Identification of Autism Spectrum Disorder using Deep Learning and the ABIDE Dataset

Anibal Sólon Heinsfeld[a], **Alexandre Rosa Franco**[b,c,d], **R. Cameron Craddock**[f,g], **Augusto Buchweitz**[b,d,e], **and Felipe Meneguzzi**[a,b]

[a]PUCRS, School of Computer Science, Porto Alegre 90619, Rio Grande do Sul, Brazil.; [b]PUCRS, Brain Institute of Rio Grande do Sul (BraIns), Porto Alegre 90619, Rio Grande do Sul, Brazil.; [c]PUCRS, School of Engineering, Porto Alegre 90619, Rio Grande do Sul, Brazil.; [d]PUCRS, School of Medicine, Porto Alegre 90619, Rio Grande do Sul, Brazil.; [e]PUCRS, School of Humanities, Porto Alegre 90619, Rio Grande do Sul, Brazil.; [f]Center for the Developing Brain, Child Mind Institute, New York, New York 10022, USA.; [g]Nathan Kline Institute for Psychiatric Research, Orangeburg, New York 10962, USA.

This manuscript was compiled on April 13, 2018

The goal of the present study was to apply deep learning algorithms to identify autism spectrum disorder (ASD) patients from large brain imaging dataset, based solely on the patients brain activation patterns. We investigated ASD patients brain imaging data from a world-wide multi-site database known as ABIDE (Autism Brain Imaging Data Exchange). ASD is a brain-based disorder characterized by social deficits and repetitive behaviors. According to recent Centers for Disease Control data, ASD affects one in 68 children in the United States. We investigated patterns of functional connectivity that objectively identify ASD participants from functional brain imaging data, and attempted to unveil the neural patterns that emerged from the classification. The results improved the state-of-the-art by achieving 70% accuracy in identification of ASD versus control patients in the dataset. The patterns that emerged from the classification show an anticorrelation of brain function between anterior and posterior areas of the brain; the anticorrelation corroborates current empirical evidence of anterior-posterior disruption in brain connectivity in ASD. We present the results and identify the areas of the brain that contributed most to differentiating ASD from typically developing controls as per our deep learning model.

Autism | fMRI | ABIDE | Resting State | Deep Learning

The primary goal of psychiatric neuroimaging research is to identify objective biomarkers that may inform the diagnosis and treatment of brain-based disorders. Data-intensive machine learning methods are a promising tool for investigating the replicability of patterns of brain function across larger, more heterogeneous data sets [1]. The first goal of the present study was to classify autism spectrum disorder (ASD) and control participants based on their respective neural patterns of functional connectivity using resting state functional magnetic resonance imaging (rs-fMRI) data. We used a deep learning method that combined supervised and unsupervised machine learning (ML) methods. The method was applied to a large population sample of brain imaging data, the Autism Imaging Data Exchange I (ABIDE I). The second goal was to investigate the neural patterns associated with ASD that contributed most to the classification; the results are interpreted in the light of the networks of regions within the brain that differentiate ASD from controls and of previous studies of ASD brain function.

ASD is associated with a range of phenotypes that vary in severity of social, communicative and sensorimotor deficits. ASD diagnostic instruments assess the characteristic social behaviors and language skills (see our result about real-world classification accuracy). Yet neuroscientific research can help bridge the gap between a clear mapping of the complexity of the spectrum of alterations in autism behavior and their neural patterns [2]. Noninvasive brain imaging studies have advanced the understanding of the neural underpinnings of brain-based disorders and their associated behavior, such as ASD and its social and communicative deficits [3–6]. The identification of patterns of activation for ASD and the association of the patterns with neural and psychological components contributes to the understanding of the etiology of mental disorders [5, 7].

One of the challenges to brain imaging studies of brain disorders is to replicate findings across larger, more demographically heterogeneous datasets that reflect the heterogeneity of clinical populations. Recently, ML algorithms have been applied to brain imaging data to extract replicable brain function patterns. These algorithms can extract replicable, robust neural patterns from brain imaging data of psychiatric disorder patients [8].

Machine-learning and disease state prediction: the next frontier for understanding the brain and psychiatric disorders

The combination of machine-learning methods with brain imaging data has allowed for the classification of mental states associated with the representation of semantic categories [9, 10], of the meanings of nouns [11–13], of emotions [14], and of learning [15]. In the case of mental disease states, studies have identified patients of brain activation associated with schizophrenia [16], with autism [5], and with depression [17]. Studies that appllied ML algorithms to ASD brain imaging data have classified individuals as autistic or control from their fMRI brain activation with up to 97% accuracy within single sites. They also identified a pattern of brain activation associated with a psychological factor (self-representation). The pattern was present in control patients and nearly absent for autistic participants [5]. In another study of classification of ASD participants [18], the authors obtained a 76.67% classification accuracy in a population sample of 178 ASD and IQ-matched typically-developing participants.

A caveat of studies that applied supervised ML to brain imaging data is their relatively small number of participants. Arbabshirani et al. [19] showed that the reliable classification accuracies of ML studies were obtained with population samples with fewer than 100 participants; higher classification accuracies, that is, above 90 percent, were only obtained with studies constrained to dozens of participants. Classification accuracy drops significantly in larger population samples and if data is from different sites [20].

