Classifying Conduct Disorder Using a Biopsychosocial Model and Machine Learning Method

Supplemental Information

Supplemental Methods and Materials

Participants

Participants were youth who completed the baseline session (at ages 9-10-years-old) and the 2-year follow-up session (ages 11-12-years-old) of the multisite ABCD Study. Details regarding the sampling strategy, sample norms, and sample composition have been described elsewhere (1). All study procedures were approved by a centralized institutional review board at the University of California San Diego and/or by each site's institutional review board (2). Caregivers provided signed informed consent and children provided written assent prior to the study. For the present analyses, participants were included if they: (a) had Conduct Disorder (CD) data available from their baseline session, (b) were not missing any data for key variables, and (c) had valid rs-fMRI data released from their baseline session that also passed the ABCD Study overall MRI quality checks (3). Further, given the large number of ABCD Study families with multiple children and/or twins that participated in the study, siblings were overrepresented in the sample (4). To help control for any family-related effects, only one, randomly selected, child per family was used in the current analyses, yielding a final sample of n = 2,368 (see Table 1 in the main text).

Model Training

The neurons of the hidden layer were calculated as follows (5):

$$y_j = f_H\left(\sum_{i=1}^{N_I} w_1(i,j)x_i + b_1\right) \text{ for } j = 1,2,\ldots,N_H,$$

where x_i was the *i*th input neuron value, y_j was the *j*th hidden neuron value, N_I was the number of input neurons, N_H was the number of hidden neurons, f_H was the hidden layer activation function, and w_1 and b_1 were the weight matrix and the bias vector connecting the input layer to the hidden layer.

We adopted the standard approach for binary classification models of defining f_H as the sigmoid activation function (6). This activation function transformed inputs into values between 0 and 1 to amplify meaningful signals and suppress noise. The calculation of the sigmoid function f_H is displayed below:

$$f_H(x) = \frac{1}{1 + e^{(-x)}}$$

Linear and non-linear transformations were then conducted on the hidden neurons y_j to calculate values in the output layer. Output nodes were calculated as follows:

$$O_K = f_O(\sum_{j=1}^{N_H} w_2(j,k)y_j + b_2) \text{ for } k = 1,2,\ldots,N_O,$$

where y_j was the *j*th hidden neuron value, O_k was the *k*th output neuron value, N_O was the number of output neurons, f_o was the output layer activation function, and w_2 and b_2 were the weight matrix and the bias vector connecting the hidden layer to the output layer. The sigmoid activation function also was used for f_o.

The output of the model was the probability that a given participant met the criteria for CD. We adopted the standard classification threshold of .5 to convert probability outputs into binary classes of "CD" or "no CD." As a result, all outputs of probability ≥ .5 were classified as meeting for CD diagnosis, and all outputs of probability < .5 were classified as not meeting for CD diagnosis (7).

We trained weight and bias parameters with the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, an iterative method for solving non-linear optimization problems (8). The BFGS algorithm conducts backpropagation to calculate the gradient of the cross-entropy loss function with respect to weight and bias parameters. Each gradient is computed one layer at a time, iterating backward from the last layer to avoid redundant calculations of intermediate terms. This method increases the efficiency of model training and updates parameters until the loss function converges at its minimum. Random batches of data were processed at a time, providing stability during model training and reliable convergence (9).

As typical with binary classification models, the cross-entropy loss L_{CE} was used to optimize parameters (10). L_{CE} measures the dissimilarity between the actual class y and the predicted class $\hat{y}$. By selecting weight and bias terms that minimize L_{CE}, the BFGS algorithm maximizes the model's classification accuracy. L_{CE} was calculated as follows:

$$L_{CE}(\hat{y}, y) = -log(p(y|x)) = -[y * log(\hat{y}) + (1 - y) * log(1 - \hat{y})]$$

The BFGS algorithm updates parameters until L_{CE} converges at its minimum (11).

Model Performance Measures

Accuracy is a performance measure that quantifies the overall precision of a classifier. This metric describes the proportion of true positive and true negative predictions among all evaluated cases.

$$Accuracy = \frac{True\ Positive + True\ Negative}{True\ Positive + True\ Negative + False\ Positive + False\ Negative}$$

Sensitivity, or the true positive rate, is the proportion of participants with CD that are correctly classified with the CD label, and specificity, or the true negative rate, is the proportion of typically developing participants that are correctly classified as not having CD.

