Received: 8 October 2019 | Revised: 10 January 2020 | Accepted: 2 February 2020
DOI: 10.1002/hbm.24947

RESEARCH ARTICLE

WILEY

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

Lijiang Wei[1] | Bin Jing[1] | Haiyun Li[1,2]

[1]School of Biomedical Engineering, Capital Medical University, Beijing, China

[2]Beijing Key Laboratory of Fundamental Research on Biomechanics in Clinical Application, Capital Medical University, Beijing, China

Correspondence
Haiyun Li, School of Biomedical Engineering, Capital Medical University, Beijing 100069 China.
Email: haiyunli@ccmu.edu.cn

Funding information
Beijing Municipal Commission of Education, Grant/Award Number: KM202010025025; Natural Science Foundation of Beijing Municipality, Grant/Award Number: L192044

Abstract

Resting-state functional connectivity (RSFC) records enormous functional interaction information between any pair of brain nodes, which enriches the individual-phenotypic prediction. To reduce high-dimensional features, correlation analysis is a common way for feature selection. However, resting state fMRI signal exhibits typically low signal-to-noise ratio and the correlation analysis is sensitive to outliers and data distribution, which may bring unstable features to prediction. To alleviate this problem, a bootstrapping-based feature selection framework was proposed and applied to connectome-based predictive modeling, support vector regression, least absolute shrinkage and selection operator, and Ridge regression to predict a series of cognitive traits based on Human Connectome Project data. To systematically investigate the influences of different parameter settings on the bootstrapping-based framework, 216 parameter combinations were evaluated and the best performance among them was identified as the final prediction result for each cognitive trait. By using the bootstrapping methods, the best prediction performances outperformed the baseline method in all four prediction models. Furthermore, the proposed framework could effectively reduce the feature dimension by retaining the more stable features. The results demonstrate that the proposed framework 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, cognitive trait, feature selection, functional connectivity, individualized prediction

1 | INTRODUCTION

Resting-state functional connectivity (RSFC) has been found to be associated with pathological conditions and physiological status of the human brain (Cribben, Haraldsdottir, Atlas, Wager, & Lindquist, 2012; Elanbari et al., 2014; Finn et al., 2015; Meskaldji et al., 2016), which therefore could be used to understand individual differences in cognition and behavior (Jiang et al., 2018). Previous studies have reported that FC patterns could be used to predict individual behavioral and cognitive phenotypes, such as attention ability (Yoo et al., 2018), verbal creativity (Sun et al., 2019), intelligence ability (Finn et al., 2015; Jiang et al., 2019), and chronological age (Dosenbach et al., 2010; Liem et al., 2017). To model the RSFC-phenotype association, four regression algorithms including connectome-based predictive

Lijiang Wei and Bin Jing have contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.
© 2020 The Authors. *Human Brain Mapping* published by Wiley Periodicals, Inc.

modeling (CPM), support vector regression (SVR), least absolute shrinkage and selection operator (LASSO), and Ridge regression have been frequently adopted for their good performances and interpretabilities (Coloigner, Phlypo, Bush, Lepore, & Wood, 2016; Cui & Gong, 2018; Dadi et al., 2019; de Vos et al., 2018; Gao, Greene, Constable, & Scheinost, 2019; Jiang et al., 2019; Meng et al., 2017; Ryali, Chen, Supekar, & Menon, 2012; Shen et al., 2017; Toiviainen, Alluri, Brattico, Wallentin, & Vuust, 2014). CPM is in fact a linear regression model and has been successfully applied in predicting intelligence (Beaty et al., 2018; Finn et al., 2015; Jiang et al., 2019), attention (Rosenberg, Finn, Scheinost, Constable, & Chun, 2017), cocaine abstinence (Yip, Scheinost, Potenza, & Carroll, 2019), and reading comprehension (Jangraw et al., 2018). SVR implements space transformation by some kernel function in order to achieve better prediction (Basak, Pal, & Patranabis, 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), temperament metrics (Jiang et al., 2018), and intelligence quotient (Jiang et al., 2019). Ridge regression adds an L2 penalization on the object function to shrink the regression coefficients, and shows good prediction performances in age (Khosla, Jamison, Kuceyeski, & Sabuncu, 2019), behavioral measures, and cognitive traits (Cui & Gong, 2018; Gao et al., 2019).

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, Plis, Sui, & Calhoun, 2017; Duangsoithong & Windeatt, 2010), and increase the computational complexity. For regression tasks, correlation analysis between RSFC and the target phenotypic measure (e.g., cognitive trait) is the commonest approach to select effective features (Dadi et al., 2019; Gabrieli, Ghosh, & Whitfield-Gabrieli, 2015). However, the correlation analysis may not be able to catch the reliable correlational relationship since resting state fMRI (rs-fMRI) data exhibit low signal-to-noise ratio (SNR) and the correlation analysis is sensitive to outliers (Wilcox, 2004, 2005). Small perturbation in the data may lead to inclusion of unrelated features or exclusion of useful features, thus the weak robustness of correlation analysis can introduce false correlations including Type I error and power problem (Rousselet & Pernet, 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 & Pernet, 2012). All these conditions may cause that the features detected from the whole dataset do not significantly correlate with the target phenotypic measure in a subset, resulting in unstable features, which may in fact contribute little to the target phenotypic 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 & Tibshirani, 1997; Hall & Robinson, 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 a 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), the bootstrapping was used as a backward feature selection method for cardiac ventricular fibrillation discrimination. Ditzler reported that the bootstrapping combined with Neyman–Pearson hypothesis test successfully detected the statistically relevant features on both synthetic and real data (Ditzler, Polikar, & Rosen, 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 the 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 the bootstrapping influenced the final prediction performance.

This study explored a bootstrapping-based feature selection framework to enhance the RSFC prediction performances in CPM, SVR, LASSO, and Ridge models, and different parameter settings in bootstrapping were 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 hypothesized that bootstrapping could improve the RSFC prediction performances of CPM, SVR, LASSO, and Ridge models.

2 | MATERIALS AND METHODS

2.1 | Resting-state fMRI data

The data came from public WU-Human Connectome Project (HCP) 1,200 release of 1,206 healthy subjects (age 22–37). All imaging data were collected on a customized 3T ConnectomeScanner adapted from a Siemens Skyra (Siemens, Erlangen, Germany) scanner using a standard 32-channel head coil. Rs-fMRI data were acquired with a multiband pulse sequence with the following scanning parameters: time repetition (TR) = 720 ms, time echo = 33.1 ms, field of view = 208 × 180 mm^2, flip angle = 52°, and voxel size = 2.0 mm isotropic cube (Smith et al., 2013). Two scanning sessions (REST1 and REST2) of high spatial–temporal resolution rs-fMRI data were

acquired during two consecutive days. Both sessions were scanned at right-to-left (run1) and left-to-right (run2) phase encoding direction, and the functional images comprised 1,200 volumes in each run. The current study adopted the 1,003 subjects with complete rs-fMRI runs (a total of 4,800 time points). Note that the HCP database has two types of reconstruction data: 191 subjects were reconstructed utilizing an original reconstruction algorithm (recon1) that performed separation of the multiband multi-slice images after transforming the acquired fully sampled data to frequency space along the readout direction, while the other 812 subjects applied an upgraded algorithm (recon2) that performed the separation in k-space. To avoid possible influences caused by reconstruction methods and validate the generalization of the proposed bootstrapping methods, we used 812 (recon2) subjects as a discovery dataset, and the remaining 191 (recon1) subjects as an independent validation dataset. For more details of inclusion and exclusion criteria on the HCP 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 four runs were concatenated to form 4,800 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 Gaussianized into Z-stats. Here, we adopted 300 ICA components as network nodes, resulting in a connectivity matrix with size 300×300 for each participant.

