Recognizing Brain States Using Deep Sparse Recurrent Neural Network

Han Wang, Shijie Zhao, Qinglin Dong, Yan Cui, Yaowu Chen, Junwei Han,
Li Xie, and Tianming Liu, *Senior Member, IEEE*

Abstract—**Brain activity is a dynamic combination of different sensory responses and thus brain activity/state is continuously changing over time. However, the brain's dynamical functional states recognition at fast time-scales in task fMRI data have been rarely explored. In this paper, we propose a novel 5-layer deep sparse recurrent neural network (DSRNN) model to accurately recognize the brain states across the whole scan session. Specifically, the DSRNN model includes an input layer, one fully-connected layer, two recurrent layers, and a softmax output layer. The proposed framework has been tested on seven task fMRI data sets of Human Connectome Project. Extensive experiment results demonstrate that the proposed DSRNN model can accurately identify the brain's state in different task fMRI data sets and significantly outperforms other auto-correlation methods or non-temporal approaches in the dynamic brain state recognition accuracy. In general, the proposed DSRNN offers a new methodology for basic neuroscience and clinical research.**

Index Terms—**Dynamic brain state, recurrent neural network, fMRI, brain networks.**

I. Introduction

UNDERSTANDING the nature of brain's activities and functions has been one of the major goals since the inception of neuroscience. During the past few decades, researchers have developed various methods to characterize and analyze the brain activity patterns, such as GLM [1], ICA [2], and sparse representation based methods [3]–[7]. The underlying assumption used in previous studies is that the brain networks and activation patterns/states are temporally stationary across the entire fMRI scan session. However, there are accumulating evidences [8]–[12] indicating that brain activities and states are under dramatic temporal changes at various time scales. For instance, it has been found that each cortical brain area runs different "programs" according to the cognitive context and the current perceptual requirements [9], and the intrinsic cortical circuits mediate the moment-by-moment functional state changes in the brain [9]. That is, in the brain's dynamic functional process, parts of the brain engage and disengage in time, allowing a person to perceive objects or scenes, and to separate remembered parts of an experience, and to bind them all together into a coherent whole [8], [13]. What's more, it is still under dynamic changes even in resting state within time scales of seconds to minutes [14]. Inspired by these observations, more and more researchers are motivated to examine the temporal dynamics of functional brain activities [15]–[19].

Currently, a dominant analysis technique used for describing the temporal dynamics of functional brain activities is the use of sliding windows [16], [20], [21]. Sliding window based approaches pre-specify the temporal resolution of the changing pattern (window size), and map the spatial distribution of the networks and provide measures of dependence, e.g., linear correlation between the timecourse for interested pair of voxels, regions, or networks [22]. These methods range from windowed versions of standard seed-based correlation or independent component analysis (ICA) techniques to new methods that consider information from individual time points [14], [16], [19]–[21], [23]–[27]. In addition to sliding window approaches, researchers also tried some alternative methods over the past few years. For instance, change point detection methods have been proposed to determine "brain state" changes based on the properties of data-driven partitioned resting state fMRI (rs-fMRI) data, such as amplitude and covariance of time series [10], [15], [18], [28]–[31]. In contrast to methods based on relationships between brain

TABLE I
PROPERTIES OF HCP TASK-FMRI DATASETS

Parameters / Task	Working Memory	Gambling	Motor	Language	Social	Relational	Emotion
# of Frames	405	253	284	316	274	232	176
Duration (Min)	5:01	3:12	3:34	3:57	3:27	2:56	2:16
# of Task Blocks	8	4	10	8	5	6	6
# of Block Labels	3	3	6	3	5	3	3
Duration of Blocks(s)	25	28	12	See Text	23	16	18

regions, event-based approaches assume that the brain activity is primarily composed of distinct events and these events can be deciphered from the BOLD fluctuations [32], [33] or through deconvolution of a given hemodynamic model from the time series [34], [35]. More recently, Vidaurre and his colleagues proposed to describe brain activity as a dynamic sequence of discrete brain states using Hidden Markov Model (HMM) to provide a rich description of the brain activities [36], [37]. In general, these approaches have enriched the description of functional brain activities and contributed to the better understanding of the temporal dynamics underlying brain activities.

However, these approaches might still have possible limitations. First, as the most widely used approach, the parameters of sliding window techniques are debatable, such as the window length and type. For example, a long temporal window will miss fast dynamics, while short window will have insufficient data to provide reliable estimation, and it is difficult to achieve convincible results without "gold standard". Second, most of these approaches make conclusions based on the relative changes between paired voxels or brain regions which ignore the temporal dependence of the whole sequence of brain activities/states. In addition, with the limited time scales, they may not be able to catch the fast temporal scale dynamics. Last but not least, many of these methods are designed for rs-fMRI data [22] and how to model the temporal dynamics of task fMRI (tfMRI) data has been rarely explored. Therefore, developing a comprehensive and systematic spatial-temporal brain dynamic activity modelling framework that can naturally recognize dynamic brain states at fast time-scales from task fMRI data is still needed.

Recently, Recurrent Neural Networks (RNNs) have shown particularly outstanding performances in many research areas, such as handwriting recognition [38], [39], language modeling [38], [40], machine translation [41], and speech recognition [42], which involve modeling sequential signals. Furthermore, a few attempts using RNNs in neural encoding [42] and responding [48] have shown that RNNs are feasible in modeling dynamic biological signals. A prominent feature of RNNs is that they can easily and effectively capture both short and long term temporal dependences in data sequences and model the inner dependent relationships using their internal memories. Specifically, RNNs maintain memory cells to preserve the information of the past sequences and learn when and how to restore these memories to make predictions, rather than predicting based only on local neighborhoods in time. Therefore, RNNs can inherently acquire the temporal dependence of the sequential data, which is quite suitable for the problem of modeling temporal dynamic activities and recognizing the fast time-scale brain states in task fMRI data.

Inspired by the great ability in modeling the temporal dependence of sequential data using RNN models, in this paper, we propose a five-layer deep sparse recurrent neural network (DSRNN) to recognize the dynamical brain states at fast time-scales in tfMRI data. Briefly, the proposed DSRNN model includes an input layer, one fully-connected layer, two recurrent layers and a softmax output layer. We tested the proposed DSRNN model with more than 800 subjects in the seven task fMRI datasets provided by the Human Connectome Project (HCP), including working memory, gambling, motor, language, social, relational, and emotion tasks [49]. Notably, the proposed DSRNN model achieved outstanding brain state recognition accuracy (over 90 percent averagely), compared with auto-correlation methods (Adaptive Autoregressive Classifier, AAR [50], [51]) and traditional non-temporal modeling approaches (Softmax and SVM). Also, the associated important brain regions in recognized brain states show meaningful consistence with traditional task-activated areas, which provides us novel insight on understanding the brain activities and functions. In general, our work demonstrate that the proposed DSRNN model has great advantage in recognizing the dynamical brain states at fast time-scales, and it offers a novel methodology for the study of dynamical functional brain activities.

II. MATERIALS AND METHODS

The proposed deep sparse recurrent neural network (DSRNN) is a five-layer deep neural network model, including an input layer, one fully-connected layer, two recurrent layers and a softmax output layer. Specially, the fully-connected layer plays a role of extracting activated brain regions and reducing data dimension, and the cascaded recurrent layers are used to capture temporal dependence in time series. In this section, we will introduce the fMRI datasets, basic theory of recurrent neural network, and the structure and mechanism of DSRNN in details.

A. Data Acquisition and Pre-Processing

In this paper, we adopted the Human Connectome Project (HCP) 900 Subjects Data Release as test beds, which include behavioral and 3T MR imaging data for over 900 healthy adult participants [52]. Seven categories of behavioral tasks are involved, including Working memory, Gambling, Motor, Language, Social, Relational, Emotion tasks. The design information of seven tasks is shown in Table 1. In total, there are 788 subjects executed all seven behavioral tasks, and

786 subjects' task fMRI (tfMRI) time series with good data quality are used in this work. All the datasets are available on https://db.humanconnectome.org. The detailed acquisition parameters were set as follows: 220 mm FOV, in-plane FOV: 208×180 mm^2, flipangle = 52, BW = 2290 Hz/Px, $2 \times 2 \times 2$ mm^3 spatial resolution, 90×104 matrix, 72 slices, TR = 0.72s, TE = 33.1 ms. The preprocessing of these tfMRI data sets includes skull removal, motion correction, slice time correction, spatial smoothing, and global drift removal (high-pass filtering). In our implementation, a $4 \times 4 \times 4$ mm^3 down-sampling is applied to reduce computational cost. Next, we will introduce these tasks briefly and more detailed procedures are available in section A of supplementary materials. Supplementary materials are available in the supplementary files /multimedia tab.

