

Bootstrapping promotes the RSFC-behavior associations: an application of individual cognitive traits prediction

Bootstrapping promotes RSFC prediction

Lijiang Wei, Bin Jing, Haiyun Li

School of Biomedical Engineering, Capital Medical University, Beijing, China

Lijiang Wei and Bin Jing are co-senior authors.

Correspondence

Haiyun Li, School of Biomedical Engineering, Capital Medical University, Beijing, China

E-mail: haiyunli@ccmu.edu.cn;

Funding information

Beijing Municipal National Science Foundation, Grant/Award Number: 4122018

ACKNOWLEDGMENTS

We are grateful to Xiangyu Ma from School of Biomedical Engineering, Capital Medical University, Beijing for valuable comments for the manuscript. This work was supported by the Beijing Municipal National Science Foundation (4122018).

Abstract

Resting state functional connectivity records enormous functional interaction information between any pair of brain nodes, which enriches the prediction of individual phenotypes. To reduce the high dimensional features in prediction, correlation analysis is a common way for feature selection. However, rs-fMRI signal exhibits typically low signal-to-noise ratio and correlation analysis is sensitive to outliers and data distribution, which may bring unstable and uninformative features to

subsequent prediction. To alleviate this problem, a bootstrapping-based feature selection framework was proposed and applied on three widely used regression models: connectome-based predictive model (CPM), support vector regression (SVR) and least absolute shrinkage and selection operator (LASSO). A large open-source dataset from Human Connectome Project (HCP) was adopted in the study and a series of cognitive traits were acted as the prediction targets. To systematically investigate the influences of different parameter settings on the bootstrapping-based framework, a total of 216 parameter combinations were evaluated through the R value between the predicted and real cognitive traits, and the best identified performance among them was chosen out as the final prediction accuracy for each cognitive trait. By using bootstrapping without replacement, the best performances of CPM with positive and negative feature sets, SVR and LASSO averagely increased by 28.0%, 33.2%, 11.6% and 24.3% in R values in contrast to the baseline method without bootstrapping. By using bootstrapping with replacement, these best performances increased by 22.1%, 22.9%, 9.4% and 19.6%. Furthermore, the bootstrapping-based feature selection methods could effectively refine the original feature sets obtained from correlation analysis, which thus retained the more stable and informative feature sets. The results demonstrate that bootstrapping-based feature selection is an easy-to-use and effective method to improve RSFC prediction of cognitive traits and is highly recommended in future RSFC prediction studies.

KEYWORDS: bootstrapping, functional connectivity, cognitive trait, individualized prediction, feature selection

1 | INTRODUCTION

Resting state functional connectivity (RSFC) has been found to be associated with pathological conditions and physiological status of human brain (Cribben et al., 2012; Elanbari et al., 2014; Finn et al., 2015; Meskaldji et al., 2016). Decoding RSFC provides an insight into understanding the neural mechanism underlying intrinsic brain activities (R. T. Jiang et al., 2018). Previous studies have reported that FC patterns could be applied to predict individual behavioral and cognitive phenotypes, such as attention ability (Yoo et al., 2018), cognition (Cui et al., 2018; Kong et al., 2019; Sun et al., 2019), intelligence ability (Finn et al., 2015; R. Jiang et al., 2019; Santarnecchi et al., 2014) and chronological age (Dosenbach et al., 2010; Liem et al., 2017). To model the RSFC-phenotype association, three prediction models including connectome-based predictive modeling (CPM), support vector regression (SVR) and least absolute shrinkage and selection operator (LASSO) have been frequently adopted for their good performances and interpretabilities (Coloigner et al., 2016; de Vos et al., 2018; R. Jiang et al., 2019; Meng et al., 2017; Ryali et al., 2012; Shen et al., 2017). CPM is in fact a linear regression model and has been successfully applied in predicting individual cognition (Finn et al., 2015; R. Jiang et al., 2019; Sripada et al., 2019), attention (Rosenberg et al., 2017), cocaine abstinence (Yip et al., 2019) and reading comprehension (Jangraw et al., 2018). SVR implements space conversion by some kernel function in order to achieve better prediction in new data space than in original data space (Basak et al., 2007), and has been utilized in prediction of mental disease (Rizk-Jackson et al., 2011), brain maturity (Dosenbach et al., 2010; Nielsen et al., 2019) and painful sensation (Tu et al., 2015). LASSO provides sparse representation by driving redundant features to zero-valued weights, and performs well in investigation of reward-related behavior (Ferenczi et al., 2016), mental state (Haufe et al., 2014) and various aspects of cognition (R. Jiang et al., 2019; R. T. Jiang et al., 2018; Rashid et al., 2014; Toiviainen et al., 2014).

For the high-dimensional RSFC, feature reduction is indispensable since redundant or

irrelevant information may confound the statistical testing significance (Bunea et al., 2011), worsen the machine learning model performance (Arbabshirani et al., 2017; Duangsoithong et al., 2010) and increase computational complexity. For the regression tasks, correlation analysis between RSFC and the target phenotypic measure (e.g. cognitive trait) is the most common approach to select features (Dadi et al., 2019; Gabrieli et al., 2015). However, correlation analysis may not be able to catch the reliable correlational relationship since rs-fMRI data exhibits low signal-to-noise ratio (SNR) and correlation analysis is sensitive to outliers (R. Wilcox, 2004; R. R. Wilcox, 2005). Small perturbation in the data may lead to inclusion of unrelated features or exclusion of useful features, thus the weak robustness of correlation can introduce false correlations including type I error and power problem (Rousselet et al., 2012). In addition, data distribution may be another easily ignored factor in the correlation analysis: it is also susceptible to the clustered points, curvature, heteroscedasticity and range (Rousselet et al., 2012). All these conditions may cause that the features detected from the whole dataset did not significantly correlate with the target phenotypic measure in a subset, resulting in unstable features, which may in fact contribute little to the target phenotype prediction. Thus, an adequate feature selection method that could detect intrinsic stable RSFC features is of great importance to the prediction task.

Bootstrapping is a statistical method that relies on the random sampling (Efron et al., 1997; Hall et al., 2009), and any focused statistical test can be examined in newly resampled datasets. After repeating the resampling process several times, the stability of the statistical significance in each resampled dataset could be accessed, which could be finally used to ascertain the statistical test by setting specific threshold. The main advantage of bootstrapping is that it does not rely on any assumption on the data, so it can ascertain the actual and stable between-group statistical significance. A few literatures have adopted bootstrapping to investigate the stability of selected features. In Alonso-Atienza et al. (2012), bootstrapping was used as a backward feature selection method for cardiac ventricular fibrillation discrimination. Ditzler reported

that bootstrapping combined with Neyman-Pearson hypothesis test successfully detected the statistically relevant features on both synthetic and real data (Ditzler et al., 2015). In Bunea et al. (2011), features which were selected most frequently by penalized least squares regression methods in bootstrapping resamples were identified as useful predictors in neuroimaging. However, few studies tried to explore the influence of different parameter settings on the bootstrapping. For example, Bunea et al. (2011) used bootstrapping with only one cutoff threshold (50% inclusion frequency) to derive the features. Abram et al. (2016) proposed a quantile threshold bootstrapping method for feature selection in penalized regression models. Hence, studies were highly needed to make known how different parameter settings in bootstrapping influenced the final prediction performance.

This study explored a bootstrapping-based feature selection framework to enhance the RSFC prediction performance in CPM, SVR and LASSO models, and different parameter settings in bootstrapping were specially compared in order to ascertain the optimal parameter setting, which mainly included two categories: bootstrapping without replacement (i.e., subsample without replacement) and bootstrapping with replacement. To verify these bootstrapping-based feature selection methods, a large public dataset from Human Connectome Project was used in this study, and 13 cognitive traits were selected as the prediction targets. We speculated that bootstrapping could improve the RSFC prediction performances of CPM, SVR and LASSO models.

2 | METHODS AND MATERIALS

2.1 | Resting-state fMRI data

The data came from publicly WU-Human Connectome Project (HCP) 1200 release. All imaging data were collected on a 3T customized Siemens scanner using a standard 32-channel head coil with a multiband pulse sequence. Two scanning sessions (REST1 and REST2) of high spatial-temporal resolution resting state fMRI (rs-fMRI)

data were acquired during two consecutive days. Both sessions were scanned at right-to-left and left-to-right phase encoding direction, and the functional images comprised 1200 volumes in one session. For more details of inclusion and exclusion criteria on the dataset, please see Van Essen et al. (2013).

2.2 | Preprocessing

The extensively preprocessed PTN (Parcellation+ Timeseries+ Netmats) data were utilized in our study. Details of preprocessing steps could be found in Smith et al. (2013). In brief, each run of rs-fMRI data underwent a minimal preprocessed pipeline and ICA+FIX to remove the potential artifacts. All 4 runs were concatenated to form 4800 volumes for each individual. After region parcellation by group-ICA, the subject-specific FC matrices, or connectomes, were calculated using Pearson's correlation and then Gaussianised into Z-stats. Here, we adopted 300 ICA components as network nodes, resulting in a connectivity matrix with size 300×300 for each participant. Notably, there were two types of fMRI data reconstruction methods available in the release, and it was not clear the potential influences induced by different reconstruction algorithms. To avoid the heterogeneity in data, the current study limited the analyses to 812 healthy subjects (ages 22-37) whose rs-fMRI data were reconstructed using a later improved version of construction method, and discarded data with original reconstruction version.