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 (Sripada et al., 2019; Stevens, 2009). Previous studies had shown that RSFC patterns were closely associated with cognitive traits (Santarnecchi, Galli, Polizzotto, Rossi, & Rossi, 2014; Sun et al., 2019). In this study, 13 measures of cognitive tests were chosen as the prediction targets that provided by HCP (Kong et al., 2019; 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 (Table S1). Seven of them measured by NIH (National Institutes of Health) Toolbox were normalized to age-adjusted scores with mean of 100 and *SD* of 15. Furthermore, some subjects with the missing cognitive measures were excluded in corresponding trait prediction analysis (see Table S1 for missing subjects).

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 be more suitable to the finite population. In this study, both bootstrapping with and without replacement (Figure 1) were investigated, and 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 an original dataset, so each subject in the resampling dataset was unique without duplication. For each cognitive measure, the bootstrapping without replacement was applied on the dataset 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), for example, BP = 50% indicated that half of all subjects were chosen from the 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 features whose FP more than 50% would be selected into the final feature set, otherwise it would be filtered as weak predictors.

2.4.2 | Bootstrapping with replacement

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

2.5 | Regression models

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

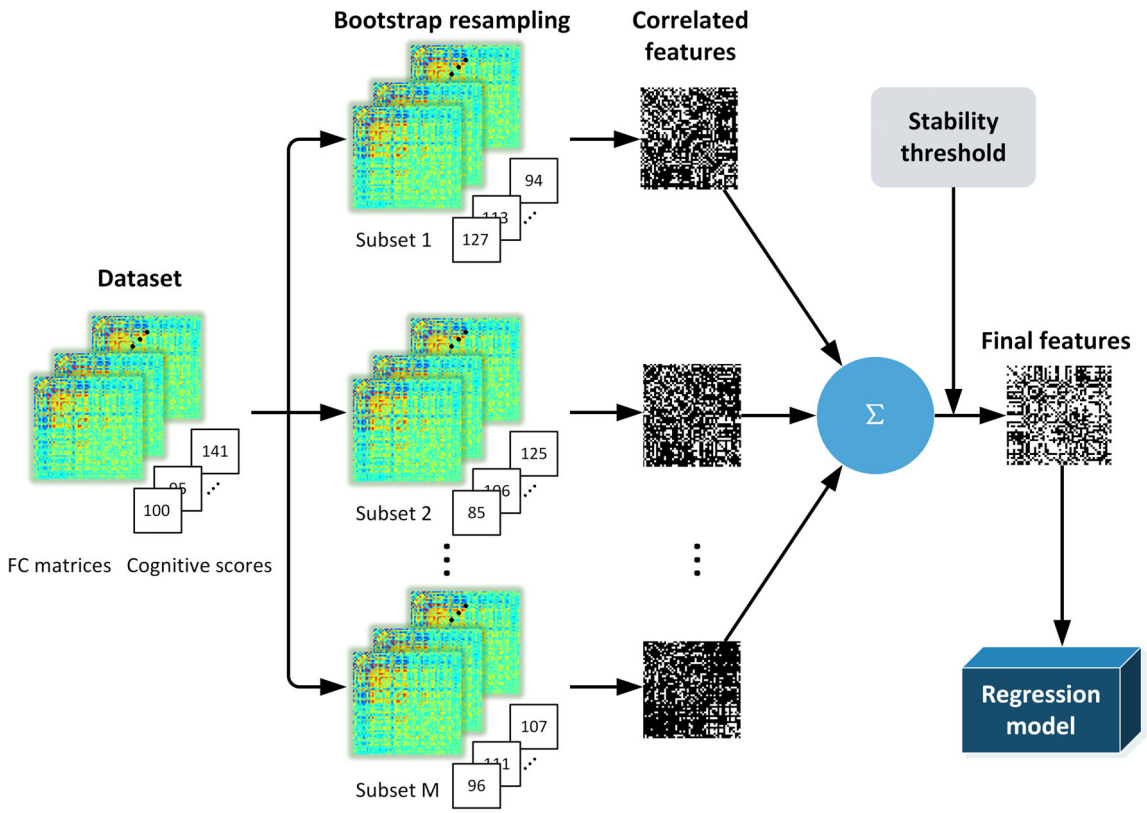

FIGURE 1 The pipeline of bootstrapping framework

2.5.1 | Connectome-based predictive modeling

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 | Support vector regression

SVR was based on statistical learning theory and fitted linear model with Vapink's ε-sensitive loss function (Smola & Scholkopf, 2004). It tried to solve an objective function $f(x)$ whose predicted values of training data deviated by no more than ε from the 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 $((x_1, y_1), \ldots, (x_N, y_N))$, the object function took the following form:

$$\min_{\beta,\xi_i,\hat{\xi}_i} \frac{1}{2}\|\beta\|^2 + C\sum_{i=1}^{m}\left(\xi_i + \hat{\xi}_i\right)$$

$$\text{s.t.} f(x_i) - y_i \le \varepsilon + \xi_i,$$

$$y_i - f(x_i) \le \varepsilon + \hat{\xi}_i,$$

$$\xi_i, \hat{\xi}_i \ge 0$$

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 object function was solved to obtain a weight for each support vector. In this study, the linear kernel was selected for SVR.

2.5.3 | Least absolute shrinkage and selection operator

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 & Hastie, 2005), thus generating robust coefficients as well as parsimonious models. This algorithm minimized the ordinary least squares plus the sum of absolute values of regression coefficients β to obtain the estimated coefficients. The object function was formed as below:

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

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

2.5.4 | Ridge

Ridge applied L2-norm regularization on ordinary least squares, which had the ability to shrink regression coefficients. The model minimized ordinary least squares and the L2-norm regularization, and the object function could be formulized as below:

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

where λ controlled the trade-off between accuracy and shrinkage strength.

