Received: 1 July 2021 | Revised: 1 February 2022 | Accepted: 13 February 2022

DOI: 10.1002/hbm.25817

RESEARCH ARTICLE

WILEY

Decoding task specific and task general functional architectures of the brain

Sukrit Gupta | Marcus Lim | Jagath C. Rajapakse

School of Computer Science and Engineering, Nanyang Technological University, Singapore

Correspondence
Jagath C. Rajapakse, School of Computer Science and Engineering, Nanyang Technological University, Singapore, 639798.
Email: asjagath@ntu.edu.sg

Funding information
Ministry of Education, AcRF Tier-2 Grant, Singapore, Grant/Award Number: 2EP20121-0003; Ministy of Education, AcRF Tier 1 Grant, Singapore, Grant/Award Number: 2019-T1-002-057

Abstract

Functional magnetic resonance imaging (fMRI) is used to capture complex and dynamic interactions between brain regions while performing tasks. Task related alterations in the brain have been classified as task specific and task general, depending on whether they are particular to a task or common across multiple tasks. Using recent attempts in *interpreting* deep learning models, we propose an approach to determine both task specific and task general architectures of the functional brain. We demonstrate our methods with a reference-based decoder on deep learning classifiers trained on 12,500 rest and task fMRI samples from the Human Connectome Project (HCP). The decoded task general and task specific motor and language architectures were validated with findings from previous studies. We found that unlike intersubject variability that is characteristic of functional pathology of neurological diseases, a small set of connections are sufficient to delineate the rest and task states. The nodes and connections in the task general architecture could serve as potential disease biomarkers as alterations in task general brain modulations are known to be implicated in several neuropsychiatric disorders.

KEYWORDS
brain decoding, deep learning, functional connectivity, functional MRI, task general architecture, task specific architecture

1 | INTRODUCTION

The study of functional self-organization of the human brain during task performance is a widely researched area in cognitive neuroscience. There is significant functional modulation in the brain during task performance (Fox et al., 2005). Two types of task evoked alterations are known to shape brain states: task specific and task general alterations (Cole, Bassett, Power, Braver, & Petersen, 2014; Düzel et al., 1999; Norman & Shallice, 1986; Nyberg et al., 1996). *Task general* alterations are task independent alterations in the brain that occur across multiple tasks whereas *task-specific* alterations depend on the particular task being performed.

Aging (Meinzer et al., 2012) and multiple neurological disorders such as major depressive disorder (MDD) (Hamilton et al., 2011; Knyazev et al., 2018), attention deficit hyperactivity disorder (ADHD) (Mills et al., 2018), mild cognitive impairment (MCI) (Melrose et al., 2018), and schizophrenia (Haatveit et al., 2016) have been characterized by anomalous task general architecture. Task general alterations are further classified as *task positive* and *task negative*, depending on whether the neurons and their connections in a region are activated or deactivated during the task performance. We model the alterations in the brain functional connectivity as *positive task general* and *negative task general* depending on the presence or absence of functional connectivity during task states relative to the rest-state.

Advances in neuroimaging have made it easier to study dynamic in vivo changes in regional relationships during different task and the rest state. Multiple studies have investigated the modulations in brain functional architecture during task performance. Betti et al. (2013) used group averaged connectivity from magnetoencephalography (MEG) and fMRI scans for 12 subjects to recover rest and task functional connectivity and study the alterations during rest and task. They concluded that although the resting-state and task network interactions had similar network topography, clear differences existed in the frequency domain. Krienen, Yeo, and Buckner (2014) acquired task fMRI data from 48 subjects for 14 distinct sub-tasks (grouped into passive, sensory, and motor tasks) and by performing correlation and seed-based analysis on the group averaged data, found that the task modulations varied around a common central tendency across all task states. Cole, Bassett, et al. (2014) used two task datasets, one containing 64 tasks and the other containing 7 tasks (from the Human Connectome Project), and found that the central tendency for each functional connection during varied tasks was the one represented during the rest state. They found that a stable intrinsic functional network architecture that is present across rest and tasks, and a few but statistically significant alterations had shaped task states. In another study, Cole, Ito, Bassett, and Schultz (2016) selected the first principal component of task activation to detect task general architecture. These studies suggest that although there are small-scale alterations in the brain functional architecture, there is an intrinsic stable functional backbone that remained unchanged during both the rest and task state (Betti et al., 2013; Cole, Bassett, et al., 2014; Krienen et al., 2014).

Although multiple recent studies (Gordon et al., 2017; Gratton et al., 2018; Xie et al., 2018) have depicted the role of subject-level characteristics, most previous efforts studying alterations in functional connectivity involved studying alterations at the group level. Gratton et al. (2018) used data for nine subjects with 10 hr of scan data per subject (10 scans of 1 hr each containing resting state and four task states) from the Midnight Scan Club (Gordon, Laumann, Gilmore, Newbold, et al., 2017) and studied the dependence of functional connectivity on task states, subjects, and scan sessions. They found that while the task states and the sessions shaped the functional connectivity patterns, common organizational principles, and subject-specific features had a higher impact. Similarly, Xie et al. (2018) used task fMRI scans and k-means clustering to separate the effects of subject-level functional connectivity and tasks to determine what caused changes in group functional connectivity. They found that subject-level functional connectivity was the most dominant factor in group-level functional connectivity variability. These results, combined with previous studies on resting state brain functional architecture pointing towards significant intersubject variability (Finn et al., 2015; Gordon et al., 2017; Gordon, Laumann, Adeyemo, & Petersen, 2017; Mueller et al., 2013; Wang et al., 2015) that necessitates consideration of subject variability in the functional connectome while studying alterations during task performance. However, earlier studies had only considered group-level functional connectivity. In this study, instead of studying group level alterations, we investigate intersubject variability by explicitly modeling subject-level alterations.

Another concern with previous studies is that they only captured linear relationships between functional connectivity at rest or at task. In a recent study, He et al. (2020) used resting state functional connectivity to predict behavior by using traditional machine learning algorithms and multiple deep neural networks (DNNs) proposed for functional connectivity based analysis. They demonstrated that kernel regression applied on functional connectivity was sufficient and DNNs did not bring additional value as far as predicting individual behavior was concerned. We test this with classification of rest and task functional connectivity by using traditional supervised learning models along with the highly parametric DNNs. We trained classifiers on both models on individual subject rest-state and task state data, and thereafter distinguished features responsible for different tasks from the trained model. We refer to identification of salient features of the network as decoding (different from the decoding performed in autoencoders). We found that although all classifiers gave a high accuracy for classification, DNNs were able to classify with a better accuracy than the shallow models such as support vector machines (SVM). We uncover the task general architecture by decoding a binary classifier trained for classifying rest and multiple task states. Using such obtained task general architecture, we demonstrated decoding of task specific brain architectures for the motor and language tasks.

Recent attempts have been made to evaluate the salience given to input features of neural network models classifying neuroimaging data. Jang, Plis, Calhoun, and Lee (2017) used the trained weights in different layers of a feedforward neural network and mapped them to four sensorimotor task states. Several works using a gradient-based decoder with different neural network architectures have culminated. Using a gradient-based decoder on a trained long short-term memory recurrent neural network, Li and Fan (2019) identified functional signatures in rest and task states. Floren, Naylor, Miikkulainen, and Ress (2015) identified brain regions involved in different visual processing tasks. However, these studies did not validate their derived feature set and issues like gradient saturation and discontinuities due to bias terms have not been addressed before using gradient-based decoders on neural networks (Shrikumar, Greenside, & Kundaje, 2017).

Recent methods such as Integrated Gradients (Sundararajan, Taly, & Yan, 2017), DeepLIFT (Shrikumar et al., 2017), and SHAP (Lundberg & Lee, 2017) attempt to ameliorate the issues with gradient-based approaches. These approaches find contributions of the neurons at each layer from those at the output layer and then backpropagate the contributions of all the layers to the input layer. DeepLIFT is one of the approaches that gives consideration to both negative and positive contributions, and computes the salience scores efficiently in a single pass, thus addressing the issues confounded in gradient based approaches. Gupta et al. (2019) used a DeepLIFT based decoder to derive salient features for distinguishing Alzheimer's disease and Autism patients from cognitively normal subjects. In the present article, we use DeepLIFT to study the microscale and mesoscale brain alterations that shape task related alterations in the brain (Cole et al., 2016). We used the state-of-the-art resting state and seven-task fMRI data from the Human Connectome Project (HCP) dataset to demonstrate our methods.

Using functional connectivity between brain regions as inputs to train DNN, we made the following novel contributions in this work:

- With our experiments on different models for rest versus task classification, we show that DNNs outperform simpler models such as SVM even with a few training samples. However, the classification accuracy of SVMs improves with more training data.
- We proposed a novel method that uncovers salient task general and task specific functional brain architectures. The task related architectures detected by our method were validated with previous studies.
- We show that although there is pronounced intersubject variability in task related brain architectures, the small subset of detected features identified by our method were able to provide near 100% classification accuracy while distinguishing tasks from the rest states.

Using feature salience scores, we identified functional connections and functional modular interactions that characterize task general, and motor and language task specific architectures of the brain. We discovered widespread system reorganization during task performance and were able to classify the rest versus all tasks, rest versus motor task, and rest versus language task with high accuracies by using a five-layer feedforward DNN. The high classification accuracy was maintained even while taking only a subset of decoded features classified as highly salient by our decoder. We confirmed that the default mode network (DMN), commonly known for its deactivation during task performance, formed positive correlations with the salience network (SN) (Elton & Gao, 2015). Similarly, we found an important role for the regions in the motor sub-systems for performing the motor task and left lateral temporal regions in performing the language task (Binder et al., 2011).

2 | METHODS

Let brain functional network $G = (\Omega, A)$ where Ω denotes the set of brain regions of interest (ROI) or nodes and $A = \{a_{ij}\}_{i,j\in\Omega}$ denotes the adjacency matrix of functional connectivity between different ROI. We obtained a symmetric A for each fMRI scan and input features x for the feedforward DNN were obtained by taking only the lower half of elements of the adjacency matrix. We used rest and task data gathered on healthy subjects in the HCP.

