Detecting Autism Spectrum Disorder with Neural Networks

Justin Alvin
University of Massachusetts Amherst
Amherst, MA
jalvin@cs.umass.edu

Abstract

We investigated the application of several different neural networks to the identification of autism spectrum disorder (ASD) based on functional and structural brain imaging patterns. Existing research has focused on the use of brain functional magnetic resonance imaging (MRI) data to identify regions in the brain that correlate with ASD. We replicated a state of the art study using functional MRIs and also explored the use of structural MRI analysis. Our research used the publicly available ABIDE dataset, containing functional and structural MRI data for over 1100 subjects. To analyze the fMRI data, we transferred weights learned by a denoising autoencoder to a fully-connected network (FCN) and achieved 0.64 accuracy in detecting ASD. However, we found that the autoencoder was unnecessary and were 0.71 accurate using only a slightly deeper fully-connected network, the highest reported accuracy for the ABIDE dataset. For structural MRI analysis, we trained 3D and 2D convolutional neural networks (CNNs) as well as a second FCN but were unable to get any signal from the test set for any of our approaches. We conclude that while we were able to improve upon the current state of the art for functional MRI analysis, further experimentation is necessary to successfully use structural MRI data.

1. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder with a range of symptoms, typically affecting speech and behavior patterns. ASD is characterized by an impairment of social function and communication as well as repetitive behaviors and interests. Recent Centers for Disease Control and Prevention (CDC) data indicates that ASD affects one in 68 children in the United States.

The neurobiological underpinnings of ASD are currently not fully understood. Most research is inconclusive and at times inconsistent, with reported increases and decreases in a variety of measures [9, 12, 14, 22]. While an analysis of brain structure or function is not necessary for the diagnosis of ASD, it can be useful in characterizing the neural basis for the disorder. The goal of magnetic resonance imaging (MRI) analysis is therefore to identify biomarkers for ASD that can inform the diagnosis and treatment of ASD and other brain disorders.

Existing research using machine learning to study ASD has largely focused on functional magnetic resonance imaging (fMRI) analysis. A number of differences in functional connectivity have been observed in ASD, including decreased cortico-cortical connectivity [17].

In the present study, we used several neural network architectures to examine functional and structural MR images of the brain. Specifically, we replicated a 2018 study by Heinsfeld *et al*. evaluating fMRI data using a denoising autoencoder [11]. Additionally, we extended the existing research by investigating structural MRI data using 3D and 2D convolutional neural networks (CNN). We explored whether cortical thickness measures in structural MRI data could be used by a CNN to detect ASD.

2. Background

A number of studies have applied machine learning techniques to disease prediction. Access to high quality MRI datasets has been a limiting factor in the progress made by this research, but recent initiatives have made large datasets for ASD, schizophrenia and Alzheimer's, among others, publicly available.

Existing studies in detecting ASD have classified subjects as either autistic or typically-developing control (TDC) based on fMRI brain activation. Murdaugh *et al*. used logistic regression to identify ASD in subjects with 0.963 accuracy [18], while Chen *et al*. used a random forest classifier to obtain 0.908 accuracy [4]. Other studies have used varying machine learning models, including SVMs and Bayesian networks [5, 1, 23].

A drawback to these studies is that they have used small, tightly controlled datasets. Because of the variable nature of both ASD and MR imaging techniques, these datasets often focus on a narrow age range or a single test site. For example, in the case of the aforementioned Chen *et al*. study, 126

ASD and 126 TDC subjects were used, while Murdaugh *et al*. analyzed MRIs from only 13 ASD and 14 TDC subjects.

The small sample size greatly limits the generalizability of these studies. Recent research using larger datasets (> 800 subjects) has been less promising, with much lower accuracies [19, 1]. The current state of the art for large datatsets uses transfer learning from an unsupervised autoencoder to a neural network of dense connections to achieve 0.70 accuracy [11].

While these studies have classified ASD using fMRI data, no machine learning-based research has investigated ASD detection using structural MRIs. However, qualitative analysis of structural MRIs has shown that the thickness of the cortex is increased in subjects with ASD [15], suggesting that there is something to be learned from structural MRI analysis.

3. The ABIDE I dataset

The present study evaluates functional and structural MRI data from the Autism Imaging Data Exchange (ABIDE I). The ABIDE dataset encompasses MRI results for 1112 subjects: 539 individuals with ASD and 573 TDC. Data was collected at 17 different imaging sites and includes resting state fMRI acquisitions, volumetric structural MRI scans and phenotypic information for each patient.

