A Multimodal Approach to Autism Spectrum Disorder Subtyping

Shreyansh Banthia
Santosh Kolagati
Yuxiao (Shawn) Sun
Yuyi (Taylor) Ding

Introduction & Literature Review

What is Autism Spectrum Disorder (ASD) ?

Developmental disability

1. **Social**
2. **Communication**
3. **Behavioral**

Subtypes – no clear clinical distinction

Literature Review

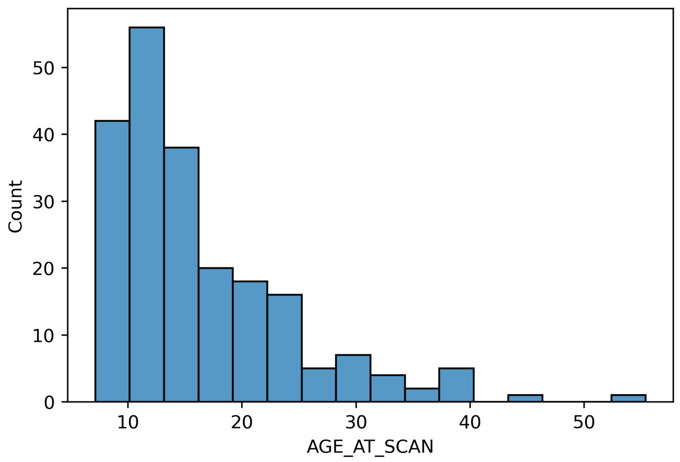
1. **Bioinformatics, MRI, and Behavior Analysis Data**
2. **Potential: Combination of IQ, ADOS, and ADI-R score**
3. **Models: RF, K-mean, and SVM**

Dataset & Modalities

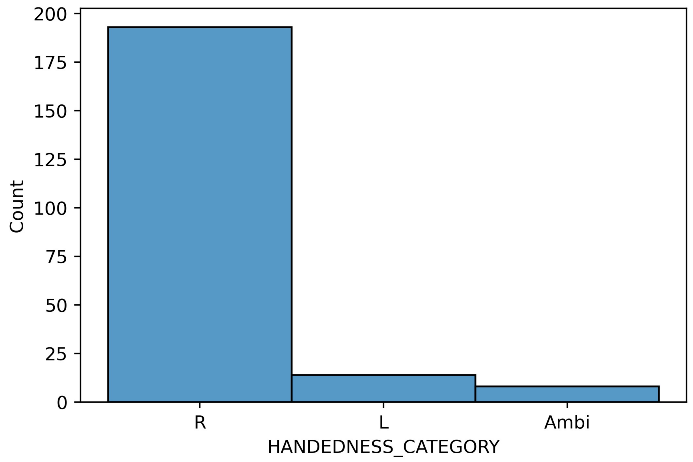
ABIDE 1

1. The data used comes from the ABIDE 1 study which ran across 17 international sites while we use data from CalTech, NYU, CMU, UCLA, Stanford and UPitt.
2. Out of the 539 subjects with confirmed ASD in the dataset, we data from 215 individuals based on the types of modalities available.
3. Below are some plots on the demographic information of the filtered cohort -

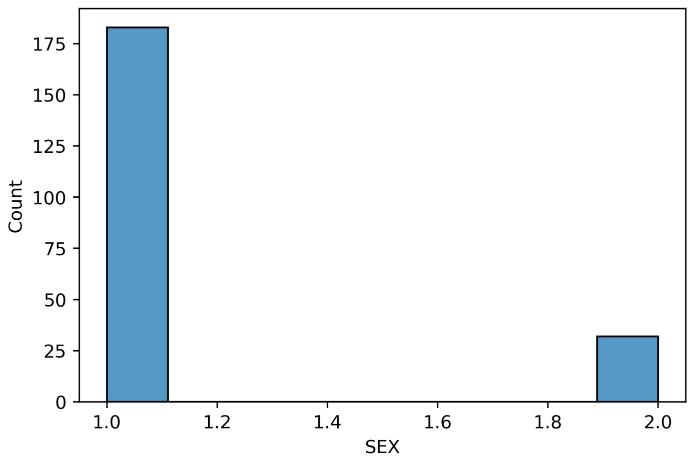
Age Distribution



Handedness Distribution



Gender Distribution



Types of modalities available

1. IQ Test Scores
2. Autism Diagnostic Interview Scores
3. Autism Diagnostic Observation Schedule Scores
4. fMRI Scans Raw Data

IQ Tests

1. The Full Scale IQ (FIQ)

   A score derived from administration of selected subtests from the **Wechsler Intelligence Scales** designed to provide a measure of an individual's overall level of general **cognitive and intellectual functioning**.

2. Verbal IQ (VIQ)

   The ability to **understand** and **reason** using **concepts framed in words**.

3. Performance IQ (PIQ)

   A measure of an individual's overall **visuospatial intellectual abilities**.