2.1 | Feedforward DNN

For each subject, we obtained (x, d) where $x = (x_i)$ is the input features and d is the task label. The output y of the DNN gives the probability of the input belonging to one of the labeled task classes from Table 1. We consider a DNN of L layers with $L - 1$ rectified linear unit (ReLU) layers and a final softmax layer.

Let the weights and biases of the layer l be given by W_l and b_l, respectively. The output h_l of layer $l \notin \{0, L\}$ is given by

$$h_l = \mathrm{ReLU}\left(W_l^T h_{l-1} + b_l\right) \tag{1}$$

For the input layer $l = 0$, $h_0 = x$. For the output softmax layer $l = L$, the output y is given by the probability of a sample x belonging to class k:

$$P(y = k|x) = \mathrm{softmax}\left(W_L^T h_{L-1} + b_L\right) \tag{2}$$

where $k \in \{1, \ldots, K\}$ represents the class label, the output layer weight $W_L = [w_{k,L}]$, and bias $b_L = (b_{k,L})$.

TABLE 1 Summary of HCP task and resting state data

Condition	Number of subjects	Number of time points	Description
Resting	756	1,200	Participants were asked to lie flat, with eyes open and maintain a relaxed fixation.
Emotion processing	806	176	Participants were tasked to match two simultaneously presented faces or shapes on two screens.
Gambling	829	253	Participants were given incentive to guess the number of a mystery card.
Language	805	316	Participants are presented with a question on short narrated stories.
Motor	826	284	Participants were asked to either move their fingers, toes or their tongue on cue.
Relational processing	802	232	Participants are presented with relational and control matching task.
Social cognition	807	274	Participants were presented with short videos and decided if the object movement were interrelated.
Working memory	828	405	Participants were asked to complete a two-back and a zero-back working memory tasks.

To learn the parameters of the network, the cross-entropy cost $J(\theta)$ was minimized:

$$J(\theta) = -E_x[\log P(y = d|x,\theta)] \tag{3}$$

where E_x is expectation taken over all the scan samples x and $\theta = \{(W_l,b_l)\}_{l=1}^{L}$ denotes all the parameters in the network. We used minibatch stochastic gradient descent learning with a fixed learning rate to learn the parameters θ.

2.2 | Salience of input features

Let f be neural network function mapping input x to output y. Let g be a simpler *explanation model* that is interpretable and an approximation of the network mapping f. Let the number of neurons in layer l be n_l. Using an appropriate reference, let us assign to each neuron i at layer l its contribution $C_{\Delta h_{i,l}\Delta h_{k,l+1}}$ to the change in the output of neuron k at layer $l+1$. Then, the change $\Delta h_{i,l}$ in the activation of the ith neuron of layer l due to the input relative to the reference is given by

$$\Delta h_{k,l+1} = \sum_{i \le n_l} C_{\Delta h_{i,l}} \tag{4}$$

when $l \in \{0,\ldots,L-2\}$ and

$$\Delta y = \sum_{i \le n_{L-1}} C_{\Delta h_{i,L-1}} \tag{5}$$

when $l = L - 1$.

In order to determine the feature salience scores, we evaluate three types of references; namely sample mean, sample mode, and taking a sample which has the least L_2 norm from the rest of the samples. We empirically determined the mean reference $\bar{x}$ to be the most representative of the samples. Given the reference input $\bar{x}$ and the original input x, we can substitute $\Delta y = f(x) - f(\bar{x})$ and $g(x) = f(x)$, giving us an equation for the model g:

$$g(x) = f(\bar{x}) + \sum_{i,l} C_{\Delta h_{i,l}\Delta y} \tag{6}$$

where the contribution of neurons in each layer l to the output y is given by $C_{\Delta h_{i,l}\Delta y}$ (Shrikumar et al., 2017):

$$C_{\Delta h_{i,l}\Delta y} = \Delta h_{i,l} \sum_{k,l} \frac{C_{\Delta h_{i,l}\Delta h_{k,l+1}}}{\Delta h_{i,l}} \frac{C_{\Delta h_{k,l+1}\Delta y}}{\Delta h_{k,l+1}} \tag{7}$$

where $C_{\Delta h_{i,l}\Delta h_{k,l+1}}$ can be computed from the Linear, Rescale, and RevealCancel rule (Shrikumar et al., 2017). We can get contribution of neurons in all layers $0 \le l < L$ to the output y by backpropagating contributions of layers to the input, using the chain-rule given by Equation (7).

The *salience scores* c for the input layer are evaluated as $c = \left(C_{\Delta h_{i,0}\Delta y}\right)$ such that each element gives the contribution of the corresponding input to the changes of the output. The DeepLIFT method (Shrikumar et al., 2017) implements computation of feature salience scores based on the changes of the output from a reference input, allowing information to propagate across the network layers even when the gradient is zero. We compute feature importance scores using Equation (6):

$$c = \underset{x}{\text{mean}} \left(f(\bar{x}) - g(x) - \sum_{i,l \in \{1,\ldots,L-1\}} C_{\Delta h_{i,l}\Delta y} \right) \tag{8}$$

The salience vector c represents the salience scores for the input features. The input features constitute the lower triangular matrix of the adjacency matrix. Vector c is mapped to a salience matrix $S = \{s_{ij}\}_{i,j\in\Omega}$ where s_{ij} represents the salience of the connectivity between ROI i and j.

While the salience matrix S from the DeepLIFT algorithm gives the salience of all connectivity features, we establish the validity of the decoded features by performing recursive feature elimination based on the salience scores of the features. If the features identified by the DeepLIFT based decoder are indeed salient, a subset of the most salient features will also give a high classification accuracy. In fact, recent work by the authors based on the premise that only a subset of regions and function connections are involved in task performance has shown promise (Gupta et al., 2021).

The input features that were insignificant were recursively removed by choosing connections whose salience $|s_{ij}| < t$ between brain regions i and j less than a threshold t and the neural network was retrained on the reduced feature set. The threshold was correspondingly set at the 90th percentile of significance. We repeated this experiment until only 0.1% initial features were left. We did not go below 0.1% as that corresponded to ≈34 features that were sufficient to give an accuracy close to the initial accuracy with all the features. The recursive elimination of irrelevant features not only tests the validity of the features identified by the decoder but also gives us the minimal subset of features needed for the classification.

2.3 | Decoding the task general brain architecture

We define decoding functional brain connectivity as identification of salient functional connectivity features that are crucial for classifying different brain states. The classification models are built using functional connectivity features for rest and task data while the DeepLIFT algorithm is used to find the salience of connectivity features.

Cole, Bassett, et al. (2014) used two task datasets, one containing 64 tasks and the other containing 7 tasks (from the Human Connectome Project), and found that the central tendency for each functional connection during varied tasks was the one represented during the rest state. They found that a stable intrinsic functional network architecture that is present across rest and tasks, and a few but statistically significant alterations had shaped task states. In another study, Cole et al. (2016) selected the first principal component of task activation to detect task general architecture. These studies suggest that although there are small-scale alterations in the brain functional

architecture, there is an intrinsic stable functional backbone that remained unchanged during both the rest and task state (Betti et al., 2013; Cole, Bassett, et al., 2014; Krienen et al., 2014).

Multiple previous studies (Betti et al., 2013; Cole et al., 2016; Cole, Bassett, et al., 2014; Krienen et al., 2014) point to the presence of an intrinsic backbone that remains unchanged during task performance. They also point to a common set of alterations that shape task states. While a task state can be considered to be amalgamation of both task general and task specific brain modulations over the rest state, the task general modulations can be represented by a common set of connectivity features that distinguish the brain at rest and during varied task states. In our formulation, connectivity features are the inputs to the DNN classifier and the task general architecture is defined by salient input features classifying the rest task versus all the tasks. We hypothesize that if we classify the brain at rest and during multiple varied task states, then the set of most salient connectivity features distinguishing the rest versus all the tasks represent the task general architecture of functional connectivity. To test this hypothesis, we identified the features with high salience scores distinguishing between rest and multiple dissimilar task states.

The sign of salience s_{ij} of a connection determines whether the *presence* or *absence* of functional connectivity between brain regions i and j leads to the differentiation of the rest-state from the task states. This information is in turn used to derive positive and negative task general architectures such that the presence of a positive task general connection during tasks is crucial in differentiating the task states from the rest state and vice-versa. We derived the task general architecture defined by the seven tasks in the HCP data.

2.4 | Decoding the task specific brain architectures

Once the task general architecture is derived, we obtain the task specific architecture with salient features obtained from a DNN trained to classify rest and the task under investigation. The salient features for the rest versus single task classification represent both task general and task specific brain modulations. We hypothesize that the salient connectivity features distinguishing a particular task from the rest represents both the task general and task specific architectures. Therefore, if we are able to compute task general architecture, we can derive the task specific architecture from the salient features of the DNN trained to classify some task and the rest state.

In order to demonstrate the determination of task specific architecture, we used the subject data from the motor and language tasks and performed classification for rest versus motor and rest versus language tasks. Let $S_{\text{rest/tasks}}$ denote the salience of connectivity features representing the task general architecture. If the salience of features in rest versus motor and rest versus language (containing both task general and task specific modulations) are denoted by $S_{\text{rest/motor}}$ and $S_{\text{rest/language}}$, respectively, then the motor and language task specific architectures are given by the salient features highlighted by $S_{\text{rest/motor}} - S_{\text{rest/tasks}}$ and $S_{\text{rest/language}} - S_{\text{rest/tasks}}$, respectively.

2.5 | Salient interactions between functional modules

Alterations in brain functional architecture are often discussed in terms of changes between different brain functional modules. The brain is organized into functional modules consisting of brain regions involved in specialized functional tasks. We quantified the importance of interactions among different brain functional modules with *modular salience maps*. Given functional modules a and b, the salience of interactions $m_{a,b}$ between them is obtained by

$$m_{a,b} = \underset{i \in a, j \in b}{\text{mean}} \{s_{i,j}\} \tag{9}$$

The modular salience map $\{m_{a,b}\}$ represents the average salience of the nodal connections within (when $a = b$) and between (when $a \neq b$) brain functional modules. The modular salience map provides a powerful framework to interpret and understand brain functional alterations in terms of how the interactions between different functional modules vary during different brain states (in this case, tasks).