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

2.6 | Prediction framework

For each cognitive trait, a 10-fold cross-validation (CV) framework was utilized in the discovery dataset. All individuals in the discovery dataset were sorted according to values of cognitive measure and then divided into 10 folds. To be specific, the (1st, 11th, 21st,…) subjects were assigned to the 1st fold, the (2nd, 12th, 22nd,…) subjects were assigned to the 2nd fold, and so on. This partition procedure ensured the same distribution of the 10 folds and avoided random bias and expensive computation due to random splitting (Cui & Gong, 2018). Within the prediction framework, each fold was iteratively used for testing while the remaining ninefolds were used for training. The bootstrapping feature selection methods were performed on the training set to select features for CPM, SVR, LASSO, and Ridge models (Figure 1). Significantly positive-, negative-, and combined-correlated FCs were allotted into positive, negative, and combined feature set (significant threshold p = .05). Then, the positive feature set was adopted for CPM-P, the negative feature set for CPM-N, and the combined feature set for SVR, LASSO, and Ridge. In both bootstrapping method with and without replacement, a broad sets of bootstrapping parameters including NB, BP, and FP were tested, where NB was chosen from [10, 20, 50, 100, 500, 1,000] 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, LASSO and Ridge, inner fivefold CVs were employed to determine their optimal model parameters (i.e., C and λ). Parameter C of SVR ranged from [2^{-5}, 2^{-4},…, 2^{10}] and parameter λ of LASSO and Ridge ranged from [2^{-10}, 2^{-9},…, 2^{5}] (Cui & Gong, 2018). The model performance was evaluated by the correlation value (R) and the mean square error (MSE) between predicted and actual cognitive values. In our study, the reported correlation R and MSE were averaged over 10 testing folds. Among all candidate parameter combinations, the prediction result with the largest R was served as the optimal bootstrapping parameter setting.

2.7 | Feature dimension reduction

To illustrate the feature dimension reduction with bootstrapping-based feature selection methods, the number of features selected by baseline method and bootstrapping methods was respectively compared. For each cognitive trait, the feature dimension was obtained by counting the feature number on each fold and then averaged across 10 folds. The final feature dimension was averaged across 13 cognitive traits to gain the mean feature dimension. Particularly, for the bootstrapping methods, only the feature dimension derived from the optimal bootstrapping parameter settings would be considered.

2.8 | Computational cost

Since the bootstrapping methods required multiple resampling on the original dataset, computational cost would be a noticeable concern especially for a large dataset. Among the bootstrapping parameters, concerning NB was the main source of computational cost, the running time of resampling and correlation test under each NB was quantified for bootstrapping methods. Because only the bootstrapping without replacement had BP, the running time in bootstrapping without replacement was recorded for each BP and then averaged under different NB. 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.

2.9 | Evaluation in the validation dataset

In order to test the generalization of predictive models obtained from the discovery dataset, we applied the predictive models acquired from

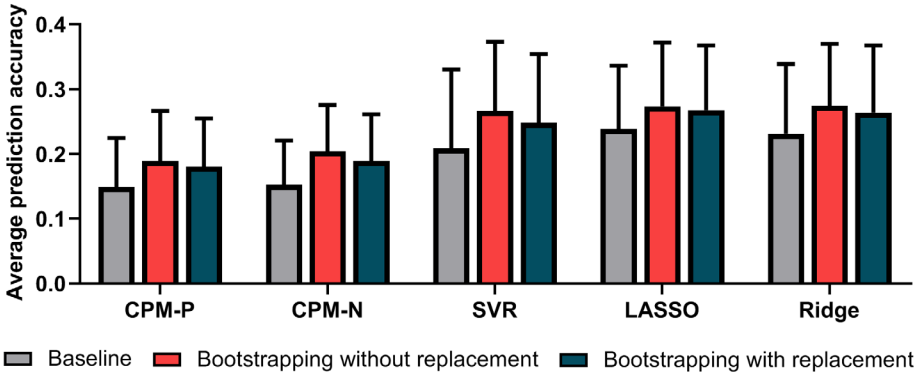

FIGURE 2 The mean and standard error of correlation *R* across 13 cognitive trait predictions. For bootstrapping methods, only the results derived from optimal parameter settings were plotted. Bootstrapping without replacement had the highest prediction accuracy. See Figure S1 for the corresponding mean square error (*MSE*)

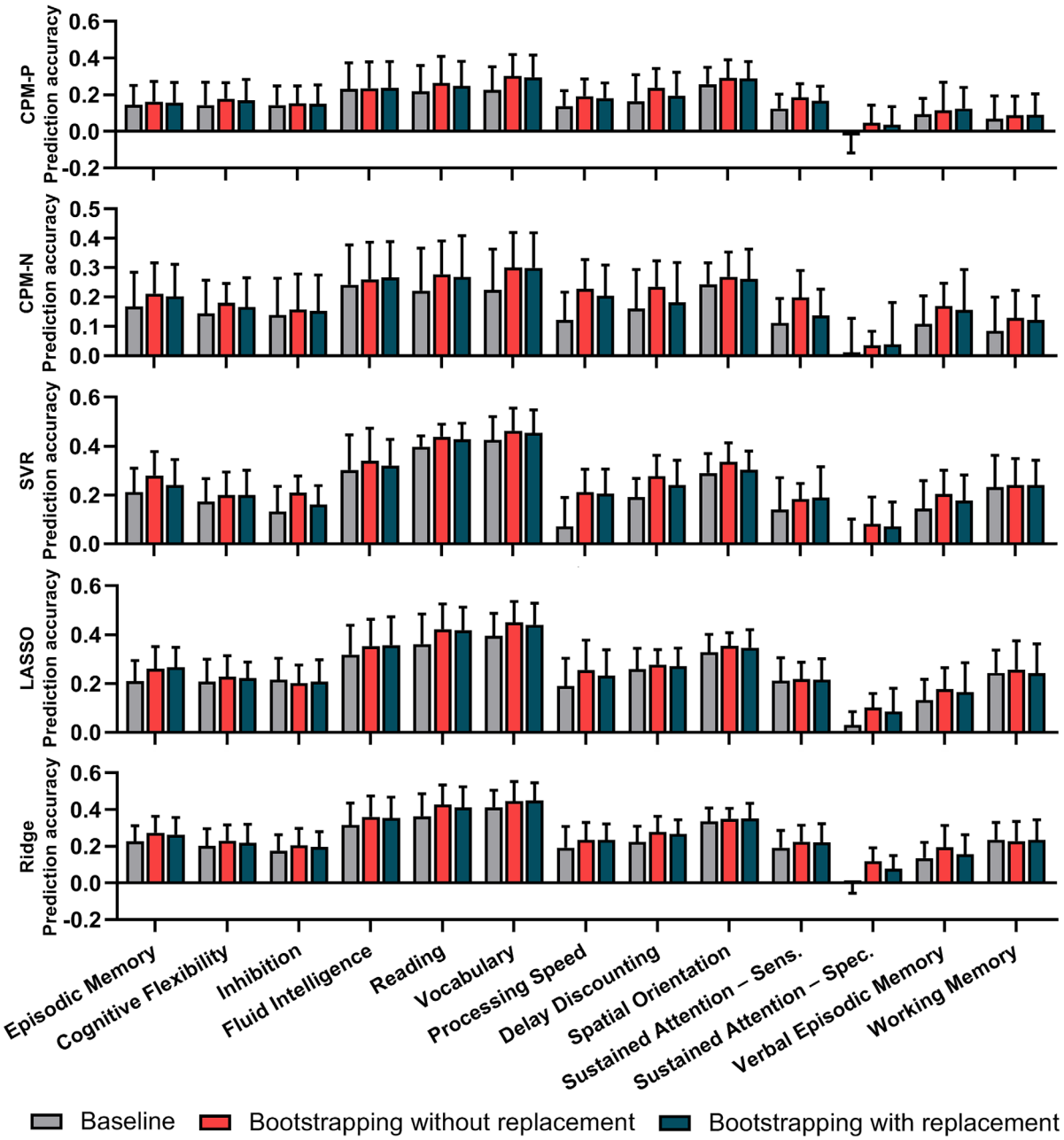

FIGURE 3 The correlation *R* of 13 cognitive trait predictions. The "bars" showed the mean and standard error of prediction correlation *R* across 10-fold cross-validation

the discovery dataset to predict subjects in the validation dataset. Notably, the predictive model was obtained with the whole discovery dataset, and features were selected by bootstrapping without replacement and baseline method, respectively. The reason we only used bootstrapping without replacement was that it displayed better prediction performance and faster computation than bootstrapping with replacement in the discovery dataset. Moreover, to simplify the bootstrapping parameter settings in the prediction model, only eight parameter combinations were explored based on the obtained optimal parameter settings in previous section: NB [50, 100], BP [0.7, 0.8], and FP [0.8, 0.9]. Likewise, the prediction performance was assessed by the correlation *R* and *MSE* between predicted and actual cognitive values.

