Functional SPECT neuroimaging using machine learning algorithms distinguishes autism spectrum disorder from healthy subjects

Daniel G. Amen1*, Mona Sharifi Sarabi2, Kristen Willeumier1, Derek Taylor1, Cyrus Raji3, Somayeh Meysami1 and Cauligi Raghavendra2

1Department of Neuroimaging and Clinical Research, Amen Clinics, Inc., 3150 Bristol, Suite 400, Costa Mesa, CA, 92626, USA
2Ming Hsieh Department of Electrical Engineering, University of Southern California, Los Angeles, CA, 90089-2562, USA
3UCLA Medical Center, Los Angeles, CA, 90095, USA

Abstract

The diagnosis of Autism Spectrum Disorder (ASD) relies on history and behavioral observation, lacking reliable biomarkers. We performed a retrospective analysis using machine learning algorithms of 928 persons with ASD (mean age: 17 ± 10.8 years; age range 4-67) obtained from a multisite psychiatric database with rest and on-task brain SPECT scans to investigate whether or not these scans distinguish ASD from healthy controls (HC, n=101; mean age: 43 ± 17.2 years; age range 13-84). Using 128 regions of interest extracts (ROIs), we applied multiple machine learning algorithms for binary classification. Due to an unbalanced sample size between ASD and controls, we then sub-sampled the data prior to feature selection and classification. Using a subsampled dataset, least absolute shrinkage and selection operator (LASSO) feature selection with Random Forest classification has been used to accurately diagnose and characterize neurodevelopmental disorders such as ASD adults only, the majority of our sample, and selected subjects in both AD and HC groups with age range of 13-67 years and found the results consistent with the combined data. These machine learning results identified potential diagnostic biomarkers differentiating ASD from HC in the regions of the cerebellum and vermis, anterior cingulate gyrus, amygdala, thalamus, frontal, and temporal lobes.

Introduction

Autism spectrum disorder (ASD) is a neurodevelopmental disorder with a prevalence rate of 1:68 [1] and is characterized by impairments across multiple domains including socio-communication ability, and restrictive and stereotyped behaviors [2]. The high heterogeneity and complexity of ASD has previously limited the capacity of neuroimaging to produce reliable and consistent biomarkers that can be applied in a standard clinical evaluation. Currently, the diagnosis of ASD includes a clinical history, physical examination, and structured screening tools including the Autism Diagnostic Interview-Revised [3] or Autism Diagnostic Observation Schedule [4,5]. As autism often involves impairments across multiple domains that affect both social and intellectual function, the time from initial clinical visit to diagnosis lasts up to 13 months [6]. Given the increasing prevalence rate of autism, there is a need to develop rapid and reliable detection tools.

The application of machine learning algorithms to neuroimaging data offers the potential to improve the precision of the diagnosis through the identification of brain based biomarkers in ASD [7]. These applications involve improving diagnostic capabilities, targeting interventions and monitoring patient outcomes [7,8]. Multivariate Analysis has shown promising clinical applicability with regards to diagnosing and characterizing neurodevelopmental disorders such as ASD [8]. Random Forest classification has been used to accurately assess differences in white matter connectivity in infants with ASD, offering the potential to assist in early diagnosis and intervention [9]. Machine learning classification has also been demonstrated to show predictive ability with regards to outcome measures for ASD, including longitudinal change in autistic traits as measured through functional network connectivity studies [10].

Machine learning classification has shown utility in identifying regional brain abnormalities in those with ASD from typical developing controls, providing the basis for candidate biomarkers. Pattern classification methods demonstrate a high predictive ability for the detection of ASD and have been performed across a variety of imaging modalities. These include structural MRI [11-13], intrinsic functional connectivity MRI (fcMRI) [14-16], diffusion tensor imaging (DTI) [17,18] and electroencephalography (EEG) [19,20]. Reported ASD abnormalities have been identified in the cerebellum [21] and cerebellar vermis [22], anterior cingulate gyrus [23], amygdala [24-26], hippocampus [24,27,28], and areas of the frontal [29-31], temporal [29,32] parietal lobes [30], caudate and putamen [29,33]. Additional imaging abnormalities in autism include impaired brain growth [24], cortical thickness [12,34], alterations in white matter architecture [18,35], and aberrant connectivity within the somatosensory, visual and default mode network [36].

