Title:

Predicting Cognitive Functioning in ADHD Using Population-Based MRI Across Large and Small Samples

Running Title:

Cognitive Prediction in ADHD Using Population MRI

Authors and Affiliations:

Farzane Lal Khakpoor[1], William van der Vliet[1], Alina Tetereva[1], Yue Wang[1], Narun Pat[1]

[1]Department of Psychology, University of Otago, Dunedin, New Zealand

Corresponding Author:

Narun Pat, also known as Narun Pornpattananangkul

Email: narun.pat@otago.ac.nz

Department of Psychology, University of Otago

William James Building, 275 Leith Walk

Dunedin 9016, New Zealand

Phone: +64 3 470 4629

Keywords:

ADHD, Cognitive Functioning, Neuroimaging, Machine Learning, Generalizability, Multimodal MRI

Abstract

Objective:

To assess whether cognitive prediction models trained on multimodal neuroimaging from a population-based cohort generalize to children with and without ADHD across internal and external datasets.

Methods:

This cross-sectional study used task-based and resting-state fMRI, structural MRI, and diffusion tensor imaging from the Adolescent Brain Cognitive Development (ABCD) Study (n = 11,747; mean age = 9.5 years) to train models predicting cognitive functioning. ADHD diagnoses were stratified into four tiers (n = 1,034 to 61), with the remaining participants classified as non-ADHD (n = 10,713). Models were trained using either single neuroimaging feature sets (e.g., task-based fMRI contrasts or cortical thickness) or combined feature sets via stacking. External generalizability was tested using an independent ADHD dataset (Lytle et al.; ADHD, n = 35; non-ADHD, n = 44; mean age = 9.0 years).

Results:

In ABCD, the stacked model integrating 81 neuroimaging types achieved comparable predictive performance for non-ADHD (r = .57) and ADHD (r = .51–.56 across tiers) groups. Key contributors included (a) fMRI contrasts from the NBack task, particularly in the ventral occipital cortex and anterior cingulate cortex (ACC), and (b) task-based functional connectivity from the dorsal attention and posterior multimodal networks. In external validation, combining different fMRI contrasts from the NBack task yielded similar performance for non-ADHD (r = .36) and ADHD (r = .42) participants.

Conclusions:

Population-trained neuroimaging models generalized well to both ADHD and non-ADHD children, underscoring the translational potential of multimodal brain-based models for predicting cognitive functioning in clinical populations.

Introduction

Machine learning holds promise for exploring brain-behavior relationships and transforming neuroimaging into a predictive tool for psychiatric phenotypes (1). Recent studies that used population-based datasets to train models demonstrated an encouraging ability of neuroimaging in predicting cognitive functioning (2–4). Particularly, those using multimodal neuroimaging, which integrates information from various neuroimaging types, tend to achieve relatively higher predictive performance (5–9). Large, population-based datasets are ideal for training such models due to their size and diversity. However, these datasets often include a mix of typically developing individuals and those with subclinical or undiagnosed conditions, which may not fully represent the neurocognitive profiles of clinical populations, In contrast, smaller clinical datasets are more specific but often lack the size needed for robust model training (1,5–7,10–13). Therefore, it is still unclear if cognitive prediction models, built from large-scale population-based neuroimaging data, generalize well to clinical populations (14), limiting the clinical utility of neuroimaging. Rigorous validation, both internal (within the same dataset) and external (across independent datasets), is essential to bridge this gap.

Cognitive functioning, encompassing attention, inhibition, working/declarative memory, and language, is a key psychiatric phenotype that predicts daily functioning and treatment outcomes (15). Frameworks like the Research Domain Criteria (RDoC) propose that cognitive functioning varies along a transdiagnostic spectrum, spanning from typical to atypical performance, reflecting a continuum of neural and behavioral characteristics across both healthy and clinical populations (16,17). Disruptions in cognitive functioning can significantly impair quality of life, particularly in neurodevelopmental conditions such as attention-deficit/hyperactivity disorder (ADHD), a highly prevalent neurodevelopmental condition (18) with deficits in several cognitive domains (19–22). These deficits are underpinned by widespread alterations in brain morphology and functions (23–27), including reduced cortical thickness in prefrontal regions, altered connectivity in attention networks, diminished activity in prefrontal and parietal regions during cognitive tasks and atypical white matter organization (28–31). This behavioral and neural pattern makes ADHD an ideal test case to assess generalisability of neuroimaging-based cognitive prediction models across populations. If models, built from multimodal neuroimaging in large-scale population-based

data, capture the transdiagnostic spectrum of cognitive functioning (16,17), they should generalize similarly to children with or without ADHD.

To explore this, we trained cognitive prediction models using all the available neuroimaging data, quantified into 81 sets of neuroimaging features, from population-based ABCD (32) children and tested their performance on children with or without ADHD. Next, to ensure that cognitive prediction models generalize well to small clinical-focused datasets, not included in the population-based study, we tested their performance using both internal validation (i.e., on children with ADHD within the ABCD) and external validation (i.e., on children with ADHD from a small-scale study (33)). Given the ADHD-related alterations in brain morphology and functions (23–27), some neuroimaging types may lead to more generalizable models than others. We benchmarked the generalizability of various models, comparing those based on a single set of neuroimaging features (e.g., a contrast from task fMRI or cortical thickness, referred to as the 'single model') RI and diffusion tensor imaging, created by a machine-learning stacking technique (6,34) referred to as the 'stacked model'). Specifically, we aimed to address four key questions: First, we investigated whether there are differences in observed cognitive functioning scores between ADHD and non-ADHD groups and, if so, whether these differences are reflected in the model predictions. Second, whether the predictive performance of these models remains consistent across ADHD and non-ADHD populations. Third, whether integrating multiple neuroimaging modalities yields better predictive accuracy in ADHD groups. Fourth, whether these findings generalize from a large, population-based dataset to a smaller, clinically focused dataset.

Materials and methods

Figure 1 shows the overall design and information about participants from the Adolescent Brain Cognitive Development (ABCD) Study (baseline from release 5.1) (32) and Lytle and colleagues' study (33). We used the population-based ABCD to a) train the cognitive prediction models and b) internally validate the models' performance on out-of-sample participants with ADHD within the same dataset (but not part of the model-building process). We used Lytle and colleagues' small-scale study on ADHD (33) to externally validate the

models trained from the ABCD. Institutional Review Boards at each data collection site approved the study protocols (33,35).

ADHD Identification

For the ABCD, we identified ADHD using Cordova and colleagues' (36) tier system which approximated the Diagnostic and Statistical Manual of Mental Illnesses (DSM-5) (37) standards and progressed from least to most stringent (see Fig 1A and Supplementary Methods). Briefly, we used parent- and teacher-reported responses and ruled out comorbid conditions. Tier 1 included participants with a current ADHD diagnosis based on parent-reported responses from the K-SADS-COMP (38). Tier 2 refined this group by excluding individuals with bipolar disorder, schizophrenia spectrum disorders, or intellectual disability. Tier 3 further incorporated teacher-reported data from the Brief Problem Monitor (BPM) (39) to confirm symptom consistency across settings. Finally, Tier 4 applied clinical cutoffs from the parent-reported Child Behavior Checklist (CBCL) attention or ADHD DSM-5 scale (40).

For the Lytle and colleagues', we followed the study's criteria, based on ADHD Rating Scale-IV (41) and Parent-Kiddie Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL) (38,42) (see Fig. 1B).

Cognitive Functioning

We operationalised cognitive functioning as follows: For the ABCD, we used the total cognitive score from the NIH Toolbox (32,43). For Lytle and colleagues', we used the total cognitive score from the Wechsler Abbreviated Scale of Intelligence, Second Edition (WASI-II) (33,44). See Supplementary Methods for details of each cognitive test.

Neuroimaging: ABCD

For the ABCD, we used pre-processed tabulated data provided by the ABCD study (45) and a subset of fMRIPrep-processed data from the ABCD-BIDS Community Collection (ABCC) (46), resulting in a variety of features. There were 81 sets of neuroimaging features. See (47) and Supplementary Methods for more details about how we processed the neuroimaging features.

Functional MRI (fMRI) Task Contrasts: 56 sets

We computed GLM contrasts, reflecting BOLD activation related to different events in three tasks: Emotional Nback (48), reward-related Monetary Incentive Delay (MID) (49) and inhibitory-controlled Stop Signal Task (SST) (50) (see Supplementary Methods). There were 26 contrasts provided by the ABCD study and 30 contrasts computed by us from the ABCC (46) (see Supplementary Fig. 1).

Functional MRI Connectivity (fMRI FC): 9 sets

Functional Connectivity (FC) was based on fMRI time series during rest and tasks. The ABCD provided 3 sets of rest FC: 1) cortical FC (the mean values of the correlations between pairs of regions within the same large-scale cortical network), 2) subcortical-to-network FC (the mean values of the correlations between subcortical regions and large-scale cortical networks) and 3) temporal variance of each brain region. Using the ABCC, we added region-to-region FC during rest and the three tasks, resulting in four more sets. Note that for the task FC, we regressed out the HRF-convolved events from the BOLD time series. Finally, following previous work (51–53), we created two sets: 1) multitask FC (region-to-region FC of all tasks combined) and 2) general FC (region-to-region FC of rest and all tasks combined) (see Supplementary Methods).

