**PAPER SUMMARY**

Different neuropsychiatric disorders have overlapping symptoms and often also overlapping neural circuits. Diagnostic criteria are able to differentiate between them, but sometimes there are some impairments about the spectrum of symptoms and the categorization into mild to severe, that could be enhanced using neuroimaging-based diagnostic classification and biomarkers to make subgroups.

A good approach is to apply supervised classification methods on functional connectivity features obtained from fMRI to identify brain-based disorders. Most of the studies use a priori clinical diagnosis to guide classification and target only one specific illness.

The authors want to explore the unsupervised methods, such as clustering, to identify neuropsychiatric disorders. They use methods that don’t require a priori specification of the number of clusters, adding also feature selection algorithms due to the high dimensionality of the feature space.

PIPELINE  
ADHD, AD, and ASD data is obtained from publicly available databases. PTSD study was carried out by the researchers.

FOR ADHD:

* 487 subjects selected from the ADHD-200 datset (ADHD-H subject were ignored)

FOR ASD

* 454 subjects selected from the Autism Brain Imaging Data Exchange database.

CONNECTIVITY MEASURES  
Each rs-fMRI image was partitioned into 200 functionally homogeneous regions of interest using spatially constrained spectral clustering. Deconvolution of ROI time series was performed to minimize inter-subject and spatial variability. Computation of 4 connectivity matrices (statistic functional connectivity (SFC), (dynamical functional connectivity (vDFC), statistic effective connectivity (SEC), variance of dynamic effective connectivity (vDEC)). Finally the variability of connectivity over time for every connection was measured and only the features with significant group differences were maintained (computed by ANOVAs testing with p<0.01).

CLUSTERING AND FEATURE SELECTION

Three clustering methods (Hierarchical Clustering, Ordering Points To Identify Clustering, Density Peak Clustering) were applied on connectivity-based features, clinical diagnostic measures and phenotypic.

SITE SPECIFIC ANALYSIS

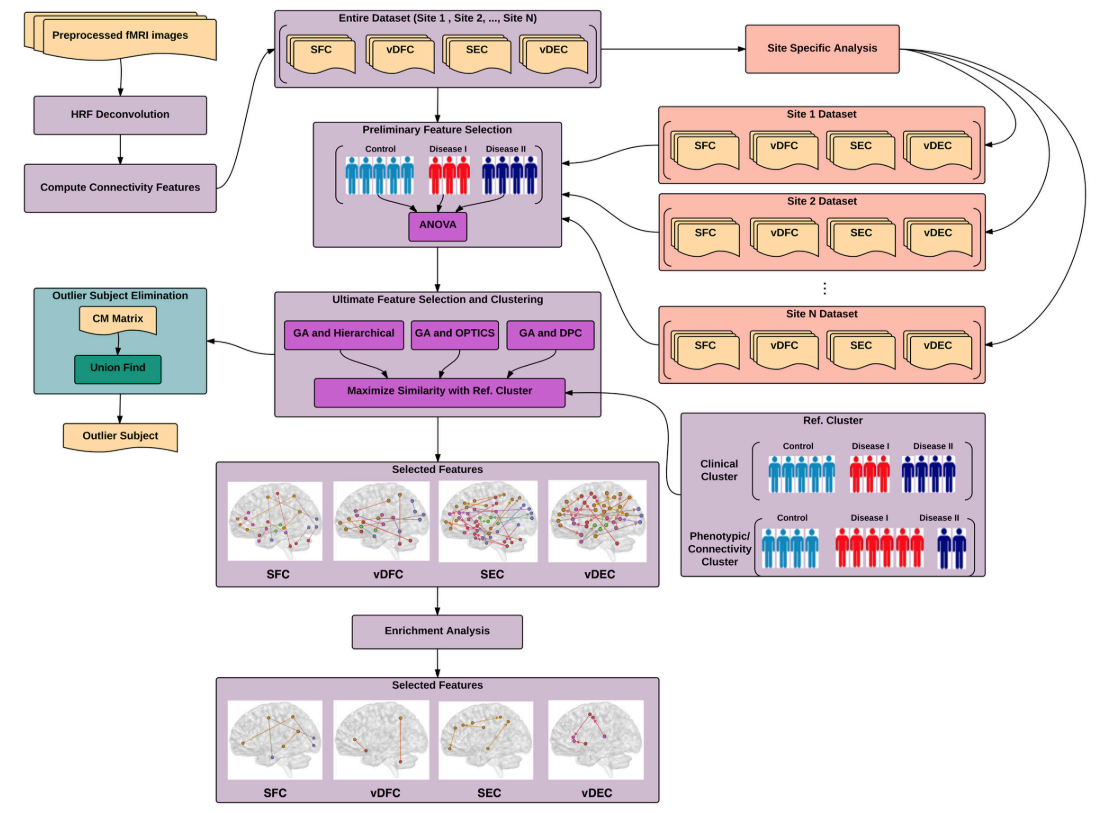
The focus of interest is on the similarity between the clusters obtained from different types of features (connectivity and phenotypic), not the clustering accuracy. However, the clustering can be enhanced eliminating the variance between datasets generated by different operators and places of scanning making a site specific analysis (feature selection for each data imaging center and intersection of the selected features).

