INVESTIGATING CONVERSION FROM MILD COGNITIVE IMPAIRMENT TO ALZHEIMER'S DISEASE USING LATENT SPACE MANIPULATION

by

Deniz Sezin Ayvaz

B.S., Industrial Engineering, Boğaziçi University, 2019

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Computer Engineering

Boğaziçi University

2023

ACKNOWLEDGEMENTS

The NACC database is funded by NIA/NIH Grant U24 AG072122. NACC data are contributed by the NIA-funded ADRCs: P30 AG062429 (PI James Brewer, MD, PhD), P30 AG066468 (PI Oscar Lopez, MD), P30 AG062421 (PI Bradley Hyman, MD, PhD), P30 AG066509 (PI Thomas Grabowski, MD), P30 AG066514 (PI Mary Sano, PhD), P30 AG066530 (PI Helena Chui, MD), P30 AG066507 (PI Marilyn Albert, PhD), P30 AG066444 (PI John Morris, MD), P30 AG066518 (PI Jeffrey Kaye, MD), P30 AG066512 (PI Thomas Wisniewski, MD), P30 AG066462 (PI Scott Small, MD), P30 AG072979 (PI David Wolk, MD), P30 AG072972 (PI Charles DeCarli, MD), P30 AG072976 (PI Andrew Saykin, PsyD), P30 AG072975 (PI David Bennett, MD), P30 AG072978 (PI Neil Kowall, MD), P30 AG072977 (PI Robert Vassar, PhD), P30 AG066519 (PI Frank LaFerla, PhD), P30 AG062677 (PI Ronald Petersen, MD, PhD), P30 AG079280 (PI Eric Reiman, MD), P30 AG062422 (PI Gil Rabinovici, MD), P30 AG066511 (PI Allan Levey, MD, PhD), P30 AG072946 (PI Linda Van Eldik, PhD), P30 AG062715 (PI Sanjay Asthana, MD, FRCP), P30 AG072973 (PI Russell Swerdlow, MD), P30 AG066506 (PI Todd Golde, MD, PhD), P30 AG066508 (PI Stephen Strittmatter, MD, PhD), P30 AG066515 (PI Victor Henderson, MD, MS), P30 AG072947 (PI Suzanne Craft, PhD), P30 AG072931 (PI Henry Paulson, MD, PhD), P30 AG066546 (PI Sudha Seshadri, MD), P20 AG068024 (PI Erik Roberson, MD, PhD), P20 AG068053 (PI Justin Miller, PhD), P20 AG068077 (PI Gary Rosenberg, MD), P20 AG068082 (PI Angela Jefferson, PhD), P30 AG072958 (PI Heather Whitson, MD), P30 AG072959 (PI James Leverenz, MD).

This study utilizes the TADPOLE data sets [1], which were developed by the EuroPOND consortium [2] and supported by the European Union's Horizon 2020 research and innovation program under grant agreement No. 666992. The TADPOLE data were originally sourced from the Alzheimer's Disease Neuroimaging Initiative (ADNI). The data collection and sharing efforts for this project were financially supported by the Alzheimer's Disease Neuroimaging Initiative (ADNI) through the National Institutes

of Health Grant U01 AG024904 and the Department of Defense ADNI (award number W81XWH-12-2-0012). ADNI receives funding from the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, and from generous contributions by the following sources: AbbVie, Alzheimer's Association; Alzheimer's Drug Discovery Foundation; Araclon Biotech; BioClinica, Inc.; Biogen; Bristol-Myers Squibb Company; CereSpir, Inc.; Cogstate; Eisai Inc.; Elan Pharmaceuticals, Inc.; Eli Lilly and Company; EuroImmun; F. Hoffmann-La Roche Ltd and its affiliated company Genentech, Inc.; Fujirebio; GE Healthcare; IXICO Ltd.; Janssen Alzheimer Immunotherapy Research & Development, LLC.; Johnson & Johnson Pharmaceutical Research & Development LLC.; Lumosity; Lundbeck; Merck & Co., Inc.; Meso Scale Diagnostics, LLC.; NeuroRx Research; Neurotrack Technologies; Novartis Pharmaceuticals Corporation; Pfizer Inc.; Piramal Imaging; Servier; Takeda Pharmaceutical Company; and Transition Therapeutics. The Canadian Institutes of Health Research (CIHR) are supplying financial resources to assist ADNI clinical centers located in Canada. Contributions from the private sector are facilitated through the Foundation for the National Institutes of Health (FNIH) [3]. The organization receiving the grant is the Northern California Institute for Research and Education, and the study's management is overseen by the Alzheimer's Therapeutic Research Institute at the University of Southern California. The distribution of ADNI data is managed by the Laboratory for Neuro Imaging, also affiliated with the University of Southern California.

This thesis was supported by Boğaziçi University Research Fund under grant number 17004 and by the Scientific and Technological Research Council of Turkey ARDEB 3501 under grant number 121E594.

The images used in this thesis, originally featured in my conference proceedings article [4], are included under the Attribution-NonCommercial-NoDerivs 4.0 International license, which protects the article. I thank my thesis supervisor Assist. Prof. İnci Meliha Baytaş for her guidance, and Assoc. Prof. Arzucan Özgür and Assoc. Prof. Öznur Taştan for being on the thesis defense jury and providing valuable feedback.

ABSTRACT

INVESTIGATING CONVERSION FROM MILD COGNITIVE IMPAIRMENT TO ALZHEIMER'S DISEASE USING LATENT SPACE MANIPULATION

Alzheimer's disease (AD), a progressive neurological disorder, is the most common cause of dementia, affecting millions worldwide. It severely affects cognitive abilities, such as speech and memory. Due to its multi-modal effects, the disease has a heterogeneous progression pattern, which makes tracking the patient's cognitive impairment level challenging. Mild Cognitive Impairment (MCI) is considered an intermediate stage before AD. Early prediction of the conversion from MCI to AD is crucial to taking necessary precautions for decelerating the disease progression and developing suitable treatments. The problem of early prediction of AD among MCI patients has often been posed as a classification problem. However, the classification models may not be sufficient to disclose the underlying factors of the transition from MCI to AD. This thesis proposes an intuitive framework to investigate the underlying causes and risk factors of the conversion from MCI to AD. The proposed deep learning-based framework employs latent space manipulation techniques to obtain principal directions toward AD diagnosis for an MCI patient and analyze the changes the patient's attributes undergo during disease progression. The predictive ability of the proposed framework is evaluated by correlating the magnitude of the manipulation in the latent space of a variational auto-encoder trained with MCI and AD patients with the possibility of conversion in later stages. Experimental results show promising quantitative and qualitative outcomes on two publicly available and commonly used AD neuroimaging data sets.

ÖZET

SAKLI UZAY MANİPÜLASYONU KULLANARAK HAFİF BİLİŞSEL BOZUKLUKTAN ALZHEIMER HASTALIĞINA DÖNÜŞÜMÜN İNCELENMESİ

İlerleyici bir nörolojik bozukluk olan Alzheimer hastalığı (AH), dünya çapında milyonlarca insanı etkileyen demansın en yaygın nedenidir. Konuşma ve hafıza gibi bilişsel yetenekleri ciddi şekilde etkiler. Çok yönlü etkileri nedeniyle hastalığın heterojen bir ilerleme örüntüsüne sahip olması, hastanın bilişsel bozukluk düzeyini takip etmeyi zorlaştırmaktadır. Hafif Bilişsel Bozukluk (HBB), AH'den önce bir ara aşama olarak kabul edilir. HBB'den AH'ye dönüşümün erken tahmin edilmesi, hastalığın ilerlemesinin yavaşlatılması ve uygun tedavilerin geliştirilmesi için gerekli önlemlerin alınması açısından büyük önem taşımaktadır. HBB hastalarında AH'nin erken tahmin edilmesi sorunu, sıklıkla bir sınıflandırma problemi olarak ortaya konulmuştur. Ancak sınıflandırma modelleri HBB'den AH'ye geçişin altında yatan nedenleri açıklamakta yeterli olmayabilir. Bu tez, HBB'den AH'ye dönüşümün altında yatan nedenleri ve risk faktörlerini araştırmak için sezgisel bir çerçeve önermektedir. Önerilen derin öğrenmeye dayalı çerçeve, bir HBB hastası için AH tanısına yönelik temel talimatları elde etmek ve hastalığın ilerlemesi sırasında hastanın niteliklerinde meydana gelen değişiklikleri analiz etmek için saklı uzay manipülasyon tekniklerini kullanır. Önerilen çerçevenin öngörme yeteneği, HBB ve AH hasta verisi ile eğitilmiş bir değişimsel otokodlayıcının saklı uzayındaki manipülasyonun büyüklüğünü, daha sonraki aşamalarda dönüşüm olasılığı ile ilişkilendirerek değerlendirilir. Deneysel sonuçlar, halka açık ve yaygın olarak kullanılan iki AH nörogörüntüleme veri kümesinde umut verici nicel ve nitel sonuçlar göstermektedir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii
ABSTRACT v
ÖZET vi
LIST OF FIGURES ix
LIST OF TABLES xiii
LIST OF ACRONYMS/ABBREVIATIONS xv
1. INTRODUCTION 1
2. LITERATURE REVIEW 4
2.1. Clinical Studies on Alzheimer's Disease 5
2.2. Deep Learning on Disease Progression 6
2.3. Deep Learning for Alzheimer's Disease 7
2.4. Latent Space Manipulation 9
3. PROPOSED METHOD 10
3.1. Research Question 10
3.2. Model Architecture 11
3.2.1. Binary Classifier 12
3.2.2. Variational Auto-Encoder 12
3.3. Latent Space Manipulation 13
3.3.1. Unsupervised Latent Space Manipulation Method 14
3.3.2. Supervised Latent Space Manipulation Method 14
3.3.3. Latent Space Manipulation Algorithm 15
3.4. Predictive Power of the Method 16
4. EXPERIMENTS AND RESULTS 19
4.1. Data Sets 19
4.1.1. TADPOLE Data Set 19
4.1.2. NACC Data Set 22
4.2. Feature Selection 24
4.2.1. TADPOLE Data Set 24

4.2.2. NACC Data Set 26
4.3. Data Preprocessing 26
4.3.1. TADPOLE Data Set 26
4.3.2. NACC Data Set 27
4.4. Training Scheme 28
4.5. Model Performances 29
4.5.1. VAE Performance 29
4.5.2. Classifier Model Performances 30
4.6. Supervised and Unsupervised Latent Space Manipulation 32
4.7. Choice of Principal Components 35
4.8. Investigating the Effect of α 35
4.8.1. TADPOLE Data Set 36
4.8.2. NACC Data Set 39
4.9. Visualizing the Classifier's Latent Space 39
4.10. Clinical Interpretation 41
4.10.1. TADPOLE Data Set 42
4.10.2. NACC Data Set 45
4.11. Predictive Performance 49
4.11.1. TADPOLE Data Set 50
4.11.2. NACC Data Set 53
4.12. Data Generation Performance 58
5. CONCLUSION AND FUTURE WORK 60
REFERENCES 63

LIST OF FIGURES

Figure 3.1. List of the main research questions. 11

Figure 3.2. Proposed framework. MCI and AD patient data are used to train a VAE and a classifier. Unit direction vectors derived from latent representations are employed for manipulation, leading to the decoding of manipulated representations. Subsequently, the classifier labels the manipulated patient data. 11

Figure 3.3. Latent space manipulation algorithm. 16

Figure 3.4. Prediction performance measurement of the method. 17

Figure 4.1. Distribution of patient age at baseline visit on TADPOLE data. . 22

Figure 4.2. Distribution of patient age at baseline visit on NACC data. 23

Figure 4.3. t-SNE visualization of VAE latent space representations of the TADPOLE training data set. 31

Figure 4.4. t-SNE visualization of VAE latent space representations of the NACC training data set. 31

Figure 4.5. t-SNE visualization of PC vectors on VAE latent space representations of TADPOLE training data. 33

Figure 4.6. t-SNE visualization of PC vectors on VAE latent space representations of NACC training data. 33

Figure 4.7. t-SNE visualization of unit direction vector obtained via SVM classifier's parameter vector on VAE latent space representations of TADPOLE training data. 34

Figure 4.8. t-SNE visualization of unit direction vector obtained via SVM classifier's parameter vector on VAE latent space representations of NACC training data. 34

Figure 4.9. Manipulation effects of PC vectors for a fixed $\alpha = 5$ on TADPOLE data set. 36

Figure 4.10. Manipulation effects of PC vectors for a fixed $\alpha = 5$ on NACC data set. 36

Figure 4.11. Manipulation effect of the first PC vector with different α values for TADPOLE data set. The horizontal axis represents α values, the vertical axis represents the average probability of AD of the data points after manipulation, and the shaded region denotes the standard deviation. 37

Figure 4.12. Manipulation effect of SUDV with different α values for TADPOLE data. Horizontal axis represents α values, vertical axis represents the average probability of AD of the data points after manipulation, and shaded region denotes the standard deviation. 37

Figure 4.13. Manipulation effect of the first PC vector with different α values on NACC data set. The horizontal axis represents α values, the vertical axis represents the average probability of AD of the data points after manipulation, and the shaded region denotes the standard deviation. 38

Figure 4.14. Manipulation effect of SUDV with different α values on NACC data. Horizontal axis represents α values, vertical axis represents the average probability of AD of the data points after manipulation, and shaded region denotes the standard deviation. 38

Figure 4.15. t-SNE of the manipulated TADPOLE test data points with their original versions in the classifier's latent space, with Figure 4.17 (a) $\alpha = 2$ and Figure 4.17 (b) $\alpha = 4$. Light blue and dark blue represent AD and MCI data points before manipulation and red and pink colors represent AD and MCI data points after manipulation. . . . 40

Figure 4.16. t-SNE of the manipulated NACC test data points with their original versions in the classifier's latent space, with Figure 4.18 (a) $\alpha = 0.5$ and Figure 4.18 (b) $\alpha = 1$. Light blue and dark blue represent AD and MCI data points before manipulation and red and pink colors represent AD and MCI data points after manipulation. 40

Figure 4.17. Correlations of the average changes in variables of the MCI patients who convert to AD after manipulation with the first PC with varying α for TADPOLE data set. 43

Figure 4.18. The ratio of conversion detection with different α values in TADPOLE data set. 43

Figure 4.19. The ratio of the number of manipulated patients with different α values in TADPOLE data set. 44

Figure 4.20. Correlations of the average changes in variables of the MCI patients who convert to AD after manipulation with the first PC with varying α for NACC data set. 47

Figure 4.21. The ratio of conversion detection with different α values in NACC data set. 47

Figure 4.22. The ratio of the number of manipulated patients with different α values in NACC data set. 48

Figure 4.23. ROC-AUC, sensitivity, and specificity metrics for different α values with the unsupervised manipulation method, with visit order = 1, and forecast horizon = 3, on TADPOLE data set. 53

Figure 4.24. ROC-AUC, sensitivity, and specificity metrics for different α values with the unsupervised manipulation method, with visit order = 1, and forecast horizon = 3, on NACC data set. 57

LIST OF TABLES

Table 4.1. Set of different biomarker types with their definitions in TADPOLE data set. 21

Table 4.2. Set of variables with their statistics and percentage of missing values in TADPOLE data set. 25

Table 4.3. Set of variables with their statistics and percentage of missing values in NACC data set. 26

Table 4.4. Performance metrics of classifiers demonstrating VAE performance for TADPOLE and NACC data sets. 30

Table 4.5. Performance metrics of classifiers for TADPOLE and NACC data sets. 32

Table 4.6. Average changes between the manipulated and original time steps labeled AD for a fixed $\alpha = 3$, for TADPOLE data set. 45

Table 4.7. Average changes between the manipulated and original time steps labeled AD for a fixed $\alpha = 1.2$, for NACC data set. 48

Table 4.8. Performance metrics for each visit order - forecast horizon pair of the predictions of our method, logistic regression model, and gradient boosting model as baseline models for TADPOLE data set. Sens: sensitivity, spec: specificity, LR: logistic regression, GB: gradient boosting model. 51

Table 4.9. Average grace period, number of AD patients, and the optimal α values for each visit order - forecast horizon pair of the predictions of our method, for TADPOLE data set. 52

Table 4.10. p-values of two-sample t-test for each biomarker value, by comparing the values of each patient with its manipulated data and its future visit data, for TADPOLE data set. 54

Table 4.11. Performance metrics for each visit order - forecast horizon pair of the predictions of our method, logistic regression model, and gradient boosting model as baseline models for NACC data set. Sens: sensitivity, spec: specificity, LR: logistic regression, GB: gradient boosting model. 55

Table 4.12. Average grace period, number of AD patients, and the optimal α values for each visit order - forecast horizon pair of the predictions of our method, for NACC data set. 56

Table 4.13. p-values of two-sample t-test for each biomarker value, by comparing the values of each patient with its manipulated data and its future visit data, for NACC data set. 57

Table 4.14. Performance metrics of the classifier trained with a mix of actual and VAE-generated data, demonstrating VAE data generation performance for TADPOLE and NACC data sets. 59

LIST OF ACRONYMS/ABBREVIATIONS

3D	Three Dimensional
cmm	Cubic Milimeters

1. INTRODUCTION

Neurodegenerative diseases are a common concern, threatening millions of lives worldwide. Alzheimer's Disease (AD) is the most well-known cause of dementia that affects cognitive functions [5]. It is a chronic disease that demonstrates heterogeneity in progression [6]. Although most patients show the same clinical outcomes in the final stage of the disease, multiple subtypes exist [7]. One reason for that is that AD is affected by many factors. The indicators of those factors serve as a great deal of multi-modal data for disease progression and patient diagnosis tasks, which are mainly grouped as clinical, cognitive, neurophysiological test results, and genotype [8]. Therefore, data-driven techniques enable potential solutions to progression and diagnosis tasks.