Structural MRI (sMRI) and Diffusion Tensor Imaging (DTI): 16 sets

sMRI reflects anatomical morphology. Using the ABCD and ABCC, we extracted 15 sets of features: cortical thickness, surface area, FreeSurfer summations, cortical volume, sulcal depth, white-matter averaged intensity, grey-matter averaged intensity, T1 normalized intensity, T1 subcortical averaged intensity, T1 subcortical volume, T1 summations, T2 normalized intensity, T2 subcortical averaged intensity, T2 subcortical volume and T2 summations. DTI reflects white matter integrity. We focused on fractional anisotropy (FA) of 23 white matter tracts (54) (see Supplementary Methods).

Neuroimaging: Lytle and colleagues'

Compared to the ABCD(32), Lytle and colleagues' study (33) provided much more limited neuroimaging modalities: two versions of the working-memory Nback task and sMRI(55). See Supplementary Methods for more details about how we processed these neuroimaging features.

fMRI Task Contrasts: 6 sets

Unlike the ABCD, Lytle and colleagues used two versions of the working-memory Nback: verbal and spatial, and they also varied the reward size and feedback presence in each task(42). We computed GLM contrasts from the two versions and averaged across the reward-size and feedback conditions. This left 3 contrasts for each version: 2back, 1back and 2-1back.

fMRI FC: 3 sets

Similar to the ABCD, we generated task FC by regressing out the HRF-convolved events. We then computed region-to-region FC separately for the verbal Nback, the spatial Nback, and the combined tasks as multitask FC.

sMRI: 4 sets

We extracted four sets of features: cortical thickness, surface area, subcortical volume, and FreeSurfer summations.

Cognitive prediction models

We trained cognitive prediction models from neuroimaging data in the ABCD using a nested leave-one-site-out cross-validation approach across 21 sites to ensure robustness and minimize site-specific biases (56). Each site served as a test set once, with the remaining sites used for training. Within each training set, features were standardized (z-scored), and the same standardization parameters were applied to the test set. Hyperparameters were optimized using an inner five-fold cross-validation, with negative mean squared error as the evaluation metric. We then applied the models to children with or without ADHD: within the ABCD (32) who were not part of the model training (internal validation) and outside the ABCD using a separate study by Lytle and colleagues'(33) (external validation). The prediction models were based either on each single set of neuroimaging features (single model) or some combinations across sets (stacked model). Single models were developed using Elastic Net regression to predict cognitive functioning scores from each of the 81 feature sets derived from task-based fMRI, resting-state fMRI, sMRI , DTI. For FC data, Partial Least Squares (PLS) regression was applied as a dimensionality reduction step (57) before Elastic Net, optimizing the number of components (up to 30 for larger sets) via grid

search to retain predictive information while managing high-dimensional data. Elastic Net balanced L1 and L2 regularization to handle multicollinearity and prevent overfitting (58), with hyperparameters (regularization strength and L1/L2 balance) tuned via grid search. Stacked models integrated predictions from single models into 15 configurations, such as the "stacked all" model (including all 81 sets of features) and the "stacked task contrasts" model (combining task-based fMRI contrasts like Nback, MID, and SST). These models used Random Forest regression, with 1,000 decision trees and tuned parameters (e.g., maximum tree depth, number of features per split) to capture non-linear relationships and enhance predictive accuracy (59). Missing data were handled using an opportunistic stacking approach (13), where missing predictions were imputed with large positive and negative values (±1000) to retain participants with partial data while informing the model about missingness (see Supplementary Fig. 1 for the list of single and stacked models).

Due to the differences in feature availability between the ABCD and Lytle and colleagues', we were restricted to the sets of features similar in both. Specifically, we used single and stacked models trained on a subset of Nback contrasts (0back, 2back and 2-0back), which manipulated working memory load—collectively referred to as WM Load contrasts for the stacked model trained on these three—along with FC, multitask FC, general FC, and sMRI data from the ABCD study. These models were then applied to single and stacked sets of features, including verbal and spatial Nback contrasts (1-back, 2-back, and 2-1-back), as well as FC, multitask FC, and sMRI data, in Lytle and colleagues' study (see Supplementary Table 1). We compared the models' predictive performance between children with or without ADHD. Bootstrapping compared performance between groups, with 95% confidence intervals for differences in metrics (2,000 iterations) and distributions of observed versus predicted cognitive scores (10,000 resamples), using Hedges' g to quantify effect sizes (60). Feature importance was assessed using Elastic Net coefficients for single models, PLS

component weights for FC, and SHapley Additive exPlanations (SHAP) values for Random Forest stacked models to quantify individual feature contributions (61,62). See Fig. 1 C&D and Supplementary Methods for more details about the prediction models and evaluation procedure. All utilized code is available in the public domain at https://github.com/HAM-lab-Otago-University/ADHD-MultiModal.

Results

We provide the predictive performance of all single and stacked models in Supplementary Tables 2-5.

Differences in Observed and Predicted Cognitive Functioning Scores between ADHD Groups

Internal Validation. Observed cognitive functioning scores were statistically significantly higher for non-ADHD compared to each ADHD tier. Predicted scores from the top-performing model (stacked all) replicated this difference, with the magnitude of the group differences (quantified by Hedges' g) being comparable between observed and predicted scores (Fig. 2A).

External Validation. Observed cognitive functioning scores were also statistically significantly higher for non-ADHD than ADHD. The predicted scores from the top-performing Stacked Spatial Nback Contrasts model showed a significant group difference, though the effect size was numerically smaller than that of the observed scores (Fig. 2B).

Predictive Performance Across Groups

Internal Validation (i.e., trained and tested in the ABCD). The stacked model that integrated predictions from all 81 neuroimaging sets of features consistently outperformed

other single and stacked models, achieving an overall out-of-sample correlation of r=.57 (95%*CI*: .56–.59). Bootstrapping analyses did not reveal any statistically significant differences in predictive performance between non-ADHD and any ADHD tiers. This comparable performance across non-ADHD and ADHD of any tiers held for 10 of 15 stacked models (Fig. 3A) and 27 of 81 single models (Supplementary Fig. 3A). Notably, unlike FC stacked models (e.g., All FCs, task FCs, rest FCs), task contrasts models exhibited reduced predictability in stricter ADHD tiers (e.g., Tiers 3 or 4) compared to non-ADHD (e.g., Nback Contrasts: $\Delta r_{\text{ADHD-Tier4-vs-non-ADHD}}$-.11; 95%*CI*: -0.18–-0.04).

External Validation (i.e., trained in the ABCD and tested in Lytle and colleagues'). The top-performing model was the stacked Nback WM load contrasts applied to spatial Nback contrasts (2back, 0back, and 2-0back), achieving an overall correlation of r=.44 (95%*CI*: .19–.52), closely aligning with its internal validation performance (r=.45; 95%*CI*: .46–.50). The second-best model, trained on all applicable sets of features (sMRI, general FC, and Nback WM load contrasts) and applied to sMRI, multitask FC and spatial Nback contrasts, attained r=.37 (95%*CI*: .03–.41). Neither model displayed statistically significant differences in performance between ADHD and non-ADHD groups (see Fig. 3B). Among single sets of features, the spatial and verbal 2back performed best (rs>.35), with no significant difference revealed across groups (see Supplementary Fig. 3B).

Enhancement via Stacking

Internal Validation. The stacked all model outperformed the best single models (Fig. 3A, Supplementary Tables 2–3) across ADHD tiers. Stacked FC and task contrast models also showed higher average correlations than single models. Among the single models, task and rest FC, along with the Nback 2-0back contrast, were top performers (Fig. 3A, Supplementary Tables 2–3). Fig. 4A illustrates the correlation between observed and

predicted cognitive functioning scores for the best-performing stacked and single models across task contrast, FC, and sMRI sets of features.

External Validation. Stacked spatial Nback models (*r*=.44) outperformed single models (e.g., spatial 2-back: *r*=.38, 95% CI: .17–.56; Fig. 3B, Supplementary Tables 4–5), mirroring the prediction boost seen in internal validation. sMRI models showed modest but consistent performance across datasets (*r*=.16–.34). Fig. 4B displays the correlation between observed and predicted cognitive functioning scores for top-performing stacked and single models in each measurement set.

Feature Importance

Feature importance analysis (Fig. 5) highlights key contributors to predicting cognitive functioning scores in the best-performing models. In the stacked-all model that included every neuroimaging set of features, the Nback 2-0back contrast and task FC exhibited the greatest average marginal contributions to the model's predictions, indicated by absolute SHAP at *M*=.30 (*SD*=.002) and *M*=.14 (*SD*=.05), respectively (Fig. 5A). Among single models, task contrasts parcellated with the Glasser atlas showed a higher average correlation between observed and predicted values compared to those based on the Destrieux atlas (Supplementary Table 2), though the latter ranked higher in the stacked model's importance ranking. The FC between regions in the Dorsal Attention and Posterior Multimodal networks had the highest average positive predictive weight in the MID FC (best single FC model) (Supplementary Table 6), while Elastic Net weights for the Nback 2-0back contrast (best single contrast model) emphasized ventral occipital and anterior cingulate cortex (ACC) regions (Supplementary Table 7). For T2 grey matter intensity (best single model among sMRI and DTI), the Paracentral Gyrus and Sulcus and Cingulate Marginal Sulcus were the most positively weighted regions (Supplementary Table 8).