Autism Diagnostic Observation Schedule (ADOS)

<- Social Interaction ->

1. Communication
2. Social
3. Stereotypical behavior
4. Gotham Scores

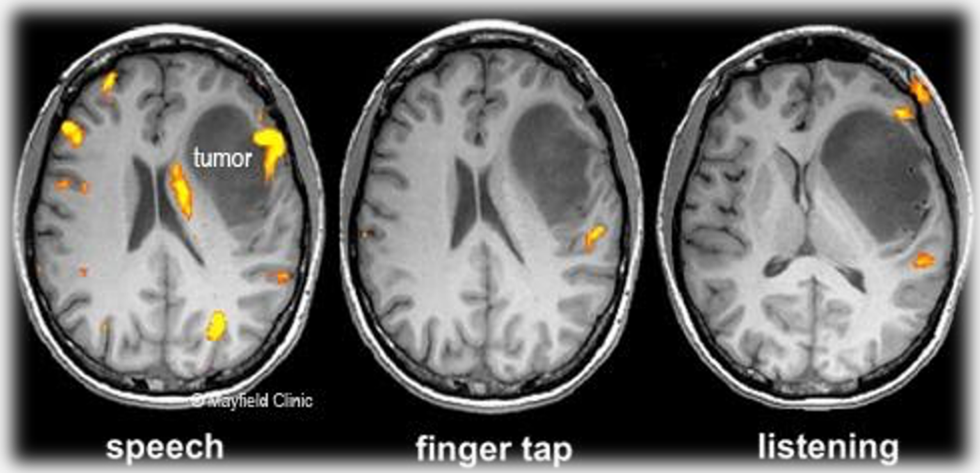
Autism Diagnostic Interview - Revised (ADI)

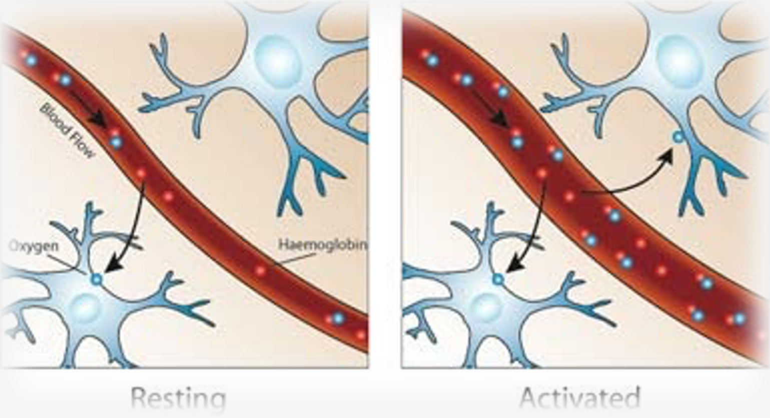
<- Interview ->

1. Social
2. Verbal
3. Restricted & Repetitive Behaviors
4. Onsite Total

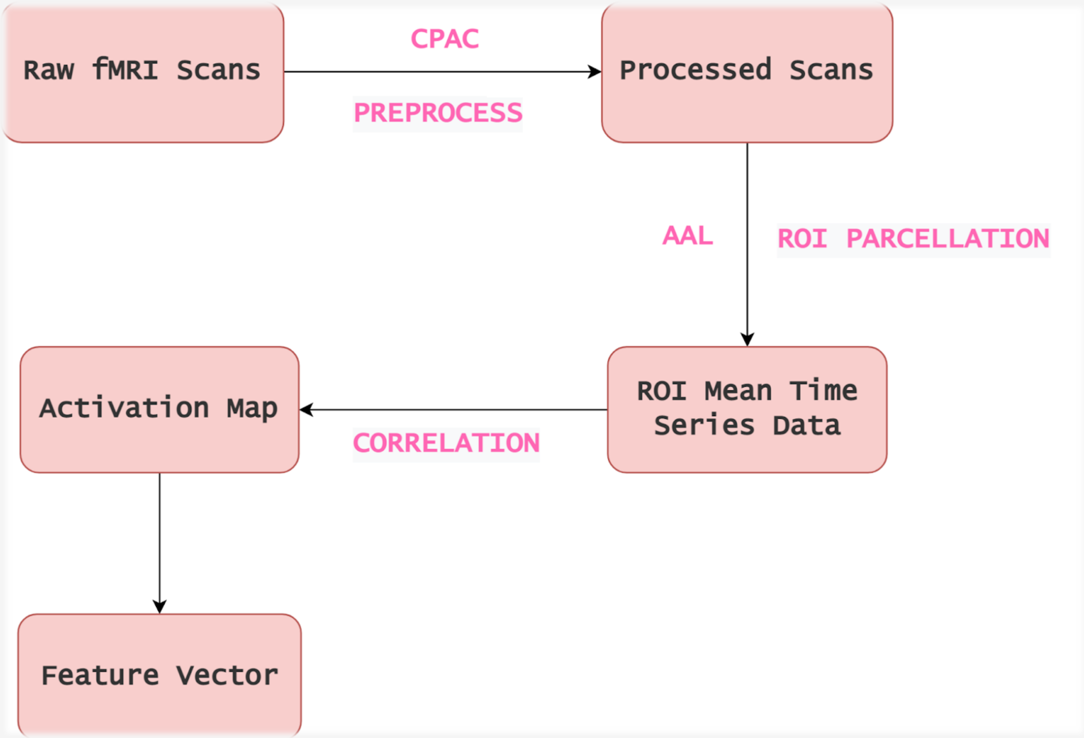
fMRI Scans - Context

1. Measures and maps the brain activity using blood flow
2. Brain requires energy to perform tasks
3. Blood contains glucose & oxygen
4. BOLD: Blood-Oxygen-Level-Dependent imaging
5. Types: Resting state and Task-based.

