Prediction of Autism Treatment Response from Baseline fMRI using Random Forests and Tree Bagging

Nicha C. Dvornek, Daniel Yang, Archana Venkataraman, Pamela Ventola, Lawrence H. Staib, Kevin A. Pelphrey, and James S. Duncan

ML-CDS 2016: Multimodal Learning for Clinical Decision Support
Athens, Greece
October 17, 2016

Yale University

Autism Spectrum Disorder (ASD)

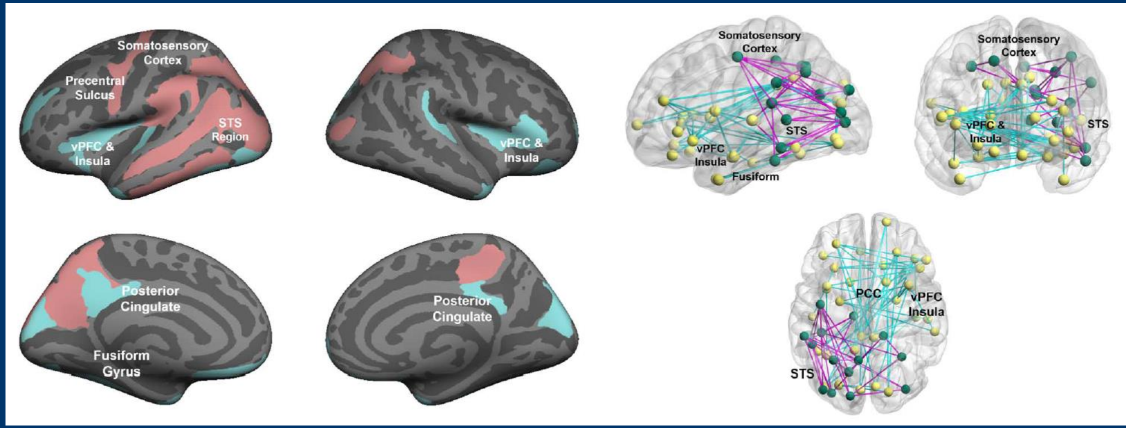
- Neurological developmental disorders characterized by impaired social interactions, difficulties in communication, and repetitive behaviors
- Promising treatment: Intensive behavioral therapies
  - e.g., Pivotal Response Therapy
  - Large commitment from patient and families
  - Early intervention is important
- However, ASD is complex!
  - No "one size fits all" treatment
  - Currently, choose therapy by trial and error

→ Need for *precision medicine*

www.autismspeaks.org

Goal: Predict Autism Treatment Outcome from Baseline fMRI

- fMRI has aided understanding of ASD pathophysiology

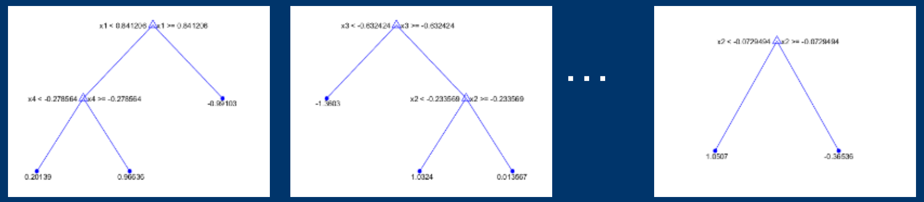


Venkataraman et al., TMI 2016

- fMRI for prediction
  - Changes in autistic traits [Plitt et al., PNAS 2015]
  - Treatment outcomes in other brain disorders [Ball et al., Neuropsych 2014]
  - →We propose first use of fMRI for predicting ASD treatment response

Goal: Predict Autism Treatment Outcome from Baseline fMRI

- Challenge: “large *p*, small *n*”
  - Large number of possible fMRI-derived inputs
  - Small number of subjects in autism studies
- Good candidate for Random Forests



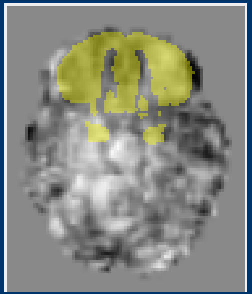
Random Forest

$\longrightarrow Y_{\text{ens}}$

- However…
  - Very noisy inputs degrade prediction accuracy
  - Small samples reduce strength of each tree

Learning Pipeline Overview

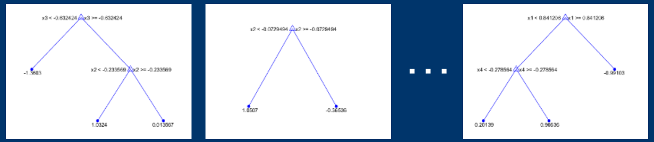
Inputs: fMRI Biomarkers



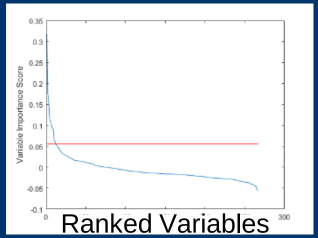
Outputs: Behavioral Treatment Response

Proposed Learning Pipeline

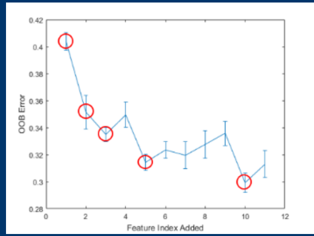
Candidate Variable Selection

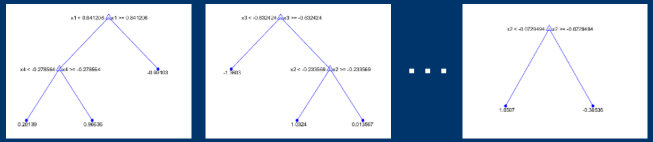


Random Forest



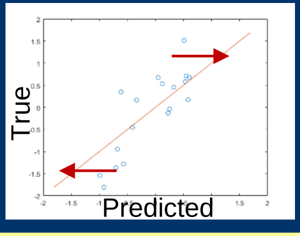
Stepwise Variable Refinement

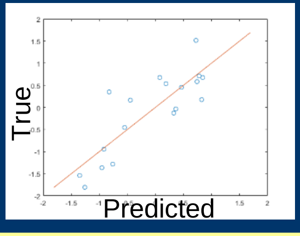




Bagged Ensemble

Bias Correction





Learned Predictive Model

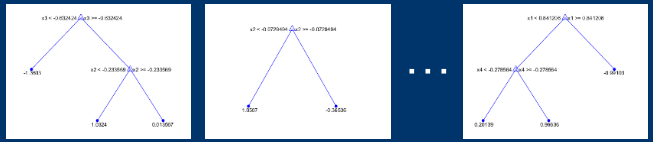
Learning Pipeline Overview

Inputs: fMRI Biomarkers

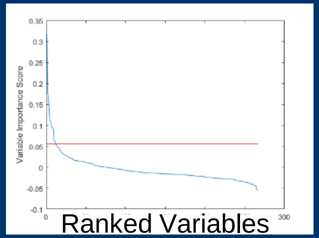
Outputs: Behavioral Treatment Response

Proposed Learning Pipeline

Candidate Variable Selection

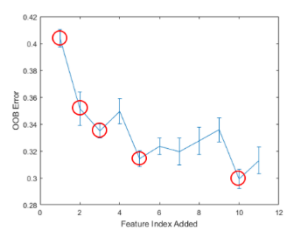


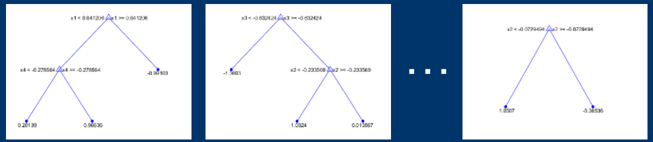
Random Forest



Ranked Variables

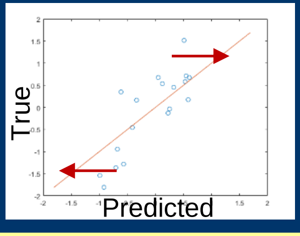
Stepwise Variable Refinement

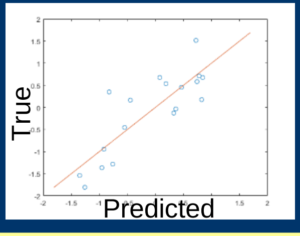




Bagged Ensemble

Bias Correction

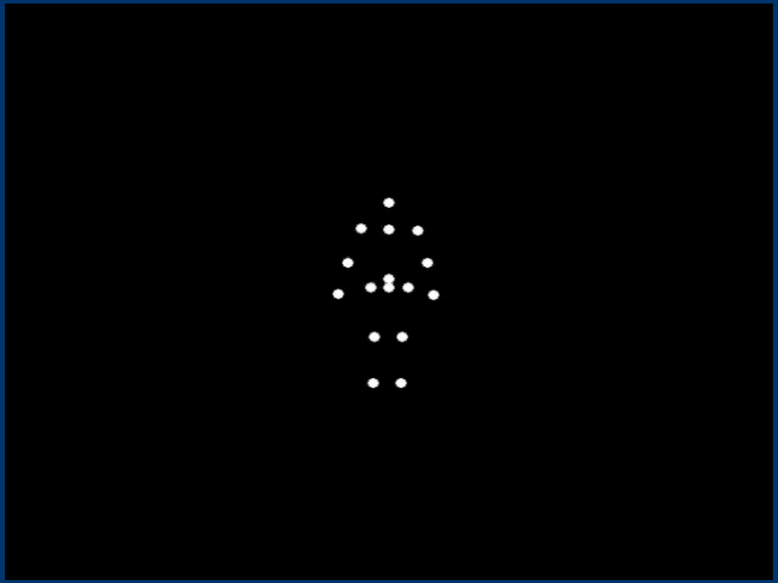


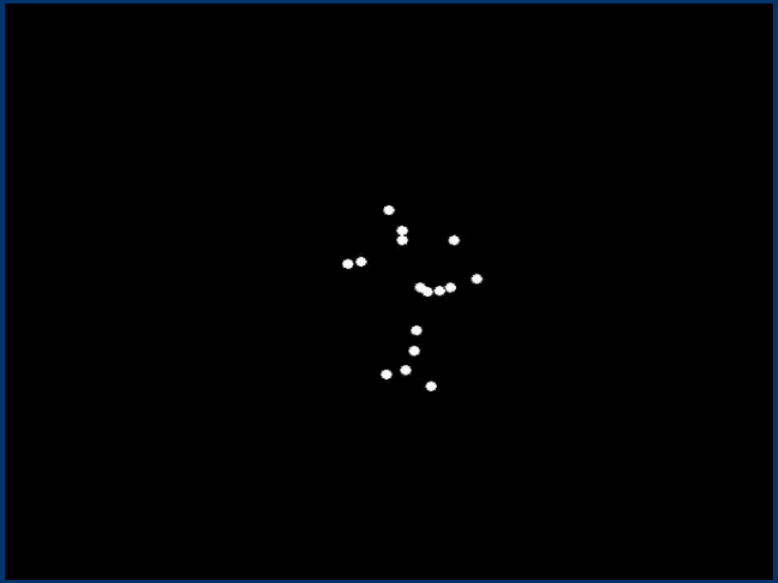


Learned Predictive Model

Inputs: Baseline fMRI-Derived Parameters

- Biopoint: Biological motion perception paradigm





- Focus on brain regions associated with social motivation: Orbitofrontal cortex, ventromedial prefrontal cortex, amygdala, and ventral striatum