Most studies that combine brain imaging and machine learning have applied supervised learning methods, such as support vector machine (SVM) or Gaussian naïve Bayes (GNB) classifiers. The subjectivity of feature selection procedures for supervised machine learning methods may be an obstacle for the comparison of results across studies. In supervised methods, class labels are assigned to a set of data used as the training data set; other data points (test data set) are classified in relation to the patterns found in the training data (using the given labels). In other words, the algorithm operates to classify pre-established labels (that is, they rely on feature selection, or feature engineering).The choice of these labels and of the features depends on a priori hypothesis or exploratory procedures; hence, they depend on a level of subjectivity. For example, the number of voxels used for classification of brain imaging data has been empirically selected on the basis of exploring sets of 100, 200, 400 and more voxels and identifying the set size that works best for the classification [11–13].

In the present study, we address the issues of generalizability and of subjectivity by classifying a psychiatric disorder using a large data set and an unsupervised machine learning method. Reduction of subjectivity in feature extraction may provide a new window into brain function that is less experimenter-dependent and more data-driven.

Classification of the ABIDE dataset

ABIDE data have previously been used by Nielsen et al. [20] to classify autism versus control subjects based on brain connectivity measurements. The authors reproduced an approach reported in Anderson [21] with modifications that included datasets from multiple sites. BOLD signal from non-overlapping, grey matter ROIs (SPM8 mask grey.nii) formed by seed voxels separated by at least 5mm was computed for the 964 subjects used. Voxels that were Euclidean-close to a specific ROI's seed voxel were included in this ROI. Based on data from the 7266 generated ROIs, Nielsen et al. [20] computed a connectivity matrix with size of 7266 x 7266, by calculating the pair-wise correlation between each ROI. Using a leave-one-out approach, a general linear model was fit to each group (ASD and control) to associate the connectivity matrix with subject-related variables: age, gender and handedness. The value of each connection was estimated for the left out subject based on the variables. It was then adjusted by using the difference between one site's mean value for the connection and another site's mean value for the same connection. This procedure mitigated between-site differences that could bias results, such as different scanners and variations in scanning parameters and protocols.

The authors attempted to accommodate multi-site data and sources of variance present in the ABIDE data. For the left-out subject, the actual value for each connection was then subtracted from the estimated values obtained from the autism model and from the control model. The average of this subtraction across all 7266 ROIs was computed, and the average values of ROIs were added up. Positive values were classified as ASD and negative values, as controls. Nielsen et al. [20] obtained as high as 60% accuracy for the classification ASD versus controls. Recently, Abrahams *et al.* [22] achieved the highest classification up to the present paper. By building participant-specific functional connectivity matrices (connectomes), the authors achieved 67% accuracy in the full ABIDE dataset. In the present study, we aimed to improve that highest accuracy obtained.

Neuroimaging and Deep learning algorithms

Koyamada et al. [23] investigated brain states from measurable brain activities by using Deep Neural Networks (DNN). They trained a artificial neural network with two hidden layers and a softmax output layer to classify task-based fMRI data from 499 subjects into seven categories related to the tasks: Emotion, Gambling, Language, Motor, Relational, Social and Working Memory. Deep models allowed for better results (mean accuracy of 50.74%) compared to supervised learning methods (mean accuracy of 47.97%) such as Linear Regression and Support Vector Machine. Plis et al. [24] used deep learning and structural T1-weighted images in order to classify patients with schizophrenia versus matched healthy controls, using data from four different sites; the authors also classified patients with Huntington disease versus healthy controls, using data combined by the PREDICT-HD project (www.predict-hd.net). First, they attempted to classify 198 schizophrenic patients and 191 controls from four different studies conducted by Johns Hopkins University (JHU), the Maryland Psychiatric Research Center (MPRC), the Institute of Psychiatry, London, UK (IOP), and the Western Psychiatric Institute and Clinic at the University of Pittsburgh (WPIC). Plis et al. trained a Deep Belief Network with 3 depths (50 hidden units in the first layer, 50 in the second layer, and 100 in the top layer). They achieved 90% classification accuracy using features extracted from three DBMs in comparison to 68% classification accuracy using raw data in a Support Vector Machine. The authors concluded that deep learning holds great potential for clinical brain imaging applications.

The second part of the work used data collected from healthy controls and patients with Huntington disease of the PREDICT-HD project. The study aimed to identify the disease using deep learning techniques, and assess the levels of severity of the disease (low, medium and high). The study used a large dataset of T1-weighted structural scans from 32 sites from different countries. The set included 2641 images from patients and 859 from healthy controls. The authors applied a Deep Belief Network (DBM) with 3 depths: 50-50-100 hidden units in the first, second and the top layer respectively. The t-Distributed Stochastic Neighbor Embedding (t-SNE) [25] technique was used to reduce the resultant data to a 2-dimensional version; the results showed a linearly separable projection into patients and control.

Deep learning algorithms take classification of brain imaging data a step further than strictly supervised methods. The algorithms use complex data representations in the learned model. Deep learning algorithms rely on minimal human intervention for extracting relevant features by using unsupervised learning methods. Classification of clinical populations using unsupervised methods may allow for exploratory search of neural patterns of psychiatric disorders that is less dependent on generating hypotheses for feature selection; it may be thus less susceptible to category errors. In supervised methods, presumed labels are used to train the classifier and find patterns of brain activation or connectivity associated with the labels (e.g. a clinical population and a control population sample). In unsupervised methods, the classifier explores population samples for patterns in the brain which may be associated with a clinical population; again, the subjectivity involved in label selection is avoided [24]. It is suggested that less subjective and possibly more unrestrained, deep learning algorithms hold promise for the application of machine learning to big data sets from multi-site repositories.