2.6 | Intersubject variability in task related functional architectures

Multiple studies have pointed out the pronounced intersubject variability in the brain functional architecture both during rest (Finn et al., 2015; Gordon, Laumann, Adeyemo, Gilmore, et al., 2017; Gordon, Laumann, Adeyemo, & Petersen, 2017; Mueller et al., 2013; Wang et al., 2015) and task performance (Gordon, Laumann, Gilmore, Newbold, et al., 2017; Gratton et al., 2018; Xie et al., 2018). It is, therefore, important that the detected task general and task specific architectures consist of features that are stable across subjects. We measure the intersubject variability in the feature salience scores for different subjects as:

$$v = \underset{x}{\text{std}} \left(f(\bar{x}) - g(x) - \sum_{i,l \in \{1,\ldots,L-1\}} C_{\Delta h_{i,l} \Delta y} \right) \tag{10}$$

where std measures the standard deviation of the salience scores over multiple subjects. The mean of the salience scores is given by salience vector c (given by Equation (8)). We perform regression on the mean of c and the variability v of the salience scores to understand the relationship between the two sets of values.

3 | EXPERIMENTS AND RESULTS

3.1 | Dataset

We experimented with data from the Human Connectome Project (HCP) (van Essen et al., 2013), comprising of fMRI data from 897 healthy adults (mean age = 28.1 years, 390 females). The HCP currently hosts one of the largest open databases of fMRI data. We

used both the task evoked fMRI (tfMRI) for seven tasks and the rest-state fMRI (rfMRI) data of healthy subjects, collected from a Siemens 3 T Skyra scanner with TR = 720 ms, TE = 33.1 ms, flip angle = 52°, FOV = 208 × 180 mm, 2 mm × 2 mm isotropic voxel and 72 slices. We used the preprocessed data release that included removal of spatial artifacts, motion correction, intensity normalization, and denoising (Glasser et al., 2013). We further processed the task data to fit a finite impulse response (FIR) model to each of the 24 subtasks for the seven tasks (emotion processing, gambling, language, motor, relational processing, social cognition, and working memory) and removed the mean task evoked activation from time-series data. This step is necessary as these task related activations have been found to inflate task state functional connectivity inappropriately (Cole et al., 2019), especially with fMRI data. We also removed the rest periods from the task related time series before computing functional correlations between them.

Table 1 summarizes the available data for each task and the rest-state. The average number of time points for the task fMRI scans was 277. We divided each fMRI resting state BOLD scan of 1,200 time points into four scans of 300 time points each. From each of these truncated BOLD time series, we obtained four functional connectivity matrices. We obtained 690 subjects having scans for all seven tasks and the resting state since not all subjects had been able to complete every task and resting state scans. For each subject, by considering one for each encoding direction, we obtained eight resting state functional connectivity matrices and 14 task state matrices.

For each functional network, we selected anatomically and functionally diverse 264 ROIs identified by Power et al. (2011) and calculated the mean time series of all voxels within a sphere of radius 2.5 mm to represent activation of each ROI. We considered the Pearson correlation between time-series to derive the functional connectivity between the 264 ROIs covering the entire cerebral cortex.

3.2 | Building the classifiers

Using the functional connectivity between 264 ROIs identified by Power et al. (2011) as inputs, we trained classifiers on both task fMRI (tfMRI) and rest-state (rfMRI) data. We ensured that all the resting state and task scans of a subject are either in the test or the train set. We used Tensorflow (Abadi et al., 2016) and Keras (Chollet, 2018) python libraries to implement the DNN. Since there was a class imbalance with ≈9,650 task samples and ≈8,300 resting state samples, we also computed the mean absolute error (MAE), the mean of the absolute errors between the predicted outputs and the ground-truth labels. We first split the dataset into a train and test set in the ratio 4:1 and then performed fivefold cross validation to select the best model parameters on the train set. We ensured that all the samples for each subject were either in the train or test set (and note that the samples from the same subject were in the same fold during fivefold cross validation). We implemented different classifiers using the feedforward neural networks (FFN), convolution neural network (CNN) (Brown, Kawahara, & Hamarneh, 2018; Meszlényi, Buza, & Vidnyánszky, 2017) and support vector machines (SVM). The details of the different configurations for the models can be found in the Appendix S1.

For all the experiments (rest vs. all tasks and rest vs. specific tasks), we found that all the classifiers achieved classification accuracy above 90% (refer Table 2). However, SVM based classifiers gave lower accuracies while the FFN and CNN based classifiers gave similar accuracies for the same number of parameters. We, therefore, used FFN based classifiers with optimal hyperparameters for feature selection and decoding. We used the best FFN model after performing hyperparameter tuning on the train set for five different FFN parameter initializations. Since the accuracies for all the classifiers were high, we performed additional experiments to determine the time taken for error convergence for the SVM and the FFN classifiers and their performances for different subset of training samples (refer corresponding figures in the Appendix S1). We found that the test performance for the FFN converged faster than the corresponding SVM classifier for all the cases (refer corresponding figures in the Appendix S1).

3.3 | Decoding the task general brain architecture

We obtained the best set of hyperparameters for the rest versus all tasks classification as described in the previous section. For the best FFN model, the configuration is given in Table 3. Using the trained classifier, we computed the salience for the input connectivity features by using DeepLIFT. The salience scores were computed using samples from the train set. It was observed that the majority of features had positive salience (number: 31,381, range: 0 to 1891.6,

TABLE 2 Performance of the FFN and the SVM (on the test set) for rest versus all task classification using different fractions of features (MAE denotes mean absolute error)

	FFN		SVM	
% features	Accuracy	MAE	Accuracy	MAE
100%	99.81	0.0048	98.35	0.0164
10%	99.88	0.0022	98.80	0.0118
1%	99.98	0.0004	98.38	0.0161
0.1%	99.98	0.0011	93.43	0.0657

TABLE 3 Parameters of the four layer FFN used for classification of rest versus all tasks and rest versus language task

Parameter	Value
Input nodes	34,716
Hidden nodes (each layer)	34,716–100–32–2
Learning rate	0.0001
Batch size	8
Optimizer	Adam
Dropout	0.1

average: 81.96 ± 124.2) in comparison with features with negative salience (number: 3,335, range: 0 to −259.60, average: −29.62 ± 31.42). The histogram for distribution of salience scores and the feature salience score matrix $S_{rest/tasks}$ are available in Appendix S1 (Figure S1a,b).

In order to validate the decoded task general architecture, we recursively eliminated the input features, using the salience scores obtained from the FFN on the train set (refer Table 4). The connections in the task general architecture are shown in Figure 1. The modular salience maps highlighting the task general connections are shown in Figure 2a,b. The functional modules were derived from Power et al. (2011). Connections with positive and negative salience scores in $S_{rest/tasks}$ correspond to positive task general and negative task general connections, respectively. Using salient modular interaction scores, we computed statistically significant interactions having p-value <.05. We report significant task positive and task negative modular interactions in Table 5. It can be seen that the DMN, FP task control, and dorsal attention networks are involved in both task positive and task negative interactions.

3.4 | Decoding motor task specific brain architecture

In the motor task adapted from Buckner, Krienen, Castellanos, Diaz, and Yeo (2011), participants were asked to either tap their left or right

TABLE 4 Performance of different classification models on the HCP data with five-fold cross-validation (the individual tasks were classified against the rest task)

Classes	FFN accuracy	FFN MAE	CNN accuracy	CNN MAE	SVM accuracy	SVM MAE
Rest versus all tasks	99.75	0.0066	99.98	0.0011	98.26	0.0164
Emotion	99.70	0.0053	99.93	0.0020	97.06	0.0274
Gambling	99.78	0.0043	99.91	0.0017	97.31	0.0229
Language	99.88	0.0096	99.96	0.0017	98.78	0.0052
Motor	99.63	0.0068	99.94	0.0024	97.25	0.0217
Relational	99.57	0.0051	99.97	0.0015	97.08	0.0249
Social	99.72	0.0051	99.92	0.0025	97.40	0.0225
Working memory	99.53	0.0055	99.95	0.0021	97.18	0.0237

