Accounting for temporal variability in functional magnetic resonance imaging improves prediction of intelligence

Yang Li1, Xin Ma2, Raj Sunderraman1, Shihao Ji1, Suprateek Kundu3

1Department of Computer Science, Georgia State University, Atlanta, Georgia, USA
2Department of Biostatistics, Columbia University, New York, New York, USA
3Department of Biostatistics, The University of Texas at MD Anderson Cancer Center, Houston, Texas, USA

Correspondence
Xin Ma, Department of Biostatistics, Columbia University, New York, NY 10027, USA.
Email: xm2141@cumc.columbia.edu

Funding information
National Cancer Institute, Grant/Award Number: R01CA244845; National Institute of Mental Health, Grant/Award Number: R01MH120299; National Institute on Aging, Grant/Award Number: R01AG071174

Abstract
Neuroimaging-based prediction methods for intelligence have seen a rapid development. Among different neuroimaging modalities, prediction using functional connectivity (FC) has shown great promise. Most literature has focused on prediction using static FC, with limited investigations on the merits of such analysis compared to prediction using dynamic FC or region-level functional magnetic resonance imaging (fMRI) times series that encode temporal variability. To account for the temporal dynamics in fMRI, we propose a bi-directional long short-term memory (bi-LSTM) approach that incorporates feature selection mechanism. The proposed pipeline is implemented via an efficient algorithm and applied for predicting intelligence using region-level time series and dynamic FC. We compare the prediction performance using different fMRI features acquired from the Adolescent Brain Cognitive Development (ABCD) study involving nearly 7000 individuals. Our detailed analysis illustrates the consistently inferior performance of static FC compared to region-level time series or dynamic FC for single and combined rest and task fMRI experiments. The joint analysis of task and rest fMRI leads to improved intelligence prediction under all models compared to using fMRI from only one experiment. In addition, the proposed bi-LSTM pipeline based on region-level time series identifies several shared and differential important brain regions across fMRI experiments that drive intelligence prediction. A test–retest analysis of the selected regions shows strong reliability across cross-validation folds. Given the large sample size of ABCD study, our results provide strong evidence that superior prediction of intelligence can be achieved by accounting for temporal variations in fMRI.

KEYWORDS
deep neural networks, feature selection, intelligence prediction, neuroimaging analysis

1 | INTRODUCTION

There is great interest in understanding the neural underpinnings of individual differences in intelligence, because it is one of the most
important predictors of long-term life success. Intelligence may be measured via cognitive measures that may include fluid intelligence, defined as the ability to use inductive and deductive reasoning (independent of previously acquired knowledge) to solve new problems (Kyllonen et al., 2017), or crystallized intelligence that involves knowledge that comes from prior learning and past experiences, among others. There is an ongoing debate on whether variations in such intelligence measures are more correlated with the brain structure, or the brain function, and investigations are underway to discover the optimal set of neuroimaging features that are most predictive of these intelligence levels. However, it is a major challenge to relate structural and functional properties of the brain to complex behavioral expression or function (Bullmore & Sporns, 2009; le Bihan et al., 2001; Raichle et al., 2001). Traditional investigations in literature have used structural neuroimaging-derived features such as whole brain volume, regional gray and white matter volumes, or regional cortical volume/thickness and diffusion indices at the whole brain level. While useful, these features may smooth over discriminative features at a finer resolution resulting in inadequate prediction of intelligence (Chen et al., 2020; Paul et al., 2016; Ritchie & Tucker-Drob, 2018; Yuan et al., 2018). Prediction using structural connectivity data derived from diffusion tensor imaging was proposed by Kawahara et al. (2017) who developed an approach involving convolutional neural networks.

Although traditional literature focused on structural brain measures for predicting intelligence, more recent studies have started to investigate prediction strategies for intelligence based on functional MRI (Dubois et al., 2018; Ferguson et al., 2017; He et al., 2020; Kashyap et al., 2019; Liu et al., 2018; Shen et al., 2017). While the majority of intelligence prediction approaches involve linear regression methods, recently emerging studies have focused on nonlinear approaches including polynomial kernel SVR (Wang et al., 2015), and kernel ridge regression (He et al., 2020; Li et al., 2019) methods. A recent review by Vieira et al. (2022) involving 37 studies concluded that while there is a plethora of studies showing correlations between brain function and intelligence, there is only a recently emerging trend of predicting intelligence based on functional brain features using machine learning algorithms. A recent paper by Abrol et al. (2021) systematically showed that deep neural networks when trained on raw data outperform classical linear and nonlinear machine learning models in the prediction of age, gender, and Mini-Mental State Examination scores. This advantage of the deep learning framework is potentially due to the ability to capture nonlinear and complex patterns of relationships between the fMRI time series signals and behavioral traits, which may not be adequately represented by linear models. Extensive numerical studies by He et al. (2020) validated that the predictive performance of kernel ridge regression approaches based on FC features was essentially comparable to prediction using deep learning methods. The review by Vieira et al. (2022) further concluded that while fMRI was the most often used modality for predicting intelligence, most methods used resting state static functional connectivity (FC) derived from the fMRI time series to predict intelligence (Li et al., 2019). Indeed, FC-based differences have been validated in individuals with varying cognitive abilities (Hearne et al., 2016). The overwhelming majority of intelligence prediction approaches based on fMRI data has relied on resting state static FC as input features in the prediction model. In addition, there are a handful of recent studies have shown that combining FC from task fMRI and resting-state fMRI modalities resulted in superior prediction of intelligence measures (Chen et al., 2022; Elliott et al., 2019; Gao et al., 2019) compared to prediction using FC derived from any single fMRI modality. These studies discovered shared and unique network features that were common across tasks and rest, which were able to explain the variations in intelligence.

One major drawback of the prediction pipelines based on static FC is that they do not have the ability to incorporate temporal variations in fMRI which may encode important information regarding brain activity influencing cognition and intelligence. For example, summary measures of blood oxygenation level-dependent (BOLD) signal variability were found to be associated with total composite cognitive score (Sheng et al., 2021). Static FC features averaging over such temporal variations may lead to information loss. There is some limited and recent related literature on prediction approaches for intelligence based on dynamic FC, which capture the temporal variations in the brain network over time. For example, recent work by Sen and Parhi (2020) used tensor decomposition to extract features from dynamic brain networks and subsequently used these features for predicting intelligence in the Human Connectome Project (HCP) study. A handful of other related studies include Liu et al. (2018), who analyzed data on 105 HCP subjects using summary network features, and Omidvarnia et al. (2021) who discovered nonrandom correlations between temporal complexity of resting state networks and fluid intelligence, based on an analysis of 987 HCP individuals. While useful, these articles used summary measures derived from the dynamic networks that potentially result in information loss, which in turn may compromise prediction performance. Most recently, Fan et al. (2020) directly used the resting-state dynamic FC features to predict intelligence based on approximately 1200 subjects from the HCP study, using a deep neural network model.

Apart from the above few articles, there is a scarcity of prediction methods based on dynamic brain networks, which is in sharp contrast to the rich literature identifying potential correlates of task-based dynamic functional connectivity with behavior (Kundu et al., 2018; Kundu et al., 2021; Thompson et al., 2013). Moreover, in-depth comparisons between the predictive ability of static and dynamic FC are fairly limited in literature, with the exception of the recent work by Sen and Parhi (2020) which provides some insights based on 475 subjects from the HCP study. Their analysis relies on tensor-based feature extraction from the dynamic FC features that were derived from a small number of brain regions (85) extracted using the Freesurfer cortical parcellation atlas (Desikan et al., 2006). While useful, the chosen atlas may not be best suited for brain network-based comparisons. Indeed, it may be preferable to use more recent atlases such as the one proposed by Gordon et al. (2016) that uses FC boundary maps to define parcels that represent putative cortical areas with highly homogeneous resting-state FC patterns within a given parcel. Moreover, larger sample sizes are desirable for more robust and
reproducible results, especially given the high dimensionality of brain networks. In summary, a detailed large-scale analysis comparing predictive abilities of static versus dynamic FC features based on more refined parcellations is warranted and much needed.

Another important aspect to consider when using dynamic FC features for predicting cognition is that these features are essentially engineered from region-level fMRI time series and hence cannot capture the temporal variability in brain function that is encoded in the original fMRI data. In fact, the relationship between functional brain connectivity strengths and resting state fMRI temporal complexity changes over time scales (Omidvarnia et al., 2021), and hence may not be straightforward. Therefore, it is not immediately clear whether FC features are primed for optimal intelligence prediction using machine learning pipelines. This is especially true for prediction approaches based on deep neural networks, which can automatically identify suitable representations from minimally preprocessed data (such as fMRI time series) via a succession of hidden layers embedded in an end-to-end learning module. There are other aspects to be taken into consideration when using brain FC features for prediction. Static FC features derived from fMRI data are expected to be sensitive to the choice of the methodology (e.g., pairwise vs. partial correlations) and tuning parameters (e.g., for controlling sparsity of brain networks), which can affect the ensuing predictive analysis. Similarly, results under dynamic FC analysis can be sensitive to the choice of the window lengths (Lindquist et al., 2014). Moreover, the noise level in the FC features may be unpredictable, which may compromise the quality of the analysis. For example, Tian and Zalesky (2021) illustrated that the test–retest reliability of connectome feature weights is generally poor across a range of predictive models, even when the predictive accuracy itself is moderately high. Therefore, it is necessary to investigate whether one can directly use the observed region-level fMRI time series data for intelligence prediction under a deep learning framework, and if and when that provides any advantages compared to prediction based on engineered brain FC features.