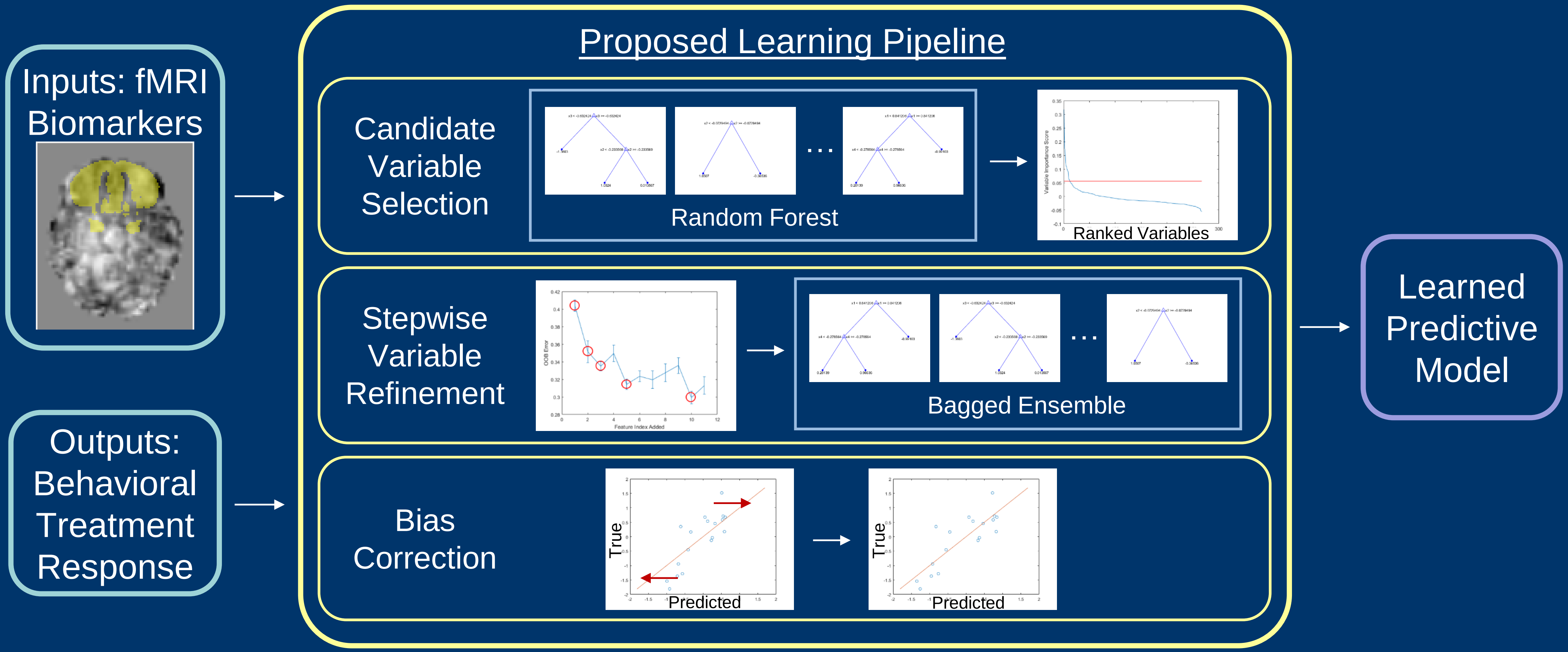
→ Inputs: t-statistics for biological motion > scrambled motion contrast

Outputs: Behavioral Treatment Outcome

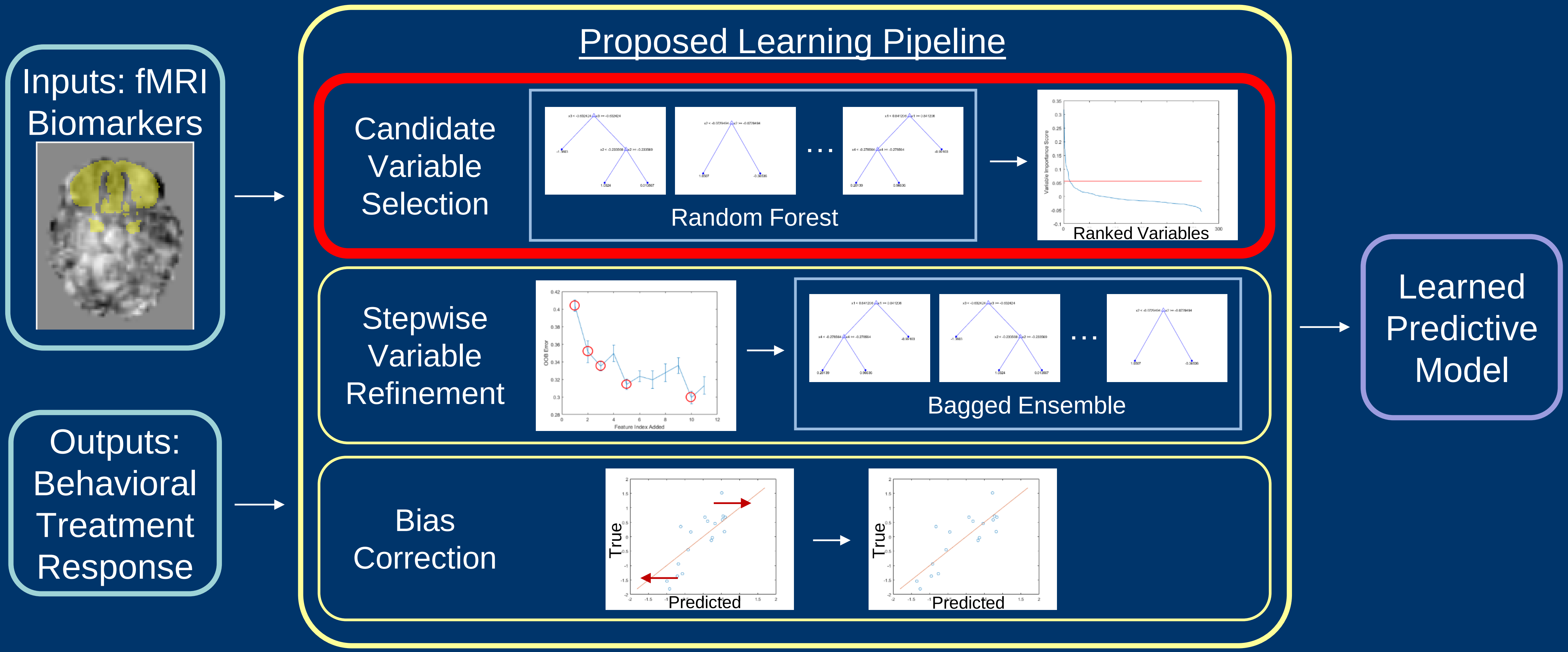
- Social Responsiveness Scale, Second Edition (SRS) Score
  - Measures severity of social impairment in ASD
  - Lower SRS score → Better function
- Measure SRS score at baseline and post-treatment

→ Outputs: Normalized change in SRS Score (ΔSRS)

