Machine Learning to Infer Neurocognitive Testing Scores Among Adolescents and Young Adults with Congenital Heart Disease

Mohammad Arafat Hussain Sheng He
Heather R. Adams Evdokia Anagnoustou
David C. Bellinger Martina Brueckner Wendy K. Chung
John Cleveland Bruce D. Gelb Elizabeth Goldmuntz
Donald J. Hagler, Jr. Hao Huang Patrick McQuillen
Thomas A. Miller Ami Norris-Brilliant
George A. Porter Jr. Nina Thomas Madalina E. Tivarus
Duan Xu Yufeng Shen Jane W. Newburger
P. Ellen Grant Sarah U. Morton Yangming Ou

May 30, 2025

1 Improved Forward Inclusion Backward Elimination (FIBE)

The FIBE pipeline[1] is an iterative feature selection method (see Fig. 1) designed to optimize the performance of a model set M with K predictive models by refining the feature set F_U. The pseudo-code for the proposed FIBE pipeline for a regression task is presented in Algorithm 1. Initially, the data D is divided into N outer folds. For each outer fold i, the data D is split into training/validation data d_i and test data $test_i$. An empty feature set placeholder F_i and an initial performance metric m_0 are initialized.

Forward Inclusion Phase (lines 8–24): In the forward inclusion phase, each feature f not already in F_i is temporarily added to F_i. The algorithm performs N inner folds on the training/validation data d_i. For each inner fold j, the data d_i is split into training data $train_{ij}$ and validation data val_{ij}. Each model p_k of M is trained on $train_{ij}$ and validated on val_{ij}, resulting in an estimated performance metric $temp_{ijk}$, where $k \in K$. The mean performance metric m_s is calculated across the N inner folds and K models. If m_s improves upon the previous metric m_{s-1}, i.e., $m_s < m_{s-1}$ for regression tasks, the feature f is retained in F_i; otherwise, it is removed.

[1] `https://github.com/i3-research/fibe`

Algorithm 1 Forward Inclusion Backward Elimination (FIBE) Pipeline for Regression Task

Input: Data D with feature set F_U, Model set M with K models, Number of outer/inner folds N, Number of iterations I, voting criterion v
Output: Selected Feature Set F_S, N-folds test performance metrics $metric_t$ $(t \in [1, N])$
for $i = 1$ to N **do** ▷ Loop over outer folds
    Split data D for outer folds training/validation d_i and test $test_i$
    Initialize empty Feature Set placeholder F_i, lowest error metric $m_0 = +\infty$
    for $iter = 1$ to I **do**
        Initialize step counter $s \leftarrow 1$
        for each feature f in feature set $F_U \backslash F_i$ **do** ▷ Forward inclusion process starts
            Add $f \rightarrow F_i$ temporarily
            for $j = 1$ to N **do** ▷ Loop over inner folds
                Split d_i into training data $train_{ij}$ and validation data val_{ij}
                for predictive model p_k and model index k in model set M **do**
                    Train p_k with feature set F_i of $train_{ij}$
                    Validate on val_{ij} and estimate error metric $temp_{ijk}$
                end for
            end for
            Estimate mean error metric $m_s = \frac{1}{NK} \sum_{j=1}^{N} \sum_{k=1}^{K} temp_{ijk}$
            if m_s is better than m_{s-1} **then**
                keep f in F_i
            else
                remove f from F_i
            end if
            $s \leftarrow s + 1$
        end for
        for each feature f in feature set F_i **do** ▷ Backward elimination process starts
            Remove f from F_i temporarily
            for Loop over inner folds: $j = 1$ to N **do**
                for predictive model p_k and model index k in model set M **do**
                    Train p_k with feature set F_i of $train_{ij}$
                    Validate on val_{ij} and estimate error metric $temp_{ijk}$
                end for
            end for
            Estimate mean error metric $m_s = \frac{1}{NK} \sum_{j=1}^{N} \sum_{k=1}^{K} temp_{ijk}$
            if m_s is better than m_{s-1} **then**
                Keep current F_i with f removed
            else
                Put back f to F_i
            end if
            $s \leftarrow s + 1$
        end for
    end for
end for
$g \leftarrow$ maximum length among F_i $(i \in [1, N])$
for $i = 1$ to N **do** ▷ Building N dictionaries with features as keys
    $g \rightarrow temp$
    for $j = len(F_i)$ to 1 **do**
        Add $temp$ as a 'value' to feature f_{ij} as 'key'
        $temp \leftarrow temp - 1$
    end for
end for
Create a unified dictionary U_{dict} from N dictionaries by including each feature (f) once and summing its values across all dictionaries.
$F_S \leftarrow$ Select features f from U_{dict} that satisfies $U_{\text{dict}}[f] \geq g$ and appears $\geq v$ times in F_i $(i \in [1, N])$
for $i = 1$ to N **do** ▷ Final testing for outer folds with selected feature set F_S
    Split data D for outer folds training d_i and test $test_i$
    Train M with feature set F_S of d_i
    Test on $test_i$ and estimate error metric $metric_i$
end for
Return: F_S, U_{dict}, $metric_i$ $(i \in [1, N])$