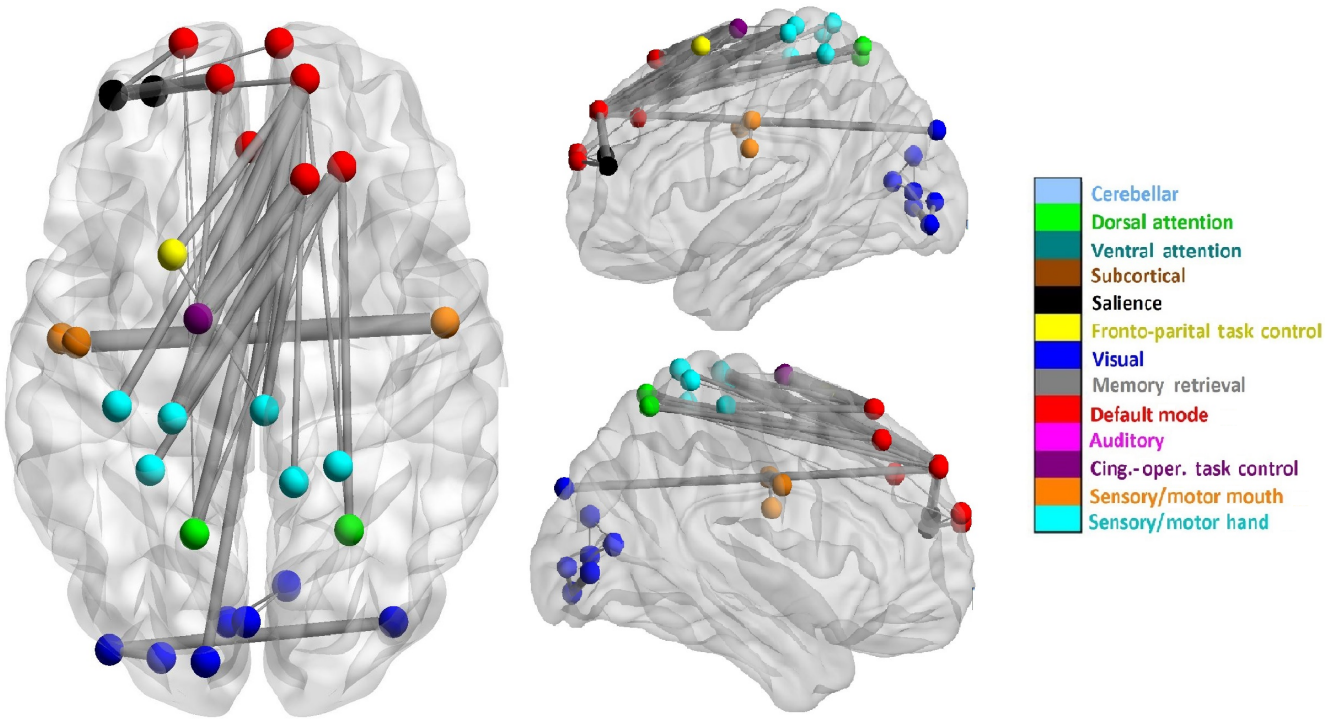
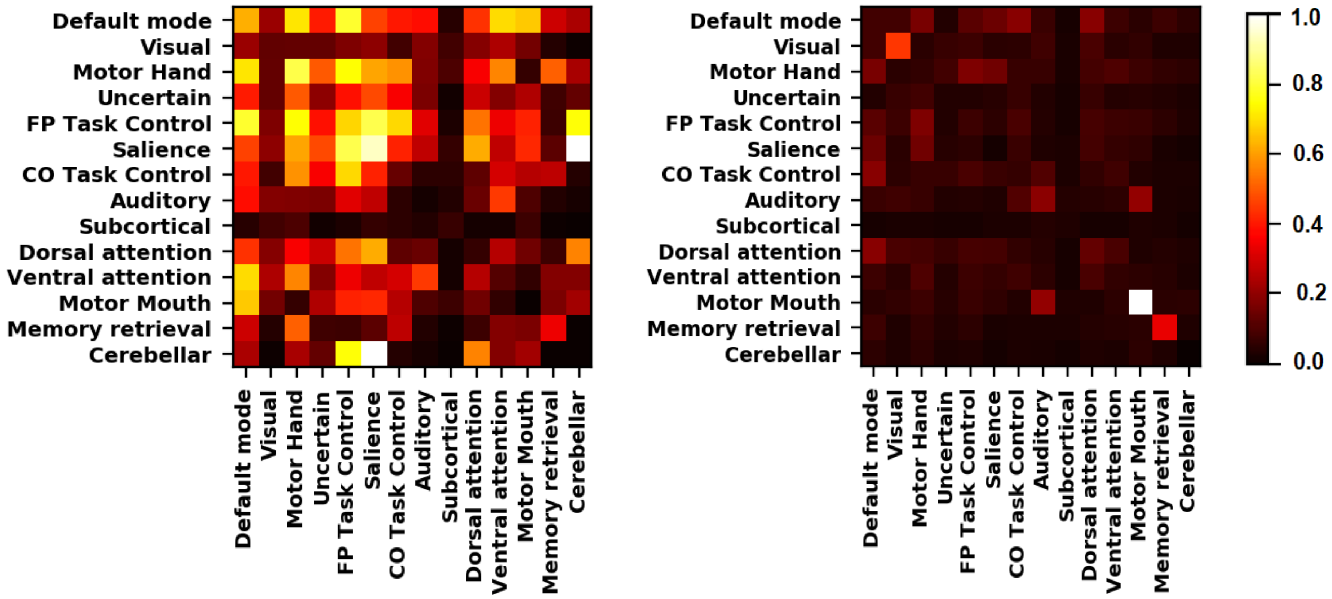


FIGURE 1 The decoded task general brain architecture obtained from consensus over multiple runs. Edge thickness corresponds to the number of runs these connections were classified as relevant for distinguishing rest and task states, and the node color corresponds to the node's modular membership



(a) Negative task general (b) Positive task general

FIGURE 2 The task general architectures in the brain at the modular level. The part figures (a: negative task general) and (b: positive task general) depict the modular salience matrices for negative and positive task general connections, respectively

TABLE 5 Significant task general modular interactions

Module	Module	*p*-value
Positive task general		
Motor mouth	Motor mouth	$<10^{-3}$
Visual	Visual	$<10^{-3}$
Memory retrieval	Memory retrieval	$<10^{-3}$
Negative task general		
Cerebellar	Salience	$<10^{-3}$
Salience	Salience	$<10^{-2}$
Salience	FP task control	<0.02
Motor hand	Motor hand	<0.02
Default mode	FP task control	<0.02
Cerebellar	FP task control	<0.03
FP task control	Motor hand	<0.03
Default mode	Motor hand	<0.04
Default mode	Ventral attention	<0.05
CO task control	FP task control	<0.05
FP task control	FP task control	<0.05

fingers, or squeeze their left or right toes, or move their tongue on cue. Each block of movement style lasted 12 s and is preceded by a 3 s cue. In each run, there were 13 blocks of which there were two tongue movements, four hand movements, four foot movements and three 15 s fixation blocks.

The configuration of the best FFN model derived for classifying rest versus motor task is given in Table 6. We achieved an average

TABLE 6 Parameters of the five layer FFN used for classification of rest versus motor task

Parameter	Value
Input nodes	34,716
Hidden nodes (each layer)	34,716–20–10–8–2
Learning rate	0.0001
Batch size	8
Optimizer	Adam
Dropout	0.1

accuracy of 99.65 and 97.82% by using all the features for the rest versus motor task with the FFN and SVM (linear kernel, $C = 0.001$) models, respectively. We show the results for the recursive feature elimination experiments with both FFN and SVM in Table 7. In order to obtain the task specific architecture, we used the salience scores of the connections for the rest versus motor task and subtracted the task general scores. We averaged the feature salience scores $S_{\text{rest/motor}}$ for all the seeds and subtracted $S_{\text{rest/tasks}}$ from $S_{\text{rest/motor}}$. In Figure 3, we show the top 0.1% relevant task specific features corresponding to the motor task computed as consensus over multiple runs.

Using the task specific salience scores, we obtained the module salience maps for both the positive and negative task specific features for the motor task (refer Figure S2a,b) and obtained the significant modular interactions (with *p*-value <.05). The significant task specific motor modular interactions involved the somatomotor mouth, somatomotor hand, CO Task Control, Dorsal attention, and the FP Task Control modules (Table 8).

3.5 | Decoding language task specific brain architecture

The language task was proposed by Binder et al. (2011), where participants were presented with short narrated stories and given a two-alternative forced-choice question about the topic of the story. The two runs that were interleaved with four blocks of a story task and four blocks of a math task. Story blocks presented brief sentences from Aesop's fables, followed by two-choice questions querying the topic of the story, whereas math blocks were presented aurally and required the subject to compute addition and subtraction problems and chose the result among the two choices by a finger tap feedback.

For the rest versus language task, the best FFN configuration was the same as rest versus all tasks (given in Table 3). For five seeds, we achieved an average accuracy of 99.86% using all the features for the language task classification. For the selected SVM model (linear kernel, $C = 1.0$), we were able to achieve a high classification accuracy with the whole feature set. We averaged the feature salience scores $S_{\text{rest/language}}$ for all the seeds and obtained the language task specific feature salience scores by subtracting $S_{\text{rest/tasks}}$ from $S_{\text{rest/language}}$. The result of the recursive feature elimination experiments is given in Table 9. We computed the top 0.1% relevant language task specific features (shown in Figure 4).

Since the language network is not delineated clearly in the Power Atlas networks, we also computed the salience of connections between different anatomical regions by mapping regions from the Power atlas to regions in the Crossley atlas (Crossley et al., 2013), with known anatomical labels, using their Euclidean distance. We computed the salience between the functional connectivity of different regions by averaging the salience for connections between multiple Power atlas regions mapped to the same region from the Crossley atlas (shown in Figure 5). Using the task specific salience scores, we obtained the module salience maps for both the positive and negative task specific features for the language task (refer Figure S2c,d) and obtained the significant modular interactions (with p-value <.05, shown in Table 10).

TABLE 7 Performance of the FFN and the SVM for rest versus motor task classification using different fraction of features

	FFN		SVM	
% features	Accuracy	MAE	Accuracy	MAE
100%	99.65	0.0051	97.82	0.0217
10%	99.63	0.0042	98.14	0.0185
1%	99.81	0.0025	98.51	0.0149
0.1%	99.94	0.0010	97.58	0.0242

Abbreviation: MAE, mean absolute error.

3.6 | Intersubject variability in task related architectures

We measured the intersubject variability in the feature salience scores by considering the mean of the salience scores, c, and the intersubject variability v in the salience scores. The scatter plots for the normalized