Given the advances that have been made with machine learning techniques and ASD, there is a need to continue developing objective methods for detection of ASD within the research community, and to identify diagnostic markers of the disorder. Single photon emission computed technology (SPECT) is an imaging modality that has been

Correspondence to: Daniel G. Amen, MD. Amen Clinics at Costa Mesa, CA. daniel@amenclinics.com

Key words: Neuroimaging, autism spectrum disorder, single photon emission computed tomography, biomarkers, machine learning, and classification.

Received: March 20, 2017; Accepted: April 07, 2017; Published: April 10, 2017
applied to better understand the neuropathology of ASD; an excellent systematic review of this work can be read in Zurcher et al. [37]. Since 1992, there have been many studies using SPECT and PET imaging, totaling over 900 ASD patients. The findings have primarily shown lower perfusion in the prefrontal cortex, temporal lobes, parietal lobes and cerebellum, compared to controls [38-43]. However, most of the individual studies had small sample sizes and none closely examined the potential sensitivity and specificity of using SPECT or PET as a diagnostic tool to distinguish between healthy and non-ASD patient controls. In the present study, we implemented multiple machine learning algorithms to brain SPECT images acquired at rest and on-task to a large cohort of ASD subjects in order to evaluate the ability to predict ASD from HC using region of interest extracts. We then use feature selection methods like least absolute shrinkage and selection operator (LASSO) and minimum redundancy maximum relevance (mRMR) to identify the key features which might serve as biomarkers in delineating between ASD and controls.

Methods

Study subjects

A sample of persons with autism (n = 928; mean age: 17 ± 10.8 years; age range: 4-67; 4:1 male to female ratio) were obtained from a large, multisite, clinical psychiatric database comprised of 27,756 patients at the Amen Clinics. All subjects in the database were evaluated at one of nine outpatient branches of the Amen Clinics (Newport Beach, Costa Mesa, Fairfield, and Brisbane, CA; Tacoma and Bellevue, WA; Reston, VA; Atlanta, GA; and New York, NY) from 1995-2015. Each participant had rest and concentration SPECT scans as part of their evaluation. An Autism diagnosis was established by a board certified or board eligible psychiatrist, using a detailed clinical history, mental status examination, and DSM-IV or DSM-V criteria. The Institutional Review Board function was conducted by an External Review Organization (IntegReview) that approved a retrospective review of anonymous data which was used by the researchers for this study (IRB #004).

A group (n = 101; mean age: 43 ± 17.2 years; age range: 13-84, 7:1 male to female ratio) of age and gender matched controls with rest and on-task SPECT studies were included in this study. Exclusion criteria for the healthy subjects were: 1) current or past evidence of psychiatric illnesses as determined by a detailed clinical history, mental status examinations, and the Structured Clinical Interview for Diagnosis for DSM-IV (SCID-IV); 2) current reported medical illnesses or medication; 3) history of brain trauma; 4) current or past drug and alcohol abuse; and 5) family history of a first degree relative with a psychiatric illness.

Brain SPECT Imaging

SPECT scans of the brain were obtained in conjunction with clinical assessments before any intervening treatment. Brain SPECT imaging is performed as described below and is standardized across all Amen Clinics [44,45]. For each SPECT scan procedure, the patient initially has an intravenous catheter placed in their arm. The subject then rests comfortably for approximately 15 minutes at which point an age- and weight-appropriate dose of technetium Tc99m exametazime is administered while the subject remains at rest in the uptake room. For the rest scans, after injection, patients sit in a dimly lit room with eyes open and low ambient noise. Approximately 30 minutes after the injection, subjects are scanned for 30 minutes. Subjects return on a separate day to undergo a second SPECT scan in which the subject is injected while performing the Conner’s Continuous Performance Test (C-CPT). For the on-task scans, patients are injected three minutes after starting the C-CPT and then perform the task for another 10 minutes. Patients are then scanned approximately 30 minutes after injection. The SPECT images are captured using a high-resolution Picker (Phillips) Prism XP 3000 triple-headed gamma camera with fan beam collimators with data collected in 128x128 matrices, yielding 120 images per scan with each image separated by three degrees spanning 360 degrees. First level processing of images from the raw data are performed using the software package Odyssey FX V8.90 developed by Picker Image System for Phillips/GE SPECT cameras. A low pass filter (Butterworth filter) is applied with a high cutoff (0.25 cycles per pixel). Chang attenuation correction is then performed [46]. Transaxial slices oriented horizontal to the AC-PC line are created along with coronal and sagittal images (6.6mm apart, unsmoothed). Three dimensional reformats are generated for review based on Odyssey image visualization software.

