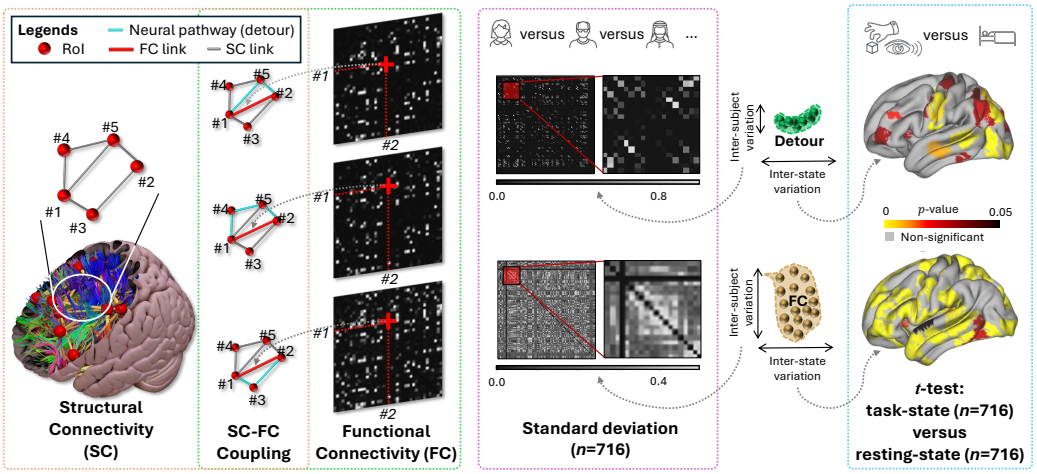
Graphical Abstract

NeuroDetour: A Neural Pathway Transformer for Uncovering Structural-Functional Coupling Mechanisms in Human Connectome

Ziquan Wei, Tingting Dan, Jiaqi Ding, Paul J. Laurienti, Guorong Wu



Highlights

***NeuroDetour*: A Neural Pathway Transformer for Uncovering Structural-Functional Coupling Mechanisms in Human Connectome**

Ziquan Wei, Tingting Dan, Jiaqi Ding, Paul J. Laurienti, Guorong Wu

- A novel multivariate SC-FC coupling allows us to develop an explainable deep model based on neural detour.
- A Transformer model that can learn path in human connectome without preprocessing.
- Validate NeuroDetour on large-scale public datasets with a total of 10,886 fMRI scans.

NeuroDetour: A Neural Pathway Transformer for Uncovering Structural-Functional Coupling Mechanisms in Human Connectome

Ziquan Wei[a], Tingting Dan[a], Jiaqi Ding[a], Paul J. Laurienti[b], Guorong Wu[a]

[a] *University of North Carolina at Chapel Hill, 334 Emergency Room Dr, Chapel Hill, 27599, NC, USA*

[b] *Wake Forest University School of Medicine, Winston-Salem, 27109, NC, USA*

Abstract

Although modern imaging methods enable in-vivo examination of connections between distinct brain areas, we still lack a comprehensive understanding of how anatomical structure underpins brain function and how spontaneous fluctuations in neural activity give rise to cognition. At the same time, many efforts in machine learning have focused on modeling the complex, nonlinear relationships between neuroimaging signals and observable traits. Yet, current machine learning techniques often overlook fundamental neuroscience insights, making it difficult to interpret transient neural dynamics in terms of cognitive behavior. To bridge this gap, we turn our attention to the interplay between structural connectivity (SC) and functional connectivity (FC), reframing this open question in network neuroscience as a graph representation learning task centered on neural pathways. In particular, we introduce the notion of a "topological detour" to describe how a given instance of FC (i.e., a direct functional connection) is physically supported by underlying SC pathways (the detour), forming a feedback loop between brain structure and function. By considering these multi-hop detour routes that mediate SC-FC coupling, we design a novel multi-head self-attention mechanism within a Transformer architecture. Building on these ideas, we present a biologically inspired deep-learning framework, *NeuroDetour*, that extracts connectomic feature representations from large-scale neuroimaging datasets and can be applied to downstream tasks such as task classification and disease prediction. We validated *NeuroDetour* on extensive public cohorts, including the Human Connectome Project (HCP) and UK Biobank (UKB), using both supervised learning and zero-shot settings. In all scenarios, *NeuroDetour* achieves state-

of-the-art results.

Keywords: Functional MRI, Network Neuroscience, Structure-function Coupling, Deep Learning, Graph Transformer, Neural Circuits

1. Introduction

The human brain, comprised of intricate networks of white matter fibers, forms a complex system [1]. Along with these structural connections are neuronal oscillations, which orchestrate functional variations across widespread neural circuits, crucial for various cognitive functions essential to daily life [2]. As neuroimaging techniques become increasingly prevalent in neuroscience research, striking efforts have been made to understand the role of structural connectivity (SC) and functional connectivity (FC) from a holistic perspective, recognizing that the dynamic nature of the human brain drives spontaneous functional fluctuations [3]. SC is a static graph and physically wired by neuronal fibers extracted from Diffusion Weighted Imaging (DWI), and FC from functional Magnetic Resonance Imaging (fMRI) of the brain is a dynamic graph representing the correlation between regional Blood-Oxygen-Level-Dependent (BOLD) signals [1].

Multiple neuroscience studies indicate that high-level cognition and behavior emerge from a close coupling between structural and functional neuroimaging. Consequently, understanding how SC and FC interplay has become a crucial step toward reverse engineering the human mind from the brain connectome [5]. Over the past decade, researchers have proposed various computational models to characterize this relationship, including graph harmonic approaches [6, 7], network communication frameworks [8], multivariate statistical methods [9], and deep learning techniques for structure-function mapping [10, 11, 12, 13, 14]. Despite these advances, the precise link between SC and FC remains elusive [15]. In particular, many pairs of regions that show strong functional connectivity (FC, green box in Fig. 1) lack a direct structural pathway (orange box in Fig. 1) [16, 17, 18]. Although SC and FC topologies do not always align, there is a growing consensus that each FC connection is in fact supported by a subgraph of SC. In other words, when two regions exhibit a functional link (red link in Fig. 1), the SC network contains an indirect path (blue links in the orange box) that physically underlies that FC. Therefore, developing high-order topological representation is essential for decoding the SC-FC coupling mechanism in the human brain.

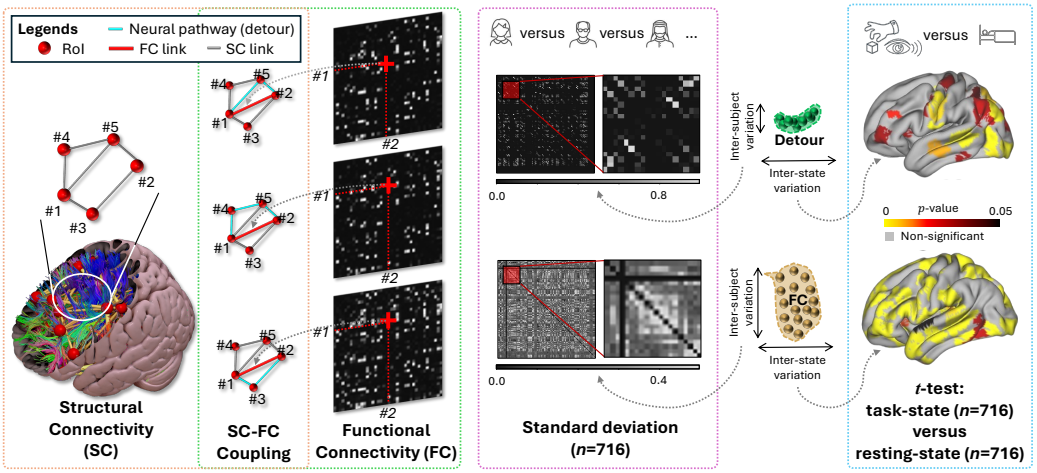


Figure 1: Embedding a novel multivariate SC–FC coupling mechanism into an interpretable deep model. Orange box: Structural connectivity (SC), illustrated by gray edges, reflects the anatomical strength of white-matter fibers linking two brain regions. Because anatomical wiring changes slowly, SC can be treated as a relatively static substrate. Green box: Functional connectivity (FC), in contrast, captures the dynamic coordination of neural activity and is often regarded as the brain's time-varying network topology [4]. Overlap (Orange × Green): Our multivariate SC–FC coupling mechanism. Instead of assuming a one-to-one correspondence between SC and FC, we model each functional connection as being supported by multiple interacting structural pathways (i.e., detour) that jointly facilitate information flow even when the direct SC between two regions is weak/absent. The comparison between inter-subject (purple box) and inter-cognitive state (blue box) variations of FC and detour adjacency (see Sec. 3) matrices.

Previous work has explored how to derive meaningful representations for subgraph topology, enabling a range of downstream applications. In knowledge graphs, such representations have supported reasoning-path discovery [19] and cycle-basis extraction [20]. In molecular modeling, subgraph encoding has been advanced through neural fingerprints [21], junction-tree autoencoders [22], and cellular Weisfeiler–Leman (WL) tests [23], each capturing distinct aspects of local structural organization. Especially, graph data in the realm of neuroscience research provides additional insight that underlines particular neurobiological mechanisms of interest, such as substructure [24, 25] and snapshot embeddings [26, 27] in SC. Nevertheless, learning the representation of physical neural pathways by combining SC and FC is not yet considered. Despite reviews [28, 29] of Graph Neural Networks (GNN) for neuroimaging showing various methods for SC-FC mapping, the downstream application

has not been extended to disease diagnosis or cognition recognition.

On the other hand, it is not straightforward to apply a generic method that learns paths in a graph to our topic. Although network path is an important substructure, unfortunately, finding all paths is an NP-hard (Nondeterministic Polynomial time-hard) problem. Recently, Path Neural Networks (PathNN) [30] first proposed to learn the representation of every existing simple path with limited length using a Multi-Layer Perceptron (MLP), and [31] brings the same idea to the WL algorithm, demonstrating the theoretical power of PathNN-like models distinguishing topology. Although they achieved state-of-the-art (SOTA) performance on chemical and molecule datasets to distinguish graphs of small size, the need for a larger graph, like a brain network, cannot be fulfilled. To address the computational costs, existing works extract structural representatives from all paths, e.g., the shortest path distance encoded in various graph transformers and message-passing neural networks [32, 33, 34, 35, 36]. There is still a preprocessing step that cannot be skipped. The gateway to applying the above methods to the human connectome is the dense connectivity resulting in unavoidable verbose preprocessing as listed in Table 1, where H denotes the limited length of the path. Even though the superior performance demonstrated in previous studies cannot be directly applied to network neuroscience due to seconds or minutes of preprocessing per graph, while our *NeuroDetour* is designed to model paths without calculating or processing any type of distance, path, or handcrafted topological transformations beyond the standard SC/FC matrices produced by fMRIPrep [37] and QSIPrep [38].

Table 1: Preprocessing by GNNs for network paths of the SC network ($N_{node} = 116$) extracted from a single UKB subject.

	Type	Time
PathNN [30]	All simple paths	5.23s ($H = 4$), 650s ($H = 5$)
Graphormer [34]	Shortest distance	270ms ($H = 7$)
NeuroDetour	None	-

To overcome the above challenges and advance our understanding of SC-FC coupling, we move beyond the traditional notion of univariate coupling, which focuses solely on the one-to-one correspondence between a direct structural link and a direct functional link, and introduce a multivariate framework. In this new paradigm, each observed FC connection is associated not just

with its immediate SC counterpart, but with an entire "detour" pathway through the SC network. For example, in the green box of Fig. 1, the red line marks a direct functional link between regions #1 and #2, while the series of blue lines traces the indirect structural route that supports that functional interaction. Although the overall SC topology remains fixed, the FC patterns can arise from different multi-hop SC pathways over time and only the shortest pathway and SC absence have been studied in the previous works [16, 39]. Importantly, the direct FC link and its SC detour together form a cyclic loop that reflects both structural and functional organization [40, 41]. This perspective of indirect SC, derived from neuroscience studies, shows that synchronized activity between two regions depends on the underlying structural circuitry, rather than on a single, direct fiber bundle.

From the perspective of machine learning, the primary objective is to learn a mapping from SC-FC to cognitive and clinical outcomes. We propose a novel deep model, coined *NeuroDetour*, to (1) enhance predictive performance in cognitive status classification by incorporating both SC and FC information, and (2) elucidate the underlying SC-FC coupling mechanism. To achieve the goals, we posit that a comprehensive understanding of cognitive neuroscience necessitates explicit modeling of how functional fluctuations are supported by structural pathways. Concretely, *NeuroDetour* exposes the SC-FC relationship by linking each observed FC to its candidates of SC detour and then a weighted-sum detour representation for phenotypic prediction. This design leads to a combined feature representation of all multi-hop detours (Sec. 3.1) and an association between the detour representation and cognitive (Sec. 2) or clinical outcomes (Sec. 4.7). To embed this conceptual framework within an end-to-end deep architecture, we focus on designing a new multi-head self-attention (MHSA) module that is explicitly informed by our multivariate SC–FC coupling paradigm. Within each attention layer, *NeuroDetour* learns multiple "detour" paths in the SC graph that are associated with individual FC links. Instead of learns from FC with smoothed inter-subject variation, as shown in Fig. 1 purple box, each attention head processes one of the detour paths with a more sparse inter-subject difference. Thereby, *NeuroDetour* enables learning how the neural pathway (aka. SC detour) of varying lengths jointly underpin functional interactions under less noisy subject-wise bias (see Fig. 1 blue box). Following the methodology presentation in the conference version [42], we applied it to the experiments (1) testing the associations between detour pathways and brain disordering and aging among two datasets, and (2) twelve challenging cognitive tests among UKB phenotypic fields.

