Assessing Autism Spectrum Disorders by Deep Learning Using MRI

Laurynas Kalesinskas*, Rohan Paul*, and Nielson Weng†‡

*Department of Biomedical Informatics, Stanford University; †Department of Medicine, Stanford University School of Medicine; ‡Department of Chemistry, Stanford University
Email: {lkalesin, ropaul, nweng}@stanford.edu

INTRODUCTION

Autism spectrum disorder (ASD) is a group of developmental disabilities characterized by impaired social communication and interaction accompanied by restricted, repetitive patterns of behavior, interests, or activities (1). It was once estimated to be a rare disorder affecting fewer than 1 in 1000 children, but recent studies have estimated the prevalence to be as much as 1 in 68 (2). Even though our understanding and clinical characterization of these disorders have progressed immensely since it was first described in 1943, the fundamental molecular pathways involved in ASD are still largely unknown (3). Consequently, the diagnostic gold standard remains as clinical diagnosis what is based on behavioral assessment.

This subjective assessment in the clinical diagnosis of ASD and the heterogeneity of ASD complicates our progress in understanding of its biology and the development of treatments. Behavioral diagnosis may not be sensitive enough to appropriately classify various disorders within ASD, and it may not be sensitive enough to identify potential initial-phase therapy and treatment that only improve the symptoms incrementally during only stage of development. An objective diagnostic tool based on physiologic changes is still lacking, and a robust predictive model based on pathophysiological findings provides a solution.

In additional to the classic forms of ASD, many syndromes and genetic disorders cause ASD or similar neurological disabilities. For examples, patients with genetic disorders such as phenylketonuria (4) and mutations in BCKDK (5) suffer from a high risk of developing ASD if they consume diets high in metabolites that they cannot break down. Furthermore, some children develops normally but regress to fail to reach developmental milestones later on in life (6). All of these non-classical presentation of ASD create a demand for a predictive model that can detect the regression toward ASK.

The spatial and temporal resolution of functional magnetic resonance imaging have proven to be useful in providing physical evidence of physiologic differences in people with ASD compared to the general population and mechanistic insight to the pathophysiology of ASD (7). Therefore, we proposed to use a deep learning model to classify autism spectrum disorder using fMRI images. Our work will pave the way for a more robust, objective diagnostic methodology that is based on pathophysiology of the disorders. In addition, we may also uncover features that provide mechanistic hypothesis to the pathogenesis in terms of brain regions, neurological circuits, or cellular pathways.

MATERIALS AND METHODS

Data Source

ABIDE (Autism Brain Imaging Data Exchange) is a publicly available dataset with 1114 subjects. It contains data from 521 patients (ASD positive) and 593 controls (ASD negative) with ages ranging from

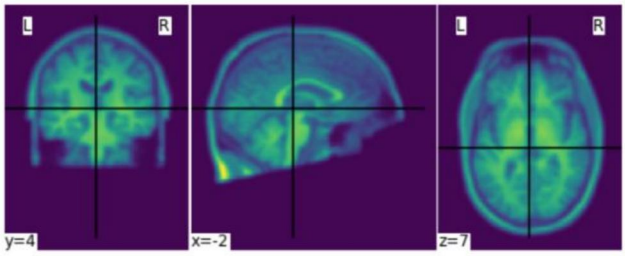

Figure 1. Smoothed MRI image shown from three different planes for a single subject.

