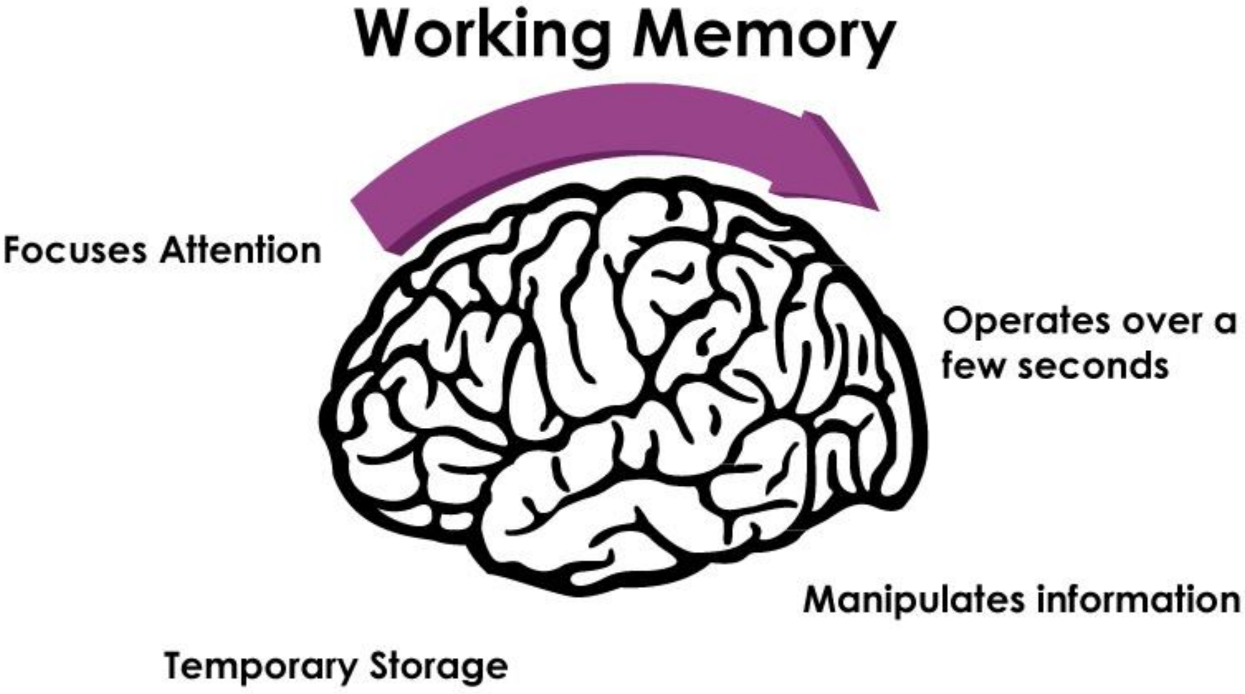
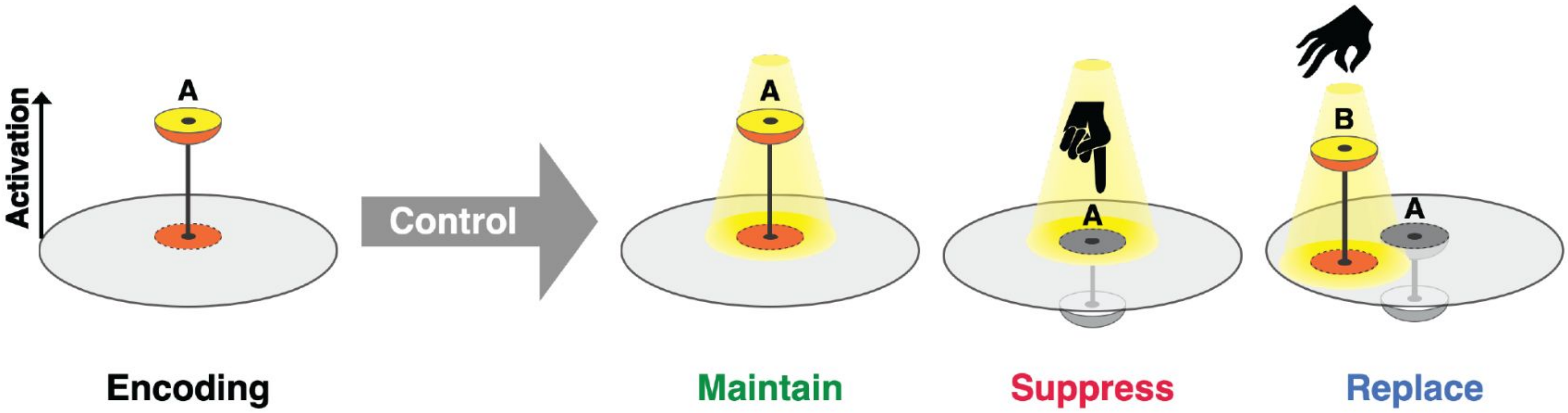
Classifying Complex Cognitive Operations from fMRI Data