Multiple significant differences regarding detour length and passing-by brain regions of our new findings further enhance our work. In summary, our contribution is threefold.

1. **Multivariate SC-FC coupling framework.** We extend the conventional univariate concept of structure-function correspondence to a multivariate formulation in which each functional link is supported by a subgraph of SC. This approach is grounded in both neuroscientific evidence and graph theoretic principles, thus yielding an interpretable model architecture.
2. **Transformer-based path representation learning.** By embedding the SC-FC coupling mechanism within a Transformer, *NeuroDetour* is capable of learning path level representations directly from connectome graphs, obviating the need for handcrafted preprocessing, and scaling effectively to large neuroimaging cohorts.
3. **Large-scale validation across public datasets.** We evaluate *NeuroDetour* on a total of 10,886 fMRI sessions drawn from the Human Connectome Project (HCP), UK Biobank (UKB), Alzheimer's Disease Neuroimaging Initiative (ADNI), and Open Access Series of Imaging Studies (OASIS). Experimental results demonstrate (i) superior accuracy in predicting healthy cognitive states, disease risk of Alzheimer's and ICD10, and cognitive tests prediction; (ii) enhanced interpretability, as evidenced by the identification of latent neural pathways from SC-FC coupling associated with disordering and aging; and (iii) clinical utility through robust performance and successful zero-shot learning across distinct datasets.

2. Inter-subject Variability in Human Connectome Project

Inter-subject variability in functional connectivity (FC) frequently obscures task-specific activation patterns. FC patterns naturally differ across individuals due to factors such as aging, vascular differences, and baseline neural dynamics, causing FC degree to fluctuate widely even within the same cognitive condition. As a result, the variability across subjects can be as large as, or larger than, the task-evoked modulation itself, making it difficult to isolate changes driven by the cognitive status rather than by undesired individual differences. Even in the Human Connectome Project Aging (HCPA) cohort, where image acquisition is standardized to minimize heterogeneity,

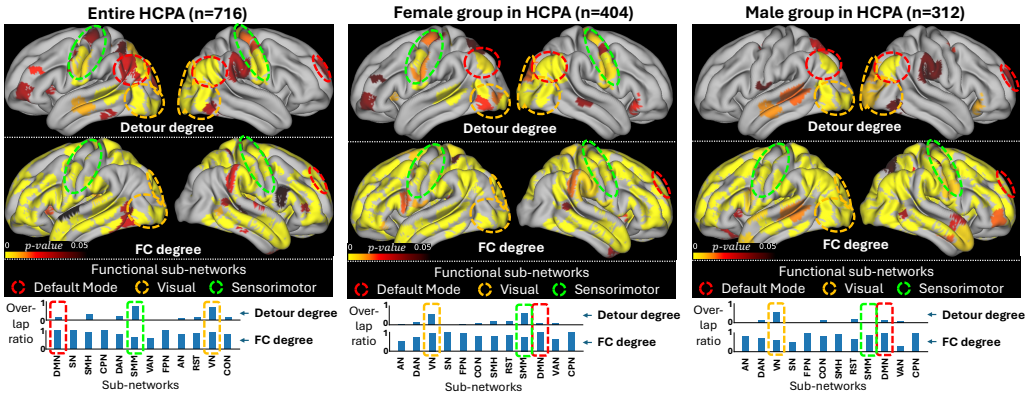


Figure 2: Top: Brain regions showing significant differences between resting-state and VISMOTOR task state are identified using both detour degree and FC degree. The color scale indicates the corresponding p-values ($0 \leq p \leq 0.05$) from a paired t-test. Bottom: The bar plot displays the overlap ratio (y-axis) between the number of significant regions identified and the total number of regions within each predefined subnetwork (x-axis) among the Gordon333 atlas. The default mode (red circle/box), visual (orange circle/box), and sensorimotor (green circle/box) networks are highlighted due to their well-established involvement in differentiating resting-state and VISMOTOR.

this challenge remains evident. To illustrate, Fig. 2 highlights regions that differ significantly between resting-state and VISMOTOR task conditions, as assessed by both detour degree and FC degree. Here, "detour degree" for a node is defined as the number of distinct topological detour pathways (with a maximum radius 6) that end at that node. Such pathways in a SC graph are identified using a depth-first search algorithm[1] between each pair of nodes that are connected by an FC edge.

Specifically, we calculated detour degree for each of the Gordon333 parcellation regions in both resting and VISMOTOR states, across all 716 HCPA subjects. For each parcel, we then constructed two paired samples, one from the resting state and one from the task state, by aligning detour degree values from the same subject under different states. A paired t-test was then applied to these paired detour degree vectors to identify parcels with significant task-

[1]We used the NetworkX implementation, `networkx.all_simple_paths` (`https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.simple_paths.all_simple_paths.html`), where input graph is the SC, source and target nodes are every node pair that has an FC connection.

related changes (threshold: $p < 0.05$). In parallel, we calculated FC degree for each parcel. They were similarly subjected to a paired t-test between resting and task conditions, with the same significance threshold. To quantify the significance level in functional subnets, we proposed an overlap ratio within each predefined subnetwork of the Gordon333 atlas. For each subnetwork i, the overlap ratio is defined as N_{sign_i}/N_{net_i}, where N_{sign_i} is the number of parcels of i-th community have $p \leq 0.05$ and N_{net_i} is the total number of parcels in i-th subnetwork. This overlap ratio quantifies the extent to which each functional subnetwork is implicated in task-related reorganization, as detected by detour and FC degree metrics.

In Fig. 2, significant parcels are overlaid on cortical surfaces and colored according to their p-values. Each of these parcels is then assigned to one of 12 functional subnetworks, which are predefined by the Gordon333 atlas. The overlap ratios are displayed as bar plots beneath the cortical renders (x-axis: subnetwork identity; y-axis: overlap ratio). Previous studies indicate that resting-state activity primarily involves the default mode network [43] (highlighted by the red circle), whereas the VISMOTOR task engages both visual (orange circle) and sensorimotor [44] (green circle) regions. In this context, detour degree demonstrates superior discrimination relative to FC degree: most significant parcels identified by detour degree align with the expected subnetworks (default mode, visual, and sensorimotor), while producing fewer false positives in unrelated subnetworks.

Aside from the t-test on the entire HCPA dataset, t-tests on diverse populations by gender are shown in Fig. 2, indicating the observation in the entire population is not biased by diversity in terms of inter-subject variation. Specifically, both groups show statistical significance solely on brain sub-networks related to functions instead of every region by FC degree. Note that the female group shows the most similar overlap ratio to the entire dataset because its sample size ($n = 404$) is closest to the total sample size ($n = 716$), reducing sampling variability relative to the male and youngest groups..

The results in Fig. 2 highlight the critical role of SC detours in identifying putative functional biomarkers, motivating us to embed the SC–FC coupling mechanism directly into the design of *NeuroDetour*.

3. Methods

Our goal is to couple neuroscience insight with deep learning capacity by explicitly encoding topological detours, multi-hop SC pathways that support functional interactions, within a graph-transformer framework.

Notations. Suppose SC and FC are denoted by adjacency matrices $\mathbf{A^S} \in \mathbb{R}^{N\times N}$ and $\mathbf{A^F} \in \mathbb{R}^{N\times N}$, respectively, where $A_{ij}^{(\cdot)}$ is the connectivity between i^{th} and j^{th} region $(i, j = 1, ..., N)$. In practice, SC and FC are calculated by subject-wise normalized white matter (WM) fiber counts and the Pearson correlation coefficient of hemodynamic signals, respectively. Furthermore, we use $\hat{\mathbf{A}}^{(\cdot)}$ to denote the binary adjacency matrix after high-pass filtering and adding self-loop for either SC or FC. The high-pass filter removes edges with weights below a fixed threshold. Specifically, for FC we retained only correlations with absolute value $\geq$0.5, and for SC we retained streamline-normalized weights $\geq$0.1. This thresholding procedure removes weak or noisy edges and functions as a high-pass filter on the weighted graph. Node attribute is denoted by $\mathbf{X} \in \mathbb{R}^{N\times C}$, where C is the feature dimension.

To capture multi-scale topological structure, we perform random walks on the SC graph, generating a sequence of multi-hop detour adjacency matrices as follows:

Definition 3.1 (Detour adjacency matrix)**.** $\mathbf{D}^h$, a binary matrix of shape $N \times N$ and stores whether a link of FC is associated with an h-hop topological detour, where a 1-hop detour is equivalent to an edge. It is obtained by element-wise production between binary matrices $\mathbf{D}^h := \left((\hat{\mathbf{A}}^{\mathbf{S}})^h > 0\right) \cdot \left(\hat{\mathbf{A}}^{\mathbf{F}}\right)$, where $1 \leq h \leq H$, and H is the maximum length of detour.

Note that $\mathbf{D}^h$ avoids finding all simple paths by our model, it significantly reduces computational costs with sufficient modeling power for neural pathways as discussed in Sec. 3.1.

NeuroDetour adopts a twin-branch architecture as a proof-of-concept design [42]. Each branch implements an identical MHSA) module based on the Transformer encoder [45]. The twin-branch setup enables aligned feature representations, allowing the model to jointly encode SC-derived detour information and FC link patterns.

Given a set of learnable parameters $\bar{\mathbf{W}}, \hat{\mathbf{W}} \in \mathbb{R}^{(HC)\times C}$ and $\bar{\boldsymbol{\alpha}}_h, \bar{\boldsymbol{\beta}}_h, \bar{\boldsymbol{\gamma}}_h, \hat{\boldsymbol{\alpha}}_h, \hat{\boldsymbol{\beta}}_h, \hat{\boldsymbol{\gamma}}_h \in \mathbb{R}^{C\times C}$ where $h = 1, \ldots, H$, H is head number of MHSA, and C denotes feature

dimension, the branch TD-MHSA is then defined as

$$\begin{aligned} f_{TD}(\mathbf{X}) &= \texttt{Concat}\left(\bar{\mathbf{X}}_1, \ldots, \bar{\mathbf{X}}_H\right)\bar{\mathbf{W}}, \\ \bar{\mathbf{X}}_h &= \texttt{Softmax}\left((\mathbf{X}\bar{\boldsymbol{\alpha}}_h)(\mathbf{X}\bar{\boldsymbol{\beta}}_h)^T + f_{mask}(\mathbf{D}^h)\right)(\mathbf{X}\bar{\boldsymbol{\gamma}}_h). \end{aligned} \tag{1}$$

Similarly, the FC-MHSA branch is defined as

$$\begin{aligned} f_{FC}(\mathbf{X}) &= \texttt{Concat}\left(\hat{\mathbf{X}}_1, \ldots, \hat{\mathbf{X}}_H\right)\hat{\mathbf{W}}, \\ \hat{\mathbf{X}}_h &= \texttt{Softmax}\left((\mathbf{X}\hat{\boldsymbol{\alpha}}_h)(\mathbf{X}\hat{\boldsymbol{\beta}}_h)^T + f_{mask}(\hat{\mathbf{A}}^{\mathbf{F}})\right)(\mathbf{X}\hat{\boldsymbol{\gamma}}_h). \end{aligned} \tag{2}$$

Applying masks to attention maps offers a principled way to restrict self-attention to the nodes of interest. To implement this, we define f_{mask} as an operation that assigns negative infinity to the false entries and zero to the true entries of a binary adjacency matrix. This ensures that the `Softmax` function effectively ignores the masked (false) positions in both branches.

The training loss integrates the downstream task objective and a consistency term $\mathcal{L}_{TD}$, defined as

$$\mathcal{L}_{TD} = ||f_{TD}(\mathbf{X}) - f_{FC}(\mathbf{X})||^2. \tag{3}$$

Taking classification as an example, the final loss $\mathcal{L} = \texttt{CELoss}\,(label, \mathbf{Y}) + \mathcal{L}_{TD}$, where the logits $\mathbf{Y} = \boldsymbol{\rho}^{-\frac{1}{2}}\hat{\mathbf{A}}^{\mathbf{F}}\boldsymbol{\rho}^{-\frac{1}{2}} f_{TD}(\mathbf{X})\boldsymbol{\Theta}$ with learnable parameters $\boldsymbol{\Theta} \in \mathbb{R}^{C\times n}$, n is class number and $\boldsymbol{\rho}$ is degree of brain network adjacency $\hat{\mathbf{A}}^{\mathbf{F}}$. This consistency constraint is designed to follow the SC-FC coupling insight of neuroscience by restricting the difference of representations from FC and TD branches so that the mapping between *NeuroDetour* representations and functional states of the human connectome is not a one-to-many scheme as mentioned in Section 1.

3.1. Path Representation

TD-MHSA strictly follows the pathway of topological detour after SC-FC coupling to produce node features hop-by-hop corresponding to each head of self-attention. Then, the following fact can be announced after expanding Eq. 1.

Fact 3.1. *The top-1 path is obtained by pathway weights* $\arg\max_{j,h}\left(\frac{1}{h}\sum_{j\in\mathbf{p}}\mathbf{S}_{ij}\bar{\boldsymbol{\gamma}}_h\bar{\mathbf{W}}_{i\sim j}\right)$, *where* $\mathbf{S}$ *denotes the value of softmax self-attention,* $\mathbf{p} \subset \mathbf{P}_i^H$ *is a set of node index of a path and* $\mathbf{P}_i^H$ *is the node collection of neural pathways within* H*-hop starting at* i*-th node.*

In plain words, we seek the top-ranked neural pathways by sorting weights for all filtered node indexes in TD-MHSA of *NeuroDetour*. The proof of the above Fact can be found in the Appendix. Taken together, *NeuroDetour* aggregates the representation of simple paths to recognize a brain graph with $O(N^2)$ computational complexity while being free of any pre-processing as listed in Table 1.

