Identifying Relationships in Functional and Structural Connectome Data Using a Hypergraph Learning Method

Brent C. Munsell[1], **Guorong Wu**[2], **Yue Gao**[3], **Nicholas Desisto**[1], and **Martin Styner**[4]

[1]Department of Computer Science, College of Charleston, Charleston, SC, USA

[2]Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[3]School of Software, Tsinghua University, Beijing, China

[4]Department of Psychiatry, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

Abstract

The brain connectome provides an unprecedented degree of information about the organization of neuronal network architecture, both at a regional level, as well as regarding the entire brain network. Over the last several years the neuroimaging community has made tremendous advancements in the analysis of structural connectomes derived from white matter fiber tractography or functional connectomes derived from time-series blood oxygen level signals. However, computational techniques that combine structural and functional connectome data to discover complex relationships between fiber density and signal synchronization, including the relationship with health and disease, has not been consistently performed. To overcome this shortcoming, a novel connectome feature selection technique is proposed that uses hypergraphs to identify connectivity relationships when structural and functional connectome data is combined. Using publicly available connectome data from the UMCD database, experiments are provided that show SVM classifiers trained with structural and functional connectome features selected by our method are able to correctly identify autism subjects with 88 % accuracy. These results suggest our combined connectome feature selection approach may improve outcome forecasting in the context of autism.

1 Introduction

Improvements in computational analyses of neuroimaging data now permit the assessment of whole brain maps of connectivity, commonly referred to as the brain connectome [7]. The brain connectome provides unprecedented information about global and regional conformations of *neuronal network architecture* (or *network architecture* for short) that is particularly relevant as it relates to neurological disorders. For this reason, the brain connectome has recently become instrumental in the investigation of network architecture organization and its relationship with health and disease, notably in the context of neurological conditions such as epilepsy, autism, Alzheimer's, and Parkinson's. In general,

Correspondence to: Brent C. Munsell.

two connectome categories exist: (1) a *structural connectome* that is reconstructed using white matter fiber tractography from diffusion tensor imaging (DTI), and (2) a *functional connectome* that is reconstructed using resting-state time-series signal data from blood oxygen level dependent (BOLD) functional MRI (rsfMRI).

In mathematical terms, a connectome is a weighted undirected graph, where nodes in the graph represent brain regions (defined in an anatomical parcellation, or brain atlas), and the edge that connects two different nodes is weighted by a value that represents the level of neural-connectivity, or information exchange. To better understand how the brain network is organized, network analysis algorithms [4] are applied to the connectome to reveal the underlying network architecture of the brain, which can then be used to quantify the differences between healthy and disease conditions. Currently, network analysis techniques have mainly been applied to just structural or functional connectivity data. However, research that combines both types of data [1,5,6,8] to better understand functional and structural connectivity relationships has gained attention in recent years.

Here a novel *combined connectome feature selection* technique is proposed that uses hypergraphs to discover latent relationships in node-based graph theoretic measures found in structural and function connectomes. The primary rational behind selecting features where structural and functional connectivity agree, is that fiber density and signal synchronization similarities are likely to be correlated, and when combined these similarities may be easier to identify and quantify. More specifically, for each diagnosis label (i.e. disease and healthy) the proposed feature selection technique uses a hypergraph learning algorithm to find a hypergraph Laplacian graph that combines structural and functional node-based connectivity measures. A hierarchical partitioning algorithm is then applied to the hypergraph Laplacian, which in turn creates a code vector that encodes structural and functional connectivity similarities. The resulting code vectors are then used to create a binary weight vector that only selects brain regions associated with structural or functional node-based connectivity measures capable of differentiating the disease condition from the healthy one. Lastly, the selected structural and functional connectome features are used to train a SVM classifier that can predict diagnosis label of subjects not included in the training procedure.

2 Materials and Methods

2.1 Participants and MRI Data Acquisition

All participant data was acquired from the publicly available University of Southern California (USC)/University of California Los Angeles (UCLA) multimodal connectivity database[1] (UMCD). In particular, high-functioning children and adolescents with an autism spectrum disorder (ASD), and healthy control (HC) children and adolescents were recruited. In total, the autism study has 70 participants (35 ASD and 35 HC) that had both rsfMRI and DTI scan data. A complete list of all the demographic data, including the scan parameters, from the original study can be found at [5].

---

[1]http://umcd.humanconnectomeproject.org.

