Check for updates

Selectively predicting the onset of ADHD, oppositional defiant disorder, and conduct disorder in early adolescence with high accuracy

Nina de Lacy[1,2]* and Michael J. Ramshaw[1,2]

[1]Huntsman Mental Health Institute, Salt Lake City, UT, United States, [2]Department of Psychiatry, University of Utah, Salt Lake City, UT, United States

Introduction: The externalizing disorders of attention deficit hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), and conduct disorder (CD) are common in adolescence and are strong predictors of adult psychopathology. While treatable, substantial diagnostic overlap complicates intervention planning. Understanding which factors predict the onset of each disorder and disambiguating their different predictors is of substantial translational interest.

Materials and methods: We analyzed 5,777 multimodal candidate predictors from children aged 9−10 years and their parents in the ABCD cohort to predict the future onset of ADHD, ODD, and CD at 2-year follow-up. We used deep learning optimized with an innovative AI algorithm to jointly optimize model training, perform automated feature selection, and construct individual-level predictions of illness onset and all prevailing cases at 11−12 years and examined relative predictive performance when candidate predictors were restricted to only neural metrics.

Results: Multimodal models achieved ∼86−97% accuracy, 0.919−0.996 AUROC, and ∼82−97% precision and recall in testing in held-out, unseen data. In neural-only models, predictive performance dropped substantially but nonetheless achieved accuracy and AUROC of ∼80%. Parent aggressive and externalizing traits uniquely differentiated the onset of ODD, while structural MRI metrics in the limbic system were specific to CD. Psychosocial measures of sleep disorders, parent mental health and behavioral traits, and school performance proved valuable across all disorders. In neural-only models, structural and functional MRI metrics in subcortical regions and cortical-subcortical connectivity were emphasized. Overall, we identified a strong correlation between accuracy and final predictor importance.

Conclusion: Deep learning optimized with AI can generate highly accurate individual-level predictions of the onset of early adolescent externalizing disorders using multimodal features. While externalizing disorders are frequently co-morbid in adolescents, certain predictors were specific to the onset of ODD or CD vs. ADHD. To our knowledge, this is the first machine learning study to predict the onset of all three major adolescent externalizing disorders with the same design and participant cohort to enable direct comparisons, analyze >200 multimodal features, and include many types of neuroimaging metrics. Future study to test our observations in external validation data will help further test the generalizability of these findings.

Introduction

Attention deficit hyperactivity disorder (ADHD), oppositional defiant disorder (ODD), and conduct disorder (CD) are common mental health conditions in adolescence, often collectively referred to as externalizing disorders. Among the most common youth mental health conditions, externalizing behaviors are the most frequent reason for referral to mental health services and a strong predictor of adult psychopathology (1). In school-age youth (K-12), 10–24% meet the criteria for externalizing disorders, with ADHD and ODD being the most common; (2) ADHD affects 7–10% of youth <18 years of age (years), with prevalence showing a strong uptick in early adolescence, peak in mid-late adolescence, and decline into adulthood. Some ∼2% of children ≤5 years are affected (vs.) ∼10% at 6–11 years and ∼13% at 12–17 years, with ∼4% of adults having clinical ADHD (2–4). In contrast, ODD and CD (collectively the disruptive externalizing disorders) affect ∼5% of youth ≤17 years, growing to ∼10–12% of adults, where in the latter, they are associated with increased risk for later co-morbid mental health and substance use disorders (5–8). Among youth with ADHD, ∼30–50% may also exhibit disruptive externalizing behaviors consistent with ODD and CD, with this association growing with increasing age and linked to later poor academic and life outcomes such as school dropout, substance abuse, and involvement with the justice system (9–13). Thus, early adolescence is a period of considerable interest in understanding which risk factors predict the onset of externalizing disorders and disambiguating those that may differentially predict the development of ADD vs. ODD and CD.

Adolescent externalizing disorders have attracted a range of research approaches. Historically, these have predominantly been cross-sectional studies quantifying group-level associations, frequently assessing neuroimaging metrics. More recently, machine learning (ML) classification techniques have been applied increasingly to large-scale datasets. Such approaches offer the advantages of providing individual-level case predictions from high-dimension and/or multimodal data, thereby bridging from extant work focused on identifying statistical associations at a group level to a pathway toward personalized medicine (14, 15). Appropriately constructed ML algorithms can simultaneously analyze hundreds or thousands of candidate predictors and enlarge the solution space. Such work has been further fueled by the increasing availability of large-scale, open science datasets incorporating multimodal variables. In peri-adolescence, the flagship initiative of this type is the ongoing population-level, longitudinal ABCD study ($n = 11,800$) used in the present study that enrolled children at ages 9–10 years and collected data from many knowledge domains, including multiple neuroimaging types (16–18). While a number of ML predictive studies have been performed in adolescent externalizing disorders, these have largely (though not exclusively) been cross-sectional and focused on predicting prevailing cases at a particular age in a single disorder. Few ML studies have predicted the future onset of disease in longitudinal data or applied a consistent analytic architecture across the three major adolescent externalizing disorders in the same population cohort to enable direct comparisons.

In the present study, we extend prior work with an ML design that analyzes a large number of multidomain candidate predictors to predict new onset cases of ADHD, ODD, and CD in early adolescence in the same design and youth cohort. We aimed to identify the best-performing predictors and compare these across these three related disorders to understand whether there were shared or unique predictors underpinning ADHD, ODD, and CD. Given the large prior literature related to brain structure and function motifs in externalizing disorders, we also wanted to compare the relative predictive ability of models composed purely of neuroimaging metrics derived from MRI with multimodal models. By leveraging an AI algorithm that jointly optimizes ML model training and performs automated feature selection, we were able to analyze 5,777 candidate predictors spanning demographics, developmental and medical history, white and gray matter brain structure, neural function (cortical and subcortical connectivity, three tasks), brain volumetrics, physiologic functions (e.g., sleep, hormone levels, pubertal stage, and physical function), cognitive and academic performance, social and cultural environment (e.g., parents, friends, and bullying), activities of everyday life (e.g., screen use and hobbies), living environment (e.g., crime, pollution, and educational and food availability), and substance use. We used features assessed at 9–10 years (107–132 months) to predict future new onset cases of ADHD, ODD, and CD at 11–12 years with deep learning with artificial neural networks, which incorporates non-linear relationships among predictors and is resistant to multicollinearity. Since extant work is more focused on predicting prevailing rather than new onset cases, we performed additional experiments to predict all prevailing cases at 11–12 years to provide comparisons with the existing literature. Our AI approach allowed us to render fully interpretable predictions, quantify relative predictor importance at both the group and individual levels, and examine the relationship between model accuracy and predictor importance across all models. All results presented are from testing for generalization in holdout, unseen data.

Materials and methods

Terminology and definitions

Terms used in the quantitative analysis may be shared among different fields with variant meanings. Here, we use ML conventions throughout (19–21). "Prediction" means predicting the quantitative value of a target variable by analyzing patterns in input data. The set of observations used to train and validate models is referred to as the "training set" and the unseen holdout set of observations is termed the "test set". We refer to the set of all input data used in training as containing "features" or "candidate predictors" and those identified in final, optimized models after testing in held-out data (presented in Results section) as "final predictors". We use "generalizability" to refer to the ability of a trained model to adapt to new, previously unseen data drawn from the same distribution, i.e., model fit in the test set. "Precision" refers to the fraction of positive predictions that were correct, "Recall" refers to the proportion of true positives that were correctly predicted, and "Accuracy" refers to the number of accurate predictions as a fraction of total predictions. Receiver Operating

Characteristic curves (ROC curves) are provided that quantify classification performance at different classification thresholds plotting true positive vs. false positive rates, where the Area Under the Curve (AUROC) is defined as the two-dimensional area under the ROC curve from (0,0) to (1,1). This paragraph defining terminology usage is adapted from our prior study.

Data and data collection in the ABCD study

We use data from the ABCD study, an epidemiologically informed prospective cohort study that recruited 11,880 children (52% male; 48% female) aged 9–10 years in 21 sites across the United States, intending to follow this youth for the next decade. Participants in the cohort included 800 twin pairs ($n = 800$) and/or non-twin siblings. These data are made available to qualified researchers at no cost from the National Institute of Mental Health Data Archive and are released periodically. The present study uses data from Release 4.0, the 42-month follow-up date. Fuller descriptions of the overall design of the ABCD study, as well as recruitment procedures and the participant sample, may be found in the studies by Jernigan et al., Garavan et al., and Volkow et al. (22–24). This study has been reviewed and deemed not human subjects research by the University of Utah Institutional Review Board.

ABCD collects a wide range of information from youth participants and their parents, comprising phenotypic, demographic, psychometric, physiologic, and developmental data, as well as multiple modalities of MRI neuroimaging. Barch et al. and Lisdahl et al. detail the phenotypic and substance abuse assessment protocols, respectively (25, 26). Here, we utilize data from assessments of physical and mental health, substance use, neurocognition, school performance, quality, culture, and environment performed for youth and their parents, as well as biospecimens (e.g., pubertal hormone levels) and environmental toxin exposure. A summary description of assessments performed and environmental and school-related variables derived from geocoding at age 9–10 years that we analyzed may be inspected in Supplementary Table 1.

Brain imaging incorporates optimized 3D T1, 3D T2, diffusion tensor imaging, resting state functional MRI (rsfMRI), and 3 task MRI (tfMRI) protocols harmonized across acquisition sites. The tfMRI protocol comprises the monetary incentive delay (MID) and stop signal tasks (SST) and an emotional version of the n-back task, which collectively measure reward processing, motivation, impulsivity, impulse control, working memory, and emotion regulation. In the present study, we utilized ABCD-provided fully processed metrics from each of these imaging types that are computed after quality control. Detailed descriptions of the requisite acquisition, pre-processing, quality control, and analytic protocols used to generate metrics may be inspected by Casey et al. and Hagler et al. (27, 28). We utilized all available processed metrics that have passed quality control from diffusion fullshell; cortical and subcortical Gordon correlations (connectivity); structural, volumetric, and all three tfMRI tasks, as well as corresponding head motion statistics for each modality. For certain modalities, such as

TABLE 1 Demographic characteristics of participant sample at ages 9–10 years.

Characteristic	Number	Percent
Sex		
Male	2,771	51.7%
Female	2,584	48.3
Gender Identity		
Male	2,768	51.7%
Female	2,577	48.1
Gender non-conforming	7	0.1
Do not know/did not answer	4	0.1
Race		
Black/African American	873	16.3%
Asian	353	6.6
White	4,236	79.1
Native American/Alaska Native	187	3.5
Other	334	6.2
Ethnicity		
Hispanic/Latino/Latinx	1,070	20.0%
Non-Hispanic	4,224	78.9
Not indicated	62	1.2

Sex refers to the sex assigned at birth on the original birth certificate. Gender refers to the youth's gender identification. Race and ethnicity refer to the parents' view of the youth's race or ethnicity. More than one race or ethnicity identification may be selected and therefore percentages may sum to >100%.

rsfMRI, multiple scans were attempted or completed. In such cases, we use metrics computed from the first scan.

Study inclusion criteria and sample partitioning for machine learning

We included youth from the larger ABCD cohort in the present study if they were (a) participants enrolled in the study at baseline (9–10 years) who were still enrolled in the ABCD study at 2-year follow-up at 11–12 years ($n = 8,085$) who had (b) complete data for all neural imaging types for at least one scan in each modality listed above that passed ABCD quality control ($n = 6,178$) and were (c) youth participants unrelated to any other youth participant in the study ($n = 5,355$). If a youth had sibling(s) present in the cohort, we selected the oldest sibling for inclusion. Demographic characteristics of this sample at age 9–10 years, the age which corresponds to input data used to make predictions, are presented below in Table 1.

Physiologic and cognitive characteristics of the same participant sample at 9–10 years may be viewed in Table 2.

The final participant sample ($n = 5,356$ participants), after inclusion criteria were applied, was randomly partitioned into a training set comprising 70% of the sample ($n = 3,749$) and a

TABLE 2 Physiologic and cognitive characteristics of participant sample at ages 9–10 years.

Characteristic	Range	Mean	Median
Age in months	107.0–132.0	119.9	120.0
Pubertal development stage	1–5	2	2
Height (inches)	36.6–81.0	55.4	55.4
Weight (pounds)	11.0–255.0	82.3	76.8
Waist circumference (cm)	17.0–61.0	26.4	25.5
Handedness			
Writing	−100.0 to 100.0	76.5	100.0
Throwing	−100.0 to 100.0	67.1	100.0
Spoon	−100.0 to 100.0	62.4	100.0
Vocabulary	51.0–208.0	109.0	109.0
Attention and inhibition	65.0–171.0	96.5	97.0
Working memory	46.0–194.0	102.0	103.0
Executive function	68.0–181.0	98.0	94.0
Processing speed	20.0–185.0	95.2	95.0

Characteristics of the study sample at 9–10 years. Pubertal development is measured with the Pubertal Development Scale (adapted from the Petersen scale) in a sex-specific manner. Height is measured twice with the average of these values presented. We note a range of 11.0–255.0 pounds for weight, which is the range present in the original ABCD data. Handedness is assessed with the Edinburgh Handedness Inventory. Cognitive metrics are assessed with the NIH Toolbox and are all age-corrected scores. Vocabulary is measured with the Picture Vocabulary test; Attention and inhibition with the Flanker Inhibitory Control & Attention Task; Executive Function with the Dimensional Change Card Sort Test; and Processing Speed with the Pattern Comparison Processing Speed Test.

holdout, unseen test set comprising 30% of the sample ($n = 1{,}607$, Figure 1). This partitioning was effected before pre-processing either input features (candidate predictors) or predictive targets to minimize bias.