Mild cognitive impairment (MCI) is the intermediate state between a cognitively normal individual and an AD patient [9]. The critical assessment is whether the patient with MCI will be diagnosed with AD. Approximately 12% of older adults with MCI are likely to develop AD in the future, compared to the 1% to 2% of cognitively normal elderly who develop dementia each year in the United States [10]. Some patients who exhibit MCI symptoms develop AD in the future. However, some may remain in the MCI stage or progress very slowly [11]. For this reason, early detection of the conversion between MCI and AD is essential for timely treatment planning that improves the life quality of the patients and their families.

For the early diagnosis, it is imperative to take the clinical history of a patient and their families, perform a neurological examination and neuropsychological tests suitable to the education level, and examine the brain tissue using neuroimaging techniques like Magnetic Resonance Imaging (MRI), and Positron Emission Tomography (PET). Today, machine learning and deep learning techniques can assist clinicians in simultaneously analyzing multiple modalities and developing predictive models for early AD diagnosis. Most previous work has focused on either time series forecasting

or classification on cross-sectional data [12, 13]. However, it is also crucial to know the underlying reasons behind the behavior of a model. These studies lack an explanation for this shift in diagnosis. By contrast, one of our goals is to effectively demonstrate what changes lead a patient with an MCI diagnosis to become an AD patient. This improvement discovers the latent space of healthcare data and has a social impact on guiding healthcare professionals on disease progression.

In this thesis, we aim to decipher the changes in multi-modal measurements, including numerical measurements from MRI data and cognitive scores, which lead to the conversion of MCI to AD by combining a deep learning framework with the latent space manipulation method adapted from the image generation domain. Our study was the first attempt to apply latent space manipulation to healthcare data [4].

Our proposed approach serves as a tool for capturing the progression from MCI to AD by simulating comparable changes in MCI patient data that align with their eventual AD-diagnosed data. This method also enables short and long-term predictions, facilitating the study of the correlation between the magnitude of manipulation and disease progression. In addition, we assess its reliability in generating synthetic patient data in situations with limited available data.

In summary, our approach leverages a variational auto-encoder (VAE) network, which is trained to capture the latent representations of both MCI and AD patient data, entirely devoid of any labels. This latent space serves as the foundation for applying supervised and unsupervised manipulation techniques to the representations of MCI patients. Subsequently, the VAE decoder transforms these manipulated latent representations into entirely new patient data.

Our proposed methodology not only reveals significant associations between directions in the latent space and AD diagnosis but also establishes strong correlations between the generated AD patients via latent space manipulation and actual MCI patients who are later diagnosed with AD. This outcome reinforces the clinical validity

of our approach, as it effectively simulates a realistic progression from MCI to AD. Consequently, our framework proves to be more informative than a binary classifier in uncovering the intricacies of this conversion process.

Furthermore, we substantiate the predictive capacity of our method for patients transitioning from MCI to AD, both in the short-term and long-term. In addition to its predictive prowess, our generative network demonstrates potential utility as a synthetic patient data generator.

2. LITERATURE REVIEW

Dementia is a debilitating condition characterized by memory, language, emotional, cognitive, and behavioral impairments, significantly hindering one's ability to perform everyday tasks [14]. One of the most well-known causes and the most common type of dementia that affects cognitive abilities is AD [5]. It is linked to synaptic loss and neurodegeneration, which causes memory loss and other cognitive issues. There is currently no recognized medication that can stop the development of this condition.

The World Health Organization estimates 50 million dementia patients worldwide [15]. By 2050, it is anticipated that there will be more than 131 million AD patients in the world [16]. Researchers have been looking into the cause of AD as the number of sufferers increases. Fortunately, years of study have clarified many uncertainties. For instance, it is well recognized that AD can be brought on by the environment, age, and medication [5]. However, many questions remain about the underlying causes and treatments for AD. Finding biomarkers (medical parameters suggesting a disease) that can predict future clinical decline and be used as outcome indicators in clinical trials of disease-modifying medications is urgently needed to encourage the formation of new therapeutics [17].

MCI is a transitional stage between normal aging and dementia [18]. When enough brain damage has accumulated over time, cognitive signs and disability appear. Initially described in 1999, this situation has been named in different ways, such as predementia AD and MCI, a condition in which people often only have minor memory impairment with relative preservation of other cognitive domains and functioning abilities. These people do not meet the criteria to be diagnosed as dementia [19] or as the prodromal state AD [20].

Since it is crucial to identify MCI patients likely to develop AD for specialized treatment therapies, the body of literature on the subject is expanding. While some

studies only consider imaging methods, others also consider the outcomes of cognitive and physiological assessments. This section reviews clinical studies with traditional machine learning models and studies in the computer science domain that propose novel methods to solve the problem.

2.1. Clinical Studies on Alzheimer's Disease

The diagnosis of AD traditionally relies on a combination of biological and behavioral markers, incorporating MRI and PET measurements, cognitive assessments, and clinical findings [21]. It is worth noting that screening for abnormal tau and amyloid proteins, while essential for precise diagnosis, may not always be feasible [22]. One hindrance is the cost associated with measuring amyloid levels, which typically requires amyloid-sensitive PET scans [23]. This underscores the pressing need for innovative approaches to diagnose AD accurately and delve deeper into its progression.

Despite the presence of common symptoms, significant heterogeneity among AD patients is adding to the complexity [6]. Consequently, a multi-modal approach that incorporates multiple biomarkers proves more effective than a single-modality approach in addressing this heterogeneity and comprehending the longitudinal progression of the disease [24].

Researchers have been actively seeking noninvasive distinguishing identifiers and optimal techniques for combining multiple markers to diagnose AD. For instance, Lu *et al.* propose integrating AD biomarkers with the CARE index, an event-based probabilistic framework [22]. Another study by Sun *et al.* suggest a hierarchical Bayesian model to jointly identify underlying elements from atrophy patterns and cognitive impairments [6].

Some research try to extract new biomarkers instead of directly using the clinical measurements. For instance, Kung *et al.* propose a biomarker called the ratio of principal curvatures (RPC), which may be used to identify the conversion from MCI

and AD [25]. Another study by Kuang *et al.* investigate a different data set gathered through a cognitive questionnaire. The authors employ simpler models like Multi-Layer Perceptron (MLP) and logistic regression. Although the models utilized in the study may be more interpretable, relying solely on the questionnaire data set is a restriction [26].

2.2. Deep Learning on Disease Progression

Deep learning models are proven to be effective in disease progression modeling. Within the scope of disease progression modeling, various studies have harnessed the capabilities of deep models to extract latent representations. Baytaş *et al.* [27] used LSTM autoencoder to extract the latent representations, which effectively summarize the time series and then cluster patients with K-Means using these representations as input. In another approach called Deep-Care [28], a vector representation for each visit was learned, and then LSTM handled multiple visit data. Doctor AI model [29] used a vector representation for each timestamp as input to forecast both diagnosis and time until the next visit.

Using a technique that comprises many networks, including an encoder, a selector, a predictor, and an embedding dictionary, Lee and Schaar [30] have proposed a deep learning strategy for temporal phenotyping of disease progression. These elements work together by optimizing joint loss functions to deliver the cluster assignment and the accompanying centroid based on a given sequence of observations. This promotes uniform future outcomes for each cluster. In the context of this thesis, we propose harnessing a VAE network to derive distributions of latent representations. This technique fosters the convergence of patient visit samples with similar conditions within the latent space.

2.3. Deep Learning for Alzheimer's Disease

In the realm of computer science, addressing the progression of neurodegenerative diseases and early detection of MCI to AD conversion presents both supervised and unsupervised learning challenges [13, 31]. Some of these challenges stem from the temporal, diverse, multivariate, multi-view, and nonlinear nature of patient data. As a result, conventional machine learning approaches may not always suffice. Therefore, deep learning methods are preferred in the healthcare field due to their ability to uncover intricate data relationships.

In deep learning research aimed at early AD diagnosis, it is common to encounter CNN architectures designed for feature extraction and segmentation [12], [32–39]. These studies harness robust deep learning frameworks to predict AD diagnosis by analyzing MRI and PET images.

Another area of extensive research in AD diagnosis and progression centers around the creation of deep frameworks leveraging a rich array of multi-modal data, including neuroimaging, clinical, cognitive, neurophysiological test results, and genotype information [8], [40–43]. For instance, Spasov *et al.* [44] employ baseline assessments encompassing MRI, demographic, genetic, and cognitive data. They then integrate feature maps at various levels within a CNN architecture to predict the transition from MCI to AD.

Certain studies use autoencoders for feature representation. Suk and Shen develop a multi-step model including a stacked auto-encoder, using MRI and PET images together with some cognitive test scores. They achieve 75% accuracy in the binary classification of MCI Converter and MCI Non-Converter (to AD) patients [45]. Another study by Venugopalan *et al.* use a stacked denoising auto-encoder and extract features from different modalities, including clinical, genomic, and neuroimaging [8]. The authors use a feature masking technique to address the interpretability problem, whereby one feature is eliminated simultaneously while a decline in prediction accuracy is ob-

served. The research shows that the hippocampus, amygdala brain areas, and Rey Auditory Verbal Learning Test are the most critical characteristics for early AD identification. Venugopalan *et al.* [8] offer insights aligned with the theme of this thesis, exploring the impact of biomarkers on AD progression using autoencoders. However, our proposed framework adopts a broader perspective. It seeks to comprehensively understand the root causes of the transition from MCI to AD by examining the overarching shifts in patient data.

Some studies seek to extract interpretable features by combining deep learning architectures with machine learning models. The approach of Bi *et al.* [46] utilizes unsupervised CNN, PCANet, to achieve feature learning of MRI images. PCANet can learn the filters in CNN through traditional unsupervised machine learning algorithms. Li *et al.* [47] utilize a Long Short-Term Memory (LSTM) model to predict the risk of progression from MCI to AD, using cognitive measures and features extracted from the MRI of the hippocampus. They apply an LSTM on baseline (first visit), sixth-month, and 12th-month visits' cognitive measures, and feed the extracted features with the hippocampus MRI and demographic information into a cox regression.

Wegmayr *et al.* [48] take the analysis of the progression to another level by training a recursive generator architecture to synthesize the following stages of a patient's brain image. They handle MRI data with generative models, develop a recursive generator model for brain image time series, and estimate the future brain image with a Generative Adversarial Network (GAN). This follow-up image is evaluated by a pretrained MCI-AD CNN classifier, and the output is utilized to obtain the AD probability for the follow-up image. It is worth noting that while this study aligns with the theme of estimating disease progression using generated data, in this thesis, the generation process is intentionally controlled and manipulated for a specific purpose.

The studies presented in this section primarily concentrate on training predictive models to classify the conversion from MCI to AD. Although deep learning models have shown impressive performance, uncovering the biomarkers and modalities that

contribute to predictive accuracy remains a challenge. In this thesis, we employ both supervised and unsupervised latent space manipulation techniques to delve into the underlying factors behind the conversion between MCI and AD. In addition to accurate classification, this thesis also places emphasis on comprehending the specific changes that take place during the MCI to AD conversion. This advancement not only unveils the latent space dynamics within longitudinal healthcare data but also carries significant social implications by providing guidance to healthcare professionals regarding disease progression.

2.4. Latent Space Manipulation

Latent space manipulation is a relatively recent idea that is typically used with GANs. The resulting output can be specifically reshaped by managing the latent space. One of the computer vision studies, for instance, uses latent space manipulation to create face images that appear more feminine, older, or covered in makeup [49]. For high-resolution medical image synthesis utilizing GANs, another study in the healthcare field manipulates latent space [50]. As in this study, learning to control the latent space can be done both during and after training (using a pre-trained model).

Both supervised and unsupervised learning frameworks can be used to learn the manipulation of the latent space. An attribute classifier is trained in the supervised techniques to discover relevant directions within the latent space [51]. Data points sampled from a latent space can also be subjected to Principal Component Analysis (PCA) to identify significant directions [52]. On the other hand, DeVries *et al.* suggest applying straightforward alterations to the latent representations discovered by an auto-encoder, such as extrapolation, interpolation, and adding Gaussian noise [53]. In a more recent study, extrapolation and interpolation are used to control the latent space in an automatic, modality-ignorant approach [54]. Studies using straightforward transformations focus on data augmentation instead of generating significantly different samples from the input.

3. PROPOSED METHOD

3.1. Research Question

As can be seen in the former studies, conversion from MCI to AD is mostly handled as a supervised learning problem. A classifier is constructed in different ways to learn to predict the future outcomes of patients. This thesis focuses on the changes during the disease progression, which biomarkers affect the conversion from MCI to AD, and whether it can be confidently claimed that a particular patient's cognitive situation will worsen. The considered biomarkers will be integrated into the patient data set as attributes, and these attributes will be mapped to a multi-dimensional latent space.

Anticipated outcomes include the proximity of latent space representations for patients with similar conditions. Within this framework, we hypothesize that there should exist directions leading from the MCI group toward the AD group in the latent space. Therefore, this problem is posed as an exploration of the meaningful directions in the latent space where the patients are represented. For this purpose, we propose implementing a novel method in this field.

While applying latent space manipulation to tabular data is unconventional, the impact of such manipulation can be evaluated using a classifier, diverging from the visual techniques commonly employed in computer vision. We aim to identify the direction within the latent space corresponding to specific diagnostic shifts in the patient data. We aim to analyze the changes in feature subsets and quantify the extent to which these changes contribute to the observed shifts. The main objective is to predict the likelihood of an individual with MCI being diagnosed with AD in the short and long term by examining patterns within the specific feature subsets. To summarize, the main research questions to be answered throughout this thesis are listed in Figure 3.1.

- How can latent directions help us understand the conversion from MCI to AD?
- Which attributes significantly affect the conversion from MCI to AD?
- How do these significant attributes change during the transition between MCI and AD?
- Can we propose a predictive approach to detect progressive MCI patients?

Figure 3.1. List of the main research questions.

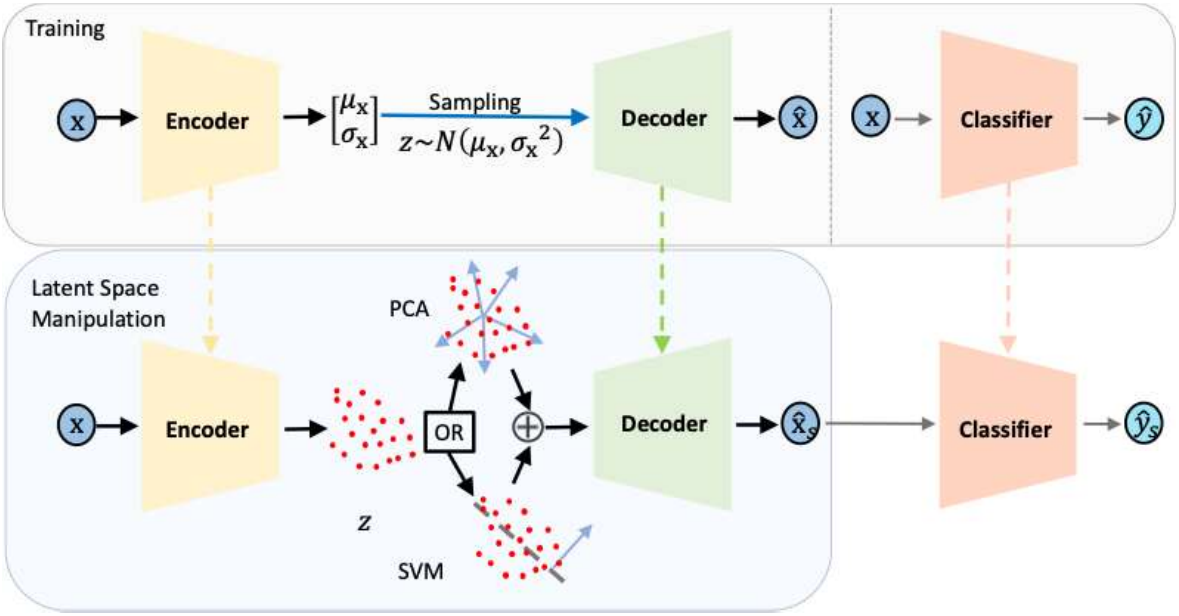


Figure 3.2. Proposed framework. MCI and AD patient data are used to train a VAE and a classifier. Unit direction vectors derived from latent representations are employed for manipulation, leading to the decoding of manipulated representations. Subsequently, the classifier labels the manipulated patient data.

3.2. Model Architecture

This section presents the proposed framework in Figure 3.2. The proposed model consists of a VAE and a binary classifier trained on MCI and AD patient data. The details of the framework will be given in the next subsections.

3.2.1. Binary Classifier

A binary classifier is trained to distinguish patient visits with MCI and AD diagnoses. We do not consider temporal modeling at this stage. The number of time steps in publicly available data sets is often insufficient to capture a temporal pattern. Therefore, an MLP network is trained with the patient visits annotated as MCI and AD diagnoses. The classifier is trained to minimize the binary cross-entropy loss below

$$\mathcal{L}_C = \frac{1}{N}\sum_{i=1}^{N} -\{y_i \log \hat{y}_i + (1-y_i)\log(1-\hat{y}_i)\}, \tag{3.1}$$

where N denotes the number of training samples, y_i is the ground truth label of i^{th} sample, $\hat{y}_i$ is the predicted label of the i^{th} sample.

The binary classifier is used to evaluate the output of the manipulation procedure. Otherwise, any feedback from the binary classifier is not used during the training of the VAE. The probabilistic outcome of the classifier also provides insight into the amount of manipulation towards the direction of AD. For instance, if an MCI patient is classified as AD with high confidence after the manipulation, this means that the latent representation of the MCI patient is moved toward the AD region in the latent space, resulting in crossing the AD side of the decision boundary in the classifier's latent space.

3.2.2. Variational Auto-Encoder

The objective of the VAE is to learn a latent space for the input data in an unsupervised way [55]. In contrast to conventional autoencoders, the encoder outputs a distribution rather than projecting the input to a latent vector of a fixed size. As a result, the VAE prevents the overfitting of particular patients in the training set while capturing patient distributions for MCI and AD. By reconstructing the input from a sample taken from a normal distribution, the decoder acquires a generative quality. Therefore, VAE's latent space is more versatile for latent space manipulation than the traditional autoencoder.