2.3 | Cognitive traits

Given that the cognitive activities were functionally modulated by the brain, prediction of cognitive traits by RSFC was a feasible task (Stevens, 2009). Previous studies had shown that RSFC patterns were closely associated with cognitive traits (Beaty et al., 2018; Sun et al., 2019). In this study, 13 measures of cognitional tests were chosen as the prediction targets that provided by HCP (Li et al., 2019). These tests included Picture Sequence Memory, Dimensional Change Card Sort, Flanker Task, Oral Reading Recognition, Picture Vocabulary, Pattern Completion Processing

Speed, List Sorting, Penn Progressive Matrices, Delay Discounting, Variable Short Penn Line Orientation, Short Penn Continuous Performance and Penn Word Memory. Seven of them measured by NIH (National Institutes of Health) Toolbox were normalized to age-adjusted scores with mean of 100 and standard deviation of 15. Furthermore, not all subjects had the whole 13 cognitive scores and subjects with missing cognitive data were excluded in subsequent analysis.

2.4 | Bootstrapping-based feature selection methods

The bootstrapping-based feature selection methods sought to find stable features that were consistently identified in the resampling subsets. Conventional bootstrapping method mainly focused on resample with replacement, but it usually required a high amount of resample times, which seemed not efficient enough for feature selection. Moreover, medical imaging dataset was usually not large enough, in this condition, bootstrapping without replacement might suit the finite population. In this study, a wide range of parameter settings on bootstrapping with and without replacement were tested and evaluated to make sure the optimal parameter setting of bootstrapping.

2.4.1 | Bootstrapping without replacement

Bootstrapping without replacement took out samples without replacement from the original dataset, so each subject in the resampling dataset was unique without duplication. For each cognitive measure, the bootstrapping without replacement was applied on training set to extract the feature vector. Let NB (number of bootstrapping) denote the sampling times. Within each resampling iteration, a portion of subjects were randomly selected without repetition. The proportion of sampling was defined as bootstrap percentage (BP), e.g., $BP = 50\%$ indicated half of all subjects were chosen from original dataset to form a resample subset. Spearman correlation analysis was then conducted in each resampling set between RSFC and cognitive measure, and significant features were selected as predictors. After all bootstrapping iterations finished, a stability threshold for frequency percentage (FP) of feature was used to

determine final feature sets in *NB* bootstrapping datasets. For example, if the stability threshold was set to 50%, then feature whose *FP* more than 50% would be selected in the final feature set, otherwise it would be filtered as a weak predictor.

2.4.2 | Bootstrapping with replacement

In bootstrapping with replacement method, resamples were obtained with replacement from original samples, so the sample size of resampling subset was same as original dataset size. There were only two parameters in this method, i.e. *NB* and *FP*. After performing Spearman correlation analysis in *NB* bootstrapping subsets, features with *FP* larger than stability threshold were chosen as predictors.

2.5 | Regression models

Three widely used regression models were employed in the study to predict the cognitive traits with RSFC, including CPM, SVR and LASSO.

2.5.1 | CPM

Briefly, CPM summarized the relevant RSFCs for each individual and fitted the summary values with cognitive measures in a linear regression model. The detail steps of CPM were as follows:

- 1) Find features in RSFC matrices that correlated with the cognitive measure, and significantly positive or negative features were respectively selected.
- 2) Summarize selected feature values of positive feature set and negative feature set separately for each individual.
- 3) Perform univariate linear regression between the summary value and the cognitive measure for positive and negative feature sets separately.
- 4) Test the model performance with the unseen testing data.

Here, label the CPM model using positive or negative feature sets respectively as CPM-P or CPM-N.

2.5.2 | SVR

SVR was based on statistical learning theory and fitted linear model with Vapnik's ε -sensitive loss function (Smola et al., 2004). It tried to solve an objective function $f(x)$ whose predicted values of training data deviated by no more than ε from actual values and flatness of regression line was maximized. Samples that deviated by more than ε from their actual values were called support vectors. Given N training samples $((x_1, y_1), \dots, (x_N, y_N))$, the objective function took the following form:

$$\begin{aligned} \min_{\beta, \xi_i, \hat{\xi}_i} & \frac{1}{2} \|\beta\|^2 + C \sum_{i=1}^m (\xi_i + \hat{\xi}_i) \\ \text{s.t.} & f(x_i) - y_i \leq \varepsilon + \xi_i, \\ & y_i - f(x_i) \leq \varepsilon + \hat{\xi}_i, \\ & \xi_i, \hat{\xi}_i \geq 0 \end{aligned}$$

where β was the regression coefficient vector for features, m was the number of support vectors, ξ_i and $\hat{\xi}_i$ were slack variables and C was the penalty parameter that controlled the trade-off of penalty between variance and bias. The objective function was solved to obtain a weight for each support vector. In this study, the linear kernel was selected for SVR.

2.5.3 | Lasso

Lasso was a penalized regression method by adding L1-norm regularization to ordinary least squares (Tibshirani, 1996). The method could attenuate regression coefficients of correlated features except one feature among them to zeros (Zou et al., 2005), thus generated robust coefficients as well as parsimonious models. This algorithm minimized ordinary least squares plus the sum of absolute values of regression coefficients β to obtain the estimated coefficients. The objective function was formed as below:

$$\min_{\beta} \sum_{i=1}^N (f(x_i) - y_i)^2 + \lambda \sum_{j=1}^P |\beta_j|$$

where λ was the tuning parameter which controlled balance between accuracy and

sparsity.

At last, CPM, SVR and LASSO accompanied with correlation analysis on the whole dataset were marked as the baseline method, and the corresponding models combined with the bootstrapping-based feature selection were regarded as the proposed method.

2.6 | Prediction Framework

For each cognitive score, a 10-fold cross-validation (CV) framework was utilized in the study. All individuals were sorted according to values of cognitive measure and then divided into 10 folds. To be specific, the (1st, 11st, 21st ...) subjects were assigned to 1st fold, the (2^{ed}, 12^{ed}, 22^{ed} ...) subjects were assigned to 2^{ed} fold and so on. This partition procedure ensured the same distribution of 10 folds and avoided random bias and expensive computation due to random splitting (Cui et al., 2018). Within the prediction framework, each fold was iteratively used for testing while the remaining 9 folds were used for training. The bootstrapping feature selection methods were performed on training set to select features for CPM, SVR and LASSO models (Figure 1). Significantly positive-, negative- and combined-correlated FCs were separated into positive, negative and combined feature sets (significant threshold $p=0.05$), which were respectively used to build CPM-P, CPM-N, SVR and LASSO models. In both bootstrapping method with and without replacement, a broad set of parameters including *NB*, *BP* and *FP* were tested, where *NB* was chosen from [10, 20, 50, 100, 500, 1000] and *FP* from [50%, 60%, 70%, 80%, 90%, 100%]. Besides, in bootstrapping without replacement, *BP* was chosen from [25%, 50%, 60%, 70%, 80%]. All parameter settings resulted in 180 combinations in bootstrapping without replacement and 36 combinations in bootstrapping with replacement. For SVR, an inner 5 fold-CV was employed to determine optimal parameters *C*, choosing from $[10^{-10}, 10^{-9} \dots 10^{10}]$. For LASSO, an inner 10 fold-CV was employed to determine the optimal λ from range 10^{-4} to 1 with step size 10^{-2} . The model performance was evaluated by the *R* value between predicted and actual cognitive values. In the study,

the reported R values were averaged over 10 testing folds. All procedures were carried out with a multi-core CPU (Intel(R) Xeon(R) E5-2630 v4 @2.20GHz, 10 cores, 20 threads) with 128 GB memory.

3 | RESULTS

3.1 | Bootstrapping improved prediction of cognitive traits

Figure 2 summarized the best prediction accuracies of 13 cognitive traits prediction achieved by bootstrapping methods and baseline method. All models using the refined feature sets derived from bootstrapping method outperformed the baseline method. Compared with baseline method (average $r=0.15$, 0.15 , 0.24 and 0.18 for CPM-P, CPM-N, SVR and LASSO separately), the bootstrapping without replacement method increased the predictive correlation values by 28.0%, 33.2%, 11.6% and 24.3%. Among four models, SVR performed best and reached an average $r=0.27$, LASSO achieved suboptimal performance at average $r=0.23$ and CPM-P and CPM-N achieved similar performances of average $r=0.19$ and $r=0.20$. In bootstrapping with replacement, the accuracy increased by 22.1%, 22.9%, 9.4% and 19.6% for CPM-P, CPM-N, SVR and LASSO separately. Similar to bootstrapping without replacement, bootstrapping with replacement displayed best accuracies in SVR (average $r=0.26$), suboptimal result in LASSO (average $r=0.22$) and similar results in CPM-P (average $r=0.18$) and CPM-N (average $r=0.19$). The best prediction accuracies of 13 cognitive measures were illustrated in Figure 3 for bootstrapping without and without replacement.