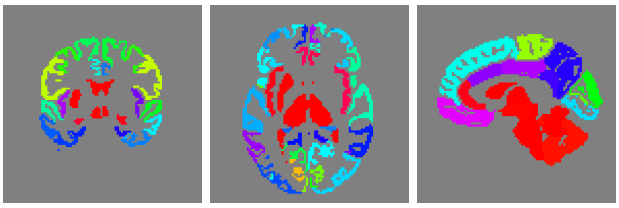
Functional and structural MRIs were preprocessed using separate pipelines, each with the goal of reducing noise and limiting the variance between outputs. Finally, brain images were delineated into regions of interest (ROIs) to isolate common areas of functional connectivity and structure.

After preprocessing and ROI segmentation, we were left with 884 samples, 408 with ASD and 476 TDC. Before evaluating our methods, we removed 100 samples from our dataset to use as test data. This data was not accessed until after training and optimizing each of our networks. Our training set consisted of the remaining 784 samples. Table 1 summarizes the distributions of ASD and TDC for our training and test sets, as well as other phenotypic information.

3.1. Functional MRI preprocessing

Functional MRI data was preprocessed using the C-PAC pipeline, provided by the Preprocessed Connectomes Project [6]. Basic preprocessing steps included slice timing correction, motion realignment, voxel intensity normalization and band-pass filtering (0.010.1 Hz). To clean possible confounding noise due to head motion and other physiological processes, each fMRI signal was processed using nuisance variable regression with 24 motion parameters, CompCor with 5 components, low-frequency drifts and global signal regressors. Finally, registration was performed by transforming each fMRI into template MNI152 space.

Each fMRI had a volumetric resolution of $61 \times 73 \times 61$ pixels. We divided the brain into spatially close regions of homogeneous functional activity using the CC200 parcellation atlas [7], yielding 200 regions of interest (ROIs).



(a) Coronal (b) Horizontal (c) Sagittal

Figure 1: **ROI cortical thickness measures**. Regions of interest in the cortex are color coded for visualization. Example coronal (a), horizontal (b) and sagittal (c) slices were taken at the midpoint of of each plane.

3.2. Structural MRI preprocessing

Cortical thickness measures were extracted from structural MRI data using the ANTs pipeline [2]. Preprocessing outputs were 3D volumes containing voxel-wise measures of cortical thickness of size $108 \times 146 \times 128$. Finally, ROIs were defined using the OASIS-TRT-20 joint fusion atlas in OASIS-30 space [16, 21, 3], resulting in 97 regions of interest in the cortex. Figure 1 shows coronal, horizontal and sagittal slices from a sample ROI atlas.

4. Functional MRI analysis

To analyze resting state fMRI data, we first replicated the experiments performed by Heinsfeld *et al*. [11] to use as a baseline. They reported 0.70 accuracy predicting ASD using a denoising autoencoder and fully-connected network (FCN) trained on the ABIDE dataset fMRIs. This was a binary classification task, classifying each subject as either ASD or TDC.

4.1. Connectivity matrix

We first extracted the mean time series for each of our 200 regions of interest. We then computed the Pearson correlation coefficient for the mean time series of each region, to create a connectivity matrix. Each cell in this matrix indicates the correlation between two ROIs, ranging from -1 (highly anti-correlated) to 1 (highly correlated). We removed the upper triangle of the matrix, which is just a repeat of the lower triangle, and the diagonal, which is the correlation of each region with itself. By flattening the remaining cells, we formed a 19900×1 vector of features for input to our neural network.

Before training our autoencoder, we applied artificial noise to our input by randomly setting elements in our input vector to zero.

Table 1: **Dataset and phenotype summary**. Diagnosis, age and sex distributions for training and test sets, as well as for total dataset after preprocessing.

Dataset	Samples	ASD	TDC	Age Avg (SD)	Age Range	Male	Female
Training	784	362	422	17.2 (8.0)	6.5-64.0	664	120
Test	100	46	54	17.6 (8.9)	7.0-50.2	82	18
Total	884	408	476	17.2 (8.1)	6.5-64.0	746	138

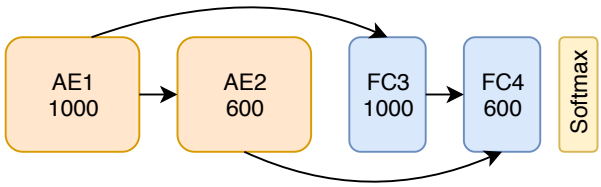


Figure 2: **Denoising autoencoder architecture** Our network consists of two stacked denoising autoencoders with 1000 and 600 hidden units respectively. Weights learned by the encoder layers of each autoencoder are transferred to the two layers of a neural network with dense connections of 1000 and 600 units.