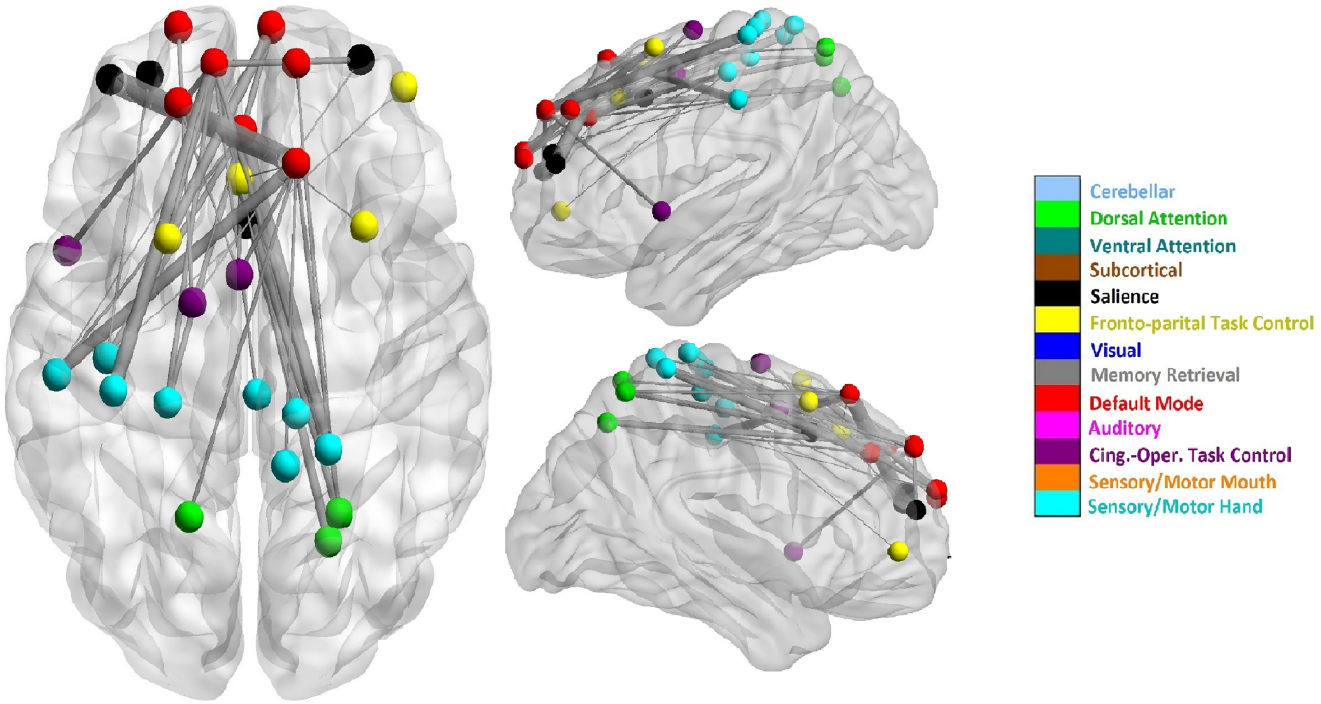


FIGURE 3 The motor task specific architecture in the brain. Figure shows the top 0.1% task specific connections for the motor task. The edge thickness corresponds to the salience score and the node color corresponds to the node's modular membership

TABLE 8 Significant modular interactions specific to the motor task

Module	Module	*p*-value
Positive motor task specific		
Motor hand	FP task control	$<10^{-3}$
Motor mouth	Motor mouth	$<10^{-3}$
Dorsal attention	Default mode	$<10^{-3}$
Motor hand	Default mode	$<10^{-2}$
Memory retrieval	Memory retrieval	$<10^{-2}$
CO task control	Default mode	$<10^{-2}$
Visual	Visual	$<10^{-2}$
Motor hand	Salience	$<10^{-2}$
CO task control	FP task control	<0.05
Negative motor task specific		
Visual	Visual	$<10^{-3}$
FP task control	FP task control	$<10^{-2}$
Dorsal attention	Dorsal attention	$<10^{-2}$
Motor mouth	Auditory	$<10^{-2}$
Memory retrieval	Memory retrieval	$<10^{-2}$
Motor mouth	Motor mouth	<0.02
Motor mouth	Motor hand	<0.03
Motor mouth	FP task control	<0.05

TABLE 9 Performance of the FFN and the SVM for rest versus language task classification using different fraction of features

	FFN		SVM	
% features	Accuracy	MAE	Accuracy	MAE
100%	99.86	0.0067	99.47	0.0082
10%	99.79	0.0098	99.51	0.0098
1%	99.82	0.0055	99.34	0.0064
0.1%	99.92	0.0030	98.10	0.0189

Abbreviation: MAE, mean absolute error.

mean and standard deviation scores are given in Figure 6. While the green markers represent all the features, the decoded task general (Figure 6a) and task specific architectures (Figure 6b for motor and (Figure 6c for language task) are represented by features in blue color. Furthermore, we regressed salience scores against decoded task specific and task general feature values and found that a quadratic polynomial model (sum of squared errors for task general: 24.14, language task specific: 27.13, motor task specific: 47.19) was a better fit than a linear model (sum of squared errors for task general: 25.00, language task specific: 29.87, motor task specific: 49.33). The best fitted quadratic polynomials are shown with red curves in the figures. On computing the correlation between the mean c and the intersubject variability c_{std} in the salience scores, we found that these two are weakly but positively correlated (Pearson's correlation for task general = 0.43, motor task specific = 0.56, and language task specific = 0.21). It can be observed that the decoded task related architectures consist of features that have a high salience score and low variability across subjects. However, some features that were on the lower end of the salience scores and had a low variability across subjects were also included. This can be discerned from the fact that most of the decoded features are below the quadratic polynomial curve fit for the respective task.

4 | DISCUSSION

4.1 | Classification of rest and task brain functional scans

We performed classification for rest versus all tasks and rest versus individual tasks (for tasks besides motor and language, refer to Appendix S1) and found that deeper models are able to achieve a higher accuracy even with a small subset of connectivity features than the shallow models such as SVM. This points to the presence of nonlinear and hierarchical relationships between functional connectivity and the tasks, which were better captured by DNN. The decoded task related functional brain architecture is composed of a small set of functional connections that were realized through elimination of insignificant features. The small fraction of salient connections between brain regions was found to be capable of accurately distinguishing brain resting and task states.

Recursive feature elimination experiments with SVM models deteriorated in classification performance when the connectivity features were reduced. This could be due to the fact that SVM models are not able to learn complex brain architecture due its shallow representations while DNN capture the complex and hierarchical relationships among functional connections. Our results differ from the results from He et al. (2020) which showed that simple models trained on functional connectivity could achieve comparable accuracy as complex deep models for their behavior prediction task. Also, our previous works (Gupta et al., 2019, 2021) using functional connectivity for disease prediction have shown that DNNs had a significantly better performance than SVM based shallow models for disease detection.

Although we removed the resting state time segments from the time series of the task data, a limitation of this study arises from the fact that we did not consider the different subtasks in each task state separately. The subtasks exist in each state, were assumed to be similar to each other, and involved in similar cognitive processing. But these subtasks may invoke different brain regions. If we consider component tasks separately, we are left with fewer time points that leads to poor signal-to-noise ratio and could ultimately degrade detection of functional activity (Birn et al., 2013). We, therefore, used the whole task scan as one specific task for our modeling.

Recent studies have used salience backpropagation to prune the neural network neurons by removing irrelevant features to prevent network overfitting (Gupta et al., 2019, 2021). However, application of such methods for brain decoding does not succeed if individual variations are not properly accounted for. This is due to the

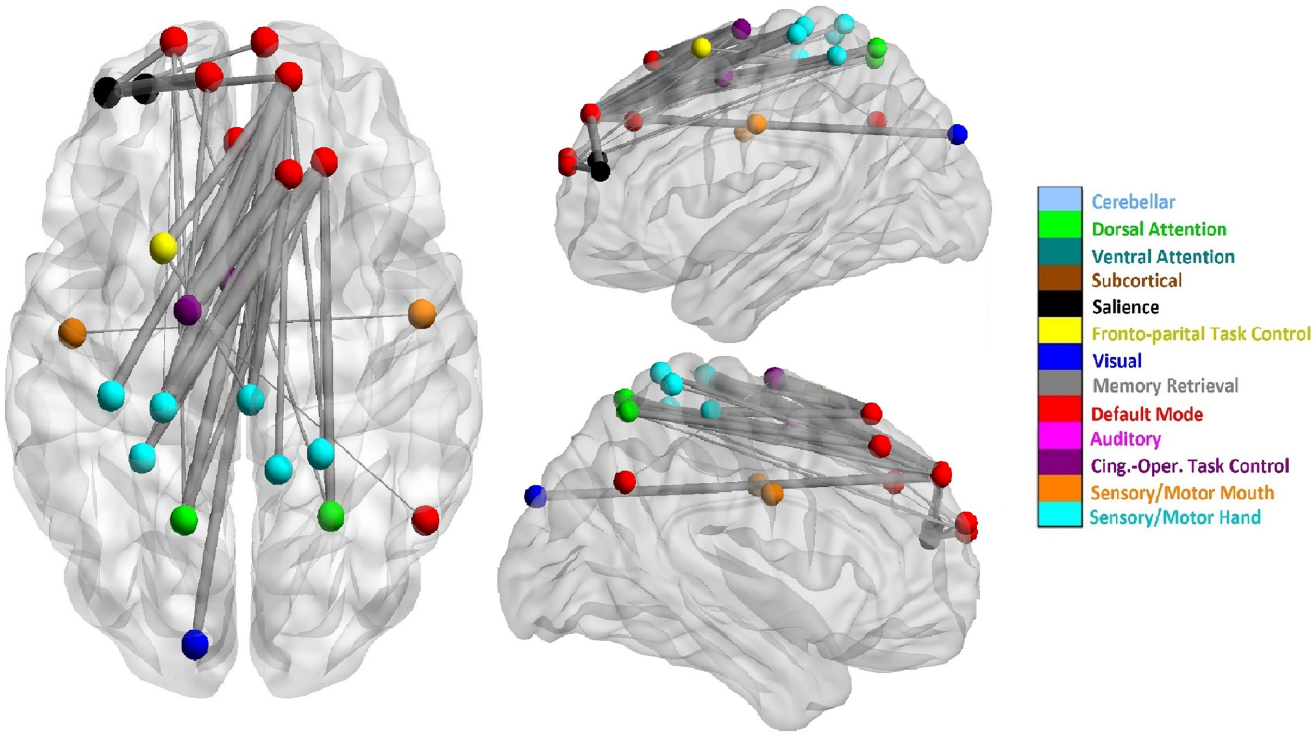
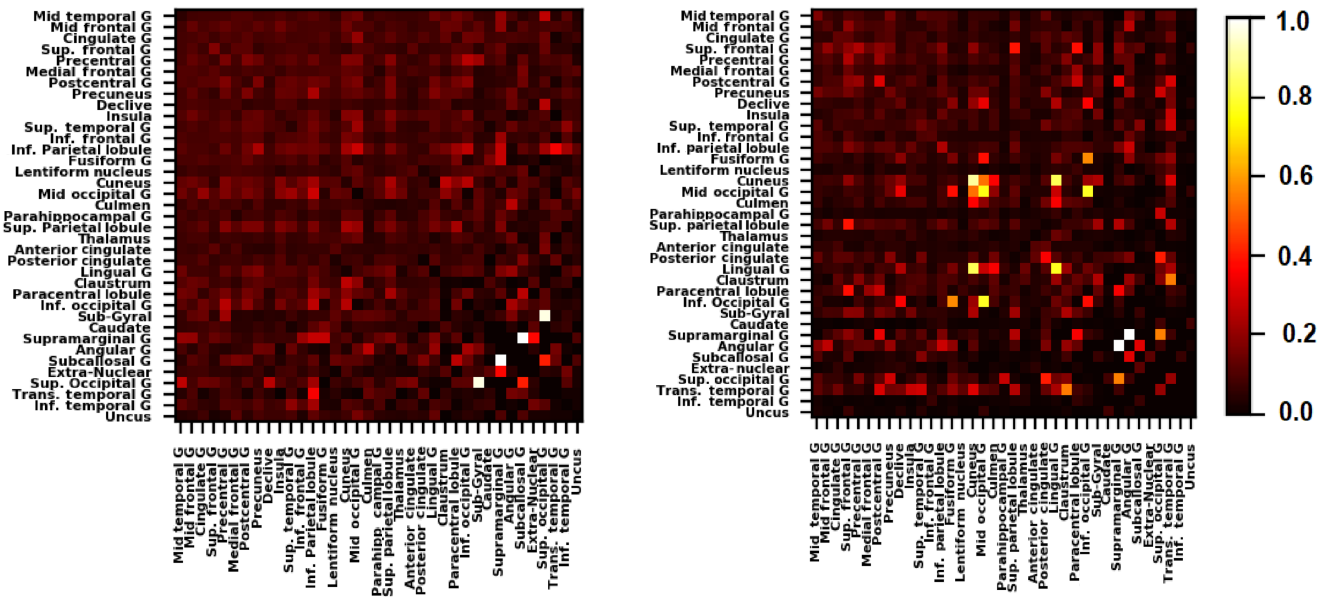