Materials and Methods

Participants. The present study was carried out using rs-fMRI data from the Autism Imaging Data Exchange (ABIDE I). ABIDE is a consortium that provides previously collected rs-fMRI ASD and matched controls data for the purpose of data sharing in the scientific community [26]. We included data from 505 ASD individuals and 530 matched controls (typical controls, TC). The ABIDE datasets were collected at 17 different imaging sites and include rs-fMRI images, T1 structural brain images and phenotypic information for each patient, which is summarized in Table 4. Table 4 contains key phenotypical information,[1] including distribution of ASD and TC by sex and age and the ADOS score for ASD subjects, as well as the Mean Framewise Displacement (FD) quality measure.[2]

Resting state and feature selection. Resting state fMRI provides neural measurements of the functional relationship between areas of the brain. Rs-fMRI data is particularly useful for investigation of clinical populations. It allows for investigation of the disruption brain networks without the added complexity of variation associated with task-related brain activation [18, 27]. It may be applied in the investigation of mental states, memory and the recall of events, clinical populations, among others [28, 29]. Rs-fMRI has been shown to be highly reproducible and provides data sets that can be easily compared across studies [30, 31]. The correlation of low frequency fluctuations on resting-state fMRI arises from fluctuations in blood oxygenation or flow. It is a manifestation of functional connectivity of the brain [32]. To investigate brain connectivity, a correlation is calculated for the average of the time series of the regions of interest. The correlation is used to build a connectivity matrix.

Data Preprocessing. Previously preprocessed rs-fMRI data was downloaded from the Preprocessed Connectomes Project (http://preprocessed-connectomes-project.org/). Data was selected from the C-PAC preprocessing pipeline. The fMRI data was slice time corrected, motion corrected, and the voxel intensity was normalized. Nuisance signal removal was performed using 24 motion parameters, CompCor with 5 components [33], low-frequency drifts (linear and quadratic trends), and global signal as regressors. Functional data was band-pass filtered (0.01-0.1Hz) and spatially registered using a nonlinear method to a template space (MNI152).

The mean time series for regions of interest was extracted for each subject. The CC200 functional parcellation atlas of the brain [34] was used to reduce the features vector size (see below). This atlas was generated by a data-driven parcellation of the whole brain into spatially close regions of homogeneous functional activity, totalizing in 200 regions.

Feature selection: functional connectivity of ROIs. Functional connectivity was used to classify subjects as ASD and TC. Functional connectivity provides an index of the level of co-activation of brain regions based on the time-series of rs-fMRI brain imaging data. Each cell in the connectivity matrix contains a Pearson correlation coefficient. The coefficient is an index of the correlation between two areas of the brain, and it ranges from 1 to -1: values close to 1 indicate that the time series are highly correlated; values close to -1 indicate that time series are anti-correlated. To upper triangle values were removed for use of the values in the correlation matrix as features. These values repeat the values of the lower triangle. We also removed the main diagonal of the matrix, since it represents an area correlating to itself. Later, we flattened the remaining triangle (i.e. collapse it in a one-dimension vector) to retrieve a vector of features, with the purpose of using it for subject classification. The number of resultant features is defined by the following equation, $S = \dfrac{(N-1)N}{2}$, in which N is the number of correlated voxels or regions. The CC200 ROI atlas was used, and the procedure resulted in 19,900 features.

Classification method. Denoising autoencoders were used to train the predictive model for better generalization; i.e. accurate classification of new subjects outside the initial pool of participants. Denoising autoencoders reconstruct input based on a corrupted version of the input [35]: stochastically, some positions of the vector derived from a functional connectivity matrix are set to zero before training the model. The corruption modules applied to corrupt data are based on binomial distributions. The goal of the technique is to make the model sufficiently accurate for predictions using novel data [36].

In the present study, we used two stacked denoising autoencoders for the unsupervised pre-training stage to extract a lower-dimensional version from the ABIDE data. We achieved the best optimization for the validation set using reconstruction loss (mean squared error); the following configuration was used in a cross-validation k-fold schema. The input and output layers have 19,900 features fully connected to a bottleneck of 1,000 units from the hidden layer. The probability of data corruption for the first autoencoder is set to 20% (for the binomial distribution: n=1, p=0.8). The second autoencoder maps 1,000 inputs from the output of the previous autoencoder to outputs through a hidden layer of 600 units. The second autoencoder corruption module is parameterized to corrupt a feature with a probability of 30% (for the binomial distribution: n=1, p=0.7).

Unsupervised training of autoencoders is carried out one layer at a time. To utilize the knowledge extracted with the autoencoders, we applied the encoders weights to a multilayer perceptron (MLP) with the configuration: 19,900-1,000-600-2. In other words, the MLP assumes an input space of 19,900 features and an output space of 2 numbers, explained below. Between the input and output layers, the network has two hidden layers with 1,000 and 600 units. This process is illustrated in Figure 1 and 2: the blue and green weights contain the unsupervised-trained encoders; Figure 2 contains the supervised-trained multi-layer perceptron that uses previous knowledge from autoencoder training. The MLP contains adjusted weights based on the autoencoder encoders; thus, its supervised training is called fine-tuning. The goal of fine-tuning is to adjust the MLP weights to output the expected classes and minimize prediction error on the supervised task. The output layer contains two output units: each unit rep-

[1] For further phenotypical information, refer to http://preprocessed-connectomes-project.org/abide/quality_assessment.html

[2] Mean Framewise Displacement is a measure of subject head motion, which compares the motion between the current and previous volumes.