Preparation of predictive targets

Predictive targets of ADHD, ODD, and CD cases were derived from the Child Behavior Checklist for youth ages 4–18 years (CBCL), known as the "ABCD Parent Child Behavior Checklist Scores Aseba (CBCL)" in ABCD study nomenclature. The CBCL is a standardized instrument in widespread clinical and research use. It forms part of the Achenbach System of Empirically Based Assessment (ASEBA), "designed to facilitate assessment, intervention planning and outcome evaluation among school, mental health, medical and social service practitioners who deal with maladaptive behavior in children, adolescents and young adults" (29). To score the CBCL, parents rate their child on a 0-1-2 scale on 118 specific problem items such as "Acts too young for age" over the prior 6 months. Answers are aggregated into raw, T, and percentile scores for eight syndrome subscales (anxiety,

somatic problems, depression, social problems, thought problems, attention problems, rule breaking, and aggressive behavior) derived from principal components analysis of data from 4,455 children referred for mental health services. The CBCL is normed in a U.S. nationally representative sample of 2,368 youth ages 4–18 years that take into account differences in problem scores for "males vs. females". It exhibits excellent test–retest reliability of 0.82–0.96 for the syndrome scales with an average r of 0.89 across all scales. Content and criterion validity are strong, with referred vs. non-referred children scoring higher on 113/188 problem items and significantly higher on all problem scales, respectively.
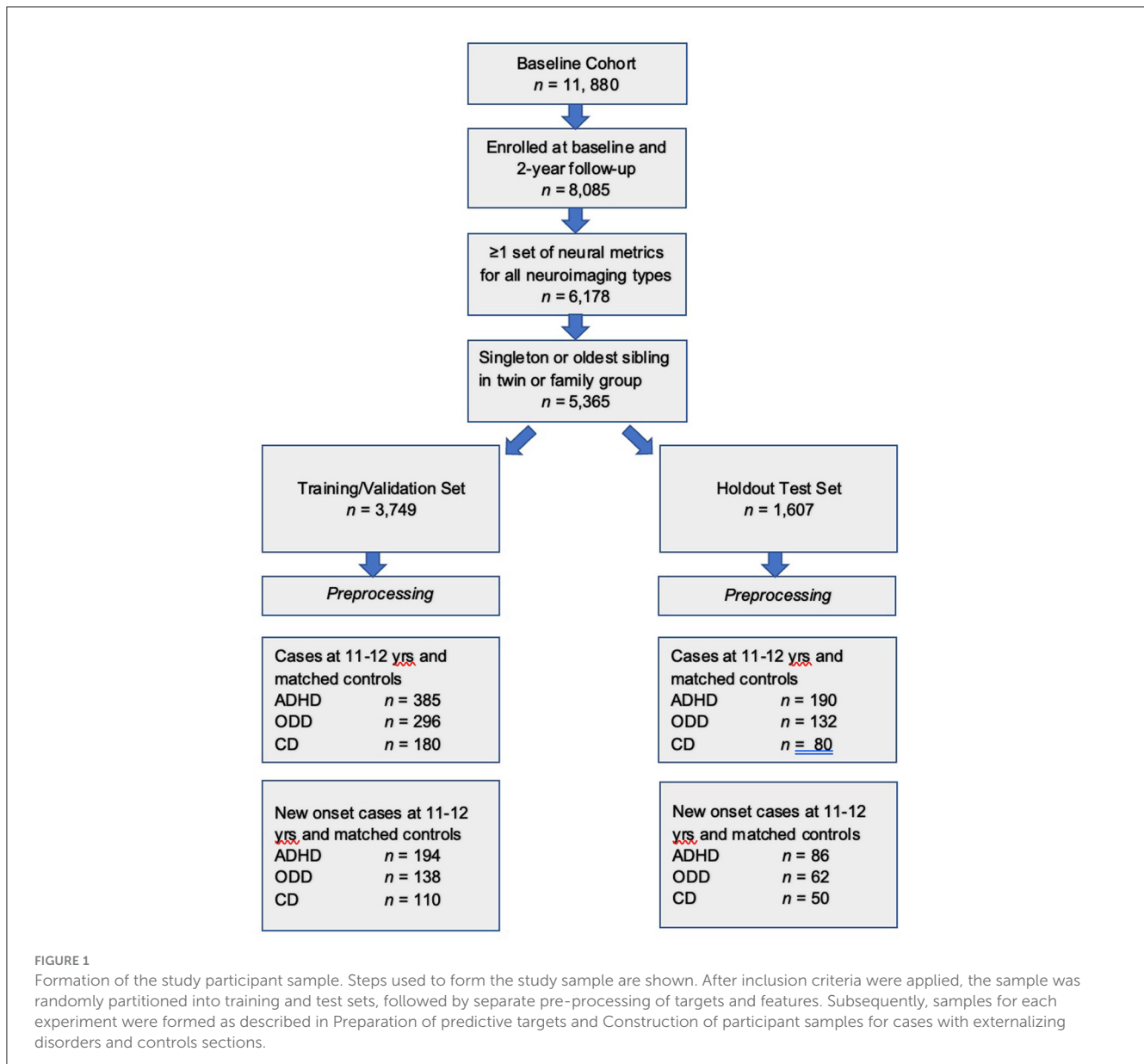
To form binary classification targets, we thresholded CBCL subscale T scores for ADHD ("attention problems"), ODD ("aggressive behavior"), and CD ("rule breaking") using cut-points established by ASEBA for clinical practice. Specifically, a T score of 65–69 (95th to 98th percentile) is considered in the "borderline clinical" range and scores ≥ 70 are considered in the "clinical range." Accordingly, we discretized T scores for each of the three subscales under consideration by deeming every individual with a T score ≥ 65 as a "case" or [1] and every individual with a score <65 as "not case" or [0]. This process was performed separately in the training and test sets for participant CBCL scores at 9–10 and 11–12 years.

Construction of participant samples for cases with externalizing disorders and controls

We formed two participant samples in each of ADHD, ODD, and CD in the training and test sets (Figure 1). The first sample type comprised all cases of ADHD, ODD, and CD present in the larger sample at 11–12 years. The second sample type comprised only new onset cases at 11–12 years. A new onset case was defined as a youth who met the criteria for ADHD, ODD, or CD as defined by the ASEBA CBCL cut-points at 11–12 years and who did not meet the criteria for the requisite disorder at 9–10 years. Thus, six participant samples in total were constructed. In all samples, we formed a balanced sample of cases and controls. The latter were youth with the lowest possible scores on the relevant syndrome scale selected from the eligible study population (see: Baseline inclusion criteria and sample partitioning for machine learning) and matched with cases for age in months and sex/gender.

Preparation of candidate predictors (input features)

We assembled a feature set for input into predictive algorithms that comprised the majority of phenotypic, demographic, psychometric, physiologic, and developmental variables available from the ABCD study (including data collection site) and all available neural metrics, including head motion statistics with the exception of temporal variance measures (Supplementary Table 1). We used only metrics collected at 9–10 years. In continuous phenotypic features, we used subscale or total scores where available, for example, subscale scores exemplifying different

FIGURE 1
Formation of the study participant sample. Steps used to form the study sample are shown. After inclusion criteria were applied, the sample was randomly partitioned into training and test sets, followed by separate pre-processing of targets and features. Subsequently, samples for each experiment were formed as described in Preparation of predictive targets and Construction of participant samples for cases with externalizing disorders and controls sections.

types of sleep-related disorders from the Munich Chronotype Questionnaire. Metrics directly quantifying mental health symptoms were excluded since we aimed to predict cases of mental illness without using symptoms, as the latter tends to inflate predictive performance and narrow the utility of findings. The feature set was then partitioned into training and test sets conforming at the participant level with case/control partitions described above (Construction of participant samples for cases with externalizing disorders; Figure 1). Pre-processing of features was then performed separately in the training and test sets to minimize bias. First, features with >35% missing values were discarded, where prior research shows that good results may be obtained with ML methods with imputation of up to 50% missing data (30). Nominal or ordinal variables were one-hot encoded to transform them into discrete variables. Continuous variables were trimmed to [mean $\pm$ 3] standard deviations to remove outliers, and all features were scaled in the interval

[0, 1] with MinMaxScaler. Missing values were imputed using non-negative matrix factorization (NNMF), a mathematically proven imputation method that minimizes the cost function of missing data rather than assuming zero values. It captures both global and local structures in the data effectively and is particularly suitable for large-scale multimodal data, having been demonstrated to perform well regardless of the underlying pattern of missingness (31–33). Supplementary Table 2 shows the number and percentage of observations in each variable trimmed and filled with NNMF for the training and test sets. After imputation with NNMF, phenotypic variables lacking summary scores were reduced to a summary metric or index using feature agglomeration to produce a final set of ($n = 763$) non-neural metrics. As described above, neural metrics ($n = 5,014$) had already been processed and underwent quality control by the ABCD study team and were, therefore, not pre-processed with the exception of scaling with the MinMaxScaler, again performed separately in the training and

test partitions. There were no missing neural features. The final combined, multimodal feature set, including all feature types, contained 5,777 features.

Overview of predictive analytic pipeline

We used deep learning with artificial neural networks to predict cases of ADHD, ODD, and CD at 11–12 years. In total, we performed 12 experiments, predicting new onset and all prevailing cases for each of the three disorders using (a) all available multimodal features and (b) only neural features. Deep learning models were implemented with k-fold cross-validation and trained by an AI algorithm that jointly performed feature selection and optimized across the hyperparameters in an automated manner. Typically, ∼40,000 model fits were performed during training in each experiment. Model training was terminated based on the Bayes Information Criterion (BIC), an information-theoretic metric. After training, final models obtained from the optimized training process were tested for their ability to generalize in the holdout, unseen test set and performance statistics of AUROC, accuracy, precision, and recall, and ROC curves computed and reported for these final, optimized models. We also computed and reported the relative importance of final predictors to making case predictions using the Shapley additive explanation (SHAP) technique. Detailed explanations of these methods are provided below. Code for predictive analytics may be accessed at the de Lacy Laboratory GitHub: https://github.com/delacylab/integrated_evolutionary_learning.

Coarse feature selection

We performed coarse feature selection individually for each of the six experimental samples before beginning model training to reduce the number of features entering the deep learning pipeline in a principled, optimized manner. This identified subsets of 5,777 features with non-zero relationships with the predictive target. First, a simple filtering process was performed in which χ^2 (categorical features) and ANOVA (continuous features) statistics and mutual information metric (all features) were computed to quantify the relationship between all features and the target, where the target (ADHD, ODD, and CD) was represented by a categorical vector in [0, 1]. Any feature with a non-zero relationship (either positive or negative) with the target was retained. Further feature selection was then performed on these filtered feature subsets using the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm. This popular regularization technique based on linear regression efficiently selects a reduced set of features by forcing certain regression coefficients to zero. The LASSO algorithm has a hyperparameter (commonly called the α) that governs the degree of penalization (shrinkage) that will be imposed on the features and thereby influences results. To optimize across this hyperparameter, we implemented the LASSO with our AI meta-learning algorithm integrated evolutionary learning (IEL) to tune

TABLE 3 Feature sets after coarse feature selection for each experiment.

	Number of features after filtering	Number of features after selection with LASSO
ADHD, new onset 11–12 years	4,272	54
ODD, new onset 11–12 years	4,314	76
CD, new onset 11–12 years	4,303	77
ADHD, all prevailing cases 11–12 years	4,388	63
ODD, all prevailing cases 11–12 years	4,401	81
CD, all prevailing cases 11–12 years	4,460	59

The total set of 5,777 multimodal input features was reduced via coarse feature selection in a two-step process of filtering followed by optimized regularization with the LASSO algorithm. This table displays the number of remaining features after filtering and regularization for each target (ADHD, ODD, and CD) and participant sample type (new onset cases and all prevailing cases at age 11–12 years). Detailed tables showing the univariate coefficients between each feature selected by LASSO-based regularization and each target may be viewed in Supplementary Tables 3a–f.

TABLE 4 Hyperparameter settings optimized with integrated evolutionary learning.

Hyperparameters	Range	Mutation shift
Learning rate	0.00001–0.01	0.0001
Beta 1	0.9–0.999	0.001
Beta 2	0.9-0.999	0.001

Optimization across the hyperparameters of learning rate, Beta 1 and Beta 2 was conducted for deep learning with artificial neural networks within the ranges shown.

the α hyperparameter in the same manner as described below in Integrated Evolutionary Learning for deep learning optimization.

The number of features retained for each of the six experimental samples after each step described above for the coarse feature selection process may be seen in Table 3. Specific features selected in the optimized LASSO regularization and the resulting univariate coefficients between each of these features and the target vectors (ADHD, ODD, and CD) for each participant sample (new onset and all prevailing cases at 11–12 years) may be viewed in Supplementary Tables 3a–f. Each feature set selected by the LASSO then entered the deep learning pipeline.

Deep learning with artificial neural networks

We used deep learning to predict cases of ADHD, ODD, and CD in each type of participant sample (new onset and all prevailing cases at 11–12 years). To predict only future cases of externalizing disorders, candidate predictors collected at 9–10 years were solely used to predict cases at 11–12 years. We further recapitulated each experiment after restricting the set of candidate predictors to

5,014 neural features to construct neural-only models to compare their performance to that obtained with multimodal features. In each case, we trained artificial neural networks using the AdamW algorithm with three layers, 300 neurons per layer, early stopping (patience = 3, metric = validation loss), and the Relu activation function. The last output layer contained a conventional softmax function. Learning hyperparameters (Table 3) were tuned with IEL as detailed below. Deep learning models were encoded with TensorFlow embedded in custom Python code.

Integrated evolutionary learning for optimization across hyperparameters and fine feature selection

ML algorithms typically have hyperparameters that control learning, where their settings can strongly affect performance. In many approaches, these hyperparameters are used at their default settings or manually tuned using "rules of thumb" and a restricted number of model fits are explored, introducing the possibility of bias and potentially limiting the solution space (34–36). To address this issue, we previously developed an AI technique called Integrated Evolutionary Learning (IEL), which can improve the performance of ML predictive algorithms in tabular data by up to 20–25% vs. the use of default model hyperparameters (37). IEL is a form of computational intelligence or metaheuristic based on an evolutionary algorithm that instantiates the concepts of biological evolutionary selection in computer code. It optimizes across the hyperparameters of the deep learning algorithm by adaptively breeding models over hundreds of learning generations and selecting for improvements in a fitness function (here, the Bayes Information Criterion, BIC).

