CLARITY ROI Analysis

February 9, 2015

1 Overview

The goal of this analysis is to show the statistical differences between the different classes of CLARITY brains. For each ROI, fourteen different properties were computed. The analysis shows that it is possible to demonstrate statistical differences between the various classes of CLARITY brains when using unsupervised clustering methods.

2 Results

None of the methods we experimented with (see Appendix A) achieved perfect clustering between the three classes. Therefore, we explored to clustering pairs of classes. When looking at pairs of classes at a time, it is apparent that in some cases, only a single feature is needed for classification. The following figure shows all three pairwise combinations. Larger plots can be found in Appendix C.

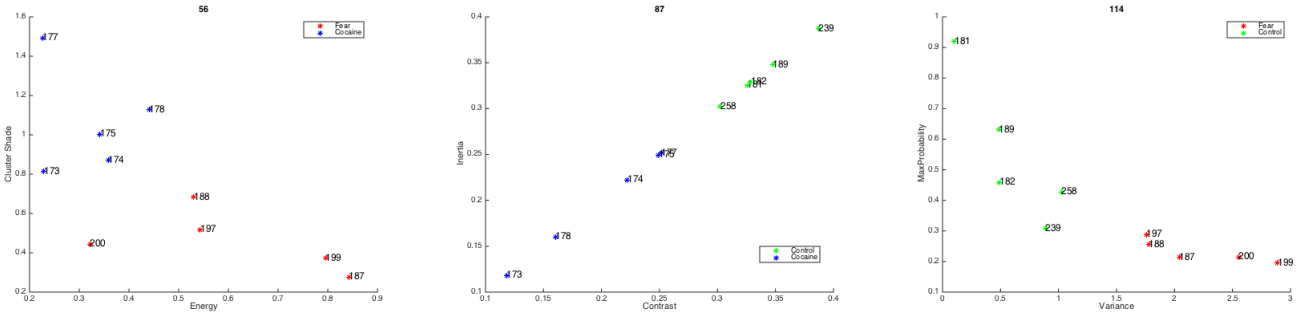

Figure 1: From left to right: Fear vs Cocaine, Control vs Cocaine, Fear vs Control

Cocaine Fear
For the plot of ACB, the ARI score is 1. The p-value is $p < 0.0004$.

Cocaine Control
For the plot of PVHmpd, the ARI score is 1. The p-value is $p < 0.0004$.

Control Fear
Unlike the last two pairs, there were no pairs of features within the ROI's listed in the appendix which showed clear separation between control and fear. LSv had the highest ARI score, 0.597. However, when searching the entire volume, ROI 114 - "Superior olivary complex, lateral part',' was one of many ROIs which had an ARI score of 1. The p-value is $p < 0.0004$.

3 Data

The list of ROIs used in the analysis is included in the appendix.

Each CLARITY Brain volume has 770 slices. Each voxel size is [9.36, 9.36, 9.98] microns, or approximately 10 microns isotropic. There are three classes of brains - Cocaine, Fear and Control. Within Control, there are two subclasses, Control A and Control B (home cage). The table below lists each brain with it's class label.

Class	BrainNo
Cocaine	174
Cocaine	175
Cocaine	178
Cocaine	173
Cocaine	177
Control Homecage	239
Control Homecage	258
Control A	181
Control A	189
Control A	182
Fear	187
Fear	199
Fear	188
Fear	197
Fear	200

4 Methods

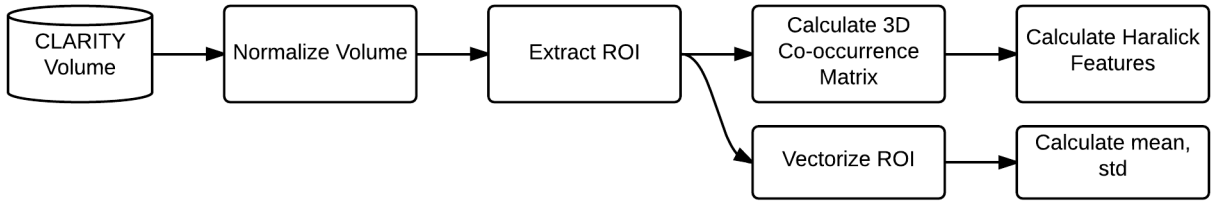

Figure 2: Overview of Analysis Method.

