ADAPTING HFMCA TO GRAPH DATA: SELF-SUPERVISED LEARNING FOR GENERALIZABLE FMRI REPRESENTATIONS

Jakub Frąc[1] *Alexander Schmatz*[1] *Qiang Li*[2] *Guido Van Wingen*[3] *Shujian Yu*[1]

[1]VU Amsterdam [2]TReNDS Center [3]Amsterdam UMC

ABSTRACT

Functional magnetic resonance imaging (fMRI) analysis faces significant challenges due to limited dataset sizes and domain variability between studies. Traditional self-supervised learning methods inspired by computer vision often rely on positive and negative sample pairs, which can be problematic for neuroimaging data where defining appropriate contrasts is non-trivial. We propose adapting a recently developed Hierarchical Functional Maximal Correlation Algorithm (HFMCA) to graph-structured fMRI data, providing a theoretically grounded approach that measures statistical dependence via density ratio decomposition in a reproducing kernel Hilbert space (RKHS), and applies HFMCA-based pretraining to learn robust and generalizable representations. Evaluations across five neuroimaging datasets demonstrate that our adapted method produces competitive embeddings for various classification tasks and enables effective knowledge transfer to unseen datasets. Codebase and supplementary material can be found here: `https://github.com/fr30/mri-eigenencoder`

***Index Terms*—** HFMCA, self-supervised learning, graph transformer, fMRI, representation learning

1. INTRODUCTION

Functional magnetic resonance imaging (fMRI) provides crucial insights into human brain dynamics, with resting-state functional connectivity serving as an important biomarker for neurological and psychiatric conditions [1–3]. Yet, deep learning applications face major challenges due to limited dataset sizes, heterogeneous preprocessing protocols, and persistent domain shifts across scanning centers.

Contrastive self-supervised learning (SSL) has offered promising solutions by adapting approaches from computer vision to neuroimaging data [4–7]. While some methods operate directly on temporal blood oxygen level-dependent (BOLD) signals [4, 6], graph-based approaches that model functional connectivity matrices offer distinct advantages for neuroimaging applications. Functional connectivity graphs provide a more structured and interpretable representation of brain organisation, capturing pairwise statistical dependencies between brain regions whilst reducing the dimensionality and complexity of the raw temporal signal space. Additionally, graph-based augmentations such as node sampling and edge perturbation preserve the underlying network topology.

A recently proposed **H**ierarchical **F**unctional **M**aximal **C**orrelation **A**lgorithm (HFMCA) [8], originally used for image data, provides a theoretically principled approach with strong potential to address these neuroimaging-specific challenges. Unlike contrastive methods that rely on explicit positive–negative sample construction [9, 10], HFMCA measures statistical dependence between low- and high-level features across multiple views without being limited to traditional two-view frameworks [11, 12]. This multi-view capability enables the capture of richer hierarchical dependencies, leading to greater feature diversity and more generalizable representations. HFMCA operates on graph-structured connectivity rather than raw BOLD signals, exploiting brain network topology to integrate complementary views of neural activity, making it well-suited for neuroimaging representation learning.

Contributions: (1) We adapt HFMCA to graph-structured fMRI data, representing the first application and extension of this framework to brain connectivity networks. (2) We demonstrate that HFMCA-pretrained encoders produce competitive embeddings for neuroimaging classification tasks across diverse datasets. (3) We show effective transfer learning capabilities, particularly in scenarios where limited labelled data is available. 4) We evaluate neural scaling laws in the context of fMRI graph encoders, showing that naive pretraining data scaling may induce negative transfer.

2. BACKGROUND

2.1. Functional Maximal Correlation Algorithm

The Functional Maximal Correlation Algorithm [8] measures statistical dependence between random variables X and Y through their probability distributions. For distributions $p(X)$ and $p(Y)$, statistical dependence is characterised by:

$$\rho(X,Y) := \frac{p(X,Y)}{p(X)p(Y)}. \tag{1}$$

FMCA approximates this dependence by decomposing it into eigenvalues σ_k and corresponding orthogonal eigenfunctions:

$$\rho(X,Y) = \frac{p(X,Y)}{p(X)p(Y)} \approx \sum_k \sqrt{\sigma_k}\varphi_k(X)\psi_k(Y) \quad (2)$$

$$\mathbb{E}_X[\varphi_k(X)\varphi_{k'}(X)] = \mathbb{E}_X[\psi_k(Y)\psi_{k'}(Y)] = \begin{cases} 1, & k = k' \\ 0, & k \neq k' \end{cases} \quad (3)$$