Table 1. Phenotype summary. M: Male, F: Female.
ADOS score: † means site did not have this information.

	ASD			TC		FD
Site	Age Avg (SD)	ADOS (SD)	Count	Age Avg (SD)	Count	
CALTECH	27.4 (10.3)	13.1 (4.7)	M 15, F 4	28.0 (10.9)	M 14, F 4	0.07
CMU	26.4 (5.8)	13.1 (3.1)	M 11, F 3	26.8 (5.7)	M 10, F 3	0.29
KKI	10.0 (1.4)	12.5 (3.6)	M 16, F 4	10.0 (1.2)	M 20, F 8	0.17
LEUVEN	17.8 (5.0)	† (†)	M 26, F 3	18.2 (5.1)	M 29, F 5	0.09
MAX MUN	26.1 (14.9)	9.5 (3.6)	M 21, F 3	24.6 (8.8)	M 27, F 1	0.13
NYU	14.7 (7.1)	11.4 (4.1)	M 65, F 10	15.7 (6.2)	M 74, F 26	0.07
OHSU	11.4 (2.2)	9.2 (3.3)	M 12, F 0	10.1 (1.1)	M 14, F 0	0.10
OLIN	16.5 (3.4)	14.1 (4.1)	M 16, F 3	16.7 (3.6)	M 13, F 2	0.18
PITT	19.0 (7.3)	12.4 (3.3)	M 25, F 4	18.9 (6.6)	M 23, F 4	0.15
SBL	35.0 (10.4)	9.2 (1.7)	M 15, F 0	33.7 (6.6)	M 15, F 0	0.16
SDSU	14.7 (1.8)	11.2 (4.3)	M 13, F 1	14.2 (1.9)	M 16, F 6	0.09
STANFORD	10.0 (1.6)	11.7 (3.3)	M 15, F 4	10.0 (1.6)	M 16, F 4	0.11
TRINITY	16.8 (3.2)	10.8 (2.9)	M 22, F 0	17.1 (3.8)	M 25, F 0	0.11
UCLA	13.0 (2.5)	10.9 (3.6)	M 48, F 6	13.0 (1.9)	M 38, F 6	0.19
UM	13.2 (2.4)	† (†)	M 57, F 9	14.8 (3.6)	M 56, F 18	0.14
USM	23.5 (8.3)	13.0 (3.1)	M 46, F 0	21.3 (8.4)	M 25, F 0	0.14
YALE	12.7 (3.0)	11.0 (†)	M 20, F 8	12.7 (2.8)	M 20, F 8	0.11

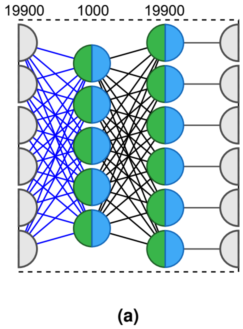

Fig. 1. Two autoencoders structure. We reduce the number of units in order to ease the visualization of the structures. (a): 19,900-1,000-19,900; (b): 1,000-600-1,000

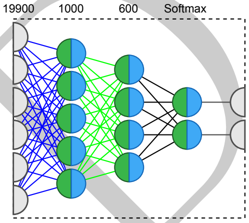

Fig. 2. Transfer learning from autoencoders AE1 and AE2 to a neural network classifier.

resents the probability of an input to be from an ASD or a TC subject. This type of output is called one-hot: during fine-tuning only one of the outputs is expected to have an activation value of 1 (and the others, 0); the output is obtained applying a softmax function. Softmax functions normalize the output distribution, so outputs denote complementary probabilities of being one class (i.e. a sum one of probabilities of being ASD or TC; for example, an output of probability of being ASD: 80%, and of being TC: 20%).

Classifier Evaluation. To evaluate the results obtained with deep learning, the performance of the model was compared with results of classifiers trained using Support Vector Machine (SVM) and Random Forest (RF). The evaluation of all models is based on a 10-fold cross-validation schema, which mixes data from all 17 sites while keeping the proportions between the different sites. Data dimensionality reduction was achieved using the encoder layers of a pre-training, unsupervised process. Table 2 summarizes the results, which we describe below. We report the accuracy, sensitivity, and specificity for all classifiers, as well as the total time taken to train each model.[3]

Results and Discussion

The deep neural network achieved a mean classification accuracy of 70% (sensitivity 74%, specificity 63%) from cross-validation folds, and a range of accuracy of 66% to 71% in individual folds. Based on the literature, this is the highest classification achieved so far. The SVM classifier achieved mean accuracy of 65% (from 62% to 72%, sensitivity 68%, specificity 62%); while the Random Forest classifier achieved mean accuracy of 63% (sensitivity 69%, specificity 58%). The results show that the deep learning algorithm classified ASD and typical participants above chance in the multi-site ABIDE data. The results also show that the algorithm outperformed the other supervised methods used for comparison. The superior accuracy of the trained model came at a substantial cost in training time, despite the use of a dedicated GPU to accelerate training. Training of the entire model took over 32 hours using an Two Intel Xeon E5-2620 processors with 24 cores running at 2GHz and 48 GB of RAM and 1 Tesla K40 GPU with 2880 CUDA cores and 12 GB of RAM.