4.1 Normalization

Even within classes, the intensity values for the CLARITY brains have unique means and variations. Therefore, we used z-standardization to normalize the volumes. After masking out each brain volume with the Allen Mouse Brain Atlas, the overall mean of the brain is subtracted out and the volume is divided by the overall standard deviation of the brain.

```
normalized_volume = (volume - mean(volume)) / std(volume);
```

We choose z-standardization because its merit with normalizing MRI data (Milham, 2013).

4.2 ROI Extraction

Using the Allen Mouse Brain Atlas (AMBA), each ROI is extracted within a bounding cube.

4.3 Statistical Properties

The statistics were computed on the extracted ROIs. For each ROI, volumetric co-occurrence matrixes were computed in all relevant (thirteen) directions. By averaging the matrixes across all directions, the features become approximately rotation invariant.

4.4 Mean and Standard Deviation

Mean and standard deviation were calculated using the built-in functions in MATLAB:

```
mean_roi = nanmean(roi);
std_roi = nanstd(roi);
```

4.5 Gray Level Co-occurrence Matrices (GLCM)

"A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level co-occurrence matrix. The GLCM functions (also known as Haralick Features) characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix" (MATLAB). For example, when considering diagonal intensity value pairs, entry c_{ij} in the co-occurrence matrix is a count of the number of times that $F(x, y) = i$ and $F(x + 1, y + 1) = j$, where F(x, y) is a pixel in the image being analyzed.

Notation: In the following formulas, the variable 'i' is the row index and the variable 'j' is the column index of the co-occurrence matrix. p(i, j) is a value in the co-occurrence matrix. σ_i and σ_j are the standard deviation of the co-occurrence matrix of in the row and column directions, respectively. μ_i and μ_j are the co-occurrence matrix of in the row and column directions, respectively. The explicit formulas are given below. All summations are over the entire co-occurrence matrix (a square matrix with length N). N corresponds to the number of gray levels being analyzed. In this analysis, N = 16.

Prior to calculating any features from the co-occurrence matrix, it must be normalized. For the given co-occurrence matrix p (indexing from 1 to N):

$$\text{norm matrix } = \frac{p}{\sum_{i=1}^{N}\sum_{j=1}^{N} p(i,j)}$$

Then,

$$\mu_i = \sum_{i=1}^{N}\sum_{j=1}^{N} i \cdot p(i,j)$$

$$\mu_j = \sum_{i=1}^{N}\sum_{j=1}^{N} j \cdot p(i,j)$$

The standard deviation follows:

$$\sigma_i = \sum_{i=1}^{N}\sum_{j=1}^{N} (i-\mu_i)^2 p(i,j)$$

$$\sigma_j = \sum_{i=1}^{N}\sum_{j=1}^{N} (j-\mu_j)^2 p(i,j)$$

The following are the 12 Haralick features:
Energy: Measures the joint probability occurrence of the specified pixel pairs.

$$Energy = \sum_{i}^{N}\sum_{j}^{N} p(i,j)^2$$

Entropy: Measures the randomness of a gray-level distribution

$$Entropy = -\sum_{i}^{N}\sum_{j}^{N} p(i,j)\log(p(i,j))$$

Correlation: Measures the joint probability occurrence of the specified pixel pairs.

$$Correlation = \sum_{i}^{N}\sum_{j}^{N} \frac{(i-\mu_i)(j-\mu_j)p(i,j)}{\sigma_i, \sigma_j}$$

Contrast: Measures the local variations in the gray-level co-occurrence matrix.

$$Contrast = \sum_{i}^{N}\sum_{j}^{N} |i-j|^2 p(i,j)$$

Variance: variance of gray levels in the co-occurrence matrix

$$Variance = \frac{1}{2}\sum_{i}^{N}\sum_{j}^{N} (i-\mu_i)^2 p(i,j) + (j-\mu_j)^2 p(i,j)$$

SumMean: Provides the mean of the gray levels in the image.

