Politecnico di Torino

Ingegneria Informatica (Computer Engineering)

Graduate session of March 2026

Multimodal Contrastive Geometric Deep Learning for Parkinson's Disease Assessment

Supervisors: Daniele Apiletti, Simone Monaco

Candidate: Emanuele Barbieri

Acknowledgements

I am profoundly grateful to the Graduation Examining Committee for evaluating my work, to the Database and Data Mining Group (DBMG) for providing a stimulating research environment, and to the following individuals for their personal impact:

My Supervisors
Daniele Apiletti
Simone Monaco
For their invaluable guidance, expertise, and unwavering support.

My Family
Jacopo Barbieri
Valerio Blaskovic
Chiara Parrotto
(through whom I extend my gratitude to the Parrotto family)
Sofia Davi
Leonilde Callocchia
Vincenzo Barbieri
Maria Curitti
Claudio Callocchia
Maria Luigia Callocchia
Bojan Blaskovic
For their unconditional love, patience, and endless encouragement throughout my academic career, and for being the constant pillars of strength who believed in me every step of the way.

My Friends
Giulio Delle Fratte
Andrea De Matteis
(through whom I extend my gratitude to the De Matteis family)
Reem Khattar
Roberto Cavallero
Gabriele Cassinelli
Mario Attrovio
Chiara De Rinaldis
For sharing the challenges of academic life and the joy of our time together.

To all of you, I dedicate these closing words. While they capture the essence of my research, I hope their truth resonates in every aspect of your own lives:

"Nothing in life is to be feared, it is only to be understood.
Now is the time to understand more, so that we may fear less."
— Marie Curie

Abstract

PD is one of the fastest-growing neurological diseases, with projections estimating 25.2 million people living with it by 2025 (a 112% increase from 2021) mainly due to the aging of the world's population. Diagnosing PD at early stages is crucial to administer targeted treatment to patients and deep learning can aid PD diagnosis. However, SOTA models often fail to merge multiple neuroimaging modalities or to compensate for the absence of these in patient data. Furthermore, existing severity and progression prediction methods often do not take full advantage of data-rich neuroimaging modalities, favoring the analysis of peripheral symptoms such as speech, movement, and handwriting. While useful for prediction, these symptomatic markers fail to provide an understanding of the underlying brain structures and functions that drive PD and the worsening of its conditions.

To address these limitations, this thesis proposes a method to combine four different neuroimaging modalities (MRI, fMRI, DTI and SPECT) from the PPMI dataset by representing them in a common embedding space using Graph Neural Networks. We develop methods to construct an embedding space using contrastive learning and leverage generative AI to enrich patient data which often has missing modalities. We then detail techniques to analyze this constructed embedding space to diagnose PD and predict its severity and progression.

Our model demonstrates a $\sim 9\%$ improvement in diagnosis accuracy, and matches the $R^2 = 0.45$ in MDS-UPDRS Part III score prediction compared to SOTA methods. Furthermore, it establishes a novel baseline of $R^2 = 0.31$ for predicting PD progression for MDS-UPDRS Part III through geometric deep learning, leveraging exclusively neuroimaging modalities. Finally, recognizing that AI-driven neurodiagnostics must implement explainability to progress medical research, our proposed method features a technique to trace model outputs back to relevant brain substructures, which can be composed of multiple structurally or functionally connected regions. The results suggest that our model provides an accurate way to quantify PD severity and predict progression while providing useful biological insights, bridging the gap between "black-box" AI and actionable medical research.

Keywords: PD, Neuroimaging, Graph Neural Networks, Contrastive Learning, Generative AI, Diagnosis, Severity, Progression, Explainability.

Contents

List of Figures

List of Tables

Glossary

***Movement Disorder Society-Unified Parkinson's Disease Rating Scale* (MDS-UPDRS)** Scale developed to evaluate and score various aspects of Parkinson's disease. ii, v–vii, 1, 4, 6, 17, 22, 32, 44–46, 49, 58, 59, 62, 67–69, 73

***Substantia Nigra Pars Compacta* (SNpc)** Crucial region of the midbrain (or mesencephalon) rich in dopaminergic neurons that regulate movement, learning, emotion, and reward system. 1, 5, 14, 28

b-values Parameter representing the strength and timing of the diffusion-sensitizing gradients in DTI. 30

b-vecs Vectors representing the spatial direction of the diffusion measurement in DTI. 30

Biomarker Measurable biological trait which indicates a physiological or pathological condition. vi, 6, 16, 17, 21, 48, 49, 62, 66, 72, 74

Blood Oxygen Level-Dependent (BOLD) Method used in fMRI to observe blood-oxygenation-level in different areas of the brain. 2, 13, 31, 33

Bradykinesia Slowness of movement that makes daily actions such as walking or speaking challenging; typical symptom of PD. 1

Brain Atlas 3D volume of the brain that defines shape and location of brain ROIs in a common coordinate space. 2, 5, 14, 15, 23, 26

Desikan-Killiany (DK) Atlas A standard anatomical parcellation atlas mapping the human cerebral cortex, commonly used in FreeSurfer pipelines. 74

Diffusion Tensor Imaging (DTI) Neuroimaging technique used to measure structural integrity of the brain. ii, v–vii, 2, 6, 13–16, 21, 22, 28–33, 35–39, 42, 43, 50, 60, 65, 67, 69, 70, 72, 74

Fractional Anisotropy (FA) Value from 0 to 1 which indicates health and density of white matter tracts. 13, 29–31

Functional Connectivity (FC) Measures of synchronized activity between different ROIs of the brain. v, 2, 15, 16, 22, 24–27, 33, 72

functional MRI (fMRI) Neuroimaging technique to measure functional relationship between ROIs of the brain. ii, v–vii, 2, 6, 13–16, 21, 22, 24–26, 28–31, 33, 35–39, 42, 43, 50, 60, 62, 63, 65, 66, 68, 70, 72, 74

Generative Reconstruction Machine learning technique which aims to reconstruct missing or partial samples from available data in a shared latent space. vi, 20, 35, 51, 61

Graph Attention Network (GAT) Specific type of GNN which implements an attention mechanism to weigh neighbor importance. 8–10

Graph Convolutional Network (GCN) Specific type of GNN that uses graph convolution operations and symmetric normalization. 8, 9, 36, 37, 50

Graph Isomorphism Network (GIN) Specific type of GNN designed to maximize expressiveness and distinguish between different graph structures. 8, 12, 37, 50

Graph Neural Network (GNN) A class of deep neural networks specialized in the analysis of graph structures which consists of nodes and edges. ii, 1, 3, 5–12, 15–17, 31, 72–76

Graph Recurrent Network (GRN) Specific type of GNN that incorporates recurrent neural networks to capture temporal dependencies. 8, 10, 11

InfoNCE Loss A specific contrastive loss function that optimizes the similarity between positive pairs while pushing away negative samples within a batch. 18–20

Kullback-Leibler (KL) Divergence A statistical measure of how one probability distribution differs from a reference distribution, heavily utilized as a regularizer in VAEs. 20, 43, 56, 57

Magnetic Resonance Imaging (MRI) Neuroimaging technique used to detail brain anatomical structure. ii, v–vii, 2, 6, 13–17, 21–25, 31–33, 35–37, 42, 43, 47, 48, 50, 60, 62, 66, 68, 71, 73, 74

Masked Variational Autoencoder (MVAE) A generative architecture that learns to reconstruct missing modalities by predicting their embeddings from a shared latent space. 20

Mean Diffusivity (MD) Average rate of water movement in the brain reflecting tissue structural integrity. 14, 29

Message Passing (Neighborhood Aggregation) Method through which a node in a GNN aggregates information from its neighbors. 3, 7–12, 16, 36, 37

MNI152 Template A standard brain template created by the Montreal Neurological Institute by averaging 152 healthy structural MRI scans, used as a universal coordinate space. 25

Motor Cortex Brain region in the frontal lobe which controls voluntary muscle movement. 14, 62, 63, 65

Neuroimaging Field of neuroscience that uses medical imaging to visualize structure and activity of the brain. ii, 1, 2, 4–7, 13, 14, 16, 17, 21–23, 32, 49, 50, 58, 72, 73, 76

Over-Smoothing Issue of GNNs that occurs when message passing mechanism aggregates neighbors too much making all nodes of the graph nearly identical. 3, 7, 10

Over-Squashing Issue of GNNs that occurs when edges of the graph compress node information into fixed-size vectors that are not large enough to preserve critical information. 3, 7, 8

Parkinson's Disease (PD) A progressive neurodegenerative disorder affecting movement due to the lack of dopamine in the brain. ii, vi, 1–6, 14, 16, 17, 21, 31, 32, 44, 62–67, 72, 74, 75

Parkinson's Progression Markers Initiative (PPMI) One of the most important longitudinal studies on Parkinson's, aimed at identifying progression biomarkers. ii, 5, 6, 21–23, 25, 28, 74

Polygenic Risk Score (PRS) Calculated estimate of a patient's risk to a certain disease based on their genetic makeup. 17

Radiotracer Chemical substance containing radioactive atoms used to trace the path of a biological process by leveraging radiation emission of the substance itself. 2, 14, 22

Region of Interest (ROI) Specific defined area within the brain defined to analyze localized neural activity relating to specific tasks or functions. vi, 2, 3, 5, 14–17, 23–28, 30, 33, 48–50, 66, 68, 69, 71, 74

Repetition Time (TR) Time between two successive radiofrequency pulses applied to the same slice in an MRI scanner. 26

Rician Noise Specific type of signal-dependent noise that follows a Rician distribution, common in magnitude MRI images. 25

Single-Photon Emission Computed Tomography (SPECT) Neuroimaging technique used to measure dopaminergic neuron density and health. ii, v–vii, 2, 6, 14–16, 21, 22, 31–33, 35–37, 42, 43, 50, 60, 66, 67, 69, 71, 73, 74

Striatal Binding Ratio (SBR) Quantitative metric from SPECT imaging that measures dopaminergic neuron density in the striatum. 22, 31

Striatum Crucial subcortical brain structure that regulates voluntary movement and action selection. 14, 22

Structural Connectivity (SC) Set of white matter tracts that link different ROIs of the brain. v, 2, 15, 16, 22, 28, 30, 33, 72

Symmetric Diffeomorphic Normalization with Rigid-Affine initialization (SyNRA) Advanced non-linear registration method to align brain images to a standard template. 25, 30

Tractography Technique used to map brain's white matter fiber pathways creating 3D tractograms. 2, 14, 15, 28

Under-Reaching Issue of GNNs that occurs when it is not deep enough to aggregate messages from distant neighbors that contain critical information. 3, 7, 10

Variational Autoencoder (VAE) Class of generative models that learn to encode data into a continuous probabilistic distribution rather than fixed vectors. 19, 20

1 Introduction

1.1 Background

1.1.1 Neurodegenerative Diseases: Parkinson's

Parkinson's Disease (PD) [1, 2] is a neurodegenerative disorder characterized by the death of dopaminergic neurons within the *Substantia Nigra Pars Compacta* (SNpc). This region of the brain is also known as black matter and it plays a crucial role in the regulation of movement, learning, emotion, and motivation. PD is one of the fastest-growing neurological diseases, with projections estimating 25.2 million people living with it by 2025 (a 112% increase from 2021) mainly due to the aging of the world's population [3].

Diagnosis and severity evaluations consist mainly of physical examinations where Bradykinesia (slowness of movement), rigidity, and resting tremors may indicate that a patient is suffering from PD. Patients are administered medication and their response to the medication can further confirm the presence of the disease, while neuroimaging is mainly utilized to rule out other pathologies that may cause symptoms that are similar to PD such as strokes or tumors.

Even though many severity rating systems exist such as the *Hoehn & Yahr* scale and the *Webster* scale, the gold standard for assessing PD severity is the *Movement Disorder Society-Unified Parkinson's Disease Rating Scale* (MDS-UPDRS) [4]. This scale is divided in four parts and evaluates patients according to the following categories:

- **Part I - Non-Motor Aspects of Experiences of Daily Living (nM-EDL):** focuses on how non-motor symptoms (such as mental and emotional issues) impact the patient's daily life.
- **Part II - Motor Aspects of Experiences of Daily Living (M-EDL):** focuses on how motor symptoms affect the patient's daily life (this part is self-evaluated by the patient)
- **Part III - Motor Examination:** is conducted by a clinician who objectively assesses the severity of the symptoms of the disease.
- **Part IV - Motor Complications:** the clinician assesses the complications that come from the long-term use of medication.

Scores for each part are either assigned by the clinician through examination and assessment or they are reported by the patient suffering from the disease in the form of a questionnaire. This makes the scale highly subjective, and research efforts are aiming to map neuroimaging and biological data to the disease severity rating scales. By using a Graph Neural Network (GNN), it is possible to model the connectivity of the brain and predict PD severity without relying on subjective exams.

1.1.2 Neuroimaging and Brain Connectivity

Neuroimaging is an extremely useful tool, but it is not sufficient on its own for PD diagnosis and severity evaluation. Current neuroimaging modalities can detect dopamine deficiency and be used to rule out conditions such as strokes but they do not have the specificity to distinguish PD from other syndromes that may appear similar on medical images. These modalities are also not precise enough to detect the subtle structural and functional changes typical of the early stages of PD, therefore they are not able to aid in the diagnosis of PD if it is in the early stages. Four major neuroimaging modalities—Magnetic Resonance Imaging (MRI) [5], functional MRI (fMRI) [6], Diffusion Tensor Imaging (DTI) [7], and Single-Photon Emission Computed Tomography (SPECT) [8]—are at the foundation of brain connectivity analysis.

MRI provides detail of anatomical structure, allowing to measure volume, position, and various metrics of different regions of interest (ROIs) of the patient's brain. DTI measures diffusion of water molecules, allowing to map white matter tracts and evaluate structural integrity. fMRI measures Blood Oxygen Level-Dependent (BOLD) signal, capturing the flow of blood in response to neural activity to measure functional relationship between ROIs. The latter three are all obtained by placing the patient in an MRI scanner which contains a powerful magnet. The magnet may pulse in different directions or the patient may be asked to perform simple tasks in order to obtain all the different types of neural images. Lastly, SPECT, is obtained by administering a specific radiotracer (Ioflupane-123) to the patient. The radiotracer binds to dopamine transporters in the brain, allowing the visualization of dopaminergic neuron density and health through a rotating gamma camera that is able to detect gamma rays.

These modalities are the basis for constructing brain connectivity networks, which can be modeled as graphs. The process typically involves:

1. **Parcellation:** Using MRI [9] to map the brain to a brain atlas, defining Regions of Interest (ROIs) that act as the *nodes* of the graph.

2. **Structural Connectivity (SC):** Using DTI and tractography to quantify the physical bundles connecting ROIs and defining structural *edge weights* [10, 11, 12].

3. **Functional Connectivity (FC):** Using fMRI to calculate strength between the BOLD time-series of different ROIs, defining the functional *edge weights* [13].

SPECT has the limitation of having low temporal resolution which makes connectivity estimation challenging. Recent research has explored fMRI integration with SPECT to compensate for the low resolution and allowing researchers to estimate functional networks from SPECT as well. By representing the brain as a set of interconnected nodes (ROIs) and edges (SC/FC), we can leverage Graph Neural Networks to analyze brain networks.

1.1.3 Graph Neural Networks

Deep learning architectures such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) are suitable for structured data such as grids that could represent images or sequences that might represent text. Brain connectivity data is characterized by more complex topology, relationships and irregular structure compared to the grid and sequence data that is suitable to be fed to CNNs and RNNs. GNNs are an extremely useful tool in addressing irregular structure and they extend deep learning operations on graphs. The brain can be represented as a graph where a set of nodes represents ROIs and a set of edges represents structural and functional connectivity between ROIs.

The mechanism that makes GNNs function is the Message Passing (Neighborhood Aggregation) mechanism. In each layer of a GNN, a node updates its features by aggregating information from its neighbors. This mechanism allows GNNs to see the neighbors of each ROI and by having deeper networks, a GNN can have a greater understanding of the topology of the brain network.

GNNs [14, 15] allow for the preservation of brain topology by working directly with the natural graph-like structure of the brain rather than with a flattened version of it that would be suitable for a CNN or an RNN. GNNs can also perform graph-level operations making them useful for graph-level regression and classification and have an understanding of the global structure of the brain rather than focusing on only certain parts of it at a time. Patterns that are indicative of PD may be invisible to localized analysis and may require to analyze the global context of the brain.

As powerful as GNNs may be they do have some limitations. GNNs are subject to over-smoothing, which occurs when a deep GNN aggregates the neighbors of the nodes so much that critical information is lost, making all node representations nearly identical. On the contrary, under-reaching occurs when a GNN is too shallow and important long-distance dependencies are not taken into account. Another issue with GNNs is over-squashing, which occurs when the edges of the graph bottleneck the information from a node's distant neighbor by compressing the information in vector of fixed size that is not large enough to preserve critical data; ultimately leading to information loss. These main contrasting limitation make it a challenge to find the ideal architecture configuration of GNNs for certain tasks.

1.2 Problem Statement and Motivation

Despite the usefulness of severity rating scales such as MDS-UPDRS and neuroimaging, the tools have the potential to play a more crucial role in PD diagnosis and severity evaluation. The main methods of diagnosis, disease evolution prediction and treatment prescription in PD originate from clinical exams and observation while the previous mentioned tools serve as an aid as there is no current, widely recognized, computational technique to leverage them for PD analysis. The main reasons that the previously mentioned tools don't currently play crucial roles in PD assessment are that they suffer from three main limitations:

1. **Subjectivity:** Clinical scores in the MDS-UPDRS are based on observations by medical staff and PD patients. These scores can vary significantly between different clinicians and patients and they are not objective.

2. **Sensitivity:** Neuroimaging analysis often fails to detect the subtle alterations in brain connectivity that occur during the early stages of PD and does not provide the tools to accurately diagnose or evaluate the status of a patient with PD.

3. **Dimensionality:** Traditional machine learning methods suitable to process structured grid or sequence-like data such as images or text are not able to extract meaningful features from neuroimaging modalities without losing the underlying topological information of the brain.

The main problem addressed in this thesis is the lack of a robust, objective, and automated system that is able to map complex, multi-modal brain connectivity patterns to the disease severity scale defined by the MDS-UPDRS and that is able to capture the underlying brain patterns that characterize neurodegenerative disorders, making predictions of the development of this class of diseases for a certain patient.

The motivation for this research is clinical necessity and technical advancement. A system that predicts MDS-UPDRS scores would allow for more standardized assessments, making personalized treatments easier to prescribe and providing more accurate prognosis for patients with an outlook on the progression of the disease. However, beyond accurate predictions, another critical element frequently missing from state-of-the-art approaches is model transparency. Explainability is a significant necessity in the field of AI in healthcare as clinicians require more than a "black-box" output and need to understand the workings of machine learning systems in order to progress research. A system that performs the aforementioned tasks in an explainable way could provide a significant contribution to the ongoing and future research both in the medical and computational field. Graph Neural Networks allow to perform deep learning operations on the brain while preserving its structure and nature. This thesis aims to help in filling the gap between clinical neurology and geometric deep learning, ultimately improving the precision of PD severity evaluation and contributing to the broader field of AI-driven neurodiagnostics.

1.2.1 Technical Challenges

Applying Graph Neural Networks (GNNs) to neurodegenerative disease assessment involves several technical difficulties that distinguish it from standard and more explored machine learning tasks.

A significant difficulty is **Data Scarcity and Small Sample Sizes**; more explored and researched deep learning domains work on datasets that contain millions of samples, while neuroimaging datasets like the Parkinson's Progression Markers Initiative (PPMI) often contain only a few hundred to a thousand subjects due to the relatively small number of patients suffering from PD specifically, the even smaller number of volunteers willing to make medical data public, and the novelty of the field of AI-driven neurodiagnostics. Small sample sizes when training GNN models make learning extremely challenging and can lead to overfitting, which occurs when the model memorizes data instead of learning from it. However, several techniques exist to compensate for limited data such as oversampling and other targeted architectural choices.

Furthermore, **Multi-modal Data Fusion** may be challenging; integrating data from different neuroimaging modalities into a single data structure suitable to be fed to a GNN requires capturing the complex relationships between all the modalities and choosing a data structure configuration that is appropriate to represent the data in the most expressive and informative way for the given task.

Another challenge is represented by **Graph Construction and Atlas Dependency**, as there are various anatomical brain atlases to choose from and an infinite number of values to choose from for thresholding of connectivity matrices; these two parameters heavily impact the resulting graph and may impact the learning of the model. Different brain atlases contain different ROIs that may be relevant for certain types of analysis more than others. For example, a brain atlas that focuses on mapping accurately the ROIs in the SNpc may be more suitable to perform analysis on PD as compared to an atlas that focuses on mapping accurately the ROIs in the hippocampus or entorhinal cortex which are regions mostly affected by AD.

A further challenge is **GNN Architecture Definition**; a GNN architecture suitable for analyzing brain connectivity data is required in order to be able to analyze the brains of patients. The GNN architecture must be carefully designed in order to maintain critical features and brain topology, preserving the biological significance of brain connectivity graphs.

Finally, for the model proposed to be relevant in the medical field, it must address **Interpretability and Clinical Trust**. A "black-box" output that simply takes in data and displays results is not sufficient for clinicians attempting to understand how certain neurodegenerative diseases work and what can be done to treat patients suffering from these diseases. Explainability—making the results and outputs of the model understandable to humans—is essential, especially in the medical field, and it is required for the progression of research.

1.2.2 Objectives and Contributions

The main goal is to develop and evaluate a Graph Neural Network framework, leveraging contrastive learning, capable of assessing Parkinson's disease severity and predicting its progression through the integration of multi-modal neuroimaging data. To achieve this, the following specific objectives have been defined:

1. **Multi-modal Graph Construction:** To design a stable pipeline for converting raw neuroimaging data from the PPMI dataset into data structures in which brain topology and its biological significance is preserved.
2. **Development of a Severity-Prediction GNN:** To implement and optimize a GNN architecture tailored for regression tasks, specifically mapping brain connectivity features to disease severity rating scales such as MDS-UPDRS scores.
3. **Progression Analysis:** To investigate the model's ability to predict future disease states by analyzing changes in brain connectivity graphs coming from the same patient in different periods of time.
4. **Interpretability and Biomarker Identification:** To implement explainability techniques to identify the specific neural sub-networks and biomarkers that serve as the most significant predictors of the disease, its severity and its progression.

This thesis provides several contributions to the fields of AI-driven healthcare and neuroimaging, including a multi-modal fusion strategy for integrating four different neuroimaging modalities (MRI, fMRI, DTI, and SPECT), a framework for severity quantization to reduce subjectivity, and a model able to predict disease progression. Additionally, we contribute to topological biomarker discovery by identifying brain regions susceptible to neurodegeneration and provide a benchmarking of the model against other baselines in the same field to demonstrate superior performance in capturing the nature of brain data.

1.3 Thesis Structure

The remaining five chapters of the thesis are structured as follows. **Chapter 2: Related Works** provides an overview of the current state-of-the-art in neuroimaging analysis for neurodegenerative disease and explores existing machine learning approaches. It also provides a comprehensive review of the literature on the subject of the thesis. **Chapter 3: Methodology** details the technical framework developed for this research, including the construction of structural and functional brain and the specific architecture of the proposed models that are used in severity prediction and progression evaluation. The details and results of the experiments performed along with the quantitative findings and qualitative analysis of learned biomarkers are described in **Chapter 4: Results**. Finally, **Chapter 5: Conclusion** summarizes the primary findings, acknowledges the limitations of the current study, and proposes directions for future research in AI-driven neurodegenerative disease assessment.

2 Related Works

The following section provides a review of the literature at the intersection of deep learning and neuroimaging. The review provides a theoretical context on GNNs. It then establishes the foundations of brain network construction from different neuroimaging modalities and the analysis of the constructed brain networks. Finally, it emphasizes the emerging role of contrastive learning [16] in modern research.

2.1 Graph Neural Networks

GNNs [17] represent a broad class of deep learning architectures design to process data represented as graphs $G = (V, E)$ [18] where V is the set of vertices of the graph and E is the set of edges. Typically, the vertices of a graph are represented by an index and the edges are represented by a pair of indices representing the two vertices that the edge connects. GNNs function thanks to a mechanism called message passing, represented by the following equation:

$$h_v^{(k)} = \sigma\left(W^{(k)} \cdot \text{AGGREGATE}\left(\{h_u^{(k-1)} : u \in N(v) \cup \{v\}\}\right)\right) \tag{1}$$

In detail, the terms of the equation are:

- $h_v^{(k)}$: the feature representation h of a vertex v at the layer index k of the GNN architecture.
- σ: an activation function (typically non-linear) that allows to model complex relationships between patterns in data.
- $W^{(k)}$: the learnable matrix of weights whose values are updated during training to minimize error.
- AGGREGATE: a permutation invariant aggregation function to combine data from a node's neighbors into a fixed-size vector. The function may be represented for example by mean, sum, or max pooling.
- $h_u^{(k-1)} : u \in N(v) \cup \{v\}$: the feature representation h of a vertex u at the previous layer index $k-1$ where vertex u is in the set of neighbors $N(v)$ of vertex v (including v itself)

Therefore, we can say that during the message passing process, node v collapses its own features and its neighbors' features into a fixed-size vector which is multiplied by weights W. Then the result of the product is passed through an activation function σ which adds non-linearity and the final result is the node representation of v at the current layer.

As discussed previously, GNNs have some limitations that derive from the message passing mechanism they implement. GNNs are subject to over-smoothing, under-reaching, and over-squashing. Different GNN [17] class architectures have been developed to address the classic limitations of the message passing

framework as well as to make GNNs a more flexible and expressive tool. The main architecture variations include but are not limited to: Graph Convolutional Network (GCN) [19], Graph Attention Network (GAT) [20], Graph Recurrent Network (GRN) [21], and Graph Isomorphism Network (GIN) [22]. The following subsections detail the main differences between these previously mentioned architectures and traditional GNNs, highlighting their functioning.

2.1.1 Graph Convolutional Network

GCNs [19] are the most widely recognized implementation of GNNs. This type of architecture performs graph convolution (inspired by CNNs) and addresses two main problems of traditional GNNs:

- Exploding/vanishing values: differences in number of neighbors that a certain node has can lead the fixed-vector representation required in message passing mechanism to not preserve critical data (as in over-squashing) or to represent few neighbors in vectors that are too large giving them too much importance.
- Lack of hierarchy: the features of a node with many neighbors can be easily drawn out by the features of its large amount of neighbors.

GCNs solve the problem by introducing symmetric normalization; in which messages in the message passing mechanism are scaled based on the degree of both the sender and the receiver. The message passing in GCNs is represented by the equation:

$$h_v^{(k)} = \sigma\left(W^{(k)} \cdot \sum_{u \in \mathcal{N}(v) \cup \{v\}} \frac{1}{\sqrt{\tilde{d}_v \tilde{d}_u}} h_u^{(k-1)}\right) \tag{2}$$

As we may observe, multiple terms of the equation are similar to the terms of equation 1 relating to GNNs. However, the differences lie in the replacement of the $h_u^{(k-1)} : u \in N(v) \cup \{v\}$ term with following term $\sum_{u \in \mathcal{N}(v) \cup \{v\}} \frac{1}{\sqrt{\tilde{d}_v \tilde{d}_u}} h_u^{(k-1)}$ composed of:

- $\tilde{d}_u$: representing the number of nodes that u is connected to incremented by 1 (to account for u's connection to itself).
- $\tilde{d}_v$: which represents the number of nodes that u is connected to incremented by 1.

The degree of the node can have different meanings based on the graph type. In undirected graphs it is simply the number of nodes that a certain node is connected to. For directed graphs it may be split in in-degree, which represents the number of edges pointing to the node, and out-degree, indicating the number of edges pointing from the node.