4.2. Network architecture

A denoising autoencoder takes a noisy input as training data and reconstructs a denoised version as output as an unsupervised task. We transferred the weights learned by two stacked denoising autoencoders to a two-layer FCN. Each autoencoder was trained separately.

Both autoencoders contain two fully-connected layers: the first to encode the input to a new size and the second to decode it to its original dimensions. Each encoder layer is followed by a tanh non-linearity, while the decoder layers are linearly activated. Both autoencoders were trained to minimize mean squared error (MSE) such that:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \,. \tag{1}$$

Our first autoencoder contains 1000 hidden units and encodes our input of shape $N \times 19900 \times 1$ to a shape of $N \times 1000 \times 1$, where N is the number of samples. The decoder layer reconstructs our $N \times 19900 \times 1$ input.

The second autoencoder contains 600 hidden units. It takes the output of the first autoencoder's encoder layer as input and compresses it to a shape of $N \times 600 \times 1$. The decoder layer reconstructs the $N \times 1000 \times 1$ input.

Weights learned by the encoder layers of the autoencoders were used to initialize the two layers of a shallow FCN with 1000 and 600 hidden units respectively. Figure 2 shows the complete architecture of our autoencoder network. Each dense layer is followed by a tanh non-linearity. Dropout regularization [20] is applied after each layer with a probability of keeping a neuron of 0.6 for the first layer and 0.8 for the second. The final layer outputs class scores using the softmax function. As with the autoencoders, the FCN was trained to minimize MSE.

4.3. Training hyperparameters

The model described by Heinsfeld *et al*. was evaluated using 10-fold cross-validation on the entire dataset. However, because best practices in machine learning encourage the separation of the training and test sets during parameter optimization, we trained our network on our training set of 784 samples and evaluated it on the test set of 100 samples. Before classification of our test data, we measured the effectiveness of our model and our selection of hyperparameters using 10-fold cross-validation over the training set.

For both autoencoders, we optimized MSE using gradient descent with the Adam update rule and a learning rate, $\eta = 1 \times 10^{-4}$. We initialized all biases to 0 and weights, W using the Xavier uniform distribution such that:

$$\text{Var}(W) = \frac{1}{n_{in}}\,, \tag{2}$$

where n_{in} is the number of units in the previous layer [8]. We trained the first autoencoder for 700 epochs with a batch size of 100 and the second for 2000 epochs with a batch size of 10.

The FCN optimized MSE using the momentum update rule with a momentum, $\mu = 0.9$ and learning rate, $\eta = 5 \times 10^{-4}$. We trained our network over 100 epochs with a batch size of 10.

4.4. Experimental results

Our validation accuracy over 10-fold cross-validation was 0.67 ± 0.05, while our accuracy on the test set was 0.64. These results are slightly lower but comparable to the 0.70 accuracy that Heinsfeld *et al*. reported over 10-fold cross-validation on the entire dataset. It is likely that our

results are a more accurate representation of the classification power of this network and generalize better because we evaluated our model on previously unseen test data. Table 2 details the final accuracies and losses on the validation and test sets for each of the models we tested.

4.5. Fully-connected network

In the present study, we replicated the findings of Heinsfeld *et al*., suggesting that theirs is a valid approach to detecting ASD from fMRI data.

However, we hypothesize that the denoising autoencoders in our network are unnecessary. Heinsfeld *et al*. suggest that the unsupervised training of the autoencoders helps to reduce the noise caused by human error in selecting labels. We propose that because the autoencoders themselves are made up of layers of dense connections, a network of only fully-connected layers trained in a supervised manner should be able to learn the same features.

To test our hypothesis, we implemented a three-layer FCN. In shorthand notation, the architecture of our FCN is $FC(256) - BN - FC(128) - BN - FC(64) - BN$, where $FC(n)$ is a fully connected layer with n nodes. BN is a batch normalization layer, as described by Ioffe and Szegedy [13]. Each dense layer is followed by a tanh non-linearity. The final layer outputs class scores using the softmax function. We applied dropout with a keep probability of 0.25.