The above arguments lay the groundwork for the unmet need to perform a detailed investigation into the prediction properties based on static FC, compared to region-level fMRI time series data as well as dynamic FC features when modeling intelligence. We note that region-level fMRI time series data is naturally more attractive to work with, given that they capture temporal variations in brain activity and are inherent in much smaller dimensions compared to brain networks. Directly using region-level fMRI time series data as input features in a deep neural network immediately alleviates several of the issues encountered in the case of network-based prediction algorithms. For example, it greatly reduces the number of model parameters, thus ameliorating concerns about over-fitting and bypassing computational challenges that may arise in the presence of tens of thousands of edges in the brain network. It also helps in interpretability as the important brain regions associated with the clinical outcomes can be more easily understood and visualized. In contrast, a network-based analysis often reports local (e.g., edge-level) network features that can be extremely granular and noisy, or global network features that are often abstract and difficult to interpret. In addition, it is desirable to perform the proposed predictive analysis using deep neural network modeling given their success (Abrol et al., 2021) in neuroimaging studies, and given the fact that it is designed to leverage complex relationships between intelligence scores and fMRI data. For example, a deep neural network using region-level fMRI time series as inputs may be able to leverage a succession of hidden layers to capture important associations between brain network features and cognitive outcomes, where the brain network features arise in an automated manner within the deep neural network without having to use brain FC features as inputs.

Motivated by the above discussions, our goals in this article are two-fold. First, we provide detailed analysis to evaluate whether static FC-based predictive approaches have any definitive advantages over prediction using (i) region-level fMRI time series data; and (ii) dynamic FC features, derived from resting state as well as task fMRI data. Our in-depth analysis spans multiple types of intelligence measures (fluid, crystallized, and total composite intelligence) and leverages multiple fMRI modalities including resting state fMRI and three types of task fMRI (MID, SST, nback; see details in Section 2.1) experiments. Most importantly, our analysis is based on a large sample size of slightly less than 7000 individuals from the Adolescent Brain Cognitive Development (ABCD) study. This provides a larger scale analysis compared to the overwhelming majority of existing neuroimaging-based prediction studies for intelligence that have primarily focused on data sets with much smaller sample sizes, with the exception of a recent network-based study by Chen et al. (2022). As a result, the findings in our article are much more robust and expected to be far more generalizable. In order to produce a successful prediction pipeline involving temporally varying neuroimaging features including region-level fMRI time series and dynamic FC, we use a novel deep neural network approach based on bi-directional long short-term memory (bi-LSTM) model (Schuster & Paliwal, 1997). On the other hand, the prediction based on static FC features is implemented via a kernel ridge regression model that has shown success in multiple neuroimaging studies (Chen et al., 2022; He et al., 2020). We implement the bi-LSTM approach via an extremely efficient graphics processing unit (GPU) computation scheme and make the code public (via Github) for practical usage by the broader community. Our second goal involves identifying shared and unique brain regions in task and rest fMRI whose time-varying activity is significantly associated with intelligence. By focusing on brain function instead of networks, our results provide complimentary findings to recent work aimed at discovering shared and differential brain network features pertaining to task and rest functional connectivity that drive variations in intelligence (Chen et al., 2022).

We identify such important brain regions using a feature selection approach involving importance weights that are adaptively learned via an $L_0$-norm regularization on the input layer of the bi-LSTM pipeline. While there is limited precedent for using $L_0$ penalization for feature selection in imaging genetics studies involving structural MRI data (Chen et al., 2021), our proposal is one of the first to adapt this idea to a deep neural network framework involving temporally varying functional neuroimaging features under a bi-LSTM framework.
We note that while the proposed approach has some limited technical resemblance with the recent work by Hebling Vieira et al. (2021), there are also fundamental differences. First, unlike their method, the proposed approach provides feature selection capabilities via an $L_0$ regularization on the input features, which is of paramount importance in neuroimaging studies and especially critical for our second goal of discovering shared and unique brain regions (pertaining to task and rest fMRI) that are related to intelligence. Second, in addition to prediction based on region-level time series data, we also extend our deep learning framework to predict intelligence using dynamic FC, which is not considered in Hebling Vieira et al. (2021). Third, we investigate the capability of both resting state and task fMRI experiments to predict a slew of intelligence measures, which provides a richer analysis compared to the resting state fMRI-based prediction presented in that article. Fourth, our analysis is based on the ABCD cohort involving close to 7000 samples, which is considerably larger (by one order of magnitude) compared to the much smaller HCP cohort that was used for analysis in Hebling Vieira et al. (2021). Such large sample sizes elicit more reproducible estimates and produce results with greater generalizability, which is consistent with recommendations in the literature (Marek et al., 2022). Finally, our scientific focus is distinct, in that, we seek to investigate whether incorporating neuroimaging features with temporal variability (either region-level fMRI time series data or dynamic FC features) led to superior prediction of intelligence compared to the routinely used static FC features, and to discover shared and differential brain regions across fMRI modalities that are related to intelligence. Therefore, our treatise provides a more comprehensive and large-scale investigation into unique perspectives that are not necessarily addressed in current literature, to the best of our knowledge.

## 2 MATERIALS AND METHODS

### 2.1 Data

We utilized the preprocessed resting-state and task-based fMRI time series from the ABCD-BIDS Community Collection (ABCC: ABCD Collection 3165: [https://github.com/ABCD-STUDY/nda-abcd-collection-3165](https://github.com/ABCD-STUDY/nda-abcd-collection-3165)). More specifically, the parcelled cifti images under Gordon atlas (Gordon et al., 2016) from ABCC release 1.1.0 were downloaded for all the available individuals, including 9608 for the resting-state session, 8101 for the monetary incentive delay (MID) task session, 7953 for the stop signal task (SST) session, and 7955 for the emotional n-back (nback) task session. We only included the 6835 individuals who had fMRI data for all four experiments in the final analysis for a fair comparison. The typical numbers of timepoints in the fMRI time series are 822, 890, 740, and 1532 for the MID, SST, nback tasks, and resting-state sessions respectively. We also collectively refer to these tasks and resting-state fMRI experiments as “brain states” in the subsequent text.

The fMRI data in the ABCD study have been preprocessed with the Human Connectome Project's minimal preprocessing pipeline (Glasser et al., 2013) and the Developmental Cognition and Neuroimaging (DCAN) Labs resting state fMRI analysis tools (Fair et al., 2020) to get the parcelled cifti images, which include fMRI time series from 333 surface regions as defined by the Gordon atlas (Gordon et al., 2016) and also 19 volumetric regions in the subcortical area, adding up to 352 regions in total. Table 1 contains the detailed information of the 13 functional modules assigned to the 333 brain surface regions in the Gordon atlas.

| Module name          | Abbreviation | Number of regions | Module name          | Abbreviation | Number of regions |
|----------------------|--------------|------------------|----------------------|--------------|------------------|
| Auditory             | Aud          | 24               | Retrospenial temporal| RT           | 8                |
| Cingulo opercular    | CO           | 40               | Salience             | Sal          | 4                |
| Cingulo parietal     | CP           | 5                | Sensory motor hand   | SMh          | 38               |
| Default mode         | DM           | 41               | Sensory motor mouth  | SMm          | 8                |
| Dorsal attention     | DA           | 32               | Ventral attention    | VA           | 23               |
| Frontoparietal       | FP           | 24               | Visual               | Vis          | 39               |
| None                 | None         | 47               |                      |              |                  |

We consider three types of neuroimaging features in our analysis. The first one is region-level fMRI time series data as coming directly from the parcelled cifti images described earlier. The second one is the static functional connectivity matrix data, which is obtained by calculating the pairwise Pearson correlations of the region-level fMRI time series. The third and last type of feature we consider is the dynamic functional connectivity, obtained by calculating the sliding window pairwise Pearson correlations of the time series, with window size selected at 50 s and window stride at 5 as suggested in Sen and Parhi (2020).

In addition to the imaging data, we also extracted the demographic and neurocognition measurements for the corresponding 6835 individuals from the ABCD release 3.0 (Karcher & Barch, 2021) year 1 data. Individual’s age is included in all models as a control variable. Our outcomes of interest, the intelligence scores, were derived from the NIH toolbox neurocognition battery for the youth (Luciana et al., 2018). Specifically, we consider three types of intelligence metrics including fluid, crystallized, and total composite intelligence.
The potential confounders for intelligence prediction include age, brain volume, gender, and handedness as considered in previous literature (Dubois et al., 2018). For the ABCD samples in our analysis, the correlations between age and fluid, crystallized, and total composite intelligence metrics are 0.263, 0.260, and 0.305, respectively. In contrast, the correlations between brain volume and the three types of intelligence metrics are orders of magnitude lower at 0.005, 0.019, and 0.012, respectively. Moreover, unlike age, there are no discernible differences in intelligence distributions when stratified by gender and handedness categories. Thus, in our analysis, we choose to only control for the age variable while leaving the others out of the model.

Other confounders related to fMRI preprocessing that were considered in the literature (Dubois et al., 2018) included motion correction, which was already included in the preprocessing steps for the ABCC repository used in our analysis, and hence was not included explicitly in our modeling. In particular, the ABCC repository incorporated additional preprocessing steps using the DCAN BOLD Processing (DBP) on top of the minimal preprocessing pipeline from HCP, which corrects for respiratory artifacts in MRI head motion estimates and the motion artifacts with motion censoring and global signal regression to the extent possible (Fair et al., 2020).

2.2 | Schematic structure

We illustrate the workflow of our analysis in Figure 1. We consider different models for the three types of neuroimaging features extracted from the fMRI images. In the following Sections 2.3 and 2.4, we present our deep neural network prediction pipeline based on temporally varying input features and an integrated feature selection mechanism that can produce feature importance scores. The proposed pipeline can cater to both region-level fMRI time series input data, as well as dynamic FC inputs. Then in Section 2.5, we introduce two types of benchmark models including the kernel ridge regression for prediction based on static FC, and the linear penalized models for prediction based on region-level fMRI time series data. In order to evaluate the prediction performance across all the models we consider, we conducted a 5-fold cross-validation analysis. We randomly divided the 6835 individuals into five partitions. Each time we used a different partition as test data, while the remaining four partitions served as training data. Within each set of training data, we also have an inner loop of 5-fold cross-validation for hyperparameter tuning in the proposed pipeline as well as the benchmark models. We note that data from different brain states of the same subject was contained either wholly in the training set or the test set. Summary statistics of the five folds of the outer loop are shown in Table 2.

