

COMBINING PHENOTYPIC AND RESTING-STATE FMRI DATA FOR AUTISM CLASSIFICATION WITH RECURRENT NEURAL NETWORKS

Nicha C. Dvornek^{*}, Pamela Ventola[†], and James S. Duncan^{**‡}

^{*}Radiology & Biomedical Imaging and [†]Child Study Center, Yale School of Medicine, New Haven, CT

^{*}Biomedical Engineering and [‡]Electrical Engineering, Yale University, New Haven, CT

ABSTRACT

Accurate identification of autism spectrum disorder (ASD) from resting-state functional magnetic resonance imaging (rsfMRI) is a challenging task due in large part to the heterogeneity of ASD. Recent work has shown better classification accuracy using a recurrent neural network with rsfMRI time-series as inputs. However, phenotypic features, which are often available and likely carry predictive information, are excluded from the model, and combining such data with rsfMRI into the recurrent neural network is not a straightforward task. In this paper, we present several methodologies for incorporating phenotypic data with rsfMRI into a single deep learning framework for classifying ASD. We test the proposed architectures using a cross-validation framework on the large, heterogeneous first cohort from the Autism Brain Imaging Data Exchange. Our best model achieved an accuracy of 70.1%, outperforming prior work.

Index Terms— Autism Spectrum Disorders, Classification, Neural Networks, Resting-state fMRI, Phenotypic Data

1. INTRODUCTION

Resting-state functional magnetic resonance imaging (rsfMRI) has been an important tool for investigating the neuropathophysiology of autism spectrum disorders (ASD), revealing changes in brain activity between individuals with ASD compared to neurotypical controls [1]. Finding the differences in brain activity which accurately distinguish ASD and neurotypical individuals will better characterize the underlying causes of ASD, leading to improved diagnosis and treatment.

Traditional approaches to learn the classification of ASD and neurotypical subjects from rsfMRI use resting-state connectivity as input features into learning algorithms such as logistic regression and support vector machines [2, 3]. However, recent work has shown improved classification accuracy with a deep learning approach using recurrent neural networks [4]. The method builds on the long short-term memory (LSTM) [5] architecture, which allows the rsfMRI time-series to be directly used as input.

This work was supported by NIH grants R01 NS035193 and T32 MH18268 (N.C.D.).

Although accuracy has improved, the ASD classification problem is still very challenging, largely due to the heterogeneity of ASD. With the neurodevelopmental nature of the disorder, one would expect that including phenotypic information (e.g., sex, age) would improve classification accuracy. Traditional learning algorithms can simply include such phenotypic features as additional input variables [6], but in deep learning, networks often need to be designed for the specific multimodal task. As such, combining phenotypic information into a neural network built for rsfMRI is not straightforward.

In this work, we explore methods of incorporating phenotypic data and rsfMRI data into a single neural network model for classification of individuals with ASD and neurotypical controls. We test the proposed network architectures using the first dataset from the Autism Brain Imaging Data Exchange (ABIDE), which includes phenotypic and neuroimaging data for a very large, heterogeneous sample of autistic and control subjects [7]. The models are validated using 10-fold cross-validation and compared against other recent work which tested on the majority of the ABIDE cohort.

2. METHODS

2.1. LSTM for Classification from rsfMRI

Recent work has shown promising results classifying ASD directly from rsfMRI time-series inputs using an LSTM-based architecture [4], shown in Fig. 1a. The rsfMRI time-series are used as input f to the network. The output from the LSTM cell at each timestep is fed into a dense layer with a single node (i.e., the weights for the node are shared across time). The outputs from the single nodes are then averaged, resulting in a single value that is input into the sigmoid activation function to give the probability of belonging to the ASD class. During training, dropout is applied on the LSTM weights and after the dense layer. We use this network as our baseline model for processing rsfMRI data for classification.