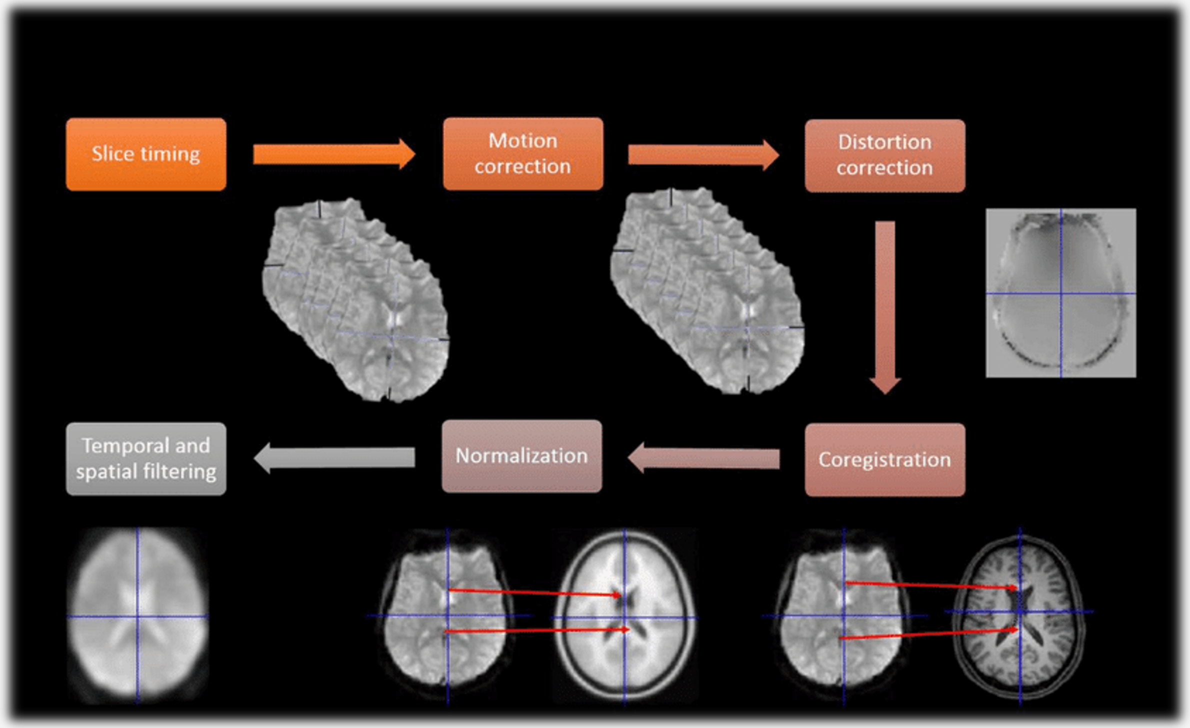




fMRI Data Preprocessing & Feature Engineering



Feature Engineering Steps



Preprocessing Steps

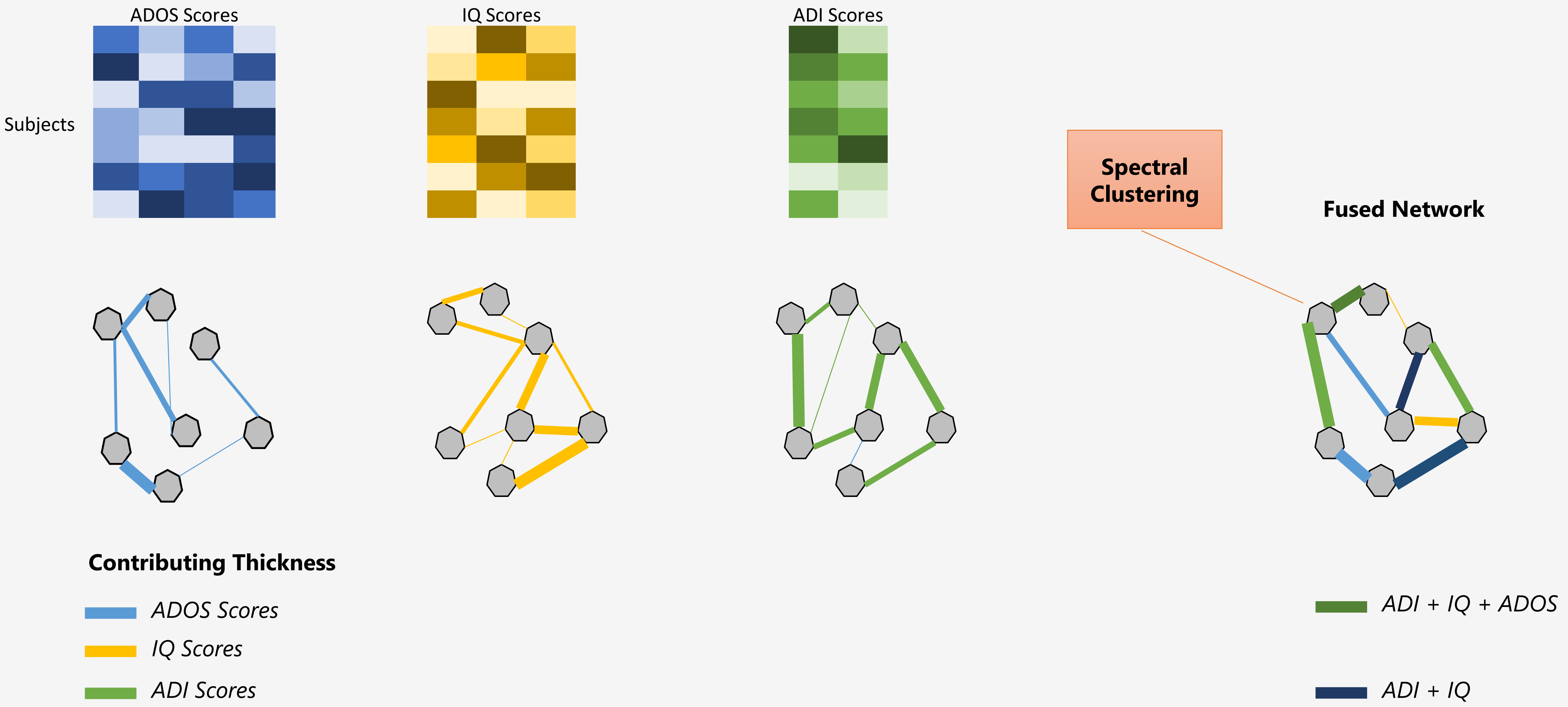
CPAC & AAL

1. CPAC — Configurable Pipeline for the Analysis of Connectomes
2. AAL — Automated Analytical Labelling
3. 116 regions of interest (ROI) are identified
4. Bold signals are averaged in an ROI
5. Activation map is created from mean time series data

Edmund T. Rolls, Marc Joliot, Nathalie Tzourio-Mazoyer, Implementation of a new parcellation of the orbitofrontal cortex in the automated anatomical labeling atlas, NeuroImage,Volume 122, 2015,Pages 1-5, https://doi.org/10.1016/j.neuroimage.2015.07.075.

Methods

Intuition Behind SNF



Math Behind SNF

1. Distance matrices created from feature matrices using a measure like Euclidean / Canberra distance

2. Distance matrices converted into Similarity Networks using a scaled exponential kernel –

$$\mathbf{W}(i,j) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{\rho^2(x_i,x_j)}{2\sigma^2}}$$

$$\sigma = \mu \frac{\overline{\rho}^2(x_i, N_i) + \overline{\rho}^2(x_j, N_j) + \rho^2(x_i, x_j)}{3}$$

3. Similarity matrices must be normalized before fusing while accounting for instability due to the self-similarity along the diagonal -

$$\mathbf{P}(i,j) = \begin{cases} \frac{\mathbf{W}_{(i,j)}}{2\sum_{k \neq i} \mathbf{W}_{(i,k)}}, & j \neq i \\ 1/2, & j = i \end{cases}$$