Learning Pipeline Overview

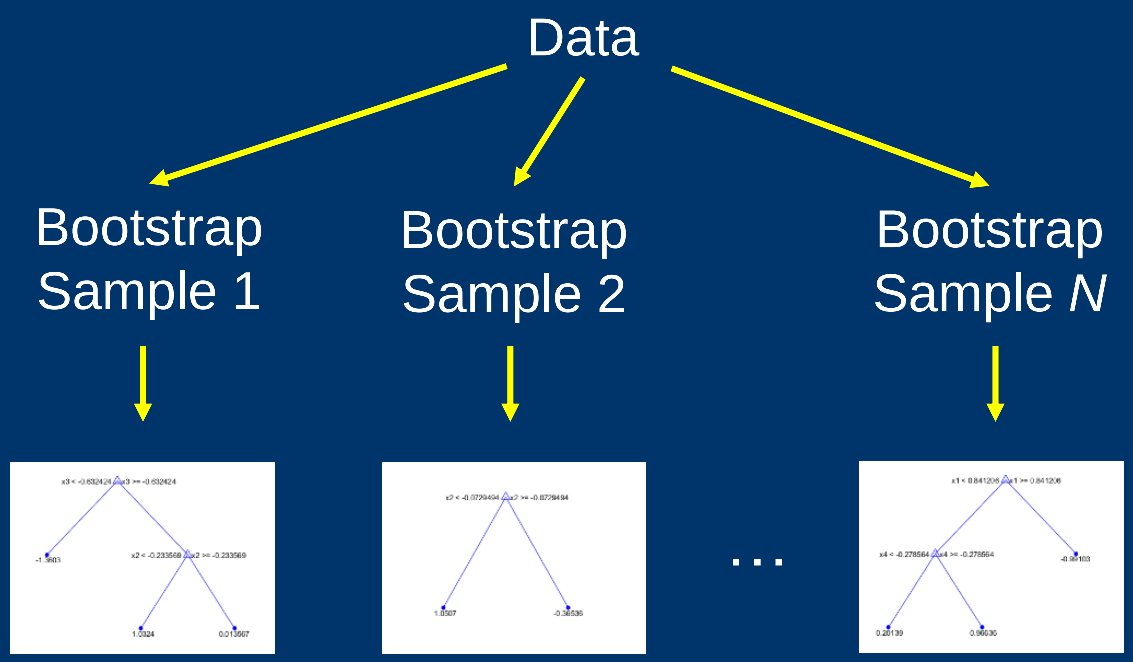


Learning Pipeline Overview



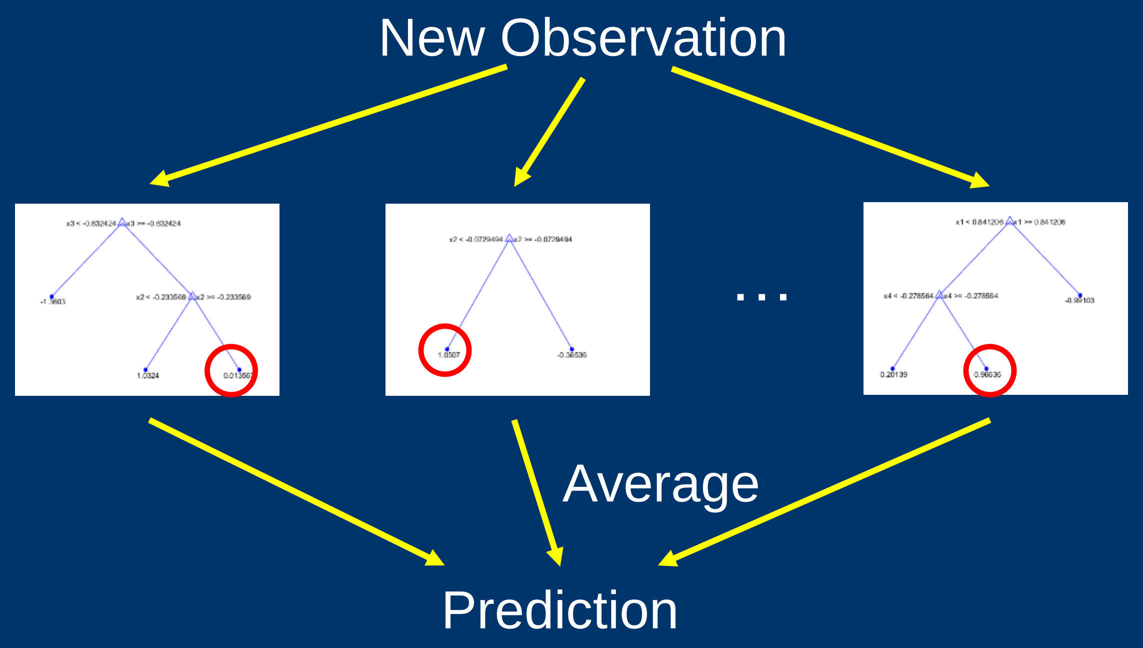
Random Forests for Regression Review

- Ensemble learning method that uses
  - Bagging



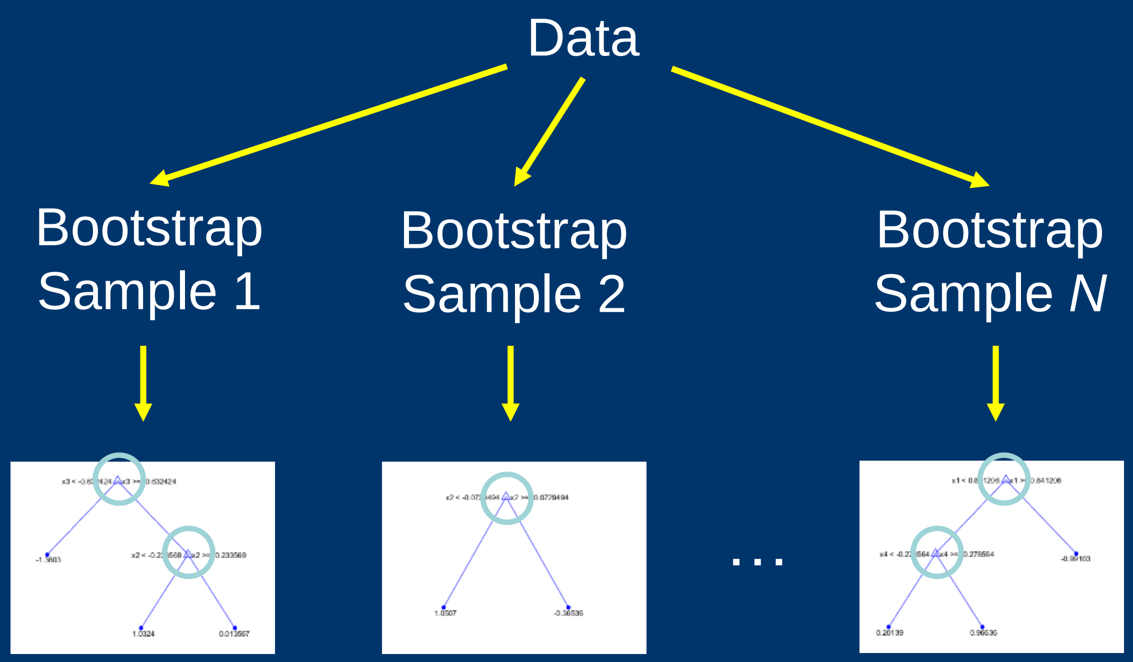
Random Forests for Regression Review

- Ensemble learning method that uses
  - Bagging



Random Forests for Regression Review

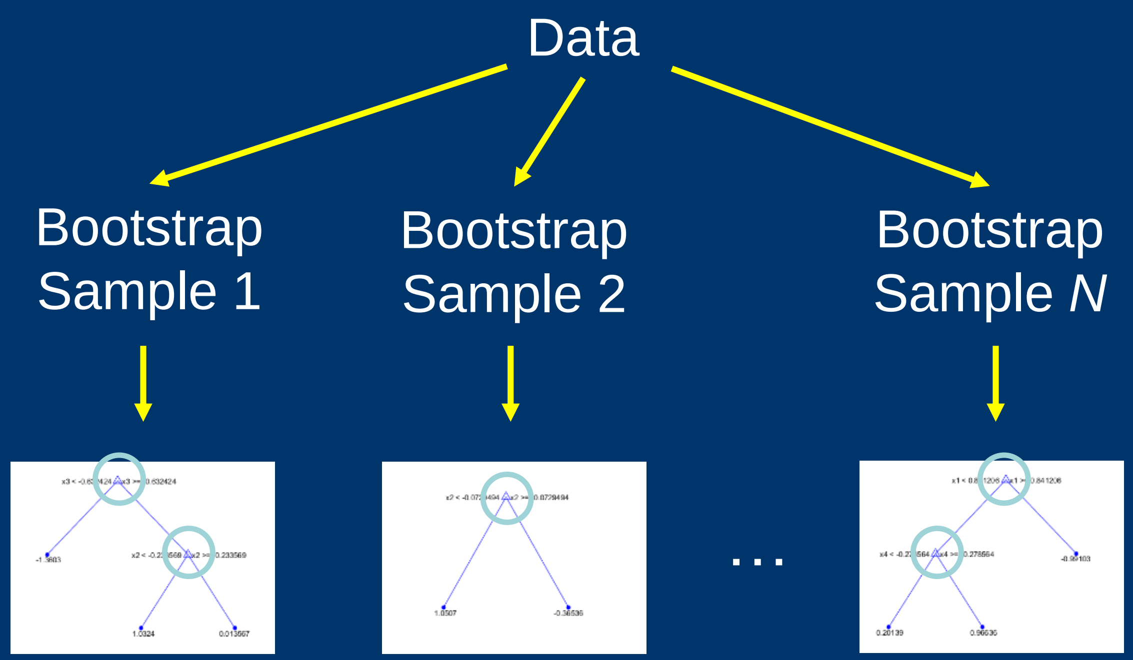
- Ensemble learning method that uses
  - Bagging
  - Random subset sampling of predictors



Best split variable chosen from *m* randomly selected predictors

Random Forests for Regression Advantages

- Reduced correlation of trees → Reduced variance of estimate
- Efficient exploration of high dimensional inputs



Best split variable chosen from *m* randomly selected predictors

Random Forests for Regression
Out-of-bag (OOB) Error

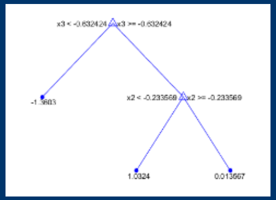
- Internal estimate of test error rate estimated by out-of-bag (OOB) error
- $E = \sum_{i}(Y_{i,oob} - Y_{i,true})^2$

