White matter connectome edge density in children with Autism Spectrum Disorders: potential imaging biomarkers using machine learning models

Article

Accepted Version

White Matter Connectome Edge Density in Children with Autism Spectrum Disorders: Potential Imaging Biomarkers Using Machine Learning Models

Seyedmehdi Payabvash,[1] Eva M. Palacios,[1] Julia P. Owen,[1,2] Maxwell B. Wang,[1,3]

Teresa Tavassoli,[4] Molly Gerdes,[4] Anne Brandes Aitken,[4] Daniel Cuneo,[1]

Elysa J. Marco, [4,5,6,]* Pratik Mukherjee.[1,7,]*

[1] Department of Radiology and Biomedical Imaging, University of California, San Francisco, CA
[2] Department of Radiology, University of Washington, Seattle, WA
[3] University of Pittsburg School of Medicine, Pittsburgh, PA
[4] Department of Neurology, University of California, San Francisco, CA
[5] Department of Psychiatry, University of California, San Francisco, CA
[6] Department of Pediatrics, University of California, San Francisco, CA
[7] Department of Bioengineering and Therapeutic Sciences, University of California, San Francisco, CA

* Dr. Elysa J. Marco and Dr. Pratik Mukherjee contributed equally to this work.

Running title: Connectome Edge Density for Autism Diagnosis

Correspondence:
Pratik Mukherjee, MD PhD
Center for Molecular and Functional Imaging,
Department of Radiology and Biomedical Imaging,
University of California, San Francisco
UCSF Box 0946, 185 Berry Street, Suite 350,
San Francisco, CA 94107, USA
Tel.: 415 514 8186
Fax: 415 353 8593
Email: Pratik.Mukherjee@ucsf.edu

Key words: Machine learning, edge density imaging, diffusion tensor imaging, autism

Abstract

Prior neuroimaging studies have reported white matter network underconnectivity as a potential mechanism for Autism Spectrum Disorder (ASD). In this study, we examined the structural connectome of children with ASD using Edge Density Imaging (EDI);

and then applied machine leaning algorithms to identify children with ASD based on tract-based connectivity metrics. Boys aged 8 to 12 years were included: 14 with ASD and 33 typically developing children (TDC). The Edge Density (ED) maps were computed from probabilistic streamline tractography applied to high angular resolution diffusion imaging (HARDI). Tract-Based Spatial Statistics (TBSS) was used for voxel-wise comparison and coregistration of ED maps in addition to conventional DTI metrics of Fractional Anisotropy (FA), Mean Diffusivity (MD), and Radial Diffusivity (RD). Tract-based average DTI/connectome metricswere calculated and used as input for different machine learning models: naïve Bayes, random forest, support vector machines (SVM), neural networks. For these models, cross-validation was performed with stratified random sampling (×1000 permutations). The average accuracy among validation samples was calculated. In voxel-wise analysis, the body and splenium of corpus callosum, bilateral superior and posterior corona radiata, and left superior longitudinal fasciculus showed significantly lower ED in children with ASD; whereas, we could not find significant difference in FA, MD, and RD maps between the two study groups. Overall, machine-learning models using tract-based ED metrics had better performance in identification of children with ASD compared to those using FA, MD, and RD. The EDI-based random forest models had greater average accuracy (75.3%), specificity (97.0%), and positive predictive value (81.5%), whereas EDI-based polynomial SVM had greater sensitivity (51.4%), and negative predictive values (77.7%). In conclusion, we found reduced density of connectome edges in the posterior white matter tracts of children with ASD; and demonstrated the feasibility of connectome-based machine-learning algorithms in identification of children with ASD.