$$SumMean = \frac{1}{2}\sum_{i}^{N}\sum_{j}^{N} (i+j) p(i,j)$$

Inertia: Measures change in intensity value pairs over the co-occurrence matrix

$$Inertia = \sum_{i}^{N}\sum_{j}^{N} (i-j)^2 p(i,j)$$

Cluster Shade: Measures of the skewness of the matrix

$$Shade = \sum_{i}^{N}\sum_{j}^{N} (i+j-\mu_i-\mu_j)^3 p(i,j)$$

Cluster tendency: Measures the grouping of pixels that have similar graylevel values

$$Tendency = \sum_{i}^{N}\sum_{j}^{N} (i+j-\mu_i-\mu_j)^4 p(i,j)$$

Homogeneity: Measures the closeness of the distribution of elements in the GLCM to the GLCM diagonal.

$$Homogeneity = \sum_{i}^{N}\sum_{j}^{N} \frac{p(i,j)}{1+|i-j|}$$

MaxProbability: Probability of the intensity pair which occurs the most

$$MaxProb = \text{ max(Co-Occurrence Matrix)}$$

Inverse Variance: Measures smoothness of the image

$$InvVar = \sum_{i}^{N}\sum_{j}^{N} \frac{p(i,j)}{(i-j)^2}, \text{for } i \neq j$$

4.6 Clustering

For this analysis, we experiment with a variety of methods, listed below. To evaluate the "goodness" of clustering, we used the Adjusted Rand Index (ARI). A score of 1 indicates perfect clustering, 0 indicates the distribution is 'random.'

$$ARI = 1 - \frac{Index - ExpectedIndex}{MaxIndex - ExpectedIndex}$$

We used K-Means to cluster the data.

5 Closing Remarks

When only considering one or two features at a time, it is difficult to separate the three classes (or four, if the control group is split into two). When considering the three classes at once, considering pairs of features from different ROIs (method three in appendix A) provides the best results. However, the binary/pair-wise classification results (e.g., Control vs Cocaine) are significantly better. The cocaine volumes are significantly different from the other two classes. Therefore, when taking the two features from one pairwise comparison, and a feature from another pairwise comparison, one gets three features which yields three class separation.

The code is available in the GitHub repo: https://github.com/ctralie/CLARITYTDA

6 Future Work

The next step is to repeat the analysis over all the ROIs present in the Allen Mouse Brain atlas using the original volumes at their native resolution prior to registration.

7 Sources

imadjust: http://www.mathworks.com/help/images/ref/imadjust.html
glcm theory: http://www.mathworks.com/help/images/gray-level-co-occurrence-matrix-glcm.html
glcm: http://www.mathworks.com/help/images/ref/graycomatrix.html
glcm properties: http://www.mathworks.com/help/images/ref/graycoprops.html
glcm properties: http://students.depaul.edu/ dxu/Papers/iasted0601.pdf
cluster shade: http://itc.ktu.lt/itc381/Ion381.pdf
Names of ROIs: Allen Mouse Brain Atlas

8 Appendix A

The following three methods explore clustering all three classes at once. Unfortunately, the highest ARI score was found using 'method 2,' was only 0.51.

8.1 Method 1

For each ROI, we searched for pair of features with the highest ARI score (clustering via K Means). The region showing the highest separation was Nucleus accumbens (ACB) when comparing cluster shade with the inverse variance. The following plot shows clear separation between between the fear and cocaine classes, but the control class is scattered throughout the plot. The ARI value is 0.40118. The p-value is less than 0.0012.