Data are exported as interfiles and converted to the NIMH Neuroimaging Informatics Technology Initiative (NiHTI) format using an ROI package developed at Amen Clinics that transforms the stereotaxic atlas into the individual scan space; further processing to region of interest values (ROI) is completed by the same package. Bilateral ROI counts are derived from the anatomical regions in the Automated Anatomical Labeling (AAL) atlas [47]. The AAL atlas consists of 128 brain regions defined across both hemispheres. ROI metrics included mean, standard deviation, minimum, maximum, 5th percentile histograms, largest maximum valued connected cluster after thresholding, and largest minimal valued connected cluster after thresholding. To account for outliers, T-score derived ROI count measurements are derived using trimmed means [48] that are calculated using all scores within the 98% confidence interval (−2.58 < Z < 2.58) for a particular scanner in the year the patient was scanned, thus correcting for any calibration differences between scanners. The ROI mean for each subject and the trimmed mean for the sample are used to calculate T with the following formula: $T = 10^k((\text{subject ROI-mean}) - [\text{trimmed Scanner-avg.}]) / [\text{trimmed scanner-st.dev.}]) + 50$. The values for both SPECT scans for each patient then become inputs for the machine learning algorithms. Figure 1 shows examples of how ROI data are outlined on SPECT images as demonstrated on a standard template brain.

The ROI data from SPECT images on patients who have been diagnosed with ASD and healthy controls aids in providing ground truth for machine learning algorithms. In addition, we also utilized data on other independent variables including age, gender, and other questions completed by all subjects analyzed in this study. In all, this aggregation of data results in 384 variables that are based on ROIs of the SPECT scans and another 10 variables from clinical questionnaires. Given the focus on classification of ASD from healthy controls based on ROI region attributes and finding the decision rules, we used 256 ROI measurements (128 rest and 128 on-task), as well as derived measures of activation (rest – on task) as input features in all analyses. We also control for age in our analyses.

Machine Learning Algorithms

Alterations in cerebral blood flow affects underlying brain functions. There are brain atlases that divide the brain into regions of interest (ROI), and the amount of blood flow in these ROIs are critical to brain function [49]. In the atlas used for ROI measurements on SPECT in this study, the Automated Anatomical Labeling (AAL) atlas, there are 128 regions defined in each brain hemisphere, and the blood
checked the data with respect to all the attributes of interest to process variables and we prepared the data to handle all of these situations. We some outliers, missing data, and inconsistencies in the values of these scanned in the same year. As with any real-world data we can expect as R-values or T-values that are relative to the mean of all other subjects.

**Data preparation**

For all the subjects in this study, we have SPECT data scaled either to healthy controls using SPM (p < 0.001 FDR Correction): 1) Occipital Superior Left; 2) Post Central Right; 3) Cingulum Anterior Right; 4) Precentral Left; 5) Cingulum Mid Right; 6) Temporal Mid Right; 7) Angular Superior Right; 8) Temporal Superior Left; 9) Frontal Superior Left; 10) Frontal Inferior Triangularis Right; 11) Frontal Superior Orbital Left; 12) Temporal Mid Left; 13) Rolandic Operculum Right; 14) Frontal Inferior Right; 15) Frontal Inferior Left; 16) Temporal Mid Left; 17) Frontal Inferior Orbital Left; 18) Temporal Inferior Left.

For inconsistent values, outliers, and redundant data records. In this study, since we have a large dataset, we removed all cases with missing values and did not apply any imputation methods in order to achieve baseline accuracies.