According to the formula above, we may state that during the message passing of a GCN, a node v performs a symmetric normalized sum of its features

and its neighbors' features based on the degrees of the nodes v and u. Then collapses them in a vector to be multiplied by the learnable weights W and the result is passed through an activation function σ. Efficiently reducing the problems of exploding/vanishing values and lack of hierarchy.

2.1.2 Graph Attention Network

Even though GCNs represent a significant step forward in improving architectures of GNNs, they still have some weaknesses, among which:

- Static weighting: in GCNs, the importance of a neighbor according to which it is aggregated with the target node is determined solely on the number of nodes that both the neighbor and the target node are connected to. The degree of a node is not representative of its features and it conveys no information other than the number of connections.
- Transduction: GCNs struggle to learn from graphs with different structure or completely new nodes and are not suitable for generalization to entirely new graphs.

GATs [20] replace the symmetric normalization component $\frac{1}{\sqrt{\tilde{d}_v \tilde{d}_u}}$ of GNNs with an attention coefficient α_{vu}. The message passing mechanism in GATs can be represented by the formula:

$$h_v^{(k)} = \sigma\left(\sum_{u \in \mathcal{N}(v) \cup \{v\}} \alpha_{vu}^{(k)} W^{(k)} h_u^{(k-1)}\right) \tag{3}$$

The attention coefficient is the output of a small learnable neural network shared across all edges which measures the importance of neighbor u with respect to node v. The calculation of the coefficient follows a process that can be divided in four steps:

- Linear projection: node features are transformed into a higher level feature space by multiplying the feature vectors of v and u by a learnable weight matrix W resulting in $\vec{z}_v = Wh_v$ and $\vec{z}_u = Wh_u$
- Concatenation: the two feature representations calculated in the previous step are concatenated $[\vec{z}_v \parallel \vec{z}_u]$
- Leaky ReLU Scoring: the dot product of the concatenated vector and a learnable attention weight vector a^T is passed through a LeakyReLU activation function introducing non-linearity to capture complex relationships $\text{LeakyReLU}\left(\vec{a}^T[Wh_v \parallel Wh_u]\right)$
- Softmax Normalization: finally the softmax function is applied to the result of the previous step to make the weights comparable across the whole neighborhood.

The calculation of attention coefficient can be summarized with the following equation:

$$\alpha_{vu} = \frac{\exp\left(\text{LeakyReLU}\left(\vec{a}^T[Wh_v \| Wh_u]\right)\right)}{\sum_{k \in \mathcal{N}(v) \cup v} \exp\left(\text{LeakyReLU}\left(\vec{a}^T[Wh_v \| Wh_k]\right)\right)} \tag{4}$$

We can say that during the GAT message passing process, node v computes a weighted sum of its own features and its neighbors' features—where each message is scaled by a learnable attention coefficient α that determines the relative importance of each neighbor—to collapse them into a single vector which is transformed by weights W. Then, the result is passed through an activation function σ which adds non-linearity, and the final result is the node representation of v at the current layer.

2.1.3 Graph Recurrent Network

Often referred to as Gated Graph Neural Network (GGNN), a GRN [21] is a specialized type of GNN architecture, inspired by RNNs, designed specifically to address the integration of temporality. A node's state in this case is not only influenced by spatial dependencies but also temporal dependencies (how the node's representation evolves from one time step to the next time step). There are multiple types of GRNs but the most common type is the above mentioned GGNN. GRNs solve two major limitations of other GNN models:

- Long range dependencies: traditional GNN models are often unable to capture long range dependencies as they require depth which leads to oversmoothing. GRNs are able to capture these dependencies through recursion.
- Sequential logic: traditional GNNs don't allow the state of nodes to evolve through multiple iterations. While GRNs can refine node representations through multiple iterations.
- Variable length propagation: a given GNN architecture is not flexible as it is suitable only for graphs of a unique size or small set of sizes. A graph too small would cause oversmoothing and a graph too large would cause under-reaching.

A GGNN's message passing mechanism can be described by the following equation:

$$a_v^{(t)} = \sum_{u \in \mathcal{N}(v)} W h_u^{(t-1)} \tag{5}$$

As it can be observed, the message passing is layer (z) independent and t denotes a point in time. The hidden state is then updated using a Gated Recurrent Unit (GRU) through which the message is passed. The GRU has the following components:

- Update gate $z_v^{(t)} = \sigma\left(W_z a_v^{(t)} + U_z h_v^{(t-1)}\right)$:
- Reset gate: $r_v^{(t)} = \sigma\left(W_r a_v^{(t)} + U_r h_v^{(t-1)}\right)$
- Candidate state $\tilde{h}_v^{(t)} = \tanh\left(W_h a_v^{(t)} + U_h(r_v^{(t)} \odot h_v^{(t-1)})\right)$:

Finally, the updated node feature representation can be written as:

$$h_v^{(t)} = (1 - z_v^{(t)}) \odot h_v^{(t-1)} + z_v^{(t)} \odot \tilde{h}_v^{(t)} \tag{6}$$

Where:

- $z_v^{(t)}$ is the update gate, and $r_v^{(t)}$ is the reset gate.
- $\odot$ denotes the element-wise (Hadamard) product.
- W and U are learnable parameter matrices.
- σ is the sigmoid function, and tanh is the hyperbolic tangent activation.

During the GRN message passing process, node v aggregates the features of its neighbors' into a message vector which is then passed—along with the node's previous hidden state—into a Gated Recurrent Unit (GRU). Then, the gating mechanism within the GRU adds non-linearity and decides how to fuse the new information with existing memory, and the final result is the node representation of v at the current time step.

2.1.4 Graph Isomorphism Network

The main issue that GINs [22] are designed to solve is expressivity; traditional GNNs are sometimes unable to distinguish graphs that have different topologies but similar neighborhood aggregation statistics. Mean aggregation fails when distinguishing between a larger neighborhood having a certain feature and a smaller neighborhood having the same feature in the same proportion. While max aggregation fails when multiple neighbors have the same strongest feature causing the model to only see one of them. GINs solve the issue by using injective mapping, which forces aggregation and update functions to be injective (which means every unique set of inputs produces a unique output). GINs achieve injective mapping by using sum aggregation and Multilayer Perceptrons (MLPs).

The message passing mechanism of GINs is detailed by the formula:

$$h_v^{(k)} = \mathrm{MLP}^{(k)}\left((1 + \epsilon^{(k)}) \cdot h_v^{(k-1)} + \sum_{u \in \mathcal{N}(v)} h_u^{(k-1)}\right) \tag{7}$$

Where:

- MLP: feedforward neural network consisting of at least three layers of nodes; an input layer, one or more hidden layers, and an output layer. Except for the input/output nodes, each node uses a non-linear activation function.

- ϵ: is a learnable parameter (which can be fixed at 0 for certain GINs) that controls how much weight is given to the node's own features with respect to the combined features of its neighbors.

The MLP transforms the computed vector and can be further divided into the following parts:

- Linear layers: each layer in the MLP performs a weighted sum of inputs. If the input is x the MLP computes $z = W \cdot x + b$. Where weights W determine the importance of each feature and bias b allows the activation function to be shifted to better fit the model's data

- Non-linear activation (σ): after the linear calculation, the result is passed through a function like ReLU which prevents neurons from collapsing into a single linear operation. Activation functions allow neurons to capture non-linearity in data.

- Layer stacking: typically, an MLP in a GIN model has two linear layers with an activation in between $\text{MLP}(x) = W_2 \cdot \sigma(W_1 \cdot x + b_1) + b_2$

A learnable weight matrix W cannot distinguish between inputs that are mathematically similar but structurally different. By using an MLP, a GIN may be able to map inputs to different areas of the vector space, allowing it to capture the structural differences that a traditional GNN would not be able to capture.

During the GIN message passing process, node v performs a summation of the features of its neighbors and adds them to its own features—weighted by a learnable parameter ϵ—to ensure the aggregation is injective; this combined vector is then transformed by a MLP, which acts as a universal approximator to preserve unique structural information, and the final result is the node representation of v at the current layer.

2.2 Brain Network Construction and Analysis

The data at the foundation of brain networks come from various neuroimaging modalities. Neural images of the brain can be preprocessed through computational pipelines which extract relevant information from each modality and build data structures that can be analyzed through machine learning methods aimed at capturing patterns in features and structure of data. We first analyze different neuroimaging modalities, then detail how brain networks are constructed and analyzed for each of them.

2.2.1 Neuroimaging

MRI is a non-invasive technique used to view the inner structure of the brain. As described in [23], the method consists in placing the patient in an MRI scanner in which a powerful magnetic field causes protons of H_2O molecules in the brain to align with the magnetic field. The scanner then emits a radio frequency pulse at the frequency of hydrogen protons which displaces them from alignment. When the radio frequency is turned off the protons naturally return into alignment and they emit their own radio signals that are detected by the machine allowing a computer to convert (based on the measured relaxation times) these signals into 2D slices of the brain which are then converted into 3D volumes.

fMRI is a specialized application of MRI that measures brain activity rather than just structure. It does so by leveraging a technique called BOLD. The technique works thanks to a unique magnetic property of hemoglobin (the protein in red blood cells that carries oxygen):

- Oxygenated hemoglobin (HbO_2): is diamagnetic, meaning it has no magnetic impact on its surroundings and blends in with brain tissue.
- Deoxygenated hemoglobin (Hb): is paramagnetic, meaning it disturbs the local magnetic field

As detailed in [24], when a region of the brain becomes more active it requires more oxygen because neurons consume oxygen when they are activated; increasing the concentration of Hb which also causes a magnetic field disturbance captured by the MRI machine which is then able to map brain activity when the patient performs simple tasks that require neuron activation.

Another specialized MRI technique is **DTI**. This modality maps the structure and integrity of white matter tracts in the brain by tracking water molecule movement direction which is restricted by physical brain structures. As stated in [25], DTI measures Anisotropic diffusion, which occurs when water molecules have a directional movement preference along white matter tracts. The magnet in the MRI scanner pulses in different directions, applying mathematical gradients, which are then used to compute tensors describing 3D oval shapes in which orientation aligns with fibers. Scalars derived from these tensors mainly include:

- Fractional Anisotropy (FA): values from 0 to 1 which indicate health and density of white matter tracts.

- Mean Diffusivity (MD): indicates average rate of water movement in all directions within a tissue reflecting the tissue's structural integrity
- Tractography: uses directional data to map and visualize white matter brain pathways.

Lastly, **SPECT** is a neuroimaging modality that is not a specialized MRI technique but rather a nuclear medical imaging technique. Patients (in particular PD) are injected with a radiotracer called Ioflupane-123 which specifically binds to dopamine transporters in the brain. Radioactive tracers decay by emitting gamma rays which are detected by a gamma camera that allows to create visualizations of the brain to observe dopaminergic neuron health.

In different images of a patient suffering from PD, the disease can appear in different ways. In MRI scans, the SNpc appears as a comma shaped spot, whereas for a PD patient the "tail" of the comma disappears making it appear like a circular area. Similarly to MRI, SPECT scans display a circular area instead of a healthy comma shaped one when tracking dopamine transporters in the Striatum. A PD brain in fMRI scans displays abnormally low activity in the motor cortex while in DTI scans, white matter tracts in the SNpc begin to break down.

2.2.2 Brain Network Construction

The previously mentioned neuroimaging modalities are all useful when building brain networks. It may be appropriate to convert brain images to different representations in order to analyze them through computational techniques rather than simply manually or visually. Various research efforts have been directed towards building brain networks with the ultimate objective of analyzing brain structure to uncover patterns that characterize neurodegenerative diseases. Methods can be categorized in uni-modal (which exploit a single neuroimaging modality to construct a brain network) and multi-modal (leveraging data from multiple modalities).

The most widely recognized pipeline for brain parcellation from MRI is the ***FreeSurfer*** [26] suite. This set of tools provides computational methods to construct a detailed 3D model of a brain given an MRI image. The process of brain parcellation through *FreeSurfer* can be divided into three main stages:

- Volumetric processing: where motion correction fixes head tilts, skull stripping removes bone and skin from the MRI, and subcortical segmentation labels deep structures.
- Surface reconstruction: finds the boundary between white and gray matter and between the brain and the cerebrospinal fluid, measuring the distance between these surfaces and every single vertex of the brain.
- Parcellation: uses a process called spherical registration to align the patient's brain with the brain atlas used, labeling the ROIs.

Typically, brain parcellation through MRI results in a set of ROIs with features describing each ROI's positioning and volume and these regions represent nodes of a brain graph. *FreeSurfer* is not the only available tool to perform brain parcellation; other tools include *FastSurfer* [27] and *QuickNAT* [28] which use deep learning architectures to replicate the *FreeSurfer* segmentation in shorter times. However *FreeSurfer* has years of validation and is a more commonly used tool.

FreeSurfer has been integrated with ***fMRIPrep*** [29]; a pipeline to preprocess fMRI data. The pipeline is considered the gold standard in cleaning fMRI data, however, it is not suitable for brain network construction. To build a functional connectivity matrix from an fMRI [30], a framework like ***Nilearn*** [31] or *XCP-D* [32] is required. The latter is a standardized pipeline, while *Nilearn* contains a set of tools to build a pipeline to process clean fMRI images into different data structures and it is considered the SOTA for building brain graphs to be used in machine learning and GNNs. It contains tools to perform signal extraction, denoising, and finally matrix calculation. FC matrix construction also requires a brain atlas containing labels of different ROIs; while the common practice was using standard atlases, recent research has shifted towards a multi-modal approach; integrating the volumetric model of a patient's brain built using MRI parcellation to identify the various ROIs of the brain. This latter technique has the advantage of yielding more precise FC matrices that take into account for the patient's individual brain rather than using a general map.

In brain network construction, DTI images are useful in computing structural connectivity matrices. The most widely used tool is ***QSIPrep*** [33]. This tool is considered the gold standard because it integrates functionalities from major software suites (such as *FSL*, *MRtrix3*, *DSI Studio* and *ANTs*) into one single workflow. Unlike fMRI processing which requires two distinct tools, *QSIPrep* allows for both preprocessing and matrix construction. The preprocessing steps for DTI images are fundamentally different from the steps to preprocess fMRI images as DTI preprocessing attempts to solve a different problem (preservation of orientation of fiber vectors). The preprocessing includes denoising, Gibbs unringing, distortion correction, and Eddy current & motion correction. With preprocessed data, *QSIPrep* is then able to perform tractography, mapping white matter streamlines, and it is able to build an SC matrix representing streamline density in each ROI. SC matrix construction also requires a brain atlas and can benefit from using patient MRI image instead of a general atlas.

As discussed SPECT suffers from low temporal resolution which makes connectivity estimation challenging and there is no standard tool to perform it. A recent breakthrough is represented by the ***NeuroMark*** [34] pipeline which researchers used to propose a multi-modal fMRI-guided SPECT brain network construction method. The method consists in leveraging Spatially Constrained Independent Component Analysis (SC-ICA). It takes in input the SPECT scan of a patient and 53 spatial priors from *NeuroMark* representing different brain networks that were obtained by analyzing a large dataset of fMRI scans. The algorithm tries to match a patient's SPECT with the shape of a prior and it results in an individual patient map in which each prior network is assigned a

weight representing how active it is.

SPECT is not the only modality that benefits from multi-modal integration. Other research efforts have been made to attempt to integrate multiple neuroimaging modalities in single data structures. The most common method of multi-modal fusion includes integrating brain parcellation, FC matrix and SC matrix in one single graph structure. The node and the features such as cortical thickness and surface area are derived from MRI brain parcellation, while FC and SC matrices are computed with the methods stated above. One SOTA approach to combine resulting data from the three modalities is represented by ***IBrainGNN***[35], which assigns features to nodes through MRI as stated and then uses FC and SC matrices to assign two different edge weights to each edge between two nodes. The algorithm then uses the Message Passing mechanism of GNNs which takes into account the weights of the edges when aggregating neighbors. Finally, it collapses the graph into a latent embedding representing the state of the brain.

2.2.3 Brain Network Analysis

Constructing brain networks has the ultimate goal of providing data structures that are suitable to be analyzed through computational techniques. As neuroimaging modalities by themselves are rich in meaningful data that is too large to be analyzed manually and risks of being overlooked if analyzed visually. Various studies have aimed to analyze brain networks modeled as graphs through GNNs [36, 37]. The main tasks in the field include but are not solely limited to: diagnosis [38, 39] and classification, severity and progression prediction, biomarker identification, disease evolution [40] and treatment monitoring.

A major research effort that combines both brain disease localization and diagnosis through brain classification is ***BrainMAP*** [41]. *BrainMAP* is a model that combines fMRI and DTI, using machine learning techniques to extract disease relevant regions from the two modalities by evaluating their utility in diagnosing PD. The two modalities are then aligned and collapsed into an embedding that is suitable to be fed to a GNN for classification and is able to accurately classify a brain into PD, healthy, or SWEDD. However, collapsing these modalities into a single embedding can obscure the explicit topological connections between distinct brain regions. Addressing this limitation by explicitly preserving and filtering the network's connections, another method with a similar task is ***BrainHGT*** [42]. The model maps the brain as a graph in which each node (identified by an atlas ROI) is assigned a feature vector representing its FC and SC to other nodes obtained through fMRI and DTI analysis. The model then forces the weight of non-predictive connections to 0, effectively sparsifying [43] the graph and maintaining only disease-relevant regions. The brain architecture is then suitable to be analyzed by a GNN for classification.

While *BrainHGT* effectively filters the network for classification, it does not inherently provide transparent, biologically interpretable insights into which specific regions act as indicators of the disease. Addressing this need for explain-

ability, one of the main research efforts in biomarker identification is represented by ***BrainGNN*** [44], which introduces components useful in making GNNs explainable including:

- ROI-aware Graph Convolution (Ra-GConv): which uses different weights for each ROI allowing the model to learn features specific to the biology of each brain region
- ROI-selection Pooling (R-Pool): useful in selecting most relevant nodes in a brain graph.
- Specific regularization losses: TPK (TopK) Loss, which forces the model to select a small sparse set of important regions. And GLC (Group-level Consistency) Loss, which ensures that the model identifies the same set of biomarkers for patients in a specific group.

Although the aforementioned models excel at diagnosis and identifying structural biomarkers from static scans, predicting the longitudinal progression of the disease presents a distinct challenge. Shifting the focus toward future trajectory, ***PDualNet*** [45] is a model that addresses two main tasks; identifying which disease progression subtype (rapid, moderate or slow) each PD patient belongs to, and predicting the patient's future MDS-UPDRS scores. The model doesn't leverage neuroimaging and GNNs; it uses an autoencoder to create feature embeddings for each patient by analyzing MDS-UPDRS scores and then a transformer encoder which operates regressively to analyze how the embeddings relate to each other over time. Then an MLP uses the transformer encoder embeddings to predict disease progression subtype. However, by relying entirely on clinical scores, *PDualNet* omits the rich structural insights provided by neuroimaging. Reintegrating these elements to predict PD disease trajectory is represented by ***AdaMedGraph*** [46]. Researchers integrated three distinct types of data: MDS-UPDRS scores, MRI neuroimaging, and genetic Polygenic Risk Score (PRS). The modalities are fused into a single node of the graph which represents a patient and the model learns which patients are the most similar and connects them in the graph. GNNs are then used to predict a certain patient's future by analyzing the trajectories of other patients that are close in the graph.

Numerous successful efforts have been directed towards disease diagnosis, and biomarker and disease relevant region identification by leveraging neuroimaging modalities [47] and GNNs. However, research shows that preferred approaches to address the tasks of disease severity and progression prediction prefer approaches that don't rely on neuroimaging and GNNs. A model that addresses the latter mentioned task using those techniques would be a novelty in the field.

2.3 Contrastive Learning and Generative AI

2.3.1 Contrastive Learning

Contrastive learning is a powerful paradigm in modern machine learning. The base of the paradigm is learning by comparison. A contrastive learning model is typically self-supervised, as it learns to map data into an embedding space by attempting to map similar data close and dissimilar data far in that embedding space.

The most widely recognized contrastive learning model is the ***CLIP*** [48] model. *CLIP* consists of two separate neural networks trained concurrently:

- Image encoder: typically a ViT or a ResNet which takes raw image pixels and transforms them into an embedding.
- Text encoder: a transformer whose task is to transform a sentence into an embedding.

During training of a batch with N pairs, the model builds a matrix of $N \times N$ combinations between image and text. N of which are correct pairs (the diagonal of the matrix) and $(N^2 - N)$ incorrect pairs. *CLIP* then uses contrastive loss to maximize the cosine similarity between positive pairs and minimize the similarity between negative pairs. The contrastive loss (*InfoNCE*) can be represented by the following formula:

$$\mathcal{L}_{\text{InfoNCE}} = -\log \frac{\exp(\text{sim}(q, k^+)/\tau)}{\exp(\text{sim}(q, k^+)/\tau) + \sum_{i=1}^{K} \exp(\text{sim}(q, k_i^-)/\tau)} \tag{8}$$

Where:

- q: represents the text embedding of the evaluated sample.
- k^+: represents a positive image match to q in the form of an embedding (this can also be an augmentation of a positive sample).
- k_i^-: represent samples that are unrelated to q.
- sim: is a similarity function (usually cosine similarity $\frac{u \cdot v}{\|u\|\|v\|}$).
- τ: temperature hyperparameter that scales similarity scores.
- K: number of negative samples

AudioCLIP [49] is an extension of the original *CLIP* model that adds audio to the embedding space. In addition to the image and text encoders it uses an audio encoder which leverages *EsResNeXt*, a CNN architecture designed for environmental sound classification. Even though *AudioCLIP* has more than two modalities, it still compares each modality pairwise by computing a symmetric cross entropy loss across all pairs:

$$L_{total} = L_{Image-Text} + L_{Image-Audio} + L_{Text-Audio} \tag{9}$$

Using contrastive learning to map more than two modalities to the same shared embedding space can be a simple way to integrate multiple modalities into a CLIP-like model. However, this solution has one main issue: drifting modalities. For instance, it may occur in the *AudioCLIP* model that audio aligns to text and text aligns to images but this doesn't guarantee that audio and images align and they may drift apart in the shared embedding space.

TriCoLo [50] enforces cyclic consistency across all three modalities simultaneously, ensuring that the 3D-Voxel and 2D-Image representations are structurally aligned, rather than just independently anchored to text. Symmetric bimodal loss is defined by:

$$L(v,t) = \frac{1}{N}\sum_{j=1}^{n} \alpha l_j^{v\to t} + (1-\alpha) l_j^{t\to v} \tag{10}$$

Where:

- $l_j^{v\to t}$: is the *InfoNCE* (Equation 8) loss of modality v towards t.
- $l_j^{t\to v}$: is the *InfoNCE* loss of modality t towards v.
- α: is a hyperparameter balancing the importance of the two directions.

The model uses the symmetric loss to compute a sum between all modalities:

$$L_{tri} = L(v,i) + L(v,t) + L(i,t) \tag{11}$$

Contrastive learning models can also implement different architectures when aligning modalities. For instance ***ImageBind*** [51] defines a hub-and-spoke architecture in which a hub acts as a central modality and all other modalities (spokes) are aligned to the hub.

2.3.2 Generative AI

A **Variational Autoencoder (VAE)** [52] is a class of generative model and it consists of three main components:

- The Encoder (Inference Model): This neural network takes a data point x (like an image) and outputs the parameters of a probability distribution. Specifically, it outputs two vectors: a mean vector μ and a standard deviation vector σ. These define where the input is located in the latent space and how much uncertainty surrounds it.
- Latent Space: This is the bottleneck layer. Instead of passing a deterministic value, the model samples a random point z from the distribution defined by μ and σ.
- The Decoder (Generative Model): This neural network takes that sampled point z and attempts to reconstruct the original input, outputting x'.

To train the VAE, the model minimizes a loss function that balances two competing goals: accurately reconstructing the input and keeping the latent space organized. The loss function consists of two parts:

$$\mathcal{L} = -\mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] + D_{KL}(q_\phi(z|x) \,\|\, p(z)) \tag{12}$$

Reconstruction Loss measures how well the decoder can reconstruct the original input x from the latent sample z. It penalizes the model if x' does not look like x. Kullback-Leibler (KL) Divergence acts as a regularizer. It measures the difference between the learned distribution $q_\phi(z|x)$ and a prior distribution $p(z)$. By forcing the distributions to closely resemble a standard bell curve, the KL Divergence ensures the latent space remains continuous and densely packed, preventing gaps where the decoder wouldn't know how to generate a realistic output.

Because the latent point z is drawn randomly, the network cannot calculate the gradient to update the encoder's weights. VAEs solve this using the reparameterization trick. Instead of sampling z directly from the distribution defined by μ and σ, the model moves the randomness to an external variable ϵ. It samples ϵ from a standard normal distribution $\mathcal{N}(0, I)$, and then calculates z deterministically:

$$z = \mu + \sigma \odot \epsilon$$

Because ϵ handles all the randomness, the pathway from z back to μ and σ is now completely deterministic. The gradients can easily flow backward through the network, allowing the encoder to update and learn.

2.3.3 ProM3E

At the intersection of contrastive learning and generative AI we find the ***ProM3E*** [53] model. The model's main goal is to learn embeddings from missing modalities in a shared latent space by aligning modalities with contrastive learning and using generative reconstruction. It trains a contrastive model using symmetric *InfoNCE* (Equation 8) and an *ImageBind*-like hub-and-spoke architecture. Then the embeddings go through a Masked Variational Autoencoder (MVAE) which shows the model only a few modalities while masking the other and the model learns to reconstruct the contrastive embedding vector of the missing modalities through the VAE loss (Equation 12).

3 Methodology

In this chapter we detail the technical framework developed for this thesis project. We start from describing how the data was acquired and preprocessed. Then explain the architecture of the model, and how it relates to the objectives of this thesis. Source code is available at: https://github.com/EmanueleBarbieri2/GeNeuro.git

3.1 The Dataset: PPMI

PPMI [1] [2] [54]. is an initiative launched in 2010 by the *Michael J. Fox Foundation* and is widely considered the most important clinical study in PD research. Its main goal is to find biomarkers of PD. PPMI is an observational study meaning that it purely collects data and makes it widely available. As of 2026:

> *"PPMI is expanding to more than 4,000 volunteers, including 2,000 prodromal participants, recruited at nearly 50 international sites. The study continues to follow its earliest volunteers, enrolled beginning in 2010."*

The study collects a wide range of multimodal data including patient exam and assessment scores, and MRI (which DTI and fMRI can be considered subsets of), SPECT, PET, EGG and CT neuroimaging scans. Furthermore, the study is longitudinal, meaning that it aims to collect data of its volunteers over time, making it suitable for predicting a patient's future status. Efforts have also been directed towards publishing data derivatives, allowing researchers to skip data preprocessing either partially or entirely and ultimately making preprocessing simpler and less computationally intensive. For the previously mentioned reasons the PPMI dataset is the gold standard for machine learning applications in PD.

One of the challenges that can be encountered when working with PPMI data as well as other neuroimaging data cohorts is represented by missing modalities; it occurs often that a patient may have one or a few neuroimaging scans associated while scans from different modalities are missing. The issue calls for a robust model able to combine a varying number of available modalities.

[1] Data used in the preparation of this thesis were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/data). For up-to-date information on the study, visit www.ppmi-info.org.

[2] PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners, including: 4D Pharma, AbbVie, AcureX, Allergan, Amathus Therapeutics, ASAP, AskBio, Avid Radiopharmaceuticals, BIAL, BioArctic, Biogen, Biohaven, BioLegend, BlueRock Therapeutics, Bristol Myers Squibb, Calico Labs, Capsida, Eli Lilly, Gain Therapeutics, GE HealthCare, Genentech, GSK, Golub Capital, Handl Therapeutics, Insitro, Jazz Pharmaceuticals, Johnson & Johnson, Lundbeck, Merck, Meso Scale Discovery, Mission Therapeutics, Neurocrine Biosciences, Neuron23, Neuropore, Pfizer, Piramal, Prevail Therapeutics, Roche, Sanofi, Servier, Takeda, Teva, UCB, Vanqua Bio, Verily, and Voyager Therapeutics.