3.2 | Bootstrapping methods largely reduced feature dimension

The feature dimension refined by bootstrapping methods decreased enormously compared with the baseline method (Figure 4). Using bootstrapping without replacement, the average feature dimension decreased from 2574 to 209 in CPM-P, from 2648 to 136 in CPM-N, from 5222 to 1359 in SVR and from 5221 to 351 in LASSO. Using bootstrapping with replacement, the average feature dimension

decreased to 521 in CPM-P, 188 in CPM-N, 2892 in SVR and 1771 in LASSO.

Generally, bootstrapping without replacement reduced more features than bootstrapping with replacement.

3.3 | Time consumption was dominated by number of bootstrapping

The time consumption of the bootstrapping methods gradually increased with NB (figure 5). Under the same NB , bootstrapping with replacement consumed more time than bootstrapping without replacement, which demonstrated the efficiency of bootstrapping without replacement. When NB exceeded 500, the average consumption time became significantly larger, implying that NB was preferably set to be less than 500.

3.4 | Optimal parameter settings were not constant for different cognitive traits

The parameter sets (i.e., NB , FP and BP) which achieved best prediction performance for each cognitive measure were depicted in Figure 6. Here, a dot in the parameter space referred to the best parameter setting for one cognitive measure. The results showed most of the cognitive scores could be best estimated by small values of NB , and 83% and 88% of best performances were respectively achieved with NB less than 100 for bootstrapping without and with replacement. As for FP and BP , they varied in different models for each cognitive trait. For CPM, large BP (around 0.8) and moderate FP (0.6-0.8) achieved good performances in bootstrapping without replacement, while high FP (0.8-1.0) was optimal for bootstrapping with replacement. For SVR, it appeared that there were no obvious rules in BP and FP . For LASSO, a large BP (0.7-0.8) seemingly obtained good performances in bootstrapping without replacement.

4 | DISCUSSION

In order to improve the performance of RSFC prediction, this study developed bootstrapping based feature selection framework for RSFC prediction of cognitive traits. Using a large sample from HCP dataset, the prediction of cognitive traits with bootstrapping methods significantly outperformed the baseline CPM, SVR and LASSO models, and the dimension of selected features decreased dramatically using the bootstrapping methods. In addition, bootstrapping without replacement displayed superior improvements and higher running efficiency in contrast to bootstrapping with replacement.

4.1 | Bootstrapping enhanced the RSFC-phenotype associations.

Modeling cognitive traits with RSFC is a persistent pursuit in modern cognitive neuroscience studies (Petersen et al., 2015; Poldrack et al., 2015). A previous study shows that the diversity of individual cognition can be represented by the inter-subject variations in the FC patterns (Mueller et al., 2013). However, the FC patterns may be noisy and unreliable (Braga et al., 2017; Gordon et al., 2017; Gratton et al., 2018; Mueller et al., 2013), thus confound the prediction precision. In recent years, improving RSFC-phenotype association is a major concern in the domain, and several methodologies have been proposed from different perspectives. The first way is to use individual-specific RSFC patterns to predict the behavioral traits. In Qin et al. (2019) and Kashyap et al. (2019), they decompose RSFC (or timecourse) to extract individual-specific RSFC (or timecourse) with different methods, which obtained obvious improvements in prediction. A second way is to combine various kinds of information with RSFC to enhance the prediction, such as task-fMRI based FC (Elliott et al., 2019; Gao et al., 2019; Xiao et al., 2019) and dynamic FC (Liegeois et al., 2019; Lim et al., 2018; Park et al., 2018), which can provide complementary information to the conventional FC. A third way is to decrease the influence of the possible noise in rs-fMRI signal, for instance, global signal regression (Li et al., 2019) and motion artifact correction (Nielsen et al., 2019) have been reported to advance the

RSFC-behavior prediction. The last way is to use the bagging strategy (Breiman, 1996) to improve the prediction with RSFC (Jollans et al., 2019).

Our proposed method provides another perspective to promote the RSFC-phonotype prediction, and displays several advantages. First, the method does not need any collection of additional data (e.g. task fMRI) or any other processing (e.g. computation of dynamic FC and FC decomposition), so it is an easy-to-use and effective method. Second, the current method can purify the selected feature sets, making the prediction model more succinct and interpretable. It is known that reducing feature dimension is always beneficial to attenuate overfitting in neuroimaging data (Jollans et al., 2019). At last, previous studies (Alonso-Atienza et al., 2012; Ditzler et al., 2015) used bootstrapping or bagging in a computation expensive manner, which requires the model construction and optimization in each bootstrapping resample dataset. This may lead to an unacceptable running time for the heuristic searching of optimal parameter setting in large neuroimaging dataset, and only have to use constant parameter setting to conduct such methods.

As we know, good stability of biomarkers is thought as important as high classification performances (He et al., 2010). Currently, the importance of the stability of the rs-fMRI signal (e.g. test-retest reliability) and the stability of selected biomarkers across sites (e.g. multi-center dataset) have been widely accepted (Marchitelli et al., 2016; Mueller et al., 2015; Noble, Scheinost, et al., 2017; Noble, Spann, et al., 2017), however, there is seldom study special focusing on the stability of selected biomarkers across subsamples in one dataset. Indeed, heterogeneity exists in nearly all human neuroimaging datasets regardless of whether the samples are healthy or not, which may therefore result in instability in feature sets derived from subsamples and further weaken the prediction performance. It is necessary to eliminate the heterogeneity in the dataset and to obtain stable representations standing for neural activities. Figure 4 clearly shows that the amounts of false correlated feature sets are large, implying that it is very beneficial to take into account the

stability of feature selection for machine learning studies with neuroimaging dataset. Our study demonstrated that the proposed bootstrapping-based framework is an effective way to eliminate the heterogeneity in the dataset and to obtain feature sets with good stability.

4.2 | Bootstrapping with replacement vs. bootstrapping without replacement

Bootstrapping with and without replacement were compared in the study. Though bootstrapping with replacement was referred more frequently than bootstrapping without replacement, it was reported not to guarantee reliable results sometimes. In Strobl et al. (2007), feature selection in random forest was biased by bootstrapping with replacement, while importance measures of predictors could be accessed more reliably by bootstrapping without replacement. This phenomenon was not hard to explain, bootstrapping with replacement ensured the resample size as same as the original dataset, so the replication samples might introduce possible bias into the final feature selection, which was especially obvious in a finite dataset (e.g. neuroimaging data). For bootstrapping without replacement, the resamples was a subset of original dataset, so they could be regarded as sampling from the same population, and properties of statistical test in resamples were constant with the original data (Rospleszcz et al., 2016). Moreover, bootstrapping without replacement obtained greater prediction improvements, reduced more features that were redundant and ran more efficiently than bootstrapping with replacement. Hence, we recommended bootstrapping without replacement as a feature selection technique to draw reliable predictors.

4.3 | Comparison of CPM, LASSO and SVR

For the prediction performance, CPM achieved the worst performance compared with SVR and LASSO. There were three possible aspects accounted for it: 1. CPM is a simple linear regression model, while SVR and LASSO are models with L1 and

L2-norm regularization, which may therefore obtain better prediction performances and also increase the model generalization. 2. The correlated feature sets were respectively divided into positive and negative feature sets in CPM, while SVR and LASSO used their combination, so such division may weaken the CPM performance significantly. 3. CPM assigned same weight to all features, which might underestimate key predictors and overestimate weak predictors. Notably, CPM achieved the largest improvement in prediction and the most obvious reduction of features using bootstrapping methods, suggesting future studies with CPM should consider the bootstrapping methods.

LASSO achieved suboptimal prediction performance in three models and it ensured sparseness by randomly selecting one feature from correlated features. LASSO was shown to perform well in challenging situation like potential features size exceeded the sample size (Abraham et al., 2017; Bunea et al., 2007a; Bunea et al., 2007b; Zhang et al., 2008). However, when confronting the condition that the feature number was remarkably larger than the sample size, it might meet such a problem: LASSO could only retain no more than $N-1$ features (N is the sample size), thus some informative features might be discarded in the model (Efron et al., 2004; Ryali et al., 2012).

SVR displayed the best performance among three models. Compared with LASSO, SVR performed slightly better and the similar trends were also revealed by Cui et al. (2018) and Dadi et al. (2019) on RSFC predictions. However, efficiency was a crucial problem for high-dimensional data while SVR was tremendously expensive in computation for model fitting (Cui et al., 2018; Shen et al., 2017).