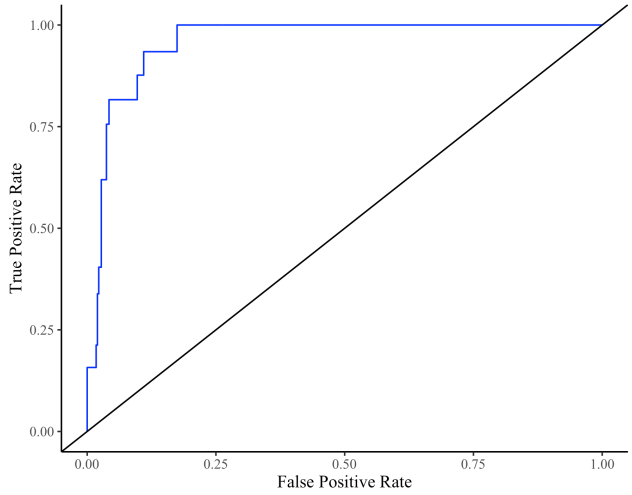
$$Sensitivity = true\ positive\ rate = \frac{True\ Positive}{True\ Positive + False\ Negative}$$

$$Specificity = true\ negative\ rate = \frac{True\ Negative}{True\ Negative + False\ Positive}$$

The strongest classifiers find an optimal balance between sensitivity and specificity. A model with high sensitivity may lack clinical relevance if it demonstrates low specificity and is biased towards true positive and false positive classifications. Conversely, a classifier may yield high specificity at the cost of low sensitivity. The receiver operating characteristic (ROC) curve plots sensitivity against 1 − specificity, also termed the false positive rate, at all possible classification thresholds. Thus, ROC curves that closely approach the left corner represent models that optimize true positive and true negative rates and maximize overall accuracy. The area under the ROC curve (AUC) is a metric that quantifies the balance between sensitivity and specificity and the overall diagnostic accuracy of the model. The AUC is

calculated via standard integration and adopts values between 0 and 1, where 0 indicates a perfectly inaccurate classifier and 1 reflects a perfectly accurate classifier. An AUC of .5 represents a ROC curve that falls on the diagonal line and displays no discriminatory ability (12). The ROC curve for the present model approaches the (0,1) point of perfect prediction, suggesting that this classifier performed well at maximizing sensitivity and specificity (Supplemental Figure S1).

Supplemental Figure S1. ROC Curve

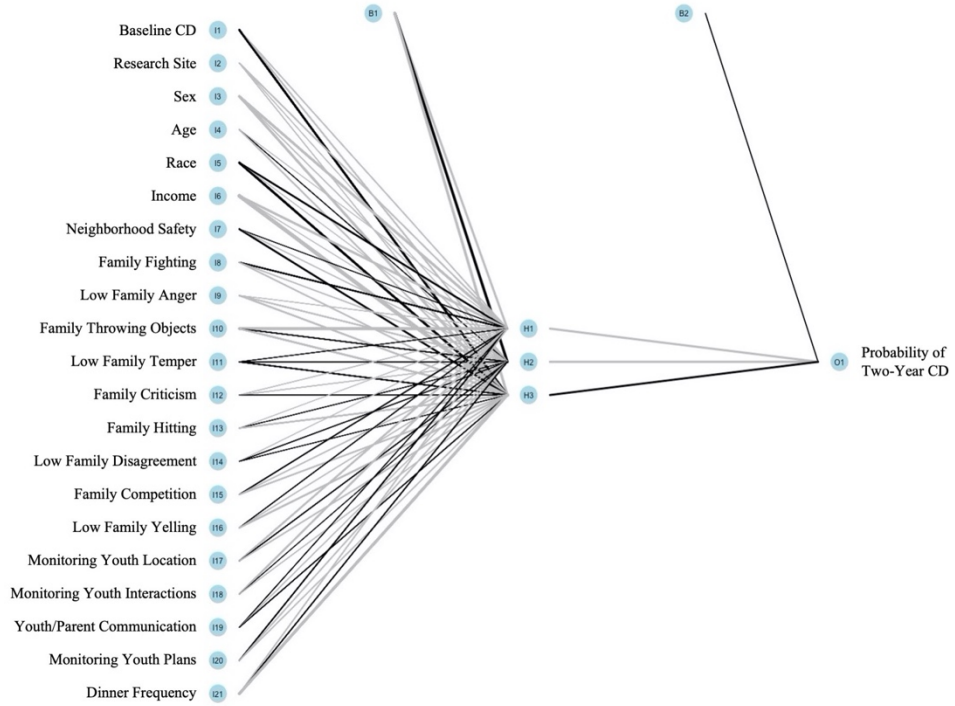


Note. The ROC curve displays the tradeoff between specificity and sensitivity at various classification thresholds. The coordinate (0,1) in the upper left corner represents a perfect classification with 100% sensitivity and specificity. The diagonal line depicts random predictions with 50% sensitivity and specificity that fail to meaningfully discriminate between classes.

Models within Risk Factor Domain

To assess the predictive ability of neighborhood and family risk factors, the social model was trained. Participants with missing neighborhood or family risk factors were excluded, yielding a final sample of 3,347 for training and testing. The architecture of the social model consisted of 21 input neurons, three hidden neurons, and one output neuron to calculate the probability that a participant met the criteria for 2-year CD (Supplemental Figure S2).

