Improving Prediction of Cognitive Abilities through Integrated Resting-State and Task fMRI

Yujing M. Jiang[1], Victor K. S. Chan[1], Jing Jun Wong[1], Nicole H. L. Wong[1], Bolton K. H. Chau[1,2,3]
[1]Department of Rehabilitation Sciences, [2]Mental Health Research Centre, [3]University Research Facility in Behavioral and Systems Neuroscience
The Hong Kong Polytechnic University, Hong Kong

Abstract

Using resting-state fMRI (rs-fMRI) and task-fMRI are two common approaches of predicting cognitive abilities, yet conventional methods offer limited accuracy. We propose a transformer-based framework that unifies rs-fMRI and t-fMRI, drawing inspiration from Large Language Models (LLMs) known for integrating diverse sequential inputs. Instead of treating these modalities separately, our approach encodes both as continuous temporal sequences and applies self-attention to learn Dynamic Activity Signatures that capture neural processes common to spontaneous and task-evoked activity. This unified latent space produces consistent predictions across rs-fMRI and t-fMRI, even in the absence of task data, while eliminating the need for multiple models. Evaluated on the Human Connectome Project (HCP) and Alzheimer's Disease Neuroimaging Initiative (ADNI) datasets, our framework demonstrated high predictive accuracy and robust generalization, proving the effectiveness of a unified model that seamlessly integrates rs-fMRI and t-fMRI.

Keywords: fMRI; Deep Learning; Cognitive Prediction.

Introduction

There has been great interest in applying functional magnetic resonance imaging (fMRI) data to predict cognitive abilities (Tavor et al., 2016; Rasero et al., 2018). One approach of making such predictions is by using resting-state fMRI data, which captures neural dynamics without specifying any task and has the advantage of being purpose general. Another approach is by using task fMRI data, which contains neural activity elicited by specific task paradigms. This generally offers better predictions, but the scope of prediction is limited to the specific cognitive ability corresponding to the task and only to individuals who are capable of performing the tasks (Canario, Chen, & Biswal, 2021; Tetereva, Li, Deng, Stringaris, & Pat, 2022). Moreover, traditional methods rely on static maps, such as independent components from rs-fMRI and contrast maps from t-fMRI, that overlook the predictabilities of continuous neural dynamics (Cole, Ito, Bassett, & Schultz, 2016; Ngo, Khosla, Jamison, Kuceyeski, & Sabuncu, 2022). To overcome these limitations, we propose a transformer-based framework that integrates rs-fMRI and t-fMRI data by training the model to distinguish between them while learning a unified dynamic encoding that captures both spontaneous and task-induced neural processes. Our results showed prediction correlations above 0.90 for working memory and relational task performances. Importantly, to assess cross-dataset generalizability, the model, which was trained using the Human Connectome Project (HCP) dataset (Van Essen et al., 2013), was tested on rs-fMRI data from the ADNI dataset (Rasero et al., 2018). The model not only generalized well but also produced predictions consistent with the clinical profiles of individuals with mild cognitive impairment and Alzheimer's disease.

Methods

We used rs-fMRI and t-fMRI data from the 3T HCP S1200 dataset (1,053 healthy young adults) to train the model for working memory and relational tasks. To test cross-dataset generalizability, we applied the HCP-trained model to rs-fMRI data from ADNI (618 cognitively normal, 296 mild cognitive impairment, and 172 Alzheimer's disease participants). This combination of healthy and clinical populations enables evaluation across diverse imaging protocols and participant characteristics.

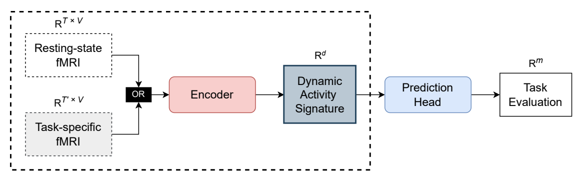

Figure 1: Schematic illustration of the proposed framework.

Model Architecture

Overview. Our framework (Figure 1) employs a transformer-based approach that fuses rs-fMRI and t-fMRI into a single latent space. By jointly encoding spontaneous and task-evoked brain signals, the model enhances cognitive prediction and captures unified neural dynamics.

Encoder. Each 3D fMRI volume is partitioned into patches, enabling the model to extract localized spatial features using a SwinTransformer module (Liu et al., 2021). These spatial features, which reflect regional brain activity, are then organized into a temporal sequence and processed by a TimeSformer (Bertasius, Wang, & Torresani, 2021) that learns long-range temporal dependencies across the 4D data. This sequential encoding is key to understanding how brain activity evolves over time and under different cognitive conditions.