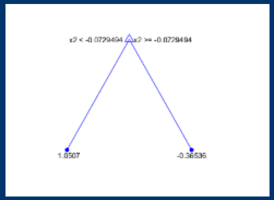
Observation *i*

OOB Sample 1

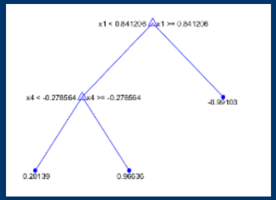
Not OOB Sample 2

OOB Sample N





…



Average

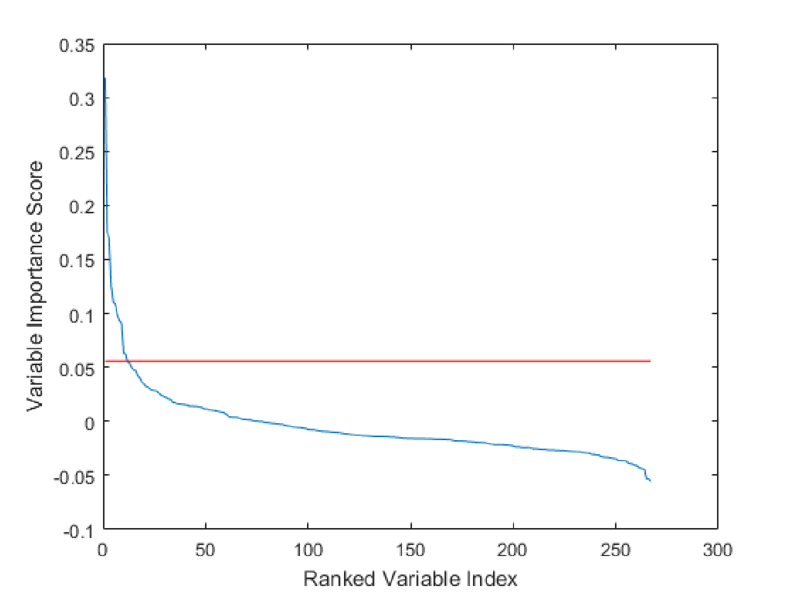
$Y_{i,oob}$

Random Forests for Regression Variable Importance

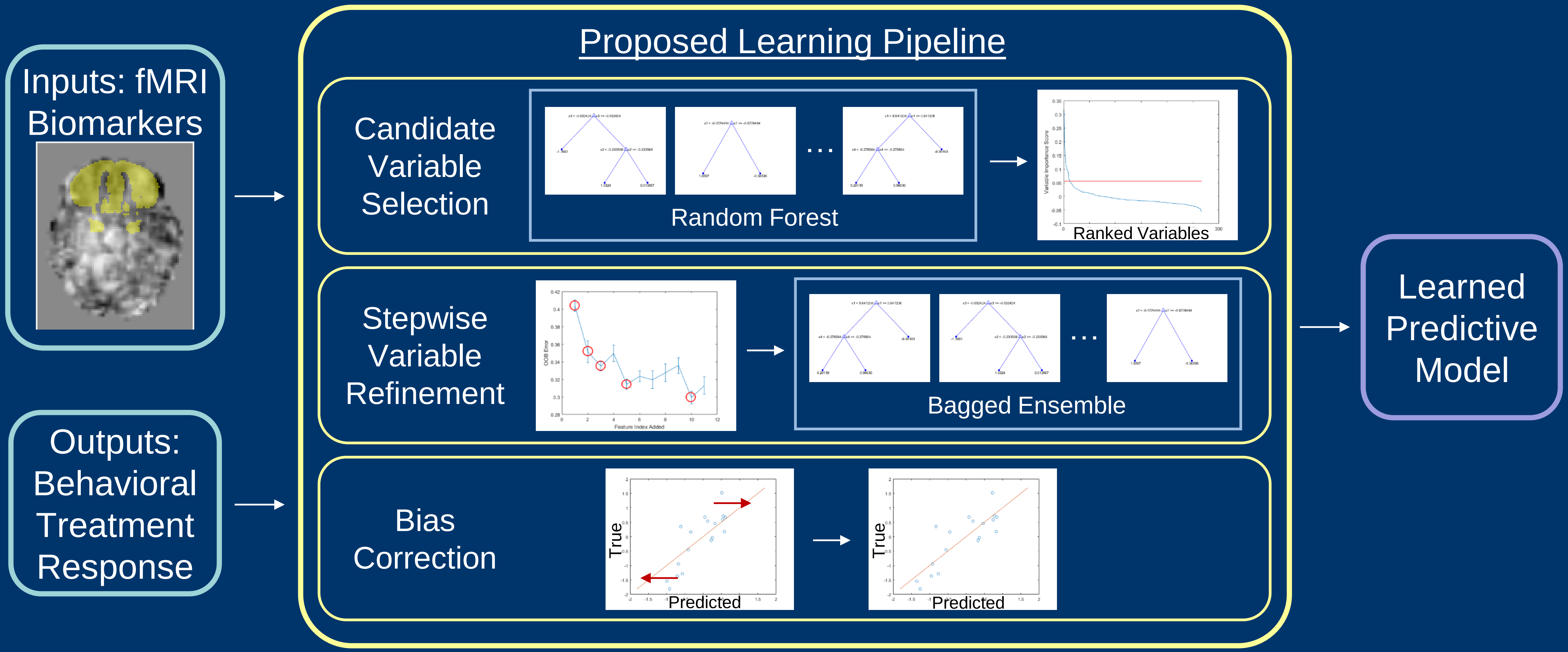
- For each tree and variable
  - Randomly permute values for the OOB samples
  - Calculate change in prediction error
- Importance score: Average change in error over all trees
- Bigger increase in error → higher variable importance
- Note: small negative scores possible due to randomness

Candidate Variable Selection Using Variable Importance

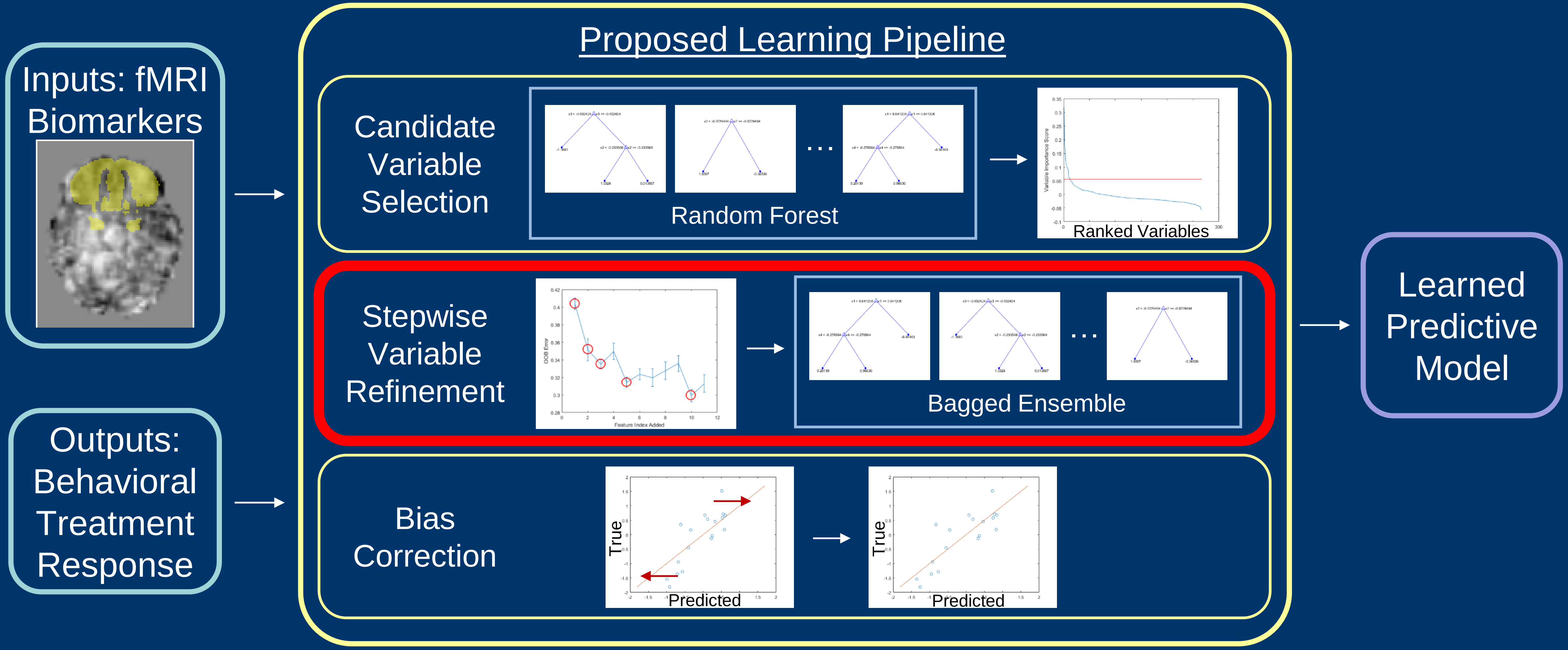
- Run random forests to obtain variable importance scores
- Retain voxels with score > absolute value of lowest negative score
  - Intuition: Irrelevant variables have low scores that fluctuate around 0