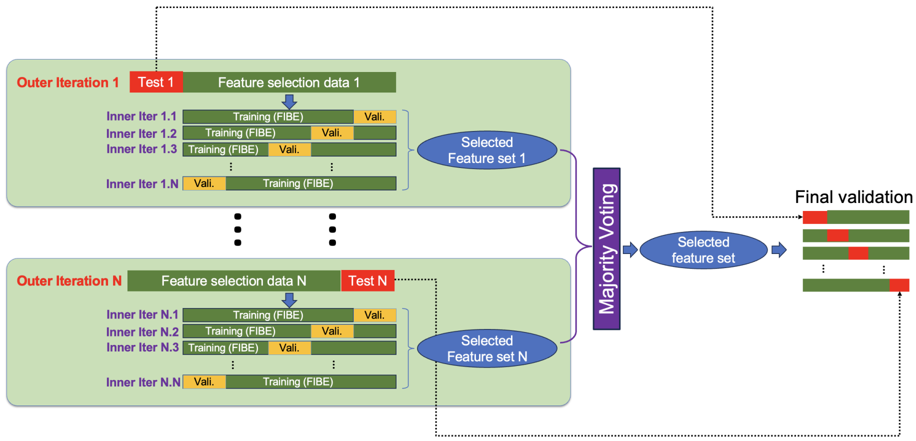

Figure 1: Schematic diagram of our improved forward inclusion backward elimination (FIBE) technique.

Backward Elimination Phase (lines 25–40): In the backward elimination phase, each feature f in F_i is temporarily removed. The inner fold cross-validation process is repeated, with each model p_k of M trained and validated on the modified feature set. The mean performance metric m_s is recalculated. If m_s improves upon the previous metric, the feature is permanently removed from F_i; otherwise, it is reinstated. This iterative process ensures that only features contributing positively to the model's performance are retained.

Feature Aggregation and Selection (lines 43–52): After completing the specified number of iterations I (lines 6–41), the features appearing in the feature sets F_i across all outer folds are evaluated. To identify important features, we construct feature dictionaries for each fold and create a unified dictionary as follows:

1. Let g denote the maximum length among the feature sets F_i ($i \in [1, N]$). For each fold i, initialize a temporary variable *temp* with the value of g. Iterate over the features in F_i in reverse order of their ranking (from $len(F_i)$ to 1). Assign the current value of *temp* as the 'value' to the feature f_{ij} (the j-th feature in F_i), where the feature serves as the 'key'. Decrease *temp* by 1 at each step.

2. After constructing individual dictionaries for all N folds, create a unified dictionary U_{dict} by including each unique feature f only once. For features that appear in multiple dictionaries, their corresponding values are summed across all dictionaries and assigned as a single value $U_{\text{dict}}[f]$.

3. Finally, select the feature set F_S from U_{dict} that satisfies two criteria: (i) the feature f's aggregated value $U_{\text{dict}}[f]$ must be greater than or equal to g, and (ii) the feature f must appear in at least v folds, where v is a predefined threshold. This ensures that the selected features are consistently important

across folds and have sufficient aggregated importance.

Final Testing (lines 53–57): Finally, the algorithm tests the model M using the selected feature set F_S on the outer test sets $test_i$. The performance metrics $metric_i$ are estimated for each outer fold i. The algorithm returns the selected feature set F_S and the test performance metrics $metric_i$.