Two neural network encoders $f_\theta : \mathcal{X} \to \mathbb{R}^K$ and $g_\omega : \mathcal{Y} \to \mathbb{R}^K$ approximate the eigenfunctions φ and ψ respectively. The autocorrelation matrices are computed as:

$$R_F = \mathbb{E}_X[f_\theta(X)f_\theta^\top(X)], \quad R_G = \mathbb{E}_Y[g_\omega(Y)g_\omega^\top(Y)] \quad (4)$$

$$P_{FG} = \mathbb{E}_{X,Y}[f_\theta(X)g_\omega^\top(Y)] \quad (5)$$

The joint autocorrelation matrix combines these components:

$$R_{FG} = \begin{bmatrix} R_F & P_{FG} \\ P_{FG}^\top & R_G \end{bmatrix} \quad (6)$$

The FMCA objective maximises statistical dependence by minimising:

$$\mathcal{L}_{FMCA} = \log\det R_{FG} - \log\det R_F - \log\det R_G \quad (7)$$

This formulation encourages orthogonal feature learning (maximising $\log\det R_F$ and $\log\det R_G$) while aligning representations between views (minimising $\log\det R_{FG}$).

Recently, FMCA has been employed in self-supervised image representation learning via hierarchical mutual information maximization [8], and in cross-modal representation learning for EEG and EMG [13]. These applications demonstrate FMCA's capacity to capture statistical dependencies across different data modalities and feature hierarchies. We extend this framework to develop a self-supervised learning scheme for functional correlation matrix graphs, where the encoder learns semantically meaningful latent representations of brain connectivity patterns.

3. METHODS

3.1. Problem formulation

Given a small labelled clinical dataset $D_c = \{(X_i, Y_i)\}_{i=1}^{N_c}$ and a large population dataset $D_p = \{X_j\}_{j=1}^{N_p}$ with $N_p \gg N_c$, each subject is represented by a functional connectivity matrix $X_i \in \mathbb{R}^{|V|\times|V|}$, computed over $|V|$ atlas-defined brain regions. Each X_i encodes the pairwise statistical dependencies of regional BOLD signals and is modelled as a weighted graph on $|V|$ nodes. The clinical label $Y_i \in \{0, 1\}$ indicates a binary diagnosis assigned by clinicians for a specific disorder. The objective is to learn a predictive model $f : \mathbb{R}^{|V|\times|V|} \to \{0, 1\}$ defined on the clinical dataset. Training f directly on D_c is susceptible to overfitting due to the limited sample size and variability in labels and data quality.

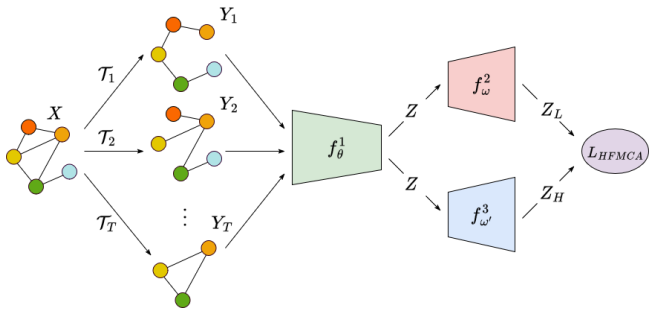

Fig. 1. HFMCA learns representations by maximising the statistical dependence between low- and high-level features of the graph. Each subsampling augmentation (e.g., random walk sampling) is processed through a shared backbone f_θ^1, producing low-level features. These features are either projected individually through a local head $Z^L = [f_\omega^2 \circ f_\theta^1(Y_1) \dots f_\omega^2 \circ f_\theta^1(Y_T)]$ or jointly aggregated through a high-level head $Z^H = \sum_i^T f_{\omega_i'}^3 \circ f_\theta^1(Y_i)$. In pretraining, both projection heads are used to enforce multi-view consistency, while only the backbone is retained for downstream tasks.

To address this, we pretrain a neural encoder f_θ^1 on D_p with a self-supervised objective to obtain semantically meaningful latent representations Z from connectivity graphs, which are then fine-tuned on D_c for diagnostic prediction.