This study uses fully connected encoder and decoder architectures, and the VAE network is trained to minimize the loss function expressed as

$$\mathcal{L}_{\text{VAE}} = \frac{1}{N}\sum_{i=1}^{N} ||\mathbf{x}_i - \hat{\mathbf{x}}_i||_2^2 + \gamma \text{KL}\left[N\left(\boldsymbol{\mu}_x, \boldsymbol{\sigma}_x^2\right), N\left(0,1\right)\right], \tag{3.2}$$

where $\mathbf{x}_i \in \mathbb{R}^d$ is the input, $\hat{\mathbf{x}}_i \in \mathbb{R}^d$ is the reconstructed input, $\boldsymbol{\mu}_x, \boldsymbol{\sigma}_x^2$ denote the mean and variance of the distribution learned by the encoder, $\text{KL}\left[\cdot\right]$ is the Kullback-Leibler divergence, and γ controls the effect of KL divergence that measures the distance between the standard normal and encoder's distribution. The first term of the equation is Mean Squared Error (MSE), which penalizes the reconstruction error. The second term regularizes the learning to have a normally distributed latent space.

3.3. Latent Space Manipulation

Latent space manipulation techniques are commonly used with generative models to solve computer vision problems. A growing number of studies are looking into ways to control a GAN's latent space in order to create new images [49], [51, 52]. Here, we hypothesize that a VAE should capture the MCI and AD distributions, resulting in two groupings in the latent space. Thus, a latent direction toward the AD group should be detected for the MCI patients in the latent space. Therefore, we propose to employ latent space manipulation to examine the factors that lead MCI patients to develop AD. To the best of our knowledge, the proposed study is the first attempt to benefit the directions in the latent space of tabular patient data to answer questions regarding the conversion between MCI and AD. After the manipulation in the VAE's latent space, the trained binary classifier is used to observe the changes in the predictions of the reconstructed manipulated samples.

The primary objective of this thesis is to gain a comprehensive understanding of the underlying factors contributing to conversion from MCI to AD diagnosis. For this purpose, we omit the inclusion of a control group, and we focus on modifying the latent representations of MCI patients. We begin by utilizing the trained VAE to generate reconstructed vectors of MCI patients. Subsequently, for the latent space manipulation

procedure, we specifically select MCI patients whose reconstructed samples have been accurately classified as MCI by the binary classifier. By employing this approach, we ensure the inclusion of correctly reconstructed data points, enabling us to obtain a clearer perspective on the effects of manipulation on label shifts. We follow one unsupervised method and one supervised method for latent space manipulation, which are described in detail in the following sections.

3.3.1. Unsupervised Latent Space Manipulation Method

Inspired by the unsupervised latent space manipulation technique proposed by [52], we investigate the latent direction in the latent space of the VAE by applying PCA [56] to the latent representations of the patient visit data. Thus, the principal components (PC) can capture the directions containing the most variation in the latent space. Since only MCI and AD patients are considered in this study, the first PC is expected to be associated with the direction of the disease progression.

3.3.2. Supervised Latent Space Manipulation Method

We also conducted experiments with a supervised manipulation technique [57] to be able to control the manipulation direction directly. An SVM [58] classifier with a linear kernel is trained on the latent representations of MCI and AD patients. According to [57], this method is predicated on the idea that a hyperplane exists in latent space for every binary characteristic such that all samples from the same side of the hyperplane have the same class label. This assumption is to be evaluated in Section 4.6.

The parameter vector of the linear SVM classifier holds a distinctive quality: it stands perpendicular to the decision boundary [58]. In the context of segregating samples labeled as MCI and AD, the decision boundary aims to create a division, with the parameter vector directed towards the positive class, which, in this case, is MCI. Conversely, the vector directed opposite to the parameter vector leads towards the AD

class within the latent space. Therefore, this vector is normalized and used as a unit direction vector in the manipulation step.

3.3.3. Latent Space Manipulation Algorithm

Since the manipulation is applied in the latent space, latent representations of the patient visit data must be obtained as a first step. The latent representations are obtained with the encoder of the trained VAE, which can be written as

$$\mathbf{z} = \text{encoder}\left(\mathbf{x}\right), \tag{3.3}$$

where $\mathbf{z}$ is the latent representation learned by the VAE, and $\mathbf{x}$ is the patient visit vector comprising numerical measurements obtained from MRI measurements and cognitive tests. The latent representation of a patient is manipulated by adding either a PC vector or the unit direction vector of the SVM classifier, multiplied by a varying coefficient α. This affine transformation can be expressed as

$$\mathbf{z}^* = \mathbf{z} + \alpha\mathbf{u}, \tag{3.4}$$

where $\mathbf{z}^*$ is the manipulated latent representation of the patient, $\mathbf{u}$ is one of the latent directions, and α denotes the coefficient that controls the amount of manipulation. Once the manipulated latent representation of the patient $\mathbf{z}^*$ is obtained, the decoder of the VAE is used to synthesize a new patient from the manipulated latent representation. The decoding procedure can be expressed as

$$\hat{\mathbf{x}}_s = \text{decoder}\left(\mathbf{z}^*\right), \tag{3.5}$$

where $\hat{\mathbf{x}}_s$ denotes the manipulated patient data. Next, the effect of manipulation on the diagnosis of the manipulated patient $\hat{\mathbf{x}}_s$ is evaluated using the binary classifier. Evaluating the new label can be expressed as

$$\hat{y}_s = \text{classifier}\left(\hat{\mathbf{x}}_s\right), \tag{3.6}$$

where $\hat{y}_s$ is the diagnosis of the manipulated patient $\hat{\mathbf{x}}_s$. The algorithm in Figure 3.3 summarizes the latent space manipulation framework presented above.

initialize:

$\mathbf{z}_{\text{selected}} \leftarrow$ an empty array

$i \leftarrow 1$

$\mathbf{z}_i = \text{encoder}\left(\mathbf{x}_i\right),\ i = 1, \cdots, N$

if unsupervised manipulation method **then**

$\mathbf{PC} = \text{PCA}_n\left(\mathbf{z}_1, \cdots, \mathbf{z}_N\right)$

$\mathbf{u} = \mathbf{PC}_j,\ \forall j \leq n$

else

$\mathbf{w} = \text{SVM}\left(\mathbf{z}_1, \cdots, \mathbf{z}_N | \mathbf{y}_1, \cdots, \mathbf{y}_N\right)$

$\mathbf{u} = -1 * \frac{\mathbf{w}}{||w||}$

end if

while $i \leq N$ **do**

$\hat{\mathbf{x}}_s = \text{decoder}\left(\mathbf{z}_i\right)$

$\hat{y}_s = \text{classifier}\left(\hat{\mathbf{x}}_s\right)$

if $\hat{y}_s = \text{MCI} \wedge y_i = \text{MCI}$ **then**

$\mathbf{z}_{\text{selected}} \leftarrow \mathbf{z}_{\text{selected}} + \mathbf{z}_i$

end if

$i \leftarrow i + 1$

end while

$\mathbf{z}^*_{\text{selected}} = \mathbf{z}_{\text{selected}} + \alpha\mathbf{u}$

$\hat{\mathbf{x}}_s = \text{decoder}\left(\mathbf{z}^*_{\text{selected}}\right)$

$\hat{\mathbf{y}}_s = \text{classifier}\left(\hat{\mathbf{x}}_s\right)$

Figure 3.3. Latent space manipulation algorithm.

3.4. Predictive Power of the Method

The predictive power of the manipulation method is evaluated by building an analogous relationship to the results of a classification model. Prediction tasks are divided based on the patient data point's visit order and the forecast horizon's length.

initialize:

TP $\leftarrow 0$

TN $\leftarrow 0$

FP $\leftarrow 0$

FN $\leftarrow 0$

confusion matrix $\leftarrow$ an empty array

for $visit = 1$, $visit{+}{+}$, $visit < 4$ **do**

  for $horizon = 1$, $horizon{+}{+}$, $horizon < 4$ **do**

    for $patient = 1$, $patient{+}{+}$, $patient < N$ **do**

      if $\hat{\mathbf{y}}_{patient_{visit}} =$ AD **then**

        if $\hat{\mathbf{y}}_{patient_{visit}} = \mathbf{y}_{patient_{visit+horizon}}$ **then**

          TP $\leftarrow$ TP $+ 1$

        else

          FP $\leftarrow$ FP $+ 1$

        end if

      else

        if $\hat{\mathbf{y}}_{patient_{visit}} = \mathbf{y}_{patient_{visit+horizon}}$ **then**

          TN $\leftarrow$ TN $+ 1$

        else

          FN $\leftarrow$ FN $+ 1$

        end if

      end if

    end for

    confusion matrix$_{visit_{horizon}} \leftarrow$ TP, TN, FP, FN

  end for

end for

Figure 3.4. Prediction performance measurement of the method.

A confusion matrix is created for each visit order and the forecast horizon. When the visit order is selected as one, and the forecast horizon is selected as one, the first visit data for each MCI patient in the test data set is manipulated with different α values. Then, their new label assigned by the pre-trained binary classifier is compared with the actual label belonging to their consecutive visit data in the test set. When the visit order is selected as one, and the forecast horizon is selected as two, the labels of the manipulated versions of the first visit data points for each MCI patient are compared with the actual labels of the third visit data point. When the visit order is selected as two, and the forecast horizon is selected as one, the labels of the manipulated versions of the second visit data points for each MCI patient are compared with the actual labels of the third visit data point. The prediction performance based on each visit order - forecast horizon pair indicates the method's early prediction ability of future short and long-term AD diagnoses correctly.

For this purpose, the first, second, and third visit data points are considered. Forecast horizon values are selected as one, two, and three visits ahead. We evaluate prediction performance using sensitivity, specificity, and ROC-AUC values. For each visit order and forecast horizon pair, various α values yield different prediction results. We select the result with the smallest discrepancy between sensitivity and specificity values. This approach ensures a balanced assessment of the method's capability to identify both prospective AD patients and non-patients. The algorithm in Figure 3.4 summarizes the proposed prediction framework.

In the framework, true positive refers to the number of MCI patients whose manipulated visit data and their visit data after the specified forecast horizon are classified as AD. False negative refers to the number of MCI patients whose manipulated visit data is classified as MCI, but their visit data after the specified forecast horizon is labeled AD. False positive refers to the number of MCI patients whose manipulated visit data is classified as AD, but their visit data after the specified forecast horizon is labeled as MCI. True negative refers to the number of MCI patients whose manipulated visit data and their visit data after the specified forecast horizon are classified as MCI.

4. EXPERIMENTS AND RESULTS

4.1. Data Sets

Experiments are conducted with two different data sets to assess the predictive performance objectively. The first data set is the TADPOLE data set [24]. The second data set is the National Alzheimer's Coordinating Center (NACC) data set [59].

4.1.1. TADPOLE Data Set

The TADPOLE data used in the preparation of this study were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database [60]. The ADNI was launched in 2003 as a public-private partnership led by Principal Investigator Michael W. Weiner, MD. With the help of a Grand Opportunities grant (ADNI-GO) in 2009, ADNI was extended by two more years. Additional competitive ADNI-1 grant extensions arrived in 2011, and 2016 with ADNI-2 [61] and ADNI-3 [62], respectively. For up-to-date information, see [60].

The primary goal of ADNI has been to test whether serial MRI, PET, other biological markers, and clinical and neuropsychological assessment can be combined to measure the progression of MCI and early AD. The other objectives are examining the rate of progression for both MCI and AD, as well as creating a substantial database of clinical and imaging data [63].

TADPOLE is a challenge designed to predict which people with a risk of having AD will start to experience symptoms in one to five years. It focuses on rollover individuals in the ADNI study. The challenge is using historical patient measurements to predict disease progression [24]. The challenge involves three kinds of data sets: training data set, prediction data set, and test data set. The training data set (D1) includes multi-modal measurements with the target information corresponding to the

diagnosis. ADNI's historical longitudinal data is used to create the TADPOLE D1 training data set. Every individual who has contributed data to ADNI on at least two separate visits (different dates) for ADNI 1, ADNI GO, and ADNI 2 has a set of measurements in this data set. The TADPOLE standard prediction set (D2) contains available data from the ADNI rollover individuals with available time points. A single (most recent) time point and a constrained number of variables from each rollover individual are included in the TADPOLE cross-sectional prediction set D3 (same set of individuals as D2) [24].

Evaluating quantitative biomarkers and cognitive tests is extremely valuable for diagnosing AD [24]. The updated diagnostic criteria for AD and MCI related to AD include the five most frequently studied biomarkers [64, 65]. These five indicators can be separated into two groups: those that analyze amyloid beta protein levels and those that analyze nerve cell damage. For the first group, cerebrospinal fluid (CSF) puncture or amyloid positron emission tomography (PET) can be used to assess amyloid beta [66]. For the second group, damage to nerve cells can be indirectly assessed by measuring the amount of tau protein in the CSF or utilizing tau-PET. It can also be assessed directly by measuring brain metabolism with fluoro-deoxyglucose (FDG) PET or atrophy with magnetic resonance imaging (MRI) [66].

TADPOLE data set comprises multi-modal data from seven different biomarker types such as MRI, PET, CSF, diffusion tensor imaging (DTI), cognitive tests, genetic and demographic information from the ADNI database, as shown in Table 4.1. D1-D2 merged data consist of visit information of 1737 subjects, of which 780 are female, and 957 are male. The measurements and test results are displayed as features, with each row representing a patient visit. Each participant has an average of 6.68 separate visits over six years following their first visit. Between the first and last time intervals, there are, on average, 3.5 ± 2.5 years. The patient age distribution at their baseline visit is given in Figure 4.1.

Table 4.1. Set of different biomarker types with their definitions in TADPOLE data set.

Biomarker Type	**Definition**
Main cognitive tests (excluding subtypes)	Neuropsychological tests administered by a clinical expert, such as Clinical Dementia Rating Scale Sum of Boxes (CDR-SB), Alzheimer's Disease Assessment Scale (ADAS) 11, ADAS-13, Mini-Mental State Examination (MMSE), Rey Auditory Verbal Learning Test (RAVLT), Montreal Cognitive Assessment (MoCA), Functional Activities Questionnaire (FAQ), and ElectroCochleoGraphy (ECoG)
MRI ROIs, obtained via Freesurfer [67]	Measures of brain structural integrity, such as volumes, cortical thicknesses, and surface areas
FDG PET ROI averages	Measures for cell metabolism, where cells affected by AD show reduced metabolism
AV45 PET ROI averages	Measures for amyloid-beta load in the brain, where amyloid-beta is a protein that misfolds (its 3D structure is not properly constructed), which results in AD
AV1451 PET ROI averages	Measures for tau load in the brain, where tau is another protein which, when abnormal, damages neurons and leads to AD
DTI ROI measures	Measures for microstructural parameters related to cells and axons, such as: mean diffusivity, axial diffusivity, radial diffusivity
CSF biomarkers	Amyloid and tau levels in the CSF
APOE status	Information regarding a gene that is a risk factor for AD
Demographics	Age, gender, education
Diagnosis	Cognitively normal (CN), mild cognitive impairment (MCI), or Alzheimer's disease (AD)

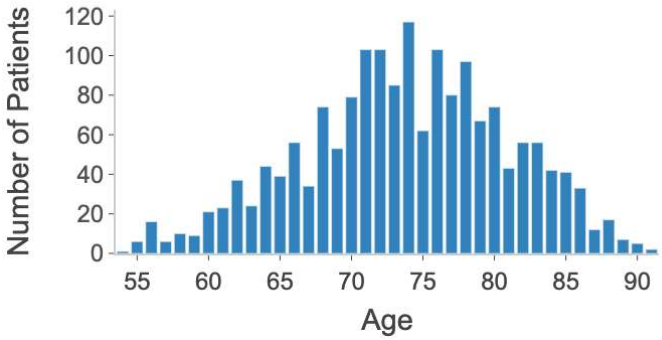


Figure 4.1. Distribution of patient age at baseline visit on TADPOLE data.

4.1.2. NACC Data Set

The National Institute on Aging established the National Alzheimer's Coordinating Center (NACC) in 1999 to support collaborative research. A sizeable relational database of standardized clinical and neuropathological research data has been created and maintained by NACC using information gathered from Alzheimer's Disease Research Centers (ADRCs) across the United States [59]. NACC data include a Uniform Data Set (UDS) with longitudinal clinical evaluation data, MRI and PET data, CSF data, autopsy data, and genomic data. To assemble a data set akin to the information extracted from the TADPOLE challenge data for our experiments, we secured access to NACC MRI and PET data, as well as UDS records.

UDS version 3.0, released in March 2015, is used in our analysis. In UDS v3, the National Institute on Aging and the Alzheimer's Association (NIA-AA) criterion is used for AD diagnosis [68]. Each visit is annotated as MCI and AD based on their NACCUDSD value, a NACC-derived variable for cognitive status at a UDS visit. NACCUDSD equal to three and four denote MCI and AD, respectively.

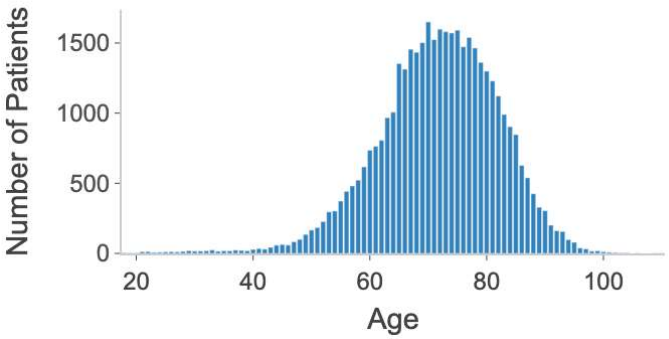


Figure 4.2. Distribution of patient age at baseline visit on NACC data.