4. Assuming that local similarities are more reliable than distant ones, a more sparse weight matrix is calculated -

$$\mathbf{S}(i,j) = \begin{cases} \frac{\mathbf{W}_{(i,j)}}{\sum_{k \in N_i} \mathbf{W}_{(i,k)}}, & j \in N_i \\ 0, & \text{otherwise} \end{cases}$$

Math Behind SNF (cont.)

5 The matrix **P** -> Gives a patient's similarity to all other patients

The matrix **S** -> Gives a patient's K most similar neighbors

6 The matrices are then fused, where we try to make them more similar to each other in each iteration -

$$\mathbf{P}^{(v)} = \mathbf{S}^{(v)} \times \frac{\sum_{k \neq v} \mathbf{P}^{(k)}}{m-1} \times (\mathbf{S}^{(v)})^T, v = 1, 2, ..., m$$

7 Fusion stops once the matrices converge or until a pre-defined number of iterations

North Star Metric, Hyperparameter Tuning & Feature Ranking

1. We use a modified Silhouette Score to assess the quality of clusters –

$$\frac{(b-a)}{\max(b,a)}$$

b -> mean nearest-cluster affinity

a -> mean intra-cluster affinity

2. For hyperparameter tuning, we used a custom implementation of the 'Random Search' algorithm

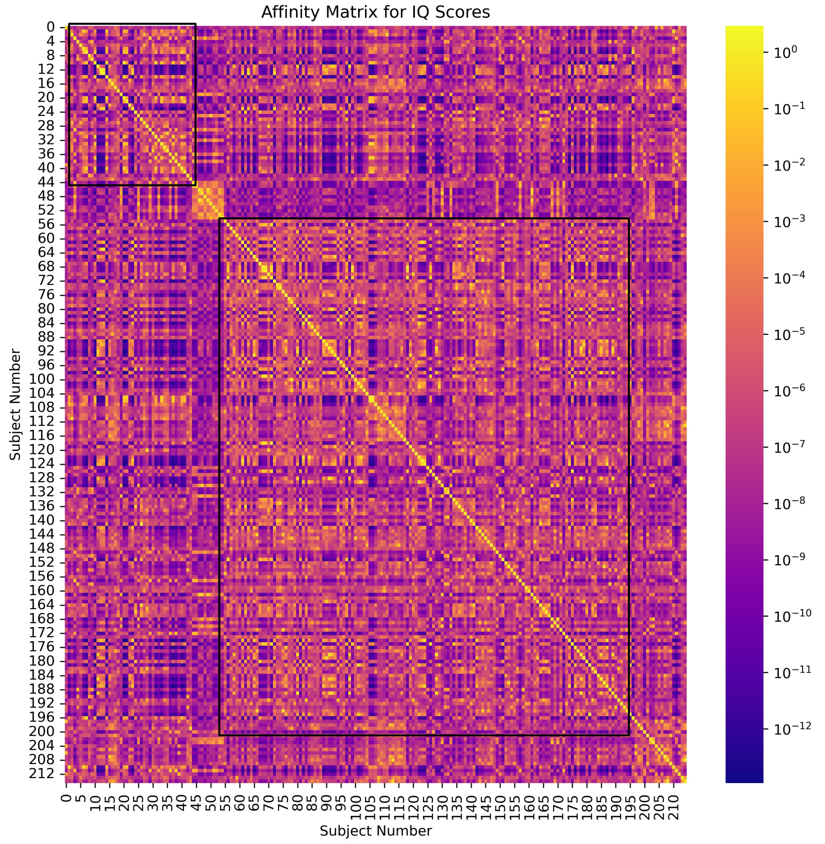
3. To rank features based on their contributions to the independent similarity matrices of each modality we use the idea of Normalized Mutual Information (V-Score) using the pseudo-algorithm given below –