3.2 Preprocessing

To detail how each of the four neuroimaging modalities were preprocessed we can divide this section into four further sections (one for each modality). Thanks to the availability of data derivatives of PPMI, the processing of MRI and SPECT was found to be trivial, while fMRI and DTI had no useful derivatives for our purpose so they required a carefully designed pipeline able to convert them respectively into FC and SC matrices.

3.2.1 SPECT

The PPMI initiative occasionally publishes *Curated data cuts* in the form of `.xlsx` files. These files collect a wide range of parameters including but not limited to: patient identification, subgroup, MDS-UPDRS scores, and SPECT values for a few regions of the brain. SPECT images are preprocessed according to the *Molecular Imaging And Kinetic Analysis Toolbox (MIAKAT)* [55] pipeline for PET scans. The software includes tools for motion correction, region of interest parcellation, and radiotracer modeling.

PPMI's **`Curated data cut - (v.2025-11-12)`** contains Striatal Binding Ratio (SBR) of the striatum which can be divided into caudate and putamen. The data cut also distinguishes brain regions in the left and right hemisphere, allowing us to define a tree-like hierarchical structure (as seen in Figure 1) in which the *striatum* is the root node and is further divided into left and right *striatum*. The sons of the left *striatum* are the left putamen and caudate while the sons of the right *striatum* are the right putamen and caudate. Each node has its own SBR value associated and the edges have unitary weight.

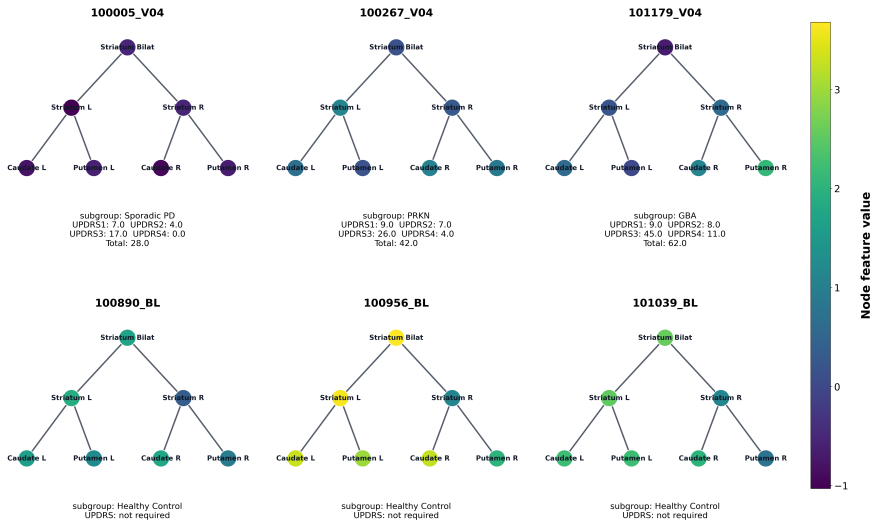

Figure 1: Examples of SPECT graphs built during data preprocessing

3.2.2 MRI

Another category of neuroimaging scans from PPMI which were found to have sufficiently detailed derivatives were the MRI images. 1716 patients were processed using *FreeSurfer* and the *fMRIPrep* pipeline which includes parcellation steps useful for extracting metrics from MRI. The latest derivatives include files:

- **`loni_fs7_aseg_volumes.csv`**: containing a set of patients for which volumes of subcortical brain regions identified by the *ASEG* atlas are reported considering the baseline visit of the patient.
- **`loni_fs7_aparc_cth.csv`**: containing a set of patients for which cortical thickness of cortical brain regions identified by the *APARC* atlas are reported considering the baseline visit of the patient.
- **`loni_fs7_aparc_sa.csv`**: containing a set of patients for which surface area of cortical brain regions identified by the *APARC* atlas are reported considering the baseline visit of the patient.

.

In the first step of our MRI preprocessing pipeline, each ROI of *FreeSurfer*'s **`aparc+aseg.mgz`** brain atlas (which combines cortical and subcortical regions from *APARC* and *ASEG* atlases) is used to identify a set of brain regions. Then each region from the atlas is associated to a region in the PPMI derivatives mentioned and a fully connected graph with all 113 ROIs is built by assigning a set of three features to each node. For subcortical nodes the features include two zeroes used for padding and a value representing volume. While for cortical regions features include one zero for padding, cortical thickness, and surface area.

In the final step, *FreeSurfer*'s atlas is used to compute (x, y, z) coordinates of the centroids of each ROI. The coordinates are then used to compute the normalized distance $d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$ between every pair of regions. Each edge in the fully connected graph is then assigned a weight equal to $1 - (\frac{d}{d_{max}})$.

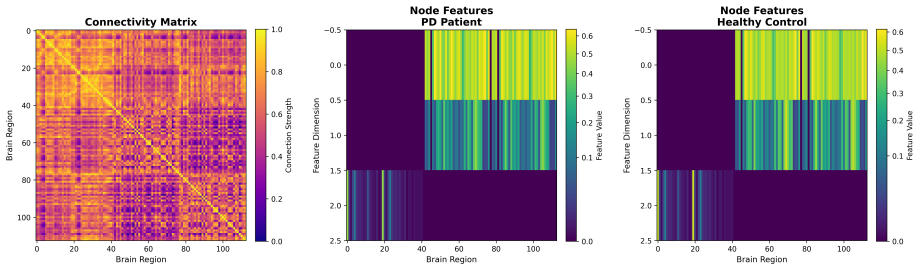

Figure 2: Examples of MRI graphs built during data preprocessing

3.2.3 fMRI

fMRI derived data was found insufficient to build FC matrices as it contained no detailed information about ROIs, only aggregate metrics. The gold standard to build FC matrices from fMRI scans is represented by the *fMRIPrep* pipeline which is included in the *FreeSurfer* suite. The pipeline includes steps such as head motion correction, slice-timing correction, susceptibility distortion correction, and co-registration with MRI. The operations performed are extremely expensive computationally, and require from 12 to 24 hours along with enough CPU cores, RAM and storage for one single fMRI to be processed. Due to the lack of appropriate resources to follow the standard, a pipeline was defined using the ***ANTsPy*** Python library in an attempt to replicate the steps of *fMRIPrep* as closely as possible without excessive processing times.

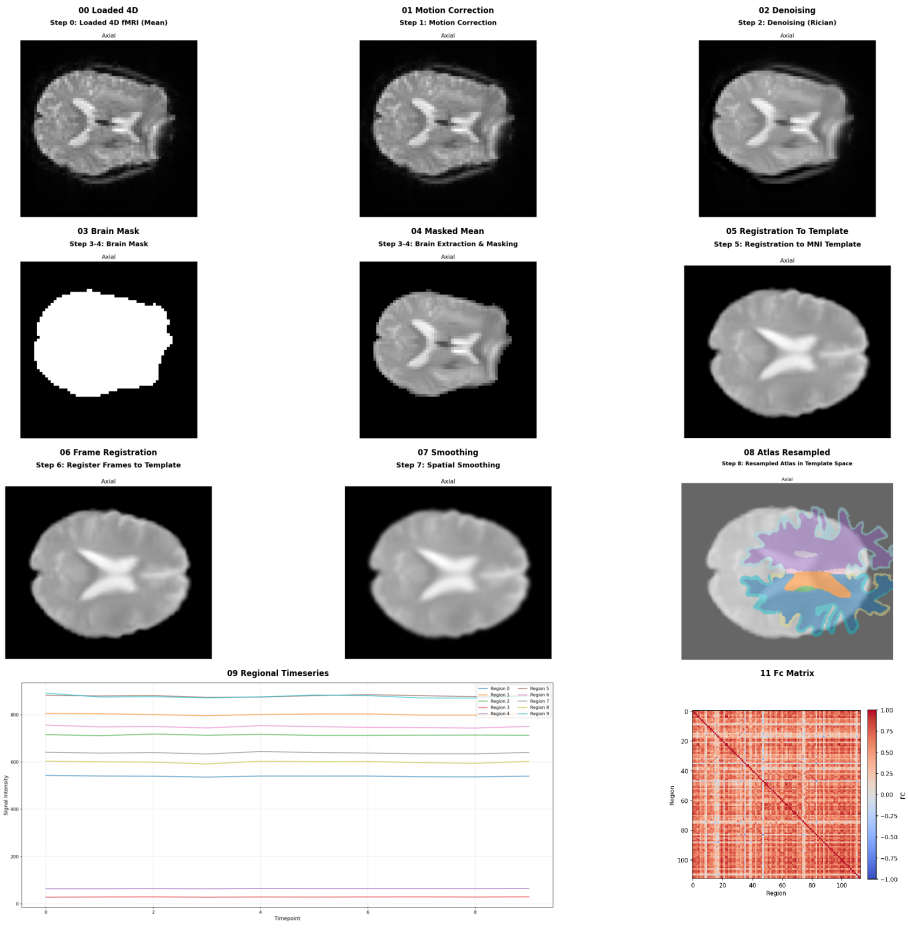

Figure 3: Steps of the fMRI preprocessing pipeline

Before running the defined pipeline, DICOM format 2D slices from the PPMI database are converted to NIfTI volumes using the `dcm2niix` tool. The pipeline then begins with **Motion Correction** which addresses the problem of the subject's head moving during the scan. fMRI scans are 4D (x, y, z, t) meaning they are a set of (x, y, z) volumes obtained over time t. If a subject's head is in a different position at time t than it was at time $t-1$ then the wrong voxels will be assigned to a certain ROI. The middle frame of the set of scans is used as a reference and a rotation matrix and translation vector that minimize the distance between each frame and the reference frame is computed; successfully correcting the motion distortion caused by the patient's movement.

MRI scanners are subject to noise. **Denoising** involves accounting for random fluctuations in MRI scanner receiver coils which cause rician noise represented by the equation:

$$p(x|\nu, \sigma) = \frac{x}{\sigma^2} \exp\left(-\frac{x^2 + \nu^2}{2\sigma^2}\right) I_0\left(\frac{x\nu}{\sigma^2}\right) \tag{13}$$

where:

- x: is the observed magnitude of the signal.
- v: is the true noiseless signal intensity.
- σ: is the standard deviation of the noise.
- I_0: is the modified Bessel function of the first kind with order 0: $I_0(z) = \frac{1}{\pi}\int_0^\pi e^{z\cos\theta} d\theta$

Once the noise level is estimated through the function, a rician-aware denoising filter is applied to find similar patches of voxels across images and average them, effectively reducing noise.

Brain Mask Extraction represents the next step in the pipeline. fMRI scans contain skull, skin, and cerebrospinal fluid which are all areas not of interest when building brain FC matrices. Values in fMRI that are under the magnitude of 0.2 are typically not part of the brain. These values are masked out from the volume by setting a threshold, keeping only the brain.

Every subject has different brain anatomy therefore in different patients the same anatomical region may be at different voxel coordinates. **Registration To MNI Template** involves transforming the brain to a common anatomical space, the *MNI152*, which was obtained by averaging the brains of 152 subjects. The registration method is Symmetric Diffeomorphic Normalization with Rigid-Affine initialization (SyNRA) which involves the following steps:

- Rigid-Affine Pre-alignment: finds rotation matrix R, translation vector t, and scaling factor s to align and orient brain centroids.
- Diffeomorphic Refinement: finds a smooth deformation field $\Phi(x, y, z)$ such that (x, y, z) of the warped image is equal to $\Phi(x, y, z)$ of the subject image. The optimization goal is minimizing: $\|\Phi_{subject} - template\|^2 + \lambda\|\Delta\Phi\|^2$

Spatial Smoothing aims to further remove noise and it is performed after the registration because performing it before would blur signals through the boundaries of the ROIs. The discrete spatial smoothing of each voxel v is calculated as a weighted sum of its local neighborhood $N(v)$:

$$I_{\text{smooth}}(v) = \frac{\sum_{n \in N(v)} I(n) \cdot w(n)}{\sum_{n \in N(v)} w(n)} \tag{14}$$

where the weight $w(n)$ for a neighbor n at a distance $d = \|n - v\|$ is determined by the Gaussian kernel:

$$w(n) = \exp\left(-\frac{\|n - v\|^2}{2\sigma^2}\right) \tag{15}$$

After an intermediate step that resamples the `aparc+aseg.mgz` brain atlas to the MNI template space to associate a set of voxels to each brain regions, **Time Series Extraction** takes place. This step extracts a ROI-wise time series from the voxel based time series by averaging all voxels across a region.

The fMRI time series contains unwanted frequency components originating from scanner drift (low frequency), and thermal noise (high frequency). A **Temporal Bandpass Filtering** is applied to filter out all signals that are not in the frequency range of interest. The Nyquist frequency is computed as $\frac{1}{2 \cdot TR}$ where TR is the repetition time of the scan (time between pulses). The low normalized and high normalized frequencies derived from Nyquist frequency are then used by a Butterworth filter to discard irrelevant frequencies.

The final step of the pipeline is **FC Matrix Computation**. For regions i and j, the pearson correlation is:

$$r_{ij} = \frac{\sum_{t=1}^{T}(X_i[t] - \mu_i)(X_j[t] - \mu_j)}{\sqrt{\sum_{t=1}^{T}(X_i[t] - \mu_i)^2}\sqrt{\sum_{t=1}^{T}(X_j[t] - \mu_j)^2}} \tag{16}$$

where:

- $X_i[t]$: activity in region i at time point t.
- μ_i: mean activity in region i.
- t: time point index.

A high correlation between two ROIs means that they are active simultaneously when the patient performs a task, therefore they are functionally connected. A patient may have multiple fMRI scans from the same session. These scans are combined using a Fisher Z-Transform:

$$z = \operatorname{arctanh}(r) = \frac{1}{2}\ln\left(\frac{1 + r}{1 - r}\right) \tag{17}$$

then averaging the correlation scores of the same ROI from different scans in the Z-space.

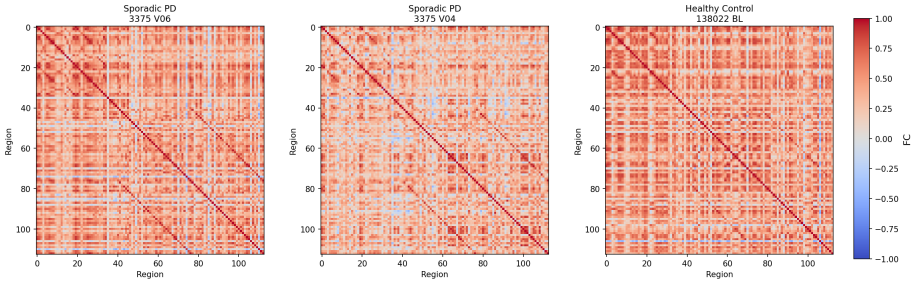

Figure 4: Examples of FC matrices built during data preprocessing

The final step to obtain functional connectivity graphs is trivial as it involves defining PyG graph structures in which nodes have feature vector (x, y, z) centroid coordinates for each ROI and edges are weighted according to the FC matrix.

3.2.4 DTI

DTI derived data is insufficient to build SC matrices as it only contained information about a few ROIs in the SNpc. The gold standard to build SC matrices from DTI scans is represented by the *QSIPrep* pipeline. The pipeline includes steps such as head motion correction, eddy current correction, susceptibility denoising, Gibbs unringing, and co-registration. Unlike *fMRIPrep*, *QSIPrep* also includes reconstruction functions to perform tensor fitting and tractography. The operations performed are computationally expensive, and require from 4 to 6 hours along with enough CPU cores, RAM and storage for one single DTI to be processed. Due to the lack of appropriate resources to follow the standard, a pipeline was defined using the ***ANTsPy*** Python library in an attempt to replicate the steps of *QSIPrep* as closely as possible without excessive processing times.

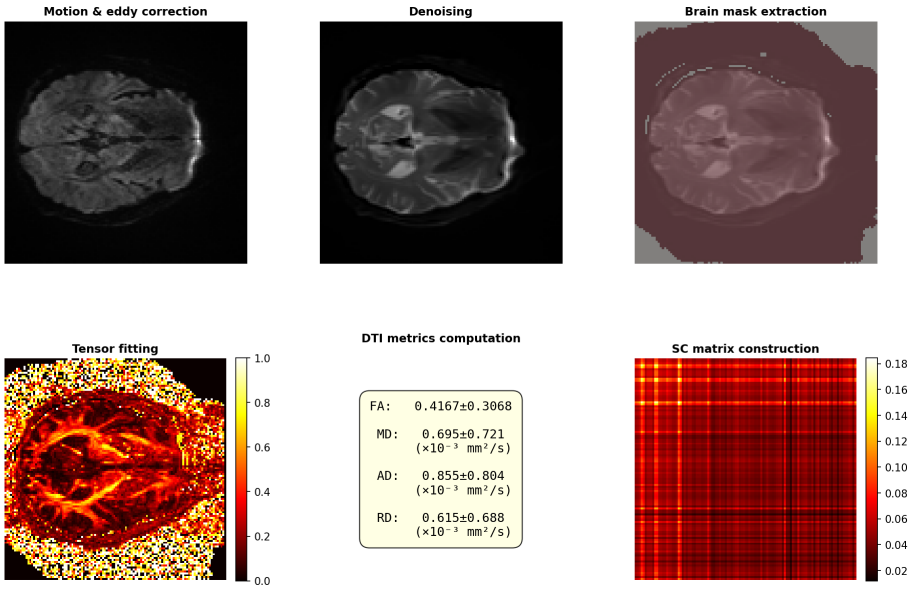

Figure 5: Steps of the DTI preprocessing pipeline

2D slices from the PPMI database are converted to NIfTI volumes using the `dcm2niix` tool, the pipeline begins with **Motion & Eddy Current Correction**. An eddy current is a loop of electrical current induced by changing the magnetic field in the conductor's vicinity. The process is similar to the motion correction process used for fMRI scans but the main differences are:

- Reference image: the reference image (*B0*) is acquired when no diffusion gradients are applied.
- Scaling: the rotation matrix accounts for scaling as well.

The pipeline continues with **Denoising** (described by Equation 13) and **Brain Mask Extraction** which follow the same procedure for DTI as they do for fMRI.

Tensor Fitting involves solving the Stejskal-Tanner equation for the unknown matrix D. Diffusion in biological tissue is anisotropic (directionally dependent). The diffusion tensor D can be represented as a 3×3 symmetric positive-definite matrix:

$$D = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{pmatrix} \tag{18}$$

The Stejskal-Tanner equation relates DTI signal to diffusion:

$$S(b, g) = S_0 \exp(-b \cdot g^T D g) \tag{19}$$

where:

- $S(b, g)$: diffusion-weighted signal intensity in the direction of g.
- S_0: non-diffusion-weighted signal intensity (*B0* image).
- b: diffusion weighting factor (b-value), defined by the pulse sequence parameters (strength, duration, and spacing of gradients).
- g: unit vector representing the direction of the applied diffusion gradient.

Once D is computed we can derive the eigenvalues of the diffusion tensor $\lambda_1, \lambda_2, \lambda_3$ and **Extract DTI Metrics** with the following equations:

$$\text{Mean Diffusivity} \qquad MD = \frac{\lambda_1 + \lambda_2 + \lambda_3}{3} \tag{20}$$

$$\text{Fractional Anisotropy} \qquad FA = \sqrt{\frac{3}{2}} \frac{\sqrt{\sum_{i=1}^{3} (\lambda_i - MD)^2}}{\sqrt{\sum_{i=1}^{3} \lambda_i^2}} \tag{21}$$

$$\text{Axial Diffusivity} \qquad AD = \lambda_1 \tag{22}$$

$$\text{Radial Diffusivity} \qquad RD = \frac{\lambda_2 + \lambda_3}{2} \tag{23}$$

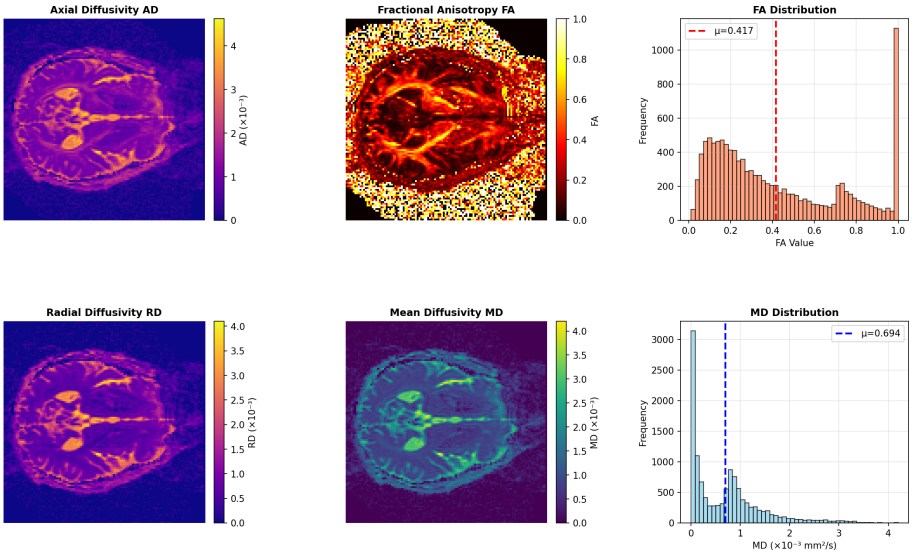

Figure 6: Visualization of DTI metrics

An intermediate merging step concatenates all scans from the same visit along with b-values (which represent the strength of the diffusion measurement) and b-vecs (which represent measurement direction) along the 4D dimension before performing **SC Matrix Construction**. Similarly to the processing of fMRI volumes, FA are registered to the MNI template using SyNRA. The structural connectivity of two ROIs is:

$$SC_{ij} = FA_i \times FA_j \tag{24}$$

A high product between two ROIs signifies healthy and dense white matter fiber pathways, meaning they have a strong structural connection.

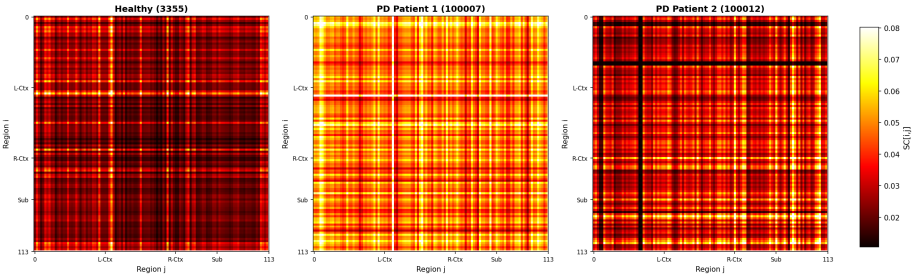

Figure 7: Examples of SC matrices built during data preprocessing

The final step to obtain structural connectivity graphs is trivial as it involves defining PyG graph structures in which nodes have feature vector (x, y, z) centroid coordinates for each ROI and edges are weighted according to the SC matrix.

3.2.5 Dataset Summary

In the following paragraphs we provide an overview of the built dataset and discuss observed metrics.

Table 1: Intersection of `(patient, visit)` Pairs by Modality

	SPECT	**MRI**	**fMRI**	**DTI**
SPECT	**3989**	973	744	1386
MRI	973	**1635**	454	470
fMRI	744	454	**1544**	604
DTI	1386	470	604	**1526**

Table 1 shows the number of `(patient, visit)` pairs for each modality along the main diagonal. The intersections represent pairs that appear in both modalities. This will appear significant in the following chapter where we provide detail of our *ImageBind*-like contrastive architecture in which all modalities are bound to a single one for contrastive learning.

Table 2: Structural and Connectivity Specifications by Neuroimaging Modality

Modality	**Type**	**Nodes**	**Node Features**	**Edges**	**Edge Metrics**
SPECT	Graph	7	SBR	6	Unitary
MRI	Graph	113	Subcortical Volume; Cortical Area & Thickness	12769	$1 - (d/d_{max})$
fMRI	Graph	113	(x, y, z)	12769	BOLD r
DTI	Graph	113	(x, y, z)	12769	FA Product

Table 2 displays the final data structures obtained from data preprocessing. GNNs tend to benefit from sparse graphs therefore connectivity graph sparsification will be explored as a hyperparameter.

Table 3: Number of `(patient, visit)` Pairs by Group

	SPECT	**MRI**	**fMRI**	**DTI**
PD	331	779	727	228
Control	3658	205	66	1298
Prodromal	0	651	751	0

Table 3 reports group counts for each modality.

The plots represented by Figure 8 show the distributions of MDS-UPDRS scores for each part based on neuroimaging modality. Due to the nature of non-motor symptoms in PD, part-1 shows evenly distributed scores. Non-motor symptoms often stabilize early in disease progression, therefore the absence of the prodromal group for SPECT and DTI scans (as shown in Table 3), doesn't impact the data distribution for this part of the scale. Distributions for part-2 and part-3 show higher values for DTI and SPECT scans. This once again aligns with the nature of the disease because motor symptoms worsen with disease progression and the absence of a prodromal group for the two modalities causes data to be distributed towards higher disease severity. Scores for part-4 become once again evenly distributed across modalities because they increase towards the most advanced stages of PD therefore they might be very similar for prodromal and PD patients.

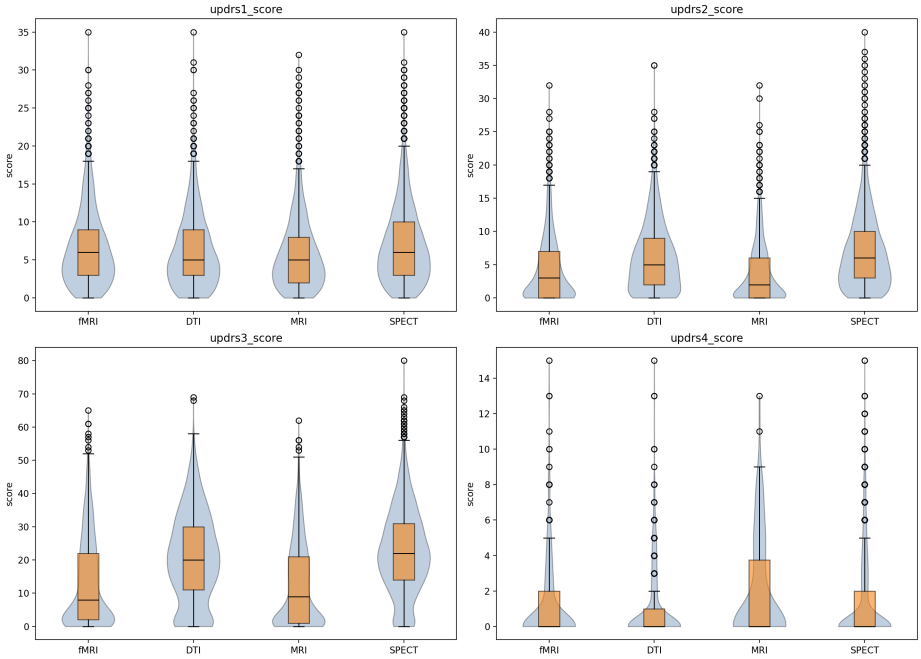

Figure 8: Plots Displaying MDS-UPDRS Distributions by Modality

Figure 9 show distributions of node features for each feature in MRI and SPECT graphs. MRI feature 1 relates to surface area and follows a gaussian distribution. Feature 2 represents the typical distribution of cortical thickness values, highlighting that different lobes of the brain (e.g. temporal and frontal) have different thickness. The subcortical region is small compared to the rest of the brain and this can be seen in feature 3, where the distribution is strongly skewed left. The widely spread gaussian distribution of the SPECT node features properly reflect the varying degrees of dopaminergic neuron loss typical of PD.

