HHS Public Access

Author manuscript

Mach Learn Med Imaging. Author manuscript; available in PMC 2017 November 03.

Published in final edited form as:

Mach Learn Med Imaging. 2017 September ; 10541: 362–370. doi:10.1007/978-3-319-67389-9_42.

Identifying Autism from Resting-State fMRI Using Long Short-Term Memory Networks

Nicha C. Dvornek[1], **Pamela Ventola**[2], **Kevin A. Pelphrey**[3], and **James S. Duncan**[1,4,5]

[1]Department of Radiology & Biomedical Imaging, Yale School of Medicine, New Haven, CT, USA

[2]Child Study Center, Yale School of Medicine, New Haven, CT, USA

[3]Autism and Neurodevelopmental Disorders Institute, George Washington University and Children's National Medical Center, Washington, DC, USA

[4]Department of Biomedical Engineering, Yale University, New Haven, CT, USA

[5]Department of Electrical Engineering, Yale University, New Haven, CT, USA

Abstract

Functional magnetic resonance imaging (fMRI) has helped characterize the pathophysiology of autism spectrum disorders (ASD) and carries promise for producing objective biomarkers for ASD. Recent work has focused on deriving ASD biomarkers from resting-state functional connectivity measures. However, current efforts that have identified ASD with high accuracy were limited to homogeneous, small datasets, while classification results for heterogeneous, multi-site data have shown much lower accuracy. In this paper, we propose the use of recurrent neural networks with long short-term memory (LSTMs) for classification of individuals with ASD and typical controls directly from the resting-state fMRI time-series. We used the entire large, multi-site Autism Brain Imaging Data Exchange (ABIDE) I dataset for training and testing the LSTM models. Under a cross-validation framework, we achieved classification accuracy of 68.5%, which is 9% higher than previously reported methods that used fMRI data from the whole ABIDE cohort. Finally, we presented interpretation of the trained LSTM weights, which highlight potential functional networks and regions that are known to be implicated in ASD.

1 Introduction

Investigating the pathophysiology of autism spectrum disorders (ASD) with functional magnetic resonance imaging (fMRI) holds promise for identifying objective biomarkers of the neurodevelopmental disorder. Discovering biomarkers for ASD would potentially lead to better understanding the underlying causes of ASD. This would have far-reaching implications, aiding in diagnosis, the design of improved therapies, and monitoring and predicting treatment outcomes.

Recent efforts have focused on investigating ASD biomarkers based on measures of functional connectivity, computed from resting-state fMRI (rsfMRI). Functional connectivity measures are used as predictors for classifying ASD v.s. neurotypical control, using popular learning methods such as support vector machines, random forests, or ridge

regression [13,3,1]. Pairwise connections deemed important for accurate classification are then potential biomarkers for ASD.

While high accuracies have been reported for identifying ASD from rsfMRI, these results were found using small, homogeneous datasets gathered from a single [15] or a few [13] imaging sites and likely do not generalize well to the larger, heterogeneous ASD population. To aid in discovering more generalizeable fndings, the Autism Brain Imaging Data Exchange (ABIDE) gathered neuroimaging and phenotypic data from 1112 subjects across 17 sites for their first publicly shared dataset, ABIDE I [7]. While larger datasets are usually helpful in achieving higher classification accuracy, the heterogeneity of ASD has proved to be a challenge; recent methods which trained on large portions of this diverse dataset have demonstrated much lower classification accuracy [12,9].

We propose a new approach in which we learn the ASD classification directly from the rsfMRI time-series, rather than from precomputed measures of functional connectivity. Since the fMRI data represents dynamic brain activity, we hypothesize that the time-series will carry more useful information than single, static functional connectivity measures. To learn directly from the rsfMRI time-series, we base our approach on Long Short-Term Memory networks (LSTMs), a type of deep neural network designed to handle very long sequence data [10].

In this paper, we investigate the use of LSTMs for identifying individuals with ASD from rsfMRI time-series. To the best of our knowledge, this is the first use of LSTMs for classifying fMRI data. We train and test the developed LSTM models on the entire ABIDE dataset and compare classification accuracy against previous studies that classified the ABIDE subjects from rsfMRI. Finally, we interpret the best model, identifying brain regions important for distinguishing ASD from typical controls. We hypothesize the learned LSTM weights will encode potential networks that have previously been implicated in ASD.

