BrainFIBRE: A Foundation Model via Information Decomposition for Brain Microstructure

Zijian Dong*, Yi Lin*, Fang Ji*, Jianxiong Zhou, Kwun Kei Ng, and Juan Helen Zhou†

National University of Singapore, Singapore
*Equal contribution †Corresponding author
{zijian_dong, helen.zhou}@nus.edu.sg, lin_yi@u.nus.edu

Abstract. Diffusion MRI is widely used to probe brain microstructure, with particular sensitivity to early cerebrovascular and neurodegenerative changes. Neurite Orientation Dispersion and Density Imaging (NODDI) decomposes the diffusion signal into three biophysically interpretable maps — neurite density index (NDI), orientation dispersion index (ODI), and free water fraction (FWF) — capturing neurite packing, fiber coherence, and extracellular fluid, respectively. These 3D maps provide a rich substrate for learning transferable microstructural representations for the human brain, yet effectively integrating them remains an open challenge: standard representation learning struggles to disentangle the unique information carried by each of the three brain microstructure maps (*i.e.,* NDI, ODI, and FWF) from their shared and synergistic interactions. Here, we present **BrainFIBRE** (**Brain F**oundation Model via **I**nformation Decomposition for **BR**ain Microstructur**E**), the first foundation model for brain microstructure, pretrained on three NODDI-derived microstructure maps from the UK Biobank dataset (55,592 participants). To achieve this, we propose Self-supervised Partial Information Decomposition (SPID), which extends PID-guided multimodal learning to the self-supervised regime for the first time. A novel Counterfactual Candidate Construction (CCC) paradigm perturbs inter-modality alignment through modality dropping and swapping, providing the contrastive signal for a Mixture-of-Experts (MoE) architecture to disentangle unique, synergistic, and redundant information without any downstream label. Evaluated on both Caucasian and Asian cohorts, our model achieves state-of-the-art performance across diverse downstream tasks predicting age, sex, cerebrovascular disease (CeVD) and neurodegenerative markers, and cognitive performance, while yielding neurobiologically interpretable representations that reveal task- and cohort-specific interaction patterns among microstructural compartments. BrainFIBRE establishes a versatile foundation for neuroimaging analysis at the microstructural level. Code is available at https://github.com/hzlab/BrainFIBRE

Keywords: Brain foundation model · Diffusion MRI · NODDI · Partial Information Decomposition · Mixture-of-Experts

1 Introduction

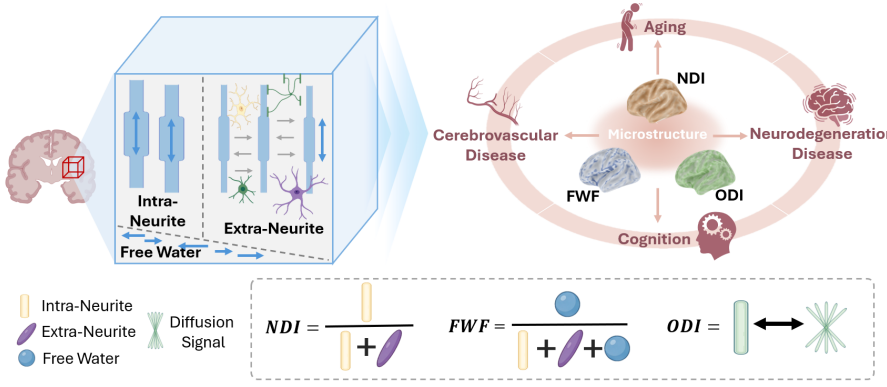

Fig. 1: Overview of NODDI-derived microstructural maps (the input to BrainFIBRE). NODDI is a biophysical model that partitions tissue microstructure into three compartments (intra-neurite, extra-neurite, and free water) to derive three corresponding parametric maps: Neurite Density Index (NDI), Orientation Dispersion Index (ODI), and Free Water Fraction (FWF). Compared to conventional diffusion maps, these biologically interpretable maps offer greater sensitivity to microstructural changes, providing critical insights into brain aging, cerebrovascular pathology, neurodegenerative processes, and interindividual variation in cognitive performance.

The human brain is built upon a highly complex microstructural architecture. Unraveling this structure is essential to understanding the biological mechanisms behind cognition, healthy aging, and neurological disease. Diffusion Magnetic Resonance Imaging (dMRI) has emerged as a cornerstone technology for this purpose, offering a unique, non-invasive window into tissue microstructure by measuring the restricted random motion of water molecules within biological tissues [19]. Despite its widespread clinical and research utility, conventional measures derived from diffusion fundamentally lack biological specificity. They summarize water diffusion across all encoding directions within a voxel, rather than isolating distinct microstructural compartments. For example, mean diffusivity (MD) quantifies the average rate of molecular diffusion, and fractional anisotropy (FA) reflects the degree of diffusion anisotropy, yet both are sensitive to multiple underlying tissue features without specifying their individual contributions. Consequently, they conflate distinct biophysical phenomena [17].

Neurite Orientation Dispersion and Density Imaging (NODDI) provides a more biologically specific characterization of tissue microstructure by modeling the voxel-wise diffusion signal into intra-neurite, extra-neurite, and free water compartments (Fig.1). This framework derives three parametric maps: neurite density index (NDI), orientation dispersion index (ODI), and free water fraction (FWF), which respectively characterize neurite packing, fiber coherence, and ex-

tracellular fluid. This microstructural modelling approach disentangles processes that are typically conflated in conventional diffusion metrics, such as neurite loss, demyelination, edema, and neuroinflammation, thereby yielding more directly interpretable tissue parameters [37].

Despite the biological specificity in NODDI, harnessing its rich, multiparametric maps for predictive modeling remains a profound computational challenge: how to effectively fuse these complex, three-dimensional spatial maps without re-entangling their distinct biophysical meanings. Addressing this challenge is crucial, as NODDI-derived maps capture clinically consequential signals that are inaccessible to macroscopic contrasts: they reveal systematic yet regionally heterogeneous microstructural trajectories across aging in all three compartments [10, 23, 26, 36], provide mechanistically grounded sensitivity to cerebrovascular injury (*e.g.*, white matter hyperintensities (WMH)) and cognitive decline [9, 13], and detect early Alzheimer's-related microstructural alterations [23, 36].

Recent brain foundation models have revolutionized neuroimaging analysis by leveraging self-supervised learning on large-scale, unlabeled datasets to extract transferable representations [3, 6, 7, 27]. However, they mainly focus on either macroscopic structural MRI (*e.g.*, T1- and T2-weighted scans) or functional MRI, leaving a model dedicated to tissue microstructure conspicuously absent. Furthermore, effectively integrating the NODDI maps into a unified learning framework is itself non-trivial. Each map encodes a distinct biophysical process, yet previous multi-modal fusion strategies in brain imaging [7,18,27] collapse different modalities into a monolithic latent space, risking the re-entanglement of the very processes that the NODDI compartmental model was explicitly designed to decouple. A principled framework is therefore needed to systematically characterize the interactions among neurite density, orientation dispersion, and free water fraction. Such a framework must disentangle whether these maps provide complementary, redundant, or synergistic information for different neurological tasks, while preserving the biophysical specificity afforded by NODDI.