2.2 Preprocessing and Connectome Reconstruction

Functional preprocessing steps were performed using the FSL[2] and AFNI[3] software libraries. In general, the following steps were performed: skull stripping, slice timing correction, motion correction with rigid-body alignment using MCFLIRT, geometric distortion correction using FUGUE. Structural pre-processing steps were performed using the FSL and Diffusion toolkit[4] software libraries. In general, the following steps were performed: skull stripping, eddy current correction, motion correction with rigid-body alignment using MCFLIRT, voxel-wise fractional anisotropy (FA) estimation, and fiber track assignment using FACT algorithm. A complete overview of all the preprocessing steps can be found at [5].

The FSL FEAT query function is then applied to the functional and structural preprocessed images. In particular, the atlas proposed in Power *et al.* [2] defines $m = 264$ ROIs that are represented by a 10 mm diameter sphere. A symmetric $m \times m$ functional connectivity matrix C_f is constructed using the extracted ROIs, where each element in the functional connectivity matrix reflects the *signal synchronization* between two ROIs, which is estimated by computing the correlation between two discrete time-series rsfMRI signals. Likewise, a symmetric $m \times m$ structural connectivity matrix C_s is constructed using the same ROIs, where each element reflects the average number of fiber tracks, or *fiber density*, that connect the two ROIs.

2.3 Node-Based Connectome Feature Vector

The next step is to convert the values defined in C into node-based connectome feature vector $\mathbf{c}_a = (c_{a1}, \ldots, c_{am})$ using the betweenness centrality graph-theoretic connectivity measure, where $a = s$ represents a node-based structural connectome feature vector, and $a = f$ represents a node-based functional connectome feature vector. Betweenness centrality is a global measure that represents the fraction of shortest paths that go through a particular node (or brain region) defined in the connectome. The betweenness centrality measure for node i is

$$c_{\alpha i} = \frac{1}{(m-1)(m-2)} \sum_{h,\, j \in m} \frac{\rho_{hj}^i}{\rho_{hj}}, \quad (1)$$

where $h \neq j, h \neq i, j \neq i$, The number of shortest path between node h and j is represented by ρ_{hj}, the number of these shortest paths going through node i is represented by ρ_{hj}^i. This is normalized to a value in [0 1], where $(m-1)(m-2)$ is the highest score attainable in the network.

2.4 Combined Connectome Feature Selection

Given a training data set $A = \left\{\mathbf{a}_\phi\right\}_{\phi=1}^{n}$ of n ASD subjects we compute set of graph Laplacians $\left\{L_\phi\right\}_{\phi=1}^{n}$, where $\mathbf{a}_\phi = (\mathbf{c}_{\phi s} \mid \mathbf{c}_{\phi f})$ is a $2m$ dimension feature vector that combines structural and functional node-based connectome values for subject ϕ. To do so, we first create a complete bipartite graph $G_\phi = (\mathbf{c}_{\phi s}, \mathbf{c}_{\phi f}, E_\phi)$ for each subject, where the edge that connects structural node i to functional node j in the bipartite graph is weighted by $w_{ij} = 1 - |c_{si}\, c_{fj}|$. The proposed edge weight strategy has a very straight forward and intuitive meaning: If two brain regions *both have* similar connectivity values then $w_{ij} \approx 1$, conversely if two brain regions *do not* have similar connectivity values then $w_{ij} \approx 0$.

A $2m \times m^2$ dimension hypergraph incidence matrix H_ϕ for subject ϕ is then created using G_ϕ. Because we use bipartite graph, it's important to note that each hyper-edge only represents the structural-functional relationship between two node-based connectome features. Once H_ϕ is found, the normalized hypergraph Laplacian[5]

$$L_\phi = I - D_v^{-1/2} H_\phi D_e^{-1} H_\phi^t D_v^{-1/2} \quad (2)$$

is computed [11], where D_v is a diagonal matrix that defines the strength for each vertex in H_ϕ, D_e is a diagonal matrix that defines the strength for each edge in H_ϕ, and I is the identity matrix. In general, our design has two advantages: (1) we only identify functional and structural connectivity relationships just between two different regions in the brain, and (2) the resulting hypergraph Laplacian is very sparse. A median hypergraph Laplacian L_m is then found using each subject specific hypergraph Laplacian in $\left\{L_\phi\right\}_{\phi=1}^{n}$, where $L_m(i, j) = median(\{L_1(i, j), L_2(i, j), \ldots, L_n(i, j)\})$.

