Getting CLARITY

Albert Lee
Sahithi Garikapati
Wei Wang

The Art of Data Science
3/10/2016

OPPORTUNITY

- **CLARITY**
  - Chemical transformation of intact biological tissues into a hydrogel-tissue hybrid
  - Studied under light
  - Retains fine structure and native biological molecules
  - Extracts information without disassembly

- Analyzing **CLARITY** data will provide the much-needed across-scales standardization on a global perspective, leading to fundamental insights into psychiatric disease.

http://www.nature.com/nprot/journal/v9/n7/full/nprot.2014.123.html

SIGNIFICANCE

- Psychiatric disease represents the leading cause of disability both in the U.S. and worldwide!
- Lack of circuit-level understanding of psychiatric diseases is causing pharmaceutical companies to shut down psychiatry programs.
- The effort made to study **CLARITY** models has the potential to influence the lives of millions of people across the globe!!!

https://www.nami.org/Learn-More/Mental-Health-By-the-Numbers

GAP

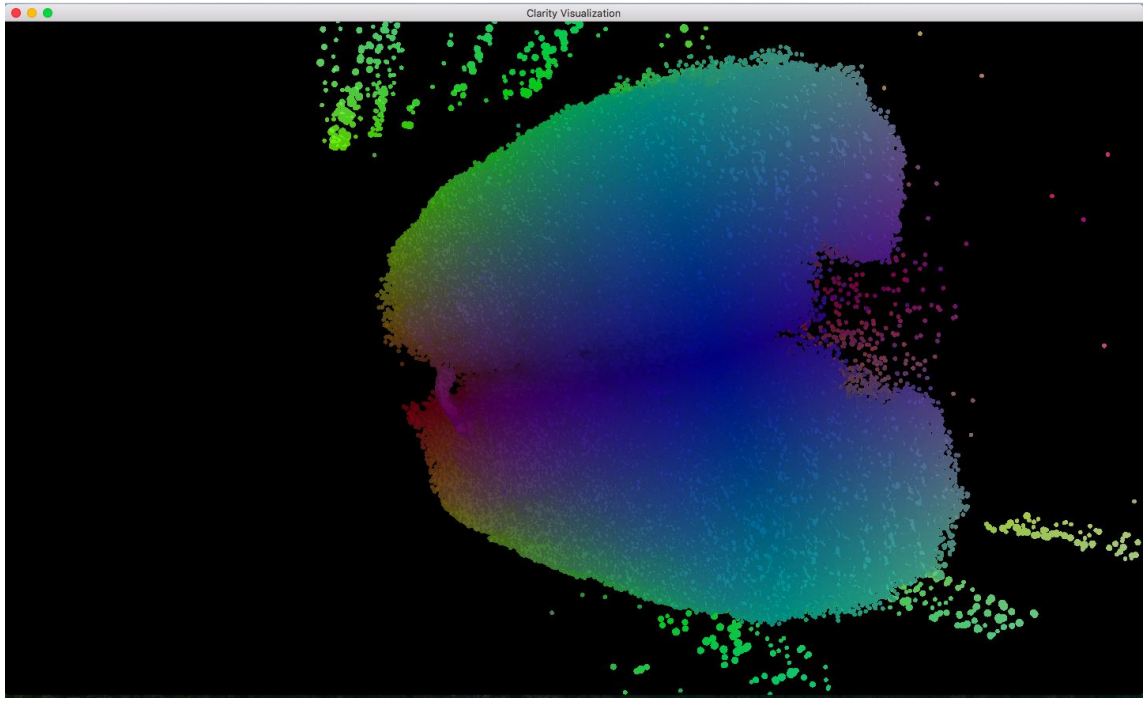
- Presently no other imaging techniques have the ability to link causal information with behavioral significance of connections in the brain.
- No standardized parameters exist to distinguish behavior based on neural structural dynamics.