Acronyms

ADIR= Autism Diagnostic Interview-Revised; ADOS= Autism Diagnostic Observation Schedule; ASD = Autism Spectrum Disorder; AUC= Area Under the Curve; BEDPOSTX=

Bayesian Estimation of Diffusion Parameters Obtained using Sampling Techniques; DTI= Diffusion Tensor Imaging; ED= Edge Density; EDI= Edge Density Imaging; FA= Fractional Anisotropy; FMRIB= Functional Magnetic Resonance Imaging of the Brain; FSL= FMRIB Software Library; GLM= General linear model; HARDI= High Angular Resolution Diffusion Imaging; MD= Mean Diffusivity; NPV= negative predictive value; PPV= positive predictive value; RD= Radial diffusivity; ROC= Receiver Operating Characteristics; TBSS= Tract-Based Spatial Statistics; TFCE= Threshold-Free Cluster Enhancement; TDC= typically developing children; SCQ= social communication questionnaire; SVM= support vector machines

Introduction

Autism spectrum disorder (ASD) represents a complex, heterogeneous neurodevelopmental condition characterized by deficits in social communication, as well as repetitive behaviors and atypical sensory reactivity (1). A 2013 survey by Centers for Disease Control and Prevention's National Center for Health Statistics showed that the prevalence of parent reported ASD among children aged 6-17 has continued to increase from 1.16% in 2007, to 2.00% in 2011-2012, likely due to broader diagnostic criteria, increased awareness of the disorder among parents and providers, increased parental age, and environmental contributors affecting epigenetic factors (2). It is now clear that inherited and *de novo* genetic changes, including copy number variations and single nucleotide variants, in neurodevelopment genes contribute to the phenotype in 25-40% of cases with an evolving understanding of polygenetic and epigenetic factors (3, 4). In addition to learning about genetic and epidemiologic factors, there has been increasing evidence from neuroimaging research suggesting that alterations in white matter microstructure and connectivity contribute to cognitive and behavioral deficits in affected children (5). Neuroimaging studies not only help illuminate the underlying mechanism of ASD phenotype in general (6, 7), but also could provide objective

biomarkers for timely identification of the ASD (8) as well as providing a marker for change with practice-based intervention (9).

Diffusion tensor imaging (DTI) and fiber tractography have provided quantitative evaluation of white matter microstructure and connectivity in children with ASD (10). The structural connectome, representing the whole-brain network of macro-scale white matter connectivity, has emerged during the past decade as a powerful formalism for the study of neurological and psychiatric diseases. The connectome is particularly relevant for ASD, which is hypothesized to result from short-range overconnectivity and long-range underconnectivity (11-13). However, to date there are no studies of ASD examining regional connectomic properties within white matter. Edge Density Imaging(EDI) has recently been introduced as a framework to represent the anatomic embedding of connectome edges within the white matter (14, 15). In EDI, the edges or links of the white matter connectome – from probabilistic tractography – are constrained to network nodes based on standard atlas parcellation of the cortical and deep gray matter nuclei (14, 15).

Machine learning analyses are also gaining popularity for pattern recognition and development of classification (or regression) models based on multidimensional data. These algorithms seem particularly suitable for devisingclassifiers based on multitude of variables extracted from diffusion and connectivity maps. In this study, we compared the white matter connectome and microstructure between children with ASD and typically developing children (TDC) using voxel-wise analysis. Then, we applied different machine-learning algorithms for identification of ASD based on the white matter tract-based average Edge Density (ED) and conventional DTI metrics.

Subjects and Methods

Participants and assessment

The participants in this study were recruited and prospectively enrolled through the UCSF Sensory Neurodevelopment and Autism Program clinical sites and research database (16-18). Children with ASD were diagnosed according to the Autism Diagnostic Interview-Revised (ADI-R) (19), Autism Diagnostic Observation Schedule (ADOS-G) (20), social communication questionnaire (SCQ) (21), and based on Diagnostic Statistical Manual – IV criteria (1). In addition, all participants were screened and interviewed by a senior pediatric neurologist (EJM) with expertise in neurodevelopmental disorders. The exclusion criteria were history of premature birth (<34 weeks), known genetic disorder associated with autism at time of enrollment (e.g. fragile X syndrome), or other neurological conditions that can potentially affect neurodevelopment (e.g. epilepsy). TDC did not meet diagnostic criteria for ASD or sensory processing disorders. In order to limit the confounding effects of age and gender, only boys aged 8 to 12 years were included in our analysis. Under the institutional review board approved protocol, informed consent was obtained from the parents or legal guardians, with the assent of all participants.