Since we specifically focus on the conversion from MCI to AD, the NACCALZP variable of the dementia group is also investigated. NACCALZP is also a NACC-derived variable implying that AD is the primary, contributing, or non-contributing cause of cognitive impairment. Thus, the AD group is sampled from patients whose NACCUDSD equals to four, and NACCALZP equals to one, corresponding to primary. For more information about the variables, please review the UDS Researchers Data Dictionary [69].

The NACC data set features an annual visit interval for each subject, different from the visit schedule found in the TADPOLE data set [59]. Another notable distinction between the two data sets is the use of the Dementia label, which, in the case of NACC, does not necessarily equate to AD. Conversely, in the TADPOLE data set, the terms Dementia and AD are employed interchangeably in data labels.

The average sequence length per patient is 3.26 for UDS data, while it is 1.83 for MRI data. How this difference is handled is explained in more detail in Section 4.2.2. The distribution of patient age at their baseline visit on the NACC data set is shown in Figure 4.2.

4.2. Feature Selection

4.2.1. TADPOLE Data Set

The TADPOLE challenge provides essential guidance for predicting AD progression. It recommends key variables, including diagnosis, neuropsychological test scores, and anatomical features derived from T1 MRI, PET, and CSF markers. The primary target for prediction is the diagnosis variable [24].

Cognitive assessments are a significant component, with recommended tests like CDR-SB, ADAS-11, ADAS-13, MMSE, RAVLT, and FAQ. The recommended MRI measures are all measured in cmm and represent the numerical value of the volume belonging to a particular area in the brain. Hippocampi volume is represented under the variable name Hippocampus, whole brain volume as WholeBrain, entorhinal cortical volume as Entorhinal, middle temporal cortical volume as MidTemp, intracranial volume as ICV, and ventricles volume as Ventricles in the TADPOLE data set. The recommended PET measures are FDG and AV45, while the recommended CSF measures are amyloid-beta level in CSF, tau level, and phosphorylated tau level. To enhance predictive models, the challenge suggests considering risk factors like APOE4 status and age [24].

The selected subset of variables is given in Table 4.2 with their original names in the data set. We did not use age and genetic information in our data set, due to their inability of distinguishing AD patients from MCI patients effectively. We initiated our feature selection by excluding variables with over 50% missing values. Additionally, as part of our data imputation strategy, we assessed whether a crucial measurement contained non-null values for the majority of patients during their first visit. If this criterion was not met, we omitted the measurement from the feature subset.

The RAVLT is a cognitive test which gauges a patient's ability to recall words from a word sequence. This test comprises five consecutive trials, with a free-recall

test following each one. The RAVLT immediate score represents the sum of correctly recalled words across all five trials. In contrast, the RAVLT learning score quantifies the difference between the number of words correctly recalled after the first and final trials. Lastly, the RAVLT forgetting score captures the disparity between the words correctly recalled after the final trial and those recalled during the delayed recall test [70]. Since each RAVLT score assesses distinct cognitive abilities, they are included as separate attributes within our variable subset.

Table 4.2. Set of variables with their statistics and percentage of missing values in TADPOLE data set.

Variable	**AD** **mean ± std**	**MCI** **mean ± std**	**Missing Value** **Percentage**
CDR-SB	$5.99 \pm 3.07 \times 10^{0}$	$1.62 \pm 1.11 \times 10^{0}$	%29.65
RAVLT immediate	$2.03 \pm 0.81 \times 10^{1}$	$3.41 \pm 1.12 \times 10^{1}$	%30.69
RAVLT forgetting	$4.23 \pm 1.93 \times 10^{0}$	$4.66 \pm 2.49 \times 10^{0}$	%30.89
RAVLT learning	$1.66 \pm 1.72 \times 10^{0}$	$3.99 \pm 2.63 \times 10^{0}$	%30.69
ADAS-11	$2.29 \pm 1.01 \times 10^{1}$	$0.99 \pm 0.51 \times 10^{1}$	%30.06
ADAS-13	$3.35 \pm 1.13 \times 10^{1}$	$1.61 \pm 0.71 \times 10^{1}$	%30.74
FAQ	$1.69 \pm 0.75 \times 10^{1}$	$0.34 \pm 0.43 \times 10^{1}$	%29.41
MMSE	$2.12 \pm 0.46 \times 10^{1}$	$2.74 \pm 0.22 \times 10^{1}$	%29.89
Ventricles	$5.32 \pm 2.54 \times 10^{4}$	$4.19 \pm 2.32 \times 10^{4}$	%41.57
Hippocampus	$5.51 \pm 1.11 \times 10^{3}$	$6.71 \pm 1.12 \times 10^{3}$	%46.61
WholeBrain	$0.95 \pm 0.11 \times 10^{6}$	$1.02 \pm 0.11 \times 10^{6}$	%39.65
Entorhinal	$2.69 \pm 0.72 \times 10^{3}$	$3.51 \pm 0.77 \times 10^{3}$	%49.23
Fusiform	$1.49 \pm 0.26 \times 10^{4}$	$1.74 \pm 0.28 \times 10^{4}$	%49.23
MidTemp	$1.65 \pm 0.32 \times 10^{4}$	$1.95 \pm 0.29 \times 10^{4}$	%49.23
ICV	$1.54 \pm 0.18 \times 10^{6}$	$1.54 \pm 0.16 \times 10^{6}$	%37.58

4.2.2. NACC Data Set

We have identified analogous cognitive test scores in the UDS data set, aiming for a resemblance to those in the TADPOLE data set. Nevertheless, it is worth noting that certain cognitive test scores, such as ADAS-11 and ADAS-13, are not present in the NACC data set. Table 4.3 provides a comprehensive list of the variables selected from the UDS data set, guided by their availability and minimal ratio of missing values. Notably, due to the high incidence of missing values, MRI and PET measurements, with an average sequence length of 1.83, have been excluded from the feature subset.

Table 4.3. Set of variables with their statistics and percentage of missing values in NACC data set.

Variable	**AD** **mean ± std**	**MCI** **mean ± std**	**Missing Value** **Percentage**
CDR-SB	$8.34 \pm 4.99 \times 10^{0}$	$1.35 \pm 1.13 \times 10^{0}$	%0.00
FAQ	$1.87 \pm 0.92 \times 10^{1}$	$0.31 \pm 0.45 \times 10^{1}$	%0.00
MMSE	$1.88 \pm 0.68 \times 10^{1}$	$2.71 \pm 0.25 \times 10^{1}$	%42.59

4.3. Data Preprocessing

4.3.1. TADPOLE Data Set

D1 and D2 merged data sets from the TADPOLE challenge are used for the experiments. To handle missing data, we employ a forward filling approach, which replaces missing values with the most recent measurement recorded for the patient before the gap. In cases where forward filling isn't applicable, we resort to utilizing an average value derived from visits sharing the same label.

We applied standard scaling to the numerical data, wherein the scaler was fitted to the training data. Subsequently, both the training and test data were transformed using this trained scaler. Additionally, measurements obtained from PET and MRI scans, encompassing areas like the entorhinal, fusiform, middle temporal cortical, and whole brain regions, were adjusted for the patient's intracranial volume. Since these measures are inherently dependent on an individual's head size and lack established standards, standardization was performed as described in previous works [71, 72].

During our experiments, it became evident that certain measurements displayed anomalies, rendering their values clinically implausible. To address this, we applied a filtering process that identified entries deviating by more than three standard deviations from the average value for each measurement, and these outliers were subsequently removed from the data set.

Given the specific focus of this thesis on analyzing the transition from MCI to AD, we restrict our experiments to patients with either MCI or AD labels. Consequently, we have a data set comprising 1343 data points with AD labels and 4021 data points with MCI labels. This results in a ratio of 33% for AD data points to MCI data points. For our evaluation, we concentrate exclusively on correctly classified MCI data points in the test set, which includes 177 patients. Of these, 28 patients (16%) convert to AD in the future. We employ a 0.2 train-test ratio, ensuring that visit data from the same individual do not appear in both the training and test data sets.

4.3.2. NACC Data Set

Missing data imputation, anomaly detection, normalization, and the process of splitting the data into training and test sets are managed using the same methods as previously outlined for the TADPOLE data set. For this data set, there are 26751 data points with AD labels and 23941 data points with MCI labels, resulting in a ratio of 112% for AD data points to MCI data points. During the evaluation phase, our attention is directed solely towards accurately classified test MCI data points, which

encompasses 1869 patients. Among these, 548 patients (29%) undergo the conversion to AD in the future.

4.4. Training Scheme

Google Colab [73] and Tensorflow [74] are used to implement the proposed approach. Both the encoder and decoder are designed as a three-layer fully connected network. ReLU activation functions [75] are used after each encoder and decoder layer. Hyperparameter optimization is employed with five-fold cross-validation to train the VAE model.

The hyperparameter space for the hidden layer dimension is $[256, 128, 64, 32, 16, 8]$ for the first dense layer. The second and third layers have the dimension number divided by two after each layer. After the cross-validated evaluation, the dimensions are $[256, 128, 64]$ for the first, second, and third layers. After conducting hyperparameter optimization separately, the hidden layer dimension of the first dense layer is decided as 256 for both TADPOLE and NACC data sets.

Another hyperparameter to be tuned is γ in Equation (3.2). Lower values lead to more robust reconstruction with less structured latent space, while higher values result in more structured latent space with weaker reconstruction [76]. The hyperparameter space of the parameter γ is set to $[1, 0.75, 0.5, 0.25, 0.1]$. After conducting hyperparameter optimization separately, γ is set to 0.1 for both TADPOLE and NACC data sets.

The binary classifier is designed as a two-layer, fully connected network. The same hyperparameter space for hidden layer dimension $[256, 128, 64, 32, 16, 8]$ is used for the classification model's hyperparameter optimization with five-fold cross-validation method. The second layer has the hidden dimension number divided by two after the first layer. Batch normalization [77] and dropout with a probability of 0.2 are used for regularization. After conducting hyperparameter optimization separately, hidden

dimensionality is determined as 128 and 64 for the two layers with the TADPOLE data set and 64 and 32 for the two layers with the NACC data set.

The binary classifier and VAE models are trained with a batch size of 128 and validation set split ratio of 0.1. Maximum number of epochs is set to 100 and the training is terminated early if the validation performance does not progress during five consecutive epochs. Adam optimizer [78] is used with a learning rate scheduler where the decay rate is 0.1, and the initial learning rate is set to 0.001. During training and data preprocessing, the random seed is fixed and set to zero.

SVM classifier is trained with a linear kernel. Hyperparameter space for C parameter is decided as $[1, 0.1, 0.01, 0.001, 0.0001]$. After five-fold cross validation method is followed for hyperparameter optimization, C parameter is determined as one.

4.5. Model Performances

4.5.1. VAE Performance

In assessing VAE performance, we employ a method from [79] that involves two classifiers. Initially, both Classifier Number One and the VAE are trained using the training data. Then, artificial data is generated from the original training data set using the VAE. This artificial data is used as input for Classifier Number One, which assigns predictions to create an artificial training data set. Classifier Number Two, identical to Classifier Number One, is trained using this artificial training data set.

To evaluate the generative model's performance in an unbiased manner, we assess the performance of Classifier Number Two using the test data, considering both Classifier Number One's performance on the generated training data and Classifier Number Two's performance on the test data [79]. The performance metrics are demonstrated in Table 4.4.

Table 4.4. Performance metrics of classifiers demonstrating VAE performance for TADPOLE and NACC data sets.

Model	**Data Set**	**Data**	**ROC-AUC**	**F1**
Classifier 1	TADPOLE	Generated Train Data	0.90	0.90
Classifier 2	TADPOLE	Test Data	0.85	0.87
Classifier 1	NACC	Generated Train Data	0.92	0.92
Classifier 2	NACC	Test Data	0.91	0.91

These metrics demonstrate that VAE can generate realistic patient data with our data sets, with small reconstruction error. Moreover, the mean and standard deviation of the VAE latent space of the TADPOLE test data are -0.03 and 0.96, and the NACC test data are -0.002 and 0.99. This indicates that the distribution of the latent space is close to standard normal distribution for both data sets, as intended [55].

Besides quantitative performance measurement, we also visualized the latent space of the VAE colored by data labels for the TADPOLE training data set in Figure 4.3 and NACC training data set in Figure 4.4. It is shown that the VAE can separate patient data in an unsupervised way for both data sets.

4.5.2. Classifier Model Performances

In this subsection, we demonstrate the performance metrics for the binary classifier and SVM classifier. To correctly evaluate the latent space manipulation effects, the binary classifier should not perform poorly. In addition, to make sure the parameter vector of the SVM classifier is directed towards the positive class, the SVM classifier's ability to divide the latent space accurately needs to be investigated. The performance metrics for the binary classifier and SVM classifier are demonstrated in Table 4.5.

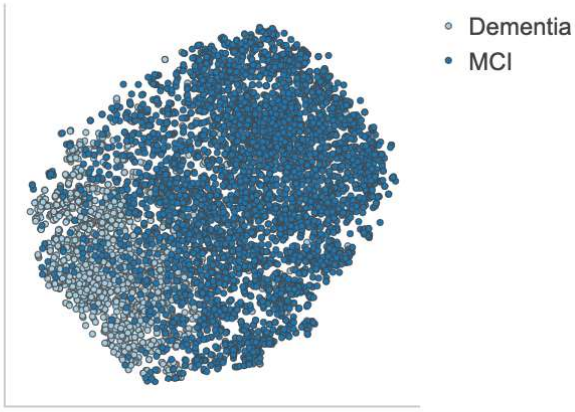


Figure 4.3. t-SNE visualization of VAE latent space representations of the TADPOLE training data set.

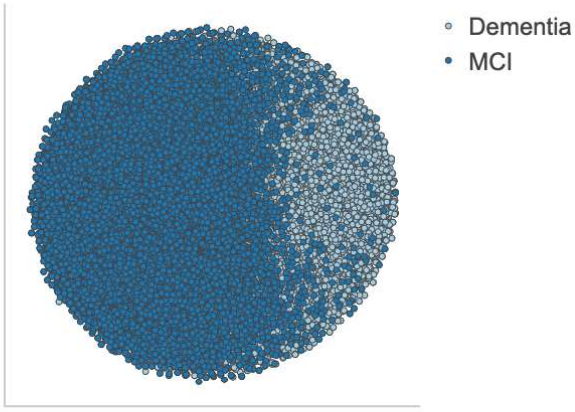


Figure 4.4. t-SNE visualization of VAE latent space representations of the NACC training data set.

Table 4.5. Performance metrics of classifiers for TADPOLE and NACC data sets.

Model	**Data Set**	**Data**	**ROC-AUC**	**F1**
Binary Classifier	TADPOLE	Test Data	0.85	0.90
SVM Classifier	TADPOLE	VAE Latent Space Representations of Test Data	0.84	0.86
Binary Classifier	NACC	Test Data	0.91	0.91
SVM Classifier	NACC	VAE Latent Space Representations of Test Data	0.91	0.91

4.6. Supervised and Unsupervised Latent Space Manipulation

In our workflow, both the VAE and the binary classifier are trained using the training data set. We then apply PCA and SVM to the latent space representations of the training data set, extracting unit direction vectors for manipulation. Subsequently, we conduct manipulation experiments on the latent space representations of data points labeled as MCI in the test data set, that are classified as MCI by the binary classifier.

Unit direction vectors, as detailed in Figure 3.3, are obtained through the described methods. For our experiments, we choose to utilize the first five PC vectors. Notably, the cosine similarity value between the first PC vector and the unit direction vector from the supervised manipulation method (SUDV) is 0.64 for TADPOLE, and 0.96 for NACC data sets. These values are the highest among other cosine similarity values with the remaining PC vectors. This observation suggests significant similarity in direction between the first PC vector, inherently a unit direction vector, and the SUDV. We anticipate that this similarity will be reflected in the outcomes of our manipulation experiments.

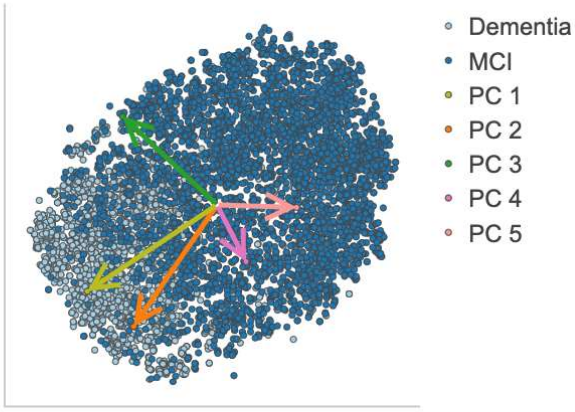


Figure 4.5. t-SNE visualization of PC vectors on VAE latent space representations of TADPOLE training data.

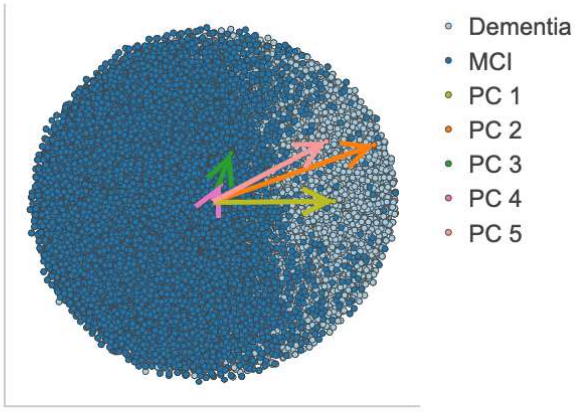


Figure 4.6. t-SNE visualization of PC vectors on VAE latent space representations of NACC training data.

The t-SNE visualizations of manipulation vectors are demonstrated in Figure 4.5 and Figure 4.6 for unsupervised manipulation, and in Figure 4.7 and Figure 4.8 for supervised manipulation. The endpoints of vector arrows represent the PC vectors and supervised unit direction vector (SUDV).