Team 2
Caleb Jerinic-Brodeur, Owen Friend, Lingwei Ouyang
SDS 384 - Scientific Machine Learning

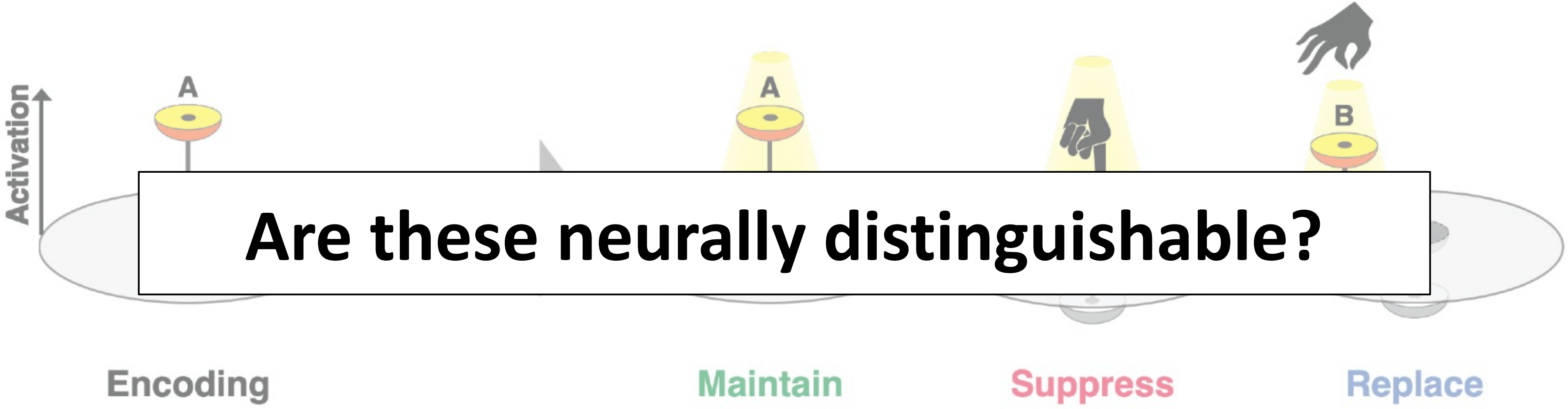
Introduction



Introduction

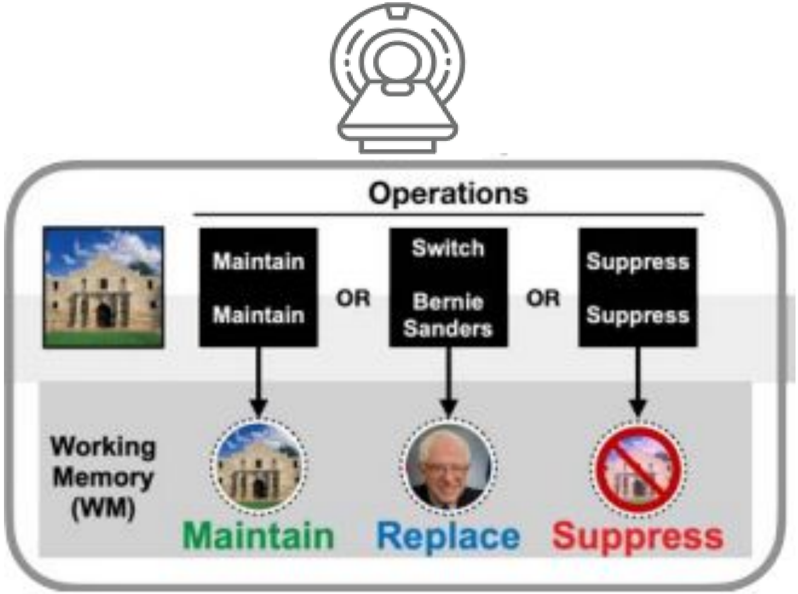


Introduction



Are these neurally distinguishable?