![Figure 1](image-url)  
**Figure 1**  Analysis workflow. The static and dynamic FCs are derived from the region-level fMRI time series. These neuroimaging features are then served as input for different models for prediction of the three intelligence metrics.

| TABLE 2 | Summary of demographic information and intelligence metrics of the five cross-validation folds. |
|---------|----------------------------------------------------------------------------------------------------------------------------------|
| N       | Age (months) | Sex (M/F) | Fluid Intelligence (mean ± SD) | Crystallized intelligence (mean ± SD) | Total comp intelligence (mean ± SD) |
|---------|--------------|-----------|-------------------------------|--------------------------------------|------------------------------------|
| 1367 × 5| [107, 132]   | 673/694   | 92.35 ± 10.25                 | 86.90 ± 6.81                         | 87.01 ± 8.70                       |
|         | [107, 132]   | 679/688   | 92.72 ± 10.06                 | 87.10 ± 6.45                         | 87.36 ± 8.41                       |
|         | [107, 132]   | 683/684   | 92.39 ± 10.42                 | 86.99 ± 6.50                         | 87.10 ± 8.54                       |
|         | [107, 132]   | 693/674   | 92.42 ± 9.99                  | 87.16 ± 6.72                         | 87.20 ± 8.52                       |
|         | [107, 132]   | 758/609   | 92.32 ± 10.48                 | 86.90 ± 6.61                         | 86.99 ± 8.74                       |
2.3 | Bi-directional LSTM model

Recurrent neural networks (RNNs) have shown great capability in processing sequential data, such as audio, video, traffic data, and so forth. RNNs contain cyclic connections that repeatedly feed the network activations from a previous step as inputs to the current step, which allows the network to capture long-range dependencies. Long short-term memory (LSTM; Hochreiter & Schmidhuber, 1997) is a specific RNN architecture that introduces control gates: input gates, output gates, and forget gates. Compared to conventional RNNs, LSTM alleviates vanishing gradient problems and therefore usually leads to a better performance. Bi-directional LSTM (bi-LSTM; Schuster & Paliwal, 1997) combines an LSTM that takes input in a forward direction with another LSTM that takes input in a backward direction. Such mechanism allows the model to build dependency among the whole input sequence that is conducive for fMRI data.

In our study, we use a two-layer bi-LSTM model. A schematic representation of this model for prediction based on region-level fMRI time-series is depicted in Figure 2. As shown on the right side of Figure 2, the model consists of two stacked LSTM modules and a fully connected layer. The input time series data for each individual is a matrix of size $352 \times T$, where 352 is the number of cortical surface and sub-cortical regions, and $T$ denotes the number of time points. We denote the column vectors from this data matrix as $x_1, x_2, \ldots, x_T$, where $x_t$ $(t = 1, \ldots, T)$ is a vector of length 352 for the time point $t$. Inputs $x_1$ through $x_T$ are fed into the bottom LSTM cells recursively, in both forward and backward directions. The outputs from both directions are concatenated (denoted as $\oplus$ in Figure 2) and then fed into the top LSTM layer.

Between each two layers, we use dropout which randomly drops some data in order to prevent overfitting issue. The final prediction is made by a fully connected layer which takes as input the mean aggregated LSTM outputs and the age of the individual, for predicting intelligence. The bi-LSTM model can be extended to use the dynamic connectivity data as input when predicting intelligence. As these connectivity matrices are symmetric, we only take the upper triangle of each matrix and flatten it as a vector before feeding it into the bi-LSTM model. In this case, the input matrix for the bi-LSTM model is of size $352 \times (352 - 1)/2$ by $T^*$, where $T^*$ is the total number of dynamic matrices, which depends on the window size and window stride when calculating the dynamic connectivity (see Section 2.1).

The description above applies to the models using any single-task or resting-state fMRI data as input. When performing an integrative analysis that combines rest and task fMRI data, we adopt a feature-concatenation approach. In particular, the input matrix has $352 \times 4$ rows and $T_0$ columns when integrating the region-level time series across rest and the three tasks, where $T_0$ is the minimum number of time points across the four fMRI experiments. Such a feature concatenation requires limited trimming of the time series across the different fMRI datasets that may result in minimal information loss, but this does not adversely affect performance in our experience. The bi-LSTM approach is applied to this matrix of concatenated time series for intelligence prediction. Similarly, when using dynamic FC features across rest and task experiments, the concatenated data matrix has $352 \times (352 - 1)/2 \times 4$ rows and $T_{0}^*$ columns, where $T_{0}^*$ represents the minimum number of dynamic connectivity matrices among all the individuals and all the fMRI experiments.

**FIGURE 2** The architecture of the model is trained with feature selection. Left: feature selection module. Right: the architecture of the two-layer bi-LSTM model.
We train the model to minimize the difference between the predicted and observed intelligence scores. More specifically, the regularized optimization target is

$$\arg\min_{\theta} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(f(X_i; \theta), y_i) + \lambda \| \theta \|_2,$$  \hspace{1cm} (1)

where $f$ denotes the bi-LSTM model parameterized by $\theta$, $X_i$ is the matrix of fMRI features (either region-level time series, or dynamic connectivity) for individual $i$, $y_i$ is the corresponding observed intelligence score, $N$ is the number of individuals in the training dataset, and $\mathcal{L}(a, b) = (a - b)^2$ is the squared loss function. The second term of Equation 1 is the $L_2$ regularization (weight decay) on model parameters and $\lambda$ controls the regularization strength.

2.3.1 | Implementation

We implement the model with PyTorch (Paszke et al., 2017). The hidden size of the LSTM is set to 80 for fMRI time series and 500 for dynamic connectivity data. The dropout rate between the two LSTM layers is set to 20%. We trained the model with backpropagation through time (BPTT; Werbos, 1990) using ADAM (Kingma & Ba, 2014) as the optimizer. The Adam optimization algorithm is an extension to stochastic gradient descent and has been widely used in deep learning field. For dynamic connectivity data, we use ADAM with SAM (Foret et al., 2020). SAM aims to improve model generalization by seeking parameters that lie in neighborhoods having uniformly low loss. We try using SAM in all of our experiments and it turns out that SAM can significantly improve model performance with dynamic connectivity data, which is a novel discovery of independent interest. The learning rate for ADAM optimizer is set to 0.0001 and the weight decay regularization parameter $\lambda$ is set to 0.01. The model is trained for 10 epochs. In terms of validation, we use 5-fold cross-validation as discussed in Section 2.2.

2.4 | Feature selection and importance score learning

Feature selection is widely used in machine learning to avoid the curse of dimensionality. By removing or down-weighting the redundant or irrelevant features, issues such as overfitting can be avoided. Additionally, feature selection is naturally equipped to generate feature importance scores and identify the most informative features in prediction tasks, which is of paramount importance in neuroimaging studies. However, feature selection in black box methods such as deep learning is not straightforward and is an ongoing area of research development. While ad-hoc feature selection methods such as inversion are often used in neuroimaging literature (Chen et al., 2022), it is desirable to propose more principled and systematic approaches for feature selection under the deep learning framework.

To enable feature selection in the bi-LSTM model using region-level fMRI time series as input, we introduce a binary mask vector $z = (z_1, \ldots, z_p)^T$ where $p = 352$ is the number of brain regions and $z_k \in \{0, 1\}$ for $k = 1, \ldots, p$. When $z_k = 0$, the corresponding time series data for region $k$ is zeroed out of the target optimization function. In order to select a small set of important regions, we add to the objection function an $L_0$ regularization term for $z$ which explicitly penalizes the number of nonzero elements in $z$ and encourages parsimonious models. The optimization target turns into the following format:

$$\arg\min_{\theta, z} \frac{1}{N} \sum_{i=1}^{N} \mathcal{L}(f(D_i \otimes X_i; \theta), y_i) + \lambda_1 \| \theta \|_2 + \lambda_2 \| z \|_0,$$  \hspace{1cm} (2)

where $\otimes$ denotes the matrix multiplication, $D_i$ is a diagonal matrix of size $p$ by $p$ with the elements of $z = (z_1, \ldots, z_p)^T$ as its diagonal elements, and $\lambda_1$ and $\lambda_2$ are two hyper-parameters which control the strength of weight decay and feature selection, respectively. However, the binary mask $z$ cannot be optimized directly in neural networks since the elements in the vector are discrete variables. Hence, we have to introduce a trainable continuous parameter for $z$. A natural choice is to define the elements in $z$ as random variables drawn from Bernoulli distribution parameterized by probability vector $\pi$. By incorporating the random variables, the model becomes stochastic so we have to minimize the expectation over $z$ and the optimization target changes into:

$$\arg\min_{\theta, \pi} \frac{1}{N} \sum_{i=1}^{N} \mathbb{E}_{\text{Bern}(z|\pi)} [\mathcal{L}(f(D_i \otimes X_i; \theta), y_i) + \lambda_1 \| \theta \|_2 + \lambda_2 \| z \|_0].$$  \hspace{1cm} (3)

Note that in Equation 3, we optimize over $\theta$ and $\pi$. However, the discrete nature of the Bernoulli distribution blocks the gradient backpropagation to $\pi$ in the bi-LSTM pipeline. Inspired by sparsity literature in deep learning, we utilize Gumbel–Softmax (Jang et al., 2016; Maddison et al., 2016) as a surrogate for Bernoulli distribution. Gumbel–Softmax was proposed to make categorical variables learnable in neural networks. Instead of sampling the elements of $z$ from Bernoulli distribution with probability vector $\pi$, we sample a vector $s = (s_1, \ldots, s_p)^T$ from the Gumbel–Softmax distribution as follows:

$$u \sim \text{Uniform}(0, 1), s_k = \sigma((\log u - \log(1 - u) + \alpha_k)/\beta), k = 1, \ldots, p,$$  \hspace{1cm} (4)

where $\sigma(a) = [e^a/(1 + e^a)]$ denotes the logistic function (also known as the Sigmoid function), $\alpha = (\alpha_1, \ldots, \alpha_p)^T$ is linked to the probability vector $\pi$ through $s_k = \sigma(\alpha_k)$ for $k = 1, \ldots, p$, and $\beta$ is called the temperature which controls the shape of the distribution. As the temperature approaches 0, $s_k$’s become binary (0 or 1). Note that $s_k$’s cannot be exact 0 or 1 if sampled according to Equation 4, which is not preferable in feature selection because it is not able to completely zero out a feature. Fortunately, Louizos et al. (2017) proposed a method that stretches the Gumbel–Softmax samples to the interval of $(\gamma, \zeta)$ where $\gamma < 0$ and $\zeta > 1$, and then clamps the samples to be between 0 and 1:

$$z_k = \text{clamp}(s_k, \gamma, \zeta),$$  \hspace{1cm} (5)

where clamp is a function that clamps the samples to the interval of $(\gamma, \zeta)$.