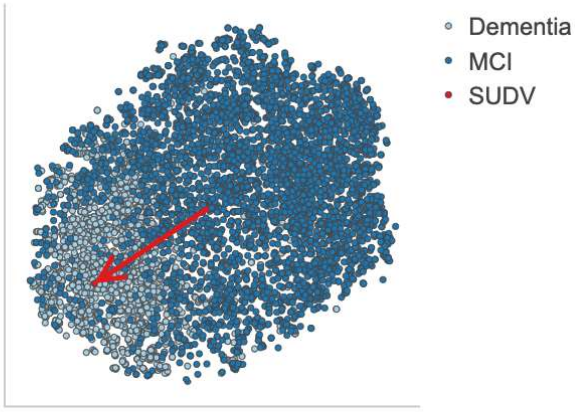


Figure 4.7. t-SNE visualization of unit direction vector obtained via SVM classifier's parameter vector on VAE latent space representations of TADPOLE training data.

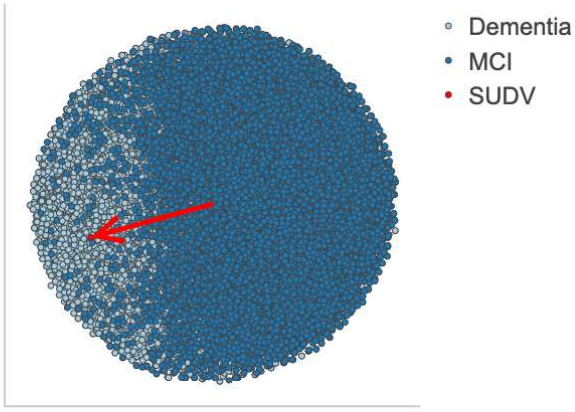


Figure 4.8. t-SNE visualization of unit direction vector obtained via SVM classifier's parameter vector on VAE latent space representations of NACC training data.

t-SNE model is fit for patient data vectors concatenated with PC vectors and SUDV. Then, an arrow is drawn from the origin point $(0, 0)$ towards each unit direction vector representative point for visualization. As can be seen, both the SUDV and the first PC vector are directed toward AD, and they are approximately perpendicular

to the linear boundary between MCI and AD data points for both data sets. This indicates that manipulation steps applied with these unit direction vectors push the visit data points labeled as MCI toward AD.

4.7. Choice of Principal Components

We extract the top five PC vectors from the patient visit latent data representations using PCA. Each PC is added to the latent representations of the MCI patients as in Equation (3.4). The manipulation effect of each PC is evaluated by the average softmax probabilities, indicative of the likelihood of being AD, for the binary classifier outputs of the manipulated data points. In Figure 4.9 and Figure 4.10, for a fixed value of $\alpha = 5$, we plot the average probability of the classifier for AD computed using manipulated MCI patients against the five different principal directions, for TADPOLE and NACC data sets, respectively [4]. This empirical evidence underscores the pivotal role of the first PC in capturing the most informative direction, leading to the most significant transformation in shifting the latent representation of an MCI patient towards AD. Subsequently, the influence of manipulation decreases considerably when considering the other PC vectors, a pattern observed in both data sets. To visualize this trend, please refer to Figure 4.5 and Figure 4.6. Hence, based on these findings, we designate the first PC as the unit direction vector for all subsequent experiments involving the unsupervised latent space manipulation method.

4.8. Investigating the Effect of α

The parameter α in Equation (3.4) allows us to control the manipulation magnitude. During the manipulation experiments, we explore a range of α values, specifically within the interval $[0.025, 10]$, with intervals of 0.025. Notably, we focus on positive values for α as our objective is to steer the latent representation towards the direction of the unit direction vector. To assess the impact of varying α values, we examine the average softmax probabilities, indicative of the likelihood of being AD, for the binary classifier outputs of the manipulated data points with varying α values.

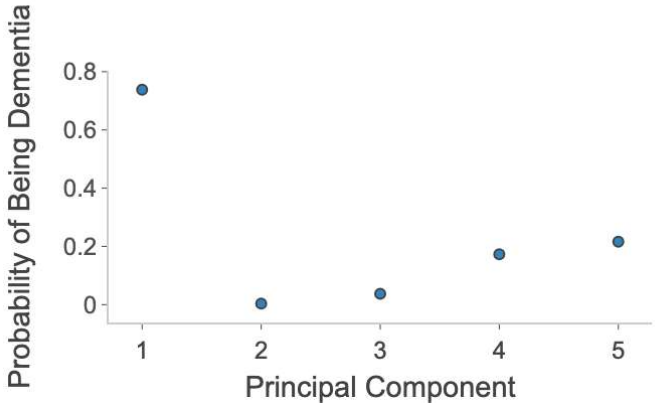


Figure 4.9. Manipulation effects of PC vectors for a fixed $\alpha = 5$ on TADPOLE data set.

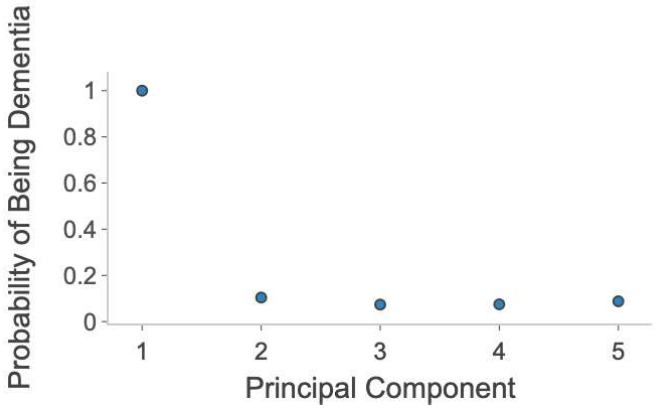


Figure 4.10. Manipulation effects of PC vectors for a fixed $\alpha = 5$ on NACC data set.

4.8.1. TADPOLE Data Set

Figure 4.11 illustrates the graph for the first PC, while Figure 4.12 represents the supervised unit direction vector for the TADPOLE test data [4]. Both graphs exhibit a non-linear behavior when it comes to converting MCI patients to AD using different α values. Upon examining the first derivative of this almost S-shaped graph, we discern that the manipulation effect peaks around α value equal to four, before gradually declining with further increases or decreases in the α value. In Figure 4.12, we reach to the same conclusion with SUDV.

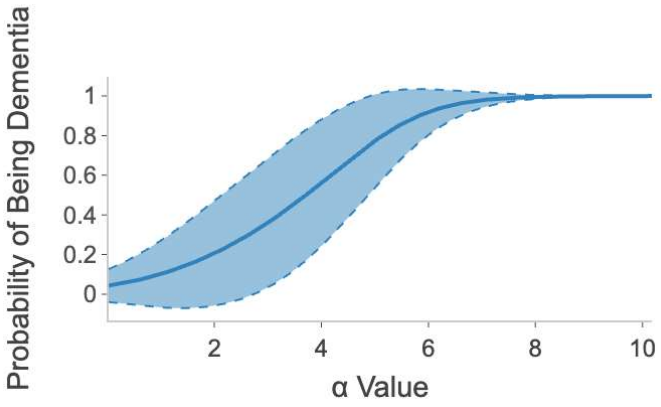


Figure 4.11. Manipulation effect of the first PC vector with different α values for TADPOLE data set. The horizontal axis represents α values, the vertical axis represents the average probability of AD of the data points after manipulation, and the shaded region denotes the standard deviation.

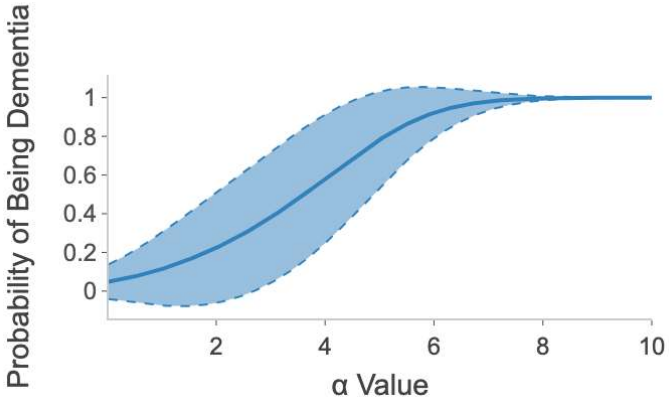


Figure 4.12. Manipulation effect of SUDV with different α values for TADPOLE data. Horizontal axis represents α values, vertical axis represents the average probability of AD of the data points after manipulation, and shaded region denotes the standard deviation.

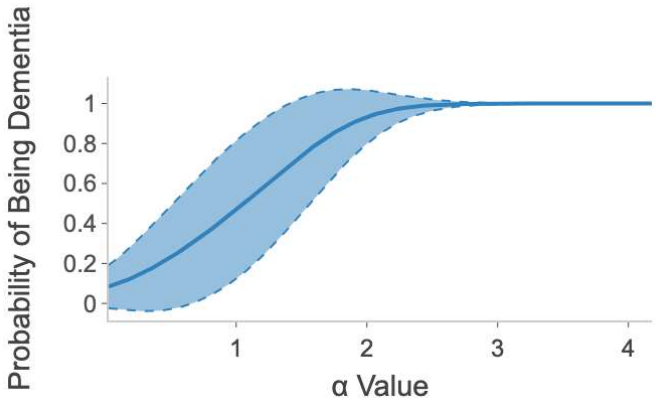


Figure 4.13. Manipulation effect of the first PC vector with different α values on NACC data set. The horizontal axis represents α values, the vertical axis represents the average probability of AD of the data points after manipulation, and the shaded region denotes the standard deviation.

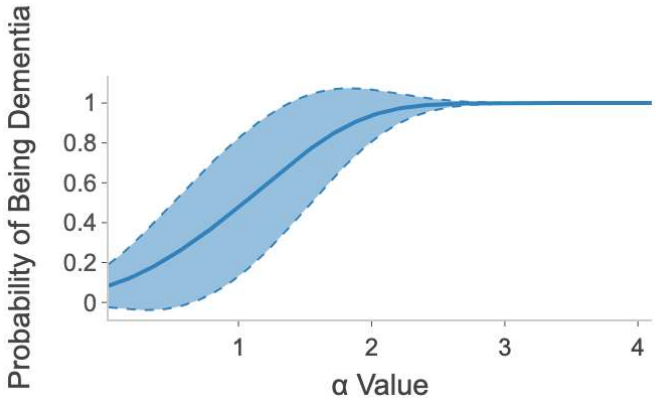


Figure 4.14. Manipulation effect of SUDV with different α values on NACC data. Horizontal axis represents α values, vertical axis represents the average probability of AD of the data points after manipulation, and shaded region denotes the standard deviation.

4.8.2. NACC Data Set

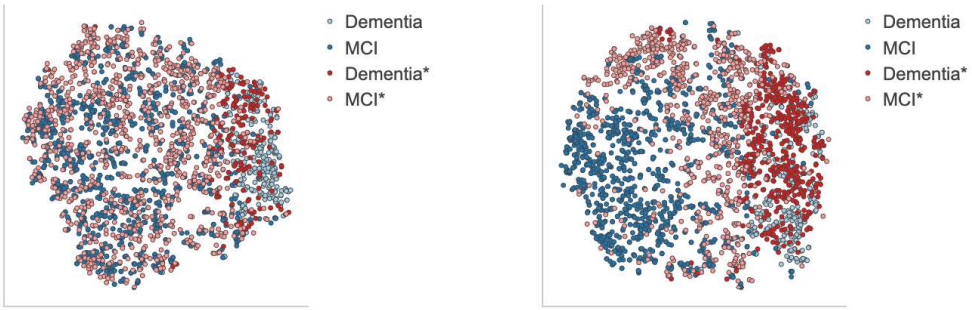
Figure 4.13 demonstrates the graph for the first PC and Figure 4.14 for the supervised unit direction vector for the NACC test data set. These figures demonstrate that similar non-linear effects of manipulation are observed on the NACC test data set with both the unsupervised and supervised manipulation methods. Considering the first derivative of this nearly S-shaped graph, the incremental manipulation effect climaxes near α equal to 1.5 and diminishes as the α value increases or decreases. Other observations for NACC data set were consistent with TADPOLE data set.

The saturation points for α values are less than three in both Figure 4.13 and Figure 4.14. In comparison to Figure 4.11 and Figure 4.12, the saturation points are notably lower in the case of the NACC data set. This behavior can be elucidated by examining Figure 4.5 and Figure 4.6. It is evident that the data points in Figure 4.6 are denser and less dispersed from the center. This particular arrangement allows the manipulation method to shift a greater number of data points toward the AD region using smaller α values.

The figures for both data sets reveal that unsupervised and supervised manipulation methods exert similar effects on patient data. The remaining experiments and analyses are carried out using both methods. Interestingly, the results show no significant disparities. As a result, for the sake of simplicity, we will present the results obtained through the unsupervised manipulation method.

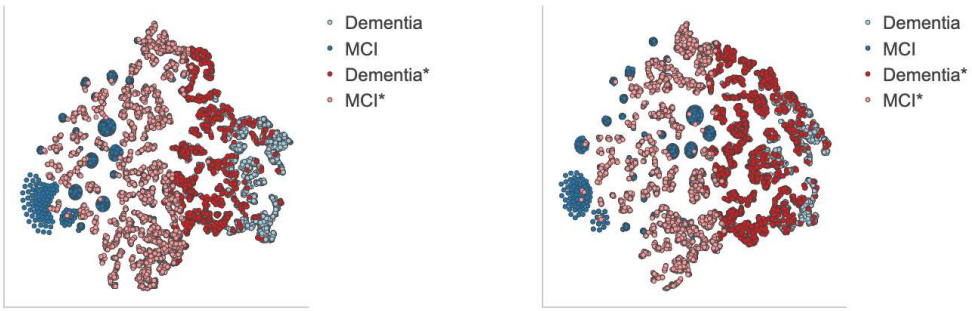
4.9. Visualizing the Classifier's Latent Space

Figure 4.15 and Figure 4.16 provide comparisons between the manipulated data points and their original counterparts in the binary classifier's latent space for the TADPOLE and NACC data sets, respectively. Here, Dementia* and MCI* refer to the manipulated versions of MCI data points, as manipulation is exclusively applied to MCI-labeled data points.



(a) $\alpha = 2$ (b) $\alpha = 4$

Figure 4.15. t-SNE of the manipulated TADPOLE test data points with their original versions in the classifier's latent space, with Figure 4.17 (a) $\alpha = 2$ and Figure 4.17 (b) $\alpha = 4$. Light blue and dark blue represent AD and MCI data points before manipulation and red and pink colors represent AD and MCI data points after manipulation.



(a) $\alpha = 0.5$ (b) $\alpha = 1$

Figure 4.16. t-SNE of the manipulated NACC test data points with their original versions in the classifier's latent space, with Figure 4.18 (a) $\alpha = 0.5$ and Figure 4.18 (b) $\alpha = 1$. Light blue and dark blue represent AD and MCI data points before manipulation and red and pink colors represent AD and MCI data points after manipulation.

The visualizations vividly demonstrate the diversity among the generated data points. Interestingly, some of the generated AD patients are positioned near the boundary separating the MCI and AD regions. Even with higher α values, these points do not become tightly clustered in the middle, which is a desirable outcome [4]. This experiment provides compelling evidence that the suggested method is well-suited for generating realistic synthetic AD patient data, particularly when dealing with limited data resources. Further investigation into the data generation performance of this proposed method can be found in Section 4.12.

4.10. Clinical Interpretation

Our analysis begins by examining the associations between changes in variables as the manipulation effect varies. We focus specifically on the visit data of MCI patients who later develop AD. To do this, we explore two key metrics: the proportion of test patients who convert to AD and the proportion of patient visit data points that change their diagnosis to AD during the manipulation process with varying α values. This analysis aims to establish a connection between these two outcomes, shedding light on the predictive potential of the method.

Additionally, we delve into the individual changes in variables following manipulation to assess their alignment with reality. To achieve this, we manipulate MCI patients using the first PC with a fixed α value. The choice of α value is based on the linkage analysis previously mentioned, considering the ratio of detected future AD patients and the ratio of manipulated patients. This analysis is carried out separately for each data set.

After applying affine transformation with a specific α value, we examine the data points that have converted to AD. These data points represent specific visit information for a patient. We extract the patient's visit information after the manipulated data point's visit time. Subsequently, we compare the average changes between the manipulated and original time steps after the visit time of the manipulated data points.

4.10.1. TADPOLE Data Set

Figure 4.17 presents a heatmap illustrating the correlation between the average changes in variables following manipulation with the first PC for increasing α values on the TADPOLE data set. In the heatmap, dark cells represent a positive correlation, while light cells signify a negative correlation. It is important to note that Figure 4.17 doesn't explicitly indicate whether variables increase or decrease after manipulation, this information can be found in Table 4.6.

The observation of positively correlated variable sets aligns with the information extracted from the original data (Table 4.2). For instance, variables like ADAS-11, ADAS-13, FAQ, ICV, and Ventricles exhibit higher average values for patients diagnosed with AD. This indicates that the addition of the first PC to the latent representations realistically changes the biomarker values. Furthermore, the heatmap illustrates that the values of ADAS-11, ADAS-13, FAQ, ICV, and Ventricles tend to increase together, while the whole brain volume behaves inversely. Another notable example is the positive relationship observed between MMSE and RAVLT measurements. As patients progress towards AD, both MMSE and RAVLT measurements decrease together.

Figure 4.18 depicts our examination of the ratio of detected test patients who convert to AD through the manipulation method [4]. The figure clearly illustrates that the number of detected converted patients increases in correspondence with the magnitude of manipulation.

In Figure 4.19, we illustrate the ratio of patients labeled as AD by the classifier after manipulating the test set. It is important to note that visits from different time points of the same MCI patient can be manipulated simultaneously using a fixed α value. Therefore, we consider a unique number of patients for this figure. One important consideration is that if the α value is set too high, the manipulation may cover nearly all patients in the test set, which is not desirable for real-life applications where the focus should be on detecting patients who will convert to AD.