Data



- fMRI data (unpublished)
- 5 subjects
  - (1098, 151424)/subject

Models

- Logistic Regression (L2 regularized)
- Random Forest (no max-depth)
- Random Forest (max-depth = 10)
- XGBoost (max-depth = 10)

Exploratory Analyses & Hypothesis Generation

- Random Forest Classification may perform better than Logistic Regression due to complexity, non-linearity of neuroimaging data
- Alternatively, may lead to major overfitting due to noise of the BOLD signal

Compare XGBoost and RF model

- XGBoost has been used to classify fMRI language networks from patients with and without epilepsy. (Torlay et al., 2017)
- Similarity: XGBoost is also a non-linear classifier similar to RF model.
- Difference: XGBoost is better at handling class imbalance by downweight less represented categories.

Modeling and Validation

- Train classifiers on BOLD data for complex cognitive operations
  - Can we classify subtle differences in behavior based on neural signal?
  - Does a RF, XGBoost, or Logistic Regression model perform better?

Classification Setup

```
ps = PredefinedSplit(subject_sample)
for train, test in ps.split():
    train_data = bold_data[train]
    test_data = bold_data[test]
    train_label = op_labels[train]
    test_label = op_labels[test]

    # feature selection
    Fselect_fpr = SelectFpr(f_classif, alpha=0.001).fit(train_data, train_label)
    bold_train_subject = Fselect_fpr.transform(train_data)
    bold_test_subject = Fselect_fpr.transform(test_data)

    # train with selected penalty
    log_reg = LogisticRegression(penalty="l2", solver="lbfgs", C=50, max_iter=1000)
    log_reg.fit(bold_train_subject, train_label)

    # now test on the held out subject
    score = log_reg.score(bold_test_subject, test_label)
    decision_score = log_reg.decision_function(bold_test_subject)
```