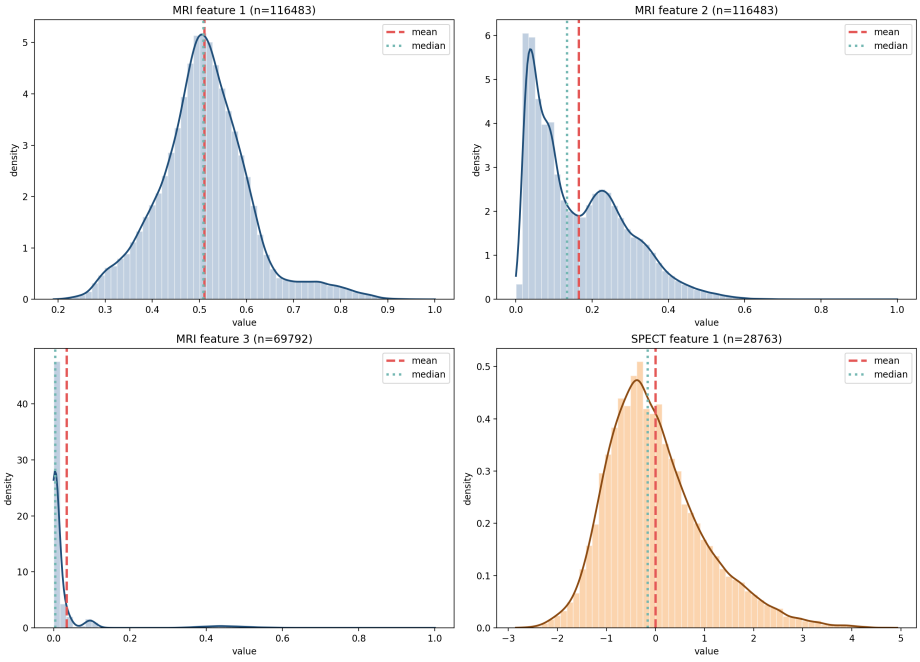

Figure 9: Distributions of Node Features in MRI and SPECT Graphs

Figure 10 displays the distributions of edge weights for DTI and fMRI graphs. The first distribution reflects SC matrices that are known to be sparse and negative connection wouldn't have a physical meaning when measuring white matter tract integrity. FC matrices have more strong connections and BOLD signals in different ROIs that can be negatively correlated.

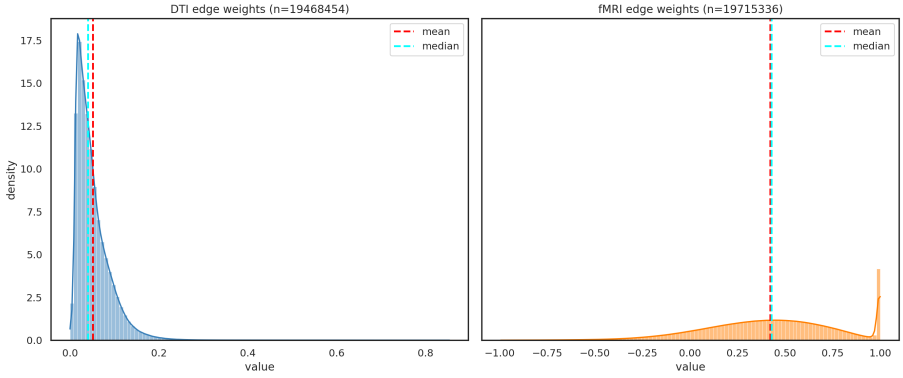

Figure 10: Distributions of Edge Weights in fMRI and DTI Graphs

3.3 Model Architecture

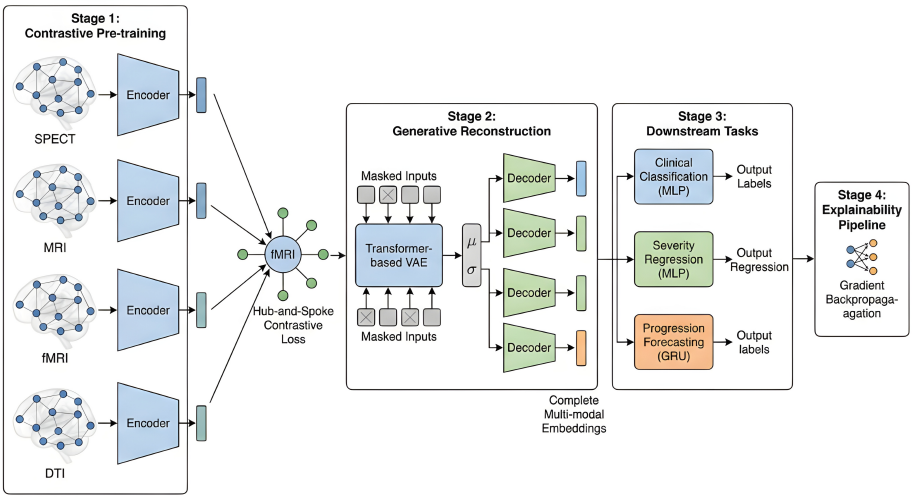

Figure 11: Visual representation of developed architecture

Our model architecture, represented in Figure 11, is deeply inspired by *ProM3E*. The *ProM3E*-like architecture is the most suitable for the objectives of this thesis for the following features:

- Contrastive learning: one of the advantages of contrastive learning is that it is able to learn embeddings that are expressive for downstream tasks even with a small number of samples. Furthermore, the architecture proposed by *ProM3E* follows an *ImageBind*-like architecture, binding every modality to a central hub. This proves useful due to the fact that the total intersection of our four modalities is small compared to pair-wise intersections.
- Modality reconstruction: the *ProM3E* proposes an architecture able to reconstruct embeddings of missing modalities from modalities that exist. This would allow us to build a larger or complete dataset in which a `(patient_visit)` pair has more or every associated modality.
- Explainability: even though it is not directly implemented in the architecture of *ProM3E*, the model is highly suitable to implement explainability because it permits us to use backpropagation and compute downstream task prediction gradients with respect to input features.

Global Notation Setup: To ensure mathematical clarity and consistency across the different modules of the architecture (Encoders, Contrastive Learning, and Generative Reconstruction), we define the following global notations:

- B: The batch size.
- $i, j \in \{1, \ldots, B\}$: Indices representing specific patients (or graph samples) within a batch.
- $\mathcal{M}$: The set of all possible modalities (SPECT, MRI, fMRI, DTI), indexed by $m \in \{1, 2, 3, 4\}$.
- $D = 1024$: The final embedding dimension.
- N: The number of nodes in a given brain graph, with individual nodes indexed by $u, v \in \{1, \ldots, N\}$.
- $\mathbf{z}_i^{(m)} \in \mathbb{R}^D$: The D-dimensional embedding for patient i's modality m.
- $\mathbf{Z}_{in} \in \mathbb{R}^{B \times 4 \times D}$: The combined input tensor containing the embeddings for an entire batch.
- $mask^{(m)} \in \{0, 1\}$: A boolean indicator where $mask^{(m)} = 1$ denotes that modality m is missing or artificially masked.

3.3.1 Encoders

Encoders must minimize intra-modality similarity. If the embeddings for a given modality are too similar to one another, the contrastive objective will fail to pair scans of the same patient, preferring to group embeddings by modality and creating four separate clusters. Given our preprocessed data, where SPECT and MRI graphs have edge weights that are identical across every patient, it is appropriate to define encoders that capture node features. Whereas DTI and fMRI graphs have identical node features but differ by edge weights, therefore, the encoders should capture variance in edge weights.

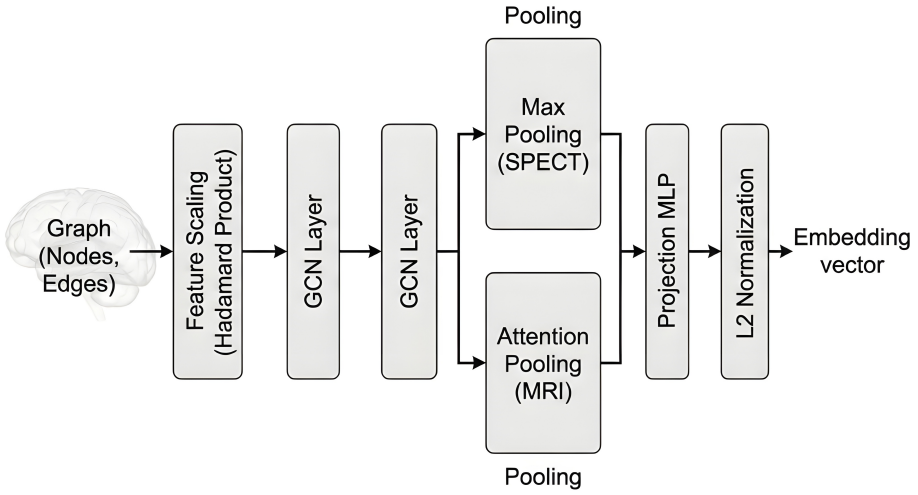

Figure 12: Visual representation of general encoder architecture implemented for SPECT and MRI graphs

The **SPECT Encoder** represented in Figure 12 implements feature scaling, which allows the model to learn relative importance of the features to each node relative to a baseline. The network multiplies the input features by a learned parameter. The scaled feature matrix $\mathbf{X}^{(0)}$ is obtained via element-wise (Hadamard) multiplication with the learned scaling matrix $\mathbf{S} \in \mathbb{R}^{N \times d_{in}}$:

$$\mathbf{X}^{(0)} = \mathbf{X} \odot \mathbf{S}$$

For an individual node v, this is simply $\mathbf{x}_v^{(0)} = \mathbf{x}_v \cdot s_v$. Two GCN layers with message passing described by Equation 2 updates node features of the scaled feature matrix. A pooling layer then converts the node representations into a single 128-dimension representation for the entire graph by taking the element-wise maximum across all N nodes of the graph $\mathbf{z} = \max_{v \in \{1,\ldots,N\}} \mathbf{x}_v^{(2)}$. Finally the vector is projected to a dimensional space of desired length by passing through an MLP and the model projects the vector onto a unit hypersphere by dividing it by its L2 norm.

Similarly to the SPECT encoder, the **MRI Encoder** implements feature scaling and a 2 layer GCN. The main difference is that the MRI encoder uses a global attention pooling strategy instead of the max attention implemented in the SPECT encoder. Each node vector $\mathbf{x}_v^{(2)}$ is passed through a linear layer ($\mathbf{W}_{gate} \in \mathbb{R}^{128\times1}$, $b_{gate} \in \mathbb{R}$) and a Sigmoid activation (σ) to get a bounded raw score $c_v = \sigma(\mathbf{x}_v^{(2)}\mathbf{W}_{gate} + b_{gate})$. Softmax is applied over all nodes u in the graph to convert these raw scores into attention weights α_v that sum to 1:

$$\alpha_v = \frac{\exp(c_v)}{\sum_{u=1}^{N} \exp(c_u)}$$

The final graph-level embedding $\mathbf{z} \in \mathbb{R}^{128}$ is the attention-weighted sum of all node features:

$$\mathbf{z} = \sum_{v=1}^{N} \alpha_v \mathbf{x}_v^{(2)}$$

Projection and L2 normalization remain unchanged.

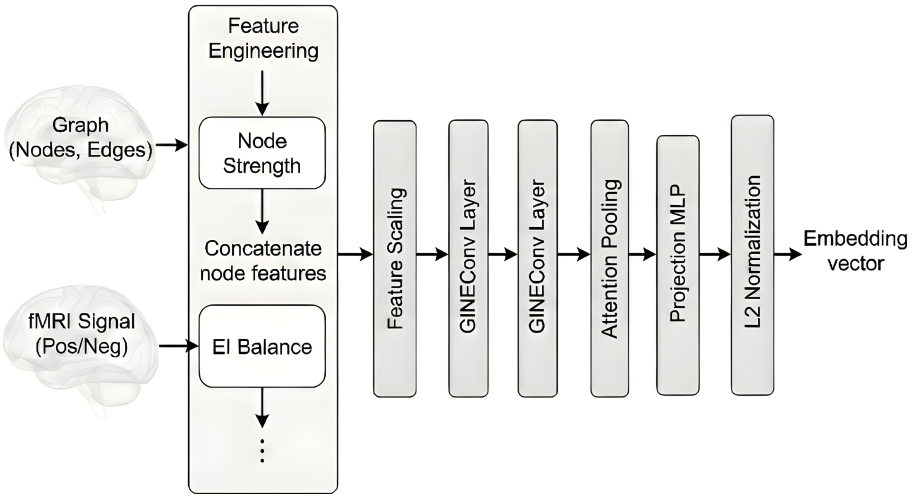

Figure 13: Visual representation of general encoder architecture implemented for DTI and fMRI graphs

The **DTI Encoder** (Figure 13) starts by calculating the total incoming structural weight for each node $w_{\text{deg},v} = \sum_{u\in\mathcal{N}(v)} e_{uv}$, concatenating the value to the original feature vector for each node. After feature scaling, the encoder performs GINEConv with message passing:

$$\mathbf{h}_v^{(gine)} = f_\Theta\left((1+\epsilon_{train})\mathbf{x}_v^{(2)} + \sum_{u\in\mathcal{N}_{mask}(v)} \text{ReLU}\left(\mathbf{x}_u^{(2)} + \mathbf{e}_{uv}^{(emb)}\right)\right)$$

and concludes with attention pooling, projection, and L2 normalization.

The **fMRI Encoder** has to handle negative values since negative correlation between brain activity is biologically feasible. It does so by separating incoming signal to each node in negative and positive and aggregates node strength similarly to the DTI encoder. Additionally it computes Excitatory-Inhibitory balance:

$$EI_v = \frac{w_{pos,v} - w_{neg,v}}{w_{pos,v} + w_{neg,v} + 10^{-6}}$$

And concatenates the three new features (negative node strength, positive node strength, EI balance) with the original features of each node. The subsequent steps to derive an embedding are identical to the DTI encoder.

3.3.2 Contrastive Learning

The contrastive learning module depicted in Figure 14 follows a Hub-and-Spoke architecture designating one modality as a central "hub" and aligning the other three "spoke" modalities to it. We specialize our notation for this phase:

- Let H denote the Hub modality (e.g., fMRI) and S denote a Spoke modality (e.g., DTI).
- Let $f_H(\cdot)$ and $f_S(\cdot)$ be the respective neural network encoders.
- Let $\mathbf{X}_H^{(i)}$ and $\mathbf{X}_S^{(i)}$ be the input graphs for the i-th patient in the batch for the Hub and Spoke modalities.
- Let $\mathbf{z}_H^{(i)} = f_H(\mathbf{X}_H^{(i)})$ and $\mathbf{z}_S^{(i)} = f_S(\mathbf{X}_S^{(i)})$ be the resulting D-dimensional embeddings.

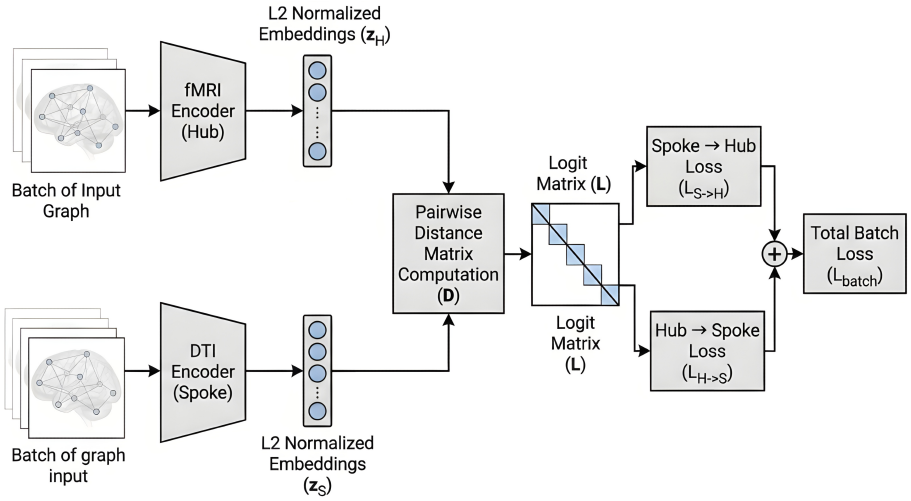

Figure 14: Visual representation of the contrastive training phase implemented by the model

Training proceeds with the following steps: **L2 Normalization** is repeated for safety even though encoders already perform it:

$$\mathbf{z}_S^{(i)} = \frac{\mathbf{z}_S^{(i)}}{\|\mathbf{z}_S^{(i)}\|_2}, \quad \mathbf{z}_H^{(i)} = \frac{\mathbf{z}_H^{(i)}}{\|\mathbf{z}_H^{(i)}\|_2}$$

Distance Matrix Computation: calculates the pairwise Euclidean (L2) distance between every Spoke embedding and every Hub embedding in the batch. Let $\mathbf{D} \in \mathbb{R}^{B \times B}$ be the distance matrix, where an element d_{ij} is the distance between the i-th patient's Spoke scan and the j-th patient's Hub scan:

$$d_{ij} = \|\mathbf{z}_S^{(i)} - \mathbf{z}_H^{(j)}\|_2$$

Logit Scaling: The distances are converted into logits (unnormalized log-probabilities) using two hyperparameters: α (scale) and β (shift). Let $\mathbf{L} \in \mathbb{R}^{B \times B}$ be the logit matrix:

$$l_{ij} = \alpha \cdot d_{ij} + \beta$$

Bidirectional Cross-Entropy Loss: The goal of the contrastive loss is to ensure that the i-th Spoke embedding is closest to the i-th Hub embedding (the diagonal of the matrix $\mathbf{L}$). It applies standard Cross Entropy in both directions (Spoke $\rightarrow$ Hub, and Hub $\rightarrow$ Spoke). For a specific row i (Spoke $\rightarrow$ Hub), the loss is:

$$\mathcal{L}_{S \to H}^{(i)} = -\log \left(\frac{\exp(l_{ii})}{\sum_{j=1}^{B} \exp(l_{ij})} \right)$$

For a specific column j (Hub $\rightarrow$ Spoke), the loss is:

$$\mathcal{L}_{H \to S}^{(j)} = -\log \left(\frac{\exp(l_{jj})}{\sum_{i=1}^{B} \exp(l_{ij})} \right)$$

The final symmetric loss for the batch is the average of the two directions over all B samples:

$$\mathcal{L}_{batch} = \frac{1}{2B} \sum_{i=1}^{B} \mathcal{L}_{S \to H}^{(i)} + \frac{1}{2B} \sum_{j=1}^{B} \mathcal{L}_{H \to S}^{(j)}$$

During **Training** we define Spokes and at every epoch, for every Spoke modality:

- Sample a batch of paired graphs $(\mathbf{X}_S, \mathbf{X}_H)$.
- Apply Graph Augmentations. This introduces small random variations to the graphs (masking edges, jittering node features) to prevent overfitting and force the model to learn robust representations.
- Compute embeddings $\mathbf{Z}_S$ and $\mathbf{Z}_H$.
- Calculate the contrastive loss $\mathcal{L}_{batch}$.
- Compute gradients: $\nabla \mathcal{L}_{batch}$.
- Gradient Clipping: To prevent exploding gradients, the norm of the gradients is bounded. If $\|\nabla \mathcal{L}\|_2 >$ clip_val (default 1.0), then $\nabla \mathcal{L} \leftarrow \frac{\nabla \mathcal{L}}{\|\nabla \mathcal{L}\|_2} \cdot$ clip_val.
- Update the weights of both f_S and f_H using the AdamW optimizer.

3.3.3 Missing Modality Reconstruction

The module (represented in Figure 15) responsible for reconstructing embeddings of missing modalities for a certain `(patient_visit)` pair from the modalities that the sample contains acts as a Masked Variational Autoencoder (VAE) built on a Transformer backbone. Its primary job is to take the embeddings generated by the encoders previously weighted by the contrastive learning module, mask out some of those modalities, and then force the network to reconstruct the missing modalities from a shared probabilistic latent space.

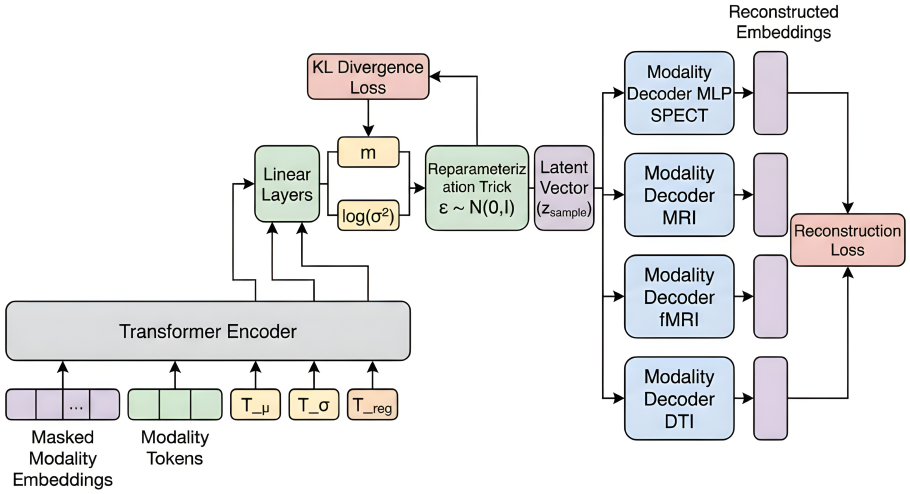

Figure 15: Visual representation of the generative module implemented

Architecture: First, the model adds a learnable modality token to distinguish which embedding belongs to which modality. Let $\mathbf{T}_{mod}^{(m)} \in \mathbb{R}^D$ be the modality token for modality m. Each modality is then passed through its own specific MLP projector $f_{proj}^{(m)}$:

$$\mathbf{h}^{(m)} = f_{proj}^{(m)}\left(\mathbf{z}_{in}^{(m)} + \mathbf{T}_{mod}^{(m)}\right)$$

To train the model to infer missing data, the embeddings of the masked modalities are replaced entirely by a shared, learnable mask token $\mathbf{T}_{mask} \in \mathbb{R}^D$:

$$\tilde{\mathbf{h}}^{(m)} = \begin{cases} \mathbf{T}_{mask} & \text{if } mask^{(m)} = 1 \\ \mathbf{h}^{(m)} & \text{if } mask^{(m)} = 0 \end{cases}$$

The model constructs a sequence to feed into the Transformer. It prepends several special learnable tokens to the modality sequence:

- $\mathbf{T}_\mu \in \mathbb{R}^D$: A token dedicated to aggregating the latent mean.

- $\mathbf{T}_\sigma \in \mathbb{R}^D$: A token dedicated to aggregating the latent variance.
- $\mathbf{T}_{reg} \in \mathbb{R}^{R \times D}$: A set of $R = 4$ "register tokens". These give the Transformer extra blank scratchpad space to process global information without corrupting the modality representations.

The final input sequence $\mathbf{X}_{seq}$ for a single patient has a length of $S = 1+1+4+4 = 10$ tokens:

$$\mathbf{X}_{seq} = [\mathbf{T}_\mu \parallel \mathbf{T}_\sigma \parallel \mathbf{T}_{reg} \parallel \tilde{\mathbf{h}}^{(1)} \parallel \tilde{\mathbf{h}}^{(2)} \parallel \tilde{\mathbf{h}}^{(3)} \parallel \tilde{\mathbf{h}}^{(4)}]$$

The sequence is passed through a standard Transformer Encoder. Let the output sequence be $\mathbf{Y}_{seq} \in \mathbb{R}^{10 \times D}$:

$$\mathbf{Y}_{seq} = \text{TransformerEncoder}(\mathbf{X}_{seq})$$

To create a generative probabilistic space, the network extracts the first two outputs of the Transformer sequence (corresponding to the $\mathbf{T}_\mu$ and $\mathbf{T}_\sigma$ tokens). Let these be $\mathbf{y}_\mu$ and $\mathbf{y}_\sigma$. These are passed through linear layers to compute the parameters of a multivariate Gaussian distribution:

$$\boldsymbol{\mu} = \mathbf{y}_\mu \mathbf{W}_\mu + \mathbf{b}_\mu$$

$$\log \boldsymbol{\sigma}^2 = \text{clamp}(\mathbf{y}_\sigma \mathbf{W}_\sigma + \mathbf{b}_\sigma, \text{min} = -10, \text{max} = 10)$$

Using the Reparameterization Trick, the model samples a single global latent representation $\mathbf{z}_{sample}$ by drawing standard normal noise $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \mathbf{I})$:

$$\mathbf{z}_{sample} = \boldsymbol{\mu} + \boldsymbol{\epsilon} \odot \exp\left(\frac{1}{2} \log \boldsymbol{\sigma}^2\right)$$

Finally, this single unified latent vector $\mathbf{z}_{sample}$ is used to reconstruct all four modalities, regardless of whether they were originally masked. The latent vector is passed through four separate modality-specific decoder MLPs $f_{dec}^{(m)}$:

$$\hat{\mathbf{z}}^{(m)} = f_{dec}^{(m)}(\mathbf{z}_{sample})$$

The output is the reconstructed embeddings for SPECT, MRI, fMRI, and DTI.

Training: To train this system, the total loss $\mathcal{L}_{total}$ balances two competing objectives: accurate reconstruction and a well-behaved latent space. The model uses a cross-entropy contrastive loss. For every active modality m, it computes the pairwise distance d_{ij} between all predictions $\hat{\mathbf{z}}_i^{(m)}$ and all ground truths $\mathbf{z}_{gt,j}^{(m)}$ in the batch:

$$d_{ij} = \|\hat{\mathbf{z}}_i^{(m)} - \mathbf{z}_{gt,j}^{(m)}\|_2$$

These distances are scaled into logits l_{ij}. The model is heavily penalized if a patient's reconstruction is closer to a different patient's ground truth:

$$\mathcal{L}_{recon}^{(m)} = -\frac{1}{B} \sum_{i=1}^{B} \log\left(\frac{\exp(l_{ii})}{\sum_{j=1}^{B} \exp(l_{ij})}\right)$$

To ensure the latent space is continuous and can be sampled from, the distribution defined by $\boldsymbol{\mu}_i$ and $\log \boldsymbol{\sigma}_i^2$ is forced to match a standard normal prior $\mathcal{N}(0, \mathbf{I})$. The model computes KL Divergence:

$$\mathcal{L}_{KL} = -\frac{1}{2B} \sum_{i=1}^{B} \sum_{d=1}^{D} \left(1 + \log \sigma_{i,d}^2 - \mu_{i,d}^2 - \exp(\log \sigma_{i,d}^2)\right)$$

To prevent posterior collapse, the divergence penalty is slowly scaled up using a warmup factor λ_{warmup} over the first few epochs:

$$\mathcal{L}_{total} = \mathcal{L}_{recon} + \lambda_{warmup} \cdot \mathcal{L}_{KL}$$

Reconstruction: The final step to reconstruct missing modalities involves masking out the actually missing modalities. The module initializes an empty input tensor $\mathbf{X}_{in} \in \mathbb{R}^{1 \times 4 \times D}$ and a full boolean mask $\mathbf{mask} \in \{0, 1\}^4$. For each modality $m \in \{1, 2, 3, 4\}$, it populates the tensor and updates the mask based on what data the patient actually has available ($\mathcal{M}_{avail}$):

$$\mathbf{X}_{in}^{(m)} = \begin{cases} \mathbf{z}_{real}^{(m)} & \text{if } m \in \mathcal{M}_{avail} \\ \mathbf{0} & \text{if } m \notin \mathcal{M}_{avail} \end{cases}$$

$$mask^{(m)} = \begin{cases} \text{False (0)} & \text{if } m \in \mathcal{M}_{avail} \\ \text{True (1)} & \text{if } m \notin \mathcal{M}_{avail} \end{cases}$$

The populated tensor and mask are passed through the Generator with gradients disabled. As explained in the previous architecture, the Generator replaces the missing indices with the $\mathbf{T}_{mask}$ token, funnels the sequence through the Transformer, samples from the latent space defined by $\boldsymbol{\mu}$, and decodes it back into four complete embeddings:

$$\hat{\mathbf{Z}}, \boldsymbol{\mu}, \_ = \text{Generator}(\mathbf{X}_{in}, \mathbf{mask})$$

Even though the Generator reconstructs all modalities, its reconstruction of an MRI will never be quite as perfect as the actual MRI the patient took. To create the highest-fidelity profile for the patient, the script performs a "Smart Reconstruction." It iterates through the results and builds a final composite dictionary by keeping the ground truth where it exists, and only using the hallucinated data where there are gaps:

$$\mathbf{z}_{final}^{(m)} = \begin{cases} \mathbf{z}_{real}^{(m)} & \text{if } m \in \mathcal{M}_{avail} \text{ (Keep Real)} \\ \hat{\mathbf{z}}^{(m)} & \text{if } m \notin \mathcal{M}_{avail} \text{ (Use Hallucinated)} \end{cases}$$

These completed profiles are then saved to disk and the final result is a dataset where every single patient has a complete set of SPECT, MRI, fMRI, and DTI embeddings, ready to be used for downstream tasks like disease classification or survival prediction.

