

# MLSP 2014 Schizophrenia Classification Challenge

Machine Learning

December 10, 2019

Tayler Sindhu

# Goal

- Predict schizophrenia classification based on brain imaging data (functional network connectivity correlation values and source-based morphometry standardized weight)

# MRI in Schizophrenia

## fMRI

- Patients with schizophrenia at rest exhibit differences in functional network connectivity compared to controls<sup>1</sup>
- Associated brain regions: bilateral dorsal medial prefrontal cortices, lateral **parietal lobes**, **inferior temporal gyri**
- Areas involved in information processing

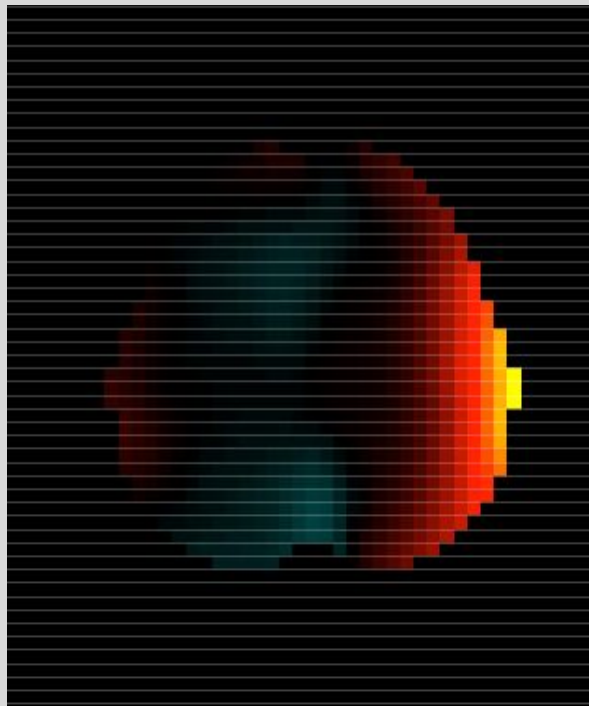
## Structural MRI

- Significant difference in bilateral **temporal lobes** (most significant), basal ganglia, **parietal lobes**, and frontotemporal regions compared to controls<sup>2</sup>

<sup>1</sup>Zhou Y, Liang M, Tian L, et al. Functional disintegration in paranoid schizophrenia using resting-state fMRI. Schizophr Res 2007;97:194-205.

<sup>2</sup>Xu L, Groth KM, Pearlson G, Schretlen DJ, Calhoun VD. Source-based morphometry: the use of independent component analysis to identify gray matter differences with application to schizophrenia. Hum Brain Mapp 2009;30:711-24.

# Functional Network Connectivity Features: Are Resting States Different for Individuals with Schizophrenia?



fMRI performed on subjects at rest

fMRI: voxels and time points

Decomposition of fMRI data with group independent component analysis

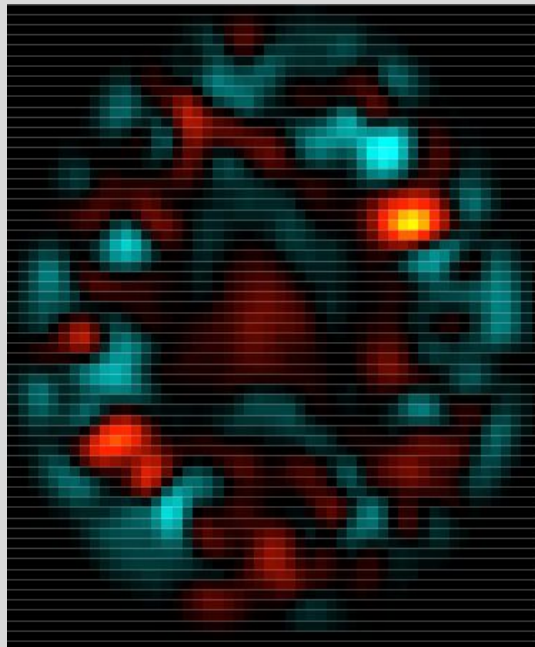
Maximizes statistical independence

Results in 28 aggregate components (brain maps) and corresponding timecourses

Correlation between timecourses calculated for individual subjects

Eichele T, Rachakonda S, Brakedal B, Eikeland R, Calhoun VD. EEGIFT: group independent component analysis for event-related EEG data. Comput Intell Neurosci 2011;2011:129365.

