earlier research methods, such as locality-constrained linear coding (LLC), directly align with locality-based quantization principles, particularly for image classification tasks. Unlike traditional vector quantization, which suffers from high quantization error due to its reliance on a single basis vector for reconstruction, LLC employs local-coordinate projections to preserve geometric relationships between descriptors. This approach aligns with the broader goal of locality-based quantization to maintain data manifold structures in high-dimensional spaces.

Soft quantization methods address codebook collapse by reformulating the discrete latent representation learning process through differentiable relaxations of hard assignment operations (Jang *et al.*, 2016; Agustsson *et al.*, 2017). These approaches fundamentally reconceptualize quantization as a probabilistic or stochastic mapping rather than a deterministic argmax operation, enabling gradient-based optimization while mitigating the tendency of models to underutilize codebook vectors. The theoretical underpinnings of these methods often draw from variational inference, where the quantization process is framed as an approximation of an intractable posterior distribution over discrete latent codes. Gumbel-Softmax quantization (Jang *et al.*, 2016), for instance, introduces a continuous relaxation of categorical distributions through temperature-controlled sampling from the Gumbel extreme value distribution. In contrast, soft-to-hard vector quantization frameworks (Agustsson *et al.*, 2017) adopt an annealing strategy that progressively sharpens the softmax weighting of codebook similarities. These methods parameterize the assignment probability as

$$\phi(z) = \text{softmax}\left(-\beta\|z - c_j\|^2\right),$$

where β serves as an inverse temperature parameter that is gradually increased during training. Theoretically, this creates a smooth optimization landscape during the early stages of training, allowing the codebook vectors to adapt more effectively to the distribution of encoder outputs. As β increases, the softmax distribution becomes sharper, leading to near-deterministic assignments and tighter clustering around specific codewords.

Despite the probabilistic relaxations of these soft quantization methods, they largely overlook the intrinsic geometric structure of the latent space in their distance metrics. Conventional approaches universally rely on Euclidean distance for measuring encoder-codebook alignment, implicitly assuming isotropic Gaussian distributions in the latent space. This simplification fails to account for the anisotropic covariance structure of real-world data distributions, where features exhibit varying scales and inter-dimensional correlations. The Euclidean paradigm forces codebook vectors to adapt to this mismatched geometry, which might create suboptimal quantization boundaries that cause collapse in regions of high data curvature. Moreover, these methods can lack adaptive metric learning mechanisms to dynamically adjust the similarity measure during training as the latent space evolves.

4.3 Proposed Framework

In this section, we introduce the locality constrained vector quantization (LCVQ) as an improved replacement of VQ that addresses the challenge of codebook collapse discussed in Sec.4.2.1.

4.3.1 Locality Constrained Vector Quantization

Zeghidour *et al.* (2021) proposed that the codebook should be initialized by the K-means centroids of the first batch in the training process. Building on that, the

idea behind the proposed method is to allow each cluster (codebook element) to have its own "shape" implied by the covariance matrix of the data within that cluster. During quantization, each codebook element would attract those data points that enhance the shape and reject those that do not fit. In other words, each vector x will be assigned to the cluster to which it best integrates, regardless of the Euclidean separation between that data point and that codebook element.

In the proposed quantization approach, we aim to find an optimal neighborhood of codebook elements to a given data point x that takes into account the variance and correlation among the features of codebook elements. This means that it effectively "whitens" the data, so the neighborhood around a point becomes an ellipse whose shape reflects the local data distribution. In practice, this allows Mahalanobis distance to capture local structure more accurately in anisotropic or correlated datasets, as it scales distances based on the underlying statistical variability rather than raw geometric separation.

In this approach, we compute the **Mahalanobis distance** between a set of query vectors and a codebook, where the Mahalanobis distance takes into account the covariance structure of the codebook vectors. It emphasizes local data structure by normalizing distances based on local covariance. It helps in selecting codebook elements that are locally relevant, which is key for methods like LLC (Wang *et al.*, 2010).

Let the codebook vectors be denoted as $\mathbf{w}_i \in \mathbb{R}^D$ for $i = 1, 2, \ldots, N$, where D is the dimensionality of each codebook vector and N is the size of the codebook. Similarly, let the input vectors (queries) be denoted as $\mathbf{x}_j \in \mathbb{R}^D$ for $j = 1, 2, \ldots, B$, where B is the batch size.

First, we compute the **covariance matrix** of the codebook vectors. To do so, we first subtract the mean of the codebook vectors from each vector, producing a

mean-centered codebook matrix $\mathbf{W}_{\text{centered}}$, where:

$$\mathbf{W}_{\text{centered}} = \mathbf{W} - \frac{1}{N}\sum_{i=1}^{N}\mathbf{w}_i$$

The covariance matrix Σ is then computed as:

$$\Sigma = \frac{1}{N-1}\mathbf{W}_{\text{centered}}^T\mathbf{W}_{\text{centered}}$$

where $\Sigma \in \mathbb{R}^{D\times D}$ is the covariance matrix, and the division by $N-1$ ensures unbiased estimation. To calculate the Mahalanobis distance between each input vector $\mathbf{x}_j$ and each codebook vector $\mathbf{w}_i$, we first compute the **difference** vector $\mathbf{d}_{ji} = \mathbf{x}_j - \mathbf{w}_i$, and then the Mahalanobis distance $D_{\text{Mahalanobis}}(\mathbf{x}_j, \mathbf{w}_i)$ is given by:

$$D_{\text{Mahalanobis}}(\mathbf{x}_j, \mathbf{w}_i) = \sqrt{\mathbf{d}_{ji}^T\Sigma^{-1}\mathbf{d}_{ji}}$$

where $\Sigma^{-1} = \text{pinv}(\Sigma)$ is the inverse covariance matrix (or pseudo-inverse in case of ill-conditioning) computed to ensure numerical stability. This formula computes the distance by normalizing the difference vector $\mathbf{d}_{ji}$ with respect to the covariance matrix Σ.

After calculating the Mahalanobis distances, the next step is to find the **top-K nearest neighbors** for each input vector. This is done by sorting the distances and selecting the smallest K values for each query $\mathbf{x}_j$. The indices of the top-K nearest codebook vectors are denoted as $\mathbf{I}_j$, where:

$$\mathbf{I}_j = \text{argsort}(D_{\text{Mahalanobis}}(\mathbf{x}_j, \mathbf{w}_i))[:K]$$

Finally, we **quantize** the input vector $\mathbf{x}_j$ by averaging the embeddings corresponding to the K nearest codebook vectors, which gives the quantized representation $\mathbf{q}_j$ of the input vector, where each $\mathbf{q}_j$ is the mean of the K nearest codebook vectors:

$$\mathbf{q}_j = \frac{1}{K} \sum_{i \in \mathbf{I}_j} \mathbf{w}_i$$

By using the Mahalanobis distance metric, our quantizer adapts to the distribution of the codebook itself, i.e., training data distribution (Zeghidour *et al.*, 2021), which is crucial in high-dimensional, correlated data. Additionally, selecting top-K codewords and averaging them avoids hard assignment, leading to smoother representations and better generalization (similar to soft quantization techniques). It is known that the Mahalanobis distance naturally penalizes dimensions with high variance and reduces the influence of noisy features. Instead of relying on randomness (like stochastic sampling), our method encourages utilization by leveraging feature covariance, leading to more semantically meaningful usage and hence doesn't need hand-designed policies for replacing codewords—codebook usage emerges naturally from the distance metric.

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

4.4 Experiment Design

Our method builds upon a standard VQ-VAE architecture for 3D MRI, introducing a modified quantization step based on Mahalanobis distance and k-nearest

codebook selection. In this section, we describe the experimental setup, including baseline line models used for performance comparison, model architecture, and training procedure (which closely follow prior VQ-VAE implementations except for our novel quantization), and the datasets used for model training and evaluation.

For baseline methods, we adopt two state-of-the-art VQ-based models as baselines: (1) the morphology-preserving VQ-VAE introduced by Tudosiu *et al.* (2024) and (2) an affine reparameterization approach proposed by Huh *et al.* (2023) to address the codebook collapse issue. Both models address critical limitations of vanilla VQ-VAE and are highly relevant for medical imaging tasks. The affine reparameterization method improves the gradient flow during codebook updates by replacing the non-differentiable nearest neighbor operation with a learnable affine projection. The morphology-preserving model, on the other hand, integrates perceptual and frequency-domain losses to ensure anatomical accuracy in brain MRI reconstructions, making it a strong benchmark for evaluating structural fidelity. In both baselines, we initialize the codebook using k-means clustering on encoder outputs, a common strategy shown to promote better initial codebook diversity and mitigate early codebook collapse (Zeghidour *et al.*, 2021). These baselines represent two complementary perspectives: optimization robustness and anatomical preservation, and together provide a rigorous standard against which to assess the benefits of our proposed locality-constrained quantization.

4.4.1 Datasets

Similar to Sec.3.4.1, we prepared a lifespan cohort of healthy controls (HC) to train the VQ-VAE model, aggregating MRI scans from public data sources: NACC, OASIS, ICBM, ABIDE, and IXI. However, we modified our pre-processing pipeline[2] along

[2]https://github.com/ysu001/PUP

with an improved quality control and ensuring no subject overlap across training, validation, and testing splits. Lifespan cohort consisted of $7,932$ T1w MRI scans aged 18 to 93 years (60.41 ± 18.97 years). Whereas, the discovery cohort (mix of healthy and unhealthy subjects) consisted of $9,913$ T1w MRI scans from the ADNI database aged 49 to 98 years (75.22 ± 7.45 years). Specifically, it consisted of 1,886 AD patients (age: 76.54 ± 7.47 years), 4,345 mild cognitive impairment (MCI) patients (age: 74.65 ± 7.71 years), and 3,450 healthy subjects (age: 75.43 ± 6.84 years). The lifespan cohort included here has a major overlap with the cohort used in the previous chapter (Sec.3.4.1). However, the discovery cohort compiled here has a significantly larger number of samples for an in-depth analysis on AD/MCI samples. The number of samples and age range per cohort are summarized in Tab.4.1.

All datasets were preprocessed and skull-stripped using FreeSurfer to rigidly register all data to a common MNI space. The final images were cropped, resulting in a head volume of $176 \times 308 \times 176$ voxels with a voxel size of $1mm^3$. During training, the images were min-max normalised to [0,1] and ran through an augmentation pipeline that consisted of random affine transformations, random contrast adjustments, random intensity shifts, and random noise injections. The data augmentation is based on MONAI2 version 0.5.3.

4.4.2 Implementation Details

Our proposed VQVAE-based method consists of an encoder E and D adopted from (Tudosiu *et al.*, 2024), which inherits from (Pinaya *et al.*, 2022b). The encoder consists of four downsamplings, with strided convolutions with stride 2 and kernel size 4, giving the downsampling factor $f = 24$. The encoder maps input volumes $X \in \mathbb{R}^{H \times W \times D}$ to a latent representation $Z \in \mathbb{R}^{h \times w \times d \times n_z}$, where $h = H/16$, $w = W/16$, $d = D/16$, and $n_z = 256$. This results in a compression factor of $4,096$. After the downsampling

Table 4.1: Age Range with Distribution and Number of Samples for Each Cohort. The Lifespan Cohort Comprises NACC, OASIS, ICBM, IXI, and ABIDE. Whereas the Discovery Cohort Consists of Samples from the ADNI Cohort.

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

layers, there are three residual blocks ($3 \times 3 \times 3$ Conv, ReLU, $1 \times 1 \times 1$ Conv, ReLU). The decoder mirrors the encoder and uses transposed convolutions with stride 2 and kernel size 4. All convolution layers have 256 kernels. The codebook size was 2048 while each element's size was 32. The encoder and decoder are interconnected with each other with a non-differentiable quantization step and a codebook C.

To maintain codebook stability and prevent collapse, we adopt the exponential moving average (EMA) update mechanism as described in Tudosiu *et al.* (2024). However, we initialize the codebook using k-means clustering on the encoder outputs, ensuring a diverse and representative set of code vectors, which is used as a baseline in recent VQ approaches (Zeghidour *et al.*, 2021; Huh *et al.*, 2023). Similar to (Tudosiu *et al.*, 2024), to ensure high-quality reconstructions and preserve anatomical details, we employ a composite loss function comprising: MSE between the input and reconstructed images, MSE between the amplitude spectra of the input and reconstructed images, computed via fast Fourier transform (FFT)-to enhance

the sharpness of reconstructions, learned perceptual image patch similarity (LPIPS) loss using a pre-trained AlexNet model (given the 3D nature of the data, LPIPS is computed slice-wise on randomly selected 2D slices from each anatomical plane-axial, sagittal, coronal), and adversarial loss using a PatchGAN discriminator is employed alongside a least squares GAN loss to encourage the generation of realistic intensity patterns.

We train the VQ-VAE using the Adam optimizer with a learning rate of 1×10^{-4} and a batch size of 6. The model is trained for 250 epochs, with early stopping based on validation loss. Data augmentation techniques, including random rotations and intensity scaling, are applied to improve generalization.

4.4.3 Evaluation

To quantify models' reconstruction performance, we use MSE, MAE, and multi-scale structural similarity index (MS-SSIM) in a pair-wise fashion between the generated synthetic samples as in earlier studies (Takida *et al.*, 2022; Huh *et al.*, 2023; Tudosiu *et al.*, 2024). MSE measures the average squared difference between the original and reconstructed voxel intensities. It is sensitive to large deviations and provides a direct measure of pixel-wise fidelity. Multi-Scale Structural Similarity Index (MS-SSIM) assesses perceptual similarity by evaluating luminance, contrast, and structural information across multiple spatial scales. Unlike MSE, which focuses on absolute differences, MS-SSIM is more aligned with human perception and is particularly useful for measuring the preservation of anatomical structures (Zhao *et al.*, 2016). We also report MAE, which measures the average of the absolute differences between reconstructed and ground truth voxel intensities, since it is less sensitive to outliers than MSE.

To evaluate codebook usage and detect potential collapse, we leverage the per-

plexity metric based on the entropy of the empirical code usage distribution across a batch, whose upper bound is equal to the number of codebook vectors. Given a distribution q over codebook entries, the perplexity is defined as:

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

Higher perplexity values indicate more uniform codebook usage, while lower values suggest that only a small subset of codebook entries are being selected (i.e., potential codebook collapse). Similar to Sec.3, MAE is used to evaluate performance on the brain age prediction task. It measures the average absolute difference between the predicted and chronological age of each subject. Lower MAE implies more accurate predictions and better downstream utility of the compressed images.

To rigorously evaluate the effectiveness of our proposed Locality-Constrained Vector Quantization (LCVQ) method, we design a comprehensive set of experiments across four model variants: (1) a standard VQ-VAE baseline (Tudosiu *et al.*, 2024) (2) VQ-VAE with affine reparameterization (AR) (Huh *et al.*, 2023), (3) VQ-VAE with our proposed LCVQ module, and (4) VQ-VAE incorporating both AR and LCVQ. This setup allows us to disentangle the individual and combined contributions of two complementary strategies: optimization-aware codebook learning (via AR) and structure-aware quantization (via LCVQ). All models share the same backbone architecture and training protocol to ensure fair comparison. Our rationale is to isolate the effect of LCVQ in enhancing local structure preservation and its compatibility with existing optimization improvements like affine reparameterization.

4.4.4 Downstream Task: Brain Age Prediction

To evaluate the utility of the compressed representations on downstream clinical tasks, we train a regression model on the quantized image representations to predict

brain age and compare it to a bigger model trained on raw imaging data for the same task. ResNet-18 was used as the model backbone in both scenarios, and MSE loss was used to update model weights. To compare the absolute performance at predicting brain age of healthy subjects, we assess MAE on predicted vs. actual age on the test of the lifespan cohort. And to further assess the generalization and disease detection capabilities of these models trained using different approaches, we report MAE and predicted brain age delta (predicted age - actual age) on discovery cohort consisting of AD, MCI and healthy controls subjects (see Sec.4.4.1 for more details).

4.5 Results and Analyses

We compare the image reconstruction accuracy and codebook utilization of our proposed quantization method to the baseline VQVAE framework for medical imaging data (Tudosiu *et al.*, 2024). The reconstruction accuracy at voxel-level is measured by MSE, MAE, MS-SSIM, and codebook utilization is measured by perplexity of latent variables (Sec.4.4.3). In Tab.4.2, VQVAE trained with LCVQ achieved best reconstruction performance across all metrics and significantly higher codebook usage with and without affine reparameterization. This suggests that incorporating statistical locality via Mahalanobis-based K-nearest quantization enhances the representational precision of latent embeddings. Notably, while the AR+LCVQ variant does not further improve reconstruction quality, it exhibits slightly higher codebook utilization, as measured by perplexity. This implies that the affine reparameterization mechanism may facilitate more effective exploration of the codebook space during training, particularly when combined with LCVQ's locality-aware soft assignments. Taken together, these results suggest that LCVQ alone provides the greatest benefit in terms of reconstruction, while the combination with affine reparameterization offers marginal gains in codebook efficiency without compromising performance. Fig.4.1

shows a qualitative comparison of raw imaging data before and after the proposed quantization step.

Methods	**MSE** ↓ (10^{-3})	**MS-SSIM** ↑	**Perplexity** ↑	**MAE**↓ (10^{-2})
VQVAE	2.01	0.9421	66.5	1.75
VQVAE + AR	2.12	0.9419	63.1	1.80
VQVAE + LCVQ (k=15)	**1.20**	**0.9684**	**368.9**	**1.43**
VQVAE + AR + LCVQ (k=15)	<u>1.52</u>	<u>0.9602</u>	<u>354.9</u>	<u>1.53</u>

Table 4.2: Comparison Between Various Quantization Approaches on Image Reconstruction Task. All methods use the same base architecture and codebooks $(1024, 32)$ were initialized with K-means across all implementations. AR: affine reparameterization, LCVQ: locality constrained vector quantization.

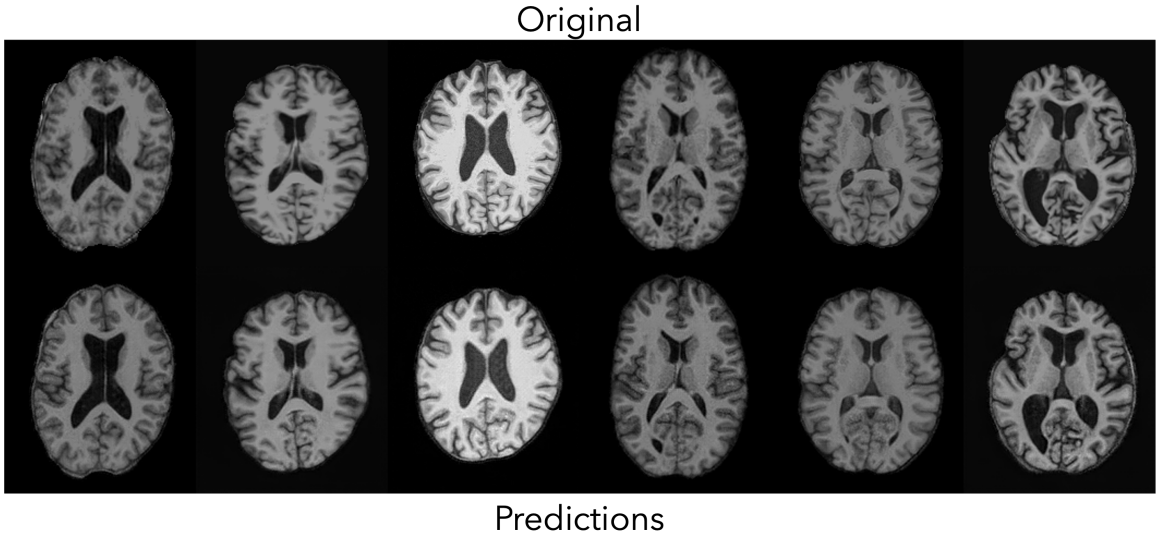

Figure 4.1: Visual Comparison of Original T1w MRI Scans and Reconstructed MRI Scans after Quantization Using the Proposed LCVQ Approach.