3.3.4 Downstream Tasks

The ultimate objective of our framework is to utilize the aligned, imputed multimodal embeddings to perform clinically relevant predictions. Our framework implements three distinct downstream tasks (as represented by Figure 16): Clinical Classification (diagnosis), Severity Regression (current UPDRS scoring), and Disease Progression Forecasting (predicting future UPDRS scores).

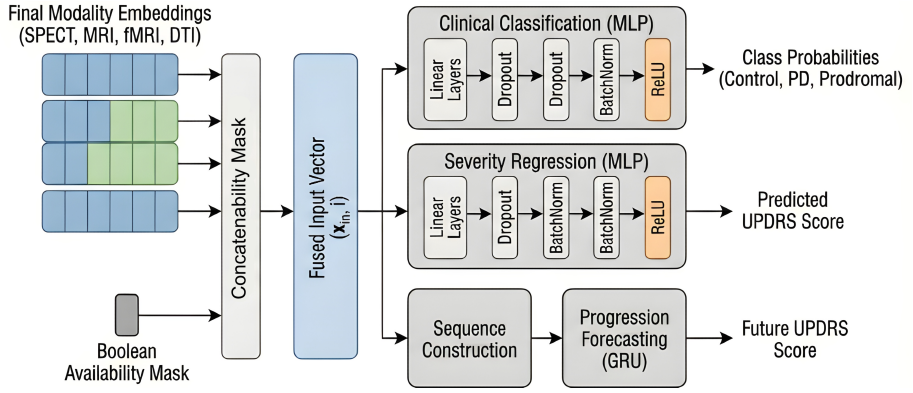

Figure 16: Visual representation of downstream tasks implemented

Data Fusion and Mask Appending: Before feeding data to any of the downstream networks, the patient profile is flattened into a single, comprehensive feature vector. For a given patient i, the embeddings for all $M = 4$ modalities are concatenated sequentially:

$$\mathbf{f}_{concat,i} = [\mathbf{z}^{(1)}_{final,i} \parallel \mathbf{z}^{(2)}_{final,i} \parallel \mathbf{z}^{(3)}_{final,i} \parallel \mathbf{z}^{(4)}_{final,i}]$$

Because each embedding has a dimension of $D = 1024$, $\mathbf{f}_{concat,i} \in \mathbb{R}^{4096}$.

To provide the network with explicit context regarding which data is biologically real and which was hallucinated by the generative module, a boolean availability mask $\mathbf{mask}_i \in \mathbb{R}^4$ is appended to the feature vector. The final input vector $\mathbf{x}_{in,i}$ for a single patient visit becomes a 4100-dimensional vector:

$$\mathbf{x}_{in,i} = [\mathbf{f}_{concat,i} \parallel \mathbf{mask}_i] \in \mathbb{R}^{4100}$$

Clinical Classification (Diagnosis): The classification module aims to categorize patients into one of three cohorts: Control, PD, or Prodromal. The model compresses the 4100-dimensional input through a Multi-Layer Perceptron (MLP) utilizing batch normalization and heavy dropout to prevent overfitting.

Let $\mathbf{W}_1 \in \mathbb{R}^{4100\times512}$ and $\mathbf{W}_2 \in \mathbb{R}^{512\times128}$ be the weight matrices for the hidden layers, with corresponding biases $\mathbf{b}_1, \mathbf{b}_2$. The hidden representations are computed as follows:

$$\mathbf{h}_{1,i} = \text{Dropout}_{0.5}(\text{ReLU}(\text{BatchNorm}(\mathbf{x}_{in,i}\mathbf{W}_1 + \mathbf{b}_1)))$$

$$\mathbf{h}_{2,i} = \text{Dropout}_{0.3}(\text{ReLU}(\text{BatchNorm}(\mathbf{h}_{1,i}\mathbf{W}_2 + \mathbf{b}_2)))$$

The final logits $\mathbf{y}_{logits,i} \in \mathbb{R}^3$ are obtained via a linear projection $\mathbf{W}_3 \in \mathbb{R}^{128\times 3}$:

$$\mathbf{y}_{logits,i} = \mathbf{h}_{2,i}\mathbf{W}_3 + \mathbf{b}_3$$

To counteract the inherent class imbalance in medical datasets, the module employs Inverse Class Weighting and Label Smoothing. A static weight w_c is calculated for each class c based on the total training samples N_{train} and the class frequency N_c: $w_c = \frac{N_{train}}{3 \cdot N_c}$. The ground truth label y_c is smoothed with a penalty margin $\alpha = 0.1$ to yield $y_c^{ls} = (1-0.1)y_c + \frac{0.1}{3}$. The network is optimized using Weighted Cross-Entropy Loss:

$$\mathcal{L}_{CE} = -\frac{1}{B}\sum_{i=1}^{B}\sum_{c=1}^{3} w_c y_{c,i}^{ls} \log(p_{c,i})$$

where $p_{c,i}$ is the softmax probability of class c for patient i. Evaluation is performed using Balanced Accuracy, defined as the unweighted macro-average of the recall for each class.

Severity Regression: The regression module predicts continuous clinical severity scores. It utilizes the exact same MLP architecture as the classifier, but the final linear layer projects $\mathbf{h}_{2,i}$ to a single scalar output $\hat{y}_i \in \mathbb{R}$.

To ensure neural network stability during training, the ground truth UPDRS scores y_i are scaled down by a factor of 100 ($y_{scaled,i} = y_i/100$). The network is optimized using Masked Mean Squared Error (MSE), computing loss only on non-missing targets:

$$\mathcal{L}_{MSE} = \frac{1}{N_{valid}} \sum_{i \in \text{valid}} (\hat{y}_i - y_{scaled,i})^2$$

The regressor is evaluated using the Coefficient of Determination (R^2), which measures the proportion of variance in the dependent variable predictable from the features:

$$R^2 = 1 - \frac{\text{MSE}}{\text{Var}(y_{scaled})}$$

Disease Progression Forecasting: Unlike the static tasks above, progression forecasting leverages chronological patient histories to predict future clinical decline. Let a patient have a history of T past visits. The input is a sequence of fused embeddings $\mathbf{X}_{hist} = [\mathbf{x}_{in,1}, \mathbf{x}_{in,2}, \ldots, \mathbf{x}_{in,T}]$.

First, each visit vector is independently compressed to a 256-dimensional space using a linear layer and LayerNorm:

$$\mathbf{c}_t = \text{Dropout}_{0.5}(\text{ReLU}(\text{LayerNorm}(\mathbf{x}_{in,t}\mathbf{W}_c + \mathbf{b}_c))) \in \mathbb{R}^{256}$$

To account for irregular intervals between clinical visits, the time elapsed since the previous visit $\delta_t = \text{year}_t - \text{year}_{t-1}$ is explicitly encoded into a 32-dimensional temporal embedding:

$$\tau_t = \text{ReLU}(\delta_t \mathbf{W}_\tau + \mathbf{b}_\tau) \in \mathbb{R}^{32}$$

The compressed feature and temporal embedding are concatenated to form the input $\mathbf{u}_t = [\mathbf{c}_t \parallel \tau_t] \in \mathbb{R}^{288}$ for a 2-layer Gated Recurrent Unit (GRU). The GRU updates its hidden state sequentially:

$$\mathbf{h}_t = \text{GRU}(\mathbf{u}_t, \mathbf{h}_{t-1}) \in \mathbb{R}^{256}$$

To forecast a future UPDRS score at a target time interval Δt_{pred} in the future, the final hidden state $\mathbf{h}_T$ (representing the patient's entire clinical trajectory) is concatenated with Δt_{pred}. This vector is passed through a predictive MLP head to output the scaled future score:

$$\hat{y}_{future} = f_{head}([\mathbf{h}_T \parallel \Delta t_{pred}])$$

This sequence model is similarly trained using MSE and evaluated using the R^2 metric.

3.3.5 Explainability and Global Biomarker Aggregation

To transition from a "black-box" predictive model to a clinically interpretable tool, we implemented a unified gradient-based feature attribution module. Because the computational graph remains intact from raw anatomical inputs through the modality-specific encoders, across the generator, and up through the downstream predictors, we can compute the sensitivity of any clinical prediction with respect to the original brain structures and their specific features.

Notation Setup: We define the interpretability metrics as follows:

- $\mathcal{L}_{task}$: The scalar output from the downstream task. For classification, this is the pre-softmax logit of the predicted class; for regression, the predicted clinical score ($\hat{y}$).
- $\mathbf{x}_{v,d}^{(m)} \in \mathbb{R}$: The raw input for node v, modality m, and feature dimension d (e.g., $d \in \{\text{Area, Thickness, Volume}\}$ for MRI).
- $e_{uv}^{(m)} \in \mathbb{R}$: The raw input edge weight between node u and node v in modality m.
- $\mathbf{g}_{v,d}^{(m)} = \frac{\partial \mathcal{L}_{task}}{\partial \mathbf{x}_{v,d}^{(m)}}$: The partial derivative of the task output with respect to a specific biological feature.

To determine which neurological regions and features drive a prediction, we utilize backpropagation to trace gradients from $\mathcal{L}_{task}$ back to the input graphs. The module calculates three distinct metrics to interpret the biological drivers:

Feature-Level Sensitivity (Gradient): Unlike traditional methods that sum gradients across dimensions, we preserve the directional sensitivity for each feature d:

$$\text{Grad}_{v,d}^{(m)} = \mathbf{g}_{v,d}^{(m)}$$

A highly positive gradient implies that an increase in that specific feature (e.g., increased Volume) correlates with an increase in the predicted score.

Feature Importance (Absolute Saliency): To identify the most impactful elements regardless of direction, we compute the absolute saliency:

$$I_{v,d}^{(m)} = \left|\mathbf{g}_{v,d}^{(m)}\right|, \quad I_{edge,uv}^{(m)} = \left|\frac{\partial \mathcal{L}_{task}}{\partial e_{uv}^{(m)}}\right|$$

Local Contribution (Input $\times$ Gradient): We utilize the Input $\times$ Gradient method to measure the true localized contribution of a patient's specific anatomy to their prediction:

$$C_{v,d}^{(m)} = \mathbf{x}_{v,d}^{(m)} \cdot \mathbf{g}_{v,d}^{(m)}$$

This identifies whether the model is reacting to the *presence* or *absence* (atrophy) of a signal.

Longitudinal and Generative Attribution: For the progression forecasting task, node and edge metrics are averaged across all available historical time steps T to derive a unified importance score:

$$I_{node,v,d}^{(m,global)} = \frac{1}{T} \sum_{t=1}^{T} I_{node,v,d,t}^{(m)}$$

Furthermore, if a modality is missing, the gradient is backpropagated through the Transformer generator to the available "anchor" modalities. This allows us to interpret how the model utilizes available context to synthesize missing biomarkers for final prediction.

Global Biomarker Aggregation: To uncover overarching pathophysiological patterns, we aggregate individual explanations into population-level biomarkers. We define the Global Absolute Saliency as the expected value of the absolute gradient across the cohort $\mathcal{P}$:

$$\bar{I}_{v,d}^{(m)} = \frac{1}{|\mathcal{P}|} \sum_{i \in \mathcal{P}} I_{v,d}^{(m,i)}$$

Biomarker Robustness (Z-Score): A robust clinical biomarker must be consistent across the population. We quantify statistical reliability using the ratio of the Global Directional Contribution ($\bar{C}$) to its standard deviation:

$$Z_{v,d}^{(m)} = \frac{\bar{C}_{v,d}^{(m)}}{\sigma_{C,v,d} + \epsilon}$$

Regions with high absolute Z-scores represent stable global biomarkers that consistently influence clinical predictions in the same direction across the cohort.

Anatomical Mapping: The numerical importance indices are dynamically mapped to human-readable anatomy. For MRI, the module identifies the top-K (v, d) pairs and appends suffixes (e.g., `_Thickness`, `_Volume`) to the FreeSurfer ROI names. This automated reporting allows for the granular distinction between diagnostic thinning and prognostic atrophy, providing actionable insights for precision medicine.

4 Results

Achieved results (Table 4) demonstrate our model's performance in downstream tasks. Comparison with state-of-the-art models in Table 5 displays above SOTA performance for the classification task, SOTA performance for the static severity regression task, and establish a novel baseline for the progression prediction task. Studies that integrate data modalities that do not derive from neuroimaging or that do not model the brain as graphs by defining ROIs are not considered comparable studies.

Numerous studies that integrate patient clinical data, including previous severity scores, achieve significantly higher performance when predicting a patient's progression. *SAFN* [56] attributes up to 60% of performance to clinical data. *MultimodalCNN-PD* [57], which achieves the highest classification accuracy doesn't preprocess brain volumes into graphs and it feeds every voxel of the brain to a CNN. While the latter two methods achieve better performance metrics, they do not preserve the biological significance of neuroimaging data, making the process of biomarker identification through explainability more challenging. For instance, an RNN that predicts future patient severity score by using past scores may not be learning the dependency between brain structures/functions and neural degradation, relying mostly on past scores and overlooking neuroimaging data. A CNN that analyzes the brain voxel by voxel without defining ROIs increases data granularity, but loses useful biological information that characterizes each brain region.

Table 4: 5-Fold Cross-Validation Performance across Downstream Tasks ($n = 5$). Results are reported as Mean ± Standard Deviation.

Task & Metric	**Score (Mean ± SD)**
Clinical Classification	
Balanced Accuracy	0.9519 ± 0.0077
Macro F1-Score	0.8829 ± 0.0155
Macro AUC	0.9830 ± 0.0040
Static Severity Regression (R^2)	
UPDRS-II (ADL)	0.2464 ± 0.0264
UPDRS-III (Motor)	0.4440 ± 0.0341
Disease Progression Forecasting (R^2)	
UPDRS-II (ADL)	0.1722 ± 0.0478
UPDRS-III (Motor)	0.3542 ± 0.0356

Table 5: Comparison of our model performance against state-of-the-art baselines across downstream tasks. Dashes (–) indicate that a model was not evaluated on that specific task. Best results within the directly comparable (ROI-based) category are highlighted in bold. Values with (*) are reported from other studies.

		Classification	**Severity (R^2)**		**Progression (R^2)**	
Model	**Modality**	**Accuracy**	**Part II**	**Part III**	**Part II**	**Part III**
Broad SOTA (Clinical, Telemonitoring & Non-ROI Imaging)						
CTESM (2025)	EEG	*0.99	–	–	–	–
Context-Aware DNN (2026)	Voice+Clinical	–	–	*0.99	–	–
mPower-CNN (2024)	Smartphone	–	*0.62	*0.58	–	–
Smartwatch-XGB (2025)	Actigraphy	–	*0.68	–	–	–
LSTM-Lag (2024)	Clinical	–	–	–	–	*0.35
SFNFA-MVD	Multi	–	–	–	*0.38	*0.48
MultimodalCNN-PD (2026)	Multi (MRI+Clin)	*0.98	–	–	–	–
Directly Comparable Baselines (ROI-Based Neuroimaging)						
GCN	SPECT	0.93	-0.09	-0.13	0.09	0.06
GCN	MRI	0.45	-0.04	-0.03	–	–
GINE	fMRI	0.50	-0.01	-0.02	-0.12	0.08
GINE	DTI	0.74	-0.04	-0.06	0.11	0.06
GCN&GINE	Multi	0.94	0.11	-0.12		
BrainMAP	Multi	*0.86	–	–	–	–
BrainMAP (our dataset)	Multi	0.92	–	–	–	–
Chudzik et al. (2024)	MRI	*0.85	–	–	–	–
Martinez-Murcia et al. (2024)	Multi	–	–	*0.44	–	–
GeNeuro	**Multi**	**0.95**	**0.26**	**0.44**	**0.17**	**0.35**

We outline the configuration of the infrastructures used in conducting our experiments:

- Preprocessing (SPECT, MRI, fMRI): *AMD Ryzen 5 9600X* (6 cores/12 threads) and 32GB of DDR5 RAM.
- Preprocessing (DTI): *Google Cloud Compute Engine* VM instance configured with a `n2-standard-32` (32 vCPU, 128 GB of RAM) CPU along with a *Google Cloud Storage Bucket*.
- Hyperparameter tuning and ablation studies: two *NVIDIA RTX PRO 6000*
- Explainability pipeline: *AMD Ryzen 5 9600X* (6 cores/12 threads) and 32GB of DDR5 RAM.

4.1 Hyperparameter Tuning

Table 6 reports the final optimized hyperparameter configuration of the model.

Table 6: Optimized Hyperparameter Configuration

Stage	Parameter	Optimized Value
Contrastive Pre-training	`-contrastive_epochs`	100
	`-contrastive_lr`	0.00015
	`-contrastive_batch_size`	32
	`-contrastive_alpha`	-8.66
	`-contrastive_beta`	8.32
	`-hub_name`	fMRI
	`-aug_mask`	0.24
	`-aug_jitter`	0.01
	`-clip_val`	2.3
	`-hidden_dim`	256
	`-embed_dim`	1024
	`-threshold`	0.60
	`-no_contrastive_aug`	False
Generative Reconstruction	`-generator_epochs`	100
	`-generator_lr`	0.000012
	`-generator_weight_decay`	0.004
	`-generator_alpha`	-4.41
	`-generator_beta`	1.41
	`-generator_lambda`	0.0004
	`-generator_divergence`	kl
	`-gen_keep_prob`	0.5
	`-gen_kl_warmup`	12
	`-gen_hidden_dim`	1024
	`-gen_num_heads`	8
	`-gen_num_layers`	5
	`-gen_num_registers`	0
	`-gen_mlp_depth`	3
	`-gen_dropout`	0.28
Downstream Tasks	`-downstream_lr`	0.01
	`-cls_epochs`	100
	`-prog_epochs`	100
	`-updrs_epochs`	100

Given the high dimensionality of the hyperparameter space, a sequential block-wise optimization strategy was employed. Hyperparameter tuning was conducted using the *Optuna* Python library with Bayesian optimization. The model was optimized to maximize classification accuracy for the downstream classification and R^2 for the downstream severity and progression prediction tasks.

4.1.1 Phase 1: Architectural Baseline and Base Learning Rates

In the first 100 trials, the search space focused on fundamental model capacity and the initial learning trajectory. The parameters tuned in this phase were:

- **Capacity:** `-hidden_dim`, `-embed_dim`, and `-gen_num_layers`.
- **Optimization:** `-contrastive_lr`, `-generator_lr`, and `-contrastive_batch_size`.
- **Initial Regularization:** `-gen_dropout` and `-generator_lambda`.

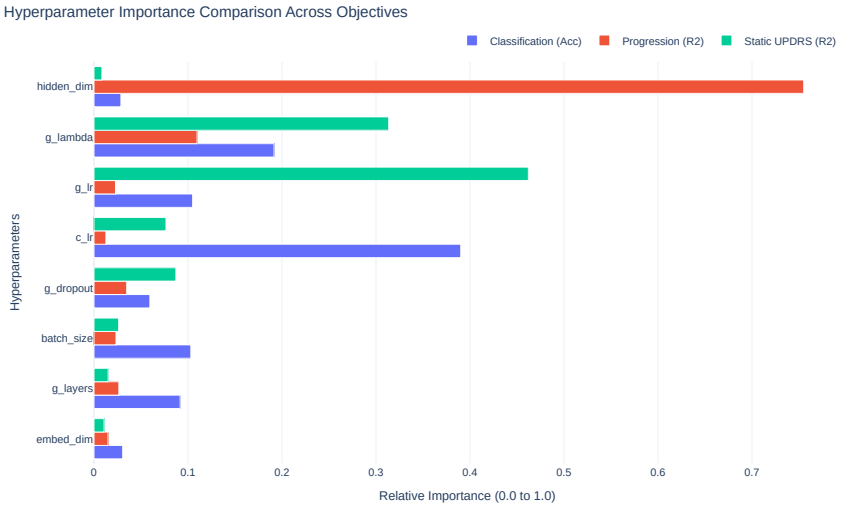

Figure 17: Comparative hyperparameter importance across the three primary downstream objectives (Phase 1)

The importance analysis (detailed in Figure 17) reveals a clear specialization of hyperparameters across the three downstream objectives. `hidden_dim` was identified as the critical driver for longitudinal progression prediction—accounting for 76.9% of the objective variance.

Classification accuracy was primarily driven by the contrastive learning rate (`c_lr`, 44.0%), indicating that the initial categorical separation is established during the pre-training phase. The prediction of static disease severity was most sensitive to the generator's optimization schedule, with `g_lr` (53.2%) and the loss-balancing term `g_lambda` (20.2%) dominating the performance.

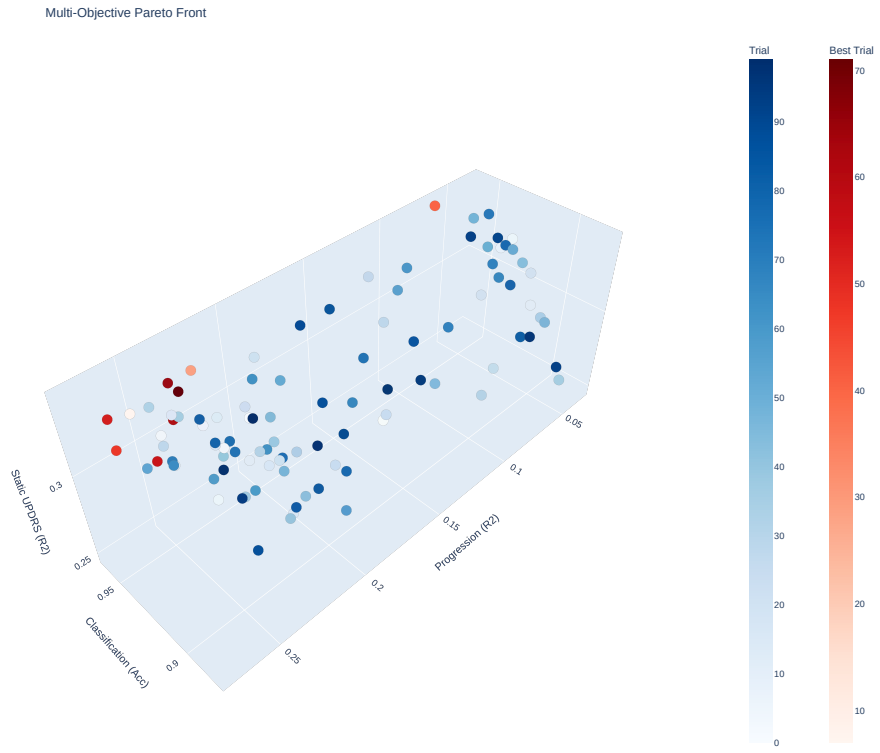

Figure 18: Pareto front illustrating the trade-off between downstream objectives (Phase 1)

`g_lambda` appeared as a top-three contributor for all three tasks balancing the reconstruction of the latent manifold against downstream task utility. Parameters representing architectural capacity, such as `g_layers` and `embed_dim`, showed moderate but stable importance (1.3% to 8.1%). Following the Pareto analysis (Figure 18), these structural parameters were anchored at 1024 for embedding dimensions and 5 for generator layers, providing a stable backbone for the regularization-focused tuning in Phase 2.

4.1.2 Phase 2: Regularization Refinement and Parameter Anchoring

The second phase introduces weight decay while narrowing the search bounds for learning rates based on the "anchors" found in Phase 1. This phase evaluates:

- **New Parameter:** `-generator_weight_decay`.
- **Narrowed Anchors:** `-contrastive_lr`, `-generator_lr`, `-generator_lambda`, and `-hidden_dim`.
- **Interaction Check:** `-gen_dropout` was retained to evaluate its codependence with weight decay.

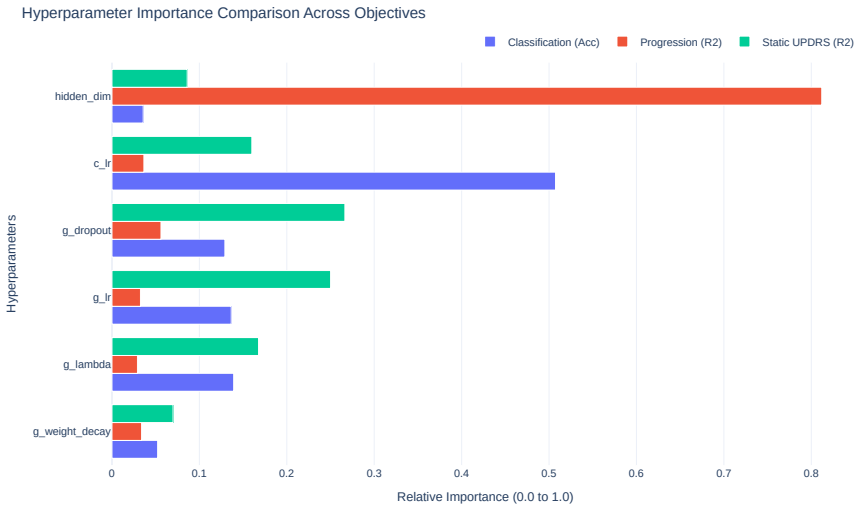

Figure 19: Comparative hyperparameter importance across the three primary downstream objectives (Phase 2)

The introduction of `-generator_weight_decay` significantly stabilized the variance observed in the regression tasks. As illustrated in Figure 19, the reliance on the generator learning rate (`g_lr`) for Static severity prediction dropped from 53.2% to 30.2%, with explicit regularization parameters (`g_dropout` at 22.0% and `g_weight_decay` at 10.1%). Furthermore, `hidden_dim` increased to an 82.1% importance share for Progression R^2, indicating that longitudinal forecasting relies primarily on architectural capacity rather than optimization speed.