We found that the best results for our network were achieved using a learning rate, $\eta = 5 \times 10^{-6}$ and a batch size of 32 over 25 epochs. Hyperparameters were selected using random search over values for η and the possible inclusion of a regularization term, which was found to have no effect. Weights were initialized using a Xavier uniform distribution and gradients were optimized with the Adam update rule. We again minimized MSE during training.

We evaluated our classifier over 10-fold cross-validation and achieved a mean accuracy of 0.67 ± 0.06 on the validation set. Our fully-trained model had an accuracy on the test set of 0.71, higher than the test accuracy of the autoencoder and the results reported by Heinsfeld *et al*. These findings validate our hypothesis that the denoising autoencoders are unnecessary for successful ASD detection. We suggest that future research focus on deeper fully-connected models without the use of autoencoders given that performance is comparable (if not better) and training time is potentially lower depending on the complexity of the network.

5. Structural MRI analysis

In addition to training models to detect ASD using fMRI data, we trained several different networks on structural MRI data. Ours is the first approach to use structural MRIs to train a machine learning model for ASD classification.

Figure 3: **3D CNN architecture** Our network has 3 convolution, 3 max-pooling and 2 fully-connected layers, followed by a softmax output. All 3D convolution kernels are $3 \times 3 \times 3$ with stride 1 and dilation 2. Each box indicates the number of filters. The pooling kernels are $2 \times 2 \times 2$ with stride 2. Each fully-connected layer has 1024 hidden units.

Before training each model, we downsampled our MRI volumes to size $54 \times 73 \times 64$, which was necessary given the large amount of memory required for our network to store each sample. Volumes were then scaled such that:

$$x_i = \frac{x_i - \min(x)}{\max(x) - \min(x)} \,. \tag{3}$$

5.1. 3D convolutional neural network

In shorthand notation, the architecture of our 3D CNN is $C(16,3,1) - BN - P - C(32,3,1) - BN - P - C(32,3,1) - BN - P - FC(1024) - BN - FC(1024) - BN$, where $C(d,f,s)$ indicates a 3D convolutional layer with d filters of size $f \times f \times f$ and stride s. P represents a $2 \times 2 \times 2$ pooling layer with stride 2. Each convolutional and dense layer is followed by a rectified linear unit (ReLU) non-linearity. The final layer outputs class scores using the softmax function.

Given that we expect very local regions in the brain to correlate with ASD, we wanted to increase the receptive field size to improve detection of these local features. To do this, we dilated each convolution by 2, thereby ensuring that we more fully account for 3D spatial information over a small number of layers. Figure 3 visualizes the complete architecture of our network.

As described by He *et al*., we initialized all biases to be 0 and the weights, W at each layer to have variance:

$$Var(W) = \frac{2}{n_{in}} \,, \tag{4}$$

where n_{in} is the number of input nodes [10]. We trained our model using Adam optimization of the cross-entropy loss, L such that:

$$L_i = -f_{y_i} + \log \sum_j e^{f_j} \,, \tag{5}$$

where f_j indicates the $j - th$ element of the vector of class scores, f. We used a learning rate, $\eta = 3 \times 10^{-6}$ over 35 epochs with a batch size of 32.

Because we trained our model on a relatively small dataset with a large number of parameters due to its 3D nature, we expected overfitting to be a substantial problem.

Table 2: **Test and validation results**. Final prediction accuracies and losses for our validation and test sets for each model we evaluated. Our FCN trained on fMRI data achieved the highest accuracy on the test set, at 0.71. None of our models trained on structural MRI data were able to reach an accuracy much better than random. The current state of the art performance on the ABIDE dataset is 0.70, trained on fMRI data [11].

Model	**MRI Type**	**Accuracy**		**Loss**	
		Validation (SD)	**Test**	**Validation (SD)**	**Test**
State of the art	functional	-	0.70	-	-
Autoencoder	functional	0.67 (0.05)	0.64	0.24 (0.03)	0.23
FCN	functional	0.67 (0.06)	**0.71**	0.21 (0.03)	0.19
CNN (3D)	structural	0.54 (0.02)	0.52	0.97 (0.05)	0.94
CNN (2D)	structural	0.56 (0.02)	0.57	0.84 (0.07)	0.95
FCN	structural	0.59 (0.03)	0.54	0.79 (0.07)	0.75

Consequently, we implemented dropout with a keep probability of 0.5 after each of the two fully-connected layers and regularized weights using L2 regularization with a constant, $\lambda = 1 \times 10^{-5}$.