4. Results

We test *NeuroDetour* with three evaluations: (1) Clinical utility for Alzheimer's disease and International Classification of Diseases (ICD) classification, (2) cognitive state classification, and (3) cognitive tests prediction. Note that the second task serves as a fundamental benchmark for evaluating whether a model can capture meaningful functional reconfiguration of the brain. Task-evoked neural activity produces robust, well-characterized changes in functional connectivity compared to resting-state. Therefore, successful classification of task-related versus resting-state indicates that the model is sensitive to state-dependent FC patterns and can detect network-level shifts driven by cognitive engagement. This provides a controlled setting in which to assess the model's ability to leverage SC–FC coupling before applying it to more heterogeneous and clinically complex datasets such as ADNI and OASIS.

4.1. Datasets

Four public datasets are used in our experiments as listed in Table 2 with the profile of the functional connectivity (FC). The static data refers to FC being extracted from the entire session of fMRI. Dynamic data is split from the full scan of fMRI into 100-length slices, causing more graphs and different connectivity.

Table 2: Data profiles, where $|G|$ denotes the number of graphs, $|C|$ denotes the number of classes, and $avg(D)$ denotes the average degree of brain networks.

	HCPA		UKB		ADNI		OASIS	
	static	dynamic	static	dynamic	static	dynamic	static	dynamic
$\|G\|$	4,863	18,306	5,890	22,600	138	294	402	1,678
$\|C\|$	4	4	2	2	2	2	2	2
$avg(D)$	6.53	13.52	12.85	36.07	44.44	43.89	56.47	59.36

The Lifespan Human Connectome Project Aging (HCPA) dataset [46] offers a rich resource for investigating task-related brain activity across the adult lifespan. It comprises 717 participants, with 4,863 fMRI sessions and 716 DWI scans. Imaging includes three task paradigms targeting memory and sensorimotor processes—VISMOTOR, CARIT, and FACENAME, as well as resting-state acquisitions. In our framework, these four states are treated as a four-class classification and (for zero-shot learning) as a binary classification between resting and task states. For most experiments, brain regions are parcellated using the AAL atlas [47], while the Gordon atlas [48] is reserved for ablation analyses.

The United Kingdom Biobank (UKB) dataset provides a large-scale collection of MRI data, processed through the same pipeline applied to HCPA. It contains 5,483 fMRI and 3,162 DWI scans, all of which include a single task engaging both cognitive and sensorimotor domains [49, 50]. In our studies, we formulate this as a two-class classification and use the Gordon atlas [48] for parcellation.

The Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [51] offers preprocessed imaging from 138 fMRI and 135 DWI sessions, with AAL parcellation applied. Participants carry clinical labels across five categories: cognitively normal (CN), subjective memory complaints (SMC), early-stage mild cognitive impairment (EMCI), late-stage mild cognitive impairment (LMCI), and Alzheimer's disease (AD). To mitigate class imbalance, we merge CN, SMC, and EMCI into a single "CN" category, while LMCI and AD are grouped as the "AD" category, enabling a binary classification setup. Noted that although this subject dichotomization is intended to facilitate a binary classification framework for our analysis, it may introduce confounds, including increased heterogeneity within combined classes and obscured transitional patterns (e.g., EMCI→LMCI). Since the merged groups span multiple disease stages with differing degrees of pathology, atrophy, and connectivity disruption [52, 53], there is a potential for reducing class separability in both SC and FC features [54]. Nonetheless, binary grouping makes our cross-dataset application between different datasets available while mitigating severe class imbalance.

The Open Access Series of Imaging Studies (OASIS) dataset [55] encompasses data from 924 individuals, totaling 3,322 fMRI sessions. For our analysis, we select 402 fMRI and 362 DWI scans that have been processed with the Destrieux parcellation [56]. We conduct a binary classification task by assigning subjects in preclinical or dementia-related stages to the "AD"

group, and those who remain cognitively unimpaired to the "CN" group.

All datasets are split using 5-fold cross-validation based on subject indices to prevent data leakage across training and validation sets. For all experiments, FC is computed using the Pearson correlation coefficient with a threshold of 0.5 by default, while SC is derived from normalized white-matter fiber counts, thresholded at 0.1. In constructing brain networks, node attributes can be defined using either the raw Blood-Oxygen-Level-Dependent (BOLD) signals (in $\mathbb{R}^{N\times T}$) or the pairwise correlation (CORR) between BOLD signals (in $\mathbb{R}^{N\times N}$). Prior work has shown that the effectiveness of these choices varies across datasets [4]. Additionally, the temporal resolution of the BOLD signal, whether full-length (static network) or short-length segments ($T = 100$, dynamic network), also significantly influences deep model performance [4]. To ensure a comprehensive evaluation of *NeuroDetour* across diverse brain network representations, we test all four combinations of node features (BOLD vs. CORR) and temporal settings (static vs. dynamic) in our experiments.

4.2. Baseline Setting

For comparison, we evaluate *NeuroDetour* against a range of competitive methods, including two baseline models (MLP and GCN), three state-of-the-art brain network models (BrainGNN [24], BNT [25], and BolT [27]), and two state-of-the-art generic graph transformers (Graphormer [34] and NAGphormer [33]). The MLP and GCN baselines each consist of a single vanilla layer followed by batch normalization and an activation function for feature extraction, and one graph convolution layer for prediction. All training and validation configurations, including random seeds, learning rates, and other hyperparameters, are kept identical across methods. Codes can be acquired via this GitHub repository[2].

4.3. Performance of Neural Activity Classification

Performance on datasets HCPA and UKB is illustrated in Figure 3 first row. Except for UKB BOLD dynamic, our *NeuroDetour* is ranked in the top two places under all data settings in terms of accuracy and the weighted F1 score. Our *NeuroDetour* achieves the highest neural activity classification accuracy across all UKB settings at 99.59%. In comparison, the best performance by other models is 99.31%, achieved by the GCN. In the HCP-A dataset, SOTA

[2] `https://github.com/Chrisa142857/neuro_detour`

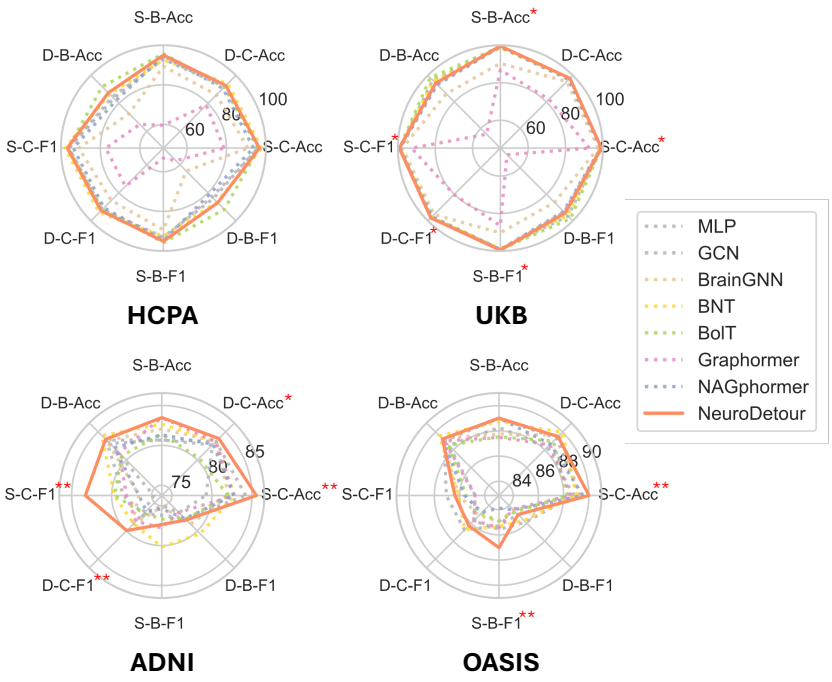


Figure 3: Performance comparison on four datasets. Experimental settings are abbrevated as S, D, C, and B denoting Static, Dynamic, CORR, and BOLD, respectively. $**$ and $*$ indicates $p \leq 10^{-3}$ and $10^{-3} < p \leq 0.05$, respectively.

brain models use more than twice the number of parameters compared to our model. Nonetheless, our *NeuroDetour* secures at least the second place across all data settings in HCP-A, largely due to the explicit formulation of physical neural pathways in model design. In contrast, generic graph transformers, lacking neuroscience knowledge, perform even worse than vanilla MLP and GCN models.

4.4. Performance on Disease Diagnosis

Fig. 3 second row shows AD vs CN classification in the ADNI and OASIS cohorts under four distinct data modalities (static BOLD, static FC, dynamic BOLD, and dynamic FC). Across these eight experiments, *NeuroDetour* consistently ranks among the top three methods, with the sole exception of the F1 score on dynamic OASIS BOLD data (D-B-F1 in Fig. 3). In particular, the incorporation of our multivariate SC-FC coupling mechanism enables *NeuroDetour* to attain the highest classification accuracy in five out

of eight cases and the highest F1 score in three out of eight cases. Notably, *NeuroDetour* is the only model to exceed 90% accuracy in any of these experiments. Although class imbalance within ADNI and OASIS leads to F1 scores that are lower than their corresponding accuracy values, *NeuroDetour* still achieves the best F1 performance across both datasets, with scores of 83.29% (ADNI) and 87.02% (OASIS) under all data settings. As expected, performance declines slightly on dynamic graphs, where the increased number of neural pathways yields a higher graph degree, yet *NeuroDetour* remains competitive with, or superior to, existing SOTA approaches. In summary, *NeuroDetour* demonstrates clear advantages over alternative methods when applied to static graph representations, a benefit that we attribute directly to its principled modeling of multivariate SC-FC coupling.

4.5. Performance on Cognitive Tests and ICD

To further evaluate, we demonstrate 12 more difficult tasks of UKB cognitive tests and ICD10 labels[3]. Note that the evaluation is measured by Mean Squared Error (MSE), Pearson correlation (r) and Accuracy (Acc), F1 score for continuous and discrete value, respectively. [Min, max] denotes the value range, and the number denotes the number of classes in the last column in the table below. ICD10 is the medical history of a subject and therefore is a multiclass classification. We simply convert ICD10 evaluation into binary classification on multiple selected codes and chapters with sufficient positive samples (>10%). Area Under the receiver operating characteristic Curve (AUC) is used to measure ICD10 with an 80%:20% train:test setting, and others are all 5-fold cross-validation. Table 4 shows results. Across a broad suite of cognitive, behavioral, and health-related prediction tasks, *NeuroDetour* consistently outperforms all competing baselines, often by very large margins. Traditional models (MLP, GCN, BrainGNN) and SOTA brain/graph models (BolT, BNT, Graphormer, NAGphormer) show highly unstable or near-chance performance across many tasks, with several models producing negative or near-zero feature importances or negligible predictive power. In contrast, *NeuroDetour* delivers substantially higher and far more stable scores across nearly every task, including FI, Trail, Puzzle, Tgame, Sleep, ICD-Ch, and ICD-top, where it significantly surpasses all baselines

[3]Fill the Field_ID of this url from Table 3: https://biobank.ndph.ox.ac.uk/ukb/field.cgi?id={Field_ID}, to access the detailed description of tests and labels

Table 3: Classification and regression tasks on cognitive tests and ICD labels in UKB.

Abbr.	Full name of test	Field_ID	fMRI-DWI pair #	Class #	Value range
Nmem	Numeric memory	4259	7101	-	[0, 2]
FI	Fluid intelligence	20016	7234	13	-
Trail	Trail making	6351	6933	10	
Puzzle	Matrix pattern completion	6373	6869	13	-
SymSub	Symbol digit substitution	23324	6865	-	[0, 35]
Vocab	Vocab level	26302	6357	-	[0.1084, 0.6161]
Tgame	Tower game	21004	6800	-	[0, 18]
Wmatch	Word matching	20197	6933	-	[0, 10]
Alcohol	Alcohol frequency	1558	7886	6	-
Sleep	Sleeplessness	1200	7886	3	-
ICD-Ch	ICD10 – Chapter level	41270	7927	22×binary	-
ICD-top	ICD10 – A00 to Z99 with $>$ 10% cases	41270	7927	4×binary	-

with minimal variance. Its dominance is especially pronounced in tasks with strong cognitive or behavioral relevance (e.g., FI, Trail, Puzzle), where it improves performance by 30 to 60 plus points relative to the best alternative. Overall, these results demonstrate that incorporating multi-hop SC detours provides a robust, interpretable, and consistently superior predictive signal across diverse domains compared to both conventional and SOTA models.

Fig. 4 below show ROC curves for Hypertension binary classification in ICD-top and Chapter VI - Diseases of the nervous system in ICD-Ch, respectively. Across both clinical prediction tasks, *NeuroDetour* demonstrates competitive or superior performance relative to existing graph- and transformer-based baselines, though the overall classification difficulty differs substantially between conditions. For Hypertension, all models show only modest discriminative ability, as reflected by AUC values hovering near chance, but *NeuroDetour* (AUC = 60.80) outperforms all baselines, indicating that its detour-based features provide performance boost. In contrast, prediction of past diseases of the nervous system proves even more challenging: all methods perform close to random (AUC ranges from 45 to 50), and the ROC curves of all models, including *NeuroDetour* (AUC = 50.19), cluster tightly around the diagonal. This pattern suggests that nervous-system disorders in UKB exhibit limited correspondence with resting-state connectivity patterns, whereas systemic traits like hypertension may retain modest but detectable neurofunctional signatures that *NeuroDetour* captures more effectively than existing architectures.

