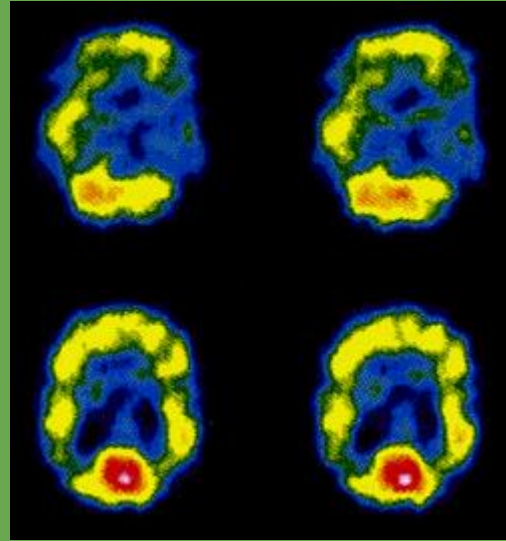


# Insights from SPECT imaging data

Yating Jing  
Srini Suresh



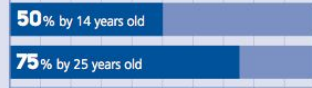
# Significance

## 01. Mental health problems develop at a young age.

1 in 5 children have a mental health problem in any given year.<sup>8</sup>



First experience of mental health problems in those suffering lifetime mental health problems.<sup>9</sup>



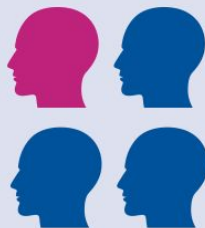
## 04. Mental health impacts on life expectancy.

Average life expectancy in England and Wales for people with mental health problems is 60 years behind the national average.<sup>12</sup>



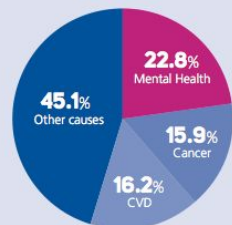
## 02. Mental health is widespread & common.

Every year 1 in 4 adults experience at least one mental disorder.<sup>10</sup>



## 03. Mental health is a significant burden.

Mental ill health is the single largest cause of disability in the UK.<sup>11</sup>



## 05. People with mental health problems have worse physical outcomes.

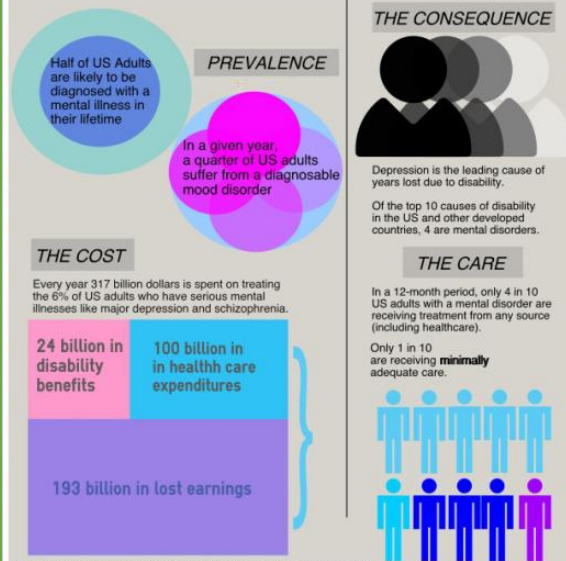
People with mental illness are at increased risk of the top five health killers, including heart disease, stroke, liver and respiratory diseases and some cancers.

### PEOPLE WITH SCHIZOPHRENIA ARE:

**2x** more likely to die from cardiovascular disease,  
**3x** more likely to die from respiratory disease.

## The State of Mental Illness

Over the course of a lifetime, around 50% of US adults are likely to be diagnosed with some form of mental illness. The cost, consequence, and prevalence of mental illness is enormous.