2 Implementation Details

In our FIBE employment, we use a F_U feature set consisting of 517 features. Our model set M consists of three predictive models (i.e., $K = 3$): regression forest (100 trees with a maximum depth of 5), linear support vector regressor (SVR), and the Gaussian SVR. We set inner and outer folds, $N = 5$, and the number of iterations, $I = 3$. We also set the voting threshold criterion $v = round(0.6 \times N)$, i.e., we select those features as dominant for a particular neurocognitive score that appears $\geq v$ times over N outer folds. For calculating the error between the ground truth and predicted scores in steps 17 and 33 in Algorithm 1, we use the mean absolute error, m_s, loss function defined as:

$$m_s = \frac{1}{NK} \sum_{j=1}^{N} \sum_{k=1}^{K} temp_{ijk} = \frac{1}{NKA} \sum_{j=1}^{N} \sum_{k=1}^{K} \left\{ \sum_{a=1}^{A} |G_{j,k,a} - P_{j,k,a}| \right\}_i, \quad (1)$$

where A is the total number of predictions accumulated for K models over N inner folds within each outer fold i, and G and P are the ground truth and the predicted scores, respectively. We ran our experiments in Python version 3.6 on the E2 cluster of Boston Children's Hospital using an Intel E5-2650 v4 Broadwell 2.2 GHz processor, and 16 GB of RAM.

3 Features Importance Ranking in Selected Feature Subset

Our selected feature set F_S comes with a dictionary U_{dict} that contains values $U_{\text{dict}}[f]$ for a feature key f, and a higher $U_{\text{dict}}[f]$ value is proportional to the feature f's early selection in forward inclusion stage as well as most frequent selections across N-folds. Thus, we can calculate the relative importance of a feature f in the feature set F_S. We define a feature weight metric f_{weight} for a particular feature f in the unified dictionary U_{dict} as:

$$f_{\text{weight}}(\%) = \frac{U_{\text{dict}}[f]}{\sum_{i=1}^{n} U_{\text{dict}}[f_i]} \times 100, \quad (2)$$

where f_i represents each feature in U_{dict}, and n denotes the total number of features in the dictionary. Note that this relative feature weight $f_{\text{weight}}(\%)$ is calculated for better feature importance visualization purposes. These weights are not used in any predictive tasks.

Table 1: Pearson Correlation for Full Scale IQ

Feature	**Correlation**	**P-value**
DAD_ED_LEV_STD	0.15	0.17
cort_thick-ctx_rh_G_front_sup	-0.28	0.0074
cort_thick-ctx-rh-fuzzy12_dorsomedialfrontal	-0.25	0.019
cort_area-ctx_lh_S_postcentral	0.38	0.00023
gwcsurf_LD_gm-ctx_lh_G_pariet_inf-Supramar	0.35	0.00069
fiber_vol-R_IFO	0.32	0.0023
gwcsurf_LD_gm-ctx-lh-superiortemporal	0.31	0.0033

Table 2: Pearson Correlation for Digit Span

Feature	**Correlation**	**P-value**
gwcsurf_FA_wm-ctx_lh_S_orbital_lateral	0.37	0.00044
cort_sulc-ctx-lh-postcentral	0.34	0.0013
fiber_LD-R_pSCS	-0.27	0.012
cort_area-ctx_lh_S_temporal_transverse	-0.19	0.067
cort_vol-ctx_lh_G_pariet_inf-Supramar	0.22	0.035
cort_sulc-ctx-rh-frontalpole	0.31	0.0029
gwcsurf_TD_gm-ctx_lh_G_front_inf-Orbital	0.26	0.014
cort_T2w_white-0.2-ctx_lh_G_temp_sup-G_T_transv	-0.20	0.064
gwcsurf_LD_wm-ctx-rh-isthmuscingulate	-0.19	0.071
gwcsurf_FA_wm-ctx_lh_G_temporal_inf	0.28	0.0086
cort_vol-ctx_lh_S_postcentral	0.38	0.0002
gwcsurf_MD_wm-ctx_rh_S_temporal_transverse	-0.24	0.026
cort_area-ctx_lh_S_postcentral	0.42	5.3e-05