2 Methods

2.1 Network Architecture

LSTMs are a special type of recurrent neural network, composed of repeated cells that receive input from the previous cell as well as the data input x_t for the current timestep t. Each LSTM cell contains a cell state c_t and hidden state h_t, which are modulated by 4 neural network layers that control the flow of information into and out of cell memory. The equations governing an LSTM are:

$$i_t=\sigma(W_i x_t+U_i h_{t-1}+b_i) \quad (1)$$

$$f_t=\sigma(W_f x_t+U_f h_{t-1}+b_f) \quad (2)$$

$$\tilde{c}_t = \tanh\left(W_c x_t + U_c h_{t-1} + b_c\right) \quad (3)$$

$$c_t = i_t * \tilde{c}_t + f_t * c_{t-1} \quad (4)$$

$$o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o) \quad (5)$$

$$h_t = o_t * \tanh\left(c_t\right) \quad (6)$$

W matrices contain weights applied to the current input, U matrices represent weights applied to the previous hidden state, b vectors are biases for each layer, and σ is the sigmoid function. The input gate i_t (eq. (1)) decides what information from the current estimated cell state is updated. The forget gate f_t (eq. (2)) controls what information from the previous cell state is kept. Next, the estimated current cell state (eq. (3)) and previous cell state are combined with restrictions from the input and forget gates, respectively, to update the cell state (eq. (4)). Finally, cell state information is filtered with the output gate o_t (eq. (5)) to update the hidden state (eq. (6)), which is the output of the LSTM cell.

We propose an LSTM architecture which takes the rsfMRI time-series as input x and connects the output of each repeating cell to a dense layer with a single node (Fig. 1). This gives the signal at every time point a more direct say in how to classify the signal, compared to the traditional approach of looking at the final output after the whole sequence is analyzed (h_T). We believe this will be more robust to the noisy fMRI data. The outputs of the single nodes are then averaged across the entire sequence and fed to a sigmoid activation function to produce the probability of an ASD label. For regularization, during training, we apply dropout to the LSTM weights as described in Gal et al. [8] and add a standard dropout layer between the single-node dense layer and pooling layer. In the following, we also investigate a two-layer LSTM model, in which the hidden states output from the first layer are used as the input sequence into a second LSTM layer, after which the architecture is the same as in the single-layer model.

2.2 Dataset and Preprocessing

The ABIDE I dataset includes rsfMRI for 539 individuals with ASD and 573 typical controls from 17 international sites. To further enhance the data-sharing effort, the Preprocessed Connectomes Project released preprocessed ABIDE data using a number of popular pipelines and several calculated derivatives [5]. We chose the data processed through the Connectome Computation System, without global signal regression but with band-pass filtering. See the ABIDE Preprocessed website [14] for more preprocessing details.

The preprocessed ABIDE data includes extracted mean time-series from regions of interest defined by several atlases. Here, we utilized the mean time-series from the Craddock 200 atlas [6], which was provided for 1100 subjects. Each time course was normalized to represent percent change from the average signal for that region of interest. Further, since different sites used different acquisition protocols, we resampled each time-series using an interval of 2 s to bring the data to the same time scale. The preprocessed mean time courses from the 200 atlas regions were used as input x into the LSTM.

2.3 Data Augmentation

While the ABIDE dataset has a large number of subjects for a neuroimaging database, training neural networks often requires many more samples to prevent overfitting. Furthermore, the ABIDE time courses have different lengths depending on the site. Thus, we propose cropping the input time courses to a fixed sequence length for all subjects and augmenting the number of inputs for each subject to make the most use of the full time-series. Based on the length of the shortest time-series, we chose a sequence length of $T = 90$, which represents 3 minutes of imaging. For each subject, we crop 10 sequences of length T from the time-series, randomly varying the starting time of each cropped sequence. This augmented our dataset by a factor of 10 to a total of 11,000 input sequences.

3 Experiments

3.1 Experimental Methods

The LSTM training and testing were performed using Keras [4]. Models were trained using the binary cross-entropy loss function and the Adadelta optimizer with the default parameter values. The dropout rate during training was fixed to 0.5. Models were initialized using default Keras settings.

We explored the impact of parameters and variations of the proposed architecture as well as training conditions. We tested not augmenting data, varying the number of hidden nodes (8, 16, 32, or 64) in the LSTM, and removing dropout. We also tested variations on the base network: connecting only the final LSTM cell's output (h_T) to a single dense node, and stacking LSTM layers.