Note that in Equation 5, $\gamma$ and $\zeta$ can be set to any values in the interval $(0, 1)$, but the choice of values will affect the performance of the model. In our experiments, we set $\gamma = 0.1$ and $\zeta = 1.5$ as they provide a good balance between sparsity and smoothness.
\[ z_k = s_k(\zeta - \gamma) + \gamma, z_k^* = \min(1, \max(0, z_k)), k = 1, \ldots, p. \]  

and then \( z^* = (z_1^*, \ldots, z_p^*)^T \) can serve as a surrogate of the binary mask vector \( z \) in the pipeline. Louizos et al. (2017) provides the conditional expectation of \( \| z^* \|_0 \) given \( \alpha \) as \( \sum_{k=1}^p \sigma(\alpha_k - \beta \log z^*_k). \)

Taken all these points together, we substitute in Equation 3 the Bernoulli distribution with the Gumbel–Softmax distribution, apply the stretching trick, and turn the optimization target function into this final format:

\[
\arg\min_{\theta, \alpha} \frac{1}{N} \sum_{i=1}^N \mathbb{E}_{q(s_i)}[L(f(D_y(s) \otimes X; \theta), y_i)] + \lambda_1 \| \theta \|_2 \\
+ \lambda_2 \sum_{k=1}^p \sigma \left( \alpha_k - \beta \log \frac{z^*_k}{\zeta} \right)
\]

where \( q \) is the Gumbel–Softmax distribution with \( s = (s_1, \ldots, s_p)^T \) generated as in Equation 4 and \( D_y(s) \) is a diagonal matrix with diagonal elements from \( g(s) = \min(1, \max(0, s)). \) After training the model and obtaining the estimates \( \hat{\alpha} \), we can calculate the feature importance scores as \( \min(1, \max(0, \sigma(\zeta - \gamma) + \gamma)). \)

We show the data flow of the feature selection module on the left panel of Figure 2. The feature selection module is jointly trained with the bi-LSTM module on the right panel. Note that we train the network stochastically by sampling the surrogate mask vector \( z^* \) from the stretched Gumbel–Softmax distribution while generating the importance scores, we use the conditional expectation of \( z^* \) instead. We use the same bi-LSTM model configurations as described in Section 2.3. The model is trained for 20 epochs. We use ADAM as the optimizer with the initial learning rate of 0.0001. The learning rate is multiplied by 0.1 at epoch 10 and 15. In terms of the hyperparameters, we set \( \gamma = 0.1, \zeta = 1.1, \beta = 2/3, \lambda_1 = 0.01, \) and \( \lambda_2 = 0.25. \) Our source code for the proposed method is available online at https://github.com/leo-yangli/abcd_time_series.

### 2.5 Benchmark comparisons

#### 2.5.1 Kernel regression methods based on static FC

We compare the performance of prediction using bi-LSTM approach based on region-level fMRI time series and dynamic FC with kernel ridge regression (KRR) based on static FC features that is considered state of the art (He et al., 2020)—see appendix A1 of He et al. (2020) for more details. The predictors for this approach involve static functional connectivity matrix coming from the resting-state fMRI or any single-task fMRI corresponding to unimodal analysis. We also implement the multi-KRR method as described in Chen et al. (2022, see supplementary methods S3) for multi-task analysis that involves static FC features from the resting-state as well as task fMRI (MID, SST, nback) experiments. As shown in He et al. (2020), the kernel-based methods have comparable prediction performance as a deep learning framework. Thus, these single- and multi-KRR methods serve as deep learning benchmark for using static FC as input features. For this reason, we did not train any specific deep learning structure beyond these KRR methods for this type of fMRI features.

#### 2.5.2 Linear penalized regression approaches using region-level fMRI time series

We also consider three types of linear regression models using region-level time series data, which are fit using penalized approaches to serve as additional benchmark methods for prediction performance. The methods include Lasso with \( \ell_1 \)-norm penalty (Tibshirani, 1996), ridge regression with \( \ell_2 \)-norm penalty (Hoerl & Kennard, 1970), and elastic net with hybrid \( \ell_1/\ell_2 \)-norm penalty (Zou & Hastie, 2005). The temporal variations of the fMRI time series for each individual are summarized into a few leading principal component (PC) scores that account for at least 95% variation derived from the region-level fMRI time series. The principal component analysis yields varying number of PCs per individual and the minimum number of PCs across all individuals that account for at least 95% of the variability is chosen. This number corresponded to 76, 80, 66, and 105 principal components (PCs) for the MID, SST, nback task, and resting-state experiments, respectively. The PCs are then stacked into vectors and used as predictors for the linear regression models.

### 2.6 Prediction performance evaluation

The evaluation criteria for prediction performance are based on normalized MSE (NMSE) and Pearson correlation (corr) between predicted and observed intelligence scores in the testing samples, which can be calculated as

\[
\text{NMSE} = \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2}, \quad \text{and} \quad \text{corr} = \frac{\sum_{i=1}^N (\hat{y}_i - \bar{y})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (\hat{y}_i - \bar{y})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}.
\]

respectively, where \( N \) is the total number of individuals in the testing samples, \( y_i \) and \( \hat{y}_i \) denote the observed and predicted intelligence scores, and \( \bar{y} \) and \( \bar{p} \) denote the mean of observed and predicted intelligence scores.

### 2.7 Test–retest reliability

Test–retest reliability is an important indicator of the robustness of the bi-LSTM model with importance score reporting (Tian & Zalesky, 2021). Essentially, a strong test–retest reliability (TRR) for the features included in the model indicates the robustness of the
approach, in terms of being able to consistently identify the important neuroimaging features driving the prediction of the outcome variable. In order to investigate the test–retest reliability under the proposed approach using region-level fMRI time series data, we compute the intraclass correlation coefficient (ICC) metric across all features and cross-validation folds (we use 5-fold cross-validation), which captures the agreement across folds in terms of importance rating of the features. We choose to report the ICC of absolute agreement using two-way random effects model with “average rater” unit (Koo & Li, 2016). This procedure reflects a type of within-sample TRR that has been previously used in neuroimaging literature (Chen et al., 2022; Dhamala et al., 2021; Jiang et al., 2020), and includes overlapping training samples between splits. This type of TRR metric is more common in literature compared to the out-of-sample TRR, which would need to be calculated based on a two-fold cross-validation scheme. See more details in Tian and Zalesky (2021). As the raw importance scores are obtained with different selection of regularization parameter under different experiments, they may not be on the same scale between 0 and 1. Thus, before computing the ICC scores, we perform a normalization process on each set of importance scores $\{z_1, \ldots, z_p\}$ with the following formula for $k = 1, \ldots, p$:

$$z_k = \frac{z_k - z_0}{z_1 - z_0}, z_1 = \max\{z_1, \ldots, z_p\}, z_0 = \min\{z_1, \ldots, z_p\}$$

### TABLE 3 Prediction performance of bi-LSTM and KRR models across fMRI experiments and intelligence metrics.

| Type of fMRI experiment | Fluid intelligence | Crystalized intelligence | Total comp intelligence |
|------------------------|-------------------|--------------------------|------------------------|
|                        | NMSE              | corr                     | NMSE                   | corr                     |
| **Bi-LSTM model with region-level fMRI time series** |                   |                          |                        |
| MID                    | 0.806* (0.010)    | 0.448* (0.010)           | 0.747 (0.019)          | 0.526* (0.018)           | 0.693* (0.008)          | 0.559* (0.006)          |
| SST                    | 0.844 (0.008)     | 0.403 (0.009)            | 0.798* (0.017)         | 0.468* (0.014)           | 0.756* (0.008)          | 0.500* (0.006)          |
| nback                  | 0.786* (0.010)    | 0.472* (0.011)           | 0.760 (0.022)          | 0.511* (0.020)           | 0.676* (0.012)          | 0.575* (0.011)          |
| rest                   | 0.835* (0.015)    | 0.407* (0.016)           | 0.778 (0.015)          | 0.484 (0.011)            | 0.733* (0.010)          | 0.517* (0.010)          |
| 3 tasks + rest         | 0.779 (0.016)     | 0.483 (0.011)            | 0.704 (0.034)          | 0.554 (0.026)            | 0.652 (0.013)           | 0.594 (0.010)           |
| **Bi-LSTM model with dynamic FC** |                   |                          |                        |
| MID                    | 0.816 (0.023)     | 0.447* (0.035)           | 0.690 (0.013)          | 0.568* (0.019)           | 0.689* (0.017)          | 0.563* (0.025)          |
| SST                    | 0.870 (0.011)     | 0.389* (0.015)           | 0.765 (0.006)          | 0.489 (0.006)            | 0.770* (0.013)          | 0.492* (0.013)          |
| nback                  | 0.802 (0.021)     | 0.469 (0.037)            | 0.714 (0.011)          | 0.542 (0.009)            | 0.673* (0.022)          | 0.576* (0.022)          |
| rest                   | 0.848 (0.027)     | 0.408* (0.020)           | 0.719 (0.016)          | 0.536 (0.011)            | 0.728* (0.017)          | 0.529* (0.010)          |
| 3 tasks + rest         | 0.763 (0.020)     | 0.507 (0.022)            | 0.628 (0.012)          | 0.615 (0.010)            | 0.611 (0.021)           | 0.624 (0.014)           |
| **KRR model with static FC** |                   |                          |                        |
| MID                    | 0.840 (0.013)     | 0.401 (0.018)            | 0.752 (0.028)          | 0.498 (0.028)            | 0.754 (0.016)           | 0.496 (0.016)           |
| SST                    | 0.892 (0.007)     | 0.328 (0.010)            | 0.819 (0.011)          | 0.426 (0.013)            | 0.825 (0.011)           | 0.418 (0.013)           |
| nback                  | 0.810 (0.014)     | 0.439 (0.017)            | 0.762 (0.019)          | 0.492 (0.021)            | 0.732 (0.004)           | 0.523 (0.004)           |
| rest                   | 0.870 (0.011)     | 0.363 (0.012)            | 0.775 (0.028)          | 0.478 (0.026)            | 0.793 (0.020)           | 0.458 (0.020)           |
| 3 tasks + rest         | 0.768 (0.016)     | 0.482 (0.016)            | 0.660 (0.014)          | 0.583 (0.012)            | 0.650 (0.015)           | 0.592 (0.013)           |