Discussion

Our study demonstrated that cognitive prediction models, trained on multimodal neuroimaging data from large-scale population-based neuroimaging data, effectively predicted cognitive functioning in children with and without ADHD. These models generalized well to children with ADHD within the same large-scale dataset (internal validation) and to an independent small-scale dataset (external validation).

Generalizability of Stacked Models to ADHD

The observed differences in cognitive functioning scores between non-ADHD and ADHD groups across tiers are consistent with prior reports of variations in cognitive performance in ADHD populations (20,21,63,64). The stacked all model accurately captured these differences in predicted cognitive scores, reflecting its sensitivity to the neural patterns underlying cognitive processes that vary across individuals with and without ADHD.

Despite this distributional difference, the majority of stacked models showed comparable performance between ADHD and non-ADHD groups, supporting the hypothesis that these models capture cognition-related neural activities that are transdiagnostic and applicable to both typical and ADHD populations. This generalizability suggests that integrating multimodal neuroimaging features—such as task-based fMRI contrasts, functional connectivity, and structural MRI—effectively represents variability in cognitive processes like attention, working memory, and executive function, which are relevant to ADHD but not exclusive to its deficits (15,16,21,65,66). This aligns with prior studies using single-modality neuroimaging to predict cognitive performance in ADHD (23,28,67), but our multimodal approach enhanced robustness by combining complementary neural signals. These findings indicate that population-based models can serve as scalable tools for capturing cognitive

functioning in clinical populations, potentially informing broader neurodevelopmental research without focusing solely on ADHD-specific impairments.

Performance Boost Through Multimodal Stacking

Stacking significantly enhanced prediction accuracy over single sets of features in ADHD groups, a finding replicated across internal and external validations. This improvement highlights the value of integrating complementary information from task contrast activities, functional connectivity, and structural MRI features, consistent with studies in healthy populations (5–9). By combining neuroimaging features, stacked models capture a richer representation of neural processes underlying cognition, such as FC and structural morphology, which are critical for cognitive tasks across diverse populations, including those with ADHD (6,29,48,51). Task-based fMRI contrasts reflect activation patterns during cognitive demands, while FC captures intrinsic network dynamics, and structural features provide insights into brain morphology that may influence cognitive capacity (68–70). This multimodal synergy enhances the models' ability to generalize to ADHD populations, where cognitive functioning varies widely but is not exclusively tied to disorder-specific alterations in particular brain regions, networks, or structures (23–27).

External Validation

The robust performance of stacked Nback models in an independent clinical dataset, despite differences in demographics, tasks, and cognitive measures, underscores the resilience of these models in capturing core neural signatures of cognitive functioning across diverse contexts. This aligns with prior work showing that brain-based cognitive predictions can transcend diagnostic boundaries (17). However, our study advances research by applying trained models directly to an external dataset without adaptation—a stricter test of

generalizability relevant to clinical settings where model tuning may be impractical. In doing so, we achieved higher correlation values on average than earlier studies (17,67). However, models trained on ABCD's WM Nback contrasts performed better when applied to the spatial Nback task in the external dataset compared to verbal Nback task, likely due to greater similarity in task design and cognitive demands. This task-specific performance suggests that while our models capture generalizable cognition-related neural activities, alignment in task paradigms enhances predictive accuracy. These findings emphasize the potential of stacked models to serve as generalisable tools for predicting cognitive functioning in clinical ADHD populations, provided that task paradigms are carefully considered in model application.

Feature Contributions and Practical Implications

The prominence of FC and Nback task contrasts (particularly, working memory related activity) in the stacked all model's feature importance and as top-performing single models aligns with prior findings that indicated their potential as predictors of cognitive functioning (2,5,12,17,67,71). Nback contrasts, which capture working memory load and cognitive control, reflect neural activation patterns in prefrontal and parietal regions critical for attention and executive function, processes relevant to cognitive performance across populations, including those with ADHD (72). Functional connectivity, particularly task FC, provides insights into dynamic network interactions that support cognitive flexibility and sustained attention, which are vital for cognitive tasks (51,69,73). The significant contribution of Dorsal Attention-Posterior Multimodal and Ventral Multimodal connectivity, and importance of the anterior cingulate and ventral occipital regions during Nback, are consistent with their roles in sustaining attention, integrating sensory information, facilitating semantic processing and emotional and cognitive regulation (67,74–79). These findings suggest that machine learning models prioritizing these features can capture neurocognitive

signatures relevant to cognitive functioning in ADHD populations, without focusing on disorder-specific deficits. For instance, these models could inform personalized interventions by identifying neural markers of cognitive performance, guide treatment planning, or enhance diagnostic precision by complementing behavioral assessments with neuroimaging-based predictions. However, clinical implementation requires further validation in diverse clinical settings and optimization for computational efficiency to ensure practical utility (80).

Limitations and Future Directions

Our results span all ADHD tiers, but we observed variation in model performance across stricter tiers (e.g., Tiers 3/4). To explore this, we benchmarked the generalizability of different sets of neuroimaging features, finding that some—such as stacked all or stacked FCs—exhibited promising consistency across ADHD and non-ADHD groups, while others, like task contrast models (e.g., task contrasts, Nback contrasts), showed reduced performance in stricter tiers, possibly due to distinct neurocognitive profiles or smaller sample sizes increasing variability. The ABCD dataset's proxy ADHD criteria and the limited external validation sample, with few overlapping features, further constrain broad generalizability. Future research should test these models in larger, clinically diagnosed ADHD cohorts using standardized protocols across datasets to clarify whether tier differences reflect diagnostic subtypes or methodological artifacts, prioritizing feature sets that generalize well. Additionally, the attenuated effect size difference between observed and predicted scores in the external dataset underscores the need for further validation or tuning in diverse ADHD populations before relying heavily on population-based models.

Conclusion

By demonstrating robust generalizability—both within the expansive ABCD dataset and across an independent, smaller clinical cohort—our study underscores the promise of population-based multimodal neuroimaging models for capturing cognitive functioning in ADHD. While the standout performance of our stacked models (internal-validation out-of-sample r=.57) that weaved together diverse structural and functional neuroimaging features, may not yet qualify as a fully-fledged clinical tool, our study signals a bright future for the generalizability of this approach across typical and atypical neurodevelopment. This holistic brain-based approach illuminates the intricate neural underpinnings of cognition, paving the way for future research to refine these models with emerging machine learning tools, potentially enhancing their precision and utility for ADHD and other neurodevelopmental conditions in real-world settings.

Acknowledgments

This research was supported by the New Zealand Health Research Council (Grant numbers: 21/618 and 24/838; awarded to Narun Pat), the University of Otago, and the Neurological Foundation of New Zealand (Grant number: 2350 PRG). The authors acknowledge the use of New Zealand eScience Infrastructure (NeSI) high-performance computing facilities and support services, funded by NeSI collaborator institutions and the Ministry of Business, Innovation & Employment.

This work used publicly available data from two sources. The Adolescent Brain Cognitive Development (ABCD) Study®, supported by the National Institutes of Health and federal partners, provided large-scale, multimodal neuroimaging and cognitive data. A full list of grant numbers is available at https://abcdstudy.org/federal-partners.html, and a list of participating institutions is available at https://abcdstudy.org/consortium_members. ABCD data were accessed via NIMH Data Archive under DOIs: http://dx.doi.org/10.15154/z563-zd24 and http://dx.doi.org/10.15154/1523041.

The second dataset, Working Memory and Reward in Children With and Without Attention-Deficit/Hyperactivity Disorder, was originally published by Lytle and colleagues and is accessible on OpenNeuro (https://openneuro.org/datasets/ds002424/versions/1.2.0).

Author contributions are as follows: Khakpoor and Pat conceived and designed the study. Data acquisition, analysis, and interpretation were conducted by Khakpoor, Pat, Wang, van der Vliet, and Tetereva. Khakpoor and Pat drafted the manuscript and performed statistical analyses. All authors critically revised the manuscript. Funding was acquired by Pat. Pat also supervised the

project. All authors approved the final version and are accountable for the accuracy and integrity of the work.

The authors thank Michael Demidenko (Stanford University) for providing fMRI task event extraction scripts. This manuscript reflects the authors' views and does not necessarily represent those of any funding body or the ABCD Consortium.

Data Availability: ABCD Study data are accessible via https://nda.nih.gov/abcd upon NDA approval. The Lytle dataset is available at https://openneuro.org/datasets/ds002424. All analysis code is available at https://github.com/HAM-lab-Otago-University/ADHD-MultiModal.

DISCLOSURES

Farzane Lal Khakpoor, William van der Vliet, Alina Tetereva, Yue Wang, and Narun Pat report no biomedical financial interests or potential conflicts of interest.