3 | RESULTS

3.1 | Bootstrapping improved prediction of cognitive traits

Figure 2 summarized the mean correlation *R* (across 13 cognitive traits) of each model using the baseline method and bootstrapping method in discovery dataset, and only results derived from the optimal parameters were plotted for bootstrapping methods. All models using the refined feature sets derived from the optimal parameters outperformed the baseline method. Compared with the baseline method (mean correlation *R* = 0.15, 0.15, 0.21, 0.24, and 0.23 for CPM-P, CPM-N, SVR, LASSO, and Ridge, respectively), the boo-

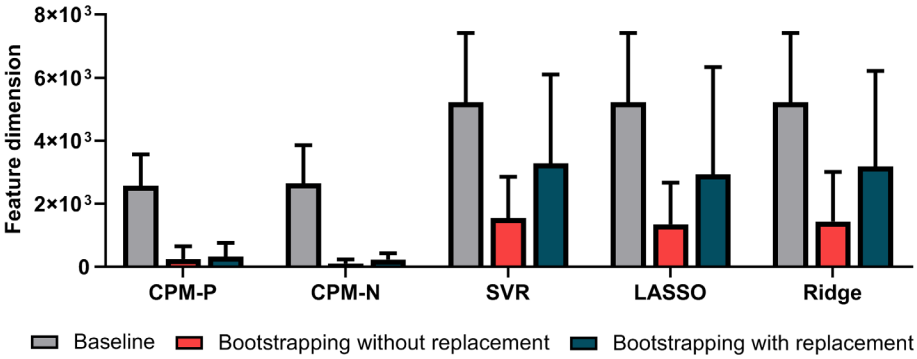

FIGURE 4 The comparison of mean feature dimension across 13 cognitive traits, where the feature dimension of bootstrapping methods were obtained under optimal parameter settings

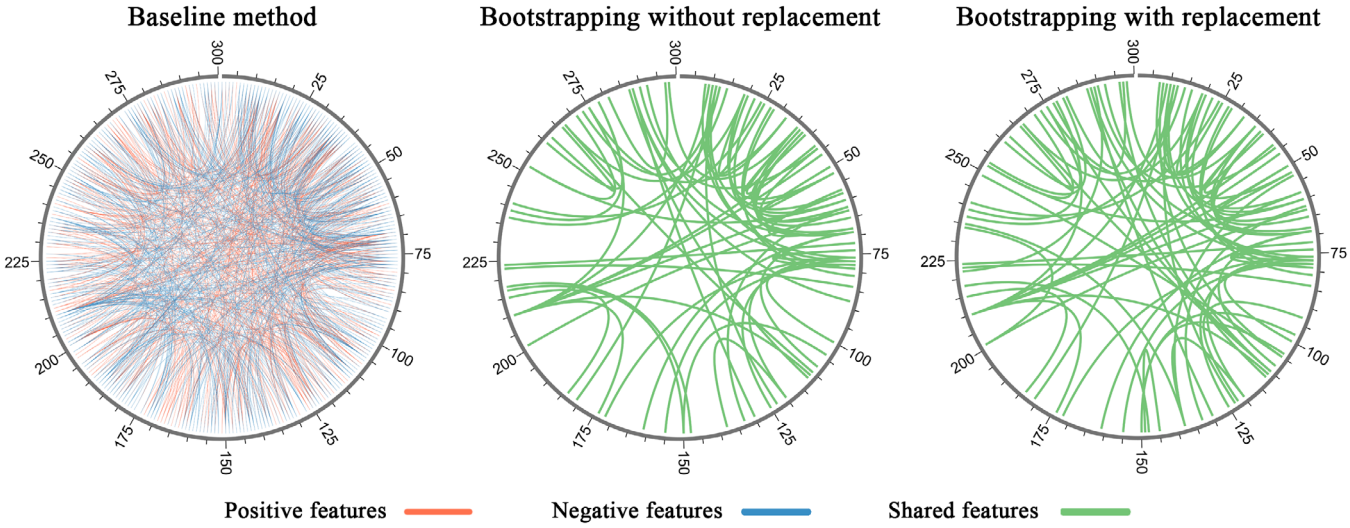

FIGURE 5 Resting-state functional connectivity (RSFC) features of the Working Memory selected by the baseline method and the bootstrapping methods in all 10-fold cross-validation. Features selected by the bootstrapping methods were derived from NB = 100, BP = 70%, and FP = 80%. Orange and blue links denoted positive and negative features, respectively, and green links in bootstrapping methods denoted shared features, that is, features that were overlapping between the baseline method and each of bootstrapping methods. Note that all features selected by the bootstrapping methods overlapped with the baseline method

tstrapping without replacement method increased the predictive correlation values by 27.0, 33.6, 27.5, 14.8, and 18.5%. Among five models, SVR, LASSO, and Ridge performed best and the mean correlation *R* was up to 0.27, while CPM-P and CPM-N achieved similar performances of mean correlation *R* = 0.19 and 0.20, respectively. In the bootstrapping with replacement, the accuracy increased by 21.2, 23.9, 18.9, 12.1, and 14.0% for CPM-P, CPM-N, SVR, LASSO, and Ridge, respectively. Similar to the bootstrapping without replacement, the bootstrapping with replacement showed best accuracies in SVR (mean correlation *R* = 0.25), LASSO (mean correlation *R* = 0.27), and Ridge (mean correlation *R* = 0.26), and similar results in CPM-P (mean correlation *R* = 0.18) and CPM-N (mean correlation *R* = 0.19). The best prediction accuracies of 13 cognitive measures were illustrated in Figure 3 for bootstrapping with and without replacement. Figure S1 showed the *MSE* (across 13 cognitive traits) of each model using the baseline method and bootstrapping methods, and only results derived from the optimal parameters were plotted for bootstrapping methods.