Partial Information Decomposition (PID) offers a theoretical foundation for understanding multimodal interactions [1, 32]. It decomposes the total information that multiple sources carry about a target into unique, redundant, and synergistic components. Uniqueness isolates what each modality contributes alone, redundancy captures information shared across modalities, and synergy reveals information that emerges only from their joint observation. Recently, a Mixture-of-Experts (MoE) architecture operationalizing PID has been proposed, assigning dedicated experts to each interaction type and extending the framework beyond the two-modality setting [33]. It achieves interpretable predictions on Alzheimer's Disease classification with four input modalities including imaging, genetic, clinical, and biospecimen data. However, it approximates PID interaction types through binary label agreement among unimodal classifiers. This coupling strictly ties the decomposition to classifier quality and, crucially, precludes the large-scale, self-supervised pretraining essential for foundation models.

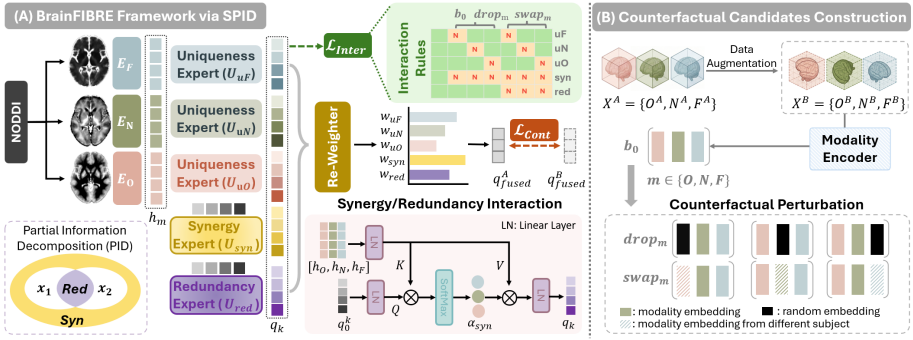

Fig. 2: Overview of BrainFIBRE. (**A**) By the design of Self-supervised Partial Information Decomposition (SPID), BrainFIBRE extends classical PID to isolate unique, synergistic, and redundant information for self-supervised pretraining. Three NODDI maps are processed by unimodal encoders into embeddings h_m, which are routed to five interaction experts: three for uniqueness and two for synergy and redundancy. A Re-Weighter adaptively aggregates expert embeddings q_k via learned weights w_k into fused representation q_{fused}. Pretraining is optimized by an interaction loss $\mathcal{L}_{Inter}$ guided by interaction rules, and a global contrastive loss $\mathcal{L}_{Cont}$. (**B**) Counterfactual Candidate Construction (CCC) generates perturbed triplets (base b_0, modality dropout ($drop_m$), and modality swap ($swap_m$)) to provide self-supervised signals for SPID.

In this work, we introduce **BrainFIBRE** (**Brain F**oundation Model via **I**nformation Decomposition for **BR**ain Microstructur**E**), the first foundation model dedicated to brain tissue microstructure (Fig.2). It was pretrained on NODDI-derived microstructure maps from 55,592 participants in the UK Biobank, the largest publicly available diffusion MRI dataset to date. BrainFIBRE treats the three NODDI maps — NDI, ODI, and FWF — as distinct modalities and learns unified multimodal representations that capture their joint predictive power while preserving the biophysical specificity of each individual compartment. To achieve this, we propose *Self-supervised Partial Information Decomposition (SPID)*, a novel pretraining objective that extends PID-guided multimodal learning to the self-supervised setting. SPID drives an MoE architecture comprising three Uniqueness experts, one Redundancy expert, and one Synergy expert, each trained to isolate its targeted information component without requiring downstream labels. Central to SPID is a *Counterfactual Candidate Construction (CCC)* paradigm that systematically generates counterfactual views through modality dropping and swapping, providing the contrastive signal needed to disentangle information in a purely self-supervised manner.

We evaluate BrainFIBRE on a comprehensive suite of downstream tasks spanning prediction of age, sex, cerebrovascular disease & neurodegenerative markers, and cognition across both Caucasian and Asian cohorts. It achieves state-of-the-art performance across tasks while yielding neurobiological interpretations that align with known microstructural signatures of aging and disease.

Our contributions can be summarized as follows:

- We present BrainFIBRE, the first brain foundation model for tissue microstructure, pretrained on NODDI-derived maps from UK Biobank.
- We propose Self-supervised Partial Information Decomposition (SPID) with a Counterfactual Candidate Construction (CCC) paradigm, enabling PID-guided self-supervised pretraining of MoE without any downstream label.
- We demonstrate state-of-the-art performance across diverse downstream tasks on both Caucasian and Asian cohorts, with neurobiological interpretations.

2 Related Work

2.1 Brain Foundation Model

Recent advances in brain foundation models have demonstrated the power of self-supervised learning to extract transferable representations from large-scale neuroimaging. These efforts have spanned diverse modalities. In the domain of electroencephalography (EEG), among the most representative works, LaBraM [16] pretrains a large brain model via masked EEG modeling, while NeuroLM [15] further integrates EEG signals into a large language model to enable unified multi-task inference through instruction tuning. In functional MRI (fMRI), BrainLM [3] and Brain-JEPA [6] pioneer representation learning for brain dynamics through masked prediction and joint-embedding architectures, respectively. In structural MRI (sMRI), BrainMVP [27] proposes self-supervised pretraining on 3D brain volumes, learning cross-parametric correspondences among multi-modal MRI. More recently, multimodal brain foundation models that jointly fuse structural and functional neuroimaging have also been proposed [7, 18]. Despite this progress, existing brain foundation models operate exclusively on macroscopic contrasts, and none address the tissue microstructures. A related line of diffusion MRI work studies learning-based microstructural parameter estimation, such as codebook-driven estimation of tissue characterization indices from diffusion measurements [31]. BrainFIBRE is complementary: rather than deriving microstructural maps from raw DWI, it learns transferable representations from already-derived NODDI maps and models the interactions among ODI, NDI, and FWF. Thus, interaction-aware representation learning over NODDI-derived tissue microstructure remains largely unexplored.

2.2 Multimodal Interaction MoEs

Existing multimodal MoEs typically rely on heuristic routing [11, 28], offering limited insight into how different modalities interact with each other [21, 29]. Modeling multimodal interactions provides a principled and natural inductive bias for guiding expert specialization in MoEs. In MMoE [35], dedicated experts are assigned to capture predetermined types of multimodal interactions. Building on this idea, I^2MoE introduces interaction-type categorization into the MoE training process and generalized the framework to support an arbitrary number of modalities [33]. However, its approximation of PID interaction types relies on

label agreement among unimodal classifiers, coupling the decomposition to classifier quality and, crucially, ruling out the large-scale self-supervised pretraining on which foundation models depend. General multimodal self-supervised frameworks such as M3-JEPA [20] use MoE-style latent alignment across heterogeneous modalities, but do not assign experts to explicit PID interaction atoms. In contrast, BrainFIBRE uses counterfactual modality dropping and swapping to specialize experts toward unique, redundant, and synergistic information in a label-free manner.