References

1. Sui J, Jiang R, Bustillo J, Calhoun V. Neuroimaging-based Individualized Prediction of Cognition and Behavior for Mental Disorders and Health: Methods and Promises. Biol Psychiatry 1969. 2020;88(11):818–28.

2. Finn ES, Shen X, Scheinost D, Rosenberg MD, Huang J, Chun MM, et al. Functional connectome fingerprinting: Identifying individuals based on patterns of brain connectivity. Nat Neurosci. 2015;18(11):1664–71.

3. Sripada C, Angstadt M, Rutherford S, Taxali A, Shedden K. Toward a “treadmill test” for cognition: Improved prediction of general cognitive ability from the task activated brain. Hum Brain Mapp. 2020;41(12):3186–97.

4. Vieira BH, Pamplona GSP, Fachinello K, Silva AK, Foss MP, Salmon CEG. On the prediction of human intelligence from neuroimaging: A systematic review of methods and reporting. Intell Norwood. 2022;93:101654.

5. Tetereva A, Knodt A, Melzer T, Van Der Vliet W, Gibson B, Hariri A, et al. Improving Predictability, Test-Retest Reliability and Generalisability of Brain-Wide Associations for Cognitive Abilities via Multimodal Stacking. bioRxiv. 2025;

6. Tetereva A, Li J, Deng JD, Stringaris A, Pat N. Capturing brain□cognition relationship: Integrating task□based fMRI across tasks markedly boosts prediction and test□retest reliability. 2022;

7. Marek S, Tervo-Clemmens B, Calabro FJ, Montez DF, Kay BP, Hatoum AS, et al. Reproducible brain-wide association studies require thousands of individuals. Nat Lond. 2022;603(7902):654–60.

8. Rasero J, Sentis AI, Yeh FC, Verstynen T. Integrating across neuroimaging modalities boosts prediction accuracy of cognitive ability. PLoS Comput Biol. 2021;17(3):e1008347–e1008347.

9. Rosenberg MD, Casey BJ, Holmes AJ. Prediction complements explanation in understanding the developing brain. Nat Commun. 2018;9(1):589–13.

10. Poldrack RA, Huckins G, Varoquaux G. Establishment of Best Practices for Evidence for Prediction: A Review. JAMA Psychiatry Chic Ill. 2020;77(5):534–40.

11. Varoquaux G. Cross-validation failure: Small sample sizes lead to large error bars. NeuroImage Orlando Fla. 2018;180(Pt A):68–77.

12. Wang Y, Anney R, Pat N. The relationship between cognitive abilities and mental health as represented by cognitive abilities at the neural and genetic levels of analysis. medRxiv. 2024 Jan 1;2024.02.09.24302602.

13. Pat N, Wang Y, Anney R, Riglin L, Thapar A, Stringaris A. Longitudinally stable, brain□based predictive models mediate the relationships between childhood cognition

and socio demographic, psychological and genetic factors. Hum Brain Mapp. 2022;43(18):5520–42.

14. Samea F, Soluki S, Nejati V, Zarei M, Cortese S, Eickhoff SB, et al. Brain alterations in children/adolescents with ADHD revisited: A neuroimaging meta-analysis of 96 structural and functional studies. Neurosci Biobehav Rev. 2019;100:1–8.

15. Abramovitch A, Short T, Schweiger A. The C Factor: Cognitive dysfunction as a transdiagnostic dimension in psychopathology. Clin Psychol Rev. 2021;86:102007.

16. Morris SE, Cuthbert BN. Research Domain Criteria: cognitive systems, neural circuits, and dimensions of behavior. Dialogues Clin Neurosci. 2012;14(1):29–37.

17. Chopra S, Dhamala E, Lawhead C, Ricard JA, Orchard ER, An L, et al. Generalizable and replicable brain-based predictions of cognitive functioning across common psychiatric illness. Sci Adv. 2024;10(45):eadn1862.

18. Sacco R, Camilleri N, Eberhardt J, Umla-Runge K, Newbury-Birch D. A systematic review and meta-analysis on the prevalence of mental disorders among children and adolescents in Europe. Eur Child Adolesc Psychiatry. 2024;33(9):2877–94.

19. Cohen Gilbert JE, Nickerson LD, Seraikas AM, Oot EN, Rieselbach MM, Schuttenberg EM, et al. Large scale brain network activation during emotional inhibitory control: Associations with alcohol misuse in college freshmen. Alcohol Clin Exp Res. 2022;46(3):410–21.

20. Duff CT, Sulla EM. Measuring Executive Function in the Differential Diagnosis of Attention-Deficit/Hyperactivity Disorder: Does It Really Tell Us Anything? Appl Neuropsychol Child. 2015;4(3):188–96.

21. Kofler MJ, Irwin LN, Soto EF, Groves NB, Harmon SL, Sarver DE. Executive Functioning Heterogeneity in Pediatric ADHD. J Abnorm Child Psychol. 2019;47(2):273–86.

22. Senkowski D, Ziegler T, Singh M, Heinz A, He J, Silk T, et al. Assessing Inhibitory Control Deficits in Adult ADHD: A Systematic Review and Meta-analysis of the Stop-signal Task. Neuropsychol Rev. 2024;34(2):548–67.

23. Cortese S, Aoki YY, Itahashi T, Castellanos FX, Eickhoff SB. Systematic Review and Meta-analysis: Resting-State Functional Magnetic Resonance Imaging Studies of Attention-Deficit/Hyperactivity Disorder. J Am Acad Child Adolesc Psychiatry. 2021;60(1):61–75.

24. LUDERS E, NARR KL, THOMPSON PM, TOGA AW. Neuroanatomical correlates of intelligence: Intelligence and the Brain. Intell Norwood. 2009;37(2):156–63.

25. Mooney MA, Hermosillo RJM, Feczko E, Miranda-Dominguez O, Moore LA, Perrone A, et al. Cumulative Effects of Resting-State Connectivity Across All Brain Networks Significantly Correlate with Attention-Deficit Hyperactivity Disorder Symptoms. J Neurosci. 2024;44(10):e1202232023.

26. Norman LJ, Carlisi C, Lukito S, Hart H, Mataix-Cols D, Radua J, et al. Structural and

Functional Brain Abnormalities in Attention-Deficit/Hyperactivity Disorder and Obsessive-Compulsive Disorder: A Comparative Meta-analysis. JAMA Psychiatry Chic Ill. 2016;73(8):815–25.

27. Seidman LJ, Valera EM, Makris N. Structural Brain Imaging of Attention-Deficit/Hyperactivity Disorder. Biol Psychiatry 1969. 2005;57(11):1263–72.

28. Samea F, Soluki S, Nejati V, Zarei M, Cortese S, Eickhoff SB, et al. Brain alterations in children/adolescents with ADHD revisited: A neuroimaging meta-analysis of 96 structural and functional studies. Neurosci Biobehav Rev. 2019;100:1–8.

29. Hoogman M, Muetzel R, Guimaraes JP, Shumskaya E, Mennes M, Zwiers MP, et al. Brain Imaging of the Cortex in ADHD: A Coordinated Analysis of Large-Scale Clinical and Population-Based Samples. Am J Psychiatry. 2019 Jul 1;176(7):531–42.

30. Emond V, Joyal C, Poissant H. [Structural and functional neuroanatomy of attention-deficit hyperactivity disorder (ADHD)]. L'Encephale. 2009 Apr;35(2):107–14.

31. Dickstein SG, Bannon K, Castellanos FX, Milham MP. The neural correlates of attention deficit hyperactivity disorder: an ALE meta-analysis. J Child Psychol Psychiatry. 2006 Oct;47(10):1051–62.

32. Jernigan TL, Brown SA, Dale AM, Tapert S, Sowell ES, Herting M, et al. Adolescent Brain Cognitive Development Study (ABCD) - Annual Release 5.1 [Internet]. Adolescent Brain Cognitive Development (ABCD) Study; 2024. Available from: https://wiki.abcdstudy.org/release-notes/start-page.html

33. Lytle MN, Hammer R, Booth JR. A neuroimaging dataset on working memory and reward processing in children with and without ADHD. Data Brief. 2020;31:105801.

34. Engemann DA, Kozynets O, Sabbagh D, Lemaître G, Varoquaux G, Liem F, et al. Combining magnetoencephalography with magnetic resonance imaging enhances learning of surrogate-biomarkers. eLife. 2020;9.

35. Clark DB, Fisher CB, Bookheimer S, Brown SA, Evans JH, Hopfer C, et al. Biomedical ethics and clinical oversight in multisite observational neuroimaging studies with children and adolescents: The ABCD experience. Dev Cogn Neurosci. 2018;32:143–54.

36. Cordova MM, Antovich DM, Ryabinin P, Neighbor C, Mooney MA, Dieckmann NF, et al. Attention-Deficit/Hyperactivity Disorder: Restricted Phenotypes Prevalence, Comorbidity, and Polygenic Risk Sensitivity in the ABCD Baseline Cohort. J Am Acad Child Adolesc Psychiatry. 2022;61(10):1273–84.

37. American Psychiatric Association issuing body. Diagnostic and statistical manual of mental disorders : DSM-5-TR. 5th edition, text revision. Washington, DC: American Psychiatric Association Publishing; 2022. (DSM-5-TR).