In working memory (WM) task, a version of the N-back task was used to assess working memory [53]. It was reported that the associated brain activations were reliable across subjects [53] and time [54]. Gambling task was adapted from Delgado *et al.* [55], and prior works demonstrated that the task elicits activations in the striatum and other reward related regions which are robust and reliable across the subjects [55]–[58]. Motor task was developed by Buckner and his colleagues which has been proven that it could identify effector specific activations in individual subjects [59]. Language task was developed by Binder *et al.* [60]. There are two subtasks: story and math where the math blocks was designed to provide a comparison task that was attentional demanding. In Social task, an engaging and validated video task was chosen as a measure of social cognition, giving evidence that it generates robust task related activation in brain regions associated with social cognition and it is reliable across subjects [61]–[64]. Relational task was adapted from the one developed by Smith *et al.* [65] which demonstrated to localize activation in anterior prefrontal cortex in individual subjects. Emotion task was adapted from Hariri and his colleagues who have shown this task could be used as a functional localizer [66] with moderate reliability across time [67].

B. Recurrent Neural Networks

Recurrent neural networks (RNNs) are feedforward neural networks with edges connecting adjacent time steps. This distinct architecture is similar to biological neuronal networks, where lateral and feedback interconnections are widespread. The internal states of RNNs preserve memories of previous state information and capture the temporal dependences of input signals. RNNs have been applied successfully in many diverse sequence modeling tasks [38]–[48], [68]. Fig. 1(a) illustrates a specific and basic recurrent cell unit and its recursive equation is as follows:

$$h^t = \tanh\left(Uh^{t-1} + W x^t + b\right) \tag{1}$$

where U, W are the weight matrices for hidden states and input features and b is bias parameters. h^t denotes the hidden state. Typically, vanilla RNNs (established by basic recurrent cell units) are very difficult to train due to the vanishing gradient problem [69]. In contrast, LSTM
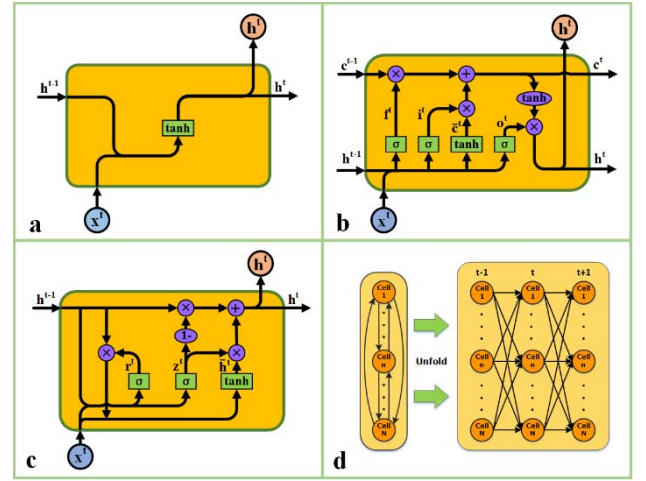

Fig. 1. A schematic illustration of RNN cell models. (a) The nonlinear basic recurrent cell unit; (b) Long short-term memory unit; (c) Gated recurrent unit; (d) The interconnections in a common recurrent hidden layer. Neurons in recurrent layer are fully interconnected and new hidden states can be influenced by all former states. Squares indicate linear combination and nonlinearity. Circles indicate element wise operations. Gates in the units control the information flow between adjacent time points.

(Long Short-Term Memory) units [70] and GRU (Gated Recurrent Unit) units [41] were specifically designed to overcome this problem and have since become the most widely-used cell architectures.

Each LSTM unit maintains a cell state that acts as its internal memory by storing information from previous time series. The contents of the cell state are controlled by the gates of the unit, and determine unit's hidden state. The first-layer hidden state of an LSTM unit is defined as follows:

$$h^t = o^t \odot \tanh\left(c^t\right) \tag{2}$$
$$o^t = \sigma(U_o h^{t-1} + W_o x^t + b_o) \tag{3}$$

where $\odot$ denotes elementwise multiplication, c^t is the cell state, and o^t is the output gate activation. The output gate controls what information would be retrieved from the cell state. The cell state of an LSTM unit is defined as:

$$c^t = f^t \odot c^{t-1} + i^t \odot \bar{c}^t \tag{4}$$
$$f^t = \sigma\left(U_f h^{t-1} + W_f x^t + b_f\right) \tag{5}$$
$$i^t = \sigma(U_i h^{t-1} + W_i x^t + b_i) \tag{6}$$
$$\bar{c}^t = \tanh\left(U_c h^{t-1} + W_c x^t + b_c\right) \tag{7}$$

where f^t is the forget gate activation, i^t is the input gate activation, and $\bar{c}^t$ is an auxiliary variable. Forget gate controls what old information would be abandoned from the cell state, while input gate controls what new information would be stored in cell state. Furthermore, U_f, U_i, U_c and W_f, W_i, W_c are the corresponding weights and b_f, b_i, b_c are the biases (i.e., the learnable parameters of the model).

The GRU unit is a simpler alternative to LSTM unit. They combine hidden state with cell state, and input gate with forget gate. The first-layer hidden state of GRU unit is defined as
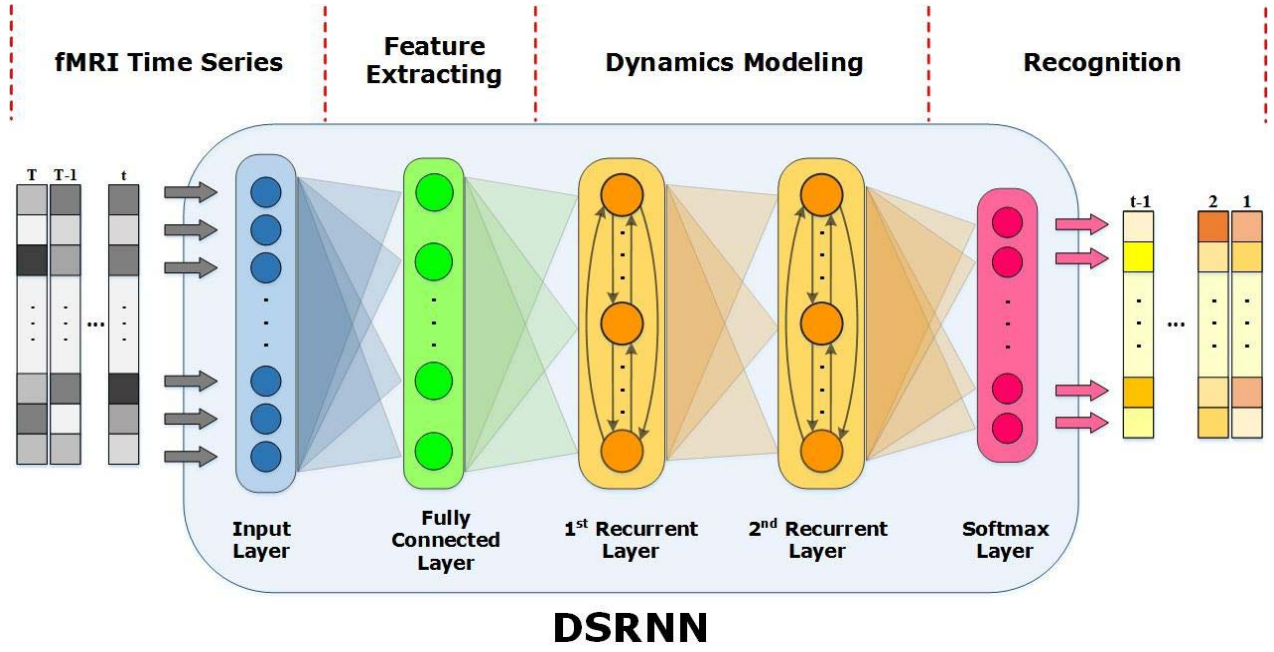
Fig. 2. Overview of the DSRNN model. The stretched fMRI vector sequences are fed to the input layer of DSRNN model. One fully connected layer is used to extract activated brain regions. Two recurrent layers with dropout are cascaded to model temporal dynamics. Finally, a softmax classifier is arranged for the brain state recognition.

