

LOW DIMENSIONAL BRAIN CONNECTOME REPRESENTATIONS FOR
ALZHEIMER'S DISEASE DEMENTIA

by

Güneş Bayır

B.S., Electrical and Electronics Engineering, İhsan Doğramacı Bilkent University,
2010

M.S., Electrical and Electronics Engineering, İhsan Doğramacı Bilkent University,
2013

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

Graduate Program in Electrical and Electronics Engineering
Boğaziçi University
2025

ACKNOWLEDGEMENTS

I would like to begin by expressing my heartfelt gratitude and sincere respect to my supervisor, Prof. Burak Acar, whose unwavering support, patience, motivation, and enthusiasm have guided and encouraged me throughout every stage of my research. This dissertation would not have been possible without his dedicated mentorship and continuous assistance.

I would also like to extend my sincere thanks to Assoc. Prof. Alkan Kabakçioğlu for his valuable contributions and continuous guidance through our regular meetings and discussions. His support has played a significant role in shaping the direction and development of my work, ultimately culminating in this dissertation. I am equally grateful to Koç University for providing access to their computing cluster, as well as to the staff involved for their technical support and assistance, which were essential to carrying out this research.

I am also grateful to Prof. Ali Taylan Cemgil for the insightful guidance and support provided during the earlier phases of my work.

I would also like to express my sincere thanks to VAVLAB members, Demet, Göktekin and Gurur for their generous support and contributions throughout my work. Their readiness to help and insightful input, especially during critical phases of both data processing and prototyping, were truly appreciated.

I would also like to express my gratitude to our neurologist collaborators for sharing their expertise and getting into insightful discussions, particularly Prof. Dr. Hakan Gürvit and the late Dr. Zerrin Yıldırım, Dr. Emre Harı, and Dr. Ulaş Ay, for generously sharing their clinical expertise. Dr. Yıldırım's invaluable contributions and dedication will always be fondly remembered.

This thesis was partially supported by the Turkish Directorate of Strategy and Budget under the TAM Project (Project No. 2007K12-873), the TÜBİTAK-ARDEB 1003 Program (Grant No. 114E053), and the Boğaziçi University Research Fund (Grant No. 16862).

I would also like to express my sincere gratitude to Prof. Cem Ersoy and the TETAM community for their valuable insights, constructive suggestions, and engaging discussions throughout the course of this research.

I must also mention that without the support of my parents, Arslan and Vahide Bayır, I wouldn't be able to find the strength and stamina to finish my work. They always encouraged me to succeed. My heartfelt thanks to you both. I love you.

Last but certainly not least, I would like to express my heartfelt thanks to my beloved wife, Gülşen, for her unwavering love, patience, and emotional support throughout this journey. Her understanding and strength made it possible for me to stay focused even during the most challenging times. I am also deeply grateful to my in-laws, Nuray and Yüksel Bal, for their kindness, encouragement, and generous support along the way. And of course, my dear children, Kuzey and Ada — thank you for filling my life with joy and laughter, even when I wasn't always there to enjoy it. I truly look forward to making up for the time we've missed together. I love you all.

ABSTRACT

LOW DIMENSIONAL BRAIN CONNECTOME REPRESENTATIONS FOR ALZHEIMER’S DISEASE DEMENTIA

Alzheimer’s Disease Dementia (ADD) progresses along a continuum, typically beginning with Subjective Cognitive Impairment (SCI), advancing through Mild Cognitive Impairment (MCI), and culminating in dementia. Capturing this trajectory in a compact and interpretable form remains a key challenge in brain network analysis. This thesis introduces a Graph Neural Network (GNN)-based framework that embeds structural brain connectomes into a two-dimensional manifold, offering a low-dimensional yet informative representation of disease staging.

The architecture integrates a Graph Isomorphism Network (GIN) and a Graph Attention Network (GAT), followed by attention-based pooling to generate graph-level embeddings. These are projected using Principal Component Analysis (PCA) to construct a two-dimensional staging space. The learned manifold reveals a coherent spatial organization of ADD, MCI, and SCI populations, with MCI cases generally located between the two clinical extremes—consistent with a gradual model of cognitive decline. The model is validated through a detailed ablation study and random analysis, supporting the reliability of the learned embeddings. Additionally, key anatomical regions—such as the precuneus, angular gyrus, and superior parietal lobule—are identified as dominant contributors to variance across the manifold, in line with established biomarkers of ADD. Together, the proposed framework offers a robust and interpretable approach for visualizing neurodegenerative progression and supports biologically grounded models of disease staging.

ÖZET

ALZHEIMER HASTALIĞI DEMANSI İÇİN DÜŞÜK BOYUTLU BEYİN KONNEKTOM TEMSİLLERİ

Alzheimer Hastalığı Demansı (AHD), genellikle Subjektif Bilişsel Bozulma (SBB) ile başlayıp Hafif Bilişsel Bozulma (HBB) evresinden geçerek demansla sonuçlanan sürekli bir ilerleme süreci izler. Bu ilerleyişin kompakt ve yorumlanabilir bir biçimde temsil edilmesi, beyin ağı analizinde önemli bir zorluk oluşturmaktadır. Bu tezde, yapısal beyin bağlantılarını iki boyutlu bir manifold üzerine gömen Çizge Sinir Ağı (GNN) tabanlı bir çerçeve sunulmaktadır. Bu yapı, hastalık evrelemesini düşük boyutlu ancak bilgi açısından zengin bir biçimde temsil etmeyi amaçlamaktadır.

Geliştirilen mimari, Çizge Eşyapılılığı Ağı (GIN) ve Çizge Dikkat Ağı (GAT) katmanlarını entegre eder; ardından, dikkat temelli havuzlama yöntemiyle çizge düzeyinde gömme vektörleri oluşturulur. Bu vektörler, Başlıca Bileşenler Analizi (PCA) kullanılarak iki boyutlu bir evreleme uzayına yansıtılır. Öğrenilen manifold, AHD, HBB ve SBB gruplarının uzayda tutarlı bir şekilde konumlandığını göstermekte; HBB örneklerinin genellikle iki klinik uç arasında yer alması, bilişsel gerilemenin kademeli doğasıyla örtüşmektedir. Model, ayrıntılı bir ablation (bileşen çıkarım) çalışması ve rastgelelik analizleri ile doğrulanmış olup, öğrenilen temsillerin güvenilirliğini desteklemektedir. Ayrıca, precuneus, açısal girus ve üst parietal lob gibi anatomik bölgeler, manifold boyunca gözlenen varyansın başlıca belirleyicileri olarak tanımlanmış ve bu bulgular, AHD'nin bilinen biyobelirteçleriyle uyumluluk göstermiştir.

Sonuç olarak, önerilen bu çerçeve, nörodejeneratif ilerleyişin görselleştirilmesine yönelik sağlam ve yorumlanabilir bir yaklaşım sunmakta; hastalık evrelemesine dair biyolojik olarak temellendirilmiş modelleri desteklemektedir.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	iii
ABSTRACT	v
ÖZET	vi
LIST OF FIGURES	ix
LIST OF TABLES	xi
LIST OF SYMBOLS	xii
LIST OF ACRONYMS/ABBREVIATIONS	xiii
1. INTRODUCTION	1
2. LITERATURE REVIEW	5
3. METHODOLOGY	17
3.1. Brain Networks	17
3.2. Graph Neural Networks	18
3.2.1. Graph Isomorphism Network	19
3.2.2. Graph Convolutional Network	20
3.2.3. Graph Attention Network	21
3.2.4. Summary of Different Graph Layers	22
3.2.5. Graph Readouts	22
3.2.5.1. Simple Readout Mechanisms	23
3.2.5.2. Global Attention Readout	24
3.2.5.3. Graph Attention Readout (GARO)	24
3.3. Attention Based Low Dimensional Brain Connectome Embedding	25
3.3.1. Architectural Modifications Explored	27
3.3.1.1. Subnet GAT	28
3.3.1.2. Subnet GAT with Subnet Attention	29
3.3.1.3. Subnet GARO	30
4. EXPERIMENTS AND RESULTS	31
4.1. Data Acquisition	31
4.2. Clinical Assessment	32

4.3. Brain Connectome Construction	33
4.4. Low Dimensional Disease Manifold	35
4.5. Classification Performance and Predictive Evaluation	46
4.6. Ablation Study	48
4.7. Random Experiments	52
4.8. Architecture Search	56
4.9. Multimodal and Unimodal Experiments With Different Networks	58
5. DISCUSSION	63
6. CONCLUSION	69
REFERENCES	71

LIST OF FIGURES

Figure 3.1.	Architecture of the Low dimensional manifold constructor	27
Figure 3.2.	GNN Block after Subnet GAT modification	29
Figure 3.3.	GNN Block after Subnet GAT and Subnet Attention modification	30
Figure 4.1.	Low dimensional manifold constructed by ADD vs. SCI classifica- tion task	37
Figure 4.2.	Low dimensional manifold constructed by ADD vs. MCI classifica- tion task	38
Figure 4.3.	Low dimensional manifold constructed by MCI vs. SCI classifica- tion task	40
Figure 4.4.	Gaussian fits to the SCI, MCI and ADD clusters represented by the 2D manifold	41
Figure 4.5.	Overlay of cluster regions onto the 2D manifold	42
Figure 4.6.	Cortical surface visualizations of SCI and ADD groups	44
Figure 4.7.	Cortical surface visualizations of SCI and MCI groups	45
Figure 4.8.	Likelihood decrease over the 2D manifold when moving from SCI to ADD	46

Figure 4.9. Random samples projected onto the low-dimensional manifold constructed by ADD vs. SCI classification task	53
Figure 4.10. Training and projecting with all random samples	55
Figure 4.11. Siamese-style GNN	61



LIST OF TABLES

Table 3.1.	Comparison of GNN Variants by Aggregation, Update, and Matrix Form Equations.	22
Table 3.2.	Architecture parameters of the Attention-Based Low Dimensional Brain Connectome Embedding.	26
Table 4.1.	Summary of the cohort of participants.	31
Table 4.2.	Common MRI Acquisition Parameters by Modality.	31
Table 4.3.	Tractography Parameters.	34
Table 4.4.	Summary of Training Parameters.	36
Table 4.5.	Predictive performance of the low-dimensional manifold constructor framework.	47
Table 4.6.	Summary of Ablation Study Parameters.	50
Table 4.7.	ROC-AUC performance of different model variants (mean $\pm$ standard deviation for ADD/SCI classification task.	57
Table 4.8.	ROC-AUC scores for three binary classification tasks (ADD/SCI, ADD/MCI, MCI/SCI) across different network types and sizes. . .	59
Table 4.9.	Ablation Study.	62

LIST OF SYMBOLS

a_{ij}	Connection weight between vertices/nodes i and j
$\mathbf{A}$	Adjacency matrix
D	Nodal feature dimension
$\mathbf{D}$	Degree matrix
e_{ij}	Edge connecting vertices/nodes i and j
E	Edge Set
F	Number of features
f_k	End-point of fiber k
G	Graph
$\mathbf{K}$	Key matrix
N	Number of nodes/vertices
$\mathcal{N}(v)$	Neighbor set of vertex/node v
$\mathbf{q}$	Query matrix
R_i	Cortical region i
r_i	A point in cortical region i
v_i	Vertex number i
V	Vertex/node Set
V_i	Volume of cortical region i
$\mathbf{W}$	Feature transformation matrix
W_{ij}	Measure of association between i -th region and k -th fiber
$\mathbf{x}_v$	Nodal feature vector of vertex/node v
$\mathbf{X}$	Nodal feature matrix
$\alpha_{u,v}$	Attention coefficient between vertices u and v
$\boldsymbol{\alpha}$	Attention matrix
μ	Mean
σ	Standard deviation

LIST OF ACRONYMS/ABBREVIATIONS

AAL	Automated Anatomical Labeling
AD	Alzheimer's Disease
ADD	Alzheimer's Disease Dementia
ADHD	Attention Deficit Hyperactivity Disorder
ADNI	Alzheimer's Disease Neuroimaging Initiative
AHD	Alzheimer Hastalığı Demansı
ASD	Autism Spectrum Disorder
AUC	Area Under Curve
BD	Bipolar Disorder
BOLD	Blood-Oxygen-Level Dependent
CDR	Clinical Dementia Rating
CFI-S	Cognitive Functions Instrument – Subject form
CFI-SP	Cognitive Functions Instrument - Study Partner form
CI	Cue Index
CNN	Convolutional Neural Network
DAN	Dorsal Attention Network
DLB	Dementia with Lewy Bodies
DMN	Default Mode Network
DWI	Diffusion-Weighted Imaging
EMCI	Early Mild Cognitive Impairment
eNet	Elastic Networks
EPI	Echo Planar Imaging
FCSRT	Free and Cued Selective Reminding Test
FFE	Fast Field Echo
fMRI	Functional Magnetic Resonance Imaging
fMRIB	Functional Magnetic Resonance Imaging of the Brain
fNet	Functional Networks
FOV	Field of View

FSL	FMRI Software Library
FTD	Frontotemporal Dementia
GARO	Graph Attention Readout
GAT	Graph Attention Network
GCN	Graph Convolutional Network
GCRNN	Graph Convolutional Recurrent Neural Network
GIN	Graph Isomorphism Network
GNN	Graph Neural Network
HBB	Hafif Bilişsel Bozulma
HC	Healthy Control
HCP	Human Connectome Project
HIV	Human Immunodeficiency Viruses
k-NN	k-Nearest Neighbors
<i>l</i> Net	Learned Functional Network
LOOCV	Leave-one-out Cross Validation
LSTM	Long Short-Term Memory
MCI	Mild Cognitive Impairment
MDD	Major Depressive Disorder
MLP	Multi-layer Perceptron
MRI	Magnetic Resonance Imaging
NC	Normal Control
NIA-AA	National Institute on Aging and Alzheimer's Association
PC	Principal Component
PC1	First Principal Component
PD	Parkinson's Disease
PDD	Parkinson's Disease Dementia
PCA	Principal Component Analysis
PET	Positron Emission Tomography
RK4	Fourth-order Runge-Kutta
RNN	Recurrent Neural Network
ROC	Receiver Operating Characteristic

ROI	Region of Interest
rs-fMRI	Resting State Functional Magnetic Resonance Imaging
SBB	Subjektif Bilgişel Bozulma
SCI	Subjective Cognitive Impairment
sNet	Structural Networks
SOB	Sum of boxes
T1w	T1-weighted
TE	Echo Time
TFE	Turbo Field Echo
TFR	Total Free Recall
TR	Repetition Time
UAT	Universal Approximation Theorem
VGAE	Variational Graph Auto-encoder
WL	Weisfeiler-Lehman

1. INTRODUCTION

Dementia is a broad clinical syndrome characterized by progressive deterioration in cognitive functioning, including memory, language, planning, judgment, and decision-making. It is often accompanied by behavioral symptoms such as mood disturbances, confusion, and personality changes. Rather than representing a single disease, dementia serves as an umbrella term encompassing various underlying etiologies. These include, but are not limited to, Alzheimer’s Disease Dementia (ADD), Parkinson’s Disease Dementia (PDD), Frontotemporal Dementia (FTD), Dementia with Lewy Bodies (DLB), Vascular Dementia, and Huntington’s Disease-related Dementia.

Among these, Alzheimer’s Disease Dementia (ADD) is the most prevalent form, accounting for approximately 60–80% of dementia cases globally. As of 2019, the number of people living with dementia was estimated to exceed 55 million, and this figure is projected to rise substantially in the coming decades due to aging populations worldwide [1,2]. The societal burden of ADD is particularly severe in low- and middle-income countries, where healthcare infrastructure and research funding remain inadequate to meet the growing clinical demand. Understanding and addressing this burden is thus both a scientific and public health imperative.

Although the precise pathophysiological mechanisms underlying ADD are not yet fully understood, several hypotheses have been proposed. One of the most widely accepted theories involves the accumulation of misfolded proteins in the brain, most notably tau protein tangles and extracellular amyloid-beta ($A\beta$) plaques. These molecular alterations are associated with a cascade of structural and functional changes in neural tissue that ultimately impair cognitive function [3].

The progression of ADD can be conceptualized as a continuum comprising multiple stages: cognitively normal or healthy controls (often referred to as NC or HC), Subjective Cognitive Impairment (SCI), Mild Cognitive Impairment (MCI), and fi-

nally, overt Alzheimer’s Disease Dementia (ADD). While healthy individuals show no cognitive complaints or clinical signs, individuals at the SCI stage report subjective memory difficulties despite the absence of measurable deficits on neuropsychological tests. MCI is characterized by more pronounced cognitive deficits that are observable through clinical assessments, although daily functional abilities are largely preserved. ADD, in contrast, represents the terminal stage, marked by profound cognitive and functional impairments that significantly disrupt everyday life.

Given the complex and progressive nature of ADD, neuroimaging modalities have emerged as indispensable, non-invasive tools for monitoring brain changes across the disease spectrum. Among these modalities, magnetic resonance imaging (MRI) and positron emission tomography (PET) are widely used. While PET imaging provides valuable molecular and metabolic insights, its capacity to delineate structural transformations remains limited [4]. Conversely, MRI offers high-resolution anatomical information and has gained considerable traction for its ability to detect structural and functional abnormalities associated with neurodegeneration.

One of the most powerful concepts derived from MRI data is that of the brain connectome—a graph-based representation of the brain in which nodes correspond to anatomical regions and edges denote connections between them. These networks can be derived from either structural MRI (e.g., diffusion MRI) or functional MRI (e.g., resting-state fMRI). Although microscale connectomes at the neuronal level offer fine-grained detail, they are not currently feasible at scale due to technical and computational limitations. Therefore, most current research utilizes macroscale connectomes constructed from standardized brain parcellation schemes, also known as anatomical atlases. Notable atlases used in the literature include the Destrieux atlas [5], the Schaefer atlas [6], and the Automated Anatomical Labeling (AAL) atlas [7], which segment the brain into functionally or anatomically meaningful regions suitable for graph construction.

In recent years, Graph Neural Networks (GNNs) have emerged as a powerful class of deep learning models capable of operating directly on graph-structured data. Their ability to model relational dependencies and topological hierarchies has made them particularly well-suited for connectome-based analysis. Motivated by these capabilities, the present thesis investigates how GNN-based representations can be leveraged to characterize the spectrum of cognitive decline in ADD. Specifically, this work proposes a novel framework that integrates connectome data into a unified GNN-based pipeline, with the goal of learning low-dimensional embeddings that preserve disease-relevant information to help characterize the disease spectrum.

The central aim of this research is to identify and model the latent manifold underlying the ADD disease trajectory using non-invasive imaging data. By capturing the continuum of cognitive deterioration in a compact and interpretable form, the proposed framework not only enhances diagnostic modeling but also opens new avenues for understanding disease staging, individualized prognosis, and biomarker discovery.

The remainder of this thesis is structured as follows.

- Chapter 2 presents a comprehensive review of the literature on brain connectomes, traditional graph-theoretical approaches, and recent developments in applying Graph Neural Networks (GNNs) to neuroimaging data, with a particular focus on Alzheimer’s Disease Dementia (ADD) and some other neurological conditions.
- Chapter 3 details the proposed methodology, including preprocessing of neuroimaging data, construction of structural and functional brain networks, design of the GNN architecture, and the techniques used to construct and analyze low-dimensional cognitive decline manifolds.
- Chapter 4 describes the experimental setup and presents key results, including classification performance, manifold visualizations, statistical significance analyses, and neurobiological interpretations of the learned representations.
- Chapter 5 discusses the broader implications of the findings, critically evaluates the limitations of the current study, and connects the results to ongoing challenges

and debates in the field of computational neuroscience.

- Chapter 6 concludes the thesis by summarizing the main contributions, emphasizing the novelty and applicability of the proposed framework, and outlining promising directions for future research and potential clinical translation.



2. LITERATURE REVIEW

The term “brain connectome” refers to a comprehensive graph-based representation of the brain, capturing its intricate network of neural connections. Since its conceptual introduction, the connectome has become a foundational framework for understanding the structural and functional organization of the brain [8]. Connectomes can be studied at different scales, most notably at the microscale and macroscale levels. The microscale connectome aims to map the complete network of individual neurons and synapses, which is an immense challenge—considering that the human cerebral cortex contains approximately 9×10^{10} neurons connected by nearly 10^{14} synapses [9]. Such scale poses considerable barriers, including limitations in data availability, imaging resolution, annotation capabilities, graph analysis algorithms, and computational infrastructure. Addressing these challenges has become an active area of research, focusing on the development of advanced data acquisition techniques [10] and the creation of scalable, high-throughput frameworks for neurological data analysis [11].

Given these difficulties, macroscale connectomes—which abstract the brain into larger anatomical or functional regions—have become more practical and widely studied. One of the most influential initiatives in this area is the Human Connectome Project (HCP), which seeks to construct comprehensive structural and functional maps of the human brain [12]. Macroscale connectome construction typically relies on brain parcellations, or brain atlases, which define standardized regions of interest (ROIs). A variety of atlases have been proposed, ranging from early anatomical atlases such as the Brodmann Atlas [13] to more recent and computationally grounded alternatives including the Automated Anatomical Labeling (AAL) atlas [7], Destrieux Atlas [5], Schaefer Atlas [6], and the Brainnetome Atlas [14]. These atlases vary in their parcellation criteria—such as anatomical landmarks, functional similarity, or connectivity patterns—but they all share the common objective of creating a coherent macro-level segmentation of the brain to facilitate downstream analysis.