Eigen decomposition is applied to L_m creating a $2m$ dimension *embedding space* and then a hierarchical partition is performed as illustrated in Fig. 1. More specifically, each embedding space partition in the hierarchy defines three cluster groups: (1) clusters that only have DTI brain regions, (2) clusters that only have rsfMRI brain regions, and (3) clusters that have both DTI and rsfMRI brain regions. At each partition the three cluster groups are found using the well-known normalized spectral clustering technique in [10]. However, instead of using a k-means algorithm the density estimation algorithm in [3] is applied, primarily because the number of clusters is automatically found and outliers can be automatically recognized and excluded. As shown in Fig. 1, the DTI and rsfMRI brain region cluster becomes the search space for the next partition in the hierarchy, and terminates when a DTI and rsfMRI brain region cluster does not exist.

In our approach, each partition level in the hierarchy represents a unique integer code, and partitions at the top of the hierarchy represent brain regions that show low structural and functional connectivity similarities (i.e. low code value), and partitions near the bottom of

---

[5]In our approach each hyper edge has the same influence, therefore W is the identity matrix and is omitted in Eq. (2).

the hierarchy represent brain regions that show high structural and functional connectivity similarities (i.e. large code value). Lastly, a code vector $\mathbf{x}_{ad} = (x_{s1}, x_{s2}, \ldots, x_{sm}, x_{f1}, x_{f2}, \ldots, x_{fm})$ is created using the code values in partition hierarchy, and then normalized by simply dividing all the code values by the height of the partition hierarchy.

This exact same procedure outlined above is then applied to a training set of HC subjects, and a HC code vector $\mathbf{x}_{hc}$ is produced. Next, a weight vector

$$\mathbf{w} = |\mathbf{x}_{ad} - \mathbf{x}_{hc}| \quad (3)$$

is created, where a weight value close to one represents structural or function brain regions that have dramatically different code values, which suggests these regions may better differentiate the disorder from the normal condition. On the other hand, a weight value close to zero represents structural or function brain regions that have the same (or very similar) code values, which suggests these regions may not be able to differentiate the disorder from the normal condition. Lastly, we make $\mathbf{w}$ binary by applying a threshold, i.e. $w_i \geq t_h = 1$ and $w_i < t_h = 0$. The primary motivation behind making the weight vector binary was to reduce the number of dimensions, which in turn will reduce the amount of error that may be introduced into the chosen classifier.

2.5 Linear SVM Classifier

Using a training data set $A = \left\{ \mathbf{a}_\phi \right\}_{\phi=1}^n$ that now includes both ASD and HC subjects, the binary diagnosis labels $\mathbf{y} = (y_1, y_2, \ldots, y_n)$, e.g. ASD = 1 and HC = 0, and the binary weight vector $\mathbf{w}$, a linear two-class SVM classifier based on the LIBSVM library[6] is trained. In particular, the binary values in $\mathbf{w}$ is applied to each feature vector in A, creating a new sparse training data matrix $\tilde{A}$. Finally, a SVM classifier is trained using $\tilde{A}$. Once the SVM classifier is trained, the diagnosis label of a subject not included in the training data set can be predicted as follows: First compute $\mathbf{a} = (\mathbf{c}_s \mid \mathbf{c}_f)$, then create sparse feature vector $\tilde{\mathbf{a}} = (a_1 w_1, a_2 w_2, \cdots, a_{2m} w_{2m})$ by applying learned binary weights, and lastly calculate the predicted diagnosis label y using trained SVM classifier, where the sign of the y (i.e., $y \geq 0$ or $y < 0$) determines the diagnosis label.

Since the proposed combined connectome feature selection has two free parameters, i.e. number of Eigen-values (or dimensions) used by cluster algorithm (d) and binary weight threshold (t_h) a grid search procedure is performed that uses 10-fold cross validation strategy. Specifically, an independent two-dimension grid-search procedure is performed for each left-out-fold, where the value stored at grid coordinate (d, t_h) are the mean and standard deviation values for the accuracy (ACC), sensitivity (SEN), specificity (SPC), negative predictive value (NPV), and positive predictive value (PPV) measures. In particular, d is adjusted at increments of 1 starting at 1 and ending at $2m$, and t_h is adjusted at increments of 0.05 starting at 0.1 and ending at 1.0. Lastly, when the grid-search procedures completes the parameter values that have the highest ACC and PPV scores are selected.

