Fingerprinting using Machine Learning on fMRI Data

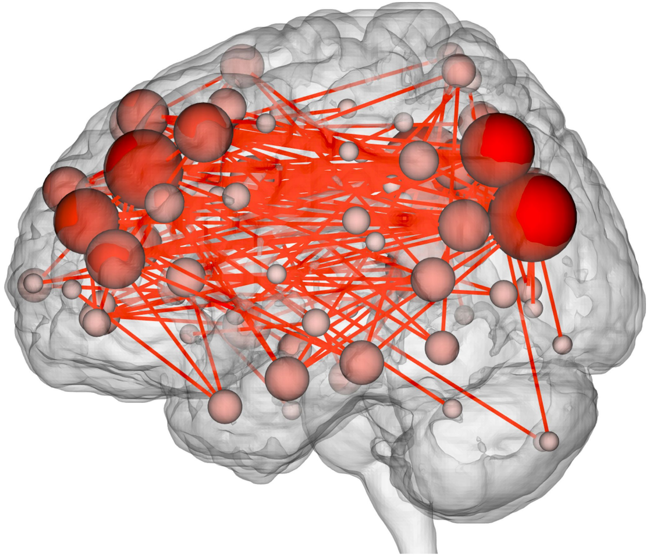
Shivansh Chandra Tripathi

What is fingerprint?

Examples:

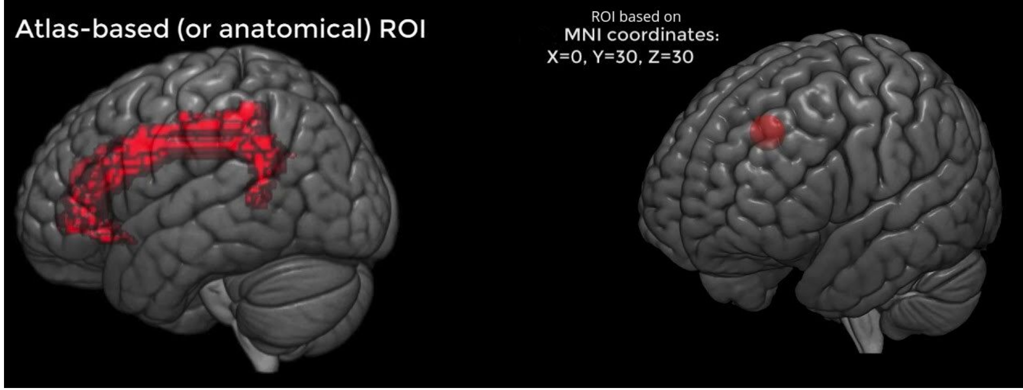
Hand based Fingerprint

DNA Fingeprinting -
Unique DNA patterns



Brain fingerprint - Unique brain connectivity graph

What is fMRI Fingerprint?



- **Region of Interests(ROI):** a voxel or a cluster of voxels in the same vicinity defining some functionality of brain.
- Functional connectivity is defined as the temporal dependency of spatially separated brain regions or as a correlation between two ROI's in mathematical terms.
- A set of carefully chosen ROI's and their Functional Connectivity defines a matrix or a graph.
- Hypothesis: This graph represents a fingerprint and can uniquely identify individuals from a group.

Literature Review

[1] employed an SVM based approach for fingerprinting.

[1] made further remarks that Fingerprints are quite stable from childhood to adulthood and that fingerprints vary with decreasing genetic similarity.

[2] employed a Deep Learning based pipeline for Fingerprinting.

[3] Proposed using an SVM based approach for task-state fMRI.

[4] utilized an autoencoder based approach to enhance the inter subject variability by attempting to remove shared neural activities.

SVM's efficiency is well established, Deep Learning methods have only started to occur recently, hence I used SVM as our classifier..