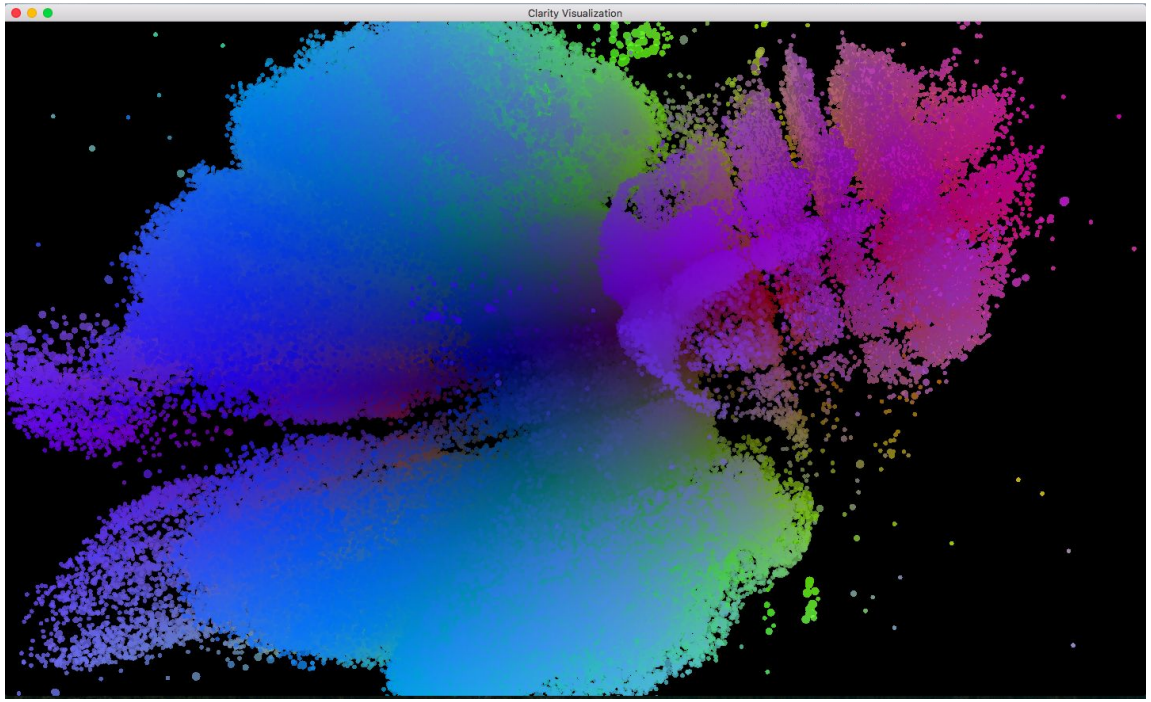
CHALLENGE

- Massive datasets leading to computational burden
- Small number of datasets
- Lack of prior studies on **CLARITY** data

Obtaining statistically significant results may be difficult.

CLARITY image visualization of a rat brain under the influence of cocaine.