We selected appropriate values for η, dropout and λ using random search, cross-validating the loss and accuracy of each set of hyperparameters over a 5-fold scheme. Our mean validation accuracy was 0.54 ± 0.02, not much better than random labeling, suggesting that our network greatly overfit to the training set. Looking at the accuracy and loss curves shown in Figures 4a and 4d, we can see that while the training set converges, the validation accuracy fluctuates around 0.5 and the loss only decreased slightly.

To account for this, we varied regularization strength up to $\lambda = 10$ for both L2 and L1 regularization and decreased the dropout keep probability to 0.25. We saw no improvement to validation accuracy or loss. Adjusting the learning rate as well as using different weight initialization strategies and alternatives to the Adam update rule had no noticeable effect. For all sets of hyperparameters used, training accuracy would eventually reach 1.0 while validation accuracy remained close to random.

Additionally, we explored several alternative network architectures. These included a deeper network of $C(32,3,1) - BN - P - C(64,3,1) - BN - P - C(128,3,1) - C(128,3,1) - BN - P - C(256,3,1) - C(256,3,1) - BN - P - FC(1024) - BN - FC(1024) - BN$ and a shallow network with a dilation rate of 3 and smaller fully-connected layers of $C(16,5,1) - BN - P - C(32,5,1) - BN - P - FC(128) - BN - FC(128) - BN$. Similar to our hyperparameter adjustments, changing the network architecture did not improve our validation accuracy. The accuracy of our final 3D CNN model on the test set was 0.52 or essentially random.

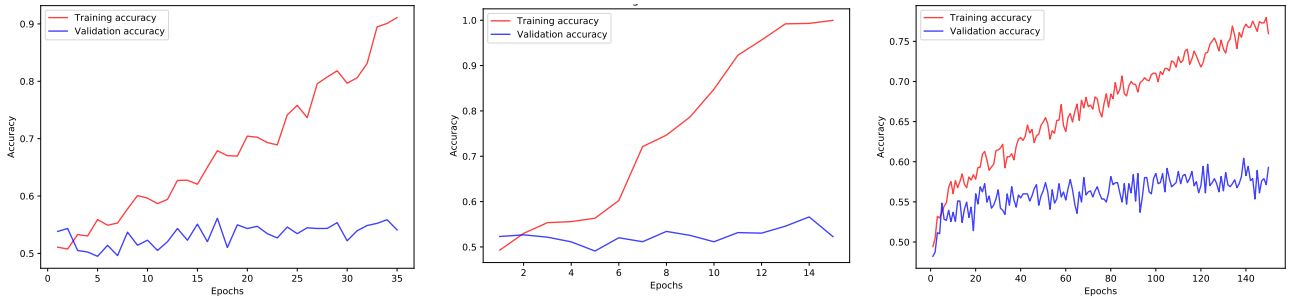
We conclude that a 3D CNN is an ineffective model to use on the ABIDE dataset for ASD classification. Our model was unable to detect any signal and instead fit to noise in the training set. This is possibly due to using a small dataset for a complicated classification task. Additionally, we may be increasing the problem of overfitting by using a 3D CNN network that has too many parameters relative to the size of our dataset.

5.2. 2D convolutional neural network

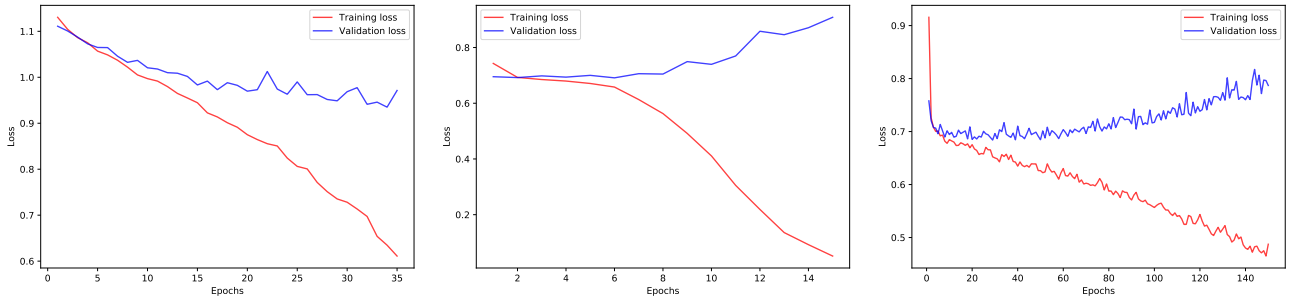
We explored using a CNN with 2D convolutions instead of 3D to lower the complexity of our model and reduce overfitting. Each of our MRI volumes contains of a single grayscale color channel, allowing us to train a 2D CNN using the third dimension of each volume as the number of channels. In other words, a CNN expecting inputs of shape $H \times W \times C$, where H, W and C are the height, width and number of channels respectively, receives an input of shape $H \times W \times D$, where D is the volume depth. We oriented our volumes such that H, W and D correspond to the sagittal, horizontal and coronal planes of the brain respectively.