For each experiment, the deep learning algorithm was nested inside IEL, which initialized the first generation of 100 models with randomized hyperparameter values or "chromosomes". Hyperparameter settings (Table 4) were subsequently recombined, mutated, or eliminated over successive generations. In recombination, "parent" hyperparameters were arithmetically averaged to form "children". In mutation, settings were shifted with the range of possible values shown in Table 4. After the first training generation, the BIC was computed for each of the 100 solutions. The 60 best models (highest BIC) were identified, and 40 of these recombined by averaging the hyperparameter setting after a pivot point at the midpoint to produce 20 'child' models. The remaining 20 were mutated to produce the same number of child models by shifting the requisite hyperparameter by the mutation shift value (Table 4). The remaining 40 models were discarded. The next generation of models was then formed by adding 60 new models with randomized settings and adding these to the 40 child models retained from the initial generation. Thereafter, IEL continued to recombine, mutate, and discard 100 models per generation in a similar fashion to minimize the BIC until the latter fitness function plateaued. With 100 models fitted per generation, IEL typically fits ~40,000 models per experiment over ~400 learning generations.

IEL jointly performs this optimization process across hyperparameter settings with automated feature selection,

mitigates the risk of overfitting, and identifies predictors that perform best. For each experiment, IEL selects among available candidate predictors after coarse feature selection (Coarse feature selection, Supplementary Table 3). A random number of features in the range [2–50] was randomly seeded for each model in the initial learning generation. After computing the fitness function, feature sets from the best-performing 60 models were allocated to child models, and other feature sets were discarded. As with hyperparameter tuning, this process was repeated for succeeding generations until the BIC plateaued.

IEL implements recursive learning to facilitate computational efficiency. After training until the BIC plateaued, we determined the elbow of the fitness function plotted vs. the number of features and re-started learning with a warm start. The feature set available after this warm start is constrained to that subset of features, thresholded by their importance, corresponding to the fitness function elbow. Learning then proceeds by thresholding features available for learning at the original warm start feature importance + 2 standard deviations. In addition, the number of models per generation is reduced to 50, 20 models are recombined, and 10 models are mutated. Otherwise, training after the warm start uses the same principles as detailed above.

Cross-validation

Deep learning models were fit within IEL using stratified k-fold cross-validation, i.e., every one of the 100 models in each learning generation within IEL was individually trained and validated using cross-validation. As described above, IEL allows the number of features used to fit each model to differ within each model in every generation in the range [2–50]. Accordingly, k (the number of splits) was set as the nearest integer above [sample size/number of features]. Cross-validation was implemented with the scikit-learn stratified K-fold function.

Testing for generalization in holdout, unseen test data and performance measurement

Finally, optimized models generated in the IEL-supervised training process were tested on the held-out, unseen test set for each sample and disorder by applying the requisite hyperparameter settings and selected features to the test set. The area under the receiver operating curve (AUROC), accuracy, precision, and recall were computed for test set models using standard scikit learn libraries. The most accurate models are presented in the Results section. The threshold for prediction probability was 0.5, and receiver operating characteristic (ROC) curves are also provided for each experiment (Supplementary Figures 1, 2).

Feature importance determination

Shapley additive explanation (SHAP) values were computed using the SHAP toolbox (https://shap.readthedocs.io/en/latest/) to

TABLE 5 Performance of deep learning optimized with integrated evolutionary learning in predicting cases of ADHD using multimodal and neural-only feature types.

Age of case determination	Accuracy (%)	Precision (%)	Recall (%)	AUROC
(A)				
New onset at age 11–12 years	86.1	81.7	83.7	0.919
All cases at age 11–12 years	93.7	91.8	91.6	0.985
(B)				
New onset at age 11–12 years	79.1	73.9	74.4	0.841
All cases at age 11–12 years	63.7	59.5	49.5	0.654

Performance statistics of accuracy, precision, recall, and the AUROC are shown for the most accurate model obtained with deep learning optimized with integrated evolutionary learning using **(A)** multimodal features and **(B)** only neural features. We used features obtained at 9–10 years of age to predict new onset cases of ADHD at 11–12 years of age, as well as all prevailing cases at 11–12 years of age. Corresponding ROC curves may be viewed in Supplementary Figures 1, 2.

determine the relative importance of each feature to predicting cases in each experiment for ADHD, ODD, and CD. SHAP is a game theoretic approach commonly used in ML to explain the output of any ML model, including "black box" estimators such as artificial neural networks, and is resistant to multicollinearity (38). It unifies prior methods such as LIME, Shapley sampling values, and Tree Interpreter.

Results

Overview

All study results detailed below are from testing the final model obtained after IEL optimization for generalization in a holdout, unseen test dataset for each experiment. We present parallel sets of results for each disorder (ADHD, ODD, and CD) in predicting new onset cases at 11–12 years and all prevailing cases at 11–12 years. Only features collected at 9–10 years are input to deep learning to make predictions. Therefore, all results represent predictions of future case status. For each disorder, results are presented for standard ML performance metrics and quantification of feature importance for (a) multimodal models constructed using all types of input features and (b) neural-only models as follows:

- Performance statistics: accuracy, precision, recall, and the AUROC. ROC curves may be viewed in Supplementary Figures 1, 2.
- Final predictors are ranked in the order of importance by their group-level SHAP score (average absolute value across the participant sample) and the mean predictor importance (group-level SHAP score) for the requisite experiment.
- Summary SHAP plots that graph individual-level final predictor importance (SHAP scores) for each member of the participant sample. SHAP summary plots are also used to determine the directionality of the relationship between the predictor and case status.

ADHD

Using multimodal data obtained at 9–10 years, deep learning optimized with IEL predicted future new onset cases of ADHD at 11–12 years with ∼86% accuracy, 0.92 AUROC, and precision and recall >80% (Table 5). When predicting all prevailing cases at 11–12 years, performance improved to ∼94% accuracy, ∼0.99 AUROC, and precision and recall >90%. When only neural features were used, performance fell by ∼6–9% in predicting new onset cases and up to 40% in prevailing cases. Neural-only models predicted new onset cases moderately well with 79% accuracy, 0.841 AUROC, and ∼74% precision and recall. Performance in predicting prevailing cases with neural-only features was poor, with ∼64% accuracy, 0.654 AUROC, and <60% precision and recall.

The presence of a disorder of excessive somnolence was the most important predictor of new onset case status in ADHD, with parent–child conflict present to a lesser degree (Table 6). The model that predicted all prevailing cases at 11–12 years was more complex. The most important predictors were whether the child had received mental health or substance abuse services before assessment at 9–10 years and the total level of parental behavioral problems. This was followed by conflict between parent and child, the presence of a sleep–wake transition or excessive somnolence disorder, and the level of parental externalizing behaviors. For both new onset and all prevailing cases, how well the child functioned at school and specifically having excellent grades in school had an inverse predictive relationship with ADHD case status. In prevailing cases, this was joined by the child's level of prosocial behaviors. In multimodal models where all feature types were available, the optimization process ran by IEL preferentially selected psychosocial features with no cognitive, neural, or biological metrics present in final, optimized models. Group-level importances for multimodal model predictors (averaged across the participant sample) were in the range [0.009, 0.20] and the mean importance for each experiment was in the range [0.06, 0.12].

In interpreting the neural-only experiments, we observed little overlap between the final, optimized models for new onset and all prevailing cases of ADHD. The only common feature was a negative relationship between case status and SST contrast in

TABLE 6 Final predictors of cases of ADHD at ages 11–12 years.

(A) Age of case determination	Ranked final predictors	Importance
New onset at age 11–12 years	Disorder of excessive somnolence	0.2038
	Does very well at school	0.1423
	Parent reports some conflict with child	0.0796
	Youth gets excellent grades at school	0.0567
	Mean	0.1206
All cases at age 11–12 years	Child has received MH/SU services in last 6 months	0.1638
	Parent total behavioral problems	0.1061
	Parent reports some conflict with child	0.0725
	Disorder of sleep-wake transition	0.0665
	Youth performs very well in school	0.0617
	Disorder of excessive somnolence	0.0375
	Youth receives excellent grades in school	0.0234
	Prosocial behaviors mean score	0.0147
	Parent externalizing problems	0.0091
	Mean	0.0617
(B) Neural data type	Ranked final predictors	Importance
New onset at age 11–12 years	T1 intensity in brain stem ROI	0.0890
	Correlation between ventral attention network and right ventral diencephalon ROI	0.0532
	SST any stop vs. correct go contrast in left pars opercularis ROI	0.0411
	SST incorrect stop vs. correct go contrast in left lingual ROI	0.0364
	T1 intensity WM for left lateral occipital ROI	0.0309
	Cortical thickness in mm of right transverse temporal ROI	0.0304
	MID loss anticipation vs. neutral contrast in right supramarginal ROI	0.0271
	Average FA in GM right caudal ACC ROI	0.0147
	Mean	0.0400
All cases at age 11–12 years	GM FA in right caudal middle frontal ROI	0.0446
	Cortical area in mm^2 of left inferior parietal ROI	0.0259
	MID small loss vs. neutral contrast in right inferior temporal ROI	0.0192
	SST incorrect go vs. incorrect stop contrast in left lingual ROI	0.0134
	Cortical thickness in mm of left pars triangularis ROI	0.0129
	SST incorrect stop vs. correct go in left lingual ROI	0.0046
	Mean	0.0200

Final predictors of all prevailing cases of ADHD at 11–12 years, as well as new onset cases only at 11–12 years of age, are shown for the most accurate models obtained using deep learning optimized with IEL obtained with **(A)** multimodal features and **(B)** only neural features. Final predictors are ranked in the order of importance, where the relative importance of each predictor is computed with the Shapley additive explanation technique and presented here averaged across all participants in the sample. Features in red indicate an inverse relationship with ADHD verified with the Shapley method. MH, mental health; SU, substance use; SST, Standard Stop Signal task; MID, Monetary Incentive Delay task; ROI, region of interest; FA, fractional anisotropy; WM, white matter; GM, gray matter.

the left lingual ROI, though the contrast effect differed between incorrect stop vs. correct go (new onset) and incorrect go vs. incorrect stop (prevailing cases). In new onset cases, the most prominent positive predictor was the correlation between the ventral attention network and right ventral diencephalon ROI, followed by SST contrast in the left pars opercularis and cortical thickness in the right transverse temporal ROI. Structural differences in the brain stem, left lateral occipital white matter, and right caudal ACC, along with MID contrast in the right supramarginal ROI, were negative predictors of new onset case status. The neural-only model of all prevailing ADHD cases was less reliable, with an AUROC of 0.654, but we found that structural features in the right caudal middle frontal and left pars triangularis ROIs predicted case status with inverse relationships with cortical area of the left parietal ROI and MID loss contrast in the right inferior temporal ROI. Group-level importances for neural-only model predictors were in the range [0.02, 0.04] and the mean importance for each experiment was in the range [0.0046, 0.089], both representing lower importance ranges than multimodal models.

We further computed and plotted individual-level SHAP values to quantify the dispersion of importances across individuals and assess the directionality of the relationship between final predictors and clinical case status (Figure 2). In these summary plots, each data point represents an individual participant, and the colorization reflects the original value of the predictor as an input feature. Thus, discrete-valued features appear as red or blue, whereas a continuous feature appears as a color gradient from low to high.

Individual-level importances in multimodal predictive models of both new onset and prevailing cases of ADHD were typically more widely dispersed than in neural-only models. Furthermore, wider dispersions across the participant samples were observed for the more important predictors.

Oppositional defiant disorder

In ODD, predictive models performed strongly using multimodal features (Table 7). In new onset cases, we achieved an accuracy of ∼97%, AUROC of 0.996, and precision and recall of ≥94%, and a 96% accuracy, AUROC of 0.988, and precision and recall of ≥95% when predicting all prevailing cases at 11–12 years. In neural-only models, we observed similar phenomena as in ADHD: performance fell substantially, with relatively better performance in predicting new onset vs. prevailing cases. When only neural features were used, performance fell by ∼20% in predicting new onset cases and up to ∼40% in prevailing cases. Neural-only models predicted new onset cases moderately well with 74% accuracy, 0.792 AUROC, and precision and recall ≥65%. Performance in predicting prevailing cases with neural features was poor, with ∼56% accuracy, 0.567 AUROC, and <55% precision and recall.