follows:

$$h^t = \left(1 - z^t\right) \odot h^{t-1} + z^t \odot \bar{h}^t \qquad (8)$$

$$z^t = \sigma\left(U_z h^{t-1} + W_z x^t + b_z\right) \qquad (9)$$

$$r^t = \sigma\left(U_r h^{t-1} + W_r x^t + b_r\right) \qquad (10)$$

$$\bar{h}^t = \tanh\left(U_h \left(r^t * h^{t-1}\right) + W_h x^t + b_h\right) \qquad (11)$$

where z^t is update gate activation, r^t is reset gate activation, and $\tilde{h}^t$ is an auxiliary variable. Like the gates in LSTM unit, those in GRU unit also control the information flow. As LSTM, trainable parameters, such as weights U_z, U_r, U_h and W_z, W_r, W_h, and biases b_z, b_r, b_h, determine the activations of the gates.

As shown in Fig. 1(d), a recurrent neural network can be unfolded as a deep feedforward network along time. Conversely, when the numbers of units and connections are limited, a desirable function computed by a very large feedforward network might alternatively be approximated by recurrent computations in a smaller network. Recurrent network can expand computational power by multiplying the limited physical resources for computation along time.

C. Deep Sparse Recurrent Neural Network

As described beforehand, the feedforward sweep alone might not provide the sufficient evidences that the output layer needs to confidently recognize brain states. Instead, an ideal model should integrate later-arriving lateral and top-down signals to converge on its ultimate response. Therefore, in order to better capture the temporal dependences preserved in fMRI time series and recognize the dynamical brain states, a five-layer DSRNN model is proposed, as shown in Fig 2.

Adjacent layers are fully connected, and the connections between layers are forward. The fully connected layer plays a role as a filter of activated brain regions, which aims to winnow and combine significant fMRI volumes with proper weight values. Then, two recurrent layers are employed following the same structure, as shown in Fig.2, so that various scales of temporal dependences can be captured. The hidden state of second recurrent layer is defined similarly to that of first recurrent layer, except for replacing the input with the hidden state of first layer. Finally, we apply a softmax layer to obtain a vector of class probabilities:

$$\hat{y}_t = softmax\left(U_y h^t + b_y\right) \qquad (12)$$

$$p\left(y_{tk} = 1 | x_1, x_2, \cdots, x_t\right) = \hat{y}_t \qquad (13)$$

where $softmax\left(x\right) = exp\left(x\right) / \sum_i exp\left(x_i\right)$.

In the training process of DSRNN, the optimization objective is made up of three components shown as follows:

$$J\left(W, b\right) = CE\left(y \| \hat{y}\right) + \beta \left\| O_{Full} \right\| + \lambda \left\| W_{Full} \right\| \qquad (14)$$

The first term is the overall error cost. In the case of multiclass classification, the error cost is defined as the cross entropy (CE) between the true labels $\{y\}$ and the predicted labels $\{\hat{y}\}$. The cross entropy for the discrete distributions P and Q over a given set is defined as follows:

$$CE\left(P, Q\right) = -\sum_i P\left(i\right) log Q\left(i\right) \qquad (15)$$

The second term is the sparseness penalty of the activation (output) O_{Full} of fully connected layer, and L_1 norm helps to concentrate task related volumes in only a few patterns. β controls the weight of sparseness penalty term.

The last term is also a sparseness penalty which forces W_{Full} to be sparse. W_{Full} is the weight matrix of fully-connected layer. We use L_1 norm as the penalty cost, and λ is the weight parameter of this penalty term. Because activated brain regions are extracted from W_{Full}, proper sparseness of W_{Full} can help to remove those brain areas which are not sensitive to task events and to make significantly activated brain regions stand out.

DSRNN adopts the feedforward sweep to extract elementary features, and then uses subsequent recurrent computations to explain the nonlinear interactions of these features. In such a process, the information maintained in fMRI signals would be gradually uncluttered in the low-level representation and contextualized by the high-level representation. Therefore, the higher level brain states are recognized accurately and brain dynamics are modeled gradually.

The recognition accuracy is defined as the proportion of correctly recognized labels, and the calculation is as follows:

$$Accuracy = \frac{N(Labels_{Rec} == Labels_{GT})}{SequenceLength} \qquad (16)$$

where $Labels_{Rec}$ are the recognized label sequences, and $Labels_{GT}$ are the ground truth labels. Therefore, numerator is the number of time points which are correctly recognized.

In this paper, the DSRNN model is implemented with Tensorflow. Truncated back propagation through time is employed in conjunction with AdamOptimizer to train the models by iteratively minimizing the objective function (Equation (14)). Besides, dropout operation [71] was applied on the inputs of each recurrent layer, which enhances the generalization of DSRNN model. The DSRNN model was run on the computing platform with two GPUs (Nvidia GTX 1080 Ti, 12G Graphic Memory) and 64 G memory.

D. Random Fragment Strategy

In this paper, the datasets we used are from the Human Connectome Project (HCP) 900 Subjects Data Release. In HCP, all subjects followed the same task designs, and the recorded event label sequences were almost the same. However, if the whole fMRI series and event label sequences were used for training, it would cause a severe overfitting problem, that is, the DSRNN would only learn the task designs, rather than the temporal dependences hidden in fMRI time series, and the training process would be meaningless.

In order to overcome this drawback of datasets, we propose a novel training strategy: random fragment strategy (RFS). In this strategy, sequence fragment with random length is cut from random location of the sequence. Thus, in each iteration, new different sequence fragments are fed to DSRNN. In this paper, the fragment length ranges from 30%~50% of the whole sequence length. By creating huge amounts of distinct sequences with different lengths, RFS enriches the diversity of datasets and makes sure that the DSRNN learns the temporal dependences preserved in fMRI time series, rather than changeless arrangement of task events.

III. RESULTS

A. Overview of the DSRNN Model

In order to validate the proposed DSRNN framework, we tested the DSRNN model on the Human Connectome Project (HCP) 900 Subjects Data release. Specifically, there are seven task fMRI datasets with more than 800 subjects which is one of the biggest publicly available tfMRI dataset. In the following sections, we will first go over the details of the DSRNN model and then discuss the brain state recognition results and associated important brain areas and the effects of different parameter settings. For convenience, we take motor task as an example, which includes the maximum stimulus events in HCP data sets.

As shown in Fig. 3(a)~3(b), the 3-D whole brain fMRI recordings of time t, where brain state is hidden, are sent to the input layer of DSRNN after being stretched into a 1-D (28549*1) vector. Then, the fully connected layer extracts activated brain regions, and converts the original large dimension (28549) recordings into small size (32) activations. Specially, activated brain regions are gradually filtered during the whole training process, and represented by the weights of fully connected layer, as shown in Fig. 3(d) ~ 3(e). As determined beforehand, we arranged 32 neuron units in fully connected layer, therefore, we obtained 32 groups of activated brain regions. After all the trainable parameters of DSRNN are settled, the activated brain regions are confirmed. In Fig. 3(e), it is easy to observe that there are a few specific brain areas activated in each group. In general, original recordings contain raw brain information, and they are definitely redundant. After this filtering process of fully connected layer, only those significant and distinctive information are preserved, and the redundancy is reduced largely (from 28,549 to 32). Therefore, the brain state corresponding to each task event can be indicated by only a few activated brain region groups. The activations of these region groups make up the outputs of fully connected layer, as shown in Fig. 3(f), where the vector size is much smaller than those in Fig. 3(b).