Note: Reported results include averaged values across the five cross-validation folds and standard deviation in braces. Paired two-sample $t$-tests are conducted on the evaluation metrics across the five folds at significance level 0.05 with Bonferroni correction for multiple testings. Bold numbers: significantly better result than the other two models. * marked numbers: significantly better result than the KRR model using static FC only.

### 3 RESULTS

#### 3.1 Prediction performance

The results presented in Tables 3 and 4 capture the comprehensive prediction analysis performed on the ABCD data. The evaluation metrics are averaged over the five cross-validation folds. The standard deviation values across the five folds are also reported for all the results. In Table 3, the numbers in bold denote the significantly improved performance compared to the other two methods presented in the table, while those marked with "**" represent significantly better performance compared to the KRR model using static FC in Table 3, based on paired two-sample $t$-tests conducted across the five-folds of testing samples. All the hypothesis tests were conducted with the significance level set at 0.05 and were adjusted for multiple testing using Bonferroni correction. We compare the prediction performance between the bi-LSTM pipeline using region-level fMRI time series as well as dynamic FC, along with the KRR approaches that use static FC. Our analysis involves MID, SST, and nback tasks along with resting state fMRI data. Our goal is to investigate which prediction pipeline performs the best, and how the relative performance changes across intelligence metrics and varies by the fMRI modality used for prediction.
When predicting fluid intelligence based on a single brain state, the bi-LSTM method with region-level fMRI time series has the best predictive performance and registers significant improvements compared to prediction based on static FC across all fMRI experiments, as well as compared to prediction based on dynamic FC extracted from the SST task. Further, when predicting total composite intelligence, the performance of the bi-LSTM approaches using region-level fMRI time-series as well as dynamic FC features are comparable, and both methods have significantly improved prediction performance compared to the KRR approach using static FC features. On the other hand, the bi-LSTM approach using dynamic FC has the best prediction performance when predicting crystallized intelligence based on a single brain state, which is significantly improved compared to the other two types of approaches in Table 3. Such a superior performance is seen across the three fMRI task modalities as well as resting state fMRI. Additionally, the bi-LSTM approach using dynamic FC has significantly improved performance when predicting intelligence based on multi-task analysis that combines the three types of task and rest fMRI data, compared to the other two approaches in Table 3. These results illustrate the potential of dynamic FC features for multi-task analysis. Taken together, the above analysis clearly illustrates the potency of temporally varying fMRI features in terms of predicting intelligence based on single or multiple fMRI experiments via a deep learning pipeline, particularly in contrast to prediction based on routinely used static FC features.

Further, the prediction accuracy under the MID and the nback tasks was consistently higher compared to resting-state fMRI time series data as well as the SST fMRI experiment across all three intelligence metrics, which is consistent with the findings pertaining to static FC-based prediction in Chen et al. (2022). In general, we note that combining data from multiple brain states may lead to improvements in prediction compared to analysis based on just one fMRI experiment, which is also consistent with the observations in Chen et al. (2022).

Finally, we also evaluate the usefulness of the deep learning prediction approach compared to a more routinely used linear model-based prediction using penalization (see Table 4) based on fMRI time series features. The results clearly indicate a dismal performance under the linear regression approaches consistently across all task modalities as well as resting-state fMRI data. These results suggest that intelligence prediction using flexible nonlinear models provide a far superior performance compared to routinely used linear regression modeling based on region-level fMRI time series.

### 3.2 Comparison of computation efficiency

We report the computation times for bi-LSTM models using single-task and resting-state data input in Figure 3. The bi-LSTM models are executed on GPU machine with Intel Xeon Gold 6242 CPU at 2.80GHz and NVIDIA Tesla V100 GPU. We can see from Figure 3 that the training speed for bi-LSTM model with fMRI time series data is typically around 20 times faster compared to the speed with dynamic FC data. For bi-LSTM models using combined task and rest

| Table 4 | Fluid intelligence | Crystallized intelligence | Total comp intelligence |
|---------|---------------------|---------------------------|------------------------|
|         | NMSE                | corr                      | NMSE                   | corr                  | NMSE                  | corr                  |
| **Lasso** |                     |                           |                        |                       |                       |                       |
| MID     | 0.997 (0.001)       | 0.059 (0.014)             | 0.998 (0.003)          | 0.050 (0.025)         | 0.995 (0.003)         | 0.078 (0.029)         |
| SST     | 0.996 (0.003)       | 0.068 (0.026)             | 0.995 (0.002)          | 0.073 (0.020)         | 0.993 (0.001)         | 0.094 (0.020)         |
| nback   | 1.000 (0.003)       | 0.020 (0.037)             | 1.001 (0.002)          | -0.003 (0.017)        | 1.001 (0.002)         | 0.002 (0.020)         |
| rest    | 0.993 (0.001)       | 0.090 (0.014)             | 0.997 (0.002)          | 0.053 (0.023)         | 0.990 (0.001)         | 0.108 (0.010)         |
| 3 tasks + rest | 0.993 (0.002) | 0.095 (0.017)             | 0.996 (0.002)          | 0.077 (0.020)         | 0.989 (0.003)         | 0.119 (0.026)         |
| **Ridge** |                     |                           |                        |                       |                       |                       |
| MID     | 0.998 (0.002)       | 0.044 (0.019)             | 0.999 (0.004)          | 0.036 (0.035)         | 0.998 (0.003)         | 0.046 (0.030)         |
| SST     | 0.997 (0.003)       | 0.052 (0.029)             | 0.999 (0.001)          | 0.030 (0.016)         | 0.997 (0.003)         | 0.056 (0.025)         |
| nback   | 0.999 (0.002)       | 0.026 (0.020)             | 1.001 (0.002)          | 0.009 (0.016)         | 1.000 (0.002)         | 0.024 (0.026)         |
| rest    | 0.996 (0.002)       | 0.058 (0.023)             | 0.998 (0.001)          | 0.038 (0.011)         | 0.996 (0.001)         | 0.063 (0.014)         |
| 3 tasks + rest | 0.995 (0.002) | 0.070 (0.021)             | 0.998 (0.004)          | 0.043 (0.026)         | 0.995 (0.004)         | 0.072 (0.029)         |
| **Elastic net** |   |                     |                        |                       |                       |                       |
| MID     | 0.997 (0.001)       | 0.053 (0.013)             | 0.998 (0.004)          | 0.046 (0.028)         | 0.995 (0.002)         | 0.082 (0.028)         |
| SST     | 0.996 (0.003)       | 0.059 (0.017)             | 0.995 (0.002)          | 0.069 (0.020)         | 0.993 (0.002)         | 0.090 (0.017)         |
| nback   | 0.999 (0.002)       | 0.020 (0.037)             | 1.000 (0.002)          | 0.006 (0.027)         | 1.001 (0.003)         | 0.020 (0.028)         |
| rest    | 0.995 (0.005)       | 0.075 (0.022)             | 0.998 (0.002)          | 0.046 (0.021)         | 0.990 (0.002)         | 0.100 (0.011)         |
| 3 tasks + rest | 0.994 (0.001) | 0.083 (0.018)             | 0.995 (0.003)          | 0.078 (0.021)         | 0.989 (0.003)         | 0.114 (0.027)         |

Note: Reported results include averaged values across the five cross-validation folds and standard deviation in braces.
data input, the computation time is 0.3 h for fMRI time series and increases dramatically to around 36 h for dynamic FC input. The dynamic FC data has much higher dimensional features as input that encumbers the training efficiency and increases the computing time, and at the same time requires greater memory. In contrast, the rapid computation times for prediction using fMRI time series data contribute to its appeal for predicting intelligence.

The KRR and linear penalized models are executed on high-performance computing (HPC) environment with Intel Xeon CPU at 2.80 GHz. For KRR model with single task or resting state data, the training time is around 1 day, while for multi-KRR model with three tasks plus resting state data, the training time can last several days. Compared to these computation runtimes in the HPC setting, the bi-LSTM analysis pipeline utilizing GPU computing shows a clear advantage in terms of computational efficiency and scalability. Further, the linear penalized models can be trained quickly, in under 5 min for Lasso and ridge penalty models, and under 1 h for elastic net penalty model. The quick computation times result from efficient implementation of these linear models via existing software packages, as well as due to the use of principal components summarizing the variability in the fMRI time series, instead of directly using the original time series data. However, the lack of prediction power rules out this type of model as a viable model for predicting intelligence using region-level fMRI time series data.