Whether the youth had ever received mental health or substance abuse services before assessment at age 9–10 years was the most important predictor of new onset case status in ODD, followed by the presence of a disorder of excessive somnolence or sleep–wake transition (Table 8). Additional important predictors
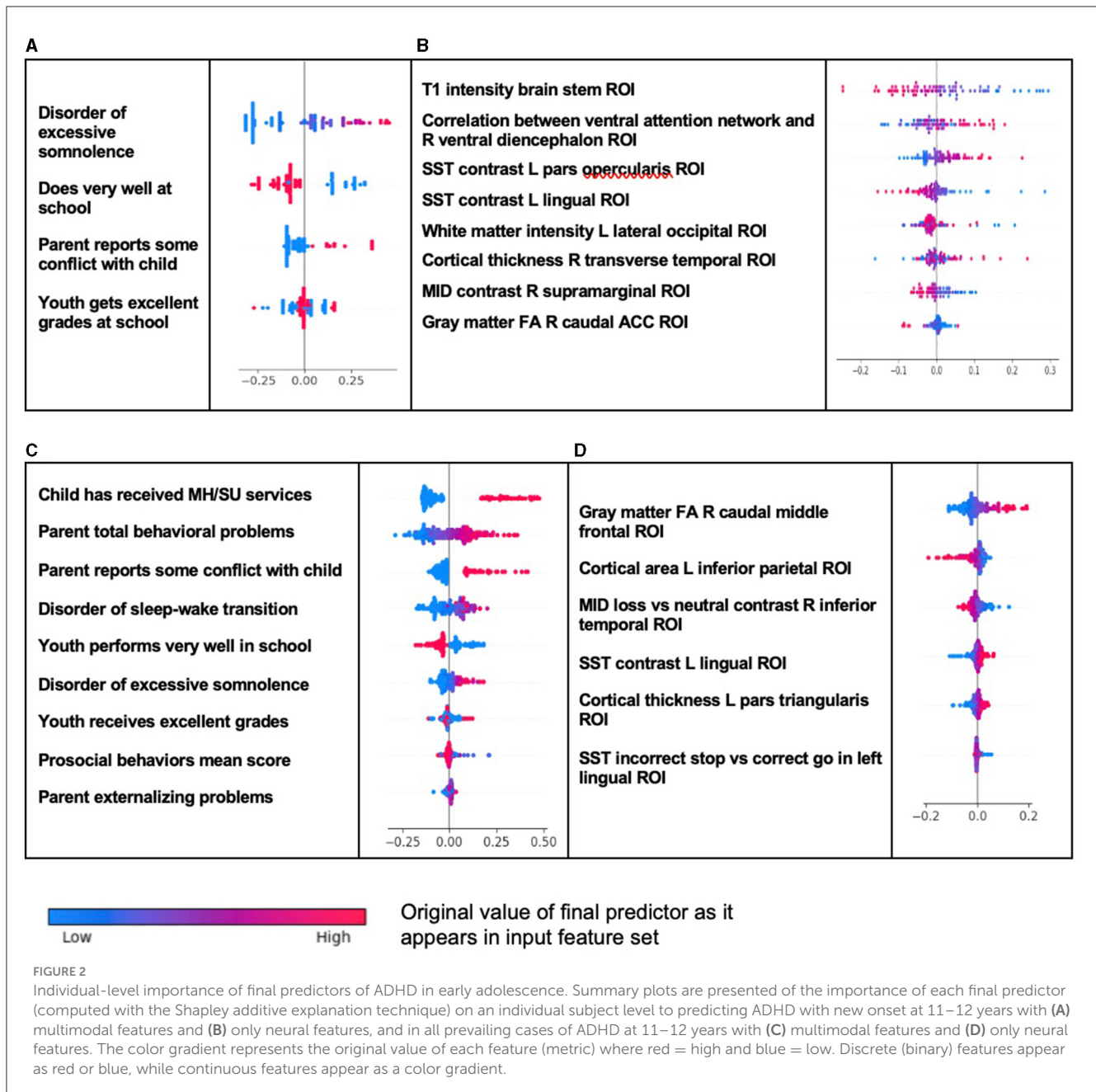
were parental factors: the presence of nerves or a nervous breakdown problem and levels of externalizing or aggressive behaviors. Youth prosocial behaviors exhibited an inverse relationship with case status. Features that predicted all prevailing cases at 11–12 years included a number of final predictors that were the same or thematically similar: whether the child had received mental health or substance abuse services in the last 6 months (the most important predictor), total sleep disturbances, disorder of sleep–wake transition, parent externalizing behaviors, and an inverse relationship with prosocial behaviors. The final predictors that differed in this model were the youth's mother having a depression problem and whether either parent had sought treatment for a mental or emotional problem. Of note, the latter predictor had an inverse relationship with case status, suggesting it was related to (an) untreated mental problem(s). In multimodal models, where all feature types were available, the optimization process run by IEL preferentially selected psychosocial features with no cognitive, neural, or biological metrics present in final, optimized models. Group-level importances for multimodal model predictors (averaged across the participant sample) were in the range [0.003, 0.18] and the mean importance for each experiment was at ∼0.07.

In neural-only models, the future onset of ODD at 11–12 years was predicted by a model (with moderately strong performance at AUROC = 0.79) containing only rsfMRI-derived correlations. Strikingly, every final predictor represented a correlation metric between a cortical network and subcortical ROI, emphasizing networks involved in salience, executive function, spatial memory, and task performance. Of note, all neural features with positive relationships with the onset of ODD were in the left hemisphere and those with inverse relationships with case status in the right hemisphere. As noted above, the neural-only model predicting all prevailing cases of ODD at 11–12 years exhibited poor performance (AUROC = ∼0.567) and cannot be considered reliable. It consisted of two structural gray matter features: fractional anisotropy of the right lateral orbitofrontal ROI and cortical area of the left inferior parietal ROI. Group-level importances for neural-only model predictors (averaged across the participant sample) were in the range [0.0007, 0.075] and the mean importance for each experiment was in the range [0.0026, 0.0410].

As observed in ADHD, individual-level importances in multimodal predictive models of both new onset and prevailing cases of ODD were typically more widely dispersed than in neural-only models (Figure 3). Furthermore, wider dispersions across the participant samples were observed for the more important predictors.

Conduct disorder

Deep learning optimized with IEL predicted future new onset cases of CD at 11–12 years with ∼90% accuracy, 0.92 AUROC, and precision and recall >85% using multimodal features assessed at 9–10 years (Table 9). In predicting all prevailing cases at 11–12 years, performance improved further to ∼96% accuracy, ∼0.99 AUROC, and precision and recall ≥95%. This strong predictive performance represented the best overall performance among the

FIGURE 2
Individual-level importance of final predictors of ADHD in early adolescence. Summary plots are presented of the importance of each final predictor (computed with the Shapley additive explanation technique) on an individual subject level to predicting ADHD with new onset at 11–12 years with **(A)** multimodal features and **(B)** only neural features, and in all prevailing cases of ADHD at 11–12 years with **(C)** multimodal features and **(D)** only neural features. The color gradient represents the original value of each feature (metric) where red = high and blue = low. Discrete (binary) features appear as red or blue, while continuous features appear as a color gradient.

three externalizing conditions. When only neural features were used, performance fell by ∼10% in predicting new onset cases and up to 20% in prevailing cases. However, this is in the context of neural-only models achieving moderately strong performance in predicting new onset cases with 80% accuracy, 0.808 AUROC, and precision and recall >70%. Performance in predicting prevailing cases with neural-only features was also moderately strong with ∼78% accuracy, 0.816 AUROC, and >70% precision and recall.

The interpretation of predictive models for CD was particularly intriguing (Table 10). Unlike ADHD and ODD, final predictors of both new onset cases and all prevailing cases at 11–12 years using multimodal data did include neural features. New onset cases of CD were predicted by psychosocial features also found in ADHD and ODD (tenor of parent–child relationship, sleep disturbances, and mental health treatment before the age of 9–10 years) but here these psychosocial factors interacted in an inverse relationship with structural disturbance in the left hippocampal ROI. Similarly, final predictors of all prevailing cases of CD at 11–12 years comprised psychosocial features common to ADHD and ODD (prior mental health treatment, tenor of parent–child relationship, sleep disturbances, and school performance), but these interacted with structural features in the left transverse temporal white matter and left caudal anterior cingulate cortex gray matter (inverse relationship). A further interesting facet of this latter model was that parent somatization traits were a driver of CD, where parent aggressive traits had an inverse relationship with case status. Somatization refers to the expression of mental phenomena as physical (somatic) symptoms seek medical care for

TABLE 7 Performance of deep learning optimized with integrated evolutionary learning in predicting cases of ODD using multimodal and neural-only feature types.

Age of case determination	Accuracy (%)	Precision (%)	Recall (%)	AUROC
(A)				
New onset at age 11–12 years	96.8	96.8	93.5	0.996
All cases at age 11–12 years	96.2	94.8	95.5	0.988
(B)				
New onset at age 11–12 years	74.2	69.4	64.5	0.792
All cases at age 11–12 years	56.1	53.9	27.2	0.567

Performance statistics of accuracy, precision, recall, and the AUROC are shown for the most accurate model obtained with deep learning optimized with integrated evolutionary learning using **(A)** multimodal features and **(B)** only neural features. We used features obtained at 9–10 years of age to predict new onset cases of ODD at 11–12 years of age, as well as all prevailing cases at 11–12 years of age. Corresponding ROC curves may be viewed in Supplementary Figures 1, 2.

them and placement of an undue focus on the distress caused by physical complaints.

In neural-only models, which performed relatively well in CD, prominent predictors of new onset cases were structural features in the right rostral middle frontal ROI, left hippocampus (as also found in the multimodal model), and right caudate. Less important features included the correlation between the cinguloopercular network and the left amygdala (also observed in ODD) and left transverse temporal ROI (also observed in the multimodal model). Final neural-only predictors of prevailing cases of CD were dominated by cortical-subcortical connectivity features comprising the cinguloopercular network with the left amygdala (also important to new onset prediction), auditory network with right hippocampus, and default mode network with right ventral diencephalon. This model was rounded out with structural gray matter differences in the left caudal ACC, also observed in the multimodal model. In both new onset and prevailing cases, there was an emphasis on subcortical structural features and connectivity between cortical networks and subcortical ROIs.

As observed in both ADHD and ODD, individual-level importances in multimodal predictive models of both new onset and prevailing cases of CD were typically more widely dispersed than in neural-only models (Figure 4). Furthermore, wider dispersions across the participant samples were observed for the more important predictors.

The relationship between accuracy and final predictor importance

We computed the mean predictor importance for each experiment to explore the relationship between model accuracy in testing in held-out, unseen data and final predictor importance after optimized, automated feature selection, for example, the average importance of final predictors of new onset ADHD at 11–12 years (Table 6). These data may be inspected in Supplementary Table 4. Furthermore, we computed the correlation and R^2 of the relationship between accuracy and mean predictor importance for each experiment described in the present study. Across all experiments, the correlation between accuracy and predictor importance in final, optimized models tested in held-out,

unseen data was 72.7% and the R^2 was 52.8%. This is summarized in Figure 5, where mean final predictor importance is shown plotted against log(accuracy) to improve scale interpretation, though we note that the reported correlation and R^2 were computed with accuracy.

Discussion

General observations across externalizing disorders

Using an AI-guided feature selection process, we were able to distill ∼6,000 candidate predictors contributed by children 9–10 years and their parents into robust, individual-level models predicting the later (11–12 years) onset of ADHD, ODD, and CD. This extended prior work in ML prediction of externalizing disorders in adolescence by assessing ∼30× more candidate predictors spanning a wider variety of knowledge domains (cognitive, psychosocial, biological, and multiple neural types). By imposing a common pre-processing and analytic design across all three major externalizing disorders in the same participant cohort, we were able to directly compare results, quantify the relative predictive performance of multimodal vs. neural features, and examine the relationship between predictor importance and model accuracy across multiple experiments. To the best of our knowledge, this is the first study using ML to predict the onset of all three major adolescent externalizing disorders and include many types of neural predictors (rsfMRI connectivity; task fMRI effects; diffusion and structural metrics), analyze >200 multimodal features, and quantify the relationship between predictor importance and accuracy.

Comparing experiments, we found that relative predictive performance varied according to disorder and predictor type (psychosocial vs. neural). Overall, deep learning optimized with IEL applied to multimodal features achieved strong performance with ∼86–97% accuracy, 0.919–0.996 AUROC, and ∼82–97% precision and recall in testing in held-out, unseen data. With multimodal features, performance was slightly stronger in predicting prevailing over new onset cases in ADHD and CD but equivalent in ODD, with the strongest performance overall in ODD, followed by CD and then ADHD. Further targeted experiments specifically assessed

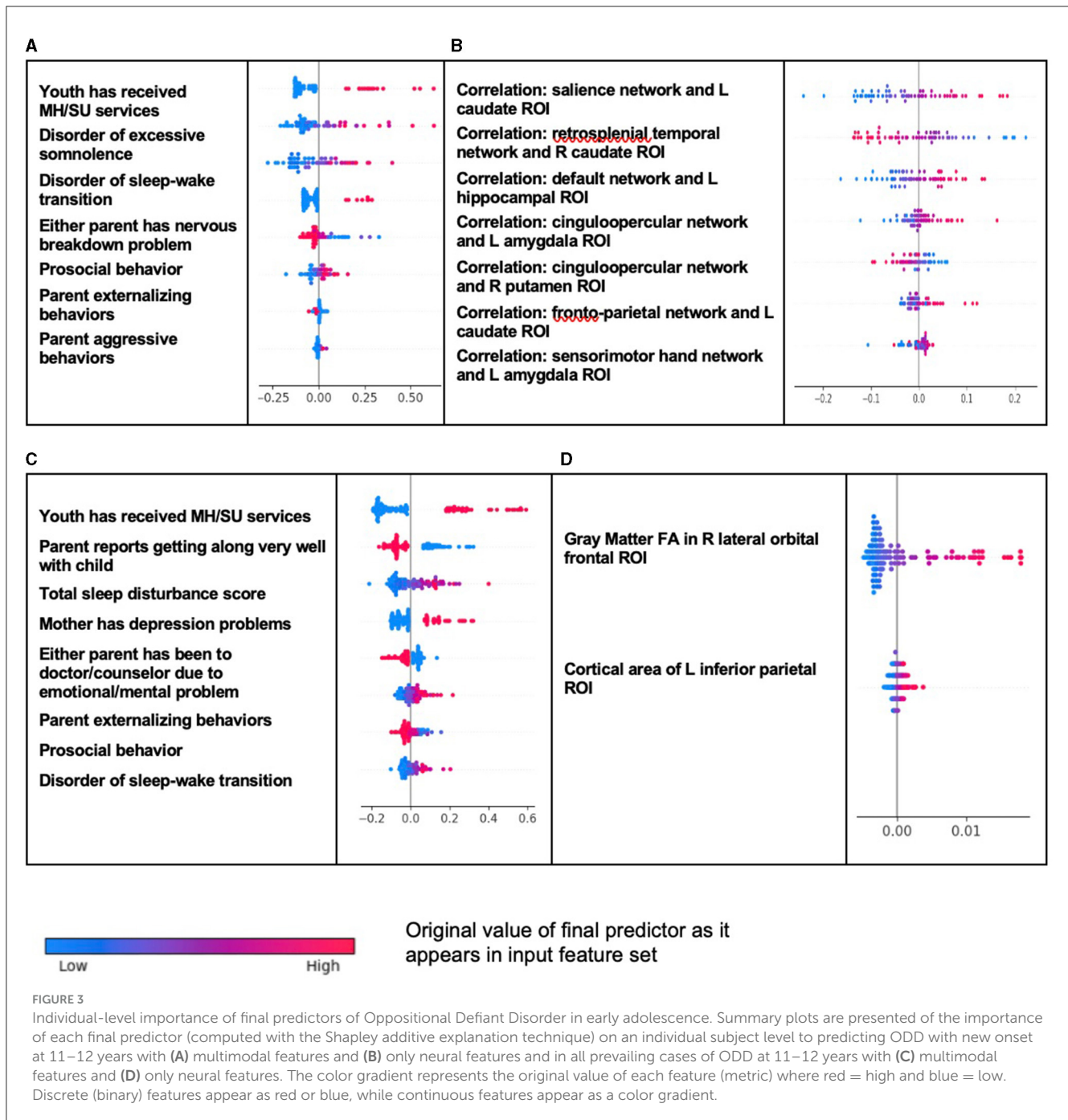TABLE 8 Final predictors of cases of ODD at ages 11–12 years.