Graph-theoretical approaches are widely used to analyze brain connectomes, offering quantitative insights into their topological organization. These methods compute network metrics at both global and local (nodal) scales. Global measures—such as average path length, assortativity, clustering coefficient, transitivity, global efficiency, and small-worldness—reflect integrative and segregative properties of the network. Local measures—including degree statistics, local clustering, assortativity, betweenness centrality, and local efficiency—capture node-level importance and connectivity. Extensive evidence shows that the brain exhibits small-world organization and hub-centric topology, consistently observed across imaging modalities and frequency bands [15, 16]. These properties have motivated the application of classical graph processing and network analysis tools in modern neuroscience.

These methods have been applied to study brain disorders, revealing associations between altered network connectivity and conditions such as schizophrenia and Alzheimer’s disease dementia (ADD) [17]. In the case of ADD, a substantial body of research has examined topological abnormalities as potential biomarkers for diagnosis and disease monitoring. Studies using both structural and functional neuroimaging consistently report changes in global efficiency, clustering, hub structure, and modular organization [18–21].

However, graph-theoretical approaches present several limitations. As classical, primarily descriptive methods, they tend to be inherently static and lack adaptability to task-specific objectives. Feature extraction in these approaches typically requires manual engineering, and their expressiveness is constrained by the selection and interpretation of predefined metrics. Consequently, compared to more flexible, data-driven methods such as graph neural networks (GNNs), they demand a significant degree of domain expertise and handcrafted design. This reliance on fixed descriptors makes them less modular and less scalable, particularly when applied to complex or high-dimensional neuroimaging data.

GNNs, on the other hand, have been chosen as an important tool to apply in the investigation of neurological disorders. The studies mainly differ in how they construct the graphs to be processed by the GNN layers. These include population graphs, cohort based common graphs and subject graphs.

In population graph approaches, a single large graph is constructed to represent an entire cohort, where each node corresponds to a subject rather than a brain region. The brain data—structural, functional, or a combination thereof—is embedded into the nodal feature vectors, while edges are typically defined using phenotypic or demographic similarity. Early studies applied Graph Convolutional Networks (GCNs) for disease classification tasks, including Alzheimer’s disease dementia (ADD) and Autism Spectrum Disorder (ASD) [22,23], with later extensions to Parkinson’s Disease (PD) [24]. While population graphs introduced a compelling framework for group-level inference, several limitations have emerged. These include sensitivity to population structure and graph construction quality, as well as scalability issues—wherein the addition of a new subject requires modification of the entire graph and recomputation of graph-level metrics, such as the Laplacian. Moreover, compressing individual brain connectomes into vector representations may lead to a loss of subject-specific topological information, limiting the capacity to capture nuanced neurobiological variations.

As an alternative to population graphs, cohort-based common graph approaches have been proposed to address several of their limitations. In this framework, a shared graph structure is constructed using the brain connectomes of all subjects, thereby incorporating brain network topology directly, rather than summarizing it into feature vectors as in population graphs. Studies such as [25] and [26] implemented k-nearest neighbor (k-NN) based graph construction and applied a fixed graph across all subjects, varying only the nodal features. While this method improves biological interpretability and avoids some scalability issues, it remains suboptimal. The use of a common graph across subjects fails to capture individual-level topological variations, which are especially critical for identifying early-stage or subject-specific disease patterns. This constraint limits the expressive power of the model and may reduce sensitivity to novel

or subtle changes in brain network organization during disease progression.

Subject-graph-based methods have emerged as the latest and arguably most prominent alternative to earlier approaches such as population-level and cohort-based common graphs. In these methods, a separate graph is constructed for each individual subject, typically derived directly from their structural or functional brain connectome. This individualized approach offers several key advantages. First, it inherently captures subject-specific topological characteristics, allowing models to learn fine-grained patterns that may be critical for tasks such as early disease detection or personalized diagnostics. Second, subject-level graphs alleviate scalability concerns, as each subject graph can be processed independently—there is no need to modify or reconstruct a global graph structure when new data is introduced, as required in population graphs. This makes the method more modular and compatible with modern machine learning pipelines.

Importantly, subject-level graph construction enables direct analysis of the topology of each subject’s brain network, facilitating deeper neurobiological interpretations and more precise localization of pathological changes. This is especially relevant in the study of progressive neurological conditions such as Alzheimer’s disease dementia (ADD), where subtle and early-stage network alterations may vary across individuals and are critical to detect.

However, despite these advantages, the current body of literature remains dominated by studies utilizing functional connectomes, particularly resting-state fMRI data. While functional networks offer valuable insights into brain dynamics, relatively few studies have explored the use of nectomes (e.g., derived from diffusion MRI) or multi-modal graphs that integrate both structural and functional modalities. This underrepresentation is notable not only in general GNN applications but also in ADD-focused research, where structural degradation plays a key role in disease progression. Incorporating structural or fused structural-functional graphs in subject-level GNN models represents an important and underexplored avenue for future research, potentially im-

proving both classification performance and neurobiological interpretability.

The use of subject-level brain connectomes has proven to be a powerful approach for investigating various neurological and psychiatric disorders. Functional and structural connectomes, in particular, serve as rich sources of information that can be leveraged through machine learning techniques to facilitate clinical diagnosis. A growing body of literature has explored the potential of graph-based and deep learning models to classify disorders such as Autism Spectrum Disorder (ASD), Attention Deficit Hyperactivity Disorder (ADHD), Parkinson’s Disease (PD), Schizophrenia, Bipolar Disorder (BD), Major Depressive Disorder (MDD), HIV-associated disorders, and Alzheimer’s Disease (AD).

Zhang et al. [27] utilized contrastive learning applied to functional connectomes to distinguish between normal controls and individuals with ASD and ADHD, highlighting the power of self-supervised learning in this domain. In a similar vein, Wen et al. [28] employed an auto-encoder-based self-supervised method to enhance functional connectome representations, significantly improving classification performance for ASD, BD, and MDD. Liu et al. [29] proposed a hierarchical architecture that performs pooling to create coarse representations from fine-grained functional connectivity features, achieving promising results in both ASD and ADHD classification.

Li et al. [30] integrated Graph Convolutional Networks (GCNs) with a Transformer architecture to address ASD diagnosis using functional connectomes. In parallel, Zhao et al. [31] proposed a novel convolution mechanism for ADHD classification that aggregates nodal features based on Euclidean distance, further enriching the model’s representational capacity. Huang et al. [32] took a multimodal approach by utilizing both structural and functional brain networks and devised a diffusion model capable of processing higher-order neighbor information, applied to schizophrenia and epilepsy diagnosis. Xia et al. [33] also utilized multimodal brain graphs, combining structural and functional connectivity data as well as nodal features from both, to classify Parkinson’s Disease.

In one of the few studies using only structural networks, Cui et al. [34] developed a masked graph learning strategy to highlight critical edges in structural connectomes, successfully applied to classification tasks involving PD, HIV, and BD. Liu et al. [35] proposed a hybrid model incorporating Long Short-Term Memory (LSTM) networks, spatial graph convolutions, and segments of the BOLD signal for functional networks to classify ADHD, MDD, and BD. Zheng et al. [36] addressed multiple classification tasks—including ASD, MDD, and schizophrenia—by leveraging Variational Graph Auto-Encoders (VGAEs) on thresholded functional connectomes. Similarly, Yu et al. [37] applied Graph Attention Networks (GATs) on higher-order correlation-based functional networks to address schizophrenia classification.

In the context of Alzheimer’s Disease and Mild Cognitive Impairment (MCI), several studies have adopted multi-modal and temporal strategies. Zhang et al. [38] proposed a Graph Convolutional Recurrent Neural Network (GCRNN) that integrates structural networks (sNets) as the underlying graph and functional networks (fNets) as nodal features to differentiate MCI from Normal Control (NC) subjects. Xing et al. [39] classified AD patients and NCs using a GCN-LSTM hybrid model applied to the ADNI-II dataset, while also extending their method to demographic attribute prediction such as age and sex. Chen et al. [40] introduced entropy and correlation-based efficient networks, which were processed using spatio-temporal convolutions and Transformers for MCI diagnosis.

Ma et al. [41] fused static and dynamic functional connectivity matrices to create a comprehensive representation, which was subsequently used to classify ADD and ASD. Park et al. [42] proposed an innovative method utilizing Hodge Laplacians to extract topological insights from structural networks, thereby enhancing ADD classification. Xu et al. [43] addressed multiple disorders—ADD, ASD, and PD—by proposing a novel differentiable pooling strategy called contrastive pooling, which not only facilitated more accurate classification but also enhanced interpretability by emphasizing inter-cohort distinctions. Finally, Li et al. [44] constructed functional connectivity matrices via Kalman filtering applied to BOLD signals and combined these with graph attention

mechanisms to classify different stages of MCI, including early and late stages relative to normal controls.

Subject-level brain connectomes have emerged as critical representations in predictive modeling tasks across a variety of other neurocognitive domains. These include the prediction of cognitive scores, clinical outcomes, neuropsychological indices, behavioral phenotypes, and demographic attributes such as age and sex. Functional and structural connectivity profiles—often extracted from large-scale neuroimaging datasets such as the Human Connectome Project (HCP)—have been increasingly utilized in conjunction with graph neural networks, attention mechanisms, and topological techniques to improve the accuracy and interpretability of these predictive tasks.

Jia et al. [45] proposed a hierarchical spatial graph convolutional network for predicting cognitive performance from resting-state fMRI data obtained from the HCP dataset. Their model constructs functional connectomes using a k-nearest neighbor (k-NN) pruning strategy, which facilitates the manifestation of small-world properties in the connectivity graph. This structural constraint enhances the graph’s biological plausibility and supports more stable learning.

In another study, Qu et al. [46] addressed cognitive score prediction by fusing structural and functional connectomes, with a particular focus on the degree of disentanglement between the two modalities. Their approach aims to isolate modality-specific contributions while preserving shared information, thus improving the generalization capacity of the learned features across cognitive domains.

Huang et al. [47] introduced a novel topological framework based on the Hodge Laplacian to capture differential connectivity patterns in functional networks. Their model was specifically tailored for predicting individual intelligence scores, exploiting cycle-based features that go beyond pairwise correlations. A follow-up study by Huang et al. [48] extended this approach by incorporating cycle convolutions and edge positional encodings, further refining the model’s ability to capture higher-order topological

information relevant to intelligence prediction.

Kim et al. [49] combined both resting-state and task-based fMRI data from the HCP to predict sex and cognitive task labels. Their approach leveraged spatio-temporal attention mechanisms alongside Graph Isomorphism Networks (GINs) to extract expressive node and graph-level features from functional connectomes. These mechanisms enhanced the model’s sensitivity to dynamic patterns in neural activity across time and space.

Azevedo et al. [50] proposed a hybrid architecture that integrates graph convolution operations on the edges of the functional connectivity graph and temporal convolutions on BOLD signal sequences. This model enhances the temporal dynamics of functional networks and was shown to perform well on binary sex classification tasks using HCP datasets. The authors applied thresholding to sparsify the functional networks, thereby reducing noise and improving the stability of the learned graph representations.

In a related direction, Sivgin et al. [51] utilized thresholded functional connectomes to predict both age and biological sex. Their primary contribution was the development of a model that learns lagged and dynamic representations of brain activity, thereby improving temporal sensitivity in prediction tasks. By modeling delayed dependencies in BOLD signals, the approach was able to capture more nuanced temporal patterns associated with developmental and demographic variation.

As discussed, a substantial body of research within the fields of neuroimaging and computational neuroscience focuses on classification and regression tasks. This methodological focus is also evident in studies related to ADD. Numerous investigations have been devoted to distinguishing individuals with ADD from cognitively normal controls, or from SCI and from those diagnosed with MCI. A subset of studies further attempts to classify sub-stages of MCI—such as early versus late MCI—or perform multiclass classification across normal cognition, MCI, and ADD.

Nevertheless, despite the proliferation of such diagnostic models, the majority of existing research remains constrained to discrete diagnostic categories and overlooks the underlying continuum of cognitive decline. The progression from normal cognition to dementia in Alzheimer’s disease is inherently gradual and heterogeneous, yet this continuous spectrum is rarely modeled explicitly. Current approaches often emphasize static classification boundaries without adequately capturing the transitional dynamics or neurobiological variations across the disease trajectory. As a result, while classification frameworks may achieve high diagnostic accuracy, they provide limited insight into the latent dimensions of disease progression—leaving the rich and clinically relevant spectrum of cognitive deterioration in ADD substantially underexplored.

Moreover, the effective characterization of this spectrum through the use of intrinsic, low-dimensional latent representations has received even less scholarly attention. While graph embedding techniques have been widely applied in classification and regression contexts, they have not yet been extensively utilized to map or represent the broader cognitive decline continuum. Several studies have employed low-dimensional embeddings of magnetic resonance imaging (MRI) data for the purpose of classification [52], discriminative analysis of patients diagnosed with ADD [53], and for the investigation of structural-functional brain associations [54]. One particularly noteworthy contribution [55] proposes a discrepancy analysis framework not only aimed at enhancing the diagnostic accuracy for ADD and Mild Cognitive Impairment (MCI), but also at capturing translational patterns along the cognitive deterioration continuum. However, such integrative and spectrum-oriented approaches remain scarce.

Despite these advancements, the majority of existing studies either fail to provide a rigorous and continuous exploration of disease states or lack sufficient granularity in modeling the dynamic progression of cognitive impairment. As a result, the cognitive decline spectrum—especially in the context of ADD—remains a significantly understudied frontier in computational neuroscience and clinical neuroimaging research.

Building upon these findings, a growing number of studies have expanded the use of connectome-based graph representations beyond clinical classification, leveraging them for a variety of continuous prediction tasks including cognitive performance, behavioral phenotypes, and demographic variables. This shift highlights the generalizability of brain graphs as modeling substrates in both clinical and normative neuroscience.

These studies collectively illustrate the expanding scope of connectome-based modeling and graph neural networks (GNNs) in neuroscience, highlighting that brain connectivity patterns—whether derived from functional or structural imaging—have been utilized not only for the classification of neurological and psychiatric disorders, but also for a wide range of other predictive tasks related to cognition, behavior, and demographic profiling. This breadth of application underscores the generalizability and flexibility of connectome representations as a modeling framework. While functional connectomes have been widely used due to their accessibility and ability to capture dynamic co-activation patterns, structural connectomes, derived from diffusion MRI, offer a more direct and biologically grounded representation of the brain’s anatomical architecture. Their relative sparsity and neuroanatomical interpretability make them particularly well-suited for graph-based learning. Recent studies have demonstrated that connectome-derived features, when processed through GNNs or related architectures, can successfully capture individual variability in cognitive performance, intelligence, personality traits, sex, and age. These findings collectively position graph-based connectome modeling as a promising and generalizable tool for both clinical and non-clinical neuroimaging applications.

While functional brain networks are generally easier to obtain than structural networks—primarily because they do not require computationally intensive tractography or diffusion-weighted imaging pipelines—they present their own unique set of challenges. Functional networks are typically derived from Blood Oxygen-Level Dependent (BOLD) signals, which reflect changes in neural activity by measuring fluctuations in oxygenated blood flow. The construction of these networks is not direct

but rather inferred through statistical associations between time-series signals recorded from different brain regions.

Multiple statistical methods have been proposed to infer functional connectivity, including Pearson correlation, Spearman correlation, partial correlation, and mutual information, among others. However, these methods often produce connectivity matrices that are overly dense, capturing both meaningful and spurious associations. The high density of these networks complicates both their interpretability and their utility in machine learning and graph-theoretic analyses. As a result, researchers frequently apply sparsification techniques to emphasize the most salient connections. These techniques range from simple thresholding and binarization to more sophisticated strategies such as minimum spanning trees, k-nearest neighbors filtering, and adaptive thresholding based on statistical significance, [56–58].

However, the introduction of sparsification steps adds a layer of complexity to the network construction process. The choice of sparsification technique and its parameters can significantly influence the topological properties of the resulting networks, and by extension, the outcomes of downstream analysis tasks such as classification, clustering, or biomarker discovery. This variability introduces a form of methodological noise and can dilute the consistency and reproducibility of findings across studies, posing a significant challenge for the broader generalization of machine learning models trained on such data.

In contrast, structural brain networks, typically derived from diffusion MRI and tractography, are inherently sparse due to the physical constraints of axonal connections in the human brain. The white matter tracts that constitute structural connectivity impose anatomical limitations on inter-regional communication, resulting in sparser and often more stable connectivity matrices. This sparsity not only reflects underlying neurobiological reality but also reduces the need for aggressive post-hoc sparsification, thereby simplifying the preprocessing pipeline and potentially enhancing the interpretability of derived graph representations.

This thesis aims to address a significant and persistent gap in the current understanding of the cognitive decline spectrum associated with Alzheimer’s Disease Dementia (ADD), particularly in relation to its complex and progressive interactions with brain network topology. Despite extensive research on diagnostic classification, the continuum of cognitive deterioration—its underlying geometry, structural variability, and neurobiological correlates—remains poorly characterized in the literature.

To this end, the thesis introduces a novel deep learning framework grounded in graph neural networks (GNNs) with integrated attention mechanisms. This architecture is specifically designed to capture subject-level brain connectivity patterns and model the continuous trajectory of cognitive decline, moving beyond discrete diagnostic labels. The proposed method identifies and exploits a low-dimensional latent manifold, revealing that the high-dimensional neuroimaging data associated with Alzheimer’s progression can be meaningfully and reliably embedded in a two-dimensional space. This manifold captures subtle yet systematic variations in structural brain connectivity, suggesting that complex topological changes occur in a stable and interpretable fashion across the cognitive decline continuum.

The findings contribute not only to methodological innovation but also to clinical neuroscience by offering a novel framework for both disease staging and early diagnosis. The proposed approach facilitates enhanced interpretability and may provide valuable insights into the progression of ADD at a population and individual level. In doing so, this work lays the foundation for future applications in personalized medicine, biomarker discovery, and mechanistic modeling of neurodegenerative disease trajectories.

3. METHODOLOGY

3.1. Brain Networks

A weighted, undirected graph with no self-loops can be defined by the following components:

- A set of nodes/vertices, where $V = \{v_i\}$, $|V| = N$,
- A set of edges, where $E = \{e_{ij}\}$, and
- A corresponding adjacency matrix, $\mathbf{A} \in \mathbb{R}^{N \times N}$, with elements a_{ij}

The adjacency matrix satisfies

$$a_{ji} = a_{ij}, \quad a_{ii} = 0, \quad a_{ij} = 0 \quad \text{if } e_{ij} \notin E. \quad (3.1)$$

This implies that the adjacency matrix is symmetric and contains no self-loops (i.e., self-connections). Moreover, it fully encodes the information represented by the edge set E .

Each node/vertex v_i may have an associated feature vector x_i , and all node features form the feature matrix, $\mathbf{X} \in \mathbb{R}^{N \times F}$, where F is the number of features per node/vertex.

The complete graph with structural and feature information is therefore given by

$$G = (V, E, \mathbf{A}, \mathbf{X}). \quad (3.2)$$

In the context of brain networks, such a graph structure is used to represent neural systems, where:

- Nodes correspond to cortical regions
- Edges represent connections between regions based on functional or structural relationships

Both nodes and edges may carry features or embeddings that describe their characteristics. However, this study considers only node features.

3.2. Graph Neural Networks

Graph Neural Networks (GNNs) are a natural choice for processing graph-structured data and, consequently, for analyzing brain networks. There exist various types of GNN architectures. Among those employed in this thesis are Graph Isomorphism Networks (GINs) [59], Graph Convolutional Networks (GCNs) [60], Graph Attention Networks (GATs) [61], as well as several graph readout mechanisms [49]. The subsequent subsections provide detailed explanations of how these specific GNN layers function. However, before delving into those details, firstly, the general operational principles of GNNs are presented, as they all share a common computational framework.

Most contemporary GNN models adopt the message-passing paradigm [62]. In this framework, node features are iteratively updated by aggregating information from their neighbors. The process consists of two primary stages: aggregation and update. The message-passing mechanism can be formally described by

$$\mathbf{a}_v = \text{Aggregate}(\{\mathbf{x}_u : u \in \mathcal{N}(v)\}), \quad \mathbf{x}_v^{\text{new}} = \text{Update}(\mathbf{x}_v, \mathbf{a}_v). \quad (3.3)$$

Here, $\mathcal{N}(v)$ denotes the set of neighboring nodes of node v , and $\mathbf{x}_v$ represents the feature vector associated with node v .

These generalized equations illustrate the flow of information within a single GNN layer, which can be interpreted as one iteration of message propagation across the graph. Although different GNN architectures may implement the aggregation and update functions in distinct ways, most of them adhere to this fundamental message-passing framework.

3.2.1. Graph Isomorphism Network

The Graph Isomorphism Network (GIN) layer serves as a representative example of the message-passing framework, featuring explicitly defined aggregation and update functions.

In GIN, the aggregation operation is implemented as a simple summation over the features of neighboring nodes, while the update step is performed using a multi-layer perceptron (MLP) [63]. Formally, the propagation rules are defined as

$$\mathbf{a}_v = \text{Aggregate}(\{\mathbf{x}_u : u \in \mathcal{N}(v)\}) = \sum_{u \in \mathcal{N}(v)} \mathbf{x}_u, \quad (3.4)$$

$$\mathbf{x}_v^{\text{new}} = \text{Update}(\mathbf{x}_v, \mathbf{a}_v) = \phi((1 + \epsilon)\mathbf{x}_v + \mathbf{a}_v). \quad (3.5)$$

Here, ϕ denotes an MLP with a single hidden layer, and ϵ is a learnable scalar parameter. The use of an MLP is motivated by the Universal Approximation Theorem (UAT) [64], which ensures sufficient expressiveness for nonlinear transformations. The decision to make ϵ learnable is based on empirical and theoretical findings showing that it enhances the representational capacity of the model by allowing flexible weighting of self-node features [59].

