2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)
April 4-7, 2018, Washington, D.C., USA

PREDICTION OF PIVOTAL RESPONSE TREATMENT OUTCOME WITH TASK FMRI USING RANDOM FOREST AND VARIABLE SELECTION

Juntang Zhuang[1], *Nicha C. Dvornek* [2,4], *Xiaoxiao Li*[1], *Daniel Yang*[2,3], *Pamela Ventola*[2], *James S. Duncan*[1,4,5]

[1] Biomedical Engineering, Yale University, New Haven, CT USA
[2] Child Study Center, Yale University, New Haven, CT USA
[3] Autism and Neurodevelopmental Disorders Institute, The George Washington Univiersity, DC, USA
[4] Radiology & Biomedical Imaging, Yale School of Medicine, New Haven, CT USA
[5] Electrical Engineering, Yale University, New Haven, CT USA

ABSTRACT

Behavior intervention has shown promise for treatment for young children with autism spectrum disorder (ASD). However, current therapeutic decisions are based on trial and error, often leading to suboptimal outcomes. We propose an approach that employs task-based fMRI for early outcome prediction. Our strategy is based on the general linear model (GLM) and a random forest, combined with feature selection techniques. GLM analysis is performed on each voxel to get t-statistic of contrast between two tasks. Due to the high dimensionality of predictor variables, feature selection is crucial for accurate prediction. Thus we propose a two-step feature selection method: a "shadow" method to select all-relevant variables, followed by a stepwise method to select minimal-optimal set of variables for prediction. A few columns of random noise are generated and added as shadow variables. Regression based on the random forest is performed, and permutation importance of each variable is estimated. Candidate voxels with higher importance than the shadow are kept. Surviving voxels are fed into stepwise variable selection methods. We test both forward and backward stepwise selection. Our method was validated on a dataset of 20 children with ASD using leave-one-out cross-validation, and compared to other standard regression methods. The proposed pipeline generated highest accuracy.

Index Terms— fMRI analysis, Brain, Autism spectrum disorder

1. INTRODUCTION

Autism spectrum disorder (ASD) is a neurodevelopmental syndrome characterized by impaired social interaction, difficulty in communication and repetitive behavior [1]. Behavior based treatment is a widely used therapies for ASD, and Pivotal Response Treatment (PRT) is a empirically-supported treatment [2]. PRT is aimed at improving social communication skills by addressing core deficits in social motivation. However, PRT requires long-time commitment and change of lifestyle, yet the effect is mainly by trial and error. Therefore, prediction of treatment outcome in early stage of treatment is valuable.

fMRI has been used to characterize abnormalties of brain network in ASD [3]. Different tasks can be performed during fMRI scan to reveal changes of the brain, including communication tasks, perceptual learning tasks, visual search tasks, and global precedence tasks[4]. In this study, children with ASD underwent fMRI viewing coherent and scrambled point-light animation of biological motion. This task has been shown to have strong connection with change in the brain of ASD patients [5].

While fMRI has been employed in classification of autistic traits [6] and treatment outcome in other brain disorders [7], it's difficult to predict PRT outcome with high accuracy. Outcome of PRT can be measured by change in social responsiveness scale (SRS) [8]. Yang et.al [9] performed prediction of change in SRS from fMRI imaging, but training and test data were not separated in cluster selection, and it might caused inflated accuracy. Based on the work in [10] by Dvornek et.al, we implemented a carefully constructed cross-validation, and proposed new pipelines to predict treatment outcome with high accuracy.

One challenge in this project is the high dimensionality of input variables, which is called a "small n, large p" problem. We propose a two-step feature selection method to solve this problem. The first step selects all relevant variables [11], and the second step performs stepwise variable selection to find the minimal optimal set of predictor variables. In the first step, independent random noise variables are added as new predictor variables; we name them "shadow" variables. Predictor variables and shadow variables are fed into regression random forest, and permutation importance of all variables are measured by change in out-of-bag (OOB) prediction accuracy. Predictor variables with higher importance than shadow variables are kept. Surviving variables then go through stepwise variable selection and bias correction. Compared with methods in [10], we propose a two-step variable selection

method that is more conservative in the first step, and has lower risk of eliminating truly predictive variables.