ELIMINATION OF OUTLIER SUBJECTS

Co-association matrices between different types of clustering in different data can give better results (more homogeneous). The authors decided to use that approach to find outliers, using the union-find algorithm to eliminate isolated subjects.

ENRICHMENT ANALYSIS

Functional interpretation method, to make quantitative statistical measurements on the association between selected connectivity features and pre-defined functional brain networks.



RESULTS

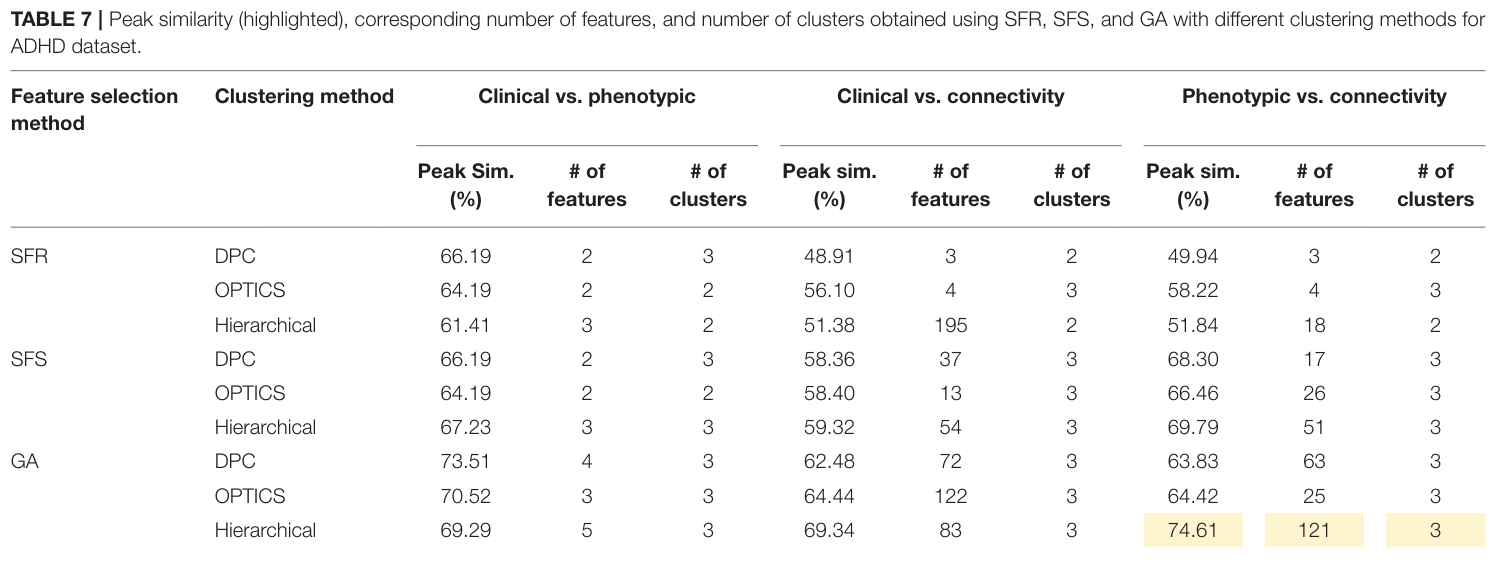
GA selected features (15) gave the highest similarity between cluster obtained from clinical diagnoses, fMRI-based connectivity and phenotypic variables, followed by SFS selected features (84). Both methds show a converfence in the function of the number of iterations in the daraset.