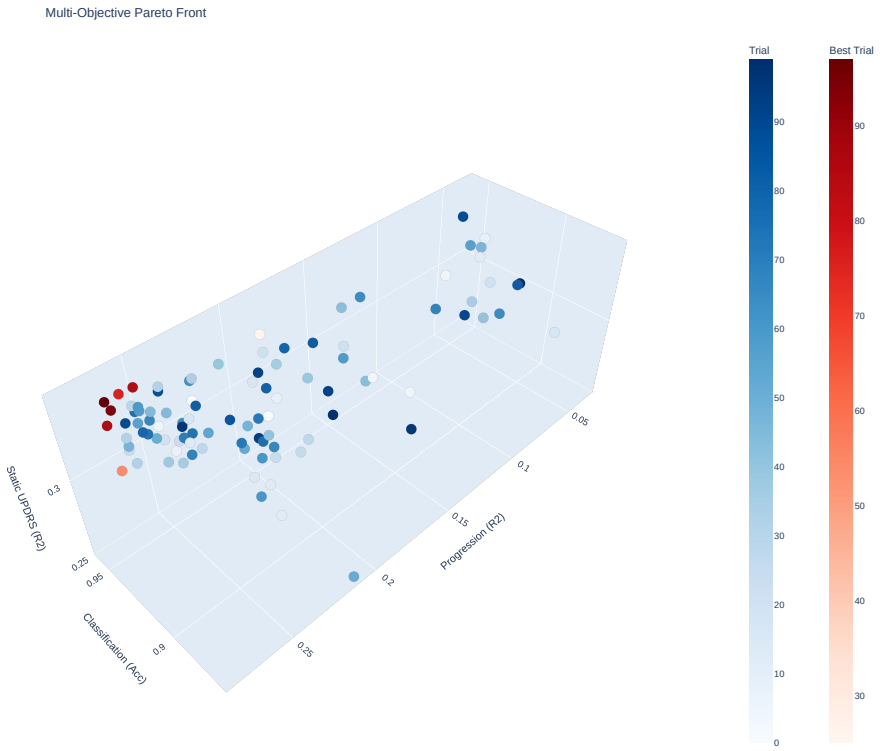

Figure 20: Pareto front illustrating the trade-off between downstream objectives (Phase 2)

By evaluating the Pareto optimal candidates (Figure 20), a compromise solution was selected that achieved 95.1% classification accuracy while maintaining robust regression performance. This locked in `hidden_dim` at 256, established heavy regularization (`g_dropout` = 0.280, `weight_decay` = 0.004164), and finalized the core learning rates for subsequent stages.

4.1.3 Phase 3: Loss Logic and Latent Space Alignment

Following the stabilization of regularization, the third phase optimized the internal weights and connectivity thresholds of the generative model. The search space was defined as follows:

- **Loss Weights:** `-contrastive_alpha`, `-contrastive_beta`, `-generator_alpha`, and `-generator_beta`.
- **Geometry & Logic:** `-threshold`, `-generator_divergence`, and `-gen_kl_warmup`.
- **Architecture Fine-tuning:** `-gen_hidden_dim`, `-gen_num_heads`, `-gen_num_registers`, and `-gen_mlp_depth`.

The resulting parameter importance analysis (Figure 21) demonstrated divergence between structural hyperparameters and loss. Variations in the internal Transformer wiring (attention heads, supplementary register tokens, and KL warmup epochs) accounted for less than 1% of objective variance across all tasks. The configuration of the loss functions, however, heavily dictated model performance.

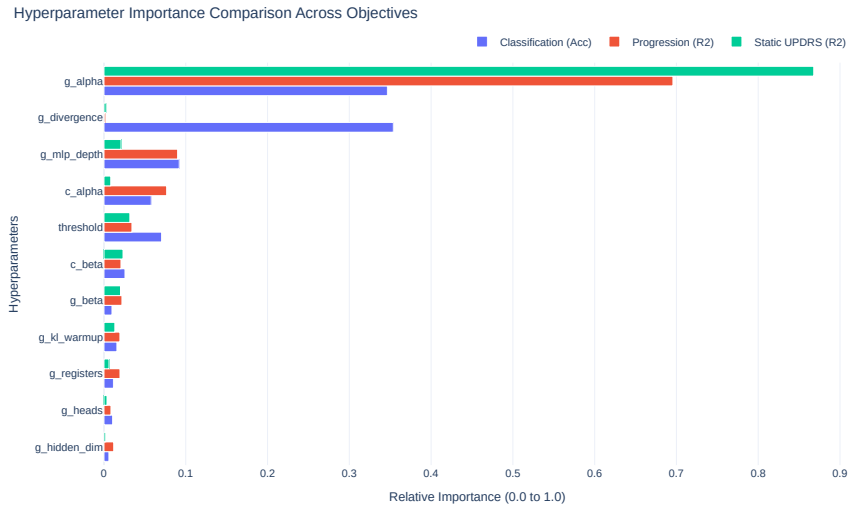

Figure 21: Comparative hyperparameter importance across the three primary downstream objectives (Phase 3)

Classification accuracy was predominantly driven by the choice of `generator_divergence` (41.8%). While empirical results showed unscaled Maximum Mean Discrepancy (MMD) achieving marginal accuracy gains by relaxing the strict Gaussian prior, Kullback-Leibler (KL) divergence was specifically

selected from the Pareto front to enforce a true Variational Autoencoder (VAE) architecture. This prevents the formation of isolated modality-specific clusters and ensures robust generative capabilities.

For the continuous regression tasks, the margin constraint `generator_alpha` exhibited absolute dominance, accounting for 64.8% of variance in progression prediction and 87.1% in static severity prediction.

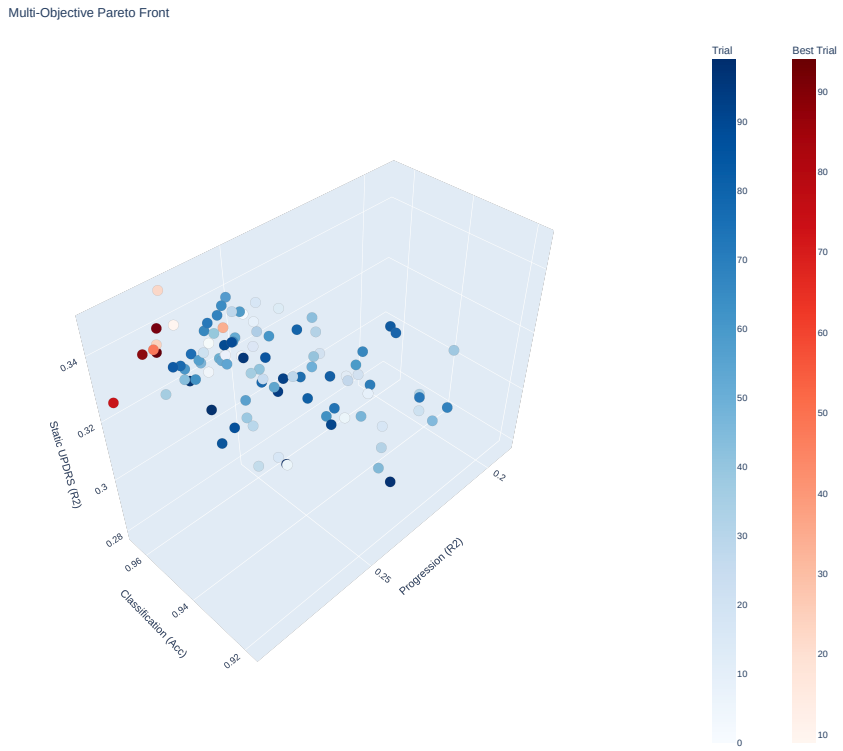

Figure 22: Pareto front illustrating the trade-off between downstream objectives (Phase 3)

The optimal compromise candidate expanded the generative capacity (`gen_hidden_dim` = 1024, `gen_num_heads` = 8) while relying purely on native embeddings without supplementary register tokens (`gen_num_registers` = 0). This locked-in KL-based configuration achieved a peak classification accuracy of 96.2%, alongside a mean static severity R^2 of 0.338 and a mean progression R^2 of 0.279, proving that a unified, continuous latent space can be achieved without sacrificing downstream clinical predictive power.

4.1.4 Phase 4: Augmentation and Final Polish

The final phase served to extract the maximum utility from the locked-in generative architecture by addressing input data robustness and the convergence of downstream task heads. This phase specifically tuned:

- **Data Robustness:** `-aug_mask` and `-aug_jitter` to regularize the contrastive pre-training against noisy neuroimaging inputs.
- **Optimization Stability:** `-clip_val` to stabilize the ultimate phase of gradient descent.
- **Downstream Specialization:** `-downstream_lr` to calibrate the learning rate of the final classifiers and regressors independent of the generative backbone.

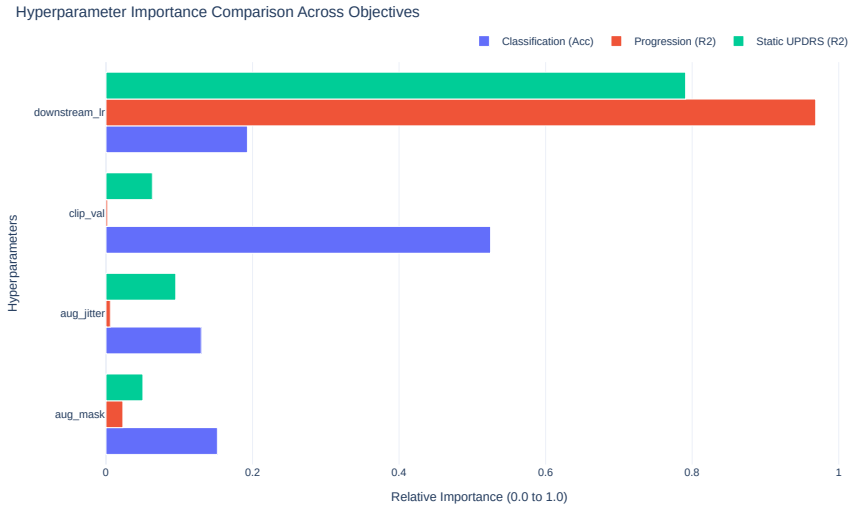

Figure 23: Comparative hyperparameter importance across the three primary downstream objectives (Phase 4)

The parameter importance analysis (detailed in Figure 23) revealed a contrast in optimization requirements between categorical and continuous tasks. Classification accuracy was primarily stabilized by gradient clipping (`clip_val`, 50.0%), which mitigated exploding gradients and prevented the model from overshooting decision boundaries during cross-entropy optimization. The continuous regression tasks (Progression and Static UPDRS) were driven by the downstream learning rate (`downstream_lr`, 95.3% and 78.5%, respectively), highlighting the necessity of step-size calibration for continuous severity mapping.

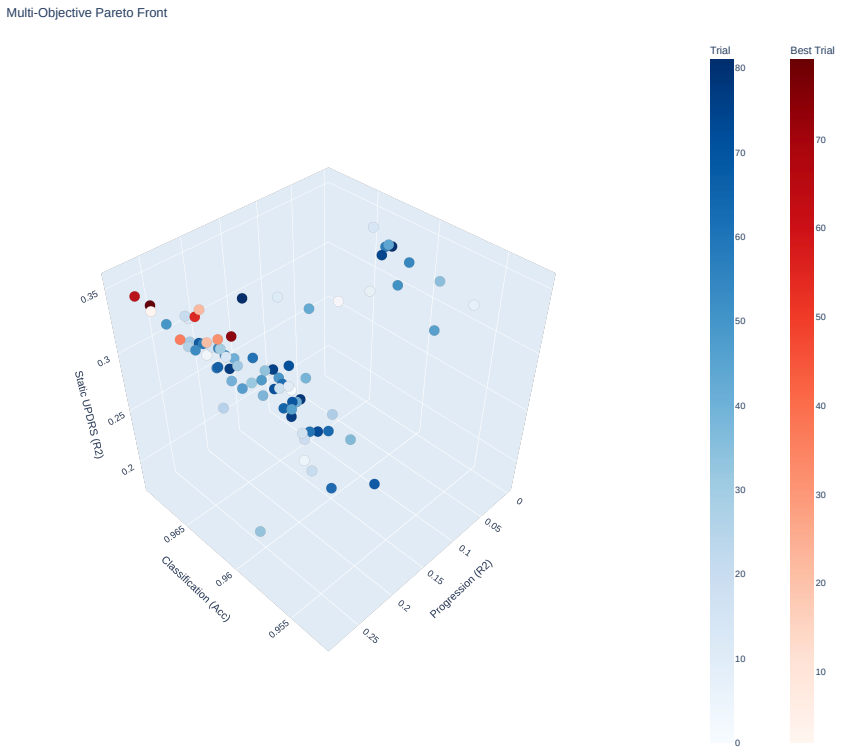

Figure 24: Pareto front illustrating the trade-off between downstream objectives (Phase 4)

Evaluating the Pareto front (Figure 24) yielded a final master configuration that introduced aggressive input masking (`aug_mask` = 0.24) alongside low feature jitter (`aug_jitter` = 0.01) and a finalized downstream learning rate of 0.01. This combination successfully forced the network to rely heavily on the missing modality hallucination module, drastically improving generalizability.

This final locked-in configuration achieved a peak classification accuracy of 96.7%, while simultaneously maintaining highly competitive continuous prediction capabilities with a Progression R^2 of 0.277 and a Static UPDRS R^2 of 0.345.

4.2 Ablation Studies

The following experiments validate the developed framework as a whole by excluding independent modules/modalities and comparing the achieved results of the latter with the results of the complete model.

4.2.1 Modality Ablation

To demonstrate the utility of combining multiple modalities we report in Table 7 the performance of the model with different combinations of used data modalities.

Table 7: Comparison of our complete model performance against the model with ablated modalities

		Classification	**Severity (R^2)**		**Progression (R^2)**	
Model	**Modality**	**Accuracy**	**Part II**	**Part III**	**Part II**	**Part III**
GeNeuro	(MRI, DTI, fMRI)	0.93	0.17	0.32	0.10	0.20
GeNeuro	(MRI, DTI, SPECT)	0.94	0.21	0.41	0.16	0.33
GeNeuro	(MRI, SPECT, fMRI)	0.95	0.24	0.44	**0.18**	0.34
GeNeuro	(SPECT, DTI, fMRI)	0.95	0.24	0.44	0.17	0.35
GeNeuro	**Full Model**	**0.95**	**0.26**	**0.44**	0.17	**0.35**

It can be observed that the model performs significantly worse without SPECT and fMRI and the increase in performance when utilizing DTI outweighs the slight decrease in the downstream progression task for part II when excluding it.

Furthermore, Table 8 confirms the chosen fMRI hub and defines SPECT as a competitor. Even though there is no clear winner between the two modalities, we can observe that MRI and DTI are not suitable hubs.

Table 8: Comparison of our complete model performance against the model with different hubs

		Classification	**Severity (R^2)**		**Progression (R^2)**	
Model	**Modality**	**Accuracy**	**Part II**	**Part III**	**Part II**	**Part III**
GeNeuro	(MRI hub)	0.93	0.23	0.42	0.17	0.34
GeNeuro	(DTI hub)	0.94	0.23	0.43	0.16	0.35
GeNeuro	(SPECT hub)	0.95	0.26	**0.45**	**0.18**	**0.36**
GeNeuro	**fMRI Hub**	**0.95**	**0.26**	0.44	0.17	0.35

4.2.2 Contrastive Learning Ablation

To demonstrate the utility of contrastive learning we report in Table 9 the performance of the model without the contrastive learning module.

Table 9: Comparison of our complete model performance against the model without contrastive learning

		Classification	**Severity (R^2)**		**Progression (R^2)**	
Model	**Modality**	**Accuracy**	**Part II**	**Part III**	**Part II**	**Part III**
GeNeuro	(No CL)	0.93	0.22	0.42	0.15	0.34
GeNeuro	**Full Model**	**0.95**	**0.26**	**0.44**	**0.17**	**0.35**

4.2.3 Generative Reconstruction Ablation

To demonstrate the utility of generative reconstruction we report in Table 10 the performance of the model without the use of generative reconstruction.

Table 10: Comparison of our complete model performance against the model without generative reconstruction

		Classification	**Severity (R^2)**		**Progression (R^2)**	
Model	**Modality**	**Accuracy**	**Part II**	**Part III**	**Part II**	**Part III**
GeNeuro	(No GR)	**0.96**	0.25	**0.46**	0.13	0.33
GeNeuro	**Full Model**	0.95	**0.26**	0.44	**0.17**	**0.35**

We can observe comparable performance between the model with and without generative reconstruction across the tasks of classification and severity regression. However, the generative module proves itself particularly useful for the downstream progression task where the completeness of a patient's history is more significant.

To validate the internal mechanics of the architecture, we conducted a visual analysis of the generated embeddings and downstream task subspaces (see Appendix A for all corresponding figures).

4.3 Learned Biomarkers

4.3.1 Case Study: Subjects 55468 (PD) and 100122 (Prodromal)

To demonstrate the utility of the framework, we compared (Table 11) two subjects with identical available modalities (MRI and fMRI only).

Table 11: Comparative Biomarker Profiles: PD vs. Prodromal

Clinical Metric	Subject 55468_BL (PD)	Subject 100122_BL (Prodromal)
Cohort	Cohort 1 (Diagnosed PD)	Cohort 4 (Prodromal)
UPDRS-III Score	31.0 (Moderate Severity)	0.0 (Asymptomatic)
MRI Primary Driver	Thickness Loss (Cortical Thinning)	Volume Stability (Midbrain/Temporal)
fMRI Connectivity	Limbic-Motor Disruption	Executive Hub Reorganization
Dominant Hub	Left Amygdala → motor cortex	Right Lateral Orbitofrontal Cortex

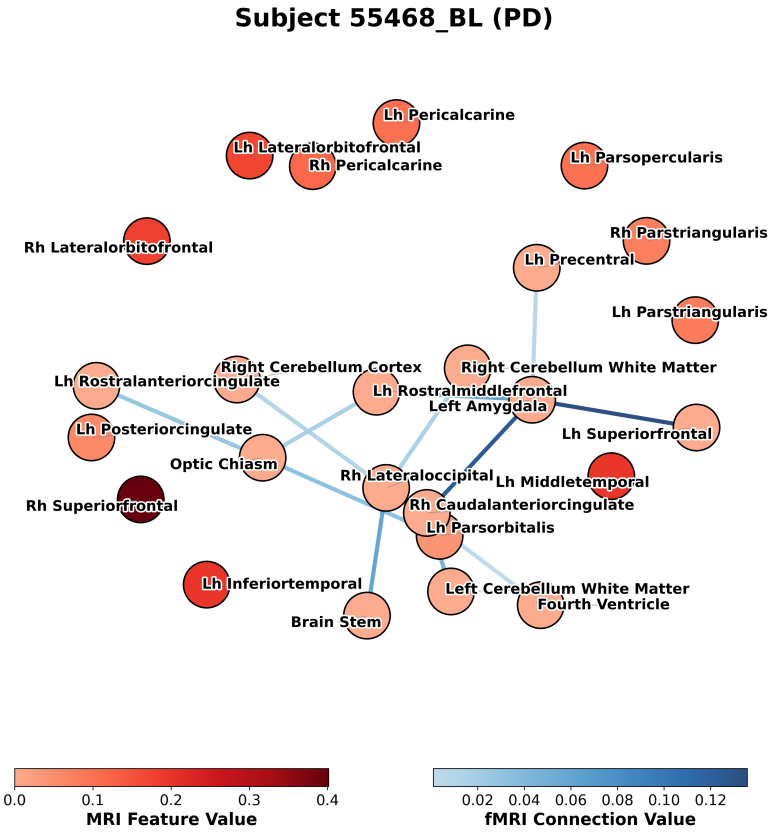

Figure 25: Brain graph containing most important nodes and edges of the brain graph of subject 55468_BL

Subject 55468_BL: The Structural Atrophy Signature The subject (whose brain graph is displayed in Figure 25) presents a moderate-to-severe motor impairment. The local explanation report reveals that the model's prediction is driven by established neurodegeneration.

The model identifies significant "negative impact" from cortical thickness in the `rh_lateralorbitofrontal` and `lh_middletemporal` regions. This suggests that the cortical thinning in these areas is the primary structural justification for the predicted motor severity.

In the fMRI domain, the top influential edges involve the left amygdala connecting to the precentral gyrus (Primary Motor Cortex) and Superior Frontal regions. The high importance of limbic-to-motor edges suggests that this patient's motor control has become heavily reliant on extra-striatal limbic networks (such as the amygdala). This compensatory routing is a common finding in advanced PD, where emotional and stress-response circuits are recruited to bypass severely degraded dopaminergic motor loops.

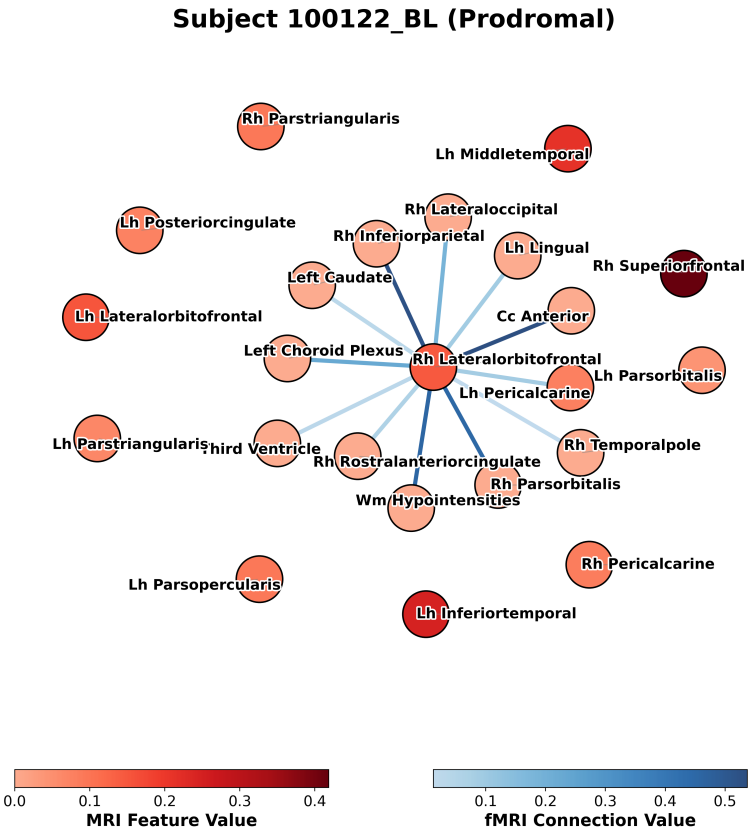

Figure 26: Brain graph containing most important nodes and edges of the brain graph of subject 100122_BL

Subject 100122_BL: Detecting the Prodromal "Silent" Signal The subject (whose brain graph is displayed in Figure 26) represents the "at-risk" population. Despite a clinical motor score of 0.0, the model identifies a high-risk profile through functional network reshuffling.

Unlike the PD case, the prodromal profile is dominated by a single hub: the Right Lateral Orbitofrontal Cortex. This region appears in nearly 90% of the top influential edges for this subject. In the absence of motor symptoms, the model appears to be detecting functional compensation, where executive frontal networks are rerouted to maintain motor stability despite early-stage pathological stress.

While the PD case showed thickness loss, this subject is flagged for changes in the Temporal Poles. Changes in temporal pole volume are often early markers of non-motor symptoms (e.g., olfactory dysfunction or mood disorders) that precede the "motor tremor" phase of Parkinson's.

Clinical Discussion and Treatment Implications For the PD patient, the model suggests Symptomatic Management, targeting the limbic-motor fluctuations with optimized Levodopa therapy and visual-motor physical therapy. For the Prodromal patient, the model provides a window for Neuroprotective Intervention, suggesting high-intensity aerobic exercise to maintain the functional integrity of the orbitofrontal hubs before structural atrophy occurs.

4.3.2 Global Biomarkers

Classification The following tables display the most important learned global biomarkers reporting relative saliency (signal-to-noise ratio) and clinical cohort contributions in the classification task. Each table is followed by a brief discussion on the relevance of the findings.

Table 12: Top 15 Global Connections (Edges) for Classification using DTI

Rank	Connection (Edge)	Rel. Saliency	PD	HC	Prod.
1	lh_postcentral ↔ rh_rostralanteriorcingulat	196.7x (100.0%)	-1.00×10^{-6}	-2.30×10^{-5}	0.00
2	lh_caudalmiddlefrontal ↔ rh_rostralanterior	150.0x (100.0%)	-1.00×10^{-6}	-2.30×10^{-5}	0.00
3	lh_precentral ↔ rh_rostralanteriorcingulate	139.2x (99.9%)	-1.00×10^{-6}	-2.10×10^{-5}	0.00
4	lh_postcentral ↔ rh_caudalanteriorcingulate	128.7x (99.9%)	-1.00×10^{-6}	-8.00×10^{-6}	0.00
5	lh_precentral ↔ rh_caudalanteriorcingulate	126.5x (99.9%)	0.00	-1.10×10^{-5}	0.00
6	lh_caudalmiddlefrontal ↔ rh_caudalanteriorc	120.4x (99.8%)	-1.00×10^{-6}	-1.00×10^{-5}	0.00
7	lh_postcentral ↔ lh_rostralanteriorcingulat	103.4x (99.8%)	-1.00×10^{-6}	-1.40×10^{-5}	0.00
8	lh_supramarginal ↔ rh_rostralanteriorcingul	100.6x (99.8%)	-1.00×10^{-6}	-1.30×10^{-5}	0.00
9	lh_supramarginal ↔ rh_caudalanteriorcingula	99.4x (99.7%)	0.00	-8.00×10^{-6}	0.00
10	lh_rostralmiddlefrontal ↔ rh_rostralanterio	93.2x (99.7%)	0.00	-1.20×10^{-5}	0.00
11	lh_rostralmiddlefrontal ↔ rh_caudalanterior	89.4x (99.7%)	0.00	-7.00×10^{-6}	0.00
12	lh_parstriangularis ↔ rh_caudalanteriorcing	89.0x (99.7%)	0.00	-6.00×10^{-6}	0.00
13	lh_frontalpole ↔ lh_rostralanteriorcingulat	88.0x (99.6%)	-1.00×10^{-6}	-1.10×10^{-5}	0.00
14	lh_frontalpole ↔ rh_rostralanteriorcingulat	82.3x (99.6%)	0.00	-1.10×10^{-5}	0.00
15	lh_rostralanteriorcingulate ↔ lh_rostralmid	81.4x (99.6%)	-1.00×10^{-6}	-1.20×10^{-5}	0.00

DTI: White Matter Integrity and Motor-Limbic Tracts As detailed in Table 12, the most critical structural connectivity biomarkers involve pathways connecting the primary motor cortex (precentral and postcentral gyri) to the anterior cingulate. These top connections are nearly 200 times more salient than the network median. The disparity in impact scores between Healthy Controls (HC) and PD patients highlights the model's reliance on these specific white matter tracts. The degradation of structural connectivity in the motor-limbic loop (which allows executive planning and emotion to regulate physical movement) serves as the key discriminator, allowing the model to rule out the healthy state based on the microstructural loss of these fibers [58].

