Machine Learning Algorithm Detecting Schizophrenia via fMRI and sMRI Data

By Kathleen Williams
CMSE 410

Current Diagnosis Method for Schizophrenia

Process of Elimination

- Estimate possibilities for the presence of other illnesses through serum tests [2,3]

Misdiagnoses

- Error during serum tests
- Similar symptomologies
- Comorbidities

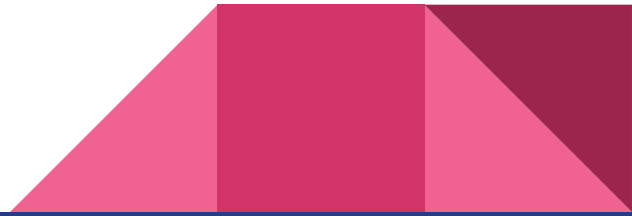
Project Outline

Data from fMRI and sMRI scans

Two groups: schizophrenic and non-schizophrenic

Question: Can fMRI and sMRI data be used to detect schizophrenia?

Solution: Create a machine learning (ML) algorithm to distinguish data from the two groups

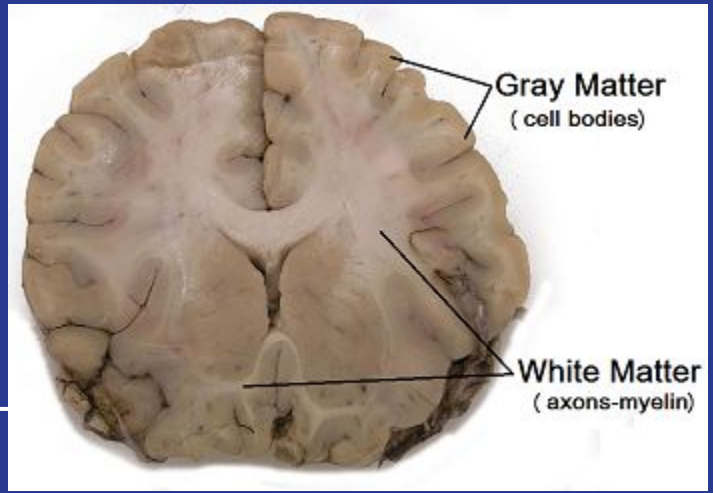


fMRI

- dynamic physiological information [4]
- blood oxygen level dependent technique, perfusion (whether by endogenous or exogenous contrast), blood flow, and cerebrospinal fluid (CSF) pulsation measurements [4]

sMRI

- static anatomical information [4]
  - gray matter



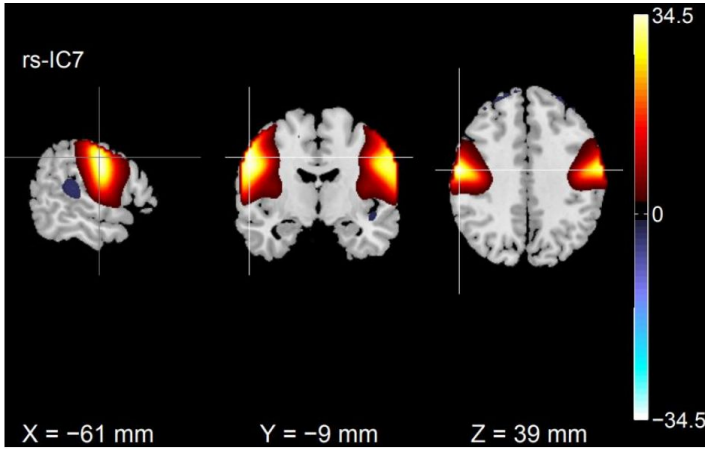
[6]

Dataset

"MLSP 2014 Schizophrenia Classification Challenge" [5]

Available on Kaggle

fMRI

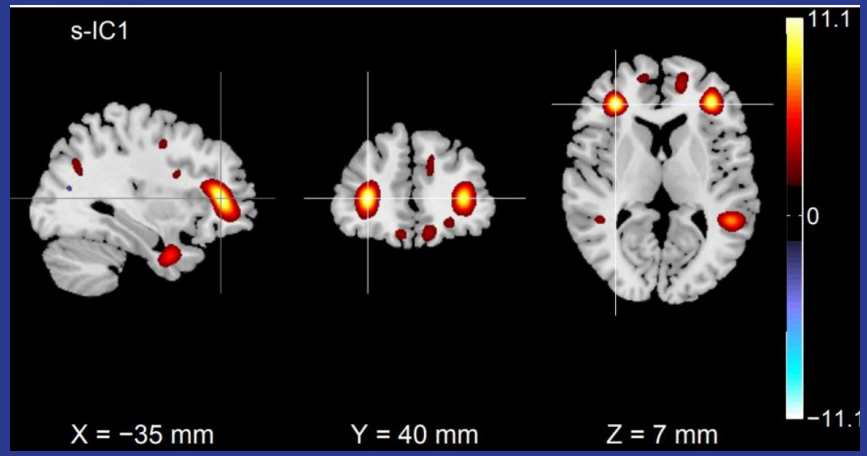


[5]