Since this dataset lacks the necessary number of subjects, compared to the number of attributes used (typically one requires samples at least 10 times the number of attributes) [50], we applied techniques to reduce the number of features/attributes used without losing too much information present in the ROI data. Specifically, we utilized LASSO to select the top few ROIs as features for our analyses. LASSO is a regression analysis approach that performs feature selection in order to improve prediction accuracy [51]. LASSO, known as embedded feature selection or regularization methods, are also called penalization methods that introduce additional constraints into the optimization of a predictive algorithm.

We compared the performance of LASSO with mRMR feature selection method and also feature extraction methods like Principle Component Analysis (PCA) that transforms the data in the high-dimensional space to a space of fewer dimensions. PCA performs a linear mapping of the data to a lower-dimensional space in such a way that the variance of the data in the low-dimensional representation is maximized.

**Training and testing dataset**

A key objective in applying machine learning algorithms in this study was to build a suitable training group from our dataset. Our dataset was comprised of 928 ASD subjects and 101 HC subjects which resulted in unbalanced data with the distribution shown in Figures 2 and 3 for both male and female subjects, therefore care should be taken in selecting appropriate training sets. We divided our patient data into two sets: one for training and the other for validation and testing. The training set should be large enough for the machine learning models to learn all the interdependencies among the variables. We carefully selected the appropriate size for training data to include the proper ratios of patients with respect to age, gender, and other relevant information. Construction of the proper training set takes time and entails an iterative process with feedback from experts. A number of experiments were conducted with different sets of features with our selected machine learning algorithms to finalize the training set and features.

To ensure that the model building is complete with each selected machine learning algorithm, we used 5-fold cross validation and
Equivalently, we can use a smaller mRMR feature set to effectively cover the results with known labels (ground truth). In this cross validation phase, we divided the training set into 5 parts and iteratively used 4 out of 5 parts for training and then tested the models on the 5th part. The results of specificity and sensitivity for all 5 experiments were averaged to see how robust the current trained model is. This process is repeated by changing the training set and conducting experiments with different sets of features to achieve the best possible result. The details of our methodology are shown in Figure 3 and in the heuristic algorithm. The age distribution of sub-sampled datasets for stage (3) in Figure 3 is shown in Figures 5-7.

### Results

**Autism versus healthy controls**

We conducted two sets of experiments – the first experiment is with subsampled data matching by age ranges and the second experiment is extremely matched by age and gender. In the first experiment we used the subsampling approach and obtained a subsample which has 168 instances. This dataset was achieved by removing all samples with missing values, mainly in ROI features and also limiting the age range of samples to those between 13-67 years old. In experiment 2, we applied the same sub-sampling method to extremely matched subjects with respect to age and gender. This dataset has 108 samples with 48 subjects of ASD and 48 subjects of HC group. The details of distribution of subjects based on age and gender are shown in Figures 3-7.

### Experiment 1: Subsampling by age matching

In this experiment, we ran multiple classifiers on the dataset of 168 samples (84 healthy samples and 84 autism samples). We follow the steps of this framework to perform feature selection and model generation. For feature selection stage, we compared multiple approaches like LASSO, mRMR and PCA and achieved the most promising results by features selected with LASSO.

mRMR provides the feature set by requiring that features are maximally dissimilar to one another, for example, their mutual Euclidean distances are maximized, or their pair-wise correlations are minimized. These minimum redundancy criteria are supplemented by the usual maximum relevance criteria such as maximal mutual information with the target phenotypes. The benefits of this approach can be realized in two ways. (1) With the same number of features, we expect the mRMR feature set to be more representative of the target phenotypes, therefore leading to a better generalization property. (2) Equivalently, we can use a smaller mRMR feature set to effectively cover the same space as a larger conventional feature set [52]. The mRMR method is designed to work with discrete data and since we have continuous values in our dataset, before applying this method we need to discretize the data into multiple bins. This step is implemented by using Matlab built-in functions and we also use the available implementation of mRMR by the inventors of this method.