In shorthand notation, the architecture of our 2D CNN is $C(64,3,1) - BN - C(64,3,1) - BN - P - C(128,3,1) - BN - C(128,3,1) - BN - P - FC(1024) - BN - FC(1024) - BN$, where $C(d,f,s)$ indicates a 2D convolutional layer with d filters of size $f \times f$ and stride s. For this network, P represents a 2×2 pooling layer with stride 2. Each batch normalization layer (BN) is followed by a rectified linear unit (ReLU) non-linearity. The final layer outputs class scores using a softmax classifier. Figure 5 shows the architecture of our 2D network.

As with the 3D CNN model, we initialized all biases and the weights using He initialization. To prevent overfitting, we applied dropout after each fully-connected layer with a keep probability of 0.75. We trained our model using Adam optimization, minimizing the cross-entropy loss with



(a) 3D CNN acc. (b) 2D CNN acc. (c) FCN acc.



(d) 3D CNN loss (e) 2D CNN loss (f) FCN loss

Figure 4: **Accuracy and loss curves**. Comparision of training (red) and valdiation (blue) accuracy and loss curves for 3D CNN, 2D CNN and FCN models for structural MRI analysis. For the CNN models, the validation accuracy fluctuates around 0.5, while slightly improving for the the FCN model.

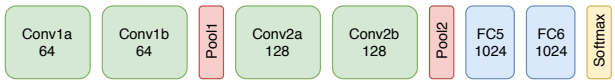


Figure 5: **2D CNN architecture** Our network has 4 convolution, 2 max-pooling and 2 fully connected layers, followed by a softmax output layer. All 2D convolution kernels are 3×3 with stride 1. Each box indicates the number of filters. The pooling kernels are 2×2 with stride 2. Each fully connected layer has 1024 output units.

a learning rate, $\eta = 1 \times 10^{-5}$ over 15 epochs with a batch size of 32.

We evaluated our model over 5-fold cross-validation, selecting hyperparameters using random search. Our mean validation accuracy was 0.56 ± 0.02 and our test accuracy was 0.57. These results were slightly higher than those for the 3D CNN. However, if we consider the accuracy and loss curves for our 2D model, shown in Figures 4b and 4e, we see that these results may be misleading. Our 2D CNN does not fair much better than the 3D version: while the training accuracy reaches 1.0, the validation accuracy again fluctuates around 0.5.

We implemented L2 regularization to reduce overfitting but found that it did not improve our validation accuracy or loss. As with the 3D CNN, we were unable to register any signal on the test set and overfit on the training set.

5.3. Fully-connected network

Our final approach to structural MRI analysis draws on the comparative success of our functional MRI analysis. Because we have already segmented each 3D volume into the same 97 ROIs, we used the raw coritical thickness measures for those regions directly without mapping them to 3D space. Our input for each subject is simply a 97×1 vector of cortical thickness values for each ROI.

We trained a fully-connected network to detect ASD based on this input. Our network consists of two densely-connected layers, each with 64 hidden neurons followed by a batch normalization layer and ReLU non-linearity. In shorthand notation, our architecture is $FC(64) - BN - FC(64) - BN$. The final layer outputs class scores using a softmax classifier.

This was a much smaller and shallower network compared to the other models we tested. A benefit of this simplified approach is that we have reduced the number of parameters to account for overfitting.

We trained our model to minimize cross-entropy loss using Adam optimization with a learning rate of 7×10^5 and batch size of 32 over 150 epochs. Biases were again initial-

ized to 0 and weights with the He initialization method.

