Identification of Infants at Risk for Autism Using Multi-parameter Hierarchical White Matter Connectomes

Yan Jin, Chong-Yaw Wee, Feng Shi, Kim-Han Thung, Pew-Thian Yap, and Dinggang Shen for the Infant Brain Imaging Study (IBIS) Network

Department of Radiology and BRIC, School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

Yan Jin: yjinz@med.unc.edu; Chong-Yaw Wee: chongyaw_wee@med.unc.edu; Feng Shi: fengshi@med.unc.edu; Kim-Han Thung: khthung@med.unc.edu; Pew-Thian Yap: ptyap@med.unc.edu; Dinggang Shen: dinggang_shen@med.unc.edu

Abstract

Autism spectrum disorder (ASD) is a variety of developmental disorders that cause life-long communication and social deficits. However, ASD could only be diagnosed at children as early as 2 years of age, while early signs may emerge within the first year. White matter (WM) connectivity abnormalities have been documented in the first year of lives of ASD subjects. We introduce a novel multi-kernel support vector machine (SVM) framework to identify infants at high-risk for ASD at 6 months old, by utilizing the diffusion parameters derived from a hierarchical set of WM connectomes. Experiments show that the proposed method achieves an accuracy of 76%, in comparison to 70% with the best single connectome. The complementary information extracted from hierarchical networks enhances the classification performance, with the top discriminative connections consistent with other studies. Our framework provides essential imaging connectomic markers and contributes to the evaluation of ASD risks as early as 6 months.

1 Introduction

Autism spectrum disorder (ASD) is a type of complex brain developmental disorders characterized by repetitive behaviors, both verbal and non-verbal communication difficulty, and social interaction obstacle. About 1% of the world population is affected by ASD. In the US, it is estimated that one in 68 children is affected by ASD. It is a life-long disease involving an annual healthcare cost of around $250 billion. Therefore, early diagnosis and medical intervention will significantly improve the life quality of subjects, and reduce the financial burden borne by the society. Unfortunately, so far, there is no single medical test for ASD diagnosis. Instead, it is based on the evaluations made by specially trained physicians and psychologists on specific behavioral tests, typically after the age of two [1]. On the other hand, studies have shown that a number of brain structural deficits may emerge in as early as the first year of life [2]. For example, white matter (WM) abnormalities have been observed over multiple locations such as corpus callosum [3] and the reduction of global network efficiency [4] was found in the brains of ASD infants between 7 months and
2 years of age. Nevertheless, few studies explored the ASD risk in infants before toddlerhood, especially in the first year of life.

Computer-aided diagnosis using features obtained from medical images have been successfully applied to identifying various clinical groups [5–8]. A number of studies attempted to classify autistic children, using features on regional structural MRI [9] and diffusion parameters of WM regions [10]. However, the subjects involved in these studies were over 7 years old, when ASD has progressed considerably. For early intervention, it is desirable to identify ASD at a much earlier stage, preferably even before the first trace of symptomatic behaviors. However, identifying ASD in infants is challenging and not well studied because of the difficulty of image acquisition from infants and also the lack of obvious symptoms at this stage.

In this paper, we propose a novel multi-channel machine-learning based classification framework to identify the six-month-old infants at high-risk for ASD. The major contributions of this study include: firstly, we develop a novel brain parcellation strategy to partition a publicly available atlas “infant AAL” [11] into anatomical meaningful regions of interest (ROIs) with adaptive sizes; secondly, unlike [10], we propose to use the features from a hierarchical set of whole-brain WM connectivity networks (i.e., connectomes), instead of conventional region-based features, to identify ASD infants; finally, we utilize an effective two-stage feature selection scheme and multi-kernel SVM classifier that can incorporate the complementary information from multi-channel sources to optimize the classification accuracy.

2 Multi-Parameter Hierarchical Connectome Classification

2.1 Overview

We employ in this study multi-parameter hierarchical WM connectivity networks as multi-source information for identification of infants who are at risk for ASD. The diagram of the complete workflow for our proposed method is shown in Fig. 1. To define the nodes in network, we start with the publicly available infant AAL atlas [9], and parcellate its 90 cerebral ROIs into 203 and 403 ROIs, respectively, for constructing more detailed brain networks (Section 2.2). Then, we define connections (edges) of the network using multiple diffusion properties, such as fractional anisotropy (FA), mean diffusivity (MD), and fiber length, over the fiber tracts connecting each pair of ROIs (nodes). Thus, three sets of hierarchical WM connectivity networks (nine in total) are constructed for each subject (Section 2.3). Finally, the relevant features selected from these networks using t-test and LASSO logistic regression are fed into a multi-kernel SVM classifier for patient identification (Section 2.4).