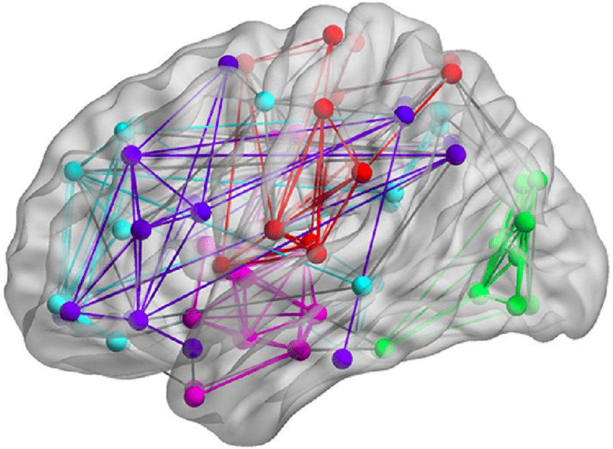
Machine Learning Pipeline for identification using Fingerprint

Preprocessing

All bold images preprocessed using FSL which includes

- removing first 10 volumes
- Band-pass filtering (0.0009 Hz, 0.08 Hz)
- Motion Correction
- Spatial Smoothing (4 mm fwhm)
- Intensity normalization
- Registration to 2mm MNI atlas space (using 12DOF and Linear search).

Feature Extraction

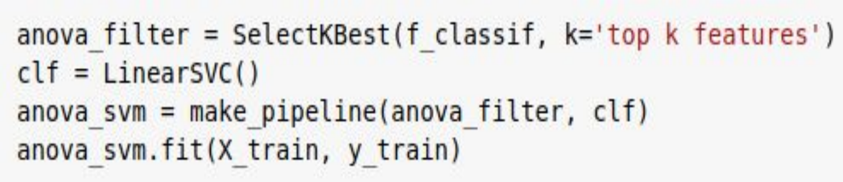


- [1] defines 264 most putative functionally connected Regions of Interests(ROI) as 264 MNI co-ordinates.
- For each Bold image, a sphere of 10 mm radius is drawn
- Within each sphere, all time series are averaged giving a total of 264 time series.
- Pearson Correlation is calculated for each of these time series pair which can be put in a 264x264 matrix.
- This matrix is called a Functional Connectivity(FC) matrix giving a total of 264x264 features for each bold image.
- Can represent this matrix as a graph called the Brain connectivity graph.

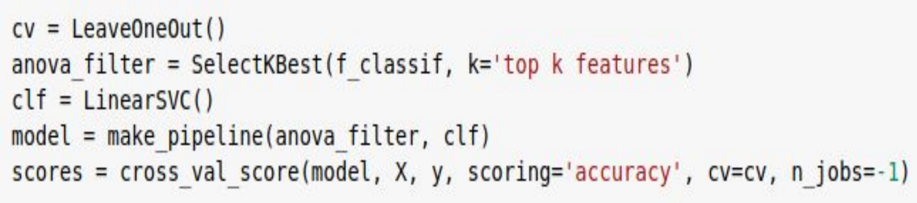
[1] Functional network organization of the human brain- Power et al., Neuron, 2011.

Feature Selection and Classification

- FC matrices are symmetric and diagonal entries are always one.
- Variable and useful features are [264x264-264(diagonal)]/2 = 34716 features
- Then top k features from 34716 are selected using ANOVA feature selection.
- A nxk data matrix is formed where n is the number of samples and divided into train-test splits or trained with LeaveOneOut Cross Validation.
- A Linear SVM is trained on the data.



```
anova_filter = SelectKBest(f_classif, k='top k features')
clf = LinearSVC()
anova_svm = make_pipeline(anova_filter, clf)
anova_svm.fit(X_train, y_train)
```



```
cv = LeaveOneOut()
anova_filter = SelectKBest(f_classif, k='top k features')
clf = LinearSVC()
model = make_pipeline(anova_filter, clf)
scores = cross_val_score(model, X, y, scoring='accuracy', cv=cv, n_jobs=-1)
```

Training and Testing Data Set Used

1) Midnight Scan Club Data

Description- 10 subjects, 10 sessions, 1 rfMRI image + some tfMRI image per session.