<sup>8</sup>Source: [www.nimh.nih.gov/health/publications/the-numbers-count-mental-disorders-in-america/](http://www.nimh.nih.gov/health/publications/the-numbers-count-mental-disorders-in-america/)  
[www.nimh.nih.gov/health/publications/the-numbers-count-mental-disorders-in-america/](http://www.nimh.nih.gov/health/publications/the-numbers-count-mental-disorders-in-america/)  
[www.who.int/mediacentre/news/releases/2004/bip11\\_tcm000001.html](http://www.who.int/mediacentre/news/releases/2004/bip11_tcm000001.html)

# Gap

- Most of the case studies are based on small sets of patient records, such as the SPECT heart dataset
- Currently, no known existing statistical or machine learning tools is widely applied for SPECT imaging analysis
- Most of the case studies are based on a number of clinical observations or limited clinical experience
- Currently, the machine learning studies on SPECT are based on small sets of patient records (SPECT heart dataset)

# Challenge

- SPECT

- Large number of dimensions (754)
- Noisy observations (lots of NAs etc.)
- Mixed information, needs cleaning and segregation
- Data nature is fat and short post cleaning, 2790 samples with 754 dimensions

# Formal statement of the problem

Patient info and Labels observed:

$$X_i, Y_i \sim \mathcal{F} = F_{X,Y}(\cdot; \theta) : \theta \in \Theta$$

Y is binary label vector:

$$Y_i = 0, 1$$

We estimate loss based on the number of correct assignments.

Loss functions differ for different classification models.

<b>LG</b>	$L = -\log P(y g(X)) = -(y \cdot \log(g(X)) + (1 - y) \cdot \log(1 - g(X)))$
<b>KNN</b>	$L = \sum_{i=1}^n D(X_i y_i = 1, X y = 1) + \sum_{i=1}^n D(X_i y_i = 0, X y = 0)$
<b>SVM</b>	$L = (\max\{0, 1 - y \cdot g(x)\})^2$
<b>RF</b>	$L = \sum_{i=1}^n (g(X_i) - y_i)^2$
<b>QDA</b>	$L = \max\{0, 1 - y \cdot g(x)\}$

$$X_i, Y_i \stackrel{i.i.d.}{\sim} F \in \mathcal{F}$$

$$X_{ij} \sim F_j \in \mathcal{F}$$

## Model Assumptions

We assume that the patients were sampled i.i.d.

We also additionally assume that the features are not identically distributed.

# Formal Statement of Algorithm

- PCA and Sparse PCA
  - 754 → 27
- Post this we trained with a series of classifiers and tested their accuracies with LOOCV.
  - Logistic Regression
  - KNN
  - Linear SVM
  - Random Forest
  - QDA

