Efficient Deep Learning Models for Predicting Individualized Task Activation from Resting-State Functional Connectivity

Soren J. Madsen[1*], Young-Eun Lee[1*], Shaun K. L. Quah[1],
Lucina Q. Uddin[3], Jeanette A. Mumford[2], Deanna M. Barch[4], Damien A. Fair[5],
Ian H. Gotlib[2], Russell A. Poldrack[2], Amy Kuceyeski[6], Manish Saggar[1†]

[1]Department of Psychiatry, Stanford University, USA

[3]Department of Psychiatry, University of California, Los Angeles, USA

[4]Department of Psychology, Washington University in St. Louis, USA

[5]Department of Pediatrics, University of Minnesota, USA

[6]Department of Radiology, Weill Cornell Medicine, USA

[*] The authors contributed equally to this work.

[†]Correspondence: saggar@stanford.edu

November 25, 2025

Keywords: deep learning, resting-state, task contrast, functional MRI

Abstract: Deep learning has shown promise in predicting task-evoked brain activation patterns from resting-state fMRI. In this study, we replicate the state-of-the-art BrainSurfCNN model using data from the Human Connectome Project, and explore biologically motivated frameworks to improve prediction performance and computational efficiency. Specifically, we evaluate two model variants: BrainSERF, which integrates a Squeeze-and-Excitation attention mechanism into the U-Net backbone, and BrainSurfGCN, a lightweight graph neural network architecture that leverages mesh topology for efficient message passing. Both models yield comparable prediction performance to BrainSurfCNN, with BrainSERF offering modest improvements in subject identification accuracy and BrainSurfGCN delivering substantial reductions in model size and training time. We also investigate factors contributing to inter-

individual variability in prediction accuracy and identify task performance and data quality as significant modulators. Our findings highlight new architectural avenues for improving the scalability of brain decoding models and underscore the need to consider individual variability when evaluating prediction fidelity.

1 Introduction

One of the foundational questions in cognitive neuroscience is whether the brain's intrinsic organization, observable during rest, encodes information predictive of how that same brain will respond during active cognitive tasks. If intrinsic functional connectivity reflects stable traits that constrain task-evoked activation, then models trained on resting-state data should be able to accurately predict task contrast maps at the individual level (Cole et al., 2016; Ngo et al., 2022; Tavor et al., 2016). This hypothesis has significant implications for understanding brain-behavior relationships, characterizing individual differences, and developing precision neuroimaging tools, particularly for populations that are unable to perform scanner-based tasks. At the same time, validating such models provides a data-driven approach to test mechanistic assumptions about the relationship between intrinsic and evoked neural states. In this work, we build on recent foundational research (Ngo et al., 2022) to investigate whether biologically inspired architectures can enhance the efficiency or fidelity of these predictions and why rest-to-task prediction performs better for some individuals and task contrasts than others.

Prior work has demonstrated that the spatial organization of resting-state networks closely mirrors the brain's task-evoked architecture (Smith et al., 2009; Smith-Collins et al., 2015; Sui et al., 2009), suggesting that intrinsic functional connectivity may serve as a latent scaffold for cognitive function. This raises the possibility that resting-state functional magnetic resonance imaging (rsfMRI) contains sufficient information to reconstruct task activation maps at the individual level, effectively using the brain at rest to infer how it behaves in action. In addition to its practical utility in cases where task data are unavailable or infeasible (e.g., clinical or pediatric populations), such predictions offer a powerful test of the stability, specificity, and behavioral relevance of intrinsic connectivity (Chow et al., 2025; Hearne et al., 2021; R. Jiang et al., 2020; Savage et al., 2024; Serin et al., 2025).

This line of inquiry also has significant implications for translational research frameworks such as the NIMH Research Domain Criteria (RDoC), which seek to redefine mental illness in terms of dimensional, neurobiologically grounded constructs (Cuthbert and Insel, 2013; Insel, 2014). Task-fMRI has traditionally played a central role in identifying brain circuits linked to RDoC domains, yet no single individual completes all relevant tasks (Quah, Madsen, et al., 2025; Quah, Jo, et al., 2025). Accurate rest-to-task

prediction could enable virtual approximation of these activation patterns, potentially allowing comprehensive, individualized neural profiling without requiring data acquisition across all RDoC task domains. Thus, predictive models offer both a mechanistic window into brain organization and a scalable tool for next-generation clinical neuroscience.

Foundational work by Tavor et al., 2016 and Cole et al., 2016 demonstrated that task activation maps could be predicted from resting-state functional connectivity using linear models, establishing a baseline for rest-to-task inference. These studies showed that intrinsic connectivity patterns contain behaviorally relevant information. Building on this foundation, Ngo et al., 2022 introduced BrainSurfCNN, a deep learning model that integrates U-Net-style architecture with surface-based mesh convolutions to significantly improve individual-level prediction accuracy. While technically impressive, such models raise two central neuroscience questions that remain unresolved: (1) Can biologically motivated changes to model architecture enhance the fidelity or efficiency of these predictions? And (2) why are some individuals or task contrasts more predictable than others?

To address the first question, we examine whether biologically motivated architectural changes can enhance the fidelity or efficiency of rest-to-task predictions. We introduce two new models built upon distinct principles. The first, BrainSERF, augments BrainSurfCNN with squeeze-and-excitation (SE) modules (Hu et al., 2018), which implement dynamic channel-wise feature reweighting. While not directly neuroanatomical, this mechanism draws inspiration from biological attention systems that modulate information flow based on contextual salience. The second model, BrainSurfGCN, utilizes graph convolution (Kipf and Welling, 2016) and moves closer to anatomical realism by treating the cortical surface as a graph and explicitly modeling localized, neighborhood-based information flow, thereby more closely mirroring the brain's topological and network-level architecture. Although all models achieved comparable prediction accuracy, BrainSurfGCN significantly reduced computational cost, highlighting how biologically grounded structural designs may offer scalable, efficient, and interpretable paths forward for individualized brain modeling.

To address the second question, i.e., why rest-to-task prediction succeeds for some individuals or task domains but not others. We examine whether variability in model performance may reflect meaningful differences in cognitive engagement, neural signal reliability, and the degree to which intrinsic connectivity scaffolds task-evoked activity. To address this, we evaluate how behavioral performance and scan quality influence prediction accuracy, and we examine the cortical distribution of prediction errors in relation to known anatomical noise profiles. These analyses aim to reveal when and where brain dynamics and individual traits facilitate or constrain successful decoding of task-related activity from resting-state data.

Together, these two lines of inquiry: biologically inspired model design and individual variability in rest-

to-task prediction, allow us to probe both the when and how of successful decoding from intrinsic connectivity. This dual approach aims to advance both mechanistic understanding and methodological innovation in precision neuroimaging.

2 Results

Here, we show that two biologically motivated model variants, BrainSERF and BrainSurfGCN, match or exceed the predictive fidelity of the original BrainSurfCNN while substantially reducing computational cost.

We used resting state and task fMRI data from the Human Connectome Project (HCP) Young Adult cohort (Van Essen et al., 2013). After applying standard HCP preprocessing pipelines including ICA-FIX, we extracted eight 600-frame samples from the four resting-state runs to construct individual connectivity fingerprints. Task fMRI provided 47 contrasts across seven task domains, and 919 individuals with complete resting and task data were included for model training and validation. An independent set of 39 participants who completed two task fMRI sessions was held out for testing, where the second session served as a repeat measurement to estimate an empirical upper bound of predictability.

To predict individual task activation maps from resting-state connectivity, we evaluated three surface-based deep learning models that operate on multi-channel fsLR meshes: BrainSurfCNN, a convolutional neural network (CNN); BrainSERF, an SE-enhanced variant of this CNN architecture; and BrainSurfGCN, a graph convolutional network (GCN) within the broader class of graph neural networks (GNNs). These architectures capture spatial dependencies through mesh convolutions or graph-based message passing, as summarized in Figure 1. Models were trained using mean square error (MSE) followed by a reconstructive contrastive loss to promote subject-specific representations. Performance was assessed using spatial correlation, Dice overlap across multiple thresholds, and subject identification accuracy. Full methodological details are provided in the Methods section.

In this section, we address two core questions central to our investigation: (1) Can biologically inspired architectural modifications improve the fidelity and efficiency of predicting task-evoked brain activation from resting-state connectivity? and (2) Why does prediction performance vary across individuals and cognitive tasks?