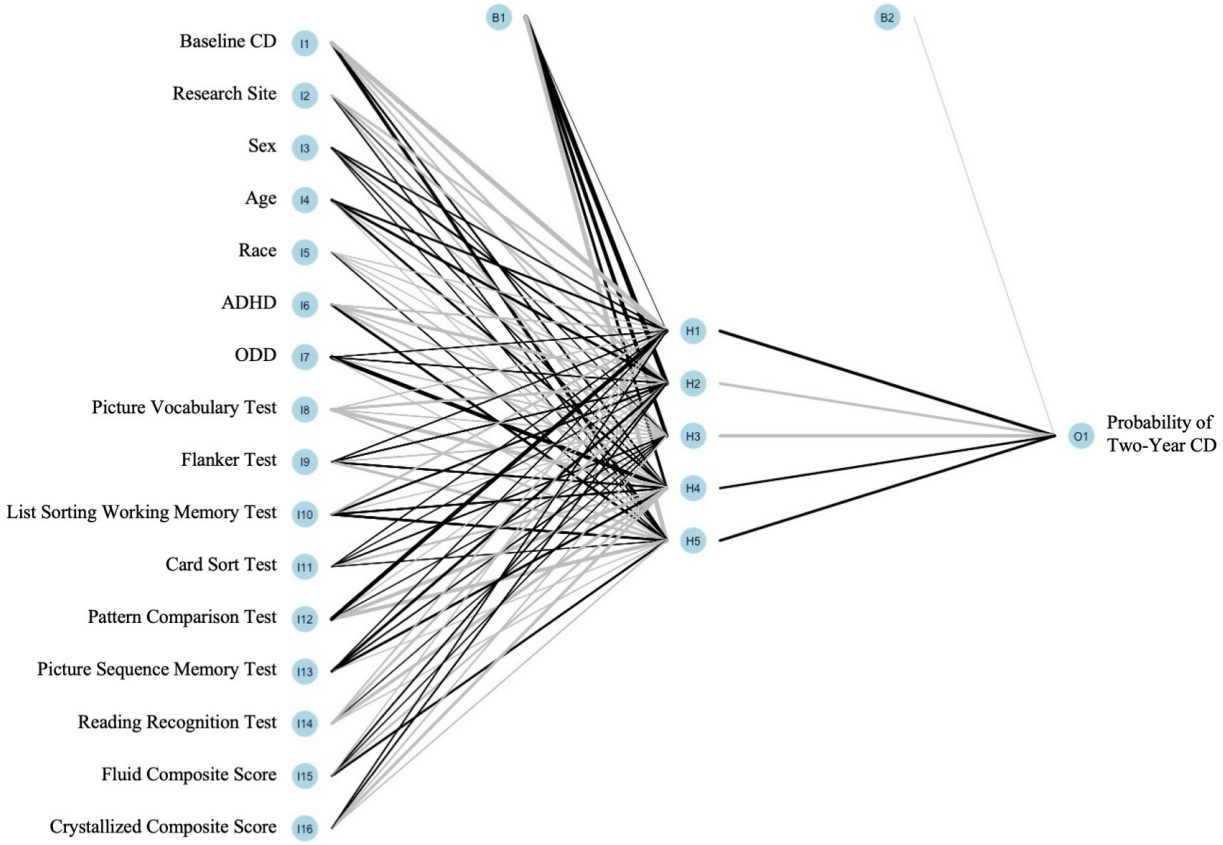
Supplemental Figure S2. Network Architecture for the Social Model



Note. The social model consists of 21 input neurons, three hidden neurons, and one output neuron. Line thickness is proportional to the magnitude of each weight and bias term. Black lines indicate positive parameters and grey lines indicate negative parameters.

For the within-domain psychological model (including ADHD, ODD, and neuropsychological indictors) a sample of 3,244 participants was available. We specified 17 input neurons, five hidden neurons, and one output neuron for this model (Supplemental Figure S3).