FORMAL STATEMENT OF PROBLEM

Model

$F(i,j) \sim$ Clarity image of a particular section

$F(I,k)=\{F(I|k)*F(k)\}$

$F(k)=\{0,1,2\}$ --> 0=Control Image, 1=Cocaine Image, 2=Fear Image

Histogram Equalization

$$F(i,j)=\text{floor}((L-1)\sum F(i,j)Pn)$$

Pn = Number of pixels of intensity n / Total number of pixels

F(I)=mean(F(i,j)) % mean gray value in the image

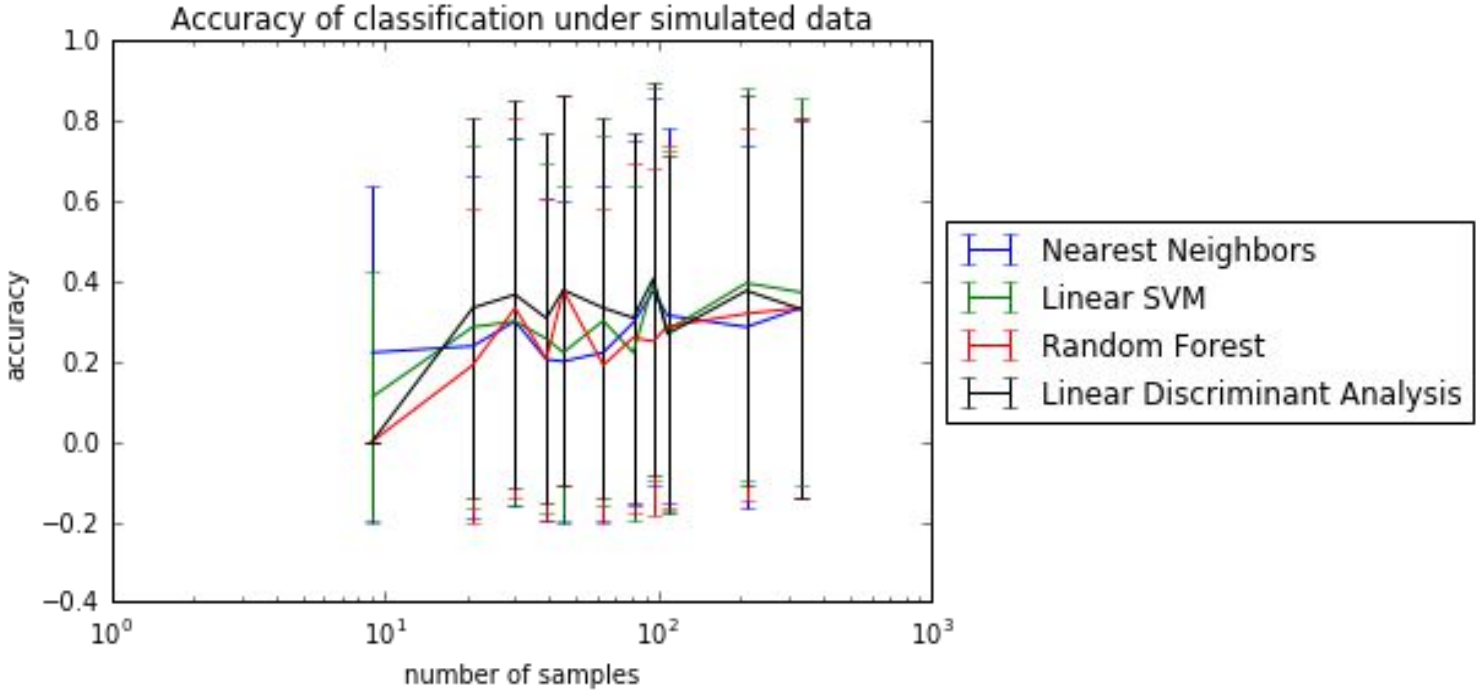
MODEL ASSUMPTIONS

- Histograms are sampled according to: xi~iidF
- Independent data points: FX|0=Norm(μ0,σ0)V×V.
- Class conditional difference across conditions= {Control, Cocaine, Fear}

FORMAL STATEMENT OF ALGORITHM

- Linear Discriminant Analysis = LDA
- Random Forest = RF
- K-Nearest Neighbors = KNN
- Support Vector Machine = SVM

RESULTS



<u>Accuracy of classification with real data:</u>

Accuracy of Nearest Neighbors: 0.00 (+/- 0.00)

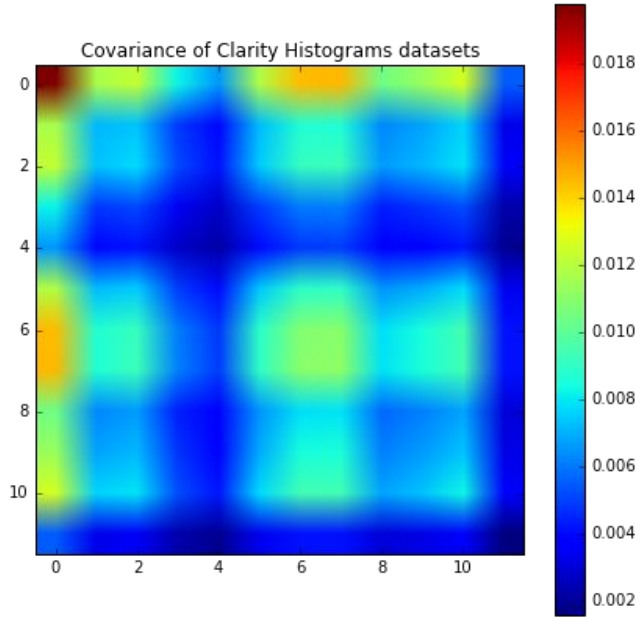
Accuracy of Linear SVM: 0.25 (+/- 0.87)