After assembling the output vectors (Fig. 3(f)) along time, we can obtain an activation trace map, as shown in Fig. 3(g), where the brain state shift can be illustrated more clearly. In this paper, we regard task design as ground truth since different task stimulus events are presented to the participants according to the task design paradigm. Comparing activation traces with the ground truth at the top of Fig. 3(g), it is easy to recognize matching relations between them. In addition, some activated region groups have similar shapes in activation trace, which indicates that these groups of regions may be activated simultaneously and they functionally cooperate with each other. Then, we further cluster all the region groups into several classes based on their activation traces, and the clustered trace map is shown in Fig. 3(h). In motor task, there are five different motion events: left foot, right foot, left hand, right hand, tongue. In Fig. 3(h), it is easy to observe that there are 5 clustered activation traces matching 5 motion events closely. We pick up the brain activation map corresponding to each motion, and show them in Fig. 3(i). In this way, the whole brain activation maps of each task event are obtained.
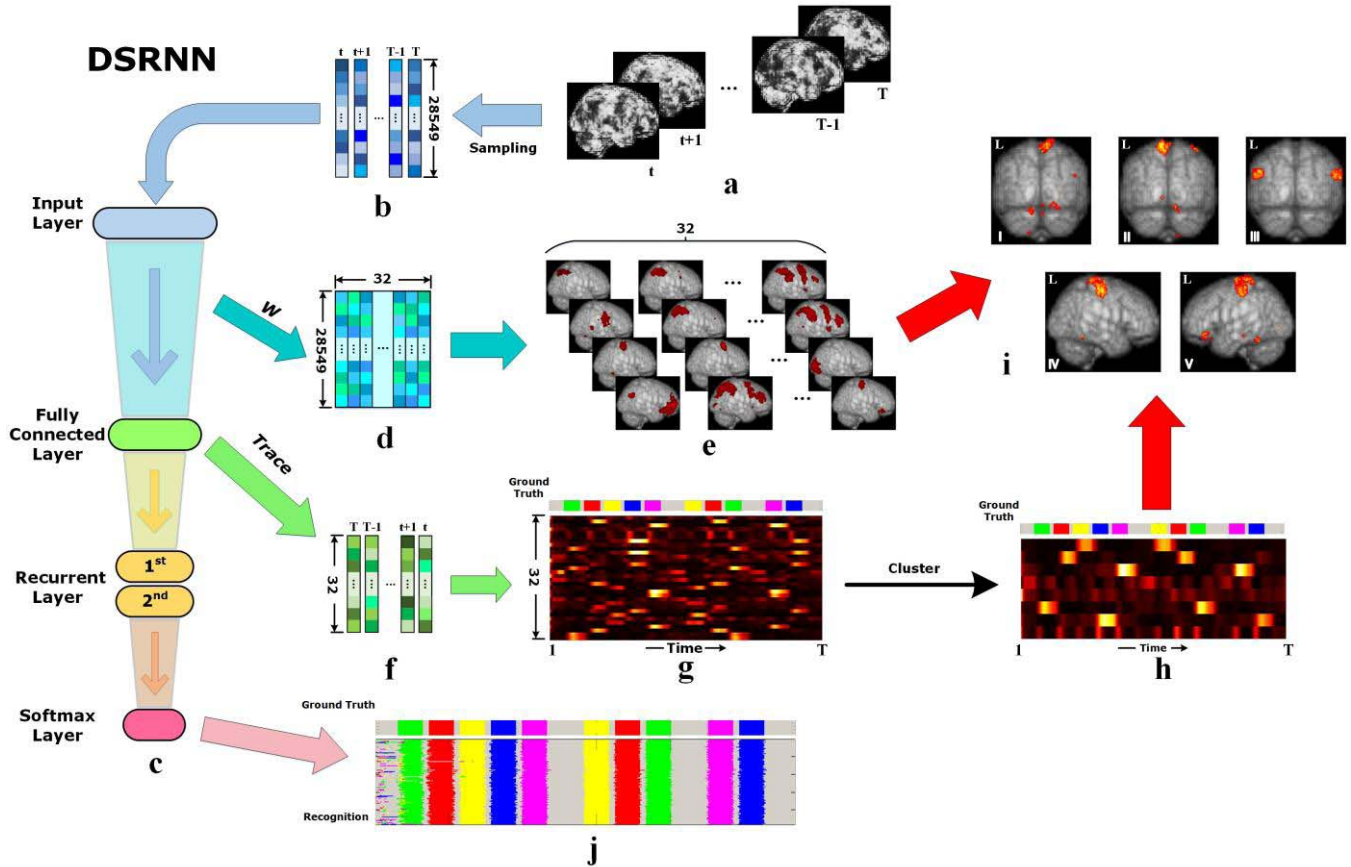
Fig. 3. Illustration of DSRNN modeling process. (a) Raw brain tfMRI images series. (b) Vectorization of tfMRI image series after sampling. (c) Hierarchy diagram of DSRNN model. (d) Weight matrix between input layer and fully connected layer. (e) Visualization of activated brain region groups represented by weight matrix of fully connected layer. (f) Output time series of fully connected layer. Each column vector indicates the activation of 32 distinctive brain region groups. (g) Visualization of fully time-scale output time series of fully connected layer, also the activation traces of 32 brain region groups. Motor task is used as an example, and the ground-truth temporal distribution of event blocks is illustrated at the top. (h) Visualization of clustered activation traces. (i) Visualization of brain activation maps corresponding to 5 motions, based on 32 brain region groups (e) and their clustered activation traces (h). Brain activation maps from I to VI are corresponding to the motion events of right foot, left foot, tongue, left hand and right hand, respectively. (j) Recognition of brain states.

In addition, the activation traces are fed to recurrent layers to capture temporal dependences, which is vitally important to model dynamic system. With the outputs of the 2nd recurrent layer, softmax classifier can make an accurate recognition of brain states, as shown in Fig. 3(j).

B. Brain State Recognition via DSRNN

In this paper, we adopted HCP 900 Subjects data release as testbeds, including 786 subjects with good data quality, which guarantees the diversity and universality of the experiment. Although a $4 \times 4 \times 4$ mm^3 down-sampling is applied to reduce training data size, we still cannot put all subjects' data in our computing platform's memory (64G Memory, 24G Graphic Memory in total). Thus, we have to randomly pick up as many subjects as memory can hold. The quantity of training and testing subjects of each task is shown in Table 2.

In Fig. 3(f) $\sim$ 3(g), activation traces are obtained from fully connected layer, before where all processes have no relationships with temporal dependences processing. Since the human brain is a deep and complicated neural network [44], in which feedforward, lateral, and feedback connections are

TABLE II
TRAINING AND TESTING SUBJECT NUMBERS OF SEVEN TASKS

Task	Train	Test
WM	240	240
GAMBLING	320	320
MOTOR	320	320
LANGUAGE	300	300
SOCIAL	320	320
RELATIONAL	320	320
EMOTION	400	360

widespread, we suppose that there must be recurrent information flow of dynamics hidden in the fMRI time series. This is also the reason why we employed recurrent layers in DSRNN model. To validate this supposition, we compared the recurrent classifier (DSRNN model) with an auto correlative classifier (Adaptive Autoregressive Classifier, AAR) and two non-temporal classifiers (Softmax and SVM) respectively, as shown in Fig. 4.

After training hundreds of subjects, DSRNN model achieved outstanding performance in test datasets, and the results are illustrated in Fig.5 and Fig. 6. From Fig. 5, it is
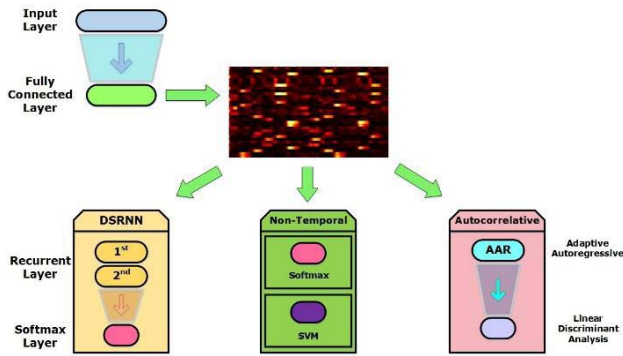
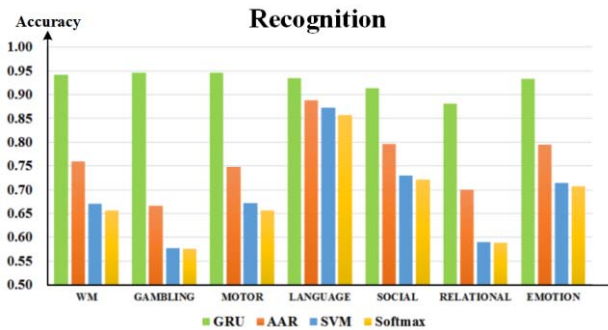Fig. 4. Brain states recognition with recurrent layers, Softmax, SVM and AAR based on activation traces.