3.2 | Bootstrapping methods largely reduced feature dimension

The feature dimension refined by the bootstrapping methods decreased enormously compared with the baseline method (Figure 4). Additionally,

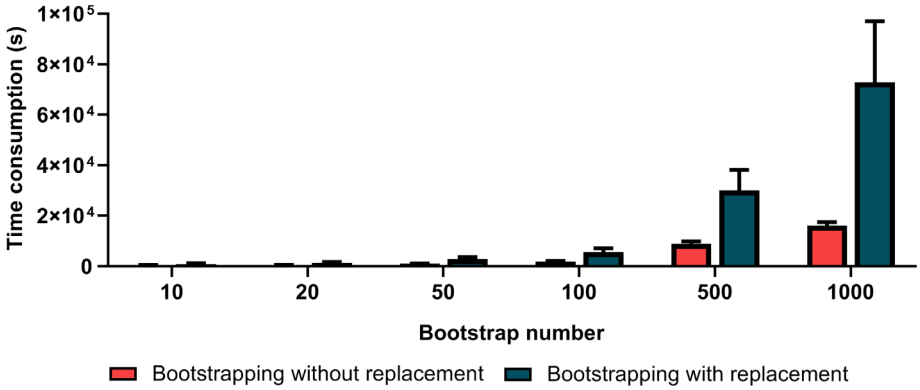

FIGURE 6 The mean time consumption of the bootstrapping methods across 13 cognitive traits

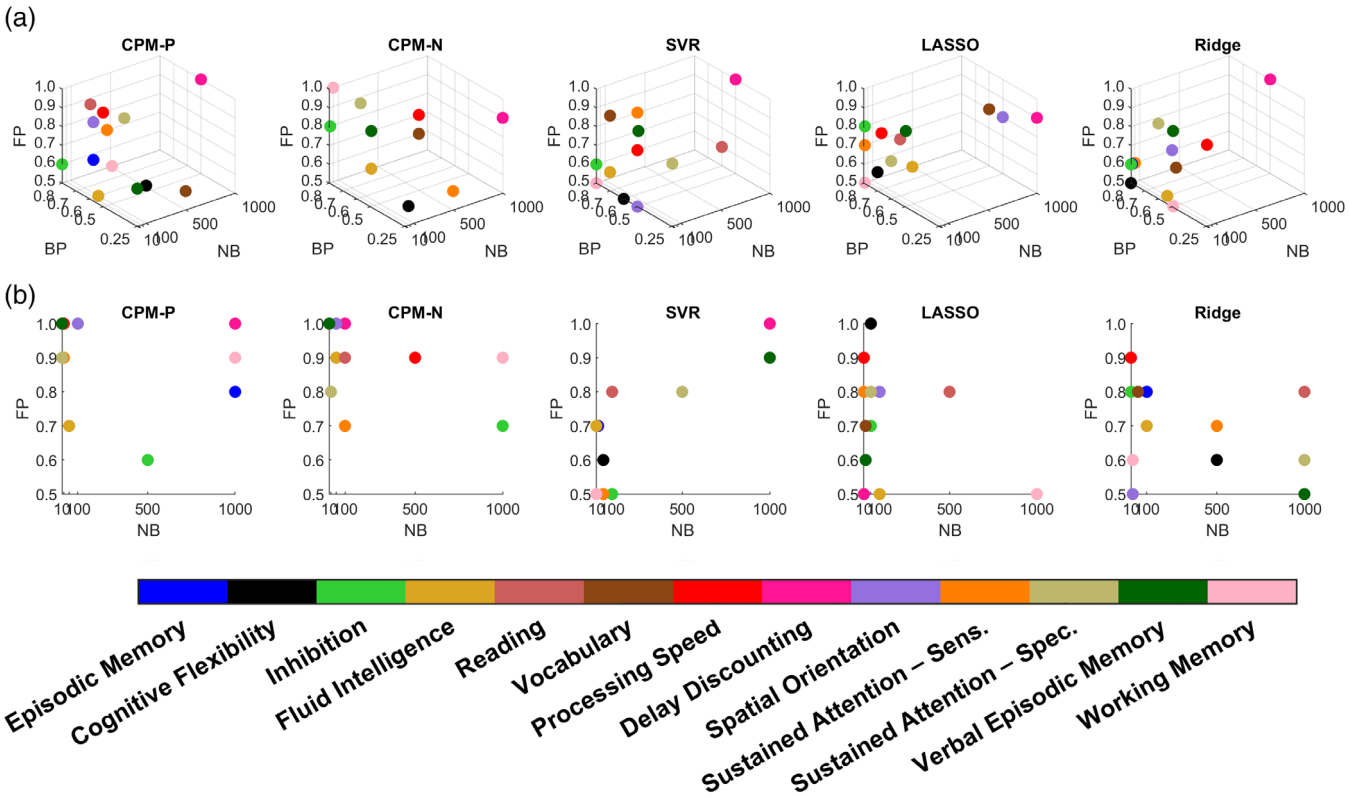

FIGURE 7 The optimal parameter setting for each cognitive measure in CPM-P, CPM-N, SVR, LASSO, and Ridge when using (a) bootstrapping without replacement and (b) bootstrapping with replacement. The colored dots denoted different types of cognitive scores

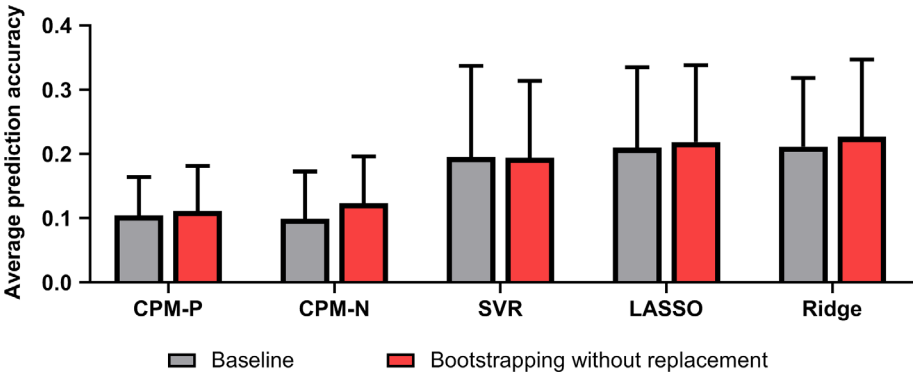

FIGURE 8 The mean correlation *R* averaged across 13 cognitive traits in the validation dataset using bootstrapping without replacement. The prediction accuracy was calculated by averaging the correlation *R* of each trait-specific model across 8 predefined parameter settings and then averaging across 13 predictive models. See Figure S2 for the corresponding mean square error (*MSE*)