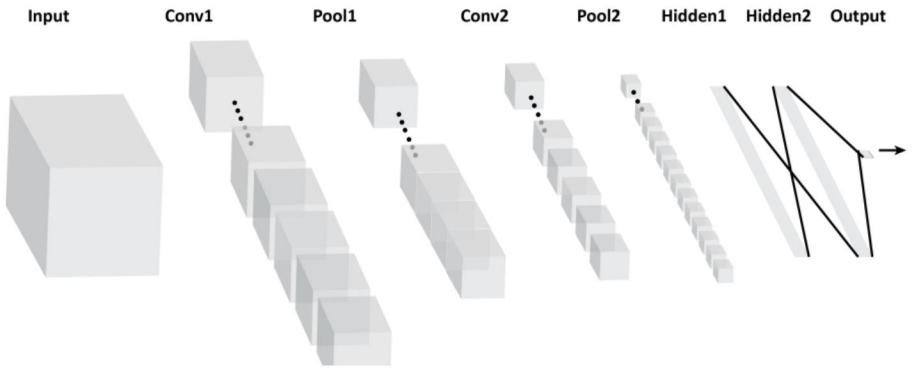

Figure 2. Architecture of the 3D CNN model for classification of Autism Spectrum Disorders.

5 - 64 years. It contains MRI and resting state fMRI data for most examples.

ABIDE is an amalgamation of data from 19 research labs across the world, as such, the dimensions of data are not entirely consistent between labs due to the difference in equipment used. Accurate labeling is available categorizing subjects into ASD and non-ASD categories. Further information such as demographic data, "handedness" (left or right), medications etc. are also available.

Data Processing

We downloaded the entire ABIDE dataset from ABIDE servers. In total, there were 1,433 fMRI and 1,383 MRI images from 19 sites. Due to ABIDE being a multi-site project, there were differences in the metadata annotation as well as the machines and techniques used to generate the data. In particular, many MRIs and fMRIs were different sizes, both over space and over time. To facilitate vectorization of the data, we first processed the MRI and fMRI images using the Nilearn package (8). The MRIs were down-sampled to a 128x128x128 size with a target affine of 1, whereas the fMRI images were down-sampled to a 64x64x31x85 size, also using a target affine of 1.

The train/dev/test sets were split randomly using a 70/15/15 split. This was done to make sure every site is roughly evenly represented in each of the train/dev/test sets and to have a robust model that would work across human heterogeneity and that the model would focus on biological differences of the brains of autism patients and typical controls.

Architecture Selection

Due to the wealth and heterogeneity of the data, we selected several network architectures and implmentations to test before focusing on an implementation which showed the most promise. We considered fMRI and MRI data separately and built the following networks to test these. All models were run on an NVIDIA GTX 1060 GPU with 6GB of VRAM.

1. Use one slice from an MRI image using a traditional CNN. This is analogous the common task of object detection (using a 2D CNN)
2. Use multiple slices to construct an estimated "3D" image from an MRI where the number of slices corresponds to channels in a traditional image. Apply 3D convolutions to extract information. This model was built in order to reduce computing time and increase depth of the network
3. Use a 3D convolutional network with an entire MRI scan
4. Use one slice with multiple frames of an fMRI scan to construct a "video" and then apply 3D CNN
5. Use a 3D CNN to extract input features from fMRI scans and pass them to an RNN.

Refer to Supplementary Table I for detailed architectures of the above models. After implementing each of the models and testing on a subset of the data, the 3D CNN model performed the best overall. As such, we focused further adjustments to this approach.

RESULTS

We chose to evaluate our model on the accuracy of prediction since our dataset was

Convolution 1 (L2 regularization λ = 0.01, Dropout rate = 0.1)					
Filters	Filter size	Stride	Padding	Activation	MaxPool
20	5 x 5 x 5	3	same	ReLU	2 x 2 x 2
Convolution 2 (Dropout rate = 0.1)					
Filters	Filter size	Stride	Padding	Activation	MaxPool
32	5 x 5 x 5	3	same	ReLU	2 x 2 x 2
Dense 1					
	Units		Activation		Dropout
	100		ReLU		0.1
Dense 2					
	Units		Activation		Dropout
	100		ReLU		-

Table I. Detailed Description of the Best Network of the Set.

relatively balanced with a 54:46 split (typical control:autism).

After initial tuning, we obtained a model which achieved 96% training and 62% dev set accuracy. To alleviate the overfitting, we added Dropout regularization (rate = 0.15) after the first dense layer. Results are shown below (Figure 3).