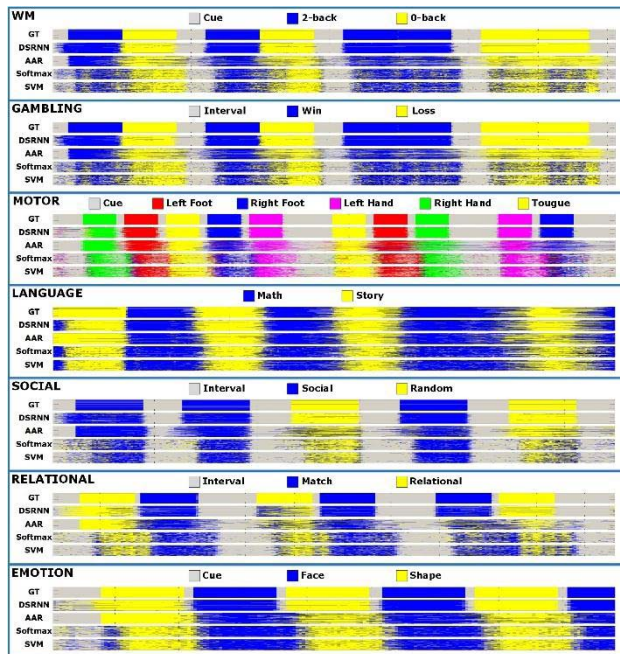Fig. 5. Brain state recognition accuracies of seven tasks.

Fig. 6. Brain state series of seven tasks. In each subgraph, five state series from top to bottom are ground truth (GT), and series recognized by DSRNN, AAR, Softmax and SVM, respectively.

easy to appreciate that DSRNN model recognized brain states of seven tasks with the highest accuracies (above 90%, except relational task 88%). The rank of recognition accuracy from high to low is: DSRNN > AAR > Softmax ≈ SVM. Fig. 6 illustrates the recognized brain states using different methods across all subjects. It is easy to observe from Fig. 6 that DSRNN model detected every subtask block in each task

accurately. It accurately recognized not only the task durations, but also the change points of states. In DSRNN's series, event blocks are very clear, even for the "Interval/Cue" blocks. In contrast with DSRNN, auto-correlation classifier (AAR) and non-temporal classifiers (Softmax and SVM) perform much worse at the state change points, and there are also plenty of noises which make event blocks ambiguous. Besides, DSRNN's recognition of the first event block in each task is not so accurate, compared with other blocks behind. This is quite reasonable, because DSRNN needs some input sequences to initially build its preservation of memory.

Before maintaining enough temporal information of previous series, the recognition cannot reach the best performance. This result further indicates the importance of temporal dependences in brain state recognition.

To conclude, our works demonstrated that DSRNN has better performance in brain state recognition than other models, especially those non-temporal classifiers.

C. Brain Activation Maps

An important characteristic of the proposed DSRNN model is that we can also obtain the important activated brain areas involved in brain state recognition. During brain state recognition, distinctive activated brain regions for each task event were extracted (illustrated in Fig. 3(i)). In this paper, we applied DSRNN model on seven HCP tasks, therefore, seven groups of brain activation maps of each task were obtained. In order to interpret these associated brain areas involved in brain state recognition, we compared these spatial maps with the activation maps obtained by GLM method, we find that interesting correspondence between these maps. For convenience, we also take motor task as an example and put other task's result into the section B of supplementary materials. Supplementary materials are available in the supplementary files /multimedia tab.

The activation maps for the motor task are shown in Fig. 7. Events of Left foot and right foot show great activations in the posterior portion of the frontal lobe. For left hand and right hand, besides these distinctions, activations in precentral cortex are very clear. In addition, bilateral precentral cortex are activated during tongue movement. In motor task, we can see clear spatial differentiation of the activations in motor cortex: left and right hand/foot motion have expected contralateral activation locations, whereas tongue motion activates bilateral regions.

D. Effects of Hyper Parameters in DSRNN

In DSRNN, there are a few hyper parameters, such as cell unit number of each layer, recurrent layer number, and recurrent cell unit type, which determine the performance of DSRNN. Therefore, before the experiment, we need to determine all these hyper parameters carefully. In this paper, we picked up 100 subjects from working memory task datasets as test samples. By five-fold cross validation (80 subjects for training and 20 for test in each trial), we explored the effects on DSRNN's recognition performance of each hyper parameter, and the exploration results are shown in Fig. 8 and Fig. 9.
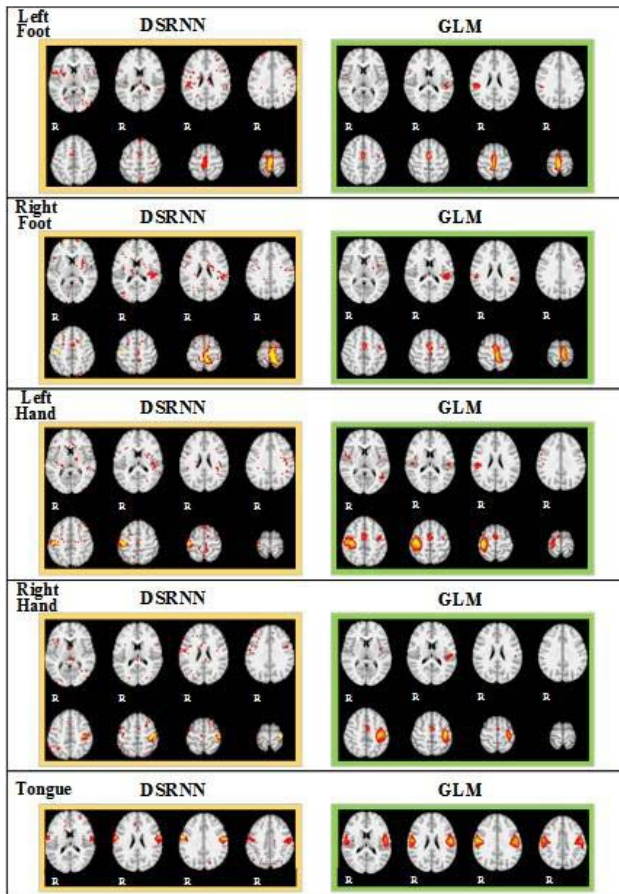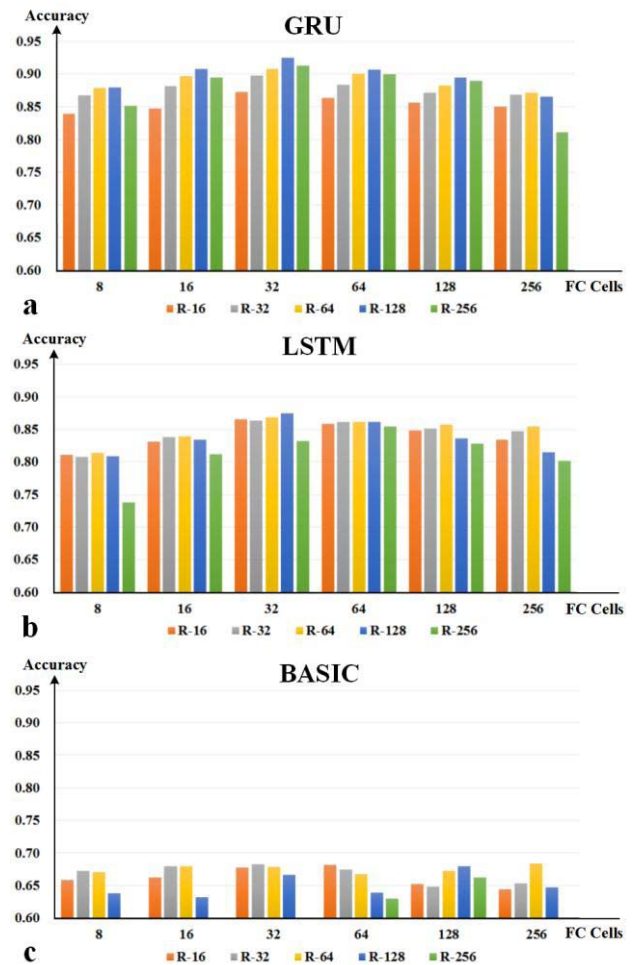
Fig. 7. Activation maps of motor task.