# Source-Based Morphometry Features: “Computational Power”



MRI brain performed on subjects



Decomposition of fMRI data with group independent component analysis



Results in 32 aggregate components  
(grey matter concentration maps)



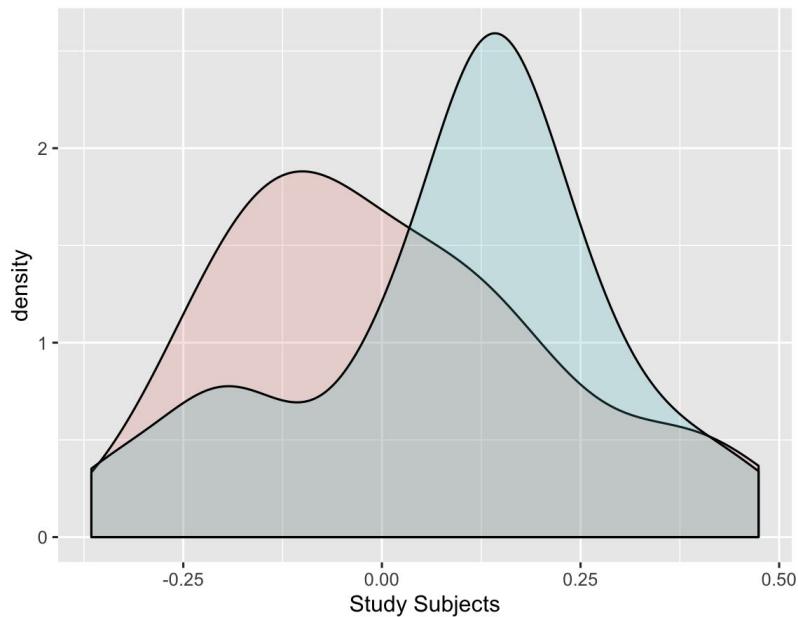
Individual expression levels by brain  
region for individual subjects

# Data Distribution

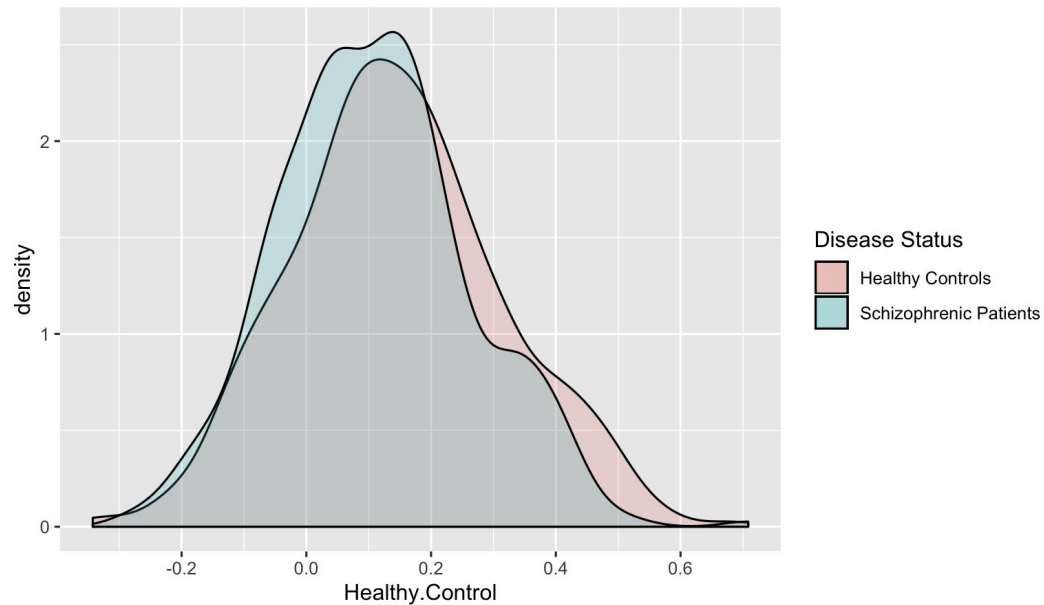
- No missing values
- Training: 86 samples (46 cases, 40 controls)
- Test: 119,748 samples (superficially expanded from 58 patients)
- 410 predictors (32 SBM, 378 FNC)
- High dimensionality