For the downstream brain age prediction, we use the trained VQVAE + LCVQ model and extract quantized representations of the same splits of imaging data from the lifespan cohort used to train the model. Tab.4.3 illustrates the trade-off between accuracy and efficiency in brain age prediction when using raw imaging data versus quantized latent representations generated by our LCVQ model. The model trained on raw 3D imaging achieves a marginally lower MAE of 5.10 years but requires approximately 4 days of training due to the high dimensionality and computational demands of volumetric data. In contrast, the model trained on compressed LCVQ representations completes training in under 15 minutes, with only a slight increase in MAE to 5.32 years. This result underscores the effectiveness of the quantization scheme: despite operating on a representation compressed by a factor of 4096, the model retains competitive performance while achieving a 99% reduction in training time. Moreover, the LCVQ encoder is trained offline and amortized across downstream tasks, making this pipeline particularly suitable for scenarios requiring rapid model deployment, such as federated learning, continual monitoring, or low-resource clinical settings (Xin and Fan, 2021). These findings demonstrate that this approach can potentially offer a practical and scalable alternative to full-resolution modeling without sacrificing substantial predictive accuracy.

More importantly, Tab.4.4 shows a comparative evaluation of brain age prediction performance on an independent discovery cohort (ADNI), highlighting prediction error (MAE) and aging-derived severity prediction (Brain Age Delta) across clinical subgroups. Despite being trained on highly compressed LCVQ representations (4096× reduction), the quantized model achieves comparable or improved accuracy relative to the raw-imaging baseline across all subgroups. Specifically, the quantized model attains a lower MAE in the MCI and HC groups (4.53 vs. 4.93; 4.29 vs. 4.58), and a modest reduction in overall MAE (4.73 vs. 5.00), demonstrating strong

Data	**MAE** (yrs) (on lifespan healthy test set)	**Training time**
Raw Imaging (w/o quantization)	5.10	∼4 days
Quantized Imaging (w/ LCVQ quantization)	5.32	∼15 minutes

Table 4.3: Comparison of Brain Age Prediction Performance on a Test of the Lifespan Cohort, with and Without Quantization. Here, training time measures the time taken solely to train the brain age prediction model.

generalization across cognitive states. Notably, while the MAE for AD subjects is slightly higher in the quantized model (6.01 vs. 5.93), this reflects a desirable property: increased deviation from chronological age in AD patients is consistent with accelerated neurodegeneration, and thus, higher brain age error in this group may indicate greater sensitivity to disease-related aging effects. Supporting this, the brain age delta for AD increased from 3.01 to 3.43 years, while remaining close to zero for MCI and HC. These findings affirm that LCVQ-based quantization preserves clinically meaningful information despite extreme compression and improves the model's ability to detect pathological aging signatures while benefiting from faster training and reduced storage.

4.5.1 Discussion

The proposed LCVQ introduces two main ideas: (i) using the Mahalanobis distance for codeword selection, and (ii) combining the top-K nearest codewords by averaging to represent an encoding, combining adaptive metric learning and soft as-

Data	**MAE** (yrs)				**Brain Age Delta** (yrs)			
	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 (w/ LCVQ quantization)	6.01	4.53	4.29	4.73	3.43	0.50	-1.56	0.36

Table 4.4: Comparison of Brain Age Prediction Performance on an Independent Discovery Cohort (ADNI), with and Without Quantization. ResNet-18 models were trained to predict brain age using MSE loss on raw imaging and quantized data. Brain Age delta measures the (predicted age - actual age), and MAE measures its absolute value on AD, MCI, and healthy control subjects. AD:alzheimer's disease, MCI:mild cognitive impairment, HC:healthy controls

signment.

Euclidean vs. Mahalanobis distance in Vector Quantization

Quantization with Euclidean distance (as in the standard VQ-VAE) works well if the true data distribution in latent space is roughly spherical in all directions, but can misbehave if the distribution is stretched or rotated. Quantization with Mahalanobis distance can reduce quantization error because codebook vectors can specialize along principal variance directions of the data instead of redundantly covering the same high-density region. Empirical studies support this: for example, Younis *et al.* (1996) introduced a VQ clustering algorithm with dynamically adjusted Mahalanobis distance and showed improved distortion performance on anisotropic data clusters. In modern deep learning parlance, researchers have begun to explore richer distance met-

rics for codebook learning. One approach is to treat codebook entries as Gaussian centroids with covariance. For instance, GMVQ (Yan *et al.*, 2024) models the latent space as a mixture of Gaussians rather than point centroids. The *distance* in this case is essentially a Mahalanobis distance to each Gaussian mean (with a shared or individual covariance). Such an approach was shown to mitigate codebook under-utilization and drastically improve reconstruction error. The improvement stems from aligning the quantization metric with the data distribution. Compared to that, LCVQ's top-K averaging is a distinguishing feature: rather than modeling the posterior as a categorical distribution (one code per latent) or mixture probability, LCVQ deterministically blends multiple code vectors. This resembles a local linear reconstruction of the latent, which is conceptually similar to locality-constrained coding in classical vision (Yang *et al.*, 2009).

This approach was also motivated by the intuition that latent features in MRI have anisotropic distributions, where some directions in feature space are more important to preserve precisely than others. By using the covariance-informed Mahalanobis metric, the proposed quantizer *learns* which differences are significant. Our reconstructions retained brain structure detail at a level where downstream analyses (like brain age prediction) remained almost as accurate as with the original data. This is promising for medical image compression as it suggests we do not necessarily have to trade off analytical usefulness for compactness.

Previous efforts to improve VQ-VAE focused on codebook learning tricks or multi-stage architectures, we instead revisited the fundamental distance measure for quantization. Our approach required only a modest change to the quantization step and a covariance calculation, yet yielded significant gains. In broader terms, it highlights that distribution-aware representations can enhance both reconstruction and inference. In addition to brain age prediction, we anticipate that other tasks, such as

disease classification (e.g., distinguishing AD vs healthy from compressed MRI) or segmentation, would similarly benefit. Essentially, the compressed latent is equally, if not more informative, and the reconstructions are more faithful, which should help any downstream model that relies on neuroimaging patterns.

4.5.2 Future work

Ablation on Fixed-K Selection

We performed an ablation study to evaluate the influence of different fixed K values used for top-K averaging in the LCVQ framework. Specifically, we tested $K = 5, 15, 25, 50, 75, 100$ (see Tab.4.5. Results demonstrated consistently stable performance across all tested values, indicating that the LCVQ method is robust with respect to variations in neighborhood size. Nevertheless, minor variations suggested a nuanced trade-off, where smaller K values risk limited neighborhood expressiveness, and larger K values may dilute locality, potentially reducing the effectiveness of Mahalanobis weighting. These observations motivated us to investigate adaptive strategies that dynamically select the optimal K value based on local data characteristics. In the following section, we introduce and evaluate such an adaptive method leveraging a chi-squared (χ^2) statistical criterion.

Adaptive Neighborhood Selection via Chi-Square Thresholding (Future Work

To explore more flexible neighborhood selection in our quantization scheme, we investigated an adaptive method that leverages the statistical properties of the Mahalanobis distance. The squared Mahalanobis distance between the query vector and

K **in VQVAE+LCVQ**	**MSE** ↓ (10^{-3})	**MS-SSIM** ↑	**Perplexity** ↑	**MAE**↓ (10^{-2})
K=5	1.40	0.9652	**391.9**	1.48
K=15	**1.20**	**0.9684**	<u>368.9</u>	<u>1.43</u>
K=25	<u>1.21</u>	0.9678	330.5	1.44
K=50	1.32	<u>0.9679</u>	329.2	**1.42**
K=75	1.32	0.9672	258.7	1.45
K=100	1.44	0.9627	295.3	1.55
Adaptive (K=15)	1.81	0.9544	316.94	2.09

Table 4.5: Ablation Study of the Top-k Parameter in the Lcvq Framework along with an Adaptive k Selection Method. All methods use the same base architecture and codebooks $(1024, 32)$ were initialized with K-means across all implementations.

each codebook element is defined as:

$$d_M^2(\mathbf{x}, \mathbf{c}_i) = (\mathbf{x} - \mathbf{c}_i)^\top \mathbf{\Sigma}^{-1} (\mathbf{x} - \mathbf{c}_i)$$

Under the assumption that query vectors are approximately sampled from a multivariate normal distribution $\mathcal{N}(\boldsymbol{\mu}, \mathbf{\Sigma})$, it is known that this squared Mahalanobis distance follows a chi-squared distribution with D degrees of freedom [3] :

$$d_M^2(\mathbf{x}, \boldsymbol{\mu}) \sim \chi_D^2$$

Based on this property, we define a statistical threshold τ^2 as the $(1 - \alpha)$ quantile of the chi-squared distribution:

$$\tau^2 = \chi_D^2(1 - \alpha)$$

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

This allows us to construct an adaptive neighborhood for each query vector:

$$\mathcal{N}_\tau(\mathbf{x}) = \left\{ \mathbf{c}_i \in \mathcal{C} \;\middle|\; d_M^2(\mathbf{x}, \mathbf{c}_i) \leq \tau^2 \right\}$$

Only codebook vectors whose Mahalanobis distance from the query is below this threshold are selected for quantization. This provides a statistically grounded and data-distribution-aware mechanism for codeword selection. To ensure robustness when no codewords fall below the threshold, we implement a fallback to selecting the top-K nearest codewords (e.g., $K = 15$). Preliminary experiments (see Tab.4.5 demonstrated that this adaptive method outperforms baseline methods but did not surpass the performance of the carefully selected fixed $K = 15$ setting. One possible explanation for this outcome is that the selected chi-squared threshold may either be overly restrictive, leading to smaller-than-optimal neighborhoods. Or occasionally overly inclusive, reducing the advantage of locality constraints. To fully assess the potential of this adaptive approach, further experimentation is required, particularly with varying upper limits on neighborhood size or more sophisticated statistical or learned gating mechanisms. This investigation remains an important direction for future work.

Evaluating LCVQ on Natural Image Datasets (Future Work)

We further evaluated our LCVQ method on natural image datasets CIFAR-10 and MNIST to assess its generalizability and impact on codebook utilization (see detailed quantitative results in Appendix E, Tab.E.1). Results demonstrated that while reconstruction accuracy remained similar to baseline and affine reparameterization methods, codebook utilization significantly increased. For instance, on CIFAR-10, codebook utilization rose dramatically from around 5% at baseline to approximately 79% with LCVQ. Similar improvements were observed on the MNIST dataset. These

findings suggest that LCVQ considerably improves the representational richness and diversity of the learned codebook vectors. Higher perplexity implies enhanced utilization of the codebook, enabling potentially more versatile embeddings for downstream tasks beyond direct reconstruction. Nevertheless, the absence of clear gains in reconstruction fidelity highlights a nuanced trade-off and motivates future work to investigate strategies that leverage improved codebook utilization to explicitly enhance reconstruction or downstream predictive performance.

Despite these promising initial findings, important questions remain regarding the optimal neighborhood size (K) and adaptive selection criteria in different data domains. Future work will investigate whether adaptive methods such as our chi-square-based thresholding or alternative data-driven selection mechanisms can similarly enhance quantization performance on general vision datasets. This exploration is essential to fully understand the versatility and scalability of the LCVQ framework beyond specialized medical imaging applications.

4.6 Conclusion

We show that locality constrained vector quantization (LCVQ), a structure-aware quantization framework, can improve representation quality by incorporating statistical locality into codeword selection. Results suggest that revisiting fundamental assumptions in quantization, such as the choice of distance metric and the rigidity of hard assignments, can lead to meaningful gains without increasing model complexity. More broadly, this highlights how integrating data geometry into discrete representation learning enables both efficient compression and better downstream performance. While our experiments focused on brain MRI, the core principles of locality and distribution-aware encoding could be applicable to multimodal fusion, federated learning, and low-resource computing. In future work, we aim to explore dynamic or

learned notions of locality, adapting LCVQ to non-Euclidean latent spaces, and investigate its role in tasks beyond compression, including out-of-distribution detection and zero-shot transfer.

Chapter 5

MRI-GUIDED PET SUPER-RESOLUTION

5.1 Motivation

The previous chapters introduced novel deep learning based techniques to effectively capture neuro-imaging-based aging patterns from T1-weighted MRI scans. While MRIs quantify macroscopic atrophy patterns through brain age prediction, amyloid PET captures the molecular pathology of amyloid-β (Aβ) plaques—a hallmark of AD that precedes structural changes by decades (Pontecorvo and Mintun, 2011). This duality mirrors the clinical progression of AD: amyloid accumulation begins 15–20 years before symptom onset, while brain age deviations reflect later-stage neurodegenerative processes. Due to the complementary roles of structural MRI and amyloid PET in early detection of AD, in this chapter, we now extend our investigation into the realm of multimodal imaging and how to best harness the complementary strengths of MRI and PET modalities for early diagnosis and monitoring of Alzheimer's disease.

Specifically, amyloid imaging is a crucial tool in the diagnosis and research of AD. It allows for the non-invasive detection of Aβ plaques in the brain, which is a core neuropathological feature of AD (Chapleau *et al.*, 2022). Detecting Aβ pathology at the earliest stages of AD, before the onset of clinical symptoms, is critical for understanding disease progression, developing intervention techniques, and potentially improving patient outcomes. However, accurate quantification of amyloid using positron emission tomography (PET) imaging is limited due to the low spatial resolution of PET scans (Thomas *et al.*, 2011), which is typically around 5 mm and

varies across scanner models and reconstruction algorithms. This causes the partial volume effect (PVE) (Hoffman *et al.*, 1979), resulting in a spill-out of signal from target gray matter regions and spill-in from surrounding areas. The magnitude of the PVE depends on the size of the target region relative to the spatial resolution of the scans. In the context of amyloid PET imaging, the size of the target regions varies across subjects and often decreases as the subject ages or with disease progression. Therefore, PVE reduces the accuracy, precision, and statistical power of quantitative amyloid PET measurements. Another well-recognized issue of amyloid PET imaging is harmonizing data acquired using different scanners, tracers, and analytical pipelines. To minimize the inter-scanner variabilities, a scanner-specific harmonization filter is often applied at the cost of further reduced spatial resolution (Joshi *et al.*, 2009). To minimize the variability of amyloid PET measurements from different analytical pipelines, acquisition protocols, and tracers, a Centiloid scale was defined to linearly transform a particular measurement to this scale (Klunk *et al.*, 2015). However, this Centiloid approach is designed for standardizing global measures and does not improve the between-measure agreements in terms of their shared variance (Chen *et al.*, 2024; Shah *et al.*, 2022a, 2023a). We hypothesize that effective methods for spatial resolution recovery will improve PET quantification and reduce inter-tracer variabilities in amyloid PET measurements, and in this research, we propose a deep learning approach to achieve the goal.

5.2 Related Works

5.2.1 Partial Volume Correction

Several partial volume correction (PVC) methods have been proposed in the literature to mitigate the PVE issue using anatomical information from MRI and CT

(Thomas *et al.*, 2011; Alessio and Kinahan, 2006; Baete *et al.*, 2004; Erlandsson *et al.*, 2016; Meltzer *et al.*, 1990; Müller-Gärtner *et al.*, 1992; Rousset *et al.*, 1998; Shidahara *et al.*, 2009). Different from that, Tohka and Reilhac (2008) showed that Richardson–Lucy, an iterative deconvolution-based method to recover spatial resolution in PET imaging and an alternative to MR-based approaches, offered comparable accuracy with reduced sensitivity to registration and segmentation errors. However, deconvolution-based correction methods are shown to amplify the image noise (Golla *et al.*, 2017). Different correction methods can also produce varying results, making standardization and comparison across studies challenging. Deep learning-based techniques (Matsubara *et al.*, 2022; Azimi *et al.*, 2024) have recently been explored to tackle some of these challenges. Deep models can better learn complex patterns of tissue heterogeneity and can perform image denoising, potentially addressing noise amplification issues (Azimi *et al.*, 2024). Deep models trained on diverse datasets may generalize better to different scanners and acquisition protocols (Matsubara *et al.*, 2022), potentially improving the standardization and consensus among multi-center studies.

5.2.2 Image Super-Resolution

Instead of focusing on partial volume correction directly, as the PVC methods reviewed above do, an alternative is image super-resolution (SR), which refers to the task of rendering a high-resolution image from its low-resolution counterpart. We contend that PVE may be tackled during the process of rendering high-resolution PET from low-resolution PET. SR is a well-studied research problem in computer vision and image processing (Moser *et al.*, 2023; Ahmad *et al.*, 2022). Use cases of SR span a broad spectrum, improving existing computer vision tasks (Dai *et al.*, 2016; Li *et al.*, 2022; Haris *et al.*, 2021) by improving image spatial resolution and perceptual

quality, improving surveillance (Rasti *et al.*, 2016), and enhancing diagnostic accuracy in medical research using different imaging modalities (Li *et al.*, 2021b; Greenspan, 2009; Isaac and Kulkarni, 2015). Traditional methods for image SR heavily rely on image statistics (Sun *et al.*, 2008; Chang *et al.*, 2004; Yang *et al.*, 2010), which has been shown to generate blurry and noisy artifacts in their high-resolution outputs (Moser *et al.*, 2023). With the advent of deep learning algorithms, several end-to-end architectures have been proposed where the models learn the mapping of low-resolution to high-resolution images through regression-based learning (Wang *et al.*, 2020). However, these methods fail to recover high-frequency details mainly because they learn an average mapping from the training dataset (due to L_1/L_2 loss functions), resulting in overly smooth model outputs and lacking spatial details (Ahmad *et al.*, 2022).

To address these limitations, generative models have been explored for SR in recent years. Generative models learn to transform a latent variable z with a tractable prior distribution to a learned data space. Generative adversarial networks (GANs), flow-based methods, and diffusion models are three common generative models used to generate synthetic data. They differ in their core approach: GANs (Goodfellow *et al.*, 2014) are trained in an adversarial setting with generator and discriminator networks, flow-based methods (Rezende and Mohamed, 2015) rely on invertible transformations to manipulate data distributions, while diffusion models (Ho *et al.*, 2020) iteratively add and then learn to remove noise to generate data. GANs are known to suffer from mode collapse (Li *et al.*, 2021a), resulting in unstable training and limiting the diversity of generated samples. Flow-based methods can impose topological constraints on the mapping between latent and data spaces, limiting their flexibility in modeling complex data distributions (Zhang and Chen, 2021). Except for longer sampling times, diffusion models have shown superior performance in generating high-

fidelity medical imaging datasets (Khader *et al.*, 2023; Müller-Franzes *et al.*, 2023; Pinaya *et al.*, 2022a). Unlike GAN priors, diffusion model priors can preserve more information to generate realistic data. Motivated by this, we propose an SR solution based on the diffusion model to improve PET quantification.

The success of any SR methods (e.g., GAN, diffusion) heavily relies on the quantity and quality of the training data. Synthetic data has been substantially useful in medical AI research to alleviate issues such as a lack of datasets, annotations, privacy concerns, and high acquisition costs (D'amico *et al.*, 2023; Rajotte *et al.*, 2022; Thambawita *et al.*, 2022). Data samples are typically artificially generated using domain knowledge or modeling techniques to mimic the characteristics and structure of real data without being directly derived from actual observations. It can be used to train AI models where target data is unavailable or scarce and provides a promising alternative to making AI models generalized to real-world datasets (Gao *et al.*, 2023; Wang *et al.*, 2022; Lyu *et al.*, 2022). These studies mainly focus on improving detection and segmentation from high-resolution imaging. However, its applicability in enhancing PET quantification remains unexplored. In this study,

1. We develop a new latent diffusion model for resolution recovery (**LDM-RR**) in PET imaging (Shah *et al.*, 2024a). Instead of training the diffusion model to minimize loss on the noise scale, we introduce a composite loss function with three terms: L_1/L_2, and MS-SSIM at the noise and image scale to improve MRI-guided reconstruction.

2. We develop a synthetic data generation pipeline to generate PET digital phantoms mimicking high-resolution PET scans for model training.

3. We evaluate the performance of our LDM-RR model in improving the statistical power of detecting longitudinal changes.

4. We evaluate the ability of the LDM-RR model to improve the agreement of amyloid PET measurements acquired using different tracers.

5.3 Proposed Framework

5.3.1 LDM-RR: MRI-guided PET Resolution Recovery

We use a latent diffusion model to generate synthetic high-resolution FBP scans given standard low-resolution FBP and matching MRI scans. Fig.5.1 & Fig.5.2 give an overview of the training process. Diffusion models have shown impressive results in generating 2D images (Saharia *et al.*, 2022). However, they are computationally demanding at the training and inference stages. Medical imaging modalities, such as MRI and PET, are more complex as they capture spatial information in 3D. Latent diffusion models operate at a lower-dimensional latent space by compressing useful information from these high-dimensional imaging data.

Our proposed LDM-RR is built upon a state-of-the-art LDM originally proposed to generate 3D brain MRIs (Pinaya *et al.*, 2022a). Specifically, it has a 2-stage training process and three different components: an encoder, a diffusion U-Net (Rombach *et al.*, 2022), and a decoder model. The encoder compresses high-dimensional data into a low-dimensional latent representation, diffusion U-Net converts simFBP to simDP in the latent space through a denoising process, and the decoder upsamples the low-dimensional simDP to its original image space. Trained models and implementation code are made available for reproducibility and further research[1] .