Lasso is a feature selection and also a regularization method that was originally introduced in the context of least squares. This method, the sum of the absolute value of the regression coefficients are forced to be less than a fixed threshold value, which consequently leads to setting certain coefficients to zero. This approach is motivated by ridge regression, in which the coefficients sum of the squares is forced to be less than a fixed threshold value. Ridge regression shrinks the size of the large regression coefficients to reduce overfitting, but doesn’t perform covariate selection which consequently will not be able to build more interpretable models. Compared to mRMR, LASSO gave us more informative features in this experiment.

We also evaluated our designed model by the abstract features implemented by the inventors of this method.

We also evaluated our designed model by the abstract features implemented by the inventors of this method.
that has 28 female subjects in each of the HC and ASD samples and 26 male subjects in each group that make a dataset of 108 subjects in total. This dataset classification performance is evaluated by the same classifiers in experiment 1. The classification performance results are shown in Tables 2A,2B. Table 3 lists the top 12 ROI features in distinguishing ASD from controls:

Discussion

In the present study, ROI data from rest and on-task SPECT images were used to inform classification of ASD from HC using machine learning algorithms, attaining a maximum classification accuracy of 74.4% with Random Forest using subsampling methods with the top 12 features. This algorithm performed with a high level of accuracy on a robust dataset comprised of 168 with 84 ASD individuals and 84 HC, with a sensitivity and specificity respectively of 72% and 76% using only ROI features and 78% and 83% sensitivity and specificity by utilizing ROI features and age attributes.

Implementing several machine learning tools with varied feature sizes, we discovered that several algorithms reached an accuracy level to provide reliable diagnostic utility. SVM and Logistic Regression was also tested on the subsampled dataset with ROI regions and age, and resulted in accuracies of >76%. This work has not previously been performed with brain SPECT imaging data, so these results contribute to the growing body of literature in the neuroimaging field that have utilized pattern recognition with machine learning classification to improve distinguishing autism from typical developing subjects. Other methods include functional connectivity MRI [10,36,53-58], voxel based morphology [59], EEG [19,20,60,61] and diffusion tensor imaging studies [18].

Table 1. Comparison of the classification performance between machine learning algorithms based on optimal classifier settings. Table 1A. Classification results of SVM, LR and RF based on based on ROI features. Table 1B. Classification results of SVM, LR and RF based on based on ROI+Age features

<table>
<thead>
<tr>
<th>Specificity</th>
<th>Sensitivity</th>
<th>AUC</th>
<th>Accuracy</th>
<th>Error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.693</td>
<td>0.552</td>
<td>0.754</td>
<td>0.636</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>0.686</td>
<td>0.769</td>
<td>0.792</td>
<td>0.732</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.724</td>
<td>0.760</td>
<td>0.817</td>
<td>0.744</td>
</tr>
</tbody>
</table>

Table 1B.

<table>
<thead>
<tr>
<th>Specificity</th>
<th>Sensitivity</th>
<th>AUC</th>
<th>Accuracy</th>
<th>Error rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM</td>
<td>0.838</td>
<td>0.677</td>
<td>0.828</td>
<td>0.762</td>
</tr>
<tr>
<td>Logistic Regression</td>
<td>0.777</td>
<td>0.743</td>
<td>0.805</td>
<td>0.762</td>
</tr>
<tr>
<td>Random Forest</td>
<td>0.788</td>
<td>0.826</td>
<td>0.894</td>
<td>0.809</td>
</tr>
</tbody>
</table>

Experiment 2: Subsampling by extreme subject matching based on age and gender

In experiment 2 we followed the same steps as in experiment 1 and simply replaced the input dataset with the extreme subsamples dataset generated by PCA that leads to lower classification performance, compared to LASSO. Since we are mostly interested in finding highly correlated features representing the blood flow of 128 ROI regions, we stick to feature selection methods that promote sparsity as discussed earlier.

In this study we applied different approaches to find the best classification model that separates ASD patients from HC. We experimented with several well-known machine learning algorithms, including Random Forest (RF), Logistic Regression (LR) and Support Vector Machine (SVM). For all our experiments we used the scikit-learn package that implements all these algorithms (http://scikit-learn.org/stable/index.html)

Figures 8 and 9 show receiver operating characteristic (ROC) plots which is a graphical plot that illustrates the performance of a binary classifier system and provides a visual overview for comparing the performance of classifiers. Table 1 shows the average specificity, sensitivity and accuracy across classifiers and demonstrates that the best classifier for both experiments is Random Forest.