MRI protocol

All brain imaging was performed on a 3T MRI scanner (Siemens, Erlangen, Germany), using a 12-channel head coil. Anatomical scans were acquired using a three-dimensional T1-weighted MPRAGE sequence (field of view=256 mm, 1 mm cubic voxels, time to repeat/echo time/inversion time = 2300 ms/2.98 ms/ 900 ms, flip angle=9°). The whole brain high angular resolution diffusion imaging (HARDI) scan was acquired using a multislice 2D single-shot twice-refocused diffusion-weighted echoplanar imaging sequence (repetition time, 8000 ms; echo time, 109 ms; 100×100 matrix; field of view, 220 mm; voxel size, 2.2×2.2×2.2 mm; 64 non-collinear diffusion directions, uniformly distributed around a unit sphere with *b* value of 2000 s/mm^2; and 1 brain volume with no diffusion weighting (16-18).

DTI post processing

We used the software packages included in the Functional Magnetic Resonance Imaging of the Brain (FMRIB) Software Library (FSL) version 5.0.8 (http://www.fmrib.ox.ac.uk). The initial quality assurance involved eddy current and motion corrections, which was followed by removal of non-brain tissue. The FSL's DTIFIT toolbox was used to compute the fractional anisotropy (FA), mean diffusivity (MD), and radial diffusivity (RD) maps. The Bayesian Estimation of Diffusion Parameters Obtained using Sampling Techniques (BEDPOSTX2) package from FSL was used for estimation of diffusion parameters at each voxel, and modeling of multiple fiber orientations per voxel (16-18). Figure 1 summarizes the image post-processing pipeline.

Edge Density Imaging

For computation of ED maps, the T1-weighted series were first parcellated into 68 cortical and 14 subcortical regions based on the Desikan–Killiany atlas from FreeSurfer software (22). These 82 regions served as the connectome nodes (Figure 1). Then, the T1-weighted volumes, and subsequently the 82 cortical and subcortical regions, were registered to the diffusion space. Using EDI methods reported previously (14, 15, 23) the cortical and subcortical regions were employed as seed and target regions for probabilistic tractography using the FSL probtrackx2 algorithm (14). The total number of structural connectome edges passing through each voxel in white matter was calculated as the EDvalue for that voxel (Figure 1).

Tract-Based Spatial Statistics (TBSS), and voxel-wise analysis

For TBSS, each FA map was registered to all other FA maps to identify the most representative map of the cohort, and use this representative FA map as the target image. This target image was then affine-aligned into MNI-152 standard space, and then the rest of FA maps were transformed into MNI-152 space, combining the nonlinear transform to the target and the affine transform from the target to standard space. The mean of aligned FA maps was used to create a skeletonized image representing the center of white matter tracts across all subjects. This white matter skeleton was thresholded to exclude voxels

with FA values less than 0.2 (Figure 2), which may represent regions of high inter-subject variability. The ED maps were projected to MNI-152 using the registration matrix from corresponding FA maps and threshold. For non-parametric voxel-wise statistics, we used "randomise" from FSL with 5000 permutations and applying Threshold-Free Cluster Enhancement (TFCE) for multiple voxel-wise comparison correction. General linear model (GLM) designs were applied to correct for subjects' age as a covariate. For the ED maps, both ASD>control and control<ASD contrast designs were tested. The final statistical maps at a p value < 0.05 threshold were created, and corrected for multiple comparisons (Figure 2). For anatomic localization, we used the John Hopkins University white matter tractography atlas (ICBM-DTI-81) incorporated in FSL.

Extraction of tract-based DTI and connectome metrics

For univariate, and multivariate analysis as well as machine learning models, the average FA, MD, RD, and ED of white matter tracts from the ICBM-DTI-81 atlas were calculated. For this purpose, the ICBM-DTI-81 template was warped into each subject's native diffusion space applying the inverse spatial transformations from coregistration step described above (16-18). The average DTI/connectome metrics was thecalculated for all 48 white matter tracts in the ICBM-DTI-81 atlas for each subject.