2. MATERIAL AND METHODS

2.1. Participants and Imaging tasks

Twenty children with ASD (mean age=5.90 years, SD=1.07 years; 7 females, 13 males; mean IQ=105, SD=16.8) participated in 16 weeks of treatment and fMRI scanning. Baseline SRS scores (mean=82.7, SD=22.6) and post-treatment SRS scores mean=66.5, SD=23.5) were measured.

A T1-weighted MP-RAGE structural MRI ($1mm^3$ resolution) and BOLD T2*-weighted fMRI (164 volumes, $3.44 \times 3.44 \times 4.00mm^3$) were acquired. During fMRI scan, coherent (BIO) and scrambled (SCRAM) point-light biological motion movie were presented to participants in alternating block with 24s duration. The coherent biological motion displays movements related to early childhood experiences, and scrambled motion combines trajectories of 16 randomly selected points from coherent motion[5].

2.2. Image Pre-processing

The T1-weighted image was segmented into grey matter, white matter and cerebrospinal fluid with SPM. fMRI images were processed in FSL with ICA-AROMA (ICA-based strategy for Automatic Removal of Motion Artifacts) [12]. The first 4s of data were discarded to establish T1 equilibrium. fMRI data were then registered to the Montreal Neurological Institute (MNI152) standard brain based on individual structural MRI scan. We performed pre-whitening with FILM in FSL to remove time series autocorrelation. Block design curves were convolved with a gamma function (phase=0s, sd =3s, mean lag=6s) with temporal derivatives. T-statistic of contrast BIO $>$ SCRAM was calculated from GLM in FSL [12].

This t-statistic image was used as input to the prediction pipeline, and output was change in SRS score (post-treatment score minus pre-treatment score). To reduce dimension of the input, we down-sampled the t-statistic image by 4 times in each direction, and chose voxels in brain regions associated with social motivation, including the orbitalfrontal cortex, ventromedial prefrontal cortex, amygdala and ventral striatum [13].

2.3. Prediction Pipeline

Prediction pipeline is demonstrated in figure 1. T-statistics within ROI and corresponding SRS score are training examples, after candidate selection with shadow method, stepwise feature selection is performed. Bias correction is finally performed to improve accuracy.

2.3.1. Fit post-treatment SRS to pre-treatment SRS

Previous studies revealed a strong linear correlation between post-treatment SRS score and pre-treatment SRS score [14]. Therefore, we fit a linear model between post and pre-treatment SRS score from training data, and use residual of this linear model as prediction target for the following pipeline. For test data, residual of SRS score change is predicted from our pipeline, and sum of predicted residual and prediction from linear model is the final prediction of change in SRS score.

$$\widehat{SRS}_{post} = \alpha_0 + \alpha_1 \times SRS_{pre} \tag{1}$$

$$r = SRS_{post} - \widehat{SRS}_{post} \tag{2}$$

where α_0, α_1 are linear model parameters fitted from training data, SRS_{pre}, SRS_{post} are SRS scores before and after PRT, $\widehat{SRS}_{post}$ is estimate of post-treatment SRS score, r is residual of linear fitting.

2.3.2. Candidate variable selection with shadow method

By convention in statistics, input data are reshaped into a 2D matrix, with each column representing a predictor variable, and each row is input data for a participant. To reduce dimension of input and avoid overfitting, we perform candidate variable selection. We generate 50 columns of independent Gaussian random noise, and append them to input data matrix as "shadow variables". The new input matrix is fed into regression random forest, and only variables with higher importance than that of shadow variable importance are kept.