3 BrainFIBRE

We propose BrainFIBRE, a brain microstructure foundation model driven by the proposed Self-supervised Partial Information Decomposition (SPID) (Fig. 2.). For each participant, the input is a 3D microstructural triplet, $\mathbf{X} = \{X_m\}_m$, where $m \in \{\text{O, N, F}\}$ and each $X_m \in \mathbb{R}^{H\times W\times D}$ corresponds to a specific NODDI-derived map (O:ODI, N:NDI, F:FWF). To disentangle the profound relationships among modalities, we design a Mixture-of-Experts (MoE) architecture to explicitly model distinct types of modality interactions (Sec. 3.1) while extending PID to the self-supervised setting (Sec. 3.2). The pretraining objective is anchored by a Counterfactual Candidate Construction (CCC) strategy, which generates perturbations to provide optimization signals.

3.1 Architecture

Disentangled Encoding of Modality Interaction As shown in Fig. 2A, BrainFIBRE consists of modality encoders and interaction experts. First, three 3D Vision Transformers (ViT) [8] encoders, denoted as E_O, E_N, and E_F, are applied on each modality map in the input triplet $\mathbf{X}$ to extract unimodal embeddings $h_m \in \mathbb{R}^{1\times d_{enc}}$.

Guided by PID, we categorize interactions into three types: *uniqueness*, *synergy* , and *redundancy*, and define five distinct experts to extract specific interaction embeddings q_k: three uniqueness experts (U_{uO}, U_{uN}, U_{uF}) for modality-exclusive features, one synergy expert (U_{syn}) for emergent multimodal features, and one redundancy expert (U_{red}) for shared information. The entire disentangled encoding process is formulated as:

$$h_m = E_m(X_m) \in \mathbb{R}^{1\times d_{enc}}, \quad q_k = U_k(\text{InterRule}([h_O, h_N, h_F]) \in \mathbb{R}^{1\times d_{exp}} \quad (1)$$

where $k \in \mathcal{K} = \{\text{uO}, \text{uN}, \text{uF}, \text{syn}, \text{red}\}$, d_{enc} and d_{exp} are the dimension of unimodal and expert embeddings, and InterRule$(\cdot)$ denotes expert-specific perturbations based on interaction rules (detailed in Sec. 3.2).

Specifically, uniqueness experts capture the information unique to each modality via linear projection heads. On the other hand, the synergy and redundancy experts model more complex interactions via query-based information encoding (Fig.2.A). For either of these two, a learnable query token $q_0^k \in \mathbb{R}^{1\times d_{enc}}$ is initialized as the Query (Q). The three unimodal embeddings are concatenated and

linearly projected to form the Key ($K \in \mathbb{R}^{3\times d_{enc}}$) and Value ($V \in \mathbb{R}^{3\times d_{enc}}$). The expert embedding q_k is computed via scaled dot-product attention:

$$\alpha_k = \text{Softmax}\left(\frac{q_0^k K^\top}{\sqrt{d_{enc}}}\right) \in \mathbb{R}^{1\times 3}, \quad q_k = \text{Linear}(\alpha_k V) \in \mathbb{R}^{1\times d_{exp}} \tag{2}$$

where α_k is the attention weights and Linear$(\cdot)$ is a linear projection head. This design allows the synergy and redundancy experts to dynamically aggregate information across all modalities driven by their respective learned queries.

Adaptive Fusion of Expert Embeddings To integrate the expert embeddings, we introduce a Re-weighter module to dynamically assign importance to each expert. Implemented as a Multi-Layer Perceptron (MLP), the Re-weighter takes the concatenated unimodal embeddings to generate a probability distribution over the five interaction experts. This yields a set of expert weights w_k, which are used to produce the final fused representation q_{fused}:

$$\mathbf{w} = \text{Softmax}(\text{MLP}([h_O, h_N, h_F])) \in \mathbb{R}^{1\times 5}, \quad q_{fused} = \sum_{k\in\mathcal{K}} w_k \cdot q_k \tag{3}$$

3.2 Self-supervised Partial Information Decomposition (SPID)

Counterfactual Candidate Construction (CCC) While PID-informed MoE frameworks enable interpretable multimodal learning [33], they are typically trained in a supervised manner, limiting their generalizability for large-scale pre-training. To address this, we propose Self-supervised Partial Information Decomposition (SPID). The core challenge of extending PID to a self-supervised setting lies in isolating unique, redundant, and synergistic interactions without external supervision. Standard spatial augmentations that preserve the aligned modality triplet fail to provide the necessary teaching signal to disentangle these interactions, as the model is never forced to distinguish between modality-specific and the natural co-occurrence among modalities.

To address this, we introduce Counterfactual Candidate Construction (CCC) (Fig. 2B) to systematically disrupt biophysical alignment and generate the contrastive signals, enabling the model to isolate and learn each specific PID interaction in an entirely self-supervised manner. For a minibatch of N participants, let $\mathbf{X}_i$ denote the input for the i-th participant. We generate two augmented views, $\mathbf{X}^A(i)$ and $\mathbf{X}^B(i)$, each comprising a triplet of NODDI maps. These views are processed by unimodal encoders to yield their respective embeddings:

$$\mathbf{H}^v(i) = \{h_m^v(i)\}_m = \{E_m(X_m^v(i))\}_m, \quad m \in \{O, N, F\} \quad \text{and} \quad v \in \{A, B\} \tag{4}$$

We use $\mathbf{H}^A$ to obtain the anchor embedding via the five interaction experts. On the other hand, $\mathbf{H}^B$ is used to construct a set of counterfactural perturbations:

1. Base Candidate (b_0): The complete and unperturbed embedding derived from $\mathbf{X}^B(i)$:

$$b_0(i) = U_k([h_O^B(i), h_N^B(i), h_F^B(i)]) \tag{5}$$

2. Modality Dropout ($drop_m$): Simulates the absence of a specific modality. $drop_m(i)$ replaces the m-th component of $\mathbf{H}^v$ with Gaussian noise. For instance:

$$drop_O(i) = U_k([\text{noise}, h_N^B(i), h_F^B(i)]) \tag{6}$$

3. Modality Swap ($swap_m$): Constructs a counterfactual sample to destroy biophysical correspondances. $swap_m(i)$ replaces the m-th component of $\mathbf{H}^v$ with the same unimodal embedding from another randomly selected participant j in the same batch. For instance:

$$swap_O(i) = U_k([h_O^B(j), h_N^B(i), h_F^B(i)]) \tag{7}$$

Together, these perturbations form a candidate pool $\mathcal{C}_i = \{b_0(i)\} \cup \{drop_m(i)\}_m \cup \{swap_m(i)\}_m$ for each participant. **Crucially, identical spatial transformations are applied across the entire batch for each view, guaranteeing that swapped embeddings maintain strict anatomical alignment to prevent trivial solutions.** Since human brains share highly conserved macroscopic anatomy, swapping a modality with one from another participant yields a counterfactual that appears ***anatomically valid*** but is fundamentally ***microstructurally mismatched***. By forcing the model to distinguish these subtle disruptions, SPID moves beyond recognizing global brain shapes to genuinely learning voxel-level biophysical correspondences among neurite density, orientation dispersion, and free water.