(A) Age of case determination	Ranked final predictors	Importance
New onset at age 11–12 years	Youth has ever received MH/SU services	0.1581
	Disorder of excessive somnolence	0.1290
	Disorder of sleep-wake transition	0.0968
	Either parent has nerves or nervous breakdown problem	0.0871
	Prosocial behavior scale score	0.0516
	Parent externalizing problems	0.0290
	Parent aggressive behavior problems	0.0032
	Mean	0.0694
All cases at age 11–12 years	Child has received MH/SU services in last 6 months	0.1856
	Parent reports getting along very well with child	0.0951
	Total sleep disturbance score	0.0827
	Mother has depression problem	0.0819
	Either parent has been to doctor or counselor due to emotional/mental problem	0.0420
	Parent externalizing problems	0.0367
	Prosocial behavior mean score	0.0358
	Disorder of sleep-wake transition	0.0347
	Mean	0.0743
(B) Neural data type	**Ranked final predictors**	**Importance**
New onset at age 11–12 years	Correlation between salience network and left caudate ROI	0.0750
	Correlation between retrosplenial temporal network and right caudate ROI	0.0743
	Correlation between default network and left hippocampal ROI	0.0509
	Correlation between cinguloopercular network and left amygdala ROI	0.0262
	Correlation between cinguloopercular network and right putamen ROI	0.0256
	Correlation between fronto-parietal network and left caudate ROI	0.0207
	Correlation between sensorimotor hand network and left amygdala ROI	0.0158
	Mean	0.0410
All cases at age 11–12 years	Average FA in GM in right lateral orbitofrontal ROI	0.0044
	Cortical area in mm^2 of left inferior parietal ROI	0.0007
	Mean	0.0026

Final predictors of all prevailing cases of ODD at 11–12 years, as well as new onset cases only at 11–12 years of age, are shown for the most accurate models obtained using deep learning optimized with IEL obtained with **(A)** multimodal features and **(B)** only neural features. Final predictors are ranked in the order of importance, where the relative importance of each predictor is computed with the Shapley additive explanation technique and presented here averaged across all participants in the sample. Features in red indicate an inverse relationship with ODD verified with the Shapley method. MH, mental health; SU, substance use; ROI, region of interest; GM, gray matter.

the standalone predictive ability of multiple neural feature types derived from MRI. After restricting the candidate predictors to 4,777 neural features, we observed that predictive performance dropped substantially across all three disorders, most prominently when predicting all prevailing cases. The small number of prior ML studies in adolescent externalizing disorders that have directly compared the utility of psychosocial vs. neural predictors have obtained similar results and performance differentials (14). However, we would highlight that neural-only features were for the most part able to predict new onset cases with accuracy and AUROC of ~80%. While not as strong as with multimodal features, this performance compares favorably with the existing literature

using ML and biobehavioral features to predict externalizing disorders in adolescents. Table 11 provides an overview of selected comparable studies.

To the best of our knowledge, this is the first study to provide directly comparable predictive models of all three major externalizing disorders. In adolescence, ADHD, ODD, and CD frequently co-occur in the population, and in adulthood, they are increasingly co-morbid with mental health conditions such as internalizing and personality disorders and substance use. It is, therefore, challenging to assemble a longitudinal cohort where participants have only ADHD, ODD, or CD without any co-morbidities, and we are not aware that such a sample exists in

FIGURE 3
Individual-level importance of final predictors of Oppositional Defiant Disorder in early adolescence. Summary plots are presented of the importance of each final predictor (computed with the Shapley additive explanation technique) on an individual subject level to predicting ODD with new onset at 11–12 years with **(A)** multimodal features and **(B)** only neural features and in all prevailing cases of ODD at 11–12 years with **(C)** multimodal features and **(D)** only neural features. The color gradient represents the original value of each feature (metric) where red = high and blue = low. Discrete (binary) features appear as red or blue, while continuous features appear as a color gradient.

adolescence with sufficient participants to enable rigorous ML analyses. Moreover, allowing naturalistic sample overlap among the externalizing conditions may improve translational relevance because it reflects the clinical population. Here, we adopted a design where all three disorders are predicted in the same cohort using the same methods to allow head-to-head comparison of final predictors and enable the identification of common vs. specific predictors across ADHD, ODD, and CD in the same population. We found that each set of final predictors was a unique combination of features and differentiated both (a) ADHD, ODD, and CD from each other and (b) future new onset from all prevailing cases. However, there were cross-cutting themes. In predicting case onset,

sleep disorders (excessive somnolence, sleep–wake transition, and total disturbances) were common, prominent predictors across ADHD, ODD, and CD. Sleep disturbances may affect up to ~40% of elementary school-age children and youth, with both internalizing and externalizing disorders at elevated risk (43, 44). Sleep disturbances have been shown to "precede, predict and significantly contribute" to behavioral issues in ADHD and worsen disruptive behaviors, ODD, and CD in adolescence, though links with sleep latency and duration have been variable (45–48). Here, our findings add to a growing body of work suggesting that sleep disturbances may be important intervention targets in elementary school-age youth to reduce the later onset of clinical ADHD, ODD,

TABLE 9 Performance of deep learning optimized with integrated evolutionary learning in predicting cases of conduct disorder using multimodal and neural-only feature types.

Age of case determination	Accuracy (%)	Precision (%)	Recall (%)	AUROC
(A)				
New onset at age 11–12 years	90.0	85.4	92.0	0.922
All cases at age 11–12 years	96.3	95.1	95.0	0.990
(B)				
New onset at age 11–12 years	80.0	75.7	72.0	0.808
All cases at age 11–12 years	77.5	71.7	75.0	0.816

Performance statistics of accuracy, precision, recall, and the AUROC are shown for the most accurate model obtained with deep learning optimized with integrated evolutionary learning using (**A**) multimodal features and (**B**) only neural features. We used features obtained at 9–10 years of age to predict new onset cases of CD at 11–12 years of age, as well as all prevailing cases at 11–12 years of age. Corresponding ROC curves may be viewed in Supplementary Figures 1, 2.

and CD. Moreover, we found that daytime somnolence and sleep–wake transition were emphasized in predicting the externalizing disorders in adolescence and not sleep latency or duration. Other themes were shared by two of the three disorders: conflict between parent and child was shared in ADHD and ODD, and in the more behaviorally severe disorders (ODD, CD), youth appeared to have come to clinical attention before age 9–10 years. In our models for all prevailing cases, shared themes were recent mental health treatment for the youth, sleep disturbances and parental burden of various types of behavioral problems, and parent–child conflict. Unsurprisingly, therefore, there are thematically common predictors across all three externalizing disorders that also reflect the extant literature. These may present opportunities to leverage both conventional interventions and newly emerging therapies, such as digital therapies using mobile devices and applications in early adolescents at risk for externalizing disorders (49–54). However, disorder-specific predictors did exist that may aid in disambiguating the onset of these conditions. Most strikingly, CD was marked by the importance of structural brain features that interacted with psychosocial predictors and appeared in neither ADHD nor ODD in multimodal models. As well, neural-only models achieved their best performance in CD over ADHD or ODD. This highlights a potential role for structural neuroimaging in identifying youth at risk for CD, the most severe and disabling of the three disorders, vs. ADHD or ODD. In terms of the latter two conditions, school performance was a prominent predictor of the onset of ADHD vs. an emphasis on lower levels of prosocial behaviors and parent mental health issues in ODD.

Recent studies sugges t that inflated effect sizes in neuroimaging studies of psychopathology and cognitive traits may be responsible for generalization failure, particularly in group-level association studies and smaller participant samples (55). While there is no exact equivalent to group-level effect size in the individual-level models provided by deep learning with artificial neural networks, predictor importance in the context of accuracy is conceptually similar. We, therefore, investigated predictor importance at both the group and individual level and its relationship with model performance in generalization testing, finding a moderately strong relationship (R^2 ∼53%) between predictor importance and accuracy. Psychosocial predictors in multimodal models had larger importance and wider inter-individual dispersions than those in neural-only experiments, even after extensive optimization and

principled feature selection. Collectively, these results suggest that the smaller importance of neural features and their more restricted inter-individual variability were at least related to their weaker performance in predicting cases. Future work will be required to determine whether these phenomena are seen in other disorders and participant samples or if other types of neural features might perform differently in predicting cases of externalizing disorders.

Predicting the onset of ADHD in early adolescence

ADHD affects up to 10% of school-age children and is characterized by inattention, impulsivity, and hyperactivity. It is a developmental disorder that shows markedly increasing prevalence from late elementary school through adolescence and is treatable. Thus, the early detection of children at risk for new onset is of substantial interest. There have been a number of ML multimodal predictive studies in adolescent ADHD, predominantly cross-sectional. National-level cohorts have offered large sample sizes to enable ML but typically a smaller range of psychosocial/demographic candidate predictors. For example, Garcia-Argibay et al. analyzed 22 candidate predictors in Swedish registry data ($n = 238,696$), achieving moderate performance with deep learning (accuracy: 69%, AUROC: 0.75) and identifying top predictors of having a parent with criminal convictions or relative with ADHD, male sex, number of academic subjects failed, and speech/learning disabilities (39). In a Japanese sample ($n = 45,779$), Maniruzzaman et al. identified family structure, insurance age, sex, medical conditions, and mental health symptomatology as significant among 19 psychosocial candidate predictors (accuracy: 86%, AUROC: 0.94) (40). Using a British school-based cohort, Ter-Minassian et al. were able to access a wider range of 68 candidate predictors and found school attendance, social-emotional development level, writing performance, male sex, and problem-solving/reasoning to be most important in predicting ADHD (AUROC: 0.72) (41). Analyzing ∼6,000 candidate multimodal predictors, we found that the onset of ADHD in early adolescence was robustly (accuracy: ∼86%, AUROC: 0.919) predicted by a simple model comprising the

TABLE 10 Final predictors of cases of conduct disorder at ages 11–12 years.

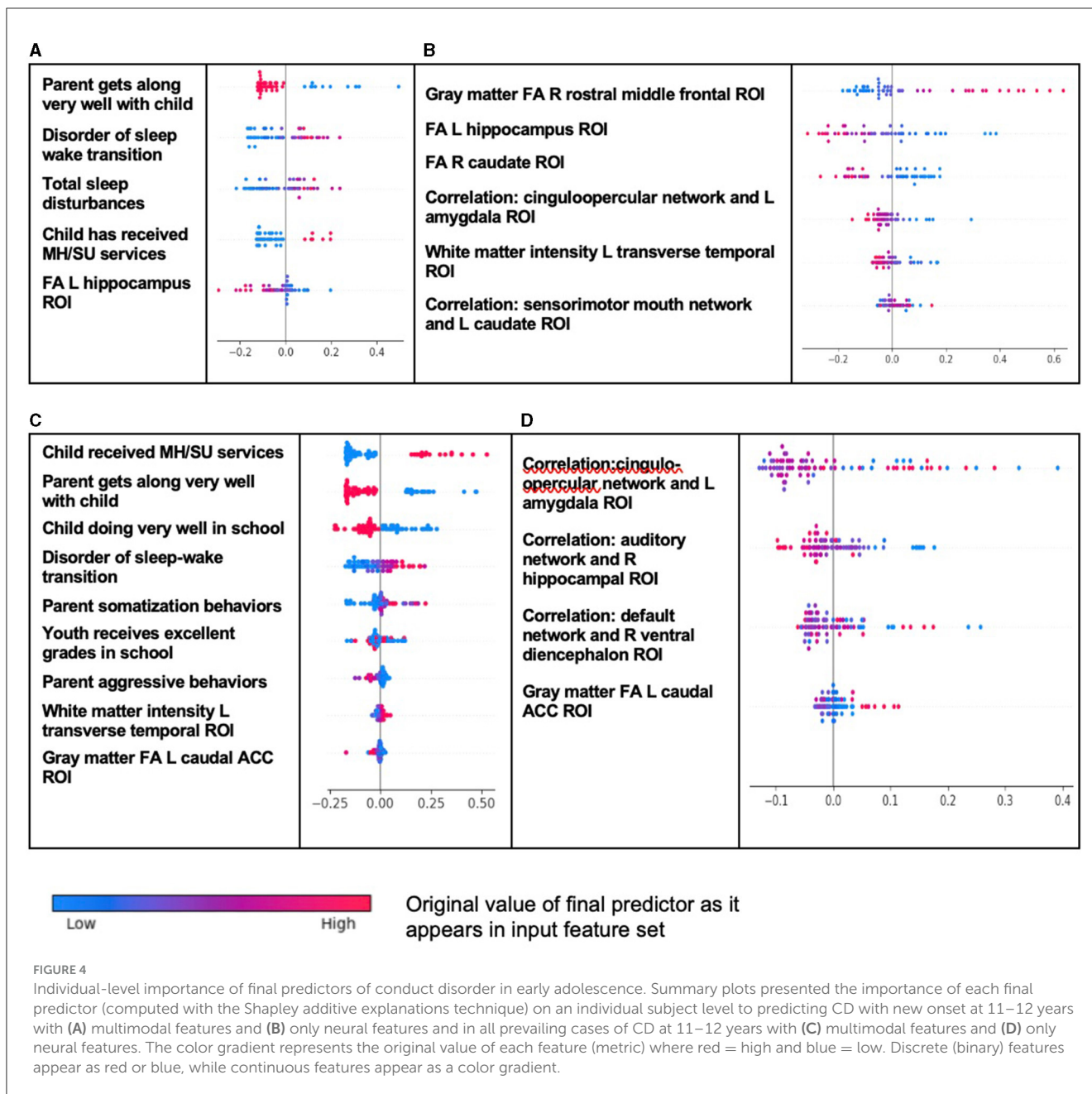| (A) Age of case determination	Ranked final predictors	Importance
New onset at age 11–12 years	Parent reports gets along very well with child	0.1126
	Disorder of sleep wake transition	0.0996
	Total sleep disturbances	0.0949
	Child has ever received MH/SU services	0.0927
	Average FA left hippocampus ROI	0.0642
	Mean	0.0928
All cases at age 11–12 years	Child has received MH/SU services in last 6 months	0.1593
	Parent reports getting along very well with child	0.1442
	Child is doing very well in school	0.0958
	Disorder of sleep-wake transition	0.0806
	Parent somatization behaviors score	0.0455
	Youth receives excellent grades in school	0.0408
	Parent aggressive behavior score	0.0235
	T1 intensity WM left transverse temporal ROI	0.0130
	Average FA GM left caudal ACC ROI	0.0123
	Mean	0.0683
(B) Neural data type	**Ranked final predictors**	**Importance**
New onset at age 11–12 years	Average FA GM right rostral middle frontal ROI	0.1628
	Average FA left hippocampus ROI	0.1348
	Average FA right caudate ROI	0.1043
	Correlation between cinguloopercular network and left amygdala ROI	0.0531
	T1 intensity WM left transverse temporal ROI	0.0490
	Correlation between sensorimotor mouth network and left caudal ROI	0.0333
	Mean	0.0890
All cases at age 11–12 years	Correlation between cinguloopercular network and left amygdala ROI	0.0960
	Correlation between auditory network and right hippocampal ROI	0.0456
	Correlation between default network and right ventral diencephalon ROI	0.0444
	Average FA GM in left caudal ACC ROI	0.0192
	Mean	0.0513