38. KAUFMAN J, BIRMAHER B, BRENT D, RAO U, FLYNN C, MORECI P, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial Reliability and Validity Data. J Am Acad Child Adolesc Psychiatry. 1997;36(7):980–8.

39. Achenbach TM, Rescorla LA. Brief Problem Monitor–Teacher Form (BPM-T) for Ages 6-18 [Internet]. Burlington, VT: ASEBA; 2001. Available from: https://aseba.org/

40. Achenbach TM, Ruffle TM. The Child Behavior Checklist and Related Forms for Assessing Behavioral/Emotional Problems and Competencies. Pediatr Rev. 2000;21(8):265–71.

41. Pappas D. Test Reviews: ADHD Rating Scale-IV: Checklists, Norms, and Clinical Interpretation. J Psychoeduc Assess. 2006;24(2):172–8.

42. Hammer R, Cooke GE, Stein MA, Booth JR. Functional neuroimaging of visuospatial working memory tasks enables accurate detection of attention deficit and hyperactivity disorder. NeuroImage Clin. 2015;9(C):244–52.

43. Bauer PJ, Zelazo PD. NIH TOOLBOX COGNITION BATTERY (CB): SUMMARY, CONCLUSIONS, AND IMPLICATIONS FOR COGNITIVE DEVELOPMENT. Monogr Soc Res Child Dev. 2013;78(4):133–46.

44. Wechsler D. Abbreviated scale of intelligence. San Antonio TX Psychol Corp. 1999;

45. Yang R, Jernigan T. Adolescent Brain Cognitive Development Study (ABCD)—Annual Release 4.0 [Internet]. NIMH Data Repositories; Available from: https://nda.nih.gov/edit_collection.html?id=2573

46. ABCD-BIDS Community. ABCD-BIDS Community Collection (ABCC) [Internet]. NIMH Data Repositories; Available from: https://nda.nih.gov/edit_collection.html?id=3165

47. Hagler DJ, Cornejo MD, Makowski C, Dick AS, Sutherland MT, Harms MP, et al. Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. NeuroImage Orlando Fla. 2019;202:116091–116091.

48. Barch DM, Burgess GC, Harms MP, Petersen SE, Schlaggar BL, Corbetta M, et al. Function in the human connectome: Task-fMRI and individual differences in behavior. NeuroImage Orlando Fla. 2013;80:169–89.

49. Knutson B, Westdorp A, Kaiser E, Hommer D. FMRI Visualization of Brain Activity during a Monetary Incentive Delay Task. NeuroImage Orlando Fla. 2000;12(1):20–7.

50. Whelan R, Conrod PJ, Poline JB, Lourdusamy A, Banaschewski T, Barker GJ, et al. Adolescent impulsivity phenotypes characterized by distinct brain networks. Nat Neurosci. 2012;15(6):920–5.

51. Elliott ML, Knodt AR, Cooke M, Kim MJ, Melzer TR, Keenan R, et al. General functional connectivity: Shared features of resting-state and task fMRI drive reliable and heritable individual differences in functional brain networks. NeuroImage Orlando Fla. 2019;189:516–32.

52. Cole MW, Bassett DS, Power JD, Braver TS, Petersen SE. Intrinsic and Task-Evoked Network Architectures of the Human Brain. Neuron Camb Mass. 2014;83(1):238–51.

53. Cole MW, Ito T, Schultz D, Mill R, Chen R, Cocuzza C. Task activations produce

spurious but systematic inflation of task functional connectivity estimates. NeuroImage Orlando Fla. 2019;189:1–18.

54. Hagler Jr DJ, Ahmadi ME, Kuperman J, Holland D, McDonald CR, Halgren E, et al. Automated white-matter tractography using a probabilistic diffusion tensor atlas: Application to temporal lobe epilepsy. Hum Brain Mapp. 2009;30(5):1535–47.

55. Hammer R, Tennekoon M, Cooke GE, Gayda J, Stein MA, Booth JR. Feedback associated with expectation for larger-reward improves visuospatial working memory performances in children with ADHD. Dev Cogn Neurosci. 2015;14(C):38–49.

56. Cawley GC, Talbot NLC. On over-fitting in model selection and subsequent selection bias in performance evaluation. J Mach Learn Res. 2010;11:2079–107.

57. Mateos-Aparicio G. Partial Least Squares (PLS) Methods: Origins, Evolution, and Application to Social Sciences. Commun Stat Theory Methods. 2011;40(13):2305–17.

58. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. ElasticNet - Scikit-learn documentation [Internet]. 2024. Available from: https://scikit-learn.org/1.5/modules/generated/sklearn.linear_model.ElasticNet.html

59. Breiman L. Random Forests. Mach Learn. 2001 Oct 1;45(1):5–32.

60. National Institute of Standards and Technology (NIST). Hedges' g Effect Size [Internet]. Available from: https://www.itl.nist.gov/div898/software/dataplot/refman2/auxillar/hedgeg.htm

61. Bifarin OO. Interpretable machine learning with tree-based shapley additive explanations: Application to metabolomics datasets for binary classification. Public Library of Science; 2023 p. e0284315. Report No.: 1932–6203.

62. Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. 2017;

63. Frazier TW, Demaree HA, Youngstrom EA, Becker JT. Meta-Analysis of Intellectual and Neuropsychological Test Performance in Attention-Deficit/Hyperactivity Disorder. Neuropsychology. 2004;18(3):543–55.

64. Pievsky MA, McGrath RE. The Neurocognitive Profile of Attention-Deficit/Hyperactivity Disorder: A Review of Meta-Analyses. Arch Clin Neuropsychol. 2018;33(2):143–57.

65. East-Richard C, R-Mercier A, Nadeau D, Cellard C. Transdiagnostic Neurocognitive Deficits in Psychiatry: A Review of Meta-Analyses. Can Psychol Psychol Can. 2020;61(3):190–214.

66. McTeague LM, Goodkind MS, Etkin A. Transdiagnostic impairment of cognitive control in mental illness. J Psychiatr Res. 2016;83:37–46.

67. Rosenberg MD, Finn ES, Scheinost D, Papademetris X, Shen X, Constable RT, et al. A neuromarker of sustained attention from whole-brain functional connectivity. Nat Neurosci. 2016;19(1):165–71.

68. Litwińczuk MC, Trujillo-Barreto N, Muhlert N, Cloutman L, Woollams A. Relating Cognition to both Brain Structure and Function: A Systematic Review of Methods. Brain Connect. 2023 Apr 1;13(3):120–32.

69. Zimmermann J, Griffiths JD, McIntosh AR. Unique Mapping of Structural and Functional Connectivity on Cognition. J Neurosci. 2018;38(45):9658–67.

70. Mather M, Cacioppo JT, Kanwisher N. How fMRI Can Inform Cognitive Theories. Perspect Psychol Sci. 2013 Jan 1;8(1):108–13.

71. Sripada C, Angstadt M, Rutherford S, Taxali A, Shedden K. Toward a "treadmill test" for cognition: Improved prediction of general cognitive ability from the task activated brain. Hum Brain Mapp. 2020;41(12):3186–97.

72. Owen AM, McMillan KM, Laird AR, Bullmore E. N-back working memory paradigm: A meta-analysis of normative functional neuroimaging studies. Hum Brain Mapp. 2005;25(1):46–59.

73. Stevens WD, Spreng RN. Resting-state functional connectivity MRI reveals active processes central to cognition. Wiley Interdiscip Rev Cogn Sci. 2014;5(2):233–45.

74. Power JD, Cohen AL, Nelson SM, Wig GS, Barnes KA, Church JA, et al. Functional Network Organization of the Human Brain. Neuron Camb Mass. 2011;72(4):665–78.

75. Bush G, Luu P, Posner MI. Cognitive and emotional influences in anterior cingulate cortex. Trends Cogn Sci. 2000 Jun 1;4(6):215–22.

76. Osher DE, Brissenden JA, Somers DC. Predicting an individual's dorsal attention network activity from functional connectivity fingerprints. J Neurophysiol. 2019;122(1):232–40.

77. Dosenbach NUF, Fair DA, Cohen AL, Schlaggar BL, Petersen SE. A dual-networks architecture of top-down control. Trends Cogn Sci. 2008;12(3):99–105.

78. Castellanos FX, Proal E. Large-scale brain systems in ADHD: beyond the prefrontal–striatal model. Trends Cogn Sci. 2012 Jan 1;16(1):17–26.

79. Shulman GL, Astafiev SV, Franke D, Pope DLW, Snyder AZ, McAvoy MP, et al. Interaction of stimulus-driven reorienting and expectation in ventral and dorsal frontoparietal and basal ganglia-cortical networks. J Neurosci Off J Soc Neurosci. 2009 Apr 8;29(14):4392–407.

80. Arbabshirani MR, Plis S, Sui J, Calhoun VD. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. Individ Subj Predict. 2017 Jan 15;145:137–65.

81. Glasser MF, Coalson TS, Robinson EC, Hacker CD, Harwell J, Yacoub E, et al. A multi-modal parcellation of human cerebral cortex. Nat Lond. 2016;536(7615):171–8.