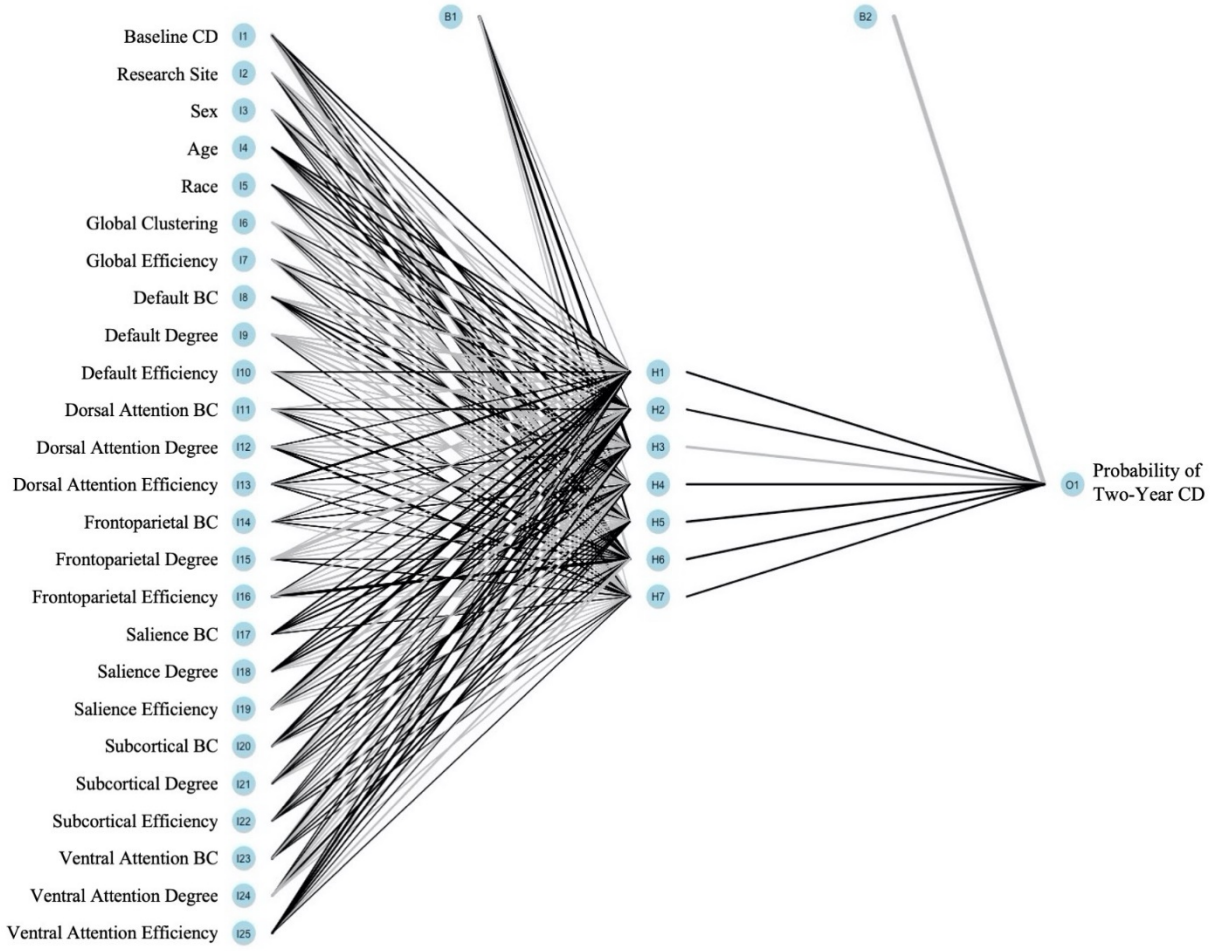
Supplemental Figure S3. Network Architecture for the Psychological Model



Note. The psychological model consists of 17 input neurons, five hidden neurons, and one output neuron. Line thickness is proportional to the magnitude of each weight and bias term. Black lines indicate positive parameters and grey lines indicate negative parameters.

To assess the predictive ability of resting-state brain topography, the biological model was trained on 2,867 participants. We determined the optimal architecture of this model to contain 25 input neurons, seven hidden neurons, and one output neuron (Supplemental Figure S4).

Supplemental Figure S4. Network Architecture for the Biological Model



Note. The biological model consists of 25 input neurons, seven hidden neurons, and one output neuron. Line thickness is proportional to the magnitude of each weight and bias term. Black lines indicate positive parameters and grey lines indicate negative parameters.

Supplemental Results

Traditional Logistic Regression Model

One purported advantage of machine learning methods is to maximize prediction by improving on traditional analytical techniques (e.g., logistic regression). To test inferential statistical models, we ran a logistic regression model including all baseline variables to predict CD diagnosis at the 2-year follow-up session. The logistic regression model demonstrated an accuracy of 80.16%, an AUC of .8674, model

sensitivity of 88.28%, and model specificity of 82.10%. These results indicate that the FNN model outperformed the logistic regression model, improving classification by approximately 11%.

Model Fit for Classifying Other Psychiatric Disorders

CD and Oppositional Defiant Disorder (ODD) are considered externalizing disorders that share latent characteristics and premorbid risk factors (13, 14). To test whether the risk factors included in our FNN model capture risk factors related to CD specifically or externalizing disorders more broadly, we explored the ODD diagnosis at the 2-year follow-up session as the outcome instead of CD. We found that using the same risk factors as in the primary analysis, and controlling for baseline ODD symptomatology, the model predicting ODD achieved 73% accuracy, 64% sensitivity, and 81% specificity. Evidence of reasonable accuracy and specificity speaks to the shared risk factors between disorders, however the stronger metrics, particularly for sensitivity, for the CD model highlights how the risk factors used in our FNN model provided a stronger prediction of CD than ODD.

We also examined whether our FNN model was stronger for CD, an externalizing disorder, than other disorders typically considered on the internalizing spectrum. We conducted two additional analyses with the goals of predicting depression or anxiety at the 2-year follow-up session. One model included K-SADS Major Depressive Disorder diagnosis at the 2-year session as the outcome, all predictors from the main model, and research collection site, biological sex at birth (dichotomously coded, male vs. female), race (dichotomously coded, white vs. non-white), age, and baseline Major Depressive Disorder as covariates. This model achieved poor metrics with 63.68% prediction accuracy, 71.39% sensitivity, and 56.36% specificity. Another model included a composite score of meeting diagnostic criteria for any K-SADS Anxiety Disorder (panic disorder, agoraphobia, separation anxiety disorder, social phobia) at the 2-year session as the outcome, all predictors from the main model, and research collection site, biological sex at birth (dichotomously coded, male vs. female), race (dichotomously coded, white vs. non-white), age, and anxiety disorder diagnoses at baseline as covariates. This model also had poor metrics with 64.07% prediction accuracy, 58.01% sensitivity, and 69.83% specificity. Together, these results suggest

that the risk factors selected in the main model are most strongly and accurately predictive of CD and not ODD, Major Depressive Disorder, or anxiety disorders.