Final predictors of all prevailing cases of CD at 11–12 years, as well as new onset cases only at 11–12 years of age, are shown for the most accurate models obtained using deep learning optimized with IEL obtained with **(A)** multimodal features and **(B)** only neural features. Final predictors are ranked in the order of importance, where the relative importance of each predictor is computed with the Shapley additive explanation technique and presented here averaged across all participants in the sample. Features in red indicate an inverse relationship with CD verified with the Shapley method. MH, mental health; SU, substance use; ROI, region of interest; FA, fractional anisotropy; WM, white matter; GM, gray matter.

presence of a disorder of excessive somnolence, two metrics of poor school performance and parent–child conflict.

Sleep disturbances are widely reported in ADHD, including longer sleep latency, frequent awakenings, non-restorative sleep, decreased sleep, and daytime somnolence (46, 56). Although many children with ADHD are treated with stimulants, the evidence that this disrupts sleep is inconclusive, though sleep disturbances are thought to worsen neurocognitive outcomes (57). In the present study, we included many types of sleep disorders and metrics as candidate predictors (Supplementary Table 1) and identified excessive somnolence as the most important predictor of the future onset of ADHD at 11–12 years when measured in children with

ages 9–10 years who have not been diagnosed with ADHD or taken stimulants. Thus, our findings extend prior work by suggesting that excessive daytime somnolence rather than other sleep metrics may be a helpful predictor of future ADHD case status. Excessive somnolence could be caused by a variety of developmental or environmental factors in school-age children, and future work may provide a mechanistic explanation of how it predicts ADHD onset. As noted above, poor school performance and family dysfunction have previously been identified as a predictor of ADHD and are well-associated with the disorder. Here, we add to this literature by identifying poor school performance and parent–child conflict as prospective predictors of ADHD onset in early adolescence.

FIGURE 4

Individual-level importance of final predictors of conduct disorder in early adolescence. Summary plots presented the importance of each final predictor (computed with the Shapley additive explanations technique) on an individual subject level to predicting CD with new onset at 11–12 years with **(A)** multimodal features and **(B)** only neural features and in all prevailing cases of CD at 11–12 years with **(C)** multimodal features and **(D)** only neural features. The color gradient represents the original value of each feature (metric) where red = high and blue = low. Discrete (binary) features appear as red or blue, while continuous features appear as a color gradient.

Of note, final predictors of new onset ADHD were essentially a subset of those that predicted all prevailing cases, whereas in the latter parent, behavioral traits of total and externalizing problem behaviors were also present.

We found that the prospective prediction of ADHD onset in early adolescence was not improved by neural features. However, our neural-only model of ADHD onset did achieve moderately strong performance (accuracy: ~79%, AUROC: 0.841) and is of interest. The neural substrate of ADHD has been extensively studied in group-level associative work. More recently, the construction of ML classifiers with neural features was stimulated by the formation of the aggregated ADHD-200 dataset and associated Global Competition, though many resultant studies have been criticized for reporting "inflated" performance statistics based on cross-validated training rather than testing for generalization in held-out, unseen data (58). Among the latter, performance has varied widely, with accuracy rarely surpassing 80% and most studies analyzing ADHD-200 cross-sectional data with a wide age span. We are not aware of other studies using ML for prospective prediction of the onset of adolescent ADHD using a comparably large number of neural features across multiple MRI types in a standardized cohort. In new onset cases, we found the most prominent predictor was the correlation between the ventral attention network and right ventral diencephalon ROI, followed by SST contrast in the left pars opercularis (Brodmann Area 44) and cortical thickness in the right transverse temporal ROI (linked with the processing of incoming auditory information). The ventral attention network is one of the primary attention networks in the
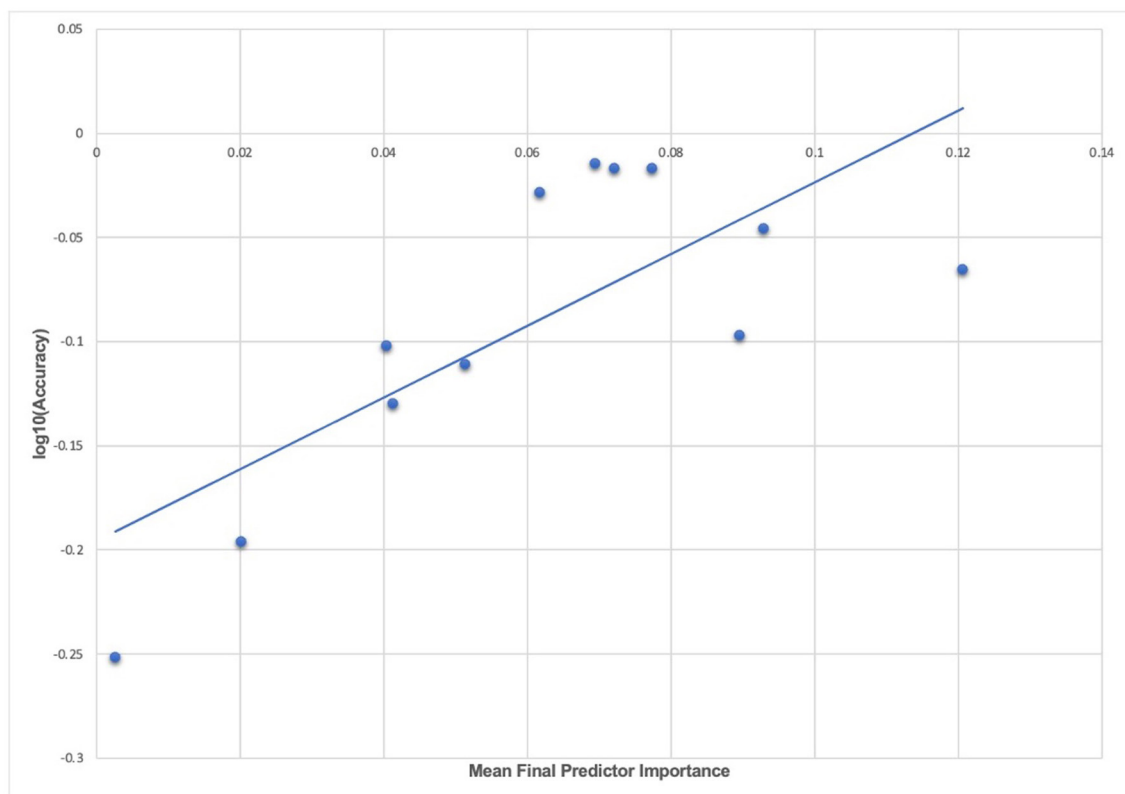
FIGURE 5
The relationship between accuracy and final predictor importance. Average variable importance computed with the Shapley additive explanations technique is shown plotted against the log of prediction accuracy in testing in held-out data for each experiment in the study. The line of best fit obtained with a linear regression is also displayed. The underlying data for this chart may be inspected in Supplementary Table 4.

TABLE 11 Overview of selected comparable studies.

Condition	References	Sample size	Type	Algorithm	Accuracy	Precision	Recall	AUROC
ADHD	Garcia-Argibay et al. (39)	238,696	CS	DL	69%	NR	72%	0.75
ADHD	Maniruzzaman et al. (40)	45,779	CS	RF	86%	NR	86%	0.94
ADHD	Ter-Minassian et al. (41)	4,178	CS	LR	81%	70%	66%	0.72
Disruptive behaviors	Menon and Krishnamurthy (42)	1,100	CS	CNN	72%	NR	70%	0.74
CD	Chan et al. (14)	2,368	Long	FNN	91%	NR	93%	0.96

Summary of comparable studies using machine learning to predict externalizing disorders in youth that are reviewed in this manuscript. Numbers after the authors' names correspond to the citation reference. CS, cross-sectional; Long, longitudinal; DL, deep learning; RF, random forest; LR, logistic regression; CNN, convolutional neural network; FNN, feedforward neural network.

brain and directs attention to unexpected stimuli. It has been very well-associated with ADHD symptomatology in both children and adults, as have differences in subcortical structures (59–63). Among subcortical structures, the diencephalon was historically less studied in ADHD. However, the thalamus, a primary component of the diencephalon that modulates and filters interfering stimuli, has recently attracted much attention to structural thalamic differences identified in youth with ADHD (64–66).

Predicting the onset of oppositional defiant disorder in adolescence

Oppositional defiant disorder is characterized by a pattern of uncooperative, defiant, and angry behavior toward authority figures that causes significant problems at home or school. Similar to ADHD, a proportion of youth with ODD "grow out" of the condition, and ∼50% of youth with ODD have

ADHD. Of those in whom ODD persists, CD may evolve, and in adulthood, ~40% go on to develop antisocial personality disorder and/or other mental health or substance abuse problems. Since the prevalence of ODD climbs markedly in elementary school and empirically based treatment is available, identifying specific prospective predictors of the disorder in late childhood and early adolescence—particularly those that differentiate it from ADHD—is of considerable importance. ODD and CD are often grouped as "disruptive disorders", and unfortunately, few large-scale ML studies have approached ODD in isolation. To the best of our knowledge, this study represents the first to analyze a large number of multimodal predictors, including multiple types of neuroimaging, to prospectively predict ODD as distinct from CD in early adolescence. We found that deep learning optimized with IEL prospectively predicted the onset of ODD with strong performance using multimodal features (accuracy: ~97%, AUROC: 0.996) in held-out, unseen data. While sleep disorders were the final predictors that ODD shared with ADHD, ODD had a more complex predictive model that additionally included several measures of parental mental health problems (either parent has depression, i.e., nerves or nervous breakdown problem; parent externalizing and aggressive problems) but did not include the metrics of school performance that predicted ADHD onset. Indeed, the most important predictor was whether the child had already come to clinical psychiatric attention before age 9–10 years. Here, our study is concordant with extant group-based studies in ODD, which associate case status with stress and conflict, parental depression, and other parental factors such as hostility, support, and scaffolding, and further suggest that symptoms are present in preschool and "cascade" toward eventual diagnosis with parental mental health problems significantly moderating treatment outcome (67–71).

As with ADHD, we found that biological and physiologic metrics were not selected in multimodal prospective prediction of the onset of ODD and that neural-only models sacrificed substantial performance. However, our neural-only model still obtained moderately strong performance (accuracy: 74%, AUROC: 0.792) and a striking result worthy of examination. While 5,777 neural features across multiple neuroimaging types were analyzed, ODD onset at 11–12 years was predicted by a markedly homogenous combination of features that were all rsfMRI metrics representing connectivity between cortical networks and subcortical ROIs, in particular limbic regions of the amygdala and caudate and putamen (dorsal striatum). Limbic regions in the left hemisphere predicted case status, while those in the right hemisphere had an inverse relationship with ODD. Moreover, cortical networks selected as final predictors had intuitive relationships with ODD symptomatology, being associated with navigating and integrating learned social rules, hierarchies, and contingencies (salience); empathy and introspection (default); efficient task switching (cinguloopercular); and executive control (fronto-parietal) (72–78). Many have known limbic nodes where the latter structures are associated with fear and threat detection and the autonomic "fight or flight" response (amygdala) reinforcement learning and action selection (dorsal striatum) (79–81). As noted above, few large-scale ML predictive studies have focused exclusively on ODD or made head-to-head comparisons

among externalizing disorders, including neuroimaging studies. Menon and Krishnamurthy predicted disruptive behaviors (collapsing ODD and CD into one category) in children aged 9–10 years in the ABCD cohort using a convolutional neural network applied to three types of neural features (diffusion, structural, and seed-based rsfMRI connectivity) obtained at 9–10 years to examine the relative predictive power of each type of imaging (42). They obtained moderate performance (accuracy: 0.72, AUROC: 0.74) without testing in held-out data and found that a combination of modalities performed better than any single imaging type. The right superior longitudinal fasciculus, middle frontal, postcentral, middle occipital and middle temporal gyri, and inferior parietal lobule were class discriminative in disruptive behaviors. Thus, the current study suggests an intriguing jumping-off point for neural prediction of ODD development in suggesting a focus on cortical-subcortical relationships centered around connectivity between cortical control networks and limbic loci performing emotional response and action selection. In this, ODD contrasts with the attention and language processing networks and areas emphasized in ADHD onset.