[1]https://github.com/jaygshah/LDM-RR

5.3.2 Compression Models

The goal of the compression model is to create a compressed representation of high-dimensional brain images that serve as the foundation for the subsequent diffusion model. We use an autoencoder (Pinaya *et al.*, 2022a) that compresses the 3D brain images into a lower-dimensional latent representation, capturing the perceptual representation of the original images while preserving essential features to reduce complexity. In the first stage, we train three modality-specific 3D autoencoder models separately for simFBP, simDP, and MRI (see Fig.5.1), each with a combination of L_1 loss, perceptual loss, a patch-based adversarial objective, and a KL regularization of the latent space (Pinaya *et al.*, 2022a). The input to the encoder is a 3D image with dimensions $256x256x256$, and we extract smaller sub-volumes of size $64x64x64$ to fit in GPU memory. The encoder maps these sub-volumes to a latent representation of size $16x16x16$. Once trained, latent representations from these encoders are used as inputs to the diffusion U-Net. See Sec.5.3.4 for more details on autoencoder model architectures and hyperparameters used.

5.3.3 Diffusion Model

Diffusion U-Net in LDMs performs denoising by iteratively predicting and removing noise in the latent space. Typically, they are trained to minimize the L_2 loss between predicted and actual noise (Pinaya *et al.*, 2022a; Rombach *et al.*, 2022). However, for super-resolution, we found that minimizing L_2 loss does not consistently guarantee the recovery of brain structure information in generated outputs (refer to Fig.5.6(B)). Prior studies have shown that using a mix of image restoration losses can produce high-fidelity images compared to single loss functions (Zhao *et al.*, 2016). L_2 regularization is sensitive to outliers and can introduce visual artifacts since it pe-

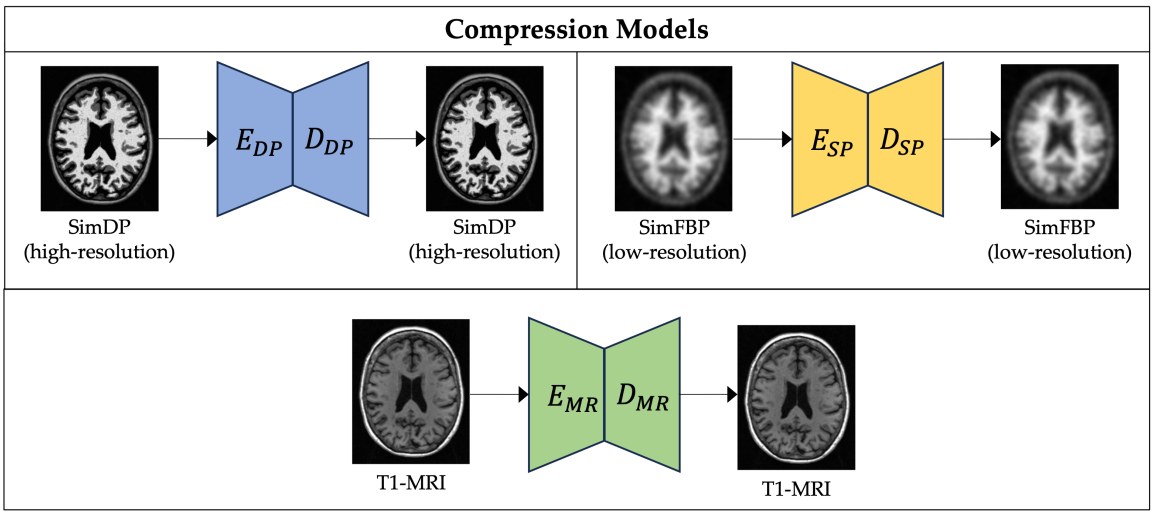

Figure 5.1: Training Three Modality-specific Autoencoder Models to Compress High-dimensional Simulated DP, Simulated FBP, and MRI Data into a Lower-dimensional Latent Representation.

nalizes high errors. L_1, on the other hand, is robust to outliers but suffers from non-differentiability at zero and slow training (Zou and Hastie, 2005). Moreover, Zhao *et al.* (2016) showed that for image restoration and SR, L_1 and L_2 penalties fail to capture structure information and proposed a multi-scale structural similarity index (MS-SSIM) metric. Voxel-level intensity has a high impact on PET quantification (López-González *et al.*, 2020).

Here, we hypothesize and show through the results that existing L_2 loss-based diffusion models (Ho *et al.*, 2020) fail to provide a clinically accurate reconstruction of PET scans. A weighted combination of L_1, L_2, and MS-SSIM losses, on the image and noise scales, can accurately generate a high-resolution image using MRI and simFBP.

To train diffusion models, a small amount of Gaussian noise is progressively added

to the data in T steps through a forward noise addition process, forming a Markov Chain (equations 1 and 2) Ho *et al.* (2020):

$$q(z_{1:T}|z_0) = \prod_{t=1}^{T} q(z_t|z_{t-1}) \tag{5.1}$$

$$q(z_t|z_{t-1}) = \mathcal{N}(z_t; \sqrt{1-\beta_t}z_{t-1}, \beta_t I) \tag{5.2}$$

Here, β_t is the fixed variance schedule and z_t follows a pure Gaussian noise distribution after many forward diffusion steps T ($T = 1000$ in our experiments). The diffusion U-Net learns the reverse diffusion process, i.e., denoising z_T to z_0 (equations 3 and 4), which can be formulated as (Ho *et al.*, 2020):

$$p_\theta(z_{0:T}) = p(z_T) \prod_{t=1}^{T} p_\theta(z_{t-1}|z_t) \tag{5.3}$$

$$p_\theta(z_{t-1}|z_t) = \mathcal{N}(z_{t-1}; \mu_\theta(z_t, t), \sigma_t^2 I) \tag{5.4}$$

$$z_t = \sqrt{\alpha_t}z_0 + \sqrt{1-\alpha_t}\epsilon \tag{5.5}$$

where μ_θ represents the denoising neural network (diffusion U-Net) and $\sigma_t^2 = \frac{1-\bar{\alpha}_{t-1}}{1-\bar{\alpha}_t}\beta_t$. Traditionally, diffusion models are trained to predict the added noise in the forward diffusion process by minimizing the L_2 loss between the predicted noise ($\hat{\epsilon}$) and the added noise (ϵ), formulated as (Rombach *et al.*, 2022):

$$L_\theta = \mathbb{E}_{x,\epsilon\sim\mathcal{N}(0,1),t}\left[||\epsilon - \epsilon_\theta(z_t, t)||_2^2\right] \tag{5.6}$$

Further, we can estimate the noise-free latent vector using the predicted noise ($\hat{\epsilon}$) from the diffusion model using Equation 5 from Ho *et al.* (2020) as:

$$\hat{z}_0 = \frac{z_t - \sqrt{1-\alpha_t}\hat{\epsilon}}{\sqrt{\alpha_t}} \tag{5.7}$$

Zhao *et al.* (2016) observed that image reconstruction performance can be improved by adding perceptual image metrics such as MS-SSIM in a network's loss function. This approach captures structural details at multiple scales while maintaining voxel-level accuracy. While this holds true for current encoder-decoder architectures, to the best of our knowledge, it has not yet been investigated for denoising diffusion networks in latent space.

Since our goal is to fuse structural information from T1-MRI to guide the reconstruction, we modify LDM's vanilla loss function (L_θ) on the noise scale to a weighted combination of L2 and MS-SSIM loss on the image scale as:

$$\text{loss}_1 = (1-\alpha)L_2(z_0, \hat{z}_0) + \alpha\text{MS-SSIM}(z_0, \hat{z}_0) \tag{5.8}$$

Here $\alpha = 0.8$ (Zhang *et al.*, 2023b) in Equation 8 is an empirically set hyperparameter. We explored $\alpha = [0.2, 0.5, 0.8]$. However, $alpha = 0.8$ resulted in the best performance in reconstructing simDP on the simulated dataset's validation set.

While L_2 allows easier optimization in diffusion training due to its convergence properties, it is known to produce an averaging effect, which forces the model to predict values closer to the mean of the training data (Zhao *et al.*, 2016). We argue that using only L_2 loss can help preserve whole image-level properties but may also produce inaccurate estimates at the voxel level. To this end, we propose an L_1 loss at the noise scale to ensure voxel-level details are preserved in the denoising process:

$$\text{loss}_2 = L_1(\epsilon, \hat{\epsilon}) = |\epsilon - \hat{\epsilon}| \tag{5.9}$$

A combined loss function using the two loss terms from Equations 8 and 9 is

defined as:

$$\text{loss}_{\text{combined}} = L_1(\epsilon, \hat{\epsilon}) + (1-\alpha)L_2(z_0, \hat{z}_0) + \alpha\text{MS-SSIM}(z_0, \hat{z}_0) \tag{5.10}$$

This combined loss is equivalent to (see Appendix B):

$$\text{loss}_{\text{combined}} = L_1(\epsilon, \hat{\epsilon}) + \gamma(1-\alpha)L_2(\epsilon, \hat{\epsilon}) + \alpha\text{MS-SSIM}(z_0, \hat{z}_0) \tag{5.11}$$

By minimizing loss on image (z) and noise (ϵ) scales, the LDM-RR model learns to reduce the disparity between the reconstructed high-resolution PET image and the target digital phantom while preserving image-level and voxel-level structural details. The latter (voxel-level details) may play an important role in correcting partial volume effects.

Fig.5.2 illustrates the second stage of training, where only the diffusion U-Net is trained, whereas the encoder and decoder model parameters are kept frozen. The input to the U-Net is a concatenation of the noisy latent representation of simDP ($z_{T(DP)}$) and the conditioning of matching MRI (z_{MR}) and simFBP (z_{SP}) latent representations. The model's predicted noise ($\hat{\epsilon}$) can be used to estimate $\hat{z}_{0(DP)}$ and calculate the combined-loss (Equation 11), which is used to update the diffusion U-Net parameters in each training epoch.

5.3.4 Implementation Details

Compression Models: Our compression models had a 3-layer 3D AutoencoderKL architecture inspired from Pinaya *et al.* (2022a). We removed all attention layers except at the model's last level. The models had 32 base channels, with a channel multiplier of $[1, 2, 2]$ and only one residual block per level. Our latent space had a dimensionality of 161616 with 3 latent channels. We trained our model over 80 epochs

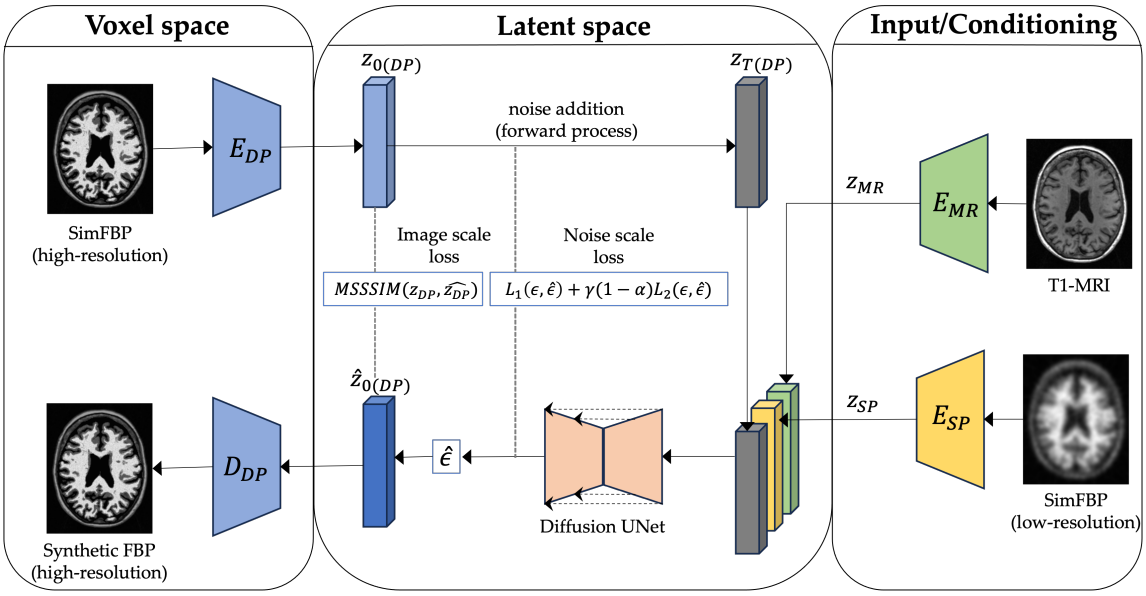

Figure 5.2: Proposed LDM-RR Framework's Training Process for PET Super-resolution. LDM is conditioned on latent representations of T1-MRI and simFBP and uses a combination of image and noise scale losses to generate corresponding high-resolution simDP.

with a minibatch of 60, with an Adam optimizer and a base learning rate of 0.0001. We used a patch-based discriminator in our adversarial loss with 32 base channels and a learning rate of 0.0001.

Diffusion Model: Our diffusion model uses a U-net architecture, with 32 base channels, a channel multiplier of $[1, 2, 2]$ and one residual block per level. The input channels are 9 (3 channels each for simFBP, simDP, and MRI latents). We used Adam optimizer with a base learning rate of 0.0001. In training, we used a DDPM scheduler with 1000 timesteps, with a linear variance schedule, from 0.0015 to 0.0195. At inference, we use a DDIM scheduler (Song *et al.*, 2020a) with 250 timesteps, which is a faster sampling method compared to DDPM while maintaining output quality.

Computational resources: All models were trained on a single NVIDIA *A*100 80

GB GPU. The inference time to generate a synthetic SR FBP was ∼10 minutes on a GPU.

5.4 Experiment Design

5.4.1 Datasets and Simulation Procedure

Imaging data from three different cohorts were used in this study to enable our experiments: 1) the Alzheimer's Disease Neuroimaging Initiative (ADNI) cohort[2] (Weiner *et al.*, 2017), 2) the Open Access Series of Imaging Studies-3 (OASIS-3) (LaMontagne *et al.*, 2019), and 3) the Centiloid Project florbetapir (FBP) calibration dataset (Navitsky *et al.*, 2018)[3] . A subset of the ADNI database containing MRI scans was utilized for data simulation to train the diffusion model, while another subset with FBP scans (Tab.5.1) was employed to evaluate the model's performance in detecting longitudinal changes. Additionally, paired FBP-PiB imaging data from the OASIS-3 and Centiloid databases (Tab.5.1) were used to further assess the model's performance in cross-tracer harmonization. Details regarding data selection and simulation are provided in the subsequent sections. The ADNI was launched in 2003 as a public-private partnership led by Principal Investigator Michael W. Weiner, MD. The original goal of ADNI was to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of mild cognitive impairment (MCI) and early Alzheimer's disease (AD). The current goals include validating biomarkers for clinical trials, improving the generalizability of ADNI data by increasing diversity in the participant cohort, and providing data concerning the diagnosis and progression of Alzheimer's disease to the scientific community.

[2]adni.loni.usc.edu

[3]http://www.gaain.org/centiloid-project

Data to Train the Diffusion Model

We utilized 3,376 MRI scans randomly selected from the ADNI database to generate simulated high-resolution digital phantoms (simDP) and simulated florbetapir (simFBP) using an MR-based procedure as previously described (Su *et al.*, 2015) and mimics the distribution of florbetapir (FBP) uptake in participants across a wide range of amyloid burden and clinical status, and the noise and spatial resolution characteristics of typical PET images. The specific set of MRIs selected as the basis for simulation does not have a major impact on subsequent experiments and, therefore, was not described in detail. The size of the dataset captures the overall distribution and variability of structural brain differences in the elderly population without losing generalizability. A detailed description of the simulation procedure is discussed in section 2.2 below. From this simulated dataset, 3,038 samples were used to train, and 338 samples were used to validate our LDM-RR model's performance.

Data for Evaluating Longitudinal Power

To evaluate LDM-RR's ability to improve statistical power to detect longitudinal changes in amyloid, we selected 167 ADNI participants with a mean age of 74.1 years (SD=6.8), who are amyloid positive at baseline using a Centiloid cutoff of 20 (Royse *et al.*, 2021) and have two-year follow-up (2.0 ± 0.06 years interval) FBP scans. The choice of these participants is to ensure they are on a trajectory to accumulate amyloid during the study period, i.e., having a positive expected rate of amyloid accumulation. Additional demographic information of this cohort is summarized in Tab.5.1.

Data for Evaluating Harmonization Performance

From the OASIS-3 database (LaMontagne *et al.*, 2019), we identified 113 pairs of FBP-PiB scans with a mean age of 68.1 years (SD=8.7), and similarly, 46 pairs

Table 5.1: Summary of Demographic Information of the Three Cohorts Included in This Study.

Cohort	**ADNI**	**OASIS-3** (yrs)	**Centiloid**
Sample Count	334 (167 baseline-followup FBPs)	113 (FBP-PIB pairs)	46 (FBP-PIB pairs)
Age (SD) years	75.1 (6.9)	68.1 (8.7)	58.4 (21.0)
Education (SD) years	16.1 (2.7)	15.8 (2.6)	N/A
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 out of 46 unknown]
PET interval (SD) years	2.0 (0.06)	N/A	15 N/A

from the Centiloid project (Navitsky *et al.*, 2018)[4] with a mean age of 58.4 years (SD=21.0). Refer to Tab.5.1 for demographic information of these two cohorts. Studies for the cohorts included here were approved by their corresponding institutional review boards, and written informed consent was obtained for each participant.

[4]http://www.gaain.org/centiloid-project

5.4.2 Image Analysis and Simulation

FreeSurfer v7.3 (Fischl, 2012) (Martinos Center for Biomedical Imaging, Charlestown, Massachusetts, USA)[5] was used to automatically segment T1-weighted MRIs to define the anatomical regions of interest (as defined in the wmparc.mgz file). PET images were processed using a FreeSurfer-dependent pipeline that included resolution harmonization filtering, inter-frame motion correction, target frame summation, PET-to-MR registration, and regional and voxel-wise SUVR calculation (Su *et al.*, 2015, 2013). A mean cortical SUVR (MCSUVR) was calculated as the summary measure of amyloid burden and used to evaluate longitudinal and harmonization performance (Su *et al.*, 2013). For comparison purposes, a Richardson-Lucy algorithm was adopted for resolution recovery (RL-RR) through iterative deconvolution (Lucy, 1974; Biggs and Andrews, 1997). In our experiment, the MATLAB (The Mathworks, Inc., v2021a) function: deconvlucy[6] was called with 20 iterations and an 8 mm full-width-half-max (FWHM) Gaussian kernel to generate the deconvolved high-resolution PET images and the corresponding MCSUVR estimation. Similar to previously described by Su *et al.* (2015), the simulation of high-resolution digital phantom (DP) and PET images (simFBP) was performed using segmented MRI as the input. For DP generation, each voxel was assigned a specific intensity value according to tissue type-specific distributions observed from actual FBP SUVR images across the aging and AD spectrum. For non-brain voxels, i.e., those not defined in the wmparc.mgz file, the voxel intensity was assigned by randomly scaling the normalized T1-MRI images to simulate moderate non-brain uptake. To generate simFBP images, the DP was smoothed and projected to the sinogram space, adding Poisson noise, and reconstructed back to the image space. We generated the simulation with a range of noise

[5]https://surfer.nmr.mgh.harvard.edu/ fswiki

[6]https://www.mathworks.com/help/images/ref/deconvlucy.html

levels as seen in real-world PET scans with a noise equivalent count rate (NECR) of 7526 kcps (Vandendriessche *et al.*, 2019; Reynés-Llompart *et al.*, 2017). The target resolution of the simFBP data is 8mm in FWHM, approximating the resolution of standardized PET data from ADNI (Joshi *et al.*, 2009). Fig.5.3 shows a visual example of a simulated digital phantom (B) and PET image (C) matching a T1-MRI image (A).