The model was able to achieve 62-67% accuracy on the held-out validation set and 63% accuracy on the test set. However, the variance of the model remained high. Attempts at regularization led to a drop in validation set accuracy.

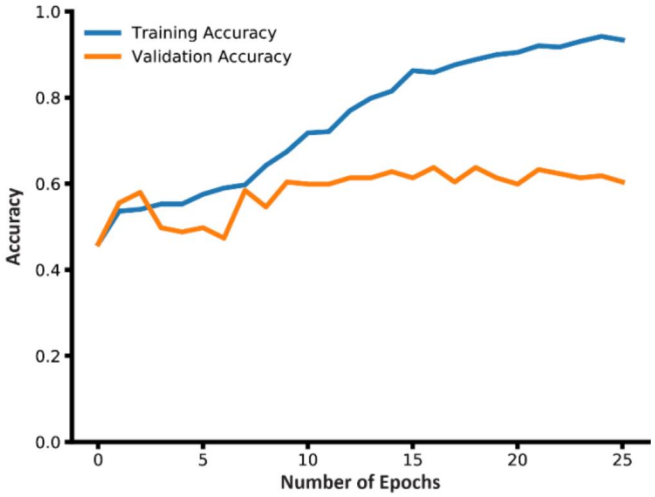

Figure 3. Train and Validation Accuracy Curve for the final model we chose.

	Training Set	Dev Set	Test Set
Loss	0.23	1.59	-
Accuracy	93%	60%	64%
Precision	-	.52	.59
Recall	-	.64	.78
F1-Score	-	.58	.67

Table II. Final Model Results.

To increase accuracy while addressing overfitting, further richness was added to the model. Increasing the number of filters in the first convolutional layer to 30 and increasing the number of hidden units in the first fully connected layer to 300. Dropout probabilities were also fine-tuned to better fit this model.

Comparing the models, our 3D CNN model with MRI data performed the best. This is likely because the 3D MRI was able to use information from across the whole brain, unlike our slice and video models, which were only able to predict based on a few slices in the z-direction. The CNN-RNN model with fMRI data also was not as accurate, likely because we had to heavily downscale the fMRIs to 64x64x64 to allow us to run the model, decreasing the resolution. Further, with the 3D CNN model, we were able to add more layers, allowing the neural network to learn more specific features.

DISCUSSION

After carrying out a number of experiments, we arrived at our current 3D CNN model which appeared to capture the most information from the MRI scans. The relative simplicity of the layers allowed us to build a deeper model with the available resources. We achieved results of 63% on the validation set with near perfect performance on our training set. The disparity between the results of the different sets shows evidence of

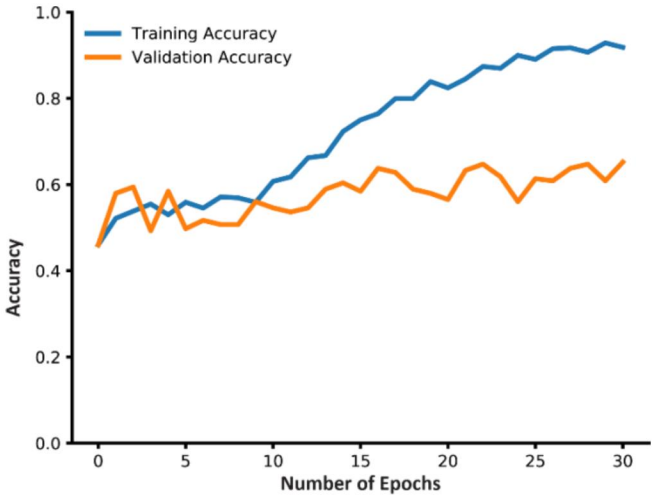

Figure 4. Train and Validation Accuracy Curve for the alternative model we generated. This model has slightly lower amount of overfitting.

overfitting to our training dataset, which we were able to mitigate using weight decay in the first convolutional layer and dropout regularization.