Replication of Feature Importance for the Biopsychosocial Model

Machine learning methods are advantageous because they model complex relationships and can identify specific predictors (features) that contribute the greatest variance to the model. However, there is debate about the degree to which weighted individual predictors should be interpreted given the complexity of the models and the inherent variation introduced into the models based on randomization (15). We re-ran our biopsychosocial model predicting CD at the 2-year follow-up session 10 times to examine the reliability of the top predictors identified by the feature importance analysis. In all 10 model iterations, greater ADHD and ODD symptomatology, greater reports of family members throwing objects, lower crystallized cognitive ability, and lower parental monitoring appeared in the top 10 features. In 90% of the model iterations, lower frontoparietal degree, lower card sorting ability, and lower subcortical efficiency were included in the top 10 features. In 70% of the model iterations, greater frontoparietal efficiency was in the top 10 features. Finally, in 60% of the model iterations, lower family-level income was present in the top 10 features. These findings indicate that many of the important features identified in the main analysis were reliably represented across model iterations.

Models within Risk Factor Domain

We compared each classifier's predictions against the known diagnostic status of participants at their 2-year follow-up assessment to construct confusion matrices (Supplemental Tables S1-S3).

Supplemental Table S1. Confusion Matrix for the Social Model

		Predicted Class	
		No CD	**CD**
True Class	**No CD**	490 (True Negative)	83 (False Positive)
	CD	163 (False Negative)	369 (True Positive)

Supplemental Table S2. Confusion Matrix for the Psychological Model

		Predicted Class	
		No CD	**CD**
True Class	**No CD**	517 (True Negative)	38 (False Positive)
	CD	239 (False Negative)	277 (True Positive)

Supplemental Table S3. Confusion Matrix for the Biological Model

		Predicted Class	
		No CD	**CD**
True Class	**No CD**	374 (True Negative)	117 (False Positive)
	CD	101 (False Negative)	355 (True Positive)

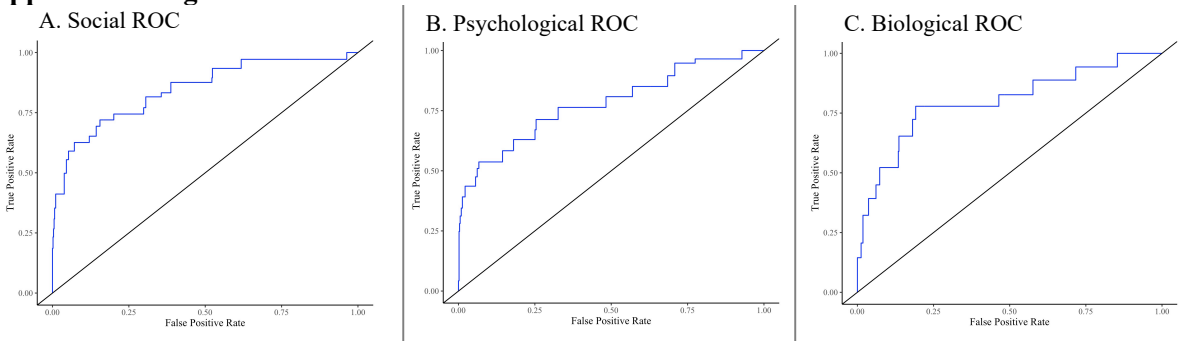
We calculated accuracy, sensitivity, specificity, and AUC measures for each classifier to compare performance across the different models (Supplemental Table S4).

Supplemental Table S4. Performance Measures

Model	Accuracy	Sensitivity	Specificity	AUC
Social	77.74%	69.36%	85.51%	. 8450
Psychological	74.14%	53.68%	93.15%	. 7796
Biological	76.98%	77.85%	76.17%	. 7990

ROC curves were generated to assess how each classifier maximized sensitivity and specificity rates (Supplemental Figure S5).