Expert-Conditional Interaction Rules Based on the candidate pool $\mathcal{C}_i$ generated via CCC, we define specific interaction rules to construct positive samples or negative samples sets for each expert k. For instance, the uniqueness expert U_O should be invariant to changes in other modalities (N, F) but sensitive to its own (O); thus, $drop_N$,$drop_F$, $swap_N$, and $swap_F$ are considered as positives, whereas $drop_O$ and $swap_O$ are negatives. The redundancy expert U_{red} should remain robust to missing partial information, making all dropout candidates positive and all swap candidates negative. For the synergy expert U_{syn}, which relies on constructive multimodal interactions, any perturbation destroys the emergent information, leaving only the base candidate b_0 as positive. Complete definitions for all experts are detailed in Table 1.

Based on these rules, we formulate the interaction objective $\mathcal{L}_{inter}$ using a multi-positive symmetric InfoNCE loss [25]. Let $\mathcal{C}_{batch}$ denotes the perturbation candidates from all the participants in the current batch. For each expert U_k and participant i, the positive candidate pool $\mathcal{P}_k(i) \subset \mathcal{C}_{batch}$, while the remainder as negatives. The objective is to maximize the similarity between the anchor $q_k^A(i) = U_k([h_O^A(i), h_N^A(i), h_F^A(i)]))$ and its positive candidates in $\mathcal{P}_k(i)$, while pushing away negative candidates and all other participants in the batch:

$$\mathcal{L}_{Inter} = \sum_{k \in \mathcal{K}} \sum_{i=1}^{N} -\log \frac{\sum_{z^+ \in \mathcal{P}_k(i)} \exp(\text{sim}(q_k^A(i), z^+)/\tau)}{\sum_{z \in \mathcal{C}_{batch}} \exp(\text{sim}(q_k^A(i), z)/\tau)} \tag{8}$$

where τ is the temperature parameter. Although formulated with View A as the anchor for brevity, $\mathcal{L}_{Inter}$ is symmetrized by averaging the losses computed in both directions (A→B and B→A). By leveraging PID principles to specialize each expert via the interaction rules, this objective enables the end-to-end training of BrainFIBRE in a purely self-supervised, label-free manner.

Pretraining Objectives To learn robust participant-specific representations, we apply the symmetric InfoNCE loss [5], denoted as $\mathcal{L}_{\text{Cont}}$, to the fused embeddings q^A_{fused} and q^B_{fused} produced by the two augmented views. This objective maximizes the similarity between the anchor q^A_{fused} and its positive counterpart q^B_{fused}, while pushing away embeddings from other participants in the batch. Like $\mathcal{L}_{Inter}$, $\mathcal{L}_{Cont}$ is symmetrized by averaging across both directions.

Training MoE frameworks in a self-supervised manner is prone to representation collapse, where the model trivially assign all samples to a single expert. To mitigate this, we introduce two auxiliary regularization terms. The Balance Loss $\mathcal{L}_{balance}$ minimizes the Kullback-Leibler (KL) divergence between the batch-averaged expert assignment probabilities and a uniform distribution, preventing global expert collapse. The Entropy Loss $\mathcal{L}_{entropy}$ minimizes the negative entropy of individual weight distributions, encouraging soft probabilistic assignments such that each participant integrates information from multiple experts rather than relying on a deterministic, one-hot routing.

The final training objective is a weighted sum of the interaction loss, global contrastive loss, and the expert regularizations:

$$\mathcal{L}_{total} = \mathcal{L}_{Inter} + \mathcal{L}_{Cont} + \beta\mathcal{L}_{balance} + \gamma\mathcal{L}_{entropy} \tag{9}$$

where β and γ are hyperparameters for expert regularization.

Theoretical grounding The CCC-based interaction rules provide a theoretical basis for using SPID as a self-supervised surrogate of PID. Specifically, *Theorem 1* in the supplementary material establishes the correctness of the expert-conditional rules: for each expert, the positive candidates are exactly those perturbations that preserve its designated latent interaction factor, while the negative candidates destroy or replace that factor almost surely. Building on this rule-level correctness, *Theorem 2* shows that the population version of the interaction contrastive task admits a PID-factorized optimum, where each expert representation can solve its contrastive task using only its corresponding factor: the three modality-unique factors, the redundant factor, or the synergistic factor. Consequently, *Corollary 1* implies that, if this factorized solution is reached, the five expert representations instantiate a coarse five-atom PID decomposition of the latent participant-specific state. This result should be interpreted as a population-level grounding of our design rather than a claim of exact recovery of the full three-source PID lattice or finite-sample identifiability. Detailed assumptions and proofs are provided in Supplementary Sec. 1.

Table 1: Expert-Conditional Interaction Rules.

Expert	Anchor	Positive	Negative
U_{uO}	q_O^A	$\{b_0, drop_N, drop_F, swap_N, swap_F\}$	$\{drop_O, swap_O\}$
U_{uN}	q_N^A	$\{b_0, drop_O, drop_F, swap_O, swap_F\}$	$\{drop_N, swap_N\}$
U_{uF}	q_F^A	$\{b_0, drop_O, drop_N, swap_O, swap_N\}$	$\{drop_F, swap_F\}$
U_{syn}	q_{syn}^A	$\{b_0\}$	$\{drop_m, swap_m\}_m$
U_{red}	q_{red}^A	$\{b_0, drop_O, drop_N, drop_F\}$	$\{swap_m\}_m$

4 Experiments

4.1 Datasets

Pretraining Dataset The UK Biobank (UKB) is a population-based prospective study of over 500,000 participants (across the United Kingdom) with extensive phenotypic, clinical, and imaging data [30]. A subset (middle-aged to older adults) underwent brain MRI as part of an ongoing imaging initiative, providing substantial variability in aging and cardiometabolic risk factors [24]. Diffusion MRI was acquired via a harmonized multi-shell protocol on identical 3T Siemens Skyra scanners, from which NODDI metrics (ODI, NDI, and FWF) were systematically estimated. BrainFIBRE was pretrained on these NODDI-derived maps from 55,592 UKB participants.