We noticed the first convolutional layer is especially sensitive to regularization. Any Dropout rate above 0.12 and L2 regularization above 0.01 destroys the model's ability to improve accuracy. This is probably due to the fact the information detected by the first layer is important and the number of neurons dedicated to this detection is too little, making it sensitive to regularization. In addition, our model was sensitive to higher dropout (dropout probability over 0.2) weights. We believe this is due to the small number of hidden units and filters. Information vital to classification is encoded in a few neurons; dropping any of them results in poor accuracy. A more effective strategy for regularization was spreading the dropout across several layers of the network.

We noticed that the first convolutional layer was also highly sensitive to other parameters including stride and pool size. This suggests that the resolution of the images is near the threshold of detecting important features. Any transformations that lower dimensionality could significantly hurt the model's performance. This also suggests that using higher resolution MRI scans could improve the performance of the architecture.

The original goal of the project was to make heavy use of fMRI images and exploit the time dimension of the scan to make the prediction. We believe that our MRI based model outperformed the

	Training Set	Dev Set	Test Set
Loss	0.23	1.59	-
Accuracy	92%	65%	63%
Precision	-	.59	.62
Recall	-	.54	.57
F1-Score	-	.56	.66

Table III. Alternative Model Results.

CNN + RNN architecture due to the higher resolution of the MRI images and the shallow depth of the model (limited by our computational infrastructure).

Upon visualization of the data, we noticed that the top and bottom part of the MRI images were less useful to our model. The top slices of the images only show a small portion of the cortex, and the bottom slices show only part of the brain stem, which mainly contributes to the lower level function of the central nervous system (such as breathing and heart rate). Neither of these structures is associated with ASD. Furthermore, the actual area occupied by these slices is relatively small and therefore adds less information than the middle slices.

Our model is comparable to other models that attempted to train on this dataset. Heinsfeld, et al achieved 70% accuracy training a deep network on fMRI data (9). Whereas, Nielsen, et al achieved 60% accuracy (10). However, there are a few key differences between our work and both of theirs. First, we used ABIDE's MRI data, as opposed to fMRI data, due to computational restrictions. With additional computational resources and appropriate network depth, it is likely that fMRI data would perform better in autism detection, due to the fact that fMRI also contains functional information. Secondly, we used raw MRI images, whereas both Nielsen and Heinsfeld featurized the fMRI using the human brain connectome. Using known, biologically-relevant features to train a network would likely improve accuracy, however, this could potentially limit the network from gaining insight into the structural and functional role of autism.

In recent years, neuroimaging and studies on the effects of psychiatric drugs revealed more pathophysiological basis of psychiatric disorders, and many of these findings are linked to structural and activation abnormalities to specific region of the brain. These anomalies can be detected

through a variety of imaging techniques including MRI and fMRI. Therefore, our work provides several models and model architectures that can be applied to a variety of psychiatric disorders not only for treatment purposes but also for research in pathogenesis of these diseases.

FUTURE DIRECTION

In the future, we will apply our model to a larger data set, acquired either through collection of new independent data from other sources or via data augmentation. This would help reduce overfitting and create a more robust model that generalizes well for autism spectrum disorder detection. Furthermore, we would also like to look into featurizing our dataset, particularly using the brain connectome to generate more relevant features for our model. A combination of all of these activities would likely improve our accuracy drastically, creating a robust model that could be used to understand the etiology of ASD.

To directly address the goal of this project – identifying parts of the brain associated with ASD, we could implement an occlusion sensitivity model which can output coordinates of the most relevant areas in a scan. It would also interesting to look at the activations of our network to see where it is differentiating between autistic and non-autistic cases

ACKNOWLEDGEMENT

The authors thank Ehsan Adeli for introducing the authors to the dataset and for providing the initial guidance on the project. The authors also thank Suraj Heereguppe for mentoring and useful discussions.