4.4 | Optimal selection of parameter setting

In bootstrapping with and without replacement, most best performances were achieved with NB less than 100, implying that bootstrapping did not require a mass of

resampling times to acquire stable features, so we recommended small NB (<100) for bootstrapping methods used in CPM, SVR and LASSO models. PF indicated the stability of the feature in the resampling subset, thus specifying an appropriate threshold for FP was a key setting. In Bunea et al. (2011), the FP threshold was set to a single value (50%) in LASSO in order to not miss any possibly relevant features, but our study found the optimal FP threshold varied for different cognitive traits, indicating that constant threshold setting could not guarantee the best performance in LASSO. For CPM, the optimal PF threshold was moderate values (0.6-0.8) but not the very high values (0.8-1) in bootstrapping without replacement, which reflected that outliers and clustered points might also exist in the bootstrapping resamples, therefore, FP threshold should not be set with very high values in CPM. BP stood for the resample percentage from the original dataset, and our results discovered that moderate values (0.7-0.8) were suitable for CPM and LASSO in bootstrapping without replacement. For SVR, there were no significant rules for all parameter settings, so the best parameter settings should be searched for each prediction target.

4.5 | Other methodological considerations

Group ICA has been found to be a better way to define the functional brain nodes than the atlas-based method for its better prediction accuracy (Dadi et al., 2019), hence, group ICA parcellation was adopted to calculate the corresponding FC in the study. There were various brain parcellations derived from group ICA provided in HCP dataset, i.e., [15, 25, 50, 100, 200, 300]. To have direct comparisons among different parcellations, the baseline CPM, SVR and LASSO methods were used to predict the cognitive functions with FC matrix under different parcellations. The results were illustrated in Figure 7, and the main tendency was that the prediction accuracy increased with dimension except LASSO, and grown slowly from 100 to 300 independent components, and similar trend was also revealed in other works (Abraham et al., 2017; Dadi et al., 2019). To investigate the bootstrapping methods on high-dimensional data, the analysis was based on 300 independent components.

Besides, significant threshold p-value was an important parameter for correlation analysis. In order to conserve as much as possible features, the p-value was set at 0.05 rather than other more stringent value in this study. In addition, concerning different head motion correction strategies and global signal regression were still controversial in functional connectivity studies, and we just focused on the bootstrapping related improvement, so we did not compare any other preprocessing steps on the preprocessed HCP data.

4.6 | Limitation and future work

The bootstrapping methods were more expensive in computation compared with the baseline method, which could be speeded up via parallel computing with multiple CPUs, making the bootstrapping methods feasible in practical use. An additional disadvantage was that the optimal parameter settings varied in different cognitive phenotype predictions, so a heuristic search was needed for new prediction target. Although the bootstrapping method could improve the RSFC-phenotype association, the prediction accuracy was still not high enough, and future work should combine other improvement ways together to make the RSFC-based prediction applicable for clinical requirement.

5 | CONCLUSION

This study presented a bootstrapping-based feature selection framework and was applied to CPM, SVR and LASSO models, the results demonstrated the bootstrapping method could not only improve the RSFC-phenotype association but also purify the selected features into a lower dimension. Future RSFC prediction works were highly recommended to use the bootstrapping method.

REFERENCES

Abraham, A., Milham, M. P., Di Martino, A., Craddock, R. C., Samaras, D., Thirion, B., & Varoquaux, G. (2017). Deriving reproducible biomarkers from multi-site

- resting-state data: An Autism-based example. *Neuroimage*, 147, 736-745.
<https://doi.org/10.1016/j.neuroimage.2016.10.045>
- Abram, S. V., Helwig, N. E., Moodie, C. A., DeYoung, C. G., MacDonald, A. W., & Waller, N. G. (2016). Bootstrap Enhanced Penalized Regression for Variable Selection with Neuroimaging Data. *Frontiers In Neuroscience*, 10.
[https://doi.org/Artn 34410.3389/Fnins.2016.00344](https://doi.org/Artn%2034410.3389/Fnins.2016.00344)
- Alonso-Atienza, F., Rojo-Alvarez, J. L., Rosado-Munoz, A., Vinagre, J. J., Garcia-Alberola, A., & Camps-Valls, G. (2012). Feature selection using support vector machines and bootstrap methods for ventricular fibrillation detection. *Expert Systems with Applications*, 39(2), 1956-1967.
<https://doi.org/10.1016/j.eswa.2011.08.051>
- Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage*, 145, 137-165. <https://doi.org/10.1016/j.neuroimage.2016.02.079>
- Basak, D., Pal, S., & Patranabis, D. C. (2007). Support Vector Regression.
- Beaty, R. E., Kenett, Y. N., Christensen, A. P., Rosenberg, M. D., Benedek, M., Chen, Q. L., Fink, A., Qiu, J., Kwapil, T. R., Kane, M. J., & Silvia, P. J. (2018). Robust prediction of individual creative ability from brain functional connectivity. *Proceedings Of the National Academy Of Sciences Of the United States Of America*, 115(5), 1087-1092.
<https://doi.org/10.1073/pnas.1713532115>
- Braga, R. M., & Buckner, R. L. (2017). Parallel Interdigitated Distributed Networks within the Individual Estimated by Intrinsic Functional Connectivity. *Neuron*, 95(2), 457-+. <https://doi.org/10.1016/j.neuron.2017.06.038>
- Breiman, L. (1996). Bagging predictors. *Machine Learning*, 24(2), 123-140.
[https://doi.org/Doi 10.1023/A:1018054314350](https://doi.org/Doi%2010.1023/A:1018054314350)
- Bunea, F., She, Y. Y., Ombao, H., Gongvatana, A., Devlin, K., & Cohen, R. (2011). Penalized least squares regression methods and applications to neuroimaging. *Neuroimage*, 55(4), 1519-1527.
<https://doi.org/10.1016/j.neuroimage.2010.12.028>
- Bunea, F., Tsybakov, A., & Wegkamp, M. (2007a). Sparsity oracle inequalities for the Lasso. *Electronic Journal Of Statistics*, 1, 169-194.
<https://doi.org/10.1214/07-EJS008>
- Bunea, F., Tsybakov, A. B., & Wegkamp, M. H. (2007b). Aggregation for gaussian regression. *Annals Of Statistics*, 35(4), 1674-1697.
<https://doi.org/10.1214/009053606000001587>
- Coloigner, J., Phlypo, R., Bush, A., Lepore, N., & Wood, J. (2016). Functional connectivity analysis for thalassemia disease based on a graphical lasso model. Paper presented at the IEEE International Symposium on Biomedical Imaging.
- Cribben, I., Haraldsdottir, R., Atlas, L. Y., Wager, T. D., & Lindquist, M. A. (2012). Dynamic connectivity regression: Determining state-related changes in brain connectivity. *Neuroimage*, 61(4), 907-920.
<https://doi.org/10.1016/j.neuroimage.2012.03.070>
- Cui, Z. X., & Gong, G. L. (2018). The effect of machine learning regression

- algorithms and sample size on individualized behavioral prediction with functional connectivity features. *Neuroimage*, 178, 622-637. <https://doi.org/10.1016/j.neuroimage.2018.06.001>
- Dadi, K., Rahim, M., Abraham, A., Chyzyk, D., Milham, M., Thirion, B., Varoquaux, G., & Initia, A. D. N. (2019). Benchmarking functional connectome-based predictive models for resting-state fMRI. *Neuroimage*, 192, 115-134. <https://doi.org/10.1016/j.neuroimage.2019.02.062>
- de Vos, F., Koini, M., Schouten, T. M., Seiler, S., van der Grond, J., Lechner, A., Schmidt, R., de Rooij, M., & Rombouts, S. A. R. B. (2018). A comprehensive analysis of resting state fMRI measures to classify individual patients with Alzheimer's disease. *Neuroimage*, 167, 62-72. <https://doi.org/10.1016/j.neuroimage.2017.11.025>
- Ditzler, G., Polikar, R., & Rosen, G. (2015). A Bootstrap Based Neyman-Pearson Test for Identifying Variable Importance. *Ieee Transactions on Neural Networks And Learning Systems*, 26(4), 880-886. <https://doi.org/10.1109/Tnnls.2014.2320415>
- Dosenbach, N. U. F., Nardos, B., Cohen, A. L., Fair, D. A., Power, J. D., Church, J. A., Nelson, S. M., Wig, G. S., Vogel, A. C., Lessov-Schlaggar, C. N., Barnes, K. A., Dubis, J. W., Feczko, E., Coalson, R. S., Pruett, J. R., Barch, D. M., Petersen, S. E., & Schlaggar, B. L. (2010). Prediction of Individual Brain Maturity Using fMRI. *Science*, 329(5997), 1358-1361. <https://doi.org/10.1126/science.1194144>
- Duangsoithong, R., & Windeatt, T. (2010). Bootstrap Feature Selection for Ensemble Classifiers.
- Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *Annals Of Statistics*, 32(2), 407-451.
- Efron, B., & Tibshirani, R. (1997). Improvements on cross-validation: The .632+ bootstrap method. *Journal Of the American Statistical Association*, 92(438), 548-560. <https://doi.org/Doi 10.2307/2965703>
- Elanbari, M., Nemmour, N., Bouhali, O., Rawi, R., Sheharyar, A., & Bensmail, H. (2014). Computational Neuro-Modeling of Visual Memory: Multimodal Imaging and Analysis. Paper presented at the International Conference on Brain Informatics & Health.
- Elliott, M. L., Knodt, A. R., Cooke, M., Kim, M. J., Melzer, T. R., Keenan, R., Ireland, D., Ramrakha, S., Poulton, R., Caspi, A., Moffitt, T. E., & Hariri, A. R. (2019). General functional connectivity: Shared features of resting-state and task fMRI drive reliable and heritable individual differences in functional brain networks. *Neuroimage*, 189, 516-532. <https://doi.org/10.1016/j.neuroimage.2019.01.068>
- Ferenczi, E. A., Zalocusky, K. A., Liston, C., Grosenick, L., Warden, M. R., Amatya, D., Katovich, K., Mehta, H., Patenaude, B., Ramakrishnan, C., Kalanithi, P., Etkin, A., Knutson, B., Glover, G. H., & Deisseroth, K. (2016). Prefrontal cortical regulation of brainwide circuit dynamics and reward-related behavior. *Science*, 351(6268). <https://doi.org/ARTN aac969810.1126/science.aac9698>