## The Logistic Function

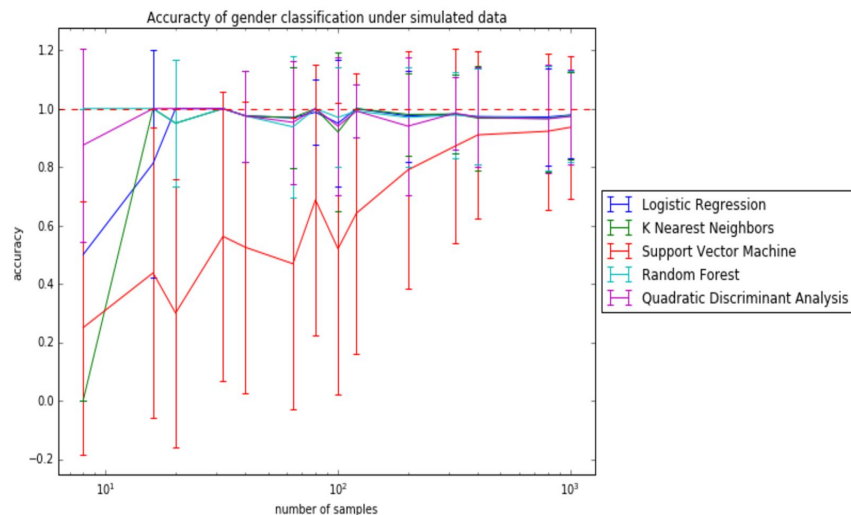
$$\sigma(t) = \frac{1}{1 + e^{-t}}$$

## Becomes Decision Rule

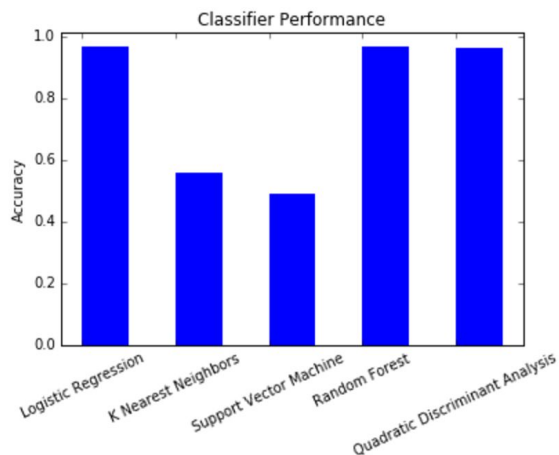
$$F(x) = \frac{1}{1 + e^{-(\beta_0 + \sum_{i=1}^n \beta_i x_i)}}$$

# Results

## Simulated



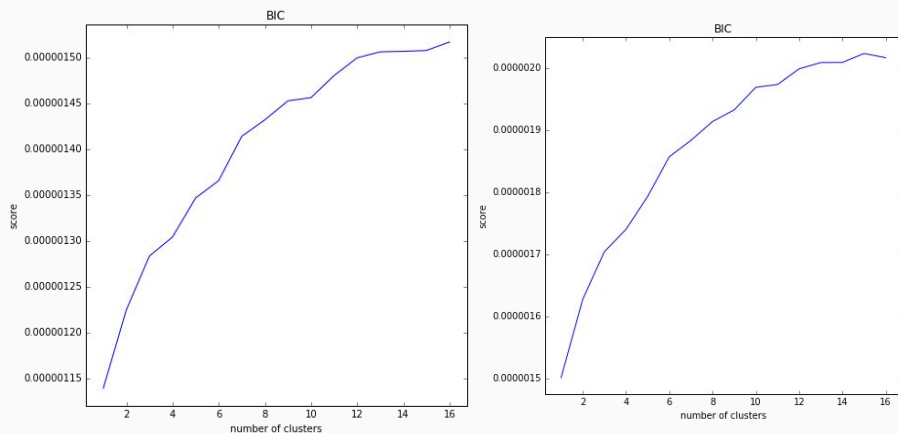
## Real Data



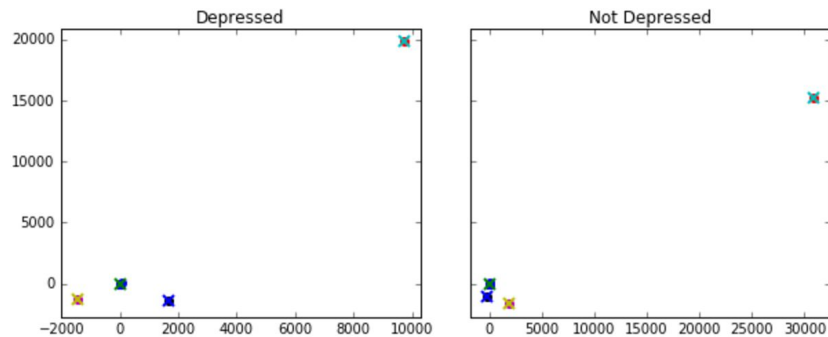
Classifier	Accuracy
LGR	0.9653 (+/- 0.183)
KNN	0.5594 (+/- 0.496)
SVM	0.4882 (+/- 0.500)
RF	0.9649 (+/- 0.184)
QDA	0.9624 (+/- 0.190)



## Sample Clusters



## Feature Clusters



## Model Checking

We assumed that our samples were i.i.d.

To verify independence we converted the reduced feature set of depressed individuals into a square matrix and then used the matrix rank to determine if the patients are linearly independent. Less than 5% of our samples were not independent.

To verify identicality we used GMM to cluster the samples and K-Means to cluster the features. Neither the samples nor the features are identically distributed.

# Resolution

- We used a systematic principled approach and we have sufficient samples, therefore we trust our results
- Look into the feature engineering steps
- Look into other relations in our data that we haven't discovered yet

# Thanks!

Yating Jing  
Srini Suresh