The results show that the algorithm applied outperformed results from previous studies of identification of autism spectrum disorder patients on ABIDE multi-site resting-state brain activation. Results using other brain parcellations are shown

[3] The source code for the training scripts is available from Github at https://github.com/lsa-pucrs/acerta-abide and archived at Zenodo [37].

Method	Accuracy	Sensitivity	Specificity	Time
SVM	0.65	0.68	0.62	1m 37s
RF	0.63	0.69	0.58	20m 55s
DNN	0.70	0.74	0.63	32h 52m 36s

Table 2. Comparison of Deep Neural Network (DNN), Random Forest (RF) and Support Vector Machine (SVM) classifiers trained using 10-fold cross-validation on the entire dataset.

in the supplementary material. The results for SVM classifications did not vary by reducing the data dimensionality with autoencoders. We applied SVM on a reduced number of dimensions learned using autoencoders without the fine-tuning process. The dimensionality reduction produced lower SVM classification results (61% accuracy using data transformation with the first autoencoder, and 63% accuracy using data transformation with first and second autoencoders). The reduced dimensions may present patterns that are too complex to be generalized by SVM and autoencoders techniques to identify ASD and TC participants in the dataset. The deep learning classification method showed a 5% increase on average in classification accuracy in comparison to SVM. The deep learning method also showed a 10% increase in classification accuracy in comparison to a previous study that attempted to classify ASD using the ABIDE multi-site data [20].

There was a loss of specificity and sensitivity in the present classification in comparison to studies that attempted to classify ASD with smaller participant samples. Studies have achieved classification accuracies above 80% and even 90% (for example, [5, 21, 38]). To assess a realistic prospect of how our model would behave in the real clinical world, we calculated two metrics: positive and negative prediction values [39] (PPV and NPV, respectively). These metrics provide an evaluation of the model generalization ability [40]. The calculation is based on the relationship between sensitivity, specificity and prevalence of ASD.

The present model achieved a PPV of 4.3% and NPV of 99%. The PPV and NPV were calculated considering that the prevalence of ASD in the United States is 2.24%, according to the 2014 surveillance estimate from the Centers for Disease Control and Prevention (CDC) [41]. The high NPV is to be expected; chances are most people are not autistic. The PPV underscores that the application of machine-learning methods to brain imaging data is not driven by diagnostic purposes. Rather, it is a data-driven approach to inform what most likely are the neural patterns associated with the disorder.

Fewer site-wise variations or the absence of such variation in the dataset work in favor of classification accuracy; but once supervised methods are presented with the challenge of classification across many sites, such as ABIDE, accuracy drops. The increase in dimensions across different datasets is a challenge to be faced by ML studies of brain imaging. The dimensions may be representative of variability that adds clinically-relevant information to the understanding of a mental disorder, such as information that results from different demographics.

ABIDE data contains sensitive variations that compromise coherence between sites. Deep learning methods encompass such variations and yield better results than shallow methods. The improvement in classification can be explained by the autoencoders's potential of coping with the latent factors from intricate structures in the raw data, and by the capacity of neural networks to encode variations in data to guide the classification process. It is suggested that the deep-learning algorithms handle complexities of multi-site, big brain imaging data sets better than SVM and the like.

In order to further evaluate the results, we performed a Wilcoxon Signed Ranks related groups test for each of the classification methods. Specifically, we compared the label of each classification method to the ground truth. For the SVM classifier, results showed a statistically significant labeling ($Z = 12.08$, $p < 0.001$). RF showed a slightly improved classification accuracy ($Z = 2.33$, $p = 0.020$); the statistical differences between labels were still significant. No statistical significant difference was shown between labels when the DNN classifier was used ($Z = 0.49$, $p = 0.624$). DNN was the only classification method to show no statistical difference between the classified labels and the ground truth.

The 70% accuracy obtained in the present study improves the state of the art. The literature thus far suggests that supervised methods are effective at classifying high-dimensional spaces in smaller population samples; deep neural networks allow the learner to represent more complex functions, especially when used with autoencoders. These networks effectively reduce the dimensionality of problems with a very large feature space [24, 42]. However, by training our model with intra-site data with the same hyperparameters and 5-fold scheme, we achieved 52% average accuracy. The amount of available data in ABIDE benefits model generalization; site variability helps to avoid overfitting across sites.

Leave-one-site-out classification. To evaluate classifier performance across sites, we performed a leave-one-site-out cross validation process. This process excluded data from one site from the training process, and used that data as the test set to evaluate the model. The rationale was to test applicability of the model to new, different sites. The results of these further analyses are reported in Table 3. Results for SVM and RF are shown in Tables 1 and 2 in the supplementary material, respectively.

Five sites showed significantly lower accuracy than the global result: SBL, MAX_MUN, STANFORD, CALTECH, and OHSU. The results suggest that the data from these sites have variability that are not present in other sites. Comparison of the accuracy scores with head motion quality measures did not show an effect of head motion on classification accuracy. The classification leaving one site out tests the global model's ability to incorporate test data and site-specific variations without losing training data specificity.