We use out-of-bag (OOB) error to test the importance of each variable in our random forest. The error rate of a random forest is estimated by mean of out-of-bag(OOB) error across different trees. When the value of a variable is permuted, change in mean OOB error represents importance of this variable. If a variable is not predictive, it's possible that permutation causes a negative importance or a large importance by chance. The first step of candidate selection should not be too strict, hence we set the threshold as the median of positive shadow importance. Predictors with higher importance than the threshold are kept. We train a regression random forest on surviving predictors, and retain variables with higher permutation importance than absolute value of lowest negative importance.

2.3.3. Stepwise variable selection

Number of variables after candidate variable selection is still large compared to dataset size, and the dataset may contain many correlated variables. Therefore, we select variables by stepwise building of tree ensembles. We tried both forward and backward stepwise variable selection.

In forward stepwise selection, starting from empty set, in each iteration an unselected variable is added, and regression

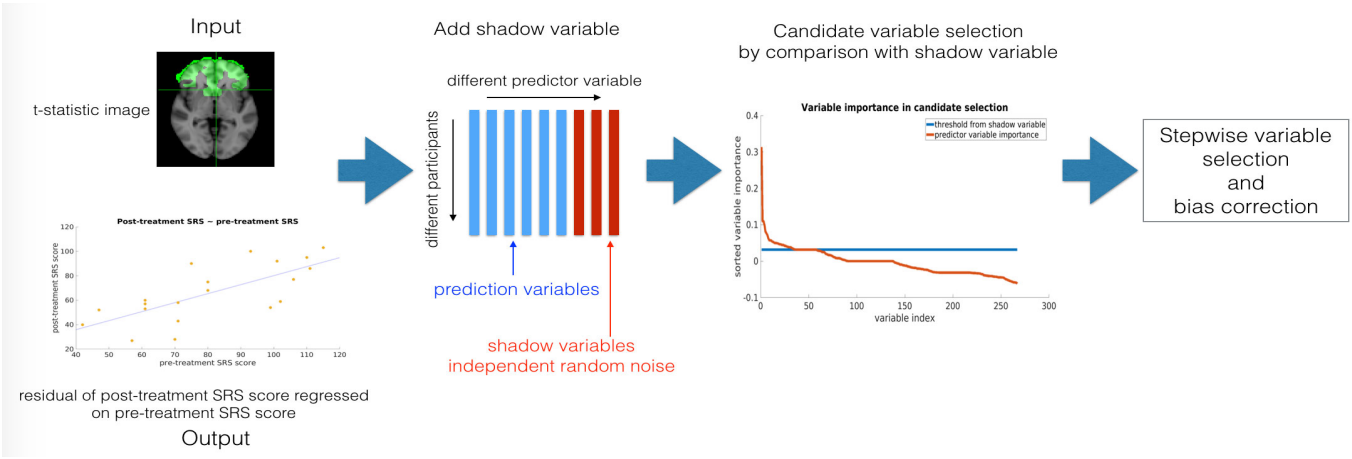


Fig. 1. Prediction pipeline. From left to right: Input training data, add shadow variable, candidate selection with threshold from shadow variables, stepwise variable selection.

Table 1. Prediction accuracy results

Name	RF	RFS	SVR	RF-BS-BC	SB	SF	SBC	**SFC**
MSE	229.7 ± 2.9	224.1 ± 5.0	207.4	136.4 ± 14.6	188.3 ± 6.9	143.7 ± 13.7	130.9 ± 14.1	$\mathbf{125.5 \pm 18.5}$
Correlation	0.42 ± 0.01	0.44 ± 0.02	0.51	0.72 ± 0.04	0.60 ± 0.03	0.72 ± 0.04	0.74 ± 0.04	$\mathbf{0.75 \pm 0.05}$

is repeated several times to generate new error rate. Error rates of regression with and without this new variable are compared using a t-test. If adding a new variable significantly improves accuracy (with $p < 0.05$), this variable is kept; otherwise is deleted.