Using 5-fold cross-validation, we evaluated the effectiveness of our network at detecting ASD. Our mean validation accuracy was 0.59 ± 0.03, slightly higher than our two CNN models. If we examine the accuracy and loss curves, shown in Figures 4c and 4f respectively, we can see that the validation accuracy of our FCN is gradually increasing but at a much lower rate than the training accuracy. We stopped training early because we found that validation loss began to explode after 150 epochs.

Applying dropout and L2 regularization to reduce this overfitting did not make any improvements. However, our test accuracy is quite low compared to validation accuracy, at 0.52, indicating that regularization may have actually been effective. Our FCN model does appear to detect some signal in the test set, unlike our two CNNs. While our final results are unimpressive, this is encouraging and suggests that a simpler model with significantly fewer parameters is better suited to our small dataset.

5.4. Discussion

None of our approaches to analyzing structural MRI data were very successful in detecting ASD. While the FCN showed some promising results, our test set accuracies for all three models showed random guessing. These findings are somewhat problematic given that qualitative analysis of structural MRIs in the ABIDE dataset has shown a correlation between cortical thickness levels and ASD [15].

Most likely we don't see any results because we do not have enough data for our network to learn anything. Because we have such a small amount of data, deeper networks may be ineffective at detecting a signal because they have too many parameters and are therefore very likely to overfit.

Unfortunately, applying data augmentation techniques to increase the size of our dataset is not possible because the specific orientation of the brain is fundamental to its structure and function. Affine transformations during data augmentation mean we lose that orientation information and our network will likely struggle to learn even more.

We conclude that at present, CNNs are ineffective at identifying ASD using structural MRI data. However, because we are confident that there is a signal in our dataset, we believe further investigation is necessary.

6. Conclusion

We evaluated several methods for identifying ASD based on functional and structural MRI data from the ABIDE dataset. While we were able to get promising results analyzing fMRI data, we were unsuccessful in using structural MRIs to detect ASD.

We replicated the approach of Heinsfeld *et al*. in analyzing fMRI data, with similar reported accuracies, but call into question their use of a denoising autoencoder. Our results show that a deep fully-connected network trained on functional connectivity measures between ROIs can outperform the current state of the art. Furthermore, because fMRIs are temporal in nature, we propose the use of a Long Short-Term Memory (LSTM) network to leverage this information more effectively.

Our results analyzing structural MRI data were inconclusive. While we expected to see a signal in our test set based on prior, qualitative research using the ABIDE dataset, we were unable to achieve accuracies much better than random for any of the models we tested. We propose that in order for neural network methods—in particular, CNNs—to be useful in predicting ASD and isolating regions in the brain that correlate with ASD, we need larger datasets.

Detection of ASD is a particularly challenging neural networks task. ASD is a spectrum of disorders rather than a binary condition and can present in many different ways, depending on the age of the subject and severity of the disorder. As such, there is a high degree of variance in ASD MRI datasets that is reflected in the wide range of findings and often conflicting results. Further confounding analysis is the possibility of noise in ground truth caused by human error in diagnosing ASD.

Existing research has attempted to control for these variables by working with smaller, focused datasets. While such approaches may yield higher classification accuracies, they do not generalize well. The present study suggests improvements on the current state of the art for large MRI dataset analysis using machine learning. We propose that with larger datasets, better imaging techniques and more controlled diagnostic methodology, neural networks can make useful contributions to the study of ASD.

References

[1] A. Abraham, M. P. Milhame, A. D. Martino, R. C. Craddock, D. Samaras, B. Thirion, and G. Varoquaux. Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. *NeuroImage*, 147:736–745, 2017.

[2] B. B. Avants, N. J. Tustison, and H. J. Johnson. Advanced normalization tools. `http://stnava.github.io/ANTs/`.

[3] R. Bakker, P. Tiesinga, and R. Ktter. The scalable brain atlas: instant web-based access to public brain atlases and related content. *Neuroinformatics*, 2015.

[4] C. P. Chen, C. L. Keown, A. Jahedi, A. Nair, M. E. Pflieger, B. A. Bailey, and R.-A. Mller. Diagnostic classification of intrinsic functional connectivity highlights somatosensory, default mode, and visual regions in autism. *Neuroimage Clinical*, 9:238–245, 2015.