	A	B	C
1	Id	Class	
2	120873	1	
3	135376	0	
4	139149	0	
5	146791	0	

	A	B	C	D
1	Id	FNC1	FNC2	FNC3
2	120873	0.34312	0.045761	-0.13112
3	135376	0.2879	0.10257	-0.32343
4	139149	0.24585	0.21662	-0.12468
5	146791	0.4209	0.33138	0.24453

sMRI



[5]

	A	B	C	D	E	F
1	Id	SBM_map1	SBM_map2	SBM_map3	SBM_map	SBM_map
2	120873	0.72506502	-0.6392542	0.35306925	-0.98171	-1.41997
3	135376	-1.32885518	0.50296975	0.0132317	1.128496	-0.07074
4	139149	0.732267758	-1.24155378	0.65494157	-0.28922	0.158316
5	146791	-0.34391707	-1.05251938	-1.15052142	0.765989	0.923129

Dataset

Training Set:

- FNC features represent the correlation values which “describe the connection level between pairs of brain maps over time”
- SBM features are the standardized weights which “describe the expression level of ICA brain maps derived from graymatter concentration”

Testing Set:

- test subject labels of having schizophrenia or not have been removed in both the FNC and SBM csv files for the testing dataset

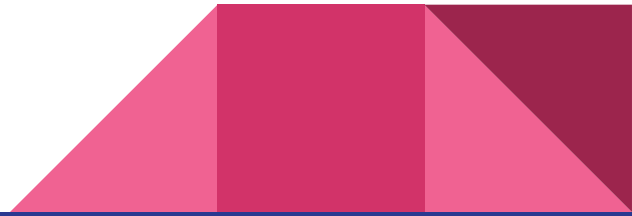
Methods and Software

Past Kaggle Challenge Winners

- Gaussian process (GP) classification,
- feature trimming/pruning
- Distance Weighted Discriminant

My Algorithm:

- Logistic regression



Start of Code

In [51]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LogisticRegression
```

In [52]:
```
# read training data
FNC_train = pd.read_csv("train_FNC.csv")
labels_train = pd.read_csv("train_labels.csv")
SBM_train = pd.read_csv("train_SBM.csv")
```

ML for fMRI and sMRI

In [58]:
```
# train model
log_reg = LogisticRegression(solver = 'lbfgs')
log_reg.fit(FNC_train, labels_train["Class"])
```

Out[58]:
```
LogisticRegression()
```

In [59]:
```
# predictions
class_prediction = log_reg.predict(FNC_test)
print(class_prediction)
```

```
[0 0 0 ... 0 0 0]
```

In [60]:
```
# score model
# every time that the score is greater than 0.5, put in class 1, and scores less than 0.5 pu in class 0
# count every time that prediction is equal to output and divide by total
log_reg.score(FNC_train, labels_train["Class"])
```

Out[60]:
```
0.5348837209302325
```

In [65]:
```
# logistic regression
log_reg2 = LogisticRegression()
log_reg2.fit(SBM_train, labels_train["Class"])
```

Out[65]:
```
LogisticRegression()
```

In [66]:
```
class_prediction = log_reg2.predict(SBM_test)
print(class_prediction)
```

```
[0 0 0 ... 0 0 0]
```

In [69]:
```
# score model
# every time that the score is greater than 0.5, put in class 1, and scores less than 0.5 put in class 0
# count every time that prediction is equal to output and divide by total
log_reg2.score(SBM_train, labels_train["Class"])
```

Out[69]:
```
0.5348837209302325
```

Conclusion: sMRI and fMRI can’t be used to detect Schizophrenia

Sources

1. AL-Ma'amari, Mohammed. “How to Choose the Best Machine Learning Model ?” *Medium*, Medium, 16 June 2020, medium.com/@mamarih1/how-to-choose-the-best-machine-learning-model-e1dbb46bdd4d.
2. “MLSP 2014 Schizophrenia Classification Challenge.” *Kaggle*, www.kaggle.com/c/mlsp-2014-mri/data.
3. “MLSP 2014 Schizophrenia Classification Challenge.” *Kaggle*, www.kaggle.com/c/mlsp-2014-mri/overview/description.
4. “Schizophrenia Diagnosis & Tests: How Doctors Know If Someone Has It.” *WebMD*, WebMD, 19 Jan. 2021, www.webmd.com/schizophrenia/schizophrenia-tests.
5. “Schizophrenia: Symptoms, Causes, Treatments.” *Cleveland Clinic*, my.clevelandclinic.org/health/diseases/4568-schizophrenia#diagnosis-and-tests.
6. “Tutorials.” *Brain Imaging Analysis Kit*, brainiak.org/tutorials/.
7. “Logistic Regression in Python Step by Step in 10 Minutes.” *YouTube*, YouTube, 26 Sept. 2019, www.youtube.com/watch?v=HYcXgN9HaTM.
8. “Logistic Regression Python Sklearn.” *YouTube*, YouTube, 31 July 2019, www.youtube.com/watch?app=desktop&v=VK6v9Ure8Lk&feature=youtu.be.