- Finn, E. S., Shen, X. L., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., Papademetris, X., & Constable, R. T. (2015). Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nature Neuroscience*, 18(11), 1664-1671. <https://doi.org/10.1038/nn.4135>
- Gabrieli, J. D. E., Ghosh, S. S., & Whitfield-Gabrieli, S. (2015). Prediction as a Humanitarian and Pragmatic Contribution from Human Cognitive Neuroscience. *Neuron*, 85(1), 11-26. <https://doi.org/10.1016/j.neuron.2014.10.047>
- Gao, S., Greene, A. S., Constable, R. T., & Scheinost, D. (2019). Combining multiple connectomes improves predictive modeling of phenotypic measures. *Neuroimage*, 201. <https://doi.org/10.1016/j.neuroimage.2019.116038>
- Gordon, E. M., Laumann, T. O., Gilmore, A. W., Newbold, D. J., Greene, D. J., Berg, J. J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H. X., Hampton, J. M., Coalson, R. S., Nguyen, A. L., McDermott, K. B., Shimony, J. S., Snyder, A. Z., Schlaggar, B. L., Petersen, S. E., Nelson, S. M., & Dosenbach, N. U. F. (2017). Precision Functional Mapping of Individual Human Brains. *Neuron*, 95(4), 791-+. <https://doi.org/10.1016/j.neuron.2017.07.011>
- Gratton, C., Laumann, T. O., Nielsen, A. N., Greene, D. J., Gordon, E. M., Gilmore, A. W., Nelson, S. M., Coalson, R. S., Snyder, A. Z., Schlaggar, B. L., Dosenbach, N. U. F., & Petersen, S. E. (2018). Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. *Neuron*, 98(2), 439-+. <https://doi.org/10.1016/j.neuron.2018.03.035>
- Hall, P., & Robinson, A. P. (2009). Reducing variability of crossvalidation for smoothing-parameter choice. *Biometrika*, 96(1), 175-186. <https://doi.org/10.1093/biomet/asn068>
- Haufe, S., Meinecke, F., Gorgen, K., Dahne, S., Haynes, J. D., Blankertz, B., & Biessmann, F. (2014). On the interpretation of weight vectors of linear models in multivariate neuroimaging. *Neuroimage*, 87, 96-110. <https://doi.org/10.1016/j.neuroimage.2013.10.067>
- He, Z. Y., & Yu, W. C. (2010). Stable feature selection for biomarker discovery. *Computational Biology And Chemistry*, 34(4), 215-225. <https://doi.org/10.1016/j.compbiolchem.2010.07.002>
- Jangraw, D. C., Gonzalez-Castillo, J., Handwerker, D. A., Ghane, M., Rosenberg, M. D., Panwar, P., & Bandettini, P. A. (2018). A functional connectivity-based neuromarker of sustained attention generalizes to predict recall in a reading task. *Neuroimage*, 166, 99-109. <https://doi.org/10.1016/j.neuroimage.2017.10.019>
- Jiang, R., Calhoun, V. D., Fan, L., Zuo, N., Jung, R., Qi, S., Lin, D., Li, J., Zhuo, C., Song, M., Fu, Z., Jiang, T., & Sui, J. (2019). Gender Differences in Connectome-based Predictions of Individualized Intelligence Quotient and Sub-domain Scores. *cerebral cortex*. <https://doi.org/10.1093/cercor/bhz134>
- Jiang, R. T., Calhoun, V. D., Zuo, N. M., Lin, D. D., Li, J., Fan, L. Z., Qi, S. L., Sun, H. L., Fu, Z. N., Song, M., Jiang, T. Z., & Sui, J. (2018). Connectome-based individualized prediction of temperament trait scores. *Neuroimage*, 183,

- 366-374. <https://doi.org/10.1016/j.neuroimage.2018.08.038>
- Jollans, L., Boyle, R., Artiges, E., Banaschewski, T., Desrivieres, S., Grigis, A., Martinot, J. L., Paus, T., Smolka, M. N., Walter, H., Schumann, G., Garavan, H., & Whelan, R. (2019). Quantifying performance of machine learning methods for neuroimaging data. *Neuroimage*, 199, 351-365. <https://doi.org/10.1016/j.neuroimage.2019.05.082>
- Kashyap, R., Kong, R., Bhattacharjee, S., Li, J. W., Zhou, J., & Yeo, B. T. T. (2019). Individual-specific fMRI-Subspaces improve functional connectivity prediction of behavior. *Neuroimage*, 189, 804-812. <https://doi.org/10.1016/j.neuroimage.2019.01.069>
- Kong, R., Li, J. W., Orban, C., Sabuncu, M. R., Liu, H. S., Schaefer, A., Sun, N. B., Zuo, X. N., Holmes, A. J., Eickhoff, S. B., & Yeo, B. T. T. (2019). Spatial Topography of Individual-Specific Cortical Networks Predicts Human Cognition, Personality, and Emotion. *cerebral cortex*, 29(6), 2533-2551. <https://doi.org/10.1093/cercor/bhy123>
- Li, J. W., Kong, R., Liegeois, R., Orban, C., Tan, Y. R., Sun, N. B., Holmes, A. J., Sabuncu, M. R., Ge, T., & Yeo, B. T. T. (2019). Global signal regression strengthens association between resting-state functional connectivity and behavior. *Neuroimage*, 196, 126-141. <https://doi.org/10.1016/j.neuroimage.2019.04.016>
- Liem, F., Varoquaux, G., Kynast, J., Beyer, F., Masouleh, S. K., Huntenburg, J. M., Lampe, L., Rahim, M., Abraham, A., Craddock, R. C., Riedel-Heller, S., Luck, T., Loeffler, M., Schroeter, M. L., Witte, A. V., Villringer, A., & Margulies, D. S. (2017). Predicting brain-age from multimodal imaging data captures cognitive impairment. *Neuroimage*, 148, 179-188. <https://doi.org/10.1016/j.neuroimage.2016.11.005>
- Marchitelli, R., Minati, L., Marizzoni, M., Bosch, B., Bartres-Faz, D., Muller, B. W., Wiltfang, J., Fiedler, U., Roccatagliata, L., Picco, A., Nobili, F., Blin, O., Bombois, S., Lopes, R., Bordet, R., Sein, J., Ranjeva, J. P., Didic, M., Gros-Dagnac, H., Payoux, P., Zoccatelli, G., Alessandrini, F., Beltramello, A., Bargallo, N., Ferretti, A., Caulo, M., Aiello, M., Cavaliere, C., Soricelli, A., Parnetti, L., Tarducci, R., Floridi, P., Tsolaki, M., Constantinidis, M., Drevelegas, A., Rossini, P. M., Marra, C., Schonknecht, P., Hensch, T., Hoffmann, K. T., Kuijper, J. P., Visser, P. J., Barkhof, F., Frisoni, G. B., Jovicich, J., & Consortium, P. (2016). Test-Retest Reliability of the Default Mode Network in a Multi-Centric fMRI Study of Healthy Elderly: Effects of Data-Driven Physiological Noise Correction Techniques. *Human Brain Mapping*, 37(6), 2114-2132. <https://doi.org/10.1002/hbm.23157>
- Meng, X., Jiang, R. T., Lin, D. D., Bustillo, J., Jones, T., Chen, J. Y., Yu, Q. B., Du, Y. H., Zhang, Y., Jiang, T. Z., Sui, J., & Calhoun, V. D. (2017). Predicting individualized clinical measures by a generalized prediction framework and multimodal fusion of MRI data. *Neuroimage*, 145, 218-229. <https://doi.org/10.1016/j.neuroimage.2016.05.026>
- Meskaldji, D. E., Preti, M. G., Bolton, T. A. W., Montandon, M. L., Rodriguez, C.,