2.2. Incorporating Phenotypic Information

We propose several methods for combining the phenotypic and fMRI data into a single network. They are listed here ap-

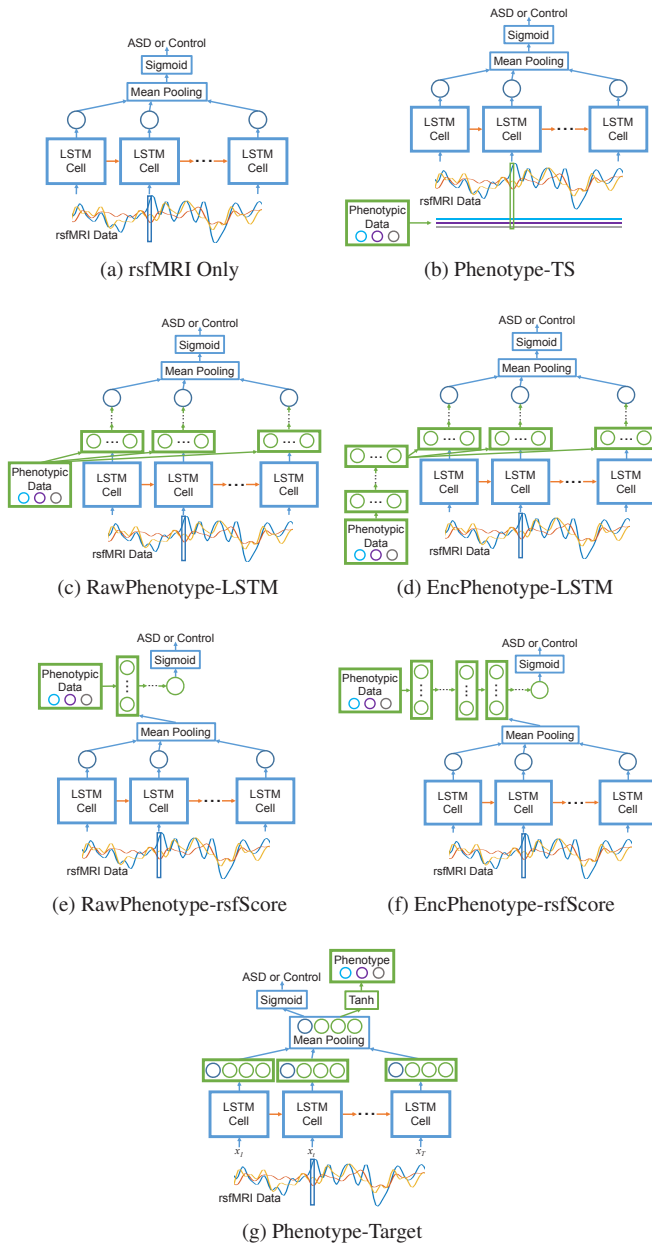


Fig. 1: LSTM-based models for combining phenotypic and rsfMRI information. Green denotes changed architecture from rsfMRI-only model. (a) Baseline model using only rsfMRI input. (b) Set phenotypic data as part of time-series input. (c) Combine phenotypic data with LSTM outputs. (d) Encode phenotypic data and combine with LSTM outputs. (e) Combine phenotypic data with score from baseline model. (f) Encode phenotypic data and combine with score from baseline model. (g) Use phenotypic data as model outputs.

proximately in order of the network depth at which the phenotypic data is incorporated with the rsfMRI information.

1. Input phenotypic data and rsfMRI time-series directly

into LSTM model (Phenotype-TS, Fig. 1b). The phenotypic data is first repeated along the time axis to have the same dimension as the rsfMRI time-series. The phenotypic data is then concatenated to the time-series data along the feature dimension and is input directly into the LSTM. This approach allows the phenotypic data to interact directly with the rsfMRI data within the LSTM cell.