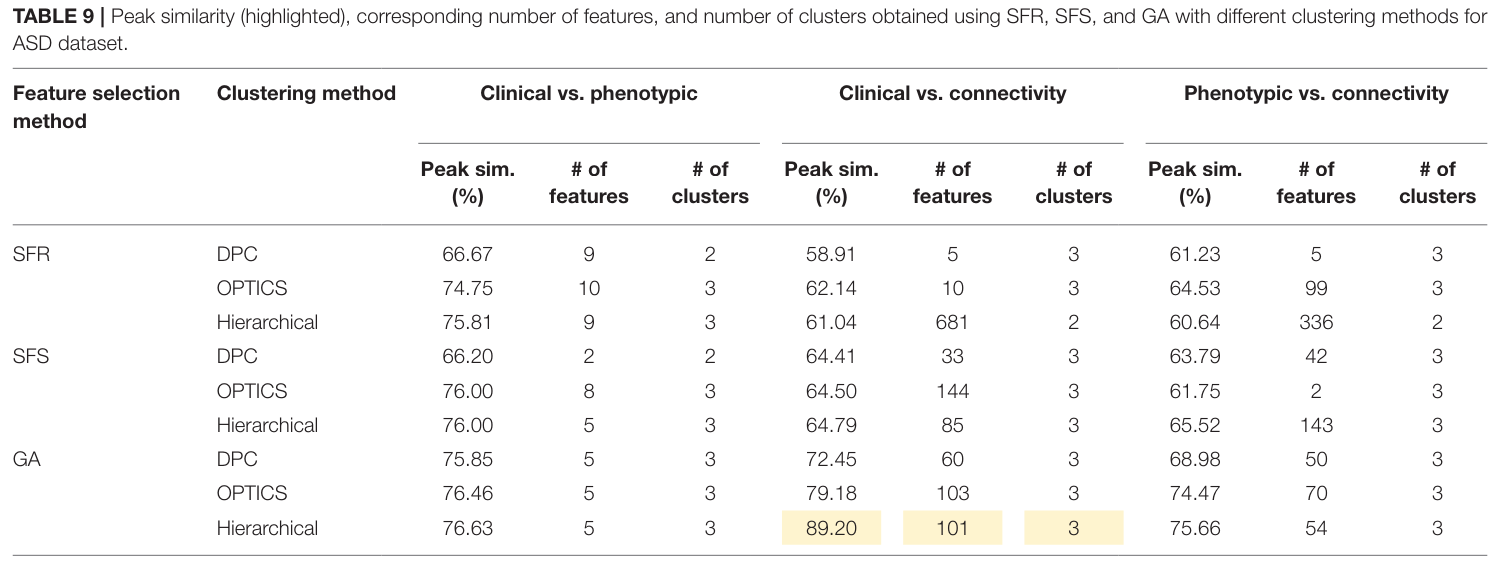
The best clustering method for the ADHD and ASD datasets was the Hierarchical Clustering, while for the other two datasets, OPTICS performed better. The BPD method was the one taking the larger time of computation.

The site-specific analysis was performed only on the ADHD dataset. The similarity of the clusters computed per each site was higher than the one obtained using the whole dataset.

For ADHD (and AD) highest similarity was achieved between connectivity-phenotypic clusters, than for clinical diagnosis-phenotypic clusters (while for ASD (and PTSD) the similarity was higher for connectivity-clinical diagnosis). This suggests that the diagnostic criteria for ADHD is not mapped well onto underlying neurobiological clusters. The authors chose to create new labels for the diagnostic based on the connectivity clusters and verify if the new grouping is valid estimating the statistical separation of phenotypic variables based on the traditional diagnostic grouping and neurobiological grouping. The results (p-value on t-test) shown that in this cases connectivity can be used to map better onto the behavioral phenotypes, than the traditional diagnostic groups.

The similarity achieved after the outlier elimination was higher.





DISCUSSION

Suggestion: Neurobiological and phenotypic biomarkers can be used to improve diagnostic precision and identify diagnostic subgroups.

Results of clustering: Selected connectivity features were split into two networks, in the first network there were significant larger functional/effective connectivities and temporal variability of constituent paths in the control group, while in the second network the significance was larger for the disease group (englobing all sub-groups for simplicity). The both networks were mapped back to the anatomical brain to make qualitative interpretation and comparisons using enrichment analysis.

Intrinsic Connectivity Networks (group of brains regions that show correlated activity at “resting” state): Default Mode Network (DMN), Visual Network (SMN), Basal Ganglia Network (BGN), Sensory Motor Network) Semantic Cognition and Attention (SCAN).

For each pathology the authors chose the features that gave higher similarity. Then using enrichment analysis they selected Network-to-Network interactions for each connectivity matrix. Then plot them divided by the features where the disease has an higher activation than in control and viceversa. In this way they could study the meaning of their findings.

(The purity of clusters can be computed regrouping objects obtained by the connectivity clusters using the diagnostic labels and reassign objects to the majority in the cluster, then accuracy is measured counting the number of correct assignations.)

CONCLUSIONS

“However, these approaches are besieged with methodological issues such as

1. a priori choice of clusters needed in k-means,
2. a stopping criterion needed in hierarchical clustering,
3. the large dimensionality of imaging data necessitates some type of dimensionality reduction for clustering to work properly and this step is either not carried out, or carried out by preselecting features not from the structure in the data, but by some external considerations such as previous findings in a given disorder, and
4. the clusters obtained from imaging data are seldom compared by those obtained from clinical diagnostic criteria or behavioral phenotypes.

To address these four issues, a general pipeline was derived on identifying different brain-based disorders using unsupervised clustering methods.”