Table 4: Performance comparison on 12 classification/regression tasks on cognitive tests and ICD in UKB CORR. Each cell has two rows of scores corresponding to the Metric in Table 3, where $*$ indicates $p \leq 0.05$.

	Metric	MLP	GCN	BrainGNN	BolT	BNT	Graphormer	NAGphormer	*NeuroDetour*
Nmem	MSE↓	0.313±$_{0.210}$	0.039±$_{0.013}$	0.164±$_{0.251}$	0.030±$_{0.012}$	**0.011**±$_{0.002}$	0.018±$_{0.002}$	0.014±$_{0.003}$	0.012±$_{0.002}$
	r↑	0.013±$_{0.023}$	0.037±$_{0.013}$	-0.014±$_{0.025}$	0.020±$_{0.015}$	**0.039**±$_{0.025}$	0.004±$_{0.017}$	0.014±$_{0.011}$	-0.008±$_{0.016}$
FI	Acc↑	66.97±$_{3.14}$	20.60±$_{0.92}$	24.27±$_{12.01}$	20.55±$_{0.63}$	25.73±$_{2.79}$	19.54±$_{1.19}$	20.46±$_{0.97}$	**87.60**±$_{0.43}$*
	F1↑	67.26±$_{2.88}$	8.11±$_{0.62}$	18.45±$_{13.11}$	9.38±$_{1.29}$	16.56±$_{3.02}$	14.29±$_{1.22}$	7.07±$_{0.73}$	**87.60**±$_{0.42}$*
Trail	Acc↑	58.37±$_{3.39}$	57.46±$_{1.39}$	50.63±$_{14.83}$	57.47±$_{1.39}$	57.47±$_{1.39}$	46.53±$_{11.08}$	57.48±$_{1.39}$	**92.04**±$_{0.30}$*
	F1↑	63.83±$_{2.88}$	41.96±$_{1.66}$	39.57±$_{6.60}$	41.96±$_{1.66}$	41.96±$_{1.66}$	38.94±$_{4.63}$	41.98±$_{1.66}$	**91.83**±$_{0.32}$*
Puzzle	Acc↑	56.28±$_{23.65}$	20.47±$_{1.18}$	24.83±$_{5.36}$	19.89±$_{0.68}$	23.96±$_{1.03}$	18.16±$_{0.68}$	20.18±$_{1.14}$	**87.61**±$_{0.40}$*
	F1↑	56.18±$_{24.44}$	8.98±$_{1.30}$	17.48±$_{9.92}$	9.58±$_{1.85}$	15.58±$_{2.04}$	13.44±$_{1.33}$	7.15±$_{0.44}$	**87.61**±$_{0.40}$*
SymSub	MSE↓	380.482±$_{19.297}$	53.636±$_{10.801}$	91.259±$_{81.809}$	31.086±$_{2.423}$	28.781±$_{1.058}$	60.121±$_{36.967}$	68.521±$_{6.051}$	**28.626**±$_{0.755}$
	r↑	-0.005±$_{0.036}$	0.083±$_{0.014}$	0.012±$_{0.025}$	0.003±$_{0.018}$	**0.116**±$_{0.012}$	-0.007±$_{0.046}$	-0.015±$_{0.015}$	0.004±$_{0.020}$
Vocab	MSE↓	0.105±$_{0.075}$	0.013±$_{0.003}$	0.029±$_{0.009}$	0.013±$_{0.003}$	**0.007**±$_{0.000}$	0.024±$_{0.007}$	0.008±$_{0.001}$	0.008±$_{0.001}$
	r↑	0.011±$_{0.014}$	0.007±$_{0.022}$	-0.005±$_{0.026}$	-0.015±$_{0.030}$	**0.031**±$_{0.021}$	0.002±$_{0.010}$	0.017±$_{0.021}$	0.017±$_{0.018}$
Tgame	MSE↓	111.103±$_{4.398}$	18.206±$_{2.630}$	13.211±$_{2.547}$	11.357±$_{0.914}$	10.163±$_{0.460}$	23.417±$_{9.972}$	13.040±$_{1.059}$	**10.024**±$_{0.161}$*
	r↑	-0.006±$_{0.030}$	0.029±$_{0.024}$	0.004±$_{0.030}$	0.007±$_{0.020}$	**0.096**±$_{0.024}$	-0.014±$_{0.022}$	-0.004±$_{0.024}$	0.012±$_{0.016}$
Wmatch	MSE↓	54.281±$_{3.946}$	8.935±$_{1.258}$	8.301±$_{0.806}$	7.354±$_{0.187}$	6.890±$_{0.262}$	10.313±$_{5.235}$	8.482±$_{0.399}$	**6.731**±$_{0.075}$*
	r↑	-0.001±$_{0.013}$	0.049±$_{0.028}$	0.023±$_{0.032}$	0.007±$_{0.032}$	**0.093**±$_{0.018}$	0.005±$_{0.027}$	-0.002±$_{0.011}$	0.021±$_{0.021}$
Alcohol	Acc↑	66.17±$_{26.56}$	31.87±$_{0.54}$	34.67±$_{6.23}$	29.73±$_{1.40}$	36.24±$_{1.45}$	29.67±$_{0.71}$	30.20±$_{0.74}$	**88.96**±$_{0.52}$*
	F1↑	65.65±$_{27.81}$	23.16±$_{0.61}$	24.59±$_{8.93}$	21.42±$_{1.58}$	27.38±$_{0.97}$	22.77±$_{0.83}$	20.80±$_{1.34}$	**88.94**±$_{0.53}$*
Sleep	Acc↑	76.55±$_{22.50}$	45.71±$_{1.26}$	51.03±$_{8.55}$	45.60±$_{1.29}$	47.23±$_{2.09}$	42.65±$_{2.52}$	45.71±$_{1.26}$	**91.11**±$_{0.66}$*
	F1↑	75.25±$_{25.15}$	28.80±$_{1.41}$	39.00±$_{13.49}$	28.92±$_{1.26}$	34.59±$_{6.32}$	35.84±$_{2.66}$	28.69±$_{1.34}$	**91.10**±$_{0.67}$*
ICD-Ch	AUC↑	52.48±$_{8.47}$	48.75±$_{4.01}$	50.82±$_{2.62}$	47.51±$_{8.20}$	51.64±$_{5.62}$	49.63±$_{1.78}$	47.84±$_{7.60}$	**53.41**±$_{10.14}$*
ICD-top	AUC↑	56.17±$_{2.46}$	51.69±$_{2.69}$	53.00±$_{1.83}$	50.05±$_{2.31}$	55.63±$_{4.36}$	52.13±$_{1.66}$	54.81±$_{3.34}$	**57.63**±$_{5.63}$*

Table 5: Zero-shot learning between four datasets using BOLD as node attributes, where F1 scores are listed, resting/tasking classification is tested for HCPA and UKB, and '**bold**' and '<u>underline</u>' denote the first and the second rank, respectively.

	HCPA→UKB		UKB→HCPA		OASIS→ADNI		ADNI→OASIS	
	static	dynamic	static	dynamic	static	dynamic	static	dynamic
Graphormer	39.09$_{\pm 28.14}$	50.97$_{\pm 4.01}$	57.78$_{\pm 14.50}$	64.36$_{\pm 7.90}$	<u>77.63$_{\pm 2.89}$</u>	77.24$_{\pm 7.34}$	<u>79.69$_{\pm 7.71}$</u>	**83.55**$_{\pm 6.90}$
NAGphormer	<u>74.49$_{\pm 4.01}$</u>	<u>70.17$_{\pm 1.31}$</u>	<u>89.77$_{\pm 0.94}$</u>	<u>73.44$_{\pm 0.70}$</u>	73.11$_{\pm 5.90}$	<u>78.09$_{\pm 7.24}$</u>	69.63$_{\pm 10.99}$	78.09$_{\pm 7.08}$
NeuroDetour	**91.29**$_{\pm 2.10}$	**72.08**$_{\pm 2.15}$	**90.61**$_{\pm 3.65}$	**75.62**$_{\pm 2.98}$	**79.78**$_{\pm 3.53}$	**81.57**$_{\pm 7.24}$	**80.03**$_{\pm 8.50}$	<u>79.65$_{\pm 6.35}$</u>

4.6. Zero-shot Learning

Although neural activity classification and cognitive disorder diagnosis are common tasks in computational neuroscience, applying zero-shot learning, where a model is trained and validated on one dataset and then tested on another, remains rare. Zero-shot learning is evaluated for HCPA/UKB (resting-versus-task classification) and ADNI/OASIS (AD-versus-CN classification), demonstrating the clinical generalizability. Specifically, we adhere to the same protocol described above: use five-fold cross-validation on the train-validation split of one dataset to select the best model parameters, then evaluate performance on the corresponding fold of the other dataset.

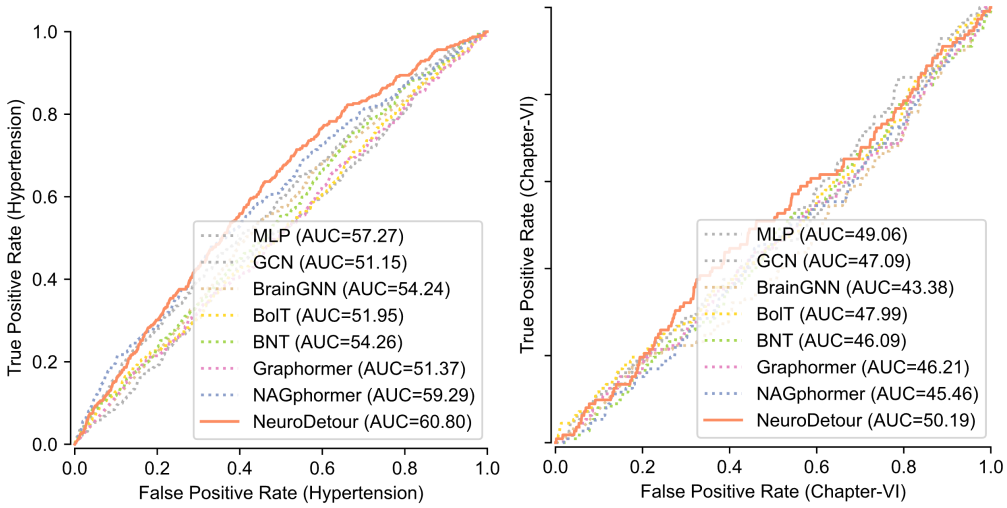


Figure 4: Receiver Operating Characteristic (ROC) curves comparing the True Positive Rate (Sensitivity) versus the False Positive Rate (1 − Specificity) across models, where Hypertension is ICD10 code I10, and Chapter-VI is ICD10 code chapter of diseases of the nervous system.

Given that the three SOTA brain models are all sensitive to the node number of the graph, they are not included in this zero-shot learning experiment. Therefore, F1 scores are listed in Table 5 in comparison against two generic graph transformers, which are designed for various vertex-cardinality. It is worth highlighting that our model outperforms all competitors under different data settings except for ADNI→OASIS dynamic. Specifically, *NeuroDetour* can surpass the best of others with a 16.8 score for HCPA→UKB static. Although the performance of zero-shot learning by *NeuroDetour* still has an observable gap to the fully supervised version shown in Fig. 3, results indicate that the neural pathway pattern learned by *NeuroDetour* is more consistent than the FC pattern across datasets.

Herein, we focus on the patterns of neural pathways that contribute to prediction in our *NeuroDetour*. As we introduced in Sec. 3.1, *NeuroDetour* is in fact weighting neural pathways to obtain the feature representation of the brain network. Therefore, the neural pathways can be sorted by their weights produced by our model. In this section, the top detour that has the top-1 weight (see Fact 3.1) per FC connection in ADNI and OASIS is analyzed with a significance score and visualized.

4.7. Detour Path Statistics

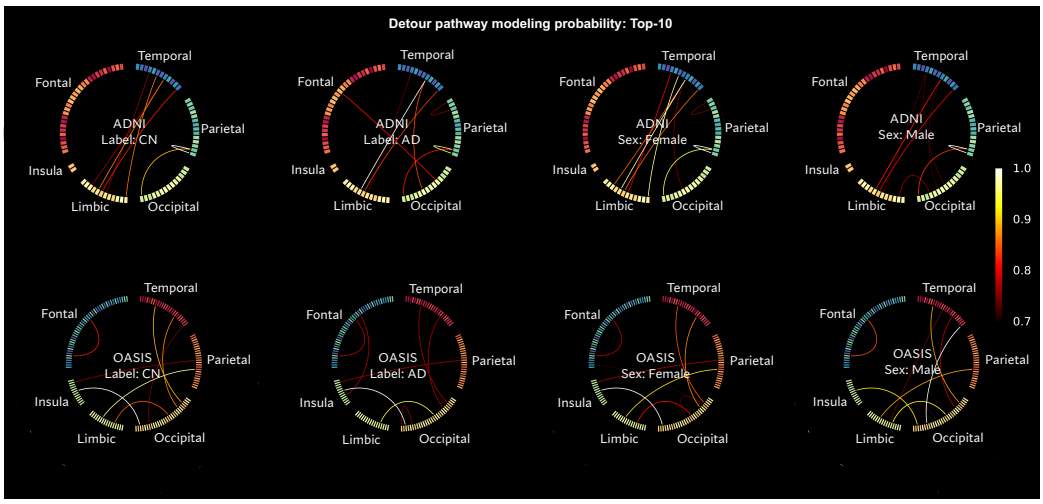


Figure 5: Circular connectograms connect two brain nodes, one of which is the endpoint of a detour path, and another has detours passing through, where link color indicates the normalized frequency and only the top-10 are displayed.