Dynamic Activity Signature and Prediction Head. A self-supervised contrastive loss aligns representations from rs-fMRI and t-fMRI of the same individual, generating a dynamic activity signature that integrates intrinsic connectivity with task-induced responses. The final prediction head then maps this signature to target outputs, such as cognitive performance scores or voxel-wise activation maps, thereby translating complex neural patterns into actionable predictions.

Results

HCP Task Performance. We first trained our model on HCP data by jointly encoding rs-fMRI and t-fMRI to learn a shared latent representation. We then evaluated its generalizability by predicting cognitive performance using only rs-fMRI. Figure 2 shows probability density functions (top) and regression plots (bottom) for WM and relational tasks. Strong correlations above 0.8, especially for the WM 0-back and 2-back tasks, demonstrate that the model reliably captures cognitive load variability. Although the WM Difference shows a slightly broader spread, its robust correlation confirms effective capture of dynamic fluctuations in task difficulty. Overall, these results validate our core premise that a unified representation of resting-state and task data can accurately predict performance even when task scans are unavailable.

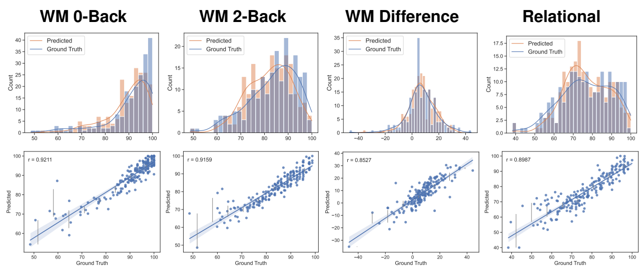

Figure 2: Model Performance in Predicting Working Memory and Relational Task Outcomes on the HCP Dataset: Probability Density Functions (Upper) and Regression Plots (Lower).

Impact of Modality Fusion and Data Representation. We next conducted further analyses (Figure 3, left) to quantify the contribution of each modality to prediction accuracy by comparing four distinct training strategies:

1. Using rs-fMRI only, which yields correlations in the 0.3–0.6 range;
2. Using t-fMRI only, which achieves moderate improvements (0.55–0.63);
3. Using synthetic t-fMRI generated from rs-fMRI via Graph Convolutional Networks (Wang, Li, & Hu, 2021), which boosts correlations to 0.7–0.75;
4. Using the proposed approach, which attains correlations above 0.90 for WM 0-back and 2-back, 0.89 for WM Difference, and 0.82 for relational tasks.

These results clearly demonstrate that integrating spontaneous connectivity from rs-fMRI with task-evoked signals is essential for achieving robust predictive performance.

Figure 3 (right) further examines how common approaches of dimension reduction of fMRI data affect prediction accuracies. We compared the high-dimensional RAW 4D data with three lower-dimensional schemes: the HCP (Glasser et al., 2013) representation, which retains 10% of voxels using a surface-based format; the Cole-Anticevic (CA) (Ji et al., 2019) parcellation, which divides the brain into functional regions; and the Multi-Modal Parcellation (MMP) (Glasser et al., 2016), which combines multiple imaging modalities. The HCP scheme consistently yielded higher correlations across tasks (often >0.9), suggesting that moderate reduction preserves key signals while reducing noise. In contrast, the more compressed CA and MMP schemes lacked the resolution to capture complex task-related patterns, lowering performance. These results highlight the importance of choosing an appropriate level of granularity, as excessive reduction can obscure critical cognitive information.

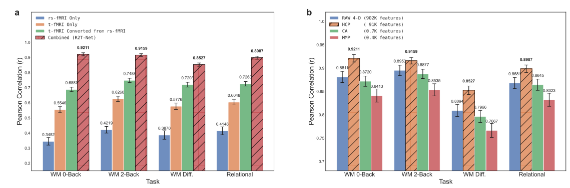

Figure 3: Evaluation of fMRI Acquisition Modalities (Left) and Neural Representation (Right) for Cognitive Performance Prediction.

Cross-Dataset Generalizability. Finally, to evaluate the robustness of our approach, we applied the HCP-trained model to rs-fMRI data from ADNI, which includes healthy controls, MCI patients, and AD patients (Figure 4). Despite variations in scanning protocols and participant characteristics, the model preserved intrinsic connectivity patterns and produced pseudo-t-fMRI predictions aligned with clinical expectations (CN $>$ MCI $>$ AD). The left panel shows correlations among predicted scores (WM 0-back, WM 2-back, and Relational), while the right panel's PCA projection separates diagnostic groups in line with known clinical trajectories. These findings suggest that our method can accurately infer cognitive performance from resting-state data alone, offering potential utility in clinical settings where task-based scans are not feasible.