82. Cole MW, Anticevic A, collaborators. ColeAnticevicNetPartition: Network partition for the human brain. 2024; Available from: https://github.com/ColeLab/ColeAnticevicNetPartition

83. Ji JL, Spronk M, Kulkarni K, Repovš G, Anticevic A, Cole MW. Mapping the human brain's cortical-subcortical functional network organization. NeuroImage Orlando Fla. 2019;185:35–57.

Figure legends

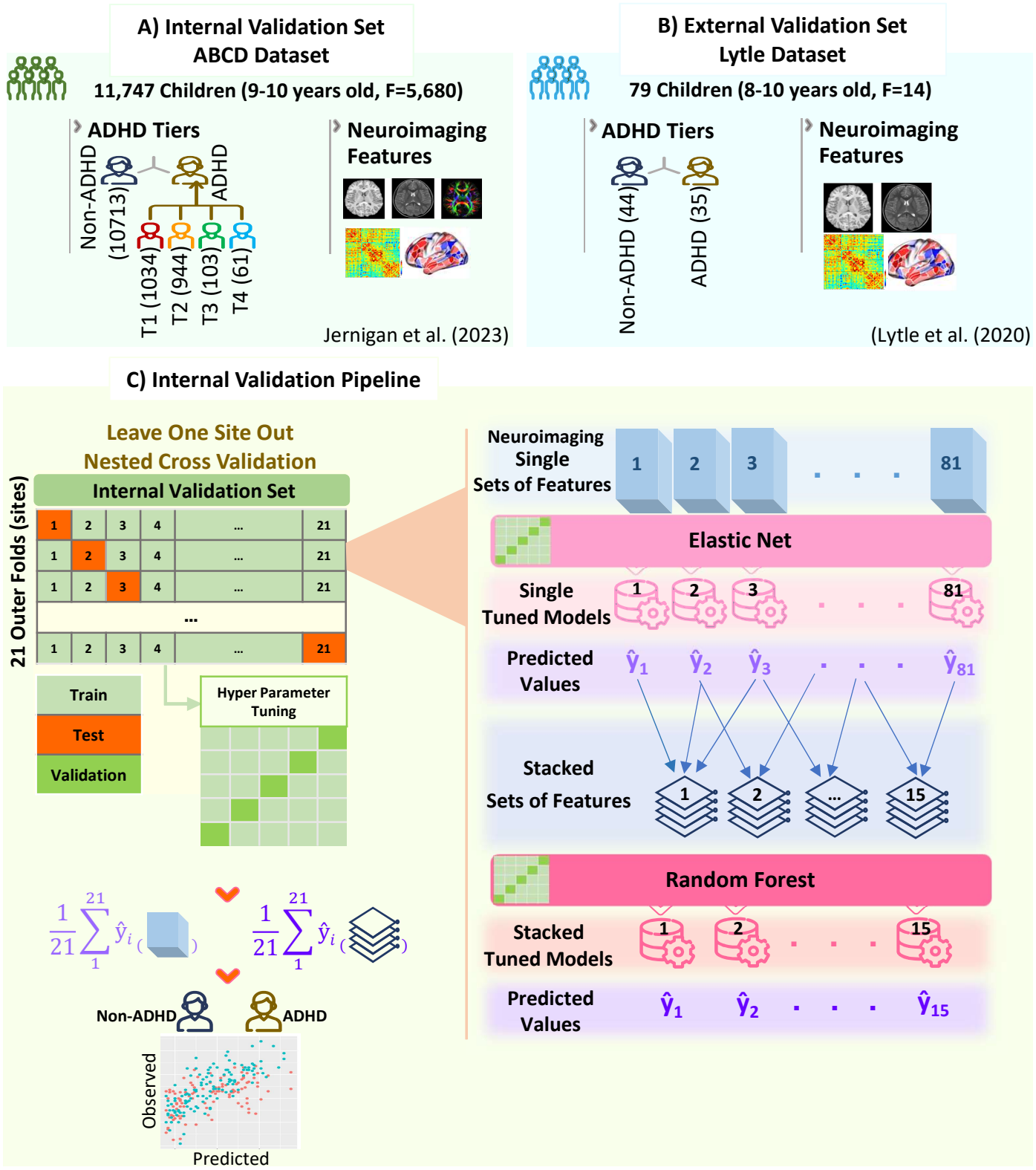
Figure 1 Overall design and information about participants. (**A**) and (**B**) indicate the data we extracted from the ABCD (i.e., internal validation dataset) and Lytle and colleagues' (i.e., external validation dataset). For the tier definitions of ADHD in the ABCD, we followed Cordova and colleagues(36). Specifically, ADHD Tier 1 included current ADHD based on parent-reported Kiddie Schedule for Affective Disorders and Schizophrenia – Computerized Version (K-SADS-COMP)(38), Tier 2 excluded cases with bipolar disorder, schizophrenia spectrum, or intellectual disability, Tier 3 filtered based on the teacher-rated Brief Problem Monitoring scale (BPM)(39) to confirm symptoms in multiple settings, and Tier 4 applied the clinical cutoff from the Child Behavior Checklist (CBCL) attention or ADHD DSM-5 scale(40). (**C**) shows our pipeline for building the cognitive prediction models and internally validating their predictive performance. (**D**) shows our pipeline for externally validating the predictive performance of our models built from the ABCD. Note that not all sets of neuroimaging features provided by the ABCB are available in Lytle and colleagues' study. As a result when conducting external validation, we could only apply the models trained from the ABCD to 13 single and nine stacked models, as opposed to 81 single and 15 stacked models used in our internal validation.

Figure 2 The distribution of observed and predicted cognitive functioning scores for the (**A**) internal and (**B**) external validation sets. The predicted scores are derived from the stacked model with the highest overall predictive accuracy based on Pearson's r. The asterisks indicate statistically significant differences between non-ADHD and each ADHD group based on a 95% bootstrapped confidence interval for the mean differences. Hg is the magnitude of the differences calculated using Hedges' g.

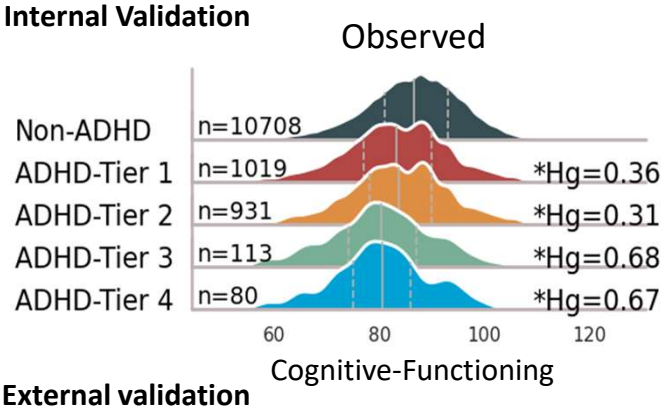
Figure 3 Predictive performance of stacked models across groups in (**A**) internal and (**B**) external validation datasets, ranked by performance in non-ADHD groups. The left panel shows Pearson's correlation between predicted and observed values for cognitive functioning scores averaged across sites. The bars show bootstrapped 95%*CI*. The right panel shows the bootstrapped confidence interval for differences in model performance between non-ADHD and each ADHD group. Intervals below/above zero indicate statistically significantly worse/better performance in ADHD groups at 95%*CI*, respectively. The highlighted colors indicate which stacked models from the internal-validation set(s) were applied to which stacked sets of features in the external-validation set(s). Five stacked models from the internal validation set were applied to the external validation set, including ABCC sMRI, WM load contrasts from the Nback task and their combination with Nback, multitask or general FCs. The four stacked models that included task-related features were applied separately to verbal and spatial Nback tasks in the external-validation set, yielding in total nine stacked models for the external-validation set. The G- prefix for multitask FC refers to the models trained on general FC in the ABCD that were then applied to multitask FC in the external-validation set.

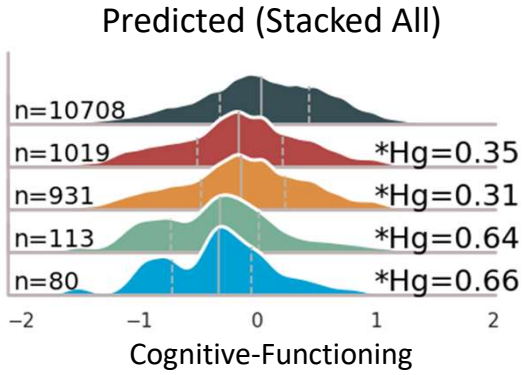
Figure 4 The scatterplots of the correlation between observed and predicted cognitive functioning scores for the best stacked model and the top-performing contrast-based, FC-based, and sMRI-DTI-based single models, in (**A**) internal and(**B**) external validation sets. Glasser Nback 2-0back indicates fMRI task contrast from the Emotional Nback task, which was parcellated using the Glasser atlas(81). Glasser MID FC indicates task FC during the Monetary Incentive Delay (MID) task, which was parcellated using the Glasser atlas(81).