In backward stepwise selection, starting from the full set of variables, in each iteration a variable is removed, and error rate before and after removal is calculated. The two error rates then go through a t-test, and a variable is removed if its removal causes significant decrease in error rate (with $p < 0.05$).

To reduce computation burden, each variable is tested once in stepwise selection, hence the sequence of addition or removal of a variable is important. Before stepwise selection, variable importance is calculated with random forest based on permutation change in OOB error, using all variables that survive candidate selection. In forward selection, variables are tested in descending order of importance; thus important variables tend to be selected early. In backward selection, variables are tested in ascending order of importance, hence unimportant variables tend to be removed early.

2.3.4. Bias correction

Small dataset size causes instability in prediction, hence bias correction is important. In this step, we calculate out-of-bag prediction of each participant with selected variables, and regress true outcome score on predicted scores. Then we use this linear model to perform bias correction on prediction of the test dataset.

$$r_{train}^{true} = \beta_0 + \beta_1 \times r_{train}^{pred} + \epsilon \tag{3}$$

$$r_{test}^{corrected} = \beta_0 + \beta_1 \times r_{test}^{pred} \tag{4}$$

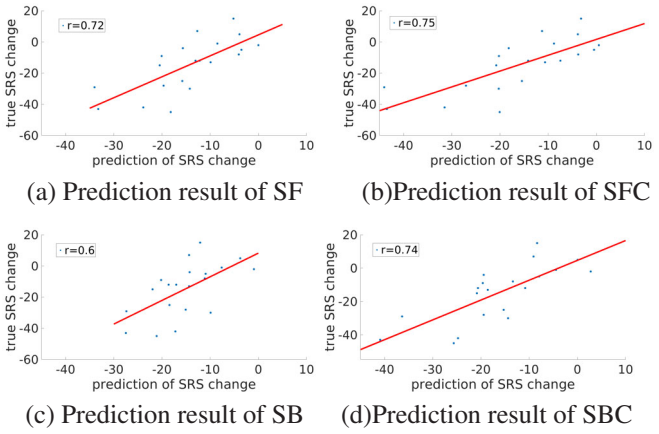
$$\widehat{SRS}_{post}^{test} = \alpha_0 + \alpha_1 \times SRS_{pre} + r_{test}^{corrected} \tag{5}$$

where β_0, β_1 are parameters from linear fitting of OOB predicted residual and true residual, ϵ is noise, r_{train}^{true} is true residual in training dataset, r_{train}^{pred} is OOB predicted residual in training dataset, and r_{test}^{pred} is prediction on test data from random forest.

3. RESULTS AND DISCUSSION

We compared proposed pipelines with other methods , including RF-BS-BC method proposed in [10], support vector regression (SVR) with a linear kernel, a standard random forest (RF), a random forest with candidate selection (RFS), candidate selection with forward selection (SF), candidate selection with foward selection and bias correction (SFC), candidate selection with backward selection (SB), candidate selection with backward selection and bias correction (SBC).

We performed 10 rounds of leave-one-out cross-validation, and measured mean square error (**MSE**) and **correlation** between predicted SRS score change and true outcomes across different rounds of cross-validation. Results are shown in table 1, note there's no randomness across rounds of SVR.



(a) Prediction result of SF (b)Prediction result of SFC

(c) Prediction result of SB (d)Prediction result of SBC

Fig. 2. Mean of prediction across rounds of validation. (a) Prediction from forward stepwise selection pipeline. (b) Prediction from forward stepwise selection pipeline and bias correction. (c) Prediction from backward selection pipeline. (d)Prediction from backward pipeline and bias correction.

Mean of prediction for each test example across rounds are shown in figure 2.

Proposed pipelines perform significantly better than standard methods such as RF and SVR, and SFC achieves highest accuracy. This is due to high dimensionality of predictor variables, our proposed pipeline carefully selects predictive variables, thus performs much better than standard methods.