FIGURE 4 The language task specific architecture in the brain. Figure shows the top 0.1% task specific connections for the language task. The edge thickness corresponds to the salience score and the node color corresponds to the node's modular membership



(a) Language positive task specific **(b)** Language negative task specific

FIGURE 5 The salient anatomical connections for the task specific language architecture in the brain. The part figures (a: Language positive task specific) and (b: Language negative task specific) depict the salient connections between regions for the positive and negative language task specific connections, respectively

heterogeneity in disease pathology across participants, whereby a small set of features cannot account for disease across all the participants in the ensemble. However from our results, we conclude that although healthy subjects have pronounced intersubject variability in functional architecture (Mueller et al., 2013), a small set of common distinguishing features characterize task related alterations across

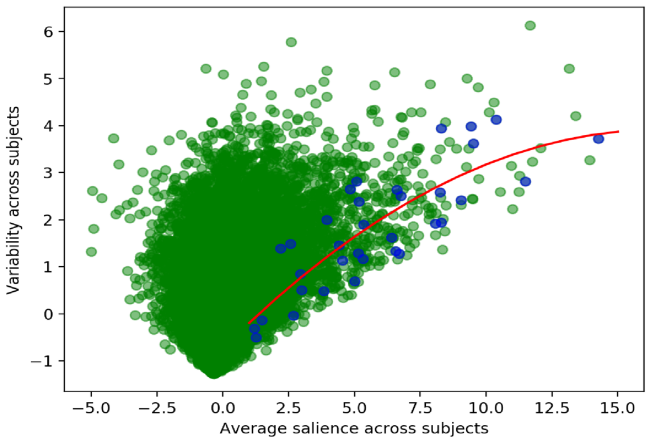
TABLE 10 Significant modular interactions specific to the language task

Module	Module	*p*-value
Positive language task specific		
Motor hand	Motor hand	$<10^{-3}$
FP task control	FP task control	$<10^{-3}$
Dorsal attention	Dorsal attention	$<10^{-3}$
Motor hand	Motor mouth	$<10^{-3}$
Salience	Salience	$<10^{-2}$
Ventral attention	Ventral attention	<0.03
Motor hand	CO task control	<0.03
Motor hand	Dorsal attention	<0.04
Negative language task specific		
Motor mouth	Motor mouth	$<10^{-3}$
Visual	Visual	$<10^{-3}$
Dorsal attention	Default mode	$<10^{-3}$
Memory retrieval	Memory retrieval	$<10^{-2}$
CO task control	Default mode	$<10^{-2}$
Motor hand	FP task control	<0.02
Motor hand	Default mode	<0.03
Motor hand	Salience	<0.04

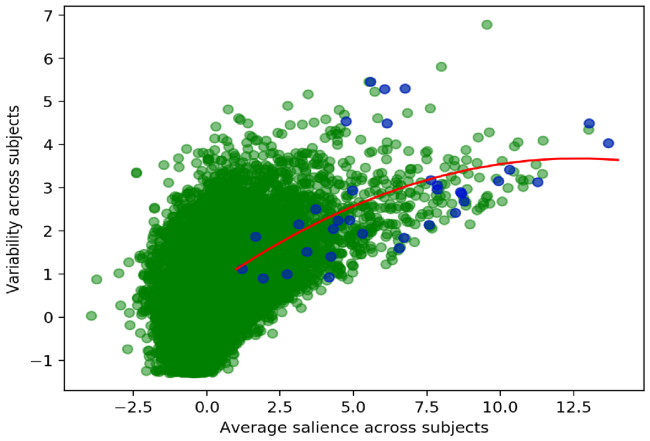
subjects. From our experiments, we infer that there are roughly two sets of features in the selected task related architectures, one that had a high salience and high variability, and another that had a moderate salience and low variability. We inferred that these two together help classify the rest and task states with an almost perfect accuracy, with the former handling variations while the latter form a stable core connectivity across all tasks. This has an important implication in aging (Meinzer et al., 2012) and detection of various neurological disorders that are characterized by impairments/alterations in the task general architecture.

4.2 | Task general architecture

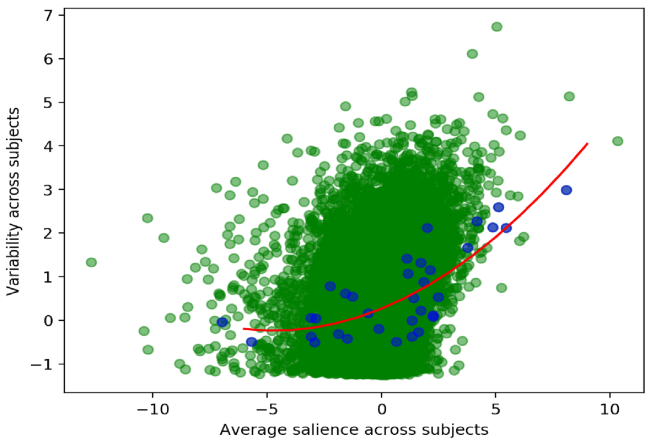
The positive and negative task general alterations found by our approach do not refer to activations/deactivations of regions but to the presence or absence of the connectivity during task performance with respect to the rest state. We found that the positive task general modulations are more wide-spread and apparent than negative task general modulations, that is, most of the salient task general connections during task performance are absent during rest (positive task general), while few salient task general connections are present during rest but absent during task (negative task general). It is to be noted that the positive and negative task general architectures only refer to the presence and absence of the connections during tasks and do not refer to the polarity (positive/negative) or magnitude (strong/weak) of the correlations.



(a) Task general features



(b) Motor task specific features



(c) Language task specific features

FIGURE 6 Intersubject variability in task related architectures. Figures (a: Task general features; b: Motor task specific features; c: Language task specific features) shows the scatter plot for the mean and variability in the task general and task specific salience scores across subjects. The blue markers represent the decoded task related features selected and the red curve represents the quadratic polynomial regression fit for the selected features

With our recursive feature elimination experiments, we observed that although the classifier performance slightly deteriorated, the classifier was able to achieve more than 99% accuracy even with just 0.1% of initial number of features. For SVM, we also performed recursive feature elimination based on the corresponding weights obtained for the best SVM model (linear kernel, $C = 0.001$). The SVM classification accuracy dropped when feature set was reduced to 0.1%. The respective MAE and accuracies for both FFN and the SVM classifiers can be discerned from Table 2. The recursive feature elimination experiments validate the correctness of the salience scores derived from our decoder. More importantly, it derived a small subset of features that was sufficient to distinguish between the rest and task states across subjects. For the FFN, we obtained the consensus of the 0.1% (34) remaining connections from each run and show the decoded features in Figure 1. As seen, the task general salient connections consist of the DMN, fronto-parietal (FP) task-control, SN, cingulo-opercular (CO) task-control, and somatomotor hand modules.

The functional brain has been shown to be composed of groups of brain regions performing specialized tasks. In order to analyze the task general changes, we performed the analysis at the modular level, whereby each module contains regions performing similar task (Gupta & Rajapakse, 2020). We analyzed the task general changes between (and within) the modules and found that different interactions from the same module could be part of negative and positive task general alterations. For negative task general alterations, we found that it was primarily the connections of the regions in the DMN (posterior cingulate cortex, precuneus, medial prefrontol cortex, and middle temporal gyrus), the fronto parietal task control (prefrontol and intraparietal sulcus) and the salience (anterior insula and dorsal anterior cingulate cortex) modules that were altered. However, for positive task general, it is primarily the connections with the somatomotor mouth, the visual and the memory retrieval modules that were altered. It is well-known that the cerebellum contains multiple somatomotor representations and relays information between the sensory neurons to the corresponding motor neurons in the cerebrum (Buckner et al., 2011). Since all tasks involved visually presented cues and the participant responses by motor movement (finger tap in all except social cognition task which involved a verbal response), our results on the task general architecture have invariably included the modulations between the cerebellar, motor, and visual systems.