---

[6]http://www.csie.ntu.edu.tw/~cjlin/libsvm.

3 Results

The grid search parameters the yielded the best ACC and PPV classification results are $d = 3$ and $t_h = 0.8$. To assess the performance of the proposed feature selection method, SVM classifiers are also trained using structural and functional connectome features in training data set A that are selected by: (1) a linear regression technique that includes ℓ_1 regularization (i.e. Lasso), and (2) no feature selection. As shown in Table 1, a SVM classifier trained using structural and functional connectome features selected by the proposed method is the most accurate at 88.3 %, can predict the disease case (i.e. PPV) approximately 87.2 % of the time, and consistently shows the highest sensitivity, specificity, and NPV.

The bar plots in Fig. 2 show the median[7] structural and functional weight values found using Eq. (3) when grid search parameter $t_h = 0.8$ is used. The SVM classifier in Table 1 is trained only using the node-based connectivity values from the selected 47 regions (24 structural regions and 23 functional regions) also shown in Fig. 2. In general, the 47 regions have largest difference in code values, which suggests the structural and functional connectivity characteristics in these brain regions are significantly different between ASD and HC subjects.

Lastly, Fig. 3 shows the median (See footnote 7) top, middle, and bottom *DTI and rsfMRI regions* in the learned partition hierarchy. Included are tables that list the brain regions in the bottom level (i.e. last partition) of the hierarchy. These regions have the most similar structural and functional connectivity characteristics. Note: The term *shared* in this figure means in this grouping the same brain region is present in both connectomes.

4 Conclusion

A novel connectome feature selection technique is proposed that uses a hypergraph learning algorithm to identify brain regions that have similar structural and functional connectivity characteristics. Compared to other well-known feature selection techniques, SVM classifiers trained using structural and functional connectome features selected by our method are significantly better than SVM classifiers trained using connectome features selected by a state-of-the-art regression algorithm. Furthermore, since our approach converts a subject specific complete bipartite graph to an incidence matrix, the resulting incidence matrix is very sparse, which in turn greatly improves the space and time complexity of our approach. Visualizations that display brain regions in the top, middle, and bottom partitions in the proposed partition hierarchy show significant structural and functional connectivity differences in ASD and HC subjects and as seen in Fig. 3. Lastly, even though the betweenness centrality node-based connectivity measure is used, our method achieved similar accuracy and PPV classification results (mean ± 3 %) when replaced by the Eigenvector centrality or clustering coefficient connectivity measures.

---

[7]Median value is found using the results from all 10 folds.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

References

1. Greicius MD, Supekar K, Menon V, Dougherty RF. Resting-state functional connectivity reflects structural connectivity in the default mode network. Cereb Cortex. 2009; 19(1):72–78. [PubMed: 18403396]

2. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, Petersen SE. Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. NeuroImage. 2012; 59(3): 2142–2154. [PubMed: 22019881]

3. Rodriguez A, Laio A. Clustering by fast search and find of density peaks. Science. 2014; 344(6191): 1492–1496. [PubMed: 24970081]

4. Rubinov M, Sporns O. Complex network measures of brain connectivity: Uses and interpretations. NeuroImage. 2010; 52(3):1059–1069. [PubMed: 19819337]

5. Rudie J, Brown J, Beck-Pancer D, Hernandez L, Dennis E, Thompson P, Bookheimer S, Dapretto M. Altered functional and structural brain network organization in autism. NeuroImage Clin. 2013; 2:79–94.

6. Saur D, Schelter B, Schnell S, Kratochvil D, Kpper H, Kellmeyer P, Kmmerer D, Klppel S, Glauche V, Lange R, Mader W, Feess D, Timmer J, Weiller C. Combining functional and anatomical connectivity reveals brain networks for auditory language comprehension. NeuroImage. 2010; 49(4):3187–3197. [PubMed: 19913624]

7. Sporns O. The human connectome: origins and challenges. Neuroimage. 2013; 80:53–61. [PubMed: 23528922]

8. Teipel SJ, Bokde AL, Meindl T, Amaro E, Soldner J, Reiser MF, Herpertz SC, Moller HJ, Hampel H. White matter microstructure underlying default mode network connectivity in the human brain. Neuroimage. 2010; 49(3):2021–2032. [PubMed: 19878723]