Neural patterns: connectivity in the autistic brain

The results for the correlation between rs-fMRI data for areas of the brain show two distinct sets of areas that were underconnected (negatively correlated) and highly connected (positively correlated) in ASD rs-fMRI data: (1) a distributed network of anterior and posterior brain areas whose activation during rs-fMRI was negatively correlated and (2) a posterior network of areas whose activation during rs-fMRI was highly correlated. The putative interpretation of these results is discussed in relation to an existing data-driven theory of anterior-posterior underconnectivity in the autistic brain.

Table 3. Leave-site-out 5-fold cross-validation results using DNN

Site-Out	Size	Accuracy	Sensitivity	Specificity
CALTECH	37	0.68	0.70	0.65
CMU	27	0.66	0.67	0.65
KKI	48	0.67	0.70	0.64
LEUVEN	63	0.65	0.63	0.67
MAX_MUN	52	0.68	0.75	0.61
NYU	175	0.66	0.66	0.65
OHSU	26	0.64	0.70	0.59
OLIN	34	0.64	0.68	0.60
PITT	56	0.66	0.69	0.62
SBL	30	0.66	0.71	0.60
SDSU	36	0.63	0.68	0.59
STANFORD	39	0.66	0.71	0.60
TRINITY	47	0.65	0.67	0.62
UCLA	98	0.66	0.69	0.63
UM	140	0.64	0.66	0.62
USM	71	0.64	0.69	0.58
YALE	56	0.64	0.69	0.59
Mean	60	0.65	0.69	0.62

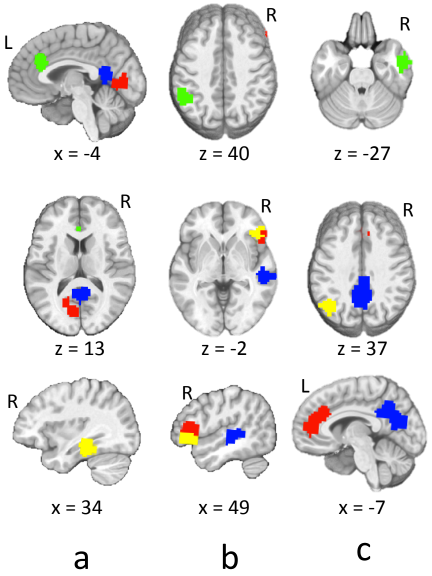

Fig. 3. Anti-correlated (underconnected) areas for ASD subjects.

The areas of the brain that showed the highest anticorrelation for ASD subjects were: Paracingulate Gyrus (Figure 3a), Supramarginal Gyrus (Figure 3b), and Middle Temporal Gyrus (Figure 3c). The anticorrelation patterns of these areas were the most relevant features for our deep learning classification. Table 4 summarizes the anticorrelated areas. The areas of the brain that showed the highest correlation for ASD subjects are shown in Figure 4. The regions with the highest correlation were all in posterior regions of the brain: Occipital Pole (Figure 4a), and Lateral Occipital Cortex; superior division (Figure 4b). The correlation patterns of these areas were the most relevant features to the deep learning classification after the anticorrelated areas. Table 5 summarizes the correlated areas.

Anterior-posterior disruption in the connectivity (correlation between time series of activation) has been shown in task-related [3, 4, 43] and rs-fMRI studies of ASD patients [44]. The characteristic of the brain function of autism patients replicated across previous studies are decreased anterior-posterior connectivity and increased local connectivity between posterior regions relative to the connectivity in the brain of controls. These studies are the basis for a brain-imaging data-driven theory of underconnectivity in autism [43]. The anterior-posterior underconnectivity theory has also been associated with indices of brain structure, more specifically, corpus callosum morphometry [45].

Previous studies of ASD brain function have suggested a disruption in anterior-posterior brain connectivity in ASD, together with increased posterior, or local, connectivity. The results for the present study suggest that there was anticorrelation in the function of anterior (paracingulate gyrus) and more posterior regions (supramarginal gyrus) and of frontal-temporal areas (e.g. middle temporal and inferior frontal; fusiform gyrus and orbital cortex) (see Table 4 for description). The interpretation we propose is that the anticorrelation reflects underconnectivity between the anterior and posterior areas of the ASD brains that contributed the most to the present classification. The trait of anterior-posterior underconnectivity underpins autistic brain function and helped discriminate between the two groups. We suggest the anterior-posterior anticorrelation result corroborates the atypical ASD brain function described in other studies.

We computed the correlation of the networks with ADOS (Autism Diagnostic Observation Schedule) results from phenotype data provided by each site to identify patterns of ASD connectivity. ADOS [46] is an assessment instrument for autism. It provides a series of social and communication tasks that are relevant to the diagnosis of ASD. The ABIDE project compiles data from 17 sites without prior coordination. Thus, ADOS data was available for 351 of the ABIDE I subjects. The analysis shows that the networks from Tables 4 and 5 did not correlate with the ADOS score.

In conclusion, the results suggest that deep learning methods may reliably classify big multi-site datasets. Classification across multiple sites has to accommodate additional sources of variance in subjects, scanning procedures and equipment in comparison to single-site datasets [20]. Such variation adds noise to the brain imaging data that challenges the ability to draw signatures from the brain activation that can classify disease states; yet the achievement of a reliable classification accuracy despite such noise generated from different equipment and demographics shows promise for machine learning applications to clinical datasets, and for future application of machine learning in the assistance of identification of mental disorders. Plitt et al. [18] stated that the overall assessment of classification of ASD using resting-state fMRI data thus far falls short of biomarker standards; such obstacle is not overcome in the present study. Yet, we suggest a step in the direction of more reliable results has been taken.