The GIN layer can be expressed in matrix form as

$$\mathbf{X}^{\text{new}} = \phi((1 + \epsilon)\mathbf{X} + \mathbf{A}\mathbf{X}). \quad (3.6)$$

Alternatively, using the adjacency matrix with self-loops, it can be written as

$$\mathbf{X}^{\text{new}} = \phi(\hat{\mathbf{A}}\mathbf{X} + \epsilon\mathbf{X}) = \phi((\hat{\mathbf{A}} + \epsilon\mathbf{I})\mathbf{X}). \quad (3.7)$$

The matrix-form of the equations essentially have the same logistics as the iterative form of the equations, written in a compact form. The symbols in these matrix-form expressions can be described as:

- $\mathbf{X} \in \mathbb{R}^{N \times F_{in}}$: Input feature matrix, where N is the number of nodes and F_{in} is the number of input features.
- $\mathbf{X}^{\text{new}} \in \mathbb{R}^{N \times F_{out}}$: Output feature matrix after applying the GIN layer.

- $\mathbf{A} \in \{0, 1\}^{N \times N}$: Adjacency matrix of the graph (without self-loops).
- $\mathbf{I} \in \mathbb{R}^{N \times N}$: Identity matrix.
- $\hat{\mathbf{A}} = \mathbf{A} + \mathbf{I}$: Adjacency matrix with added self-loops.

3.2.2. Graph Convolutional Network

Graph Convolutional Networks (GCNs) extend the principles of Convolutional Neural Networks (CNNs) [65] to non-Euclidean domains, specifically graph-structured data. They operate within the message-passing framework, wherein node features are iteratively updated based on their local neighborhood. In addition to this spatial interpretation, GCNs are also well-studied from the perspective of Spectral Graph Theory [66], providing a complementary frequency-domain analysis of their operations.

The message-passing formulation of GCNs can be expressed as follows:

$$\mathbf{a}_i = \text{Aggregate}(\{\mathbf{x}_j : j \in \mathcal{N}(i) \cup \{i\}\}) = \sum_{j \in \mathcal{N}(i) \cup \{i\}} \frac{e_{j,i}}{\sqrt{\hat{d}_j \hat{d}_i}} \mathbf{x}_j, \quad (3.8)$$

$$\mathbf{x}_i^{\text{new}} = \text{Update}(\mathbf{x}_i, \mathbf{a}_i) = \sigma(\mathbf{W}^\top \mathbf{a}_i). \quad (3.9)$$

This iterative form also conforms to the message-passing framework and has obvious similarities compared to GIN equations in the previous page. The terms in these equations can be explained as:

- $\hat{d}_i = 1 + \sum_{j \in \mathcal{N}(i)} e_{j,i}$, with $\hat{d}_i$ denoting the degree of node i including the self-loop.
- $e_{j,i}$ represents the edge weight from node j to node i (default: 1).
- $\mathbf{W}$ is the learnable weight matrix applied during the update step. Its dimensions are $F_{in} \times F_{out}$, where F_{in} is input feature dimension, and F_{out} is output feature dimension.
- $\sigma(\cdot)$ is an activation function of choice.

As described in [60], these equations can be written in more compact form as

$$\mathbf{X}^{\text{new}} = \sigma(\hat{\mathbf{D}}^{-\frac{1}{2}} \hat{\mathbf{A}} \hat{\mathbf{D}}^{-\frac{1}{2}} \mathbf{X} \mathbf{W}). \quad (3.10)$$

The terms in this equation can be described as:

- $\hat{\mathbf{A}} = \mathbf{A} + \mathbf{I}$: adjacency matrix with added self-loops.
- $\hat{\mathbf{D}}$: degree matrix of $\hat{\mathbf{A}}$, where $\hat{D}_{ii} = \sum_j \hat{A}_{ij}$.
- $\mathbf{X} \in \mathbb{R}^{N \times F_{in}}$: input node feature matrix, where N is the number of nodes and F_{in} is the number of input features.
- $\mathbf{W} \in \mathbb{R}^{F_{in} \times F_{out}}$: learnable weight matrix, mapping input features to output features.
- $\mathbf{X}^{new} \in \mathbb{R}^{N \times F_{out}}$: output node feature matrix after applying the GCN layer.

3.2.3. Graph Attention Network

The Graph Attention Network (GAT) layer also operates within the message-passing framework. It introduces a self-attention mechanism that computes attention coefficients between a node and its neighbors, enabling the model to assign different importance weights to each neighbor when aggregating features. Node features are then updated as a weighted sum of neighboring features, based on these learned attention scores.

This update rule can be written as

$$\mathbf{a}_v = \text{Aggregate}(\{\alpha_{u,v} \mathbf{z}_u : u \in \mathcal{N}(v)\}) = \sum_{u \in \mathcal{N}(v)} \alpha_{u,v} \mathbf{z}_u, \quad (3.11)$$

$$\mathbf{x}_v^{new} = \text{Update}(\mathbf{x}_v, \mathbf{a}_v) = \sigma(\mathbf{a}_v), \quad (3.12)$$

where the individual elements of the attention matrix can be characterized with

$$\alpha_{u,v} = \frac{\exp(h(\mathbf{a}^\top [\mathbf{z}_u \parallel \mathbf{z}_v]))}{\sum_{a \in \mathcal{N}(v) \cup \{v\}} \exp(h(\mathbf{a}^\top [\mathbf{z}_a \parallel \mathbf{z}_v]))}. \quad (3.13)$$

Other terms and symbols in the above equations can be described as:

- $\mathbf{z}_u = \mathbf{W}^\top \mathbf{x}_u$, where $\mathbf{W}$ is a learnable weight matrix applied to the input features.
- $\alpha_{u,v}$ is the attention coefficient indicating the importance of node u 's features to node v . It is 0 if the nodes are not connected.

- $\parallel$ denotes concatenation of feature vectors.
- $\mathbf{a}$ is a learnable attention vector.
- $h(\cdot)$ is a nonlinearity, typically chosen as LeakyReLU.
- $\sigma(\cdot)$ a nonlinearity which is often applied to $\mathbf{x}_v^{\text{new}}$ after aggregation.

In matrix form, the aggregate and update equations can be written as

$$\mathbf{X}^{\text{new}} = \sigma(\boldsymbol{\alpha}\mathbf{X}\mathbf{W}), \quad (3.14)$$

where $\boldsymbol{\alpha}$ is the calculated attention matrix with elements defined in Equation (3.13).

3.2.4. Summary of Different Graph Layers

To conclude the detailed descriptions provided in Sections 3.2.1, 3.2.2, and 3.2.3, a comparative summary of the GIN, GCN, and GAT layers are presented. The table below highlights the key components of each architecture, including their aggregation strategies, update mechanisms, and corresponding matrix form equations. This summary serves to emphasize the conceptual and operational differences among the layers and to facilitate direct comparison.

Table 3.1. Comparison of GNN Variants by Aggregation, Update, and Matrix Form Equations. Please refer to sections 3.2.1, 3.2.2 and 3.2.3 for the definitions of the variables used in this table.

Model	Aggregate	Update	Matrix Form
GIN	$\mathbf{a}_v = \sum_{u \in \mathcal{N}(v)} \mathbf{x}_u$	$\mathbf{x}_v = \text{MLP}(\mathbf{a}_v)$	$\mathbf{X}^{\text{new}} = \text{MLP}\left((\hat{\mathbf{A}} + \epsilon\mathbf{I})\mathbf{X}\right)$
GCN	$\mathbf{a}_v = \sum_{u \in \mathcal{N}(v) \cup \{v\}} \frac{e_{u,v}}{\sqrt{\hat{d}_u \hat{d}_v}} \mathbf{x}_u$	$\mathbf{x}_v = \sigma(\mathbf{a}_v)$	$\mathbf{X}^{\text{new}} = \sigma(\hat{\mathbf{D}}^{-\frac{1}{2}} \hat{\mathbf{A}} \hat{\mathbf{D}}^{-\frac{1}{2}} \mathbf{X} \mathbf{W})$
GAT	$\mathbf{a}_v = \sum_{u \in \mathcal{N}(v)} \alpha_{u,v} \mathbf{W}^T \mathbf{x}_u$	$\mathbf{x}_v = \sigma(\mathbf{a}_v)$	$\mathbf{X}^{\text{new}} = \sigma(\boldsymbol{\alpha} \mathbf{X} \mathbf{W})$

3.2.5. Graph Readouts

After the message-passing or information propagation phases within a graph neural network, the data remains in a structured form—consisting of a graph with nodes

and their updated feature representations. However, in graph-level tasks such as classification or regression, it is necessary to transform this structured representation into a fixed-size vector that captures the global information of the entire graph. This transformation allows downstream processing, such as passing the graph representation to a standard multilayer perceptron (MLP) or other classical neural architectures for prediction tasks.

To achieve this transformation, various strategies have been proposed, generally referred to as graph readout, graph pooling, or graph summarization methods. The most straightforward approaches include concatenating all node features in a predefined order or applying permutation-invariant aggregation functions such as summation, averaging, or max pooling across all nodes. While these methods are simple and effective in some scenarios, they may fail to capture complex structural or feature-based relationships across the graph.

To address these limitations, more sophisticated readout mechanisms have been developed, including Set2Set [67], SortPooling [68], Gated Attention Readout (GARO) [49], and other attention-based pooling methods [69]. These techniques aim to learn more expressive and adaptive summarizations of the node set by modeling inter-node dependencies or dynamically weighting node contributions.

3.2.5.1. Simple Readout Mechanisms. Sum, mean, and max pooling—collectively referred to as readout functions—constitute a class of straightforward yet effective aggregation strategies employed in graph neural networks (GNNs). These methods operate by computing the element-wise sum, mean, or maximum of the feature vectors associated with the neighbors of a given node or vertex. Despite their simplicity, such pooling operations are foundational components in numerous GNN architectures due to their computational efficiency and empirical effectiveness. Notably, sum pooling has been demonstrated to possess greater representational power than mean or max pooling, particularly in its ability to distinguish complex graph structures [59].

3.2.5.2. Global Attention Readout. In this thesis, the attention-based readout strategy introduced in [69] which is called Global Attention is primarily adopted, which utilizes a global attention mechanism over the node features to produce a set of importance weights. These weights are then used to calculate a convex weighted average of the node feature vectors, effectively yielding a fixed-size vector that serves as a summary of the entire graph. The resulting graph-level embedding is then passed to subsequent layers for final prediction tasks. The equation governing this transform is

$$\mathbf{x}_G = \sum_{\forall v \in V} \text{softmax}(\phi_{\text{gate}}(\mathbf{x}_v)) \odot \mathbf{x}_v, \quad (3.15)$$

where $\mathbf{x}_G$ denotes the summarized graph feature vector, and $\phi_{\text{gate}} : \mathbb{R}^{F_{in}} \rightarrow \mathbb{R}$ is a learnable neural network applied to each node's feature vector. F_{in} denotes the nodal feature dimensionality.

3.2.5.3. Graph Attention Readout (GARO). Graph attention readout, in short, GARO [49], is another attention-based readout method based on key query mechanism used in Transformer architectures [70]. The GARO method is incorporated in our experiments to evaluate the impact of different readout mechanisms. It consists of a key and query transformation implemented by the learnable key and query matrices with dimensions $D \times D$, i.e. $\mathbf{W}_{\text{key}}, \mathbf{W}_{\text{query}}$. The governing equations can be summarized with

$$\mathbf{K} = \mathbf{W}_{\text{key}} \mathbf{X}^T, \quad (3.16)$$

$$\mathbf{q} = \mathbf{W}_{\text{query}} (\phi_{\text{mean}} \mathbf{X})^T, \quad (3.17)$$

$$z_{\text{space}} = \text{sigmoid} \left(\frac{\mathbf{q}^T \mathbf{K}}{\sqrt{D}} \right), \quad (3.18)$$

$$\mathbf{x}_G = z_{\text{space}} \mathbf{X}, \quad (3.19)$$

where ϕ_{mean} is an $1 \times N$ vector of coefficients where each element being equal to $1/N$, N being the number of nodes in the graph, facilitating an averaging operation over the nodal features in the graph.

3.3. Attention Based Low Dimensional Brain Connectome Embedding

Figure 3.1 illustrates the overall architecture employed throughout this thesis to facilitate a range of downstream analyses. The design incorporates core components introduced in Section 3.2, particularly within the Graph Neural Network (GNN) module. Specifically, the GNN block is composed of three sequential layers: a Graph Isomorphism Network (GIN) layer, followed by a Graph Attention Network (GAT) layer, and subsequently a global attention-based readout mechanism for generating a graph-level representation. This readout method is herein referred to as *Attention Pooling*.

The architecture comprises two operational branches: the *online flow* and the *offline flow*. The online flow pertains to the training phase of the model, wherein the architecture is optimized for a binary classification task. The classification tasks explored in this thesis include Alzheimer’s Disease (ADD) versus Mild Cognitive Impairment (MCI), ADD versus Subjective Cognitive Impairment (SCI), and MCI versus SCI.

During the online flow, the model processes input brain networks—represented as binarized graphs with associated node features—via the GNN module. This module performs message passing using the GIN and GAT layers and subsequently computes a graph-level summary using the attention-based readout. The output of the GNN module is a fixed-dimensional graph embedding of dimension $D = 64$, which serves as input to a downstream multilayer perceptron (MLP) classifier.

The choice of GNN components is motivated by their respective strengths: the GIN layer is included due to its high expressive power in distinguishing graph structures, while the GAT layer is leveraged for its ability to dynamically weight node-level features based on learned attention coefficients, thereby allowing the model to focus on salient regions of the graph. The attention-based pooling layer refines the graph representation further by assigning higher importance to nodes that contribute more significantly to the task, thereby enhancing the model’s discriminatory capacity.

The MLP classifier consists of two fully connected layers. The first hidden layer contains 64 neurons and employs a Rectified Linear Unit (ReLU) activation function. The final layer consists of 2 output neurons corresponding to the binary classification outcome. The output is passed through a softmax function to obtain class probabilities. The model is trained end-to-end using the binary cross-entropy loss function. The key architectural parameters are summarized in Table 3.2.

Table 3.2. Architecture parameters of the Attention-Based Low Dimensional Brain Connectome Embedding.

GIN nodal feature dimension	128
GAT nodal feature dimension (D)	64
Number of neurons in the first hidden layer	64
Number of neurons in the output layer	2

Upon completion of training within the online flow, the resulting GNN model is capable of summarizing input brain networks into compact, fixed-size embeddings in a D -dimensional latent space. These learned representations are subsequently utilized in the *offline flow*, specifically within the *Low-Dimensional Manifold Constructor* module. This module applies Principal Component Analysis (PCA) to the set of graph embeddings obtained from the training data. PCA is employed to reduce the dimensionality of the data while retaining the principal components that capture the most variance. The result is a low-dimensional manifold that enables visualization and further analytical procedures.

The offline flow enables the projection of new, previously unseen brain networks into this learned latent space. To do this, a given network is passed through the trained GNN to obtain its graph-level embedding, which is then projected onto the PCA-derived manifold. This enables further downstream analyses, such as tracking trajectories across the cognitive decline spectrum or interpreting patterns associated with neurodegenerative progression.

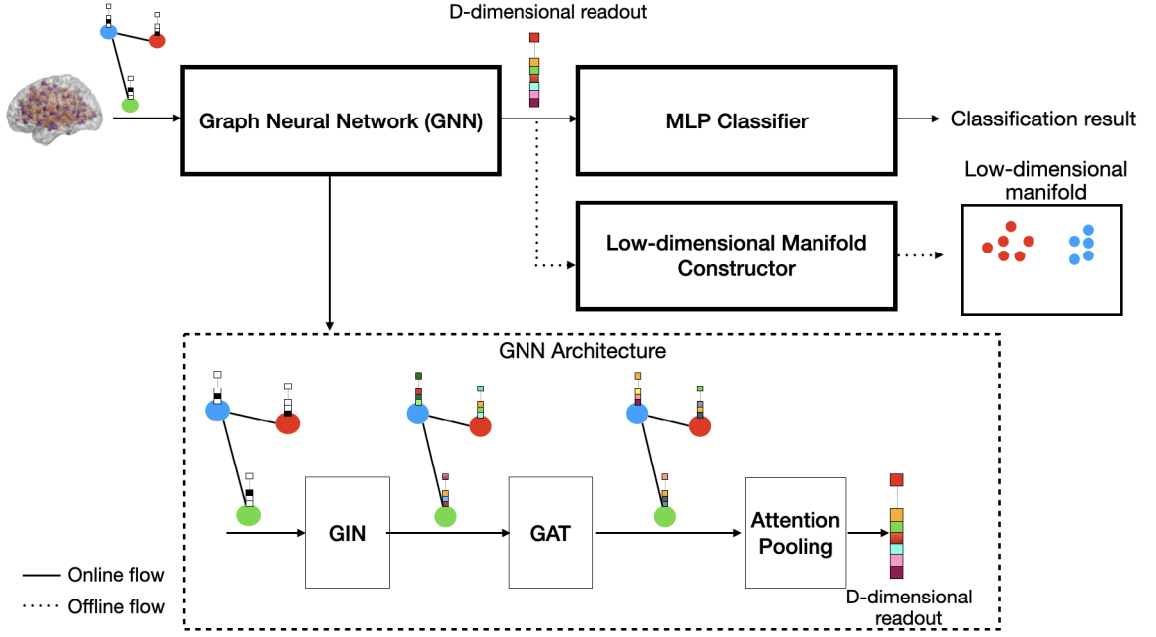


Figure 3.1. The framework includes a GNN for feature extraction and an MLP for binary classification (e.g., ADD vs. SCI). The GNN uses GIN, GAT, and attention pooling to generate D-dimensional embeddings. These are used for classification in the online phase (solid lines). In the offline phase (dashed lines), embeddings are projected into a low-dimensional manifold via PCA for visualization and analysis.

3.3.1. Architectural Modifications Explored

To enhance the interpretability of the proposed framework outlined in Figure 3.1—hereafter referred to as the original model—several architectural modifications have been investigated. These modifications are motivated by the objective of enabling scale-space analysis and facilitating biologically meaningful interpretations of brain organization. In particular, the notion of soft clustering and subnetwork-level representation has been incorporated to probe the multiscale structure of brain connectomes. By aggregating nodal features into subnetworks, the model aims to capture latent spatial patterns that may be associated with functional modules or disease-specific alterations.

The effectiveness of these architectural variants is assessed in comparison with the original model, both in terms of predictive accuracy and interpretability. The following sections describe two such modifications: *Subnet GAT*, which introduces subnetwork-level aggregation following the attention mechanism, and *Subnet GAT with Subnet Attention*, where subnetwork formation precedes attention-based feature propagation. Each variant offers a distinct perspective on feature transformation and attention over network topology.

3.3.1.1. Subnet GAT. The architectural modification introduced in the Subnet GAT block is illustrated in Figure 3.2. In this configuration, a new feature aggregation layer is appended subsequent to the Graph Attention Network (GAT) operation. Let the output of the GAT layer be denoted by $\mathbf{X}'$. This output is then aggregated across subnetworks through multiplication with a matrix $\mathbf{B}$, yielding a new feature matrix $\mathbf{X}''$.

The matrix $\mathbf{B} \in \mathbb{R}^{k \times n}$ is a learnable component, where n represents the number of nodes in the original graph and k is a hyperparameter defining the number of subnetworks (or subgraphs) to be extracted. Subnet features are derived by weighted summation of node-level features using the coefficients encoded in $\mathbf{B}$. The construction of $\mathbf{B}$ is formally defined as

$$\mathbf{B}_{k \times n} = \text{sigmoid}(\beta + \lambda \mathbf{m}_k). \quad (3.20)$$

In this formulation, $\mathbf{m}_k$ denotes the first k eigenvectors of the graph Laplacian corresponding to the lowest k eigenvalues, effectively capturing the low-frequency structural characteristics of the graph. The matrix $\beta \in \mathbb{R}^{k \times n}$ comprises learnable parameters, while λ is a scalar hyperparameter that regulates the influence of graph topology on subnet definitions. A higher value of λ biases subnet construction more heavily toward spectral properties of the graph, whereas a lower value encourages the model to rely more on optimization dynamics during training. The sigmoid activation ensures numerical stability by bounding the aggregation coefficients.

The computation sequence is summarized with

$$\mathbf{X}' = \alpha \mathbf{XW} + \mathbf{b}, \quad (3.21)$$

$$\mathbf{X}'' = \mathbf{B}\mathbf{X}' \Rightarrow \mathbf{X}'' = \mathbf{B}\alpha \mathbf{XW} + \mathbf{Bb}. \quad (3.22)$$

Collectively, the Subnet GAT mechanism can be represented by the composite attention matrix $\mathbf{B}\alpha$, which acts at the subnetwork level rather than individual nodes. Here, α represents the node-level attention matrix described in Section 3.2.3.

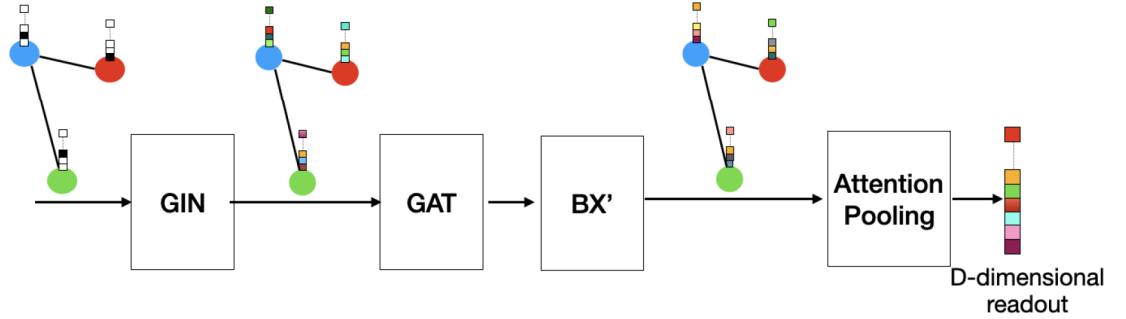


Figure 3.2. GNN Block after Subnet GAT modification. A subnetwork-level feature aggregation module is introduced following the GAT layer.