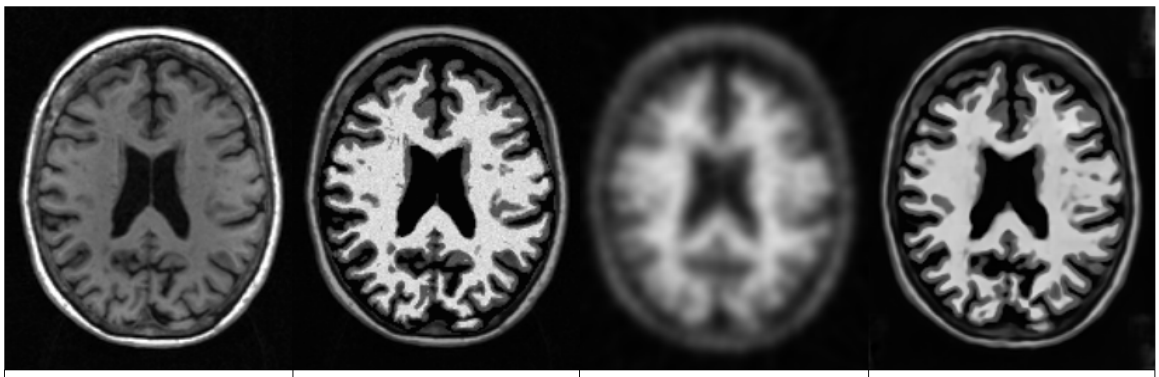

Figure 5.3: Visualization of a Simulated Digital Phantom (SimDP) and Simulated FBP (SimFBP) from the Data Simulation Pipeline Using T1-MRI and the LDM-RR Generated Synthetic Super-resolution FBP.

5.4.3 Statistical Analysis

5.4.4 Simulated Data Analysis

To evaluate our LDM-RR's actual performance at generating a high-resolution FBP scan, we compare the mean recovery coefficient (RC), measured as the ratio of synthetic high-resolution PET MCSUVR to that of ground-truth from simulated DP, and variability measured by the standard error (SE). A value closer to one indicates higher reconstruction performance. We compare our method to RL-RR, traditional

LDM with L_2 noise loss, LDM with L_2 noise and L_1 image loss, and simulated FBP scans without any correction.

5.4.5 Longitudinal Analysis

To evaluate the longitudinal performance of the LDM-RR, for each of the 167 participants, the annualized rate of amyloid accumulation was calculated by dividing the MCSUVR change from baseline to the follow-up visit by the follow-up interval, commonly used in longitudinal PET studies (Bollack *et al.*, 2024). The imaging data were acquired using standard protocols, and harmonization procedures were performed to reduce variability. The mean and standard deviation of the annualized rate of change were evaluated for each analysis method, i.e., raw measurement, with RL-RR, and with the LDM-RR. One-sample t-test (one-tail) was used to determine whether the annualized rate was significantly greater than zero. A smaller p-value is interpreted as having greater power to detect the longitudinal accumulation of amyloid burden. To further compare the statistical power of different techniques in a longitudinal setting, we estimated the number of 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 similar to previous studies (Su *et al.*, 2016; Chen *et al.*, 2015). A smaller estimated sample size (SS) indicates greater statistical power.

5.4.6 Cross-tracer Analysis

In the cross-tracer analysis, we evaluated the impact of RL-RR vs. LDM-RR on the agreement of PET-derived global amyloid burden, i.e., MCSUVR, using paired FBP-PIB data from OASIS-3 and the Centiloid project. The imaging data were also collected following standard protocols and underwent harmonization processes

to minimize variability, similar to the ADNI study. In our experiment, the raw PIB MCSUVR was used as the reference amyloid burden measurement, and we evaluated whether the corrected FBP MCSUVR is more strongly correlated with PIB MCSUVR using Steiger's test. We also test whether the LDM-RR corrected FBP MCSUVR is more strongly associated with PIB MCSUVR than the RL method.

5.5 Results and Analyses

5.5.1 Qualitative Assessments

Fig.5.4 showcases corrected FBP scans using RL (C) and LDM-RR (D) methods in comparison to the real FBP scan without any correction (A). The proposed LDM-RR model-generated synthetic FBP image has an improved spatial resolution, with a similar level of anatomical details matching T1-MRI (Fig.5.4 (B)). A similar example of LDM-RR applied to simulated data is shown in Fig.5.3 (D). Although RL-RR does not require an MRI, it generated noisier images and was not able to fully recover the high-resolution details (Fig.5.4 (C)). The LDM-RR method leverages the high-resolution structural information from MRI to guide the super-resolution process, resulting in PET images with reduced partial volume effects.

5.5.2 Evaluation on Simulated Data

A visual example of model-generated synthetic FBP from the test set of simulated data is shown in Fig.5.3 (D). The mean RC from different diffusion models compared to the RL-RR method and without any correction is shown in Fig.5.5. Our proposed LDM-RR model was able to better reconstruct target simDP (0.96, SE=0.004, $p < 0.001$) compared to RL-based correction (0.82, SE=0.005, $p < 0.001$) and without any corrections (0.76, SE=0.008, $p < 0.001$). It also performs significantly better

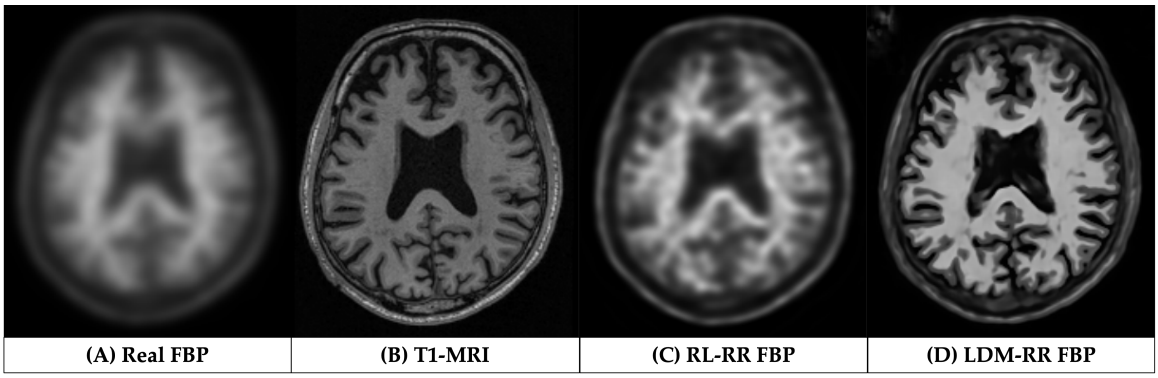

Figure 5.4: Visual Comparison of Generated FBP Scans Using RL-RR and Our LDM-RR to Real FBP and T1-MRI for a Sample from the OASIS-3 Cohort.

compared to a typical LDM architecture (Pinaya *et al.*, 2022a) for super-resolution (1.32, SE=0.08, $p < 0.001$) and other combinations of noise and image scale loss (1.58, SE=0.09, $p < 0.001$). The improvement in recovery coefficient with LDM-RR was also statistically significant ($p < 0.001$) compared to other LDM methods.

5.5.3 Evaluation on Real Longitudinal Amyloid PET Data

Tab.5.2 shows a comparison of statistical power to detect amyloid accumulation in longitudinal studies using the LDM and RL methods for resolution recovery in comparison to measurements from raw FBPs without any correction. The annualized rate of amyloid accumulation was significantly greater than zero for all three methods ($p < 0.0001$), suggesting an increase in brain amyloid burden over time as expected. Notice that the annualized rate of amyloid accumulation is the unit of SUVR/year, which is specific to the underlying quantification methods and not directly comparable. Numerically, the p-value was smallest using the LDM-RR and largest without any correction, suggesting our proposed method had the best power in detecting longitudinal changes. Additionally, the LDM-RR required a much smaller sample size estimate to detect a 25% reduction in the amyloid accumulation rate due to treatment

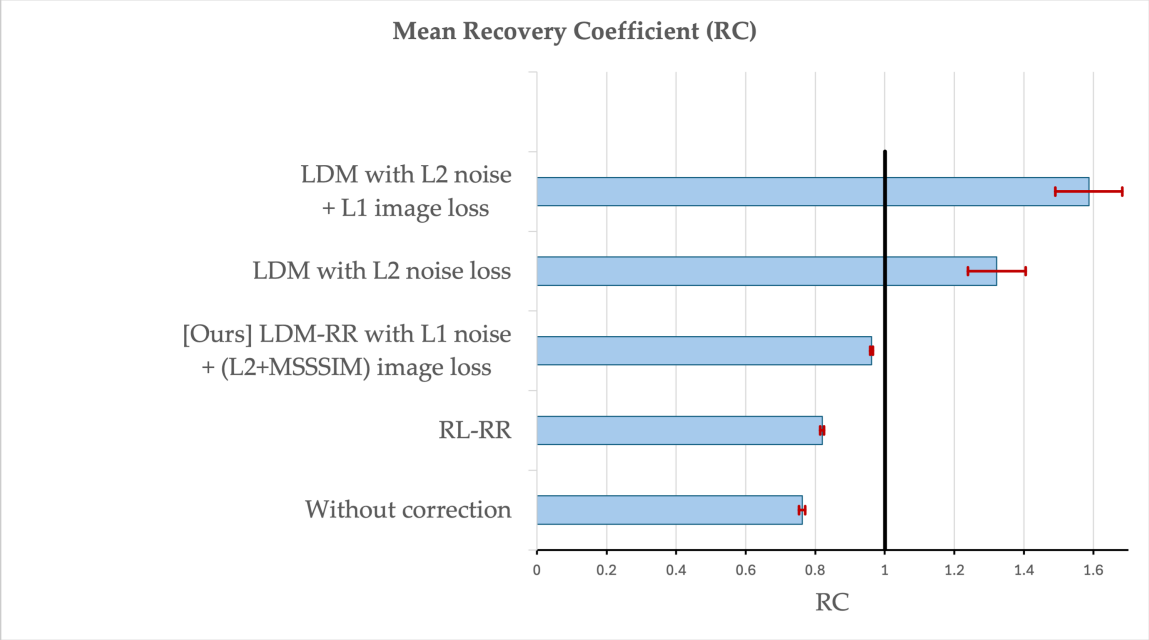

Figure 5.5: Comparison of Mean Recovery Coefficient (RC) Using Different Methods on a Held-out Test of 338 Samples Randomly Selected from the Simulated Dataset. A Value Closer to 1 Indicates High Performance.

in hypothetical anti-amyloid trials. To put the sample size estimation into context, the recently completed TRAILBLAZER-ALZ2 randomized trial of donanemab (Sims *et al.*, 2023) recruited 860 participants for the treatment arm and 876 for the placebo arm. The donanemab was able to reduce the patient's brain amyloid burden by over 80%. Therefore, our assumed treatment effect is considerably more moderate.

5.5.4 Evaluation on Real Cross-tracer Amyloid PET Data

The performance of LDM-RR and RL-RR methods at harmonizing cross-tracer global amyloid burden measurements is shown in Tab.5.3. Agreement of MCSUVR measurements between tracers significantly improved ($p < 0.001$), as shown by a higher correlation for both LDM and RL-based corrections to the reference measure. The improvements in LDM-RR-based partial volume corrections compared to RL were

Table 5.2: Statistical Power in Detecting Longitudinal Changes Measured by Mean, Standard Deviation, and P-value of an Annualized Rate of Amyloid Accumulation and Sample Size (SS) per Arm Estimates to Detect a 25% Reduction in Amyloid Accumulation Rate Due to Treatment (80% Power and a Two-tailed Type-I Error of P=0.05)

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**

also statistically significant ($p = 0.042$). The results provide evidence supporting that the inter-tracer variability in PET-derived amyloid burden measurement is at least in part related to the partial volume effect associated with lower spatial resolution and the contaminated signal from the target regions of interest. While the numerical improvement in terms of the Pearson correlation is small, they were statistically significant, suggesting that improving image resolution can be one of the strategies for reducing the variability.

5.5.5 Discussion

It is well recognized that PET imaging has inherently low spatial resolution, which leads to PVE, resulting in loss of sensitivity to focal changes and compromised accuracy due to signal contamination (Rousset *et al.*, 1998; Aston *et al.*, 2002). Many different techniques have been developed to account for PVE and improve quantitative accuracy (Rousset *et al.*, 1998; Tohka and Reilhac, 2008; Meltzer *et al.*, 1999;

Table 5.3: 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.

Method	**Pearson Correlation**	**Steiger's p-value**
Without 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)

Teo *et al.*, 2007). In the context of PET neuroimaging, commonly adopted techniques are often region-based and do not provide high-resolution images (Su *et al.*, 2015; Frouin *et al.*, 2002; Rousset *et al.*, 2008). Voxel-wise approaches do exist (Shidahara *et al.*, 2009; Tohka and Reilhac, 2008; Boussion *et al.*, 2006); however, they are known to amplify noise while having limited ability to recover the full spatial resolution or have gone through limited evaluation for targeted applications (Thomas *et al.*, 2016; Gonzalez-Escamilla *et al.*, 2017; Baker *et al.*, 2017). This study presents a new approach to improving PET quantification, leveraging latent diffusion models trained using controlled simulated data. We show that diffusion models have a strong potential to enhance PET quantification through super-resolution. Our LDM-RR model's performance on longitudinal amyloid and cross-tracer PET data demonstrates that diffusion-based super-resolution (SR) approaches can outperform traditional approaches in tackling the issue of PVE in PET imaging.

We propose an alternative to L_2 loss, which has been a *de facto* standard in train-

ing diffusion models. L_2 penalty pushes the model to reduce large errors, potentially sacrificing high-frequency details (Fig.5.6 (B)). Moreover, L_2 loss is sensitive to the scale of voxel intensities. In super-resolution tasks, where the goal is to reconstruct fine details (voxel-level), L_2 loss may not be ideal for capturing subtle differences in high-frequency information, whereas L_1 loss may help capture the voxel-level details, which are considered to be crucial to addressing PVE. In addition, it is interesting to observe the added contribution from the multi-scale structural similarity index (MS-SSIM) metric, which confirmed the research findings of Zhao *et al.* (2016). Visual comparison in Fig.5.6 shows that using a combined loss at image and noise scales (Fig.5.6(D)), the generated high-resolution FBP images have a more accurate representation of brain structure from MRI and voxel-level uptake measurements.

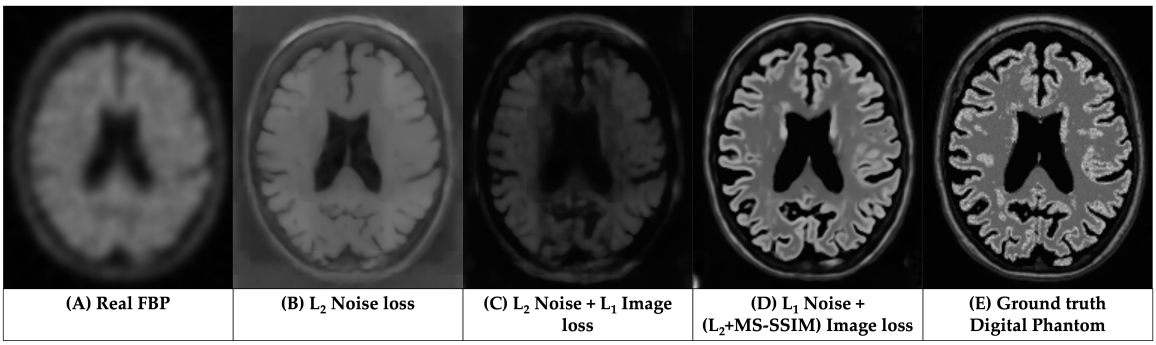

Figure 5.6: Comparisons of the Results from LDM Trained Using Different Loss Functions.

Simulation data was generated to approximate the PET imaging formation process and the distribution of tracer uptake as observed in real amyloid PET images. The simulated data were used to train the LDM-RR model and evaluate its performance against the ground truth, which is otherwise not possible. It should be recognized that simulated data cannot fully replicate the overall distribution and characteristics of real amyloid PET data, which may introduce bias to the trained model. More

sophisticated simulations can potentially be adopted to minimize this potential bias and improve model performance. Nevertheless, evaluation of model performance in a real-world setting is important, which we discuss further below.

We selected two commonly encountered scenarios in the investigation of Alzheimer's disease to evaluate the real-world utility and benefits of diffusion-based SR techniques for PET. In the longitudinal analysis, we leveraged data from the ADNI cohort in participants with a baseline visit and a 2-year follow-up to examine the sensitivity and statistical power of different correction methods. The participants were intentionally selected to have moderate to medium-high levels of pathological amyloid burden at baseline to maximize the probability of these participants accumulating amyloid plaques during the follow-up period, and therefore, we expect a positive increase in the MCSUVR measure and deviation from that reflects measurement noise. It is worth noting that both resolution recovery methods (RL-RR and LDM-RR) led to greater numerical values of the rate of amyloid accumulation, which reflects the improved recovery coefficient as expected. In the meantime, the standard deviation of the estimated rate also increased numerically, which can be a combined effect of the improved recovery coefficient and possible amplification of noise. The net effect of the correction methods is reflected by the p-values of the one-sample t-test applied to the rate data, where a smaller p-value indicates a greater statistical power, demonstrating a beneficial effect of correcting for PVE. The sample size estimation in hypothetical anti-amyloid treatment trials further confirmed the notion that correcting for PVE improves the longitudinal power. This improvement can lead to reduced experimental costs in longitudinal observational studies and clinical trials, which will facilitate treatment development. In a clinical setting, this improved power can lead to better patient management by providing more sensitive and accurate monitoring of disease progression once treatment becomes routinely available to patients. While

we chose to demonstrate the capability of our proposed technique to improve the quantitative analysis of clinical amyloid PET imaging data, the same principle and method can also be applied to the analysis of preclinical animal PET data. Previous studies have developed advanced algorithms for the analysis of preclinical PET data, e.g., (Giaccone *et al.*, 2022). Super-resolution methods, in general, and our proposed LDM-RR technique specifically, can recover the high-frequency signals lost during the image formation process by leveraging other sources of information, such as MR or prior knowledge, such as a template, and improve the quantitative accuracy of PET-derived measurements in both preclinical and clinical applications.

The second real-world application we tested in this study is the ability of PVE correction to improve agreement between PET-derived measurements from different tracers. Using amyloid PET imaging as an example, currently, there are at least five different PET tracers that are widely used in research studies, clinical trials, and patient management to measure amyloid burden. It is well recognized that the different tracers behave differently, leading to discrepancies in PET-derived amyloid burden measurements. At least part of this discrepancy is related to the contamination of the target measurement from nuance signals' spill-in to our measurements. We demonstrated that both correction methods improved the agreement, and the LDM-RR outperformed the RL method statistically. In the meantime, we recognize the improvement has not reached the level where two tracers can be used interchangeably and may not outperform some of the other techniques we have developed (Chen *et al.*, 2024; Shah *et al.*, 2022a). On the other hand, this experiment demonstrated that PVE correction can be one of the strategies we can employ to improve harmonization, and a combination of multiple techniques may be the ultimate solution to fully solve the harmonization problem. We also acknowledge that there are other approaches leveraging deep learning techniques to address issues related to the stan-

dardization and harmonization of image-derived measurements, e.g. (Kanaujia *et al.*, 2024). Further investigation in this direction is warranted.

Recent studies have shown that integrating classifier-free guidance (CFG) in diffusion models (Okada *et al.*, 2024) can improve image super-resolution by suppressing irrelevant diversity, balancing anatomical fidelity, and task-specific adaptability. CFG's core strength lies in its ability to dynamically modulate conditioning signals, such as MRI priors, without relying on external classifiers, making it particularly valuable in medical imaging where labeled data is sparse. For instance, β-CFG's adaptive guidance scaling proposed by Malarz *et al.* (2025) could optimize LDM-RR's MRI conditioning strength regionally and amplify guidance in cortical areas vulnerable to PVE while reducing it in artifact-free zones to preserve tracer-specific signals. This aligns with research done by Okada *et al.* (2024) in decomposing super-resolution into sub-tasks (e.g., blind restoration, SR) and using CFG to prioritize structural consistency. Although designed for anomaly detection, we can incorporate methods like AnoFPDM (Che *et al.*, 2025), where dynamic noise scale and threshold tuning can be used to adaptively adjust diffusion steps and loss weighting during PET reconstruction.

Limitations

This work has a few limitations: One potential limitation of our diffusion model-based framework is the computational complexity to train and validate the model on 3D imaging data. Even with a faster sampling method (Song *et al.*, 2020a), the inference time is considerably higher compared to other generative models. Due to the sequential nature of the denoising process, this is a known limitation of diffusion models and remains an active area of research. Second, the model was trained on synthetic data rather than real data, which may limit the technique's performance.

This is primarily due to the fact that the approach we adopted requires paired data with ground truth high-resolution images for training and validation, which is lacking. Examination of methods that are self-supervised or semi-supervised may allow us to overcome this limitation. We also like to point out that our intended application is narrowly focused on AD-related applications, while the underlying principles can be applied more broadly, although it will be beyond the scope of this paper.