ACKNOWLEDGMENTS. The research was supported by CAPES, Brazilian Ministry of Education (Projeto ACERTA CAPES/OBEDUC 0898/2013; number 23038.002530/2013-93.

1. Varoquaux G, Thirion B (2014) How machine learning is shaping cognitive neuroimaging. *GigaScience* 3(1):28.
2. Just MA, Keller TA, Kana RK (2013) A Theory of Autism Based on Frontal-Posterior Underconnectivity in *Development and Brain Systems in Autism.* (New York: Psycology Press), pp. 35–63.

Table 4. Anti-correlated areas in the brain

Fig.	Source area (green)	Red areas	Blue areas	Yellow areas
3 a	Paracingulate Gyrus	Middle Temporal Gyrus; posterior division	Precuneous Cortex	Temporal Fusiform Cortex; posterior division
3 b	Supramarginal Gyrus	Inferior Frontal Gyrus	Superior Temporal Gyrus	Frontal Orbital Cortex
3 c	Middle Temporal Gyrus	Paracingulate Gyrus	Precuneous Cortex, Cingulate Gyrus	Lateral Occipital Cortex

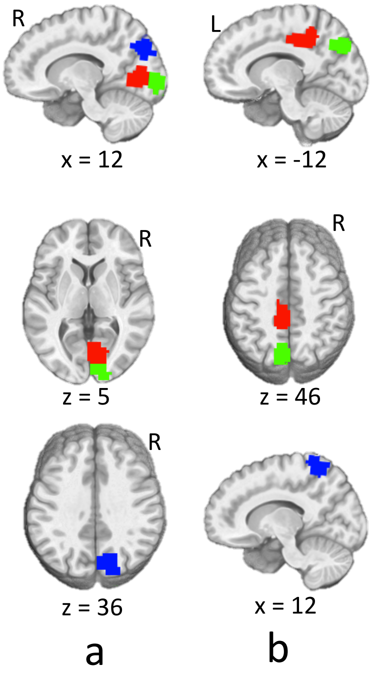

Fig. 4. Highly correlated (connected) areas for ASD subjects.

Table 5. Correlated areas in the brain

Fig.	Source area (green)	Red areas	Blue areas
4 a	Occipital Pole	Intracalcarine Cortex	Lateral Occipital Cortex; superior division
4 b	Lateral Occipital Cortex; superior division	Cingulate Gyrus; posterior division	Postcentral Gyrus