### 3.3 Test–retest reliability of the importance scores

We report the summary of ICC scores calculated from the cross-validation folds in Figure 4, based on the scaled importance scores obtained by fitting the bi-LSTM models with $L_0$ regularization using region level fMRI time series data in Section 2.4. From the results, it is clearly evident that the proposed approach has strong test–retest reliability as indicated by high ICC scores, consistently across the three task modalities and resting state fMRI. The important brain regions identified by the proposed approach under the region-level fMRI time series analysis are hence highly consistent across the five folds considered in the analysis, and this is true for predicting fluid, crystallized, and total composite intelligence. We note that it was not possible to report test–retest reliability under the other two FC-based prediction approaches. This is due to the fact that the KRR method is not naturally equipped to perform feature selection without resorting to additional inversion techniques (Chen et al., 2022), while the bi-LSTM approach involving dynamic connectivity features has an ultra-high dimensional feature space, making it computationally challenging to implement the $L_0$ regularization mechanism. In contrast, the region-level fMRI analysis naturally lends itself to the computation of importance scores via the $L_0$ regularization approach in Section 2.4, in a scalable manner that facilitates the selection of important brain regions that are more interpretable compared to brain network-based features.

### 3.4 Brain region selection based on the importance scores

We have shown in Section 3.3 that the scaled importance scores are highly consistent across the five cross-validation folds based on the ICC metric with “average rater” unit. This strongly validates our practice to average the importance scores across the five folds and rank the importance of the different brain regions according to the average scores.

#### 3.4.1 Distribution of important brain regions

The overall distribution of importance scores across functional modules are illustrated in Figure 5 containing boxplots. The frontoparietal...
(FP) module and default mode network (DMN) have some of the highest importance scores across the three intelligence metrics and brain states, although some FP and DMN regions may have low importance for crystallized intelligence. Moreover, cingulo parietal (CP) regions show high importance corresponding to rest and nback tasks, but low importance corresponding to the MID and SST tasks, when predicting intelligence. The regions in the dorsal attention (DA) module have high importance corresponding to the total composite intelligence, but these regions show relatively lower importance when predicting fluid and crystallized intelligence. Regions in the retrosplenial temporal (RT) module show low importance corresponding to the SST task as compared to other brain states, when predicting all the three intelligence metrics. Moreover, regions in the salience (Sal) network have considerably higher importance scores corresponding to the MID task, but not for the other brain states. Finally, the visual network registered high importance scores across brains states and intelligence metrics, although the relative importance of brain regions in this module was higher when predicting fluid and total composite intelligence but slightly lower when predicting crystallized intelligence.

3.4.2 | Shared and differential brain regions across brain states

Figure 6 illustrates the correlations of the averaged importance scores across brain states, and intelligence metrics. The figure contains nine panels, that illustrate correlations between the importance scores across different brain states when (i) predicting one particular intelligence metric (diagonal panels); and (ii) predicting two distinct intelligence metrics (off-diagonal panels). As evident from the diagonal panels, the correlation between the importance scores across brain states is the highest when predicting fluid intelligence, followed by total composite score and crystallized intelligence, in that order. Further, the feature importance correlations between the SST and nback tasks are seen to be lower when predicting all the three intelligence metrics, with the correlations being the lowest for crystallized intelligence prediction. In general, the important brain regions from the SST task that drive intelligence prediction have lower concordance with the regions corresponding to other tasks and resting state. In contrast, when predicting total composite intelligence, the importance scores corresponding to brain regions for resting state has strong correlations with other brain states. These results suggest the presence of shared as well as distinct brain regions across brain states that drive intelligence prediction.

To further investigate shared and differential brain regions in greater detail, we calculated the number of brain regions that were uniquely identified as important for a given brain state but not others, as well as the number of brain regions that were identified as important jointly across pairs of brain states. Moreover, we considered a brain region to be important if it belong to the top 10% of the importance scores. For greater interpretability, our analysis was stratified by functional modules, and by different intelligence metrics.

These results were illustrated in Figure 7, which presents the number of unique regions in the diagonal cells and shared regions across pairs of brain states in the off-diagonal cells, when predicting intelligence. The highest number of unique brain regions were
discovered in (i) the default mode network for all brains states except the nback task; (ii) the frontoparietal network corresponding to all the brain states except the resting state; and (iii) the visual network corresponding to nback task and resting state. Further, the ventral attention network contained a number of brain regions that were uniquely important and corresponding to the MID task when predicting intelligence. In addition, shared brain regions were also discovered. In particular, (i) resting state and nback task shared a high number of important regions in the visual network; (ii) MID and SST tasks and resting state shared common important brain regions in the default mode network; and (iii) several shared brain regions were discovered in the frontoparietal region that were common pairwise for the MID, SST, and nback tasks.

The shared brain regions common across three or more brain states are presented in Table 5 and visualized in Figure 8. The common brain regions shared across the three tasks lie in (i) the DMN (2 ROIs), the FP (1 ROI), and the DA network (1 ROI), when predicting fluid intelligence; (ii) in the FP (4 ROIs), the DMN (2 ROIs), the VA network (1 ROI), when predicting crystallized intelligence; and (iii) in the DA network (1 ROI), the FP (2 ROIs) and the DMN (1 ROI), when predicting the total composite intelligence. When comparing shared regions between two tasks and rest for intelligence prediction, these
regions were distributed in (i) the visual and VA network, along with regions in the DMN when predicting fluid intelligence; and (ii) in the DMN and the visual network when predicting crystallized intelligence.

The shared regions between two tasks and rest when predicting total composite intelligence were found to be overlapping with the discoveries for fluid and crystallized intelligence prediction, except for one region: region 167, which was assigned to the frontoparietal module and lies in the right supramarginal gyrus. Finally, regions 257 belonging to the DMN, and 277 belonging to the FP network, along with region 115 (MNI coordinates [-23.4, 61, -6.8]), were discovered as important regions that were shared between all brain states when predicting fluid intelligence. In addition to these three regions, FP region 240 and region 121 (MNI coordinates [-23.4, 61, -6.8]), were shared between all brain states when predicting crystallized intelligence. Both regions 115 and 121 were not assigned to any functional module but are located in the superior frontal gyrus of the anterior prefrontal cortex in the left hemisphere. Regions 257 and 277 were discovered as important across all brain states when predicting total composite intelligence, which have already been captured when predicting fluid and crystallized intelligence.

### 3.4.3 Shared and differential brain regions across intelligence metrics

Our previous heatmaps in Figure 6 illustrated strong importance score correlations when predicting different intelligence metrics, based on a given brain state. This would suggest the presence of common brain regions that are activated when predicting different types of intelligence metrics. Table 6 lists the top 10 regions for predicting fluid intelligence and crystallized intelligence that are divided into shared regions that show up to be important when predicting both intelligence metrics, as well as differential regions that are important for predicting fluid intelligence but not crystallized intelligence and vice versa. Each of the three tasks have more shared regions with respect to fluid and crystallized intelligence, and fewer differential regions. In contrast, important brain regions at resting state are seen to be more differential when predicting fluid and crystallized intelligence. These regions are also visually illustrated in Figure 9.

| Brain states          | Shared brain regions |
|-----------------------|----------------------|
| **Fluid intelligence**|                      |
| 3 tasks               | 52, 114, 115, 257, 277 |
| MID, SST, and rest    | 78, 79, 94, 115, 220, 240, 257, 277 |
| MID, nback, and rest  | 98, 115, 141, 257, 263, 277, 309 |
| SST, nback, and rest  | 115, 211, 257, 277 |
| 3 tasks and rest      | 115, 257, 277 |
| **Crystallized intelligence** |                |
| 3 tasks               | 78, 114, 115, 121, 240, 241, 257, 277, 320 |
| MID, SST, and rest    | 94, 115, 121, 150, 240, 257, 277 |
| MID, nback, and rest  | 115, 121, 140, 240, 257, 258, 277 |
| SST, nback, and rest  | 115, 121, 240, 252, 257, 277 |
| 3 tasks and rest      | 115, 121, 240, 257, 277 |
| **Total composite intelligence** |          |
| 3 tasks               | 52, 167, 257, 277 |
| MID, SST, and rest    | 150, 257, 277 |
| MID & nback & rest    | 98, 257, 258, 263, 277, 309 |
| SST & nback & rest    | 211, 252, 257, 277 |
| 3 tasks & rest        | 257, 277 |

| Region information     | Functional module | Region index | MNI coordinates |
|------------------------|-------------------|--------------|-----------------|
| **Left hemisphere**    |                   |              |                 |
| Frontoparietal         | 78                | [-40.3, 50.4, -4.8] |
| Dorsal attention       | 52                | [-42.9, -45.0, 43.0] |
| Default mode           | 94                | [-39.3, -73.9, 38.3] |
|                        | 114               | [-27.5, 53.6, 0.0] |
|                        | 150               | [-6.5, 54.7, 18.1] |
| Visual                 | 98                | [-34.2, -86.6, -0.5] |
|                        | 140               | [-25.2, -97.2, -7.9] |
|                        | 141               | [-22.6, -81.7, -11.7] |
| Ventral attention      | 79                | [-47.2, 39.0, -9.1] |
| None                   | 115               | [-23.8, 52.2, -12.8] |
|                        | 121               | [-23.8, 52.2, -12.8] |

| **Right hemisphere**   |                   |              |                 |
| Frontoparietal         | 167               | [47.9, -42.5, 41.5] |
|                        | 240               | [42.8, 48.3, -5.1] |
|                        | 277               | [28.4, 57.0, -5.1] |
|                        | 320               | [30.9, 52.2, 9.9] |
| Dorsal attention       | 211               | [38.8, -42.6, 40.4] |
|                        | 252               | [23.0, -66.4, 51.8] |
| Default mode           | 220               | [48.9, -53.0, 28.6] |
|                        | 257               | [7.4, -69.3, 49.9] |
| Visual                 | 258               | [35.4, -77.1, 21.1] |
|                        | 263               | [31.7, -85.2, 24] |
|                        | 309               | [20.4, -87.3, -6.6] |
| Ventral attention      | 241               | [48.1, 38.3, -9.2] |