- Morgenthaler, S., Giannakopoulos, P., Haller, S., & Van De Ville, D. (2016). Prediction of long-term memory scores in MCI based on resting-state fMRI. *Neuroimage-Clinical*, 12, 785-795. <https://doi.org/10.1016/j.nicl.2016.10.004>
- Mueller, S., Wang, D. H., Fox, M. D., Pan, R. Q., Lu, J., Li, K. C., Sun, W., Buckner, R. L., & Liu, H. S. (2015). Reliability correction for functional connectivity: Theory and implementation. *Human Brain Mapping*, 36(11), 4664-4680. <https://doi.org/10.1002/hbm.22947>
- Mueller, S., Wang, D. H., Fox, M. D., Yeo, B. T. T., Sepulcre, J., Sabuncu, M. R., Shafee, R., Lu, J., & Liu, H. S. (2013). Individual Variability in Functional Connectivity Architecture of the Human Brain. *Neuron*, 77(3), 586-595. <https://doi.org/10.1016/j.neuron.2012.12.028>
- Nielsen, A. N., Greene, D. J., Gratton, C., Dosenbach, N. U. F., Petersen, S. E., & Schlaggar, B. L. (2019). Evaluating the Prediction of Brain Maturity From Functional Connectivity After Motion Artifact Denoising. *cerebral cortex*, 29(6), 2455-2469. <https://doi.org/10.1093/cercor/bhy117>
- Noble, S., Scheinost, D., Finn, E. S., Shen, X. L., Papademetris, X., McEwen, S. C., Bearden, C. E., Addington, J., Goodyear, B., Cadenhead, K. S., Mirzakhania, H., Cornblatt, B. A., Olvet, D. M., Mathalon, D. H., McGlashan, T. H., Perkins, D. O., Belger, A., Seidman, L. J., Thermenos, H., Tsuang, M. T., van Erp, T. G. M., Walker, E. F., Hamann, S., Woods, S. W., Cannon, T. D., & Constable, R. T. (2017). Multisite reliability of MR-based functional connectivity. *Neuroimage*, 146, 959-970. <https://doi.org/10.1016/j.neuroimage.2016.10.020>
- Noble, S., Spann, M. N., Tokoglu, F., Shen, X. L., Constable, R. T., & Scheinost, D. (2017). Influences on the Test-Retest Reliability of Functional Connectivity MRI and its Relationship with Behavioral Utility. *cerebral cortex*, 27(11), 5415-5429. <https://doi.org/10.1093/cercor/bhx230>
- Petersen, S. E., & Sporns, O. (2015). Brain Networks and Cognitive Architectures. *Neuron*, 88(1), 207-219. <https://doi.org/10.1016/j.neuron.2015.09.027>
- Poldrack, R. A., & Farah, M. J. (2015). Progress and challenges in probing the human brain. *Nature*, 526(7573), 371-379. <https://doi.org/10.1038/nature15692>
- Qin, J., Shen, H., Zeng, L. L., Gao, K., Luo, Z., & Hu, D. (2019). Dissociating individual connectome traits using low-rank learning. *Brain Res*, 1722, 146348. <https://doi.org/10.1016/j.brainres.2019.146348>
- Rashid, B., Damaraju, E., Pearlson, G. D., & Calhoun, V. D. (2014). Dynamic connectivity states estimated from resting fMRI Identify differences among Schizophrenia, bipolar disorder, and healthy control subjects. *Frontiers In Human Neuroscience*, 8. <https://doi.org/10.3389/fnhum.2014.00897>
- Rizk-Jackson, A., Stoffers, D., Sheldon, S., Kuperman, J., Dale, A., Goldstein, J., Corey-Bloom, J., Poldrack, R. A., & Aron, A. R. (2011). Evaluating imaging biomarkers for neurodegeneration in pre-symptomatic Huntington's disease using machine learning techniques. *Neuroimage*, 56(2), 788-796. <https://doi.org/10.1016/j.neuroimage.2010.04.273>
- Rosenberg, M. D., Finn, E. S., Scheinost, D., Constable, R. T., & Chun, M. M. (2017). Characterizing Attention with Predictive Network Models. *Trends In*

- Cognitive Sciences, 21(4), 290-302. <https://doi.org/10.1016/j.tics.2017.01.011>
- Rospleszcz, S., Janitza, S., & Boulesteix, A. L. (2016). Categorical variables with many categories are preferentially selected in bootstrap-based model selection procedures for multivariable regression models. *Biometrical Journal*, 58(3), 652-673. <https://doi.org/10.1002/bimj.201400185>
- Rousselet, G. A., & Pernet, C. R. (2012). Improving standards in brain-behavior correlation analyses. *Frontiers In Human Neuroscience*, 6. <https://doi.org/Artn 11910.3389/Fnhum.2012.00119>
- Ryali, S., Chen, T. W., Supekar, K., & Menon, V. (2012). Estimation of functional connectivity in fMRI data using stability selection-based sparse partial correlation with elastic net penalty. *Neuroimage*, 59(4), 3852-3861. <https://doi.org/10.1016/j.neuroimage.2011.11.054>
- Santaracchi, E., Galli, G., Polizzotto, N. R., Rossi, A., & Rossi, S. (2014). Efficiency of Weak Brain Connections Support General Cognitive Functioning. *Human Brain Mapping*, 35(9), 4566-4582. <https://doi.org/10.1002/hbm.22495>
- Shen, X. L., Finn, E. S., Scheinost, D., Rosenberg, M. D., Chun, M. M., Papademetris, X., & Constable, R. T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. *Nature Protocols*, 12(3), 506-518. <https://doi.org/10.1038/nprot.2016.178>
- Smith, S. M., Vidaurre, D., Beckmann, C. F., Glasser, M. F., Jenkinson, M., Miller, K. L., Nichols, T. E., Robinson, E. C., Salimi-Khorshidi, G., Woolrich, M. W., Barch, D. M., Ugurbil, K., & Van Essen, D. C. (2013). Functional connectomics from resting-state fMRI. *Trends In Cognitive Sciences*, 17(12), 666-682. <https://doi.org/10.1016/j.tics.2013.09.016>
- Smola, A. J., & Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics And Computing*, 14(3), 199-222. <https://doi.org/Doi 10.1023/B:Stco.0000035301.49549.88>
- Sripada, C., Rutherford, S., Angstadt, M., Thompson, W. K., Luciana, M., Weigard, A., Hyde, L. H., & Heitzeg, M. (2019). Prediction of neurocognition in youth from resting state fMRI. *molecular psychiatry*, 1-9. <https://doi.org/10.1038/s41380-019-0481-6>
- Stevens, M. C. (2009). The developmental cognitive neuroscience of functional connectivity. *Brain And Cognition*, 70(1), 1-12. <https://doi.org/10.1016/j.bandc.2008.12.009>
- Strobl, C., Boulesteix, A. L., Zeileis, A., & Hothorn, T. (2007). Bias in random forest variable importance measures: Illustrations, sources and a solution. *Bmc Bioinformatics*, 8. <https://doi.org/Artn 2510.1186/1471-2105-8-25>
- Sun, J. Z., Liu, Z. W., Rolls, E. T., Chen, Q. L., Yao, Y., Yang, W. J., Wei, D. T., Zhang, Q. L., Zhang, J., Feng, J. F., & Qiu, J. (2019). Verbal Creativity Correlates with the Temporal Variability of Brain Networks During the Resting State. *cerebral cortex*, 29(3), 1047-1058. <https://doi.org/10.1093/cercor/bhy010>
- Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal Of the Royal Statistical Society Series B-Methodological*, 58(1), 267-288.
- Toivainen, P., Alluri, V., Brattico, E., Wallentin, M., & Vuust, P. (2014). Capturing the