Voxel-based morphometry (VBM)

We applied VBM to investigate voxel-wise differences in the local grey and white matter volume and topography between children with ASD versus TDC. The VBM tool included in FSL was used(https://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FSLVBM). Initially, the brain in T1-weighted images was extracted using the Brain Extraction Tool. Next, brain-extracted images were segmented into gray matter, white matter, and CSF. Then, thegray matter segments were aligned to ICBM-152 space applying the affine registration. These images were averaged to create a study-specific template, and finally all gray matter segments were nonlinearly registered and concatenated onto ICBM-152 space. These registered volume images were then modulated, and corrected for local

expansion or contraction. These modulated segmented images were smoothed with an isotropic Gaussian kernel at a sigma of 3 mm.Similar to voxel-wise statistics for TBSS, we used "randomise" from FSL with 5000 permutations and applied TFCE. We also used GLM for analysis of age as a covariate.

Machine learning

We evaluated different machine learning algorithms for predicting cohort assignment as ASD versus TDC based on tract-specific average connectome/DTI metric: naïve Bayes, random forest, support vector machine (SVM) with linear kernel, and polynomial kernel, and neural networks. Combination of these models with different DTI and connectomic metrics were evaluated using the 48 white matter average FA, MD, RD, and ED values as input for each model. In order to evaluate the performance of these algorithms, the subject cohort was randomly divided into the training and validation datasets with preservation of ASD-to-TDC ratio. The stratified random sampling for training and validation datasets was repeated 1000 times for cross-validation. For each permutation, the machine-learning model was trained on training sample, and a confusion matrix was constructed in corresponding validation sample based on the model predictions. The accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were calculated using thr confusion matrix for each validation cohort, and the average across 1000 permutations are presented. The accuracy was defined by sum of true positive and true negative subjects (correctly classified) divided by total number of subjects in each validation sample. For each confusion matrix, the area under the curve (AUC) of receiver operating characteristics (ROC) was also calculated.

For naïve Bayes, we applied the "naivebayes" "r" package with a Laplace smoothing value of 0 (https://www.r-project.org/). For random forest analysis, we applied the "randomForest" package. As recommended by the package authors, we used 500 random trees in each of the random forest models; and a 1/3 subset of variables was tried at each split. The averaged "mean decrease in Gini coefficients" is reported to depict the effect of different variables on classification accuracy of the final model. For SVM analysis, we

used the "e1071" package in "r" project. We applied both linear and polynomial kernels for data classification. Tuning the SVM models, a cost of 0.1 returned the optimal error rate, and was applied for all linear kernels. For the polynomial kernel, a sigma of 1 was applied. For neural networks, we applied the "neuralnet" package using "resilient back propagation" methods. Using 2/3 rule, we included 5 hidden layers with 32, 21, 7, 4, and 2 neurons, consequentially.

Statistics

The data are expressed as number (frequency) or average ± standard deviation, where appropriate. For univariate analysis, student's t test was performed. The Cohen's d coefficient was calculated to determine the effect size of the tract-based average DTI/connectome metrics using the "effsize" package from "r". For multivariate analysis, we applied the penalized logistic regression with stepwise forward and backward selection using the "stepPlr" package. The penalized regression is suitable for multivariate analysis with substantial collinearity between independent variables.

Results

Participant characteristics

A total of 47 boys, 8 to 12 years of age, were included from the Sensory Neurodevelopment and Autism Program neuroimaging collection. Of these, 14 (30%) boys met criteria for ASD and 33 (70%) did not. All children with ASD exceeded screening criteria on either the ADIR or SCQ parent report measures, with all these children exceeding autism score criteria on the ADOS. There was no significant difference in the average age of children with ASD (8.9 ± 2.7 years) versus the TDC (10.0 ± 3.3 years, p=0.52).