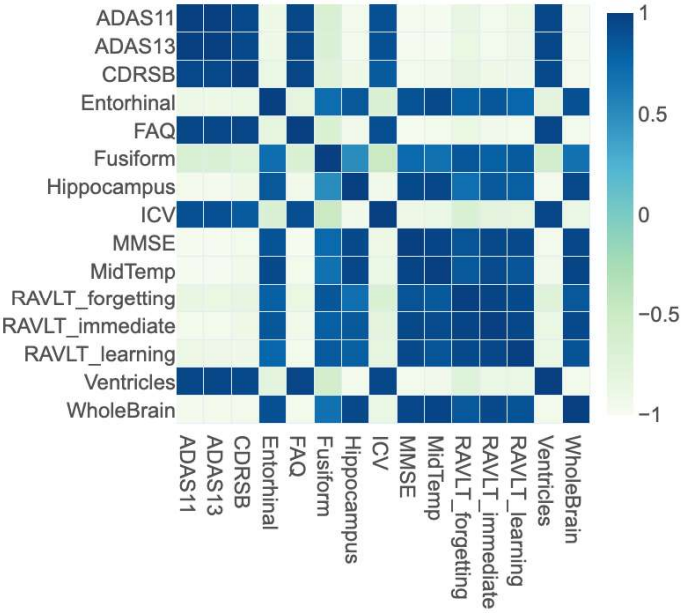


Figure 4.17. Correlations of the average changes in variables of the MCI patients who convert to AD after manipulation with the first PC with varying α for TADPOLE data set.

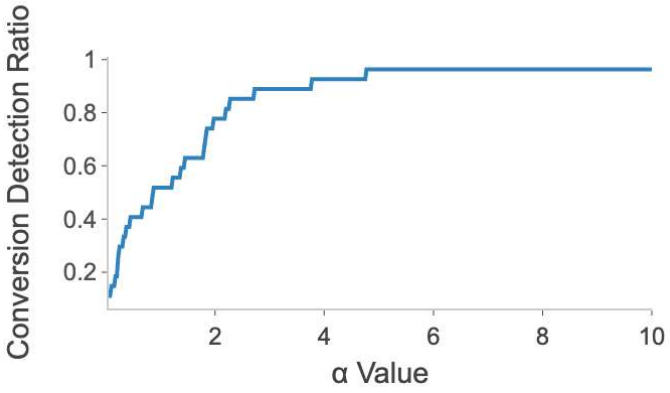


Figure 4.18. The ratio of conversion detection with different α values in TADPOLE data set.

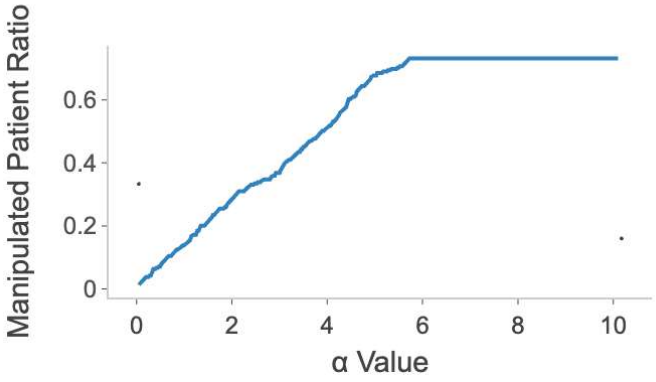


Figure 4.19. The ratio of the number of manipulated patients with different α values in TADPOLE data set.

Striking a balance between Figure 4.18 and Figure 4.19 is paramount for our method. The ultimate goal is to achieve a scenario where we can effectively detect a maximal number of progressive MCI patients, those on the verge of transitioning to AD, while minimizing the need for manipulation across the entire patient cohort. For instance, when α is set to three, we observe that approximately 89% of progressive MCI patients can be correctly identified, while the manipulation is applied to only about 36% of the entire cohort. These two figures clearly illustrate that by selecting an appropriate α value, clinicians can efficiently identify the majority of progressive MCI cases, while concentrating their efforts on a relatively small portion of the patient population.

Table 4.6 presents a comparison of the average changes between the manipulated and original time steps after the visit time of the manipulated data points for specific patients, using a fixed α value of three. This choice of α is based on the analysis in Figure 4.18 and Figure 4.19. Examination of future visit information after the manipulated data points for these patients reveals that the patient data changes due to manipulation align with reality. For example, clinically plausible results include an increase in the CDR-SB score and a decrease in the MMSE score as the disease

progresses. Although the manipulation effect may not completely capture changes in CDR-SB and FAQ variables, the alterations in other variables still enable the visit data to be classified as AD by the classifier.

Table 4.6. Average changes between the manipulated and original time steps labeled AD for a fixed $\alpha = 3$, for TADPOLE data set.

Variable	**Manipulated Data vs. Original Data**	**Future Visit Data vs. Original Data**
CDR-SB	%75	%97
RAVLT immediate	%-23	%-25
RAVLT forgetting	%-9	%-8
RAVLT learning	%-39	%-47
ADAS-11	%50	%48
ADAS-13	%45	%43
FAQ	%100	%161
MMSE	%-11	%-14
Ventricles	%18	%17
Hippocampus	%-15	%-10
Whole brain	%-6	%-4
Entorhinal	%-19	%-9
Fusiform	%-10	%-8
MidTemp	%-11	%-6
ICV	%2	%1

4.10.2. NACC Data Set

Figure 4.20 presents a heatmap illustrating the correlation between the average changes in variables following manipulation with the first PC for increasing α values on the NACC data set. It is important to note that Figure 4.20 doesn't explicitly indicate

whether variables increase or decrease after manipulation; this information can be found in Table 4.7. The observation of positively correlated variable sets aligns with the information extracted from the original data (Table 4.3). For instance, variables like FAQ and CDR-SB exhibit higher average values for patients diagnosed with AD, while MMSE has lower average values. This suggests that the addition of the first PC to the latent representations realistically alters the biomarker values.

In Figure 4.21, we investigate the ratio of detected test patients who convert to AD by the manipulation method for NACC data set. The results obtained for the NACC data set are consistent with the findings for the TADPOLE data set.

In Figure 4.22, we demonstrate the ratio of patients labeled as AD by the classifier after manipulating the NACC test set. For the concerns explained in Section 4.10.1, there needs to be a balance between Figure 4.21 and Figure 4.22. For example, when α is equal to 1.2, the ratio of detected progressive MCI patients is approximately 83%, while the percentage of manipulated patients is around 37%.

It is noteworthy that achieving the same ratio of the cohort to focus on with the NACC data set results in a slightly smaller ratio of detected patients compared to the TADPOLE data set (83% versus 89%). This observation may suggest better predictive performance with the TADPOLE data set. However, further investigation is needed, as discussed in Section 4.11. Additionally, the smoother graphs in the NACC data set can be attributed to the larger number of patients in this data set.

Table 4.7 presents a comparison of the average changes in variables after manipulation with a fixed α value for the NACC data set. In this analysis, an α value of 1.2 was chosen based on the results from Figure 4.21 and Figure 4.22. The future visit data following the manipulation of specific patients show changes that align with realistic disease progression. Notably, the manipulation effect on variables like CDR-SB and FAQ falls behind in changing these variables as expected, similar to the findings in Table 4.6.

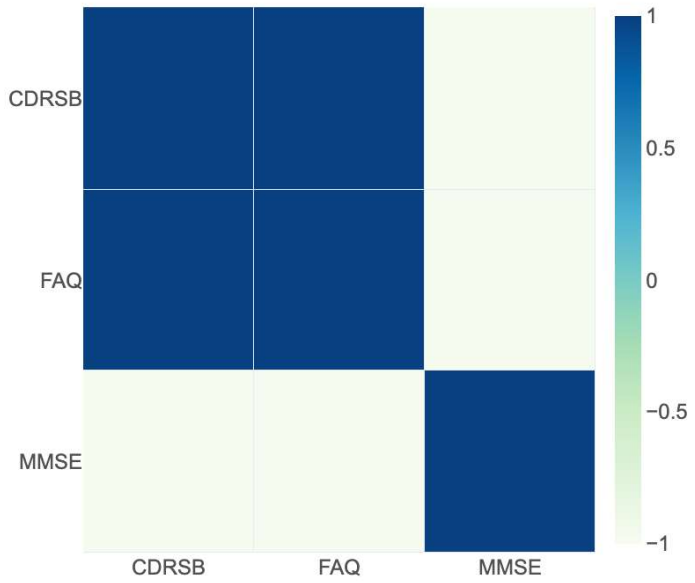


Figure 4.20. Correlations of the average changes in variables of the MCI patients who convert to AD after manipulation with the first PC with varying α for NACC data set.

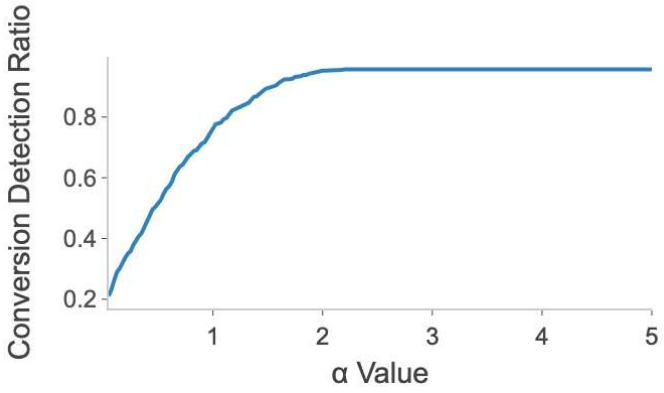


Figure 4.21. The ratio of conversion detection with different α values in NACC data set.

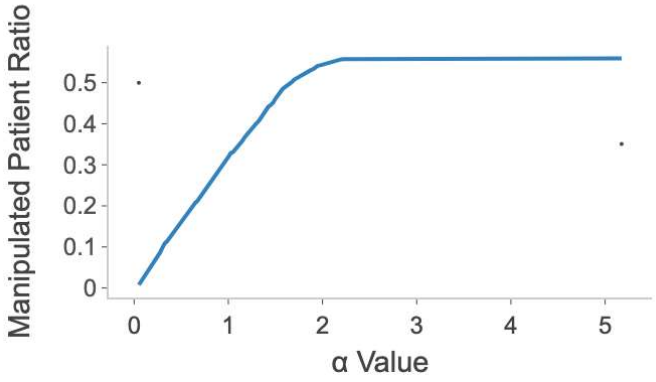


Figure 4.22. The ratio of the number of manipulated patients with different α values in NACC data set.

However, it is important to mention that the change ratios for CDR-SB and FAQ differ from those in Table 4.6. This discrepancy can be attributed to variations in data statistics between the TADPOLE and NACC data sets. While the mean and standard deviation of these variables in MCI-labeled data are not significantly different between the two data sets, there are differences in both mean and standard deviation for AD-labeled data in both data sets. These differences contribute to the distinct change ratios observed in both manipulated vs. original data and future visits vs. original data.

Table 4.7. Average changes between the manipulated and original time steps labeled AD for a fixed $\alpha = 1.2$, for NACC data set.

Variable	**Manipulated Data vs. Original Data**	**Future Visit Data vs. Original Data**
CDR-SB	%176	%159
FAQ	%217	%188
MMSE	%-14	%-12

The analyses conducted in this section suggest that the results are clinically plausible for both the TADPOLE and NACC data sets, as well as for the variables examined. These findings indicate that it is possible to identify a direction in the latent space of a deep network that encapsulates significant information about AD. The subsequent section will further explore these clinically plausible results within the context of predictive performance.

4.11. Predictive Performance

As described under Section 3.4, the method's predictive power is measured by creating an analogy with a classification procedure. In this section, we will evaluate the predictive performance of our method for the selected visit order and forecast horizon pairs. We will compare the predictive performance of our method with two baseline machine learning models: a logistic regression model and a gradient boosting model.

The evaluation metrics we will use to assess predictive performance are sensitivity and specificity. Sensitivity measures how well a model can correctly identify patient cases, while specificity measures the model's ability to accurately identify non-patient cases [80]. Both sensitivity and specificity are crucial in disease progression modeling, so we will select the trial with the best balance between these two metrics for our method.

Additionally, we will calculate the average grace period and the number of AD patients for each visit order and forecast horizon pair. The average grace period represents the average number of months between the visit data point manipulated for the selected visit order and the predicted visit data point for the chosen forecast horizon. The number of AD patients refers to the number of MCI patients during the specified visit time in the selected visit order who later converted to AD in the forecast time point.

To further validate our results, we will investigate the changes in future visit data for MCI patients who converted to AD after their manipulated visit data point shifted diagnosis. We will compare the biomarker values in their manipulated data with their average values in the actual future visit data. We will conduct paired two-sample t-tests for each biomarker value to determine if there is a significant difference between the manipulated data and the actual future visit data. The alternative hypothesis of the paired t-test states that there is significant difference between the manipulated data and actual future visit data.

4.11.1. TADPOLE Data Set

Table 4.8 presents the performance metrics for each visit order and forecast horizon pair of the predictions generated by our method, the logistic regression model, and the gradient boosting model using the TADPOLE data set. For instance, when manipulating the first visit data, our method can correctly identify 100%, 88%, and 90% of the patients who convert to AD in their second, third, and fourth visits, respectively. Additionally, our method achieves 99%, 85%, and 87% accuracy in detecting non-AD patients.

These results indicate that our method can effectively identify most of the future AD patients from their first visit while keeping the false positive rate low for future non-AD patients. As expected, sensitivity values tend to decrease as the forecast horizon increases, but they consistently remain at or above 70% for each visit order and forecast horizon pair. This suggests that our method can detect MCI patients who will convert to AD within both short and long-term forecasts.

In comparison, the gradient boosting and logistic regression models exhibit higher specificity values but lower sensitivity values. Their sensitivity values are less than or equal to 50%, indicating inferior performance compared to our proposed method. Notably, for the visit order and forecast horizon pair where the number of AD patients is one (as shown in Table 4.9), both the gradient boosting and logistic regression models

have zero sensitivity, signifying their failure to detect this particular patient. Overall, the results demonstrate that our method offers competitive performance in predicting disease progression compared to the baseline models, especially in terms of sensitivity.

Table 4.8. Performance metrics for each visit order - forecast horizon pair of the predictions of our method, logistic regression model, and gradient boosting model as baseline models for TADPOLE data set. Sens: sensitivity, spec: specificity, LR: logistic regression, GB: gradient boosting model.

Visit Order	**Forecast Horizon**	**Sens. Ours**	**Spec. Ours**	**Sens. LR**	**Spec. LR**	**Sens. GB**	**Spec. GB**
1	1	1.00	0.99	0.00	0.99	0.00	0.97
1	2	0.88	0.85	0.30	0.99	0.20	0.95
1	3	0.90	0.87	0.36	0.97	0.45	0.95
2	1	0.75	0.79	0.14	0.99	0.14	0.97
2	2	0.80	0.80	0.25	0.99	0.38	0.96
2	3	0.72	0.73	0.20	0.95	0.30	0.92
3	1	1.00	0.86	0.00	0.98	0.00	0.96
3	2	0.75	0.76	0.20	0.96	0.20	0.94
3	3	0.72	0.72	0.22	0.95	0.22	0.91

Table 4.9 provides information about the average grace period, number of AD patients, and optimal α values for each visit order and forecast horizon pair using the TADPOLE data set. A notable trend is the monotonic increase in the optimal α values as the forecast horizon increases within each visit order. This pattern is expected since, for a fixed visit order, as we project further into the future, we need a larger manipulation effect to better forecast the later stages of disease progression. This indicates that the magnitude of manipulation should be adjusted according to the forecast horizon to improve predictions.

Another interesting observation is the variability in optimal α values across different visit orders with the same forecast horizon. This suggests that the latent space representations of data from different visit orders may have varying distances to the data within the same forecast horizon. Consequently, different magnitudes of manipulation are required to achieve optimal predictions, highlighting the heterogeneity in AD progression.

In summary, these findings emphasize the importance of tailoring the manipulation magnitude to the specific visit order and forecast horizon, as they can significantly impact the predictive performance of the method. In Figure 4.23, we present an illustrative plot showcasing the performance metric values for a specific experiment with a visit order of one and a forecast horizon of three. This plot provides a visual representation of how varying α values impact the method's predictive performance.

Table 4.9. Average grace period, number of AD patients, and the optimal α values for each visit order - forecast horizon pair of the predictions of our method, for TADPOLE data set.

Visit Order	**Forecast Horizon**	$\boldsymbol{\alpha}$	**Average Grace Period**	**Number of AD**
1	1	0.575	4.69	1
1	2	3.125	9.66	8
1	3	3.125	16.33	10
2	1	3.350	5.29	8
2	2	3.400	12.17	10
2	3	4.050	18.85	11
3	1	2.825	7.23	2
3	2	3.525	14.14	4
3	3	3.775	20.50	7

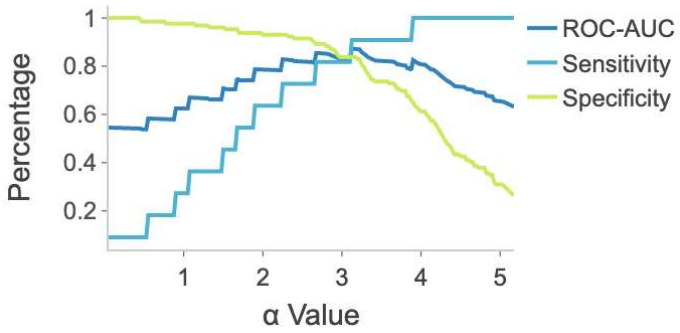


Figure 4.23. ROC-AUC, sensitivity, and specificity metrics for different α values with the unsupervised manipulation method, with visit order = 1, and forecast horizon = 3, on TADPOLE data set.