3. Kana RK, Keller Ta, Cherkassky VL, Minshew NJ, Just MA (2009) Atypical Fronto-Posterior Synchronization of Theroy of Mind Regions in Autism During Mental State Attribution. *Social neuroscience* 4(2):135–152.
4. Schipul SE, Williams DL, Keller TA, Minshew NJ, Just MA (2012) Distinctive Neural Processes during Learning in Autism. *Cerebral Cortex* 22(4):937–950.
5. Just MA, Cherkassky VL, Buchweitz A, Keller TA, Mitchell TM (2014) Identifying Autism from Neural Representations of Social Interactions: Neurocognitive Markers of Autism. *PLoS ONE* 9(12):1–22.
6. Aylward EH et al. (1999) MRI volumes of amygdala and hippocampus in non-mentally retarded autistic adolescents and adults. *Neurology* 53(9):2145–2145.
7. Jordan MI, Mitchell TM (2015) Machine learning: Trends, perspectives, and prospects. 349(6245).
8. Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: A tutorial overview. *NeuroImage* 45(1):S199–S209.
9. Haxby JV (2001) Distributed and Overlapping Representations of Faces and Objects in Ventral Temporal Cortex. *Science* 293(5539):2425–2430.
10. O'Toole AJ, Jiang F, Abdi H, Haxby JV (2005) Partially Distributed Representations of Objects and Faces in Ventral Temporal Cortex. *Journal of Cognitive Neuroscience* 17(4):580–590.
11. Buchweitz A, Shinkareva SV, Mason RA, Mitchell TM, Just MA (2012) Identifying bilingual semantic neural representations across languages. *Brain and Language* 120(3):282–289.
12. Mitchell TM et al. (2008) Predicting Human Brain Activity Associated with the Meanings of Nouns. *Science* 320(5880):1191–1195.
13. Shinkareva SV, Malave VL, Mason RA, Mitchell TM, Just MA (2011) Commonality of neural representations of words and pictures. *NeuroImage* 54(3):2418–2425.
14. Kassam KS, Markey AR, Cherkassky VL, Loewenstein G, Just MA (2013) Identifying Emotions on the Basis of Neural Activation. *PLoS ONE* 8(6):e66032.
15. Bauer AJ, Just MA (2015) Monitoring the growth of the neural representations of new animal concepts. *Human Brain Mapping* 36(8):3213–3226.
16. Yang J et al. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* 42(7):565–569.
17. Craddock RC, Holtzheimer PE, Hu XP, Mayberg HS (2009) Disease state prediction from resting state functional connectivity. *Magnetic Resonance in Medicine* 62(6):1619–1628.
18. Plitt M, Barnes KA, Martin A (2015) Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. *NeuroImage: Clinical* 7:359–366.
19. Arbabshirani MR, Plis S, Sui J, Calhoun VD (2016) Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *NeuroImage*.
20. Nielsen JA et al. (2013) Multisite functional connectivity MRI classification of autism: ABIDE results. *Frontiers in Human Neuroscience* 7(September):1–12.
21. Anderson JS et al. (2011) Functional connectivity magnetic resonance imaging classification of autism. *Brain* 134(12):3739–3751.
22. Abraham A et al. (2017) Deriving reproducible biomarkers from multi-site resting-state data: An autism-based example. *NeuroImage* 147:736 – 745.
23. Koyamada S, Shikauchi Y, Nakae K, Koyama M, Ishii S (2015) Deep learning of fMRI big data: a novel approach to subject-transfer decoding.
24. Plis SM et al. (2014) Deep learning for neuroimaging: a validation study. *Frontiers in neuroscience* 8(August):229.
25. Maaten LVD, Hinton G, van der Maaten L HG (2008) Visualizing data using t-SNE. *J. Mach. Learn. Res.* 9:2579–2605.
26. Di Martino A et al. (2014) The Autism Brain Imaging Data Exchange: Towards Large-Scale Evaluation of the Intrinsic Brain Architecture in Autism. 19(6):659–667.
27. Smith SM et al. (2009) Correspondence of the brain's functional architecture during activation and rest. *Proceedings of the National Academy of Sciences* 106(31):13040–13045.
28. Fox MD (2010) Clinical applications of resting state functional connectivity. *Frontiers in Systems Neuroscience*.
29. Shirer WR, Ryali S, Rykhlevskaia E, Menon V, Greicius MD (2012) Decoding Subject-Driven Cognitive States with Whole-Brain Connectivity Patterns. *Cerebral Cortex* 22(1):158–165.
30. Franco AR, Mannell MV, Calhoun VD, Mayer AR (2013) Impact of Analysis Methods on the Reproducibility and Reliability of Resting-State Networks. *Brain Connectivity* 3(4):363–374.
31. Shehzad Z et al. (2009) The Resting Brain: Unconstrained yet Reliable. *Cerebral Cortex* 19(10):2209–2229.
32. Biswal B, Yetkin FZ, Haughton VM, Hyde JS (1995) Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic resonance in medicine* 34(4):537–541.
33. Behzadi Y, Restom K, Liau J, Liu TT (2007) A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage* 37(1):90–101.
34. Craddock RC, James G (2012) A whole brain fMRI atlas spatial Generated via Spatially Constrained Spectral Clustering. *Human brain . . .* 33(8).
35. Vincent P, Larochelle H, Bengio Y, Manzagol PA (2008) Extracting and composing robust features with denoising autoencoders. *Proceedings of the 25th international Conference on Machine learning* pp. 1096–1103.
36. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA (2010) Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *Journal of Machine Learning Research* 11(3):3371–3408.
37. Heinsfeld AS, Franco A, Buchweitz A, Meneguzzi F (2017) lsa-pucrs/acerta-abide: Code companion to Neuroimage: Clinical submission.
38. Uddin LQ, Supekar K, Menon V (2013) Reconceptualizing functional brain connectivity in autism from a developmental perspective. *Frontiers in Human Neuroscience* 7.
39. Altman DG, Bland JM (1994) Diagnostic tests 2: Predictive values. *BMJ* 309(6947):102.
40. Castellanos FX, Di Martino A, Craddock RC, Mehta AD, Milham MP (2013) Clinical applications of the functional connectome. *NeuroImage* 80:527–540.
41. Zablotsky B, Black LI, Maenner MJ, Schieve LA, Blumberg SJ (2015) Estimated Prevalence of Autism and Other Developmental Disabilities Following Questionnaire Changes in the 2014 National Health Interview Survey. *National health statistics reports* (87):1–20.

42. Hjelm RD et al. (2014) Restricted Boltzmann machines for neuroimaging: An application in identifying intrinsic networks. *NeuroImage* 96:245–260.
43. Just MA (2004) Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity. *Brain* 127(8):1811–1821.
44. Cherkassky VL, Kana RK, Keller TA, Just MA (2006) Functional connectivity in a baseline resting-state network in autism. *NeuroReport* 17(16):1687–1690.
45. Just MA, Cherkassky VL, Keller TA, Kana RK, Minshew NJ (2006) Functional and Anatomical Cortical Underconnectivity in Autism: Evidence from an fMRI Study of an Executive Function Task and Corpus Callosum Morphometry. *Cerebral Cortex* 17(4):951–961.
46. Akshoomoff N, Corsello C, Schmidt H (2006) The Role of the Autism Diagnostic Observation Schedule in the Assessment of Autism Spectrum Disorders in School and Community Settings. *The California School Psychologist* 11(1):7–19.
47. Vapnik V (1998) The support vector method of function estimation in *Nonlinear Modeling*. (Springer), pp. 55–85.
48. Ho TK (1995) Random decision forests in *Document Analysis and Recognition, 1995., Proceedings of the Third International Conference on*. (IEEE), Vol. 1, pp. 278–282.
49. Tzourio-Mazoyer N et al. (2002) Automated anatomical labeling of activations in {SPM} using a macroscopic anatomical parcellation of the {MNI} {MRI} single-subject brain. *NeuroImage* 15(1):273 – 289.
50. Dosenbach NUF et al. (2010) Prediction of individual brain maturity using fmri. *Science* 329(5997):1358–1361.
51. Maaten Lvd, Hinton G (2008) Visualizing data using t-sne. *Journal of Machine Learning Research* 9(Nov):2579–2605.