To this end, we organize our results in a three-part sequence: (i) Model Improvement: We first evaluate predictive performance across three model architectures, including BrainSurfCNN (baseline), BrainSERF, and BrainSurfGCN, highlighting improvements in efficiency and fidelity afforded by biologically inspired

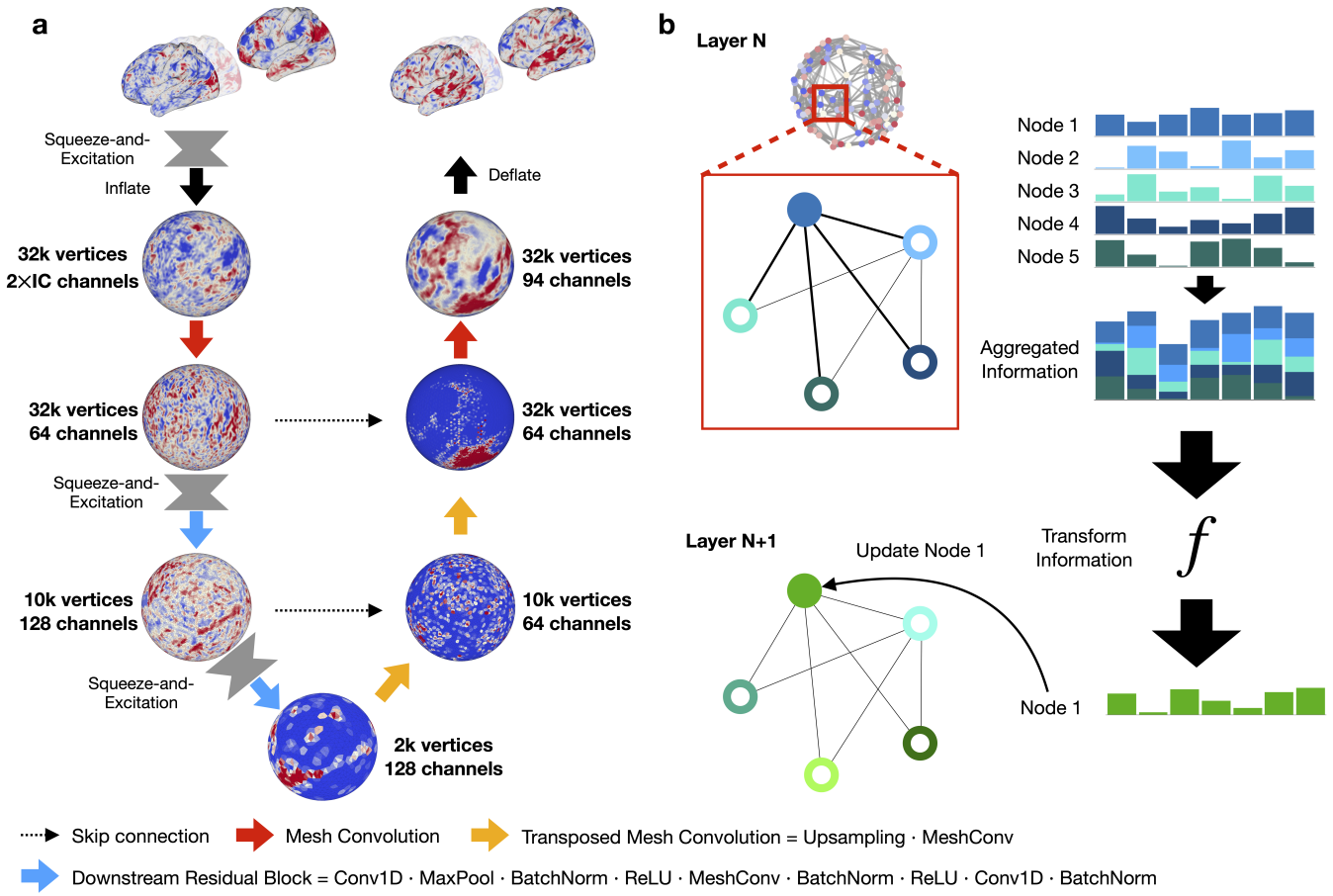

Figure 1: **Overview of BrainSurfGCN architecture for cortical mesh-based fMRI analysis.** a. The input surface-based fMRI data are processed through a hierarchical mesh convolutional network, incorporating SE modules and skip connections. Vertices and channel dimensions are progressively transformed using mesh convolution, residual blocks, and upsampling operations. b. At each BrainSurfGCN layer, node features in the cortical mesh are aggregated, transformed, and updated according to graph connectivity. Aggregated node information is passed through nonlinear transformations to generate updated node representations for subsequent layers. This framework enables multi-scale integration and hierarchical feature extraction from full-brain functional mesh data.

modifications. (ii) Model Evaluation: We then assess subject-level prediction specificity through a subject identification analysis, offering a complementary measure of how well each model captures individualized activation patterns. (iii) Mechanisms of Variability: Finally, we examine factors that drive individual and contrast-level differences in prediction performance, including task engagement, data quality, and inter-subject variability in activation patterns.

2.1 Architectural Improvements

We first evaluated how well each model, including BrainSurfCNN, BrainSERF, and BrainSurfGCN, could predict individual-level task activation maps from resting-state connectivity patterns. As shown in Figure 2 for one representative participant, all three models qualitatively captured key features of task-evoked responses, such as somatomotor activation during the Motor task and activation in association cortices during the Social task. Among them, BrainSurfGCN offered the sharpest spatial localization and highest alignment with ground truth, suggesting that biologically grounded architectural changes can improve predictive fidelity and efficiency.

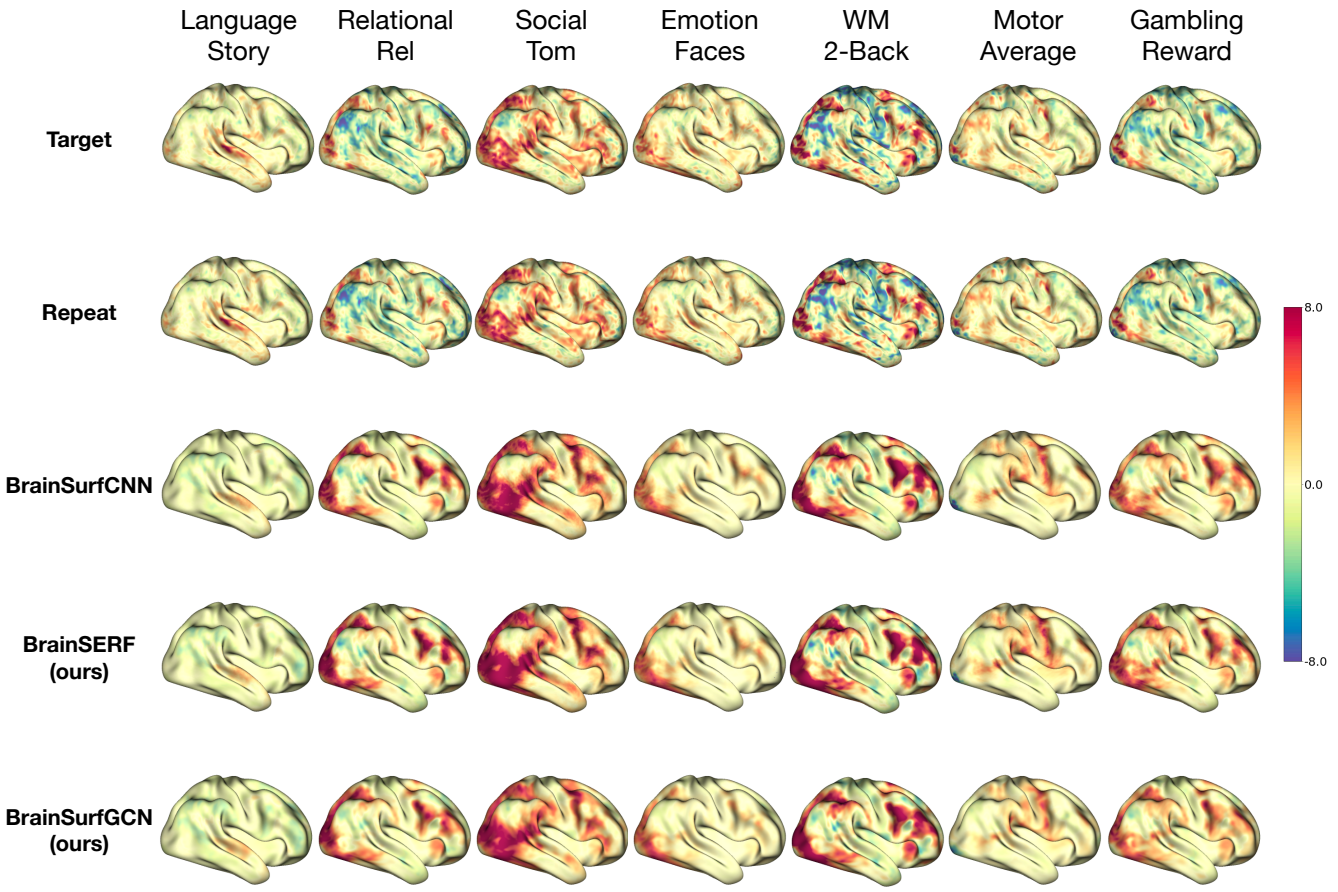

Figure 2: **Task activation prediction of 7 task contrasts.** We compare ground truth, repeat, and models' predicted activation for an individual (subject 917255). The activation maps on the right hemisphere (lateral view) represent, respectively, Language: Story, Relational Processing: Relational, Social Cognition: Theory of Mind, Emotion: Faces, Working Memory: 2-Back, Motor: Average, and Gambling: Reward.

To quantitatively evaluate model performance, we computed Dice coefficients between predicted and ground-truth activation maps across multiple thresholds (5%–50%, in 5% increments), following the

procedure established in Ngo et al., 2022. Thresholding allows us to isolate task-relevant signal while mitigating the influence of noise. Figure 3 shows example outputs for three thresholds (10%, 25%, 50%) from the Social Cognition: Theory of Mind contrast, illustrating how prediction sensitivity changes with threshold level.