3.3.1.2. Subnet GAT with Subnet Attention. A complementary architectural variant is presented in Figure 3.3, where the subnetwork-level aggregation is performed prior to the GAT operation. In this version, the same aggregation matrix $\mathbf{B}$, as defined in Equation (3.20), is employed to reduce the original node-level feature space into k subnetwork representations. The GAT layer is then applied to these aggregated features under the assumption of a fully connected graph among subnetworks.

This design shifts the attention mechanism to operate directly on subnetwork representations. Accordingly, the composite attention matrix becomes $\alpha \mathbf{B}$, in contrast to $\mathbf{B}\alpha$ used in the previous approach. This inversion reflects the change in the sequence of aggregation and attention operations.

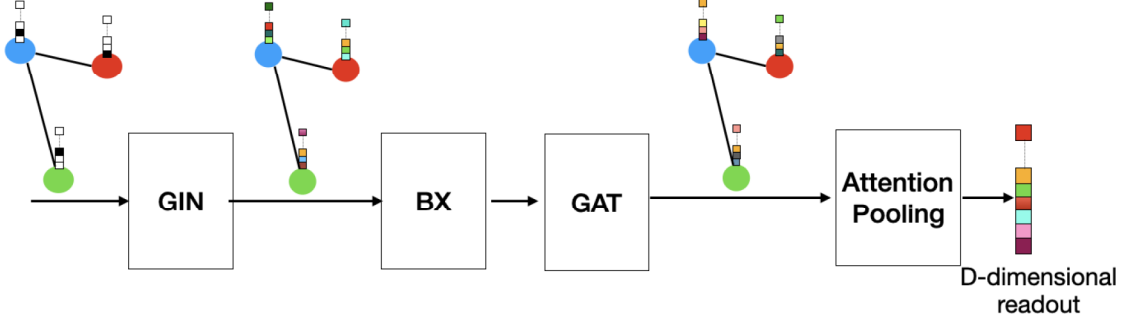


Figure 3.3. GNN Block after Subnet GAT and Subnet Attention modification.

Subnetwork aggregation is applied prior to the GAT layer.

3.3.1.3. Subnet GARO. The Graph Attention Readout (GARO) layer was further modified to facilitate scale-space analysis and the extraction of interpretable subnetwork-level representations. Specifically, the computation of the key matrix was reformulated and can be described with the equation

$$\mathbf{K} = \mathbf{X}^\top \sigma(\mathbf{W}_{\text{key}}). \quad (3.23)$$

In contrast to the original GARO formulation—where $\mathbf{W}_{\text{key}}$ is of dimension $D \times D$ —the revised approach employs a key weight matrix of dimension $n \times k$, where n denotes the number of nodes and k is the number of subnetworks. This structural change enables a redefinition of the key matrix as a soft aggregation of node features, modulated by sigmoid-activated gating functions. The resulting $\mathbf{K}$ matrix effectively encodes soft clustering coefficients that allow the construction of latent super-nodes or subnetworks composed of overlapping node subsets.

The use of the sigmoid activation function plays a crucial role by ensuring that only partial contributions from each node’s feature vector are aggregated, enabling nuanced and non-convex combinations of node representations. Although the resulting subnetworks are not mutually exclusive—i.e., overlapping in structure—they provide a meaningful abstraction of how node groupings may emerge to support task-specific representations. This modification aligns with the broader objective of uncovering multiscale and modular organizational principles within brain networks.

4. EXPERIMENTS AND RESULTS

4.1. Data Acquisition

The cohort utilized in this thesis comprises a total of 88 participants. All individuals provided written informed consent in accordance with the approvals granted by the relevant institutional ethics committees¹. Table 4.1 presents the distribution of participants by age, gender, and diagnostic group. This dataset is called as CAPA dataset.

Table 4.1. Summary of the cohort of participants.

	ADD	MCI	SCI
Number of Participants	18	46	24
Age	69.83 ± 8.52	62.04 ± 9.78	56.95 ± 7.54
Gender (M/F)	10/8 (44% F)	23/23 (50% F)	12/12 (50 % F)

The imaging data was acquired using Philips Achieve 3T MRI system (Netherlands) with a 32-channel head coil at Istanbul University, Faculty of Medicine, Neuroimaging Unit of Hulusi Behçet Life Sciences Research Laboratory. Table 4.2 shows the common parameters used for different modalities.

Table 4.2. Common MRI Acquisition Parameters by Modality.

Parameter	T1-weighted	DWI	rs-fMRI
TE/TR (ms/ms)	3.8 / 8.3	92 / 9032	30 / 3000
FOV (mm ²)	220 × 240	200 × 236	212 × 199
Voxel Size (mm ³)	1.0 ³	2.27 ³	3.31 ³
Slices	180	71	48 (axial)
Flip Angle	8°	-	80°

¹Istanbul University, Istanbul Faculty of Medicine, Ethics Committee Approval: 877/30.05.2014; Boğazici University, Ethics Committee Approval: 2014-1/17.02.2014

In addition to the acquisition parameters presented in Table 4.2, the following methodological details apply to each imaging modality:

- T1-weighted (T1w) MRI: Acquired using a 3D Fast Field Echo (FFE) pulse sequence in conjunction with a multi-shot Turbo Field Echo (TFE) imaging mode.
- Diffusion-Weighted Imaging (DWI): Data were acquired using 120 gradient directions with a maximum gradient strength of 40 mT/m and a slew rate of 200 mT/m/ms. The acquisition utilized 6 shells in q-space, and the imaging matrix was set to 112×112 .
- Resting-State fMRI (rs-fMRI): Acquired using the Fast Field Echo (FFE) technique in multi-slice mode with a single-shot Echo Planar Imaging (EPI) sequence.

Subsequently, each subject’s T1-weighted MRI, resting-state fMRI, and diffusion-weighted imaging (DWI) data were isotropically resampled to a resolution of 1.5 mm and co-registered using a combination of FreeSurfer, FSL, and Tortoise software toolboxes. The resulting co-registered volumes were then subjected to manual quality control and curation by expert neurologists to ensure anatomical accuracy and preprocessing reliability.

4.2. Clinical Assessment

Diagnostic classification and clinical staging of participants were conducted in accordance with established criteria. Initial clinical assessment was performed using the National Institute on Aging and Alzheimer’s Association (NIA-AA) criteria [71]. Participants meeting the criteria for probable Alzheimer’s disease dementia with an amnesic presentation were assigned to the Alzheimer’s disease dementia (ADD) group. Those in the Mild Cognitive Impairment (MCI) group were identified based on their performance on the Free and Cued Selective Reminding Test (FCSRT), specifically scoring 27 or below on Total Free Recall (TFR) and 0.67 or below on the Cue Index (CI). The TFR threshold of 27 was selected for its greater sensitivity in detecting dementia risk, as compared to the more conservative cutoff of 24. Participants ex-

ceeding these thresholds ($\text{TFR} > 27$ or $\text{CI} > 0.67$) were classified as having Subjective Cognitive Impairment (SCI). In addition, SCI participants were recruited through community advertisements and required to score at least 1 on either the Cognitive Function Instrument – Subject form (CFI-S) or the Study Partner form (CFI-SP).

Clinical staging was further evaluated using the Clinical Dementia Rating (CDR) scale. ADD participants had CDR scores of 0.5 or 1, corresponding to very mild or mild dementia, respectively. MCI participants had a CDR score of 0.5 and a CDR-Sum of Boxes (CDR-SOB) score of 0.5 or 1. SCI participants, representing cognitively normal individuals, had a CDR score of 0. Final diagnoses were established through comprehensive neurological and neuropsychological assessments, supported by cranial magnetic resonance imaging (MRI), and reviewed by a panel of behavioral neurologists.

Participants were excluded if they had a history of neurological or psychiatric disorders affecting cognition, alcohol or substance abuse, significant head trauma involving loss of consciousness, or moderate to severe white matter hyperintensities (Fazekas score 2–3) as observed on MRI. Individuals with contraindications to MRI scanning were also excluded from the study.

4.3. Brain Connectome Construction

In this study, structural brain networks (sNets) are primarily utilized to represent individual subjects, while complementary analyses are also conducted using multiple variants of functional brain networks (fNets). These networks are constructed based on two widely adopted parcellation schemes. The first is the Destrieux atlas [5], which yields cortical networks comprising $n = 148$ nodes. The second is the Schaefer atlas [6], which supports variable spatial granularity: it produces functional networks with $n = 100, 200, \dots, 1000$ nodes and structural networks with $n = 200, 300, 400$ nodes. This range facilitates the investigation of atlas resolution effects on downstream tasks.

Structural connectomes were generated using deterministic tractography, implemented via fourth-order Runge–Kutta (RK4) numerical integration, as outlined in [72]. The tractography parameters employed in this process are detailed in Table 4.3.

Table 4.3. Tractography Parameters.

Parameter	Value
Maximum Curvature	35°
Minimum FA Threshold (Stopping Criterion)	0.15
Minimum Fiber Length	20 mm
Number of Seeds per Voxel	30
Step Size	0.7 mm

The structural connectome edge weights, denoted as a_{ij} , are computed using a volumetric association formulation

$$a_{ij} = \frac{2}{V_i + V_j} \sum_k W_{ik} W_{jk}. \quad (4.1)$$

Here, V_i denotes the volume of cortical region i , and W_{ik} quantifies the association between the i -th region and the k -th fiber. This association metric, introduced by [73], employs a volume-normalized probabilistic framework to ensure numerical stability in graph learning models. The association strength W_{ik} reflects the likelihood of a diffusing particle originating from the endpoint f_k of a fiber and arriving at a voxel r_i in region R_i , subject to a spatial proximity constraint. The corresponding Gaussian kernel-based formulation is given by

$$W_{ik} = \sum_{r_i \in R_i, |r_i - f_k| \leq 2\sigma} \frac{1}{(2\pi)^{3/2} \sigma^3} \exp\left(-\frac{|r_i - f_k|^2}{2\sigma^2}\right). \quad (4.2)$$

In this equation, σ is set to 0.155, based on kernel density estimation as suggested by [74]. The resulting structural matrices are binarized by thresholding with zero. Specifically, the governing equations are

$$\mathbf{A}[\mathbf{A} > 0] = 1, \quad (4.3)$$

$$\mathbf{A}[\mathbf{A} = 0] = 0. \quad (4.4)$$

Although structural connectomes constitute the primary analytical focus of this thesis,

alternative forms of functional connectivity are also explored for comparative purposes. In addition to the standard fNet representation, two other variants—denoted as *l*Net and eNet—are evaluated.

Functional connectomes are derived from resting-state BOLD (Blood Oxygen Level Dependent) signals, which represent time-series activity recorded across voxels. To generate node-level signals, BOLD time-series are averaged within each parcellated region. Functional connections are traditionally inferred from statistical dependencies (e.g., Pearson correlations) between region-level time-series, $\mathbf{b}_i$ and $\mathbf{b}_j$. However, to enhance sparsity and robustness, this study adopts an optimization-based formulation proposed in [75] and later adapted for neurological applications by [55]. This formulation yields the *l*Net variant, which emphasizes local smoothness and sparsity in the connectivity structure.

The eNet variant, employed in parallel, is constructed following the methodology outlined in [76]. These functional network variants collectively facilitate a comprehensive evaluation of how different connectivity modeling approaches influence graph-based learning in the context of ADD.

4.4. Low Dimensional Disease Manifold

The architecture depicted in Figure 3.1 is initially trained on a binary classification task aimed at distinguishing between individuals diagnosed with Alzheimer’s Disease Dementia (ADD) and those with Subjective Cognitive Impairment (SCI). This training leverages a graph neural network (GNN), from which intermediate graph-level representations are extracted. These representations are obtained from the output of the attention-based graph readout layer, which integrates node-level information into a global graph embedding. The resulting latent representation is a 64-dimensional vector that encapsulates key structural and functional properties of the brain connectome.

The training process is conducted using all available ADD and SCI samples, and the optimization is carried out using the binary cross-entropy loss function, which is appropriate for binary classification tasks. The specific hyperparameters and training configurations used are detailed in Table 4.4.

After the model has been trained, principal component analysis (PCA) is applied to the collection of 64-dimensional embeddings in order to reduce their dimensionality to two principal components. These two components are selected because they capture 99% of the total variance in the learned representations, ensuring that the vast majority of the information encoded in the original latent space is preserved. While this dimensionality reduction is useful for visualization, the selection is also analytically justified by the high variance retention. Subsequently, brain graphs corresponding to individuals with Mild Cognitive Impairment (MCI)—who were not included during training—are projected onto this learned low-dimensional manifold.

As shown in Figure 4.1, the MCI samples are spatially located between the clusters formed by the ADD and SCI groups. This configuration is consistent with the clinical characterization of MCI as an intermediate stage in the progression from normal cognitive aging to Alzheimer’s disease, thereby providing further validation of the learned embeddings as biologically meaningful representations of neurodegenerative progression.

Table 4.4. Summary of Training Parameters.

Description	Value
Batch Size	6
Number of Epochs	100
Learning Rate	0.001
Weight Decay	0.001
Optimizer	Adam

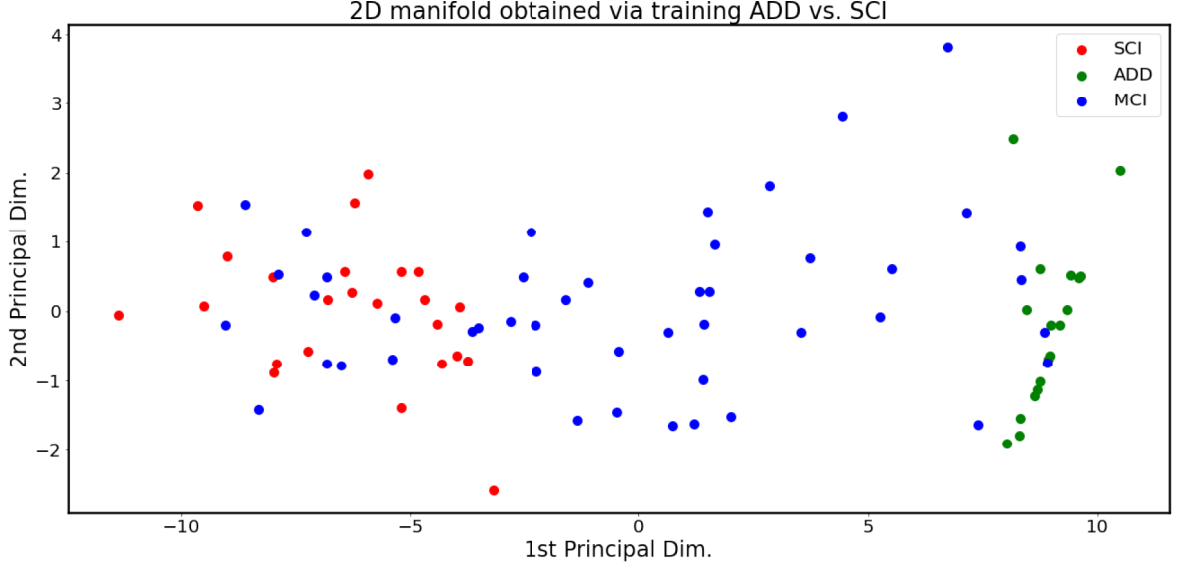


Figure 4.1. The model in Figure 3.1 is trained on ADD vs. SCI classification. After training, PCA is applied to the learned embeddings to obtain a 2D manifold. MCI samples, not seen during training, are projected onto this space and fall between ADD and SCI clusters. This aligns with clinical expectations, supporting the model’s ability to capture the cognitive decline spectrum.

Then, the same GNN architecture was applied to a second binary classification task, this time training the model to discriminate between ADD and MCI. The training procedure mirrored that of the ADD vs. SCI setting, using identical hyperparameters (Table 4.4) and the full set of available ADD and MCI samples. As before, 64-dimensional graph-level embeddings were extracted from the attention-based read-out layer. To enable comparison across diagnostic groups and reduce dimensionality, again, PCA was applied and retained the top two components, which captured 99% of the variance in the latent space.

In this setting, SCI samples—excluded from training—were projected onto the resulting two-dimensional space. As shown in Figure 4.2, these samples do not occupy a transitional position between ADD and MCI, as MCI did in the previous experiment. Instead, they are largely intermixed with the MCI cluster. This observation is consistent with the clinical understanding of SCI as a preclinical or early stage in the continuum of cognitive decline. The overlap suggests that SCI and MCI may share

underlying neurobiological characteristics captured by the model, and highlights the difficulty of sharply delineating these groups in latent space. The resulting manifold further supports the plausibility of the learned representations as reflecting meaningful stages of disease progression.

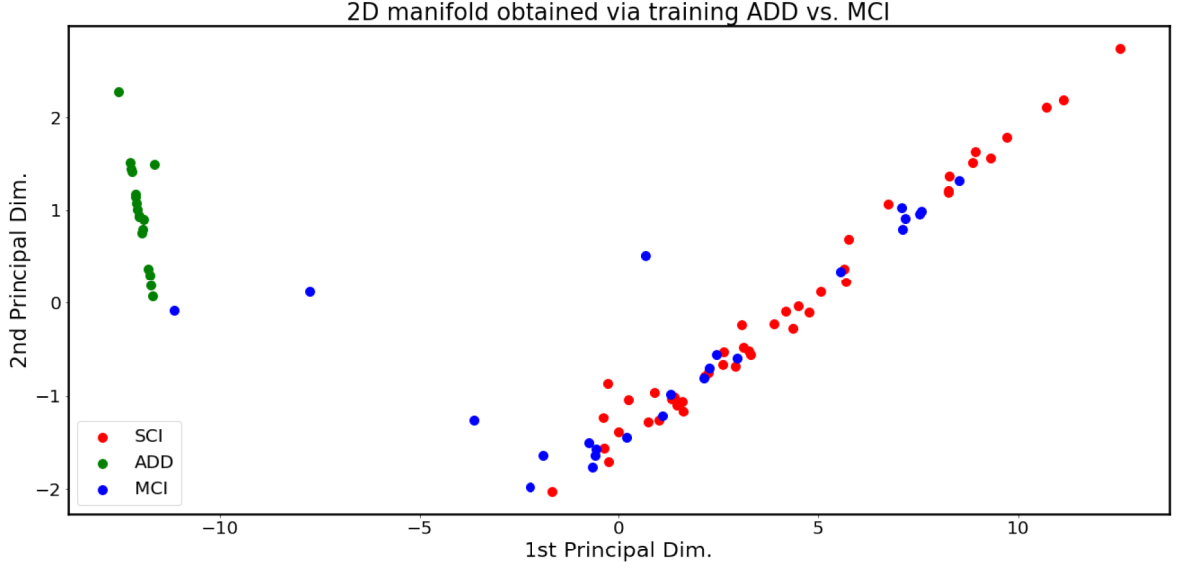


Figure 4.2. A similar experiment is performed using a model trained on ADD vs. MCI. PCA is applied to the learned embeddings, and SCI samples—unseen during training—are projected onto the 2D manifold. SCI cases blend with MCI, lacking a distinct cluster. This overlap reflects the clinical closeness of SCI and MCI and supports the model’s sensitivity to early-stage cognitive variations.

Finally, a third experiment was conducted using the same architectural setup, this time training the model to differentiate between MCI and SCI. As in the previous settings, 64-dimensional graph-level embeddings were extracted from the output of the GNN’s attention-based readout layer. These embeddings were subsequently subjected to principal component analysis (PCA), retaining the top two principal components that together accounted for 99% of the variance in the original feature space. Following this, ADD samples—excluded from training in this setting—were projected onto the resulting two-dimensional manifold. The outcome of this experiment is visualized in Figure 4.3.

In contrast to the manifolds obtained in the ADD vs. SCI and ADD vs. MCI experiments, the resulting low-dimensional space does not exhibit a clearly interpretable or clinically meaningful structure. First and foremost, the classification performance of the model on the MCI vs. SCI task was notably lower than in the other two binary classification settings. This limited predictive capacity, discussed in greater detail in Section 4.5, likely hindered the model’s ability to learn discriminative and biologically coherent representations. Consequently, the projection of MCI and SCI samples shows substantial overlap in the two-dimensional space, with no well-defined separation between the two groups. While some SCI samples appear interspersed among MCI cases, others form a loosely defined cluster, indicating inconsistency in how the model embedded SCI samples.

Moreover, when ADD samples are introduced as a third class and projected onto this manifold, they do not form a distinct cluster that is well-separated from MCI or SCI. Instead, ADD samples appear partially situated between the two overlapping groups and, in some cases, even align more closely with SCI than with MCI. This distribution contradicts the expected clinical trajectory of cognitive decline, in which ADD typically represents a more advanced stage relative to both MCI and SCI. The lack of coherent spatial organization in this manifold underscores the consequences of weak classification performance: when the model fails to accurately discriminate between classes during training, the latent space it constructs does not reflect meaningful disease-related differences.

This result underscores the critical relationship between a model’s predictive performance and the interpretability of its learned representations. When the model lacks sufficient discriminative power, it is unable to capture clinically meaningful patterns in the data, leading to embeddings that are less informative and more difficult to interpret. A detailed evaluation of the model’s performance in this classification setting is presented in Sections 4.5 and 4.9. In light of these limitations, the manifold learned from the MCI vs. SCI task will not be further examined in the remainder of this thesis, as it offers limited interpretive or clinical value.