    - Treat the labels from the fused network as the ground truth
    - Cluster using a single feature
    - Calculate the V-Score, comparing the completeness and homogeneity in the clusters formed

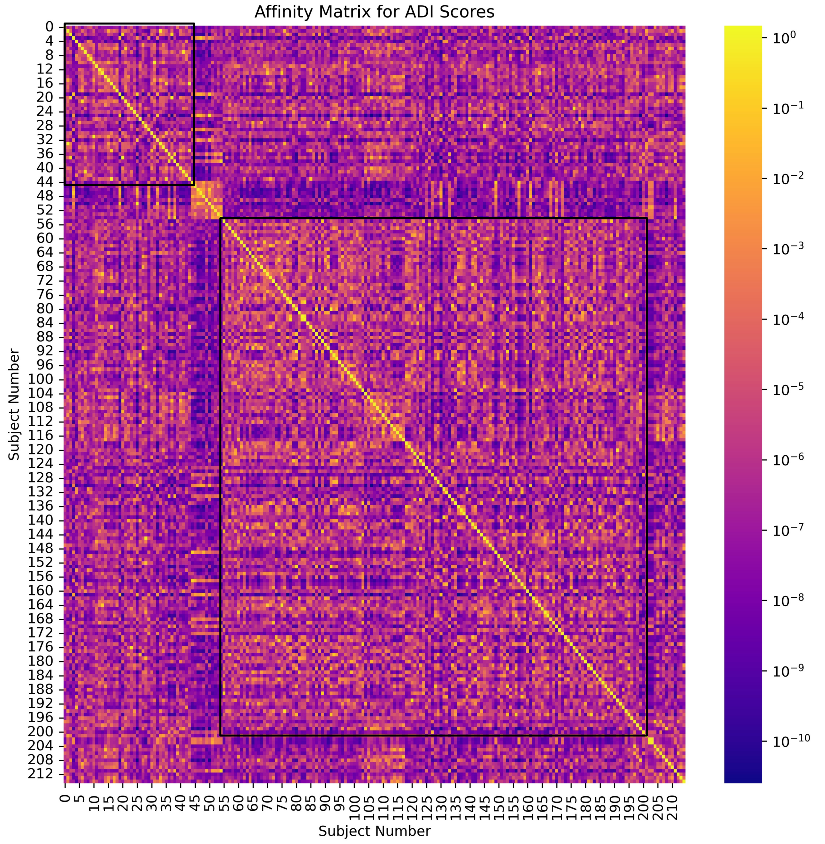
Results

Single Modality Affinity Matrices

IQ Scores

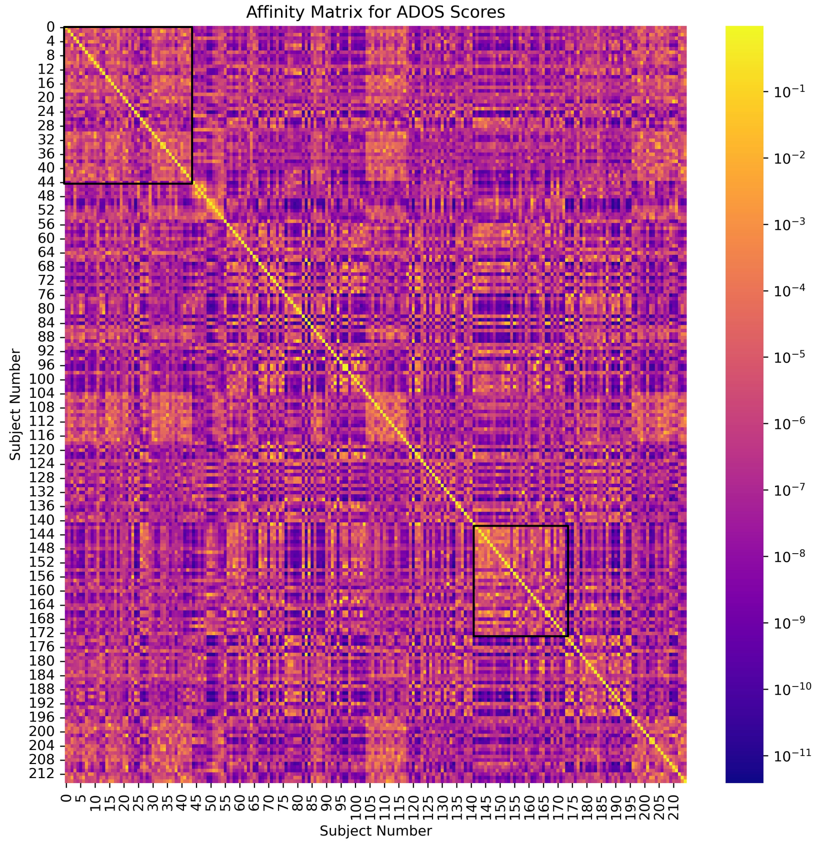


ADI Scores

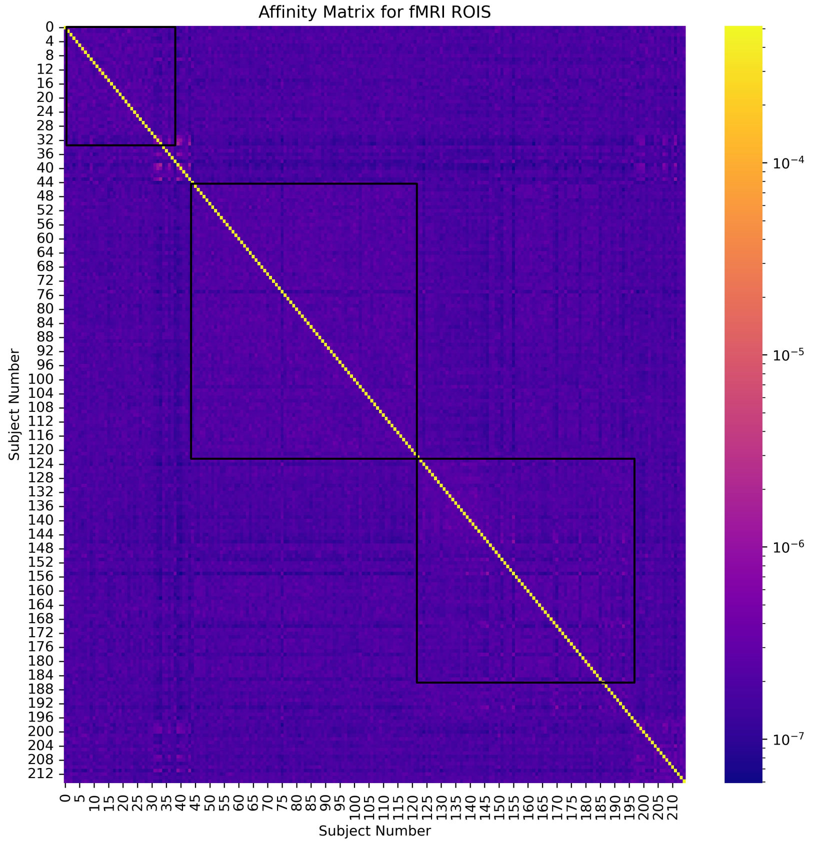


Single Modality Affinity Matrices (cont.)