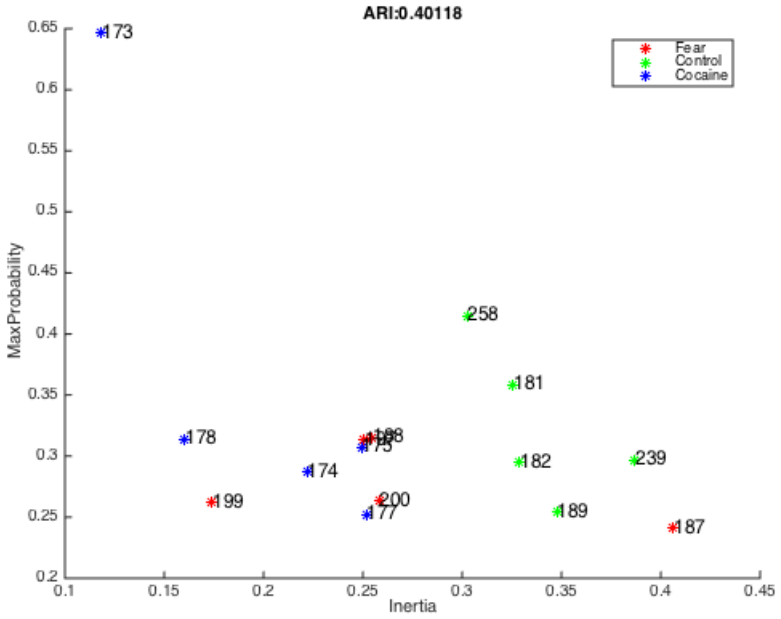

Figure 3: Scatterplot of ROI 56.

8.2 Method 2

For each feature, we searched for a pair of ROIs with the highest ARI score (clustering via K-Means). Then, we looped through all possible (feature, ROI) paris to find the pairs with the highest ARI score. There were a couple of features which received an ARI score higher than 0.4. For the plot below of inertia, the ARI score is 0.40118. The p-value is 0. In this plot, the separation between control and cocaine is clear, while fear is dispersed throughout.

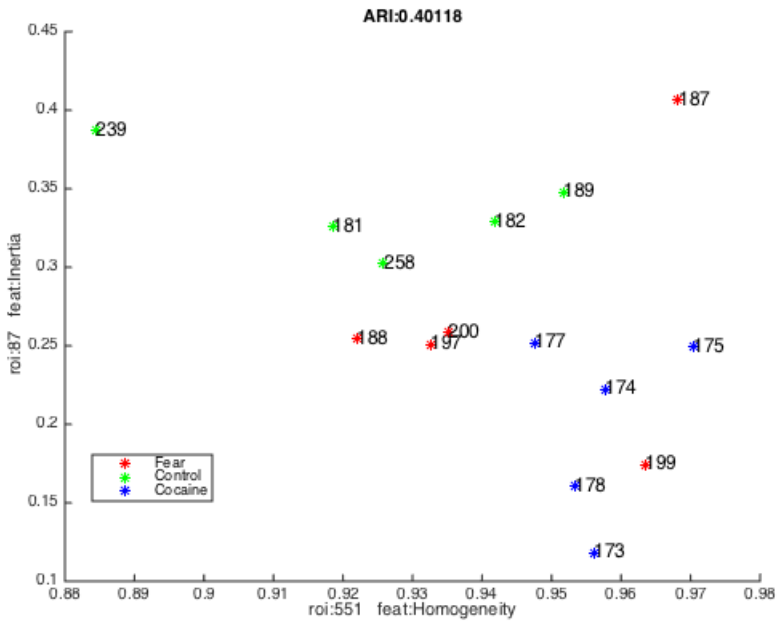

Figure 4: Scatterplot of inertia comparing ROI 551 and ROI 87.

8.3 Method 3

For each feature, we found the two ROIs with the highest ARI score (clustering). For the plot below, the ARI score is 0.411. The p-value is less than 0.0004.