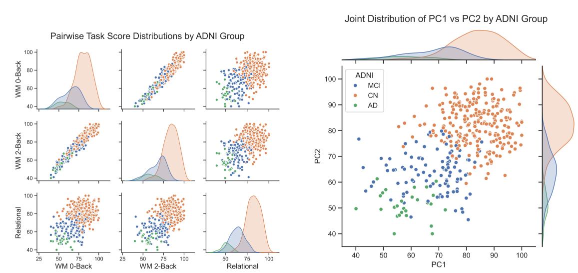

Figure 4: Predicted Cognitive Performance in ADNI: Feature Relationships (Left) and PCA Projection of Performance Metrics (Right).

Discussion

Unifying rs-fMRI and t-fMRI within a single transformer-based framework significantly improves predictive accuracy over traditional single-modality approaches. Our method learns a shared latent representation that captures both intrinsic connectivity and task-evoked dynamics, enabling robust cognitive predictions from resting-state data which is especially valuable for clinical populations unable to perform task scans. It offers deeper insight into the brain's functional organization under varying demands and advances individualized fMRI-based cognitive modeling. Experiments across diverse datasets show that combining both modalities outperforms separate analyses and synthetic mappings. Additionally, our results suggest that incorporating behavioral and structural measures in future work could further refine temporal dynamics and strengthen the multimodal framework for both research and clinical use.

Acknowledgements

This study was funded by Hong Kong Research Grants Council (C5001-23Y) and The Hong Kong Polytechnic University Project of Strategic Importance (1-ZE0E).

Data were provided [in part] by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Data used in preparation of this article were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). As such, the investigators within the ADNI contributed to the design and implementation of ADNI and/or provided data but did not participate in analysis or writing of this report. Data collection and sharing for this project was funded by the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904) and DOD ADNI (Department of Defense award number W81XWH-12-2-0012).

References

Bertasius, G., Wang, H., & Torresani, L. (2021). Is space-time attention all you need for video understanding? In *Icml* (Vol. 2, p. 4).

Canario, E., Chen, D., & Biswal, B. (2021). A review of resting-state fmri and its use to examine psychiatric disorders. *Psychoradiology*, *1*(1), 42–53.

Cole, M. W., Ito, T., Bassett, D. S., & Schultz, D. H. (2016). Activity flow over resting-state networks shapes cognitive task activations. *Nature neuroscience*, *19*(12), 1718–1726.

Glasser, M. F., Coalson, T. S., Robinson, E. C., Hacker, C. D., Harwell, J., Yacoub, E., ... others (2016). A multi-modal parcellation of human cerebral cortex. *Nature*, *536*(7615), 171–178.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., ... others (2013). The minimal preprocessing pipelines for the human connectome project. *Neuroimage*, *80*, 105–124.

Ji, J. L., Spronk, M., Kulkarni, K., Repovš, G., Anticevic, A., & Cole, M. W. (2019). Mapping the human brain's cortical-subcortical functional network organization. *Neuroimage*, *185*, 35–57.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., ... Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the ieee/cvf international conference on computer vision* (pp. 10012–10022).

Ngo, G. H., Khosla, M., Jamison, K., Kuceyeski, A., & Sabuncu, M. R. (2022). Predicting individual task contrasts from resting-state functional connectivity using a surface-based convolutional network. *NeuroImage*, *248*, 118849.

Rasero, J., Aerts, H., Ontivero Ortega, M., Cortes, J. M., Stramaglia, S., & Marinazzo, D. (2018). Predicting functional networks from region connectivity profiles in task-based versus resting-state fmri data. *PloS one*, *13*(11), e0207385.

Tavor, I., Jones, O. P., Mars, R. B., Smith, S., Behrens, T., & Jbabdi, S. (2016). Task-free mri predicts individual differences in brain activity during task performance. *Science*, *352*(6282), 216–220.

Tetereva, A., Li, J., Deng, J. D., Stringaris, A., & Pat, N. (2022). Capturing brain-cognition relationship: Integrating task-based fmri across tasks markedly boosts prediction and test-retest reliability. *NeuroImage*, *263*, 119588.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., ... others (2013). The wu-minn human connectome project: an overview. *Neuroimage*, *80*, 62–79.

Wang, L., Li, K., & Hu, X. P. (2021). Graph convolutional network for fmri analysis based on connectivity neighborhood. *Network Neuroscience*, *5*(1), 83–95.