ADOS Scores

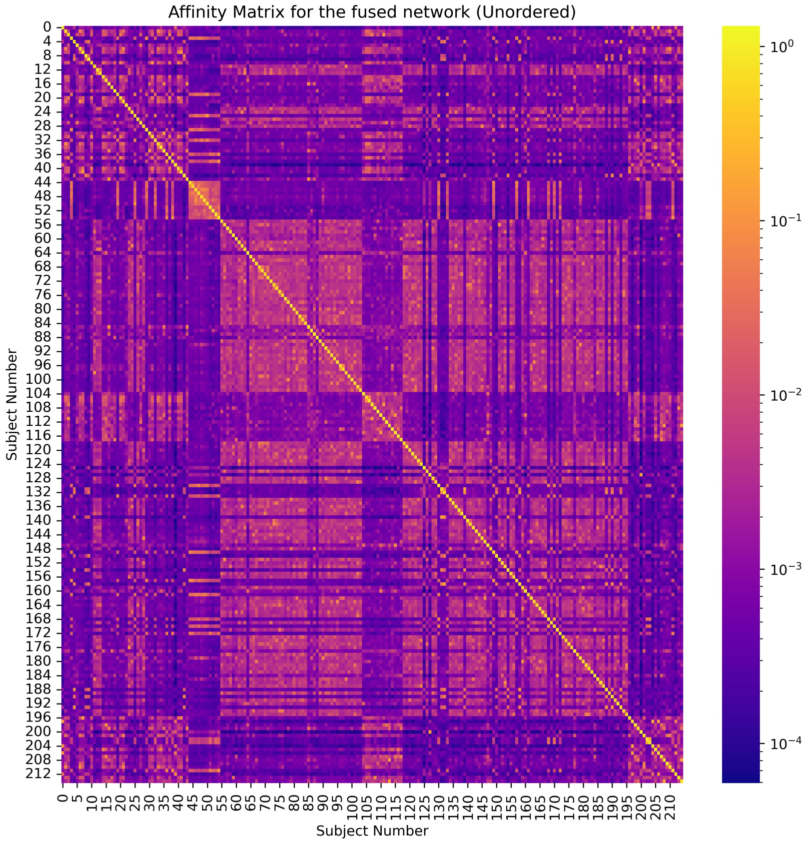


fMRI ROIs

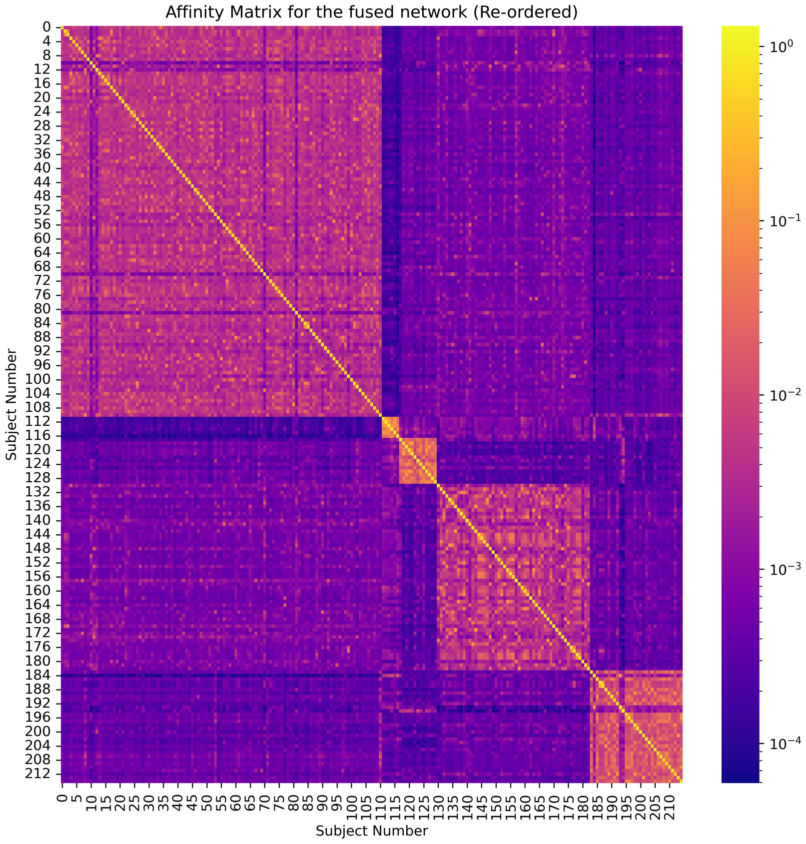


Fused Network Matrices

Fused Network (Unordered)

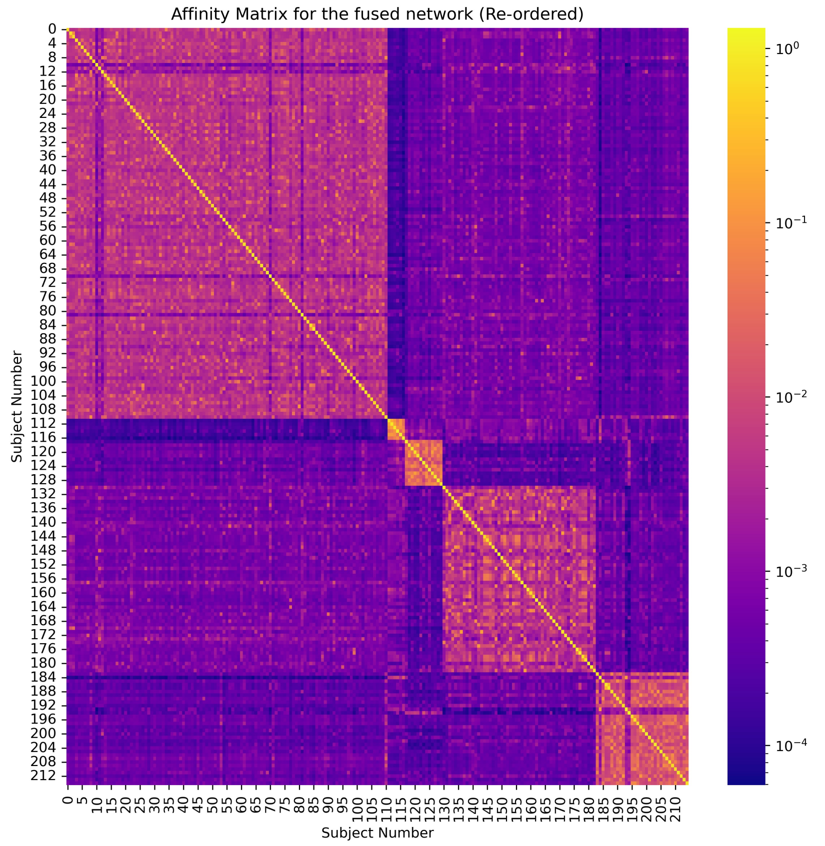


Fused Network (Ordered)



Result Metrics