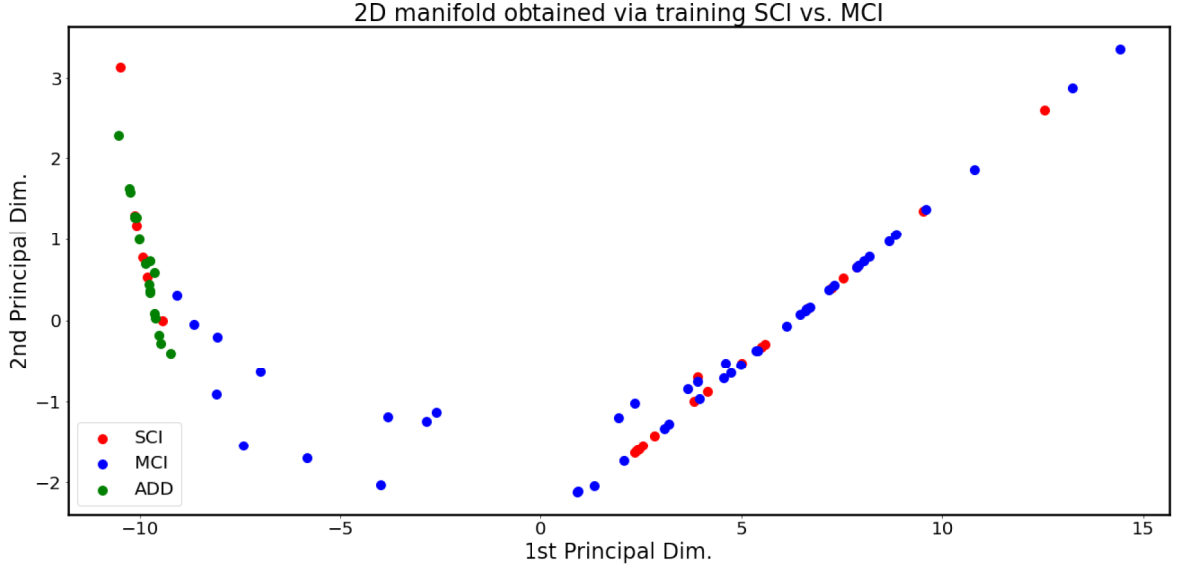


Figure 4.3. A similar experiment is conducted using a model trained on MCI vs. SCI.

The model performs poorly, and the resulting manifold shows no clear separation. Samples from both groups appear intermixed, especially on the left side of the plot.

This aligns with known challenges in distinguishing MCI from SCI. When ADD samples are projected, they also overlap, making the visualization uninterpretable.

After examining the three resulting manifolds, the most promising candidate for further analysis is the one obtained from training the model on the ADD vs. SCI classification task, as shown in Figure 4.1. This selection is motivated by the observation that the ADD and SCI samples form relatively well-separated clusters in the two-dimensional latent space, while MCI samples are situated between them, thereby offering a coherent and clinically plausible representation of the cognitive decline continuum. In contrast, the manifold derived from the ADD vs. MCI classification task fails to capture this continuum effectively, as SCI samples project onto the same region as MCI, thereby obscuring the overall disease progression structure. As previously discussed, the manifold generated from the MCI vs. SCI classification task is even less suitable for further analysis, primarily due to the model’s limited predictive performance in that setting, which resulted in poorly structured embeddings. Furthermore, this latter manifold represents only a partial spectrum of disease progression. Consequently, the ADD vs. SCI trained manifold offers the most interpretable and complete representation of cognitive decline and is thus selected as the basis for downstream

analyses.

To enable further investigation of the manifold structure, distinct regions are defined within the latent space that are predominantly associated with each diagnostic class: ADD-heavy, SCI-heavy, and MCI-heavy regions. To achieve this, Gaussian distributions are fit to the first principal component of the manifold, as this component captures the majority of the variance in the latent representations. For each diagnostic group, the empirical mean and standard deviation are computed along this component. The corresponding class-heavy regions are then defined as intervals spanning one standard deviation above and below the mean for each group, resulting in three distinct regions. The fitted Gaussian curves and corresponding standard deviation boundaries are illustrated in Figure 4.4. These boundaries are subsequently overlaid onto the original manifold visualization in Figure 4.5, thereby demarcating the regions used for subsequent normative analysis.

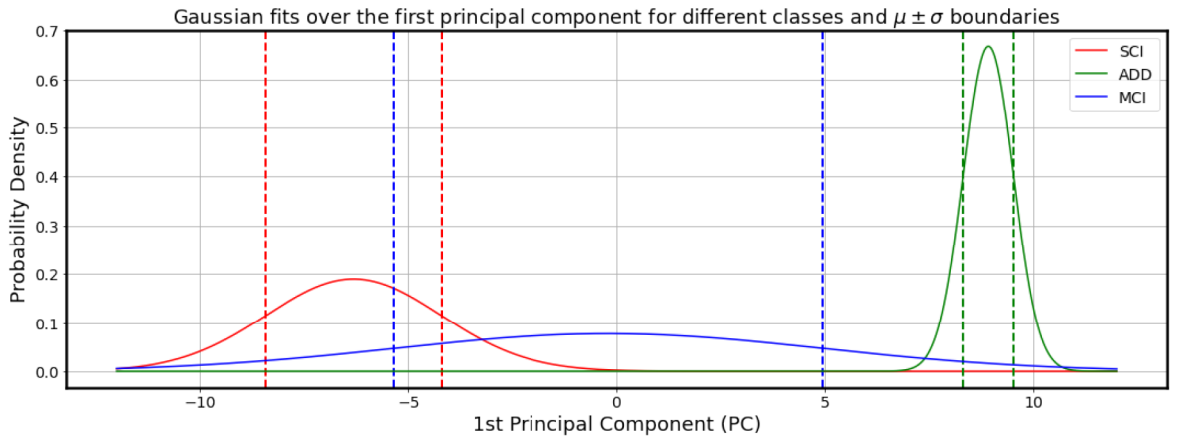


Figure 4.4. The first principal component of the manifold in Figure 4.1 is modeled using Gaussian fits for SCI, MCI, and ADD. Class-specific means and standard deviations define intervals spanning $\pm\sigma$, marked by dashed lines. These serve as reference regions for analyzing connectome changes across diagnostic groups.

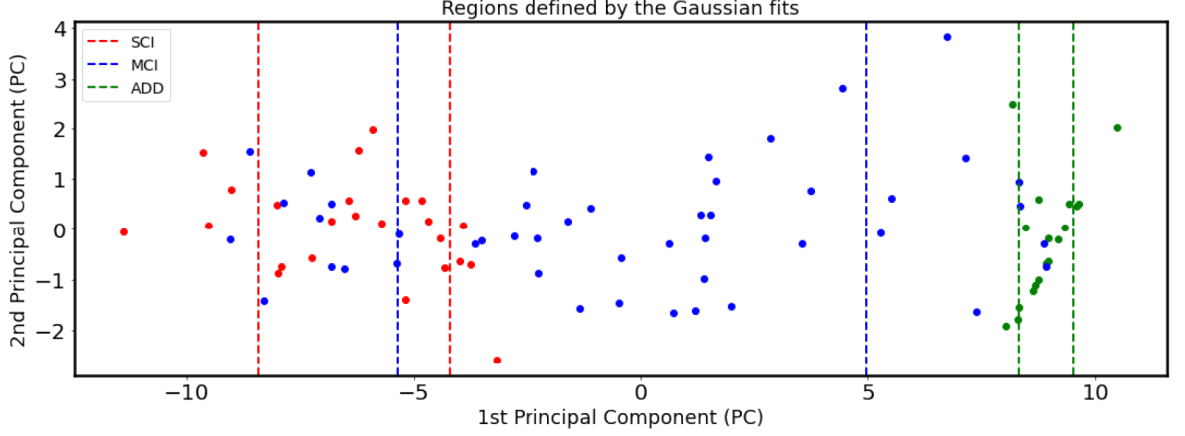


Figure 4.5. The class-specific regions defined in Figure 4.4 are overlaid onto the two-dimensional manifold presented in Figure 4.1 to provide a clearer visual delineation of the diagnostic boundaries. These regions, along with the corresponding patient samples, are subsequently utilized in downstream analyses to investigate alterations in brain connectome organization across clinical groups.

With these regions established, a normative analysis was conducted to identify deviations from a reference connectome. The SCI-heavy region was designated as the normative baseline, and the mean and standard deviation of connectome edge weights were computed for all SCI patients within this region. Assuming a Gaussian distribution of edge weights, the resulting mean SCI connectome served as a normative template against which other samples—particularly those located in the MCI-heavy and ADD-heavy regions—were compared.

For a given connectome edge, let μ_{sci} denote the mean edge value across SCI patients, and σ_{sci} its corresponding standard deviation. For a patient from a different group (e.g., MCI or ADD), let e_c be the edge value in question. Under the normal distribution assumption for SCI edge values, the probability that μ_c could have been generated from the same distribution is computed using a two-tailed Gaussian integral, which can be described as

$$p = 2 \int_{|e_c|}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_{sci}} e^{-\frac{1}{2\sigma_{sci}^2}(x-\mu_{sci})^2} dx. \quad (4.5)$$

This probability value serves as a measure of similarity between each patient’s connectome and the normative SCI connectome.

tome and the normative SCI representation on a per-edge basis. To translate these edge-level probabilities into region-level insights, the probabilities for all edges connected to each brain region (cortex) are aggregated, effectively producing a regional similarity score. These regional scores are then aggregated across all patients within each group-defined region (e.g., ADD-heavy or MCI-heavy), resulting in group-level cortical similarity maps.

To facilitate interpretation and visualization, the probability values are transformed logarithmically, producing a log-likelihood measure that ranges from 0 to negative infinity. Lower values indicate a higher degree of deviation from the normative SCI connectome, and thus, greater dissimilarity.

The results of this normative analysis are presented in Figures 4.6 and 4.7, which depict brain maps highlighting cortical regions that differ most significantly between SCI and ADD, and between SCI and MCI, respectively. In these visualizations, color intensity encodes the degree of difference: lighter colors (e.g., light blue) indicate minimal deviation, while darker shades progressing toward yellow and red represent increasing dissimilarity. These maps allow for the identification of specific cortical regions that may play a role in the early or late stages of cognitive decline.

Importantly, the most prominently differing regions, such as the precuneus, are consistent with established findings in the neuroscience and clinical literature [77], lending further credibility to the analysis. A more detailed interpretation of these brain maps, including their implications in the context of neurodegenerative disease progression, will be discussed in Section 5.

The efficacy of the learned manifold was further assessed using a sliding window analysis traversing the cognitive spectrum from Subjective Cognitive Impairment (SCI) to Alzheimer’s Disease Dementia (ADD). The procedure was initiated at the mean connectome of the SCI cohort, with the window systematically shifted toward the ADD region. Each window spanned six units along the first principal component axis

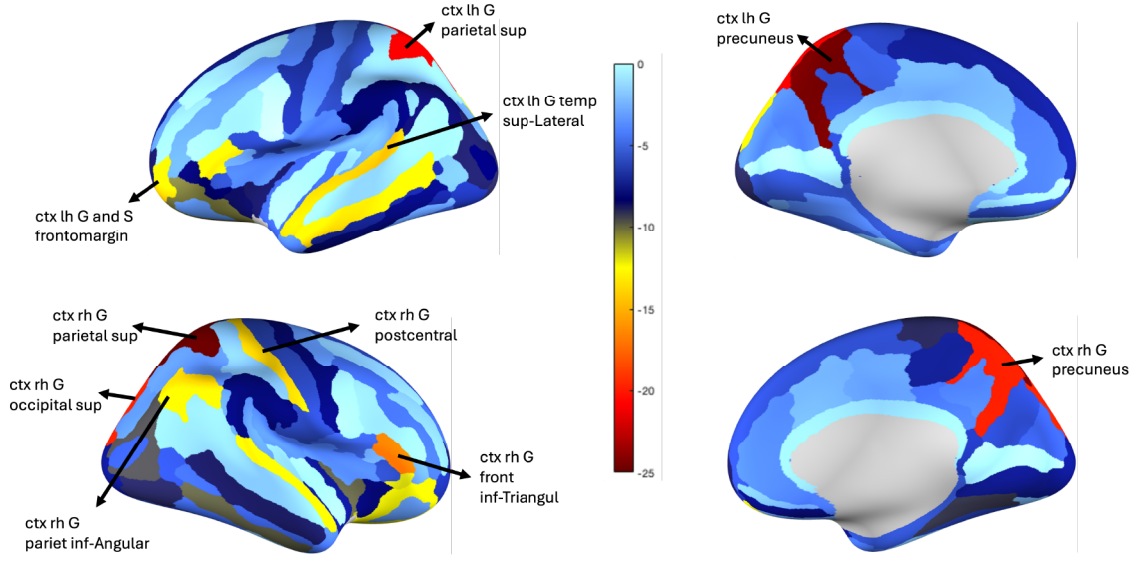


Figure 4.6. Cortical surface representations highlighting structural alterations distinguishing ADD from SCI. Relative to the SCI group, individuals with ADD exhibit substantial deviations in brain network organization, with prominent differences localized to the bilateral precuneus, superior parietal lobule, postcentral and angular gyrus.

(± 3), and for each position, an average connectome was computed following the same methodology previously applied to the SCI group. The probabilistic similarity analysis was then repeated for each window to evaluate structural deviations along the manifold trajectory.

In this context, there is no need to aggregate over individual samples within the sliding window because the connectomes have already been averaged at the window level. While the previous analysis provided cortex-wise similarity scores, the current objective is to evaluate the overall likelihood of the aggregated connectome within each sliding window with respect to the mean SCI representation. To obtain a single scalar measure of similarity for each window, the per-cortex likelihoods are aggregated by computing their product, thereby yielding a unified likelihood score for that region. We, then, take the logarithm of this measure, to obtain a better representation. This methodology allows us to quantify how similarity to the SCI mean connectome changes progressively along the SCI-to-ADD continuum. It is important to note that, for this

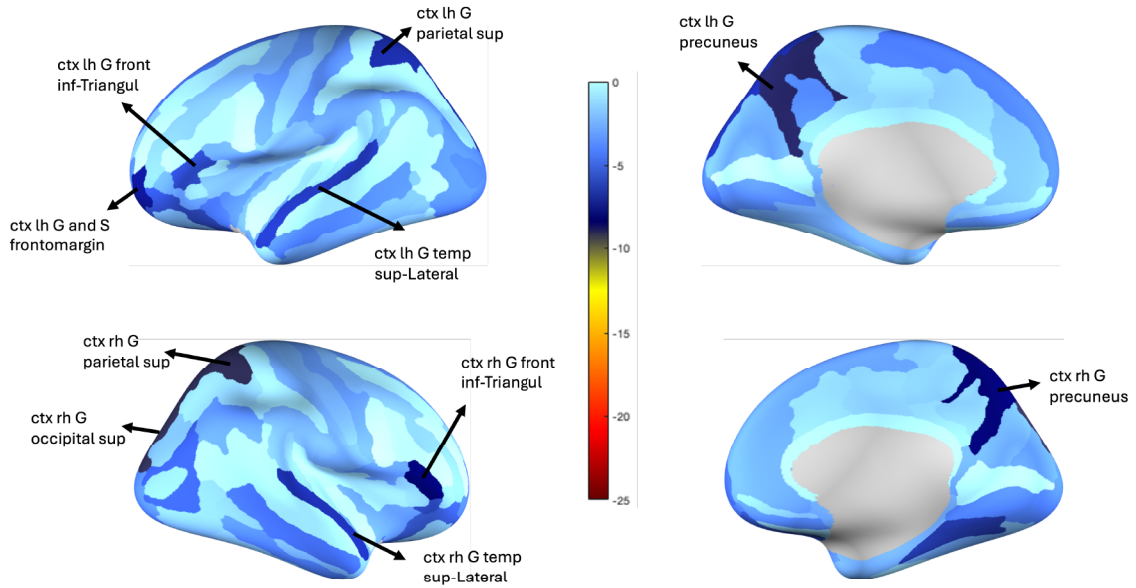


Figure 4.7. Cortical surface maps display regional deviation patterns between SCI and MCI groups. Values indicate the log-sum of edge-wise differences from the SCI reference, projected onto cortical regions. Warmer colors denote greater deviation.

Notable areas of difference include the right superior parietal lobule, precuneus, occipital cortex, and inferior frontal gyrus.

analysis, mean connectomes are used per window to ensure consistent and meaningful comparison across windows. This normalization is essential because each window may encompass a different number of subjects, which would otherwise bias the likelihood estimates. Such normalization was not strictly necessary in our earlier analysis, where the focus was on identifying brain regions that differed significantly between SCI and ADD/MCI within themselves, rather than comparing probabilistic similarities across multiple regions.

Finally, as depicted in Figure 4.8, a near monotonic decline is observed in similarity scores as the window advances toward the ADD cohort. This trend aligns with both the expected trajectory of cognitive decline and established neurobiological evidence, thereby validating the proposed approach.

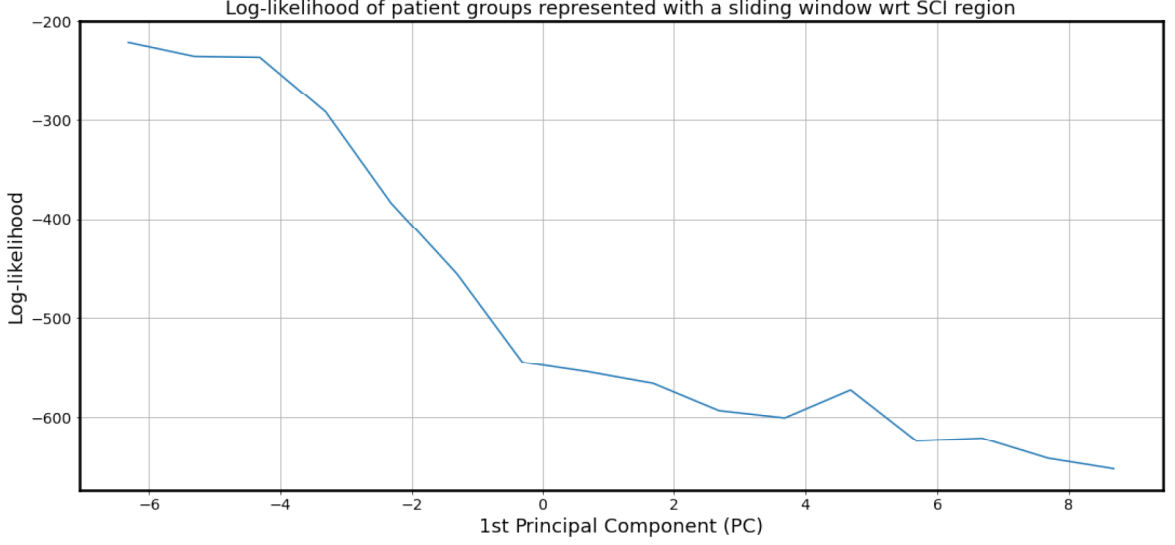


Figure 4.8. Log-likelihoods of patient samples are computed across overlapping regions along the horizontal axis of the manifold in Figure 4.1 using a sliding window. A clear decline in log-likelihood is observed from SCI to ADD, aligning with the expected clinical progression and supporting the manifold’s interpretability in reflecting disease severity.

4.5. Classification Performance and Predictive Evaluation

Building upon the manifold representations discussed in Section 4.4, it is crucial to assess the predictive power of the proposed model architecture shown in Figure 3.1. Unlike the manifold construction procedure, which utilizes the entire dataset to infer structural embeddings, the classification pipeline requires strict validation procedures to ensure generalizability. To this end, leave-one-out cross-validation (LOOCV) is employed, which is particularly suitable for small datasets, as it allows for nearly unbiased performance estimation.

Each classification experiment is repeated 10 times to mitigate the effects of stochastic variability during training. the average and standard deviation of the predictive performance across repetitions are reported. For example, in the binary classification task distinguishing ADD ($n = 18$) from SCI ($n = 24$), this procedure results in a total of $42 \times 10 = 420$ training-evaluation iterations. All experiments utilize the same

training parameters listed in Table 4.4, with LOOCV applied on top as an additional layer of validation.

While the manifold learning step utilized all three diagnostic categories (ADD, MCI, SCI) for downstream analyses, the classification tasks dealt with are strictly binary by design. Therefore, similar to how three distinct manifolds were presented in Section 4.4, three separate predictive performance results are reported corresponding to the three binary tasks: ADD vs. SCI, ADD vs. MCI, and MCI vs. SCI. This approach ensures methodological consistency and supports interpretability across both manifold and classification analyses. Notably, the models used for classification are highly comparable to those employed during manifold construction, differing only in that one additional subject may be present due to the cross-validation protocol.

Table 4.5 summarizes the predictive performance for each task, evaluated using the Receiver Operating Characteristic - Area Under Curve (ROC-AUC) metric. ROC-AUC is selected as it provides a robust evaluation of model discrimination ability, especially under class imbalance conditions. Unlike accuracy, ROC-AUC offers a more interpretable and meaningful measure in our context, where classes are often unevenly distributed.

Table 4.5. Predictive performance of the low-dimensional manifold constructor framework is evaluated across three binary tasks. The model achieves high ROC-AUC in distinguishing ADD from both MCI and SCI. However, performance is poor for MCI vs. SCI, highlighting the subtlety and difficulty of differentiating these adjacent stages of cognitive decline.