9. Tibshirani R. Regression shrinkage and selection via the lasso. J Roy Stat Soc B. 1994; 58:267–288.

10. Von Luxburg U. A tutorial on spectral clustering. Stat Comput. 2007; 17(4):395–416.

11. Zhu D, Li K, Terry DP, Puente AN, Wang L, Shen D, Miller LS, Liu T. Connectome-scale assessments of structural and functional connectivity in MCI. Hum Brain Mapp. 2014; 35(7): 2911–2923. [PubMed: 24123412]
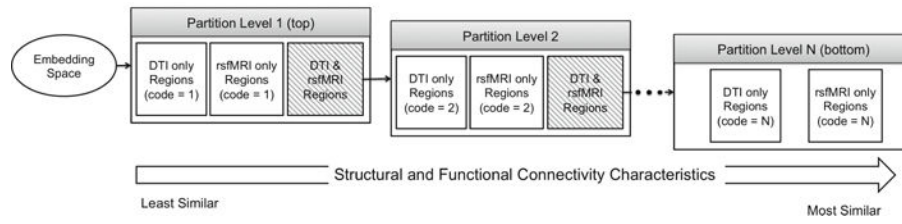
Fig. 1.
Hierarchical partition approach. Each partition level in the hierarchy has a unique integer code value that represents structural and functional connectivity similarities between brain regions.
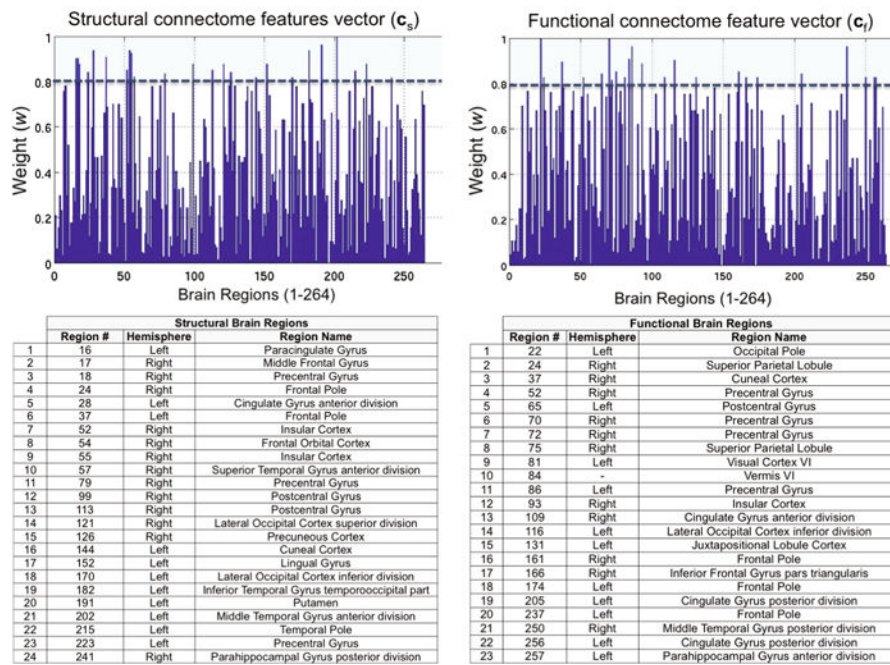
Structural connectome features vector (c_s)

Functional connectome feature vector (c_f)

	Structural Brain Regions		
	Region #	Hemisphere	Region Name
1	16	Left	Paracingulate Gyrus
2	17	Right	Middle Frontal Gyrus
3	18	Right	Precentral Gyrus
4	24	Right	Frontal Pole
5	28	Left	Cingulate Gyrus anterior division
6	37	Left	Frontal Pole
7	52	Right	Insular Cortex
8	54	Right	Frontal Orbital Cortex
9	55	Right	Insular Cortex
10	57	Right	Superior Temporal Gyrus anterior division
11	79	Right	Precentral Gyrus
12	99	Right	Postcentral Gyrus
13	113	Right	Postcentral Gyrus
14	121	Right	Lateral Occipital Cortex superior division
15	126	Right	Precuneous Cortex
16	144	Left	Cuneal Cortex
17	152	Left	Lingual Gyrus
18	170	Left	Lateral Occipital Cortex inferior division
19	182	Left	Inferior Temporal Gyrus temporooccipital part
20	191	Left	Putamen
21	202	Left	Middle Temporal Gyrus anterior division
22	215	Left	Temporal Pole
23	223	Left	Precentral Gyrus
24	241	Right	Parahippocampal Gyrus posterior division