- musical brain with Lasso: Dynamic decoding of musical features from fMRI data. *Neuroimage*, 88, 170-180. <https://doi.org/10.1016/j.neuroimage.2013.11.017>
- Tu, Y., Tan, A., Fu, Z., Hung, Y. S., Hu, L., & Zhang, Z. (2015). Supervised nonlinear dimension reduction of functional magnetic resonance imaging data using Sliced Inverse Regression. Paper presented at the Engineering in Medicine & Biology Society.
- Van Essen, D. C., Smith, S. M., Barch, D. M., Behrens, T. E. J., Yacoub, E., Ugurbil, K., & Consortium, W.-M. H. (2013). The WU-Minn Human Connectome Project: An overview. *Neuroimage*, 80, 62-79. <https://doi.org/10.1016/j.neuroimage.2013.05.041>
- Wilcox, R. (2004). Inferences Based on a Skipped Correlation Coefficient. *Journal Of Applied Statistics*, 31(2), 131-143. <https://doi.org/10.1080/0266476032000148821>
- Wilcox, R. R. (2005). *Introduction to Robust Estimation and Hypothesis Testing*.
- Xiao, L., Stephen, J. M., Wilson, T. W., Calhoun, V. D., & Wang, Y. P. (2019). Alternating Diffusion Map Based Fusion of Multimodal Brain Connectivity Networks for IQ Prediction. *Ieee Transactions on Biomedical Engineering*, 66(8), 2140-2151. <https://doi.org/10.1109/Tbme.2018.2884129>
- Yip, S. W., Scheinost, D., Potenza, M. N., & Carroll, K. M. (2019). Connectome-Based Prediction of Cocaine Abstinence. *American Journal Of Psychiatry*, 176(2), 156-164. <https://doi.org/10.1176/appi.ajp.2018.17101147>
- Yoo, K., Rosenberg, M. D., Hsu, W. T., Zhang, S., Li, C. S. R., Scheinost, D., Constable, R. T., & Chun, M. M. (2018). Connectome-based predictive modeling of attention: Comparing different functional connectivity features and prediction methods across datasets. *Neuroimage*, 167, 11-22. <https://doi.org/10.1016/j.neuroimage.2017.11.010>
- Zhang, C. H., & Huang, J. (2008). The sparsity and bias of the lasso selection in high-dimensional linear regression. *Annals Of Statistics*, 36(4), 1567-1594. <https://doi.org/10.1214/07-AOS520>
- Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net (vol B 67, pg 301, 2005). *Journal Of the Royal Statistical Society Series B-Statistical Methodology*, 67, 768-768. <https://doi.org/DOI.10.1111/j.1467-9868.2005.00527.x>

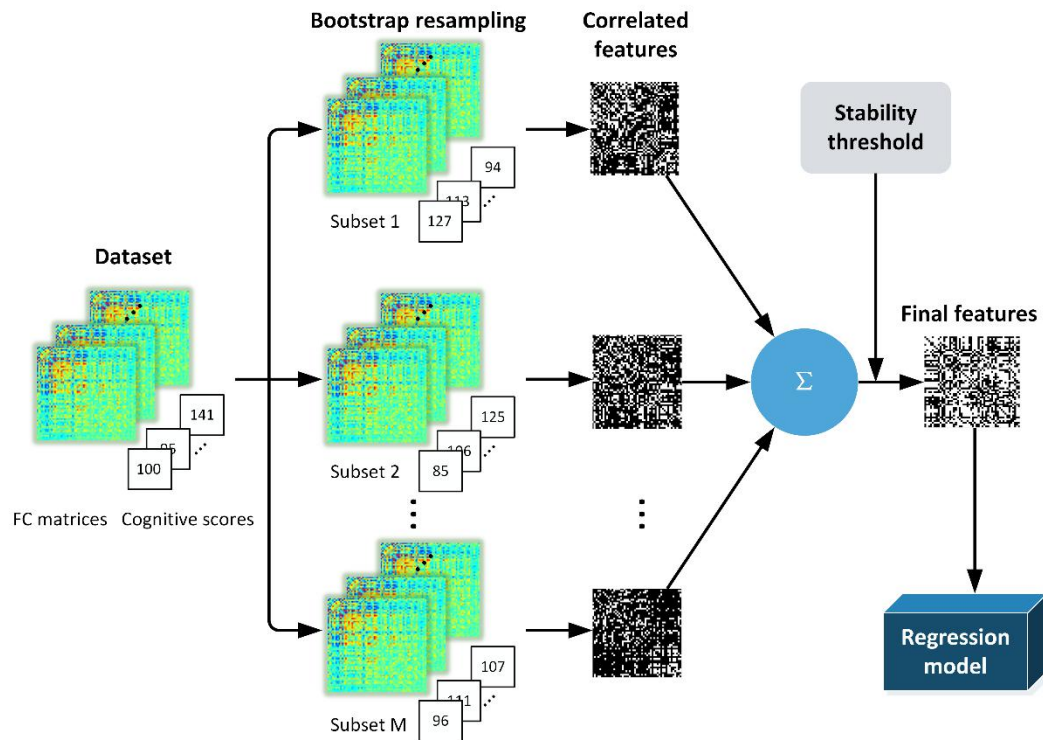


Figure 1. The pipeline of bootstrapping framework.

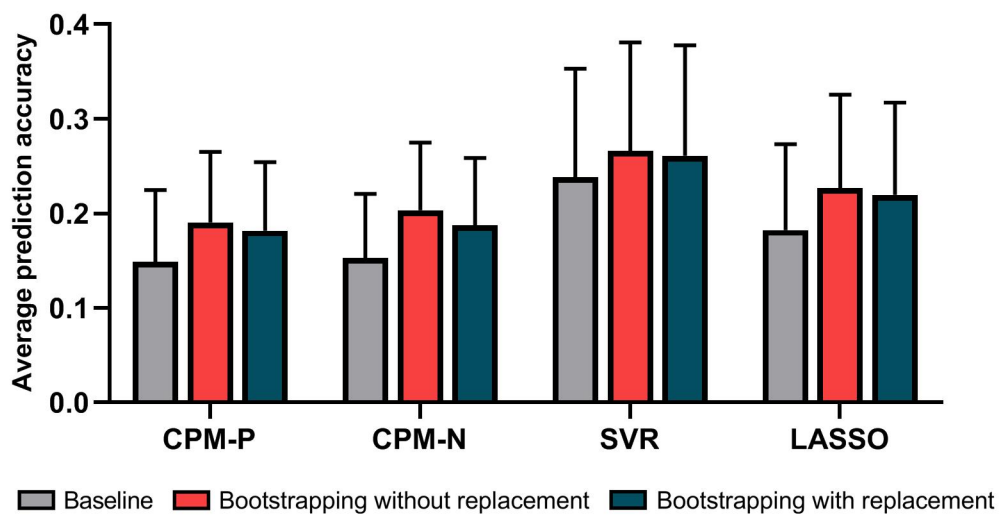


Figure 2. The average best prediction accuracies of bootstrapping methods and baseline method with CPM-P, CPM-N, SVR and LASSO model.

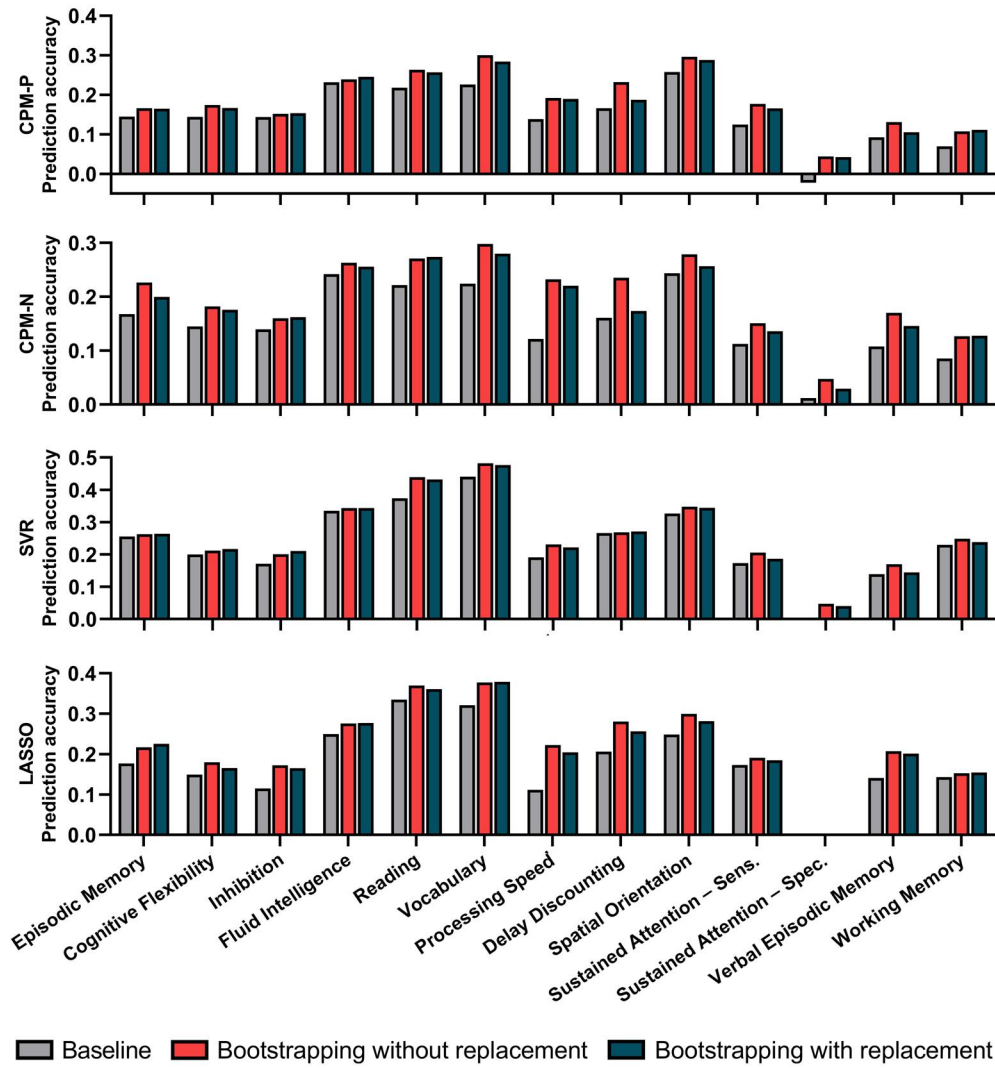


Figure 3. The best prediction performances for 13 cognitive measures with CPM-P, CPM-N, SVR and LASSO.