### 4 DISCUSSION

In this article, we have performed a comprehensive comparison of different types of fMRI features (static FC, dynamic FC, and region-level fMRI time series data) for cognitive prediction across different intelligence metrics and fMRI experiments, using the large-scale ABCD study that involved nearly 7000 individuals. We have implemented a five-fold cross-validation scheme to validate our analysis and findings. The sample sizes in both the training and testing sets are well above the recommendation from Marek et al. (2022) for achieving good reproducibility and generalizability. We used a deep neural network involving a novel bi-LSTM approach that naturally incorporated time-
### TABLE 6  Shared and differential top 10 brain regions in predicting fluid and crystallized intelligences.

| Experiment | Common regions | Unique for fluid | Unique for crystallized |
|------------|----------------|------------------|-------------------------|
| **MID**    | 78: L.FP [−40.3, 50.4, −4.8] | 38: L.SM [−35.8, −29.7, 54.5] | 53: L.SM [−51.5, −11.9, 29.7] |
|            | 79: L.VA [−47.2, 39, −9.1]    | 141: L.Vis [−22.6, −81.7, −11.7] | 114: L.DM [−27.5, 53.6, 0] |
|            | 98: L.Vis [−34.2, −86.6, −0.5] |                               |                         |
|            | 115: L.none [−23.4, 61, −6.8]  |                               |                         |
|            | 212: R.SM [53.9, −8.3, 26.1]   |                               |                         |
|            | 241: R.VA [48.1, 38.3, −9.2]    |                               |                         |
|            | 277: R.FP [28.4, 57, −5.1]     |                               |                         |
|            | 309: R.Vis [20.4, −87.3, −6.6] |                               |                         |
| **SST**    | 13: L.RT [−14.4, −57.8, 18.4] | 277: R.FP [28.4, 57, −5.1]     | 211: R.DA [38.8, −42.6, 40.4] |
|            | 94: L.DM [−39.3, −73.9, 38.3] |                               |                         |
|            | 116: LDM [−5.9, 54.8, −11.3]   |                               |                         |
|            | 117: LDM [−6.8, 38.2, −9.4]    |                               |                         |
|            | 151: LDM [−15.7, 64.7, 13.7]   |                               |                         |
|            | 167: R.FP [47.9, −42.5, 41.5]  |                               |                         |
|            | 219: R.CO [57.5, −40.3, 34.7]  |                               |                         |
|            | 241: R.VA [48.1, 38.3, −9.2]   |                               |                         |
|            | 279: R.DM [7.2, 48.4, −10.1]   |                               |                         |
| **nback**  | 89: L.CP [−12.7, −64.9, 31.8] | 103: L.CO [−55.1, −32.3, 23]   | 114: L.DM [−27.5, 53.6, 0] |
|            | 97: L.Vis [−31.3, −84.2, 9]    | 257: R.DM [7.4, −69.3, 49.9]   | 115: L.none [−23.4, 61, −6.8] |
|            | 98: L.Vis [−34.2, −86.6, −0.5] |                               |                         |
|            | 140: L.Vis [−25.2, −97.2, −7.9] |                               |                         |
|            | 141: L.Vis [−22.6, −81.7, −11.7] |                               |                         |
|            | 258: R.Vis [35.4, −77.1, 21.1] |                               |                         |
|            | 263: R.Vis [31.7, −85.7, 2.4]  |                               |                         |
|            | 309: R.Vis [20.4, −87.3, −6.6] |                               |                         |
| **rest**   | 15: L.Vis [−11.3, −83.2, 3.9]  | 1: L.DM [−11.2, −52.4, 36.5]   | 90: L.Vis [−13.7, −77.4, 26.6] |
|            | 115: L.none [−23.4, 61, −6.8]  | 97: L.Vis [−31.3, −84.2, 9]    | 255: R.Vis [17.6, −78.3, 34] |
|            | 140: L.Vis [−25.2, −97.2, −7.9] | 162: R.DM [12.3, −51.6, 34.5] | 277: R.FP [28.4, 57, −5.1] |
|            | 175: R.Vis [15.5, −74.1, 9.4]  | 310: R.Vis [5.1, −80.2, 23.1] | 311: R.Vis [14.6, −70.3, 23.3] |
|            | 252: R.DA [23, −66.4, 51.8]    |                               |                         |
|            | 258: R.Vis [35.4, −77.1, 21.1] |                               |                         |

Note: Region information in “a.b [c]” format. “a” indicates hemisphere: L—left hemisphere, R—right hemisphere. “b” indicates functional module assignment. “c” includes the coordinates of the region centroid in the MNI space.

---

**FIGURE 8**  Brain maps for common top 10% brain regions that were shared across three or more brain states.
varying features in the fMRI data. Our analysis illustrated distinct advantages in prediction based on time-varying fMRI features compared to prediction based on the routinely used static FC features across brain states and intelligence metrics. We note that the static FC-based prediction results presented in our analysis are comparable with the corresponding results from Chen et al. (2022) who examined a subset of the individuals from release 2.0 of the ABCD data. Moreover, for our analysis based on only one brain state, region-level FC features generally had the best performance when predicting fluid intelligence, and the performance was comparable to that based on dynamic FC features when predicting total composite intelligence. These results suggest the strong potential of using region-level fMRI time series data for predicting intelligence metrics based on one brain state, especially given the considerably smaller number of features (by orders of magnitude) compared to dynamic FC, which also leads to fast computation and straightforward interpretability. On the other hand, dynamic FC-based prediction yields significantly better results compared to both static FC and region-level time series-based prediction when rest and task fMRI are combined. This points to the promising potential of dynamic FC to be used as a neuroimaging biomarker for multi-task prediction analysis. One potential area of concern when using dynamic FC for multi-task prediction is the large computational resources that required specialized computational architecture such as GPU computing used in our implementation. In summary, our analysis develops and empirically validates novel temporally varying neuroimaging biomarkers for intelligence prediction based on single as well as multiple fMRI experiments.

Another innovative and practical aspect of our analysis is the feature selection mechanism that is implemented via an $L_0$ regularization on the time-varying input features in the bi-LSTM framework. A cross-validation analysis yielded high test–retest reliability of the feature importance scores generated under this approach, which showed strong reproducibility and validated the robustness of our method. We note that while more interpretable linear regression models can also be used for feature selection, we do not use these models for analyzing important brain regions due to their poor predictive ability. Using the proposed feature selection mechanism, we provided a thorough analysis of shared and differential brain regions that are discovered as important when predicting different intelligence metrics across varying brain states. Our analysis provides complimentary results to existing literature that has shown shared and unique brain network features between task and rest fMRI that drive cognition and intelligence (Chen et al., 2022; Elliott et al., 2019).

Both the default mode network and frontoparietal networks are shown to contain a number of brain regions that were common across the three tasks when predicting intelligence. Of these regions, region number 257 in the default mode network and region number 277 in the frontoparietal network were discovered as important across task and rest fMRI experiments, when predicting fluid, crystallized, and total composite intelligence. Region 257 lies in the right precuneus Brodmann area 7, which has been shown to be important in visuomotor coordination (Cavanna & Trimble, 2006). Region 277 lies in the right superior frontal gyrus of the brain, which has been associated with cognitive control of impulsive responses (Hu et al., 2016). Additionally, region number 115, which did not belong to any prespecified functional network, was also shown to be an important region across all brain states when predicting both fluid and crystallized intelligence. This region lies in the left superior frontal gyrus of the brain that has been shown to be associated with higher cognitive functions such as working memory and spatial cognition (du Boisgueheneuc et al., 2006).

The important brain regions shared across the fluid and crystallized intelligence metrics appear to be more widely distributed for the MID task (scattered across the frontoparietal, visual, ventral attention, and sensory motor networks), but they were almost exclusively concentrated within the visual network for prediction based on the nback task, and the shared regions were largely concentrated within the default mode network for prediction based on the SST task.
Moreover, the preponderance of shared important brain regions in the visual network when predicting intelligence metrics based on the nback task is not surprising given that this task involves visual patterns in the experiment, which evoke strong activations in the visual network that appear to drive changes in both fluid and crystallized intelligence. The specific brain regions in the visual network that were recruited during the nback task and are important for predicting both fluid and crystallized intelligence include regions 97 and 98 in the left middle occipital gyrus, region 140 in the left inferior occipital gyrus, and region 141 in the left lingual gyrus, which are all parts of the secondary visual cortex in the left hemisphere. Additional important regions include regions 258 and 263 in the right middle occipital gyrus, and region 309 in the right lingual gyrus, which all belong to the secondary visual cortex in the right hemisphere. These findings are consistent with existing literature that have discovered the importance of occipital regions (Miró-Padilla et al., 2020) and the visual cortex (Hallenbeck et al., 2021) in the context of the nback task.

On the other hand, shared regions that were common for predicting fluid and crystallized intelligence based on the SST task involved several brain areas from the default mode network. In the SST task, the participant responds to an arrow stimulus, by selecting one of two options, depending on the direction in which the arrow points. If an audio tone is present, the subject must withhold making that response (inhibition). The test consists of an initial practice phase, and a subsequent task phase where the auditory stop signal is generated according to the design unknown to participants. Given that the SST task involves a learning (practice) phase, and transitions between unknown stops and task, the recruitment of the default mode network in the SST task that is associated with cognitive ability is supported by previous literature. Existing work shows that the DMN is recruited in switching tasks, in the case of a demanding shift from a cognitive context to a different one (Crittenden, 2015), as well as during decision-making (Smith et al., 2021) and when subjects have to automatically apply learned rules (Vatansever et al., 2017).