2) HCP Data

1200 subjects, each subject has 4 rfMRI image(rfMRI_REST1_LR, rfMRI_REST1_RL, rfMRI_REST2_LR, rfMRI_REST2_RL), 14 tfMRI images (tfMRI_EMOTION_LR, tfMRI_EMOTION_RL, tfMRI_MOTOR_LR, tfMRI_MOTOR_RL, tfMRI_GAMBLING_LR, tfMRI_RELATIONAL_LR, tfMRI_RELATIONAL_RL, tfMRI_LANGUAGE_LR, tfMRI_LANGUAGE_RL, tfMRI_SOCIAL_LR, tfMRI_SOCIAL_RL, tfMRI_WM_LR, tfMRI_WM_RL)

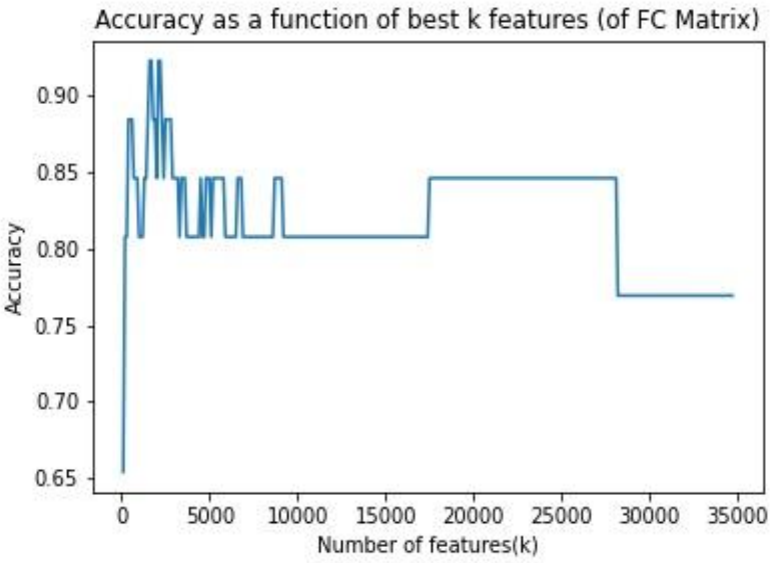
LR and RL are two different directions of phase encoding of magnetic field gradient during fMRI scan.

Results

Midnight Scan Club Data (ds000224/version 00001)

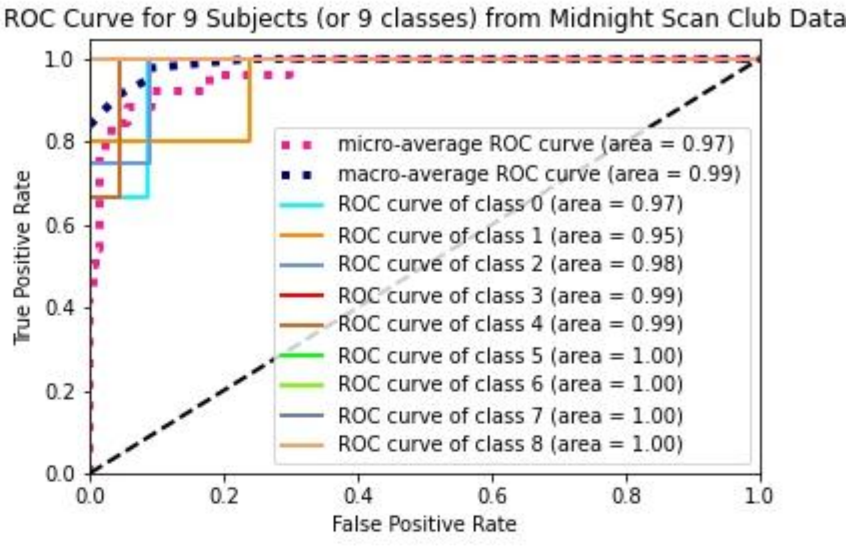
- 10 subjects, each underwent 10 sessions, each session recorded 1 resting state fMRI giving 10x10 = 100 samples.
- Erroneous samples reduced the sample size to 84 and 9 subjects.
- LeaveOneOut Cross Validation gives a mean accuracy of 0.905 with a mean standard deviation of 0.294 in accuracy.