Table 3: Pearson Correlation for Math Computation

Feature	**Correlation**	**P-value**
ndd_var	-0.11	0.3
aseg_LD-Left-Cerebellum-Cortex	0.27	0.012
cort_sulc-ctx_rh_G_cingul-Post-dorsal	-0.20	0.057
cort_sulc-ctx_rh_G_front_inf-Orbital	-0.24	0.026
sex	-0.24	0.022
cort_sulc-ctx_lh_G_occipital_sup	0.20	0.061
fiber_vol-Fmaj	0.20	0.062
gwcsurf_MD_wm-ctx_rh_S_suborbital	-0.20	0.057
cort_area-ctx_rh_G_temp_sup-Plan_polar	0.19	0.071
cort_area-ctx-rh-postcentral	0.28	0.0072
cort_sulc-ctx_rh_S_front_middle	0.20	0.062

Table 4: Pearson Correlation for Coding

Feature	**Correlation**	**P-value**
gwcsurf_TD_gm-ctx_lh_Lat_Fis-ant-Vertical	-0.21	0.044
cort_area-ctx-rh-inferiortemporal	0.22	0.038
aseg_MD-Left-Cerebellum-Cortex	0.36	0.00058
cort_thick-ctx_rh_S_circular_insula_inf	-0.41	7.4e-05
subcort_T2w-Right-Cerebellum-Cortex	0.23	0.027
cort_sulc-ctx-rh-mean	-0.34	0.0011
gwcsurf_FA_wm-ctx_lh_S_orbital_lateral	0.36	0.00051
gwcsurf_TD_gwc-ctx_rh_G_subcallosal	-0.24	0.025
fiber_FA-R_ATR	0.32	0.0023
gwcsurf_TD_gwc-ctx_lh_G_occipital_middle	-0.20	0.057
cort_thick-ctx_rh_G_temp_sup-Plan_tempo	-0.37	0.0004

Table 5: Pearson Correlation for Symbol Search

Feature	**Correlation**	**P-value**
gwcsurf_MD_wm-ctx_rh_S_suborbital	-0.22	0.042
cort_sulc-ctx_lh_G_rectus	0.18	0.094
cort_area-ctx_lh_S_circular_insula_inf	-0.27	0.0099
gwcsurf_FA_gwc-ctx_lh_S_orbital_lateral	0.19	0.067
cort_sulc-ctx_lh_G_precuneus	-0.20	0.057
subcort_T2w-Right-Accumbens-area	0.17	0.12
gwcsurf_LD_gwc-ctx_lh_S_orbital_lateral	0.22	0.036
subcort_vol-Left-Accumbens-area	0.26	0.016
cort_thick-ctx_lh_S_cingul-Marginalis	-0.16	0.14
cort_area-ctx_lh_S_front_inf	-0.17	0.12
gwcsurf_FA_gwc-ctx_lh_Pole_temporal	0.22	0.043
gwcsurf_FA_wm-ctx-rh-caudalanteriorcingulate	0.14	0.19
gwcsurf_TD_gwc-ctx_lh_S_orbital_lateral	-0.16	0.13
cort_T1w_white-0.2-ctx_rh_S_postcentral	-0.20	0.064

Table 6: Pearson Correlation for Processing Speed Index

Feature	**Correlation**	**P-value**
cort_thick-ctx_rh_S_circular_insula_inf	-0.35	0.00087
cort_sulc-ctx_lh_S_orbital_lateral	0.26	0.014
chr_var	-0.18	0.089
gwcsurf_LD_gwc-ctx_lh_S_orbital_lateral	0.16	0.13
cort_vol-ctx-rh-superiortemporal	-0.32	0.0022
gwcsurf_LD_gwc-ctx_lh_Pole_occipital	-0.24	0.024
cort_vol-ctx-rh-inferiortemporal	0.18	0.084
gwcsurf_MD_gm-ctx_lh_Pole_occipital	0.19	0.069
gwcsurf_FA_wm-ctx_rh_S_orbital-H_Shaped	-0.17	0.12
gwcsurf_TD_gwc-ctx-lh-parsorbitalis	-0.23	0.028

Table 7: Pearson Correlation for Verbal Comprehension Index

Feature	**Correlation**	**P-value**
gwcsurf_FA_gwc-ctx_lh_S_pericallosal	-0.37	0.00037
cort_area-ctx_lh_S_postcentral	0.31	0.0033
gwcsurf_LD_gwc-ctx_lh_S_postcentral	-0.43	2.9e-05
total_#_cardiac_surgery	-0.19	0.08
gwcsurf_LD_gwc-ctx_lh_S_pericallosal	-0.35	0.00076
cort_thick-ctx_rh_S_precentral-sup-part	-0.31	0.0031
gwcsurf_LD_gwc-ctx-lh-isthmuscingulate	-0.40	0.00011
CHD_diagnosis	0.17	0.11
ndd_var	-0.06	0.55
cort_thick-ctx_rh_G_front_sup	-0.35	0.00068
cort_thick-ctx-rh-fuzzy12_dorsomedialfrontal	-0.36	0.00049
gwcsurf_TD_gwc-ctx_lh_Lat_Fis-post	-0.31	0.0034