2. **Combine phenotypic data directly with LSTM outputs from rsfMRI input (RawPhenotype-LSTM, Fig. 1c).** The phenotypic data is again replicated along the time axis to have the same dimension as the rsfMRI time-series. The phenotypic data and LSTM outputs are then input into D dense layers, with one node in the last layer. This model would allow the phenotypic data to interact with important brain networks for classification [4].
3. **Encode phenotypic features, then combine with LSTM outputs from rsfMRI input (EncPhenotype-LSTM, Fig. 1d).** First, the phenotypic data is encoded by one or more dense layers. The new representation of the phenotypic data is then repeated to have the same dimension in time as the rsfMRI time-series. The transformed phenotypic information and LSTM outputs are then input into D dense layers, with one node in the last layer.
4. **Combine phenotypic data directly with score from rsfMRI input (RawPhenotype-rsfScore, Fig. 1e).** Phenotypic data and the final output of the LSTM model are input into D dense layers, with a single node in the last layer. This model gives the raw phenotypic data the most direct influence on the classification score.
5. **Encode phenotypic features, then combine with score from rsfMRI input (EncPhenotype-rsfScore, Fig. 1f).** Phenotypic information is first encoded using one or more dense layers. The encoded phenotypes and final output of the LSTM model are input into D dense layers, ending with a layer with a single node. This model has the most independent processing before combining modalities.
6. **Set phenotypic data as target variables (Phenotype-Target, Fig. 1g).** Auxiliary loss functions have improved learning of the target objective [8]. Instead of using the phenotypic information as input, the phenotypic data p are considered auxiliary target values estimated along with the ASD classification label l in a multi-task learning setting, i.e., outputs are $y = [l, p]$. Since autism classification is the primary goal, we use the following loss function L :

$$L(y, \hat{y}) = L_{ASD}(l, \hat{l}) + \lambda L_{phenotype}(p, \hat{p}), \quad (1)$$

where \hat{y} are the estimated output values, L_{ASD} is the loss for the ASD classification label, $L_{phenotype}$ is the loss for phenotypic features, and λ controls the contribution of the phenotypic loss term. While the phenotypic data does not interact directly with the rsfMRI data, this approach trains the LSTM model to be consistent with the phenotypic data.

3. EXPERIMENTS

3.1. Data and Preprocessing

We performed experiments on data from the ABIDE I cohort [7]. We used data released by the Preprocessed Connectomes Project, generated from the Connectome Computation System pipeline, without global signal regression, and with band-pass filtering [9]. This resulted in 529 autism subjects and 571 typical controls from 17 different imaging sites.

The rsfMRI input for the neural networks was derived from mean time-series extracted from regions of interest defined by the Craddock 200 atlas [10]. Each time-series was normalized to percent change from mean signal and resampled at 2 s intervals to bring the data to the same scale. Furthermore, since different sites used varying scan lengths, we randomly cropped the time-series data to length $T = 90$. We performed 10 random crops per subject, augmenting the dataset to $N = 11000$ samples. Each sample's fMRI input f_i for the neural network is then a matrix of size 90×200 .

For the phenotypic inputs, we included data that was complete or almost complete for all subjects: age, sex, handedness, full IQ, and eye status during the fMRI scan (opened or closed according to imaging protocol). Note we did not use imaging site as a variable so that the model would be applicable to new data from other locations. Handedness was coded as an ordinal variable with 3 levels $\{-1, 0, 1\}$ representing left dominant, ambidextrous, and right dominant. Subjects with missing handedness data were assigned to the right dominant category since most people are right-handed. Subjects with missing full IQ score were given a value of 100, which by definition is considered the average IQ. The data for each phenotype were normalized to lie in $[-1, 1]$. Each sample's phenotypic data p_i is thus a vector with dimension $P = 5$.

3.2. Methods Implementation and Evaluation

The neural network models were implemented using Keras [11] with the binary cross-entropy loss function, except the auxiliary targets model which was implemented with mean squared error since many phenotypes are continuous variables. The adadelat optimizer and default initializations were applied. The LSTM output dimension was fixed to 32, dropout rate during training was set to 0.5, and dense layers (except for the final output) applied the tanh activation.