[5] H. Chen, X. Duan, F. Liu, F. Lu, X. Ma, Y. Zhang, L. Q. Uddin, and H. Chen. Multivariate classification of autism spectrum disorder using frequency-specific resting-state functional connectivitya multi-center study. *Progress in Neuro-*

Psychopharmacology and Biological Psychiatry, 64:1–9, 2016.

[6] R. C. Craddock, Y. Benhajali, C. Chu, F. Chouinard, A. Evans, A. Jakab, B. S. Khundrakpam, J. D. Lewis, Q. Li, M. Milham, C. Yan, and P. Bellec. The neuro bureau preprocessing initiative: open sharing of preprocessed neuroimaging data and derivatives. *Neuroinformatics*, 2013.

[7] R. C. Craddock, G. A. James, P. E. Holtzheimer, X. P. Hu, , and H. S. Mayberg. A whole brain fmri atlas generated via spatially constrained spectral clustering. *Human Brain Mapping*, 33(8):1914–1928, 2011.

[8] X. Glorot and Y. Bengio. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of Machine Learning Research*, 9:249–256, 2010.

[9] N. Hadjikhani, R. M. Joseph, J. Snyder, and H. Tager-Flusberg. Anatomical differences in the mirror neuron system and social cognition network in autism. *Cerebral Cortex*, 6:1276–1282, 2006.

[10] K. He, X. Zhang, S. Ren, and J. Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. *IEEE International Conference on Computer Vision*, 2015.

[11] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi. Identification of autism spectrum disorder using deep learning and the abide dataset. *NeuroImage: Clinical*, 17:16–23, 2018.

[12] K. L. Hyde, F. Samson, A. C. Evans, and L. Mottron. Neuroanatomical differences in brain areas implicated in perceptual and other core features of autism revealed by cortical thickness analysis and voxel-based morphometry. *Human Brain Mapping*, 31:556–566, 2010.

[13] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *CoRR*, 2015.

[14] R. J. Jou, N. J. Minshew, M. S. Keshavan, M. P. Vitale, and A. Y. Hardan. Enlarged right superior temporal gyrus in children and adolescents with autism. *Brain Research*, 1360:205–212, 2010.

[15] B. S. Khundrakpam, J. D. Lewis, P. Kostopoulos, F. Carbonell, and A. C. Evans. Cortical thickness abnormalities in autism spectrum disorders through late childhood, adolescence, and adulthood: A large-scale mri study. *Cerebral Cortex*, 27(3):1721–1731, 2017.

[16] A. Klein, S. S. Ghosh, F. S. Bao, J. Giard, Y. Hame, E. Stavsky, N. Lee, B. Rossa, M. Reuter, E. C. Neto, and A. Keshavan. Mindboggling morphometry of human brains. *PLoS Computational Biology*, 13(3), 2017.

[17] A. D. Martino, C. Kelly, R. Grzadzinski, X.-N. Zuo, M. Mennes, M. A. Mairena, C. Lord, F. X. Castellanos, and M. P. Milham. Aberrant striatal functional connectivity in children with autism. *Biological Psychiatry*, 69(9):847–856, 2011.

[18] D. L. Murdaugh, S. V. Shinkareva, H. R. Deshpande, J. Wang, M. R. Pennick, and R. K. Kana. Differential deactivation during mentalizing and classification of autism based on default mode network connectivity. *PLoS ONE*, 7(11), 2012.

[19] J. A. Nielsen, B. A. Zielinski, P. T. Fletcher, A. L. Alexander, N. Lange, E. D. Bigler, J. E. Lainhart, and J. S. Anderson. Multisite functional connectivity mri classification of autism: Abide results. *Frontiers in Human Neuroscience*, 7(599), 2013.

[20] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014.

[21] N. J. Tustison, P. A. Cook, A. Klein, G. Song, S. R. Das, J. T. Duda, B. M. Kandel, N. V. Strien, J. R. Stone, J. C. Gee, and B. B. Avants. Large-scale evaluation of ants and freesurfer cortical thickness measurements. *NeuroImage*, 2014.

[22] G. L. Wallace, N. Dankner, L. Kenworth, J. N. Giedd, and A. Martin. Age-related temporal and parietal cortical thinning in autism spectrum disorders. *Brain*, 133:3745–3754, 2010.

[23] Y. Zhou, F. Yu, and T. Duong. Multiparametric mri characterization and prediction in autism spectrum disorder using graph theory and machine learning. *PLoS ONE*, 9(6), 2014.