Voxel-wise comparison of connectivity maps

On voxel-wise TBSS analysis, after applying TFCE, the body and splenium of the corpus callosum, bilateral superior and posterior corona radiata, and the left superior longitudinal fasciculus had significantly lower ED in children with ASD compared to TDC (Figure 2). Conversely, there were no voxels in which the TDC cohort showed lower ED relative to the children with ASD. Table 1 lists the number of voxels in each of the white matter tracts from the ICBM-DTI-81 atlas with significant differences between the ASD and TDC study groups after applying TFCE. The GLM showed that children's age – in the restricted 8-12 years range examined – had no significant effect on voxel-wise ED values. In addition, there was no voxel-wise difference between children with ASD and TDC comparing FA (lowest voxel-wise p=0.245), MD (lowest voxel-wise p=0.275), and RD (lowest voxel-wise p=0.240) maps.

Univariate and multivariate tract-based comparison

There was no significant difference between average ED and DTI microstructural metrics of select white matter tracts between ASD and TDC groups. Tables 2a through 2d tabulate the results for select 22 white matter tracts chosen based on their Cohen's d effect size and results of voxel-wise analysis (Table 1). However, in multivariate stepwise penalized regression, the average ED of the left posterior corona radiata emerged as the only independent predictor of ASD (p=0.046). In ROC analysis, the average ED of the left posterior corona radiata yielded an AUC of 0.665 (95% confidence interval: 0.491 to 0.838, p=0.077) for distinction of children with ASD from TDC.

Machine learning classification

Applying a 3:7 ratio with preservation of the ASD-to-TDC proportion, we created 1000 stratified random samples of training and validation datasets. The training datasets included 26 TDC, and 11 ASD children; while the validation datasets included 7 TDC, and 3 ASD children. Table 3 and Figure 3 demonstrate the test characteristics for different machine learning models. Overall, the machine learning models using tract-based ED had better performance in classification of children with ASD compared to those using FA, MD, or

RD. Among all combinations, the greatest accuracy (75.3%), specificity (97.0%), and PPV (81.5%) were achieved by EDI-based random forest models; and the greatest sensitivity (51.4%), and NPV (77.7%) were achieved by EDI-based SVM with polynomial kernel (Figure 3).

The averaged "mean decrease in Gini coefficients" of random forest models from stratified cross validation are reported to depict the effect of different variables on classification accuracy of the final model (Table 4). In EDI-based random forest models, the mean ED of left posterior thalamic radiation, right sperior cerebellar peduncle, left sagittal stratum, left medial lemniscus, and left superior corona radiata had the highest averaged mean decrease in Gini coefficient.

Voxel-based morphometry

The VBM analysis revealed no macrostructural difference between the two study groups. There was no significant difference in gray-white matter relative tissue concentration, or regional volume comparing ASD children with TDC, in either ASD<control or control<ASD constructs.

Discussion

In voxel-wise analysis of the white matter connectome in children with ASD, we found lower ED in the body and splenium of corpus callosum, bilateral superior and posterior corona radiata, and the left superior longitudinal fasciculus compared to TDC. This measure of regional white matter connectome ED was more sensitive than conventional DTI metrics (i.e. FA, MD, and RD maps) as well as VBM of brain macrostructure, which failed to detect significant differences between the ASD and TDC groups at the studied sample size. While the voxel-wise analysis provides crucial information regarding the microstructural underpinning of the ASD, tract-based metrics extracted based on preset atlas might provide more feasible tool for distinguishing individual subjects with ASD. In this

preliminary study, we showed the feasibility of applying different supervised machine learning algorithms for identificationof children with ASD based on tract-based DTI and connectomic metrics. While univariate tract-based variables fail to distinguish children with ASD from TDC, machine learning models could construct imaging biomarkers for identification of ASD based on multitude of topographic DTI and connectomic information. In present study, the EDI-based models had better performance in identification of children with ASD compared to conventional DTI metrics, although the results should be confirmed in larger cohorts.