Learning Pipeline Overview

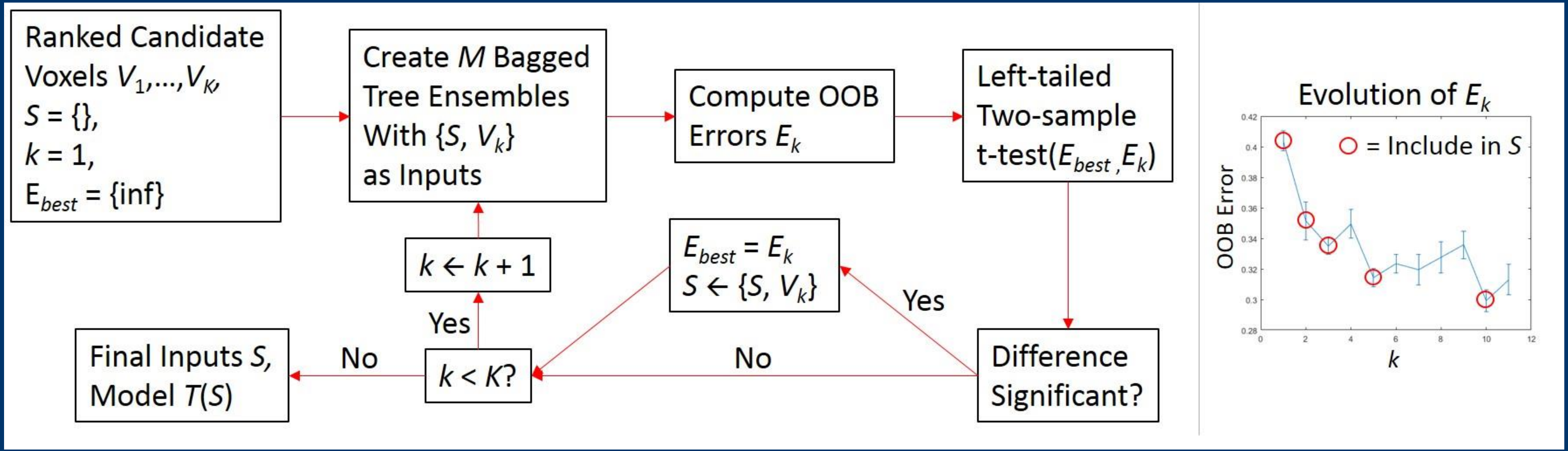


Learning Pipeline Overview

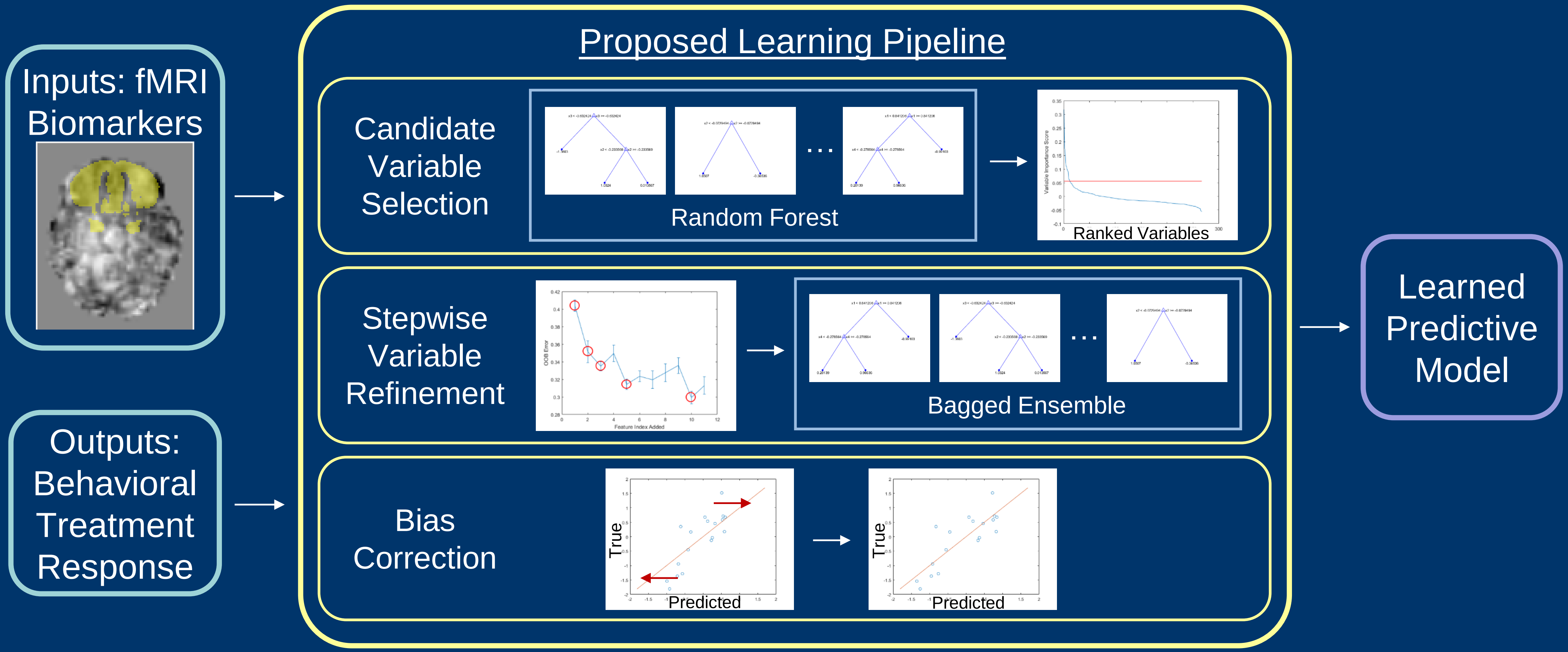


Stepwise Variable Refinement

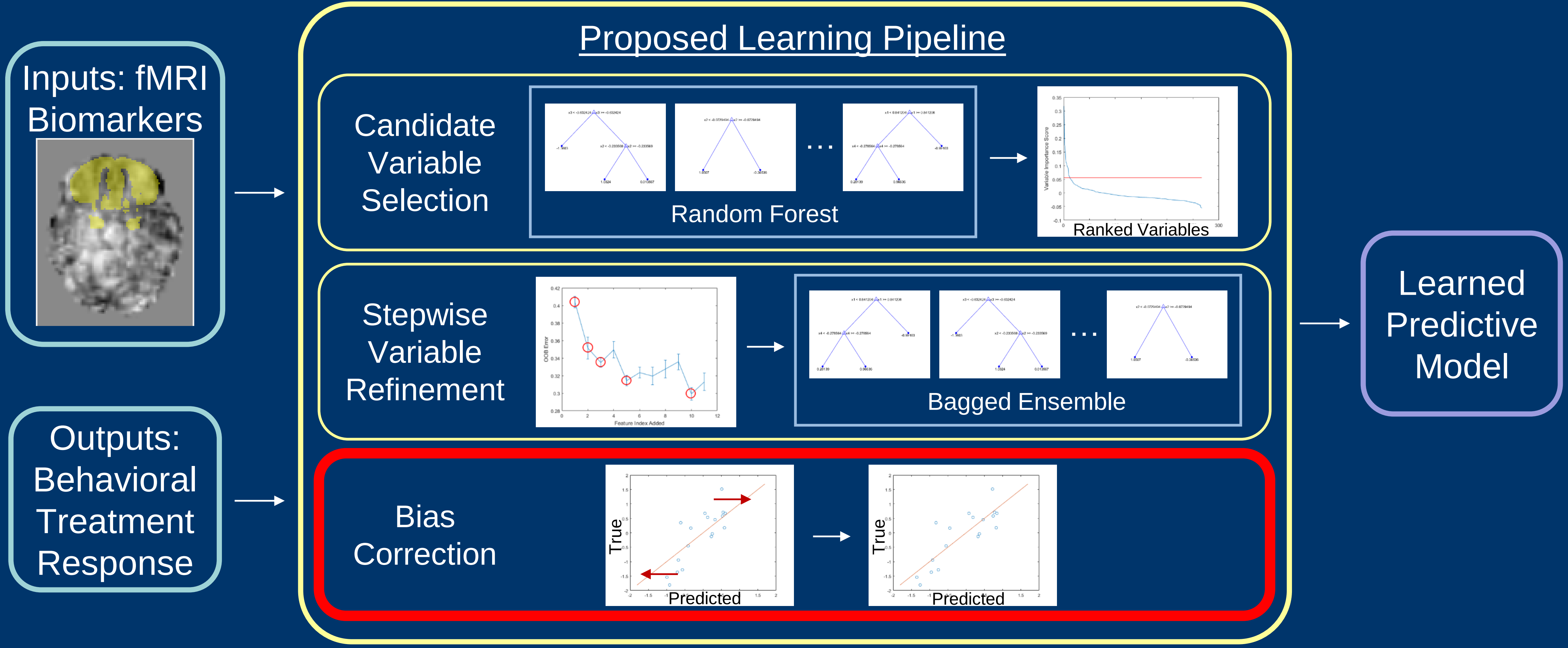
- Iteratively refine candidate input variables for bagged tree ensemble
- V_i = *i*th ranked candidate voxel, *S* = Set of best voxel inputs, *E* = OOB Error



Learning Pipeline Overview

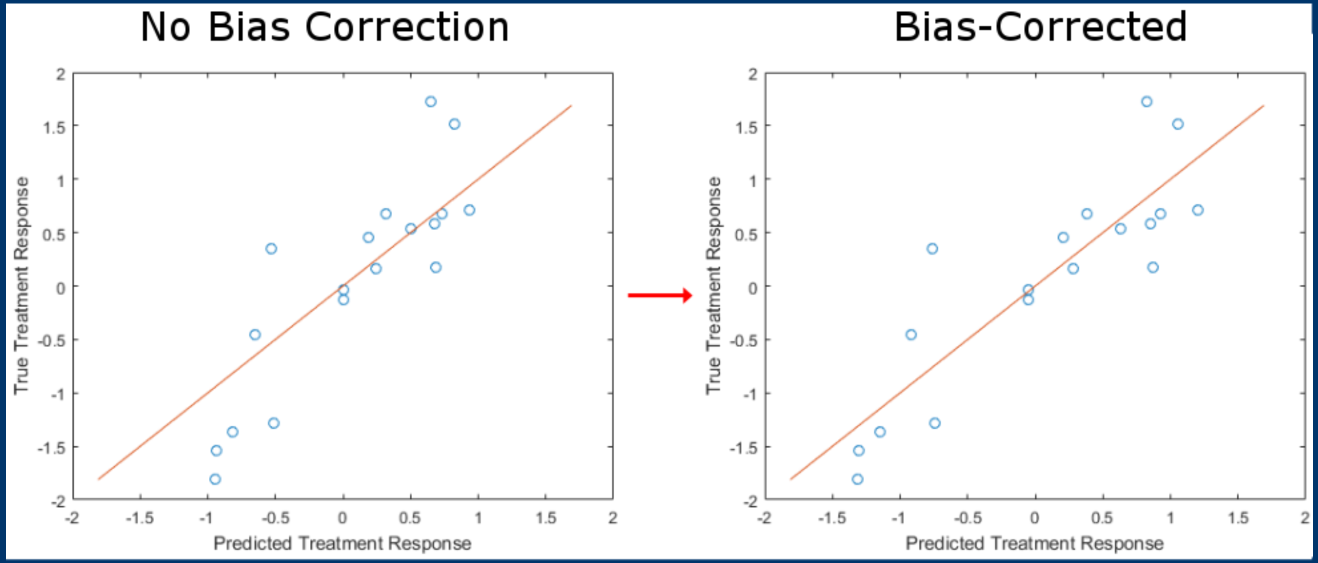


Learning Pipeline Overview



Bias Correction

- Regression tree ensembles underestimate high values and overestimate low values
- Linear model: $Y_{\text{true}} = \beta_1 Y_{\text{ens}} + \beta_0$
- Estimate parameters using OOB predictions