Supplemental Figure S5. ROC Curves



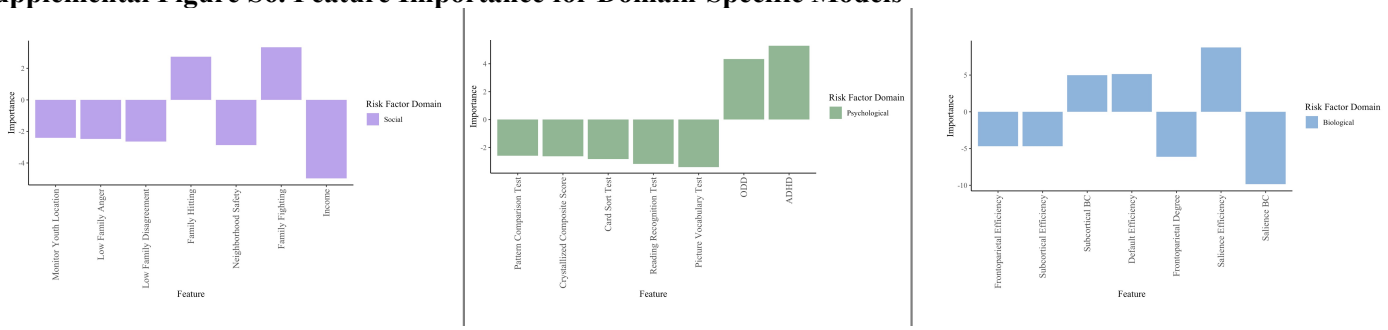
Note. The ROC curves display the tradeoff between specificity and sensitivity at various classification thresholds. The ROC curve for our model balances specificity and sensitivity by maximizing true positive rates and minimizing false positive rates.

Models within Domain: Feature Importance and Sensitivity

In the social model, income, neighborhood safety, and parental monitoring negatively correlated with CD. In contrast, family fighting, hitting, disagreement, and anger all positively predicted CD. In the psychological model, ADHD and ODD positively predicted CD. Conversely, low performance on the picture vocabulary, reading recognition, card sort, and pattern comparison tasks and reduced crystallized cognitive ability predicted CD. Lastly, in the biological model, betweenness centrality, degree, and efficiency measures in the default, frontoparietal, salience, and subcortical networks were the strongest predictors of CD. Greater efficiency in the default and salience networks and greater subcortical betweenness centrality predicted CD. Reduced efficiency in the frontoparietal and subcortical networks,

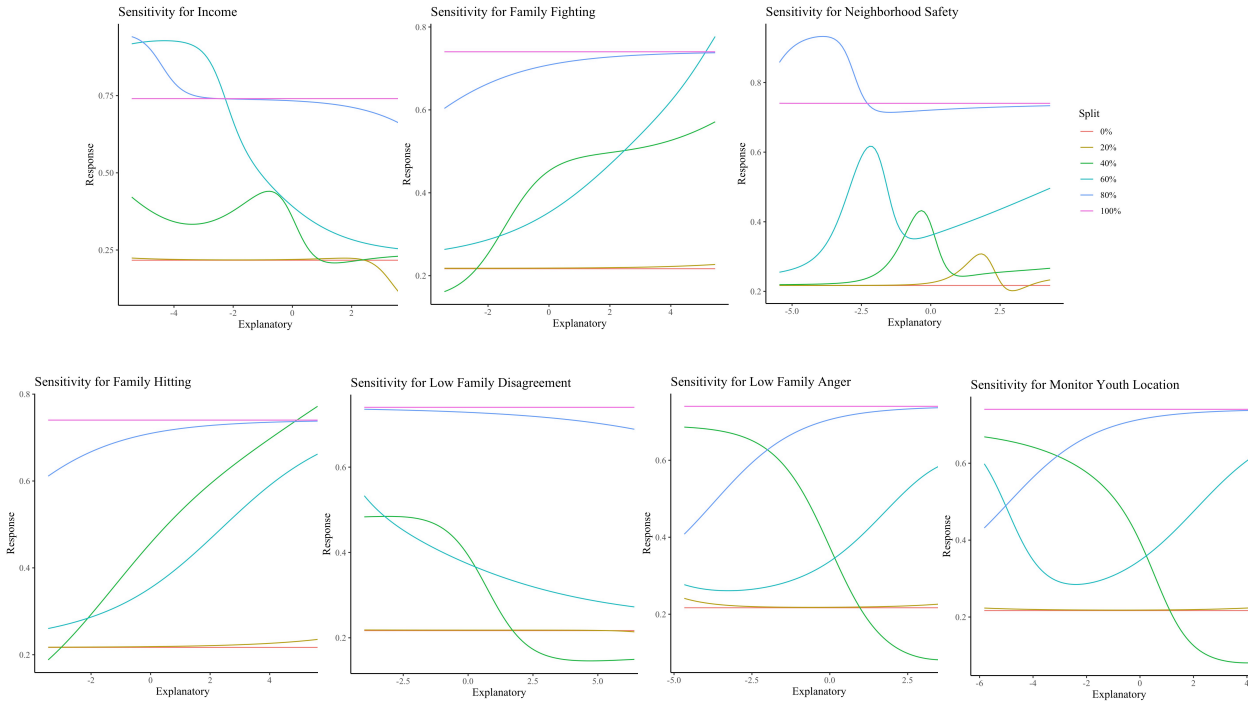
as well as lower salience betweenness centrality and frontoparietal degree correlated with CD (Supplemental Figures S6-S9).