Task	ROC-AUC (mean $\pm$ std)
ADD/MCI	0.81 $\pm$ 0.025
ADD/SCI	0.93 $\pm$ 0.005
MCI/SCI	0.55 $\pm$ 0.048

As shown in Table 4.5, the model exhibits high discriminative power in the ADD vs. SCI classification task, achieving a mean ROC-AUC of 0.93, indicating a strong separation between these two diagnostic groups. The ADD vs. MCI task yields a respectable performance of 0.81 ROC-AUC. Both of these results reinforce the interpretive clarity of the corresponding manifolds shown in Figure 4.1 and 4.2.

Conversely, the MCI vs. SCI classification task results in a ROC-AUC of only 0.55, which is marginally better than chance-level performance (0.50). This outcome suggests that the model is unable to learn distinctive features separating these two groups, likely due to the greater overlap in their clinical profiles, the limited sample size, and potential diagnostic uncertainty inherent in these adjacent stages. This observation aligns with prior findings indicating that differentiating MCI from SCI is notably more complex than distinguishing ADD from either group [78, 79].

Furthermore, this poor performance casts doubt on the reliability of the MCI/SCI manifold presented in Figure 4.3, which similarly failed to capture a coherent representation of the cognitive decline trajectory. Collectively, the classification results substantiate the utility of the trained manifolds, particularly for tasks where the diagnostic categories are well-separated, and establish a quantitative basis for the downstream analyses using these manifolds throughout this thesis.

4.6. Ablation Study

The proposed architecture, as illustrated in Figure 3.1, demonstrates strong performance in various discriminative tasks, with the exception of the classification between Mild Cognitive Impairment (MCI) and Subjective Cognitive Impairment (SCI). To systematically assess the contribution of each component within the architecture, a comprehensive ablation study is conducted. Although the MCI/SCI classification task does not yield satisfactory results, its performance metrics is included in our analysis for the sake of completeness and to provide a full account of the model’s behavior across all examined scenarios.

In this ablation study, the impact of architectural variations are investigated by either entirely removing specific layers or replacing them with alternative ones. Then, how these modifications influence the model’s performance compared to the results reported in Section 4.5 are analyzed. This allows us to draw conclusions about the relative importance and utility of each component within the overall Graph Neural Network (GNN) architecture. In particular, the following variations are considered:

- The complete removal of the Graph Isomorphism Network (GIN) layer
- The complete removal of the Graph Attention Network (GAT) layer
- Replacement of the GIN layer with a Graph Convolutional Network (GCN) layer
- Replacement of the attention-based pooling mechanism with a sum pooling operation

Each of these architectural modifications is designed to isolate and quantify the individual contribution of specific components, thereby elucidating their respective roles in enhancing the overall performance of the proposed GNN framework.

The results of the ablation study are presented in Table 4.9, which provides a comparative evaluation of the model’s performance under various architectural modifications. This analysis highlights the impact of including or excluding specific layers and pooling mechanisms on the overall classification performance. The experimental protocol employed to generate these results is consistent with that described in Section 4.5, and is based on the configuration outlined in Table 4.4, utilizing a leave-one-out cross-validation (LOOCV) framework. For each classification task, the corresponding setup is further detailed in Table 4.6.

To ensure robustness and mitigate the effects of randomness, each experiment is repeated ten times, following the same procedure previously adopted. The reported results in the ablation study include both the mean and standard deviation across these repeated runs, thereby providing a reliable estimate of performance variability.

Table 4.6. Summary of Ablation Study Parameters.

Description	Value		
	ADD/SCI	ADD/MCI	MCI/SCI
Number of Repeats	10	10	10
Number of Folds	42	64	70
Batch Size	6	6	6
Number of Epochs	100	100	100
Learning Rate	0.001	0.001	0.001
Weight Decay	0.001	0.001	0.001
Optimizer	Adam	Adam	Adam

As shown in Table 4.9, none of the ablated model configurations outperform the original architecture, with the exception of the MCI/SCI classification task, which is excluded from direct comparison due to its inherently low predictive power. The results for MCI/SCI classification remain consistently close to random performance across all architectural variations, with the highest observed ROC-AUC value reaching only approximately 0.61. This value is only marginally above the theoretical baseline of 0.5, which corresponds to random guessing. Given the limited sample size available for this task, such marginal improvements cannot be considered statistically significant, nor do they suggest a meaningful learning signal. As previously noted, the MCI/SCI classification results are included in the ablation analysis solely for the sake of completeness, to provide a comprehensive evaluation of model behavior across all targeted classification tasks.

In contrast, the ADD/MCI classification task exhibits greater sensitivity to architectural changes. Among the ablated models, the configurations employing GIN+GAT with sum pooling and GIN with attention pooling yield results that are comparable to those of the original architecture. Specifically, the ADD/MCI performance for these configurations shows a difference of approximately 1 percentage point in ROC-AUC relative to the original model, indicating only a minor performance degradation. This suggests that while attention pooling and the inclusion of the GAT layer may con-

tribute incrementally, their absence does not drastically impair performance for this particular task when a GIN layer is retained.

The ADD/SCI classification task, however, reveals more substantial disparities in performance between the original and ablated models. The original architecture achieves a ROC-AUC of approximately 0.93, whereas the best-performing ablated models, GIN+GAT+Sum Pool and GIN+Attention Pool, attain ROC-AUC scores closer to 0.90. Although the gap appears modest in absolute terms, it represents a meaningful decline in classification performance within a statistical context. The decline illustrates the additive value of the combined architectural enhancements—particularly the synergy between GIN and attention mechanisms—embedded in the original model.

A closer examination of the architectural variants reveals key patterns. The three strongest configurations—namely, the original model, GIN+GAT+Sum Pool, and GIN+Attention Pool—share two structural features: the inclusion of a GIN layer and at least one attention-based component, whether it appears in the pooling strategy or as a dedicated GNN layer. In contrast, the remaining two ablated configurations either omit the GIN layer entirely or substitute it with a GCN layer. This shift results in a marked performance degradation. Specifically, removing the GIN layer reduces ROC-AUC from 0.93 to 0.84 for the ADD/SCI task and from 0.81 to 0.74 for the ADD/MCI task. Meanwhile, replacing the GIN layer with a GCN layer further exacerbates this decline, leading to ROC-AUC scores of 0.78 for ADD/SCI and only 0.62 for ADD/MCI. The latter result is particularly concerning, as it approaches the level of random classification, indicating a substantial loss of discriminative capability.

These findings underscore the critical importance of the GIN layer in achieving high classification performance, particularly for tasks with greater structural complexity such as ADD/SCI. While attention mechanisms contribute positively, their impact is less pronounced when used in isolation. Conversely, replacing or omitting the GIN layer introduces significant performance deficits, suggesting that GIN provides a representational advantage not captured by simpler convolutional mechanisms such as

GCN. Therefore, the ablation results confirm that the architectural choices embedded in the original model are not only justified but also necessary for maintaining robust predictive performance across the majority of tasks.

These results will be further examined in Chapter 5, where the observed performance differences are interpreted in light of the underlying semantics and expressiveness of various GNN layer types. This includes a deeper exploration of why certain layers, such as GIN, contribute more substantially to model performance and how architectural choices align with the structural characteristics of the input data.

4.7. Random Experiments

To assess the statistical significance and robustness of our findings, two control experiments were conducted using synthetically generated data. In the first experiment, random structural brain networks that matched the sparsity levels observed in real connectome data were generated and inputted into the trained model. Specifically, the architecture depicted in Figure 3.1 was utilized, trained on the ADD/SCI classification task, to extract graph-level representations. Following model training on the full set of Alzheimer’s Disease (ADD) and Subjective Cognitive Impairment (SCI) samples, a two-dimensional latent manifold was constructed using principal component analysis (PCA). This dimensionality reduction preserved nearly all of the variance present in the learned representations.

The rationale for selecting the ADD/SCI classification task lies in its ability to span the full spectrum of cognitive decline, making the resulting manifold particularly informative. Furthermore, the model’s strong predictive performance on this task, as reported in Sections 4.5 and 4.9, supports the validity of the learned representation space.

To evaluate how the model interprets random network structures, the synthetic graphs were processed through the trained GNN, extracted their latent representations,

and projected these into the PCA space derived from the ADD and SCI samples. As illustrated in Figure 4.9, the random networks are projected to a compact and isolated region in the manifold, occupying an area that is distinctly distant from the clusters formed by the ADD and SCI samples. This suggests that the random networks differ substantially from biologically meaningful brain graphs.

In contrast, the MCI samples, when projected onto the same manifold (see Figure 4.1), occupy an intermediate position between the ADD and SCI clusters. In several cases, MCI projections even overlap with those of ADD or SCI, indicating that their representations are structurally and topologically closer to real clinical cases than to random noise. The concentrated and isolated placement of random networks in the latent space further supports the conclusion that the topology and manifold structure captured by the model reflect meaningful neurobiological patterns, rather than being the result of random or spurious features.

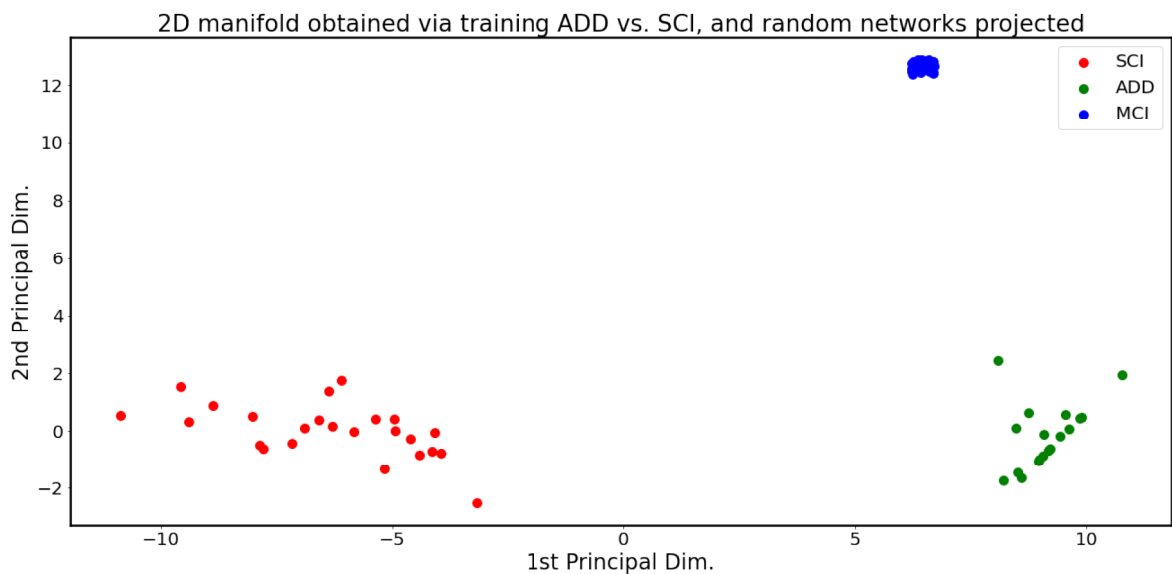


Figure 4.9. To validate the manifold’s robustness, a control experiment projects randomized brain networks—matching real data sparsity but lacking biological structure—into the PCA space. Unlike real MCI samples, these form a compact, isolated cluster, confirming that the manifold captures meaningful, disease-relevant brain organization rather than random variation.

Later, a second control experiment was conducted to further examine the structure and validity of the learned manifold. In this experiment, the model architecture presented in Figure 3.1 was trained entirely on synthetically generated random networks. These random graphs were constructed to match the sparsity levels of the real structural brain networks in order to control for basic topological properties, similar to what was done in the first control experiment above, while ensuring that no biologically meaningful information was present. The training configuration remained identical to that used in the original setup, as detailed in Table 4.4, including the use of the same number of samples and class labels. Specifically, the model was trained on an artificial ADD vs. SCI classification task, using 18 random networks labeled as ADD and 24 as SCI, mirroring the real dataset in size and class distribution.

As expected, the model failed to demonstrate any meaningful predictive capacity on this synthetic dataset. The random nature of the input data precluded the existence of discernible patterns, rendering the learning process ineffective. Notably, the model was unable even to overfit to the training data, likely due to the complete absence of structured signals. This is visually supported by the manifold illustrated in Figure 4.10, in which the latent representations of the synthetic ADD and SCI samples are densely overlapped, showing no evidence of separability or class structure. When random MCI samples—also generated independently—were subsequently projected into the same latent space, their embeddings were similarly indistinguishable from those of the synthetic ADD and SCI groups, further emphasizing the lack of any underlying manifold geometry.

To gain deeper insight into the nature of the representations learned from real versus random data, the dimensionality of their respective manifolds were analyzed by examining the variance explained by principal components. For embeddings derived from real structural brain networks, approximately 99% of the variance was captured by the first two principal components, indicating that a two-dimensional projection is sufficient to preserve the essential structural information encoded in the data. In stark contrast, embeddings obtained from random networks required at least four prin-

principal components to explain a comparable proportion of the variance. This disparity suggests that the low-dimensional structure observed in the real data is not an artifact of the dimensionality reduction technique, nor a result of arbitrary projection choices for visualization. Rather, it reflects intrinsic organizational properties of the brain connectome data that are absent in random graphs, and is meaningful in its own context.

These observations collectively validate that the manifold constructed using real brain connectome data—specifically from ADD and SCI samples—is not of random nature. Instead, it captures meaningful structural distinctions, supporting the interpretability of the learned representations and laying the groundwork for future neurological and biological investigations.

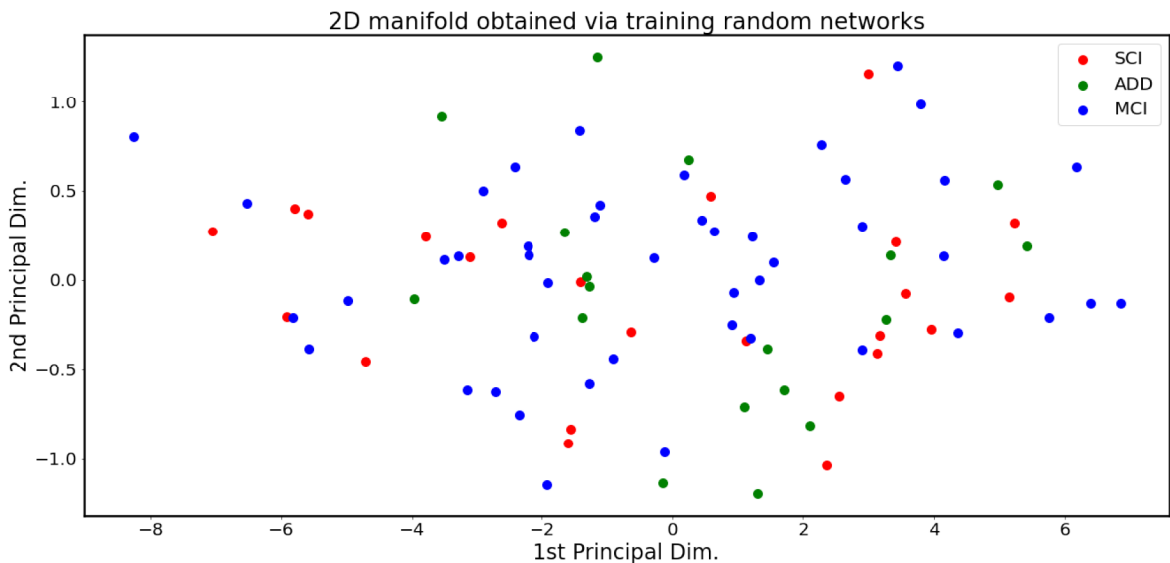


Figure 4.10. A control experiment using randomized brain networks—preserving sparsity but lacking biological structure—reveals no clear separation among ADD, MCI, and SCI in the PCA space. This contrasts with prior results and confirms that the original manifold captures meaningful, disease-relevant brain topology rather than random variation.

4.8. Architecture Search

The architectural framework illustrated in Figure 3.1 has been systematically modified to facilitate an architecture search process, specifically targeting the ADD/SCI classification task. This focus stems from the robust performance of the original model on this task. Any proposed architectural variant should demonstrate comparable performance in this setting before being extended to more challenging classification problems, such as distinguishing ADD from MCI, or MCI from SCI.

The various architectural modifications implemented are comprehensively detailed in Section 3.3.1. These alterations primarily target components surrounding the Graph Attention Layer (GAT)—including both its upstream and downstream stages—as well as the Graph Attention Readout Layer (GARO). The intent behind these modifications is to model brain networks as a form of soft clustering over cortical nodes, thereby enabling more interpretable relationships between attention coefficients, learned weights, and the resultant subnetworks. In this context, subnetworks refer to functionally or topologically cohesive clusters of nodes whose interactions contribute to classification performance.

Broadly, the modifications fall into two categories: those affecting the GAT itself and those altering the graph readout mechanism. As summarized in Table 4.7, the first three rows correspond to GAT-centric alterations, whereas the fourth row introduces a modified readout mechanism. The first three models are based on the unmodified architecture depicted in Figure 3.1, which serves as the canonical model throughout this thesis. The fifth entry presents an unmodified version of the readout-based model architecture, comprising a GIN layer followed by a GARO layer, with the remainder of the architecture consisting of standard dense layers and a softmax classifier, as is typical for classification tasks.

Table 4.7. ADD/SCI classification task: ROC-AUC performance of different model variants investigated during architectural exploration (mean $\pm$ standard deviation).

Method	ROC-AUC (mean $\pm$ std)
GIN + Subnet GAT + Attention Pool	0.90 $\pm$ 0.013
GIN + Subnet GAT with Laplacian Bias + Attention Pool	0.87 $\pm$ 0.01
GIN + Subnet GAT with Subnet Attention + Attention Pool	0.86 $\pm$ 0.01
GIN + Subnet GARO	0.84 $\pm$ 0.04
GIN + GARO	0.83 $\pm$ 0.03

The architectural exploration conducted in this section adheres to the same experimental protocol utilized in previous evaluations to ensure consistency and comparability of results. Specifically, all model variants were trained using the configuration outlined in Table 4.4, which details the learning rate, optimizer settings, batch size, and other hyperparameters. Furthermore, the evaluation framework employs a leave-one-out cross-validation strategy, repeated across ten randomized splits. This setup mirrors the methodology adopted in Sections 4.6 and 4.5, providing a robust and fair basis for performance comparison. This setup ensures a fair and reliable basis for evaluating the effects of the architectural modifications.

As observed in Table 4.7, and in comparison with the results presented in Table 4.5, the most promising architecture is the model described in the first row, which incorporates a Subnet GAT and attention pooling on top of a GIN backbone, achieving a mean ROC-AUC of 0.90. Nonetheless, this performance falls slightly short of the original model introduced in Section 3.3, which achieves a ROC-AUC of 0.93.

The second model introduces a Laplacian-based bias term into the attention mechanism of the Subnet GAT, leveraging the topological information encoded in the graph Laplacian. This approach results in a minor decrease in performance, with a ROC-AUC of 0.87, suggesting that the addition of structural bias does not enhance

predictive capability in this context. The third model employs a subnet-level attention mechanism across subgraphs of the original input, yielding a similar yet slightly lower performance of 0.86.

The final two models replace the GAT-based design with a key-query-based attention mechanism implemented through the GARO layer. The Subnet GARO variant achieves a ROC-AUC of 0.84, marginally outperforming the original GARO-based architecture (ROC-AUC of 0.83). However, both remain less effective than the GAT-based modifications, and substantially underperform relative to the original.

In summary, while these architectural modifications—particularly those introducing subnet-level semantics, attention biases, or alternative readout schemes—provide potentially valuable interpretive insights into the functional clustering and communication of cortical nodes, they do not yield improvements in classification performance. These findings suggest a trade-off between model interpretability and predictive accuracy that warrants further investigation.

4.9. Multimodal and Unimodal Experiments With Different Networks

In this section, the classification performance across various brain network types are reported: structural networks (sNets), learned functional networks (*l*Nets), and elastic networks (eNets). Detailed methodologies for constructing each network type are provided in Section 4.3. Networks were evaluated at multiple parcellation resolutions using the Schaefer atlas, ranging from 100 to 1000 nodes, in addition to the Destrieux atlas with 148 nodes.

Results for all three binary classification tasks —ADD/SCI, ADD/MCI, and MCI/SCI— are presented for completeness. They, summarized in Table 4.8, demonstrate the classification performance measured by mean ROC-AUC scores. The table reports only the mean values due to space constraints; however, the variance across trials was minimal, and thus the mean values are representative. The same experimen-

tal configuration, detailed in Table 4.4, was used throughout, employing leave-one-out cross-validation and averaging results across ten runs. Empty entries in the table indicate either unavailable data or incomplete analyses due to limited preliminary performance. Notably, experiments using structural networks are focused on node counts of $n = 148, 200, 400, 800$ due to data availability.

Table 4.8. ROC-AUC scores for three binary classification tasks (ADD/SCI, ADD/MCI, MCI/SCI) across different network types and sizes. Bold values indicate best performance at each row for the corresponding binary classification task.

Description of *l*Net and eNet are included in detail in Section 4.3.