3.2. Graph Construction from fMRI Data

We preprocess resting-state fMRI recordings by parcellating each subject's brain into 116 anatomical regions using the AAL116 atlas [14]. For each region of interest, we extract mean BOLD time series and compute Pearson correlation coefficients between all region pairs, yielding symmetric 116×116 functional connectivity matrices. To build graphs for neural network processing, we retain the top $\frac{|V|^2}{400}$ correlation coefficients as edges, using connectivity values as node features. That is, for each network G, the feature of node k is defined as $X_k = [\rho_{k1}, \dots, \rho_{k|\mathcal{V}|}]^T \in \mathbb{R}^{|\mathcal{V}|}$, where ρ_{kq} is the Pearson correlation between BOLD signals at nodes k and q.

3.3. Hierarchical FMCA with Graphs

Motivated by the original framework developed for image classification [8], we extend HFMCA to graph-structured fMRI data. HFMCA models dependencies between hierarchical feature representations of brain regions rather than pairwise correlations between different data views. Given a source graph X and T augmentations $Y = \{Y_1, Y_2 \dots, Y_T\}$ where $Y_i = \mathcal{T}_i(X)$, the method defines:

Low-level features: $Z^L = \{Z_t^L\}_{t=1}^T$ from individual augmentation representations Z.

High-level features: Z^H from aggregated augmentation

representations Z.

The hierarchical dependence is modelled as:

$$\rho(Z^L, Z^H) = \frac{p(Z^L, Z^H)}{p(Z^L)p(Z^H)}. \quad (8)$$

Three network components extract these features: backbone encoder f_θ^1, and projection heads f_ω^2 and $f_{\omega'}^3$ (Figure 1). For each brain connectivity graph X with augmentations $Y = \{Y_1, \ldots, Y_L\}$, where $Y_i = T_i(X)$, the set of embeddings is defined as $Z = \{f_\theta^1(Y_1), \ldots, f_\theta^1(Y_T)\}$. These features are passed through the projection heads:

Low-level projection: $Z_L = [f_\omega^2 \circ f_\theta^1(Y_1) \ldots f_\omega^2 \circ f_\theta^1(Y_T)]$,
High-level projection: $Z_H = \sum_{i=1}^T f_{\omega_i'}^3 \circ f_\theta^1(Y_i)$.

The hierarchical autocorrelation matrices become:

$$R_L = \mathbb{E}[Z^L(Z^L)^\top], \quad R_H = \mathbb{E}[Z^H(Z^H)^\top], \quad (9)$$

$$P_{LH} = \mathbb{E}[Z^L(Z^H)^\top], \quad R_{LH} = \begin{bmatrix} R_L & P_{LH} \\ P_{LH}^\top & R_H \end{bmatrix}. \quad (10)$$

The training objective minimises:

$$\mathcal{L}_{HFMCA} = \log\det R_{LH} - \log\det R_L - \log\det R_H. \quad (11)$$

Our augmentation strategy employs graph-specific transformations including random walk sampling, node dropping, feature masking, and edge removal. After pretraining, projection heads are discarded and only the backbone encoder f_θ^1 is retained for downstream classification tasks.

3.4. Graph Transformer Architecture

We employ a Graph Transformer backbone based on the GPS architecture [15], which combines local neighbourhood information via message-passing with global attention mechanisms. The architecture incorporates Random Walk Positional Encodings [16] to preserve structural relationships between brain regions. After processing through multiple transformer layers, node embeddings are aggregated using global mean pooling to produce graph-level representations.

4. EXPERIMENTS

4.1. Datasets and Tasks

We evaluate our approach on five neuroimaging datasets representing diverse classification challenges:
REST [17]: Major Depressive Disorder and sex classification (1642 subjects; 51.6% Healthy/MDD ratio; 61.0% Male/Female ratio)
ABIDE [18]: Autism Spectrum Disorder classification (866 subjects; 53.6% Healthy/ASD ratio)
BSNIP [19]: Schizophrenia (SZ) and Bipolar Disorder with Psychosis (BP) (1464 subjects; 43.7%/34.2%/22.1% Healthy/SZ/BP ratio)
AOMIC [20]: Sex classification (881 subjects; 51.9% Male/ Female ratio)
HCP [21]: Sex classification (443 subjects; 55.5% Male/ Female ratio) Models are pretrained on REST and ABIDE datasets (2005 subjects total), then evaluated on all datasets including previously unseen ones (BSNIP, AOMIC, HCP).