Supplemental Figure S6. Feature Importance for Domain-Specific Models



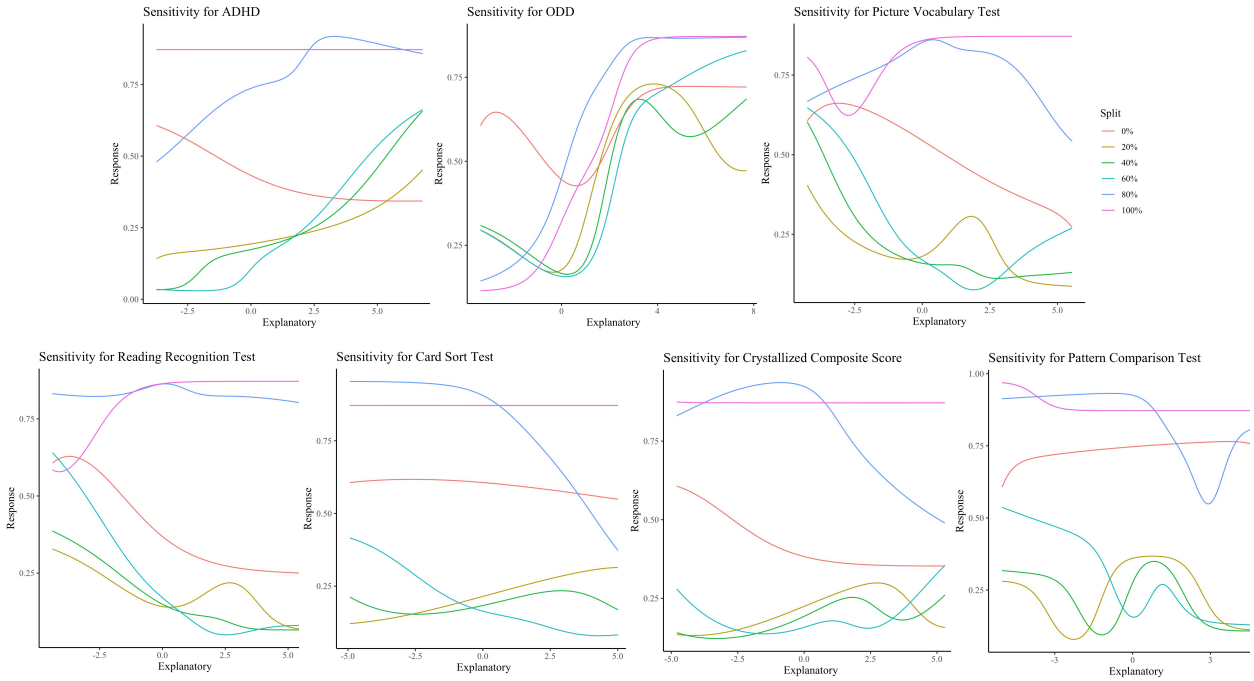
Note. Features are ordered from left to right by increasing absolute value of importance. The top seven features are represented in the social, psychological, and biological domains.

Supplemental Figure S7. Sensitivity Plots for the Social Model



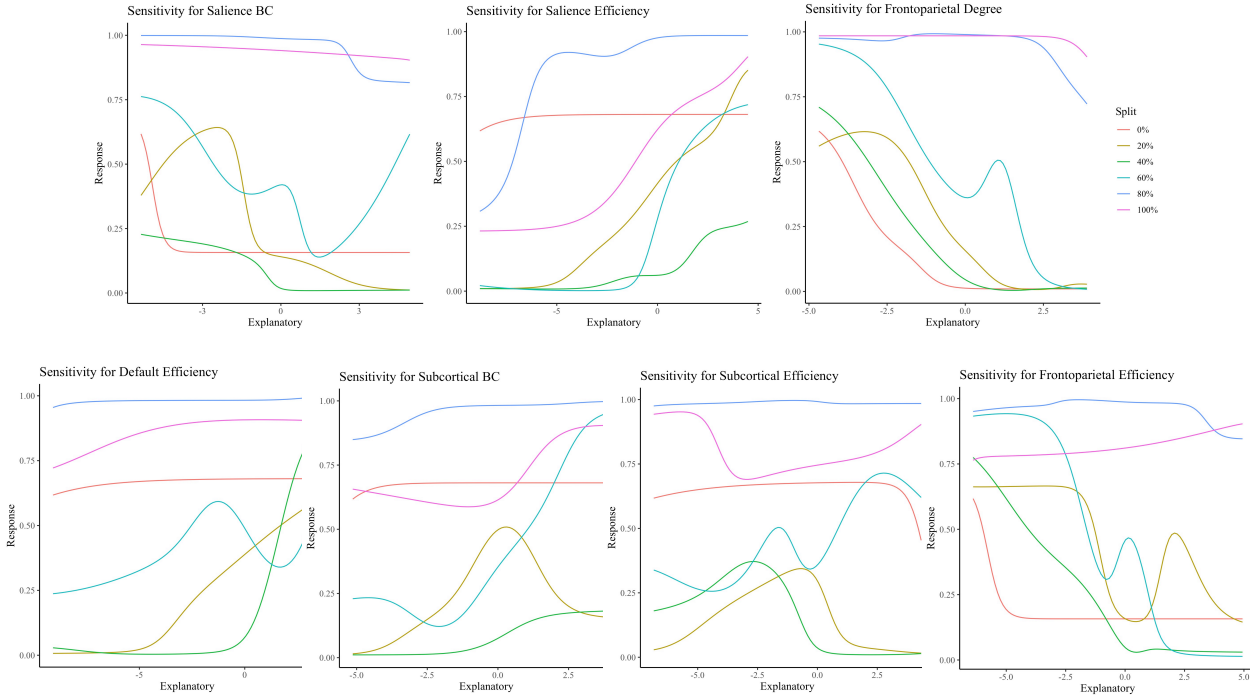
Note. The sensitivity plots display the relationship between model predictions and the risk factors of income, family fighting, neighborhood safety, family hitting, low family disagreement, low family anger, and parental monitoring respectively. The explanatory variable denotes risk factor values, and the response variable denotes the probability that a participant is diagnosed with 2-year CD. The risk factors of family fighting, hitting, disagreement, and anger positively correlate with the likelihood of developing CD, while the risk factors of income, neighborhood safety, and parental monitoring negatively correlate with the likelihood of developing CD.