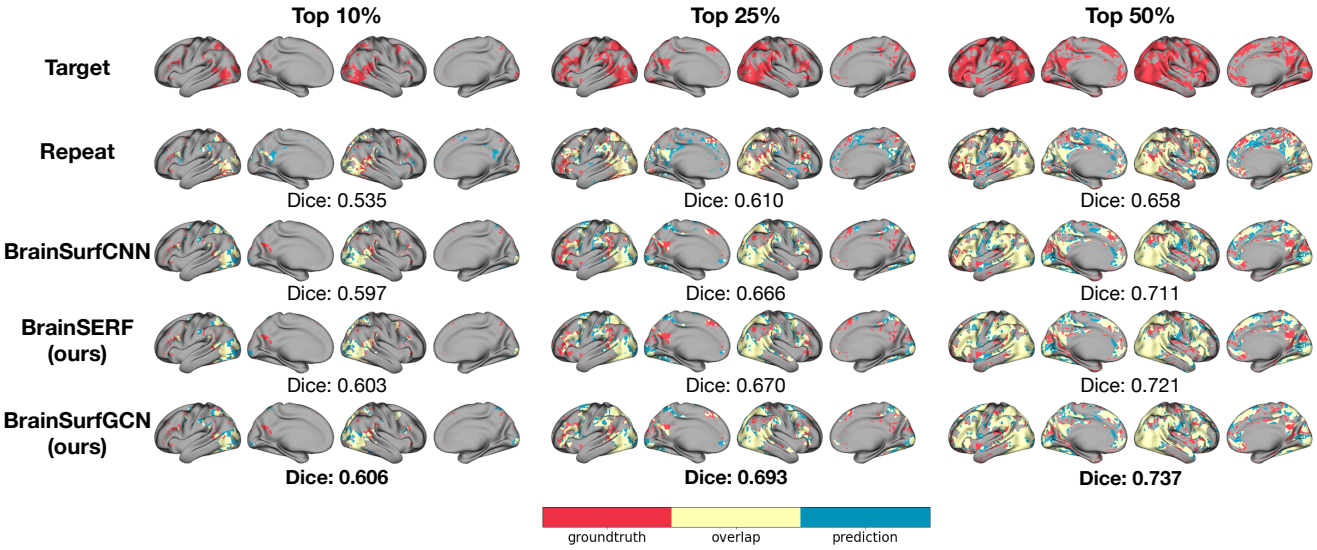

Figure 3: **Thresholded task activation maps for Social Cognition: Theory of Mind (subject 917255).** Activation maps are thresholded at three levels (top 10%, top 25%, and top 50%) to illustrate how effectively each model captures the spatial distribution of task-evoked responses. Ground truth (target), repeat scans, and predictions from BrainSurfCNN, BrainSERF, and BrainSurfGCN are shown for the right hemisphere, with colors indicating ground truth, model prediction, and overlap. Dice overlap scores are displayed beneath each row, quantifying prediction fidelity across thresholds. The highest Dice value within each column is bolded to highlight the best-performing models.

We computed Dice overlap scores across thresholds from 5% to 50% in 5 percent steps to characterize how prediction quality changes with varying activation selectivity. As shown in Figure 4, these curves provide a clearer view of model behavior across activation intensities. In several contrasts, model predictions slightly exceeded the repeat dataset, which represents the empirical noise ceiling. This pattern, also noted by Ngo et al., 2022, likely arises because repeat scans contain session-to-session variability and small spatial shifts in activation, lowering their measurable overlap. The smoother and less noisy nature of model predictions can therefore yield higher Dice scores without violating theoretical limits. All Dice scores for every final model across all task contrasts are provided in the Supplementary Figure 11.

To summarize performance across thresholds, we calculated the area under the Dice curve (Dice AUC), with a maximum possible value of 0.45 based on perfect overlap across the integration range. These values are shown in Figure 5 and provide a concise metric for comparing model performance across tasks and individuals. We also examined unthresholded predictions for baseline reference.

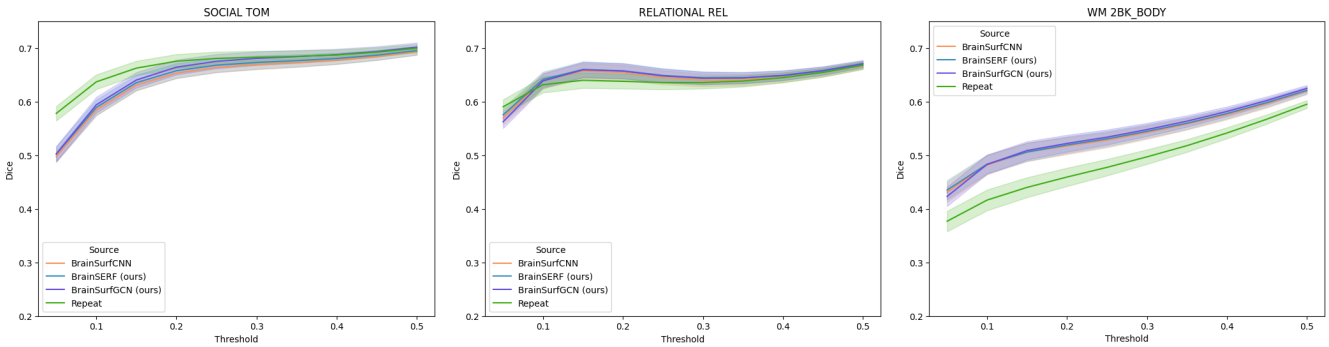

Figure 4: **Dice score plots for three selected tasks showing variations in performance across task contrasts.** Each of these plots shows a variation in model performance when compared to the repeat. The first shows asymptotic convergence to repeat contrast performance. The second shows similar performance between model predictions and repeat contrasts, and the third plot shows improved performance over the repeat contrasts. The error bars denote standard deviations in the dice scores across test subjects.

Finally, extended results, including per-task Dice AUC scores and additional visualizations, are available in Supplementary Tables 3 and 4 and Figure 12-17, providing a more granular view of how each model performs across contrasts and individuals.

2.2 Evaluation of Predictive Fidelity - Subject-Level Specificity

To evaluate the individual specificity of model predictions, we examined both the spatial correlation between predicted and ground-truth task activation maps and the subject identification accuracy. This analysis assesses whether the models capture unique, individualized features of brain function that distinguish one subject from another.

We first computed the spatial correlation difference: for each predicted contrast, we compared its correlation with the corresponding ground-truth contrast from the same subject versus those from different subjects. A high correlation difference indicates that the model more accurately captures idiosyncratic features of the target individual. As shown in Figure 6, all three models exhibit strong within-subject correlation relative to others, suggesting that they reliably capture individually localized patterns of task-evoked activity.

Next, we performed a subject identification task: for each predicted contrast, we computed its correlation with all ground-truth contrasts across the test set and evaluated whether the highest correlation matched the true subject. This classification-based metric offers a concrete measure of the identifiability of individuals from predicted brain maps. Average subject identification accuracies across task contrasts

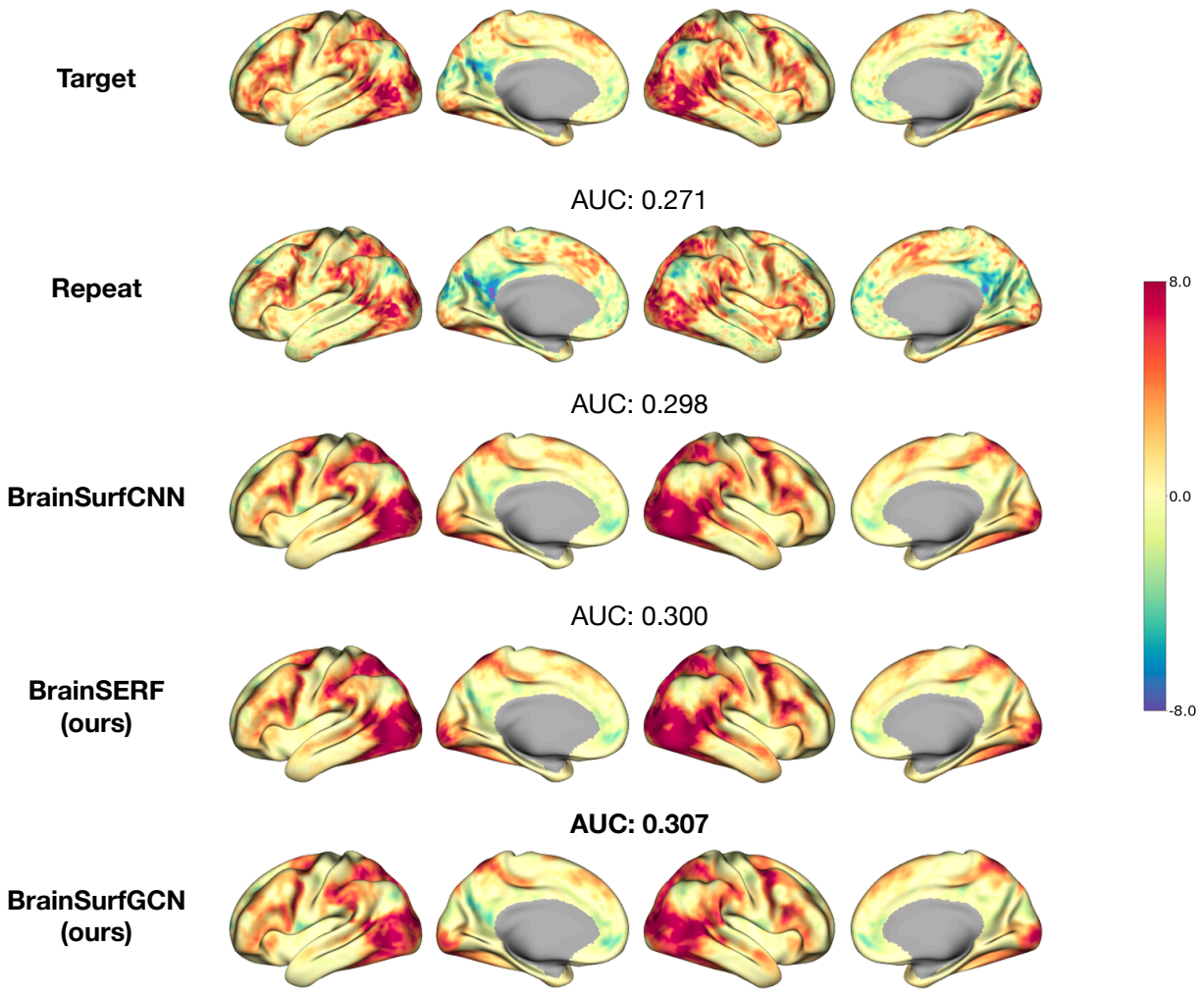

Figure 5: **Unthresholded task activation for Social Cognition: Theory of Mind** We compare ground truth, repeat, and model predicted activation for an individual (subject 917255). Model predictions are much smoother on the whole than the ground truth or repeat activations. Though smoother, the models perform higher in the Dice AUC score than the repeat contrasts.

are summarized in Table 1. We also explored a bagging approach by averaging predictions across models trained on different independent components (ICs). This ensemble-like strategy yielded modest improvements in subject identifiability, suggesting that aggregation may help stabilize model predictions across runs.