# Feature Distribution by Class

SBM Features: Distribution of Means



FNC Features: Distribution of Means



# Data Analysis Steps

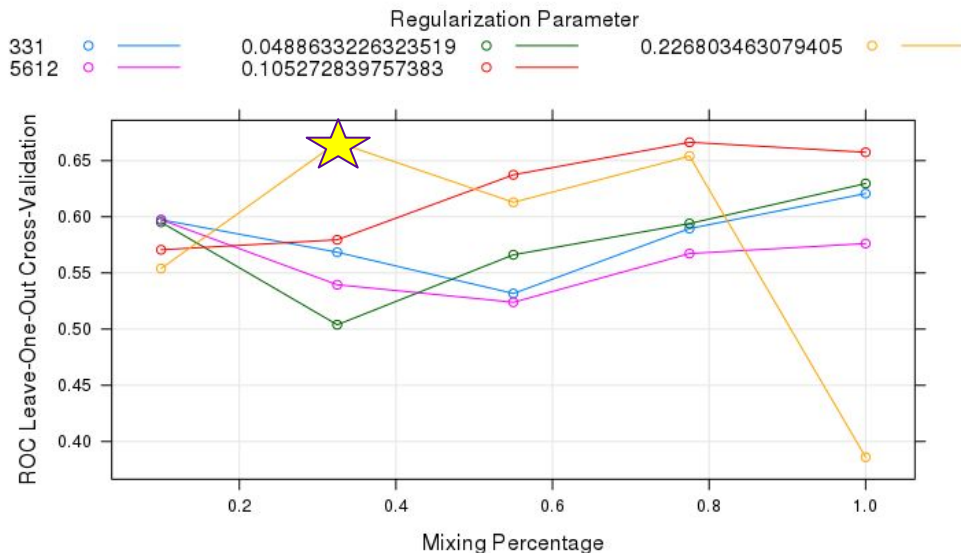
- Compile dataframes
- **Split data** into training and test sets
- Center and scale (depending on the model)
- Assess for **near-zero variance predictors** (none)
- **Identify correlated predictors** (23 features recommended for removal)
- **Find linear combinations** (324 identified)
- Create new data frame without linear combinations
- Feature selection: **recursive feature elimination**



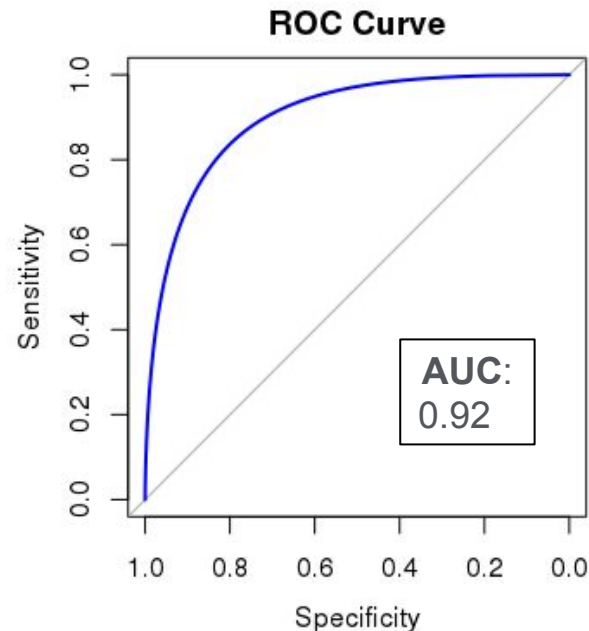
## Models attempted (feature selection not shown)