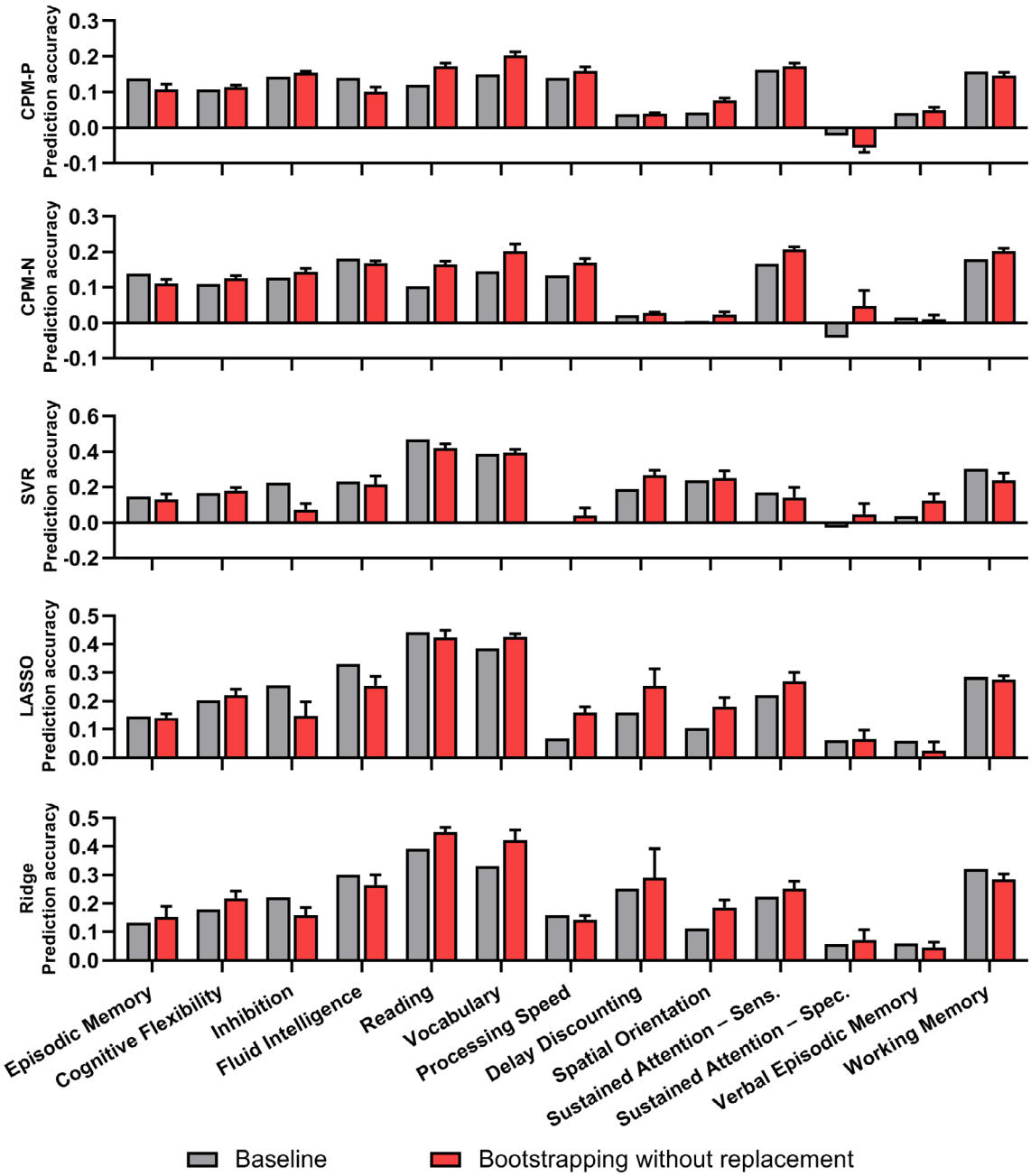

FIGURE 9 The correlation *R* of 13 cognitive trait predictions using the validation dataset. For bootstrapping without replacement, the "bars" showed the mean and standard error of correlation *R* across eight predefined parameter settings

Figure 5 illustrated an example of features for Working Memory selected by the baseline method and bootstrapping methods. More importantly, the refined feature set by bootstrapping methods overlapped with feature set selected by baseline method (represented in green links). Using the bootstrapping without replacement, the average feature dimension decreased from 2,574 to 249 in CPM-P, from 2,648 to 97 in CPM-N, from 5,221 to 1,539 in SVR, from 5,221 to 1,342 in LASSO, and from 5,221 to 1,441 in Ridge. Using the bootstrapping with replacement, the average feature dimension decreased to 320 in CPM-P, 235 in CPM-N, 3,273 in SVR, 2,921 in LASSO, and 3,185 in Ridge. Generally, the bootstrapping without replacement reduced more features than bootstrapping with replacement.

3.3 | Time consumption increased with the number of bootstrapping

The time consumption of the bootstrapping methods gradually increased with NB (Figure 6). Under the same NB, bootstrapping with replacement consumed more time than bootstrapping without replacement. When NB exceeded 500, the average consumption time became significantly large, implying that NB was preferably set to be less than 500.

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

The optimal bootstrapping parameter settings (i.e., NB, FP and BP) for each cognitive measure were depicted in Figure 7. Here, a dot in the parameter space referred to the best parameter setting for one cognitive measure. The results showed that most of the values of the optimal NBs were small, and 72% of these were less than 100 in the bootstrapping methods. As for FP and BP, they varied in different models under different bootstrapping methods. In bootstrapping without replacement, most optimal FPs were relatively high (0.7~0.9) for CPM and LASSO, relatively small (around 0.5) for SVR, and moderate (0.5~0.7) for Ridge. In bootstrapping with replacement, most optimal FPs were relatively high (0.8~1.0) for CPM, LASSO, and Ridge, and moderate (around 0.5) for SVR, while relatively low BPs (around 0.5) were suitable for CPM, SVR, and Ridge, but moderate (around 0.8) for LASSO.

3.5 | Bootstrapping also improved prediction of cognitive traits in the validation dataset

Figure 8 illustrated the mean correlation *R* (across 13 cognitive traits) of each model using the baseline method and bootstrapping (without replacement) in the validation dataset. Compared with the baseline method (mean correlation *R* = 0.10, 0.10, 0.20, 0.21, and 0.21 for CPM-P, CPM-N, SVR, LASSO, and Ridge, respectively), the bootstrapping without replacement obviously increased the predictive correlation values by 5.8, 24.5, 4.3, and 7.2% for CPM-P, CPM-N, LASSO, and Ridge separately, while the mean correlation *R* of SVR remained almost unchanged. Among all models, LASSO and Ridge performed best and reached mean correlation *R* = 0.22 and 0.23, SVR achieved suboptimal performance at mean correlation *R* = 0.19, CPM-P and CPM-N achieved similar performances of mean correlation *R* = 0.11 and 0.12, respectively. The correlation *R* of 13 cognitive measures was illustrated in Figure 9, and mean *MSE* across 13 cognitive traits was shown in Figure S2.

4 | DISCUSSION

In order to improve the performance of RSFC prediction, this study developed a bootstrapping based feature selection framework for RSFC prediction of cognitive traits. Using a large sample from HCP dataset, the prediction of cognitive traits with the bootstrapping methods significantly outperformed the baseline CPM, SVR, LASSO, and Ridge model under optimal parameters, and the dimension of selected features decreased dramatically using the bootstrapping methods. In addition, bootstrapping could improve the predictive performances in the validation dataset, demonstrating its good generalization.

4.1 | Bootstrapping enhanced the RSFC-phenotype associations

Modeling cognitive traits with RSFC is a persistent pursuit in modern cognitive neuroscience studies (Petersen & Sporns, 2015; Poldrack & Farah, 2015). A previous study showed 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 & Buckner, 2017; Gordon et al., 2017; Gratton et al., 2018; Mueller et al., 2013), thus confounding 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 decomposed RSFC (or timecourse) to extract individual-specific RSFC (or timecourse) with different methods, which obtained obvious improvements in prediction. The 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, Stephen, Wilson, Calhoun, & Wang, 2019) and dynamic FC (Liegeois et al., 2019; Lim et al., 2018; Park et al., 2018), which could provide complementary information to the conventional FC. The third way is to decrease the influence of the possible noise in rs-fMRI signal, for instance, the 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-phenotype 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 these studies had to use a constant parameter setting to conduct such methods without any optimal parameter searching. In contrast, our method is a more feasible manner and also displays good generalization that verified in the validation dataset.