Interestingly, all models achieved similar accuracy levels, and in certain tasks (e.g., Language: Math and Language: Story), they even outperformed identifiability from repeat scans, i.e., actual task activations collected from the same subject during a separate session. These repeat contrasts serve as a theoretical upper bound, offering a benchmark for model performance. The average accuracy for the repeat contrasts was 92.9% across 47 task contrasts, highlighting the strong test-retest reliability of the original data.

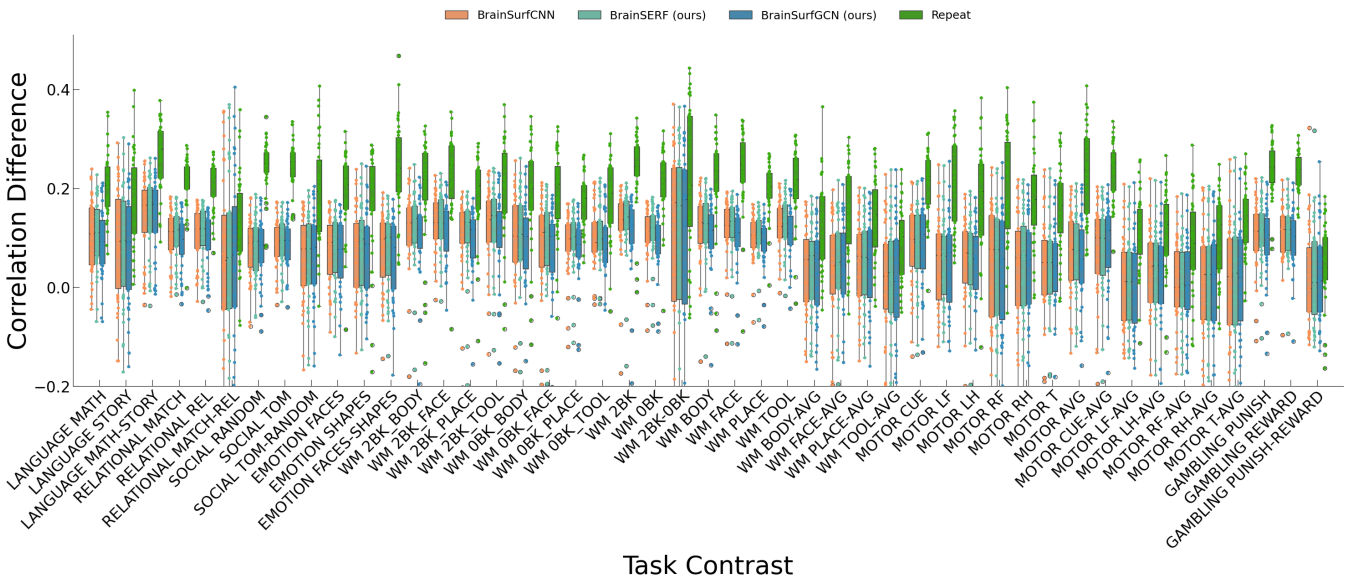

Figure 6: **Difference in spatial correlation of predicted contrast to ground truth contrast of the same subject versus the mean spatial correlation of predicted contrast to ground truth contrasts of all other test set subjects.** The correlation metric primarily measures the linear relationship between the model's predicted task contrasts and the empirically observed (ground truth) contrasts. In this analysis, we derive the metric displayed in the figure by first calculating the correlation between a prediction based on a single subject's resting-state data and the subject's ground truth task activation contrast. Subsequently, we subtract the mean correlation of the same prediction from the task activation contrasts of all other subjects. A high correlation difference value signifies that the spatial correlation of a given prediction is substantially higher when comparing the prediction to the ground truth contrast of the same subject, as opposed to the ground truth contrasts of different subjects. This indicates that the model proficiently captures the idiosyncratic features of each individual's task-related brain activity.

Table 1: Subject Identification Test Accuracy Across Varying ICs

Model	15 ICs	25 ICs	50 ICs	100 ICs	Avg
BrainSurfCNN	80.3	81.0	79.5	79.3	81.8
BrainSERF (ours)	**80.6**	**82.3**	80.4	79.8	**82.8**
BrainSurfGCN (ours)	79.5	79.5	**80.7**	**79.9**	82.4
repeat (Benchmark)			**92.9**		

Finally, we quantified the computational footprint of each model to highlight trade-offs between performance and efficiency. Table 2 reports the number of trainable parameters across architectures and ICs. Notably, BrainSurfGCN achieved substantial reductions in parameters. This is attributable to its mesh-based design, which allows parameter sharing across nodes and avoids the need to store full mesh gradients. Importantly, model size remained largely stable across different IC configurations, since

operations scale with fixed hidden-state sizes rather than the number of ICs directly.

Table 2: Number of Trainable Parameters for Each Model

No. Independent Components	BrainSurfCNN	BrainSERF	BrainSurfGCN
15 ICs	359,902	363,062	**23,742**
25 ICs	365,022	368,846	**24,382**
50 ICs	377,822	384,706	**25,982**
100 ICs	403,422	422,426	**29,182**

2.3 Mechanisms of Variability in Predictability

Having established that all three architectures can generate individualized task activation patterns, we next asked why prediction accuracy varies across individuals and contrasts. We focused on three potential sources of variability grounded in cognitive neuroscience: task engagement, resting-state data quality, and inter-subject variability in task-evoked activation.

2.3.1 Task Engagement Influences Predictability

We first examined whether behavioral task engagement modulates the success of rest-to-task prediction. If a participant performs poorly on a task, reflected in lower accuracy, slower reaction times, or greater behavioral variability, the neural representations associated with that task may be weaker or less consistently expressed, reducing the model's ability to reconstruct individual activation patterns.

For all contrasts with available behavioral data, we divided subjects into above-average and below-average performers and compared their prediction accuracies using the correlation-difference metric computed from BrainSurfGCN predictions. As shown in Figure 7, both the predicted contrasts and the empirical repeat contrasts showed significantly higher correlation-difference values for above-average performers (BrainSurfGCN: $p = 0.035$; repeat contrasts: $p = 0.018$, Mann-Whitney U Test).

At the single-task level, two contrasts, including Language: Math and Relational: Relational, showed significant group differences (uncorrected), with above-average performers exhibiting higher predictability. This pattern mirrors results from the repeat scans, where Relational: Relational also showed a significant difference. These analyses together suggest that task engagement is a meaningful determinant of predictability, and reduced behavioral performance may reflect reduced neural engagement or inconsistent cognitive strategies during the task. Because these analyses were exploratory and intended for hypothesis generation, we did not correct for multiple comparisons.

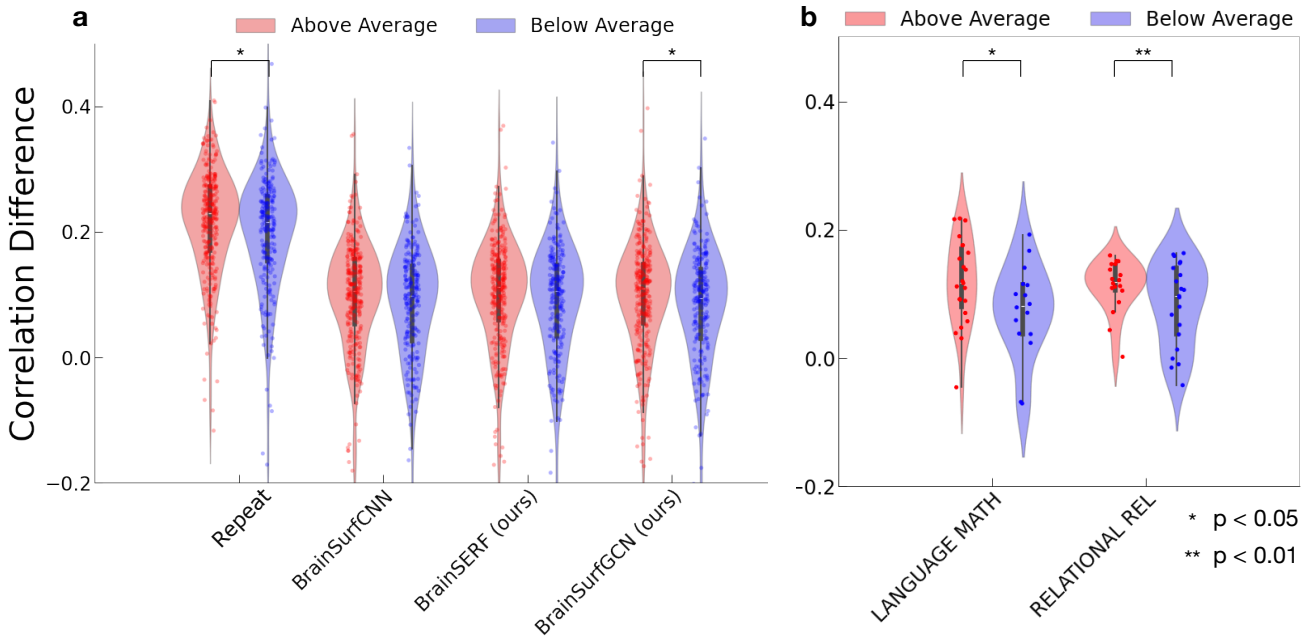

Figure 7: **Correlation difference for above-average and below-average task accuracy groups.** We computed the correlation difference with the ground truth test set for the repeat contrasts and each set of model predictions. Selecting only the values associated with tasks with a measured accuracy metric for task performance. We split these sets into above-average and below-average performers. **(a)** For the four sets of contrasts, we found that the repeat set and the predictions made by BrainSurfGCN indicate a statistical significance between the two groups in a Mann-Whitney U test. **(b)** The statistical difference in BrainSurfGCN prediction performance was driven by two specific tasks: 'Language: Math' and 'Relational: Relational.'