The DMN has been characterized as a "task negative" network and it plays a role in extended support of internal mentation processes (Fox et al., 2005; Kelly, Uddin, Biswal, Castellanos, & Milham, 2008) and observing internal and/or external environment (Gao, Gilmore, Alcauter, & Lin, 2013). The CO task control, FP task control, and the salience networks help in configuring the brain to different brain states, and configuring information processing in response to different task demands. These networks thus support varied cognitive functions including sensory perception and motor control. The FP task control network acts as a flexible hub for cognitive control facilitating special attention to trial specific information (Marek & Dosenbach, 2018) whereas the CO task control network maintains alertness during task performance and correction based on feedback from each trial (Cocchi, Zalesky, Fornito, & Mattingley, 2013). The salience network determines the importance given to different stimuli for the brain to act and is known to be widely activated across different tasks (Seeley et al., 2007). Our results are thus in line with the widely known task related deactivations that are observed in the DMN (Buckner, Andrews-Hanna, & Schacter, 2008) and the memory retrieval systems (Fox et al., 2005; Power et al., 2011) while the FP task-control, CO task-control and ventral attention systems are known for their task related activations (Cole, Repovš, & Anticevic, 2014; Corbetta, Kincade, Ollinger, McAvoy, & Shulman, 2000; Power et al., 2011).

4.3 | Task specific architectures

We investigated task specific modulations for tasks that involved more perceptual and fewer cognitive processing systems so that we did not only identify task specific modulations but also can also verify them. The task specific alterations for both the motor and the language tasks were found to be limited to primarily the somatomotor, attentional, task control, and the salience systems.

For the recursive feature elimination experiments with the motor task, the MAE improved with fewer features and the accuracy for classification with just 0.1% features was the highest (and close to 100%) for the FFN. For the SVM, the accuracy and MAE improved initially, but there was a deterioration when the feature set was reduced to 0.1%. As a result, we were able to find a set of 34 task specific connections from the FFN that were important for the classification task. These consist of correlations between the somatomotor hand, SN, dorsal attention, FP Task control, and DMN (prefrontal cortex) subsystems. Besides moving the hand and feet, the motor task also involved tongue movements. These are reflected in the connectivity of multiple systems with somatomotor hand and mouth systems being classified as crucial. While the FP task control (Zanto & Gazzaley, 2013) network is known to alter its functional connectivity with nodes of other networks based on task goals, the dorsal attention (Corbetta et al., 2000) and the CO task control (Sadaghiani & D'Esposito, 2015) networks are involved in voluntary attention and are altered when directed attention and alertness is required.

As seen, the motor task specific alterations within visual, motor mouth and memory retrieval modules are reported as both positive and negative. It is known that the motor task leads to an increased local, within module functional connectivity and a decrease in global integration (Cohen & D'Esposito, 2016). Our results for the motor modules show that there is heterogeneity in how the within motor modules connections are altered, since some connections are positive task specific while others are negative task specific. This is noteworthy since the motor areas are known for their high within module functional connectivity during rest (Biswal, Kylen, & Hyde, 1997; Biswal, Zerrin Yetkin, Haughton, & Hyde, 1995) and that there is intrinsic information processing in these regions even in the absence of tasks or during rest (Biswal et al., 1997; Cordes et al., 2000; Greicius, Krasnow, Reiss, & Menon, 2003). The present results

demonstrate that the performance of the motor task alters functional connectivity differently for different regions. It can also be seen that most of the negative motor task specific alterations were limited to connections within the respective module (motor, visual, FP task control, dorsal attention, and memory retrieval), which points to the absence of the connections during task performance.

For the language task, we performed recursive feature elimination and found that the accuracy for classification with just 0.1% features was higher than the accuracy for 100% features (refer Table 9). However, the performance with SVM deteriorated with fewer features both in terms of accuracy and MAE. The top 0.1% language task specific features (shown in Figure 4) are from the auditory (mostly left: L temporal), FP task-control (angular L and right: R), DMN (inferior parietal L and R, and frontal), SN (anterior cingulum), somatomotor hand (post central gyrus), and visual (occipital lobe) subsystems. Anatomically, the most distinguishing connections were from regions in the supramarginal gyrus, occipital lobe, angular gyrus, lingual gyrus, cuneus, temporal gyrus, and the frontal orbital lobe. While the supramarginal gyrus, angular gyrus, lingual gyrus and cuneus are well known for their role in language and visual processing tasks, multiple studies (Binder et al., 2011; Crinion, Lambon-Ralph, Warburton, Howard, & Wise, 2003; Flinker, Chang, Barbaro, Berger, & Knight, 2011; Meschyan & Hernandez, 2006) have reported extensive activations in different areas of the temporal lobe along with areas in the lateral occipital lobe and the frontal lobe during language processing. The significant task specific language modular interactions involved the visual module (regions from the occipital lobe), the SN (anterior insula and dorsal anterior cingulate cortex), somatomotor hand, somatomotor mouth, dorsal and ventral attention modules, memory retrieval (precuneus and posterior cingulate cortex) modules (Table 10).

Thus, we have not only shown that a small set of features is able to classify the rest and task states but also verified that these feature sets have been deemed important by previous studies. More importantly, our work inferred the task general and task specific architecture from fMRI scans.

5 | CONCLUSION

We detected task general and task specific brain architectures by decoding the DNNs trained to classify rest-state and cognitive task states fMRI data. Using a small subset of functional connectivity between functionally diverse brain regions as features for a five-layer feedforward DNN, we were able to reliably classify rest-state and seven tasks and rest-state from motor task scans of 690 subjects from the HCP with nearly 100% accuracy. Using a salience backpropagation based decoder, we identified connections that formed the task general and task specific brain architecture in the brain. We also studied what modular interactions are altered during task performance. We validated the results of our decoder by showing that a small subset of features deemed important by the decoder gives us a high accuracy. We also studied the intersubject variability in the task related functional architectures. Our method has a strong potential of replication in brain state classification and identification of distinctive biomarkers between the healthy and diseased brain states.

ACKNOWLEDGMENT

The authors have no conflict of interest. The authors would like to thank Yi Hao Chan for his support during the revision of the manuscript. This work was partially supported by AcRF Tier 1 grant 2019-T1-002-057 and Tier-2 grant T2EP20121-0003 from the Ministry of Education, Singapore.

CONFLICT OF INTEREST

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID

Sukrit Gupta https://orcid.org/0000-0002-8974-8482
Jagath C. Rajapakse https://orcid.org/0000-0001-7944-1658

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., Wicke, M., Yu, Y., & Zheng, X. (2016). Tensorflow: A system for large-scale machine learning. In 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16), pp. 265–283.

Betti, V., Della Penna, S., de Pasquale, F., Mantini, D., Marzetti, L., Romani, G. L., & Corbetta, M. (2013). Natural scenes viewing alters the dynamics of functional connectivity in the human brain. *Neuron*, *79*(4), 782–797.

Binder, J. R., Gross, W. L., Allendorfer, J. B., Bonilha, L., Chapin, J., Edwards, J. C., … Weaver, K. E. (2011). Mapping anterior temporal lobe language areas with fmri: A multicenter normative study. *NeuroImage*, *54*(2), 1465–1475.

Birn, R. M., Molloy, E. K., Patriat, R., Parker, T., Meier, T. B., Kirk, G. R., … Prabhakaran, V. (2013). The effect of scan length on the reliability of resting-state fmri connectivity estimates. *NeuroImage*, *83*, 550–558.

Biswal, B., Zerrin Yetkin, F., Haughton, V. M., & Hyde, J. S. (1995). Functional connectivity in the motor cortex of resting human brain using echo-planar mri. *Magnetic Resonance in Medicine*, *34*(4), 537–541.

Biswal, B. B., Kylen, J. V., & Hyde, J. S. (1997). Simultaneous assessment of flow and BOLD signals in resting-state functional connectivity maps. *NMR in Biomedicine*, *10*(4-5), 165–170.

Brown, C. J., Kawahara, J., & Hamarneh, G. (2018). Connectome priors in deep neural networks to predict autism. In 2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018), pp. 110–113.

Buckner, R. L., Andrews-Hanna, J. R., & Schacter, D. L. (2008). The brain's default network. *Annals of the New York Academy of Sciences*, *1124*(1), 1–38.

Buckner, R. L., Krienen, F. M., Castellanos, A., Diaz, J. C., & Yeo, B. T. (2011). The organization of the human cerebellum estimated by intrinsic functional connectivity. *Journal of Neurophysiology*, *106*(5), 2322–2345.

Chollet, F. (2018). *Keras: Deep Learning for humans*. Retrieved from https://github.com/keras-team/keras#keras-deep-learning-for-humans

Cocchi, L., Zalesky, A., Fornito, A., & Mattingley, J. B. (2013). Dynamic cooperation and competition between brain systems during cognitive control. *Trends in Cognitive Sciences*, *17*(10), 493–501.

Cohen, J. R., & D'Esposito, M. (2016). The segregation and integration of distinct brain networks and their relationship to cognition. *Journal of Neuroscience*, *36*(48), 12083–12094.

Cole, M. W., Bassett, D. S., Power, J. D., Braver, T. S., & Petersen, S. E. (2014). Intrinsic and task-evoked network architectures of the human brain. *Neuron*, *83*(1), 238–251.

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

Cole, M. W., Ito, T., Schultz, D., Mill, R., Chen, R., & Cocuzza, C. (2019). Task activations produce spurious but systematic inflation of task functional connectivity estimates. *NeuroImage*, *189*, 1–18.

Cole, M. W., Repovš, G., & Anticevic, A. (2014). The frontoparietal control system: A central role in mental health. *The Neuroscientist*, *20*(6), 652–664.

Corbetta, M., Kincade, J. M., Ollinger, J. M., McAvoy, M. P., & Shulman, G. L. (2000). Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience*, *3*(3), 292–297.

Cordes, D., Haughton, V. M., Arfanakis, K., Wendt, G. J., Turski, P. A., Moritz, C. H., … Meyerand, M. E. (2000). Mapping functionally related regions of brain with functional connectivity MR imaging. *American Journal of Neuroradiology*, *21*(9), 1636–1644.

Crinion, J. T., Lambon-Ralph, M. A., Warburton, E. A., Howard, D., & Wise, R. J. (2003). Temporal lobe regions engaged during normal speech comprehension. *Brain*, *126*(5), 1193–1201.