2.2 ROI Parcellation

The infant brain and the adult brain have large anatomical variances. Here, we used a 1-year-old infant AAL atlas that was adapted from the adult AAL atlas, which is a widely used single-subject high-resolution T1-weighted atlas [11]. The entire cerebrum was parcellated into 90 anatomically meaningful ROIs. However, to get more detailed connectivity information, it is necessary to divide the original ROIs into finer ROIs. We propose a novel
strategy to parcellate the original ROIs into smaller sub-ROIs based on adaptive sizes. In our case, 203 and 403 ROIs were obtained when the size was chosen at around 20 mm and 16 mm, respectively. The detailed algorithm is described in Algorithm 1. Fig. 2 illustrates the 3D views of the three scales of infant AAL ROIs.

### Algorithm 1

**ROI Parcellation**

Set the cube size $a$ and load the original infant AAL ROIs.

Divide the entire volume (including the background) evenly into cubes of size $a^3$.

for ROI $i=1:90$

compute the center of each new sub-ROI

if any inside $V_{cube} > 0.3*a^3$, e.g., $V_{cube} > 0.3*a^3$, calculate the center of the cube.

else Find a neighboring cube whose center is closest to its center and combine two regions.

Recompute the new center of the combined region.

end if

end for

for ROI $i=1:90$

assign new ROIs

Assign each inside voxel a new ROI index based on its shortest distance to an inside center.

end for

2.3 Hierarchical Connectomes

A connectome provides a comprehensive description of the complete structural connectivity of the entire brain. The features extracted from a connectome provide rich information for identifying ASD subjects due to its comprehensive characterization of the connections between brain regions. Furthermore, multi-scale connectomes may provide a range of complementary coarse-to-fine information for multi-level analysis of brain connections.

Two ROIs were considered anatomically connected if they were traversed by a common set of fibers. Connection strength between the two ROIs was computed as the average of FA, MD, and fiber length over the traversing fibers, respectively. Then, three connectivity networks (corresponding to FA, MD, and fiber length) can be constructed based on pairwise connection strengths. By using all three different scales of the infant AAL atlas with 90, 203, and 403 ROIs, respectively, we constructed three sets of hierarchical connectivity networks (nine in total) for each subject. Fig. 3 illustrates a hierarchical set of FA connectomes from a pair of low-risk and high-risk subjects, respectively. Notice that the multi-scale networks of a low-risk subject contain stronger connections (i.e., more and thicker edges) than those of a high-risk subject.

2.4 Multi-Kernel Classification

We considered each element in each connectivity network as a feature. Thus, the number of features increased with the square of the number of ROIs, with tens of thousands of features for the finest scale. Many of them would be redundant or irrelevant for classification. Thus, we performed feature selection on each network separately to determine an optimal subset of features that were relevant. An initial subset of features was first selected using feature-wise
t-tests performed between the high-risk group and the low-risk group. Features with p-values below an empirical threshold were further sieved using LASSO logistic regression.

Multi-kernel SVM has been proven to be more effective than single-kernel SVM [9] in utilizing the complementary information from multiple sources for improving classification accuracy. We applied it to the features selected from the nine networks, as described above.

Let \( \{(x_i^j, y_i), i = 1, \ldots, M, j = 1, \ldots, T\} \) be the set of training data, where \( x_i^j \in \mathbb{R}^{n_j \times 1} \) represents the feature vector of the \( i \)-th subject for network \( j \) and \( y_i \in \{1\text{ (high-risk)} \text{ or } -1\text{ (low-risk)}\} \) is the label. \( M \) is the number of training subjects and \( T \) is the number of networks (i.e., \( T = 9 \) in our study). The primal formulation of multi-kernel SVM is given as

\[
\min_{\beta_j, P_j, \xi_i} \frac{1}{2} \sum_{j=1}^{T} \beta_j^2 \|P_j\|^2 + C \sum_{i=1}^{M} \xi_i \\
\text{s. t.} \quad y_i (\sum_{j=1}^{T} \beta_j (P_j^T \phi_j(x_i^j) + b)) \geq 1 - \xi_i \quad (1)
\]

where \( \beta_j, P_j \) and \( \phi_j \) denote the weight, the normal vector of the classification hyperplane, and the kernel-induced mapping function for the \( j \)-th network, respectively. \( b \) denotes the bias term, \( \xi_i \) denotes the slack variable (for misclassification), and \( C \) is a parameter that controls the degree of misclassification. Given a test subject \( x \), the prediction of its label \( \hat{y} \) will be

\[
\hat{y} = \text{sign}(\sum_{j=1}^{T} \beta_j (P_j^T \phi_j(x) + b)).
\]

We used an open-source software package SimpleMKL [12], which could decide the weight of each network simultaneously while solving the optimization problem.