2.3.2 Resting-State Signal Quality Shapes Prediction Error

We next examined whether resting-state data quality constrains prediction performance. Specifically, we evaluated the relationship between regional temporal signal-to-noise ratio (tSNR) and prediction error, measured as MSE.

We first computed the average tSNR across subjects for each cortical vertex (Figure 8a). Regions with lower tSNR, such as orbitofrontal cortex and temporal poles, are known to be susceptible to susceptibility artifacts and physiological noise. We then compared the spatial distribution of low-tSNR regions with areas showing high prediction error, computed by comparing model predictions with repeat contrasts.

Using Dice AUC to quantify overlap, we found that contrasts with the highest error-to-tSNR overlap were predominantly delta contrasts (e.g., Language: Math > Story, Working Memory: Face > Average, Motor: Tongue > Average). All of these showed Dice AUC values above 0.13. These results indicate that rest-to-task prediction performance is partly limited by the intrinsic noise characteristics of resting-state

data, especially for contrasts requiring subtle condition-specific distinctions.

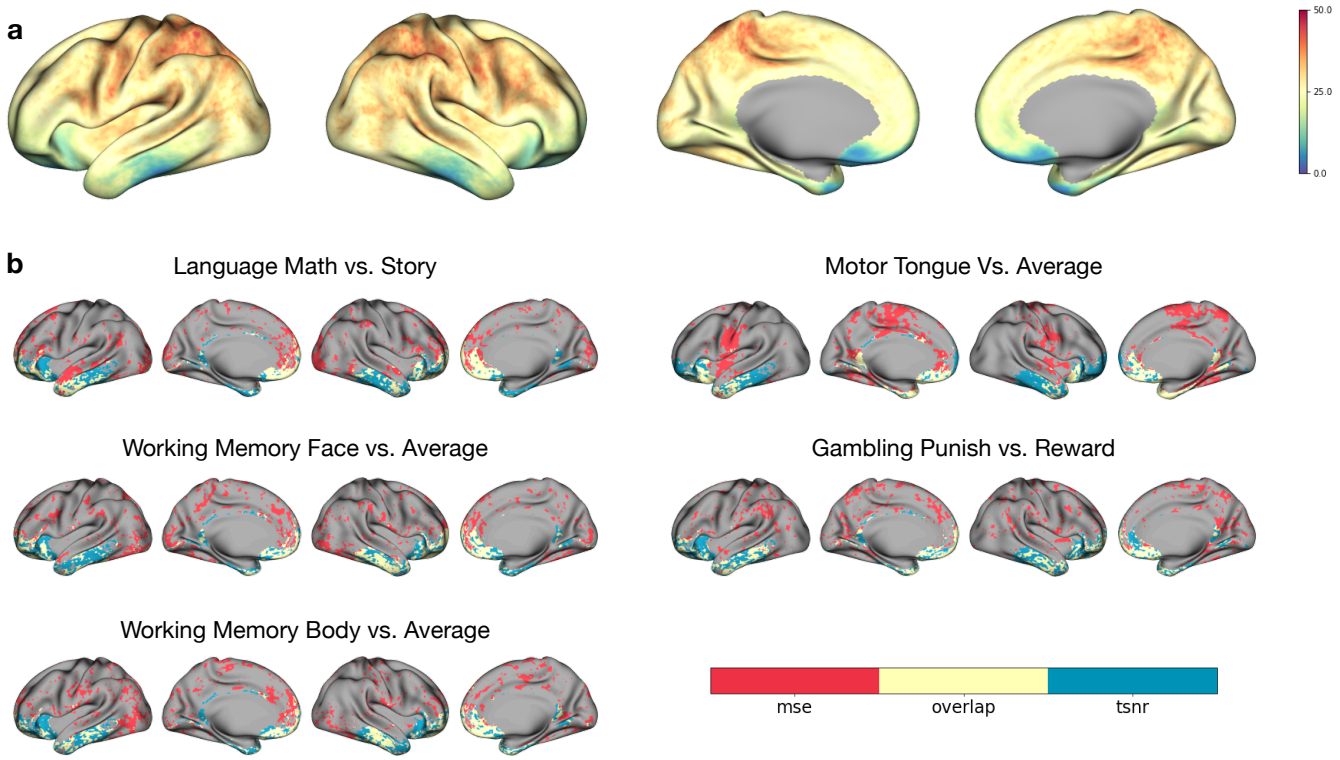

Figure 8: **Relationship between lower tSNR and higher MSE of BrainSurfGCN's predictions on the test set. (a)** We computed the average tSNR for each vertex on the mesh across the test set and plotted it on the cortex. **(b)** The top 5 contrasts with the most overlap between the MSE and tSNR using the Dice AUC metric. We measured the overlap between the regions of the highest MSE and lowest tSNR. In this depiction, these maps are shown with a 25% threshold applied to both, allowing us to analyze the specific regions associated with the highest model error.

2.3.3 Variability in Cortical Activation Across Subjects and Contrasts

Finally, we asked whether inter-subject variability in task-evoked activation constrains predictability. Prior studies have shown that contrasts involving subtle cognitive differences, particularly delta contrasts, exhibit less spatial consistency across individuals than single-condition contrasts (Seghier and Price, 2016).

To quantify this, we generated activation-frequency maps for each contrast by computing, for each subject, the top 10% of activated vertices and then calculating the proportion of subjects showing activation at each vertex. These maps were computed for the empirical data (test and repeat) and for predictions from all three models.

As illustrated in Figure 9, delta contrasts showed markedly greater spatial variability than single-condition contrasts. Even the repeat contrasts showed low spatial correlation for delta contrasts (e.g., Relational Match – Relational: $r = 0.626$), underscoring their low test–retest reliability. In contrast, single-condition contrasts such as Relational: Relational and Relational: Match showed high repeatability ($r \approx 0.97$), reflecting stable and stereotyped activation patterns.

The models reproduced these broad patterns: predictions were smoother and more spatially constrained than the empirical data and showed reduced correspondence for delta contrasts relative to single-condition contrasts. This suggests that high inter-subject variability, combined with lower reliability of delta contrasts, limits the upper bound of predictability for these contrasts, even for the empirical repeat maps.

Supplementary Tables 5 and 5 provide full spatial correlation matrices, further highlighting the dominant influence of the repeat-derived noise ceiling on contrast-level predictability.

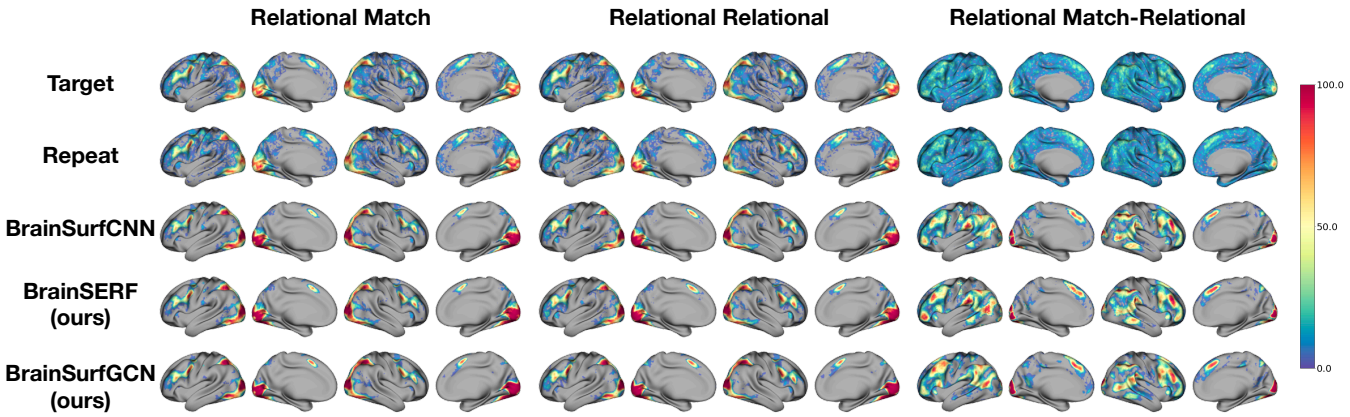

Figure 9: **Cortical frequency plots associated with relational task contrasts.** For each task contrast, we threshold at 10% and count the number of subjects that show activation on the vertex. This gives us a vertex value between 0 and 100, indicating the percentage of subjects with this vertex in the top 10% of activated vertices. Delta contrast maps indicating variability across the population in the activation maps show much more spatial variation compared to both the empirical data and the model-predicted contrasts

Together, these analyses reveal that variability in rest-to-task prediction arises from a combination of cognitive factors (task engagement), technical factors (resting-state data quality), and neurobiological factors (inter-subject variability in cortical recruitment). These findings emphasize that prediction performance is not solely determined by model architecture, but is fundamentally shaped by the stability and reliability of the underlying neural signals.