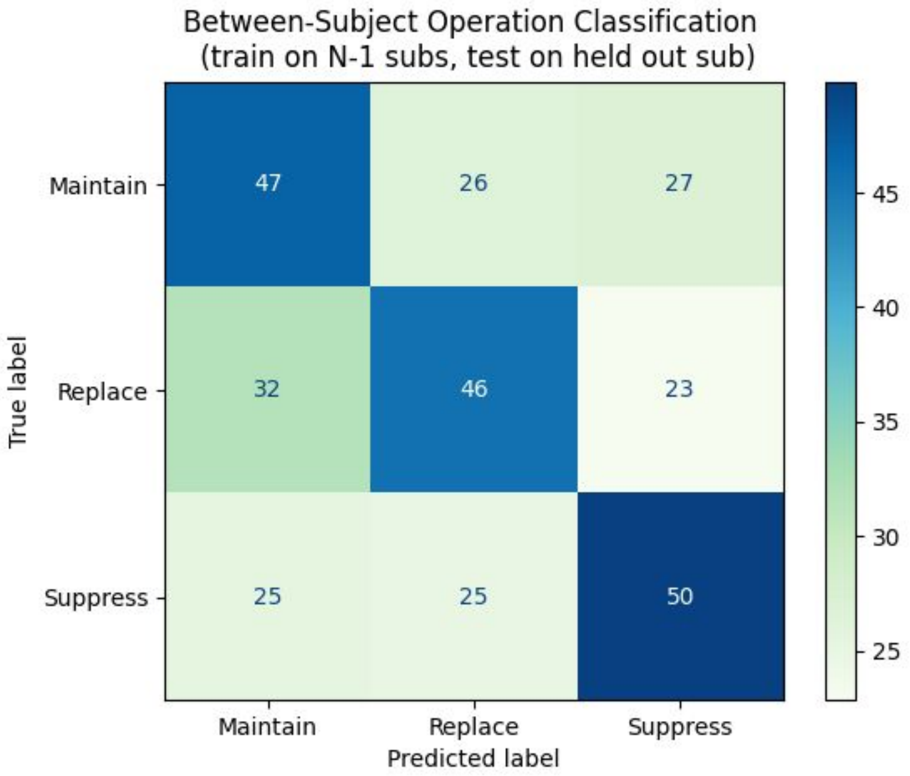
Cross-Validation Setup

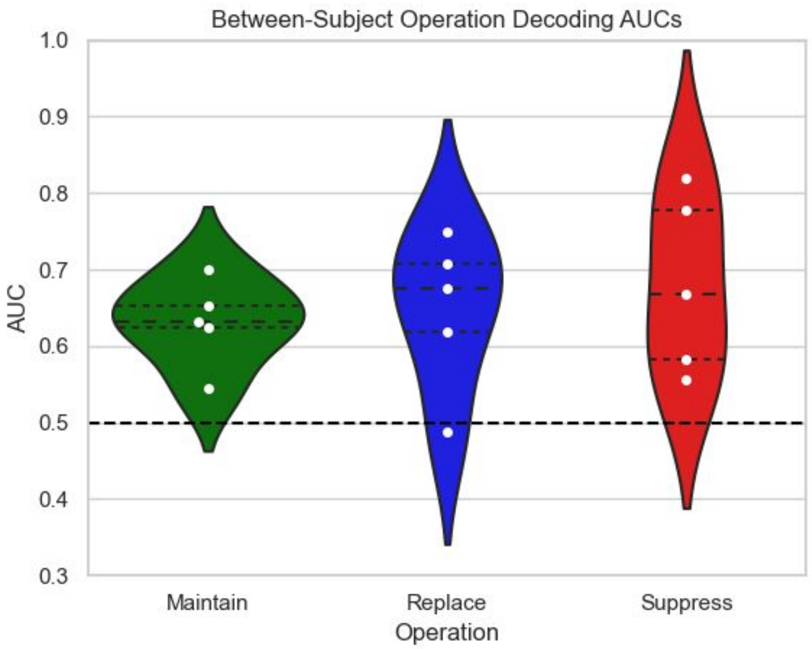
Feature Selection
[151424 ⟶ 19347]