Num. Nodes	ADD/SCI			ADD/MCI			MCI/SCI		
	sNet	<i>l</i> Net	eNet	sNet	<i>l</i> Net	eNet	sNet	<i>l</i> Net	eNet
100		0.83			0.51			0.653	
148	0.93	0.82		0.81	0.51		0.55	0.54	
200	0.74	0.791	0.614	0.771	0.536	0.58	0.455	0.67	0.54
300	0.757				0.571			0.575	
400	0.843	0.724	0.64	0.728	0.591	0.57	0.442	0.562	0.63
500	0.68				0.475			0.524	
600		0.548			0.45			0.53	
700		0.657			0.45			0.51	
800	0.816	0.584	-	0.74	0.58		0.64	0.40	0.41
900		0.48			-			0.37	
1000		0.585			-				0.61

According to Table 4.8, the most robust classification performance is achieved using sNets constructed from the Destrieux atlas ($n = 148$ nodes). This network has also been the primary configuration throughout this thesis, including its application in manifold-based modeling of cognitive decline. Across both the ADD/SCI and ADD/MCI tasks, sNets consistently outperform *l*Net and eNet counterparts. Within the Schaefer-based networks, predictive performance does not exhibit a consistent trend

with increasing resolution. For instance, in the ADD/SCI task, ROC-AUC improves from 0.74 at $n = 200$ to 0.843 at $n = 400$, but then declines and fluctuates at higher resolutions (e.g., 0.68 at $n = 500$ and 0.816 at $n = 800$). Given the growing computational complexity and potential signal noise at finer parcellations, the marginal benefits of increasing resolution appear limited. This inconsistency may be attributed to the reduced signal-to-noise ratio in high-resolution connectomes, though further validation is needed.

The *l*Net variants exhibit moderate classification performance, particularly at lower resolutions. For instance, an ROC-AUC of 0.83 is achieved at $n = 100$ for the ADD/SCI classification task, followed by an approximately monotonic decline as the number of nodes increases. Overall, higher-resolution *l*Nets do not demonstrate consistent performance improvements. Moreover, the eNet configurations—evaluated only for the Schaefer 200- and 400-node parcellations—fail to yield meaningful classification accuracy, with ROC-AUC scores generally hovering near chance level (0.5–0.6).

In the ADD/MCI classification task, sNets are again the only network type achieving notable performance. The Destrieux sNet ($n = 148$) shows the highest ROC-AUC, followed by the Schaefer 200-node variant. Performance declines at $n = 400$ and marginally improves again at $n = 800$, although it does not surpass the 148- or 200-node networks. Neither *l*Net nor eNet provide discriminative power beyond random chance in this classification task.

Results for the MCI/SCI classification task are presented for completeness, despite prior indications—discussed in Section 4.5—that this is a particularly challenging problem. As expected, none of the network types (sNet, *l*Net, or eNet) achieve statistically meaningful classification performance in this case.

Also, multi-network configurations were explored to assess whether combining diverse network types and resolutions could improve performance. Specifically, the following combinations were evaluated in a Siamese-like architecture:

- sNet ($n = 148$) + lNet ($n = 148$)
- sNet ($n = 400$) + lNet ($n = 100$)
- sNet ($n = 400$) + sNet ($n = 800$)
- sNet ($n = 800$) + lNet ($n = 100$)

These network pairs were input into a dual-branch version of the architecture shown in Figure 3.1, where the two branches are concatenated post-graph readout, as illustrated in Figure 4.11. However, none of the combined configurations resulted in performance improvements. The Destrieux-based sNets, on their own, remain the most effective choice for both ADD/SCI and ADD/MCI classification tasks.

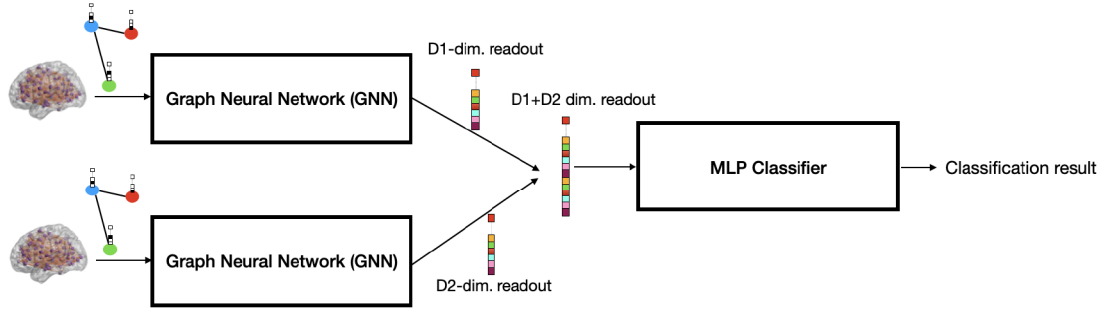


Figure 4.11. Siamese-style GNN architecture. Two parallel GNN branches—each comprising GIN, GAT, and attention pooling layers—process distinct brain networks.

Shared architecture ensures modality-invariant representation. Graph-level embeddings are concatenated and input to an MLP for binary classification.

Table 4.9. Ablation study of model components. ROC-AUC results across classification tasks using leave-one-out cross-validation, averaged over ten runs. The baseline reflects the full architecture in Figure 3.1, integrating GIN, GAT, and attention pooling. Variants omitting or altering components highlight the contribution of each module to predictive performance.

Method	Task	ROC-AUC (mean $\pm$ std)
Original Model		
GIN+GAT+Attention Pool	ADD/MCI	0.81 $\pm$ 0.025
	ADD/SCI	0.93 $\pm$ 0.005
	MCI/SCI	0.55 $\pm$ 0.0048
Ablated Models		
GIN+GAT+Sum Pool	ADD/MCI	0.80 $\pm$ 0.018
	ADD/SCI	0.90 $\pm$ 0.011
	MCI/SCI	0.55 $\pm$ 0.0057
GIN+Attention Pool	ADD/MCI	0.79 $\pm$ 0.014
	ADD/SCI	0.90 $\pm$ 0.007
	MCI/SCI	0.61 $\pm$ 0.0021
GCN+GAT+Attention Pool	ADD/MCI	0.62 $\pm$ 0.032
	ADD/SCI	0.78 $\pm$ 0.012
	MCI/SCI	0.49 $\pm$ 0.0022
GAT+Attention Pool	ADD/MCI	0.74 $\pm$ 0.010
	ADD/SCI	0.84 $\pm$ 0.016
	MCI/SCI	0.52 $\pm$ 0.0021

5. DISCUSSION

This thesis explores the capacity of graph neural networks (GNNs) to characterize the spectrum of cognitive decline in Alzheimer’s Disease Dementia (ADD) by leveraging structural, functional, and multi-modal brain connectomes. A central contribution of this research is the development of a GNN-based low-dimensional manifold construction framework, designed to capture the latent structure of neurodegenerative progression. As detailed in Section 4, this framework not only demonstrates robust predictive capabilities but also provides interpretable neurobiological insights into the cognitive decline trajectory. Conceptually, this manifold-learning tool is generalizable and holds potential applicability across different datasets and tasks that aim to uncover the underlying structure of gradual pathological transitions in the brain.

A core finding of this study is that training the proposed framework on an ADD versus Subjective Cognitive Impairment (SCI) classification task yields a two-dimensional manifold that effectively encodes the disease spectrum. As visualized in Figure 4.1, the learned manifold reveals a smooth and continuous progression from SCI to ADD. Importantly, Mild Cognitive Impairment (MCI) samples—excluded during training—naturally project onto the central region of this spectrum. This observation is particularly noteworthy, as MCI is clinically considered an intermediate stage between SCI and ADD. The unsupervised placement of MCI samples between the SCI and ADD endpoints lends further support to the interpretability and clinical validity of the learned manifold. Notably, the two latent dimensions of this manifold capture over 99% of the variance in the data, suggesting that complex structural dependencies in brain networks can be effectively summarized in a low-dimensional latent space in the context of cognitive decline.

To further test the robustness and specificity of this framework, an alternative manifold was constructed by training the model on an ADD versus MCI classification task. In this configuration, SCI samples—again withheld from training—clustered

closely with MCI samples, as shown in Figure 4.2. This pattern aligns with the known clinical relationship between SCI and MCI, where SCI is often viewed as a prodromal phase preceding MCI. The fact that SCI samples did not arbitrarily fall between ADD and MCI but instead aligned more closely with MCI supports the hypothesis that the learned manifold captures meaningful neurobiological gradients rather than reflecting superficial class separations. The model thus appears to position unseen samples not arbitrarily but in accordance with their true proximity to the trained class endpoints.

In contrast, attempts to construct a disease manifold using the MCI versus SCI classification task were unsuccessful due to poor predictive performance. These results, reported in Table 4.5, are consistent with existing literature indicating that distinguishing MCI from SCI is particularly challenging. As Bohn et al. (2025) note, the subtle clinical differences and significant overlap in features between SCI and MCI cohorts complicate accurate stratification in early-stage Alzheimer’s disease research. Our own findings corroborate this difficulty, as neither the integration of multi-modal inputs nor the use of alternate brain parcellations resulted in meaningful improvement in this task. The manifold derived from MCI–SCI classification, shown in Figure 4.7, was retained for completeness, though its clinical interpretability remains limited.

To assess the statistical significance and reliability of our findings, a series of randomization experiments were conducted. In the first such experiment, empirical brain networks were replaced with randomly generated graphs possessing similar sparsity profiles. As illustrated in Figure 4.10, the resulting manifold failed to exhibit any discernible structure: class boundaries dissolved, and samples from different groups became intermingled. Furthermore, representing 99% of the variance in these randomized networks required four dimensions—double that of real data—highlighting the inability of the model to extract low-dimensional structure from noise. This contrast underscores the nontrivial structure present in actual brain networks and supports the model’s ability to capture meaningful patterns only when grounded in biologically plausible data.

A second validation experiment involved projecting random networks onto the previously learned ADD–SCI manifold. As shown in Figure 4.9, these random samples did not fall within the disease spectrum but were instead projected far from both class-defined regions. This further confirms that the model does not simply interpolate or average unseen inputs; rather, it positions samples according to their structural coherence with learned disease patterns. Together, these control experiments support the conclusion that the discovered manifold is not an artifact of model architecture or data sparsity, but instead reflects genuine and robust neurobiological organization.

Ablation studies were conducted to systematically examine the individual contributions of key architectural components in the proposed framework. As detailed in Sections 4.5 and 4.6, the Graph Isomorphism Network (GIN) layers emerged as a particularly critical component. GINs are known for their strong representational capacity and theoretical equivalence to the Weisfeiler-Lehman (WL) graph isomorphism test, making them among the most expressive message-passing architectures currently available. Our empirical findings substantiate this theoretical strength: the removal of GIN layers from the architecture led to a significant decline in classification performance across all tested configurations. When replaced with Graph Convolutional Networks (GCNs), performance further deteriorated. This outcome aligns with well-documented limitations of GCNs, particularly their tendency toward over-smoothing, which causes the learned node representations to become indistinguishably similar across layers—an undesirable property when fine-grained structural distinctions are critical, as in brain connectome analysis.

In addition to node-level propagation layers, the impact of incorporating attention-based mechanisms at different stages of the network was evaluated. The message-passing phase employed Graph Attention Networks (GAT), which dynamically weigh neighboring node features according to learned attention coefficients. These coefficients allow the model to assign greater importance to informative neighbors while down-weighting less relevant or noisy ones. Inclusion of GAT layers consistently yielded marginal but measurable gains in classification performance, suggesting that such selec-

tive propagation improves model robustness and adaptability to individual connectomic patterns.

Moreover, the graph-level aggregation phase was also varied to evaluate different pooling strategies. Specifically, global sum pooling—a simple and permutation-invariant method that aggregates all node features equally—with global attention pooling, which learns task-specific weights for each node and performs a weighted aggregation was compared. Although sum pooling is computationally efficient and effective in some contexts, our results indicate that attention-based pooling consistently outperforms it in terms of classification accuracy. This superiority likely stems from the ability of attention mechanisms to focus on topologically or clinically salient nodes (e.g., regions most affected by neurodegeneration), rather than treating all nodes uniformly. Given that structural brain networks often exhibit heterogeneous pathological burdens across different regions, such selective aggregation is particularly advantageous.

Collectively, these ablation results underscore the importance of architectural expressivity and adaptive weighting mechanisms in modeling complex brain networks. The combination of GIN-based message passing, GAT-based local attention, and attention informed graph-level pooling provides a robust and interpretable modeling pipeline. These findings also offer practical guidance for future work aiming to apply GNNs in neuroimaging contexts: while deeper models or additional modalities may provide marginal gains, ensuring high expressivity in propagation layers and integrating biologically grounded attention mechanisms appear to yield the most substantial improvements in both performance and interpretability.

Also, the probabilistic properties of the manifold were examined to evaluate whether the spatial arrangement of samples aligns with neurobiological expectations. By computing a brain similarity likelihood between each cohort-defined region on the 2D manifold and the SCI group, a monotonic decrease in similarity from SCI to ADD was observed (Figure 4.8). This gradient of declining similarity provides strong evidence in support of the continuum hypothesis: that the cognitive decline associated

with ADD follows a gradual, progressive trajectory rather than abrupt categorical shifts. This insight opens avenues for refined disease staging and patient-specific projections. In principle, a new patient could be mapped onto the manifold to infer their pathological proximity to different disease stages, offering a compact, interpretable, and biologically grounded summary of their cognitive status.

The neurobiological plausibility of the discovered manifold was further assessed through normative brain region comparisons across clinical groups. As depicted in Figures 4.6 and 4.7, regions such as the precuneus—a core hub of the Default Mode Network (DMN)—and the bilateral superior parietal lobules—primarily associated with the Dorsal Attention Network (DAN)—exhibited pronounced structural alterations across the SCI, MCI, and ADD cohorts. The precuneus, in particular, is a well-documented marker of early Alzheimer’s-related neurodegeneration and has been consistently identified as a sensitive biomarker for structural and functional network degradation in the literature [80, 81]. Alterations in DAN-associated regions, such as the superior parietal lobule, are also of significant interest, as these areas are implicated in attentional control and may underlie early-stage attentional deficits observed in ADD [82, 83].

Beyond these core regions, additional differences were observed in the somato-motor cortex, visual networks, and frontoparietal control network. These may reflect either diffuse neurodegenerative effects or compensatory adaptations aimed at preserving function in the face of progressive damage. Importantly, the involvement of such diverse brain systems aligns with the emerging understanding that ADD pathology extends beyond traditional memory-related networks and affects multiple large-scale functional systems [84].

With regard to data modality, structural networks were emphasized throughout this thesis due to their superior performance in predictive modeling, as shown in Section 4.9. Functional connectomes, though valuable in capturing dynamic co-activation patterns, did not yield additional benefit in either unimodal or multi-modal configu-

rations in our dataset. While the architecture is fully compatible with multi-modal and Siamese extensions (see Figure 4.11), limitations in data quality and sample size precluded meaningful gains from these configurations. Further research with larger and more diverse datasets may help clarify the complementary utility of functional networks in disease manifold modeling.

Finally, our experiments suggest that brain atlas resolution plays a non-negligible role in model performance. Specifically, the Destrieux atlas, which parcellates the brain into 148 regions, provided the most effective resolution for both classification accuracy and manifold interpretability. Alternate parcellations such as the Schaefer atlas, across various resolutions, failed to outperform Destrieux in our empirical analyses. This suggests that both spatial resolution and anatomical coherence of brain parcellations are important design considerations for graph-based modeling in neuroimaging.

6. CONCLUSION

This thesis explored the application of Graph Neural Networks (GNNs) to characterize the spectrum of cognitive decline associated with Alzheimer’s Disease Dementia (ADD), utilizing structural brain connectomes. A novel GNN-based representation learning framework was introduced to construct a low-dimensional manifold that captures the complex alterations in brain connectivity across different stages of cognitive impairment.

The framework employed subject-level structural brain networks derived from diffusion MRI and was trained using a supervised classification task contrasting ADD and Subjective Cognitive Impairment (SCI). The resulting two-dimensional latent manifold captured approximately 99% of the variance in the original connectomic data. Notably, samples from the Mild Cognitive Impairment (MCI) group—excluded from training—projected onto intermediate regions of the manifold, consistent with clinical interpretations that MCI represents an intermediate stage between SCI and ADD. This spatial organization supports the validity of the manifold as a model of disease progression beyond discrete diagnostic categories.

Neurobiological assessments of the manifold revealed coherent patterns of alteration in key brain regions, including the precuneus, superior parietal lobules, somato-motor cortex, and visual networks. These findings aligned with the broader literature on neurodegenerative changes in ADD. Randomization experiments confirmed that the observed manifold structure did not arise by chance, underscoring the biological significance of the learned representations.

Architectural ablation studies confirmed the importance of specific GNN components. The use of Graph Isomorphism Network (GIN) layers proved critical for maintaining predictive performance, while attention mechanisms at both message passing and readout stages further enhanced model effectiveness. Pooling strategies also influ-

enced representation quality, highlighting the sensitivity of connectome-based models to architectural design choices.

Although derived from a limited dataset, the results strongly suggest that the complex and heterogeneous changes in brain connectomes over the course of dementia may be governed by a small number of consistent, biologically grounded patterns. The two-dimensional manifold discovered in this study provides a compact and interpretable substrate for staging dementia and offers valuable insight into the concurrent and dynamic reorganization of brain structure during disease progression. Preliminary neurological interpretations support this manifold’s clinical plausibility and encourage further investigation using larger and more diverse datasets. Future extensions may benefit from the integration of additional biomarkers, such as clinical, genetic, or physiological data, to enrich the manifold and expand its diagnostic potential.

Several key directions emerge from these findings. Enlarging the dataset to include a more representative spectrum of subjects would allow for broader validation and generalization. Incorporating longitudinal imaging data could facilitate modeling of disease trajectories over time. The inclusion of multi-modal information—including genetics, cerebrospinal fluid biomarkers, or behavioral data—may improve the interpretability and utility of the model. Finally, collaboration with clinical researchers would enable the translation of the manifold framework into diagnostic or prognostic tools with real-world applicability.

In conclusion, this thesis introduced a principled approach to modeling Alzheimer’s Disease Dementia using GNNs and low-dimensional manifold learning. The proposed framework effectively captured latent cognitive decline patterns that align with clinical observations and neurobiological expectations. These findings underscore the transformative potential of graph-based deep learning methods in advancing diagnostic capabilities and mechanistic understanding of neurodegenerative diseases.

REFERENCES

1. “2023 Alzheimer’s Disease Facts and Figures”, *Alzheimer’s & Dementia*, Vol. 19, No. 4, pp. 1598–1695, 2023.
2. Chowdhary, N., C. Barbui, K. J. Anstey, M. Kivipelto, M. Barbera, R. Peters, L. Zheng, J. Kulmala, R. Stephen, C. P. Ferri, Y. Joannette, H. Wang, A. Comas-Herrera, C. Alessi, K. Suharya (Dy), K. J. Mwangi, R. C. Petersen, A. A. Motala, S. Mendis, D. Prabhakaran, A. Bibi Mia Sorefan, A. Dias, R. Gouider, S. Shahar, K. Ashby-Mitchell, M. Prince and T. Dua, “Reducing the Risk of Cognitive Decline and Dementia: WHO Recommendations”, *Frontiers in Neurology*, Vol. 12, 2022.
3. Hanseeuw, B. J., R. A. Betensky, H. I. L. Jacobs, A. P. Schultz, J. Sepulcre, J. A. Becker, D. M. O. Cosio, M. Farrell, Y. T. Quiroz, E. C. Mormino, R. F. Buckley, K. V. Papp, R. A. Amariglio, I. Dewachter, A. Ivanoiu, W. Huijbers, T. Hedden, G. A. Marshall, J. P. Chhatwal, D. M. Rentz, R. A. Sperling and K. Johnson, “Association of Amyloid and Tau With Cognition in Preclinical Alzheimer Disease: A Longitudinal Study”, *JAMA Neurology*, Vol. 76, No. 8, pp. 915–924, 2019.
4. Mosconi, L., V. Berti, L. Glodzik, A. Pupi, S. D. Santi and M. J. de Leon, “Pre-Clinical Detection of Alzheimer’s Disease Using FDG-PET, with or without Amyloid Imaging”, *Journal of Alzheimer’s Disease*, Vol. 20, No. 3, pp. 843–854, 2010.
5. Destrieux, C., B. Fischl, A. Dale and E. Halgren, “Automatic Parcellation of Human Cortical Gyri and Sulci Using Standard Anatomical Nomenclature”, *Neuroimage*, Vol. 53, No. 1, pp. 1–15, 2010.
6. Schaefer, A., R. Kong, E. M. Gordon, T. O. Laumann, f. X.-N. Zuo, A. J. Holmes, S. B. Eickhoff and B. T. T. Yeo, “Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI”, *Cerebral Cortex*, Vol. 28, No. 9, pp. 3095–3114, 2018.

7. Tzourio-Mazoyer, N., B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer and M. Joliot, “Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcellation of the MNI MRI Single-Subject Brain”, *NeuroImage*, Vol. 15, No. 1, pp. 273–289, 2002.
8. Sporns, O., G. Tononi and R. Kötter, “The Human Connectome: A Structural Description of the Human Brain”, *PLOS Computational Biology*, Vol. 1, No. 4, p. e42, 2005.
9. Azevedo, F. A., L. R. Carvalho, L. T. Grinberg, J. M. Farfel, R. E. Ferretti, R. E. Leite, W. J. Filho, R. Lent and S. Herculano-Houzel, “Equal Numbers of Neuronal and Nonneuronal Cells Make the Human Brain an Isometrically Scaled-up Primate Brain”, *Journal of Comparative Neurology*, Vol. 513, No. 5, pp. 532–541, 2009.
10. Kleinfeld, D., A. Bharioke, P. Blinder, D. D. Bock, K. L. Briggman, D. B. Chklovskii, W. Denk, M. Helmstaedter, J. P. Kaufhold, W.-C. A. Lee, H. S. Meyer, K. D. Micheva, M. Oberlaender, S. Prohaska, R. C. Reid, S. J. Smith, S. Takemura, P. S. Tsai and B. Sakmann, “Large-scale Automated Histology in the Pursuit of Connectomes”, *Journal of Neuroscience*, Vol. 31, No. 45, pp. 16125–16138, 2011.
11. Vogelstein, J. T., E. Perlman, B. Falk, A. Baden, W. Gray Roncal, V. Chandrashekar, F. Collman, S. Seshamani, J. L. Patsolic, K. Lillaney, M. Kazhdan, R. Hider, Jr, D. Pryor, J. Matelsky, T. Gion, P. Manavalan, B. Wester, M. Chevillet, E. T. Trautman, K. Khairy, E. Bridgeford, D. M. Kleissas, D. J. Tward, A. K. Crow, B. Hsueh, M. A. Wright, M. I. Miller, S. J. Smith, R. J. Vogelstein, K. Deisseroth and R. Burns, “A Community-Developed Open-Source Computational Ecosystem for Big Neuro Data”, *Nature Methods*, Vol. 15, No. 11, pp. 846–847, 2018.
12. Van Essen, D. C., S. M. Smith, D. M. Barch, T. E. Behrens, E. Yacoub and K. Ugurbil, “The WU-Minn Human Connectome Project: An Overview”, *NeuroImage*, Vol. 80, pp. 62–79, 2013.