REFERENCES

1. APA AAP (2013) Cautionary Statement for Forensic Use of DSM-5. *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition*, p 947.
2. Autism, Investigators DDMNSY 2010 P (2010) Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2010. *Morb Mortal Wkly Rep Surveill Summ* 63:1–21.
3. Chen JA, Peñagarikano O, Belgard TG, Swarup V, Geschwind DH (2015) The Emerging Picture of Autism Spectrum Disorder: Genetics and Pathology. *Annu Rev Pathol Mech Dis* 10(1):111–144.
4. Lowe TL, Young JG, Cohen DJ, Tanaka K, Seashore MR (1980) Detection of Phenylketonuria in Autistic and Psychotic Children. *JAMA J Am Med Assoc* 243(2):126–128.
5. Novarino G, et al. (2012) Mutations in BCKD-kinase lead to a potentially treatable form of autism with epilepsy. *Science* 338(6105):394–7.
6. Daniels AM, et al. (2011) Stability of Initial Autism Spectrum Disorder Diagnoses in Community Settings. *J Autism Dev Disord* 41(1):110–121.
7. Dichter GS (2012) Functional magnetic resonance imaging of autism spectrum disorders. *Dialogues Clin Neurosci* 14(3):319–351.
8. Abraham A, et al. (2014) Machine Learning for Neuroimaging with Scikit-Learn. *NeuroImage Clin* 17:16–23.
9. Heinsfeld AS, Franco AR, Craddock RC, Buchweitz A, Meneguzzi F (2018) Identification of autism spectrum disorder using deep learning and the ABIDE dataset. *NeuroImage Clin* 17:16–23.
10. Nielsen JA, et al. (2013) Multisite functional connectivity MRI classification of autism: ABIDE results. *Front Hum Neurosci* 7:599.

SUPPLEMENTAL MATERIALS

Model	Architecture					
Single slice 2D CNN	*Convolution 1*					
	Filters	Filter size	Stride	Padding	Activation	MaxPool
	16	5 x 5	3	same	ReLU	2 x 2
	Convolution 2					
	Filters	Filter size	Stride	Padding	Activation	MaxPool
	32	5 x 5	1	same	ReLU	2 x 2
	Dense 1					
		Units		Activation		Dropout
		128		ReLU		0.2
Multiple slices 2D CNN	*Convolution 1*					
	Filters	Filter size	Stride	Padding	Activation	MaxPool
	16	5 x 5	1	same	ReLU	2 x 2
	Convolution 2					
	Filters	Filter size	Stride	Padding	Activation	MaxPool
	32	5 x 5	1	same	ReLU	2 x 2
	Dense 1					
		Units		Activation		Dropout
		128		ReLU		0.2
Single slices over 3D CNN	*Convolution 1*					
	Filters	Filter size	Stride	Padding	Activation	MaxPool
	32	5 x 5 x 5	1	same	ReLU	-
	Convolution 2					
	Filters	Filter size	Stride	Padding	Activation	MaxPool
	64	5 x 5 x 5	1	valid	ReLU	3 x 3 x3
	Dense 1					
		Units		Activation		Dropout
		128		ReLU		-
3D CNN on whole MRI	*Convolution 1 (L2 regularization λ = 0.01, Dropout rate = 0.1)*					
	Filters	Filter size	Stride	Padding	Activation	MaxPool
	16	5 x 5 x 5	3	same	ReLU	2 x 2 x 2
	Convolution 2 (Dropout rate = 0.1)					
	Filters	Filter size	Stride	Padding	Activation	MaxPool
	32	5 x 5 x 5	3	same	ReLU	2 x 2 x 2
	Dense 1 + Dense 2*					
		Units		Activation		Dropout*
		100		ReLU		0.1
CNN + RNN on fMRI	*Time Distributed Convolution 1*					
	Filters	Filter size	Stride	Padding	Activation	MaxPool
	16	5	1	same	ReLU	2 x 2 x 2
	Time Distributed Convolution 2					
	Filters	Filter size	Stride	Padding	Activation	MaxPool
	32	3	1	valid	ReLU	3 x 3 x 3
	LSTM 1					
			Units			Activation
			100			tanh

S. Table 1: Architectures of each of the proposed models used to select the primary model