As we know, good stability of biomarkers is thought to be as important as high classification performances (He & Yu, 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., multicenter 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 specially 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 participants are healthy or not, which may therefore result in instability in feature sets derived from subsamples and further weakens the prediction performance. It is necessary to eliminate the heterogeneity in the dataset and to obtain stable representations standing for neural activities. Figures 4 and 5 clearly showed that the amounts of fake correlated feature sets were large, implying that it was very beneficial to take into account the stability of feature selection for machine learning studies on 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 versus bootstrapping without replacement

The bootstrapping with and without replacement were compared in the study. Although bootstrapping with replacement was referred more frequently than bootstrapping without replacement, it was reported not to guarantee reliable results sometimes. In Strobl, Boulesteix, Zeileis, & Hothorn (2007), for example, feature selection in random forest was biased by bootstrapping with replacement, while important 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, but 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 consistent with the original data (Rospleszcz, Janitza, & Boulesteix, 2016). In our study, bootstrapping without replacement selected less features and showed better prediction performances for cognitive measures than bootstrapping with replacement. Moreover, bootstrapping without replacement required less computational cost, because it resampled less samples from original dataset than bootstrapping with replacement did. Since bootstrapping without replacement demonstrated superiority in improving prediction performance, reducing redundant features and running cost than bootstrapping with replacement, we recommended bootstrapping without replacement as a feature selection technique to draw reliable predictors.

4.3 | Comparison of CPM, LASSO, SVR, and Ridge regression

CPM had poorer prediction performance compared with SVR, LASSO, and Ridge models. There were three possible aspects accounted for it: (a) CPM is a simple linear regression model, while SVR, LASSO, and Ridge are models with L1 or L2-norm regularization, which may therefore increase the model complexity and obtain better prediction performances. (b) The correlated feature sets were respectively divided into positive and negative feature sets in CPM, while SVR, LASSO, and Ridge used their combination, which may also weaken the CPM performance significantly. (c) CPM assigned same weight to all features, which might underestimate key predictors and overestimate weak predictors. However, when using bootstrapping methods, CPM achieved relatively large improvement in prediction and the most obvious reduction of features, suggesting future studies using CPM should consider the bootstrapping methods.

SVR could produce sparse model since it only relied on support vectors, and it could effectively decrease overfitting risk and display good generalization. However, efficiency was a crucial problem for high-dimensional data while SVR was tremendously expensive in computation for model fitting (Cui & Gong, 2018; Shen et al., 2017).

LASSO ensured sparseness by randomly selecting one feature from correlated features, which was shown to perform well in challenging situation like potential features size exceeded the sample size (Abraham et al., 2017; Bunea, Tsybakov, & Wegkamp, 2007a, 2007b; Zhang & Huang, 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 features (N is the sample size), thus some informative features might be discarded in the model (Efron, Hastie, Johnstone, & Tibshirani, 2004; Ryali et al., 2012).

Ridge regression can handle the problem of multicollinearity by adding a small positive quantity in the diagonal elements of design matrix (Hoerl & Kennard, 2000), whereas Ridge regression was unable

to produce parsimonious model since it kept all predictors in the model without any selection (Zou & Hastie, 2005).

Although the four regression algorithms adopted were different, when combined with bootstrapping methods, the predictive models obtained higher prediction performances (Figure 2) and lower feature dimensions (Figure 3), suggesting that the benefits of the bootstrapping methods were not limited to specific regression algorithm. Besides, the bootstrapping methods displayed good generalization in the validation dataset, demonstrating bootstrapping methods should be adopted in future prediction tasks.

4.4 | Optimal selection of parameter setting

There were three parameters for the bootstrapping methods (NB, BP, and FP), and it is not known beforehand which parameter setting was best for a give problem. In both bootstrapping methods, most of the best performances were achieved with NB less than 100 in discovery dataset, implying that the bootstrapping did not require a mass of resampling times to acquire stable features. Besides, small NB could lessen computational cost, so we recommended small NB (<100) for the bootstrapping methods. FP indicated the stability of the features 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 a bootstrapping-based feature selection method 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. BP stood for the resample percentage of the bootstrapping without replacement from the original dataset, which also did not show obvious rules. The results from the validation dataset also demonstrated that although the optimal parameter setting for bootstrapping without replacement was not searched, it still enhanced overall prediction performances in different regression models. Therefore, in practical application, to identify optimal parameters is time consuming and seems unnecessary, to choose the parameters from a reasonable range would be also useful to achieve higher prediction accuracy.

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 methods 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, that is, [15, 25, 50, 100, 200, and 300]. To have direct comparisons among different parcellations, the baseline CPM, SVR, LASSO, and Ridge method were used to predict the cognitive traits with FC matrices under different parcellations. The results were illustrated in Figure 10, and the main tendency was that the prediction accuracy increased with component dimension, and grew slowly from 100 to 300 ICA components, and similar trend was also revealed in other studies (Abraham et al., 2017; Dadi et al., 2019). Therefore, the current analysis was only based on 300 independent components. Besides, significant threshold *p*-value was an important parameter for correlation analysis. In order to conserve features as much as possible, the *p*-value was set at .05 rather than other more stringent values 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. Although the bootstrapping methods could improve the RSFC-phenotype association, the prediction accuracy was still not high enough, and future work should combine other improvement strategies together to make the RSFC-based prediction applicable for clinical requirement. At last, although our study only focused on RSFC, the bootstrapping methods could easily extend to any correlation-based prediction tasks with any imaging data, such as task-fMRI functional connectivity, DTI and EEG.

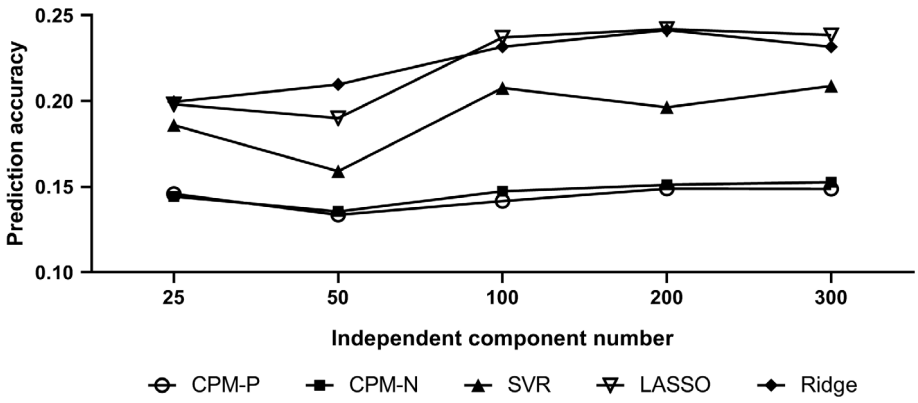

FIGURE 10 The prediction correlation *R* in different independent component number. Each point was mean correlation *R* averaging across 13 cognitive trait predictions, and only baseline method was used for feature selection. The result of 15 ICA components was not shown since no significant features were selected for some cognitive traits

5 | CONCLUSION

This study presented a bootstrapping-based feature selection framework and applied to CPM, SVR, LASSO, and Ridge models, and the results demonstrated the bootstrapping methods 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 methods especially bootstrapping without replacement.