3 Discussion

In this work, we reproduced the predictive performance of BrainSurfCNN and introduced two novel variants, including BrainSERF, which integrates a biologically inspired SE attention mechanism, and BrainSurfGCN, which implements a graph-based message-passing framework. These models were designed in response to two central questions raised in the introduction: (1) Can deep learning models of resting-state to task prediction be made more biologically grounded? And (2) What are the key drivers of variability in individual-level prediction performance?

3.1 Model innovations: biologically inspired and efficient architectures

BrainSERF incorporates a channel-wise attention mechanism (SE block) that allows the model to adaptively scale the contribution of each IC map. This mimics a form of neural gain control, potentially improving biological plausibility by enabling the model to emphasize or suppress distributed sources, analogous to dynamic reweighting in cortical circuits. While performance gains were modest, BrainSERF consistently achieved higher subject identification accuracy compared to the baseline, suggesting value in incorporating inductive biases drawn from biology.

In contrast, BrainSurfGCN focused on computational efficiency, achieving more than 15 times reduction in trainable parameters without sacrificing performance. BrainSurfCNN and BrainSERF apply spherical convolution using parameterized differential operators that rely on four mesh representations: the original mesh, gradients along the x and y surface directions, and the Laplacian. While effective, this approach increases memory and compute demands by requiring multiple large tensors to be stored and processed at each layer. In contrast, our new BrainSurfGCN model uses a graph-based message-passing framework that represents the cortical surface as a node-edge graph and updates vertex features via a shared learnable weight matrix Θ. This design eliminates the need to explicitly compute or store differential operators, while still capturing local one-hop information propagation like spherical convolutions. The result is a dramatic reduction in model size (from $\sim$100MB to $<$1MB) and training time (from $\sim$26 to $\sim$6 hours), without sacrificing predictive performance. This efficiency makes BrainSurfGCN well-suited for resource-constrained environments, rapid prototyping, and potential clinical deployment, where model size and training speed are critical considerations.

3.2 Understanding variability in model performance

Our second line of investigation focused on understanding why model performance varies across task contrasts and individuals. We observed that prediction accuracy was influenced by both task engagement and data quality.

First, we found preliminary evidence that poorer model predictions aligned with lower task accuracy and slower reaction times, particularly in tasks requiring sustained attention or complex reasoning. This relationship between poor prediction performance and poor task performance has also been shown by Gonzalez-Castillo et al., 2015, where the prediction of cognitive states from functional connectivity was impaired by a possible loss of concentration or awareness. We present preliminary evidence for our hypothesis; rigorous future work is needed to better account for task performance during prediction.

Second, we investigated why delta contrast maps (i.e., condition A vs. condition B) yielded lower prediction and test-retest reliability compared to contrasts versus baseline. An exploratory analysis of repeat error and tSNR maps revealed that delta contrasts showed greater overlap between signal dropout regions (low tSNR) and areas of high prediction error, particularly in contrasts involving subtle task manipulations (e.g., Math vs. Story, Punish vs. Reward). This suggests that delta contrasts may reflect noisier or less stable signal differences, challenging their utility as prediction targets. Future studies should consider modeling task contrasts with varying levels of reliability separately or weighting voxels based on tSNR-informed confidence.

3.3 Limitations and future direction

Our study has several limitations that open avenues for future research: (1) Architectural scope: While BrainSERF and BrainSurfGCN introduced biologically and computationally motivated design changes, the performance gains were incremental. New architectural innovations, such as equivariant GNNs, spectral attention, or hybrid encoding-decoding frameworks (Kondor, 2025), may yield greater improvements in both accuracy and interpretability. (2) Generalization across mesh structures: BrainSurfGCN was evaluated on a fixed cortical mesh; however, its flexibility to operate on graphs of varying resolution remains untested. Leveraging coarser meshes, as done in the original BrainSurfCNN, may offer further speedups with minimal performance loss. (3) Zero-shot task generalization: Our models were trained to predict 47 fixed task contrasts. Extending these models to unseen tasks remains an open challenge and may require task embedding approaches or transfer learning on task instructions. (4) Group IC derivation: Resting-state features were constructed using group-level IC maps provided by HCP, which may include some overlap with test subjects. Although this risk is minimal given the dataset size, future work should

evaluate whether using subject-specific or held-out group ICs changes prediction fidelity. (5) Reliability of delta contrasts: As highlighted, delta contrasts appear inherently noisier and harder to predict, both for our models and repeat scans. This raises questions about their interpretability and suggests a need for revised preprocessing, weighting schemes, or alternate contrast formulations.

3.4 Toward biologically grounded predictive modeling

Together, our findings underscore the importance of modeling choices that align with biological principles, such as attention-based feature modulation and spatially constrained graph computations, while also highlighting the practical value of efficient architectures for large-scale neuroimaging applications. Both BrainSERF and BrainSurfGCN demonstrate that resting-state connectivity contains latent signatures of task-evoked activity, and that carefully designed models can extract these signals in a subject-specific manner.

Moving forward, linking these predictive representations to behavior, trait variability, and clinical outcomes will be crucial for translating resting-state models into real-world neuroscience and psychiatry. Additionally, better handling of data quality, contrast selection, and task variability will improve the reliability and robustness of future models.

4 Method

4.1 Dataset and Preprocessing

We used de-identified, publically available data from the HCP dataset (Van Essen et al., 2013) to train our models in the same way as BrainSurfCNN. The study was approved by the Stanford University Institutional Review Board.

Specifically, as input to our models, we used the FIX-cleaned, 3T rsfMRI data acquired in four 14.4-minute runs, each with 1200 time points per session per subject. Data acquisition and preprocessing were done according to (Barch et al., 2013; Glasser et al., 2013; Saggar et al., 2018; Smith et al., 2013), where rsfMRI preprocessing includes highpass filtering, MELODIC ICA, and FIX for component classification. We employ the data augmentation scheme as proposed in (Ngo et al., 2022), in which samples of the resting state data are drawn from all four recordings of resting state data. This aggregation of recording data amounts to 4,800 time points split into eight samples of 600 time points from which the functional

connectivity is determined. These eight samples allow us to extend the training set of our data by uniformly sampling these eight chunks in each training epoch.

The acquisition and preprocessing methods for the HCP dataset have been described elsewhere in (Barch et al., 2013; Glasser et al., 2013; Smith et al., 2013). Group-level parcellations derived from spatial ICA were also released by the HCP, and we examined the performance of our models trained on various component parcellations for computing the functional connectomes. Since the authors in (Ngo et al., 2022) explored data with 50 ICs, we thought a natural extension would be to explore various amounts of independent components around that order of magnitude. We create new training datasets derived from 15, 25, 50, and 100 independent components using these spatial components. For our target data, HCP's tfMRI data spans seven task domains, and, following Tavor et al., 2016 and Ngo et al., 2022, we obtained 47 unique contrasts after excluding redundant task contrasts.

Additionally, we only included subjects with all four resting-state runs and all 47 task contrasts, which gave us 919 subjects for training, of which five subjects were used for a validation set (Ngo et al., 2022). The use of the validation set prevents overfitting our models during the training process. Additionally, we also carried out a 5-fold cross-validation training regime and found similar performance in our models, as seen in Supplementary Figure 10.

Our test set comprised subjects ($n = 39$) from the original HCP dataset who each participated in two separate visits under identical acquisition protocols, consistent with Ngo et al., 2022. We refer to the contrasts from the first visit as the "test set" and those from the second visit as the "repeat dataset," adopting BrainSurfCNN terminology. The test set participants do not overlap with those in our training set. We selected these 39 participants because they each provide two sets of contrasts, enabling a direct comparison of our model's predictions against a repeat contrast from the same individual in a different session. We used the repeat dataset as an optimal predictor of ground-truth task activation (i.e., non-model-based or model-free). A contrast taken from the same subject in a different session represents the best possible model for individual variability, providing a theoretical upper bound, or noise ceiling, for model performance. Importantly, no resting-state data were collected for the repeat (second) scan, so it is not used as input for prediction.

In the proposed deep learning framework, the input data are structured as multi-channel fsLR polyhedral meshes, each consisting of 32,492 vertices, to accommodate the high-dimensional spatial topology of cerebral cortex representations. Here, "fsLR" denotes FreeSurfer-derived cortical surface meshes aligned to a standardized template (fsLR template) across the left and right hemispheres. Each channel within these meshes is dedicated to an independent component derived from functional MRI data preprocessing, serving as a distinct feature for the model's input layer. For instance, in a scenario where the model is

configured to analyze 50 independent components, the total input dimensionality would amount to 100 channels, reflecting the bilateral nature of cerebral anatomy with a separate mesh for each hemisphere. The model's output mirrors the input's structural format, generating a polyhedral mesh with 32,492 vertices. However, in the output mesh, each vertex's value across the channels indicates the predicted task contrast for the corresponding location in the brain hemisphere. Given that the study focuses on predicting 47 distinct task contrasts, the model's output dimensionality is adjusted to feature 94 channels, with each pair of channels representing the predicted contrasts for the left and right hemispheres, respectively.

4.2 Models

4.2.1 BrainSurfCNN

The development of BrainSurfCNN integrates the foundational principles of a U-Net architecture (Milletari et al., 2016; Ronneberger et al., 2015), a prevalent framework in medical image segmentation, with advanced mesh convolution techniques to facilitate the analysis of brain surface data. This innovative approach is significantly influenced by the pioneering work on convolution for spherical meshes called UGSCNN, as detailed by C. Jiang et al., 2019.