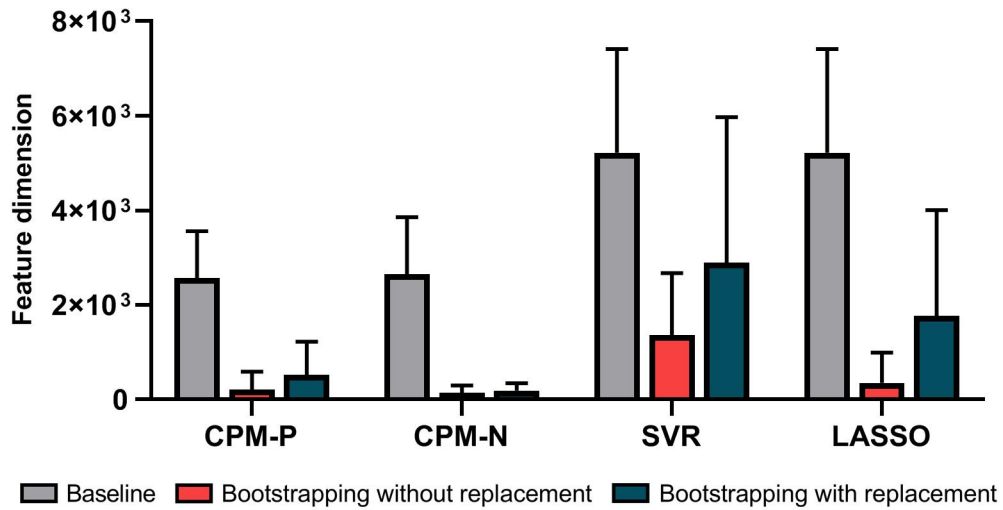


Figure 4. The comparison of feature dimensions between baseline and bootstrapping methods.

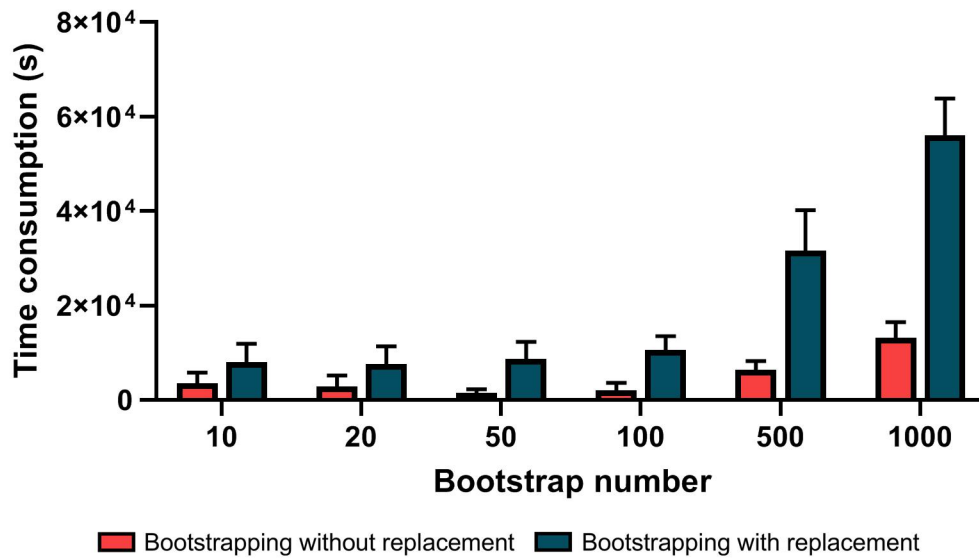


Figure 5. The average time consumption in bootstrapping methods for 13 cognitive traits.

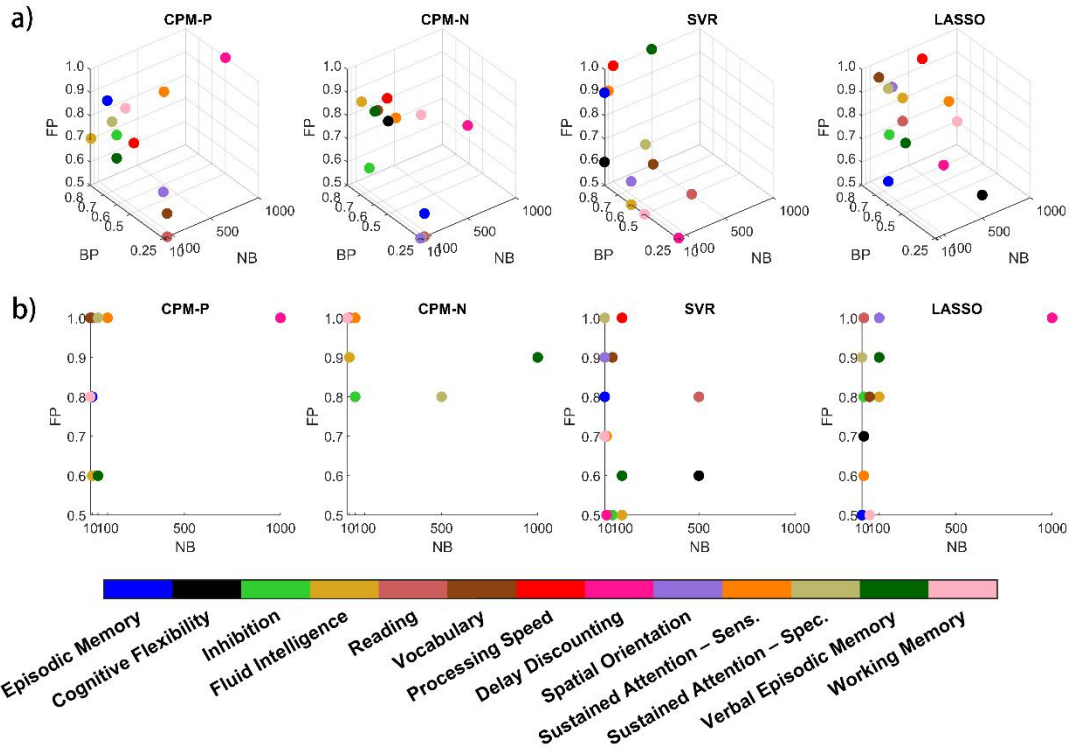


Figure 6. The optimal parameter setting for each cognitive measure in CPM-P, CPM-N, SVR and LASSO when using **a)** bootstrapping without replacement and **b)** bootstrapping with replacement. The colored dots denoted different types of cognitive scores.

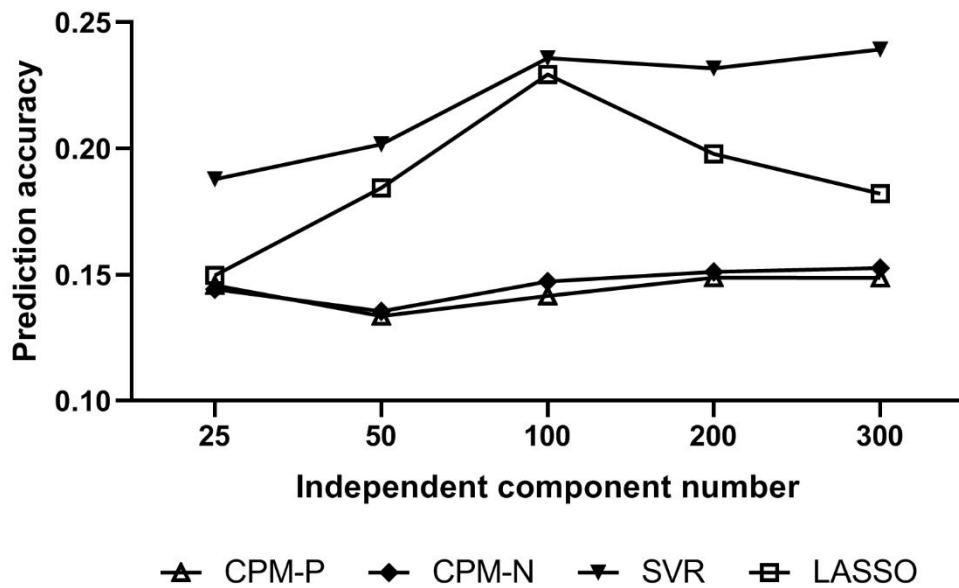


Figure 7. The prediction accuracy of different number of independent components.

The result of 15 parcels was not shown since no significant features were selected for some cognitive scores.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

DATA AVAILABILITY STATEMENT

Data were provided by the Human Connectome Project and WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University. Code for this work is freely available at the GitHub repository (<https://github.com/Lijiang-Wei/bootstrapping.git>).