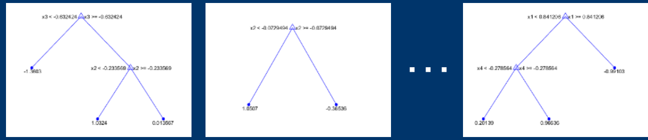
Learning Pipeline Overview

Inputs: fMRI Biomarkers

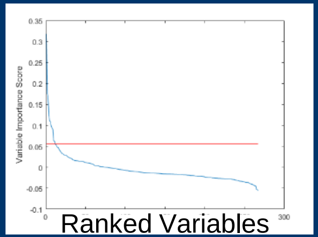
Outputs: Behavioral Treatment Response

Proposed Learning Pipeline

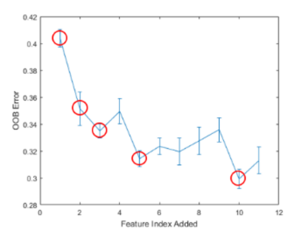
Candidate Variable Selection

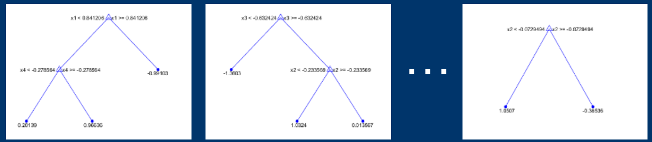


Random Forest



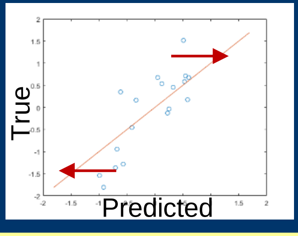
Stepwise Variable Refinement

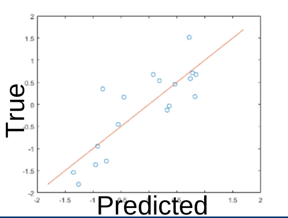




Bagged Ensemble

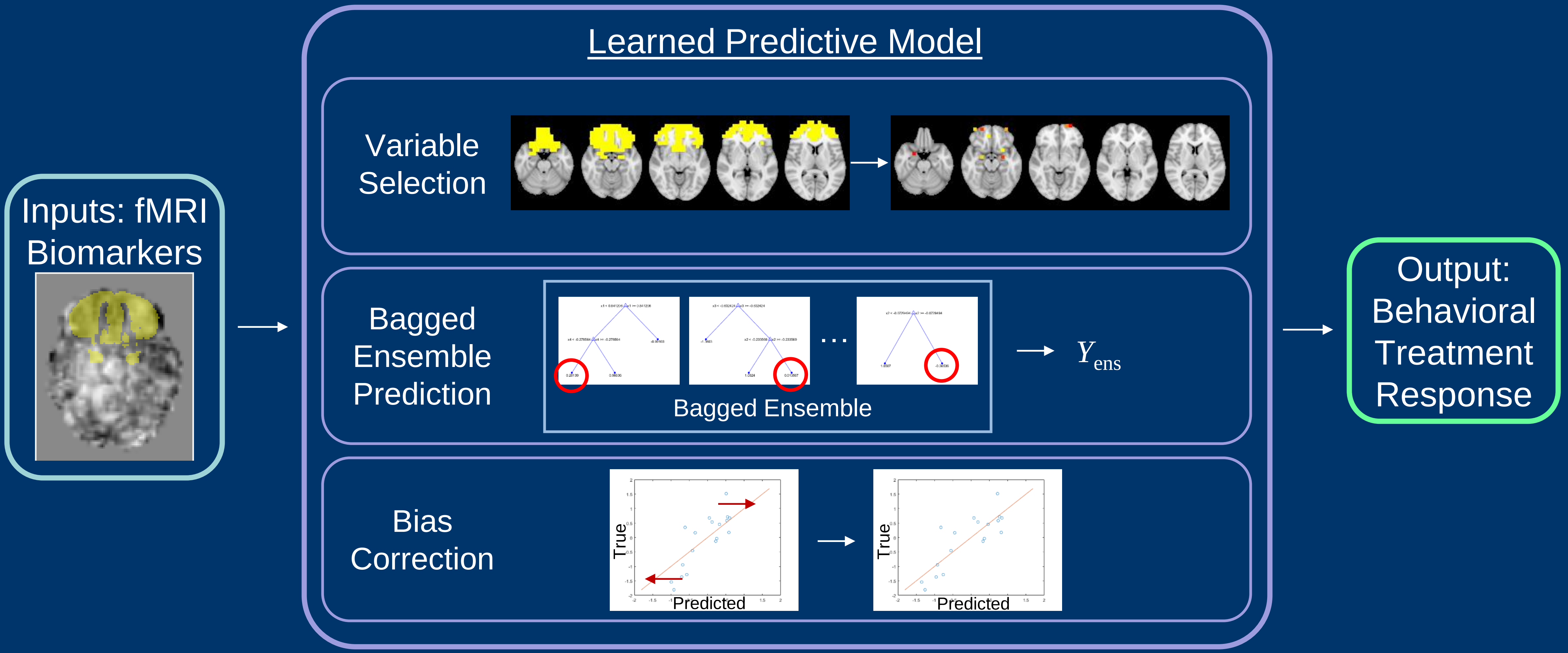
Bias Correction





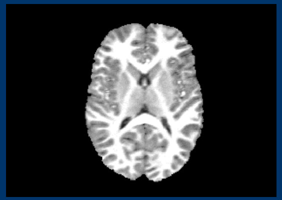
Learned Predictive Model

Predictions from New Data



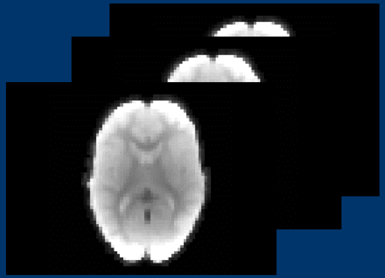
Data

- 19 ASD children underwent 16 weeks Pivotal Response Therapy, 7 hrs/week
- Imaging at baseline:
  - T1-weighted MP-RAGE structural MRI



1 x 1 x 1 mm^3

  - BOLD T2*-weighted fMRI with Biopoint paradigm

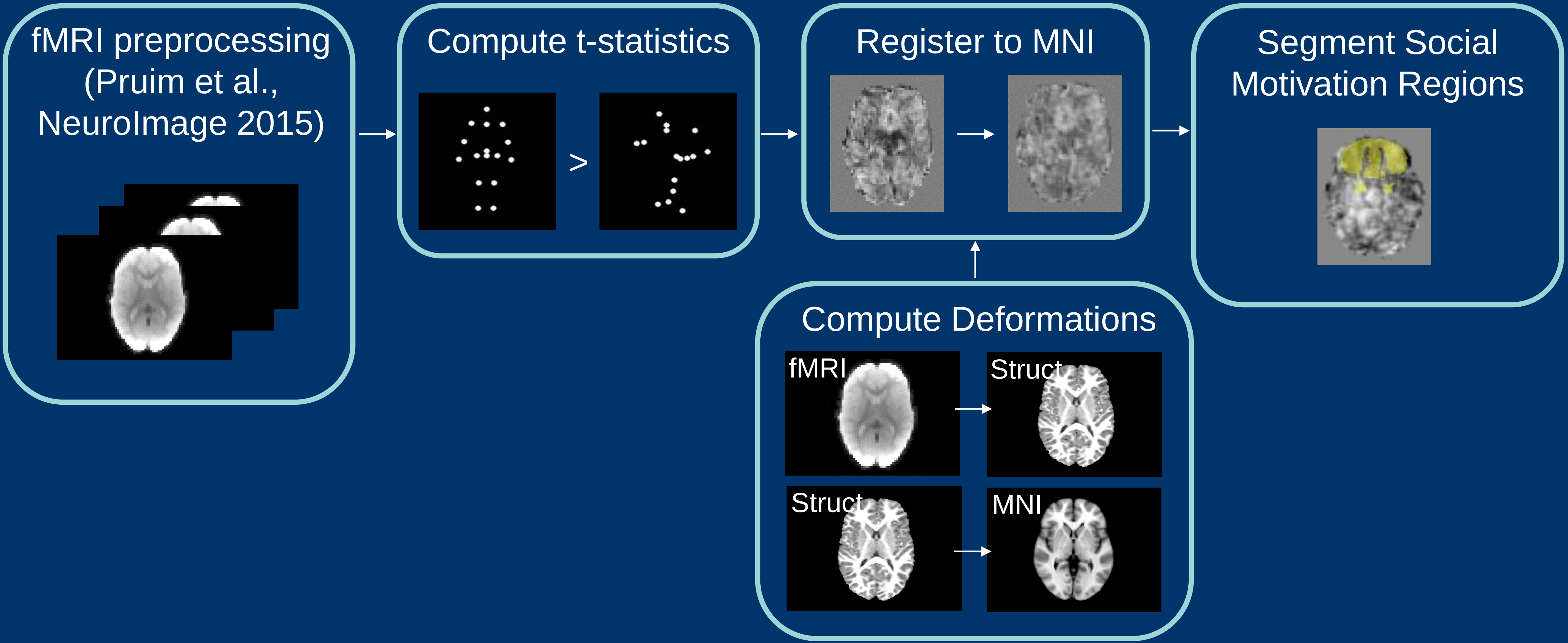


164 volumes

3.44 x 3.44 x 4.00 mm^3

- Note: Data collection involved > 2200 hours

Image Preprocessing



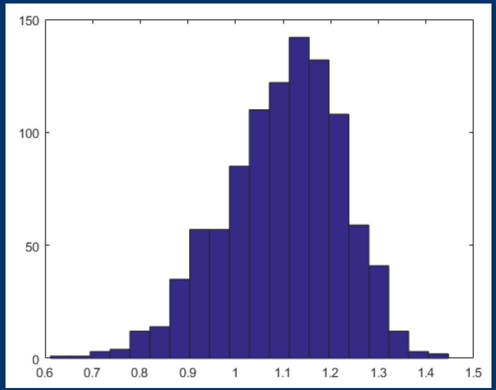
Methods Compared

1. Standard random forest
2. Standard support vector machine with linear kernel
3. Random forest variable selection → random forest
4. Random forest variable selection → bagging
5. Random forest variable selection → stepwise variable refinement
6. Random forest variable selection → stepwise variable refinement → bias correction (Proposed approach)