In experiment 2 we followed the same steps as in experiment 1 and simply replaced the input dataset with the extreme subsamples dataset generated by PCA that leads to lower classification performance, compared to LASSO. Since we are mostly interested in finding highly correlated features representing the blood flow of 128 ROI regions, we stick to feature selection methods that promote sparsity as discussed earlier.
Amen DG (2017) Functional SPECT neuroimaging using machine learning algorithms distinguishes autism spectrum disorder from healthy subjects

**Table 3. Top 12 Regions in Distinguishing ASD from Controls**

<table>
<thead>
<tr>
<th>Region Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>T_Baseline_Cerebellum_7b_L</td>
</tr>
<tr>
<td>T_Baseline_Frontal_Mid_L</td>
</tr>
<tr>
<td>T_Consentration_Cerebellum_3_L</td>
</tr>
<tr>
<td>T_Consentration_Cingulum_Ant_L</td>
</tr>
<tr>
<td>T_Concentration_Vermis_10</td>
</tr>
<tr>
<td>T_Baseline_Amygdala_R</td>
</tr>
<tr>
<td>T_Consentration_Heschl_L</td>
</tr>
<tr>
<td>T_Consentration_Cerebellum_10_L</td>
</tr>
<tr>
<td>T_Baseline_Temporal_Inf_Ant_L</td>
</tr>
<tr>
<td>T_Baseline_Frontal_Sup_L</td>
</tr>
<tr>
<td>T_Baseline_Thalamus_L</td>
</tr>
<tr>
<td>T_Concentration_Frontal_Sup_Medial_R</td>
</tr>
</tbody>
</table>

Among the top most informative features identified using LASSO, we found involvement of regions implicated in ASD pathology, including regions of the cerebellum and vermis, anterior cingulate gyrus, amygdala, thalamus, frontal, and temporal lobes. The fact that both concentration and baseline SPECT scan ROIs were important in delineating between ASD and controls suggests that both scans may be useful in a clinical setting.

These regions are correlated with the core symptom domains of ASD including socio-communication deficits, impairments in cognition, language and repetitive/stereotyped movement and are consistent with previously described neuroimaging abnormalities. One of the most notable neuropathological hallmarks of ASD includes dysfunction of the cerebellar vermis, cortices and cerebro-cerebellar circuits [22,62,21]. This finding is consistent with structural and functional studies in MRI and SPECT, demonstrating the prevalence of cerebellar dysfunction in ASD. Such studies also demonstrated behavioral impairments with regard to motor control, language, social control, affective expression and exploratory attention [63]. The left anterior inferior temporal lobe was also important in our study in distinguishing between the two groups. Prior literature has shown the right temporal parietal junction is an associative region and has been shown to respond atypically in those with autism spectrum conditions [64]. This region has been implicated in the understanding of the "theory of mind," [65] language, spatial cognition and attention that is implicated in the social-communication deficits observed in ASD. Furthermore, the amygdala was identified as a distinguishing feature. Limbic areas such as the amygdala are involved in emotional perception and regulation, and have been found to be enlarged in autistics [66]. The amygdala has been implicated in the socio-emotional impairments observed in autism, due to its widespread connectivity to cortical and subcortical structures [67]. Additional features identified include the frontal lobes, thalamus and posterior cingulate cortex which have been previously reported to show perfusion abnormalities with SPECT [41,42,68] and to correlate with the attentional dysfunction, the inability to modulate social behaviors, language and emotional processing deficits and the inability to perceive emotional expressions in autism. The fact that the left transverse temporal gyrus of Heschl on concentration SPECT was a distinguishing feature is intriguing, given the impaired auditory processing a prior MEG study found in autistic children [69]. Additional left-sided regions identified in our analysis may also related not only to the language functions of that hemisphere, but may also reflect the fact that most subjects in our sample were right-handed, thus denoting a left hemisphere dominance.

As ASD is increasingly prevalent and diagnosed in children, we wanted to ensure that age would not be a discriminating factor, therefore we applied our algorithms only to adults with ASD and the HC subjects and to an age matched range of 18-40 in both groups. In both of these experiments, our results were approximately equivalent when we used all the data in both groups.