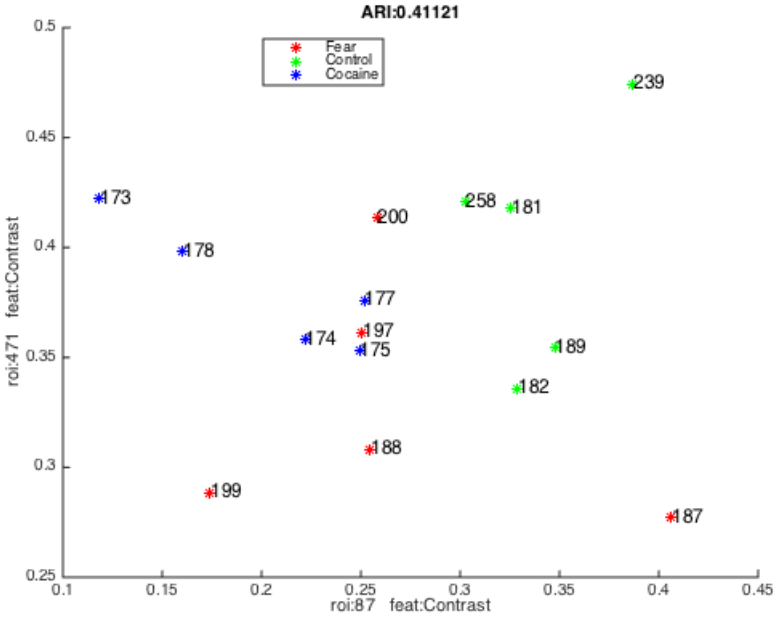

Figure 5: Scatterplots showing separation between cocaine (blue) and control (green).

9 Appendix B

ID	Acronym	Name
56	ACB	Nucleus accumbens
250	LSc	Lateral septal nucleus caudal
258	LSr	Lateral septal nucleus rostral
266	LSv	Lateral septal nucleus ventral part
131	LA	Lateral amygdalar nucleus
544	CEAc	Central amygdalar nucleus capsular part
551	CEAl	Central amygdalar nucleus lateral part
559	CEAm	Central amygdalar nucleus medial part
79	PVHmm	araventricular hypothalamic nucleus magnocellular division medial magnocellular part
652	PVHpml	Paraventricular hypothalamic nucleus magnocellular division posterior magnocellular part lateral zon
660	PVHpmm	Paraventricular hypothalamic nucleus magnocellular division posterior magnocellular part medial zon
55	PVHap	Paraventricular hypothalamic nucleus parvicellular division anterior parvicellular part
87	PVHmpd	Paraventricular hypothalamic nucleus parvicellular division medial parvicellular part dorsal zone
110	PVHpv	Paraventricular hypothalamic nucleus parvicellular division periventricular part
439	PVHdp	Paraventricular hypothalamic nucleus descending division dorsal parvicellular part
447	PVHf	Paraventricular hypothalamic nucleus descending division forniceal part
455	PVHlp	Paraventricular hypothalamic nucleus descending division lateral parvicellular part
464	PVHmpv	Paraventricular hypothalamic nucleus descending division medial parvicellular part ventral zone
50	PRC	Precommissural nucleus
67	INC	Interstitial nucleus of Cajal
587	ND	Nucleus of Darkschewitsch
872	DR	Dorsal nucleus raphe
391	CA1slm	Field CA1 stratum lacunosum-moleculare
415	CA1sr	Field CA1 stratum radiatum
471	CA3slm	Field CA3 stratum lacunosum-moleculare
479	CA3slu	Field CA3 stratum lucidum
486	CA3so	Field CA3 stratum oriens
495	CA3sp	Field CA3 pyramidal layer
504	CA3sr	Field CA3 stratum radiatum
10703	DG-mo	Dentate gyrus molecular layer
10704	DG-po	Dentate gyrus polymorph layer
632	DG-sg	Dentate gyrus granule cell layer
749	VTA	Ventral tegmental area
303	BLAa	Basolateral amygdalar nucleus anterior part
311	BLAp	Basolateral amygdalar nucleus posterior part
451	BLAv	Basolateral amygdalar nucleus ventral part

10 Appendix C

Larger plots from the results section:

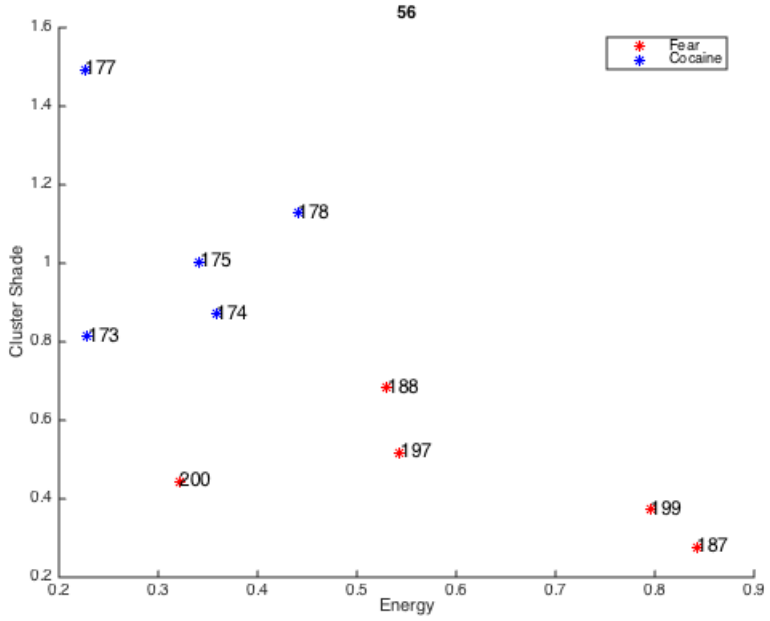

Figure 6: Fear vs Cocaine

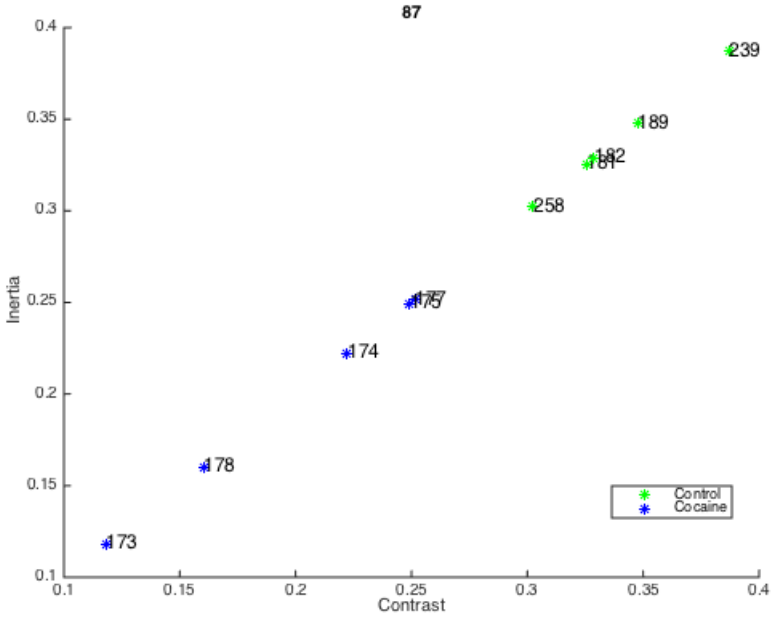

Figure 7: Control vs Cocaine

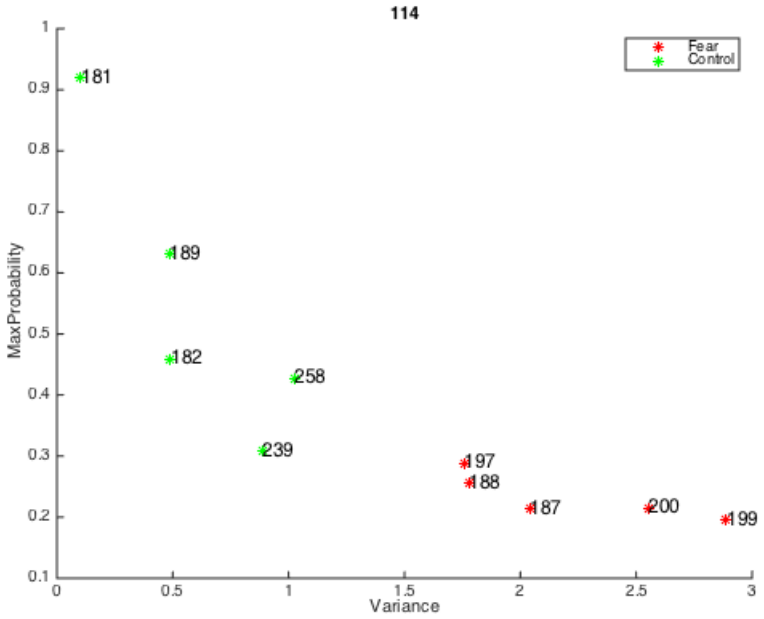

Figure 8: Fear vs Control