To validate the performance of the LSTMs, we used stratified 10-fold cross-validation, such that the proportion of subjects from each site was approximately the same in all folds. For each fold, data was split into 85% for training, 5% for validation, and 10% for testing. When using the augmented dataset, all sequences belonging to the same subject appeared in either training, validation, or testing. Training was stopped when the validation loss had not decreased in 20 epochs or when 300 epochs had been executed. The trained model was then tested on the left-out test data. Accuracy was assessed based on classification of each input sequence ("sequence accuracy") as well as classification of each subject using the average score of all input sequences from a subject ("subject accuracy"). Significance tests were performed using two-tailed, paired t-tests with $\alpha = 0.05$. We compared our approach to previous studies that trained on ABIDE rsfMRI. To better compare against these other studies which used different subsets from the ABIDE cohort, we computed the difference

between the model's accuracy and the accuracy of assigning classifications by chance within the tested dataset.

Finally, we considered interpretation of the LSTM model which resulted in the highest classification accuracy. Entries in the LSTM weight matrix $W_l(n, r)$ with large magnitude, regardless of sign, should denote that atlas region r has a strong influence on LSTM node n for layer l. We investigated regions that were considered important for each layer and for each node. First, for each layer l, we averaged the absolute values of the weights across all nodes. We then created a binary mask of important regions, defined as those regions with weight magnitudes greater than 2 standard deviations above the mean for the layer. The mask of important regions was then input into Neurosynth, a meta-analysis tool that compares a brain map to a database of approximately 10,000 fMRI studies and assigns correlations between the map and almost 3000 descriptors [16]. Similarly, for each node n, we defined important regions as those with weights greater than 2 standard deviations away from the mean in the node for each layer, aggregated the important regions across all layers into a single binary mask per node, and input the mask into Neurosynth for interpretation.

3.2 Classification Accuracy

Results from previous studies and from our LSTM models are compared in Table 1. The highest accuracy was reported by Plitt et al. [13]; however, only a small, very homogeneous subset (16%) of the ABIDE dataset was used. Chen et al. [3] showed a large improvement compared to chance, but also used a very pruned subset of the data with a single training/validation split. The two studies with the largest datasets [12,9] demonstrated lower accuracy compared to our LSTM model trained on only a single input sequence from each subject. All other LSTM models, which used the augmented dataset, performed even better. Subject accuracy was higher than sequence accuracy for all models. Among the single layer models, the highest subject accuracy was achieved for the LSTM with 32 hidden nodes (68.5%). Compared to the most competitive result using the majority of the ABIDE cohort [1], the difference between our accuracy and chance is over 3% higher, while our dataset contained more challenging, heterogeneous data with 25% more subjects. Furthermore, compared to the study with the closest number of subjects to ours [9], our model improved accuracy compared to chance by 9%. Thus, our method would likely generalize best to new data.

All tested variations of the proposed network resulted in degraded accuracy. Removing dropout regularization reduced accuracy by almost 7%. Using only the final hidden state of the LSTM sequence decreased accuracy by 4%. Finally, creating a deeper model with two stacked LSTM layers was not helpful.

3.3 Model Interpretation

We investigated the learned weights of the best model, LSTM32. Table 2 shows, for each layer, the top associated Neurosynth anatomical and functional descriptors. The input and forget gates, which modulate the cell state information, are heavily influenced by regions associated with language and communication; impairment of these functions are primary symptoms of ASD. Functional terms associated with influential regions for the current estimated cell state are important for supporting social interactions, which are difficult for

individuals with autism. The output gate is most influenced by regions associated with self-referential processing, which has been shown to be impaired in autistic individuals [11].

Finally, we explored potential brain networks encoded by each LSTM cell node. The important regions for the four nodes with greatest influence (i.e., with the largest weight magnitudes from the dense layer of the neural network) are shown in Fig. 2. These region groupings highlight neurocognitive functions affected by ASD; e.g., social reward is diminished, face processing and communication skills are impaired, and theory of mind, a leading hypothesis for social impairment in autism, is lacking in autistic individuals [2].

4 Conclusions

We have presented a method for identifying individuals with ASD from rsfMRI using LSTMs. Our model demonstrated the highest classification accuracy compared to other methods which utilized the majority of the ABIDE cohort. We contend it is important to succeed on large heterogeneous datasets, since ASD covers a wide spectrum, and image quality can be difficult to control for individuals with autism and young children. Data augmentation and choice of network structure were crucial in training an accurate model. More in depth tuning of hyperparameters, training on other parcellations, including demographic information, and combining models would likely lead to higher classification accuracy.