Downstream Datasets To rigorously evaluate the representational power and clinical generalizability of BrainFIBRE, we benchmarked the model across diverse downstream datasets and tasks. We first evaluated the model on a held-out UKB test set (4,307 participants) across demographic (age, sex), neurodegenerative (hippocampal atrophy at 5-year follow-up), and cognitive (processing speed) predictions. Next, we assessed the model on the independent Human Connectome Project in Aging (HCP-Aging) dataset [2](630 participants) to evaluate demographic and executive function (Flanker, CardSort) prediction. Crucially, to demonstrate the utility of BrainFIBRE beyond predominantly Caucasian populations and into targeted clinical applications, we evaluated it on SINGER [34]. This Asian dataset (818 participants) comprises community-dwelling older adults at risk for vascular cognitive impairment, ranging from normal cognition to mild cognitive impairment. We utilized this dataset to predict brain age, cognition, and white matter hyperintensity (WMH) volume, a key cerebrovascular biomarker. Results were averaged over three independent runs, each using a distinct random seed for data split (train/val/test ratio of 6:2:2). Details of imaging preprocessing are provided in the supplementary materials.

4.2 Implementation Details

In BrainFIBRE, we employed 3D ViT-S [8] as the backbone for all unimodal encoders, with an input volume size of (96, 112, 96) and a patch size of (16, 16,

Table 2: Performance comparisons on UKB held-out dataset. Best results are bolded; underlined values denote statistical significance against all others ($p < 0.05$).

Model	Pretrain	Age		Sex		Hipp.Atrophy		Proc.Speed	
		MAE↓	Corr↑	F1↑	Acc↑	MAE ↓	Corr↑	MAE↓	Corr↑
ViT-O	✗	4.50 ±0.07	0.67 ±0.00	78.08 ±1.30	78.11 ±1.31	110.29 ±3.06	0.16 ±0.01	0.75 ±0.01	0.20 ±0.06
ViT-N	✗	4.49 ±0.09	0.67 ±0.01	82.01 ±1.29	82.13 ±1.31	110.01 ±2.57	0.17 ±0.03	0.75 ±0.01	0.20 ±0.04
ViT-F	✗	4.24 ±0.01	0.71 ±0.01	82.18 ±0.59	82.21 ±0.60	109.85 ±3.43	0.19 ±0.01	0.71 ±0.01	0.32 ±0.02
I^2MOE [33]	✗	6.21 ±0.15	0.42 ±0.05	71.94 ±1.32	72.58 ±1.39	111.90 ±2.23	0.17 ±0.02	0.73 ±0.01	0.33 ±0.02
BrainMVP [27]	✓	4.46 ±0.27	0.68 ±0.04	78.93 ±2.57	78.65 ±2.33	109.76 ±2.96	0.19 ±0.04	0.70 ±0.01	**0.36** ±0.01
TFS	✗	4.33 ±0.02	0.69 ±0.01	79.55 ±2.02	79.62 ±1.99	111.05 ±2.08	0.18 ±0.01	0.71 ±0.01	0.32 ±0.02
BrainFIBRE	✓	**3.95** ±0.02	**0.76** ±0.00	**89.30** ±0.82	**89.33** ±0.84	**109.21** ±2.03	**0.21** ±0.01	**0.70** ±0.01	0.33 ±0.02

TFS: trained-from-scratch. Hipp.Atropy: future hippocampal atrophy. Proc.Speed: processing speed.

16). d_{enc} and d_{exp} is 384 and 128, respectively. The pretraining was conducted on 8 NVIDIA H200 (140GB) GPUs. The model was pretrained for 200 epochs with a global batch size of 280, which required 40 hours. We utilized the AdamW optimizer [22] ($\beta_1 = 0.9, \beta_2 = 0.999$) with a base learning rate of $2e^{-4}$, a weight decay of $1e^{-4}$, and a cosine annealing learning rate scheduler.

Regarding the pretraining objectives, the temperature parameters τ for $\mathcal{L}_{Inter}$ and $\mathcal{L}_{Cont}$ were empirically set to 0.2 and 0.12, respectively. The expert regularization weights were set to $\beta = 0.05$ and $\gamma = 0.01$. To prevent early expert collapse, we adopted a warmup strategy during the first 5 epochs, where uniform routing weights were explicitly assigned to all five interaction experts.

For regression tasks, we adopted Mean Absolute Error (MAE) and Pearson Correlation (Corr) as the evaluation metrics. For classification tasks, we adopted the F1 score and Accuracy (Acc) metrics. More details regarding training configurations are provided in the supplementary materials.

4.3 Downstream Evaluation

Performance Comparison with Baselines on Internal Dataset BrainFIBRE demonstrates strong representation capability and generalizability across a diverse set of downstream tasks, encompassing demographic prediction, the prediction of neurodegenerative marker (hippocampus atrophy), and cognitive performance (processing speed). Table 2 summarizes the quantitative results on the held-out test set from UKB. To assess our approach, we benchmarked BrainFIBRE against a comprehensive set of baselines: (1) three unimodal 3D ViTs (ViT-O, ViT-N, and ViT-F) trained exclusively on a single NODDI map; (2) I^2MOE, a fully supervised multimodal model; (3) BrainMVP, a state-of-the-art multimodal self-supervised foundation model; and (4) a standard train-from-scratch paradigm using the BrainFIBRE architecture. Significance was evaluated via paired t-test, annotated only when BrainFIBRE significantly outperformed all baselines ($p < 0.05$). Overall, BrainFIBRE outperforms the unimodal, supervised & self-supervised multimodal baselines. The substantial gains underscore

Table 3: Performance comparisons on HCP-A dataset. Best results are bolded; underlined values denote statistical significance against all others ($p < 0.05$).

Model	Pretrain	Age		Sex		Flanker		CardSort	
		MAE↓	Corr↑	F1↑	Acc↑	MAE↓	Corr↑	MAE↓	Corr↑
ViT-O	✗	11.92 ±0.51	0.49 ±0.09	61.33 ±6.70	47.09 ±15.40	6.14 ±0.34	0.23 ±0.07	6.10 ±0.41	0.18 ±0.05
ViT-N	✗	11.09 ±0.69	0.64 ±0.03	57.33 ±2.10	36.43 ±0.86	6.13 ±0.34	0.18 ±0.07	6.08 ±0.36	0.20 ±0.07
ViT-F	✗	6.84 ±0.57	0.82 ±0.03	57.87 ±4.20	52.72 ±4.13	5.92 ±0.27	0.33 ±0.03	5.99 ±0.40	0.30 ±0.04
I^2MOE [33]	✗	10.37 ±0.42	0.82 ±0.01	67.47 ±4.90	66.15 ±3.80	5.86 ±0.26	0.30 ±0.06	5.95 ±0.4	0.26 ±0.06
BrainMVP [27]	✓	8.26 ±3.30	0.81 ±0.04	71.20 ±9.49	79.90 ±4.69	5.92 ±0.33	0.35 ±0.01	6.12 ±0.67	0.37 ±0.04
TFS	✗	6.57 ±0.35	0.84 ±0.02	74.40 ±3.39	73.86 ±3.02	5.99 ±0.23	0.31 ±0.04	6.08 ±0.47	0.34 ±0.02
BrainFIBRE	✓	**5.54** ±0.50	**0.87** ±0.02	**83.73** ±2.29	**83.36** ±2.09	**5.75** ±0.39	**0.40** ±0.05	**5.92** ±0.29	**0.38** ±0.02

TFS: trained-from-scratch.