Training

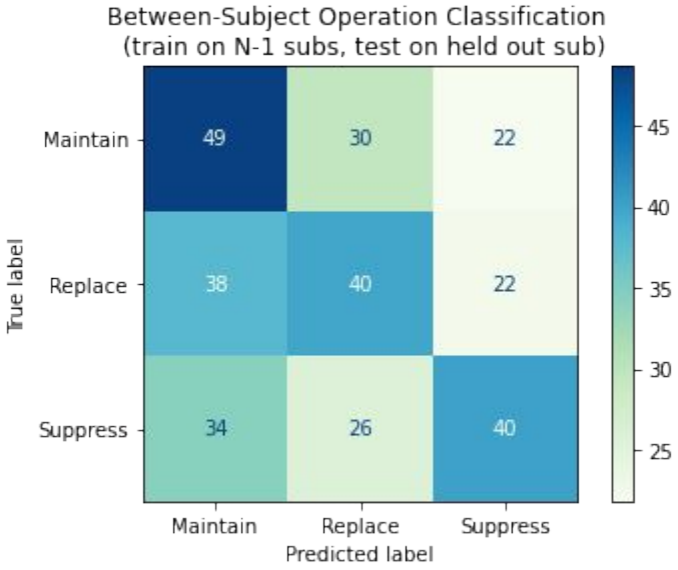
Testing

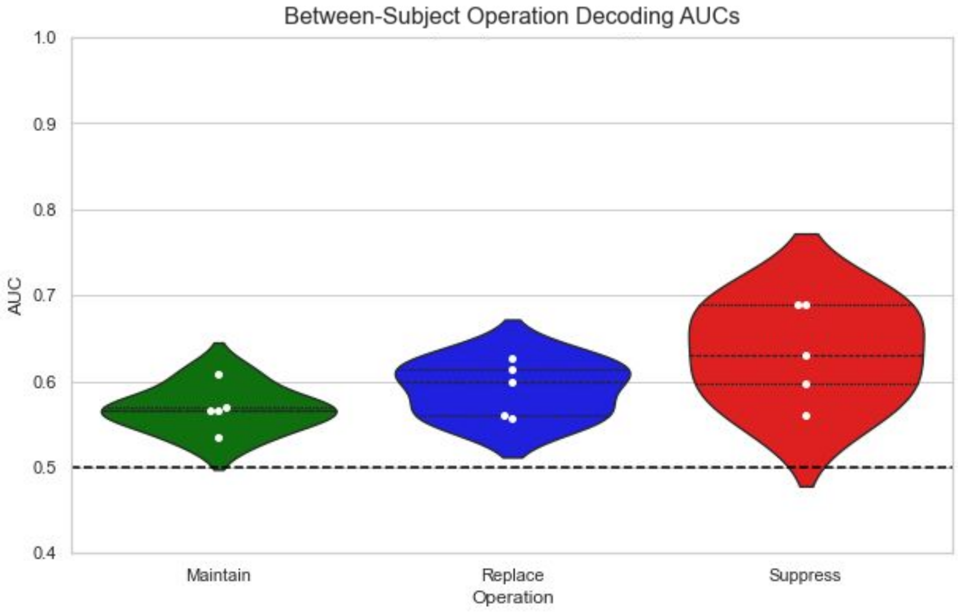
Results - Logistic Regression



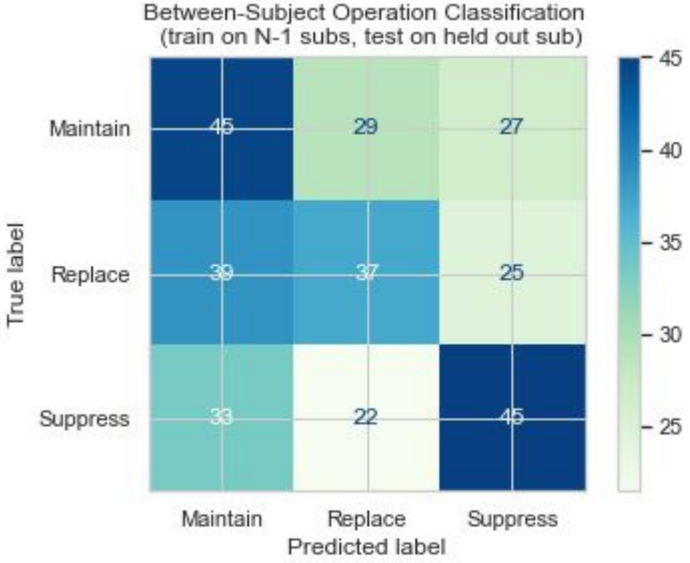


Results - Random Forest (no max depth)