Fused Network (Ordered)



1. We discovered 5 distinct clusters in our data and the cluster membership was as follows –

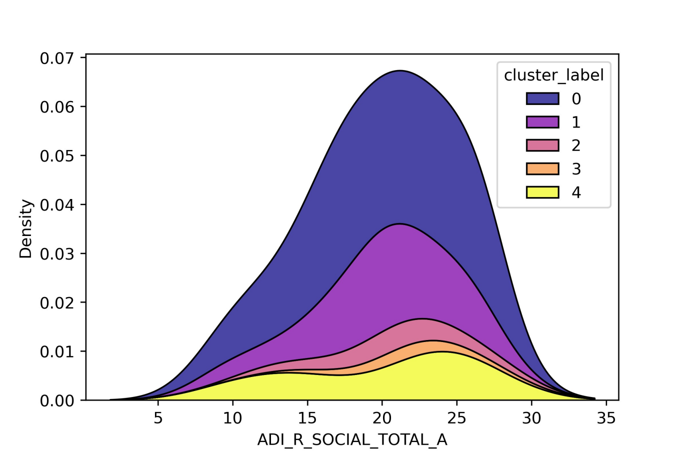
   - Cluster 0 – 111
   - Cluster 1 - 6
   - Cluster 2 - 13
   - Cluster 3 - 53
   - Cluster 4 - 32