Compared with RF-BS-BC, SFC and SBC achieve higher prediction accuracy. This validates accuracy of proposed shadow method to select all relevant variables, while RF-BS-BC method has higher risk to remove predictive variables. Number of variables fed into stepwise selection is 9.3 ± 2.8 in RF-BS-BC, and is 13.5 ± 3.7 in proposed pipelines. This is because shadow method selects variable by comparison with importance of random noise, while RF-BS-BC uses a higher importance threshold derived from predictor variables themselves.

Forward stepwise regression performs slightly better than backward stepwise regression. This is because variable selection criterion ($p < 0.05$) is strict, thus more variables are kept in backward selection pipeline. For high dimensional problems, too many correlated variables decreases prediction accuracy. Bias correction increases prediction accuracy, mainly because bias is large in small datasets.

4. CONLUSION

In this paper we proposed pipelines to predict SRS change from pre-treatment fMRI. Our pipeline achieved significantly higher accuracy than standard prediction methods such as RF and SVR, and outperformed RF-BS-BC. Results also validated accuracy of two-step variable selection based on our shadow method.

Selected variables are discrete voxels instead of regions, which is hard for biological interpretation. Future work will extend this method to select brain regions for interpretation, and predict treatment outcome simultaneously. Another direction is to predict treatment outcome from the whole-brain image, instead of from pre-selected regions. .

5. REFERENCES

[1] Catherine Lord et al., "Autism spectrum disorders," *Neuron*, vol. 28, no. 2, pp. 355–363, 2000.

[2] Lynn Kern Koegel et al., "Pivotal response intervention i: Overview of approach," *Journal of the Association for Persons with Severe Handicaps*, vol. 24, no. 3, pp. 174–185, 1999.

[3] Marcel Adam Just et al., "Cortical activation and synchronization during sentence comprehension in high-functioning autism: evidence of underconnectivity," *Brain*, vol. 127, no. 8, pp. 1811–1821, 2004.

[4] Kate Plaisted et al., "Children with autism show local precedence in a divided attention task and global precedence in a selective attention task," *The Journal of Child Psychology and Psychiatry and Allied Disciplines*, vol. 40, no. 5, pp. 733–742, 1999.

[5] Martha D Kaiser et al., "Neural signatures of autism," *Proceedings of the National Academy of Sciences*, vol. 107, no. 49, pp. 21223–21228, 2010.

[6] Tali M Ball et al., "Single-subject anxiety treatment outcome prediction using functional neuroimaging," *Neuropsychopharmacology*, vol. 39, no. 5, pp. 1254, 2014.

[7] Mark Plitt et al., "Resting-state functional connectivity predicts longitudinal change in autistic traits and adaptive functioning in autism," *Proceedings of the National Academy of Sciences*, vol. 112, no. 48, pp. E6699–E6706, 2015.

[8] Teryn P Bruni, "Test review: Social responsiveness scale–second edition (srs-2)," 2014.

[9] D Yang et al., "Brain responses to biological motion predict treatment outcome in young children with autism," *Translational psychiatry*, vol. 6, no. 11, pp. e948, 2016.

[10] Nicha C Dvornek et al., "Prediction of autism treatment response from baseline fmri using random forests and tree bagging," *Multimodal Learning for Clinical Decision Support*, 2016.

[11] Roland Nilsson et al., "Consistent feature selection for pattern recognition in polynomial time," *Journal of Machine Learning Research*, vol. 8, no. Mar, pp. 589–612, 2007.

[12] Stephen M Smith et al., "Advances in functional and structural mr image analysis and implementation as fsl," *Neuroimage*, vol. 23, pp. S208–S219, 2004.

[13] Coralie Chevallier et al., "The social motivation theory of autism," *Trends in cognitive sciences*, vol. 16, no. 4, pp. 231–239, 2012.

[14] Pamela Ventola et al., "Improvements in social and adaptive functioning following short-duration prt program: a clinical replication," *Journal of autism and developmental disorders*, vol. 44, no. 11, pp. 2862–2870, 2014.