Predicting the onset of conduct disorder in early adolescence

While conduct disorder may be grouped with ODD as the "disruptive" disorder, it is differentiated by the presence of aggression and destructive behaviors directed toward people, animals, or property, a serious violation of rules, and lack of empathy. The CD is often considered the most severe and disabling of adolescent externalizing disorders. While ~60–70% of youth will lose the diagnosis in adulthood, those that do not have a relatively poor prognosis are associated with the development of other mental health and substance abuse disorders, antisocial personality disorder, and life impairment, including involvement in the justice system. In prior group-based longitudinal studies, the development of CD has been associated with impulsivity, parental behaviors such as poor supervision and punitive discipline, cold or antisocial parental traits and parental conflict, family risk factors such as large size or low income, contextual factors such as antisocial peers, and adverse school or neighborhood environments (82). In the only multimodal ML classification study previously performed in CD specifically, Chan et al. also used data collected from children aged 9–10 years in the ABCD cohort to predict all prevailing cases of CD at 11–12 years (14). This study employed artificial neural networks and 52 candidate predictors comprising 20 graph metrics computed from rsfMRI, 16 psychosocial features selected empirically based on prior literature, four basic demographic descriptors, and nine cognitive metrics derived from psychometrics testing. In contrast to the present study, CD, ADHD, and ODD symptomatology at 9–10 years were also allowed as candidate predictors. This design achieved 91% accuracy and 0.96 AUROC compared to our 96% accuracy and 0.99 AUROC in the prospective prediction of all prevailing cases of CD at 11–12 years. They found that greater ADHD and ODD symptomatology, frontoparietal efficiency, and reports of family members throwing objects predicted future CD, while lower crystallized cognitive and card sorting ability,

subcortical efficiency, frontoparietal degree, family income, and parental monitoring were inversely related to case status. This study is comparable to our multimodal model predicting all prevailing cases at 11–12 years, though we analyzed a larger number of candidate predictors and more types of neural metrics. We similarly found parent–child conflict to be an important final predictor, but otherwise, final predictors emphasized sleep disturbance (sleep–wake transition), poor school performance, parent somatization and aggressive traits, and structural brain differences in the left transverse temporal and anterior cingulate cortex (ACC) ROIs.

When focusing on predicting new onset cases of CD with multimodal features, we identified a more parsimonious model with strong predictive performance (accuracy: 90%, AUROC: 0.922) where family conflict, sleep disturbances (sleep–wake transition disorder and total disturbances), and whether the child had come to clinical attention before 9–10 years were important predictors. It is notable that among the three adolescent externalizing disorders, CD is the only condition in which neural predictors were selected as final predictors among ~6,000 multimodal candidate predictors after extensive AI-guided feature selection. In prospectively predicting the onset of CD, structural differences in the left hippocampus ROI interacted with psychosocial factors to drive the prediction of case status. Prior group-based studies (including in the ABCD cohort) have identified associations between CD symptomatology and structural and functional differences in the limbic system (which includes the hippocampus), ACC, orbito-frontal, prefrontal, and temporal cortices, though not all studies segregate CD from ODD (83–86). Extant work has also specifically identified aberrant volumes in paralimbic structures, including hippocampal ROIs in incarcerated adults and youth with psychopathic traits (87–89). Although the hippocampus is known for its role in memory formation, it is deeply interconnected with other limbic structures and plays a prominent role in fear conditioning and affective processes (90).

While neural features proved more important in the multimodal prediction of CD vs. ADHD and ODD, when we restricted candidate predictors to only neural features, performance dropped substantially. Similarly, Chan et al. found that accuracy dropped to 77% and AUROC to 80% when only neural features were used to make prospective predictions of all prevailing CD cases at 11–12 years, a very comparable performance differential. However, the moderately strong performance was still obtained (accuracy: 80%, AUROC: 0.808), giving credence to these findings. In the neural-only model, features in good concordance with prior literature were identified, with differences in frontal, temporal, and limbic (caudate, amygdala, and hippocampus) structures and connectivity between the cinguloopercular network and amygdala appearing as important predictors of CD onset in early adolescence. While some regional ROIs, particularly limbic structures, were common to ODD and CD, we found that neural predictors of CD onset emphasized structural over connectivity features and the hippocampus appeared as limbic structural predictor that was specific to CD.

Conclusion

Taken together, our results suggest that highly accurate (>85%) prediction of the onset of each of the early adolescent externalizing disorders is possible using ML optimized with AI and that individual-level prospective prediction of ADHD, ODD, and CD benefits from the inclusion of multimodal features drawn from multiple knowledge domains, particularly psychosocial predictors related to sleep disorders, parent mental health and behavioral traits, and school performance. In CD specifically—but not ADHD or ODD—metrics derived from structural MRI interacted with psychosocial features in predicting later case onset and these neural features and may hold particular promise in identifying children at risk for this highly disabling disorder. Cognitive features derived from psychometric testing and other forms of physiologic data (e.g., hormonal levels and biometric testing) were deemphasized throughout our experiments. Among neural features, metrics related to subcortical ROIs and connectivity between cortical and subcortical ROIs were prominent and congruent with the existing literature on externalizing disorders. In terms of MRI modalities, structural gray and white matter features and rsfMRI-derived connectivity were valuable prospective predictors across all three disorders, with tfMRI only appearing in ADHD and no diffusion MRI metrics featured. We achieved high performance across all multimodal experiments and identified a strong correlation between accuracy and final predictor importance, suggesting that automated feature selection with AI techniques such as IEL can facilitate the discovery of impactful predictors among high-dimension data in a principled manner and generate robust predictive models.

Limitations

This study uses secondary data from the ABCD study. We were, therefore, unable to control for any bias during data collection, and there is a mild bias toward higher-income participant families of white race in the early adolescent cohort, though the ABCD study strived for population representation. Similarly, externalizing disorders have shown differences in population case ascertainment associated with characteristics such as sex/gender, race/ethnicity, and sociodemographic factors, which have varied over time and are still the subject of ongoing research. We do not take a position on these phenomena in this study and constructed balanced samples based on case ascertainment using the CBCL. Cases were matched with controls based on natal sex and age. However, sex, gender, race, ethnicity, and many sociodemographic and cultural factors were included as candidate predictors, as is standard practice in large-scale ML studies and which does allow for the influence of such factors to be revealed. Further work using a similar design in participant samples stratified by sex/gender or race/ethnicity could also elucidate differential effects. Data are not available before baseline (age 9–10 years) assessment, and we cannot, therefore, conclusively rule out that youth participants may have met criteria based on the CBCL for ADHD, ODD, or CD before ≤8 years. Thus, it is possible that certain cases coded as "new onset" at 11–12 years of age could have met clinical criteria at ≤8 years but not at 9–10 years. In the present study, we defined cases as an individual meeting ASEBA clinical thresholds in CBCL subscale scores pertinent to ADHD, ODD, and CD and did not exclude participants who thereby met criteria for other conditions. Thus, co-morbidity may be present in the experimental samples, as is common in clinical populations and occurs in most research studies

in early adolescence. Our study is not exhaustive. It is possible that different results could have been obtained if more or different candidate predictors were included. We tested for generalization in a holdout, unseen test set obtained by partitioning the data, a gold standard method in ML. However, methods and results should also be tested for replication in an external dataset other than ABCD.

Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

Ethics statement

Ethical approval was not required for the study involving humans in accordance with the local legislation and institutional requirements. Written informed consent to participate in this study was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and the institutional requirements.

Author contributions

NL: Conceptualization, Data curation, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Validation, Writing—original draft, Writing—review & editing. MR: Data curation, Formal analysis, Methodology, Software, Validation, Writing—review & editing.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fpsyt.2023.1280326/full#supplementary-material

References

1. Mazzucchelli TG, Sanders MR. Children with externalizing behavior problems. In: Mazzucchelli MRSTG, editor. *The Power of Positive Parenting: Transforming the Lives of Children, Parents, and Communities Using the Triple P System*. Oxford: Oxford University Press (2018). p. 85–96.

2. Danielson ML, Bitsko RH, Holbrook JR, Charania SN, Claussen AH, McKeown RE, et al. Community-based prevalence of externalizing and internalizing disorders among school-aged children and adolescents in four geographically dispersed school districts in the United States. *Child Psychiatry Hum Dev.* (2021) 52:500–14. doi: 10.1007/s10578-020-01027-z

3. Danielson ML, Bitsko RH, Ghandour RM, Holbrook JR, Kogan MD, Blumberg SJ. Prevalence of parent-reported ADHD diagnosis and associated treatment among U.S. children and adolescents, 2016. *J Clin Child Adolesc Psychol.* (2018) 47:199–212. doi: 10.1080/15374416.2017.1417860

4. Kessler RC, Adler L, Barkley R, Biederman J, Conners CK, Demler O, et al. The prevalence and correlates of adult ADHD in the United States: results from the National Comorbidity Survey Replication. *Am J Psychiatry.* (2006) 163:716–23. doi: 10.1176/ajp.2006.163.4.716

5. Boat TF, Wu JT. *Prevalence of Oppositional Defiant Disorder and Conduct Disorder. Mental Disorders and Disabilities Among Low-Income Children 13.* Washington, DC: National Academies Press (US) (2015).

6. Bitsko RH, Claussen AH, Lichstein J, Black LI, Jones SE, Danielson ML, et al. Mental health surveillance among children - United States, 2013-2019. *MMWR Suppl.* (2022) 71:1–42. doi: 10.15585/mmwr.su7102a1

7. Perou R, Bitsko RH, Blumberg SJ, Pastor P, Ghandour RM, Gfroerer JC, et al. Mental health surveillance among children–United States, 2005-2011. *MMWR Suppl.* (2013) 62:1–35.

8. Nock MK, Kazdin AE, Hiripi E, Kessler RC. Lifetime prevalence, correlates, and persistence of oppositional defiant disorder: results from the National Comorbidity Survey Replication. *J Child Psychol Psychiatry.* (2007) 48:703–13. doi: 10.1111/j.1469-7610.2007.01733.x

9. Herrero ME, Hechtman L, Weiss G. Antisocial disorders in hyperactive subjects from childhood to adulthood: predictive factors and characterization of subgroups. *Am J Orthopsychiatry.* (1994) 64:510–21. doi: 10.1037/h0079567

10. Bird HR, Canino G, Rubio-Stipec M, Gould MS, Ribera J, Sesman M, et al. Estimates of the prevalence of childhood maladjustment in a community survey in Puerto Rico. The use of combined measures. *Arch Gen Psychiatry.* (1988) 45:1120–6. doi: 10.1001/archpsyc.1988.01800360080010

11. Kuja-Halkola R, Lichtenstein P, D'Onofrio BM, Larsson H. Codevelopment of ADHD and externalizing behavior from childhood to adulthood. *J Child Psychol Psychiatry.* (2015) 56:640–7. doi: 10.1111/jcpp.12340

12. Moffitt TE, Silva PA. Self-reported delinquency, neuropsychological deficit, and history of attention deficit disorder. *J Abnorm Child Psychol.* (1988) 16:553–69. doi: 10.1007/BF00914266

13. August GJ, Stewart MA, Holmes CS. A four-year follow-up of hyperactive boys with and without conduct disorder. *Br J Psychiatry.* (1983) 143:192–8. doi: 10.1192/bjp.143.2.192

14. Chan L, Simmons C, Tillem S, Conley M, Brazil IA, Baskin-Sommers A. Classifying Conduct disorder using a biopsychosocial model and machine learning method. *Biol Psychiatry Cogn Neurosci Neuroimaging.* (2023) 8:599–608. doi: 10.1016/j.bpsc.2022.02.004

15. Poldrack RA, Huckins G, Varoquaux G. Establishment of best practices for evidence for prediction: a review. *JAMA Psychiatry.* (2020) 77:534–40. doi: 10.1001/jamapsychiatry.2019.3671

16. Alexander LM, Escalera J, Ai L, Andreotti C, Febre K, Mangone A, et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci Data.* (2017) 4:170181. doi: 10.1038/sdata.2017.181

17. Bjork JM, Straub LK, Provost RG, Neale MC. The ABCD study of neurodevelopment: identifying neurocircuit targets for prevention and treatment of adolescent substance abuse. *Curr Treat Options Psychiatry.* (2017) 4:196–209. doi: 10.1007/s40501-017-0108-y

18. Karcher NR, Barch DM. The ABCD study: understanding the development of risk for mental and physical health outcomes. *Neuropsychopharmacology.* (2021) 46:131–42. doi: 10.1038/s41386-020-0736-6

19. Ripley BD. *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press (1996).

20. Russell S, Norvig P. *Artificial Intelligence: A Modern Approach*. 3rd ed. Upper Saddle River, NJ: Pearson Education (2010).