13. Brodmann, K., *Vergleichende Lokalisationslehre der Grosshirnrinde: in ihren Prinzipien dargestellt auf Grund des Zellenbaues*, 2nd unveränderte Auflage, Barth, Leipzig, 1925.
14. Fan, L., H. Li, J. Zhuo, Y. Zhang, J. Wang, L. Chen, Z. Yang, C. Chu, S. Xie, A. R. Laird, P. T. Fox, S. B. Eickhoff, C. Yu and T. Jiang, “The Human Brainnetome Atlas: A New Brain Atlas Based on Connectional Architecture”, *Cerebral Cortex*, Vol. 26, No. 8, pp. 3508–3526, 2016.
15. Bullmore, E. and O. Sporns, “Complex Brain Networks: Graph Theoretical Analysis of Structural and Functional Systems”, *Nature Reviews Neuroscience*, Vol. 10, No. 3, pp. 186–198, 2009.
16. Buzsáki, G., *Rhythms of the Brain*, Oxford University Press, New York, USA, 2006.
17. Wu, K., Y. Taki, K. Sato, H. Qi, R. Kawashima and H. Fukuda, “A Longitudinal Study of Structural Brain Network Changes with Normal Aging”, *Frontiers in Human Neuroscience*, Vol. 7, p. 113, 2013.
18. Tijms, B., A. Wink, W. de Haan, W. van der Flier, C. Stam, P. Scheltens and F. Barkhof, “Alzheimer’s Disease: Connecting Findings from Graph Theoretical Studies of Brain Networks”, *Neurobiology of Aging*, Vol. 34, No. 8, pp. 2023–2036, 2013.
19. He, Y., Z. Chen and A. Evans, “Structural Insights into Aberrant Topological Patterns of Large-Scale Cortical Networks in Alzheimer’s Disease”, *Journal of Neuroscience*, Vol. 28, No. 18, pp. 4756–4766, 2008.
20. Yao, Z., Y. Zhang, L. Lin, Y. Zhou, C. Xu, T. Jiang and the Alzheimer’s Disease Neuroimaging Initiative, “Abnormal Cortical Networks in Mild Cognitive Impairment and Alzheimer’s Disease”, *PLOS Computational Biology*, Vol. 6, No. 11, pp.

- 1–11, 2010.
21. Supekar, K., V. Menon, D. Rubin, M. Musen and M. D. Greicius, “Network Analysis of Intrinsic Functional Brain Connectivity in Alzheimer’s Disease”, *PLOS Computational Biology*, Vol. 4, No. 6, pp. 1–11, 2008.
 22. Parisot, S., S. I. Ktena, E. Ferrante, M. Lee, R. Guerrero, B. Glocker and D. Rueckert, “Disease Prediction Using Graph Convolutional Networks: Application to Autism Spectrum Disorder and Alzheimer’s Disease”, *Medical Image Analysis*, Vol. 48, pp. 117–130, 2018.
 23. Kazi, A., S. Shekarforoush, S. Arvind Krishna, H. Burwinkel, G. Vivar, K. Kortüm, S.-A. Ahmadi, S. Albarqouni and N. Navab, “InceptionGCN: Receptive Field Aware Graph Convolutional Network for Disease Prediction”, *Information Processing in Medical Imaging*, pp. 73–85, Hong Kong, China, 2019.
 24. Ghorbani, M., A. Kazi, M. Soleymani Baghshah, H. R. Rabiee and N. Navab, “RA-GCN: Graph Convolutional Network for Disease Prediction Problems with Imbalanced Data”, *Medical Image Analysis*, Vol. 75, p. 102272, 2022.
 25. Ktena, S. I., S. Parisot, E. Ferrante, M. Rajchl, M. Lee, B. Glocker and D. Rueckert, “Metric Learning with Spectral Graph Convolutions on Brain Connectivity Networks”, *NeuroImage*, Vol. 169, pp. 431–442, 2018.
 26. Zhang, H., R. Song, L. Wang, L. Zhang, D. Wang, C. Wang and W. Zhang, “Classification of Brain Disorders in rs-fMRI via Local-to-Global Graph Neural Networks”, *IEEE Transactions on Medical Imaging*, Vol. 42, No. 2, pp. 444–455, 2023.
 27. Zhang, S., X. Chen, X. Shen, B. Ren, Z. Yu, H. Yang, X. Jiang, D. Shen, Y. Zhou and X.-Y. Zhang, “A-GCL: Adversarial Graph Contrastive Learning for fMRI Analysis to Diagnose Neurodevelopmental Disorders”, *Medical Image Analysis*,

Vol. 90, p. 102932, 2023.

28. Wen, G., P. Cao, L. Liu, J. Yang, X. Zhang, F. Wang and O. R. Zaiane, “Graph Self-Supervised Learning With Application to Brain Networks Analysis”, *IEEE Journal of Biomedical and Health Informatics*, Vol. 27, No. 8, pp. 4154–4165, 2023.
29. Liu, M., H. Zhang, F. Shi and D. Shen, “Hierarchical Graph Convolutional Network Built by Multiscale Atlases for Brain Disorder Diagnosis Using Functional Connectivity”, *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 35, No. 11, pp. 15182–15194, 2024.
30. Li, N., J. Xiao, N. Mao, D. Cheng, X. Chen, F. Zhao and Z. Shi, “Joint Learning of Multi-Level Dynamic Brain Networks for Autism Spectrum Disorder Diagnosis”, *Computers in Biology and Medicine*, Vol. 171, No. C, p. 108054, 2024.
31. Zhao, K., B. Duka, H. Xie, D. J. Oathes, V. Calhoun and Y. Zhang, “A Dynamic Graph Convolutional Neural Network Framework Reveals New Insights into Connectome Dysfunctions in ADHD”, *NeuroImage*, Vol. 246, p. 118774, 2022.
32. Huang, J., L. Zhou, L. Wang and D. Zhang, “Attention-Diffusion-Bilinear Neural Network for Brain Network Analysis”, *IEEE Transactions on Medical Imaging*, Vol. 39, No. 7, pp. 2541–2552, 2020.
33. Xia, J., Y. H. Chan, D. Girish and J. C. Rajapakse, “IMG-GCN: Interpretable Modularity-Guided Structure-Function Interactions Learning for Brain Cognition and Disorder Analysis”, *International Conference on Medical Image Computing and Computer Assisted Intervention*, p. 470–480, Marrakesh, Morocco, 2024.
34. Cui, H., W. Dai, Y. Zhu, X. Li, L. He and C. Yang, “Interpretable Graph Neural Networks for Connectome-Based Brain Disorder Analysis”, *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 324–334, Singapore, 2022.

35. Liu, L., G. Wen, P. Cao, T. Hong, J. Yang, X. Zhang and O. R. Zaiane, “BrainTGL: A Dynamic Graph Representation Learning Model for Brain Network Analysis”, *Computers in Biology and Medicine*, Vol. 153, No. C, p. 106521, 2023.
36. Zheng, K., S. Yu and B. Chen, “CI-GNN: A Granger Causality-Inspired Graph Neural Network for Interpretable Brain Network-Based Psychiatric Diagnosis”, *Neural Netw.*, Vol. 172, No. 106147, p. 106147, 2024.
37. Yu, R., C. Pan, X. Fei, M. Chen and D. Shen, “Multi-Graph Attention Networks with Bilinear Convolution for Diagnosis of Schizophrenia”, *IEEE Journal of Biomedical and Health Informatics*, Vol. 27, No. 3, pp. 1443–1454, 2023.
38. Zhang, L., A. Zaman, L. Wang, J. Yan and D. Zhu, “A Cascaded Multi-modality Analysis in Mild Cognitive Impairment”, *International Conference on Medical Image Computing and Computer Assisted Intervention*, p. 1–9, Shenzhen, China, 2019.
39. Xing, X., Q. Li, M. Yuan, H. Wei, Z. Xue, T. Wang, F. Shi and D. Shen, “DS-GCNs: Connectome Classification using Dynamic Spectral Graph Convolution Networks with Assistant Task Training”, *Cerebral Cortex*, Vol. 31, No. 2, pp. 1259–1269, 2020.
40. Chen, D. and L. Zhang, “FE-STGNN: Spatio-Temporal Graph Neural Network with Functional and Effective Connectivity Fusion for MCI Diagnosis”, *International Conference on Medical Image Computing and Computer Assisted Intervention*, p. 67–76, Vancouver, BC, Canada, 2023.
41. Ma, Y., W. Cui, J. Liu, Y. Guo, H. Chen and Y. Li, “A Multi-Graph Cross-Attention-Based Region-Aware Feature Fusion Network Using Multi-Template for Brain Disorder Diagnosis”, *IEEE Transactions on Medical Imaging*, Vol. 43, No. 3, pp. 1045–1059, 2024.

42. Park, J., Y. Hwang, M. Kim, M. K. Chung, G. Wu and W. H. Kim, “Convolving Directed Graph Edges via Hodge Laplacian for Brain Network Analysis”, *International Conference on Medical Image Computing and Computer Assisted Intervention*, p. 789–799, Vancouver, BC, Canada, 2023.
43. Xu, J., Q. Bian, X. Li, A. Zhang, Y. Ke, M. Qiao, W. Zhang, W. Khang Jeremy Sim and B. Gulyás, “Contrastive Graph Pooling for Explainable Classification of Brain Networks”, *IEEE Transactions on Medical Imaging*, Vol. 43, No. 9, pp. 3292–3305, 2024.
44. Li, Y., J. Liu, Y. Jiang, Y. Liu and B. Lei, “Virtual Adversarial Training-Based Deep Feature Aggregation Network From Dynamic Effective Connectivity for MCI Identification”, *IEEE Trans. Medical Imaging*, Vol. 41, No. 1, pp. 237–251, 2022.
45. Jiang, Y., Z. He, Z. Peng and Y. Yuan, “Hierarchical Graph Learning with Small-World Brain Connectomes for Cognitive Prediction”, *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 306 – 316, Marrakesh, Morocco, 2024.
46. Qu, G., A. Orlichenko, J. Wang, G. Zhang, L. Xiao, K. Zhang, T. W. Wilson, J. M. Stephen, V. D. Calhoun and Y.-P. Wang, “Interpretable Cognitive Ability Prediction: A Comprehensive Gated Graph Transformer Framework for Analyzing Functional Brain Networks”, *IEEE Transactions on Medical Imaging*, Vol. 43, No. 4, pp. 1568–1578, 2024.
47. Huang, J., M. K. Chung and A. Qiu, “Heterogeneous Graph Convolutional Neural Network via Hodge-Laplacian for Brain Functional Data”, *Information Processing in Medical Imaging*, pp. 278–290, Springer Nature Switzerland, Cham, 2023.
48. Huang, J., N. Chen and A. Qiu, “Topological Cycle Graph Attention Network for Brain Functional Connectivity”, *Proceedings of International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 723 – 732,

Marrakesh, Morocco, 2024.

49. Kim, B.-H., J. C. Ye and J.-J. Kim, “Learning Dynamic Graph Representation of Brain Connectome with Spatio-Temporal Attention”, *Proceedings of the 35th Conference on Neural Information Processing Systems*, pp. 4314–4327, Red Hook, NY, USA, 2021.
50. Azevedo, T., A. Campbell, R. Romero-Garcia, L. Passamonti, R. A. Bethlehem, P. Liò and N. Toschi, “A Deep Graph Neural Network Architecture for Modelling Spatio-Temporal Dynamics in Resting-State Functional MRI Data”, *Medical Image Analysis*, Vol. 79, p. 102471, 2022.
51. Sivgin, I., H. A. Bedel, S. Ozturk and T. Cukur, “A Plug-in Graph Neural Network to Boost Temporal Sensitivity in fMRI Analysis”, *IEEE Journal of Biomedical and Health Informatics*, Vol. 28, No. 9, pp. 5323–5334, 2024.
52. Akgüller, O., M. Balcı and G. Cioca, “Information Geometry and Manifold Learning: A Novel Framework for Analyzing Alzheimer’s Disease MRI Data”, *Diagnostics*, Vol. 15, No. 2, p. 153, 2025.
53. Jiang, H., P. Cao, M. Xu, J. Yang and O. Zaiane, “Hi-GCN: A Hierarchical Graph Convolution Network for Graph Embedding Learning of Brain Network and Brain Disorders Prediction”, *Computers in Biology and Medicine*, Vol. 127, p. 104096, 2020.
54. Li, Y., G. Mateos and Z. Zhang, “Learning to Model the Relationship Between Brain Structural and Functional Connectomes”, *IEEE Transactions on Signal and Information Processing over Networks*, Vol. 8, pp. 830–843, 2022.
55. Gamgam, G., Z. Yıldırım, A. Kabakçioğlu, H. Gurvit, T. Demiralp and B. Acar, “Siamese Graph Convolutional Network Quantifies Increasing Structure-Function Discrepancy over the Cognitive Decline Continuum”, *Computer Methods and Pro-*

- grams in Biomedicine*, Vol. 254, p. 108290, 2024.
56. Ding, J.-E., D. Luo, A. Zilverstand, K. Kulkarni and F. Liu, “NeuroTree: Hierarchical Functional Brain Pathway Decoding for Mental Health Disorders”, arXiv:2502.18786v3, 2025.
 57. Bloemheugel, S., J. van den Hoogen and M. Atzmueller, “Graph Construction on Complex Spatiotemporal Data for Enhancing Graph Neural Network-Based Approaches”, *International Journal of Data Science and Analytics*, Vol. 18, No. 2, pp. 1–18, 2023.
 58. Bas, E., *Extracting Structural Information on Manifolds from High Dimensional Data and Connectivity Analysis of Curvilinear Structures in 3D Biomedical Images*, Ph.D. Thesis, Northeastern University, 2011.
 59. Xu, K., W. Hu, J. Leskovec and S. Jegelka, “How Powerful are Graph Neural Networks?”, *International Conference on Learning Representations*, New Orleans, USA, 2019.
 60. Kipf, T. N. and M. Welling, “Semi-Supervised Classification with Graph Convolutional Networks”, *International Conference on Learning Representations*, Toulon, France, 2017.
 61. Veličković, P., G. Cucurull, A. Casanova, A. Romero, P. Liò and Y. Bengio, “Graph Attention Networks”, *International Conference on Learning Representations*, Vancouver, Canada, 2018.
 62. Gilmer, J., S. S. Schoenholz, P. F. Riley, O. Vinyals and G. E. Dahl, “Neural Message Passing for Quantum Chemistry”, *Proceedings of the 34th International Conference on Machine Learning*, pp. 1263–1272, Sydney, Australia, 2017.
 63. Rumelhart, D. E., G. E. Hinton and R. J. Williams, “Learning Representations by Back-propagating Errors”, *Nature*, Vol. 323, No. 6088, p. 533–536, 1986.

64. Hornik, K., M. Stinchcombe and H. White, “Multilayer Feedforward Networks are Universal Approximators”, *Neural Networks*, Vol. 2, No. 5, pp. 359–366, 1989.
65. Lecun, Y., L. Bottou, Y. Bengio and P. Haffner, “Gradient-Based Learning Applied to Document Recognition”, *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278–2324, 1998.
66. Spielman, D. A., “Spectral Graph Theory and its Applications”, *48th Annual IEEE Symposium on Foundations of Computer Science*, pp. 29–38, Providence, RI, USA, 2007.
67. Vinyals, O., S. Bengio and M. Kudlur, “Order Matters: Sequence to Sequence for Sets”, *4th International Conference on Learning Representations*, San Juan, Puerto Rico, 2016.
68. Zhang, M., Z. Cui, M. Neumann and Y. Chen, “An end-to-end deep learning architecture for graph classification”, *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, pp. 4438–4445, New Orleans, Louisiana, USA, 2018.
69. Li, Y., D. Tarlow, M. Brockschmidt and R. S. Zemel, “Gated Graph Sequence Neural Networks”, *4th International Conference on Learning Representations*, San Juan, Puerto Rico, 2016.
70. Vaswani, A., N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. u. Kaiser and I. Polosukhin, “Attention is All You Need”, *31st Conference on Neural Information Processing Systems*, p. 5998–6008, California, USA, 2017.
71. 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: The Journal of the Alzheimer’s Association*, Vol. 7, No. 3, pp. 263–269, 2011.
72. Dick, A. S. and P. Tremblay, “Beyond the Arcuate Fasciculus: Consensus and Controversy in the Connectional Anatomy of Language”, *Brain*, Vol. 135, No. 12, pp. 3529–3550, 2012.
73. Moyer, D., B. A. Gutman, J. Faskowitz, N. Jahanshad and P. M. Thompson, “Continuous Representations of Brain Connectivity Using Spatial Point Processes”, *Medical Image Analysis*, Vol. 41, pp. 32–39, 2017.
74. Durusoy, G., Z. Yıldırım, D. Y. Dal, C. Ulasoglu-Yildiz, E. Kurt, G. Bayır, E. Özacar, E. Özarslan, A. Demirtaş-Tatlıdede, B. Bilgiç, T. Demiralp, H. Gurvit, A. Kabakçioğlu and B. Acar, “B-Tensor: Brain Connectome Tensor Factorization for Alzheimer’s Disease”, *IEEE Journal of Biomedical and Health Informatics*, Vol. 25, No. 5, pp. 1591–1600, 2021.
75. Kalofolias, V., “How to Learn a Graph from Smooth Signals”, *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, pp. 920–929, Cadiz, Spain, 2016.
76. Yuksel Dal, D., Z. Yıldırım, H. Gurvit, A. Kabakçioğlu and B. Acar, “Reorganization of Brain Connectivity Across the Spectrum of Clinical Cognitive Decline”, *Neurological Sciences*, Vol. 45, No. 12, p. 5719–5730, 2024.
77. Parker, A. F., C. M. Smart, V. Scarapicchia and J. R. G. and, “Identification of Earlier Biomarkers for Alzheimer’s Disease: A Multimodal Neuroimaging Study of Individuals with Subjective Cognitive Decline”, *Journal of Alzheimer’s Disease*, Vol. 77, No. 3, pp. 1067–1076, 2020.

78. van der Flier, W. M. and P. Scheltens, “Amsterdam Dementia Cohort: Performing Research to Optimize Care”, *Journal of Alzheimer’s Disease*, Vol. 62, No. 3, pp. 1091–1111, 2018.
79. Petrazzuoli, F., S. Vestberg, P. Midlöv, H. Thulesius, E. Stomrud and S. Palmqvist, “Brief Cognitive Tests Used in Primary Care cannot Accurately Differentiate Mild Cognitive Impairment from Subjective Cognitive Decline”, *Journal of Alzheimer’s Disease*, Vol. 75, No. 4, pp. 1191–1201, 2020.
80. Hari, E., G. Kizilates-Evin, E. Kurt, A. Bayram, C. Ulasoglu-Yildiz, H. Gurvit and T. Demiralp, “Functional and Structural Connectivity in the Papez circuit in Different Stages of Alzheimer’s Disease”, *Clinical Neurophysiology*, Vol. 153, pp. 33–45, 2023.
81. Villemagne, V. L., K. E. Pike, G. Chételat, K. A. Ellis, R. S. Mulligan, P. Bourgeat, U. Ackermann, G. Jones, C. Szoek, O. Salvado, R. Martins, G. O’Keefe, C. A. Mathis, W. E. Klunk, D. Ames, C. L. Masters and C. C. Rowe, “Longitudinal assessment of A and cognition in aging and Alzheimer disease”, *Annals of Neurology*, Vol. 69, No. 1, pp. 181–192, 2011.
82. Braak, H. and E. Braak, “Neuropathological Staging of Alzheimer-Related Changes”, *Acta Neuropathologica*, Vol. 82, No. 4, pp. 239–259, 1991.
83. Mesulam, M.-M., *Principles of Behavioral and Cognitive Neurology*, Oxford University Press, New York, NY, US, 2000.
84. Seeley, W. W., R. K. Crawford, J. Zhou, B. L. Miller and M. D. Greicius, “Neurodegenerative Diseases Target Large-Scale Human Brain Networks”, *Neuron*, Vol. 62, No. 1, pp. 42–52, 2009.
85. Thapaliya, B., R. Miller, J. Chen, Y. P. Wang, E. Akbas, R. Sapkota, B. Ray, P. Suresh, S. Ghimire, V. D. Calhoun and J. Liu, “DSAM: A Deep Learning

Framework for Analyzing Temporal and Spatial Dynamics in Brain Networks”, *Medical Image Analysis*, Vol. 101, p. 103462, 2025.

86. Fah, H., L. Bohn, R. Greiner and R. A. Dixon, “Comparing Machine Learning Classifier Models in Discriminating Cognitively Unimpaired Older Adults from Three Clinical Cohorts in the Alzheimer’s Disease Spectrum: Demonstration Analyses in the COMPASS-ND study”, *Frontiers in Aging Neuroscience*, Vol. 17, 2025.