3 Experiment Results

3.1 Subjects and Data Processing

The participants used in this study were chosen from the Infant Brain Imaging Study (IBIS, http://www.ibis-network.org), an ongoing study of brain and behavioral development in infants. The high-risk infants have at least one older sibling with ASD, while the low-risk infants have no first-degree relatives with ASD. Included in this study were 40 six-month-old high-risk infants (29 males/11 females) and also 40 low-risk infants (27 males/13 females).

The DWI images were acquired with a 2 mm isotropic spatial resolution, and consisted of one non-diffusion-weighted \( b_0 \) volume and 25 diffusion-weighted volumes with \( b=1000 \) s/mm\(^2\). FA and MD images were then extracted from the data after diffusion tensor fitting. The infant AAL atlas was registered to the subjects’ FA images. With the deformation fields generated, we warped all three sets of infant AAL ROIs (i.e., 90, 203, and 403 ROIs) to the DWI image space of each subject. Whole-brain tractography was performed using deterministic streamline tractography with peaks detected from the WM orientation distribution functions. Seed points were chosen as voxels with FA > 0.3. The maximum fiber turning angle was set to 45°, and tracking was stopped when FA < 0.15.
3.2 Experiment Setting

Three sets of hierarchical connectivity networks based on mean FA, mean MD, and mean fiber length were constructed as described in Section 2.3. The performance was evaluated using nested 5-fold cross validation with 10 randomized repetitions. In feature selection, for each network, an initial $t$-test was conducted to select the features with $p$-values < 0.001. Some features were further discarded using LASSO logistic regression. More specifically, the results of the binarized regression in the inner 5-fold cross validation within the training set were compared to the ground truth. The LASSO parameters with the best accuracy were used to select the features for the test fold. Finally, selected features from each network were fed into SimpleMKL for classification. Another inner 5-fold cross validation using the training set was performed to find the best parameter $C$ in Eq. (1) and the test fold was classified with that $C$.

For performance evaluation, we used several metrics, i.e., accuracy (ACC), sensitivity (SEN), specificity (SPE), and the area under the receiver operating characteristic curve (AUC) for performance evaluation. Let TP, TN, FP, and FN denote, respectively, true positive, true negative, false positive, and false negative cases that the algorithm detects. The definitions of ACC, SEN, and SPE are:

$$\text{ACC} = \frac{TP + TN}{TP + TN + FP + FN}$$
$$\text{SEN} = \frac{TP}{TP + FN}$$
$$\text{SPE} = \frac{TN}{TN + FP}$$

3.3 Classification Results

Fig. 4 shows the performance comparison between using a single network and multiple networks. In (A–C), the performance with the hierarchical set of networks (90, 203, and 403 ROIs) is compared to that of each single network, in terms of the parameter FA, MD, and fiber length, respectively. The accuracies of multi-network are ~73%, with a gain of ~5% compared to the best single network. The AUCs are ~0.78 with a gain of ~13% against the best single network. Smaller standard deviations for most of the statistics, especially AUC, suggest a more stable performance by the hierarchical network framework. In (D), our proposed method, which includes all parameter networks, further outperforms each hierarchical network alone with an accuracy of 76% and an AUC of 0.80. To validate the significant improvements, we conducted the pair-wise $t$-tests based on ACC on the total 50 fold results. For both levels, i.e., the hierarchical networks vs. the single network with a single parameter and our proposed framework vs. those single-parameter hierarchical networks, $p < 0.05$. The results demonstrate that the complementary information provided by both multi-scale networks and multiple parameters indeed enhances the performance significantly.

3.4 Most Discriminative Connections

We summed the counts of each connection selected by our proposed method for each level of connectome (90, 203, and 403 ROIs) over the 50 folds. The counts for FA, MD, and fiber length were combined at each level. The top 20 discriminative connections are reported using connectograms, a circular representation of connectomics, in Fig. 5. For the 203-ROI and 403-ROI networks, the original 90-ROI connections that contain those sub-ROI connections are shown. Thickness of each connection reflects its selection frequency, i.e.,
the thicker the line, the higher the selection frequency. Many connections are common to all three networks. However, the 403-ROI connectogram shows more diverse connections, even inter-hemisphere connections that are not shown in the coarser scales of connectomes. The connecting regions include frontal, parietal, temporal, occipital lobes, and a few subcortical regions, such as globus pallidus and putamen. Our findings are consistent with the previous studies [4]. Those top connections may serve as possible imaging markers for ASD diagnosis.