Accuracy vs. best k-features plot



Using ANOVA feature selection and selecting k best features, and ranging k from 1 to 34716 features, the following plot is obtained, where max mean leave one out cross validation accuracy was found at k = 1600.

ROC Curve



Area Under Curve (AUC) score:

AUC = 1 implies an excellent model.

AUC = 0.5 implies that the model has no power than a random guess.

AUC = 0 implies that the model is poor, and reciprocating results.

High AUC indicates a good sensitivity or True Positive rate and a good specificity or True Negative rate at various thresholds.

All ROC values have area under curve (AUC) in the range (0.9, 1) indicating the classifier is efficient.

HCP Data (100 Subjects)

Used 100 Subjects, 48 males, 52 females- used task images tfMRI_EMOTION_LR, tfMRI_EMOTION_RL, tfMRI_MOTOR_LR, tfMRI_MOTOR_RL, tfMRI_LANGUAGE_LR, tfMRI_LANGUAGE_RL, tfMRI_WM_LR, tfMRI_WM_RL .

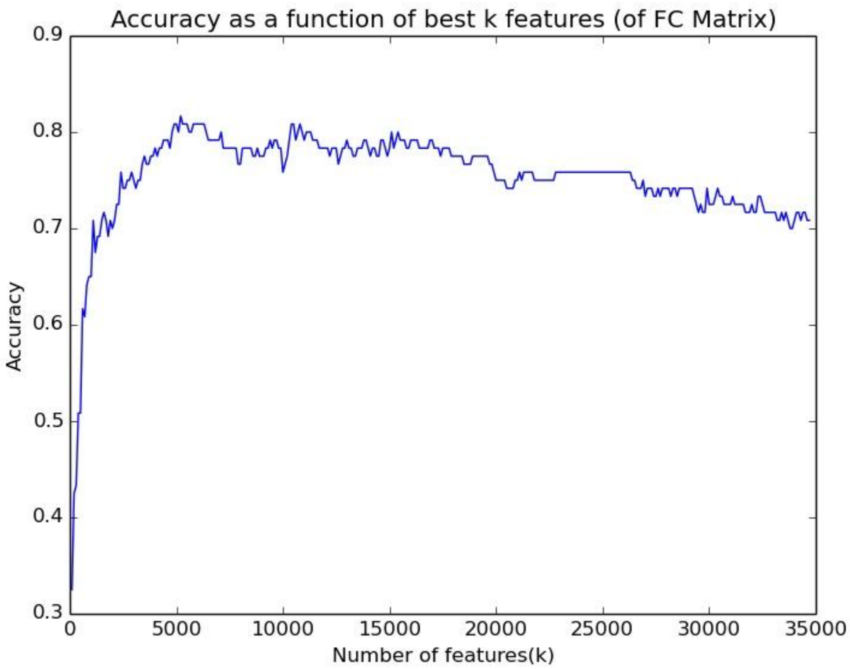
Each subject represents a class.

Giving n = 100x8 = 800 samples, i.e., 8 samples per class.

Accuracy at 30% test split is 0.68.