Fig. 8. The exploration results of hyper parameters. (a) Results of GRU-based network. (b) Results of LSTM-based network. (c) Results of basic unit-based network. The horizontal axis represents the number of neurons in fully connected layer, and the vertical axis indicates the recognition accuracy. All networks contain two recurrent layers, and bars in different colors indicate different numbers of cell units per recurrent layer.
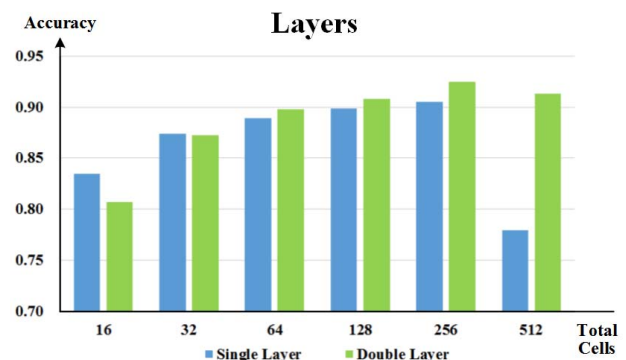
Fig. 9. The exploration results of DSRNN with different numbers of recurrent layers. The horizontal axis represents the total number of recurrent cell units, and the vertical axis indicates the recognition accuracy. Bars in different colors indicate different recurrent layer numbers.

According to previous RNN studies [48], two recurrent layers might capture multi-scale temporal dependences in series, and they outperformed the network with only single recurrent layer. So in the exploration of hyper parameters, we assumed that DSRNN with two recurrent layers might perform better than only one layer. Based on this assumption, we explored the effects of three hyper parameters: the cell number of fully connected layer, the cell unit number of each recurrent layer, and the cell type of recurrent layer. In Fig. 8(a), with GRU unit being the recurrent cell, DSRNN reached the highest performance when fully connected layer cell number was 32, and 128 GRU units were in each recurrent layer. Then, we repeated the experiment with LSTM unit and basic cell unit, and the results are shown in Fig. 8(b) and Fig. 8(c). In Fig. 8(b), network with 32 cells in fully connected layer and 128 LSTM units in each recurrent layer also performed slightly better than others. Compared with the results of GRU units, LSTM units achieved lower recognition accuracy by about 5%, on average. As to basic recurrent unit, the accuracy of recognition was even worse, which was less than 70%, and some were even below 60%. Therefore, after comparing the performances of different combinations of three hyper parameters (the cell number of fully connected layer, the cell unit number of each recurrent layer, and the cell type of recurrent layer), we determined the architecture of DSRNN, which included a 32-cell fully connected layer, and two 128-GRU-based recurrent layers.

As all other hyper parameters were confirmed, we explored the influences of different recurrent layer numbers on DSRNN's performances. In order to make a comparison, we made sure that both architectures had the same number of recurrent cell units (GRU), so that the performance would

Fig. 10. The exploration results of sparseness penalty weights: beta and lambda. The horizontal axis indicates the values of lambda, and the vertical axis denotes the recognition accuracy. Bars in different colors indicate different beta values.

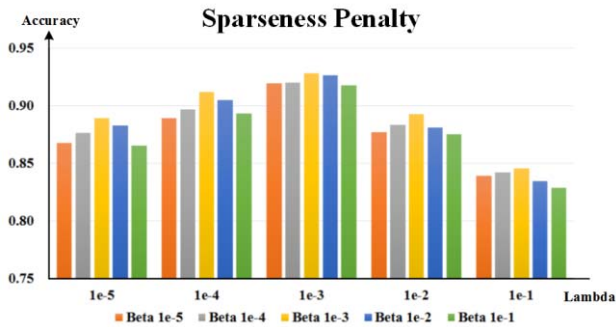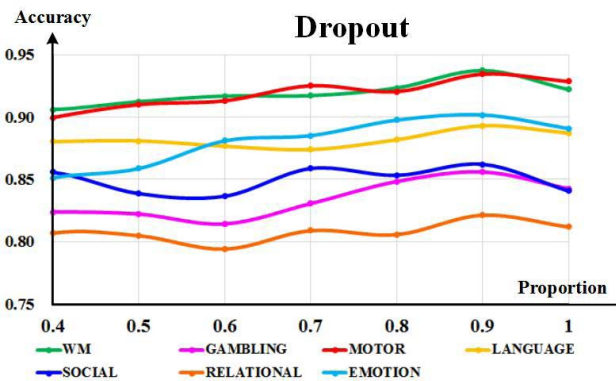Fig. 11. The exploration of dropout proportion based on seven tasks. The horizontal axis indicates the proportion, and the vertical axis denotes the recognition accuracy. Curves in different colors indicate different tasks.

be only determined by the cell arrangement. The results are shown in Fig. 9. When there were only a small number of recurrent units, all units arranging in parallel in one layer achieved better recognition accuracy than cascaded arrangement. With the increase of cell number, the accuracies of both architectures improved, and the two-layer improved more. With a total of 256 recurrent cell units, both architectures reached the top, and two-layer architecture performed better. This result demonstrated that with sufficient cell units, arrangement of two cascaded recurrent layers could capture time dependences better than using only single layer arrangement.

E. Effects of Free Parameters in DSRNN

Based on the optimal structure of DSRNN, the free parameters of DSRNN, which should be determined before formal experiments, can be explored. In this paper, there are three free parameters that we needed to focus on: two sparseness penalty weights β and λ, and keeping proportion of dropout.

To investigate β and λ, we tried orders of magnitude ranging from 1e-6 to 1e-1. The recognition results of different combinations of β and λ are shown in Fig. 10. It is easy to see that λ has greater influences on DSRNN model than β, while β provides fine-tuning to DSRNN's performance. When β and λ both were set 1e-3, DSRNN model achieved the best performance.

Dropout operation is also introduced to enhance the generalization of DSRNN model. As all other parameters are decided,

it is feasible to select the best proportion based on all seven tasks. We tried the proportion from 0.4 to 1.0, and found that DSRNN performed the best at 0.9. It can be inferred that there is high consistency among subjects in each task, and DSRNN can model fast time-scale brain states effectively and robustly.

IV. DISCUSSION AND CONCLUSION

In this paper, a 5-layer deep sparse recurrent neural network (DSRNN) model is proposed to model the dynamic brain states in task fMRI data. With outstanding capability of capturing sequence temporal dependence, DSRNN can recognizes brain states accurately. In addition, the associated brain activated regions also demonstrate meaningful correspondence with traditional GLM activation results which provide us novel insight on functional brain activities. Comparing with common auto-correlation modeling method (AAR) and traditional non-temporal modeling approaches (Softmax and SVM), DSRNN achieved obviously outstanding recognition accuracy.

Though current DSRNN model has achieved excellent performances in brain states recognition accuracy, it still can be developed in a few aspects. Firstly, although the fully connected layer applied in DSRNN extracted distinctive activated brain regions successfully, there are many other neural network structures, such as CNN, DBN, etc., to be further revisited. These models are also widely used in computer vision and feature extraction, and have more complicated structure than fully connected layer. Better performance might be achieved if these models are employed. Secondly, in this paper, we recognize the brain states of each single task and obtained seven different state recognition models. Each model can identify subtasks or brain states of only one category of task. In the future, we can train one DSRNN model with all seven tasks. In this way, more distinctive activation maps and more differences can be obtained among tasks. Thirdly, as a proposal, DSRNN model might be applied online in the future. Because once the DSRNN model is trained successfully, real-time tfMRI data time seies can be fed online and the brain behaviors can be detected in quite a short time. This application can be used widely as a brain states identifier.