2. Silhouette Score – 0.837

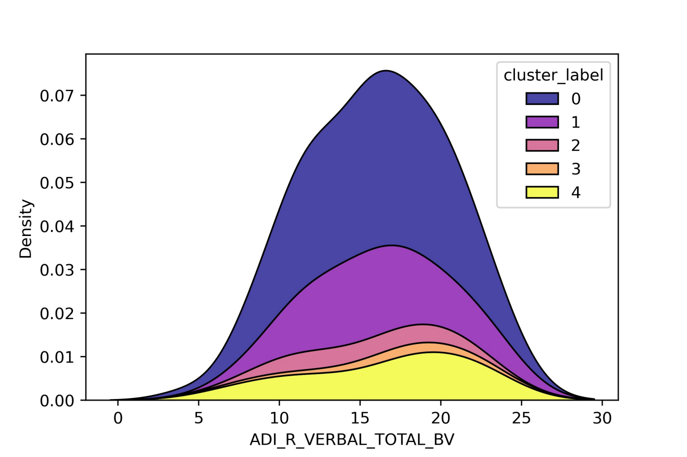
   (Score being closer to +1 means that quality clusters forming)

Cluster Analysis

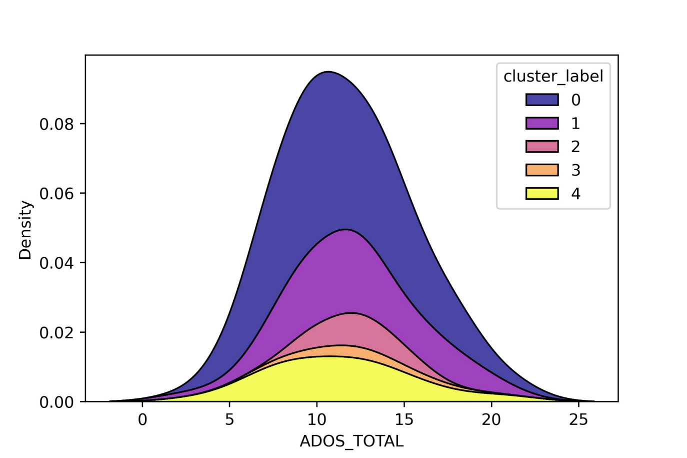
1. ADI Social Score



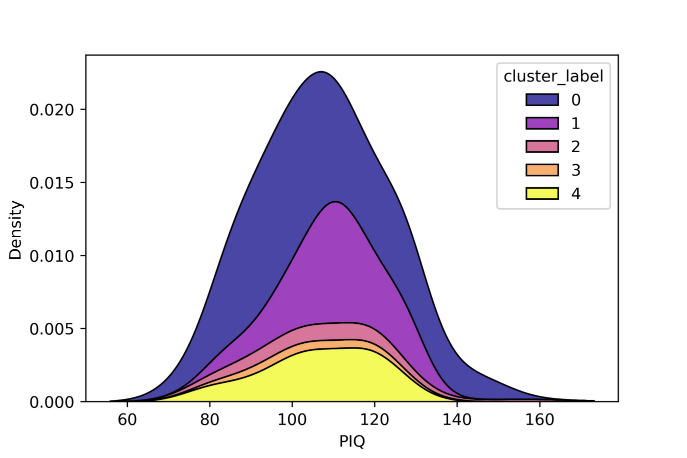
2. ADI Verbal Score



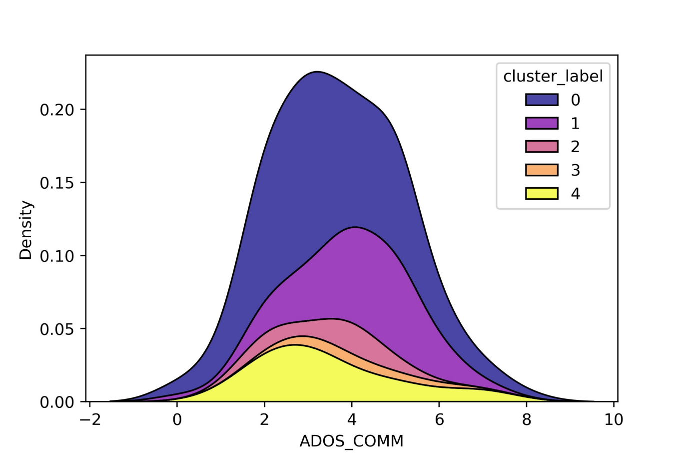
3. ADOS Total Score



4. PIQ Score



5. ADOS Comms Score



Future Work

Future Work

1. Work with clinicians to evaluate the usefulness of the subtypes found
2. Replicate the experiment on the larger ABIDE-II dataset
3. Perform ablative experiments to understand the relevance of each modality
4. Get clinician input on understanding the feature importance
5. Validate the clusters by using best practices & guidance available

Thank You!

Appendix