Figure 5 Feature Importance of stacked and single models. (**A**) Feature importance of the stacked all model that included all sets of neuroimaging features. This feature importance is based on absolute SHAP values, averaged across test sites in the ABCD. A higher absolute SHAP value indicates a higher contribution to the prediction. Error bars reflect standard deviations across sites. (**B**) Feature importance of the best performing model among FC models: the MID FC. - Feature importance of FC is based on the weight of each feature in the partial least square (PLS) components and the amount of variance explained by each component average across sites, visualized using the Cole-Anticevic Brain-wide Network Partition(82,83). (**C**) Feature importance of the best-performing model among task contrasts: the Nback 2-0back model. Feature importance of task contrast is based on Elastic Net coefficients, averaged across sites. (**D**) Feature importance for the best model among sMRI and DTI: T2 Grey Matter Intensity. Feature importance of sMRI and DTI is based on Elastic Net coefficients, averaged across sites.

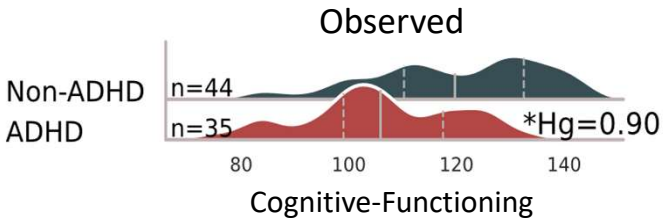

A) Internal Validation Set
ABCD Dataset
11,747 Children (9-10 years old, F=5,680)
ADHD Tiers
Non-ADHD (10713)
ADHD
T1 (1034)
T2 (944)
T3 (103)
T4 (61)
Neuroimaging Features
Jernigan et al. (2023)
B) External Validation Set
Lytle Dataset
79 Children (8-10 years old, F=14)
ADHD Tiers
Non-ADHD (44)
ADHD (35)
Neuroimaging Features
(Lytle et al. (2020)
C) Internal Validation Pipeline
Leave One Site Out
Nested Cross Validation
Internal Validation Set
21 Outer Folds (sites)
Train
Test
Validation
Hyper Parameter Tuning
Neuroimaging Single Sets of Features
Elastic Net
Single Tuned Models
Predicted Values
Stacked Sets of Features
Random Forest
Stacked Tuned Models
Predicted Values
Non-ADHD
ADHD
Observed
Predicted
D) External Validation Pipeline
Train
Whole Internal Validation Set
10 out of 81 applicable Single Tuned Models
5 out of 15 applicable Stacked Tuned Models
Test
Whole External Validation Set
13 Single
9 Stacked
Non-ADHD
ADHD
Observed
Predicted