There were several methodological challenges to address with this dataset to ensure best practices were followed for computational analysis using machine learning algorithms. Our first objective in designing this study was to obtain a large dataset of good quality with enough participants from each group for the model to learn from the training data. We then recognized that the high number of ROIs from the imaging data produced a dimensionality problem. We reduced the
dimensionality by identifying the most important attributes that could reliably separate the two classes. We then had to address the imbalance in the number of subjects in the two classes by identifying which features to use and testing a number of experimental models with different algorithms to obtain the reported classification results. We evaluated each model carefully by setting aside some data for testing to avoid overfitting and to ensure generalizability of the model. And finally, further validation of the classifier was performed by testing our machine learning models on an independent test set containing both ASD and HC subjects. We performed these steps to assess the reliability of the data, because when predictive models show interpretable features that are reliable it offers the potential for diagnostic utility in a clinical setting.

This study had several strengths but also included limitations which need to be addressed for future studies. The strengths included neuroimaging data that was acquired both at rest and on-task from a well-validated and widely available functional imaging modality. This work was performed on a large sample size of ASD subjects obtained across multiple sites, and was measured with quantitative analysis using machine learning algorithms in which the classifier was validated on an independent data set. The first limitation was a retrospective analysis of existing data, so we did not have information on IQ or language ability to assess language delay. Therefore, this data must be interpreted across a broad range of intelligence and communication abilities. As the autism spectrum includes a wide range of intellectual ability, future work will include assessing IQ in order to perform more focused studies on differentiating low functioning, versus high functioning individuals from the HC group. Second, while diagnosis of ASD was established by meeting the DSM-IV criteria for autism as assessed by an expert clinician, inclusion of the Autism Diagnostic Interview-Revised or Autism Diagnostic Observation Schedule would address the level of functionality of the patients. And finally, future work will not only include the HC group, but will also include individuals with mood disorders including attention deficit hyperactivity disorder (ADHD), anxiety and obsessive compulsive disorder. Studies with a HC allow for understanding the neurobiological underpinnings of the ASD, but symptoms of autism and psychiatric disorders such as ADHD often co-occur [70], and it will be important to explore how machine learning approaches can be used to understand the neurobiological differences between autism and co-morbid disorders, as this is what occurs in a real-world clinical setting.

The clinical relevance of this work is its identification of specific brain regions on perfusion SPECT neuroimaging that diagnostically distinguish autism from controls. Because these findings were obtained from a large database, there exists a foundation for development of computer algorithms for individualized prediction of diagnosis. For example, a single patient scan can be inputted into a machine learning program and then matched to a brain SPECT scan database of pre-existing controls and patient scans such as with autism. The patient’s scan could then be matched to either a patient or control group based on quantified imaging characteristics in combinations of regions specific for various disorders from autism to Alzheimer’s to TBI.

To our knowledge this is the first brain SPECT imaging study demonstrating the use of machine learning methods to predict ASD from a HC. These results add to the growing body of literature validating the use of machine learning approaches with functional neuroimaging data to improve prediction and classification of individuals with psychiatric disorders like autism. Given the heterogeneity of ASD, this approach has important implications in the clinical setting in both the diagnosis, intervention and monitoring of treatment outcomes.

References

18. Ingahlulikar M, Parker D, Bloy L, Roberts TP, Verma R (2011) Diffusion based abnormality markers of pathology: toward learned diagnostic prediction of ASD. *Neuroimage* 57: 918-927. [Crossref]
Amen DG (2017) Functional SPECT neuroimaging using machine learning algorithms distinguishes autism spectrum disorder from healthy subjects


61. Trevathan C, Delfafild-Butt JT (2013) Autism as a developmental disorder in intentional movement and affective engagement. *Front Integr Neurosci* 7:49. [Crossref]


63. Li W, Mai X, Liu C (2014) The default mode network and social understanding of others: what do brain connectivity studies tell us. *Front Hum Neurosci* 8: 74. [Crossref]


70. Leitner Y (2014) The co-occurrence of autism and attention deficit hyperactivity disorder in children - what do we know? *Front Hum Neurosci* 8: 268. [Crossref]