	Functional Brain Regions		
	Region #	Hemisphere	Region Name
1	22	Left	Occipital Pole
2	24	Right	Superior Parietal Lobule
3	37	Right	Cuneal Cortex
4	52	Right	Precentral Gyrus
5	65	Left	Postcentral Gyrus
6	70	Right	Precentral Gyrus
7	72	Right	Precentral Gyrus
8	75	Right	Superior Parietal Lobule
9	81	Left	Visual Cortex VI
10	84	-	Vermis VI
11	86	Left	Precentral Gyrus
12	93	Right	Insular Cortex
13	109	Right	Cingulate Gyrus anterior division
14	116	Left	Lateral Occipital Cortex inferior division
15	131	Left	Juxtapositional Lobule Cortex
16	161	Right	Frontal Pole
17	166	Right	Inferior Frontal Gyrus pars triangularis
18	174	Left	Frontal Pole
19	205	Left	Cingulate Gyrus posterior division
20	237	Left	Frontal Pole
21	250	Right	Middle Temporal Gyrus posterior division
22	256	Left	Cingulate Gyrus posterior division
23	257	Left	Parahippocampal Gyrus anterior division

Fig. 2.

Bar plots that show the median weight values for each node-based connectome feature (structural and functional) found by Eq. (2). The SVM classifier in Table 1 was trained only using the node-based connectivity values from the selected 47 regions. Approximately 91 % reduction in node-based connectome features
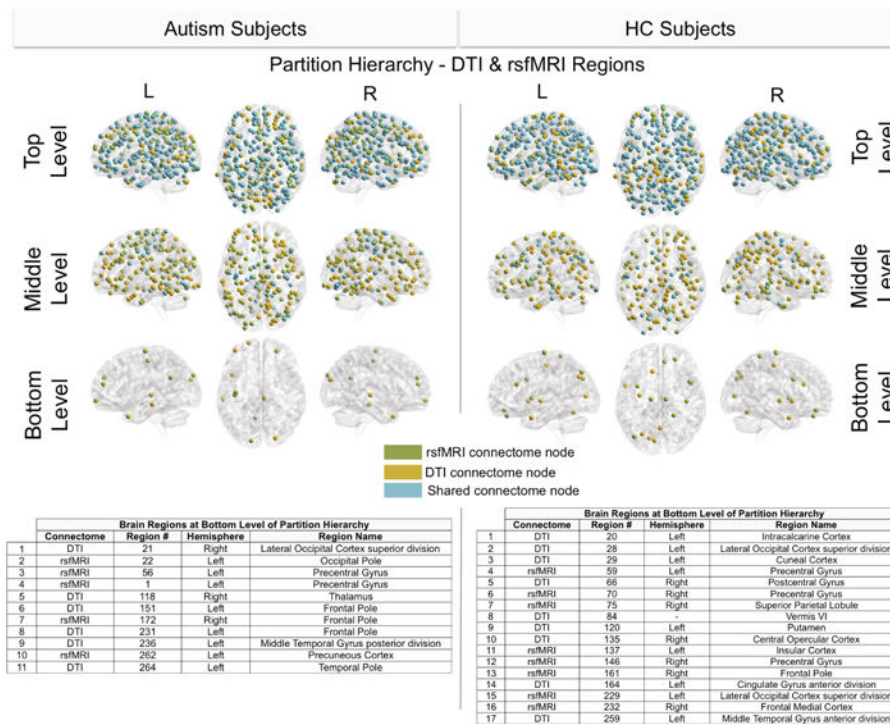
Fig. 3.
Visualizations that show the *DTI and rsfMRI regions* in the top, middle, and bottom partitions (see Fig. 1 for design of partition hierarchy). The tables summarize the brain regions in the bottom partition of the hierarchy.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1

ASD vs. HC 10-fold classification results in $\bar{x} \pm \sigma$ format. The highest performance measures are shown in bold font.

Feature Selection	PPV	ACC	NPV	SEN	SPC
Proposed	**87.2 % ± 2.9**	**88.3 % ± 2.7**	**89.4 % ± 5.8**	**89.4 % ± 6.1**	**86.2 % ± 3.3**
Lasso [9]	74.0 % ± 2.9	73.7 % ± 3.5	73.8 % ± 5.7	72.9 % ± 7.8	74.6 % ± 2.1
None (SVM only)	64.1 % ± 4.1	63.9 % ± 4.3	63.8 % ± 5.1	62.9 % ± 7.3	64.9 % ± 5.0