Finally, there are several potential applications in basic neuroscience and clinical research with the proposed DSRNN model. For instance, it has been observed that dynamic performance can be sensitive to psychiatric or neurologic disorders, and the associated brain areas in brain state recognition may provide novel insight for clinical diagnosis. Besides, many literatures suggest that the brain activities and states are under dynamical changes and the proposed model provide a useful tool to recognize the brain states at fast time-scale in task fMRI data. In general, our proposed DSRNN model offers a new methodology for basic and clinical neuroscience research.

REFERENCES

[1] K. J. Friston, A. P. Holmes, K. J. Worsley, J.-P. Poline, C. D. Frith, and R. S. Frackowiak, "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Mapping*, vol. 2, no. 4, pp. 189–210, 1994.

[2] M. J. Mckeown *et al.*, "Spatially independent activity patterns in functional MRI data during the Stroop color-naming-task," *Proc. Nat. Acad. Sci. USA*, vol. 95, no. 3, pp. 803–810, 1998.

[3] J. Lv *et al.*, "Sparse representation of whole-brain fMRI signals for identification of functional networks," *Med. Image Anal.*, vol. 20, no. 1, pp. 112–134, 2014.

[4] S. Zhao *et al.*, "Supervised dictionary learning for inferring concurrent brain networks," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 2036–2045, Oct. 2015.

[5] X. Jiang, L. Zhao, H. Liu, L. Guo, K. M. Kendrick, and T. Liu, "A cortical folding pattern-guided model of intrinsic functional brain networks in emotion processingm," *Frontiers Neurosci.*, vol. 12, p. 575, Aug. 2018.

[6] J. Lv *et al.*, "Holistic atlases of functional networks and interactions reveal reciprocal organizational architecture of cortical function," *IEEE Trans. Biomed. Eng.*, vol. 62, no. 4, pp. 1120–1131, Apr. 2015.

[7] S. Zhao *et al.*, "Decoding auditory saliency from brain activity patterns during free listening to naturalistic audio excerpts," *Neuroinformatics*, vol. 16, nos. 3–4, pp. 309–324, 2018.

[8] K. J. Friston, "Transients, metastability, and neuronal dynamics," *NeuroImage*, vol. 5, no. 2, pp. 164–171, Feb. 1997.

[9] C. D. Gilbert and A. M. Sigman, "Brain states: Top-down influences in sensory processing," *Neuron*, vol. 54, no. 5, pp. 677–696, 2007.

[10] M. A. Lindquist, C. Waugh, and T. D. Wager, "Modeling state-related fMRI activity using change-point theory," *NeuroImage*, vol. 35, no. 3, pp. 1125–1141, Apr. 2007.

[11] M. D. Fox and M. E. Raichle, "Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging," *Nature Rev. Neurosci.*, vol. 8, no. 9, pp. 700–711, Sep. 2007.

[12] S. M. Smith *et al.*, "Temporally-independent functional modes of spontaneous brain activity," *Proc. Nat. Acad. Sci. USA*, vol. 109, no. 8, pp. 3131–3136, 2012.

[13] A. A. Fingelkurts and A. A. Fingelkurts, "Making complexity simpler: Multivariability and metastability in the brain," *Int. J. Neurosci.*, vol. 114, no. 7, pp. 843–862, 2004.

[14] C. Chang and G. H. Glover, "Time–frequency dynamics of resting-state brain connectivity measured with fMRI," *NeuroImage*, vol. 50, no. 1, pp. 81–98, 2010.

[15] J. Ou *et al.*, "Atomic dynamic functional interaction patterns for characterization of ADHD," *Hum. Brain Mapping*, vol. 35, no. 10, pp. 5262–5278, 2014.

[16] X. Li *et al.*, "Dynamic functional connectomics signatures for characterization and differentiation of PTSD patients," *Hum. Brain Mapping*, vol. 35, no. 4, pp. 1761–1778, 2014.

[17] X. Zhang *et al.*, "Identifying and characterizing resting state networks in temporally dynamic functional connectomes," *Brain Topogr.*, vol. 27, no. 6, pp. 747–765, 2014.

[18] J. Zhang *et al.*, "Inferring functional interaction and transition patterns via dynamic Bayesian variable partition models," *Hum. Brain Mapping*, vol. 35, no. 7, pp. 3314–3331, 2014.

[19] J. Ou *et al.*, "Characterizing and differentiating brain state dynamics via hidden Markov models," *Brain Topogr.*, vol. 28, no. 5, pp. 666–679, 2015.

[20] D. A. Handwerker, V. Roopchansingh, J. Gonzalez-Castillo, and P. A. Bandettini, "Periodic changes in fMRI connectivity," *NeuroImage*, vol. 63, no. 3, pp. 1712–1719, 2012.

[21] E. A. Allen, E. Damaraju, S. M. Plis, E. B. Erhardt, T. Eichele, and V. D. Calhoun, "Tracking whole-brain connectivity dynamics in the resting state," *Cerebral Cortex*, vol. 24, no. 3, pp. 663–676, 2014.

[22] S. D. Keilholz, C. Caballerogaudes, P. Bandettini, G. Deco, and V. D. Calhoun, "Time-resolved resting state fMRI analysis: current status, challenges, and new directions," *Brain Connectivity*, vol. 7, no. 8, pp. 465–481, 2017.

[23] R. M. Hutchison, T. Womelsdorf, J. S. Gati, S. Everling, and R. S. Menon, "Resting-state networks show dynamic functional connectivity in awake humans and anesthetized macaques," *Hum. Brain Mapping*, vol. 34, no. 9, pp. 2154–2177, 2013.

[24] S. D. Keilholz, M. E. Magnuson, W.-J. Pan, M. Willis, and G. J. Thompson, "Dynamic properties of functional connectivity in the rodent," *Brain Connectivity*, vol. 3, no. 1, pp. 31–40, 2013.

[25] N. Petridou, C. C. Gaudes, I. L. Dryden, S. T. Francis, and P. A. Gowland, "Periods of rest in fMRI contain individual spontaneous events which are related to slowly fluctuating spontaneous activity," *Hum. Brain Mapping*, vol. 34, no. 6, pp. 1319–1329, 2013.

[26] X. Zhang *et al.*, "Characterization of task-free and task-performance brain states via functional connectome patterns," *Med. Image Anal.*, vol. 17, no. 8, pp. 1106–1122, 2013.

[27] Ü. Sakoğlu, G. D. Pearlson, K. A. Kiehl, Y. M. Wang, A. M. Michael, and V. D. Calhoun, "A method for evaluating dynamic functional network connectivity and task-modulation: Application to schizophrenia," *Magn. Reson. Mater. Phys., Biol. Med.*, vol. 23, nos. 5–6, pp. 351–366, 2010.

[28] Z. Lian *et al.*, "Exploring functional brain dynamics via a Bayesian connectivity change point model," in *Proc. IEEE Int. Symp. Biomed. Imag.*, Apr. 2014, pp. 600–603.

[29] Z. Lian *et al.*, "Generalized fMRI activation detection via Bayesian magnitude change point model," in *Proc. IEEE 11th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2014, pp. 21–24.

[30] Y. Xu and M. A. Lindquist, "Dynamic connectivity detection: An algorithm for determining functional connectivity change points in fMRI data," *Frontiers Neurosci.*, vol. 9, p. 285, Sep. 2015.

[31] S. Chen, J. Langley, X. Chen, and X. Hu, "Spatiotemporal modeling of brain dynamics using resting-state functional magnetic resonance imaging with Gaussian hidden Markov model," *Brain Connect*, vol. 6, no. 4, pp. 326–334, 2016.

[32] E. Tagliazucchi, P. Balenzuela, D. Fraiman, P. Montoya, and D. R. Chialvo, "Spontaneous BOLD event triggered averages for estimating functional connectivity at resting state," *Neurosci. Lett.*, vol. 488, no. 2, pp. 158–163, 2011.