Per dataset, there are various groups of subjects depending on the label, sex, age, or phenotype. In Fig. 5, the differences in length of detour pathway are analyzed using the independent t-test on subjects with different labels or sex. Fig. 5 shows the histogram of the top detours used by *NeuroDetour* starting at a node and passing through another node. The color of the connection denotes the normalized frequency $\in [0, 1]$ of the transmission node, and only the top 10 connections are displayed. It is clear that the top detours in both datasets (160k detours in ADNI and 943k detours in OASIS) share similar transmission lobes between different labels and sex, even though the frequency rank is different. For example, *NeuroDetour* consistently learns from detours connecting the temporal, parietal, occipital, and limbic lobes to classify AD in ADNI. While in OASIS, detours in all cortical lobes except the frontal lobe are contributing for AD prediction.

In Fig. 6 (b) and (e), the relationship between age or Mini-Mental State Examination (MMSE) and the length of the detour is analyzed using linear regression for ADNI, where the average and standard deviation of the y-axis is displayed with the Pearson correlation R, t-test p-value, and n the number of detours starting at a lobe. It is clear that age is predominantly positively

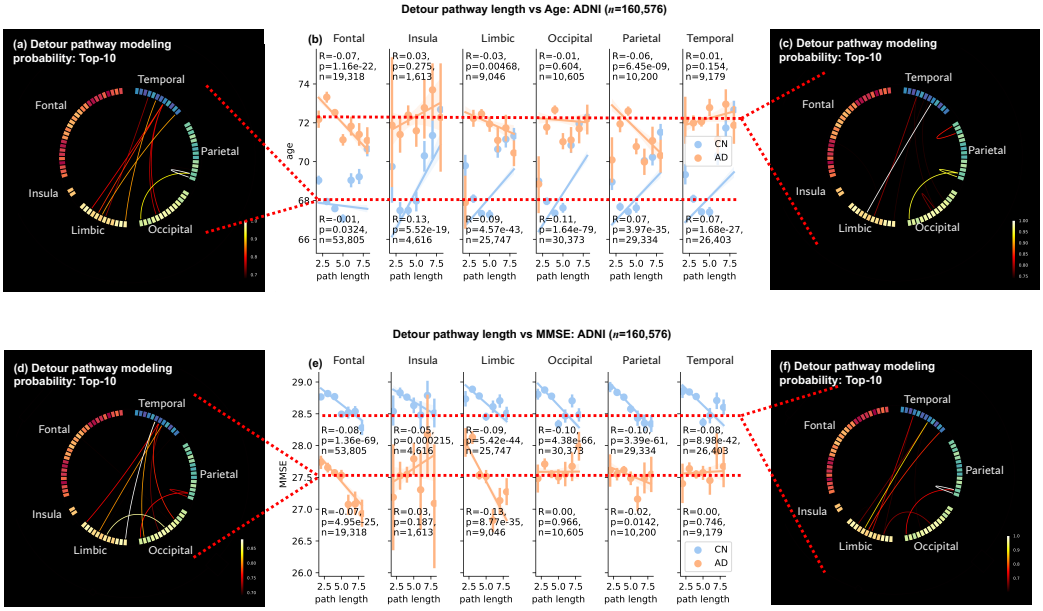


Figure 6: (a) Among CN subjects at age 65 to 70, the histogram of detour pathways that start at one node and pass through another node, where only top-10 pathways are displayed with the color indicating the probability. (b) Path length versus age within cortical lobes with the linear regression, where dots represent the mean of age, R the Pearson correlation, p the significance score, and n the number of detours. (c) Among AD subjects at age 70 to 75, the histogram of detour pathways. (d) Among AD subjects with MMSE 27 to 28, the histogram of detour pathways. (e) Path length versus MMSE. (f) Among CN subjects with MMSE 28 to 29, the histogram of detour pathways.

correlated with path length in CN subjects, while AD subjects consistently show a lower correlation $|R| \leq 0.07$ due to relatively limited detour amount. For MMSE, the correlation is the opposite since lower MMSE indicating a higher risk of AD. In (a), (c), (d), and (f), the histograms of detours within each CN and AD group are illustrated. Unlike the similar pattern shown in Fig. 5, two pairs, (a)-(c) and (d)-(f), are differentiated in terms of the highest connection (white curves) occurring within the parietal lobe and between the limbic and temporal lobes in the CN and AD groups, respectively.

Fig. 7 shows the same analysis and illustration in OASIS. Given that more brain regions are partitioned in OASIS atlas than ADNI, the number of detours for each lobe is greater ($n > 10^5$) in this analysis as shown in (b) and (e), resulting in greater significance (smaller p) and stronger correlation (larger $|R|$) in both groups for age versus length and AD group for MMSE

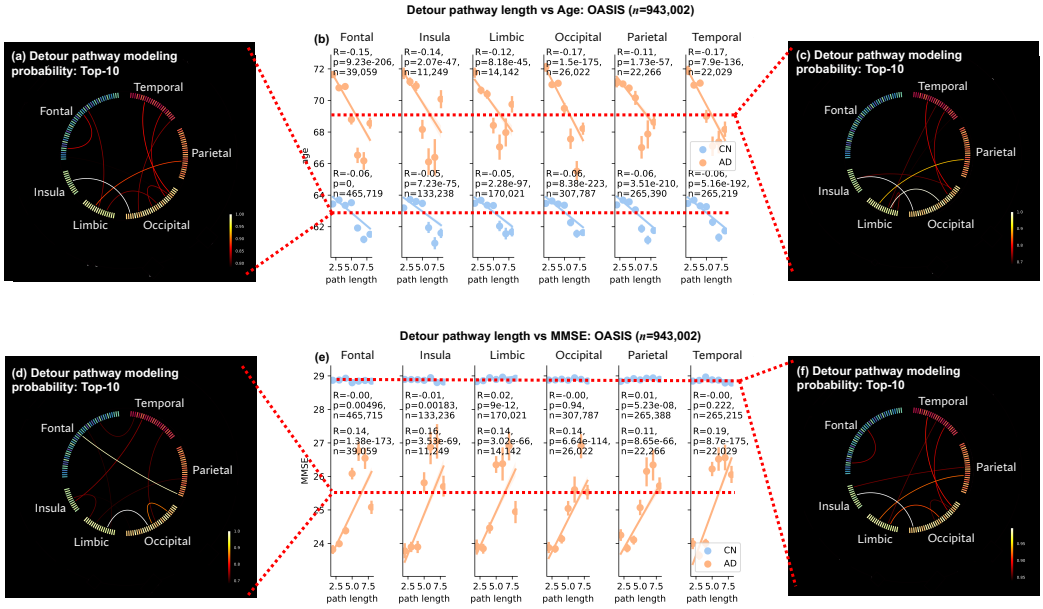


Figure 7: (a) Among CN subjects at age 60 to 65, the histogram of detour pathways that start at one node and pass through another node, where only the top-10 pathways are displayed with the color indicating the probability. (b) Path length versus age within cortical lobes with the linear regression, where dots represent the mean of age, R the Pearson's correlation, p the significance score, and n the number of detours. (c) Among AD subjects aged 65 to 70, the histogram of detour pathways. (d) Among AD subjects with MMSE 25 to 26, the histogram of detour pathways. (e) Path length versus MMSE. (f) Among CN subjects with MMSE 29 to 30, the histogram of detour pathways.

versus length. However, two pairs, (a)-(c) and (d)-(f), are less differentiated than ADNI since most connections pass through multiple lobes.

4.8. Visualization of Detour

We overlap AD and CN groups in Fig. 8 (a) to highlight the differences of detour, where we can observe long range detour from gyrus and sulcus of the mid-anterior cingulate cortex (IG_and_S_cingul.Mid.Ant) to middle occipital gyrus (IG_occipital_middle). In Fig. 8 (b), to show examples, we visualize three structural pathways corresponding to three functionally connected node pairs (highlighted in blue). Each pathway is depicted with uniformly colored links to denote a detour. These node pairs were selected on the basis of FC links that exhibited significant differences ($p \leq 0.05$) between AD and CN subjects in the OASIS dataset, as determined by a t-test. To

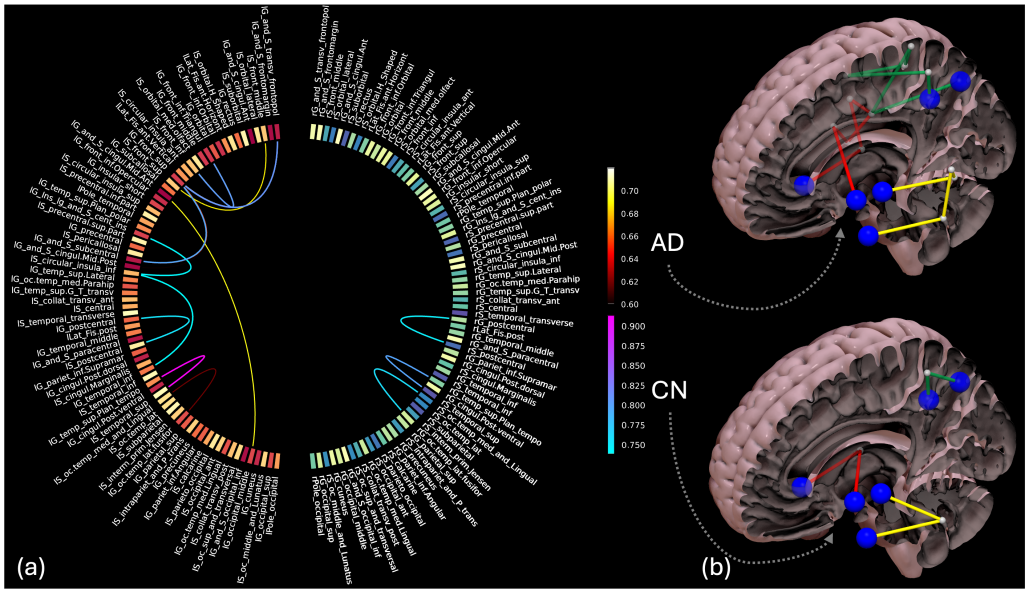


Figure 8: Examples of detour differences for AD and CN subjects in the OASIS dataset: (a) AD/CN comparison on the top frequent detours learned by *NeuroDetour*, where only the top 10 paths are displayed. (b) The visualization of top detours corresponding to three significant FC links.

mitigate the influence of inter-subject variability in FC (see Sec. 2), we restricted our analysis to FC links between regions in the subcortical structures, entorhinal cortex, occipital lobe, and parietal lobe—areas consistently associated with AD progression [57, 58].

Upon comparing the resulting detour visualizations, it becomes clear that AD subjects require longer SC detours to sustain the same direct FC connections, whereas CN subjects rely on shorter, typically two hops, pathways. This finding suggests that a diseased connectome may recruit additional SC links from unaffected regions in order to compensate for lesion-induced disruptions [59, 60]. This pattern is consistent with the notion of compensatory rerouting, e.g., neurodegenerative processes in AD disrupt direct SC link [61], forcing information flow to traverse longer alternative SC routes to preserve functional communication. Such increases in detour length align with prior reports of posterior-to-frontal recruitment in AD [62], where the brain leverages intact long-range fibers to compensate for localized structural loss.

While this interpretation is biologically plausible, we also acknowledge that several potential confounds may cause a false positive biomarker. Factors

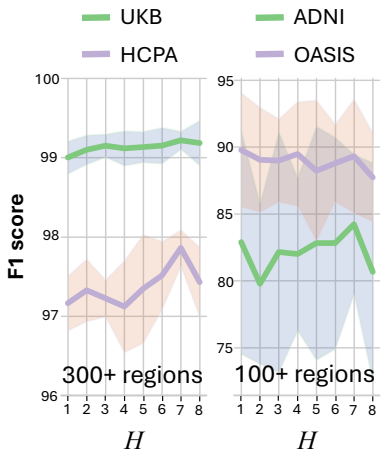


Figure 9: Ablation study of various lengths of the neural pathway that is visible to *NeuroDetour*. Static BOLD is set as node attributes in this experiment. The blue or orange shade is the range of error bars, and the lines are average F1 scores.

such as global degradation in SC [63], or diminished FC signal quality in AD [64] could also contribute to apparent increases in SC pathway length. These sources of variability may amplify or mimic compensatory changes in the SC-FC relationship. Nonetheless, the consistency of forming SC detours across multiple node pairs and two datasets suggests that the observed pattern reflects a systematic reorganization of anatomical rerouting rather than incidental measurement noise. Such evidence underscores the interpretability of *NeuroDetour* by demonstrating how alterations in SC can support observed FC under pathological conditions.

4.9. Ablation Studies

Maximum detour length. Because *NeuroDetour* constrains the maximum length of detour by the number of attention heads in the MHSA module, we performed an ablation study in which we varied the head count H from 1 to 8, keeping all other hyperparameters fixed. As shown in Fig. 9, the F1 score (green curves) on HCPA and UKB, both of which employ a fine-grained parcellation of 333 regions, reaches its maximum at $H = 7$. In contrast, for ADNI and OASIS, whose atlases contain fewer than 200 regions, the F1 score remains essentially constant across all values of H. This divergence can be explained by the fact that, in high-resolution parcellations (HCPA, UKB), reconstructing a given neural pathway often requires more hops, and

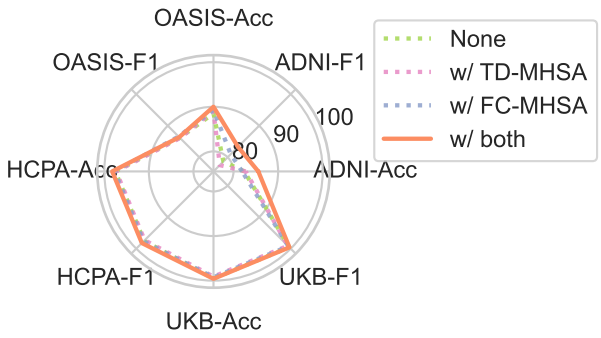


Figure 10: Ablation of none branch (a vanilla Transformer), single branch, and twin branch of *NeuroDetour* on four datasets with static BOLD as node attributes.

thus more attention heads, than in lower-resolution parcellations (ADNI, OASIS), where larger regions can span the same anatomical distances with fewer intermediate nodes.