Turning our attention to Table 4.10, we focus on visit order of two and forecast horizon of two. This pair is chosen due to its favorable predictive performance and a sufficient number of patients for analysis. This table displays the p-values obtained from the previously mentioned t-tests. Notably, a majority of the biomarker values exhibit p-values exceeding the significance threshold of 0.05. This suggests that, for biomarkers with p-values above this threshold, we cannot reject the null hypothesis positing no significant difference between the manipulated data and actual future visit data.

4.11.2. NACC Data Set

Table 4.11 provides an overview of the performance metrics associated with each visit order and forecast horizon pair for the predictions made using the NACC data set. For instance, when manipulating the first visit data, our method demonstrates the ability to identify 65%, 65%, and 64% of patients who will convert to AD during their second, third, and fourth visits, respectively. Additionally, our method achieves a corresponding accuracy of 64%, 64%, and 64% in correctly classifying non-AD patients.

Table 4.10. p-values of two-sample t-test for each biomarker value, by comparing the values of each patient with its manipulated data and its future visit data, for TADPOLE data set.

Variable	**p-value**
CDR-SB	0.23
RAVLT immediate	0.47
RAVLT forgetting	0.17
RAVLT learning	0.06
ADAS-11	0.07
ADAS-13	0.06
FAQ	0.28
MMSE	0.08
Ventricles	0.16
Hippocampus	0.05
Whole brain	0.06
Entorhinal	0.12
Fusiform	0.05
MidTemp	0.01
ICV	0.17

These results signify that our method excels at detecting a significant portion of future AD patients based on their initial visit, while simultaneously producing fewer than 50% false positive alarms for prospective non-AD patients. As expected, sensitivity values tend to decrease as the forecast horizon extends for each visit order. Nevertheless, it is worth noting that sensitivity and specificity values consistently surpass or equal %60 for all visit order and forecast horizon pairs. This underscores the method's efficacy in identifying most MCI patients who will eventually convert to AD, both in the short and long-term.

In contrast, the gradient boosting model and logistic regression model exhibit superior specificity values but fall short in terms of sensitivity, with values consistently below or equal to %55. This discrepancy underscores the superior performance of our proposed method in detecting future AD patients compared to the baseline models.

Table 4.11. Performance metrics for each visit order - forecast horizon pair of the predictions of our method, logistic regression model, and gradient boosting model as baseline models for NACC data set. Sens: sensitivity, spec: specificity, LR: logistic regression, GB: gradient boosting model.

Visit Order	**Forecast Horizon**	**Sens. Ours**	**Spec. Ours**	**Sens. LR**	**Spec. LR**	**Sens. GB**	**Spec. GB**
1	1	0.65	0.64	0.22	0.95	0.24	0.94
1	2	0.65	0.64	0.33	0.91	0.33	0.90
1	3	0.64	0.64	0.45	0.82	0.45	0.83
2	1	0.70	0.69	0.37	0.92	0.38	0.91
2	2	0.69	0.68	0.44	0.87	0.42	0.87
2	3	0.64	0.63	0.43	0.82	0.42	0.82
3	1	0.69	0.69	0.32	0.92	0.32	0.91
3	2	0.67	0.67	0.44	0.88	0.38	0.87
3	3	0.67	0.68	0.54	0.80	0.55	0.79

Table 4.12 provides insights into the average grace period, number of AD patients, and optimal α values for various visit order and forecast horizon pairs using the NACC data set. Here, we observe that as the forecast horizon increases within a fixed visit order, the optimal α value also increases. This trend is consistent with the findings for the TADPOLE data set.

Additionally, it is worth noting that the optimal α values decrease as the visit order increases within the same forecast horizon. This observation is also consistent

with the findings for the TADPOLE data set. For example, the optimal α value for the proximity between the first visit and the second visit data ($\alpha = 1$) suggests a lesser proximity compared to that of the second visit and the third visit data ($\alpha = 0.9$).

The average grace period values in the NACC data set differ notably from those in the TADPOLE data set. In the NACC data set, the average number of months between the visit order data point and the data points within the forecast horizon is nearly three times greater for each visit order and forecast horizon pair compared to the TADPOLE data set. Furthermore, for each visit order and forecast horizon pair, the number of patients who eventually convert to AD is considerably higher in the NACC data set compared to the TADPOLE data set. These observations underscore the distinctions in patient data between the two data sets. In Figure 4.24, an example plot for visit order equal to one and forecast horizon equal to three is demonstrated, as the performance metrics values for experimenting with different α values.

Table 4.12. Average grace period, number of AD patients, and the optimal α values for each visit order - forecast horizon pair of the predictions of our method, for NACC data set.

Visit Order	**Forecast Horizon**	**α**	**Average Grace Period**	**Number of AD**
1	1	1	17.66	223
1	2	1.075	32.74	252
1	3	1.1	47.45	231
2	1	0.9	16.45	112
2	2	1	31.51	135
2	3	1.1	46.48	103
3	1	0.875	15.98	68
3	2	0.95	30.96	63
3	3	0.975	46.44	67

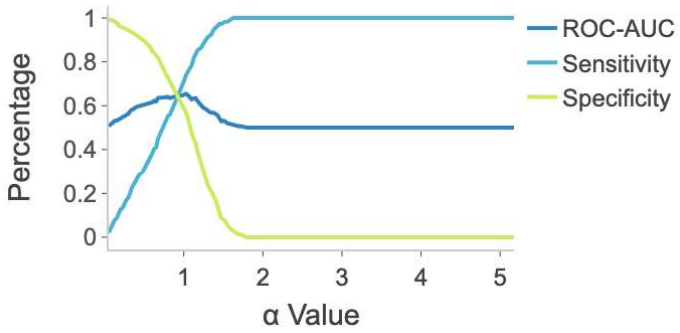


Figure 4.24. ROC-AUC, sensitivity, and specificity metrics for different α values with the unsupervised manipulation method, with visit order = 1, and forecast horizon = 3, on NACC data set.

Table 4.13 provides the p-values resulting from the t-tests mentioned earlier in this section. Specifically, we focus on visit order two and forecast horizon two due to the favorable predictive performance and a sufficiently sized patient sample. The table reveals that the majority of biomarker values exhibit p-values exceeding the significance threshold of 0.05. This implies that for biomarkers with p-values surpassing this threshold, the null hypothesis of the paired t-test, claiming no significant difference between the manipulated data and actual future visit data, cannot be rejected. The results conform with the findings for the TADPOLE data set.

Table 4.13. p-values of two-sample t-test for each biomarker value, by comparing the values of each patient with its manipulated data and its future visit data, for NACC data set.

Variable	**p-value**
CDR-SB	0.88
FAQ	0.46
MMSE	0.25

In essence, the findings suggest that patients who progress to AD undergo changes in their biomarker values that closely align with the alterations predicted by the manipulation method. This reinforces the method's ability to capture meaningful changes in patients' conditions and lends further support to its clinical plausibility.

Our method exhibits superior predictive performance in the TADPOLE data set. Several factors contribute to this discrepancy. Firstly, the NACC data set used in this study lacks MRI and PET data, and some of the cognitive test scores are missing, which results in a less comprehensive set of multi-modal attributes. Secondly, the average grace period in the NACC data set is notably higher than in the TADPOLE data set, indicating that the model must anticipate events further into the future. These two factors collectively render the prediction of AD progression more challenging in the NACC data set compared to the TADPOLE data set.

Despite the lower predictive performance in the NACC data set when compared to the results obtained from the TADPOLE data set, our proposed method still outperforms the baseline models in detecting future AD patients. Furthermore, the larger number of patients who convert to AD in the NACC data set enables a more statistically robust analysis of the relationship between α and disease progression. By comparing the values in Table 4.9 and Table 4.12, it becomes more apparent in the latter that as the forecast horizon increases with a fixed visit order, the optimal α values increase. Conversely, as the visit order increases with a fixed forecast horizon, the optimal α values decrease. This pattern provides valuable insights into the relationship between disease progression and the magnitude of manipulation, as well as the inherent heterogeneity of AD progression.

4.12. Data Generation Performance

To assess the data generation performance of our proposed method, we conducted experiments similar to those used to measure the generative performance of the VAE model in Section 4.5. The process involved training a classifier on a combination of

real and synthetic patient data, followed by evaluating its performance on the test data from both the TADPOLE and NACC data sets.

Table 4.14. Performance metrics of the classifier trained with a mix of actual and VAE-generated data, demonstrating VAE data generation performance for TADPOLE and NACC data sets.

Data Set	**ROC-AUC**	**F1**
TADPOLE	0.88	0.86
NACC	0.90	0.90

Synthetic AD patient data were generated by manipulating MCI patient visit data, which were labeled as AD by the pre-trained binary classifier. A new training data set was created by combining 100% of the actual (not manipulated) MCI patient visit data with 25% of the real AD patient visit data from the training data set. The synthetic AD data was also added to this training data set. A classifier, identical to the pre-trained binary classifier, was trained using this new training data set. The predictive performance of this newly trained classifier on the test data set was measured.

The classification performance of the new binary classifier on the test data is provided in Table 4.14 for both the TADPOLE and NACC data sets. High predictive performance on the test data of the model trained with generated AD patient visit data through manipulation suggests that our proposed method can effectively generate realistic synthetic patient visit data.

5. CONCLUSION AND FUTURE WORK

Neurodegenerative diseases like AD pose a significant global health concern, with diverse progression patterns and factors influencing their development. MCI serves as an intermediate stage between healthy individuals and AD patients, and accurately identifying MCI patients who will progress to AD is crucial for early intervention. Previous research has mainly focused on single-time point predictions or future time points. However, our goal is to decipher the changes in multi-modal measurements that lead to the conversion of MCI to AD and utilize a method to model the disease progression.

For this purpose, a framework based on a variational auto-encoder and a binary classifier is proposed. Latent space manipulation techniques are applied to the latent space of VAE trained with MCI and AD patient data. First, PCA and SVM are applied to the VAE latent space representations of the training data set for unsupervised and supervised manipulation, respectively. This process helps obtain unit direction vectors for manipulation. An affine transformation is applied to the latent space of VAE with unit direction vectors and a coefficient to control the magnitude. Then, the effect of manipulation is evaluated by a pre-trained binary classifier.

The experiments use the TADPOLE challenge data set [24] and NACC data set [59]. The TADPOLE data set was obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI), a longitudinal multicenter study focused on creating biomarkers for Alzheimer's disease [60]. The data set contains various types of measurements, including MRI, PET, CSF, cognitive tests, genetic information, and demographic data. The NACC data set, established by the National Institute on Aging, is a comprehensive database of standardized clinical and neuropathological research data gathered from Alzheimer's Disease Research Centers across the United States. The Uniform Data Set (UDS), which includes longitudinal clinical evaluation data, is used from the NACC data set [69].

The experiments are performed on the test data sets, focusing on MCI-labeled data points correctly classified as MCI by the pre-trained binary classifier. It is found that the first principal component vector and the unit direction vector obtained via the SVM classifier's parameter vector are directed toward AD in the latent space. According to the cosine similarity value of these two vectors, they have similar directions, indicating potential similarity in the results of manipulation experiments. Another observation is that among the first five principal component vectors, the first principal component vector captures the most informative direction and produces the most significant change in moving the latent representation of an MCI patient toward AD. The magnitude of manipulation is shown to have a non-linear effect on conversion from MCI to AD for both data sets and manipulation methods. Since this experiment implies similar effects for both methods, the unsupervised manipulation with the first principal component is selected to demonstrate the rest of the experimental results.

Prediction tasks are divided based on the patient data point's visit order and the forecast horizon's length. Performance metrics of the predictions of our method for each visit order and forecast horizon pair are demonstrated. The predictive tasks are conducted with various manipulation magnitude values, and the one with the best balance between specificity and sensitivity is selected. The results are compared to the baseline models, selected as a logistic regression model and a gradient boosting model. Although the results with the TADPOLE data set are better than those with the NACC data set, both results outweigh the baseline models in terms of sensitivity. Furthermore, we hypothesize the correlation between AD progression and manipulation magnitude.

The effect of manipulation is shown on the latent space of the classifier, and the realistic data generation performance of the method is further investigated, leading to promising results. The manipulated data is also investigated in terms of how it changes MCI patient data realistically. The average changes in attributes align with the actual data statistics.

The experimental results indicate that it is possible to discover a clinically meaningful direction in the latent space of a deep network. The proposed framework shows the potential to reveal underlying information captured by deep models and use this information to generate realistic synthetic patient data. The proposed framework is also proven to be utilized as a disease progression modeler and forecaster.

In terms of future work, two key recommendations are outlined. Firstly, considering the temporal dimension is pivotal. Recognizing the potential significance of temporal patterns in conveying essential information about disease progression, delving deeper into the sequential handling of patient data for latent space manipulation holds promise for further exploration. Secondly, the concepts of robustness and the extent of applicability warrant attention. The TADPOLE data set achieves greater success in predictive tasks attributed to the inclusion of MRI and PET data, along with more frequent visitations. On the other hand, the NACC data set excels in discerning the relationship between manipulation magnitude and disease progression, thanks to its substantial patient count. In forthcoming research, subjecting the proposed method to data sets associated with other diseases sharing similarities with AD would be valuable, thereby expanding its applicability across a broader spectrum of medical conditions.

REFERENCES

1. European Progression of Neurological Disease Initiative and Alzheimer's Disease Neuroimaging Initiative, "The Alzheimer's Disease Prediction Of Longitudinal Evolution (TADPOLE) Challenge", 2017, https://tadpole.grand-challenge.org, accessed on April 15, 2021.

2. European Progression of Neurological Disease Initiative, 2020, http://europond.eu, accessed on August 15, 2023.

3. Foundation for the National Institutes of Health, 2022, http://fnih.org, accessed on August 15, 2023.

4. Ayvaz, D. S. and I. M. Baytaş, "Investigating Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using Latent Space Manipulation", *The International Symposium on Health Informatics and Bioinformatics*, p. 2097, Erdemli, Mersin, Turkey, 2022.

5. Hou, Y., X. Dan, M. Babbar, Y. Wei, S. G. Hasselbalch, D. L. Croteau and V. A. Bohr, "Ageing as a Risk Factor for Neurodegenerative Disease", *Nature Reviews Neurology*, Vol. 15, No. 10, pp. 565–581, 2019.

6. Sun, N., E. C. Mormino, J. Chen, M. R. Sabuncu and B. T. Yeo, "Multi-Modal Latent Factor Exploration of Atrophy, Cognitive and Tau Heterogeneity in Alzheimer's Disease", *NeuroImage*, Vol. 201, p. 116043, 2019.

7. Devi, G. and P. Scheltens, "Heterogeneity of Alzheimer's Disease: Consequence for Drug Trials?", *Alzheimer's Research & Therapy*, Vol. 10, No. 1, pp. 1–3, 2018.

8. 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*, Vol. 11, No. 1, pp. 1–13, 2021.

9. Libon, D., L. Delano-Wood, M. Bondi and R. Au, "Mild Cognitive Impairment", *Encyclopedia of the Neurological Sciences*, pp. 72–75, Elsevier, 2014.

10. Plassman, B. L., K. M. Langa, G. G. Fisher, S. G. Heeringa, D. R. Weir, M. B. ofstedal, J. R. Burke, M. D. Hurd, G. G. Potter, W. L. Rodgers, D. C. Steffens, J. J. McArdle, R. J. Willis and R. B. Wallace, "Prevalence of Cognitive Impairment without Dementia in the United States", *Annals of Internal Medicine*, Vol. 148, No. 6, pp. 427–434, 2008.

11. Minhas, S., A. Khanum, F. Riaz, S. a. Khan and A. Alvi, "Predicting Progression from Mild Cognitive Impairment to Alzheimer's Disease Using Autoregressive Modelling of Longitudinal and Multimodal Biomarkers", *IEEE Journal of Biomedical and Health Informatics*, Vol. 22, No. 3, pp. 818–825, 2017.

12. Wang, N., M. Chen and K. P. Subbalakshmi, "Explainable CNN-Attention Networks (C-Attention Network) for Automated Detection of Alzheimer's Disease", ArXiv:2006.14135, 2020.

13. Nguyen, M., tong He, L. An, D. C. Alexander, J. Feng and B. T. Yeo, "Predicting Alzheimer's Disease Progression Using Deep Recurrent Neural Networks", *NeuroImage*, Vol. 222, p. 117203, 2020.

14. Centers for Disease Control and Prevention, "Alzheimer's Disease and Healthy Aging", 2019, https://www.cdc.gov/aging/dementia/index.html, accessed on August 31, 2021.

15. World Health Organization, "Dementia", 2020, https://www.who.int/news-room/fact-sheets/detail/dementia, accessed on August 31, 2021.

16. Jiang, P., X. Wang, Q. Li, L. Jin and S. Li, "Correlation-Aware Sparse and Low-Rank Constrained Multi-Task Learning for Longitudinal Analysis of Alzheimer's Disease", *IEEE Journal of Biomedical and Health Informatics*, Vol. 23, No. 4, pp.

1450–1456, 2018.

17. Mueller, S. G., M. W. Weiner, L. J. Thal, R. C. Petersen, C. Jack, W. Jagust, J. Q. Trojanowski, A. W. Toga and L. Beckett, "The Alzheimer's Disease Neuroimaging Initiative", *Neuroimaging Clinics*, Vol. 15, No. 4, pp. 869–877, 2005.

18. Rabin, L. a., N. Paré, A. J. Saykin, M. J. Brown, H. a. Wishart, L. a. Flashman and R. B. Santulli, "Differential Memory Test Sensitivity for Diagnosing Amnestic Mild Cognitive Impairment and Predicting Conversion to Alzheimer's Disease", *Aging, Neuropsychology, and Cognition*, Vol. 16, No. 3, pp. 357–376, 2009.

19. Petersen, R. C., R. O. Roberts, D. S. Knopman, B. F. Boeve, Y. E. Geda, R. J. Ivnik, G. E. Smith and C. R. Jack, "Mild Cognitive Impairment: Ten Years Later", *Archives of Neurology*, Vol. 66, No. 12, pp. 1447–1455, 2009.