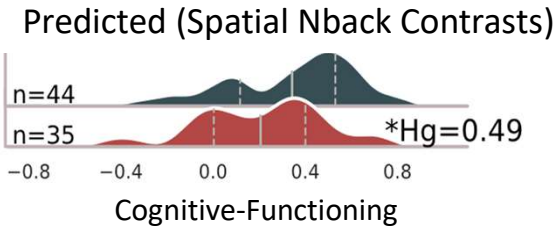
A) Internal Validation

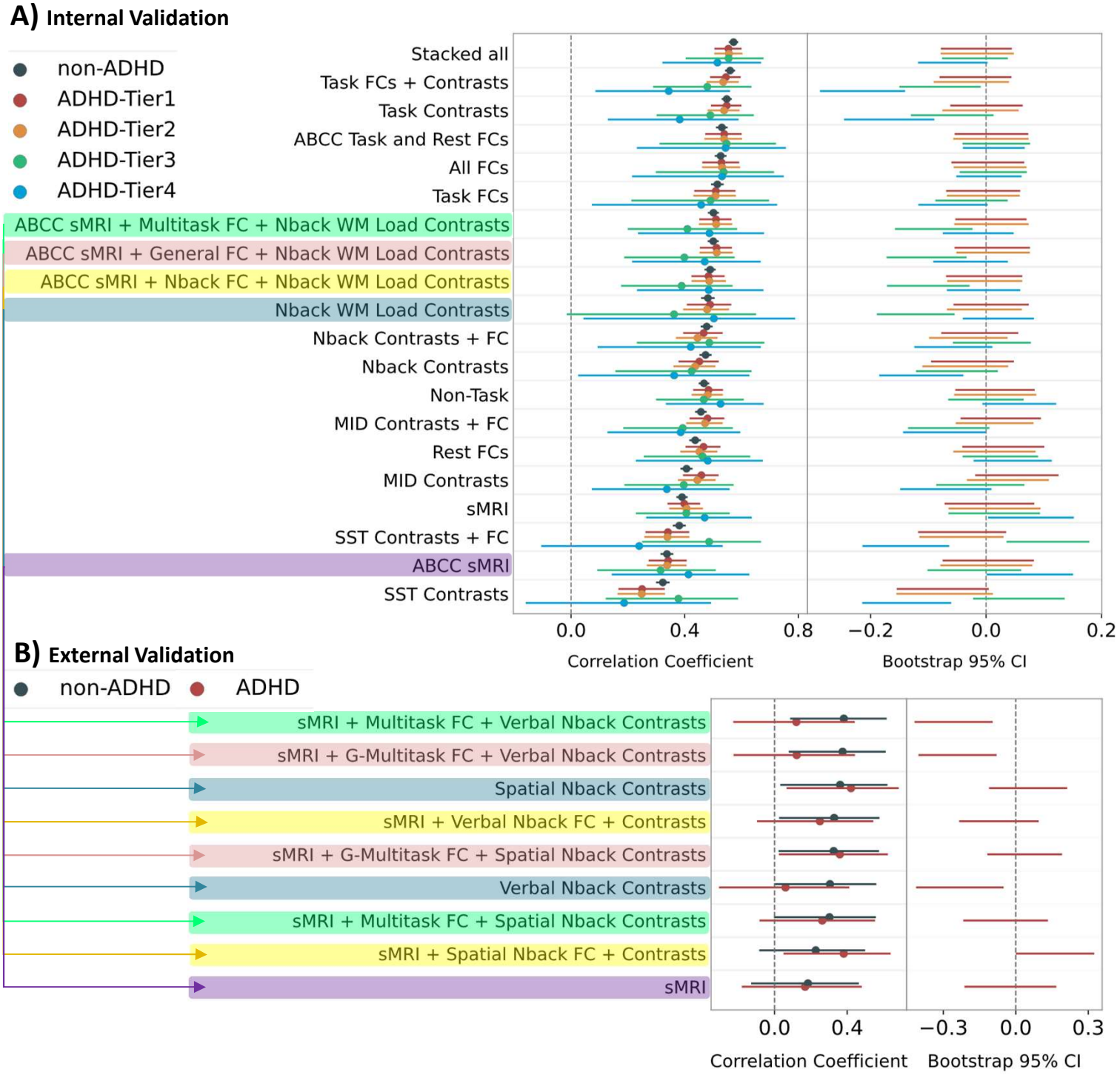


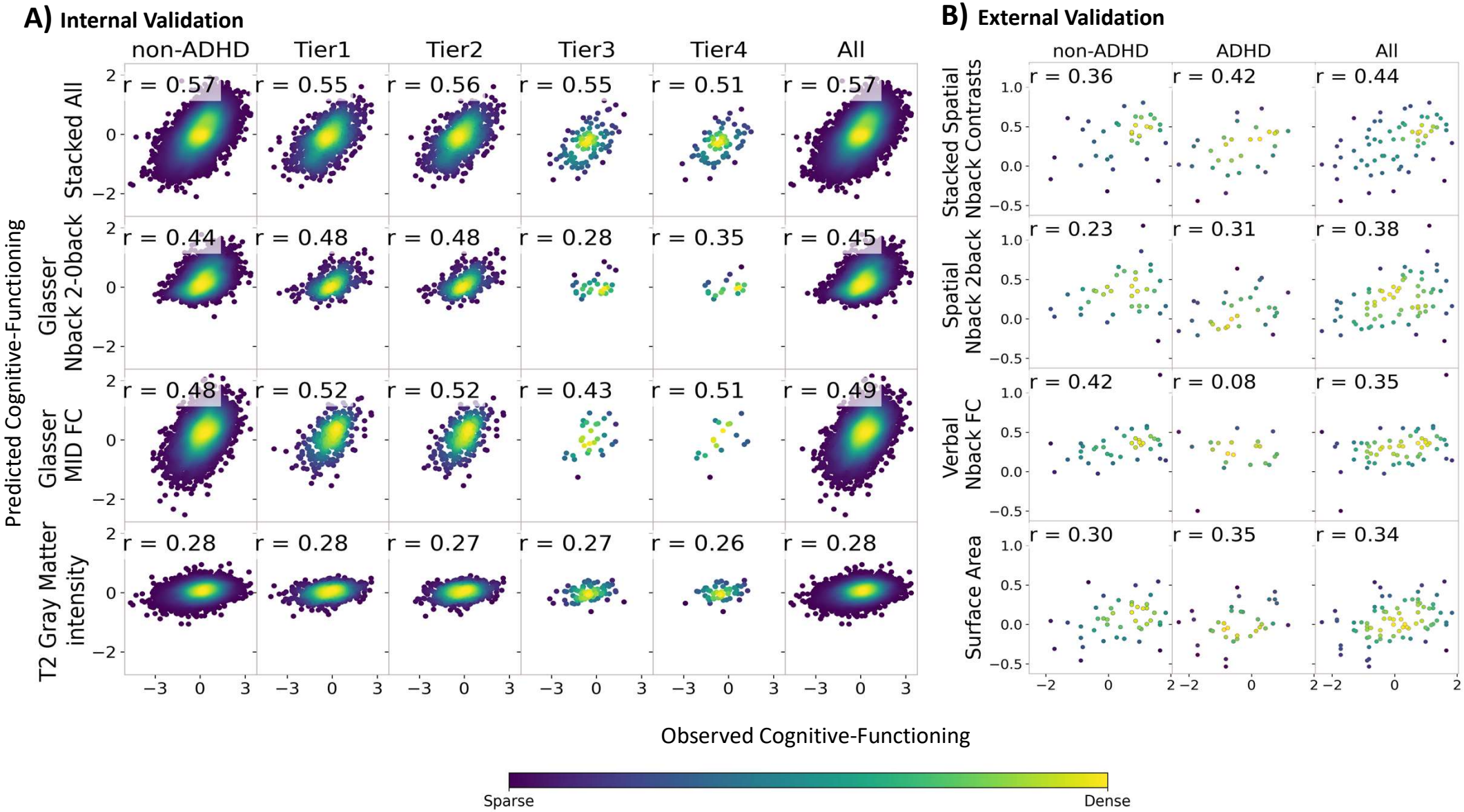


B) External validation

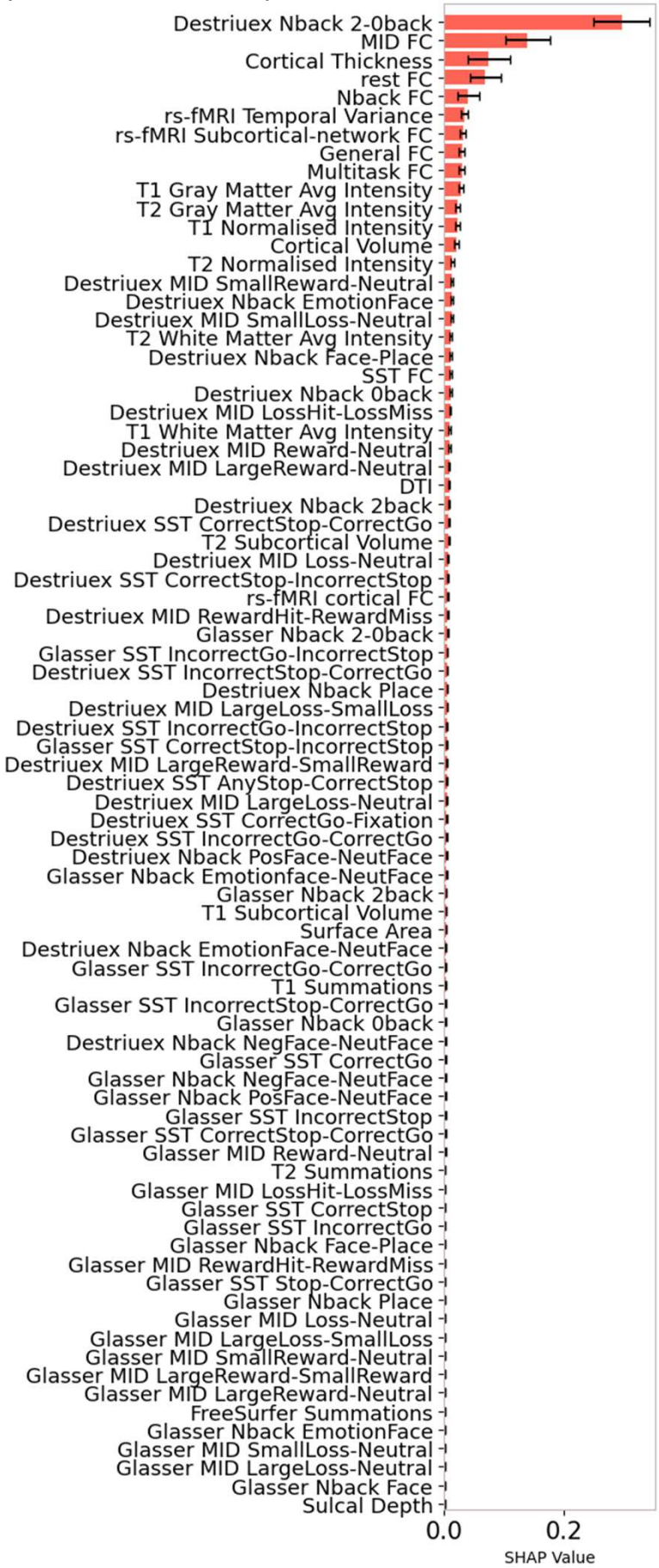




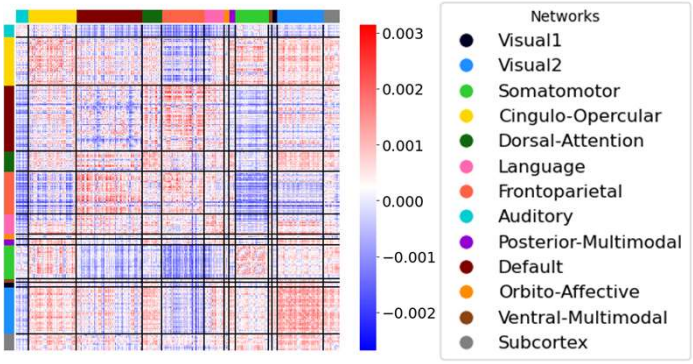

A) Internal Validation
non-ADHD
ADHD-Tier1
ADHD-Tier2
ADHD-Tier3
ADHD-Tier4
Stacked all
Task FCs + Contrasts
Task Contrasts
ABCC Task and Rest FCs
All FCs
Task FCs
ABCC sMRI + Multitask FC + Nback WM Load Contrasts
ABCC sMRI + General FC + Nback WM Load Contrasts
ABCC sMRI + Nback FC + Nback WM Load Contrasts
Nback WM Load Contrasts
Nback Contrasts + FC
Nback Contrasts
Non-Task
MID Contrasts + FC
Rest FCs
MID Contrasts
sMRI
SST Contrasts + FC
ABCC sMRI
SST Contrasts
0.0
0.4
0.8
−0.2
0.0
0.2
Correlation Coefficient
Bootstrap 95% CI
B) External Validation
non-ADHD
ADHD
sMRI + Multitask FC + Verbal Nback Contrasts
sMRI + G-Multitask FC + Verbal Nback Contrasts
Spatial Nback Contrasts
sMRI + Verbal Nback FC + Contrasts
sMRI + G-Multitask FC + Spatial Nback Contrasts
Verbal Nback Contrasts
sMRI + Multitask FC + Spatial Nback Contrasts
sMRI + Spatial Nback FC + Contrasts
sMRI
0.0
0.4
−0.3
0.0
0.3
Correlation Coefficient
Bootstrap 95% CI


A) Internal Validation
non-ADHD
Tier1
Tier2
Tier3
Tier4
All
Stacked All
Glasser Nback 2-0back
Glasser MID FC
T2 Gray Matter intensity
Predicted Cognitive-Functioning
r = 0.57
r = 0.55
r = 0.56
r = 0.55
r = 0.51
r = 0.57
r = 0.44
r = 0.48
r = 0.48
r = 0.28
r = 0.35
r = 0.45
r = 0.48
r = 0.52
r = 0.52
r = 0.43
r = 0.51
r = 0.49
r = 0.28
r = 0.28
r = 0.27
r = 0.27
r = 0.26
r = 0.28
B) External Validation
non-ADHD
ADHD
All
Stacked Spatial Nback Contrasts
Spatial Nback 2back
Verbal Nback FC
Surface Area
r = 0.36
r = 0.42
r = 0.44
r = 0.23
r = 0.31
r = 0.38
r = 0.42
r = 0.08
r = 0.35
r = 0.30
r = 0.35
r = 0.34
Observed Cognitive-Functioning
Sparse
Dense

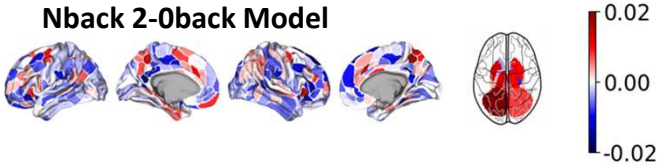
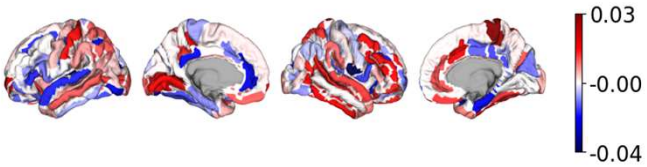
A) Feature Importance of Stacked All Model (Absolute SHAP Values)



B) Feature Importance of MID FC Model



C) Feature Importance of Nback 2-0back Model



D) Feature Importance of T2 Grey Matter Intensity Model