Table 13: Top 15 Global Connections (Edges) for Classification using fMRI

Rank	Connection (Edge)	Rel. Saliency	PD	HC	Prod.
1	left_cerebellum_cortex ↔ optic_chiasm	58.3x (100.0%)	0.00	1.00×10^{-6}	-4.40×10^{-5}
2	optic_chiasm ↔ rh_posteriorcingulate	51.1x (100.0%)	0.00	3.00×10^{-6}	3.70×10^{-5}
3	cc_mid_anterior ↔ left_cerebellum_cortex	47.3x (100.0%)	0.00	0.00	-2.70×10^{-5}
4	optic_chiasm ↔ rh_corpuscallosum	46.9x (99.9%)	0.00	1.00×10^{-6}	3.60×10^{-5}
5	left_cerebellum_white_matter ↔ optic_chiasm	43.9x (99.9%)	0.00	1.00×10^{-6}	-3.10×10^{-5}
6	cc_mid_anterior ↔ left_cerebellum_white_mat	37.7x (99.9%)	0.00	0.00	-2.20×10^{-5}
7	optic_chiasm ↔ rh_postcentral	35.6x (99.9%)	1.00×10^{-6}	2.00×10^{-6}	2.90×10^{-5}
8	left_cerebellum_cortex ↔ rh_lateraloccipita	34.4x (99.9%)	0.00	0.00	-2.10×10^{-5}
9	left_cerebellum_cortex ↔ rh_entorhinal	32.9x (99.8%)	-1.00×10^{-6}	0.00	-1.30×10^{-5}
10	optic_chiasm ↔ rh_supramarginal	32.1x (99.8%)	1.00×10^{-6}	8.00×10^{-6}	2.00×10^{-5}
11	fourth_ventricle ↔ optic_chiasm	31.5x (99.8%)	0.00	0.00	-2.50×10^{-5}
12	left_cerebellum_white_matter ↔ rh_lateraloc	30.6x (99.8%)	0.00	0.00	-2.40×10^{-5}
13	left_cerebellum_cortex ↔ left_vessel	30.0x (99.7%)	0.00	0.00	-1.90×10^{-5}
14	left_cerebellum_cortex ↔ right_accumbens_ar	29.3x (99.7%)	0.00	0.00	-1.40×10^{-5}
15	left_cerebellum_cortex ↔ rh_inferiorparieta	28.6x (99.7%)	0.00	0.00	-1.10×10^{-5}

fMRI: Functional Network Reshuffling in Prodromal Stages Table 13 reveals a relevant pattern in functional connectivity: the top edges show disproportionately high contribution scores specifically for the Prodromal class. Connections involving the cerebellar cortex, optic chiasm, and posterior cingulate drive the classification of at-risk patients. This reinforces the hypothesis that before severe structural atrophy occurs, the brain undergoes functional network reshuffling. The model utilizes these compensatory functional hubs—particularly those linked to visual-spatial processing and the Default Mode Network—as the primary biomarker for early, pre-motor disease detection [59].

Table 14: Top 15 Global MRI Biomarkers for Classification (Feature-Level)

Rank	ROI Feature	Rel. Saliency	PD	HC	Prod.
1	right_cerebral_white_matter_Volume	3.3x (100.0%)	-1.63×10^{-4}	-1.86×10^{-3}	3.75×10^{-4}
2	left_cerebral_white_matter_Volume	3.3x (97.8%)	-1.63×10^{-4}	-1.86×10^{-3}	3.62×10^{-4}
3	left_cerebellum_white_matter_Volume	3.0x (95.6%)	-1.00×10^{-5}	-1.08×10^{-4}	1.70×10^{-5}
4	right_cerebellum_white_matter_Volume	3.0x (93.3%)	-1.00×10^{-5}	-1.02×10^{-4}	1.60×10^{-5}
5	right_lateral_ventricle_Volume	3.0x (91.1%)	-8.00×10^{-6}	-8.00×10^{-5}	1.40×10^{-5}
6	brain_stem_Volume	3.0x (88.9%)	-1.40×10^{-5}	-1.49×10^{-4}	2.30×10^{-5}
7	left_lateral_ventricle_Volume	3.0x (86.7%)	-9.00×10^{-6}	-8.90×10^{-5}	1.50×10^{-5}
8	right_cerebellum_cortex_Volume	2.9x (84.4%)	-3.50×10^{-5}	-3.66×10^{-4}	5.40×10^{-5}
9	right_thalamus_Volume	2.9x (82.2%)	-5.00×10^{-6}	-4.90×10^{-5}	7.00×10^{-6}
10	left_cerebellum_cortex_Volume	2.9x (80.0%)	-3.40×10^{-5}	-3.56×10^{-4}	5.00×10^{-5}
11	rh_superiortemporal_Volume	2.9x (77.8%)	0.00	0.00	0.00
12	third_ventricle_Volume	2.9x (75.6%)	-1.00×10^{-6}	-9.00×10^{-6}	1.00×10^{-6}
13	right_ventraldc_Volume	2.9x (73.3%)	-3.00×10^{-6}	-2.70×10^{-5}	4.00×10^{-6}
14	right_putamen_Volume	2.9x (71.1%)	-3.00×10^{-6}	-3.10×10^{-5}	4.00×10^{-6}
15	rh_lateralorbitofrontal_Volume	2.9x (68.9%)	0.00	0.00	0.00

MRI: Volumetric Alterations as the Primary Structural Classifier The structural MRI biomarkers (Table 14) are strictly dominated by volumetric alterations, revealing that macro-structural changes are required to establish diagnostic baselines. The model demonstrates a clear preference for cerebral and cerebellar white matter volumes, as well as the expansion of the lateral ventricles. The high impact of ventricular volume serves as a classic surrogate marker for global brain atrophy *ex vacuo*. This indicates the model successfully learned that widespread structural volume loss, rather than localized cortical thinning, establishes the primary morphological signature of the defined clinical classes [60].

Table 15: Global Biomarkers for Classification using SPECT

Rank	ROI	Rel. Saliency	PD	HC	Prod.
1	striatum_L	1.2x (100.0%)	1.50×10^{-5}	-5.43×10^{-3}	0.00
2	striatum_R	1.1x (85.7%)	1.39×10^{-3}	6.79×10^{-2}	0.00
3	putamen_L	1.1x (71.4%)	-2.91×10^{-3}	4.88×10^{-3}	0.00
4	putamen_R	1.0x (57.1%)	-1.88×10^{-3}	4.64×10^{-2}	0.00
5	caudate_L	1.0x (42.9%)	-4.21×10^{-4}	1.98×10^{-3}	0.00
6	caudate_R	1.0x (28.6%)	-4.48×10^{-4}	4.88×10^{-2}	0.00
7	striatum_bilat	0.8x (14.3%)	9.33×10^{-3}	3.84×10^{-2}	0.00

SPECT: The Dopaminergic Gold Standard As expected from established clinical pathology, the SPECT biomarkers (Table 15) rely entirely on dopaminergic transporter signals in the basal ganglia. The left and right striatum, alongside the bilateral putamen, emerge as the most robust features. Crucially, the impact scores reveal a massive disparity between Healthy Controls and PD patients (who exhibit negative contributions due to dopamine depletion). This confirms that the attribution module successfully isolated the primary pathological hallmark of Parkinson's disease without human interference [54].

Severity The following tables display the most important learned global biomarkers reporting relative saliency in the static severity regression task. Each table is followed by a brief discussion on the relevance of the findings.

Table 16: Top 15 Global Connections (Edges) for UPDRS Regression using DTI

Rank	Connection (Edge)	Rel. Saliency	Impact	Clinical Effect
1	lh_postcentral ↔ rh_rostralanteriorcingulat	231.9x (100.0%)	-2.37×10^{-8}	Decreases (Better)
2	lh_lingual ↔ lh_rostralanteriorcingulate	189.9x (100.0%)	1.82×10^{-7}	Increases (Worse)
3	left_caudate ↔ lh_lingual	183.9x (99.9%)	2.02×10^{-7}	Increases (Worse)
4	lh_frontalpole ↔ rh_rostralanteriorcingulat	175.0x (99.9%)	-3.73×10^{-8}	Decreases (Better)
5	lh_caudalmiddlefrontal ↔ rh_rostralanterior	174.5x (99.9%)	-2.29×10^{-8}	Decreases (Better)
6	lh_postcentral ↔ lh_rostralanteriorcingulat	166.4x (99.9%)	-4.68×10^{-8}	Decreases (Better)
7	left_accumbens_area ↔ lh_lingual	163.6x (99.8%)	1.69×10^{-7}	Increases (Worse)
8	cc_mid_anterior ↔ lh_lingual	162.1x (99.8%)	1.26×10^{-7}	Increases (Worse)
9	lh_postcentral ↔ rh_frontalpole	157.5x (99.8%)	-3.28×10^{-8}	Decreases (Better)
10	lh_precentral ↔ rh_rostralanteriorcingulate	154.9x (99.8%)	-2.25×10^{-8}	Decreases (Better)
11	lh_frontalpole ↔ rh_frontalpole	151.3x (99.7%)	-2.67×10^{-8}	Decreases (Better)
12	lh_lingual ↔ lh_superiorfrontal	140.0x (99.7%)	1.66×10^{-7}	Increases (Worse)
13	lh_frontalpole ↔ lh_rostralanteriorcingulat	139.5x (99.7%)	-3.94×10^{-8}	Decreases (Better)
14	lh_rostralmiddlefrontal ↔ rh_rostralanterio	128.4x (99.7%)	-5.35×10^{-8}	Decreases (Better)
15	lh_caudalanteriorcingulate ↔ lh_lingual	128.3x (99.6%)	1.33×10^{-7}	Increases (Worse)

DTI: Structural Integrity of the Motor-Visual Axis As shown in Table 16, the severity of motor impairment is heavily predicted by the structural degradation of primary sensorimotor tracts (postcentral gyrus) connecting to the anterior cingulate. Additionally, high-saliency connections involve the `lh_lingual` gyrus connecting to the caudate and accumbens area. A positive impact score indicates that abnormal structural connectivity in these regions drives higher UPDRS scores. This supports the clinical observation that as the primary motor loop degrades, patients increasingly rely on visual-motor integration networks (via the lingual gyrus) to bypass damaged pathways, resulting in severe motor deficits when these structural connections fail [61].

Table 17: Top 15 Global Connections (Edges) for UPDRS Regression using fMRI

Rank	Connection (Edge)	Rel. Saliency	Impact	Clinical Effect
1	optic_chiasm ↔ rh_corpuscallosum	40.9x (100.0%)	-2.56×10^{-7}	Decreases (Better)
2	lh_superiorfrontal ↔ optic_chiasm	40.9x (100.0%)	-1.93×10^{-7}	Decreases (Better)
3	optic_chiasm ↔ rh_posteriorcingulate	34.6x (100.0%)	-1.13×10^{-7}	Decreases (Better)
4	left_cerebral_white_matter ↔ optic_chiasm	31.9x (99.9%)	-1.11×10^{-7}	Decreases (Better)
5	optic_chiasm ↔ rh_rostralanteriorcingulate	31.5x (99.9%)	-1.70×10^{-7}	Decreases (Better)
6	lh_rostralmiddlefrontal ↔ optic_chiasm	31.2x (99.9%)	-1.31×10^{-7}	Decreases (Better)
7	lh_precentral ↔ optic_chiasm	30.8x (99.9%)	-8.22×10^{-8}	Decreases (Better)
8	optic_chiasm ↔ right_lateral_ventricle	30.6x (99.9%)	-2.00×10^{-7}	Decreases (Better)
9	optic_chiasm ↔ right_cerebral_white_matter	30.4x (99.8%)	-1.62×10^{-7}	Decreases (Better)
10	optic_chiasm ↔ rh_supramarginal	28.6x (99.8%)	-9.54×10^{-8}	Decreases (Better)
11	lh_corpuscallosum ↔ optic_chiasm	27.8x (99.8%)	-1.30×10^{-7}	Decreases (Better)
12	left_lateral_ventricle ↔ optic_chiasm	24.5x (99.8%)	-8.39×10^{-8}	Decreases (Better)
13	lh_parstriangularis ↔ optic_chiasm	23.3x (99.8%)	-4.95×10^{-8}	Decreases (Better)
14	optic_chiasm ↔ rh_postcentral	23.0x (99.8%)	-1.21×10^{-7}	Decreases (Better)
15	left_cerebellum_cortex ↔ optic_chiasm	22.8x (99.7%)	1.18×10^{-7}	Increases (Worse)

fMRI: Visual-Cognitive Functional Uncoupling The functional connectivity markers for severity (Table 17) highlight an overwhelming reliance on the optic chiasm connecting to higher-order cognitive and motor regions (corpus callosum, superior frontal, posterior cingulate). This reflects a profound uncoupling of visual input from executive and motor execution networks. As dopaminergic striatal function declines, patients experience deficits in integrating visuospatial information, driving higher clinical severity scores when these functional resting-state pathways are disrupted.

Table 18: Top 15 Global MRI Biomarkers for UPDRS Severity (Feature-Level)

Rank	ROI Feature	Rel. Saliency	Impact	Clinical Effect
1	rh_lateralorbitofrontal_Thickn	2.5x (100.0%)	-9.01×10^{-6}	Decreases (Better)
2	rh_parstriangularis_Thickness	2.5x (97.8%)	-5.20×10^{-6}	Decreases (Better)
3	lh_pericalcarine_Thickness	2.5x (95.6%)	-4.83×10^{-6}	Decreases (Better)
4	lh_middletemporal_Thickness	2.5x (93.3%)	-1.08×10^{-5}	Decreases (Better)
5	rh_pericalcarine_Thickness	2.5x (91.1%)	-5.30×10^{-6}	Decreases (Better)
6	lh_parsorbitalis_Thickness	2.5x (88.9%)	-2.32×10^{-6}	Decreases (Better)
7	lh_posteriorcingulate_Thicknes	2.4x (86.7%)	-3.86×10^{-6}	Decreases (Better)
8	lh_lateralorbitofrontal_Thickn	2.4x (84.4%)	-8.94×10^{-6}	Decreases (Better)
9	lh_inferiortemporal_Thickness	2.4x (82.2%)	-1.16×10^{-5}	Decreases (Better)
10	lh_parstriangularis_Thickness	2.4x (80.0%)	-4.37×10^{-6}	Decreases (Better)
11	lh_parsopercularis_Thickness	2.4x (77.8%)	-5.23×10^{-6}	Decreases (Better)
12	rh_superiorfrontal_Thickness	2.4x (75.6%)	-2.33×10^{-5}	Decreases (Better)
13	rh_inferiortemporal_Thickness	2.4x (73.3%)	-1.10×10^{-5}	Decreases (Better)
14	rh_rostralmiddlefrontal_Thickn	2.4x (71.1%)	-1.92×10^{-5}	Decreases (Better)
15	rh_middletemporal_Thickness	2.4x (68.9%)	-1.16×10^{-5}	Decreases (Better)

MRI: Cortical Thinning in Frontal and Temporal Lobes Table 18 demonstrates reliance on Cortical Thickness to predict severity. The model focuses on the orbitofrontal, temporal, and pericalcarine cortices (lower thickness equals worse clinical score). This confirms that gray matter atrophy in associative frontal-temporal areas mirrors worsening of the disease's symptoms [62].

Table 19: Global Biomarkers for UPDRS Regression using SPECT

Rank	ROI	Rel. Saliency	Impact	Clinical Effect
1	striatum_L	1.3x (100.0%)	-2.39×10^{-4}	Decreases (Better)
2	striatum_R	1.2x (85.7%)	-2.72×10^{-4}	Decreases (Better)
3	putamen_L	1.0x (71.4%)	7.86×10^{-5}	Increases (Worse)
4	putamen_R	1.0x (57.1%)	9.07×10^{-6}	Increases (Worse)
5	caudate_L	1.0x (42.9%)	-6.68×10^{-4}	Decreases (Better)
6	caudate_R	1.0x (28.6%)	-7.06×10^{-4}	Decreases (Better)
7	striatum_bilat	0.9x (14.3%)	-2.03×10^{-4}	Decreases (Better)

SPECT: Asymmetric Striatal Depletion The SPECT biomarkers for UPDRS regression (Table 19) focus exclusively on the precise anatomical subregions of the basal ganglia. The left and right striatum act as the strongest anchors. The negative impact scores assigned to the striatum and caudate indicate that a decrease in dopaminergic signal in these regions directly drives a higher (worse) motor severity prediction. Notably, the model effectively captures the asymmetric nature of Parkinsonian dopaminergic cell loss, assigning distinct importance weights to the left and right hemispheres based on their varying rates of progressive depletion [63].

Progression The following tables display the most important learned global biomarkers reporting relative saliency in the severity progression task. Each table is followed by a brief discussion on the relevance of the findings.

Table 20: Top 15 Global Connections (Edges) for Progression Forecasting using DTI

Rank	Connection (Edge)	Rel. Saliency	Impact	Clinical Effect
1	lh_postcentral ↔ rh_rostralanteriorcingulat	830.0x (100.0%)	1.38×10^{-8}	Increases (Worse)
2	lh_precentral ↔ rh_rostralanteriorcingulate	715.0x (100.0%)	4.95×10^{-8}	Increases (Worse)
3	lh_caudalmiddlefrontal ↔ rh_rostralanterior	713.4x (100.0%)	1.47×10^{-8}	Increases (Worse)
4	brain_stem ↔ rh_caudalanteriorcingulate	672.4x (99.9%)	-4.91×10^{-8}	Decreases (Better)
5	lh_frontalpole ↔ rh_rostralanteriorcingulat	548.1x (99.9%)	-2.54×10^{-8}	Decreases (Better)
6	left_cerebellum_cortex ↔ rh_rostralanterior	531.5x (99.9%)	-5.95×10^{-8}	Decreases (Better)
7	lh_postcentral ↔ rh_frontalpole	522.8x (99.9%)	2.25×10^{-9}	Increases (Worse)
8	rh_caudalanteriorcingulate ↔ rh_parsorbital	508.6x (99.8%)	-9.16×10^{-8}	Decreases (Better)
9	lh_rostralmiddlefrontal ↔ rh_rostralanterio	505.5x (99.8%)	5.98×10^{-8}	Increases (Worse)
10	lh_supramarginal ↔ rh_rostralanteriorcingul	501.7x (99.8%)	4.42×10^{-8}	Increases (Worse)
11	rh_caudalanteriorcingulate ↔ rh_temporalpol	471.8x (99.8%)	-1.13×10^{-7}	Decreases (Better)
12	rh_caudalanteriorcingulate ↔ rh_inferiortem	458.6x (99.7%)	-6.83×10^{-8}	Decreases (Better)
13	lh_frontalpole ↔ rh_frontalpole	441.2x (99.7%)	-2.84×10^{-8}	Decreases (Better)
14	lh_superiorparietal ↔ rh_rostralanteriorcin	437.8x (99.7%)	5.36×10^{-8}	Increases (Worse)
15	lh_caudalmiddlefrontal ↔ rh_frontalpole	405.8x (99.7%)	-2.04×10^{-9}	Decreases (Better)

DTI: The Collapse of Executive Motor Tracts Progression forecasting (Table 20) isolates the structural connection between the sensorimotor cortex (postcentral/precentral gyri) and the anterior cingulate. Remarkably, this connection is identified as being over 800 times more salient than the network median. The anterior cingulate is a critical hub for executive function, emotional regulation, and cognitive motor control. The degradation of these central tracts serves as a massive predictor of rapid future decline, suggesting that the widespread structural uncoupling of cognitive and executive networks is the primary driver of longitudinal progression [58].

Table 21: Top 15 Global Connections (Edges) for Progression Forecasting using fMRI

Rank	Connection (Edge)	Rel. Saliency	Impact	Clinical Effect
1	optic_chiasm ↔ rh_rostralanteriorcingulate	124.7x (100.0%)	-1.73×10^{-7}	Decreases (Better)
2	optic_chiasm ↔ right_lateral_ventricle	120.3x (100.0%)	-1.49×10^{-7}	Decreases (Better)
3	lh_lateralorbitofrontal ↔ optic_chiasm	114.3x (100.0%)	-1.09×10^{-7}	Decreases (Better)
4	lh_lateraloccipital ↔ rh_superiorfrontal	100.7x (99.9%)	-8.05×10^{-8}	Decreases (Better)
5	lh_supramarginal ↔ optic_chiasm	99.8x (99.9%)	-1.03×10^{-7}	Decreases (Better)
6	optic_chiasm ↔ rh_caudalanteriorcingulate	96.2x (99.9%)	-1.08×10^{-7}	Decreases (Better)
7	rh_lateraloccipital ↔ right_caudate	93.6x (99.9%)	-6.33×10^{-8}	Decreases (Better)
8	lh_rostralanteriorcingulate ↔ optic_chiasm	92.0x (99.9%)	-1.83×10^{-7}	Decreases (Better)
9	lh_lateraloccipital ↔ rh_caudalanteriorcing	86.0x (99.8%)	-5.23×10^{-8}	Decreases (Better)
10	lh_superiorfrontal ↔ optic_chiasm	83.8x (99.8%)	-9.96×10^{-8}	Decreases (Better)
11	lh_caudalmiddlefrontal ↔ optic_chiasm	81.0x (99.8%)	-1.41×10^{-7}	Decreases (Better)
12	rh_caudalanteriorcingulate ↔ rh_cuneus	79.5x (99.8%)	-1.44×10^{-7}	Decreases (Better)
13	rh_caudalanteriorcingulate ↔ right_vessel	77.2x (99.8%)	-3.71×10^{-8}	Decreases (Better)
14	optic_chiasm ↔ rh_posteriorcingulate	76.8x (99.7%)	-6.11×10^{-8}	Decreases (Better)
15	rh_caudalanteriorcingulate ↔ rh_lateralocci	75.6x (99.7%)	-6.49×10^{-8}	Decreases (Better)

fMRI: Visual and Executive Uncoupling in Progression Table 21 reveals that longitudinal progression is heavily forecasted by the functional disconnect between visual pathways (optic chiasm, lateral occipital) and frontal executive regions (rostral anterior cingulate, orbitofrontal cortex). This uncoupling signifies that early compensatory visual-motor circuits begin to fail as the disease advances, predicting a steeper trajectory of clinical decline when higher-order frontal coordination is lost.

Table 22: Top 15 Global MRI Biomarkers for Progression Forecasting (Feature-Level)

Rank	ROI Feature	Rel. Saliency	Impact	Clinical Effect
1	rh_lateralorbitofrontal_Thickn	2.2x (100.0%)	-7.65×10^{-6}	Decreases (Better)
2	rh_parstriangularis_Thickness	2.2x (97.8%)	-4.23×10^{-6}	Decreases (Better)
3	lh_pericalcarine_Thickness	2.2x (95.6%)	-3.82×10^{-6}	Decreases (Better)
4	lh_middletemporal_Thickness	2.2x (93.3%)	-9.37×10^{-6}	Decreases (Better)
5	rh_pericalcarine_Thickness	2.2x (91.1%)	-4.19×10^{-6}	Decreases (Better)
6	lh_parsorbitalis_Thickness	2.2x (88.9%)	-2.01×10^{-6}	Decreases (Better)
7	lh_posteriorcingulate_Thicknes	2.2x (86.7%)	-3.15×10^{-6}	Decreases (Better)
8	lh_lateralorbitofrontal_Thickn	2.2x (84.4%)	-7.66×10^{-6}	Decreases (Better)
9	lh_inferiortemporal_Thickness	2.2x (82.2%)	-9.92×10^{-6}	Decreases (Better)
10	lh_parstriangularis_Thickness	2.2x (80.0%)	-3.58×10^{-6}	Decreases (Better)
11	lh_parsopercularis_Thickness	2.2x (77.8%)	-4.17×10^{-6}	Decreases (Better)
12	rh_superiorfrontal_Thickness	2.2x (75.6%)	-2.01×10^{-5}	Decreases (Better)
13	rh_inferiortemporal_Thickness	2.2x (73.3%)	-9.49×10^{-6}	Decreases (Better)
14	rh_rostralmiddlefrontal_Thickn	2.1x (71.1%)	-1.63×10^{-5}	Decreases (Better)
15	rh_middletemporal_Thickness	2.1x (68.9%)	-9.94×10^{-6}	Decreases (Better)

MRI: Progressive Cortical Atrophy The MRI features in Table 22 demonstrate that while Classification utilized gross subcortical volumes, Progression Forecasting is almost entirely driven by Cortical Thickness. The model anchors its future predictions on continued cortical thinning in the orbitofrontal, temporal, and pericalcarine areas. This proves that while widespread volume loss is sufficient for static diagnosis, tracking the progressive micro-structural thinning of higher-order cortical processing centers is required to accurately forecast the trajectory of future neurodegeneration [64].

Table 23: Global Biomarkers for Progression Forecasting using SPECT

Rank	ROI	Rel. Saliency	Impact	Clinical Effect
1	striatum_R	1.2x (100.0%)	-3.53×10^{-4}	Decreases (Better)
2	striatum_L	1.1x (85.7%)	-1.45×10^{-4}	Decreases (Better)
3	putamen_R	1.0x (71.4%)	-1.86×10^{-4}	Decreases (Better)
4	putamen_L	1.0x (57.1%)	-5.60×10^{-5}	Decreases (Better)
5	caudate_L	1.0x (42.9%)	-1.47×10^{-4}	Decreases (Better)
6	caudate_R	0.9x (28.6%)	-3.62×10^{-4}	Decreases (Better)
7	striatum_bilat	0.7x (14.3%)	-4.55×10^{-4}	Decreases (Better)

SPECT: Measuring Dopaminergic Reserve The SPECT biomarkers for progression (Table 23) mirror the focus of the static tasks but serve a different function. The right and left striatum exhibit negative impacts (Decreases/Better). In a longitudinal context, this represents the model measuring the patient's Dopaminergic Reserve [65]. A higher signal in these regions provides a "buffer" against disease advancement, correlating with a slower progression curve over time, whereas a fully depleted baseline guarantees a steep clinical decline.

5 Conclusion

This final chapter summarizes the primary contributions of this thesis, acknowledges the underlying limitations of the developed framework, and proposes promising directions for future research in both AI-driven neurodiagnostics and clinical medicine.

5.1 Main Contributions

The core technical and clinical contributions of this study are summarized as follows:

- **Custom Neuroimaging Preprocessing:** We developed a highly tailored preprocessing pipeline utilizing the *ANTsPy* library to construct SC and FC matrices from raw DTI and fMRI scans. While inspired by current gold-standard methodologies, the brain graph construction process was custom-engineered to optimize both the available multimodal data and computational resources. The results validate the efficacy of this unique graph extraction approach.
- **Robust Multimodal Fusion:** This thesis builds upon existing architectures to unify brain graphs into a multimodal representation. This is achieved by projecting graph modalities into a shared embedding space via GNN-based encoders and contrastive learning. The framework incorporates a generative model capable of reconstructing a patient's missing modalities from their available scans, ultimately concatenating them into a fixed-length feature vector.
- **High-Fidelity Clinical Predictive Modeling:** By leveraging these fused brain representations, we demonstrated the framework's capacity to perform clinical downstream tasks. The proposed methods achieve state-of-the-art performance in PD classification and static severity regression, while establishing a highly competitive baseline for longitudinal disease progression forecasting.
- **Gradient-Based Explainability and Biomarker Identification:** Moving beyond "black-box" predictions, this thesis successfully implements a backpropagation-based explainability module. By tracing model predictions directly back to the raw input brain graphs, the framework successfully identifies biomarkers grounded in real biological brain regions and their anatomical connections.

The successful integration of these components holds significant clinical implications for the management of Parkinson's disease. By providing a robust, automated pipeline capable of mapping complex brain connectivity to objective clinical metrics like the MDS-UPDRS, this framework pushes toward a more standardized, data-driven approach for patient assessment. Crucially, the model's ability to autonomously identify established neurodegenerative patterns—such as asymmetric striatal depletion in SPECT scans and the progressive cortical thinning of associative frontal-temporal areas in MRI—validates its clinical relevance. Furthermore, the framework's capacity to detect early, pre-motor functional network reshuffling (such as visual-cognitive uncoupling) in prodromal patients demonstrates its immense potential as an early-warning diagnostic tool. The ability to extract these biologically interpretable biomarkers from multimodal networks bridges the critical gap between computational predictions and clinical neuroscience, offering clinicians tangible insights into the structural and functional alterations driving the disease's severity and future trajectory.

From a technical perspective, this work demonstrates that addressing the geometric complexities of the brain through GNNs and contrastive learning is a highly viable path for advanced neurodiagnostics. One of the most significant computational achievements of this thesis is the successful implementation of the missing modality reconstruction module. Real-world medical datasets are inherently incomplete, and the ability to leverage a shared latent space to probabilistically hallucinate missing neuroimaging data resolves a major bottleneck in AI-driven healthcare. By utilizing the designed architecture, the framework ensures that patients with sparse data can still benefit from high-fidelity, multimodal predictive modeling without reverting to basic imputation techniques that destroy underlying topological patterns.

