Demographics and categorical diagnosis

https://wiki.abcdstudy.org/release-notes/non-imaging/general.html#demographics

The KSADS is a semi-structured interview to measure current and past symptoms of mood, anxiety, psychotic, and

disruptive behavior disorders, among others, in children ages 6-18 years old.

 KSADS 2.0 mental health release notes DSM-V based symptoms and diagnoses based on the responses to individual questions.

Townsend, L, Kobak, K., Kearney, C., Milham, M., Andreotti, C., Escalera, J., Alexander, L., Gill, M.K., Birmaher, B.,Sylvester, R., Rice, D., Deep, A., Kaufman, J. (2020). Development of Three Web-Based Computerized Versions of the Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS-COMP) Child Psychiatric Diagnostic Interview: Preliminary Validity Data. Journal of the American Academy of Child and Adolescent Psychiatry, Feb;59(2):309-325.

doi:10.1016/j.jaac. PMID: 31108163.

Kaufman, J., Kobak, K., Birmaher, B., & de Lacy, N. (2021). KSADS-COMP Perspectives on Child Psychiatric Diagnostic Assessment and Treatment Planning. Journal of the American Academy of Child and Adolescent Psychiatry, 60(5), 540–542

Demographic questionnaire derived primarily from the validated computerized Kiddie Schedule for Affective Disorders and Schizophrenia for DSM-5 (KSADS-5 2.0) (Kobak et al., 2013). The KSADS-5 is a comprehensive diagnostic tool compatible with DSM-5 criteria, which facilitates the precise categorization of psychiatric diagnoses. Diagnoses may be classified as "certain," "possible," "in remission," or "not present." This categorical diagnostic approach leverages established criteria to determine the presence or absence of disruptive or abnormal behaviors, aligning closely with standards set forth in the Diagnostic and Statistical Manual of Mental Disorders (APA, 2000). Demographic variables include age, sex, race and ethnicity.

Self and Parental Reports of dimensional diagnosis

The Achenbach System of Empirically Based Assessment (ASEBA) is a comprehensive evaluation tool developed after years of research and practical use (*ASEBA*, 2019). It is designed to assess behavioral, emotional, and social aspects, along with strengths, competencies, and adaptive functioning in individuals ranging from 1½ years old to over 90. ASEBA is widely applied in diverse areas such as mental health services, education, healthcare, research, and more. The Child Behavior Checklist (CBCL) and Brief Problem Monitor (BPM), two components of the ASEBA, provides a dimensional diagnostic/assessment approach that places behaviors along a continuum of frequency and/or severity. Moreover, assessments are normed by informant, age, sex, and ethnicity. The resulting scores are reported as z-scores within a full T-score range, with a mean of 50 and a standard deviation of 10.

Parent-Reported Child Behavior Checklist

This parent-report ASEBA instrument assess youth dimensional psychopathological syndromes

Dimensional and adaptive functioning

ASEBA recommends not using summary scores for participants with more than 8

missing items. Based on ASEBA scoring,the DSM and subscale measures have a range >= 50. The three CBCL

Internalizing, Externalizing, and Total Problem scales have a full T-score range with a mean = 50 and SD = 10.

Achenbach, T. M. (2009). The Achenbach System of Empirically Based Assessment (ASEBA): Development,

Findings, Theory, and Applications. Burlington, VT: University of Vermont Research Center for Children, Youth, &

Families. Find here

The CBCL is a component of the ASEBA first published in 2001, and is used to assess children's behavioral, emotional, or social problems (Achenbach, 2001). It is a 112-item parent-reported survey, which uses a 3-point Likert scale for responses: "Very True," "Somewhat True," or "Not True," where parents are asked to rate each item based on their child's behavior "now or within the past six months." As depicted in Figure 5, the CBCL consists of several dimensions categorized into Syndrome Scales and DSM-Oriented Scales. The eight syndrome scales are established through factor analysis. They encompass clusters of common behaviors or symptoms. Meanwhile, the more recently developed seven DSM-Oriented Scales align with diagnostic categories outlined in the DSM-5 (American Psychiatric Association, 2013; Nelson et al., 2001). Furthermore, these scales are grouped into three high-level domains Internalizing, Externalizing, and Total Problems scales. These dimensions offer a detailed assessment of a child's emotional, social, and behavioral functioning, aiding in identifying areas that may benefit from therapeutic or educational interventions.

Self-Reported Brief Problem Monitor

The items, scales, and norms are based on decades of research and

practical experience, as summarized in the BPM Manual (Achenbach, McConaughy, Ivanova, & Rescorla, 2017).

Reference: Achenbach, T. M. (2009). The Achenbach System of Empirically Based Assessment (ASEBA): Development,

Findings, Theory, and Applications. Burlington, VT: University of Vermont Research Center for Children, Youth, &

Families.