20. Dubois, B., H. H. Feldman, C. Jacova, J. L. Cummings, S. T. DeKosky, P. Barberger-Gateau, A. Delacourte, G. Frisoni, N. C. Fox, D. Galasko, S. Gauthier, H. Hampel, G. A. Jicha, K. Meguro, J. O'Brien, F. Pasquier, P. Robert, M. Rossor, S. Salloway, M. Sarazin, L. C. de Souza, Y. Stern, P. J. Visser and P. Scheltens, "Revising the Definition of Alzheimer's Disease: A New Lexicon", *The Lancet Neurology*, Vol. 9, No. 11, pp. 1118–1127, 2010.

21. Chapman, R. M., M. Mapstone, J. W. McCrary, M. N. Gardner, A. Porsteinsson, T. C. Sandoval, M. D. Guillily, E. DeGrush and L. A. Reilly, "Predicting Conversion from Mild Cognitive Impairment to Alzheimer's Disease Using Neuropsychological Tests and Multivariate Methods", *Journal of Clinical and Experimental Neuropsychology*, Vol. 33, No. 2, pp. 187–199, 2011.

22. Lu, X., J. Chen, H. Shu, Z. Wang, Y.-m. Shi, Y.-g. Yuan, C.-m. Xie, W.-x. Liao, F. Su, Y.-c. Shi and Z.-J. Zhang, "Predicting Conversion to Alzheimer's Disease Among Individual High-risk Patients Using the Characterizing AD Risk Events Index Model", *CNS Neuroscience & Therapeutics*, Vol. 26, No. 7, pp. 720–729,

2020.

23. Teipel, S. J., M. Dyrba, P. a. Chiesa, F. Sakr, I. Jelistratova, S. Lista, A. Vergallo, P. Lemercier, E. Cavedo, M. O. Habert, B. Dubois, H. Hampel, M. J. Grothe, INSIGHT-preAD Study Group and the Alzheimer's Disease Neuroimaging Initiative, "In Vivo Staging of Regional Amyloid Deposition Predicts Functional Conversion in the Preclinical and Prodromal Phases of Alzheimer's Disease", *Neurobiology of Aging*, Vol. 93, pp. 98–108, 2020.

24. Marinescu, R. V., N. P. Oxtoby, A. L. Young, E. E. Bron, A. W. Toga, M. W. Weiner, F. Barkhof, N. C. Fox, P. Golland, S. Klein and D. C. Alexander, "TADPOLE Challenge: Accurate Alzheimer's Disease Prediction Through Crowdsourced Forecasting of Future Data", *International Workshop on Predictive Intelligence in Medicine*, pp. 1–10, Springer, 2019.

25. Kung, T.-H., T.-C. Chao, Y.-R. Xie, M.-C. Pai, Y.-M. Kuo and G. G. C. Lee, "Neuroimage Biomarker Identification of the Conversion of Mild Cognitive Impairment to Alzheimer's Disease", *Frontiers in Neuroscience*, Vol. 15, p. 584641, 2021.

26. Kuang, J., P. Zhang, T. Cai, Z. Zou, L. Li, N. Wang and L. Wu, "Prediction of Transition from Mild Cognitive Impairment to Alzheimer's Disease Based on a Logistic Regression–artificial Neural Network–decision Tree Model", *Geriatrics & Gerontology International*, Vol. 21, No. 1, pp. 43–47, 2021.

27. Baytaş, I. M., C. Xiao, X. Zhang, F. Wang, A. K. Jain and J. Zhou, "Patient Subtyping via Time-Aware LSTM Networks", *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 65–74, Halifax, NS, Canada, 2017.

28. Pham, T., T. Tran, D. Phung and S. Venkatesh, "Deepcare: A Deep Dynamic Memory Model for Predictive Medicine", *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pp. 30–41, Springer, 2016.

29. Choi, E., M. T. Bahadori, A. Schuetz, W. F. Stewart and J. Sun, "Doctor AI: Predicting Clinical Events via Recurrent Neural Networks", *Machine Learning for Healthcare Conference*, pp. 301–318, 2016.

30. Lee, C. and M. Van Der Schaar, "Temporal Phenotyping Using Deep Predictive Clustering of Disease Progression", *International Conference on Machine Learning*, pp. 5767–5777, 2020.

31. Termine, A., C. Fabrizio, C. Strafella, V. Caputo, L. Petrosini, C. Caltagirone, E. Giardina and R. Cascella, "Multi-Layer Picture of Neurodegenerative Diseases: Lessons from the Use of Big Data through Artificial Intelligence", *Journal of Personalized Medicine*, Vol. 11, No. 4, p. 280, 2021.

32. Basheera, S. and M. S. S. Ram, "A Novel CNN Based Alzheimer's Disease Classification Using Hybrid Enhanced ICA Segmented Gray Matter of MRI", *Computerized Medical Imaging and Graphics*, Vol. 81, p. 101713, 2020.

33. Amini, M., M. Pedram, A. Moradi and M. Ouchani, "Diagnosis of Alzheimer's Disease Severity with fMRI Images Using Robust Multitask Feature Extraction Method and Convolutional Neural Network (CNN)", *Computational and Mathematical Methods in Medicine*, pp. 1–15, 2021.

34. Parmar, H. S., B. Nutter, R. Long, S. Antani and S. Mitra, "Deep Learning of Volumetric 3D CNN for fMRI in Alzheimer's Disease Classification", *Medical Imaging: Biomedical Applications in Molecular, Structural, and Functional Imaging*, Vol. 11317, p. 113170C, International Society for Optics and Photonics, 2020.

35. AbdulAzeem, Y., W. M. Bahgat and M. Badawy, "A CNN Based Framework for Classification of Alzheimer's Disease", *Neural Computing and Applications*, pp. 1–14, 2021.

36. Xia, Z., G. Yue, Y. Xu, C. Feng, M. Yang, T. Wang and B. Lei, "A Novel End-

to-End Hybrid Network for Alzheimer's Disease Detection Using 3D CNN and 3D CLSTM", *IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, pp. 1–4, 2020.

37. Liu, M., J. Zhang, E. Adeli and D. Shen, "Joint Classification and Regression via Deep Multi-task Multi-Channel Learning for Alzheimer's Disease Diagnosis", *IEEE Transactions on Biomedical Engineering*, Vol. 66, No. 5, pp. 1195–1206, 2018.

38. Ocasio, E. and T. Q. Duong, "Deep Learning Prediction of Mild Cognitive Impairment Conversion to Alzheimer's Disease At 3 Years After Diagnosis Using Longitudinal and Whole-Brain 3D MRI", *PeerJ Computer Science*, Vol. 7, p. e560, 2021.

39. Gao, F., H. Yoon, Y. Xu, D. Goradia, J. Luo, T. Wu, Y. Su and Alzheimer's Disease Neuroimaging Initiative, "AD-NET: Age-Adjust Neural Network for Improved MCI to AD Conversion Prediction", *NeuroImage: Clinical*, Vol. 27, p. 102290, 2020.

40. Bringas, S., S. Salomón, R. Duque, C. Lage and J. L. Montaña, "Alzheimer's Disease Stage Identification Using Deep Learning Models", *Journal of Biomedical Informatics*, Vol. 109, p. 103514, 2020.

41. Li, L., Y. Yang, Q. Zhang, J. Wang, J. Jiang and Alzheimer's Disease Neuroimaging Initiative, "Use of Deep-Learning Genomics to Discriminate Healthy Individuals from Those with Alzheimer's Disease or Mild Cognitive Impairment", *Behavioural Neurology*, 2021.

42. Zhou, J., L. Hu, Y. Jiang and L. Liu, "A Correlation Analysis Between SNPs and ROIs of Alzheimer's Disease Based on Deep Learning", *BioMed Research International*, pp. 1–13, 2021.

43. Lee, G., K. Nho, B. Kang, K.-A. Sohn and D. Kim, "Predicting Alzheimer's Dis-

ease Progression Using Multi-Modal Deep Learning Approach", *Scientific Reports*, Vol. 9, No. 1, pp. 1–12, 2019.

44. Spasov, S., L. Passamonti, A. Duggento, P. Lio, N. toschi and Alzheimer's Disease Neuroimaging Initiative, "A Parameter-Efficient Deep Learning Approach to Predict Conversion from Mild Cognitive Impairment to Alzheimer's Disease", *Neuroimage*, Vol. 189, pp. 276–287, 2019.

45. Suk, H.-I. and D. Shen, "Deep Learning Based Feature Representation for AD/MCI Classification", *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 583–590, Springer, 2013.

46. Bi, X., S. Li, B. Xiao, Y. Li, G. Wang and X. Ma, "Computer Aided Alzheimer's Disease Diagnosis by an Unsupervised Deep Learning Technology", *Neurocomputing*, Vol. 392, pp. 296–304, 2020.

47. Li, H. and Y. Fan, "Early Prediction of Alzheimer's Disease Dementia Based on Baseline Hippocampal MRI and 1-Year Follow-Up Cognitive Measures Using Deep Recurrent Neural Networks", *IEEE 16th International Symposium on Biomedical Imaging*, pp. 368–371, 2019.

48. Wegmayr, V., M. Hörold and J. M. Buhmann, "Generative Aging of Brain MR-Images and Prediction of Alzheimer Progression", *German Conference on Pattern Recognition*, pp. 247–260, Springer, 2019.

49. Shen, Y., J. Gu, X. Tang and B. Zhou, "Interpreting the Latent Space of GANs for Semantic Face Editing", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9243–9252, 2020.

50. Fetty, L., M. Bylund, P. Kuess, G. Heilemann, T. Nyholm, D. Georg and t. Löfstedt, "Latent Space Manipulation for High-resolution Medical Image Synthesis via the StyleGAN", *Zeitschrift für Medizinische Physik*, Vol. 30, No. 4, pp.

305–314, 2020.

51. Goetschalckx, L., A. Andonian, A. Oliva and P. Isola, "GANalyze: Toward Visual Definitions of Cognitive Image Properties", *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 5744–5753, 2019.

52. Härkönen, E., A. Hertzmann, J. Lehtinen and S. Paris, "GANSpace: Discovering Interpretable GAN Controls", *Advances in Neural Information Processing Systems*, Vol. 33, pp. 9841–9850, 2020.

53. DeVries, T. and G. W. Taylor, "Dataset Augmentation in Feature Space", ArXiv:1702.05538, 2017.

54. Cheung, T.-H. and D.-Y. Yeung, "{MODALS}: Modality-agnostic Automated Data Augmentation in the Latent Space", *International Conference on Learning Representations*, 2021.

55. Kingma, D. P. and M. Welling, "Auto-Encoding Variational Bayes", ArXiv:1312.6114, 2013.

56. F.R.S., K. P., "LIII. on Lines and Planes of Closest Fit to Systems of Points in Space", *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, Vol. 2, No. 11, pp. 559–572, 1901.

57. Shen, Y., J. Gu, X. Tang and B. Zhou, "Interpreting the Latent Space of GANs for Semantic Face Editing", *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9243–9252, 2020.

58. Cortes, C. and V. Vapnik, "Support-Vector Networks", *Machine Learning*, Vol. 20, No. 3, pp. 273–297, 1995.

59. Beekly, D. L., E. M. Ramos, W. W. Lee, W. D. Deitrich, M. E. Jacka, J. Wu, J. L. Hubbard, T. D. Koepsell, J. C. Morris, W. a. Kukull and NIA Alzheimer's Disease

Centers, “The National Alzheimer’s Coordinating Center (NACC) Database: the Uniform Data Set”, *Alzheimer Disease & Associated Disorders*, Vol. 21, No. 3, pp. 249–258, 2007.

60. Alzheimer’s Disease Neuroimaging Initiative, “Alzheimer’s Disease Neuroimaging Initiative”, 2017, https://adni.loni.usc.edu, accessed on April 30, 2021.

61. Carrillo, M. C., C. a. Sanders and R. G. Katz, “Maximizing the Alzheimer’s Disease Neuroimaging Initiative II”, *Alzheimer’s & Dementia*, Vol. 5, No. 3, pp. 271–275, 2009.

62. 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, R. C. Petersen, J. Salazar, A. J. Saykin, L. M. Shaw, A. W. Toga, J. Q. Trojanowski and Alzheimer’s Disease Neuroimaging Initiative, “The Alzheimer’s Disease Neuroimaging Initiative 3: Continued Innovation for Clinical Trial Improvement”, *Alzheimer’s & Dementia*, Vol. 13, No. 5, pp. 561–571, 2017.

63. Petersen, R. C., P. Aisen, L. A. Beckett, M. Donohue, A. Gamst, D. J. Harvey, C. Jack, W. Jagust, L. Shaw, A. Toga, J. Trojanowski and M. Weiner, “Alzheimer’s Disease Neuroimaging Initiative (ADNI): Clinical Characterization”, *Neurology*, Vol. 74, No. 3, pp. 201–209, 2010.

64. Albert, M. S., S. T. DeKosky, D. Dickson, B. Dubois, H. H. Feldman, N. C. Fox, A. Gamst, D. M. Holtzman, W. J. Jagust, R. C. Petersen, P. J. Snyder, M. C. Carrillo, B. Thies and C. H. Phelps, “The Diagnosis of Mild Cognitive Impairment Due to Alzheimer’s Disease: Recommendations from the National Institute on Aging-Alzheimer’s Association Workgroups on Diagnostic Guidelines for Alzheimer’s Disease”, *Alzheimer’s & Dementia*, Vol. 7, No. 3, pp. 270–279, 2011.

65. McKhann, G. M., D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack Jr,

C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, R. C. Mohs, J. C. Morris, M. N. Rossor, P. Scheltens, M. C. Carrillo, B. Thies, S. Weintraub and C. H. Phelps, "The Diagnosis of Dementia Due to Alzheimer's Disease: Recommendations from the National Institute on Aging-Alzheimer's Association Workgroups on Diagnostic Guidelines for Alzheimer's Disease", *Alzheimer's & Dementia*, Vol. 7, No. 3, pp. 263–269, 2011.

66. Jack Jr, C. R., D. S. Knopman, W. J. Jagust, R. C. Petersen, M. W. Weiner, P. S. Aisen, L. M. Shaw, P. Vemuri, H. J. Wiste, S. D. Weigand, T. G. Lesnick, V. S. Pankratz, M. C. Donohue and J. Q. Trojanowski, "Tracking Pathophysiological Processes in Alzheimer's Disease: An Updated Hypothetical Model of Dynamic Biomarkers", *The Lancet Neurology*, Vol. 12, No. 2, pp. 207–216, 2013.

67. Fischl, B., "FreeSurfer", *Neuroimage*, Vol. 62, No. 2, pp. 774–781, 2012.

68. Jack Jr, C. R., D. A. Bennett, K. Blennow, M. C. Carrillo, B. Dunn, S. B. Haeberlein, D. M. Holtzman, W. Jagust, F. Jessen, J. Karlawish, E. Liuk, J. L. Molinuevol, T. Montinem, C. Phelpsn, K. P. Rankino, C. C. Rowep, P. Scheltens, E. Siemers, H. M. Synder and R. Sperling, "NIA-AA research framework: toward a biological definition of Alzheimer's disease", *Alzheimer's & Dementia*, Vol. 14, No. 4, pp. 535–562, 2018.

69. National Alzheimer's Coordinating Center , "NACC Uniform Data Set Researchers Data Dictionary", 2015, https://files.alz.washington.edu/documentation/uds3-rdd.pdf, accessed on January 30, 2023.

70. Gainotti, G. and C. Marra, "Some Aspects of Memory Disorders Clearly Distinguish Dementia of the Alzheimer's Type from Depressive Pseudo-Dementia", *Journal of Clinical and Experimental Neuropsychology*, Vol. 16, No. 1, pp. 65–78, 1994.

71. Sargolzaei, S., A. Sargolzaei, M. Cabrerizo, G. Chen, M. Goryawala, S. Noei,

Q. Zhou, R. Duara, W. Barker and M. Adjouadi, "A Practical Guideline for Intracranial Volume Estimation in Patients with Alzheimer's Disease", *BMC Bioinformatics*, Vol. 16, No. 7, pp. 1–10, 2015.

72. Voevodskaya, O., A. Simmons, R. Nordenskjöld, J. Kullberg, H. Ahlström, L. Lind, L.-O. Wahlund, E.-M. Larsson, E. Westman and Alzheimer's Disease Neuroimaging Initiative, "The Effects of Intracranial Volume Adjustment Approaches on Multiple Regional MRI Volumes in Healthy Aging and Alzheimer's Disease", *Frontiers in Aging Neuroscience*, Vol. 6, p. 264, 2014.

73. Bisong, E., *Google Colaboratory*, pp. 59–64, Apress, Berkeley, CA, 2019.

74. Abadi, M., A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mane, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viegas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu and X. Zheng, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems", ArXiv:1603.04467, 2016.

75. Agarap, A. F., "Deep Learning Using Rectified Linear Units (ReLu)", ArXiv:1803.08375, 2018.

76. Higgins, I., L. Matthey, A. Pal, C. Burgess, X. Glorot, M. Botvinick, S. Mohamed and A. Lerchner, "beta-VAE: Learning Basic Visual Concepts with a Constrained Variational Framework", *International Conference on Learning Representations*, 2017.

77. Ioffe, S. and C. Szegedy, "Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift", *International Conference on Machine Learning*, pp. 448–456, 2015.

78. Kingma, D. P. and J. Ba, "Adam: a Method for Stochastic Optimization", ArXiv:1412.6980, 2014.

79. Leontev, M., A. Mikheev, K. Sviatov and S. Sukhov, "Quality Metrics of Variational Autoencoders", *International Conference on Information Technology and Nanotechnology (ITNT)*, pp. 1–5, 2020.

80. Baratloo, A., M. Hosseini, A. Negida and G. El Ashal, "Part 1: Simple Definition and Calculation of Accuracy, Sensitivity and Specificity", *Emerg (Tehran)*, Vol. 3, No. 2, pp. 48–49, 2015.