4 Conclusion

We propose a multi-channel classification framework that utilizes complementary information from a set of multi-parameter hierarchical WM connectivity networks to enhance the performance in identifying 6-month-old infants who are at high-risk for ASD. Our result of ACC/AUC of 76%/0.8 vs. 70%/0.7 for the best single network demonstrates the effectiveness of the propose method. Our method can be potentially applied to detecting other WM related diseases as well.

Acknowledgments

This work was supported by the National Institute of Health grants EB006733, EB008374, EB009634, AG041721, MH100217, and AG042599. We thank the National Database for Autism Research (NDAR) for providing the data to this research.

References

1. Centers for Disease Control and Prevention. Prevalence of autism spectrum disorder among children aged 8 years. Surveillance Summaries. 2014; 63(2):1–21.
2. Ozonoff S, et al. A prospective study of the emergence of early behavioral signs of autism. J Am Acad Child Adolesc Psychiatry. 2010; 49(3):256–266. [PubMed: 20410715]
3. Wolf JJ, et al. Differences in white matter fiber tract development present from 6 to 24 months in infants with Autism. Am J Psychiatry. 2012; 169(6):589–600. [PubMed: 22362397]
4. Lewis JD, et al. Network inefficiencies in autism spectrum disorder at 24 months. Transl Psychiatry. 2014; 4:e388.10.1038/tp.2014.24 [PubMed: 24802306]
5. Zhu X, et al. A novel matrix-similarity based loss function for joint regression and classification in AD diagnosis. Neuroimage. 2014; 100:91–105. [PubMed: 24911377]
6. Li, J.; Jin, Y.; Shi, Y.; Dinov, ID.; Wang, DJ.; Toga, AW.; Thompson, PM. Voxelwise spectral diffusional connectivity and its applications to alzheimer’s disease and intelligence prediction. In: Mori, K.; Sakuma, I.; Sato, Y.; Barillot, C.; Navab, N., editors. MICCAI 2013, Part I. LNCS. Vol. 8149. Springer; Heidelberg: 2013. p. 655-662.
7. Zhan L, et al. Comparison of nine tractography algorithms for detecting abnormal structural brain networks in Alzheimer’s disease. Front Aging Neurosci. 2015; 7:48.10.3389/fnagi.2015.00048 [PubMed: 25926791]
8. Jin Y, et al. Automatic clustering of white matter fibers in brain diffusion MRI with an application to genetics. Neuroimage. 2014; 100:75–90. [PubMed: 24821529]
9. Wee CY, et al. Diagnosis of autism spectrum disorders using regional and interregional morphological features. Hum Brain Mapp. 2014; 35(7):3414–3430. [PubMed: 25050428]
10. Ingalhalikar M, et al. Diffusion based abnormality markers of pathology: Toward learned diagnostic prediction of ASD. Neuroimage. 2012; 57(3):918–927. [PubMed: 21609768]
11. Shi F, et al. Infant brain atlases from neonates to 1- and 2-year-olds. PLoS One. 2011; 6(4):e18746.10.1371/journal.pone.0018746 [PubMed: 21533194]
12. Rakotomamonjy A, et al. SimpleMKL. J Mach Learn Res. 2008; 9:2491–2521.
Fig. 1.
The proposed classification framework based on multi-parameter hierarchical WM connectivity networks.
Fig. 2.
The 3D views of the three scales of infant AAL ROIs.
Fig. 3.
The 3D views of the hierarchical set of connectomes thresholded by mean FA = 0.4 from a pair of low-risk and high-risk subjects. Each sphere represents the center of an ROI and its size indicates the normalized volume of the ROI. The colors of the sub-ROIs in (B), (C), (E), and (F) correspond to those of the original ROIs in (A) and (D). The thickness of edges represents the mean value of FA averaged over the fibers connecting two respective ROIs.
Fig. 4.
Performance comparison for the 90-ROI (blue), 203-ROI (cyan), and 403-ROI (green) networks with (A) the hierarchical network framework (yellow) using FA; (B) the hierarchical network framework (orange) using MD; (C) the hierarchical network framework (dark red) using fiber length; and (D) the proposed hierarchical network framework using multiple diffusion statistics (red). The error bars denote the standard deviations obtained from 10 repetitions.
Fig. 5.
Connectograms of the top 20 discriminative connections selected by our framework for (A) 90-ROI, (B) 203-ROI, and (C) 403-ROI networks, respectively. The intra- and inter-hemisphere connections are shown in red and black colors, respectively.