[33] E. Tagliazucchi, P. Balenzuela, D. Fraiman, and D. R. Chialvo, "Criticality in large-scale brain fMRI dynamics unveiled by a novel point process analysis," *Front. Physiol.*, vol. 3, p. 15, Feb. 2012.

[34] C. C. Gaudes, N. Petridou, S. T. Francis, I. L. Dryden, and P. A. Gowland, "Paradigm free mapping with sparse regression automatically detects single-trial functional magnetic resonance imaging blood oxygenation level dependent responses," *Hum. Brain Mapping*, vol. 34, no. 3, pp. 501–518, 2013.

[35] F. I. Karahanoğlu, C. Caballero-Gaudes, F. Lazeyras, and D. Van De Ville, "Total activation: FMRI deconvolution through spatio-temporal regularization," *NeuroImage*, vol. 73, pp. 121–134, Jun. 2013.

[36] D. Vidaurre *et al.*, "Discovering dynamic brain networks from big data in rest and task," *NeuroImage*, vol. 180, pp. 646–656, Oct. 2017.

[37] D. Vidaurre, A. J. Quinn, A. P. Baker, D. Dupret, A. Tejero-Cantero, and M. W. Woolrich, "Spectrally resolved fast transient brain states in electrophysiological data," *NeuroImage*, vol. 126, pp. 81–95, Feb. 2016.

[38] A. Graves. (2014). "Generating sequences with recurrent neural networks." [Online]. Available: https://arxiv.org/abs/1308.0850

[39] A. Graves, M. Liwicki, S. Fernández, R. Bertolami, H. Bunke, and J. R. Schmidhuber, "A novel connectionist system for unconstrained handwriting recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 5, pp. 855–868, 2008.

[40] I. Sutskever, J. Martens, and G. E. Hinton, "Generating text with recurrent neural networks, in book generating text with recurrent neural networks," in *Proc. 28th Int. Conf. Mach. Learn.*, 2011, pp. 1017–1024.

[41] K. Cho *et al.* (2014). "Learning phrase representations using RNN encoder-decoder for statistical machine translation." [Online]. Available: https://arxiv.org/abs/1406.1078

[42] H. Sak, A. Senior, and F. Beaufays. (2014). "Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition." [Online]. Available: https://arxiv.org/abs/1402.1128

[43] U. Güçlü and M. A. J. van Gerven, "Deep neural networks reveal a gradient in the complexity of neural representations across the ventral stream," *J. Neurosci.*, vol. 35, no. 27, p. 10005, 2015.

[44] N. Kriegeskorte, "Deep neural networks: A new framework for modeling biological vision and brain information processing," *Annu. Rev. Vis. Sci.*, vol. 1, no. 1, pp. 417–446, 2015.

[45] R. Dipietro *et al.*, "Recognizing surgical activities with recurrent neural networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2016, pp. 551–558.

[46] D. L. K. Yamins and J. J. DiCarlo, "Using goal-driven deep learning models to understand sensory cortex," *Nature Neurosci.*, vol. 19, no. 3, pp. 356–365, 2016.

[47] W. Xue, G. Brahm, S. Pandey, S. Leung, and S. Li, "Full left ventricle quantification via deep multitask relationships learning," *Med. Image Anal.*, vol. 43, pp. 54–65, Jan. 2017.

[48] U. Güçlü and M. A. J. van Gerven, "Modeling the dynamics of human brain activity with recurrent neural networks," *Frontiers Comput. Neurosci.*, vol. 11, p. 7, Feb. 2017.

[49] D. M. Barch *et al.*, "Function in the human connectome: Task-fMRI and individual differences in behavior," *NeuroImage*, vol. 80, no. 8, pp. 169–189, 2013.

[50] G. Pfurtscheller, C. Neuper, A. Schlogl, and K. Lugger, "Separability of EEG signals recorded during right and left motor imagery using adaptive autoregressive parameters," *IEEE Trans. Rehabil. Eng.*, vol. 6, no. 3, pp. 316–325, Sep. 1998.

[51] A. Schlögl, D. Flotzinger, and G. Pfurtscheller, "Adaptive autoregressive modeling used for single-trial EEG classification," *Biomed. Tech. Biomed. Eng.*, vol. 42, no. 6, pp. 162–167, 1997.

[52] *WU-Minn HCP 900 Subjects Data Release: Reference Manual*.

[53] A. Drobyshevsky, S. B. Baumann, and W. Schneider, "A rapid fMRI task battery for mapping of visual, motor, cognitive, and emotional function," *NeuroImage*, vol. 31, no. 2, pp. 732–744, 2006.

[54] A. Caceres, D. L. Hall, F. O. Zelaya, S. C. Williams, and M. A. Mehta, "Measuring fMRI reliability with the intra-class correlation coefficient," *NeuroImage*, vol. 45, no. 3, pp. 758–768, 2009.

[55] M. R. Delgado, L. E. Nystrom, C. Fissell, D. C. Noll, and J. A. Fiez, "Tracking the hemodynamic responses to reward and punishment in the striatum," *J. Neurophysiol.*, vol. 84, no. 6, pp. 3072–3077, 2000.

[56] E. E. Forbes *et al.*, "Altered striatal activation predicting real-world positive affect in adolescent major depressive disorder," *Amer. J. Psychiatry*, vol. 166, no. 1, pp. 64–73, 2009.

[57] J. C. May *et al.*, "Event-related functional magnetic resonance imaging of reward-related brain circuitry in children and adolescents," *Biol. Psychiatry*, vol. 55, no. 4, pp. 359–366, 2004.

[58] E. M. Tricomi, M. R. Delgado, and J. A. Fiez, "Modulation of caudate activity by action contingency," *Neuron*, vol. 41, no. 2, pp. 281–292, 2004.

[59] F. M. Krienen *et al.*, "The organization of the human cerebral cortex estimated by intrinsic functional connectivity," *J. Neurophysiol.*, vol. 106, no. 3, pp. 1125–1165, 2011.

[60] J. R. Binder *et al.*, "Mapping anterior temporal lobe language areas with fMRI: A multicenter normative study," *NeuroImage*, vol. 54, no. 2, pp. 1465–1475, 2011.

[61] F. Castelli, F. Happé, U. Frith, and C. Frith, "Movement and mind: A functional imaging study of perception and interpretation of complex intentional movement patterns," *NeuroImage*, vol. 12, no. 3, pp. 314–325, 2000.

[62] F. Castelli, C. Frith, F. Happé, and U. Frith, "Autism, Asperger syndrome and brain mechanisms for the attribution of mental states to animated shapes," *Brain*, vol. 125, no. 8, pp. 1839–1849, 2002.

[63] T. Wheatley, S. C. Milleville, and A. Martin, "Understanding animate agents: Distinct roles for the social network and mirror system," *Psychol. Sci.*, vol. 18, no. 6, pp. 469–474, 2007.

[64] S. J. White, D. Coniston, R. Rogers, and U. Frith, "Developing the Frith–Happé animations: A quick and objective test of theory of mind for adults with autism," *AUTISM Res.*, vol. 4, no. 2, pp. 149–154, 2011.

[65] R. Smith, K. Keramatian, and K. Christoff, "Localizing the rostrolateral prefrontal cortex at the individual level," *NeuroImage*, vol. 36, no. 4, pp. 1387–1396, 2007.

[66] A. R. Hariri, A. Tessitore, V. S. Mattay, F. Fera, and D. R. Weinberger, "The amygdala response to emotional stimuli: A comparison of faces and scenes," *NeuroImage*, vol. 17, no. 1, pp. 317–323, 2002.

[67] S. B. Manuck, S. M. Brown, E. E. Forbes, and A. R. Hariri, "Temporal stability of individual differences in amygdala reactivity," *Amer. J. Psychiatry*, vol. 164, no. 10, pp. 1613–1614, 2007.

[68] A. M. Schäfer and H.-G. Zimmermann, "Recurrent neural networks are universal approximators," *Int. J. Neural Syst.*, vol. 04, no. 17, pp. 253–263, 2007.

[69] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE Trans. Neural Netw.*, vol. 5, no. 2, pp. 157–166, Mar. 1994.

[70] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[71] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov, "Improving neural networks by preventing co-adaptation of feature detectors," *Comput. Sci.*, vol. 3, no. 4, pp. 212–223, 2012.