The learned LSTM input weights had meaningful interpretation; anatomical regions with high influence on the network have previously been shown to be abnormal in ASD. Further, meta-analysis highlighted neurocognitive processes that are affected in individuals with ASD. Inspection of network activations and hidden state weights could lead to greater insights into the mechanism of ASD.

Acknowledgments

This work was supported in part by T32 MH18268 and R01 NS035193.

References

1. Abraham A, Milham MP, Martino AD, Craddock RC, Samaras D, Thirion B, Varoquaux G. Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. Neuroimage. 2017; 147:736–745. [PubMed: 27865923]
2. Baron-Cohen S, Abraham A, Leslie M, Frith U. Does the autistic child have a “theory of mind”. Cognition. 1985
3. Chen CP, Keown CL, Jahedi A, Nair A, Pflieger ME, Bailey BA, Müller RA. Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism. Neuroimage: Clinical. 2015
4. Chollet, F. Keras. 2015. https://github.com/fchollet/keras
5. Craddock C, Benhajali Y, Chu C, Chouinard F, Evans A, Jakab A, Khundrakpam BS, Lewis JD, Li Q, Milham M, Yan C, Bellec P. The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. Neuroinformatics. 2013
6. Craddock RC, James GA, Holtzheimer PE, Hu XP, Mayberg HS. A whole brain fmri atlas generated via spatially constrained spectral clustering, human brain mapping. Human Brain Mapping. 2012
7. Di Martino A, Yan CG, Li Q, Denio E, Castellanos FX, Alaerts K, Anderson JS, Assaf M, Bookheimer SY, Dapretto M, Deen B, Delmonte S, Dinstein I, Ertl-Wagner B, Fair DA, Gallagher

L, Kennedy DP, Keown CL, Keysers C, Lainhart JE, Lord C, Luna B, Menon V, Minshew NJ, Monk CS, Mueller S, Müller RA, Nebel MB, Nigg JT, O'Hearn K, Pelphrey KA, Peltier SJ, Rudie JD, Sunaert S, Thioux M, Tyszka JM, Uddin LQ, Verhoeven JS, Wenderoth N, Wiggins JL, Mostofsky SH, Milham MP. The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular Psychiatry. 2014

8. Gal Y, Ghahramani Z. A theoretically grounded application of dropout in recurrent neural networks. NIPS. 2016
9. Ghiassian S, Greiner R, Jin P, Brown MRG. Using functional or structural magnetic resonance images and personal characteristic data to identify adhd and autism. PLOS One. 2016
10. Hochreiter S, Schmidhuber J. Long short-term memory. Neural Computation. 1997
11. Lombardo MV, Barnes JL, Wheelwright SJ, Baron-Cohen S. Self-referential cognition and empathy in autism. PLoS One. 2007
12. Nielsen JA, Zielinski BA, Fletcher PT, Alexander AL, Lange N, Bigler ED, Lainhart JE, Anderson JS. Multisite functional connectivity mri classification of autism: Abide results. Front. Hum. Neurosci. 2013
13. Plitt M, Barnes KA, Martin A. Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. Neuroimage: Clinical. 2015
14. Preprocessed Connectomes Project. ABIDE Preprocessed. http://preprocessed-connectomes-project.org/abide/
15. Uddin LQ, Supekar K, Lynch CJ, Khouzam A, Phillips J, Feinstein C, Menon V. Salience network-based classification and prediction of symptom severity in children with autism. JAMA Psychiatry. 2014
16. Yarkoni, T., Poldrack, RA., Nichols, TE., Van Essen, DC., Wager, TD. Large-scale automated synthesis of human functional neuroimaging data. Nature Methods. 2011. www.neurosynth.org