Model architecture. The twin-branch architecture is central to *NeuroDetour*, enabling the integration of SC–FC coupling into the deep model. Ablating this design by removing one or both branches highlights its importance. As shown in Fig. 10, the full twin-branch configuration consistently achieves the best performance across all datasets, with the sole exception of the F1 score on OASIS. Although TD-MHSA leverages detour-based adjacency to capture brain community structure (Sec. 2), the fused SC–FC representations enforced by the twin-branch consistency loss $\mathcal{L}_{TD}$ contribute more substantially to the performance gains.

Loss weighting. The consistency loss $\mathcal{L}_{TD}$ is added to the final loss with $1\times$ weight by default. To show the sensitivity of *NeuroDetour* to this loss weight, we evaluated the model using $0.25\times$, $0.5\times$, $1\times$, and $2\times$ under data setting of static CORR. For nonclinical applications, the neural activity classification in the HCPA and UKB datasets (first two rows in Table 6) is not sensitive to loss weighting, given the minor fluctuation of accuracy scores. In contrast, AD prediction shows an incline trend of F1 score and a peak between $0.5\times$ and $1\times$.

Model depth. Deep models have a scalable model size according to their flexible depth by stacking their components. In Fig. 11 (a), we test *NeuroDetour* and competitors with various model layer numbers. It is obvious

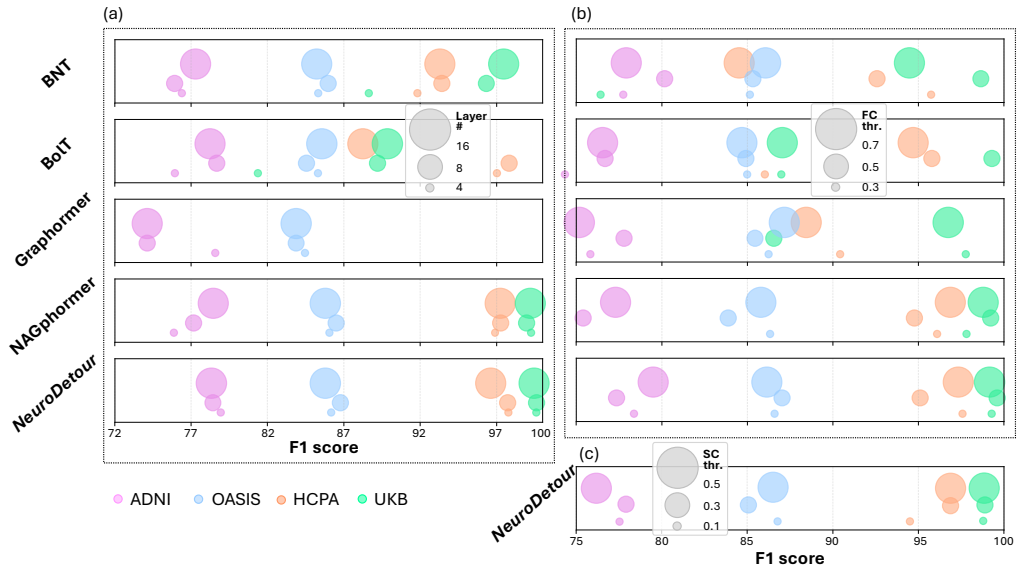


Figure 11: Bubble plots of ablation studies on (a) various model depths, (b) FC, and (c) SC thresholds (thr.) across all datasets using static BOLD as node attributes.

that *NeuroDetour* outperforms all competitors on four datasets regarding the oversmoothness problem. This is reasonable since the network architecture of *NeuroDetour* uses a graph-transformer architecture, making it far less susceptible to oversmoothing. In contrast, none of the existing methods can hold the performance rank after altering the layer number, e.g., NAGphormer mostly has the second rank except for the ADNI dataset.

FC threshold. FC is constructed as graph data by the threshold of the Pearson coefficient. The rationale behind choosing thresholds 0.3, 0.5, and 0.7 is that the graph degree is linearly correlated to the threshold (Pearson correlation $|R| > 0.9$). Therefore, evenly distributed thresholds are chosen to show the robustness rather than nearby thresholds, which lead to similar graph topology and neural pathways.

In Fig. 11 (b), different FC thresholds are used to evaluate the robustness of *NeuroDetour*. More fluctuations appeared when altering the FC threshold than the layer number. Graphormer and BNT show more than 20% drops in F1 scores on HCPA and UKB datasets, respectively, while *NeuroDetour* consistently achieves the top performance on HCPA and UKB, with only 0.5 and 0.7 FC thresholds in ADNI and OASIS where it is not ranked first.

Table 6: Performance of *NeuroDetour* using various weights of the consistency loss $\mathcal{L}_{TD}$ under data setting of static CORR.

	Metric	0.25×	0.5×	1×	2×		0.25×	0.5×	1×	2×
HCPA	Acc	$97.43_{\pm 0.73}$	$97.49_{\pm 0.33}$	$97.78_{\pm 0.17}$	$97.51_{\pm 0.33}$	ADNI	$83.48_{\pm 5.31}$	$82.00_{\pm 7.31}$	$83.48_{\pm 5.31}$	$82.07_{\pm 6.86}$
	F1	$97.44_{\pm 0.73}$	$97.49_{\pm 0.33}$	$97.78_{\pm 0.18}$	$97.51_{\pm 0.32}$		$79.63_{\pm 6.67}$	$78.33_{\pm 9.24}$	$80.69_{\pm 7.85}$	$77.54_{\pm 8.34}$
UKB	Acc	$99.07_{\pm 0.06}$	$99.08_{\pm 0.24}$	$99.17_{\pm 0.11}$	$99.05_{\pm 0.14}$	OASIS	$87.99_{\pm 5.17}$	$90.02_{\pm 3.80}$	$89.52_{\pm 3.66}$	$88.82_{\pm 4.86}$
	F1	$99.07_{\pm 0.06}$	$99.08_{\pm 0.24}$	$99.17_{\pm 0.11}$	$99.05_{\pm 0.14}$		$85.98_{\pm 5.16}$	$86.94_{\pm 5.76}$	$85.76_{\pm 5.57}$	$86.80_{\pm 4.13}$

Table 7: Performance comparison using different parcellation schemes, Gordon333 and AAL116, under data setting of static CORR, where the F1 scores are listed.

	BrainGNN		BolT		BNT		*NeuroDetour*	
	UKB	HCPA	UKB	HCPA	UKB	HCPA	UKB	HCPA
Gordon	$97.54_{\pm 0.52}$	$91.50_{\pm 3.00}$	$99.13_{\pm 0.33}$	$95.24_{\pm 2.50}$	$98.71_{\pm 0.35}$	$95.35_{\pm 1.04}$	$99.27_{\pm 0.17}$	$97.96_{\pm 0.30}$
AAL	$98.39_{\pm 0.38}$	$90.92_{\pm 1.41}$	$99.02_{\pm 0.25}$	$96.38_{\pm 0.43}$	$99.01_{\pm 0.22}$	$95.92_{\pm 0.66}$	$99.07_{\pm 0.26}$	$96.90_{\pm 0.41}$

SC threshold. SC detours are constructed based on the threshold of fiber amount (0.1 by default). To ensure that performance is not dependent on SC threshold as well, we vary the SC thresholds from 0.1 to more conservative values of 0.3 and 0.5, across all four datasets evaluating *NeuroDetour*. The results, as depicted in the Fig. 11 (c), show that performance is robust across the tested SC thresholds, where the F1 score varies slightly (less than 5% F1 score change) depending on the dataset and threshold, and does not exhibit a strong, inverse dependency on the SC threshold.

Brain parcellation. Two different parcellation schemes, Gordon333 and AAL116, are tested on UKB and HCPA datasets in Table 7. We can observe not only the best performance by our *NeuroDetour* on both datasets and parcellation schemes, but also the same trend. BrainGNN and BolT show different best parcellation schemes for UKB and HCPA. In contrast, BNT and *NeuroDetour* perform the best using AAL and Gordon, respectively, for different datasets.

In summary, our ablation studies show the rational settings for the hyper-parameter and the architecture of *NeuroDetour* along with the best robustness when altering layer number and FC threshold.

4.10. Computational Costs

Our experiments were conducted on a local computing platform equipped with two Intel(R) Xeon(R) Gold 6448Y CPUs and four NVIDIA RTX 6000

Table 8: Model computational costs.

	BrainGNN	BNT	BolT	Graphormer	NAGphormer	*NeuroDetour*
Param. number	7.30M	1.57M	1.58M	0.30M	0.26M	0.69M
Preprocess time (ms/graph)	-	-	-	270	40	-
Train time (ms/graph)	7.24	1.82	3.83	2.79	3.92	1.61
Test time (ms/graph)	2.61	0.64	1.83	0.90	1.85	0.67

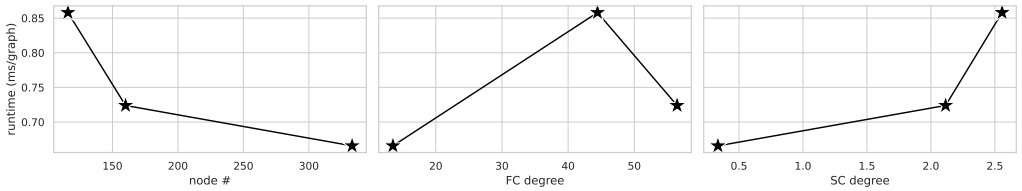


Figure 12: The correlation between testing runtime (ms/graph) and graph node number and degree.

Ada GPUs. All state-of-the-art models were run using their default hyperparameters, except for Graphormer, for which we used our own implementation. Detailed settings can be found in GitHub[4]. The computational costs can be indicated by the learnable parameter amount and computing time in OASIS dataset as listed in Table 8.

Since *NeuroDetour* learns from neural pathways, the runtime is also correlated with node number and graph degree (Fig. 12). Based on the definition of neural pathway in SC detours, a higher SC degree leads to more pathways. Although more nodes of the brain graph bring more input tokens to *NeuroDetour*, runtime is negatively correlated with node number (left panel in Fig. 12) due to the predominant impact from SC degree (right panel in Fig. 12). FC degree shows no correlation with the testing runtime.

5. Discussion

Recent studies indicate that structural and functional connectivity do not exhibit a uniform one-to-one correspondence across the human brain. Instead, it has revealed a heterogeneous relationship between structural and functional connectivity, with some brain regions exhibiting stronger coupling between structure and function and others exhibiting weaker coupling [16, 39].

[4] `https://github.com/Chrisa142857/neuro_detour`

Motivated by this heterogeneity, *NeuroDetour* models each functional interaction as potentially supported by multiple structural pathways, extending the SC-FC relationship to a multivariate, pathway-based formulation. Inspired by this, we first introduce a new connectivity metric, i.e., topological detour, using SC-FC coupling so that inter-subject variations can be weakened as we present in Section 2. A deep model, *NeuroDetour*, is then designed following the same idea, and thus it performs more accurately and robustly than existing methods for human connectome recognition. Results on zero-shot learning further demonstrate the potential of *NeuroDetour* to be a foundation model for network neuroscience.

Importantly, *NeuroDetour* can be situated within the broader context of computational neuroscience frameworks that focus on bridging structural and functional connectivity. Previously, biophysical models such as dynamic mean-field and neural mass models [65, 66] simulate the emergence of FC by modeling neural activity as the result of interactions through SC graphs, modulated by factors such as conduction delays, local dynamics, and stochastic noise. These models inherently capture indirect, multi-hop influences through dynamics and oscillatory coupling between SC and FC. While these traditional approaches provide rich mechanistic insight from simulation, they are computationally expensive and difficult to scale or generalize across datasets. In contrast, *NeuroDetour* offers a data-driven, scalable alternative that explicitly encodes multi-hop SC pathways as learnable feature representations. This distinction highlights the complementary nature of the two paradigms: Biophysical simulations yield mechanistic understanding through dynamic modeling, whereas our deep learning-based approach learns topological detours from multivariate SC-FC coupling that support neural activity prediction at scale. The structural detour thus serves as a biologically plausible yet computationally tractable proxy for complex multi-hop interactions modeled implicitly in the dynamic system of the human mind.

While *NeuroDetour* demonstrates strong empirical performance and introduces an interpretable framework for structure-function coupling via multi-hop neural pathways, several limitations warrant discussion.

First, by relying on a detour adjacency matrix rather than exhaustively searching all simple paths, our approach makes a deliberate trade-off between tractability and completeness. In practice, the number of possible neural pathways supporting a multi-hop FC link grows exponentially with network degree, particularly in dynamic connectomes, where transient connections further increase graph density. By pruning this search space to a subset

of candidate detours, *NeuroDetour* reduces computational complexity but inevitably avoids finding all simple paths containing weak connections that also contribute to functional coupling. Consequently, the feature representations learned by *NeuroDetour* may not fully disentangle overlapping or redundant pathways, limiting its theoretical capacity to distinguish among all potential SC-FC interplay. This constraint is especially pronounced in dynamic data, as evidenced by the lower performance on dynamic graphs (see Fig. 3), where the higher FC degree introduces an explosion of candidate detours.