Numerous tractography and functional MRI studies characterize ASD as a neurodevelopmental disorder due to underconnectivity between different brain regions (10, 24). The majority of prior DTI studies have demonstrated decreased FA and increased MD in white matter tracts (25), most commonly reported in the corpus callosum, and cingulum (26). The FA represents directional variation in apparent diffusion, and the MD is the average of eigenvalues measuring diffusion rate irrespective of direction. While the changes in FA and MD are sensitive measures of white matter microstructure, they are relatively nonspecific, and may represent lower axonal density, thinner axons, or less myelination. In EDI, on the other hand, quantification of network edges represents the significance of each white matter voxel in the overall connectomic framework. Driven from the probabilistic calculation of tract density, ED provides a metric sensitive to directionality of diffusion at each voxel, and theoretically more likely representative of true neural fibers given the constraints in construction of edges originating from predetermined cortical and subcortical gray matter regions as connetome nodes. Although the number of cases in the current cohort was too small to draw a firm conclusion, these results suggest that ED is potentially more sensitive than FA, MD, and RD maps for identification of microstructural connectivity differences between children with ASD and the control cohort.

In our study, the results of EDI are consistent with the theory of decreased transcallosal fiber connectivity in children with ASD (27). The commissural tracts connecting bilateral premotor, primary motor, and primary sensory cortex traverse through the body of corpus callosum (28); and the lower ED in the mid corpus callosum

may help explain sensory and motor processing deficits in ASD children. The splenium, specifically, connects occipital, parietal, and temporal regions, which are involved in visual processing. Lower ED in the splenium of corpus callosum in children with ASD may be related to deficits in visual processing (25). Notably, callosal abnormalities, such as the diminished ED noted herein, are one of the most well replicated findings for individuals with ASD, with implications for slower transmission of information leading to deleterious consequences for processing of nuanced and socially rich visual information (29, 30).

While voxel-wise comparison of EDI maps in children with ASD can elucidate the neural microstructural underpinning of autism, development of voxel-wise DTI/connectomic fingerprint for identification of individual subjects with ASD can be challenging. As a solution, atlas-based parcellation of white matter tracts can provide automated and reproducible tractometry variables for classification of individual subjects based on topographic pattern of connectivity changes. In our series, there was no significant difference in average tract-based DTI or connectomic metrics between children with ASD and TDC (Table 2). On the other hand, the average ED in the left posterior corona radiata was the only independent predictor of ASD in multivariate stepwise penalized regression. Still, the average ED of the left posterior corona radiata could not reach statistical significance in identification of children with ASD applying receiver operating characteristic analysis ($p=0.077$). Nevertheless, multidimensional data from tract-based connectivity metrics could identify individuals with ASD using machine-learning algorithms.

In recent years, there has been increasing interest in the application of machine learning for objective and reproducible decision models in diagnosis or treatment planning. These models have been applied in neurological disorders, such as the diagnosis of Alzheimer's (32) and Parkinson's disease (33), or lesion-symptom mapping in stroke patients (34). Different machine learning algorithms, however, vary in their implementation, mathematical logic, and computation. Given the small sample size, and inherent variability in results of machine learning algorithms depending on the randomly sampled training and validation datasets, we reported the averaged test characteristics among 1000 stratified samples for cross validation of each model. Thus, Table 3 results likely represent a realistic assessment of the classification accuracy for each combination of diffusion/connectomic metrics and machine learning model in our cohort. The models using ED had better performance in identification of ASD compared to those using conventional DTI metrics. Also, the accuracy of random forest models were slightly higher than SVM

models mostly due to higher PPV and specificity, although the results may not hold in a larger dataset (Table 3). It should be note that the high accuracy achieved by random forest was in part due to assigning majority of participants to control cohort, achieving higher specificity and NPV at expense of low sensitivity. This can affect clinical application of the model given the cost of assigning a child with ASD to the TDC cohort, and thus preventing him/her from receiving proper and timely treatments. It is also noteworthy that, in a larger dataset, SVM models may outperform random forests for binary classification (35). Of note, random forests models can also offer a glimpse into their classification constructs by reporting variable importance (mean decrease in Gini); however, the possible neurobiological translation of these outputs remain elusive – whether they are mere reflection of random forest algorithm computation or may indeed point out to neurobiologically important white matter tracts. Nevertheless, our preliminary results emphasize the importance of exploring different machine learning options to identify the suitable solution in the development of image-based classifiers.