Table 8: Pearson Correlation for Vocabulary

Feature	**Correlation**	**P-value**
cort_vol-ctx_lh_G_oc-temp_med-Lingual	0.27	0.012
chr_var	-0.23	0.028
cort_vol-ctx_lh_S_postcentral	0.25	0.02
cort_T2w_white-0.2-ctx_rh_S_orbital-H_Shaped	0.25	0.017
subcort_vol-Right-Hippocampus	0.26	0.012
aseg_LD-Right-Caudate	0.24	0.023
gwcsurf_FA_wm-ctx_lh_G_temp_sup-G_T_transv	0.26	0.013
gwcsurf_LD_gm-ctx_lh_G_temp_sup-Lateral	0.23	0.027
cort_thick-ctx-rh-fuzzy12_dorsomedialfrontal	-0.25	0.017
gwcsurf_MD_gm-ctx_rh_G_temp_sup-Lateral	0.33	0.0017
cort_thick-ctx_rh_G_front_sup	-0.24	0.023
cort_sulc-ctx-lh-lateraloccipital	0.29	0.0053
gwcsurf_MD_gm-ctx_lh_G_pariet_inf-Supramar	0.34	0.0012

Table 9: Pearson Correlation for Similarities

Feature	**Correlation**	**P-value**
gwcsurf_FA_gwc-ctx_lh_S_pericallosal	-0.39	0.00019
gwcsurf_LD_gwc-ctx_lh_S_pericallosal	-0.33	0.0015
cort_T2w_white-0.2-ctx_rh_G_temporal_inf	-0.32	0.0026
gwcsurf_LD_gwc-ctx-lh-parsorbitalis	-0.43	3.1e-05
gwcsurf_LD_gwc-ctx_lh_S_postcentral	-0.43	2.7e-05
gwcsurf_MD_gm-ctx-lh-caudalmiddlefrontal	0.31	0.0032
gwcsurf_LD_gm-ctx_lh_S_intrapariet_and_P_trans	0.42	4e-05

Table 10: Pearson Correlation for Reading Composite

Feature	**Correlation**	**P-value**
cort_sulc-ctx_lh_G_temporal_inf	0.34	0.0012
cort_vol-ctx_rh_S_postcentral	0.29	0.0055
cort_thick-ctx_lh_G_Ins_lg_and_S_cent_ins	-0.30	0.0042
cort_sulc-ctx-lh-postcentral	0.29	0.0064
gwcsurf_FA_wm-ctx-lh-transversetemporal	0.25	0.018
subcort_vol-Right-Hippocampus	0.27	0.01
gwcsurf_FA_wm-ctx_rh_G_temp_sup-G_T_transv	0.21	0.048
aseg_LD-3rd-Ventricle	-0.30	0.0048
cort_thick-ctx_rh_S_circular_insula_inf	-0.24	0.025
aseg_MD-3rd-Ventricle	-0.26	0.014
cort_T2w_white-0.2-ctx_lh_Lat_Fis-ant-Horizont	0.23	0.028

Table 11: Pearson Correlation for Spelling

Feature	Correlation	P-value
case	0.23	0.03
gwcsurf_FA_wm-ctx_rh_Lat_Fis-ant-Vertical	0.27	0.01
cort_vol-ctx_lh_S_postcentral	0.31	0.0028
gwcsurf_LD_gm-ctx-rh-caudalmiddlefrontal	-0.22	0.04
cort_thick-ctx_lh_G_Ins_lg_and_S_cent_ins	-0.24	0.025
cort_area-ctx_rh_G_postcentral	0.19	0.071
cort_area-ctx_lh_G_cingul-Post-dorsal	0.25	0.02
gwcsurf_FA_wm-ctx-lh-transversetemporal	0.25	0.019
cort_sulc-ctx_rh_S_oc-temp_lat	0.28	0.007
aseg_LD-Left-Cerebellum-Cortex	0.27	0.0098
gwcsurf_FA_gm-ctx_rh_G_oc-temp_lat-fusifor	0.19	0.067
cort_vol-ctx_rh_S_subparietal	0.25	0.017
cort_vol-ctx-rh-rostralanteriorcingulate	0.21	0.046