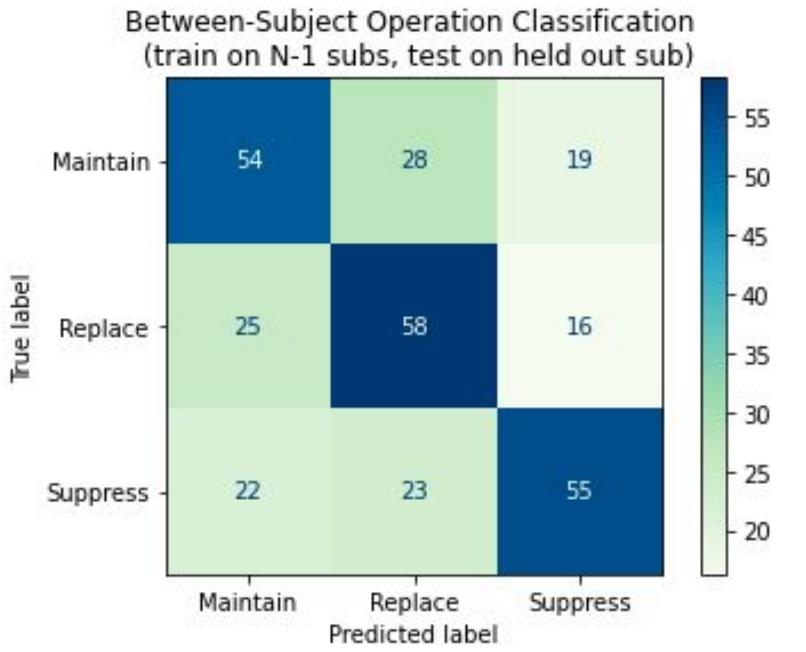


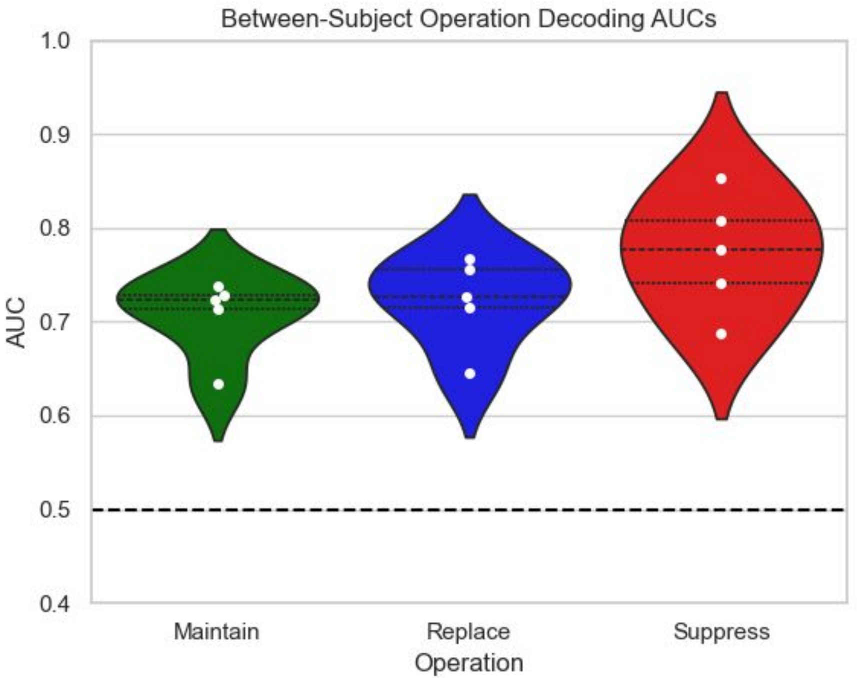


Results - Random Forest (max depth = 10)



Results - XGBoost model (max depth = 10)





auc scores: 0.74 **sd**: 0.04
Mean score: 0.55

Model Comparisons

- XGBoost outperforms all other models
  - No differences between Logistic Regression and either RF classifier

Contrast	A	B	T	dof	p-unc
Model	RF_maxdepthten	RF_nomaxdepth	-0.218238	4.0	0.837926
Model	RF_maxdepthten	XGBoost	-4.564634	4.0	0.010303
Model	RF_maxdepthten	log_reg	-2.189586	4.0	0.093735
Model	RF_nomaxdepth	XGBoost	-4.619621	4.0	0.009884
Model	RF_nomaxdepth	log_reg	-2.096883	4.0	0.104019
Model	XGBoost	log_reg	3.288295	4.0	0.030264

Discussion - Random Forest Models

- Greater max depth can be problematic with noisy neuroimaging data, leading to overfitting
- ‘Suppress’ operation may lead to noisier neural signal, resulting in worse fit with higher max depth (overfitting on the noise)

Discussion - Random Forest Models

- ‘Maintain’ appears to do the opposite
- Could reflect varying difficulty of two cognitive processes

Discussion - XGBoost model

- In Torlay et al., (2017), classification based on regional (instead of whole-brain) data reached 91%±5% accuracy. Is the low accuracy due to using whole brain? (only certain region is related)
- XGBoost is a significantly more complicated model than the other two models. Perhaps it performed better due to complexity.

Limitations

- Small sample size
  - Typical fMRI study is ~25 participants
- fMRI data is noisy
  - Increased noise when looking across whole brain?

Conclusion

- Complex behaviors can be decoded from neural signals
  - Replicates prior work
- Future work will attempt to classify these behaviors with other imaging modalities (e.g., EEG)

Thank you!

Questions?