The architecture adopts a hierarchical encoder-decoder structure composed of downsampling and upsampling pathways, which leverage mesh convolutional operations to effectively capture spatial features across multiple resolutions. Importantly, it incorporates a Residual Pooling Block (implemented as `ResPoolBlock`), which combines residual connections, mesh convolutions, and pooling operations to enhance robust multi-scale feature learning. The primary objective of BrainSurfCNN (Ngo et al., 2022) is to leverage the spatial hierarchies inherent in polyhedral meshes to predict task-related activation patterns, represented as contrast maps, from resting-state IC maps.

In our study, we replicated the BrainSurfCNN model using the source code made publicly available by the original authors on GitHub [1]. This replication process was not merely an exercise in model reconstruction but a deliberate effort to validate and extend the model's application. A cornerstone of our analysis involved a systematic examination of BrainSurfCNN's performance across a spectrum of resting-state scans characterized by varying numbers of independent components. Specifically, we conducted comparative analyses using datasets composed of 15, 25, 50, and 100 independent components to assess the model's robustness and scalability with varying levels of information. This approach allowed us to explore the impact of independent components on the model's ability to predict task contrasts accurately

[1] GitHub Link: https://github.com/ngohgia/brain-surf-cnn

on both an individual and a group level of analysis.

4.2.2 BrainSERF

The SE attention mechanism has emerged as a pivotal innovation in the field of computer vision, significantly enhancing the representational power of convolutional neural networks by enabling channel-wise feature recalibration (Hu et al., 2018). This technique introduces trainable scaling coefficients that adaptively adjust the weighting of each channel in the network's feature maps, thereby allowing the model to emphasize informative features and suppress less useful ones dynamically. The SE attention mechanism operates by first 'squeezing' global spatial information into a channel descriptor through global average pooling, followed by 'excitation' operations, fully connected layers that capture channel-wise dependencies. Given two weight matrices W_{sq} and W_{ex}, we perform the squeeze operation by globally pooling each mesh so that our representation of input X transforms from shape $\mathbb{R}^{32,492\times C}$ to $\mathbb{R}^{1\times C}$ where C represents the number of hidden channels. Next, we define a ratio r such that the matrix W_{sq} transforms the data from $\mathbb{R}^{1\times C}$ to $\mathbb{R}^{1\times \frac{C}{r}}$. We finish the 'squeezing' with a Tanh activation to capture negative values to get X'. To perform the 'excitation' step, we multiply the transformed X' by W_{ex} to expand the channel-wise axis back to a matrix in $\mathbb{R}^{1\times C}$ to get S. We applied a Tanh activation function again to capture negative values. This matrix S represents a scaling on the channel-wise axis. We channel-wise multiply $X \in \mathbb{R}^{32,492\times C}$ by S to achieve the output X_{scaled}, a channel-wise rescaled form of the input X. The intuition with this mechanism is to periodically rescale the hidden representation of the mesh channels and highlight or suppress entire latent activation patterns.

These operations produce scaling coefficients applied to the original feature maps, effectively allowing the network to perform self-attention across channels. In our network architecture, we implement SE attention before the mesh convolutional layers during the coarsening stage of the mesh, inspired by the work in Wang et al., 2021. The network additionally incorporates the Residual Pooling Block, as described in BrainSurfCNN, to further enhance hierarchical feature learning through residual connections and mesh-based pooling.

Furthermore, we have innovated beyond traditional activation functions by transitioning from Rectified Linear Unit (ReLU) activations to Hyperbolic Tangent (Tanh) activations throughout our network. This adjustment is primarily motivated by the need to accommodate negative values within the SE attention mechanism and the predicted task contrasts. The Tanh activation function enables our network to capture a broader spectrum of feature dynamics, including both positive and negative activations that are present in the task contrast maps. Additional training hyperparameters, command-line arguments, and example usage scripts are provided on our GitHub repository: https://github.com/braindynamicslab/

dl-task-contrast-prediction

4.2.3 BrainSurfGCN

In predicting task contrast activation maps from rsfMRI, employing graph neural networks offers several compelling advantages that align with the intrinsic properties of brain data and the objectives of neuroimaging analysisZhang et al., 2021. The human brain can be conceptualized as a complex network, with nodes representing different brain regions and edges representing functional or structural connections between these regions. This network-centric view is inherently compatible with the structure of GNNs. We define a graph below.

Definition 1. *An undirected graph G can be defined as $G = (\mathcal{V}, \mathcal{E})$, where $v \in \mathcal{V}$ represents a node with feature size in $\mathbb{R}^C$ and $e \in \mathcal{E}$ represents an undirected edge between two vertices v_i and v_j.*

We leverage the spherical mesh used in Ngo et al., 2022 as the graph's structure for each input, such that $|\mathcal{V}| = 32{,}492$ mesh. Thus, our input data has two parts. The node-wise representation, $\mathcal{V}$, is represented as a matrix in $\mathbb{R}^{|\mathcal{V}| \times \mathcal{C}}$, where each node $v \in \mathcal{V}$ has C learnable parameters. The edge-wise representation is a list of tuples of vertices that are connected such that edge $e \in \mathcal{E}$ connecting vertices v_i and v_j is represented as (i, j), and $\mathcal{E} \in \mathbb{Z}^{2 \times |\mathcal{E}|}$.

We build BrainSurfGCN upon the work of Kipf and Welling, 2016. The architecture we use for this study builds upon the implementation of the Graph Convolution Layer seen in PyTorch Geometric (Fey and Lenssen, 2019), and we build a network composed of what we call the BDLayer. The structure of the BDLayer includes a GCN Layer, a LeakyReLU activation, and a BatchNorm layer. The BatchNorm layer accelerates learning by reducing internal covariate shift (Ioffe and Szegedy, 2015) through the recentering of features using the formula:

$$y = \frac{x - \mathbb{E}[x]}{\sqrt{\mathsf{Var}[x]}} * \gamma + \beta \tag{1}$$

where γ and β are learnable parameters, this layer acts along a node's feature-wise axis. The GCN Layer defines the update of a node's feature space from X to X' as

$$X' = \hat{D}^{-1/2} \hat{A} \hat{D}^{-1/2} X \Theta \tag{2}$$

where $\hat{A} = A + I$ denotes an adjacency matrix with added self-loops, $\hat{D}$ denotes the diagonal degree

matrix of $\hat{A}$, and Θ denotes the learnable weight matrix. The adjacency matrix A is formed such that $A_{ij} = 1$ if $e_{ij} \in \mathcal{E}$, and $A_{ij} = 0$ if $e_{ij} \notin \mathcal{E}$. This operation allows us to pass information between vertices on the mesh at a distance of 1 hop per layer. To ensure that node information is disseminated across the entire graph, we use 8 BDLayers. Additionally, we believe that spatial information is important in predicting the activation on the mesh, so we included the XYZ coordinates of each vertex as 3 additional channels for our input data. These coordinates represent the locations of the vertices of the polyhedral sphere centered on the origin (0,0,0) with a radius of 1. Work in Liu et al., 2018 also suggests that including spatial information can improve convolution operations. With a resting-state map of 50 ICs per hemisphere, our input feature space lies in $\mathbb{R}^{32,492\times(2*50+3)}$. Figure 1 shows our architecture from a high-level perspective, complete with the input and output feature spaces.

4.3 Training Details

All the models trained in these experiments used the same set of hyperparameters and training regimes. Models were developed and trained in PyTorch (Paszke et al., 2019), and we used Adam (Kingma and Ba, 2014) for optimization. We utilized a learning rate of 0.001 and trained the models for 50 epochs using MSE loss to promote task contrast reconstruction. We used the model with the best-predicted correlation on the validation set at the end of 50 epochs for evaluation. Because we trained several types of models, we compared performance after the first round of training to decide if training should continue.

After performing some preliminary sanity checks (e.g., correlation between prediction and ground truth), we continued with the next phase of training. In this second step, we used the trained model to compute and average the MSE across the training set for same-subject reconstruction loss, α, and across-subject reconstruction loss, γ, to be used as hyperparameters in the next step of training that used a reconstructive-contrastive (RC) loss function proposed by Ngo et al., 2022. Given a mini-batch of N samples, $B = \{\hat{x}_1..\hat{x}_N\}$, in which $\hat{x}_i$ is the target multi-channel contrast image of subject i, the RC loss function is defined as

$$L_R = \frac{1}{N}\sum_{i=1}^{N}(x_i - \hat{x}_i)^2 \text{ and } L_C = \frac{1}{\frac{1}{2}(N^2 - N)}\sum_{\hat{x}_j \in B, j \neq i}^{N}(x_i - \hat{x}_j)^2 \tag{3}$$

$$L_{RC} = [L_R - \alpha]_+ + [L_R - L_C + \gamma]_+. \tag{4}$$

Using the RC loss described in Equation 4, we train the model again for 50 epochs. Before training, we initialize values for α and γ from the losses computed over the training set. These values are updated every 10 epochs such that:

$$\alpha_t = \frac{1}{2}\alpha_{t-1} \text{ and } \gamma_t = 2\gamma_{t-1} \tag{5}$$

until α reaches a minimum of 1 and γ reaches a maximum of 10. Using this RC loss function, we encouraged the model to differentiate task contrast reconstructions between subjects, thereby capturing individual differences.

4.4 Evaluation

To evaluate the efficacy of our model in predicting task contrast maps from resting-state fMRI data, we employ three distinct metrics: correlation, Dice coefficient (also referred to as F1 score), and subject identification accuracy. These metrics offer a comprehensive assessment of the model's performance, each addressing different aspects of prediction quality.