Second, the use of a detour adjacency matrix introduces a deliberate simplification to address the NP-hard problem of searching all simple paths in the connectome. Although this design substantially improves computational efficiency and enables scalable learning across large datasets, it may obscure biologically relevant pathways that do not conform to the detour criteria, e.g., SC detours for multi-hop FC paths. Indirect or weakly connected SC links, which are known to significantly influence functional connectivity [16, 8], might be underrepresented or entirely omitted among this category.

To mitigate these limitations and advance the biological validity of structure-function modeling, we identify several potential strategies for future works:

1. Empirical validation of predicted pathways: t-test the top-ranked neural pathways identified by *NeuroDetour* with SC links in large-scale studies (e.g., UK Biobank) could assess their neurobiological plausibility.
2. Subject-specific detour criteria: Altering the detour criteria on a per-subject basis, such as SC detours for k-hop FC paths (k determined by subject variations) instead of the direct FC link, could cover more biologically relevant detours.

By addressing these directions, future iterations of *NeuroDetour* could enhance both the interpretability and neurobiological fidelity of learned representations, contributing to a more accurate and mechanistic understanding of structure-function coupling in the human brain.

6. Conclusion

In this work, we present *NeuroDetour*, a graph transformer framework designed to model physical neural pathways through a novel multivariate SC-FC coupling mechanism. Grounded in neuroscience insight, *NeuroDetour* leverages a twin-branch architecture to produce feature representations that

integrate multi-hop SC pathways and direct FC links, enabling a biologically motivated modeling of brain organization. Our extensive experiments across large-scale datasets, including HCP and UKB, demonstrate that *NeuroDetour* outperforms SOTA brain graph models and generic graph transformers on tasks such as neural activity classification and cognitive disorder diagnosis. In particular, the framework achieves robust performance under various data settings, including zero-shot learning in unseen subjects, resting/task state classification, and Alzheimer's diagnosis, highlighting both its generalizability and clinical relevance.

Beyond predictive accuracy, *NeuroDetour* provides interpretable insights through pathway-level statistics in ADNI and OASIS. By identifying top-ranked SC detours that contribute to AD predictions, *NeuroDetour* offers an interpretable bridge between pathway length, the histogram of brain regions where detour pathways passing through, cognition related phenotype (MMSE), sex, and age. These findings are consistent with neuroscientific hypotheses suggesting that longer, indirect pathways may support residual function in pathological connectomes, particularly in disorders like Alzheimer's disease.

Overall, *NeuroDetour* marks a step toward foundational models in computational neuroscience that not only perform well across downstream applications, but also begin to reveal the latent structure-function relationships that underlie brain dynamics and cognition.

Acknowledgement

This work was supported by the National Institutes of Health (AG091653, AG068399, AG084375) and the Foundation of Hope.

Appendix A. Accessibility

All data can be accessible via internet (HCPA[5], UKB[6], ADNI[7], OASIS[8]). The licenses to obtain that data can also be accessed on the websites. The preprocessing algorithms are FSL software [67] and a surface seed-based

[5] `https://www.humanconnectome.org/`
[6] `https://www.ukbiobank.ac.uk/`
[7] `https://adni.loni.usc.edu/`
[8] `https://sites.wustl.edu/oasisbrains/`

probabilistic fiber tractography [68]. The codes and data split settings can be acquired via this GitHub repository[9].

Appendix B. Data Preprocessing

The preprocessing of T1w, fMRI, and DWI data follows the standardized workflow summarized in this section. All processing was performed using publicly available, community-validated pipelines to ensure reproducibility across datasets.

Appendix B.1. T1-Weighted MRI Processing

We first preprocess the anatomical T1w images to generate tissue masks and cortical parcellations that serve as the anatomical reference for both fMRI and DWI.

- **Brain extraction and tissue segmentation.** Skull stripping and segmentation of gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) were conducted with FSL [67]. Example of maps are shown in green and purple boxes of Fig.1.
- **Cortical parcellation.** The cortical surface was then parcellated into regions according to the atlases used in each experiment (e.g., Destrieux160, AAL116, and Gordon333). These regions define the nodes for both structural and functional connectivity matrices.

Appendix B.2. Functional MRI Processing (fMRIPrep)

All fMRI preprocessing was performed using fMRIPrep [37], a widely adopted and fully standardized pipeline `https://fmriprep.org/en/stable/`. The key steps include: (1) Motion, distortion, and slice-time correction using the pipeline's integrated routines. (2) Susceptibility distortion correction via fieldmaps registration, depending on dataset availability. (3) Co-registration to T1w space using boundary-based registration. (4) Nuisance regression including motion parameters, WM/CSF signals, and high-pass filtering, following fMRIPrep's default confounds model. (5) Regional time-series extraction. Mean BOLD signals were extracted within each atlas parcel after preprocessing.

[9] `https://github.com/Chrisa142857/neuro_detour`

Functional connectivity (FC). Pearson correlation coefficients were computed between all pairs of regional time series. To reduce weak or noisy correlations, we applied a threshold of 0.5, keeping only connections with absolute correlation above this value.

Appendix B.3. Diffusion MRI Processing (QSIPrep)

All diffusion-weighted image preprocessing was performed using QSIPrep [38], an established BIDS-App for diffusion pipelines `https://qsiprep.readthedocs.io/en/latest/`. QSIPrep performs: (1) Eddy-current, susceptibility, and motion correction using robust joint modeling. (2) Gradient table correction and B0 alignment across DWI volumes. (3) Tensor fitting to obtain diffusion tensor images (DTI) [69].

Structural connectivity (SC). We used surface-seeded probabilistic tractography [68] on QSIPrep outputs to generate whole-brain fiber trajectories. For each pair of atlas regions:

$$\text{SC weight} = \frac{N_{\text{connecting streamlines}}}{N_{\text{total streamlines in the brain}}}. \tag{B.1}$$

To suppress noisy or biologically implausible fibers, we applied a threshold of 0.1 to the SC weight matrix.

Appendix B.4. Final Outputs

The resulting FC and SC matrices in $\mathbb{R}^{N_{\text{node}} \times N_{\text{node}}}$ constitute the two brain networks used in all downstream modeling analyses.

Appendix C. Proof of Fact

Proof. The first is to expand Eq. (1) by using the definition of detuor adjacent matrix.

$$\begin{aligned} f_{TD}(\mathbf{X}) &= \texttt{Concat}\left(\bar{\mathbf{X}}_1, \ldots, \bar{\mathbf{X}}_H\right) \bar{\mathbf{W}} \\ &= \sum_{h=1}^{H} \bar{\mathbf{X}}_h \bar{\mathbf{W}}_{i\sim j}, i := (h-1) * C, j := h * C, \end{aligned} \tag{C.1}$$

where $\bar{\mathbf{W}}_{i\sim j}$ is a slice of weights $\bar{\mathbf{W}}$ with C entries, and we have

$$\bar{\mathbf{X}}_h = \texttt{Softmax}\left((\mathbf{X}\bar{\boldsymbol{\alpha}}_h)(\mathbf{X}\bar{\boldsymbol{\beta}}_h)^T + f_{mask}(\mathbf{D}^h)\right)(\mathbf{X}\bar{\boldsymbol{\gamma}}_h). \tag{C.2}$$

In plain words, `Softmax` in the above equation is the standard softmax operation that only involves h-hop connected columns and applies on every row, then it produces a self-attention matrix $\mathbf{S} \in \mathbb{R}^{N \times N}$, where

$$\mathbf{S}_{ij} = \begin{cases} \texttt{Softmax}\left(x_i\right)_j & \text{if } j \in \text{argtrue}\mathbf{D}^h_{i,\cdot}, \\ 0 & \text{otherwise.} \end{cases} \tag{C.3}$$

where x denotes the term inside `Softmax` in Eq. C.2. Then

$$f_{TD}(\mathbf{X})_i = \sum_{h=1}^{H} \sum_{j \in \text{argtrue}\mathbf{D}^h_{i,\cdot}} \mathbf{S}_{ij} \mathbf{X}_j \bar{\boldsymbol{\gamma}}_h \bar{\mathbf{W}}_{i \sim j}. \tag{C.4}$$

Herein, $f_{TD}(\mathbf{X})_i$ weights all nodes of H-long neural pathways by sorting results $\mathbf{P}^H_i$. By assigning each node with the learnable weights $\mathbf{S}_{ij}\bar{\boldsymbol{\gamma}}_h \bar{\mathbf{W}}_{i \sim j}$ since $\mathbf{S}_{ij}$ is a scalar that can be moved aside to matrices, each path weights is thus represented by this weight consists of learnable parameters and can be extracted to explain the contribution by neural pathways to any downstream application. This can finish the proof. □

Appendix D. Abbreviations

Please refer to Table D.9 for the full name of brain sub-network abbreviations.

References

[1] D. S. Bassett, O. Sporns, Network neuroscience, Nature neuroscience 20 (3) (2017) 353–364. doi:10.2307/jj.8306312.13.

[2] S. L. Bressler, V. Menon, Large-scale brain networks in cognition: emerging methods and principles, Trends in cognitive sciences 14 (6) (2010) 277–290. doi:10.1016/j.tics.2010.04.004.

[3] Q. K. Telesford, S. L. Simpson, J. H. Burdette, S. Hayasaka, P. J. Laurienti, The brain as a complex system: using network science as a tool for understanding the brain, Brain connectivity 1 (4) (2011) 295–308. doi:10.1089/brain.2011.0055.

Table D.9: The full name of sub-network abbreviations in Table 2.

Full Name	Abbreviation
Cingulo Parietal Network	CPN
Ventral Attention Network	VAN
Default Mode Network	DMN
Sensorimotor Mouth	SMM
Retrosplenial Temporal	RST
Sensorimotor Hand	SMJ
Cingulo-Opercular Network	CON
Fronto-Parietal Network	FPN
Salience Network	SN
Visual Network	VN
Dorsal Attention Network	DAN
Auditory Network	AN

[4] A. Said, R. Bayrak, T. Derr, M. Shabbir, D. Moyer, C. Chang, X. Koutsoukos, Neurograph: Benchmarks for graph machine learning in brain connectomics, Advances in Neural Information Processing Systems 36 (2023) 6509–6531.

[5] R. F. Betzel, L. Byrge, Y. He, J. Goñi, X.-N. Zuo, O. Sporns, Changes in structural and functional connectivity among resting-state networks across the human lifespan, NeuroImage 102 (2014) 345–357. doi:10.1016/j.neuroimage.2014.07.067.

[6] M. G. Preti, D. Van De Ville, Decoupling of brain function from structure reveals regional behavioral specialization in humans, Nature communications 10 (1) (2019) 4747. doi:10.1038/s41467-019-12765-7.

[7] H. Liu, T. Dan, Z. Huang, D. Yang, W. H. Kim, M. Kim, P. Laurienti, G. Wu, Holobrain: A harmonic holography for self-organized brain function, in: International Conference on Information Processing in Medical Imaging, Springer, 2023, pp. 29–40. doi:10.1007/978-3-031-34048-2_3.

[8] J. Goñi, M. P. Van Den Heuvel, A. Avena-Koenigsberger, N. Velez de Mendizabal, R. F. Betzel, A. Griffa, P. Hagmann, B. Corominas-Murtra, J.-P. Thiran, O. Sporns, Resting-brain functional connectivity predicted

by analytic measures of network communication, Proceedings of the National Academy of Sciences 111 (2) (2014) 833–838.

[9] B. Mišić, R. F. Betzel, M. A. De Reus, M. P. Van Den Heuvel, M. G. Berman, A. R. McIntosh, O. Sporns, Network-level structure-function relationships in human neocortex, Cerebral Cortex 26 (7) (2016) 3285–3296.

[10] V. D. Calhoun, M. F. Amin, D. Hjelm, E. Damaraju, S. M. Plis, A deep-learning approach to translate between brain structure and functional connectivity, in: 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, 2017, pp. 6155–6159. doi:10.1109/ICASSP.2017.7953339.

[11] T. Sarwar, Y. Tian, B. T. Yeo, K. Ramamohanarao, A. Zalesky, Structure-function coupling in the human connectome: A machine learning approach, NeuroImage 226 (2021) 117609. doi:10.1016/j.neuroimage.2020.117609.

[12] J. Neudorf, S. Kress, R. Borowsky, Structure can predict function in the human brain: a graph neural network deep learning model of functional connectivity and centrality based on structural connectivity, Brain Structure and Function (2022) 1–13doi:10.1007/s00429-021-02403-8.

[13] T. Dan, J. Ding, G. Wu, Explore brain-inspired machine intelligence for connecting dots on graphs through holographic blueprint of oscillatory synchronization, Nature Communications 16 (1) (2025) 9425.

[14] T. Dan, Z. Wei, W. H. Kim, G. Wu, Exploring the enigma of neural dynamics through a scattering-transform mixer landscape for riemannian manifold, in: Forty-first International Conference on Machine Learning, 2024.
URL https://openreview.net/forum?id=EYOo48YGhy

[15] J. S. Damoiseaux, M. D. Greicius, Greater than the sum of its parts: a review of studies combining structural connectivity and resting-state functional connectivity, Brain structure and function 213 (2009) 525–533. doi:10.1007/s00429-009-0208-6.

[16] C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J.-P. Thiran, R. Meuli, P. Hagmann, Predicting human resting-state functional connectivity from structural connectivity, Proceedings of the National Academy of Sciences 106 (6) (2009) 2035–2040. doi:10.1073/pnas.0811168106.

[17] B. Mišić, R. F. Betzel, A. Nematzadeh, J. Goni, A. Griffa, P. Hagmann, A. Flammini, Y.-Y. Ahn, O. Sporns, Cooperative and competitive spreading dynamics on the human connectome, Neuron 86 (6) (2015) 1518–1529.