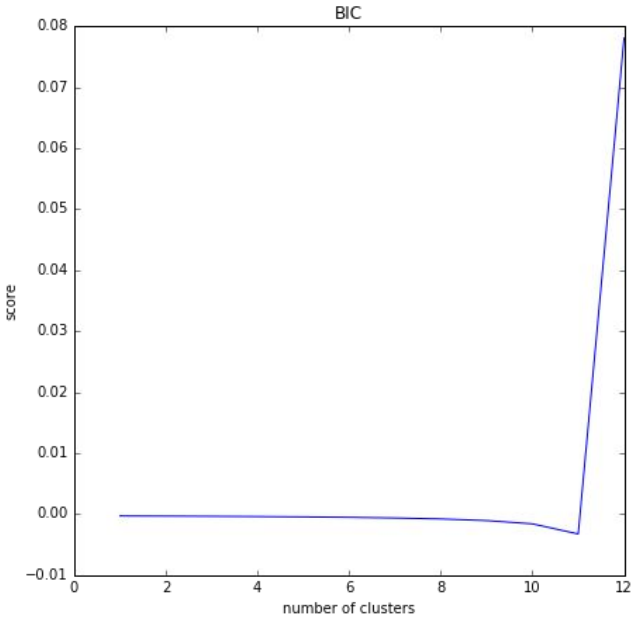
Accuracy of Random Forest: 0.08 (+/- 0.55)

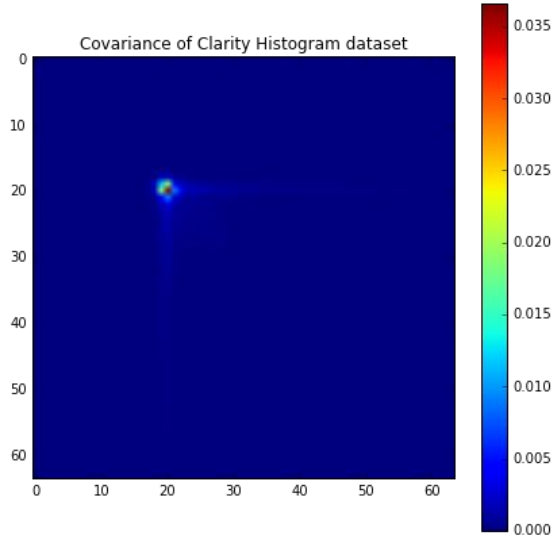
Accuracy of Linear Discriminant Analysis: 0.33 (+/- 0.94)

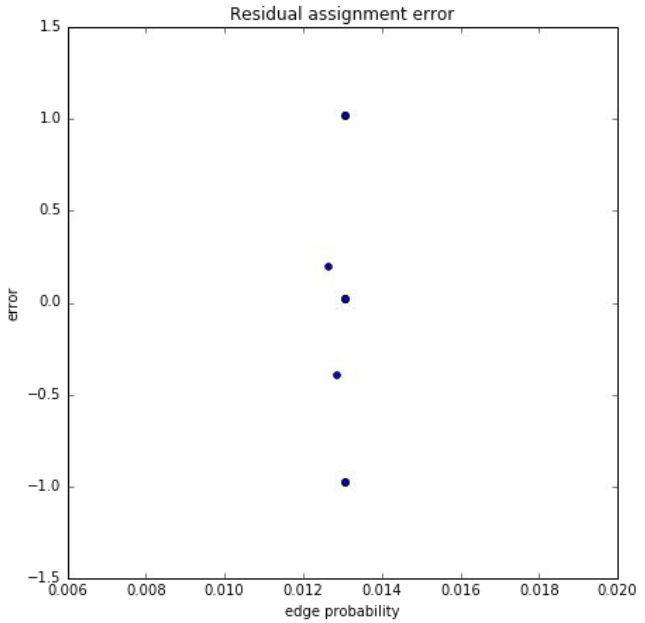
The low accuracy was expected due to the small sample size.

MODEL CHECKING









RESOLUTION

- Because of a small number of datasets most statistical tests do not produce meaningful results.
- We intend to return to exploratory data analysis and re-evaluate our model.
- Additionally, instead of histograms, 3D adjacency matrixes and/or axis-aligned graphs should be considered.