Model	REST (MDD)	REST (Sex)	ABIDE
Majority class	51.6 ± 0.0	61.0 ± 0.0	53.6 ± 0.0
Baseline_F	55.5 ± 1.0	64.8 ± 1.1	53.5 ± 1.6
VICReg_F	56.6 ± 1.0	65.4 ± 0.5	53.5 ± 1.3
BT_F	$\mathbf{57.3 \pm 0.6}$	64.7 ± 0.7	53.5 ± 1.7
SimCLR_F	55.9 ± 0.9	$\underline{65.9 \pm 0.7}$	$\underline{54.0 \pm 1.0}$
HFMCA_F	$\mathbf{57.4 \pm 0.9}$	$\mathbf{66.6 \pm 1.3}$	$\mathbf{54.7 \pm 1.5}$

Table 1. Classification accuracy (%) with frozen encoders on datasets seen during pretraining.

Model	REST (MDD)	REST (Sex)	ABIDE
Majority class	51.6 ± 0.0	61.0 ± 0.0	53.6 ± 0.0
Baseline	58.8 ± 1.2	$\underline{68.3 \pm 0.4}$	54.9 ± 1.8
VICReg	58.7 ± 1.0	65.5 ± 1.4	52.9 ± 1.7
BT	$\mathbf{59.9 \pm 1.4}$	65.2 ± 1.5	53.5 ± 1.0
SimCLR	59.3 ± 1.2	$\underline{68.4 \pm 0.9}$	$\mathbf{56.3 \pm 1.3}$
HFMCA	$\mathbf{59.8 \pm 1.2}$	$\mathbf{70.3 \pm 1.0}$	$\mathbf{56.1 \pm 1.6}$

Table 2. Classification accuracy (%) with unfrozen encoders on previously seen datasets.

4.2. Training Protocol and Evaluation

The GPS encoder is equipped with two projection heads required for HFMCA training. We train for 200 epochs using Adam optimiser with learning rate 10^{-3}, weight decay 10^{-6}, and batch size 256. The model is pretrained on a combined REST and ABIDE dataset, comprising approximately 2500 samples. After pretraining, projection heads are discarded and only the backbone encoder is retained for downstream evaluation.

We compare HFMCA against established self-supervised baselines: SimCLR [9], Barlow Twins [11], VICReg [12], and a randomly initialised Baseline. Following standard practice [12, 22], we evaluate both frozen and unfrozen scenarios using nested 5-fold cross-validation with 10 independent runs.

Model	BSNIP	AOMIC (Sex)	HCP (Sex)
Majority class	43.7 ± 0.0	51.9 ± 0.0	55.5 ± 0.0
Baseline_F	47.7 ± 0.9	56.9 ± 1.3	63.0 ± 2.4
VICReg_F	46.4 ± 0.9	60.0 ± 1.5	65.7 ± 1.5
BT_F	46.2 ± 1.2	$\mathbf{60.9 \pm 0.8}$	66.0 ± 1.5
SimCLR_F	47.9 ± 0.5	59.2 ± 1.5	$\mathbf{66.6 \pm 1.5}$
HFMCA_F	$\mathbf{48.7 \pm 0.8}$	59.4 ± 1.1	66.0 ± 1.9

Table 3. Classification accuracy (%) with frozen encoders on previously unseen datasets.

Model	BSNIP	AOMIC (Sex)	HCP (Sex)
Majority class	43.7 ± 0.0	51.9 ± 0.0	55.5 ± 0.0
Baseline	47.8 ± 0.8	62.5 ± 1.8	$\mathbf{71.1 \pm 2.2}$
VICReg	47.1 ± 1.0	60.5 ± 1.7	66.0 ± 3.2
BT	47.2 ± 1.2	60.3 ± 1.4	66.1 ± 1.1
SimCLR	$\mathbf{48.7 \pm 0.6}$	63.7 ± 1.8	68.2 ± 3.1
HFMCA	48.4 ± 0.8	$\mathbf{64.6 \pm 1.4}$	70.2 ± 0.6

Table 4. Zero-shot classification accuracy (%) with unfrozen encoders.

4.3. Quality of Learned Representations

To assess the quality of the learned representations, we attach a trainable linear classifier on top of the frozen graph encoder embeddings and evaluate performance on multiple classification tasks (Table 1). We also unfreeze the encoder and fine-tune it on the same downstream tasks (Table 2). HFMCA achieves competitive results across all benchmarks, demonstrating more consistent performance across runs.

4.4. Transfer Learning Evaluation

To evaluate the transferability of the HFMCA encoder, we fine-tune both frozen and unfrozen variants with attached linear heads on tasks outside the pretraining datasets. The results (Tables 3 and 4) show that HFMCA consistently outperforms random initialization (*Baseline*) and remains competitive with other methods. Notably, it achieves more stable performance on average, exhibiting lower variance across experimental runs.