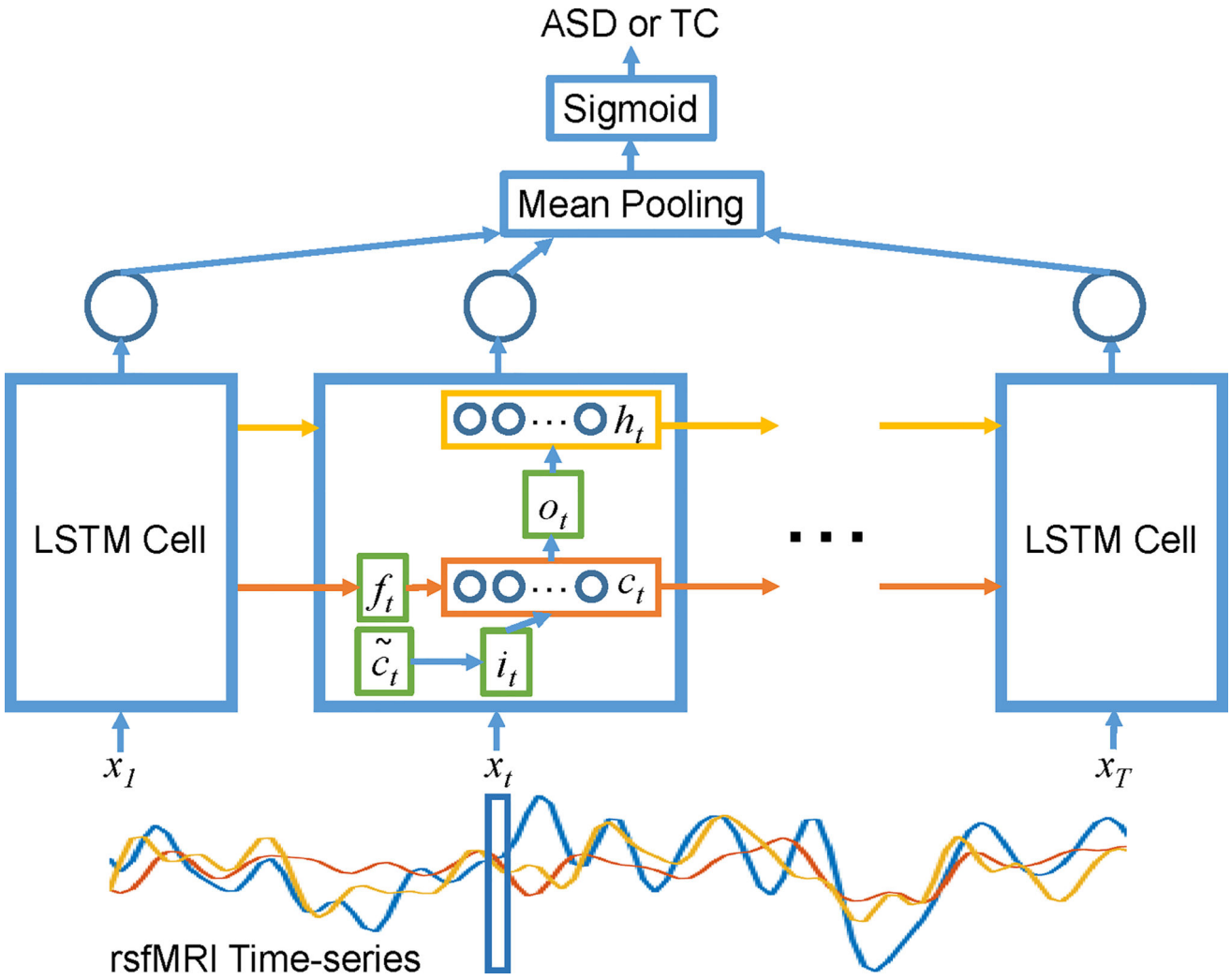

Fig. 1.

Diagram of the LSTM network for classifying ASD from rsfMRI. The recurrent neural network is visualized "unrolled" for clarity. Each green square is a neural network layer that takes x_t and h_{t-1} as inputs.