The BPM, another component of the ASEBA, was first published in 2011 (Achenbach et al., 2011). Developed to complement parental assessments, adolescents provide self-reports on higher level domains and attention. It is a 19-item self-reported survey used to assess children's behavioral and emotional functioning and their responses to interventions (RTIs). It also uses a 3-point Likert scale for responses: "Very True", "Somewhat True," or "Not True." Children are instructed to rate each item based on their behavior "currently or within the past six months." As illustrated in Figure 5, the BPM measures four scales, including Internalizing, Attention Problems, Externalizing, and Total Problems scales, paralleling the items and scales found on the more comprehensive CBCL/6-18 (Achenbach et al., 2017).



***Figure 5.*** *Structure of the ASEBA, specifically focusing on the CBCL and the BPM. The parent-reported CBCL consists of Syndrome Scales and the more recently developed DSM-Oriented Scales. The DSM-Oriented Scales are aligned with diagnostic criteria from the DSM and include categories such as OCD and Anxiety Problems. The syndrome Scales include clusters of symptoms which are further grouped into three high-level domains known as (1) Internalizing Problems, (2) Externalizing, and a (3) Total Problems score that sums all problem items. The child-reported BPM is a shorter version, provides a rapid assessment parallel to dimensions in CBCL for monitoring behavioral and emotional functioning over time.*

Data for the current study were drawn from the Adolescent Brain and

Cognitive Development (ABCD) dataset (release 5.0), a national study of

adolescent brain development. Participants (unweighted n = 11,875) were

recruited from 21 catchment sites across the United States. (Levin & Liu, 2024)

Participants

1. **Participants**
   * **Sample Size**: Indicate the number of participants in the clinical group and the healthy control group.
   * **Demographics**: Provide demographic information for both groups (e.g., age, gender, socio-economic status).
   * **Recruitment**: Describe how participants were recruited for both the clinical and healthy samples, including inclusion and exclusion criteria for both groups.
   * **Clinical Diagnosis**: Specify the criteria used to classify individuals into the clinical group (e.g., DSM criteria, assessment scales).
   * **Ethical Considerations**: Mention the ethical approvals obtained and consent processes followed.

The process of participant selection and group classification is illustrated in Figure 4. Participants were excluded for the following reasons:

Missing CBCL questionnaire, Missinng BPM questitonnaire (4) met diagnostic criteria for any other present or current psychiatric disorder assessed by the caregiver-reported Kiddie Schedule for Affective Disorders and Schizophrenia (Kaufman et al., 1997) (8) failed FreeSurfer segmentation; (9) failed T1 quality control; and/or (10) missing ROI or covariate tabulated data from the National Institutes of Mental Health databases.

Inclusion:

The Recommended Imaging Inclusion instrument ( mri\_y\_qc\_incl)

provides the simple option of include or exclude series (1 or 0) based on

automated and manual QC review per MR measure - T1w,

Specific Criteria for T1-weighted (T1w) Data:

To include T1-weighted data, the following criteria should be met:

Passed Raw QC: The raw imaging data must meet the quality control benchmarks

FreeSurfer QC: The output from FreeSurfer should not show any critical errors

Derived Results Exist: Certain derived results must exist and not be marked as NA in the image preprocessing pipeline.

lack of MRI data, lack of parental or self-report data, or a diagnosis of any other psychological disorder. The OCD group was defined as having a current diagnosis of OCD based on the KSADS. Notably, individuals with a current diagnosis of OCD are included in the OCD group regardless of any comorbid psychiatric conditions, reflecting the highly comorbid nature of OCD. Conversely, the healthy control group was defined by the absence of any current psychological diagnosis. A total of 6,513 participants were included in the study.



***Figure 4.*** *Flowchart of participant selection and subgroup classification at 2-year follow-up. Considered groups are colored in green.*

For the healthy control group (n= 6079), the mean age of participants was 9.47 years (SD = 0.51). The sex distribution, 47.3% were female (n = 2875), 0% identified as intersex-male (n = 1), and 52.7% were male (n = 3203). The racial and ethnic composition of the group was as follows: 2.2% Asian (n = 136), 13.4% Black (n = 813), 18.9% Hispanic (n = 1150), 9.8% identifying as Other (n = 597), and 55.7% White (n = 3383). For the OCD group (n= 434), the mean age of participants was 9.46 years (SD = 0.50). The sex distribution included 47.5%female (n = 206) and 52.5% male (n = 228). The racial and ethnic composition was 0.9% Asian (n = 4), 14.7% Black (n= 64), 18.7% Hispanic (n = 81), 14.3% identifying as Other (n = 62), and 51.4% White (n = 223).

**Variables of Interest**

* Define key variables being compared between clinical and healthy samples, both from sMRI data (e.g., specific brain regions, volumes) and questionnaire results (e.g., symptom severity, cognitive scores).
* Describe any composite variables or indices created for analysis

Modelling approach

All analysis and data handling were done in R version 4.3.3. Models were built with the xgboost package. Separate XGBoost models were trained for each symptom domain resulting in four prediction models: parent-reported (1) internalizing score, and (2) externalizing score, as well as child-reported (3) internalizing score, and externalizing score. For each model, the predictor sets comprised 194 variables derived from the Destrieux and Aseg brain regions. To prevent bias in predictions, the target variable was not included in the predictor set. Missing data in the subcortical atlas (n=9) was handled by mean imputation. Additionally, two variables related to Aseg lesions in the left and right hemispheres were excluded due to having over 30% missing data. All target variables were standardized using z-score transformation to facilitate transparent comparisons across datasets. The dataset was then partitioned into training and testing subsets, with 70% of the data (n=4,559) allocated for training and the remaining 30% (n=1,954) reserved for testing.