Table 4: Performance comparisons on SINGER. Best results are bolded; underlined values denote statistical significance against all others ($p < 0.05$).

Model	Pretrain	Age		Mean Thickness		WMH		Proc.Speed	
		MAE↓	Corr↑	MAE↓	Corr↑	MAE↓	Corr↑	MAE↓	Corr↑
ViT-O	✗	3.84 ±0.06	0.37 ±0.03	0.08 ±0.00	0.26 ±0.05	2.47 ±0.33	0.21 ±0.02	0.67 ±0.08	0.09 ±0.03
ViT-N	✗	3.60 ±0.04	0.47 ±0.01	0.08 ±0.00	0.36 ±0.05	2.42 ±0.28	0.37 ±0.12	0.67 ±0.07	0.10 ±0.03
ViT-F	✗	3.37 ±0.11	0.55 ±0.03	0.07 ±0.00	0.47 ±0.03	2.37 ±0.31	0.34 ±0.04	0.67 ±0.08	0.21 ±0.01
I^2MOE [33]	✗	4.10 ±0.13	0.36 ±0.07	0.07 ±0.00	0.48 ±0.04	2.66 ±0.38	0.34 ±0.00	0.66 ±0.06	0.19 ±0.03
BrainMVP [27]	✓	3.44 ±0.40	0.50 ±0.03	0.07 ±0.02	0.52 ±0.03	2.54 ±0.56	0.44 ±0.13	0.70 ±0.09	0.15 ±0.03
TFS	✗	3.56 ±0.11	0.49 ±0.03	0.07 ±0.01	0.41 ±0.10	4.49 ±0.40	0.25 ±0.16	0.86 ±0.33	0.19 ±0.01
BrainFIBRE	✓	**3.21** ±0.12	**0.60** ±0.04	**0.07** ±0.00	**0.55** ±0.02	**2.08** ±0.30	**0.59** ±0.02	**0.66** ±0.07	**0.25** ±0.03

TFS: trained-from-scratch. WMH: white matter hyperintensity volume. Proc.Speed: processing speed.

BrainFIBRE's capacity to effectively extract and integrate subtle microstructural features from diffusion imaging data.

Superior Performance and Generalizability on External Dataset across Cohorts We further evaluated BrainFIBRE on the HCP-Aging dataset on demography prediction (age, sex) and executive function tasks (Flanker and CardSort scores). As shown in Table 3, BrainFIBRE consistently achieves superior predictive performance compared to baseline methods, highlighting its capacity to capture both fundamental demographic features and complex cognitive outcomes. This suggests BrainFIBRE's strong generalizability to unseen cohorts and the sensitivity of its learned representations to clinically relevant traits.

To further evaluate cross-ethnic generalizability and disease-relevant predictive utility, we assessed BrainFIBRE's performance on SINGER across age, mean cortical thickness, white matter hyperintensity (WMH) volume, and processing speed (Proc.Speed) prediction. As shown in Table 4, BrainFIBRE outperforms all baseline models on all four tasks. This robust advantage in an ethnically distinct cohort underscores the model's resilience to population heterogeneity. Importantly, it suggests that BrainFIBRE learns representations that capture

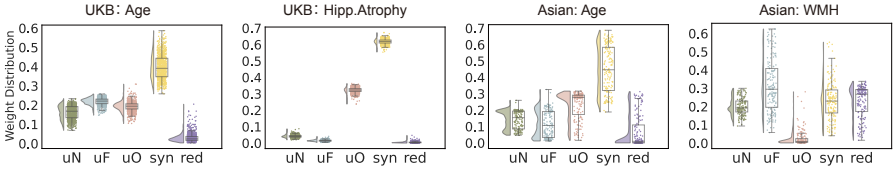

Fig. 3: Distribution of expert weights across the test set. The y-axis of each plot illustrates the weight distribution of each of the five experts for a specific task. These include three uniqueness experts (uN, uF, uO), one synergy expert (syn), and one redundancy expert (red).

fundamental, broadly transferable microstructural features, rather than leveraging cohort- or dataset-specific biases.

Task-Specific Modality Interaction Patterns To investigate whether and how the interaction patterns between the three brain microstructural measures vary across different datasets and tasks, we visualized the test-set distributions of the five expert weights assigned by the Re-Weighter across four tasks in Fig.3. In the UKB dataset, the synergy expert consistently exhibits the highest average weight, indicating a strong reliance on emergent information fused jointly from all three NODDI maps. Among the uniqueness experts, age prediction relies more heavily on FWF than NDI or ODI, suggesting free water fraction is a more prominent age-related microstructural marker. On the other hand, ODI dominates hippocampal atrophy prediction, suggesting that alterations in neurite dispersion are highly indicative of regional neurodegeneration.

The Asian dataset exhibits a more uniform expert weight distribution with greater participant variance. While age prediction remains synergy-dominated, it exhibits stronger ODI uniqueness contribution than in UKB, suggesting microstructural alterations in dispersion (potentially reflecting changes in fibre architecture and dendritic organization) provide additional aging-relevant signal in this cohort enriched for cardiometabolic risk. In contrast, the interaction pattern shifts markedly for WMH volume prediction: FWF uniqueness becomes the primary contributor, followed by redundancy and synergy, with minimal reliance on ODI. This pattern suggests that WMH burden is driven predominantly by extracellular water expansion, consistent with established neuropathological findings [13].

4.4 Ablation Study

To investigate the effectiveness of different interaction experts in BrainFIBRE, we conducted an ablation study comparing three settings: (1) *BrainFIBRE*: using the fully fused representation from all experts; (2)*Single-Expert*: evaluating each expert independently, and (3) *Dual-Modality*: omitting one NODDI map to rely on two modality encoders and four corresponding experts (two uniqueness,

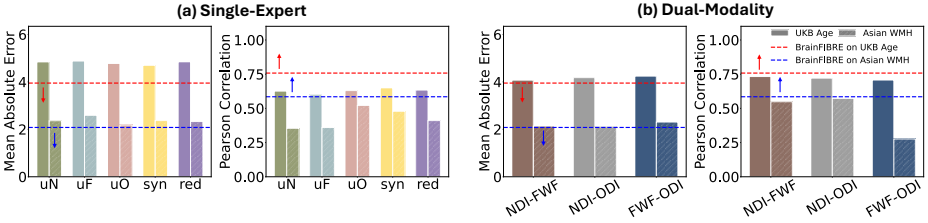

Fig. 4: Ablation study on age (UKB) and WMH volume (Asian) prediction tasks. We compared the full BrainFIBRE model (using the final fused embedding) against two settings: (a) Single-Expert (using individual expert embeddings), and (b) Dual-Modality (using fused embeddings derived from only two input NODDI maps).