Model	Test AUC	Test Accuracy (95% CI)
Random forest	0.70	0.73 (0.52, 0.88)
Oblique random forest	0.79	0.77 (0.56, 0.91)
Logistic regression	0.56	0.46 (0.27, 0.67)
Support vector machine with linear kernel	0.65	0.69 (0.48, 0.86)
Support vector machine with radial basis function kernel	0.68	0.69 (0.48, 0.86)
Glmnet	0.69	0.65 (0.44, 0.83)
AdaBoost	0.72	0.65 (0.44, 0.83)

# Best Model: glmnet (Generalized Linear Model via Penalized Maximum Likelihood)

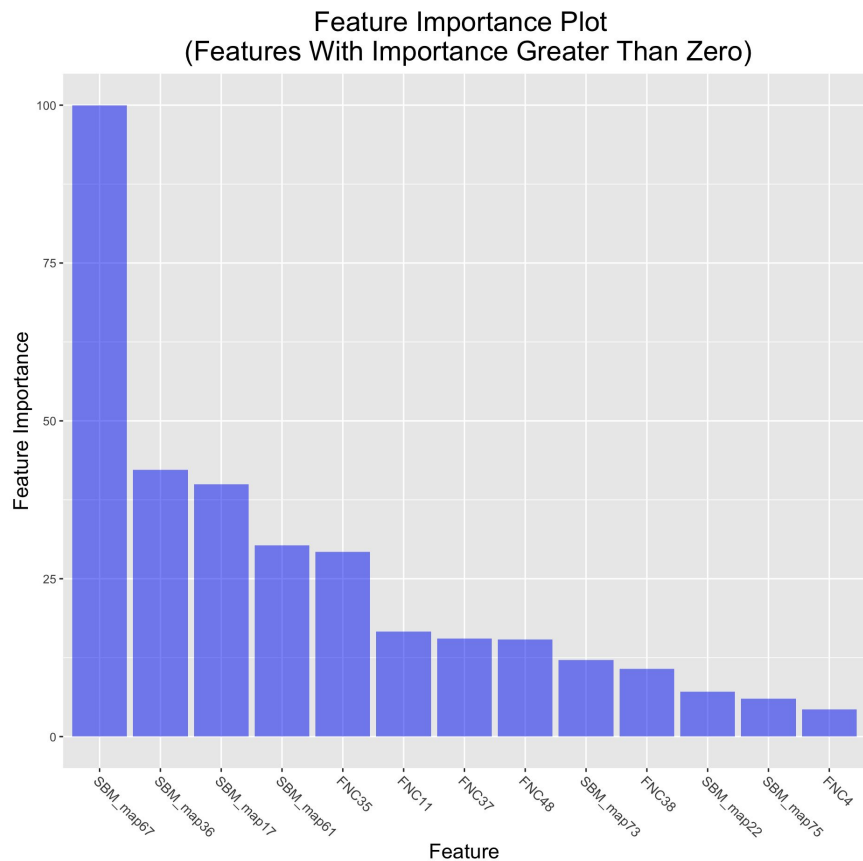


**Tuning parameters:** alpha (mixing percentage), lambda (regularization parameter)



**Test Classification Accuracy:**  
0.85 (95% CI 0.65, 0.96)

# Glmnet with Linear Combination Features Removed



# Confusion Matrix

	Schizophrenic	Healthy
Schizophrenic	10	3
Healthy	1	12

**Sensitivity:** 0.91

**Specificity:** 0.80

**Positive predictive value:** 0.77

**Negative predictive value:** 0.92

# Final Thoughts

- No user was able to surpass an AUC of 0.8 on the test set
- It is possible that clinically, resting states of fMRI data and grey matter concentration matters alone are not sufficiently predictive
- Small dataset, high dimensionality
- Leave-one-out cross-validation
- Bias-variance tradeoff