4.5. Scaling Laws

To investigate whether pretraining performance scales with dataset size, we pretrained HFMCA encoders on progressively larger datasets: **REST** (1313 subjects), **REST + ABIDE** (2179 subjects), **REST + ABIDE + HCP** (2622 subjects), and **REST + ABIDE + HCP + BSNIP** (4512 subjects). Each encoder was evaluated using frozen linear heads across all downstream tasks following the protocol from Section 4.3. Figure 2 shows that contrary to established scaling laws [23, 24], performance does not improve monotonically with increased data volume. Performance peaks with REST + ABIDE but declines when adding HCP and BSNIP datasets. This aligns with recent findings that naive scaling can produce negative transfer effects in graph foundation models [25, 26].

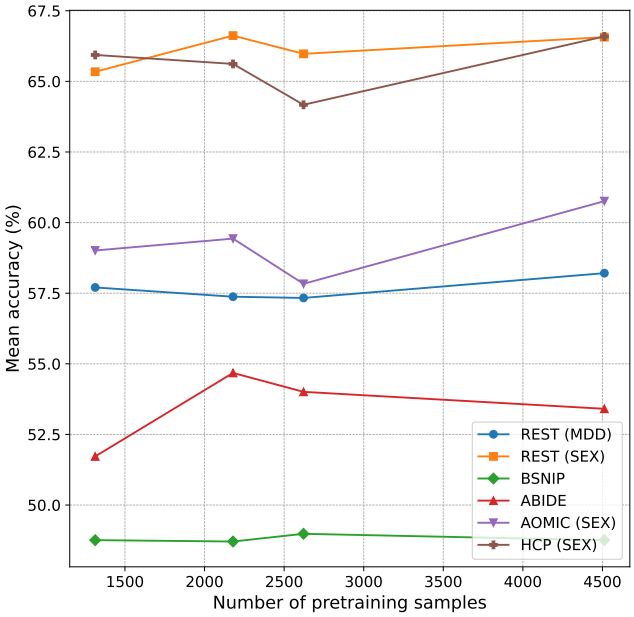

Fig. 2. Encoders were trained with HFMCA using varying amounts of pretraining data and fine-tuned with linear heads. No clear linear relationship was observed between the amount of pretraining data and downstream performance, which aligns with findings from recent studies.

5. CONCLUSION

We successfully extended HFMCA to graph-structured fMRI data, providing a theoretically principled approach to self-supervised representation learning. Our method achieves competitive performance across five neuroimaging datasets. The demonstrated transfer learning capabilities and stable training make HFMCA particularly suitable for neuroimaging applications. Future work should explore larger-scale datasets and investigate HFMCA as a component of foundational models for brain imaging. Even though the initial scaling law analysis suggests greater complexity compared to text and vision domains, the framework and transferability indicate important contributions toward generalizable computational models of brain function.

References

[1] C. S. Monk, S. J. Peltier, J. L. Wiggins, S.-J. Weng, M. Carrasco, S. Risi, and C. Lord, "Abnormalities of intrinsic functional connectivity in autism spectrum disorders,," *NeuroImage*, vol. 47, no. 2, pp. 764–772, 2009.

[2] B. de Kwaasteniet, E. Ruhe, M. Caan, M. Rive, S. Olabarriaga, M. Groefsema, L. Heesink, G. van Wingen, and D. Denys, "Relation between structural and functional connectivity in major depressive disorder," *Biological Psychiatry*, vol. 74, no. 1, pp. 40–47, 2013.

[3] M. Liang, Y. Zhou, T. Jiang, Z. Liu, L. Tian, H. Liu, and Y. Hao, "Widespread functional disconnectivity in schizophrenia with resting-state functional magnetic resonance imaging," *Neuroreport*, vol. 17, no. 2, pp. 209–213, Feb. 2006.

[4] C. Shi, Y. Wang, Y. Wu, S. Chen, R. Hu, M. Zhang, B. Qiu, and X. Wang, "Self-supervised pretraining improves the performance of classification of task functional magnetic resonance imaging," *Frontiers in Neuroscience*, vol. 17, pp. 1199312, June 2023.

[5] X. Wang, Y. Chu, Q. Wang, L. Cao, L. Qiao, L. Zhang, and M. Liu, "Unsupervised contrastive graph learning for resting-state functional mri analysis and brain disorder detection," *Human Brain Mapping*, vol. 44, no. 17, pp. 5672–5692, Dec 2023, Epub 2023 Sep 5.