Ultimately, the inclusion of a comprehensive, gradient-based explainability module ensures that these technical advancements translate into actionable medical intelligence. By moving beyond opaque "black-box" systems, the framework allows predictions to be traced directly back to patient-specific anatomical and functional anomalies. This transparency is fundamental for establishing clinical trust, facilitating the adoption of machine learning tools in standard medical practice, and enabling researchers to uncover novel pathophysiological insights. These findings not only validate the proposed geometric deep learning architecture but also lay a strong, interpretable foundation for the next generation of predictive neurodiagnostics.

5.2 Limitations

While the developed framework demonstrates strong predictive capabilities, several methodological and practical limitations must be acknowledged:

Preprocessing and Parcellation Constraints: Due to computational resource limitations, it was unfeasible to employ computationally exhaustive gold-standard preprocessing pipelines. Utilizing *FreeSurfer* in conjunction with *QSIPrep* and *fMRIPrep* would have yielded higher-fidelity data, enabling the construction of patient-tailored atlases via MRI rather than relying on a generic atlas to compute MRI edge weights and DTI/fMRI node features. Similarly, leveraging the *MIAKAT* pipeline would have significantly increased the spatial granularity of the SPECT graphs. Furthermore, the Desikan-Killiany (DK) atlas was selected as a pragmatic baseline; employing an atlas specifically optimized for mapping ROIs critical to PD pathology would likely yield superior predictive performance and uncover a more targeted set of clinical biomarkers.

Generative Imputation and Interpretability: The missing modalities reconstructed by the generative module represent the statistically most likely brain state derived from the training distribution, rather than true patient-specific biological observations. Consequently, these hallucinated embeddings had to be strictly excluded from the gradient-based biomarker identification process, as routing gradients through imputed data would yield mathematically artifactual and clinically inaccurate explanations.

Information Bottleneck and Topology Loss: To perform downstream tasks, the multimodal brain graphs are ultimately flattened into 1024-dimensional feature vectors. While the intrinsic graph topology is partially preserved within these embeddings via the GNN message-passing mechanism, explicit structural and relational data is inevitably compressed. This loss of direct topological information introduces an information bottleneck that is particularly detrimental to the longitudinal disease progression prediction task, which would benefit from tracking explicit temporal changes in specific structural edges.

Dataset Bias and Real-World Generalization: The framework was developed and evaluated using the PPMI dataset, which is highly curated and standardized. In contrast, real-world clinical data is inherently noisier, characterized by diverse scanner calibrations, missing modalities at irregular intervals, and varying imaging artifacts. The robustness and generalization capabilities of our architecture remain to be validated in a less curated, prospective hospital environment.

5.3 Future Work

Building upon the foundations established by this framework, future research in the field of AI-driven neurodiagnostics can address the limitations of the current model. While research in the field of clinical medicine would validate the explainability of the model.

Integration of Advanced Parcellation and Preprocessing: Future iterations of the framework should prioritize the integration of computationally intensive, gold-standard pipelines like *FreeSurfer*, *QSIPrep*, and *MIAKAT*. Transitioning from the generic Desikan-Killiany baseline to a high-resolution, PD-specific subcortical atlas would allow for much finer granularity. This would enable the identification of highly localized biomarkers within the basal ganglia and dopaminergic pathways.

Probabilistic Explainability for Imputed Data: To maximize the utility of the generative module, future research could explore uncertainty-aware generative models. By calculating confidence intervals or uncertainty bounds for the hallucinated embeddings, it may become possible to perform probabilistic gradient attribution. This would allow the framework to extract tentative, confidence-weighted clinical insights even from partially missing data.

Spatiotemporal Graph Neural Networks (ST-GNNs): To overcome the topological information bottleneck during longitudinal forecasting, the downstream recurrent neural network could be replaced with an ST-GNN. By processing a time series of explicit brain graphs rather than flattened latent vectors, the model could directly track the physical degradation of specific anatomical edges over time, yielding richer temporal insights into disease progression.

Prospective Multi-Center Clinical Validation: To bridge the gap between retrospective dataset evaluation and real-world medical application, the framework must be validated across diverse, multi-center hospital environments. Testing the architecture against uncurated data with varied scanner calibrations and artifacts will be critical to establishing its robustness, generalizability, and ultimate viability as an AI-driven neurodiagnostic tool. Furthermore, biomarkers that were discovered by the model but found no validation in current literature could be researched and validated.

5.4 Final Remarks

In conclusion, this thesis demonstrates that a combination of GNN-based contrastive geometric deep learning and generative artificial intelligence can meaningfully address the complexities of multimodal neuroimaging. By moving beyond opaque predictive models to provide localized, biologically grounded insights, the proposed framework not only advances the technical frontier of medical AI but also offers a bridge to close the gap between AI-driven neurodiagnostics and actionable medical research.

A Supplementary Latent Space Visualizations

This appendix provides the extended visual analysis validating the internal mechanics of the architecture.

A.1 Multi-Modal Embedding Space

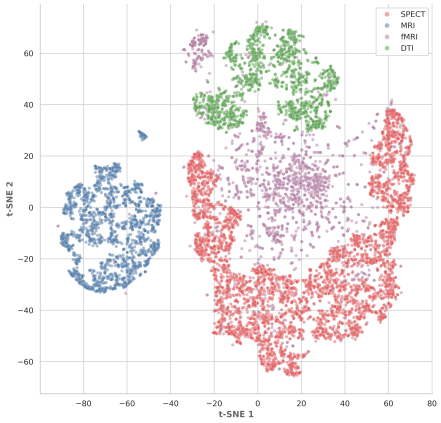

Figure 27: Embedding space of only real embeddings. Visualizations of the multi-modal embedding space demonstrated that real modalities successfully shared the same latent space without independent clustering.

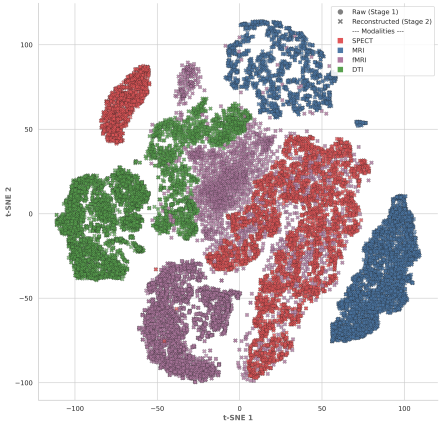

Figure 28: Joint embedding space of real and generated embeddings. Hallucinated reconstructions of missing modalities overlapped reliably with real embeddings.

A.2 Downstream Task Subspaces

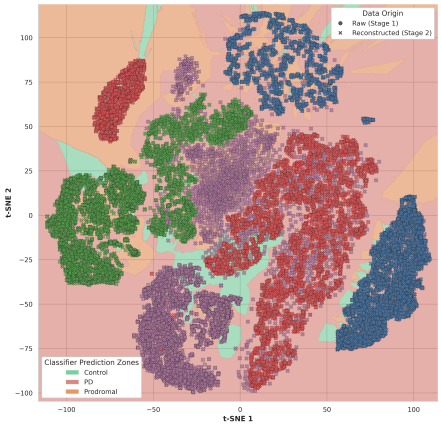

Figure 29: Classification boundaries learned by the downstream classification task. Projections of the downstream classification task onto the global embedding space confirmed that the model learned distinct, logical subspaces for classification boundaries.

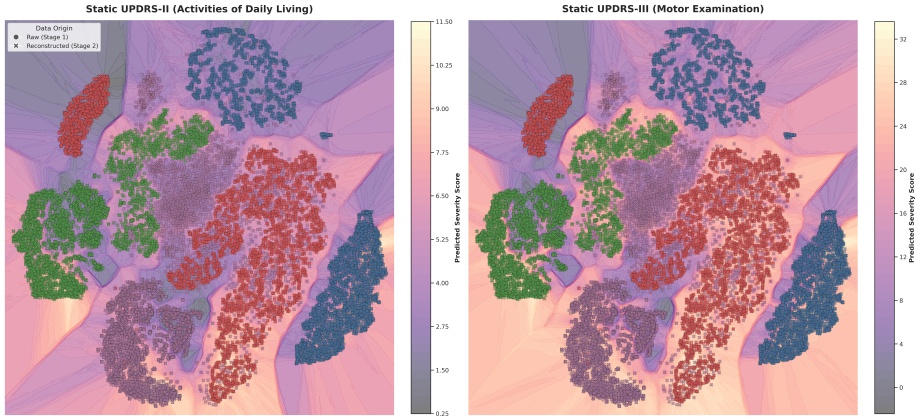

Figure 30: Regression gradients learned by the downstream severity regression task. Projections of the downstream severity regression task onto the global embedding space confirmed that the model learned distinct severity regression gradients.

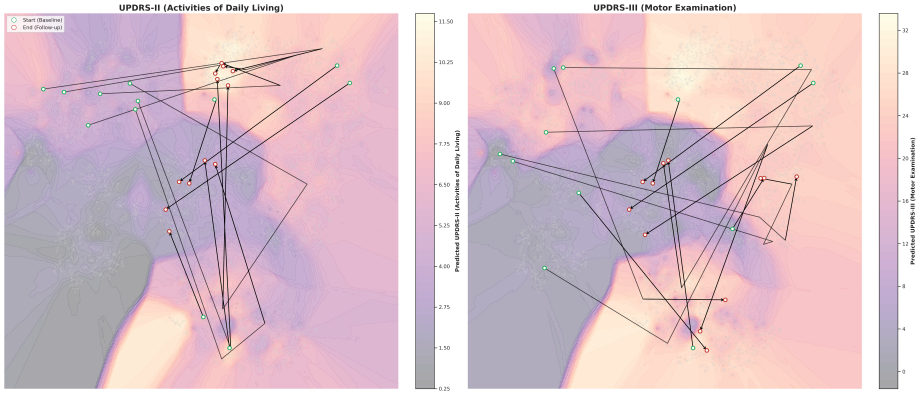

Figure 31: Severity trajectories learned by the downstream progression task. Projections of the downstream severity progression task onto the global embedding space confirmed that the model learned distinct severity progression trajectories.

References

[1] Lorraine V. Kalia and Anthony E. Lang. "Parkinson's disease". In: *The Lancet* 386.9996 (Aug. 29, 2015), pp. 896–912. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(14)61393-3.

[2] Bastiaan R. Bloem, Michael S. Okun, and Christine Klein. "Parkinson's disease". In: *The Lancet* 397.10291 (June 12, 2021), pp. 2284–2303. ISSN: 0140-6736, 1474-547X. DOI: 10.1016/S0140-6736(21)00218-X.

[3] Dongning Su et al. "Projections for prevalence of Parkinson's disease and its driving factors in 195 countries and territories to 2050: modelling study of Global Burden of Disease Study 2021". In: *BMJ* 388 (Mar. 5, 2025), e080952. ISSN: 1756-1833. DOI: 10.1136/bmj-2024-080952.

[4] Christopher G. Goetz et al. "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Scale presentation and clinimetric testing results". In: *Movement Disorders* 23.15 (2008), pp. 2129–2170. ISSN: 1531-8257. DOI: 10.1002/mds.22340.

[5] Girish Katti, Syeda Ara, and Dr Shireen. "Magnetic resonance imaging (MRI) - A review". In: *Intl J Dental Clin* 3 (Mar. 31, 2011).

[6] Amée F. Wolters et al. "Resting-state fMRI in Parkinson's disease patients with cognitive impairment: A meta-analysis". In: *Parkinsonism & Related Disorders* 62 (May 1, 2019), pp. 16–27. ISSN: 1353-8020. DOI: 10.1016/j.parkreldis.2018.12.016.

[7] Yu Zhang and Marc A. Burock. "Diffusion Tensor Imaging in Parkinson's Disease and Parkinsonian Syndrome: A Systematic Review". In: *Frontiers in Neurology* 11 (Sept. 25, 2020). ISSN: 1664-2295. DOI: 10.3389/fneur.2020.531993.

[8] T. Brücke et al. "SPECT and PET imaging of the dopaminergic system in Parkinson's disease". In: *Journal of Neurology* 247.4 (Sept. 1, 2000), pp. IV2–IV7. ISSN: 1432-1459. DOI: 10.1007/PL00007769.

[9] Priscila T. Levi et al. "The effect of using group-averaged or individualized brain parcellations when investigating connectome dysfunction in psychosis". In: *Network Neuroscience* 7.4 (Dec. 22, 2023), pp. 1228–1247. ISSN: 2472-1751. DOI: 10.1162/netn_a_00329.

[10] Xiaofei Huang et al. "Structural connectivity from DTI to predict mild cognitive impairment in de novo Parkinson's disease". In: *NeuroImage: Clinical* 41 (Jan. 1, 2024), p. 103548. ISSN: 2213-1582. DOI: 10.1016/j.nicl.2023.103548.

[11] Chun-Hung Yeh et al. "Mapping Structural Connectivity Using Diffusion MRI: Challenges and Opportunities". In: *Journal of Magnetic Resonance Imaging* 53.6 (2021), pp. 1666–1682. DOI: https://doi.org/10.1002/jmri.27188.

[12] P. Hagmann et al. “DTI mapping of human brain connectivity: statistical fibre tracking and virtual dissection”. In: *NeuroImage* 19.3 (July 1, 2003), pp. 545–554. ISSN: 1053-8119. DOI: 10.1016/S1053-8119(03)00142-3.

[13] Kaiming Li et al. “Review of methods for functional brain connectivity detection using fMRI”. In: *Computerized Medical Imaging and Graphics* 33.2 (Mar. 1, 2009), pp. 131–139. ISSN: 0895-6111. DOI: 10.1016/j.compmedimag.2008.10.011.

[14] Xi Zhang et al. “Multi-View Graph Convolutional Network and Its Applications on Neuroimage Analysis for Parkinson’s Disease”. In: *AMIA Annual Symposium Proceedings* 2018 (Dec. 5, 2018), pp. 1147–1156. ISSN: 1942-597X.

[15] Jie Zhou et al. *Graph Neural Networks: A Review of Methods and Applications*. Oct. 6, 2021. DOI: 10.48550/arXiv.1812.08434. arXiv: 1812.08434[cs].

[16] Yi Yang, Hejie Cui, and Carl Yang. *PTGB: Pre-Train Graph Neural Networks for Brain Network Analysis*. May 20, 2023. DOI: 10.48550/arXiv.2305.14376. arXiv: 2305.14376[q-bio].

[17] Gabriele Corso et al. “Graph neural networks”. In: *Nature Reviews Methods Primers* 4.1 (Mar. 7, 2024), p. 17. ISSN: 2662-8449. DOI: 10.1038/s43586-024-00294-7.

[18] Reinhard Diestel. *Graph Theory*. Springer Nature, Jan. 20, 2025. 466 pp. ISBN: 978-3-662-70107-2.

[19] Si Zhang et al. “Graph convolutional networks: a comprehensive review”. In: *Computational Social Networks* 6.1 (Nov. 10, 2019), p. 11. ISSN: 2197-4314. DOI: 10.1186/s40649-019-0069-y.

[20] Petar Veličković et al. *Graph Attention Networks*. Feb. 4, 2018. DOI: 10.48550/arXiv.1710.10903. arXiv: 1710.10903[stat].

[21] Yujia Li et al. *Gated Graph Sequence Neural Networks*. Sept. 22, 2017. DOI: 10.48550/arXiv.1511.05493. arXiv: 1511.05493[cs].

[22] Keyulu Xu et al. *How Powerful are Graph Neural Networks?* Feb. 22, 2019. DOI: 10.48550/arXiv.1810.00826. arXiv: 1810.00826[cs].

[23] *A review of the use of magnetic resonance imaging in Parkinson’s disease - Nadya Pyatigorskaya, Cécile Gallea, Daniel Garcia-Lorenzo, Marie Vidailhet, Stéphane Lehericy, 2014*. URL: https://journals.sagepub.com/doi/full/10.1177/1756285613511507 (visited on 01/12/2026).

[24] Julia J. Harris, Clare Reynell, and David Attwell. “The physiology of developmental changes in BOLD functional imaging signals”. In: *Developmental Cognitive Neuroscience* 1.3 (Apr. 27, 2011), pp. 199–216. ISSN: 1878-9293. DOI: 10.1016/j.dcn.2011.04.001.

[25] Logan R. Ranzenberger, Joe M. Das, and Travis Snyder. “Diffusion Tensor Imaging”. In: *StatPearls*. Treasure Island (FL): StatPearls Publishing, 2025.

[26] Bruce Fischl. "FreeSurfer". In: *NeuroImage* 62.2 (Aug. 15, 2012), pp. 774–781. ISSN: 1095-9572. DOI: 10.1016/j.neuroimage.2012.01.021.

[27] Leonie Henschel et al. "FastSurfer - A fast and accurate deep learning based neuroimaging pipeline". In: *NeuroImage* 219 (Oct. 1, 2020), p. 117012. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2020.117012.

[28] Abhijit Guha Roy et al. "QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy". In: *NeuroImage* 186 (Feb. 1, 2019), pp. 713–727. ISSN: 1053-8119. DOI: 10.1016/j.neuroimage.2018.11.042.

[29] Oscar Esteban et al. "fMRIPrep: a robust preprocessing pipeline for functional MRI". In: *Nature Methods* 16.1 (Jan. 2019), pp. 111–116. ISSN: 1548-7105. DOI: 10.1038/s41592-018-0235-4.

[30] Baxter P. Rogers et al. "Assessing functional connectivity in the human brain by fMRI". In: *Magnetic Resonance Imaging* 25.10 (Dec. 1, 2007), pp. 1347–1357. ISSN: 0730-725X. DOI: 10.1016/j.mri.2007.03.007.

[31] Alexandre Abraham et al. "Machine learning for neuroimaging with scikit-learn". In: *Frontiers in Neuroinformatics* 8 (Feb. 21, 2014), p. 14. ISSN: 1662-5196. DOI: 10.3389/fninf.2014.00014.

[32] Kahini Mehta et al. "XCP-D: A robust pipeline for the post-processing of fMRI data". In: *Imaging Neuroscience* 2 (Aug. 13, 2024), imag–2–00257. ISSN: 2837-6056. DOI: 10.1162/imag_a_00257.

[33] Matthew Cieslak et al. "QSIPrep: an integrative platform for preprocessing and reconstructing diffusion MRI data". In: *Nature Methods* 18.7 (July 2021), pp. 775–778. ISSN: 1548-7105. DOI: 10.1038/s41592-021-01185-5.

[34] Amritha Harikumar et al. "Fully automated estimation of fMRI-guided SPECT brain networks and their functional network connectivity in schizophrenia patients versus controls: A NeuroMark ICA approach". In: *Aperture Neuro* 5 (June 11, 2025). DOI: 10.52294/001c.138726.

[35] Gang Qu et al. "Integrated brain connectivity analysis with fMRI, DTI, and sMRI powered by interpretable graph neural networks". In: *Medical Image Analysis* 103 (2025), p. 103570. ISSN: 1361-8415. DOI: https://doi.org/10.1016/j.media.2025.103570.

[36] Hejie Cui et al. "BrainGB: A Benchmark for Brain Network Analysis with Graph Neural Networks". In: *IEEE transactions on medical imaging* 42.2 (Feb. 2023), pp. 493–506. ISSN: 0278-0062. DOI: 10.1109/TMI.2022.3218745.

[37] Qingbao Yu et al. "Application of Graph Theory to Assess Static and Dynamic Brain Connectivity: Approaches for Building Brain Graphs". In: *Proceedings of the IEEE* 106.5 (May 2018), pp. 886–906. ISSN: 1558-2256. DOI: 10.1109/JPROC.2018.2825200.

[38] Jie Mei, Christian Desrosiers, and Johannes Frasnelli. “Machine Learning for the Diagnosis of Parkinson’s Disease: A Review of Literature”. In: *Frontiers in Aging Neuroscience* Volume 13 - 2021 (2021). ISSN: 1663-4365. DOI: 10.3389/fnagi.2021.633752.

[39] Manan Binth Taj Noor et al. “Application of deep learning in detecting neurological disorders from magnetic resonance images: a survey on the detection of Alzheimer’s disease, Parkinson’s disease and schizophrenia”. In: *Brain Informatics* 7.1 (Oct. 9, 2020), p. 11. ISSN: 2198-4026. DOI: 10.1186/s40708-020-00112-2.

[40] Maria Frasca et al. *Predicting Parkinson’s disease evolution using deep learning*. Jan. 5, 2024. DOI: 10.48550/arXiv.2312.17290. arXiv: 2312.17290[eess].

[41] Nguyen Linh Dan Le et al. *BrainMAP: Multimodal Graph Learning For Efficient Brain Disease Localization*. June 12, 2025. DOI: 10.48550/arXiv.2506.11178. arXiv: 2506.11178[cs].

[42] Jiajun Ma et al. *BrainHGT: A Hierarchical Graph Transformer for Interpretable Brain Network Analysis*. Nov. 18, 2025. DOI: 10.48550/arXiv.2511.17604. arXiv: 2511.17604[cs].

[43] Gaotang Li et al. “Interpretable Sparsification of Brain Graphs: Better Practices and Effective Designs for Graph Neural Networks”. In: *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. KDD ’23: The 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Long Beach CA USA: ACM, Aug. 6, 2023, pp. 1223–1234. ISBN: 979-8-4007-0103-0. DOI: 10.1145/3580305.3599394.

[44] Xiaoxiao Li et al. “BrainGNN: Interpretable Brain Graph Neural Network for fMRI Analysis”. In: *Medical Image Analysis* 74 (Dec. 1, 2021), p. 102233. ISSN: 1361-8415. DOI: 10.1016/j.media.2021.102233.

[45] Vasiliki Rizou et al. “PDualNet: a deep learning framework for joint prediction of Parkinson’s disease progression subtype and MDS-UPDRS scores”. In: *Scientific Reports* 15.1 (Nov. 25, 2025), p. 41931. ISSN: 2045-2322. DOI: 10.1038/s41598-025-25812-9.

[46] Jie Lian et al. “Personalized progression modelling and prediction in Parkinson’s disease with a novel multi-modal graph approach”. In: *NPJ Parkinson’s Disease* 10 (Dec. 1, 2024), p. 229. ISSN: 2373-8057. DOI: 10.1038/s41531-024-00832-w.

[47] Juhyun Lee et al. “Structural MRI–Informed Multimodal Fusion for Robust Alzheimer’s Disease Prediction”. In: NeurIPS 2025 Workshop for Imageomics: Discovering Biological Knowledge from Images Using AI. Oct. 9, 2025.

[48] Alec Radford et al. *Learning Transferable Visual Models From Natural Language Supervision*. Feb. 26, 2021. DOI: 10.48550/arXiv.2103.00020. arXiv: 2103.00020[cs].

[49] Andrey Guzhov et al. *AudioCLIP: Extending CLIP to Image, Text and Audio*. June 24, 2021. DOI: 10.48550/arXiv.2106.13043. arXiv: 2106.13043[cs].

[50] Yue Ruan et al. "TriCoLo: Trimodal Contrastive Loss for Text to Shape Retrieval". In: *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. 2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Waikoloa, HI, USA: IEEE, Jan. 3, 2024, pp. 5803–5813. ISBN: 979-8-3503-1892-0. DOI: 10.1109/WACV57701.2024.00571.

[51] Rohit Girdhar et al. *ImageBind: One Embedding Space To Bind Them All*. May 31, 2023. DOI: 10.48550/arXiv.2305.05665. arXiv: 2305.05665[cs].

[52] Lucas Pinheiro Cinelli et al. "Variational Autoencoder". In: *Variational Methods for Machine Learning with Applications to Deep Networks*. Ed. by Lucas Pinheiro Cinelli et al. Cham: Springer International Publishing, 2021, pp. 111–149. ISBN: 978-3-030-70679-1. DOI: 10.1007/978-3-030-70679-1_5.

[53] Srikumar Sastry et al. *ProM3E: Probabilistic Masked MultiModal Embedding Model for Ecology*. Nov. 4, 2025. DOI: 10.48550/arXiv.2511.02946. arXiv: 2511.02946[cs].

[54] Kenneth Marek et al. "The Parkinson Progression Marker Initiative (PPMI)". In: *Progress in Neurobiology* 95.4 (Dec. 2011), pp. 629–635. ISSN: 03010082. DOI: 10.1016/j.pneurobio.2011.09.005.

[55] Roger Gunn, Christopher Coello, and Graham Searle. "Molecular Imaging And Kinetic Analysis Toolbox (MIAKAT) - A Quantitative Software Package for the Analysis of PET Neuroimaging Data". In: *Journal of Nuclear Medicine* 57 (supplement 2 May 1, 2016), pp. 1928–1928. ISSN: 0161-5505, 2159-662X.

[56] Dristi Datta et al. *A Sparse-Attention Deep Learning Model Integrating Heterogeneous Multimodal Features for Parkinson's Disease Severity Profiling*. arXiv:2601.00519 [cs]. Jan. 2026. DOI: 10.48550/arXiv.2601.00519.

[57] Tongle Zhi et al. "MultimodalCNN-PD: a Parkinson's disease diagnostics framework using multimodal convolutional neural network". English. In: *Frontiers in Aging Neuroscience* 18 (Feb. 2026). ISSN: 1663-4365. DOI: 10.3389/fnagi.2026.1733075.

[58] Gattellaro G et al. "White matter involvement in idiopathic Parkinson disease: a diffusion tensor imaging study". In: *AJNR. American journal of neuroradiology* 30.6 (June 2009). ISSN: 1936-959X. DOI: 10.3174/ajnr.A1556.

[59] Tessitore A et al. "Default-mode network connectivity in cognitively unimpaired patients with Parkinson disease". In: *Neurology* 79.23 (Apr. 12, 2012). ISSN: 1526-632X. DOI: 10.1212/WNL.0b013e31827689d6.

[60] Dalaker To et al. “Ventricular enlargement and mild cognitive impairment in early Parkinson’s disease”. In: *Movement disorders : official journal of the Movement Disorder Society* 26.2 (Jan. 2, 2011). ISSN: 1531-8257. DOI: 10.1002/mds.23443.

[61] Fling Bw et al. “Asymmetric pedunculopontine network connectivity in parkinsonian patients with freezing of gait”. In: *Brain : a journal of neurology* 136 (Pt 8 Aug. 2013). ISSN: 1460-2156. DOI: 10.1093/brain/awt172.

[62] Mak E et al. “Baseline and longitudinal grey matter changes in newly diagnosed Parkinson’s disease: ICICLE-PD study”. In: *Brain : a journal of neurology* 138 (Pt 10 Oct. 2015). ISSN: 1460-2156. DOI: 10.1093/brain/awv211.

[63] Fiorenzato E et al. “Asymmetric Dopamine Transporter Loss Affects Cognitive and Motor Progression in Parkinson’s Disease”. In: *Movement disorders : official journal of the Movement Disorder Society* 36.10 (Oct. 2021). ISSN: 1531-8257. DOI: 10.1002/mds.28682.

[64] Nemmi F et al. “Parkinson’s disease and local atrophy in subcortical nuclei: insight from shape analysis”. In: *Neurobiology of aging* 36.1 (Jan. 2015). ISSN: 1558-1497. DOI: 10.1016/j.neurobiolaging.2014.07.010.

[65] Nandhagopal R et al. “Progression of dopaminergic dysfunction in a LRRK2 kindred: a multitracer PET study”. In: *Neurology* 71.22 (Nov. 25, 2008). ISSN: 1526-632X. DOI: 10.1212/01.wnl.0000335973.66333.58.

[66] Braak H et al. “Staging of brain pathology related to sporadic Parkinson’s disease”. In: *Neurobiology of aging* 24.2 (Apr. 2003). ISSN: 0197-4580. DOI: 10.1016/s0197-4580(02)00065-9.

[67] Wu T and Hallett M. “The cerebellum in Parkinson’s disease”. In: *Brain : a journal of neurology* 136 (Pt 3 Mar. 2013). ISSN: 1460-2156. DOI: 10.1093/brain/aws360.