The correlation metric primarily measures the linear relationship between the model's predicted task contrasts and the empirically observed (ground truth) contrasts. This metric was chosen for replication to compare the results from (Ngo et al., 2022). By computing Pearson's correlation coefficient, we quantify how the predicted values co-vary with the actual task contrasts across the mesh's vertices. A high correlation coefficient indicates that the model effectively captures the spatial distribution of neural activations associated with specific tasks, mirroring the patterns observed in the ground truth data. A crucial aspect of the performance is ensuring that within-subject covariance is more similar than inter-subject covariance.

The Dice coefficient enhances our evaluation by measuring spatial overlap between the predicted and ground truth task contrast activation patterns. To calculate this metric, we first apply a series of thresholds to the activation values at each vertex, classifying them as 'activated' or 'non-activated.' For each threshold, the Dice coefficient is computed as the harmonic mean of precision and recall between the two binary sets of activated vertices. Thus, the specificity of predicted activation on the contrasts is computed over a range of threshold values, ensuring predicted regions with higher activation (from a higher threshold) overlap with the higher activations in the ground truth contrasts. Thereby, we capture another measurement, in addition to correlation, related to both the spatial distribution and the relative magnitude of the cortical activation.

To assess the model's capability in capturing individual-specific features within the task contrasts, we use an evaluation method introduced by Ngo et al., 2022 based on subject identification accuracy. This approach involves computing the correlation between each predicted task contrast and all available ground truth contrasts across subjects for the same task. The identification of the subject is deemed correct if the highest correlation corresponds to the correct subject's ground truth contrast compared to those for other subjects. This metric reflects the model's sensitivity to individual differences in brain activation patterns, highlighting its potential for personalized neuroimaging analysis. We computed this accuracy for each quantity of ICs used in training and also included a bagged average across the models trained on various ICs to give us an average prediction.

Data and Code Availability

All code used for this article is made publicly available on GitHub at https://github.com/braindynamicslab/dl-task-contrast-prediction

All data used in this article is publicly available through the Human Connectome Project.

References

Barch, D. M., Burgess, G. C., Harms, M. P., Petersen, S. E., Schlaggar, B. L., Corbetta, M., Glasser, M. F., Curtiss, S., Dixit, S., Feldt, C., et al. (2013). Function in the human connectome: Task-fmri and individual differences in behavior. *Neuroimage*, *80*, 169–189.

Chow, W. W., Seghouane, A.-K., & Seghier, M. L. (2025). A statistical characterization of dynamic brain functional connectivity. *Human Brain Mapping*, *46*(2), e70145.

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.

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of rdoc. *BMC medicine*, *11*(1), 126.

Fey, M., & Lenssen, J. E. (2019). Fast graph representation learning with PyTorch Geometric. *ICLR Workshop on Representation Learning on Graphs and Manifolds*.

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

Gonzalez-Castillo, J., Hoy, C. W., Handwerker, D. A., Robinson, M. E., Buchanan, L. C., Saad, Z. S., & Bandettini, P. A. (2015). Tracking ongoing cognition in individuals using brief, whole-brain functional connectivity patterns. *Proceedings of the National Academy of Sciences*, *112*(28), 8762–8767. https://doi.org/10.1073/pnas.1501242112

Hearne, L. J., Mill, R. D., Keane, B. P., Repovš, G., Anticevic, A., & Cole, M. W. (2021). Activity flow underlying abnormalities in brain activations and cognition in schizophrenia. *Science Advances*, *7*(29), eabf2513. https://doi.org/10.1126/sciadv.abf2513

Hu, J., Shen, L., & Sun, G. (2018). Squeeze-and-excitation networks. *Proceedings of the IEEE conference on computer vision and pattern recognition*, 7132–7141.

Insel, T. R. (2014). The nimh research domain criteria (rdoc) project: Precision medicine for psychiatry. *American journal of psychiatry*, *171*(4), 395–397.

Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *International conference on machine learning*, 448–456.

Jiang, C., Huang, J., Kashinath, K., Marcus, P., Niessner, M., et al. (2019). Spherical cnns on unstructured grids. *arXiv preprint arXiv:1901.02039*.

Jiang, R., Zuo, N., Ford, J. M., Qi, S., Zhi, D., Zhuo, C., Xu, Y., Fu, Z., Bustillo, J., Turner, J. A., Calhoun, V. D., & Sui, J. (2020). Task-induced brain connectivity promotes the detection of individual differences in brain-behavior relationships. *NeuroImage*, *207*, 116370. https://doi.org/https://doi.org/10.1016/j.neuroimage.2019.116370

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kipf, T. N., & Welling, M. (2016). Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907*.

Kondor, R. (2025). The principles behind equivariant neural networks for physics and chemistry. *Proceedings of the National Academy of Sciences*, *122*(41), e2415656122.

Liu, R., Lehman, J., Molino, P., Petroski Such, F., Frank, E., Sergeev, A., & Yosinski, J. (2018). An intriguing failing of convolutional neural networks and the coordconv solution. *Advances in neural information processing systems*, *31*.

Milletari, F., Navab, N., & Ahmadi, S.-A. (2016). V-net: Fully convolutional neural networks for volumetric medical image segmentation. *2016 fourth international conference on 3D vision (3DV)*, 565–571.

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. https://doi.org/https://doi.org/10.1016/j.neuroimage.2021.118849

Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., et al. (2019). Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, *32*.

Quah, S., Madsen, S., Pirzada, S., Jo, B., Uddin, L. Q., Mumford, J. A., Barch, D. M., Gotlib, I. H., Fair, D. A., Poldrack, R. A., et al. (2025). Refining rdoc using individual-level task fmri factor models reveals reproducible brain-wide motifs. *bioRxiv*, 2025–10.

Quah, S., Jo, B., Geniesse, C., Uddin, L., Mumford, J., Barch, D., Fair, D., Gotlib, I., Poldrack, R., & Saggar, M. (2025). A data-driven latent variable approach to validating the research domain criteria framework. *Nature communications*, *16*(1), 830.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 234–241.

Saggar, M., Sporns, O., Gonzalez-Castillo, J., Bandettini, P. A., Carlsson, G., Glover, G., & Reiss, A. L. (2018). Towards a new approach to reveal dynamical organization of the brain using topological data analysis. *Nature communications*, *9*(1), 1399.

Savage, H. S., Mulders, P. C., Van Eijndhoven, P. F., Van Oort, J., Tendolkar, I., Vrijsen, J. N., Beckmann, C. F., & Marquand, A. F. (2024). Dissecting task-based fmri activity using normative modelling: An application to the emotional face matching task. *Communications biology*, *7*(1), 888.

Seghier, M. L., & Price, C. J. (2016). Visualising inter-subject variability in fmri using threshold-weighted overlap maps. *Scientific reports*, *6*(1), 20170.

Serin, E., Ritter, K., Schumann, G., Banaschewski, T., Marquand, A., & Walter, H. (2025). Generating synthetic task-based brain fingerprints for population neuroscience using deep learning. *Communications Biology*, *8*(1), 1572.

Smith, S. M., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D. A., Griffanti, L., Harms, M. P., et al. (2013). Resting-state fmri in the human connectome project. *Neuroimage*, *80*, 144–168.

Smith, S. M., Fox, P. T., Miller, K. L., Glahn, D. C., Fox, P. M., Mackay, C. E., Filippini, N., Watkins, K. E., Toro, R., Laird, A. R., et al. (2009). Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the national academy of sciences*, *106*(31), 13040–13045.

Smith-Collins, A. P., Luyt, K., Heep, A., & Kauppinen, R. A. (2015). High frequency functional brain networks in neonates revealed by rapid acquisition resting state fmri. *Human brain mapping*, *36*(7), 2483–2494.

Sui, J., Adali, T., Pearlson, G. D., & Calhoun, V. D. (2009). An ica-based method for the identification of optimal fmri features and components using combined group-discriminative techniques. *Neuroimage*, *46*(1), 73–86.

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.

Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., & Ugurbil, K. (2013). The wu-minn human connectome project: An overview [Mapping the Connectome]. *NeuroImage*, *80*, 62–79. https://doi.org/https://doi.org/10.1016/j.neuroimage.2013.05.041

Wang, J., Lv, P., Wang, H., & Shi, C. (2021). Sar-u-net: Squeeze-and-excitation block and atrous spatial pyramid pooling based residual u-net for automatic liver segmentation in computed tomography. *Computer Methods and Programs in Biomedicine*, *208*, 106268.

Zhang, Y., Tetrel, L., Thirion, B., & Bellec, P. (2021). Functional annotation of human cognitive states using deep graph convolution. *NeuroImage*, *231*, 117847.

Acknowledgments

This work was supported by an NIH R01 MH127608 and an MCHRI Faculty Scholar Award to Manish Saggar. Data were provided 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.

Author Contributions

Soren Madsen: Conceptualization, methodology, software, validation, formal analysis, investigation, data curation, writing of original draft, reviewing and editing, and visualization. Young-Eun Lee: Methodology, software, validation, reviewing and editing, and visualization. Shaun K. L. Quah: Reviewing and editing. Lucina Q. Uddin: Conceptualization, reviewing and editing, and funding acquisition. Jeanette A. Mumford: Conceptualization, reviewing and editing, and funding acquisition. Deanna M. Barch: Conceptualization, reviewing and editing, and funding acquisition. Damien A. Fair: Conceptualization, reviewing and editing, and funding acquisition. Ian H. Gotlib: Conceptualization, reviewing and editing, and funding acquisition. Russell A. Poldrack: Conceptualization, reviewing and editing, supervision, and funding acquisition. Amy Kuceyeski: Conceptualization, validation, reviewing and editing, supervi-

sion, and funding acquisition. Manish Saggar: Conceptualization, methodology, investigation, resources, reviewing and editing, supervision, and funding acquisition.

Competing Interests

No competing interests are present among the authors of this work.