Table 12: Pearson Correlation for Word Reading

Feature	Correlation	P-value
ndd_var	-0.12	0.24
cort_T1w_white-0.2-ctx_lh_G_oc-temp_med-Lingual	0.18	0.085
cort_sulc-ctx_rh_G_front_inf-Orbital	-0.24	0.022
cort_sulc-ctx_lh_G_temporal_middle	-0.26	0.013
cort_vol-ctx_lh_Lat_Fis-post	-0.22	0.038
gwcsurf_FA_wm-ctx_rh_G_temp_sup-G_T_transv	0.27	0.0099
gwcsurf_LD_gm-ctx_rh_Pole_temporal	0.22	0.038
cort_vol-ctx_lh_G_oc-temp_med-Lingual	0.35	0.00068
gwcsurf_LD_wm-ctx_lh_S_suborbital	0.22	0.037
gwcsurf_TD_gwc-ctx_lh_G_oc-temp_med-Parahip	-0.22	0.043
gwcsurf_FA_gwc-ctx_rh_S_orbital-H_Shaped	-0.18	0.095
cort_thick-ctx-rh-fuzzy12_dorsomedialfrontal	-0.20	0.063
fiber_vol-L_CgH	0.19	0.069

Table 13: Pearson Correlation for Sentence Comprehesion

Feature	Correlation	P-value
cort_sulc-ctx_lh_G_temporal_inf	0.27	0.011
fiber_FA-R_ATR	0.16	0.13
cort_area-ctx_lh_S_postcentral	0.26	0.015
cort_area-ctx_lh_G_subcallosal	-0.16	0.13
fiber_vol-R_pSCS	0.19	0.082
cort_T2w_white-0.2-ctx_rh_G_subcallosal	0.18	0.091

Table 14: Pearson Correlation for Matrix Reasoning

Feature	**Correlation**	**P-value**
gwcsurf_MD_gwc-ctx-lh-parsorbitalis	-0.45	1.2e-05
cort_sulc-ctx-rh-fuzzy12_posterolateraltemporal	-0.36	0.00046
fiber_FA-L_CgC	0.31	0.0029
fiber_FA-R_IFO	0.42	3.6e-05
gwcsurf_LD_gwc-ctx_lh_G_Ins_lg_and_S_cent_ins	-0.41	6.4e-05
gwcsurf_TD_wm-ctx_rh_G_occipital_sup	-0.32	0.0025
cort_sulc-ctx_lh_S_orbital_med-olfact	0.39	0.00018
cort_thick-ctx-lh-posteriorcingulate	-0.39	0.00013
cort_T2w_white-0.2-ctx_rh_S_oc-temp_med_and_Lingual	-0.31	0.0031
fiber_vol-L_CgC	0.34	0.00094
gwcsurf_FA_gwc-ctx_rh_G_Ins_lg_and_S_cent_ins	0.31	0.0032
gwcsurf_MD_gwc-ctx_lh_S_circular_insula_ant	-0.46	5.2e-06
gwcsurf_TD_gwc-ctx_rh_G_rectus	-0.38	0.00029

Table 15: Pearson Correlation for Block Design

Feature	**Correlation**	**P-value**
cort_sulc-ctx_lh_G_temp_sup-Plan_polar	-0.21	0.047
cort_sulc-ctx_rh_G_precentral	-0.28	0.007
cort_T2w_white-0.2-ctx-lh-parsorbitalis	0.23	0.027
DAD_ED_LEV_STD	0.27	0.011
cort_T1w_white-0.2-ctx-rh-parsorbitalis	-0.21	0.046
cort_sulc-ctx-rh-pericalcarine	0.29	0.0056
cort_sulc-ctx_lh_S_precentral-inf-part	0.22	0.039
chr_var	-0.14	0.19
gwcsurf_LD_gm-ctx-lh-supramarginal	0.32	0.0025
aseg_FA-Left-Thalamus-Proper	-0.20	0.066
cort_sulc-ctx-lh-fuzzy12_superiortemporal	-0.20	0.058
cort_sulc-ctx-lh-pericalcarine	0.24	0.021