Crossley, N. A., Mechelli, A., Vértes, P. E., Winton-Brown, T. T., Patel, A. X., Ginestet, C. E., … Bullmore, E. T. (2013). Cognitive relevance of the community structure of the human brain functional coactivation network. *Proceedings of the National Academy of Sciences*, *110*(28), 11583–11588.

Düzel, E., Cabeza, R., Picton, T. W., Yonelinas, A. P., Scheich, H., Heinze, H.-J., & Tulving, E. (1999). Task-related and item-related brain processes of memory retrieval. *Proceedings of the National Academy of Sciences*, *96*(4), 1794–1799.

Elton, A., & Gao, W. (2015). Task-positive functional connectivity of the default mode network transcends task domain. *Journal of Cognitive Neuroscience*, *27*(12), 2369–2381.

Finn, E. S., Shen, X., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., … Constable, R. T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, *18*(11), 1664–1671.

Flinker, A., Chang, E., Barbaro, N., Berger, M., & Knight, R. (2011). Sub-centimeter language organization in the human temporal lobe. *Brain and Language*, *117*(3), 103–109.

Floren, A., Naylor, B., Miikkulainen, R., & Ress, D. (2015). Accurately decoding visual information from fMRI data obtained in a realistic virtual environment. *Frontiers in Human Neuroscience*, *9*, 327.

Fox, M. D., Snyder, A. Z., Vincent, J. L., Corbetta, M., van Essen, D. C., & Raichle, M. E. (2005). The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proceedings of the National Academy of Sciences*, *102*(27), 9673–9678.

Gao, W., Gilmore, J. H., Alcauter, S., & Lin, W. (2013). The dynamic reorganization of the default-mode network during a visual classification task. *Frontiers in Systems Neuroscience*, *7*, 34.

Glasser, M. F., Sotiropoulos, S. N., Wilson, J. A., Coalson, T. S., Fischl, B., Andersson, J. L., … WU-Minn HCP Consortium. (2013). The minimal preprocessing pipelines for the human connectome project. *NeuroImage*, *80*, 105–124.

Gordon, E. M., Laumann, T. O., Adeyemo, B., Gilmore, A. W., Nelson, S. M., Dosenbach, N. U., & Petersen, S. E. (2017). Individual-specific features of brain systems identified with resting state functional correlations. *NeuroImage*, *146*, 918–939.

Gordon, E. M., Laumann, T. O., Adeyemo, B., & Petersen, S. E. (2017). Individual variability of the system-level organization of the human brain. *Cerebral Cortex*, *27*(1), 386–399.

Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., … Dosenbach, N. U. F. (2017). Precision functional mapping of individual human brains. *Neuron*, *95*(4), 791–807.

Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., … Petersen, S. E. (2018). Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron*, *98*(2), 439–452.

Greicius, M. D., Krasnow, B., Reiss, A. L., & Menon, V. (2003). Functional connectivity in the resting brain: A network analysis of the default mode hypothesis. *Proceedings of the National Academy of Sciences*, *100*(1), 253–258.

Gupta, S., Alzheimers Disease Neuroimaging Initiative, Chan, Y. H., & Rajapakse, J. C. (2021). Obtaining leaner deep neural networks for decoding brain functional connectome in a single shot. *Neurocomputing*, *453*, 326–336.

Gupta, S., Chan, Y. H., Rajapakse, J. C., Initiative, A. D. N., & Alzheimers Disease Neuroimaging Initiative. (2019). Decoding brain functional connectivity implicated in ad and mci. In International conference on medical image computing and computer-assisted intervention, pp. 781–789.

Gupta, S., & Rajapakse, J. C. (2020). Iterative consensus spectral clustering improves detection of subject and group level brain functional modules. *Scientific Reports*, *10*(1), 1–15.

Haatveit, B., Jensen, J., Alnæs, D., Kaufmann, T., Brandt, C. L., Thoresen, C., … Westlye, L. T. (2016). Reduced load-dependent default mode network deactivation across executive tasks in schizophrenia spectrum disorders. *NeuroImage: Clinical*, *12*, 389–396.

Hamilton, J. P., Furman, D. J., Chang, C., Thomason, M. E., Dennis, E., & Gotlib, I. H. (2011). Default-mode and task-positive network activity in major depressive disorder: Implications for adaptive and maladaptive rumination. *Biological Psychiatry*, *70*(4), 327–333.

He, T., Kong, R., Holmes, A. J., Nguyen, M., Sabuncu, M. R., Eickhoff, S. B., … Yeo, B. T. (2020). Deep neural networks and kernel regression achieve comparable accuracies for functional connectivity prediction of behavior and demographics. *NeuroImage*, *206*, 116276.

Jang, H., Plis, S. M., Calhoun, V. D., & Lee, J.-H. (2017). Task-specific feature extraction and classification of fMRI volumes using a deep neural network initialized with a deep belief network: Evaluation using sensorimotor tasks. *NeuroImage*, *145*, 314–328.

Kelly, A. C., Uddin, L. Q., Biswal, B. B., Castellanos, F. X., & Milham, M. P. (2008). Competition between functional brain networks mediates behavioral variability. *NeuroImage*, *39*(1), 527–537.

Knyazev, G. G., Savostyanov, A. N., Bocharov, A. V., Brak, I. V., Osipov, E. A., Filimonova, E. A., … Aftanas, L. I. (2018). Task-positive and task-negative networks in major depressive disorder: A combined fMRI and EEG study. *Journal of Affective Disorders*, *235*, 211–219.

Krienen, F. M., Yeo, B. T., & Buckner, R. L. (2014). Reconfigurable task-dependent functional coupling modes cluster around a core functional architecture. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *369*(1653), 20130526.

Li, H., & Fan, Y. (2019). Interpretable, highly accurate brain decoding of subtly distinct brain states from functional mri using intrinsic functional networks and long short-term memory recurrent neural networks. *NeuroImage*, *202*, 116059.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, 4765–4774.

Marek, S., & Dosenbach, N. U. (2018). The frontoparietal network: Function, electrophysiology, and importance of individual precision mapping. *Dialogues in Clinical Neuroscience*, *20*(2), 133–140.

Meinzer, M., Seeds, L., Flaisch, T., Harnish, S., Cohen, M. L., McGregor, K., … Crosson, B. (2012). Impact of changed positive and negative task-related brain activity on word-retrieval in aging. *Neurobiology of Aging*, *33*(4), 656–669.

Melrose, R. J., Jimenez, A. M., Riskin-Jones, H., Weissberger, G., Veliz, J., Hasratian, A. S., … Sultzer, D. L. (2018). Alterations to task positive and

task negative networks during executive functioning in mild cognitive impairment. *NeuroImage: Clinical*, *19*, 970–981.

Meschyan, G., & Hernandez, A. E. (2006). Impact of language proficiency and orthographic transparency on bilingual word reading: An fmri investigation. *NeuroImage*, *29*(4), 1135–1140.

Meszlényi, R. J., Buza, K., & Vidnyánszky, Z. (2017). Resting state fMRI functional connectivity-based classification using a convolutional neural network architecture. *Frontiers in Neuroinformatics*, *11*, 61.

Mills, B. D., Miranda-Dominguez, O., Mills, K. L., Earl, E., Cordova, M., Painter, J., … Fair, D. A. (2018). ADHD and attentional control: Impaired segregation of task positive and task negative brain networks. *Network Neuroscience*, *2*(2), 200–217.

Mueller, S., Wang, D., Fox, M. D., Yeo, B. T., Sepulcre, J., Sabuncu, M. R., … Liu, H. (2013). Individual variability in functional connectivity architecture of the human brain. *Neuron*, *77*(3), 586–595.

Norman, D. A., & Shallice, T. (1986). Attention to action. In R.. J. Davidson, G.. E. Schwartz, & D. Shapiro (Eds.), *Consciousness and self-regulation* (pp. 1–18). Boston, MA: Springer.

Nyberg, L., McIntosh, A. R., Cabeza, R., Habib, R., Houle, S., & Tulving, E. (1996). General and specific brain regions involved in encoding and retrieval of events: What, where, and when. *Proceedings of the National Academy of Sciences*, *93*(20), 11280–11285.

Power, J. D., Cohen, A. L., Nelson, S. M., Wig, G. S., Barnes, K. A., Church, J. A., … Petersen, S. E. (2011). Functional network organization of the human brain. *Neuron*, *72*(4), 665–678.

Sadaghiani, S., & D'Esposito, M. (2015). Functional characterization of the cingulo-opercular network in the maintenance of tonic alertness. *Cerebral Cortex*, *25*(9), 2763–2773.

Seeley, W. W., Menon, V., Schatzberg, A. F., Keller, J., Glover, G. H., Kenna, H., … Greicius, M. D. (2007). Dissociable intrinsic connectivity networks for salience processing and executive control. *Journal of Neuroscience*, *27*(9), 2349–2356.

Shrikumar, A., Greenside, P., & Kundaje, A. (2017). Learning important features through propagating activation differences. In Proceedings of the 34th international conference on machine learning-volume 70, pp. 3145–3153.

Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. In Proceedings of the 34th international conference on machine learning-volume 70, pp. 3319–3328.

van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E., Yacoub, E., Ugurbil, K., & WU-Minn HCP Consortium. (2013). The WU-Minn human connectome project: An overview. *NeuroImage*, *80*, 62–79.

Wang, D., Buckner, R. L., Fox, M. D., Holt, D. J., Holmes, A. J., Stoecklein, S., … Liu, H. (2015). Parcellating cortical functional networks in individuals. *Nature Neuroscience*, *18*(12), 1853–1860.

Xie, H., Calhoun, V. D., Gonzalez-Castillo, J., Damaraju, E., Miller, R., Bandettini, P. A., & Mitra, S. (2018). Whole-brain connectivity dynamics reflect both task-specific and individual-specific modulation: A multitask study. *NeuroImage*, *180*, 495–504.

Zanto, T. P., & Gazzaley, A. (2013). Fronto-parietal network: Flexible hub of cognitive control. *Trends in Cognitive Sciences*, *17*(12), 602–603.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Gupta, S., Lim, M., & Rajapakse, J. C. (2022). Decoding task specific and task general functional architectures of the brain. *Human Brain Mapping*, *43*(9), 2801–2816. https://doi.org/10.1002/hbm.25817