5.6 Conclusion

We introduced a latent diffusion model-based resolution recovery method (LDM-RR) to enhance PET image resolution and mitigate the impact of PVE. Results demonstrate that the LDM-RR method improves spatial resolution while preserving critical amyloid and anatomical information, outperforming traditional methods like Richardson–Lucy (RL) correction. LDM-RR model showed superior performance at reconstructing high-resolution PET images, improved statistical power for detecting longitudinal amyloid accumulation, and a strong potential to improve the agreement between measurements obtained from different PET tracers, contributing to better data harmonization across multi-center studies. These findings suggest that diffusion-based super-resolution techniques offer a promising alternative to conventional PVC methods by overcoming noise amplification issues and achieving better image fidelity.

Chapter 6

FUTURE WORK & CONCLUSION

6.1 Future Work

The research presented in this dissertation opens several promising avenues for future investigation. Building upon the methodological advancements herein, future efforts will focus on extending the capabilities and applications of our core techniques. For instance, the robust brain age prediction model developed in Chapter 3 could be advanced by incorporating multi-modal data sources: genetic, cognitive, fluid biomarkers, or amyloid status (Ly *et al.*, 2020); and by adapting it to delineate disease-specific aging trajectories in various neurological conditions (Zhu *et al.*, 2023). For the MRI-guided super-resolution technique from Chapter 5, future work will aim to generalize its application to other PET tracers and imaging modality pairs. The main aim here would be to improve its robustness across diverse scanner environments through advanced domain adaptation and ensure longitudinal consistency for reliable disease progression monitoring. The locality constrained vector quantization framework (see Sec. 4.3) warrants further exploration into its theoretical properties such as convergence behavior and optimal parameter selection, alongside its systematic evaluation across a wider array of medical imaging modalities beyond T1w-MRI, and in non-medical domains requiring efficient and high-fidelity data compression. For instance, introducing coarse-to-fine quantization levels could capture both global anatomy and localized pathology within a single model (Takida *et al.*, 2023).

A significant thrust of future research will be directed towards the clinical validation of these developed methodologies on in-house datasets. This includes the design

and execution of prospective, longitudinal studies to affirm the clinical utility of the robust brain age biomarkers. Additionally, investigating the synergies between the developed techniques offers another promising research direction. For example, the efficacy of using LCVQ-compressed data as input for the robust brain age prediction framework or applying super-resolved PET data for more precise regional brain age estimations could be systematically assessed. Moreover, the potential of LCVQ to facilitate privacy-preserving federated learning by minimizing data transmission burdens warrants investigation, potentially enabling collaborative model training across institutions without compromising sensitive patient information.

It is also essential to acknowledge the inherent limitations and methodological nuances of the developed techniques. While LCVQ offers a novel approach to quantization, considerations regarding the computational overhead of Mahalanobis distance for extremely large codebooks and the sensitivity to covariance matrix estimation require ongoing attention. For the brain age prediction framework, the interpretation of brain age as a biomarker remains complex, and its deviation from chronological age is influenced by a multitude of factors that necessitate careful consideration in clinical contexts (Smith *et al.*, 2019). While our robust brain age framework reduces regression bias, the interpretability of Δage as a biomarker remains complex. Recent work demonstrates that cross-sectional Δage conflates unvarying baseline traits with true aging acceleration, limiting its prognostic value without longitudinal verification (Smith *et al.*, 2025). Future studies should prioritize multimodal decomposition of aging processes and prospective validation to isolate clinically actionable signals. Similarly, while the MRI-guided super-resolution significantly enhances image quality, the potential for introducing subtle algorithmically-induced artifacts that do not reflect underlying biology, and the dependence on co-registration of the guiding MRI, must be diligently evaluated to ensure diagnostic integrity (Markiewicz *et al.*, 2021).

We also acknowledge that the potential biases within training datasets can be learned and amplified by these models, if not proactively addressed. This can lead to disparities in performance across different demographic groups. Moreover, the "black box" nature of some deep learning components also underscores the continuing need for developing more interpretable AI models in medicine.

6.2 Conclusion

Although deep learning is poised to reshape neuro-imaging-based automated clinical diagnosis over the next decade, its full clinical promise is still impeded by inefficient representations and predictive models, along with issues of hardware-limited imaging resolution. In this dissertation, we address each barrier through three tightly coupled contributions that collectively advance data-efficient, biologically grounded analysis of the aging and diseased brain.

- In Chapter 3, we introduced a novel framework designed to improve the reliability of brain age estimation. By incorporating ordinal classification and distance regularization, we aim to reduce the systematic bias of regression to the mean in predicted brain age. This approach yields clinically more robust biomarkers of brain health, offering potential for improved tracking of neurodegenerative processes.

- Building on the need for efficient representations, Chapter 4 introduced Locality-Constrained Vector Quantization (LCVQ) that leverages Mahalanobis distance and local structure preservation to improve codebook utilization in VQ-VAE models. By selecting multiple codebook entries based on local covariance and applying adaptive sparsity, LCVQ achieves better latent encoding for medical images compared to traditional approaches. This improved representation can

preserve performance in downstream tasks, such as brain age prediction, and addresses known issues of codebook collapse and inefficiency in prior VQ-based models.

- Further expanding our contributions to quantitative imaging, in Chapter 5 we introduced a generative modeling approach for improving the resolution and accuracy of amyloid PET scans. By leveraging MRI data to guide the super-resolution of PET images through latent diffusion models, the method substantially enhances the quantitative information available for assessing amyloid burden, a key factor in the study and diagnosis of Alzheimer's disease and related dementias. We believe this work is one of the first to address the issue of partial volume effects in PET imaging using deep learning. The synergy between generative modeling and multimodal fusion provides a pathway toward more precise and less invasive imaging protocols.

Acarsoy, C., M. K. Ikram, M. A. Ikram, M. W. Vernooij and D. Bos, "Migraine and brain structure in the elderly: The rotterdam study", Cephalalgia **44**, 9, 03331024241266951 (2024).

Aggarwal, C. C., A. Hinneburg and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional space", in "Database Theory—ICDT 2001: 8th International Conference London, UK, January 4–6, 2001 Proceedings 8", pp. 420–434 (Springer, 2001).

Agustsson, E., F. Mentzer, M. Tschannen, L. Cavigelli, R. Timofte, L. Benini and L. V. Gool, "Soft-to-hard vector quantization for end-to-end learning compressible representations", Advances in neural information processing systems **30** (2017).

Ahmad, W., H. Ali, Z. Shah and S. Azmat, "A new generative adversarial network for medical images super resolution", Scientific Reports **12**, 1, 9533 (2022).

Alessio, A. M. and P. E. Kinahan, "Improved quantitation for pet/ct image reconstruction with system modeling and anatomical priors", Medical physics **33**, 11, 4095–4103 (2006).

Ambastha, A. K., T.-Y. Leong, A. D. N. Initiative *et al.*, "A deep learning approach to neuroanatomical characterisation of alzheimer's disease", in "MEDINFO 2017: Precision Healthcare through Informatics", pp. 1249–1249 (IOS Press, 2017).

Aston, J. A., V. J. Cunningham, M.-C. Asselin, A. Hammers, A. C. Evans and R. N. Gunn, "Positron emission tomography partial volume correction: estimation and algorithms", Journal of Cerebral Blood Flow & Metabolism **22**, 8, 1019–1034 (2002).

Azimi, M.-S., A. Kamali-Asl, M.-R. Ay, N. Zeraatkar, M.-S. Hosseini, A. Sanaat, H. Dadgar and H. Arabi, "Deep learning-based partial volume correction in standard and low-dose positron emission tomography-computed tomography imaging", Quantitative Imaging in Medicine and Surgery **14**, 3, 2146 (2024).

Babenko, A. and V. Lempitsky, "Additive quantization for extreme vector compression", in "Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition", pp. 931–938 (2014).

Baecker, L., J. Dafflon, P. F. Da Costa, R. Garcia-Dias, S. Vieira, C. Scarpazza, V. D. Calhoun, J. R. Sato, A. Mechelli and W. H. Pinaya, "Brain age prediction: A comparison between machine learning models using region-and voxel-based morphometric data", Human brain mapping **42**, 8, 2332–2346 (2021a).

Baecker, L., R. Garcia-Dias, S. Vieira, C. Scarpazza and A. Mechelli, "Machine learning for brain age prediction: Introduction to methods and clinical applications", EBioMedicine **72** (2021b).

Baete, K., J. Nuyts, K. Van Laere, W. Van Paesschen, S. Ceyssens, L. De Ceuninck, O. Gheysens, A. Kelles, J. Van den Eynden, P. Suetens *et al.*, "Evaluation of anatomy based reconstruction for partial volume correction in brain fdg-pet", Neuroimage **23**, 1, 305–317 (2004).

Baker, S. L., A. Maass and W. J. Jagust, "Considerations and code for partial volume correcting [18f]-av-1451 tau pet data", Data in brief **15**, 648–657 (2017).

Barisch-Fritz, B., J. Shah, J. Krafft, Y. E. Geda, T. Wu, A. Woll and J. Krell-Roesch, "Physical activity and the outcome of cognitive trajectory: a machine learning approach", European Review of Aging and Physical Activity **22**, 1, 1 (2025).

Bashyam, V. M., G. Erus, J. Doshi, M. Habes, I. M. Nasrallah, M. Truelove-Hill, D. Srinivasan, L. Mamourian, R. Pomponio, Y. Fan *et al.*, "Mri signatures of brain age and disease over the lifespan based on a deep brain network and 14 468 individuals worldwide", Brain **143**, 7, 2312–2324 (2020).

Baykal, G., M. Kandemir and G. Unal, "Edvae: Mitigating codebook collapse with evidential discrete variational autoencoders", Pattern Recognition **156**, 110792 (2024).

Beheshti, I., S. Mishra, D. Sone, P. Khanna and H. Matsuda, "T1-weighted mri-driven brain age estimation in alzheimer's disease and parkinson's disease", Aging and disease **11**, 3, 618 (2020).

Beheshti, I., S. Nugent, O. Potvin and S. Duchesne, "Bias-adjustment in neuroimaging-based brain age frameworks: A robust scheme", NeuroImage: Clinical **24**, 102063 (2019).

Belitz, K. and P. Stackelberg, "Evaluation of six methods for correcting bias in estimates from ensemble tree machine learning regression models", Environmental Modelling & Software **139**, 105006 (2021).

Beltran, J. F., B. M. Wahba, N. Hose, D. Shasha, R. P. Kline and A. D. N. Initiative, "Inexpensive, non-invasive biomarkers predict alzheimer transition using machine learning analysis of the alzheimer's disease neuroimaging (adni) database", PloS one **15**, 7, e0235663 (2020).

Bengio, Y., N. Léonard and A. Courville, "Estimating or propagating gradients through stochastic neurons for conditional computation", arXiv preprint arXiv:1308.3432 (2013).

Bethlehem, R. A., J. Seidlitz, S. R. White, J. W. Vogel, K. M. Anderson, C. Adamson, S. Adler, G. S. Alexopoulos, E. Anagnostou, A. Areces-Gonzalez *et al.*, "Brain charts for the human lifespan", Nature **604**, 7906, 525–533 (2022).

Bhagwat, N., J. Pipitone, A. N. Voineskos, M. M. Chakravarty, A. D. N. Initiative *et al.*, "An artificial neural network model for clinical score prediction in alzheimer disease using structural neuroimaging measures", Journal of Psychiatry and Neuroscience **44**, 4, 246–260 (2019).

Bhagwat, N., J. D. Viviano, A. N. Voineskos, M. M. Chakravarty, A. D. N. Initiative *et al.*, "Modeling and prediction of clinical symptom trajectories in alzheimer's disease using longitudinal data", PLoS computational biology **14**, 9, e1006376 (2018).

Biggs, D. S. and M. Andrews, "Acceleration of iterative image restoration algorithms", Applied optics **36**, 8, 1766–1775 (1997).

Bollack, A., L. E. Collij, D. V. García, M. Shekari, D. Altomare, P. Payoux, B. Dubois, O. Grau-Rivera, M. Boada, M. Marquié *et al.*, "Investigating reliable amyloid accumulation in centiloids: Results from the amypad prognostic and natural history study", Alzheimer's & Dementia **20**, 5, 3429–3441 (2024).

Boudiaf, M., J. Rony, I. M. Ziko, E. Granger, M. Pedersoli, P. Piantanida and I. B. Ayed, "A unifying mutual information view of metric learning: cross-entropy vs. pairwise losses", in "European conference on computer vision", pp. 548–564 (Springer, 2020).

Boussion, N., M. Hatt, F. Lamare, Y. Bizais, A. Turzo, C. Cheze-Le Rest and D. Visvikis, "A multiresolution image based approach for correction of partial volume effects in emission tomography", Physics in Medicine & Biology **51**, 7, 1857 (2006).

Cao, P., X. Liu, J. Yang, D. Zhao, M. Huang, J. Zhang and O. Zaiane, "Nonlinearity-aware based dimensionality reduction and over-sampling for ad/mci classification from mri measures", Computers in biology and medicine **91**, 21–37 (2017a).

Cao, Y., Z. Wu and C. Shen, "Estimating depth from monocular images as classification using deep fully convolutional residual networks", IEEE Transactions on Circuits and Systems for Video Technology **28**, 11, 3174–3182 (2017b).

Castellano, G., A. Esposito, E. Lella, G. Montanaro and G. Vessio, "Automated detection of alzheimer's disease: a multi-modal approach with 3d mri and amyloid pet", Scientific Reports **14**, 1, 5210 (2024).

Chang, C.-H., C.-H. Lin and H.-Y. Lane, "Machine learning and novel biomarkers for the diagnosis of alzheimer's disease", International journal of molecular sciences **22**, 5, 2761 (2021).

Chang, H., D.-Y. Yeung and Y. Xiong, "Super-resolution through neighbor embedding", in "Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.", vol. 1, pp. I–I (IEEE, 2004).

Chapleau, M., L. Iaccarino, D. Soleimani-Meigooni and G. D. Rabinovici, "The role of amyloid pet in imaging neurodegenerative disorders: a review", Journal of Nuclear Medicine **63**, Supplement 1, 13S–19S (2022).

Che, Y., F. Rafsani, J. Shah, M. M. R. Siddiquee and T. Wu, "Anofpdm: Anomaly detection with forward process of diffusion models for brain mri", in "Proceedings of the Winter Conference on Applications of Computer Vision", pp. 1113–1122 (2025).

Chen, K., V. Ghisays, J. Luo, Y. Chen, W. Lee, T. Wu, E. M. Reiman and Y. Su, "Harmonizing florbetapir and pib pet measurements of cortical aβ plaque burden using multiple regions-of-interest and machine learning techniques: An alternative to the centiloid approach", Alzheimer's & Dementia **20**, 3, 2165–2172 (2024).

Chen, K., A. Roontiva, P. Thiyyagura, W. Lee, X. Liu, N. Ayutyanont, H. Protas, J. L. Luo, R. Bauer, C. Reschke *et al.*, "Improved power for characterizing longitudinal amyloid-β pet changes and evaluating amyloid-modifying treatments with a cerebral white matter reference region", Journal of Nuclear Medicine **56**, 4, 560–566 (2015).

Chen, X., L. Li, A. Sharma, G. Dhiman and S. Vimal, "The application of convolutional neural network model in diagnosis and nursing of mr imaging in alzheimer's disease", Interdisciplinary Sciences: Computational Life Sciences pp. 1–11 (2021).

Choi, H., Y. K. Kim, E. J. Yoon, J.-Y. Lee, D. S. Lee and A. D. N. Initiative, "Cognitive signature of brain fdg pet based on deep learning: domain transfer from alzheimer's disease to parkinson's disease", European Journal of Nuclear Medicine and Molecular Imaging **47**, 403–412 (2020).

Chong, C. D., D. W. Dodick, B. L. Schlaggar and T. J. Schwedt, "Atypical age-related cortical thinning in episodic migraine", Cephalalgia **34**, 14, 1115–1124 (2014).

Cole, J. H., R. Leech, D. J. Sharp and A. D. N. Initiative, "Prediction of brain age suggests accelerated atrophy after traumatic brain injury", Annals of neurology **77**, 4, 571–581 (2015).

Cole, J. H., R. P. Poudel, D. Tsagkrasoulis, M. W. Caan, C. Steves, T. D. Spector and G. Montana, "Predicting brain age with deep learning from raw imaging data results in a reliable and heritable biomarker", NeuroImage **163**, 115–124 (2017).

Cole, J. H., S. J. Ritchie, M. E. Bastin, V. Hernández, S. Muñoz Maniega, N. Royle, J. Corley, A. Pattie, S. E. Harris, Q. Zhang *et al.*, "Brain age predicts mortality", Molecular psychiatry **23**, 5, 1385–1392 (2018).

Cui, R., M. Liu, A. D. N. Initiative *et al.*, "Rnn-based longitudinal analysis for diagnosis of alzheimer's disease", Computerized Medical Imaging and Graphics **73**, 1–10 (2019).

Dai, D., Y. Wang, Y. Chen and L. Van Gool, "Is image super-resolution helpful for other vision tasks?", in "2016 IEEE Winter Conference on Applications of Computer Vision (WACV)", pp. 1–9 (IEEE, 2016).

D'amico, S., D. Dall'Olio, C. Sala, L. Dall'Olio, E. Sauta, M. Zampini, G. Asti, L. Lanino, G. Maggioni, A. Campagna *et al.*, "Synthetic data generation by artificial intelligence to accelerate research and precision medicine in hematology", JCO Clinical Cancer Informatics **7**, e2300021 (2023).

Dartora, C., A. Marseglia, G. Mårtensson, G. Rukh, J. Dang, J.-S. Muehlboeck, L.-O. Wahlund, R. Moreno, J. Barroso, D. Ferreira *et al.*, "A deep learning model for brain age prediction using minimally preprocessed t1w images as input", Frontiers in Aging Neuroscience **15**, 1303036 (2024).

de Lange, A.-M. G., T. Kaufmann, D. van der Meer, L. A. Maglanoc, D. Alnæs, T. Moberget, G. Douaud, O. A. Andreassen and L. T. Westlye, "Population-based neuroimaging reveals traces of childbirth in the maternal brain", Proceedings of the National Academy of Sciences **116**, 44, 22341–22346 (2019).

Delmotte, K., J. Schaeverbeke, K. Poesen and R. Vandenberghe, "Prognostic value of amyloid/tau/neurodegeneration (atn) classification based on diagnostic cerebrospinal fluid samples for alzheimer's disease", Alzheimer's research & therapy **13**, 1–13 (2021).

Di Martino, A., C.-G. Yan, Q. Li, E. Denio, F. X. Castellanos, K. Alaerts, J. S. Anderson, M. Assaf, S. Y. Bookheimer, M. Dapretto *et al.*, "The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism", Molecular psychiatry **19**, 6, 659–667 (2014).

Domingos, P., "A few useful things to know about machine learning", Communications of the ACM **55**, 10, 78–87 (2012).

Donnelly-Kehoe, P. A., G. O. Pascariello, J. C. Gómez, A. D. N. Initiative *et al.*, "Looking for alzheimer's disease morphometric signatures using machine learning techniques", Journal of neuroscience methods **302**, 24–34 (2018).

Draper, K. and J. Ponsford, "Cognitive functioning ten years following traumatic brain injury and rehabilitation.", Neuropsychology **22**, 5, 618 (2008).

Duan, J., Y. Liu, H. Wu, J. Wang, L. Chen and C. P. Chen, "Broad learning for early diagnosis of alzheimer's disease using fdg-pet of the brain", Frontiers in Neuroscience **17**, 1137567 (2023).

Elhadad, A., M. Jamjoom and H. Abulkasim, "Reduction of nifti files storage and compression to facilitate telemedicine services based on quantization hiding of downsampling approach", Scientific Reports **14**, 1, 5168 (2024).

Erlandsson, K., J. Dickson, S. Arridge, D. Atkinson, S. Ourselin and B. F. Hutton, "Mr imaging–guided partial volume correction of pet data in pet/mr imaging", PET clinics **11**, 2, 161–177 (2016).

Fischl, B., "Freesurfer", Neuroimage **62**, 2, 774–781 (2012).

Franke, K. and C. Gaser, "Longitudinal changes in individual brainage in healthy aging, mild cognitive impairment, and alzheimer's disease", GeroPsych (2012).

Franke, K., G. Ziegler, S. Klöppel, C. Gaser, A. D. N. Initiative *et al.*, "Estimating the age of healthy subjects from t1-weighted mri scans using kernel methods: exploring the influence of various parameters", Neuroimage **50**, 3, 883–892 (2010).