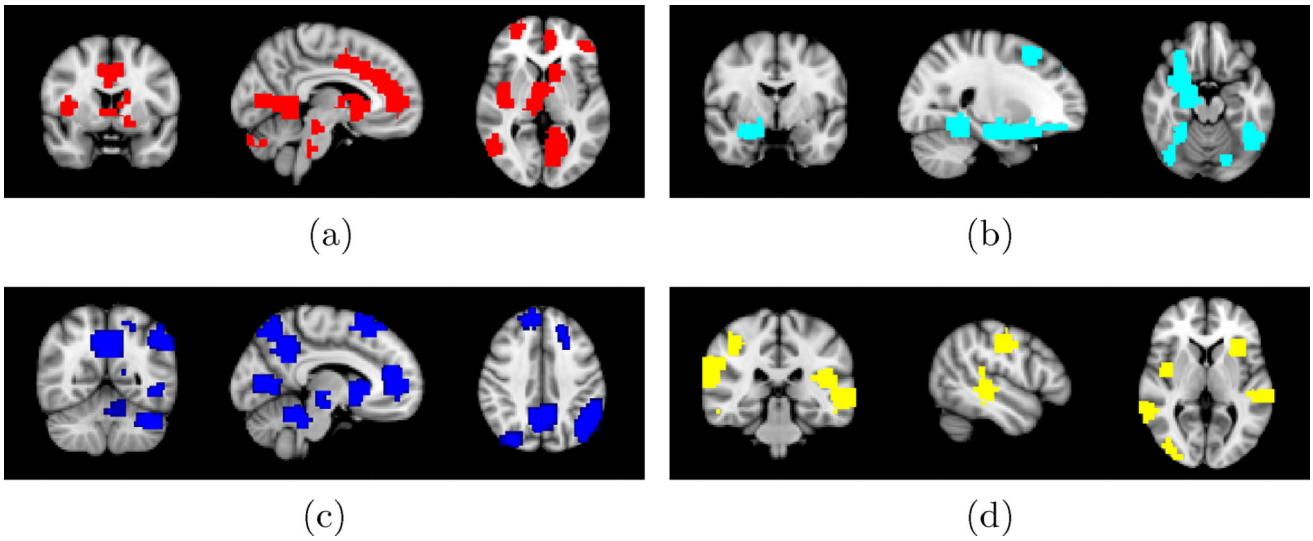

Fig. 2.

Influential brain regions for the 4 most important LSTM nodes. Top associated Neurosynth functional features include: (a) Pain, reward, anticipation, incentive. (b) Faces, objects, word form, emotional, visual. (c) Default mode, reward, listening, mental states, theory of mind. (d) Listening, sounds, theory of mind, social, speech perception.

Mach Learn Med Imaging. Author manuscript; available in PMC 2017 November 03.

Table 1

Autism classification results from methods that trained on rsfMRI from the ABIDE dataset.

Classification Method	Validation Method	Number of Subjects	Mean (SD) Sequence Accuracy (%)	Difference from Chance (%)	Mean (SD) Subject Accuracy (%)	Difference from Chance (%)
Plitt et al. [13]	CV10	178	-	-	69.7	19.7
Chen et al. [3]	Train/Val	252	-	-	66	16
Abraham et al. [1]	CV10	871	-	-	66.9 (2.7)	13.2
Nielsen et al. [12]	LOO	964	-	-	60.0	6.4
Ghiassian et al. [9]	Train/Val	1111	-	-	59.2	7.6
LSTM8	CV10	1100	65.6 (4.1)	13.7	66.7 (5.3)	14.8
LSTM16	CV10	1100	65.3 (4.8)	13.3	66.8 (5.4)	14.9
LSTM32	CV10	1100	**66.8 (4.5)**	**14.9**	**68.5 (5.5)**†	**16.6**
LSTM64	CV10	1100	65.8 (3.8)	13.9	67.5 (4.4)†	15.5
LSTM32_NoAug	CV10	1100	-	-	61.4 (4.5)*	9.5
LSTM32_NoDrop	CV10	1100	59.7 (2.3)	7.7	61.8 (4.0)*	9.9
LSTM32_Last	CV10	1100	62.2 (3.3)	10.3	64.5 (4.5)†	12.5
LSTM32×2	CV10	1100	66.3 (4.2)	14.4	67.5 (5.0)†	15.5

Classification accuracies of best LSTM-based model are in bold. LSTM# = LSTM with # hidden nodes, NoAug = No data augmentation, NoDrop = No dropout regularization, Last = Pass only the last hidden state to dense node, LSTM#×2 = Two-layer LSTM with # hidden nodes, Train/Val = Single training and validation set, CV10 = 10-fold cross-validation, LOO = Leave-one-out cross-validation, SD = Standard deviation.

†Significant difference between sequence and subject accuracies.

*Significant difference compared to best model LSTM32.

Table 2

Top Neurosynth anatomical and functional terms associated with the mask created from the brain regions with the greatest weight magnitudes for each layer.

Layer	Anatomical Terms	Functional Terms
Input	Superior Temporal Sulcus, Middle Temporal Gyrus, Planum Temporale	Sentence, Comprehension, Linguistic, Audiovisual, Language
Forget	Inferior Frontal Gyrus, Temporal Pole, Planum Temporale	Sentence, Verb, Nouns, Semantically, Sentence Comprehension
Cell	Midbrain, Thalamus, Superior Temporal Sulcus	Reward, Speaker, Voice, Audiovisual, Speech
Output	Hypothalamus, Inferior Parietal Lobe, Medial Prefrontal Cortex	Self, Sexual, Referential, Memory Retrieval, Regulation