[6] J. O. Caro et al., "BrainLM: A foundation model for brain activity recordings," in *ICLR*, 2024.

[7] I. Malkiel, G. Rosenman, L. Wolf, and T. Hendler, "Self-supervised transformers for fmri representation," in *Proceedings of the 5th International Conference on Medical Imaging with Deep Learning*, 2022, pp. 895–913.

[8] Bo Hu, Yuheng Bu, and José C Príncipe, "Learning orthonormal features in self-supervised learning using functional maximal correlation," in *2024 IEEE International Conference on Image Processing (ICIP)*, 2024, pp. 472–478.

[9] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton, "A simple framework for contrastive learning of visual representations," in *ICML*, 2020, pp. 1597–1607.

[10] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick, "Momentum contrast for unsupervised visual representation learning," in *CVPR*, 2020, pp. 9729–9738.

[11] Jure Zbontar, Li Jing, Ishan Misra, Yann LeCun, and Stéphane Deny, "Barlow twins: Self-supervised learning via redundancy reduction," in *ICML*, 2021, pp. 12310–12320.

[12] Adrien Bardes, Jean Ponce, and Yann Lecun, "Vicreg: Variance-invariance-covariance regularization for self-supervised learning," in *ICLR*, 2022.

[13] S. Ma, B. Hu, T. Jia, A. K. Clarke, B. Zicher, A. H. Caillet, D. Farina, and J. C. Principe, "Learning cortico-muscular dependence through orthonormal decomposition of density ratios," in *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024.

[14] N. Tzourio-Mazoyer et al., "Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the MNI MRI single-subject brain," *NeuroImage*, vol. 15, no. 1, pp. 273–289, 2002.

[15] Ladislav Rampášek, Michael Galkin, Vijay Prakash Dwivedi, Anh Tuan Luu, Guy Wolf, and Dominique Beaini, "Recipe for a general, powerful, scalable graph transformer," in *NeurIPS*, 2022, vol. 35, pp. 14501–14515.

[16] Vijay Prakash Dwivedi, Anh Tuan Luu, Thomas Laurent, Yoshua Bengio, and Xavier Bresson, "Graph neural networks with learnable structural and positional representations," in *International Conference on Learning Representations*, 2022.

[17] X. Chen et al., "The direct consortium and the rest-meta-mdd project: towards neuroimaging biomarkers of major depressive disorder," *Psychoradiology*, vol. 2, no. 1, pp. 32–42, 2022.

[18] C. Craddock, Y. Benhajali, C. Chu, F. Chouinard, A. Evans, A. Jakab, B. Khundrakpam, J. D. Lewis, Q. Li, M. Milham, C. Yan, and P. Bellec, "The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives," *Neuroinformatics*, 2013.

[19] C. A. Tamminga, G. Pearlson, M. Keshavan, J. Sweeney, B. Clementz, and G. Thaker, "Bipolar and schizophrenia network for intermediate phenotypes: Outcomes across the psychosis continuum," *Schizophrenia Bulletin*, vol. 40, pp. S131–S137, 02 2014.

[20] L. Snoek, M. M. van der Miesen, T. Beemsterboer, A. Van Der Leij, A. Eigenhuis, and H. S. Scholte, "The amsterdam open mri collection, a set of multimodal mri datasets for individual difference analyses," *Scientific Data*, vol. 8, no. 1, pp. 1–23, 2021.

[21] D. C. Van Essen et al., "The human connectome project: A data acquisition perspective," *NeuroImage*, vol. 62, no. 4, pp. 2222–2231, 2012, Connectivity.

[22] Ishan Misra and Laurens van der Maaten, "Self-supervised learning of pretext-invariant representations," in *CVPR*, 2020, pp. 6707–6717.

[23] J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, and D. Amodei, "Scaling laws for neural language models," *arXiv preprint arXiv:2001.08361*, 2020.

[24] Y. Bansal et al., "Data scaling laws in nmt: The effect of noise and architecture," in *ICML*, 2022, pp. 1466–1482.

[25] Weihua Hu, Bowen Liu, Joseph Gomes, Marinka Zitnik, Percy Liang, Vijay Pande, and Jure Leskovec, "Strategies for pre-training graph neural networks," in *ICLR*, 2020.

[26] Yuxuan Cao et al., "When to pre-train graph neural networks? from data generation perspective!," in *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, 2023, pp. 142–153.