Frölich, L., O. Peters, P. Lewczuk, O. Gruber, S. J. Teipel, H. J. Gertz, H. Jahn, F. Jessen, A. Kurz, C. Luckhaus *et al.*, "Incremental value of biomarker combinations to predict progression of mild cognitive impairment to alzheimer's dementia", Alzheimer's research & therapy **9**, 1–15 (2017).

Frouin, V., C. Comtat, A. Reilhac and M.-C. Grégoire, "Correction of partial-volume effect for pet striatal imaging: fast implementation and study of robustness", Journal of Nuclear Medicine **43**, 12, 1715–1726 (2002).

Gao, C., B. D. Killeen, Y. Hu, R. B. Grupp, R. H. Taylor, M. Armand and M. Unberath, "Synthetic data accelerates the development of generalizable learning-based algorithms for x-ray image analysis", Nature Machine Intelligence **5**, 3, 294–308 (2023).

Gao, F., Y. Su, J. Shah and T. Wu, "Deep residual inception encoder-decoder network for amyloid pet harmonization", US Patent 12,186,114 (2025).

Gao, Z., C. Tan, J. Wang, Y. Huang, L. Wu and S. Z. Li, "Foldtoken: Learning protein language via vector quantization and beyond", arXiv preprint arXiv:2403.09673 (2024).

Gardner, M. and J. Heady, "Some effects of within-person variability in epidemiological studies", Journal of Chronic Diseases **26**, 12, 781–795 (1973).

Gaser, C., K. Franke, S. Klöppel, N. Koutsouleris, H. Sauer and A. D. N. Initiative, "Brainage in mild cognitive impaired patients: predicting the conversion to alzheimer's disease", PloS one **8**, 6, e67346 (2013).

Geda, Y. E., R. O. Roberts, M. M. Mielke, D. S. Knopman, T. J. Christianson, V. S. Pankratz, B. F. Boeve, O. Sochor, E. G. Tangalos, R. C. Petersen *et al.*, "Baseline neuropsychiatric symptoms and the risk of incident mild cognitive impairment: a population-based study", American Journal of Psychiatry **171**, 5, 572–581 (2014).

Giaccone, P., V. Benfante, A. Stefano, F. P. Cammarata, G. Russo and A. Comelli, "Pet images atlas-based segmentation performed in native and in template space: a radiomics repeatability study in mouse models", in "International Conference on Image Analysis and Processing", pp. 351–361 (Springer, 2022).

Gill, S., P. Mouches, S. Hu, D. Rajashekar, F. P. MacMaster, E. E. Smith, N. D. Forkert, Z. Ismail and A. D. N. Initiative, "Using machine learning to predict dementia from neuropsychiatric symptom and neuroimaging data", Journal of Alzheimer's Disease **75**, 1, 277–288 (2020).

Golla, S. S., M. Lubberink, B. N. van Berckel, A. A. Lammertsma and R. Boellaard, "Partial volume correction of brain pet studies using iterative deconvolution in combination with hypr denoising", EJNMMI research **7**, 1–12 (2017).

Gonzalez-Escamilla, G., C. Lange, S. Teipel, R. Buchert, M. J. Grothe, A. D. N. Initiative *et al.*, "Petpve12: an spm toolbox for partial volume effects correction in brain pet–application to amyloid imaging with av45-pet", Neuroimage **147**, 669–677 (2017).

Goodfellow, I. J., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville and Y. Bengio, "Generative adversarial nets", Advances in neural information processing systems **27** (2014).

Gray, R., "Vector quantization", IEEE Assp Magazine **1**, 2, 4–29 (1984).

Greenspan, H., "Super-resolution in medical imaging", The computer journal **52**, 1, 43–63 (2009).

Gutiérrez, P. A., M. Perez-Ortiz, J. Sanchez-Monedero, F. Fernandez-Navarro and C. Hervas-Martinez, "Ordinal regression methods: survey and experimental study", IEEE Transactions on Knowledge and Data Engineering **28**, 1, 127–146 (2015).

Haan, M. N. and R. Wallace, "Can dementia be prevented? brain aging in a population-based context", Annu. Rev. Public Health **25**, 1–24 (2004).

Haris, M., G. Shakhnarovich and N. Ukita, "Task-driven super resolution: Object detection in low-resolution images", in "Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part V 28", pp. 387–395 (Springer, 2021).

Ho, J., A. Jain and P. Abbeel, "Denoising diffusion probabilistic models", Advances in neural information processing systems **33**, 6840–6851 (2020).

Hoffman, E. J., S.-C. Huang and M. E. Phelps, "Quantitation in positron emission computed tomography: 1. effect of object size", Journal of computer assisted tomography **3**, 3, 299–308 (1979).

Hong, A. S., D. Levin, L. Parker, V. M. Rao, D. Ross-Degnan and J. F. Wharam, "Trends in diagnostic imaging utilization among medicare and commercially insured adults from 2003 through 2016", Radiology **294**, 2, 342–350 (2020).

Horvath, S. and K. Raj, "Dna methylation-based biomarkers and the epigenetic clock theory of ageing", Nature reviews genetics **19**, 6, 371–384 (2018).

Huffman, D. A., "A method for the construction of minimum-redundancy codes", Proceedings of the IRE **40**, 9, 1098–1101 (1952).

Huh, M., B. Cheung, P. Agrawal and P. Isola, "Straightening out the straight-through estimator: Overcoming optimization challenges in vector quantized networks", in "International Conference on Machine Learning", pp. 14096–14113 (PMLR, 2023).

Hussain, A. J., A. Al-Fayadh and N. Radi, "Image compression techniques: A survey in lossless and lossy algorithms", Neurocomputing **300**, 44–69 (2018).

Isaac, J. S. and R. Kulkarni, "Super resolution techniques for medical image processing", in "2015 International Conference on Technologies for Sustainable Development (ICTSD)", pp. 1–6 (IEEE, 2015).

Ito, K., R. Fujimoto, T.-W. Huang, H.-T. Chen, K. Wu, K. Sato, Y. Taki, H. Fukuda and T. Aoki, "Performance evaluation of age estimation from t1-weighted images using brain local features and cnn", in "2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)", pp. 694–697 (IEEE, 2018).

Jang, E., S. Gu and B. Poole, "Categorical reparameterization with gumbel-softmax", arXiv preprint arXiv:1611.01144 (2016).

Jegou, H., M. Douze and C. Schmid, "Product quantization for nearest neighbor search", IEEE transactions on pattern analysis and machine intelligence **33**, 1, 117–128 (2010).

Jiang, H., N. Lu, K. Chen, L. Yao, K. Li, J. Zhang and X. Guo, "Predicting brain age of healthy adults based on structural mri parcellation using convolutional neural networks", Frontiers in neurology **10**, 1346 (2020).

Jiang, H., Z. Ma, Y. Hu, B. Yang and L. Zhang, "Medical image compression based on vector quantization with variable block sizes in wavelet domain", Computational intelligence and neuroscience **2012**, 1, 541890 (2012).

Jónsson, B. A., G. Bjornsdottir, T. Thorgeirsson, L. M. Ellingsen, G. B. Walters, D. Gudbjartsson, H. Stefansson, K. Stefansson and M. Ulfarsson, "Brain age prediction using deep learning uncovers associated sequence variants", Nature communications **10**, 1, 5409 (2019).

Joshi, A., R. A. Koeppe and J. A. Fessler, "Reducing between scanner differences in multi-center pet studies", Neuroimage **46**, 1, 154–159 (2009).

Joshi, A., M. M. R. Siddiquee, J. Shah, T. Schwedt, C. Chong, B. Li and T. Wu, "Prediction of headache improvement using multimodal machine learning in patients with acute post-traumatic headache (p5-12.002)", in "Neurology", vol. 102, p. 3391 (Lippincott Williams & Wilkins Hagerstown, MD, 2024).

Kanaujia, V. K., A. Kumar and S. P. Yadav, "Advancements in automatic kidney segmentation using deep learning frameworks and volumetric segmentation techniques for ct imaging: A review", Archives of Computational Methods in Engineering **31**, 5, 3151–3169 (2024).

Kang, M. J., S. Y. Kim, D. L. Na, B. C. Kim, D. W. Yang, E.-J. Kim, H. R. Na, H. J. Han, J.-H. Lee, J. H. Kim *et al.*, "Prediction of cognitive impairment via deep learning trained with multi-center neuropsychological test data", BMC medical informatics and decision making **19**, 1–9 (2019).

Kaufmann, T., D. van der Meer, N. T. Doan, E. Schwarz, M. J. Lund, I. Agartz, D. Alnæs, D. M. Barch, R. Baur-Streubel, A. Bertolino *et al.*, "Common brain disorders are associated with heritable patterns of apparent aging of the brain", Nature neuroscience **22**, 10, 1617–1623 (2019).

Ker, J., L. Wang, J. Rao and T. Lim, "Deep learning applications in medical image analysis", Ieee Access **6**, 9375–9389 (2017).

Khader, F., G. Müller-Franzes, S. Tayebi Arasteh, T. Han, C. Haarburger, M. Schulze-Hagen, P. Schad, S. Engelhardt, B. Baeßler, S. Foersch *et al.*, "Denoising diffusion probabilistic models for 3d medical image generation", Scientific Reports **13**, 1, 7303 (2023).

Kim, J., W. S. Ha, S. H. Park, K. Han and M. S. Baek, "Association between migraine and alzheimer's disease: a nationwide cohort study", Frontiers in Aging Neuroscience **15**, 1196185 (2023).

Klöppel, S., M. Kotschi, J. Peter, K. Egger, L. Hausner, L. Frölich, A. Förster, B. Heimbach, C. Normann, W. Vach *et al.*, "Separating symptomatic alzheimer's disease from depression based on structural mri", Journal of Alzheimer's disease **63**, 1, 353–363 (2018).

Klunk, W. E., R. A. Koeppe, J. C. Price, T. L. Benzinger, M. D. Devous Sr, W. J. Jagust, K. A. Johnson, C. A. Mathis, D. Minhas, M. J. Pontecorvo *et al.*, "The centiloid project: standardizing quantitative amyloid plaque estimation by pet", Alzheimer's & dementia **11**, 1, 1–15 (2015).

LaMontagne, P. J., T. L. Benzinger, J. C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant, J. Hassenstab, K. Moulder, A. G. Vlassenko *et al.*, "Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease", medrxiv pp. 2019–12 (2019).

Łańcucki, A., J. Chorowski, G. Sanchez, R. Marxer, N. Chen, H. J. Dolfing, S. Khurana, T. Alumäe and A. Laurent, "Robust training of vector quantized bottleneck models", in "2020 International Joint Conference on Neural Networks (IJCNN)", pp. 1–7 (IEEE, 2020).

Langdon, G. and J. Rissanen, "Compression of black-white images with arithmetic coding", IEEE Transactions on Communications **29**, 6, 858–867 (1981).

Lanitis, A., C. Draganova and C. Christodoulou, "Comparing different classifiers for automatic age estimation", IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) **34**, 1, 621–628 (2004).

Le, T. T., R. T. Kuplicki, B. A. McKinney, H.-W. Yeh, W. K. Thompson, M. P. Paulus and T. . Investigators, "A nonlinear simulation framework supports adjusting for age when analyzing brainage", Frontiers in aging neuroscience **10**, 317 (2018).

LeCun, Y., Y. Bengio and G. Hinton, "Deep learning", nature **521**, 7553, 436–444 (2015).

Lee, D., C. Kim, M. Cho and W. S. HAN, "Locality-aware generalizable implicit neural representation", Advances in Neural Information Processing Systems **36**, 48363–48381 (2023).

Lee, H. and S. Chen, "Systematic bias of machine learning regression models and correction", IEEE Transactions on Pattern Analysis and Machine Intelligence (2025).

Li, H., Y. Yang, M. Chang, S. Chen, H. Feng, Z. Xu, Q. Li and Y. Chen, "Srdiff: Single image super-resolution with diffusion probabilistic models", Neurocomputing **479**, 47–59 (2022).

Li, X., Y. Jiang, J. J. Rodriguez-Andina, H. Luo, S. Yin and O. Kaynak, "When medical images meet generative adversarial network: recent development and research opportunities", Discover Artificial Intelligence **1**, 1–20 (2021a).

Li, Y., B. Sixou and F. Peyrin, "A review of the deep learning methods for medical images super resolution problems", Irbm **42**, 2, 120–133 (2021b).

Liang, H., F. Zhang and X. Niu, "Investigating systematic bias in brain age estimation with application to post-traumatic stress disorders", Tech. rep., Wiley Online Library (2019).

Liem, F., G. Varoquaux, J. Kynast, F. Beyer, S. K. Masouleh, J. M. Huntenburg, L. Lampe, M. Rahim, A. Abraham, R. C. Craddock *et al.*, "Predicting brain-age from multimodal imaging data captures cognitive impairment", Neuroimage **148**, 179–188 (2017).

Liu, F., M. Hernandez-Cabronero, V. Sanchez, M. W. Marcellin and A. Bilgin, "The current role of image compression standards in medical imaging", Information **8**, 4, 131 (2017).

Liu, L., H. Lu, H. Xiong, K. Xian, Z. Cao and C. Shen, "Counting objects by blockwise classification", IEEE Transactions on Circuits and Systems for Video Technology **30**, 10, 3513–3527 (2019).

Liu, X., J. Wang, F. Ren and J. Kong, "Group guided fused laplacian sparse group lasso for modeling alzheimer's disease progression", Computational and mathematical methods in medicine **2020**, 1, 4036560 (2020).

Long, J. M. and D. M. Holtzman, "Alzheimer disease: an update on pathobiology and treatment strategies", Cell **179**, 2, 312–339 (2019).

López-González, F. J., J. Silva-Rodríguez, J. Paredes-Pacheco, A. Niñerola-Baizán, N. Efthimiou, C. Martín-Martín, A. Moscoso, Á. Ruibal, N. Roé-Vellvé and P. Aguiar, "Intensity normalization methods in brain fdg-pet quantification", Neuroimage **222**, 117229 (2020).

López-Otín, C., M. A. Blasco, L. Partridge, M. Serrano and G. Kroemer, "The hallmarks of aging", Cell **153**, 6, 1194–1217 (2013).

Lucy, L. B., "An iterative technique for the rectification of observed distributions", Astronomical Journal, Vol. 79, p. 745 (1974) **79**, 745 (1974).

Ly, M., Z. Y. Gary, H. T. Karim, N. R. Muppidi, A. Mizuno, W. E. Klunk, H. J. Aizenstein, A. D. N. Initiative *et al.*, "Improving brain age prediction models: incorporation of amyloid status in alzheimer's disease", Neurobiology of aging **87**, 44–48 (2020).

Lyu, F., M. Ye, J. F. Carlsen, K. Erleben, S. Darkner and P. C. Yuen, "Pseudo-label guided image synthesis for semi-supervised covid-19 pneumonia infection segmentation", IEEE Transactions on Medical Imaging **42**, 3, 797–809 (2022).

Malarz, D., A. Kasymov, M. Zieba, J. Tabor and P. Spurek, "Classifier-free guidance with adaptive scaling", arXiv preprint arXiv:2502.10574 (2025).

Mallo, S. C., S. Valladares-Rodriguez, D. Facal, C. Lojo-Seoane, M. J. Fernández-Iglesias and A. X. Pereiro, "Neuropsychiatric symptoms as predictors of conversion from mci to dementia: a machine learning approach", International psychogeriatrics **32**, 3, 381–392 (2020).

Mar, J., A. Gorostiza, O. Ibarrondo, C. Cernuda, A. Arrospide, Á. Iruin, I. Larrañaga, M. Tainta, E. Ezpeleta and A. Alberdi, "Validation of random forest machine learning models to predict dementia-related neuropsychiatric symptoms in real-world data", Journal of Alzheimer's Disease **77**, 2, 855–864 (2020).

Marcus, D. S., A. F. Fotenos, J. G. Csernansky, J. C. Morris and R. L. Buckner, "Open access series of imaging studies: longitudinal mri data in nondemented and demented older adults", Journal of cognitive neuroscience **22**, 12, 2677–2684 (2010).

Marcus, D. S., T. H. Wang, J. Parker, J. G. Csernansky, J. C. Morris and R. L. Buckner, "Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults", Journal of cognitive neuroscience **19**, 9, 1498–1507 (2007).

Markiewicz, P. J., J. C. Matthews, J. Ashburner, D. M. Cash, D. L. Thomas, E. De Vita, A. Barnes, M. J. Cardoso, M. Modat, R. Brown *et al.*, "Uncertainty analysis of mr-pet image registration for precision neuro-pet imaging", Neuroimage **232**, 117821 (2021).

Mårtensson, G., D. Ferreira, T. Granberg, L. Cavallin, K. Oppedal, A. Padovani, I. Rektorova, L. Bonanni, M. Pardini, M. G. Kramberger *et al.*, "The reliability of a deep learning model in clinical out-of-distribution mri data: a multicohort study", Medical Image Analysis **66**, 101714 (2020).

Matsubara, K., M. Ibaraki, T. Kinoshita and A. D. N. Initiative, "Deeppvc: prediction of a partial volume-corrected map for brain positron emission tomography studies via a deep convolutional neural network", EJNMMI physics **9**, 1, 50 (2022).

Meltzer, C. C., P. E. Kinahan, P. J. Greer, T. E. Nichols, C. Comtat, M. N. Cantwell, M. P. Lin and J. C. Price, "Comparative evaluation of mr-based partial-volume correction schemes for pet", Journal of Nuclear Medicine **40**, 12, 2053–2065 (1999).

Meltzer, C. C., J. P. Leal, H. S. Mayberg, H. N. Wagner Jr and J. J. Frost, "Correction of pet data for partial volume effects in human cerebral cortex by mr imaging", Journal of computer assisted tomography **14**, 4, 561–570 (1990).

Mentzer, F., D. Minnen, E. Agustsson and M. Tschannen, "Finite scalar quantization: Vq-vae made simple", arXiv preprint arXiv:2309.15505 (2023).

Mirkes, E. M., J. Allohibi and A. Gorban, "Fractional norms and quasinorms do not help to overcome the curse of dimensionality", Entropy **22**, 10, 1105 (2020).

Moguilner, S., S. Baez, H. Hernandez, J. Migeot, A. Legaz, R. Gonzalez-Gomez, F. R. Farina, P. Prado, J. Cuadros, E. Tagliazucchi *et al.*, "Brain clocks capture diversity and disparities in aging and dementia across geographically diverse populations", Nature medicine pp. 1–12 (2024).

More, S., G. Antonopoulos, F. Hoffstaedter, J. Caspers, S. B. Eickhoff, K. R. Patil, A. D. N. Initiative *et al.*, "Brain-age prediction: A systematic comparison of machine learning workflows", NeuroImage **270**, 119947 (2023).

Moser, B. B., F. Raue, S. Frolov, S. Palacio, J. Hees and A. Dengel, "Hitchhiker's guide to super-resolution: Introduction and recent advances", IEEE Transactions on Pattern Analysis and Machine Intelligence **45**, 8, 9862–9882 (2023).

Müller-Franzes, G., J. M. Niehues, F. Khader, S. T. Arasteh, C. Haarburger, C. Kuhl, T. Wang, T. Han, T. Nolte, S. Nebelung *et al.*, "A multimodal comparison of latent denoising diffusion probabilistic models and generative adversarial networks for medical image synthesis", Scientific Reports **13**, 1, 12098 (2023).

Müller-Gärtner, H. W., J. M. Links, J. L. Prince, R. N. Bryan, E. McVeigh, J. P. Leal, C. Davatzikos and J. J. Frost, "Measurement of radiotracer concentration in brain gray matter using positron emission tomography: Mri-based correction for partial volume effects", Journal of Cerebral Blood Flow & Metabolism **12**, 4, 571–583 (1992).

Myszczynska, M. A., P. N. Ojamies, A. M. Lacoste, D. Neil, A. Saffari, R. Mead, G. M. Hautbergue, J. D. Holbrook and L. Ferraiuolo, "Applications of machine learning to diagnosis and treatment of neurodegenerative diseases", Nature reviews neurology **16**, 8, 440–456 (2020).

Navitsky, M., A. D. Joshi, I. Kennedy, W. E. Klunk, C. C. Rowe, D. F. Wong, M. J. Pontecorvo, M. A. Mintun and M. D. Devous Sr, "Standardization of amyloid quantitation with florbetapir standardized uptake value ratios to the centiloid scale", Alzheimer's & Dementia **14**, 12, 1565–1571 (2018).

Niccoli, T. and L. Partridge, "Ageing as a risk factor for disease", Current biology **22**, 17, R741–R752 (2012).

of the International Headache Society (IHS), H. C. C., "The international classification of headache disorders, (beta version)", Cephalalgia **33**, 9, 629–808 (2013).