Training

Hyperparameter tuning was achieved through 5-fold cross-validation (CV) using a grid search strategy, exploring parameters such as the number of boosting rounds, tree depth, learning rate, and subsampling ratios. Model selection was optimized by minimizing the root-mean-square error (RMSE), ensuring high predictive accuracy. The cross-validation output is then plotted as the interaction between boosting rounds and CV error across different tree depths and learning rates. These plots visualize how model complexity and iteration affect performance of the model.

Subsample = high class imbalance.

Testing

The hyperparameters yielding the best performance were used to evaluate feature importance. Feature importance plots were generated for each model, allowing identification of the most impactful predictors of symptom severity. Predictions were generated using the trained model with optimized tuning parameters applied to the test dataset. Subsequently, performance metrics including, R-squared (R²) and mean absolute error (MAE) were computed with the caret package. This entailed comparing the model-generated predictions against the actual outcomes, thereby assessing the model's generalization capabilities to unseen data. The final model was also utilized to derive a feature importance analysis, identifying the relative significance of each feature. This matrix is displayed as a graphical representation of the top three features, which elucidates the key variables impacting model predictions. This methodological approach is essential for both evaluating predictive accuracy and discerning the primary factors influencing the model's decision-making process.

Results



Training

Optimal hyperparameters were selected for each model based on minimizing the RMSE for each target variable the reporting source (parent vs. child) and problem behavior domain (internalizing vs. externalizing). Optimal hyperparameter tuning are reported in Table 1., the results reveal distinct configurations of tuning parameters to optimize predictive performance across different symptom domains and reporter. In terms of internalizing symptoms, both models exhibited identical settings for tree depth, minimum child weight, column sampling, and learning rate (see table x for an overview of ting tang tang). However, the optimal number of boosting rounds was greater for the parent-reported model than for the child-reported model. Furthermore, the gamma and row sampling rates were greater in the child-reported model than in the parent-reported model. For externalizing symptoms, the tree depth, learning rate, column sampling rate, and row sampling rate were consistent across both the parent and child models. However, the gamma and minimum child weight values were higher in the child-reported model compared to the parent-reported model. Additionally, the number of boosting rounds was greater in the parent-reported model than in the child-reported model. These tuned hyperparameters reflect adjustments for model complexity and regularization that may correspond to differences in reporting sources and symptom domains.



*Table 1. Optimal hyperparameter values for models predicting internalizing and externalizing symptoms based on sMRI data. Hyperparameters were independently tuned for each model. Notable differences were observed in the number of boosting rounds, gamma, and row sampling rate across reporting sources and symptom domains. These differences reflect adjustments in model complexity and regularization to optimize predictive performance.*

Testing

Model performance was evaluated using RMSE, MAE, and R² for each of the four models. For internalizing symptoms, the parent-reported model achieved an RMSE of 0.9450, R² of 0.0015, and MAE of 0.7677. The child-reported model had a similar RMSE (0.9447) but slightly lower R² (0.0002) and MAE (0.7057), suggesting modest improvements in average prediction error but limited variance explained. For externalizing symptoms, the parent-reported model had an RMSE of 0.8862, R² of 0.0006, and MAE of 0.7360. The child-reported model showed a slightly higher RMSE (0.9056), but a higher R² (0.0038) and lower MAE (0.6507), indicating somewhat better prediction accuracy and variance capture in this domain. Overall, while all models yielded low R² values—indicating limited explained variance—the lower MAE values, particularly for child-reported externalizing symptoms, suggest better precision in predicting individual scores despite limited model fit.

****

***Table 2.***

Discussion

For scaled target variables, here's a rough guideline:

* An RMSE close to 0 suggests an excellent model fit.
* An RMSE less than 0.1 is often considered very good, indicating that the predictions are close to the actual values within the scaled range.
* An RMSE between 0.1 and 0.2 may indicate moderate performance, while values above 0.2 could suggest that the model might not be performing well. Gpt

Find normal values in cogneuro

MAE is not identical to [root-mean square error](https://en.wikipedia.org/wiki/Root-mean-square_deviation) (RMSE), although some researchers report and interpret it that way. The MAE is conceptually simpler and also easier to interpret than RMSE: it is simply the average absolute vertical or horizontal distance between each point in a scatter plot and the Y=X line. In other words, MAE is the average absolute difference between X and Y. https://en.wikipedia.org/wiki/Mean\_absolute\_error

A variety of challenges accompany efforts to process multimodal imaging data, particularly with large numbers of subjects, multiple sites, and multiple scanner manufacturers.(Hagler et al., 2019)

Head motion is a significant issue, particularly with children, as it degrades image quality, and potentially biases derived measures for each modality