For resting state, the shared important regions for predicting fluid and crystallized intelligence primarily belong to the visual network, while the distinct brain regions that show differentiation when predicting fluid and crystallized intelligence are concentrated in the visual and default mode network. These brain regions have been shown in previous literature to be connected to individual's intelligence performance (Dubois et al., 2018; Gray et al., 2003; Jiang et al., 2020; Santarencchi et al., 2015; Song et al., 2009; van den Heuvel et al., 2009; Yuan et al., 2012).

### 4.1 Potential limitations and future directions

The challenges of test–retest reliability in fMRI measurements are acknowledged, with previous studies showing poor reliability in task-related fMRI measures for adult populations (Elliott et al., 2020; Noble et al., 2021). Another recent study (Kennedy et al., 2022) examined data from the ABCD study and found poor reliability and longitudinal stability of task-fMRI measures in children. Unlike these studies that directly evaluated test–retest reliability from repeated fMRI experiments, the analysis in the present study focuses on the role of fMRI measurements and functional connectivity features in predicting intelligence. It detects important brain regions with higher weights in the prediction model and verifies their validity by computing within-sample test–retest reliability using an ICC metric obtained through 5-fold cross-validation (Kennedy et al., 2022). In contrast to the low reliability reported in previous studies, the analysis reveals higher within-sample reliability indicated by ICC scores. The consistency of the estimated feature importance scores across the five cross-validation folds allows for downstream analysis of shared and differential brain regions based on averaged importance scores. While the within-sample reliability may be inflated (which is a potential limitation), we note that the primary goal is to obtain relative rankings of importance scores and test–retest reliability estimates for different brain regions and methods. The analysis demonstrates substantial improvements in test–retest reliability compared to existing methods and we expect the relative rankings to remain stable even with a switch from a 5-fold to 5-fold cross-validation for out-of-sample reliability.

Although low test–retest reliability in preprocessed fMRI measurements presents challenges for predictive analysis and may potentially impact the validity of the findings, we do not look at this in detail in the current manuscript as it is not our focus. Nonetheless, there are some potential remedies that may be explored. For example, we would note that multi-task analysis based on whole-brain data and combining task and rest fMRI experiments is expected to boost the reliability levels as noted in Zuo et al. (2019) and Noble et al. (2019). Other possible approaches to tackle such issues include first computing the test–retest reliability based on the preprocessed fMRI measurements derived from repeated sessions for each individual, and subsequently selecting only a subset of samples with high reliability for downstream analysis. Another possible alternative may include a suitable mechanism to accommodate heterogeneity across subjects in our analysis. These strategies will be explored in more detail in future work.

An additional consideration concerns the fact that our analysis followed the routine practice of not aligning the fMRI time-series across subjects or experiments when performing the principal components-based prediction (Dubois et al., 2018; Jiang et al., 2020; Sen & Parhi, 2020; Vieira et al., 2022). We expect the effect of non-alignment to be fairly nominal (if any) given that the principal components are extracted for each individual and each session separately, which does not require strict alignment. If required, there are some proposals for aligning functional curves in the literature such as (Dai et al., 2019), which can be potentially leveraged for our analysis as an initial step before the prediction analysis. Finally, a close examination of Figure 6 reveals that for a given brain state, the importance scores for predicting different types of intelligence metrics show high correlations (off-diagonal panels). This implies the potential for transfer learning under the proposed approach where the fitted model for predicting a particular intelligence metric can be used for predicting another type of intelligence score based on the same type of fMRI.
experiment (brain state), which is of increasing interest in neuroimaging studies (Schirmer et al., 2021).

5 | CONCLUSION

In this work, our comprehensive comparison of predictive abilities across different types of fMRI features and involving several task and rest fMRI experiments, adds to the limited understanding in the current literature. The performance in predicting fluid, crystallized, and total composite intelligence using static FC is comparable to that in Chen et al. (2022), while much superior performance is achieved using dynamic FC and region-level fMRI time series. Such region-level fMRI time series features provide comparable predictive performance compared to dynamic FC features extracted from a single brain state but involve massively smaller number of features and hence are much more computationally tractable as well as interpretation friendly. On the other hand, dynamic FC features demonstrate considerable advantages in multi-task prediction analysis integrating task and rest fMRI experiments. Moreover, the predictive advantages under multi-task analysis are clearly evident compared to prediction based on just one brain state. We hope that the conclusive evidence from our large-sample analysis will result in a paradigm shift in terms of promoting the use of temporally varying fMRI features and potentially incorporating both task and rest fMRI data for intelligence prediction. It would potentially be of interest to validate this conclusion for other behavioral traits in the future.

AUTHOR CONTRIBUTIONS
Yang Li: methodology, software, writing – original draft, visualization; Xin Ma: data curation, methodology, software, writing – original draft/review and editing, visualization; Raj Sunderraman: funding acquisition; Shihao Ji: supervision, conceptualization; Suprateek Kundu: conceptualization, writing – original draft/review and editing, funding acquisition.

FUNDING INFORMATION
This work was supported by NIH awards R01AG071174, R01MH120299, and R01CA244845. We would also greatly acknowledge the support of Cisco Systems, Inc. for its university research fund to this work.

CONFLICT OF INTEREST STATEMENT
The authors declare no conflict of interests in this work.

DATA AVAILABILITY STATEMENT
The data used in the analysis was downloaded from the National Institute of Mental Health (NIMH) Data Archive (https://nda.nih.gov), which the authors have obtained permission to use.

ORCID
Yang Li https://orcid.org/0009-0006-4018-5582
Xin Ma https://orcid.org/0000-0001-6631-9223

REFERENCES
Abrol, A., Fu, Z., Salman, M., Silva, R., Du, Y., Plis, S., & Calhoun, V. (2021). Deep learning encodes robust discriminative neuroimaging representations to outperform standard machine learning. Nature Communications, 12(1), 1–17.
Boisguehenuec, F., Levy, R., Volle, E., Seassau, M., Duffau, H., Kinkingnehn, S., Samson, Y., Zhang, S., & Dubois, B. (2006). Functions of the left superior frontal gyrus in humans: A lesion study. Brain, 129(12), 3315–3328.
Bullmore, E., & Sporns, O. (2009). Complex brain networks: Graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience, 10(3), 186–198.
Cavanna, A. E., & Trimble, M. R. (2006). The precuneus: a review of its functional anatomy and behavioural correlates. Brain, 129, 564–583.
Chen, J., Li, X., Calhoun, V. D., Turner, J. A., van Erp, T. G. M., Wang, L., Andreassen, O. A., Agartz, I., Westlye, L. T., Jonsson, E., Ford, J. M., Mathalon, D. H., Macciardi, F., O’Leary, D. S., Liu, J., & Ji, S. (2021). Sparse deep neural networks on imaging genetics for schizophrenia case-control classification. Human Brain Mapping, 42(8), 2556–2568.
Chen, J., Tam, A., Kebets, V., Orban, C., Ooi, L. Q. R., Asplund, C. L., Marek, S., Dosenbach, N. U. F., Eickhoff, S. B., Bzdok, D., Holmes, A. J., & Yeo, B. T. T. (2022). Shared and unique brain network features predict cognitive, personality, and mental health scores in the ABCD study. Nature Communications, 13(1), 1–17.
Chen, P.-Y., Chen, C. L., Hsu, Y. C., Cam-CAN, & Tseng, W. I. (2020). Fluid intelligence is associated with cortical volume and white matter tract integrity within multiple-demand system across adult lifespan. NeuroImage, 212, 116576.
Crittenden, B. M., Mitchell, D. J., & Duncan, J. (2015). Recruitment of the default mode network during a demanding act of executive control. eLife, 4, 116576.
Dai, M., Zhang, Z., & Srivastava, A. (2019). Analyzing dynamical brain functional connectivity as trajectories on space of covariance matrices. IEEE Transactions on Medical Imaging, 39(3), 611–620.
Desikan, R. S., Ségonne, F., Fischl, B., Quinn, B. T., Dickerson, B. C., Blacker, D., Buckner, R. L., Dale, A. M., Maguire, R. P., Hyman, B. T., Albert, M. S., & Killiany, R. J. (2006). An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage, 31(2), 968–980.
Dhamala, E., Janmion, K. W., Jaywant, A., Dennis, S., & Kuceyeski, A. (2021). Distinct functional and structural connections predict crystallised and fluid cognition in healthy adults. Human Brain Mapping, 42(10), 3102–3118.
Dubois, J., Galdi, P., Paul, L. K., & Adolphs, R. (2018). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. Philosophical Transactions of the Royal Society, B: Biological Sciences, 373(1756), 20170284.
Elliott, M. L., Knodt, A. R., Cooke, M., Kim, M. J., Melzer, T. R., Keenan, R., Ireland, D., Ramakha, S., Poulton, R., Caspi, A., Moffitt, T. E., & Hariri, A. R. (2019). General functional connectivity: Shared features of resting-state and task fMRI drive reliable and heritable individual differences in functional brain networks. NeuroImage, 189, 516–532.
Elliott, M. L., Knodt, A. R., Ireland, D., Morris, M. L., Poulton, R., Ramakha, S., Sison, M. L., Moffitt, T. E., Caspi, A., & Hariri, A. R. (2020). What is the test-retest reliability of common task-functional MRI measures? New empirical evidence and a meta-analysis. Psychological Science, 31(7), 792–806.
Fair, D. A.-D., Miranda-Dominguez, O., Snyder, A. Z., Perrone, A., Earl, E. A., Van, A. N., Koller, J. M., Feczko, E., Tisdall, M. D., van der Kouwe, A., Klein, R. L., Mirro, A. E., Hampton, J. M., Adeyemo, B.,

Raj Sunderraman https://orcid.org/0000-0001-6822-6629
Shihao Ji https://orcid.org/0000-0002-3573-5379
Suprateek Kundu https://orcid.org/0000-0002-1767-4875