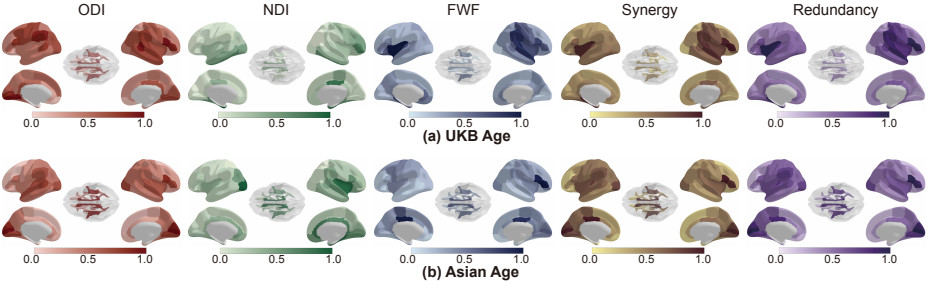

Fig. 5: Spatial attention patterns captured by different interaction experts. For both (a) UKB age and (b) Asian age prediction, salient regions are highlighted on both the cortical surface (peripheral views) and white matter tracts (central view).

one synergy, and one redundancy). As shown in Fig.4, the fused representation consistently outperforms any single expert. This indicates that our model benefits from integrating information from all modality interactions rather than relying on a single type of interaction alone. On the other hand, the dual-modality settings also underperform compared to BrainFIBRE, where dropping any of the modality would degrade the model performance.

Collectively, these results demonstrate that explicitly modeling and aggregating unique, redundant, and synergistic components leads to more robust and generalizable representations across heterogeneous clinical tasks. Additional ablations on individual loss terms are provided in the supplementary materials.

4.5 Interpretation

Salient Brain Regions Identified by Different Experts As shown in Fig.5, different experts captured distinct spatial patterns across NODDI compartments. For age prediction in UKB, in grey matter (GM), uniqueness experts revealed modality-specific patterns: NDI was localized to the medial temporal cortex, ODI to more posterior regions, and FWF spanned the broadest cortical extent, while synergy concentrated in medial temporal and insular areas. In white

matter (WM), age prediction was driven by limbic and association tracts, with FWF showing widespread sensitivity and NDI&ODI more selective involvement. Synergy and redundancy experts consistently highlighted the cingulum as a hub of multifactorial microstructural aging. These spatial patterns align with prior lifespan diffusion studies reporting medial temporal NDI decline [23], widespread free water increases [4], and preferential vulnerability of limbic association tracts such as cingulum during healthy aging [26].

For the Asian dataset, in GM, uniqueness experts showed a relative shift toward frontal-parietal, midline, and posterior cortical involvement compared with UKB, corresponding to executive control, sensorimotor, and default mode networks, while WM signatures remained largely consistent. This pattern is consistent with vascular-sensitive network vulnerability [12, 14], potentially reflecting enrichment of cardiometabolic risk factors in the cohort. We provide additional interpretative analyses for other tasks in the supplementary materials.

5 Conclusion

We introduce BrainFIBRE, the first foundation model for brain tissue microstructure, pretrained on NODDI-derived maps from UK Biobank. By treating neurite density, orientation dispersion, and free water fraction as distinct modalities, we propose Self-supervised Partial Information Decomposition (SPID) with Counterfactual Candidate Construction (CCC) to train an MoE architecture that explicitly isolates unique, redundant, and synergistic information — without supervision. Evaluations across demographically and ethnically diverse cohorts demonstrates state-of-the-art performance on predicting brain age, cerebrovascular and neurodegenerative markers, and cognitive performance, while the learned representations yield neurobiologically interpretable patterns consistent with established neuropathological findings. These results highlight the value of principled multimodal interaction modeling at the microstructural level and position BrainFIBRE as a versatile foundation for clinical neuroimaging applications.

Acknowledgments

This study was supported by the Singapore National Medical Research Council (NMRC/OFLCG19May-0035, NMRC/CIRG/1485/2018, NMRC/CSA-SI/0007/2016, NMRC/MOH-00707-01, NMRC/CG/435 M009/2017-NUH/NUHS, CIRG21nov-0007, HLCA23Feb-0004, OFIRG24Jul-0049, Human Potential Joint Grant HPJGC25-0002), A*STAR Singapore (AME Programmatic Fund A20G8b0102, Human Potential Joint Grant H25P3M00001, HHP Industry-Alignment Fund H24J4a0143), Ministry of Education (MOE-T2EP20223-0013), and Yong Loo Lin School of Medicine Research Core Funding, National University of Singapore, Singapore. We thank Christopher Chen, Nagaendran Kandiah, Kwong Hsia Yap, and their teams for participant recruitment and cognitive assessment in SINGER. We also thank Fiona Goh and other lab members for imaging data collection and quality control.

References

1. Bertschinger, N., Rauh, J., Olbrich, E., Jost, J., Ay, N.: Quantifying unique information. Entropy **16**(4), 2161–2183 (2014)
2. Bookheimer, S.Y., Salat, D.H., Terpstra, M., Ances, B.M., Barch, D.M., Buckner, R.L., Burgess, G.C., Curtiss, S.W., Diaz-Santos, M., Elam, J.S., et al.: The lifespan human connectome project in aging: an overview. Neuroimage **185**, 335–348 (2019)
3. Caro, J.O., de Oliveira Fonseca, A.H., Rizvi, S.A., Rosati, M., Averill, C., Cross, J.L., Mittal, P., Zappala, E., Dhodapkar, R.M., Abdallah, C., et al.: Brainlm: A foundation model for brain activity recordings. In: The Twelfth International Conference on Learning Representations
4. Chad, J.A., Pasternak, O., Salat, D.H., Chen, J.J.: Re-examining age-related differences in white matter microstructure with free-water corrected diffusion tensor imaging. Neurobiology of aging **71**, 161–170 (2018)
5. Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: A simple framework for contrastive learning of visual representations. In: International conference on machine learning. pp. 1597–1607. PmLR (2020)
6. Dong, Z., Li, R., Wu, Y., Nguyen, T.T., Chong, J., Ji, F., Tong, N., Chen, C., Zhou, J.H.: Brain-jepa: Brain dynamics foundation model with gradient positioning and spatiotemporal masking. Advances in Neural Information Processing Systems **37**, 86048–86073 (2024)
7. Dong, Z., Ruilin, L., Chong, J.S.X., Dehestani, N., Teng, Y., Lin, Y., Li, Z., Zhang, Y., Xie, Y., Ooi, L.Q.R., et al.: Brain harmony: A multimodal foundation model unifying morphology and function into 1d tokens. In: The Thirty-ninth Annual Conference on Neural Information Processing Systems
8. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
9. Duering, M., Finsterwalder, S., Baykara, E., Tuladhar, A.M., Gesierich, B., Konieczny, M.J., Malik, R., Franzmeier, N., Ewers, M., Jouvent, E., et al.: Free water determines diffusion alterations and clinical status in cerebral small vessel disease. Alzheimer's & Dementia **14**(6), 764–774 (2018)
10. Greenman, D., Bennett, I.J.: Aging of gray matter microstructure: A brain-wide characterization of age group differences using noddi. Neurobiology of Aging **149**, 34–43 (2025)
11. He, J., Qiu, J., Zeng, A., Yang, Z., Zhai, J., Tang, J.: Fastmoe: A fast mixture-of-expert training system. arXiv preprint arXiv:2103.13262 (2021)
12. Ji, F., Chai, Y.L., Liu, S., Kan, C.N., Ong, M., Richards, A.M., Tan, B.Y., Venketasubramanian, N., Pasternak, O., Chen, C., et al.: Associations of blood cardiovascular biomarkers with brain free water and its relationship to cognitive decline: a diffusion-mri study. Neurology **101**(2), e151–e163 (2023)
13. Ji, F., Pasternak, O., Liu, S., Loke, Y.M., Choo, B.L., Hilal, S., Xu, X., Ikram, M.K., Venketasubramanian, N., Chen, C.L.H., et al.: Distinct white matter microstructural abnormalities and extracellular water increases relate to cognitive impairment in alzheimer's disease with and without cerebrovascular disease. Alzheimer's Research & Therapy **9**(1), 63 (2017)
14. Ji, F., Wei, J.L.K., Leng, S., Zhong, L., Tan, R.S., Gao, F., Ng, K.K., Leong, R.L., Pasternak, O., Chee, M.W., et al.: Heart-brain mapping: Cardiac atrial function is associated with distinct cerebral regions with high free water in older adults. Journal of Cerebral Blood Flow & Metabolism **44**(7), 1218–1230 (2024)