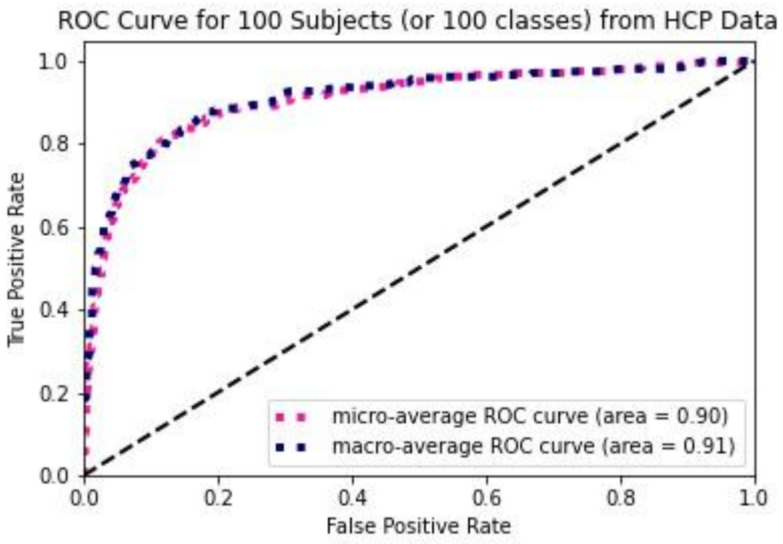
LeaveOneOut Cross Validation gives a mean accuracy of 0.83 with a mean standard deviation of 0.376 in accuracy.

Accuracy vs. best k-features plot



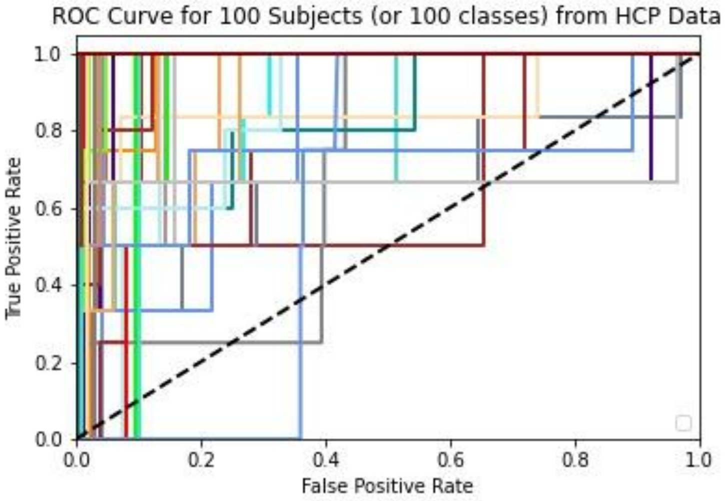
Using ANOVA feature selection and selecting k best features, and ranging k from 1 to 34716 features, the following plot is obtained, where max mean leave one out cross validation accuracy was found at k = 5200.