Okada, S., R. Yoshihashi, H. Kataoka *et al.*, "Real-srgd: Enhancing real-world image super-resolution with classifier-free guided diffusion", in "Proceedings of the Asian Conference on Computer Vision", pp. 3739–3755 (2024).

Organization, W. H. *et al.*, "Over 1 in 3 people affected by neurological conditions, the leading cause of illness and disability worldwide", World Health Organization (2024).

Palermo, F., H. Li, A. Capstick, N. Fletcher-Lloyd, Y. Zhao, S. Kouchaki, R. Nilforooshan, D. Sharp and P. Barnaghi, "Designing a clinically applicable deep recurrent model to identify neuropsychiatric symptoms in people living with dementia using in-home monitoring data", arXiv preprint arXiv:2110.09868 (2021).

Pan, H., H. Han, S. Shan and X. Chen, "Mean-variance loss for deep age estimation from a face", in "Proceedings of the IEEE conference on computer vision and pattern recognition", pp. 5285–5294 (2018).

Pang, S., Z. Lu, J. Jiang, L. Zhao, L. Lin, X. Li, T. Lian, M. Huang, W. Yang and Q. Feng, "Hippocampus segmentation based on iterative local linear mapping with representative and local structure-preserved feature embedding", IEEE transactions on medical imaging **38**, 10, 2271–2280 (2019).

Peng, H., W. Gong, C. F. Beckmann, A. Vedaldi and S. M. Smith, "Accurate brain age prediction with lightweight deep neural networks", Medical image analysis **68**, 101871 (2021).

Pereyra, G., G. Tucker, J. Chorowski, Ł. Kaiser and G. Hinton, "Regularizing neural networks by penalizing confident output distributions", arXiv preprint arXiv:1701.06548 (2017).

Pinaya, W. H., P.-D. Tudosiu, J. Dafflon, P. F. Da Costa, V. Fernandez, P. Nachev, S. Ourselin and M. J. Cardoso, "Brain imaging generation with latent diffusion models", in "MICCAI Workshop on Deep Generative Models", pp. 117–126 (Springer, 2022a).

Pinaya, W. H., P.-D. Tudosiu, R. Gray, G. Rees, P. Nachev, S. Ourselin and M. J. Cardoso, "Unsupervised brain imaging 3d anomaly detection and segmentation with transformers", Medical Image Analysis **79**, 102475 (2022b).

Pontecorvo, M. J. and M. A. Mintun, "Pet amyloid imaging as a tool for early diagnosis and identifying patients at risk for progression to alzheimer's disease", Alzheimer's research & therapy **3**, 1–9 (2011).

Qiao, H., L. Chen, Z. Ye and F. Zhu, "Early alzheimer's disease diagnosis with the contrastive loss using paired structural mris", Computer methods and programs in biomedicine **208**, 106282 (2021).

Qiao, H., L. Chen and F. Zhu, "Ranking convolutional neural network for alzheimer's disease mini-mental state examination prediction at multiple time-points", Computer Methods and Programs in Biomedicine **213**, 106503 (2022).

Qiu, S., P. S. Joshi, M. I. Miller, C. Xue, X. Zhou, C. Karjadi, G. H. Chang, A. S. Joshi, B. Dwyer, S. Zhu *et al.*, "Development and validation of an interpretable deep learning framework for alzheimer's disease classification", Brain **143**, 6, 1920–1933 (2020).

Qu, H., S. Yang, Z. Yao, X. Sun and H. Chen, "Association of headache disorders and the risk of dementia: meta-analysis of cohort studies", Frontiers in Aging Neuroscience **14**, 804341 (2022).

Rafsani, F., D. Sheth, Y. Che, J. Shah, M. M. R. Siddiquee, C. Chong, S. Nikolva, G. Dumkrieger, B. Li, T. Wu *et al.*, "Using large-scale contrastive language-image pre-training to maximize brain mri-based headache classification (p4-12.007)", in "Neurology", vol. 104, p. 2728 (Lippincott Williams & Wilkins Hagerstown, MD, 2025).

Rahman Siddiquee, M. M., J. Shah, C. Chong, S. Nikolova, G. Dumkrieger, B. Li, T. Wu and T. J. Schwedt, "Headache classification and automatic biomarker extraction from structural mris using deep learning", Brain Communications **5**, 1, fcac311 (2023).

Rahman Siddiquee, M. M., J. Shah, T. Wu, C. Chong, T. Schwedt and B. Li, "Healthygan: Learning from unannotated medical images to detect anomalies associated with human disease", in "International Workshop on Simulation and Synthesis in Medical Imaging", pp. 43–54 (Springer, 2022).

Rajotte, J.-F., R. Bergen, D. L. Buckeridge, K. El Emam, R. Ng and E. Strome, "Synthetic data as an enabler for machine learning applications in medicine", Iscience **25**, 11 (2022).

Ramzan, F., M. U. G. Khan, A. Rehmat, S. Iqbal, T. Saba, A. Rehman and Z. Mehmood, "A deep learning approach for automated diagnosis and multi-class classification of alzheimer's disease stages using resting-state fmri and residual neural networks", Journal of medical systems **44**, 1–16 (2020).

Rasti, P., T. Uiboupin, S. Escalera and G. Anbarjafari, "Convolutional neural network super resolution for face recognition in surveillance monitoring", in "Articulated Motion and Deformable Objects: 9th International Conference, AMDO 2016, Palma de Mallorca, Spain, July 13-15, 2016, Proceedings 9", pp. 175–184 (Springer, 2016).

Reynés-Llompart, G., C. Gámez-Cenzano, I. Romero-Zayas, L. Rodríguez-Bel, J. L. Vercher-Conejero and J. M. Martí-Climent, "Performance characteristics of the whole-body discovery iq pet/ct system", Journal of Nuclear Medicine **58**, 7, 1155–1161 (2017).

Rezende, D. and S. Mohamed, "Variational inference with normalizing flows", in "International conference on machine learning", pp. 1530–1538 (PMLR, 2015).

Rombach, R., A. Blattmann, D. Lorenz, P. Esser and B. Ommer, "High-resolution image synthesis with latent diffusion models", in "Proceedings of the IEEE/CVF conference on computer vision and pattern recognition", pp. 10684–10695 (2022).

Rosati, R., L. Romeo, V. M. Vargas, P. A. Gutiérrez, C. Hervás-Martínez and E. Frontoni, "A novel deep ordinal classification approach for aesthetic quality control classification", Neural Computing and Applications **34**, 14, 11625–11639 (2022).

Rothe, R., R. Timofte and L. Van Gool, "Deep expectation of real and apparent age from a single image without facial landmarks", International Journal of Computer Vision **126**, 2-4, 144–157 (2018).

Rousset, O. G., D. L. Collins, A. Rahmim and D. F. Wong, "Design and implementation of an automated partial volume correction in pet: application to dopamine receptor quantification in the normal human striatum", Journal of nuclear medicine **49**, 7, 1097–1106 (2008).

Rousset, O. G., Y. Ma and A. C. Evans, "Correction for partial volume effects in pet: principle and validation", Journal of nuclear medicine **39**, 5, 904–911 (1998).

Royse, S. K., D. S. Minhas, B. J. Lopresti, A. Murphy, T. Ward, R. A. Koeppe, S. Bullich, S. DeSanti, W. J. Jagust, S. M. Landau *et al.*, "Validation of amyloid pet positivity thresholds in centiloids: a multisite pet study approach", Alzheimer's research & therapy **13**, 1, 99 (2021).

Saharia, C., W. Chan, S. Saxena, L. Li, J. Whang, E. L. Denton, K. Ghasemipour, R. Gontijo Lopes, B. Karagol Ayan, T. Salimans *et al.*, "Photorealistic text-to-image diffusion models with deep language understanding", Advances in neural information processing systems **35**, 36479–36494 (2022).

Salvatore, C., A. Cerasa, P. Battista, M. C. Gilardi, A. Quattrone, I. Castiglioni and A. D. N. Initiative, "Magnetic resonance imaging biomarkers for the early diagnosis of alzheimer's disease: a machine learning approach", Frontiers in neuroscience **9**, 307 (2015).

Schwedt, T. J., A. A. Pradhan, M. L. Oshinsky, M. F. Brin, H. Rosen, N. Lalvani, A. Charles, M. Ashina, T. P. Do, R. Burstein *et al.*, "The headache research priorities: Research goals from the american headache society and an international multistakeholder expert group", Headache: The Journal of Head and Face Pain **64**, 8, 912–930 (2024).

Selkoe, D. J., "Alzheimer's disease–genotypes, phenotype, and treatments", Science **275**, 5300, 630–631 (1997).

Shah, J., Y. Che, J. Sohankar, J. Luo, B. Li, Y. Su, T. Wu and A. D. N. Initiative, "Enhancing amyloid pet quantification: Mri-guided super-resolution using latent diffusion models", Life **14**, 12, 1580 (2024a).

Shah, J., K. Chen, E. M. Reiman, B. Li, T. Wu and Y. Su, "Transfer learning based deep encoder decoder network for amyloid pet harmonization with small datasets", Alzheimer's & Dementia **19**, e062947 (2023a).

Shah, J., F. Gao, B. Li, V. Ghisays, J. Luo, Y. Chen, W. Lee, Y. Zhou, T. L. Benzinger, E. M. Reiman *et al.*, "Deep residual inception encoder-decoder network for amyloid pet harmonization", Alzheimer's & Dementia **18**, 12, 2448–2457 (2022a).

Shah, J., V. Ghisays, Y. Chen, J. Luo, B. Li, E. M. Reiman, K. Chen, T. Wu and Y. Su, "Mri signatures of brain age in the alzheimer's disease continuum", Alzheimer's & Dementia **18**, e061942 (2022b).

Shah, J., J. Krell-Roesch, E. Forzani, D. S. Knopman, C. R. Jack, R. C. Petersen, Y. Che, T. Wu and Y. E. Geda, "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 **1**, 11 (2025).

Shah, J., J. Luo, J. Sohankar, E. M. Reiman, K. Chen, Y. Su, B. Li and T. Wu, "A multi-class deep learning model to estimate brain age while addressing systematic bias of regression to the mean", in "Alzheimer's Association International Conference", (ALZ, 2023b).

Shah, J., M. M. Rahman Siddiquee, J. Krell-Roesch, J. A. Syrjanen, W. K. Kremers, M. Vassilaki, E. Forzani, T. Wu and Y. E. Geda, "Neuropsychiatric symptoms and commonly used biomarkers of alzheimer's disease: A literature review from a machine learning perspective", Journal of Alzheimer's Disease **92**, 4, 1131–1146 (2023c).

Shah, J., M. M. R. Siddiquee, C. Chong, T. Schwedt, J. Li, V. Berisha, K. Ross and T. Wu, "Capturing mri signatures of brain age as a potential biomarker to predict persistence of post-traumatic headache (s20. 006)", in "Neurology", vol. 102, p. 3099 (Lippincott Williams & Wilkins Hagerstown, MD, 2024b).

Shah, J., M. M. R. Siddiquee, Y. Su, T. Wu and B. Li, "Ordinal classification with distance regularization for robust brain age prediction", in "Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision", pp. 7882–7891 (2024c).

Shidahara, M., C. Tsoumpas, A. Hammers, N. Boussion, D. Visvikis, T. Suhara, I. Kanno and F. E. Turkheimer, "Functional and structural synergy for resolution recovery and partial volume correction in brain pet", Neuroimage **44**, 2, 340–348 (2009).

Siddiquee, M. M. R., J. Shah, T. Wu, C. Chong, T. J. Schwedt, G. Dumkrieger, S. Nikolova and B. Li, "Brainomaly: Unsupervised neurologic disease detection utilizing unannotated t1-weighted brain mr images", in "Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision", pp. 7573–7582 (2024).

Sims, J. R., J. A. Zimmer, C. D. Evans, M. Lu, P. Ardayfio, J. Sparks, A. M. Wessels, S. Shcherbinin, H. Wang, E. S. M. Nery *et al.*, "Donanemab in early symptomatic alzheimer disease: the trailblazer-alz 2 randomized clinical trial", Jama **330**, 6, 512–527 (2023).

Smith, S. M., K. L. Miller and T. E. Nichols, "Characterising ongoing brain aging from cross-sectional data", bioRxiv pp. 2025–01 (2025).

Smith, S. M., D. Vidaurre, F. Alfaro-Almagro, T. E. Nichols and K. L. Miller, "Estimation of brain age delta from brain imaging", Neuroimage **200**, 528–539 (2019).

Son, H. J., J. S. Oh, M. Oh, S. J. Kim, J.-H. Lee, J. H. Roh and J. S. Kim, "The clinical feasibility of deep learning-based classification of amyloid pet images in visually equivocal cases", European journal of nuclear medicine and molecular imaging **47**, 332–341 (2020).

Song, J., C. Meng and S. Ermon, "Denoising diffusion implicit models", arXiv preprint arXiv:2010.02502 (2020a).

Song, J., J. Zheng, P. Li, X. Lu, G. Zhu and P. Shen, "An effective multimodal image fusion method using mri and pet for alzheimer's disease diagnosis", Frontiers in digital health **3**, 637386 (2021).

Song, T.-A., S. R. Chowdhury, F. Yang and J. Dutta, "Pet image super-resolution using generative adversarial networks", Neural Networks **125**, 83–91 (2020b).

Sørensen, L., M. Nielsen, A. D. N. Initiative *et al.*, "Ensemble support vector machine classification of dementia using structural mri and mini-mental state examination", Journal of neuroscience methods **302**, 66–74 (2018).

Spaulding, J., H. Noda, M. N. Shirazi and E. Kawaguchi, "Bpcs steganography using ezw lossy compressed images", Pattern Recognition Letters **23**, 13, 1579–1587 (2002).

Su, Y., T. M. Blazey, C. J. Owen, J. J. Christensen, K. Friedrichsen, N. Joseph-Mathurin, Q. Wang, R. C. Hornbeck, B. M. Ances, A. Z. Snyder *et al.*, "Quantitative amyloid imaging in autosomal dominant alzheimer's disease: results from the dian study group", PLoS One **11**, 3, e0152082 (2016).

Su, Y., T. M. Blazey, A. Z. Snyder, M. E. Raichle, D. S. Marcus, B. M. Ances, R. J. Bateman, N. J. Cairns, P. Aldea, L. Cash *et al.*, "Partial volume correction in quantitative amyloid imaging", Neuroimage **107**, 55–64 (2015).

Su, Y., G. M. D'Angelo, A. G. Vlassenko, G. Zhou, A. Z. Snyder, D. S. Marcus, T. M. Blazey, J. J. Christensen, S. Vora, J. C. Morris *et al.*, "Quantitative analysis of pib-pet with freesurfer rois", PloS one **8**, 11, e73377 (2013).

Sun, J., Z. Xu and H.-Y. Shum, "Image super-resolution using gradient profile prior", in "2008 IEEE conference on computer vision and pattern recognition", pp. 1–8 (IEEE, 2008).

Sun, P., Y. Jiang, S. Chen, S. Zhang, B. Peng, P. Luo and Z. Yuan, "Autoregressive model beats diffusion: Llama for scalable image generation", arXiv preprint arXiv:2406.06525 (2024).

Takida, Y., Y. Ikemiya, T. Shibuya, K. Shimada, W. Choi, C.-H. Lai, N. Murata, T. Uesaka, K. Uchida, W.-H. Liao *et al.*, "Hq-vae: Hierarchical discrete representation learning with variational bayes", arXiv preprint arXiv:2401.00365 (2023).

Takida, Y., T. Shibuya, W. Liao, C.-H. Lai, J. Ohmura, T. Uesaka, N. Murata, S. Takahashi, T. Kumakura and Y. Mitsufuji, "Sq-vae: Variational bayes on discrete representation with self-annealed stochastic quantization", arXiv preprint arXiv:2205.07547 (2022).

Teo, B.-K., Y. Seo, S. L. Bacharach, J. A. Carrasquillo, S. K. Libutti, H. Shukla, B. H. Hasegawa, R. A. Hawkins and B. L. Franc, "Partial-volume correction in pet: validation of an iterative postreconstruction method with phantom and patient data", Journal of Nuclear Medicine **48**, 5, 802–810 (2007).

Thambawita, V., P. Salehi, S. A. Sheshkal, S. A. Hicks, H. L. Hammer, S. Parasa, T. d. Lange, P. Halvorsen and M. A. Riegler, "Singan-seg: Synthetic training data generation for medical image segmentation", PloS one **17**, 5, e0267976 (2022).

Thomas, B. A., V. Cuplov, A. Bousse, A. Mendes, K. Thielemans, B. F. Hutton and K. Erlandsson, "Petpvc: a toolbox for performing partial volume correction techniques in positron emission tomography", Physics in Medicine & Biology **61**, 22, 7975 (2016).

Thomas, B. A., K. Erlandsson, M. Modat, L. Thurfjell, R. Vandenberghe, S. Ourselin and B. F. Hutton, "The importance of appropriate partial volume correction for pet quantification in alzheimer's disease", European journal of nuclear medicine and molecular imaging **38**, 1104–1119 (2011).

Tohka, J. and A. Reilhac, "Deconvolution-based partial volume correction in raclopride-pet and monte carlo comparison to mr-based method", Neuroimage **39**, 4, 1570–1584 (2008).

Treder, M. S., J. P. Shock, D. J. Stein, S. Du Plessis, S. Seedat and K. A. Tsvetanov, "Correlation constraints for regression models: Controlling bias in brain age prediction", Frontiers in psychiatry **12**, 615754 (2021).

Trivedi, M. R., A. M. Joshi, J. Shah, B. Readhead, Y. Su, E. M. Reiman, T. Wu and Q. Wang, "Interpretable deep learning framework towards understanding molecular changes in human brains with alzheimer's disease: implication for microglia activation and sex differences in ad", bioRxiv pp. 2023–12 (2023).

Tudosiu, P.-D., W. H. Pinaya, P. Ferreira Da Costa, J. Dafflon, A. Patel, P. Borges, V. Fernandez, M. S. Graham, R. J. Gray, P. Nachev *et al.*, "Realistic morphology-preserving generative modelling of the brain", Nature Machine Intelligence **6**, 7, 811–819 (2024).

Valizadeh, S., J. Hänggi, S. Mérillat and L. Jäncke, "Age prediction on the basis of brain anatomical measures", Human brain mapping **38**, 2, 997–1008 (2017).

Van Den Oord, A., O. Vinyals *et al.*, "Neural discrete representation learning", Advances in neural information processing systems **30** (2017).

Vandendriessche, D., J. Uribe, H. Bertin and F. De Geeter, "Performance characteristics of silicon photomultiplier based 15-cm afov tof pet/ct", EJNMMI physics **6**, 1–17 (2019).

Venugopalan, J., L. Tong, H. R. Hassanzadeh and M. D. Wang, "Multimodal deep learning models for early detection of alzheimer's disease stage", Scientific reports **11**, 1, 3254 (2021).

Wang, H., Y. Zhou, J. Zhang, J. Lei, D. Sun, F. Xu and X. Xu, "Anomaly segmentation in retinal images with poisson-blending data augmentation", Medical Image Analysis **81**, 102534 (2022).

Wang, J., J. Yang, K. Yu, F. Lv, T. Huang and Y. Gong, "Locality-constrained linear coding for image classification", in "2010 IEEE computer society conference on computer vision and pattern recognition", pp. 3360–3367 (IEEE, 2010).

Wang, Z., J. Chen and S. C. Hoi, "Deep learning for image super-resolution: A survey", IEEE transactions on pattern analysis and machine intelligence **43**, 10, 3365–3387 (2020).

Weiner, M. W., D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, R. C. Green, D. Harvey, C. R. Jack Jr, W. Jagust, J. C. Morris *et al.*, "The alzheimer's disease neuroimaging initiative 3: Continued innovation for clinical trial improvement", Alzheimer's & Dementia **13**, 5, 561–571 (2017).

Whittington, A., R. N. Gunn, A. D. N. Initiative *et al.*, "Tauiq: A canonical image based algorithm to quantify tau pet scans", Journal of Nuclear Medicine **62**, 9, 1292–1300 (2021).

Xia, T., A. Chartsias, C. Wang, S. A. Tsaftaris, A. D. N. Initiative *et al.*, "Learning to synthesise the ageing brain without longitudinal data", Medical Image Analysis **73**, 102169 (2021).

Xin, G. and P. Fan, "A lossless compression method for multi-component medical images based on big data mining", Scientific Reports **11**, 1, 12372 (2021).