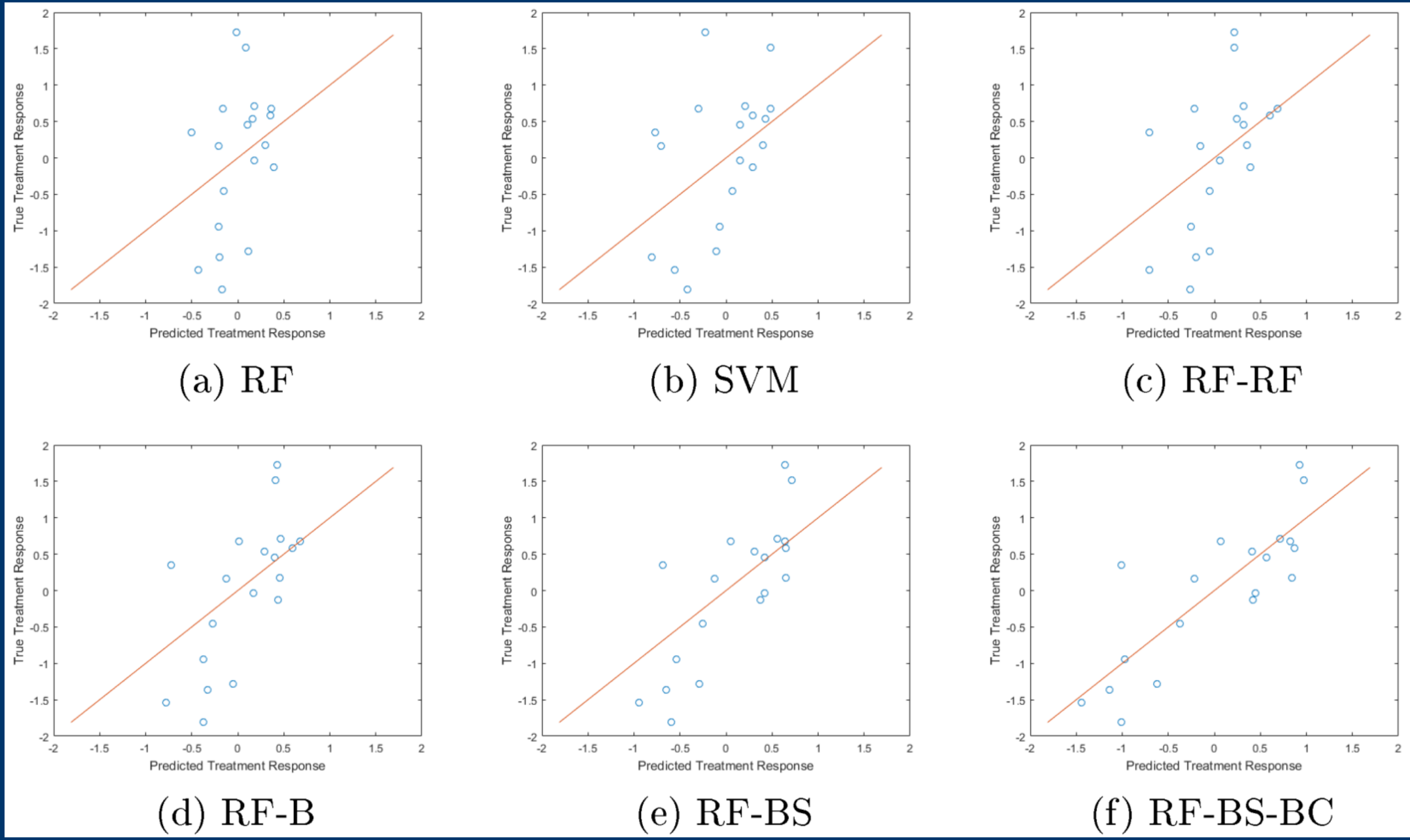
- MATLAB implementation with default parameters, except
  - 5000 trees for variable selection
  - 1000 trees for final models

Evaluation Criteria

- Leave-one-out cross-validation
- Accuracy measures for
  - Outputs (ΔSRS)
    - Mean squared error
    - Pearson's correlation coefficient
  - Predicted outcomes (Post SRS)
    - Relative absolute error
    - Mean absolute percentage error
- Significance assessed using permutation tests
  - p = (# runs with values more extreme than observed statistic) / 1000



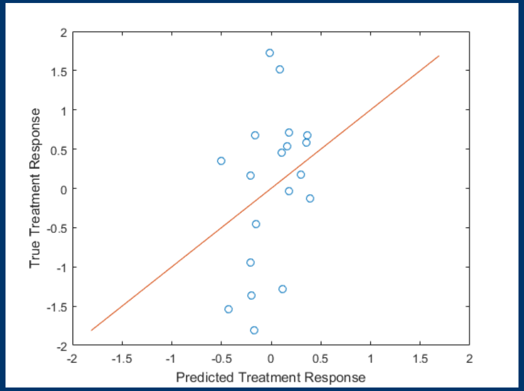
True vs. Predicted Response



Red line: Perfect prediction

Prediction Accuracy

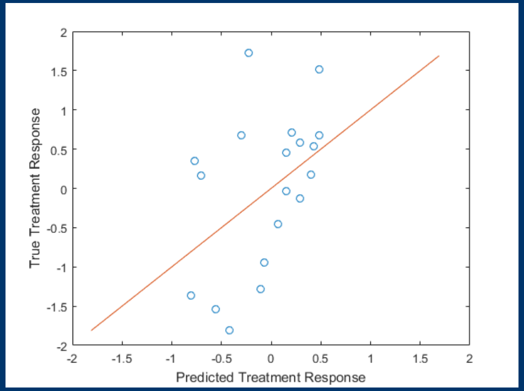
Algorithm	MSE $\pm$ SD	p_{MSE}	r	p_r	RAE	p_{RAE}	MAPE $\pm$ SD	p_{MAPE}
RF	0.82 $\pm$ 0.96	0.019	0.39	0.038	0.63	0.044	0.24 $\pm$ 0.26	0.043
SVM	0.75 $\pm$ 0.93	0.037	0.46	0.040	0.60	0.051	0.22 $\pm$ 0.22	0.051
RF-RF	0.69 $\pm$ 0.80	0.024	0.54	0.023	0.56	0.026	0.21 $\pm$ 0.25	0.030
RF-B	0.57 $\pm$ 0.67	0.012	0.68	0.006	0.50	0.012	0.20 $\pm$ 0.23	0.025
RF-BS	0.40 $\pm$ 0.45	0.001	0.80	0.001	0.44	0.005	0.17 $\pm$ 0.19	0.013
RF-BS-BC	0.29 $\pm$ 0.43	0.001	0.83	0.001	0.35	0.001	0.13 $\pm$ 0.15	0.001



Random forest:
Worst prediction accuracy

Prediction Accuracy

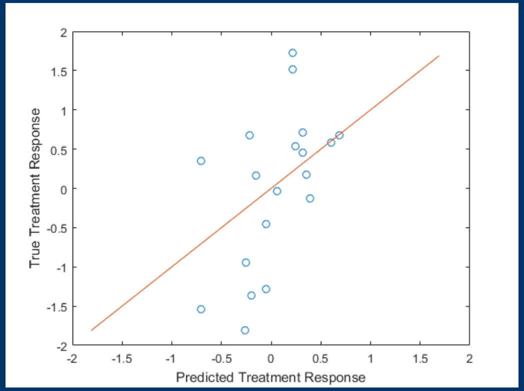
Algorithm	MSE $\pm$ SD	p_{MSE}	r	p_r	RAE	p_{RAE}	MAPE $\pm$ SD	p_{MAPE}
RF	0.82 ± 0.96	0.019	0.39	0.038	0.63	0.044	0.24 ± 0.26	0.043
SVM	0.75 ± 0.93	0.037	0.46	0.040	0.60	0.051	0.22 ± 0.22	0.051
RF-RF	0.69 ± 0.80	0.024	0.54	0.023	0.56	0.026	0.21 ± 0.25	0.030
RF-B	0.57 ± 0.67	0.012	0.68	0.006	0.50	0.012	0.20 ± 0.23	0.025
RF-BS	0.40 ± 0.45	0.001	0.80	0.001	0.44	0.005	0.17 ± 0.19	0.013
RF-BS-BC	0.29 ± 0.43	0.001	0.83	0.001	0.35	0.001	0.13 ± 0.15	0.001



Support vector machine:
Similar errors as random forest

Prediction Accuracy

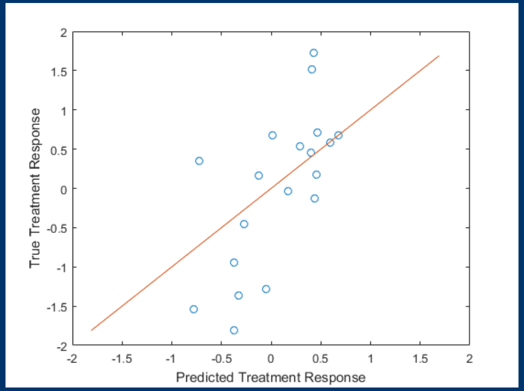
Algorithm	MSE ± SD	p_{MSE}	r	p_r	RAE	p_{RAE}	MAPE ± SD	p_{MAPE}
RF	0.82 ± 0.96	0.019	0.39	0.038	0.63	0.044	0.24 ± 0.26	0.043
SVM	0.75 ± 0.93	0.037	0.46	0.040	0.60	0.051	0.22 ± 0.22	0.051
RF-RF	0.69 ± 0.80	0.024	0.54	0.023	0.56	0.026	0.21 ± 0.25	0.030
RF-B	0.57 ± 0.67	0.012	0.68	0.006	0.50	0.012	0.20 ± 0.23	0.025
RF-BS	0.40 ± 0.45	0.001	0.80	0.001	0.44	0.005	0.17 ± 0.19	0.013
RF-BS-BC	0.29 ± 0.43	0.001	0.83	0.001	0.35	0.001	0.13 ± 0.15	0.001