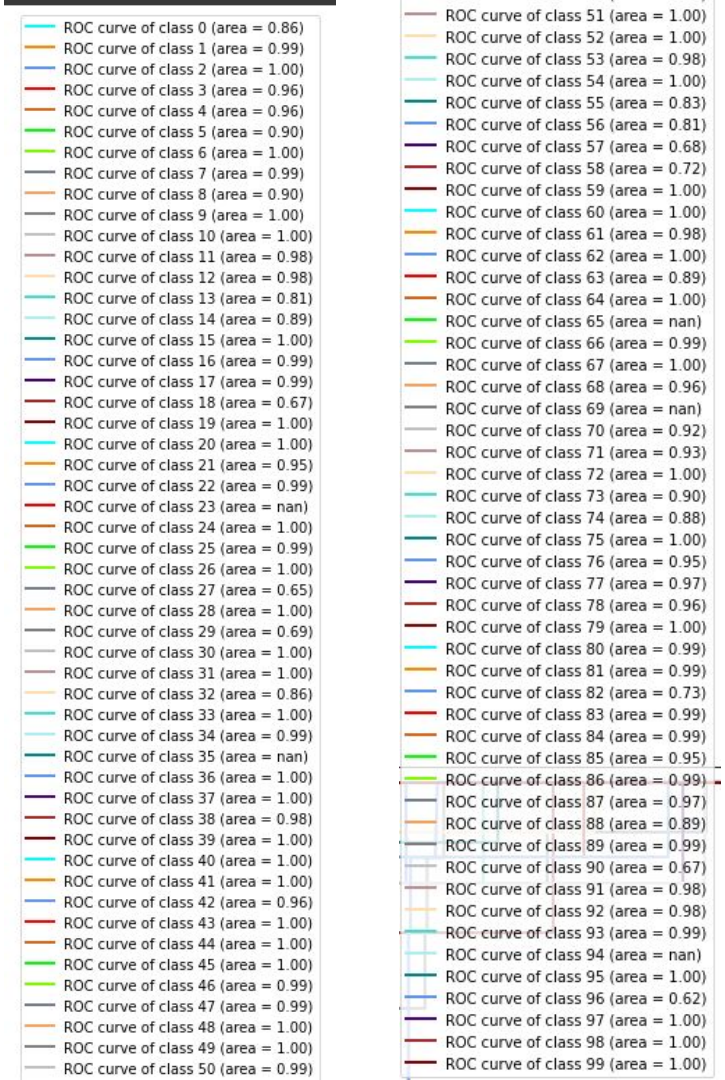
ROC Curve



AUC values are close to 1 for micro-average and macro-average plus for individual classes (except some are in range 0.6 - 0.9) indicating a powerful classifier.

ROC Curve


ROC Curve for 100 Subjects (or 100 classes) from HCP Data
True Positive Rate
False Positive Rate
0.0
0.2
0.4
0.6
0.8
1.0
0.0
0.2
0.4
0.6
0.8
1.0