ACKNOWLEDGMENTS

This research was supported by Beijing Municipal Natural Science Foundation (L192044). B.J. was supported by Beijing Municipal Commission (KM202010025025). We are grateful to Xiangyu Ma from Capital Medical University for valuable discussions for the manuscript.

AUTHOR CONTRIBUTIONS

L.W. performed the data analysis and wrote the draft. B.J. conceived and designed the experiments, and rewrote some paragraphs in Introduction and Discussion parts. H.L. revised the draft. All authors read, revised, and approved the final version of the manuscript.

CONFLICT OF INTERESTS

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).

ORCID

Lijiang Wei https://orcid.org/0000-0001-7635-3468

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*, 344. https://doi.org/10.3389/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. *Neural Information Processing-Letters and Reviews*, *11*(10), 203–224.

Beaty, R. E., Kenett, Y. N., Christensen, A. P., Rosenberg, M. D., Benedek, M., Chen, Q. L., … 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), 457–471.e5. https://doi.org/10.1016/j.neuron.2017.06.038

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1023/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*. Proceedings. IEEE International Symposium on Biomedical Imaging. https://doi.org/10.1109/ISBI.2016.7493504

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., Chyzhyk, D., Milham, M., Thirion, B., … 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., … 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., … 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. *ICDM'10 Proceedings of the 10th Industrial Conference on Advances in Data Mining: Applications and Theoretical Aspects*, *6171*, 28–41. https://doi.org/10.1007/978-3-642-14400-4_3

Efron, B., Hastie, T., Johnstone, I., & Tibshirani, R. (2004). Least angle regression. *Annals of Statistics*, *32*(2), 407–451. https://doi.org/10.1214/009053604000000067

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/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. In: Ślęzak, D., Tan, A. H., Peters, J. F., Schwabe, L. (eds), *Brain Informatics and Health. BIH 2014. Lecture Notes in Computer Science*, vol 8609. Cham: Springer. https://doi.org/10.1007/978-3-319-09891-3_3

Elliott, M. L., Knodt, A. R., Cooke, M., Kim, M. J., Melzer, T. R., Keenan, R., … 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., … Deisseroth, K. (2016). Prefrontal cortical regulation of brainwide circuit dynamics and reward-related behavior. *Science*, *351* (6268), aac9698. https://doi.org/ARTNaac969810.1126/science.aac9698

Finn, E. S., Shen, X. L., Scheinost, D., Rosenberg, M. D., Huang, J., Chun, M. M., … 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*, 116038. 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., … Dosenbach, N. U. F. (2017). Precision functional mapping of individual human brains. *Neuron*, *95*(4), 791–807.e7. 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., … Petersen, S. E. (2018). Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron*, *98*(2), 439–452.e5. 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., & Biessgmann, 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

Hoerl, A. E., & Kennard, R. W. (2000). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, *42*(1), 80–86. https://doi.org/10.2307/1271436

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. T., Calhoun, V. D., Fan, L., Zuo, N., Jung, R., Qi, S., … Sui, J. (2019). Gender differences in connectome-based predictions of individualized intelligence quotient and sub-domain scores. *Cerebral Cortex*, *00*, 1–13. https://doi.org/10.1093/cercor/bhz134

Jiang, R. T., Calhoun, V. D., Zuo, N. M., Lin, D. D., Li, J., Fan, L. 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., … 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

Khosla, M., Jamison, K., Kuceyeski, A., & Sabuncu, M. R. (2019). Ensemble learning with 3D convolutional neural networks for functional connectome-based prediction. *NeuroImage*, *199*, 651–662. https://doi.org/10.1016/j.neuroimage.2019.06.012

Kong, R., Li, J. W., Orban, C., Sabuncu, M. R., Liu, H. S., Schaefer, A., … 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., … 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

Liégeois, R., Li, J., Kong, R., Orban, C., Van De Ville, D., Ge, T., … Yeo, B. T. T. (2019). Resting brain dynamics at different timescales capture distinct aspects of human behavior. *Nature Communications*, *10*(1), 2317–2317. https://doi.org/10.1038/s41467-019-10317-7

Liem, F., Varoquaux, G., Kynast, J., Beyer, F., Masouleh, S. K., Huntenburg, J. M., … 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

Lim, J., Teng, J., Patanaik, A., Tandi, J., & Massar, S. A. A. (2018). Dynamic functional connectivity markers of objective trait mindfulness. *NeuroImage*, *176*, 193–202. https://doi.org/10.1016/j.neuroimage.2018.04.056

Marchitelli, R., Minati, L., Marizzoni, M., Bosch, B., Bartres-Faz, D., Muller, B. W., … 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., … 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., … 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., … 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., … 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., … 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

Park, B., Moon, T., & Park, H. (2018). Dynamic functional connectivity analysis reveals improved association between brain networks and eating behaviors compared to static analysis. *Behavioural Brain Research*, *337*, 114–121. https://doi.org/10.1016/j.bbr.2017.10.001

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 Research*, *1722*, 146348. https://doi.org/10.1016/j.brainres.2019.146348

Rizk-Jackson, A., Stoffers, D., Sheldon, S., Kuperman, J., Dale, A., Goldstein, J., … 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*, 119. https://doi.org/10.3389/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

Santarnecchi, 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., Beckmann, C. F., Andersson, J., Auerbach, E. J., Bijsterbosch, J., Douaud, G., … Consortium, W.-M. H. (2013). Resting-state fMRI in the human Connectome project. *NeuroImage*, *80*, 144–168. https://doi.org/10.1016/j.neuroimage.2013.05.039

Smola, A. J., & Scholkopf, B. (2004). A tutorial on support vector regression. *Statistics and Computing*, *14*(3), 199–222. https://doi.org/10.1023/B:Stco.0000035301.49549.88

Sripada, C., Rutherford, S., Angstadt, M., Thompson, W. K., Luciana, M., Weigard, A., … 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*(1), 25. https://doi.org/Artn2510.1186/1471-2105-8-25

Sun, J. Z., Liu, Z. W., Rolls, E. T., Chen, Q. L., Yao, Y., Yang, W. J., … 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.

Toiviainen, 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*. Annual International Conference of the IEEE Engineering in Medicine & Biology Society, *2015*, 2641–2644. https://doi.org/10.1109/EMBC.2015.7318934

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* (2nd ed.). San Diego, CA: Academic Press.

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., … 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*, 301–320. https://doi.org/10.1111/j.1467-9868.2005.00527.x

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Wei L, Jing B, Li H. Bootstrapping promotes the RSFC-behavior associations: An application of individual cognitive traits prediction. *Hum Brain Mapp*. 2020; 41:2302–2316. https://doi.org/10.1002/hbm.24947