In the experiments, we tested 1) the baseline model using only rsfMRI (Fig. 1a), 2) a multilayer perceptron model using only phenotypic data, and 3) the models proposed in Section 2.2. For the phenotype-only model, we used 1 hidden layer with P nodes. For models with phenotypic data encoding, we implemented 1 dense layer with P nodes (EncA), and 2 dense layers where the first layer has P nodes and the second layer has 1 node (EncB). For merging of phenotypic and rsfMRI information, we tested $D = 1$ and $D = 2$ with number of

nodes in the hidden layer equal to the number of inputs. The auxiliary targets model was trained with $\lambda = 0.1$.

We evaluated performance of the different neural network models using 10-fold cross-validation, with stratified sampling based on imaging site. The data for each fold was split using 85% for training, 5% for validation, and 10% for testing, with all samples generated from a subject assigned to one partition. Training was stopped when validation loss had not decreased in 10 epochs. Predictions for test data using the final trained model were used to assess model performance.

We assessed performance using "sample accuracy" and "subject accuracy," the accuracy of classifying each sample (of the augmented dataset) and each subject, respectively. Subject label was determined from the average score of all samples from the subject. We computed the mean and standard deviation (SD) of the accuracy across cross-validation folds. Since the same data splits were used and the number of observed results (10) is small for each model, we tested for a significant improvement in results using a paired, one-tailed t-test with $p < 0.1$. Finally, to better compare against previous works which use different subsets of the ABIDE cohort, we computed the difference between the nominal accuracy and chance level (percent of more common class in the dataset).

3.3. Results and Discussion

Table 1 summarizes the performance of the different classification approaches. The first few rows highlight previous work that tested on the majority of the ABIDE cohort and used both rsfMRI and phenotypic data. Parisot et al. [12] produced the highest accuracy result; however, they used the smallest number of subjects, pruning difficult data, and they depended on site as a phenotypic feature, which makes the model not generalizable to data obtained from new imaging sites.

Results for the implemented neural networks are shown in the bottom portion of Table 1. Subject accuracy was better than sample accuracy for almost all models, signifying the importance of data augmentation. The baseline rsfMRI-only model already performs similarly to the best published model with phenotypic information [12] (see second to last column). Using only phenotypic data results in an 8.5% increase over chance level, suggesting such information is helpful in ASD classification. Generally, models that first encoded the phenotypes into a single score (EncB) resulted in poorer accuracy, while combining the raw phenotypic data directly with baseline model outputs performed better. The best model combined the phenotypic data with the final score after rsfMRI processing using 1 hidden layer. This increased accuracy by 2.2% over the rsfMRI-only model. Further, the accuracy is 2.4% higher than the best prior work, without including imaging site which improves the generalizability of our model. While nominally the increase in accuracy appears modest, the ASD classification problem is very difficult, and small improvements will have larger effect as dataset sizes increase.