ROC curve of class 0 (area = 0.86)
ROC curve of class 1 (area = 0.99)
ROC curve of class 2 (area = 1.00)
ROC curve of class 3 (area = 0.96)
ROC curve of class 4 (area = 0.96)
ROC curve of class 5 (area = 0.90)
ROC curve of class 6 (area = 1.00)
ROC curve of class 7 (area = 0.99)
ROC curve of class 8 (area = 0.90)
ROC curve of class 9 (area = 1.00)
ROC curve of class 10 (area = 1.00)
ROC curve of class 11 (area = 0.98)
ROC curve of class 12 (area = 0.98)
ROC curve of class 13 (area = 0.81)
ROC curve of class 14 (area = 0.89)
ROC curve of class 15 (area = 1.00)
ROC curve of class 16 (area = 0.99)
ROC curve of class 17 (area = 0.99)
ROC curve of class 18 (area = 0.67)
ROC curve of class 19 (area = 1.00)
ROC curve of class 20 (area = 1.00)
ROC curve of class 21 (area = 0.95)
ROC curve of class 22 (area = 0.99)
ROC curve of class 23 (area = nan)
ROC curve of class 24 (area = 1.00)
ROC curve of class 25 (area = 0.99)
ROC curve of class 26 (area = 1.00)
ROC curve of class 27 (area = 0.65)
ROC curve of class 28 (area = 1.00)
ROC curve of class 29 (area = 0.69)
ROC curve of class 30 (area = 1.00)
ROC curve of class 31 (area = 1.00)
ROC curve of class 32 (area = 0.86)
ROC curve of class 33 (area = 1.00)
ROC curve of class 34 (area = 0.99)
ROC curve of class 35 (area = nan)
ROC curve of class 36 (area = 1.00)
ROC curve of class 37 (area = 1.00)
ROC curve of class 38 (area = 0.98)
ROC curve of class 39 (area = 1.00)
ROC curve of class 40 (area = 1.00)
ROC curve of class 41 (area = 1.00)
ROC curve of class 42 (area = 0.96)
ROC curve of class 43 (area = 1.00)
ROC curve of class 44 (area = 1.00)
ROC curve of class 45 (area = 1.00)
ROC curve of class 46 (area = 0.99)
ROC curve of class 47 (area = 0.99)
ROC curve of class 48 (area = 1.00)
ROC curve of class 49 (area = 1.00)
ROC curve of class 50 (area = 0.99)
ROC curve of class 51 (area = 1.00)
ROC curve of class 52 (area = 1.00)
ROC curve of class 53 (area = 0.98)
ROC curve of class 54 (area = 1.00)
ROC curve of class 55 (area = 0.83)
ROC curve of class 56 (area = 0.81)
ROC curve of class 57 (area = 0.68)
ROC curve of class 58 (area = 0.72)
ROC curve of class 59 (area = 1.00)
ROC curve of class 60 (area = 1.00)
ROC curve of class 61 (area = 0.98)
ROC curve of class 62 (area = 1.00)
ROC curve of class 63 (area = 0.89)
ROC curve of class 64 (area = 1.00)
ROC curve of class 65 (area = nan)
ROC curve of class 66 (area = 0.99)
ROC curve of class 67 (area = 1.00)
ROC curve of class 68 (area = 0.96)
ROC curve of class 69 (area = nan)
ROC curve of class 70 (area = 0.92)
ROC curve of class 71 (area = 0.93)
ROC curve of class 72 (area = 1.00)
ROC curve of class 73 (area = 0.90)
ROC curve of class 74 (area = 0.88)
ROC curve of class 75 (area = 1.00)
ROC curve of class 76 (area = 0.95)
ROC curve of class 77 (area = 0.97)
ROC curve of class 78 (area = 0.96)
ROC curve of class 79 (area = 1.00)
ROC curve of class 80 (area = 0.99)
ROC curve of class 81 (area = 0.99)
ROC curve of class 82 (area = 0.73)
ROC curve of class 83 (area = 0.99)
ROC curve of class 84 (area = 0.99)
ROC curve of class 85 (area = 0.95)
ROC curve of class 86 (area = 0.99)
ROC curve of class 87 (area = 0.97)
ROC curve of class 88 (area = 0.89)
ROC curve of class 89 (area = 0.99)
ROC curve of class 90 (area = 0.67)
ROC curve of class 91 (area = 0.98)
ROC curve of class 92 (area = 0.98)
ROC curve of class 93 (area = 0.99)
ROC curve of class 94 (area = nan)
ROC curve of class 95 (area = 1.00)
ROC curve of class 96 (area = 0.62)
ROC curve of class 97 (area = 1.00)
ROC curve of class 98 (area = 1.00)
ROC curve of class 99 (area = 1.00)

Inference

Good classification results indicate that the 264 ROIs indicate regions of neural firings that are most unique.

Good AUC scores indicate the classifier is powerful enough to be used for fingerprint identification.

Since only tfMRI were used in HCP Data, indicates that the underlying fingerprint is retained to a large extent when scans are changed from rest to task.

More ROIs can be explored in future and better classifiers to eliminate common neural activities making the fingerprint more unique.

Tools used

FSL for preprocessing.

Utilized fslroi command line tool for creating ROI 10mm radius masks.

Nilearn python library for creating FC matrices.

Scikit-learn python library for ANOVA F-test, LeaveOneOut Cross Validation and SVM Classification.

References

[1] Functional Connectivity Fingerprints at Rest Are Similar across Youths and Adults and Vary with Genetic Similarity- Damion V. Demeter et al., iScience 23, 2020

[2]Deep Learning Based Pipeline for Fingerprinting Using Brain Functional MRI Connectivity Data- Nicolas F. Lori et al., Elsevier, 2018.

[3]Support Vector Machine for Analyzing Contributions of Brain Regions During Task-State fMRI- Mengyue Wang et al., frontiers in Neuroinformatics, 2019.

[4]Functional connectome fingerprinting: Identifying individuals and predicting cognitive function via deep learning- Biao Cai et al., 2020.