Yan, M., J. Wu, R. Shah and D. Liu, "Gaussian mixture vector quantization with aggregated categorical posterior", arXiv preprint arXiv:2410.10180 (2024).

Yang, J., J. Wright, T. S. Huang and Y. Ma, "Image super-resolution via sparse representation", IEEE transactions on image processing **19**, 11, 2861–2873 (2010).

Yang, J., K. Yu, Y. Gong and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification", in "2009 IEEE Conference on computer vision and pattern recognition", pp. 1794–1801 (IEEE, 2009).

Yoshimura, T., A. Hasegawa, S. Kogame, K. Magota, R. Kimura, S. Watanabe, K. Hirata and H. Sugimori, "Medical radiation exposure reduction in pet via super-resolution deep learning model", Diagnostics **12**, 4, 872 (2022).

Younis, K. S., S. K. Rogers and M. P. DeSimio, "Vector quantization based on dynamic adjustment of mahalanobis distance", in "Proceedings of the IEEE 1996 National Aerospace and Electronics Conference NAECON 1996", vol. 2, pp. 858–862 (IEEE, 1996).

Yu, Y., H.-Q. Cui, S. S. Haas, F. New, N. Sanford, K. Yu, D. Zhan, G. Yang, J.-H. Gao, D. Wei *et al.*, "Brain-age prediction: Systematic evaluation of site effects, and sample age range and size", Human Brain Mapping **45**, 10, e26768 (2024).

Zeghidour, N., A. Luebs, A. Omran, J. Skoglund and M. Tagliasacchi, "Soundstream: An end-to-end neural audio codec", IEEE/ACM Transactions on Audio, Speech, and Language Processing **30**, 495–507 (2021).

Zhang, B., S. Zhang, J. Feng and S. Zhang, "Age-level bias correction in brain age prediction", NeuroImage: Clinical **37**, 103319 (2023a).

Zhang, Q. and Y. Chen, "Diffusion normalizing flow", Advances in neural information processing systems **34**, 16280–16291 (2021).

Zhang, S., L. Yang, M. B. Mi, X. Zheng and A. Yao, "Improving deep regression with ordinal entropy", arXiv preprint arXiv:2301.08915 (2023b).

Zhao, H., O. Gallo, I. Frosio and J. Kautz, "Loss functions for image restoration with neural networks", IEEE Transactions on computational imaging **3**, 1, 47–57 (2016).

Zhu, J.-D., Y.-F. Wu, S.-J. Tsai, C.-P. Lin and A. C. Yang, "Investigating brain aging trajectory deviations in different brain regions of individuals with schizophrenia using multimodal magnetic resonance imaging and brain-age prediction: a multicenter study", Translational Psychiatry **13**, 1, 82 (2023).

Zou, H. and T. Hastie, "Regularization and variable selection via the elastic net", Journal of the Royal Statistical Society Series B: Statistical Methodology **67**, 2, 301–320 (2005).

APPENDIX A

STATEMENT OF CO-AUTHOR PERMISSIONS

I, Jay Shah, hereby affirm that all co-authors of the works included in this document have granted their permission for the inclusion of the articles in my dissertation. Each co-author is aware that the articles will appear as part of this academic work and has consented to its use in this context.

APPENDIX B

LOSS FUNCTION FOR LDM-RR: PET RESOLUTION RECOVERY FRAMEWORK

Substituting z and $\hat{z_0}$ from equations 5 and 7, loss_1 can be simplified as:

$$\begin{aligned}
\text{loss}_1 &= (1-\alpha)\left\|\frac{z_t-\sqrt{1-\alpha_t}\epsilon}{\sqrt{\alpha_t}}-\frac{z_t-\sqrt{1-\alpha_t}\hat{\epsilon}}{\sqrt{\alpha_t}}\right\|_2^2+\alpha\cdot\text{MSSSIM}(z_0,\hat{z_0}) \\
&= (1-\alpha)\left(\frac{z_t-\sqrt{1-\alpha_t}}{\sqrt{\alpha_t}}\right)\|\epsilon-\hat{\epsilon}\|_2^2+\alpha\cdot\text{MSSSIM}(z_0,\hat{z_0}) \\
&= \gamma(1-\alpha)\|\epsilon-\hat{\epsilon}\|_2^2+\alpha\cdot\text{MSSSIM}(z_0,\hat{z_0})
\end{aligned}$$

And substituting loss_1, where $\gamma=\frac{z_t-\sqrt{1-\alpha_t}}{\sqrt{\alpha_t}}$ in the combined loss equation we get:

$$\text{loss}_{\text{combined}} = L_1(\epsilon,\hat{\epsilon})+\gamma(1-\alpha)L_2(\epsilon,\hat{\epsilon})+\alpha\cdot\text{MSSSIM}(z_0,\hat{z_0})$$

APPENDIX C

FIGURES FOR RIED-NET

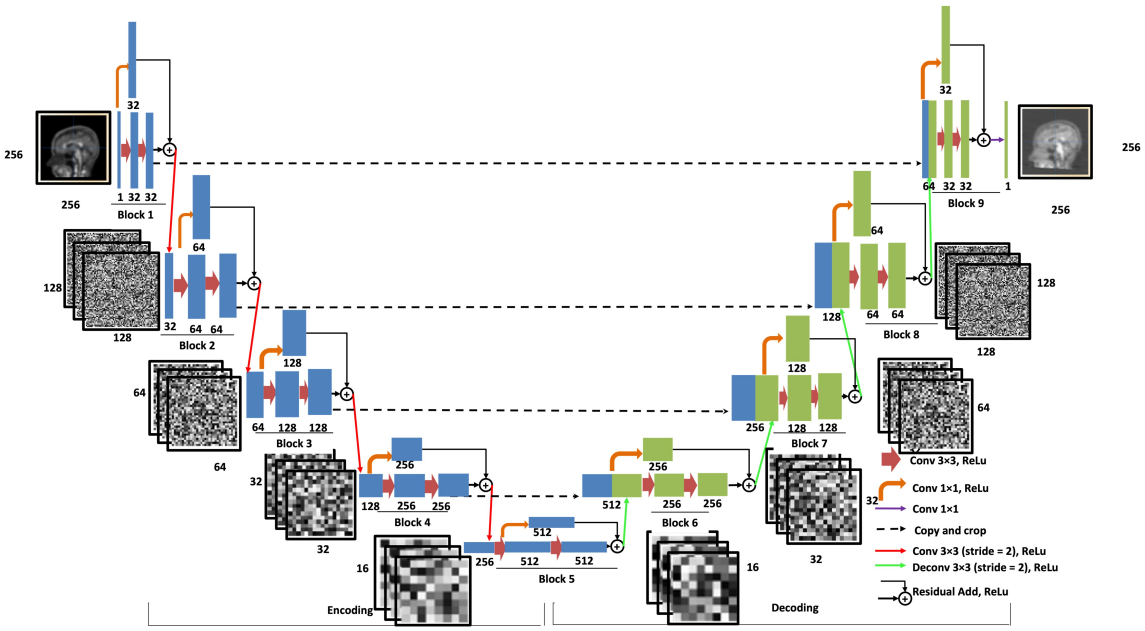

Figure C.1: Overall Architecture of RIED-Net. It uses a U-Net-like structure with a residual inception shortcut to enhance training efficiency. It includes 5 encoding and 4 decoding blocks. Input is a 2D florbetapir image slice (256×256×1). Each block processes data via 3×3 or 1×1 convolutions (ReLU), with skip connections between corresponding encoder-decoder layers. The final layer outputs a synthetic 2D PiB slice using a 1×1 convolution. Convolutional, deconvolutional, and copy operations are represented with distinct arrows.

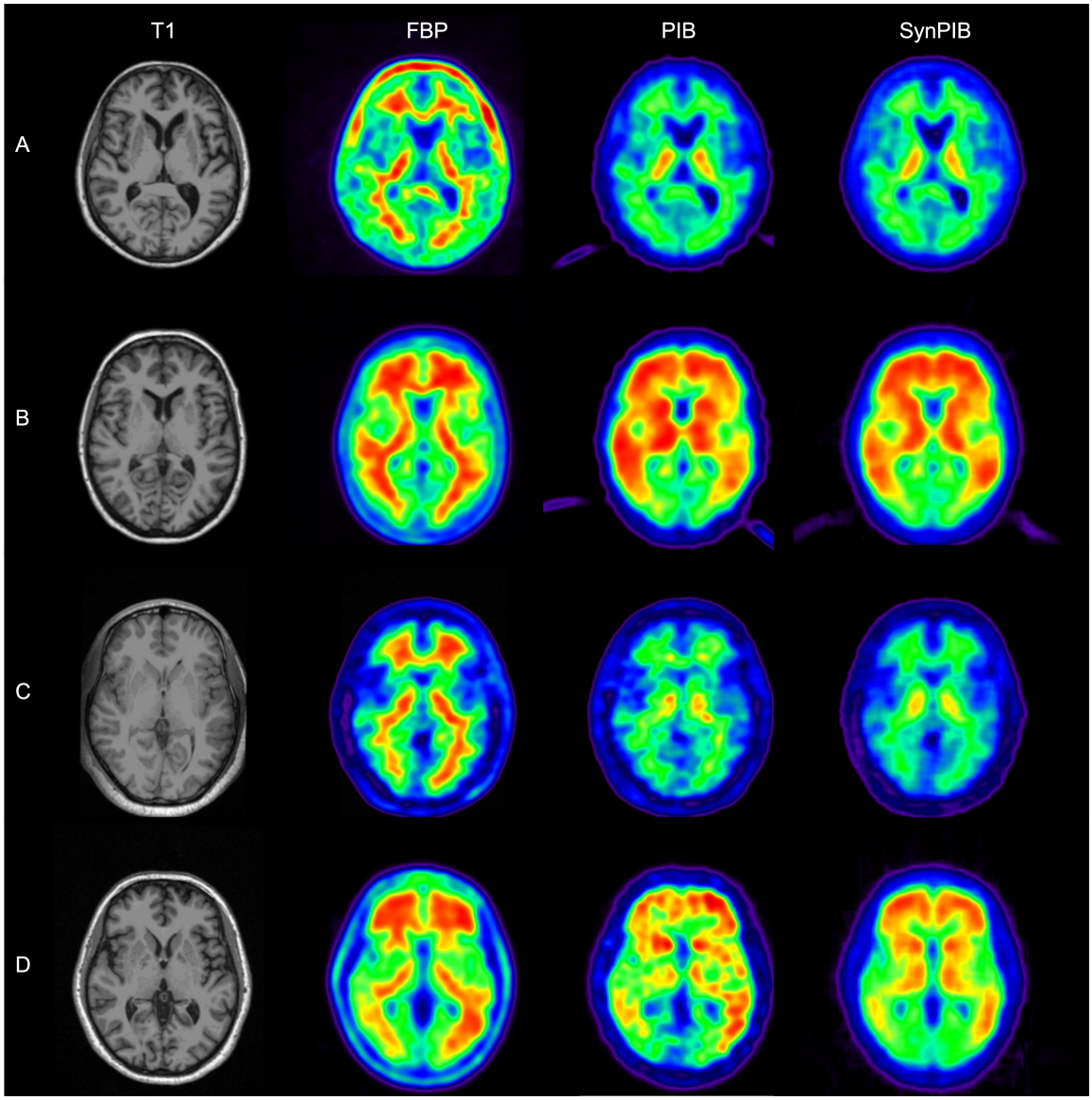

Figure C.2: Visual Comparison of Synthetic Images Generated Using RIED-net to Real Pittsburgh Compound-b (PIB) Data for the OASIS (Panels a and B) and GAAIN (Centiloid Project; Panels C and D) Datasets Used in Our Study. Panels A and C Show Representative Images from Amyloid Negative Participants, and Panels B & D Are Examples from Amyloid Positive Participants. Abbreviations: FBP, Florbetapir; SynPIB, Synthetic PIB.

APPENDIX D

FIGURES FOR PHYSICAL ACTIVITY DATA

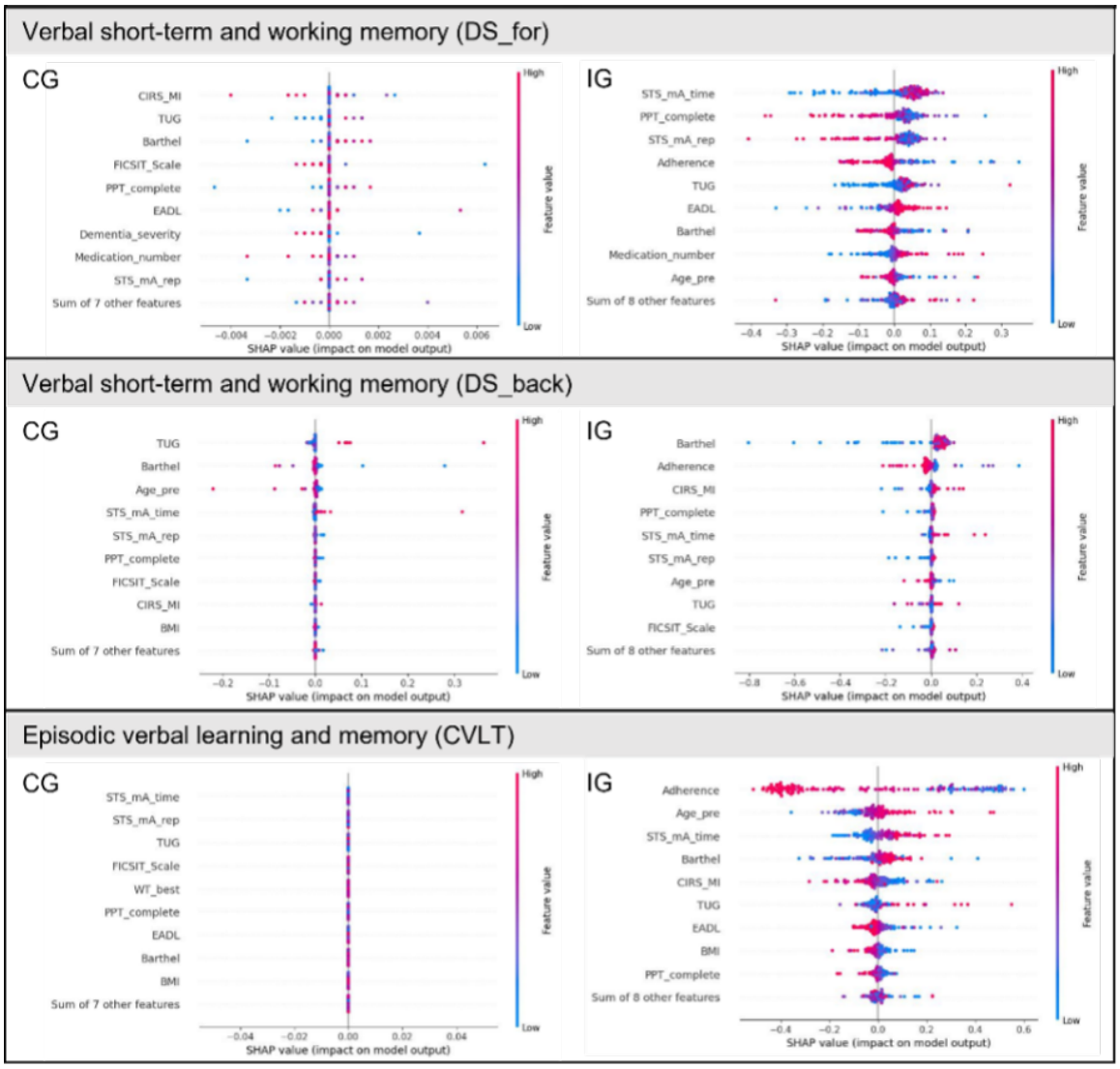

Figure D.1: Classification of Decline Vs. Non-decline in the Cognitive Outcome Variables, Presented by SHAP Plots for CG and IG.

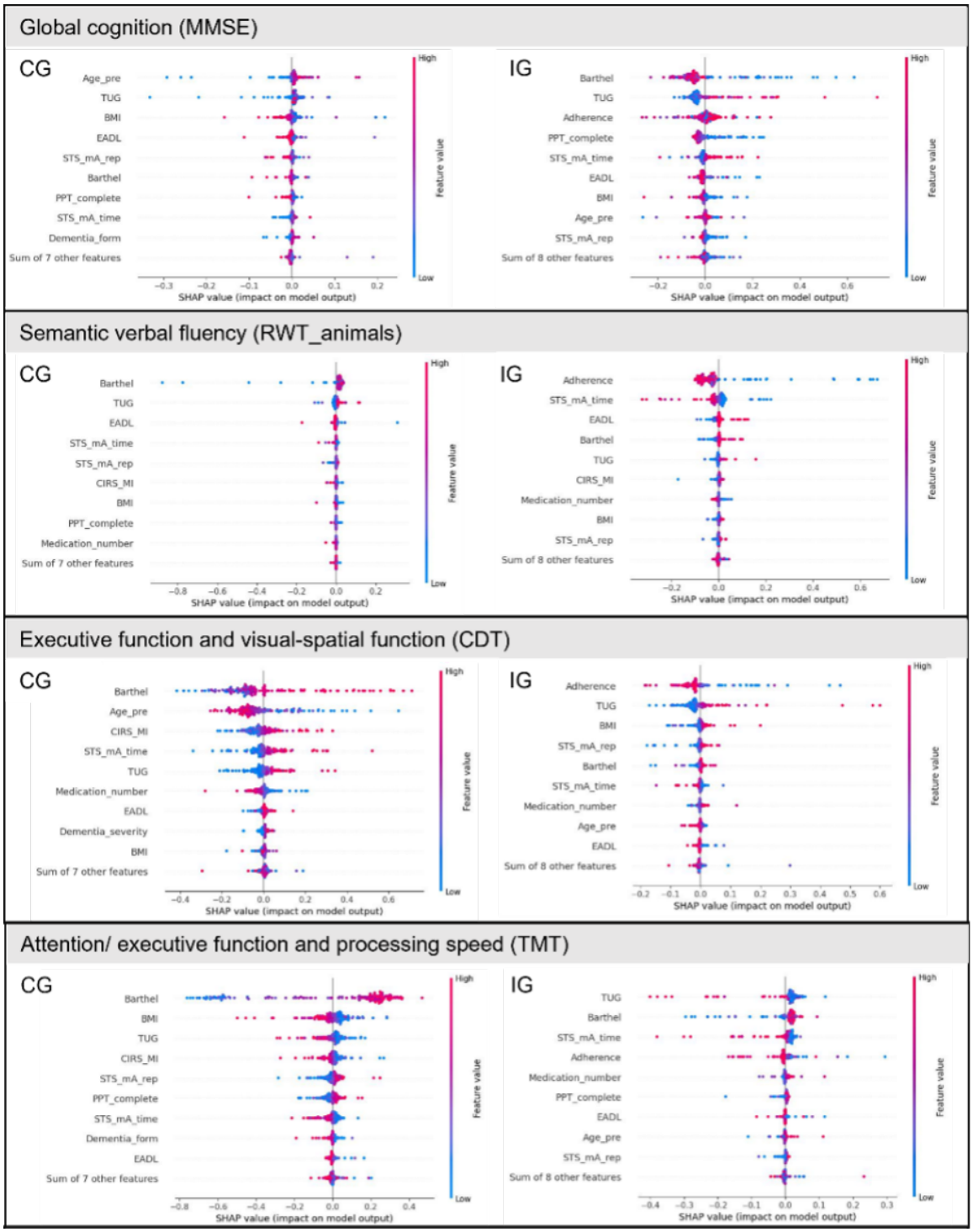

Figure D.2: (Same as the above Plot) Classification of Decline Vs. Non-decline in the Cognitive Outcome Variables, Presented by SHAP Plots for CG and IG.

APPENDIX E

RESULTS FOR LCVQ

CIFAR-10, 1024			**MNIST, 1024**		
Method	**MSE**	**PER %**	**Method**	**MSE**	**PER %**
Baseline	0.054	4.980	Baseline	0.035	4.292
AR	0.054	4.590	Affine	0.034	6.333
LCVQ	0.055	76.450	LCVQ	0.032	73.938
CIFAR-10, 2048			**MNIST, 2048**		
Baseline	0.054	2.441	Baseline	0.038	1.172
Affine	0.053	3.076	Affine	0.038	1.610
LCVQ	0.052	46.036	LCVQ	0.031	54.562

Table E.1: Comparison of Baseline Vqvae, Affine Reparameterization (AR), and Proposed LCVQ Quantization Methods on CIFAR-10 and MNIST Datasets, Evaluated by Mean Squared Error (MSE) and Codebook Utilization Using Perplexity (PER%).