Select top variables → random forest:
Variable selection improves prediction

Prediction Accuracy

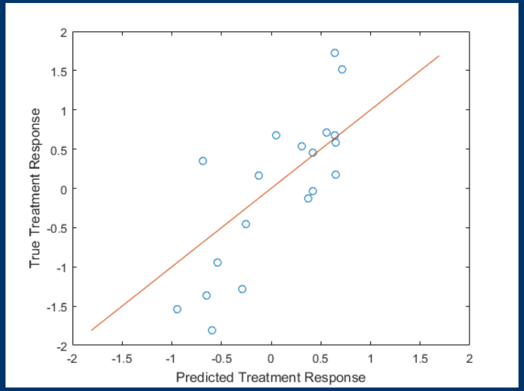
Algorithm	MSE ± SD	p_{MSE}	r	p_r	RAE	p_{RAE}	MAPE ± SD	p_{MAPE}
RF	0.82 ± 0.96	0.019	0.39	0.038	0.63	0.044	0.24 ± 0.26	0.043
SVM	0.75 ± 0.93	0.037	0.46	0.040	0.60	0.051	0.22 ± 0.22	0.051
RF-RF	0.69 ± 0.80	0.024	0.54	0.023	0.56	0.026	0.21 ± 0.25	0.030
RF-B	0.57 ± 0.67	0.012	0.68	0.006	0.50	0.012	0.20 ± 0.23	0.025
RF-BS	0.40 ± 0.45	0.001	0.80	0.001	0.44	0.005	0.17 ± 0.19	0.013
RF-BS-BC	0.29 ± 0.43	0.001	0.83	0.001	0.35	0.001	0.13 ± 0.15	0.001



Select top variables → bagging: Stronger trees reduce errors

Prediction Accuracy

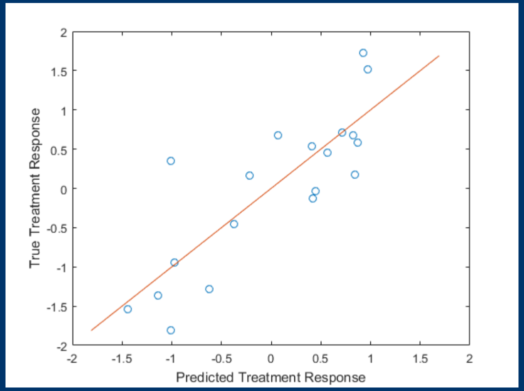
Algorithm	MSE ± SD	p_{MSE}	r	p_r	RAE	p_{RAE}	MAPE ± SD	p_{MAPE}
RF	0.82 ± 0.96	0.019	0.39	0.038	0.63	0.044	0.24 ± 0.26	0.043
SVM	0.75 ± 0.93	0.037	0.46	0.040	0.60	0.051	0.22 ± 0.22	0.051
RF-RF	0.69 ± 0.80	0.024	0.54	0.023	0.56	0.026	0.21 ± 0.25	0.030
RF-B	0.57 ± 0.67	0.012	0.68	0.006	0.50	0.012	0.20 ± 0.23	0.025
RF-BS	0.40 ± 0.45	0.001	0.80	0.001	0.44	0.005	0.17 ± 0.19	0.013
RF-BS-BC	0.29 ± 0.43	0.001	0.83	0.001	0.35	0.001	0.13 ± 0.15	0.001



Stepwise variable refinement:
Improved over bagging top variables

Prediction Accuracy

Algorithm	MSE ± SD	p_{MSE}	r	p_r	RAE	p_{RAE}	MAPE ± SD	p_{MAPE}
RF	0.82 ± 0.96	0.019	0.39	0.038	0.63	0.044	0.24 ± 0.26	0.043
SVM	0.75 ± 0.93	0.037	0.46	0.040	0.60	0.051	0.22 ± 0.22	0.051
RF-RF	0.69 ± 0.80	0.024	0.54	0.023	0.56	0.026	0.21 ± 0.25	0.030
RF-B	0.57 ± 0.67	0.012	0.68	0.006	0.50	0.012	0.20 ± 0.23	0.025
RF-BS	0.40 ± 0.45	0.001	0.80	0.001	0.44	0.005	0.17 ± 0.19	0.013
RF-BS-BC	0.29 ± 0.43	0.001	0.83	0.001	0.35	0.001	0.13 ± 0.15	0.001



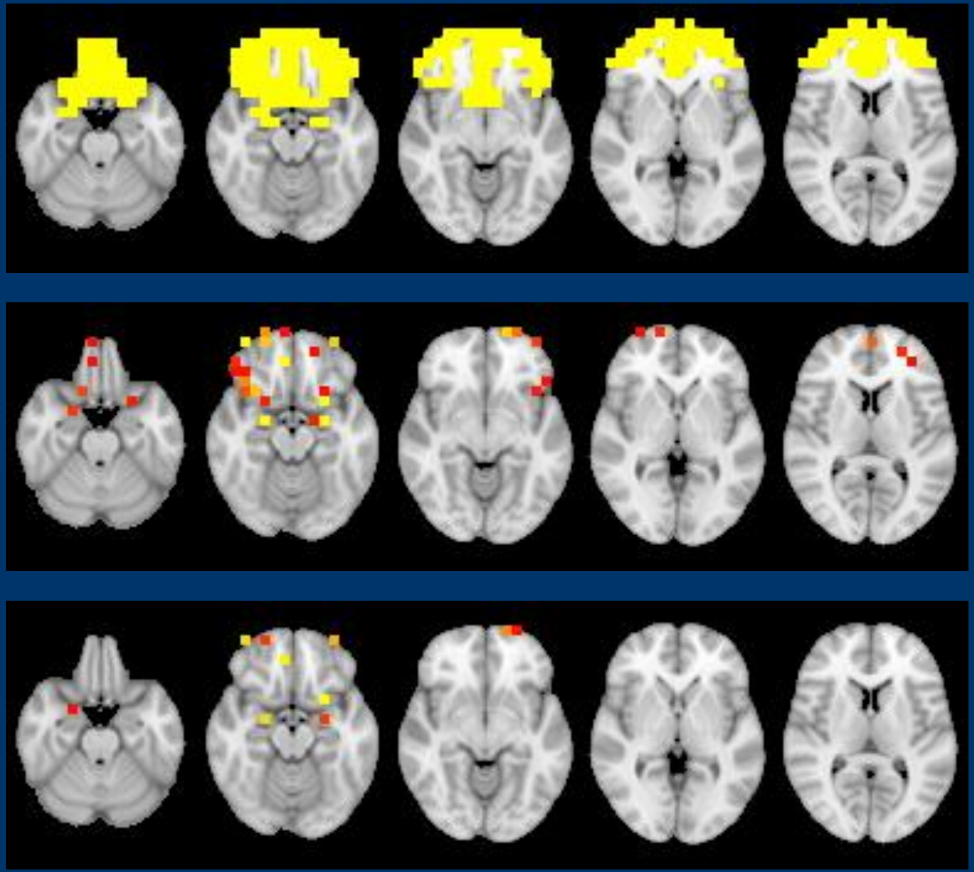
Proposed approach:
Highest prediction accuracy

Variable Selection Results

Social motivation regions (original inputs): Orbitofrontal cortex, ventromedial prefrontal cortex, amygdala, ventral striatum

Random forest candidate variable selection

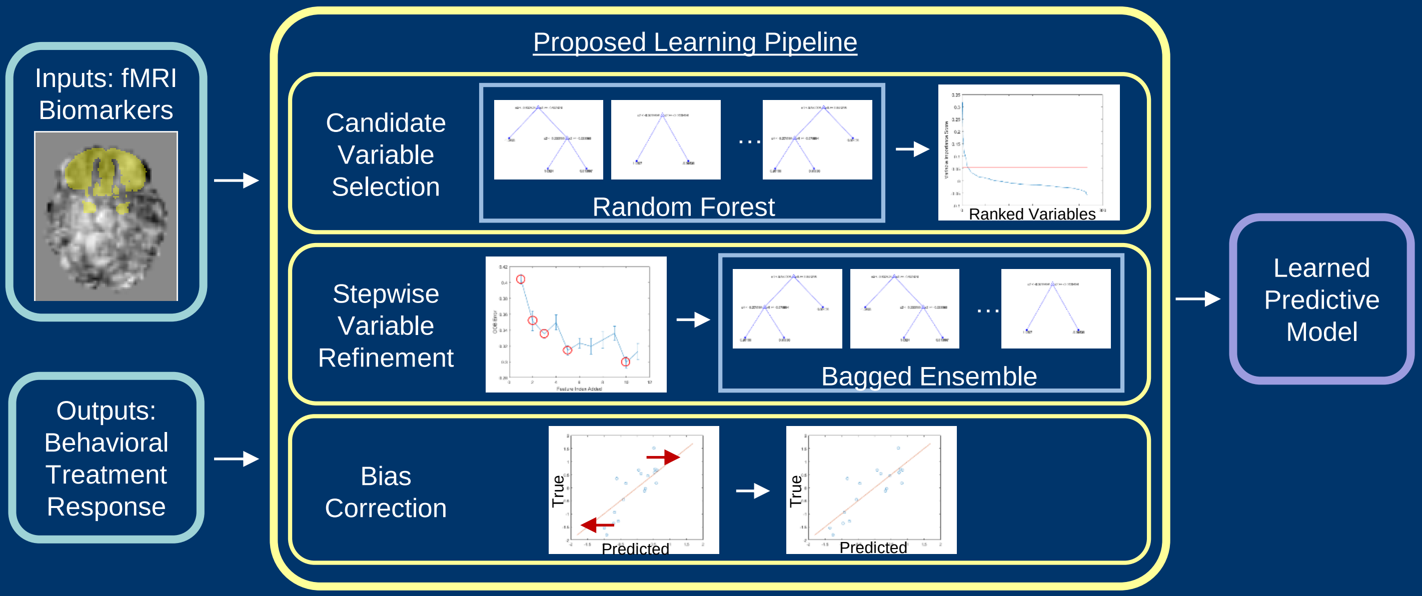
Stepwise variable refinement (final inputs)



Red → Yellow: More frequently selected across trials

Conclusions

- Developed learning pipeline to predict response to autism behavior therapy from baseline fMRI
- Move toward personalized treatment



- Future work
  - Explore other biomarkers for prediction, e.g., functional connectivity
  - More data, assess generalization

Thank You!

- Dr. Pamela Ventola (Data collection)
- Dr. Daniel Yang (fMRI preprocessing)
- NIH grants T32 MH18268 and R01 NS035193

- Contact: nicha.dvornek@yale.edu