[18] R. F. Betzel, J. D. Medaglia, D. S. Bassett, Diversity of meso-scale architecture in human and non-human connectomes, Nature communications 9 (1) (2018) 346. doi:10.1038/s41467-017-02681-z.

[19] M. Ding, C. Zhou, Q. Chen, H. Yang, J. Tang, Cognitive graph for multi-hop reading comprehension at scale, arXiv preprint arXiv:1905.05460 (2019). doi:10.18653/v1/P19-1259.

[20] Z. Yan, T. Ma, L. Gao, Z. Tang, C. Chen, Cycle representation learning for inductive relation prediction, in: International Conference on Machine Learning, PMLR, 2022, pp. 24895–24910.

[21] D. K. Duvenaud, D. Maclaurin, J. Iparraguirre, R. Bombarell, T. Hirzel, A. Aspuru-Guzik, R. P. Adams, Convolutional networks on graphs for learning molecular fingerprints, Advances in neural information processing systems 28 (2015).

[22] W. Jin, R. Barzilay, T. Jaakkola, Junction tree variational autoencoder for molecular graph generation, in: International conference on machine learning, PMLR, 2018, pp. 2323–2332. doi:10.1039/9781788016841-00228.

[23] C. Bodnar, F. Frasca, N. Otter, Y. Wang, P. Lio, G. F. Montufar, M. Bronstein, Weisfeiler and lehman go cellular: Cw networks, Advances in Neural Information Processing Systems 34 (2021) 2625–2640.

[24] X. Li, Y. Zhou, N. Dvornek, M. Zhang, S. Gao, J. Zhuang, D. Scheinost, L. H. Staib, P. Ventola, J. S. Duncan, Braingnn: Interpretable brain graph neural network for fmri analysis, Medical Image Analysis 74 (2021) 102233. doi:10.1016/j.media.2021.102233.

[25] X. Kan, W. Dai, H. Cui, Z. Zhang, Y. Guo, C. Yang, Brain network transformer, Advances in Neural Information Processing Systems 35 (2022) 25586–25599.

[26] A. Thomas, C. Ré, R. Poldrack, Self-supervised learning of brain dynamics from broad neuroimaging data, Advances in neural information processing systems 35 (2022) 21255–21269.

[27] H. A. Bedel, I. Sivgin, O. Dalmaz, S. U. Dar, T. Çukur, Bolt: Fused window transformers for fmri time series analysis, Medical Image Analysis 88 (2023) 102841. doi:10.1016/j.media.2023.102841.

[28] A. Bessadok, M. A. Mahjoub, I. Rekik, Graph neural networks in network neuroscience, IEEE Transactions on Pattern Analysis and Machine Intelligence 45 (5) (2022) 5833–5848. doi:10.1109/TPAMI.2022.3209686.

[29] H. Cui, W. Dai, Y. Zhu, X. Kan, A. A. C. Gu, J. Lukemire, L. Zhan, L. He, Y. Guo, C. Yang, Braingb: a benchmark for brain network analysis with graph neural networks, IEEE transactions on medical imaging 42 (2) (2022) 493–506. doi:10.1109/TMI.2022.3218745.

[30] G. Michel, G. Nikolentzos, J. F. Lutzeyer, M. Vazirgiannis, Path neural networks: Expressive and accurate graph neural networks, in: International Conference on Machine Learning, PMLR, 2023, pp. 24737–24755.

[31] C. Graziani, T. Drucks, F. Jogl, M. Bianchini, T. Gärtner, et al., The expressive power of path-based graph neural networks, in: Forty-first International Conference on Machine Learning, 2024.

[32] W. Zhu, T. Wen, G. Song, L. Wang, B. Zheng, On structural expressive power of graph transformers, in: Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2023, pp. 3628–3637. doi:10.1145/3580305.3599451.

[33] J. Chen, K. Gao, G. Li, K. He, Nagphormer: A tokenized graph transformer for node classification in large graphs, in: Proceedings of the International Conference on Learning Representations, 2023.

[34] C. Ying, T. Cai, S. Luo, S. Zheng, G. Ke, D. He, Y. Shen, T.-Y. Liu, Do transformers really perform badly for graph representation?, Advances in neural information processing systems 34 (2021) 28877–28888.

[35] P. Li, Y. Wang, H. Wang, J. Leskovec, Distance encoding: Design provably more powerful neural networks for graph representation learning, Advances in Neural Information Processing Systems 33 (2020) 4465–4478.

[36] J. You, R. Ying, J. Leskovec, Position-aware graph neural networks, in: International conference on machine learning, PMLR, 2019, pp. 7134–7143.

[37] O. Esteban, C. J. Markiewicz, R. W. Blair, C. A. Moodie, A. I. Isik, A. Erramuzpe, J. D. Kent, M. Goncalves, E. DuPre, M. Snyder, et al., fmriprep: a robust preprocessing pipeline for functional mri, Nature methods 16 (1) (2019) 111–116.

[38] M. Cieslak, P. A. Cook, X. He, F.-C. Yeh, T. Dhollander, A. Adebimpe, G. K. Aguirre, D. S. Bassett, R. F. Betzel, J. Bourque, et al., Qsiprep: an integrative platform for preprocessing and reconstructing diffusion mri data, Nature methods 18 (7) (2021) 775–778.

[39] Z.-Q. Liu, R. F. Betzel, B. Misic, Benchmarking functional connectivity by the structure and geometry of the human brain, Network Neuroscience 6 (4) (2022) 937–949. doi:10.1162/netn_a_00236.

[40] C. J. Honey, O. Sporns, L. Cammoun, X. Gigandet, J.-P. Thiran, R. Meuli, P. Hagmann, Networks of anatomical covariance, Neuroimage 37 (1) (2007) 162–171.

[41] M. D. Greicius, Resting-state functional connectivity in neuropsychiatric disorders, Current opinion in neurology 22 (4) (2009) 347–354. doi:10.1097/WCO.0b013e328306f2c5.

[42] Z. Wei, T. Dan, J. Ding, G. Wu, *neuropath*: A neural pathway transformer for joining the dots of human connectomes, Advances in Neural Information Processing Systems 37 (2024) 67826–67849.

[43] M. E. Raichle, The brain's default mode network, Annual review of neuroscience 38 (1) (2015) 433–447.

[44] R. J. Lepping, W. S. McKinney, G. C. Magnon, S. K. Keedy, Z. Wang, S. A. Coombes, D. E. Vaillancourt, J. A. Sweeney, M. W. Mosconi, Visuomotor brain network activation and functional connectivity among

individuals with autism spectrum disorder, Human brain mapping 43 (2) (2022) 844–859.

[45] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Advances in neural information processing systems 30 (2017).

[46] S. Y. Bookheimer, D. H. Salat, M. Terpstra, B. M. Ances, D. M. Barch, R. L. Buckner, G. C. Burgess, S. W. Curtiss, M. Diaz-Santos, J. S. Elam, et al., The lifespan human connectome project in aging: an overview, Neuroimage 185 (2019) 335–348. doi:10.1016/j.neuroimage.2018.10.009.

[47] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, M. Joliot, Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain, Neuroimage 15 (1) (2002) 273–289. doi:10.1006/nimg.2001.0978.

[48] E. M. Gordon, T. O. Laumann, B. Adeyemo, J. F. Huckins, W. M. Kelley, S. E. Petersen, Generation and evaluation of a cortical area parcellation from resting-state correlations, Cerebral cortex 26 (1) (2016) 288–303. doi:10.1093/cercor/bhu239.

[49] A. R. Hariri, A. Tessitore, V. S. Mattay, F. Fera, D. R. Weinberger, The amygdala response to emotional stimuli: a comparison of faces and scenes, Neuroimage 17 (1) (2002) 317–323. doi:10.1006/nimg.2002.1179.

[50] T. J. Littlejohns, J. Holliday, L. M. Gibson, S. Garratt, N. Oesingmann, F. Alfaro-Almagro, J. D. Bell, C. Boultwood, R. Collins, M. C. Conroy, et al., The uk biobank imaging enhancement of 100,000 participants: rationale, data collection, management and future directions, Nature communications 11 (1) (2020) 2624. doi:10.1038/s41467-020-15948-9.

[51] M. W. Weiner, D. P. Veitch, P. S. Aisen, L. A. Beckett, N. J. Cairns, J. Cedarbaum, M. C. Donohue, R. C. Green, D. Harvey, C. R. Jack Jr, et al., Impact of the alzheimer's disease neuroimaging initiative, 2004 to 2014, Alzheimer's & Dementia 11 (7) (2015) 865–884. doi:10.1016/j.jalz.2015.04.005.

[52] C. R. Jack Jr, D. S. Knopman, W. J. Jagust, R. C. Petersen, M. W. Weiner, P. S. Aisen, L. M. Shaw, J. Q. Trojanowski, et al., Tracking pathophysiological processes in alzheimer's disease: an updated hypothetical model of dynamic biomarkers, The Lancet Neurology 12 (2) (2013) 207–216.

[53] L. Pini, M. Pievani, M. Bocchetta, D. Altomare, P. Bosco, E. Cavedo, S. Galluzzi, M. Marizzoni, G. B. Frisoni, Brain atrophy in alzheimer's disease and aging, Ageing Research Reviews 30 (2016) 25–48.

[54] M. Daianu, A. Mezher, M. F. Mendez, N. Jahanshad, E. E. Jimenez, P. M. Thompson, Disrupted rich club network in behavioral variant frontotemporal dementia and early-onset a lzheimer's disease, Human brain mapping 37 (3) (2016) 868–883.

[55] P. J. LaMontagne, T. L. Benzinger, J. C. Morris, S. Keefe, R. Hornbeck, C. Xiong, E. Grant, J. Hassenstab, K. Moulder, A. G. Vlassenko, et al., Oasis-3: longitudinal neuroimaging, clinical, and cognitive dataset for normal aging and alzheimer disease, MedRxiv (2019) 2019–12doi:10.1016/j.jalz.2018.06.1439.

[56] C. Destrieux, B. Fischl, A. Dale, E. Halgren, Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature, Neuroimage 53 (1) (2010) 1–15. doi:10.1016/j.neuroimage.2010.06.010.

[57] G. L. Wenk, et al., Neuropathologic changes in alzheimer's disease, Journal of Clinical Psychiatry 64 (2003) 7–10.

[58] E. Canu, D. G. McLaren, M. E. Fitzgerald, B. B. Bendlin, G. Zoccatelli, F. Alessandrini, F. B. Pizzini, G. K. Ricciardi, A. Beltramello, S. C. Johnson, et al., Mapping the structural brain changes in alzheimer's disease: The independent contribution of two imaging modalities, Journal of Alzheimer's Disease 26 (s3) (2011) 263–274. doi:10.3233/JAD-2011-0040.

[59] P. Bonifazi, A. Erramuzpe, I. Diez, I. Gabilondo, M. P. Boisgontier, L. Pauwels, S. Stramaglia, S. P. Swinnen, J. M. Cortes, Structure–function multi-scale connectomics reveals a major role of the fronto-striato-thalamic circuit in brain aging, Human Brain Mapping 39 (12) (2018) 4663–4677. doi:10.1002/hbm.24312.

[60] C. J. Bajada, J. Schreiber, S. Caspers, Fiber length profiling: A novel approach to structural brain organization, Neuroimage 186 (2019) 164–173. doi:10.1016/j.neuroimage.2018.10.070.

[61] I. Aganj, A. Frau-Pascual, J. E. Iglesias, A. Yendiki, J. C. Augustinack, D. H. Salat, B. Fischl, Compensatory brain connection discovery in alzheimer's disease, in: 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), IEEE, 2020, pp. 283–287.

[62] C. L. Grady, A. R. McIntosh, S. Beig, M. L. Keightley, H. Burian, S. E. Black, Evidence from functional neuroimaging of a compensatory prefrontal network in alzheimer's disease, Journal of Neuroscience 23 (3) (2003) 986–993.

[63] D.-S. Wang, D. W. Dickson, J. S. Malter, β-amyloid degradation and alzheimer s disease, BioMed Research International 2006 (1) (2006) 058406.

[64] A. Badhwar, A. Tam, C. Dansereau, P. Orban, F. Hoffstaedter, P. Bellec, Resting-state network dysfunction in alzheimer's disease: a systematic review and meta-analysis, Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring 8 (2017) 73–85.

[65] G. Deco, V. Jirsa, A. R. McIntosh, O. Sporns, R. Kötter, Key role of coupling, delay, and noise in resting brain fluctuations, Proceedings of the National Academy of Sciences 106 (25) (2009) 10302–10307.

[66] G. Deco, A. Ponce-Alvarez, D. Mantini, G. L. Romani, P. Hagmann, M. Corbetta, Resting-state functional connectivity emerges from structurally and dynamically shaped slow linear fluctuations, Journal of Neuroscience 33 (27) (2013) 11239–11252.

[67] M. Jenkinson, C. F. Beckmann, T. E. Behrens, M. W. Woolrich, S. M. Smith, Fsl, Neuroimage 62 (2) (2012) 782–790. doi:10.1016/j.neuroimage.2011.09.015.

[68] P. Fillard, M. Descoteaux, A. Goh, S. Gouttard, B. Jeurissen, J. Malcolm, A. Ramirez-Manzanares, M. Reisert, K. Sakaie, F. Tensaouti, et al., Quantitative evaluation of 10 tractography algorithms on a realistic diffusion mr phantom, Neuroimage 56 (1) (2011) 220–234. doi:10.1016/j.neuroimage.2011.01.032.

[69] A. Tristán-Vega, S. Aja-Fernández, Dwi filtering using joint information for dti and hardi, Medical Image Analysis 14 (2) (2010) 205–218. doi:10.1016/j.media.2009.11.001.