Classification Method	Number of Subjects	Mean (SD) Sample Accuracy (%)	Difference from Chance (%)	Mean (SD) Subject Accuracy (%)	Difference from Chance (%)	Phenotypic Data
Parisot et al. [12]	871	-	-	69.5	15.8	Sex, Site
Nielsen et al. [2]	964	-	-	60.0	6.4	Age, Sex, Handedness
Ghiassian et al. [6]	1111	-	-	65.0	13.4	Age, Sex, Handedness, Full IQ, Performance IQ, Verbal IQ, Site, Eye status
rsfMRI-only	1100	66.7 (3.6)	14.8	67.9 (4.3) [†]	16.0	None
Phenotype-only				60.4 (3.6)	8.5	Age, Sex, Handedness, Full IQ, Eye status
Phenotype-TS		66.1 (4.0)	14.2	67.0 (3.5)	15.1	
RawPhenotype-LSTM, $D = 1$		67.4 (3.7)	15.5	68.2 (4.1) [†]	16.3	
RawPhenotype-LSTM, $D = 2$		66.5 (2.5)	14.5	67.5 (3.8) [†]	15.6	
EncAPhenotype-LSTM, $D = 1$		66.0 (2.9)	14.1	67.8 (4.3) [†]	15.9	
EncAPhenotype-LSTM, $D = 2$		66.7 (3.4)	14.8	67.5 (4.4)	15.5	
EncBPhenotype-LSTM, $D = 1$		66.5 (4.0)	14.5	68.1 (3.3) [†]	16.2	
EncBPhenotype-LSTM, $D = 2$		66.1 (3.3)	14.1	65.8 (3.8)	13.9	
RawPhenotype-rsfScore, $D = 1$		67.9 (2.9)*	16.0	69.6 (1.9) [†]	17.7	
RawPhenotype-rsfScore, $D = 2$		67.9 (2.7)	16.0	70.1 (3.2)^{†*}	18.2	
EncAPhenotype-rsfScore, $D = 1$		66.8 (3.6)	14.9	68.4 (4.7) [†]	16.5	
EncAPhenotype-rsfScore, $D = 2$		67.3 (3.9)	15.3	68.2 (4.6) [†]	16.3	
EncBPhenotype-rsfScore, $D = 1$		66.7 (4.3)	14.8	66.5 (4.9)	14.6	
EncBPhenotype-rsfScore, $D = 2$		66.2 (3.0)	14.3	67.4 (4.6) [†]	15.5	
Phenotype-Target		65.7 (3.5)	13.8	67.2 (4.0) [†]	15.3	

Table 1: ASD classification results from methods that trained on rsfMRI and phenotypic features from the ABIDE dataset. See the text (Section 2.2 and 3.2) for method abbreviations. Best model is in bold. [†] Subject accuracy is significantly better than sample accuracy. * Accuracy is significantly better compared to rsfMRI-only model

4. CONCLUSIONS

In this work, we explored several different neural network approaches for incorporating phenotypic data with rsfMRI to classify ASD and neurotypical subjects. Our best model combined the phenotypic data with the rsfMRI score, resulting in a classification accuracy of 70.1% (18.2% higher than naive labeling) on the ABIDE dataset, outperforming recent work.

Future directions include modifying the architecture to incorporate structural MRI. Also, including behavioral measures may further improve classification accuracy. Finally, the effect of including phenotypic features on the learned important functional networks for classification will be explored.

5. REFERENCES

- [1] V. L. Cherkassky et al., “Functional connectivity in a baseline resting-state network in autism,” *Neuroreport*, vol. 17, no. 16, pp. 1687–1690, 2006.
- [2] J. A. Nielsen et al., “Multisite functional connectivity mri classification of autism: Abide results,” *Front. Hum. Neurosci.*, vol. 7, pp. 599, 2013.
- [3] A. Abraham et al., “Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example,” *Neuroimage*, vol. 147, pp. 736–745, 2017.
- [4] N. C. Dvornek et al., “Identifying autism from resting-state fmri using long short-term memory networks,” in *MLMI 2017*, LNCS 10541, pp. 362–370.
- [5] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [6] S. Ghiassian et al., “Using functional or structural magnetic resonance images and personal characteristic data to identify adhd and autism,” *PLOS One*, 2016.
- [7] A. Di Martino et al., “The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism,” *Mol. Psychiatry*, 2014.
- [8] C. Szegedy et al., “Going deeper with convolutions,” in *CVPR 2015*, 2015.
- [9] C. Craddock et al., “The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives,” in *Neuroinformatics*, 2013.
- [10] R. C. Craddock et al., “A whole brain fmri atlas generated via spatially constrained spectral clustering, human brain mapping,” *Hum. Brain Mapp.*, 2012.
- [11] F. Chollet, “Keras,” <https://github.com/fchollet/keras>, 2015.
- [12] S. Parisot et al., “Spectral graph convolutions for population-based disease prediction,” in *MICCAI 2017*, 2017, LNCS 10435, Part III, pp. 177–185.