Supplemental Figure S8. Sensitivity Plots for the Psychological Model



Note. The sensitivity plots display the relationship between model predictions and the risk factors of ADHD, ODD, picture vocabulary performance, reading recognition, card sorting, crystallized cognition, and pattern comparison respectively. The explanatory variable denotes risk factor values, and the response variable denotes the probability that a participant is diagnosed with 2-year CD. The risk factors of ADHD and ODD symptomatology positively correlate with the likelihood of developing CD, while the risk factors of crystallized cognition and performance on the picture vocabulary, reading recognition, card sort, and pattern comparison tasks negatively correlate with the likelihood of developing CD.

Supplemental Figure S9. Sensitivity Plots for the Biological Model



Note. The sensitivity plots display the relationship between model predictions and the risk factors of salience betweenness centrality (BC) and efficiency, frontoparietal degree and efficiency, subcortical BC and efficiency, and default efficiency respectively. The explanatory variable denotes risk factor values, and the response variable denotes the probability that a participant is diagnosed with 2-year CD. The risk factors of default and salience efficiency and subcortical BC positively correlate with the likelihood of developing CD, while the risk factors of frontoparietal and subcortical efficiency, salience BC, and frontoparietal degree negatively correlate with the likelihood of developing CD.

Supplementary References

1. Garavan H, Bartsch H, Conway K, Decastro A, Goldstein R, Heeringa S, et al. (2018): Recruiting the ABCD sample: Design considerations and procedures. *Developmental cognitive neuroscience*. 32:16-22.
2. Clark DB, Fisher CB, Bookheimer S, Brown SA, Evans JH, Hopfer C, et al. (2018): Biomedical ethics and clinical oversight in multisite observational neuroimaging studies with children and adolescents: The ABCD experience. *Developmental cognitive neuroscience*. 32:143-154.
3. Hagler Jr DJ, Hatton S, Cornejo MD, Makowski C, Fair DA, Dick AS, et al. (2019): Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *Neuroimage*. 202:116091.
4. Iacono WG, Heath AC, Hewitt JK, Neale MC, Banich MT, Luciana MM, et al. (2018): The utility of twins in developmental cognitive neuroscience research: How twins strengthen the ABCD research design. *Developmental cognitive neuroscience*. 32:30-42.
5. Jha D, Kim J-I, Kwon G-R (2017): Diagnosis of Alzheimer's disease using dual-tree complex wavelet transform, PCA, and feed-forward neural network. *Journal of healthcare engineering*. 2017.
6. Jurafsky D, Martin JH Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.
7. Zou Q, Xie S, Lin Z, Wu M, Ju Y (2016): Finding the best classification threshold in imbalanced classification. *Big Data Research*. 5:2-8.
8. Ripley B, Venables W, Ripley MB (2016): Package 'nnet'. *R package version*. 7:700.
9. Le QV, Ngiam J, Coates A, Lahiri A, Prochnow B, Ng AY (2011): On optimization methods for deep learning. *ICML*.
10. Mohri M, Rostamizadeh A, Talwalkar A (2018): *Foundations of machine learning*. MIT press.
11. Ciaburro G, Venkateswaran B (2017): *Neural Networks with R: Smart models using CNN, RNN, deep learning, and artificial intelligence principles*. Packt Publishing Ltd.
12. Mandrekar JN (2010): Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology*. 5:1315-1316.
13. Tuvblad C, Zheng M, Raine A, Baker LA (2009): A common genetic factor explains the covariation among ADHD ODD and CD symptoms in 9–10 year old boys and girls. *Journal of abnormal child psychology*. 37:153-167.
14. Witkiewitz K, King K, McMahon RJ, Wu J, Luk J, Bierman KL, et al. (2013): Evidence for a multi-dimensional latent structural model of externalizing disorders. *Journal of abnormal child psychology*. 41:223-237.
15. Nielsen AN, Barch DM, Petersen SE, Schlaggar BL, Greene DJ (2020): Machine learning with neuroimaging: Evaluating its applications in psychiatry. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*. 5:791-798.