15. Jiang, W., Wang, Y., Lu, B.l., Li, D.: Neurolm: A universal multi-task foundation model for bridging the gap between language and eeg signals. In: The Thirteenth International Conference on Learning Representations
16. Jiang, W., Zhao, L., Lu, B.l.: Large brain model for learning generic representations with tremendous eeg data in bci. In: The Twelfth International Conference on Learning Representations
17. Jones, D.K., Knösche, T.R., Turner, R.: White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion mri. Neuroimage **73**, 239–254 (2013)
18. Khajehnejad, M., Habibollahi, F., Razi, A.: Brainsymphony: A transformer-driven fusion of fmri time series and structural connectivity. arXiv preprint arXiv:2506.18314 (2025)
19. Le Bihan, D., Mangin, J.F., Poupon, C., Clark, C.A., Pappata, S., Molko, N., Chabriat, H.: Diffusion tensor imaging: concepts and applications. Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine **13**(4), 534–546 (2001)
20. Lei, H., Cheng, X., Qin, Q., Wang, D., Fan, K., Huang, H., Gu, Q., Wu, Y., Jiang, Z., Chen, Y., et al.: M3-jepa: Multimodal alignment via multi-gate moe based on the joint-embedding predictive architecture. arXiv preprint arXiv:2409.05929 (2024)
21. Liu, A., Feng, B., Xue, B., Wang, B., Wu, B., Lu, C., Zhao, C., Deng, C., Zhang, C., Ruan, C., et al.: Deepseek-v3 technical report. arXiv preprint arXiv:2412.19437 (2024)
22. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
23. Merluzzi, A.P., Dean III, D.C., Adluru, N., Suryawanshi, G.S., Okonkwo, O.C., Oh, J.M., Hermann, B.P., Sager, M.A., Asthana, S., Zhang, H., et al.: Age-dependent differences in brain tissue microstructure assessed with neurite orientation dispersion and density imaging. Neurobiology of aging **43**, 79–88 (2016)
24. Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L., et al.: Multimodal population brain imaging in the uk biobank prospective epidemiological study. Nature neuroscience **19**(11), 1523–1536 (2016)
25. Oord, A.v.d., Li, Y., Vinyals, O.: Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748 (2018)
26. Peters, B.D., Ikuta, T., DeRosse, P., John, M., Burdick, K.E., Gruner, P., Prendergast, D.M., Szeszko, P.R., Malhotra, A.K.: Age-related differences in white matter tract microstructure are associated with cognitive performance from childhood to adulthood. Biological psychiatry **75**(3), 248–256 (2014)
27. Rui, S., Chen, L., Tang, Z., Wang, L., Liu, M., Zhang, S., Wang, X.: Multi-modal vision pre-training for medical image analysis. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 5164–5174 (2025)
28. Shazeer, N., Mirhoseini, A., Maziarz, K., Davis, A., Le, Q., Hinton, G., Dean, J.: Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. In: International Conference on Learning Representations (2017)
29. Shen, L., Chen, G., Shao, R., Guan, W., Nie, L.: Mome: Mixture of multimodal experts for generalist multimodal large language models. Advances in neural information processing systems **37**, 42048–42070 (2024)
30. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al.: Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS medicine **12**(3), e1001779 (2015)

31. Wang, T., Wan, Z., Cao, S., Yu, J., He, Y., Xie, Y., Zhang, F., Wu, Y.: Deep learning empowered microstructure codebook: New paradigm for multi-parameter tissue characterization estimation. Human Brain Mapping **47**(5), e70513 (2026)
32. Williams, P.L., Beer, R.D.: Nonnegative decomposition of multivariate information. arXiv preprint arXiv:1004.2515 (2010)
33. Xin, J., Yun, S., Peng, J., Choi, I., Ballard, J.L., Chen, T., Long, Q.: I2moe: Interpretable multimodal interaction-aware mixture-of-experts. arXiv preprint arXiv:2505.19190 (2025)
34. Xu, X., Chew, K., Wong, Z., Phua, A., Chong, E., Teo, C., Sathe, N., Chooi, Y., Chia, W., Henry, C., et al.: The singapore geriatric intervention study to reduce cognitive decline and physical frailty (singer): study design and protocol. The Journal of Prevention of Alzheimer's Disease **9**(1), 40–48 (2022)
35. Yu, H., Qi, Z., Jang, L.K., Salakhutdinov, R., Morency, L.P., Liang, P.P.: Mmoe: Enhancing multimodal models with mixtures of multimodal interaction experts. In: Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. pp. 10006–10030 (2024)
36. Yu, X., Przybelski, S.A., Reid, R.I., Lesnick, T.G., Raghavan, S., Graff-Radford, J., Lowe, V.J., Kantarci, K., Knopman, D.S., Petersen, R.C., et al.: Noddi in gray matter is a sensitive marker of aging and early ad changes. Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring **16**(3), e12627 (2024)
37. Zhang, H., Schneider, T., Wheeler-Kingshott, C.A., Alexander, D.C.: Noddi: practical in vivo neurite orientation dispersion and density imaging of the human brain. Neuroimage **61**(4), 1000–1016 (2012)