21. Kuhn M, Johnson K. *Applied Predictive Modeling*. New York, NY: Springer (2013).

22. Jernigan TL, Brown SA, Dowling GJ. The adolescent brain cognitive development study. *J Res Adolesc.* (2018) 28:154–6. doi: 10.1111/jora.12374

23. Garavan H, Bartsch H, Conway K, Decastro A, Goldstein RZ, Heeringa S, et al. Recruiting the ABCD sample: design considerations and procedures. *Dev Cogn Neurosci.* (2018) 32:16–22. doi: 10.1016/j.dcn.2018.04.004

24. Volkow ND, Koob GF, Croyle RT, Bianchi DW, Gordon JA, Koroshetz WJ, et al. The conception of the ABCD study: from substance use to a broad NIH collaboration. *Dev Cogn Neurosci.* (2018) 32:4–7. doi: 10.1016/j.dcn.2017.10.002

25. Barch DM, Albaugh MD, Avenevoli S, Chang L, Clark DB, Glantz MD, et al. Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: rationale and description. *Dev Cogn Neurosci.* (2018) 32:55–66. doi: 10.1016/j.dcn.2017.10.010

26. Lisdahl KM, Sher KJ, Conway KP, Gonzalez R, Feldstein Ewing SW, Nixon SJ, et al. Adolescent brain cognitive development (ABCD) study: overview of substance use assessment methods. *Dev Cogn Neurosci.* (2018) 32:80–96. doi: 10.1016/j.dcn.2018.02.007

27. Casey BJ, Cannonier T, Conley MI, Cohen AO, Barch DM, Heitzeg MM, et al. The Adolescent Brain Cognitive Development (ABCD) study: imaging acquisition across 21 sites. *Dev Cogn Neurosci.* (2018) 32:43–54. doi: 10.1016/j.dcn.2018.03.001

28. Hagler DJ Jr, Hatton S, Cornejo MD, Makowski C, Fair DA, Dick AS, et al. Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *Neuroimage.* (2019) 202:116091. doi: 10.1016/j.neuroimage.2019.116091

29. Mcconaughy SH. CHAPTER 10 - The Achenbach system of empirically based assessment. In: Andrews JW, Saklofske DH, Janzen HL, editors. *Educational Psychology, Handbook of Psychoeducational Assessment*. Academic Press (2001). p. 289–324. Available online at: https://www.sciencedirect.com/science/article/pii/B9780120585700500124

30. Jager S, Allhorn A, Biessmann F. A benchmark for data imputation methods. *Front Big Data.* (2021) 4:693674. doi: 10.3389/fdata.2021.693674

31. Dhillon IS, Sra S. Generalized nonnegative matrix approximations with bregman divergences. *Adv Neural Inf Process Syst.* (2006) 18. Available online at: https://www.cs.utexas.edu/~inderjit/public_papers/nnma_nips05.pdf

32. Tandon R, Sra S. *Sparse nonnegative Matrix Approximation: New Formulations and Algorithms. Max Planck Institute for Biological Cybernetics Technical Report*. Max Planck Institute for Biological Cybernetics (2010). 193 p. Available online at: https://is.mpg.de/fileadmin/user_upload/files/publications/MPIK-TR-193_[0].pdf

33. Xu J, Wang Y, Xu X, Cheng KK, Raftery D, Dong J. NMF-based approach for missing values imputation of mass spectrometry metabolomics data. *Molecules.* (2021) 26:5787. doi: 10.3390/molecules26195787

34. Goodfellow I, Bengio Y, Courville A. *Deep Learning*. Cambridge: The MIT Press (2016). xxii, 775 p.

35. Bishop CM. *Pattern Recognition and Machine Learning*. New York, NY: Springer (2006). xx, 738 p.

36. X. B, Varoquaux G. *Survey of Machine-Learning Experimental Methods at NeurIPS 2019 and ICLR.* (2020).Available online at: https://halarchives-ouvertesfr/hal-02447823 (accessed November 24, 2023).

37. de Lacy N, Ramshaw MJ, Kutz JN. Integrated evolutionary learning: an artificial intelligence approach to joint learning of features and hyperparameters for optimized, explainable machine learning. *Front Artif Intell.* (2022) 5:832530. doi: 10.3389/frai.2022.832530

38. Lundberg S, Lee S. A unified approach to interpreting model predictions. *arXiv.* (2017) 1705.07874 [cs.AI].

39. Garcia-Argibay M, Zhang-James Y, Cortese S, Lichtenstein P, Larsson H, Faraone SV. Predicting childhood and adolescent attention-deficit/hyperactivity disorder onset: a nationwide deep learning approach. *Mol Psychiatry.* (2023) 28:1232–9. doi: 10.1038/s41380-022-01918-8

40. Maniruzzaman M, Shin J, Al Mehedi Hasan M. Predicting children with ADHD using behavioral activity: a machine learning analysis. *Appl Sci.* (2022) 12:2737. doi: 10.3390/app12052737

41. Ter-Minassian L, Viani N, Wickersham A, Cross L, Stewart R, Velupillai S, et al. Assessing machine learning for fair prediction of ADHD in school pupils using a retrospective cohort study of linked education and healthcare data. *BMJ Open.* (2022) 12:e058058. doi: 10.1136/bmjopen-2021-058058

42. Menon SS, Krishnamurthy K. Multimodal ensemble deep learning to predict disruptive behavior disorders in children. *Front Neuroinform.* (2021) 15:742807. doi: 10.3389/fninf.2021.742807

43. Owens JA, Spirito A, McGuinn M, Nobile C. Sleep habits and sleep disturbance in elementary school-aged children. *J Dev Behav Pediatr.* (2000) 21:27–36. doi: 10.1097/00004703-200002000-00005

44. Alfano CA. (Re)Conceptualizing sleep among children with anxiety disorders: where to next? *Clin Child Fam Psychol Rev.* (2018) 21:482–99. doi: 10.1007/s10567-018-0267-4

45. Dimakos J, Gauthier-Gagne G, Lin L, Scholes S, Gruber R. The associations between sleep and externalizing and internalizing problems in children and adolescents with attention-deficit/hyperactivity disorder: empirical findings, clinical implications, and future research directions. *Child Adolesc Psychiatr Clin N Am.* (2021) 30:175–93. doi: 10.1016/j.chc.2020.08.001

46. Hvolby A, Jorgensen J, Bilenberg N. Actigraphic and parental reports of sleep difficulties in children with attention-deficit/hyperactivity disorder. *Arch Pediatr Adolesc Med.* (2008) 162:323–9. doi: 10.1001/archpedi.162.4.323

47. Hosokawa R, Tomozawa R, Fujimoto M, Anzai S, Sato M, Tazoe H, et al. Association between sleep habits and behavioral problems in early adolescence: a descriptive study. *BMC Psychol.* (2022) 10:254. doi: 10.1186/s40359-022-00958-7

48. Aronen ET, Lampenius T, Fontell T, Simola P. Sleep in children with disruptive behavioral disorders. *Behav Sleep Med.* (2014) 12:373–88. doi: 10.1080/15402002.2013.821653

49. Lunsford-Avery JR, Bidopia T, Jackson L, Sloan JS. Behavioral treatment of insomnia and sleep disturbances in school-aged children and adolescents. *Child Adolesc Psychiatr Clin N Am.* (2021) 30:101–16. doi: 10.1016/j.chc.2020.08.006

50. Ma ZR, Shi LJ, Deng MH. Efficacy of cognitive behavioral therapy in children and adolescents with insomnia: a systematic review and meta-analysis. *Braz J Med Biol Res.* (2018) 51:e7070. doi: 10.1590/1414-431x20187070

51. Mathews A, Gibbons N, Harrison E, Stallard P. A feasibility study to explore the use of digital treatment of sleep as a first-step intervention to improve adolescent mental health. *Behav Sleep Med.* (2023) 21:172–84. doi: 10.1080/15402002.2022.2063866

52. Jones DJ, Forehand R, Cuellar J, Parent J, Honeycutt A, Khavjou O, et al. Technology-enhanced program for child disruptive behavior disorders: development and pilot randomized control trial. *J Clin Child Adolesc Psychol.* (2014) 43:88–101. doi: 10.1080/15374416.2013.822308

53. He F, Qi Y, Zhou Y, Cao A, Yue X, Fang S, et al. Meta-analysis of the efficacy of digital therapies in children with attention-deficit hyperactivity disorder. *Front Psychiatry.* (2023) 14:1054831. doi: 10.3389/fpsyt.2023.1054831

54. Radanliev P, De Roure D. Alternative mental health therapies in prolonged lockdowns: narratives from Covid-19. *Health Technol.* (2021) 11:1101–7. doi: 10.1007/s12553-021-00581-3

55. Marek S, Tervo-Clemmens B, Calabro FJ, Montez DF, Kay BP, Hatoum AS, et al. Reproducible brain-wide association studies require thousands of individuals. *Nature.* (2022) 603:654–60. doi: 10.1038/s41586-022-04492-9

56. Cortese S, Hill CM. Sleep and ADHD: adding pieces to the puzzle. *Sleep Med.* (2023) 103:144–5. doi: 10.1016/j.sleep.2023.02.002

57. Lunsford-Avery JR, Krystal AD, Kollins SH. Sleep disturbances in adolescents with ADHD: a systematic review and framework for future research. *Clin Psychol Rev.* (2016) 50:159–74. doi: 10.1016/j.cpr.2016.10.004

58. Zhang-James Y, Razavi AS, Hoogman M, Franke B, Faraone SV. Machine learning and MRI-based diagnostic models for ADHD: are we there yet? *J Atten Disord.* (2023) 27:335–53. doi: 10.1177/10870547221146256

59. Rubia K, Alegria AA, Brinson H. Brain abnormalities in attention-deficit hyperactivity disorder: a review. *Rev Neurol.* (2014) 58(Suppl. 1):S3–16.

60. McCarthy H, Skokauskas N, Frodl T. Attention network hypoconnectivity in adults diagnosed as having attention-deficit/hyperactivity disorder in childhood. *JAMA Psychiatry.* (2014) 71:1299–300. doi: 10.1001/jamapsychiatry.2014.726

61. McCarthy H, Skokauskas N, Mulligan A, Donohoe G, Mullins D, Kelly J, et al. Attention network hypoconnectivity with default and affective network hyperconnectivity in adults diagnosed with attention-deficit/hyperactivity disorder in childhood. *JAMA Psychiatry.* (2013) 70:1329–37. doi: 10.1001/jamapsychiatry.2013.2174

62. Hoogman M, Bralten J, Hibar DP, Mennes M, Zwiers MP, Schweren LSJ, et al. Subcortical brain volume differences in participants with attention deficit hyperactivity disorder in children and adults: a cross-sectional mega-analysis. *Lancet Psychiatry.* (2017) 4:310–9. doi: 10.1016/S2215-0366(17)30049-4

63. Yerys BE, Tunc B, Satterthwaite TD, Antezana L, Mosner MG, Bertollo JR, et al. Functional connectivity of frontoparietal and salience/ventral attention networks have independent associations with co-occurring attention-deficit/hyperactivity disorder symptoms in children with autism. *Biol Psychiatry Cogn Neurosci Neuroimaging.* (2019) 4:343–51. doi: 10.1016/j.bpsc.2018.12.012

64. Bailey T, Joyce A. The role of the thalamus in ADHD symptomatology and treatment. *Appl Neuropsychol Child.* (2015) 4:89–96. doi: 10.1080/21622965.2015.1005475

65. Fu C, Chen S, Qian A, Zhou R, Zhou J, Li J, et al. Larger thalamus correlated with inattentive severity in the inattentive subtype of ADHD without comorbidity. *Psychiatry Res.* (2021) 304:114079. doi: 10.1016/j.psychres.2021.114079

66. Mills KL, Bathula D, Dias TG, Iyer SP, Fenesy MC, Musser ED, et al. Altered cortico-striatal-thalamic connectivity in relation to spatial working memory capacity in children with ADHD. *Front Psychiatry.* (2012) 3:2. doi: 10.3389/fpsyt.2012.00002

67. Dedousis-Wallace A, Drysdale SAO, McAloon J, Murrihy RC, Greene RW, Ollendick TH. Predictors and moderators two treatments of oppositional defiant disorder in children. *J Clin Child Adolesc Psychol.* (2022) 1–16. doi: 10.1080/15374416.2022.2127102 (Epub ahead of print).

68. Lavigne JV, Gouze KR, Hopkins J, Bryant FB. A multidomain cascade model of early childhood risk factors associated with oppositional defiant disorder symptoms in a community sample of 6-year-olds. *Dev Psychopathol.* (2016) 28(4pt2):1547–62. doi: 10.1017/S0954579415001194

69. Lavigne JV, Gouze KR, Hopkins J, Bryant FB, LeBailly SA. A multi-domain model of risk factors for ODD symptoms in a community sample of 4-year-olds. *J Abnorm Child Psychol.* (2012) 40:741–57. doi: 10.1007/s10802-011-9603-6

70. Lavigne JV, Lebailly SA, Hopkins J, Gouze KR, Binns HJ. The prevalence of ADHD, ODD, depression, and anxiety in a community sample of 4-year-olds. *J Clin Child Adolesc Psychol.* (2009) 38:315–28. doi: 10.1080/15374410902851382

71. Copeland W, Shanahan L, Costello EJ, Angold A. Configurations of common childhood psychosocial risk factors. *J Child Psychol Psychiatry.* (2009) 50:451–9. doi: 10.1111/j.1469-7610.2008.02005.x

72. Seeley WW. The salience network: a neural system for perceiving and responding to homeostatic demands. *J Neurosci.* (2019) 39:9878–82. doi: 10.1523/JNEUROSCI.1138-17.2019

73. Seeley WW, Menon V, Schatzberg AF, Keller J, Glover GH, Kenna H, et al. Dissociable intrinsic connectivity networks for salience processing and executive control. *J Neurosci.* (2007) 27:2349–56. doi: 10.1523/JNEUROSCI.5587-06.2007

74. Dosenbach NU, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, et al. A core system for the implementation of task sets. *Neuron.* (2006) 50:799–812. doi: 10.1016/j.neuron.2006.04.031

75. Dosenbach NU, Fair DA, Miezin FM, Cohen AL, Wenger KK, Dosenbach RA, et al. Distinct brain networks for adaptive and stable task control in humans. *Proc Natl Acad Sci USA.* (2007) 104:11073–8. doi: 10.1073/pnas.0704320104

76. Andrews-Hanna JR, Reidler JS, Sepulcre J, Poulin R, Buckner RL. Functional-anatomic fractionation of the brain's default network. *Neuron.* (2010) 65:550–62. doi: 10.1016/j.neuron.2010.02.005

77. Andrews-Hanna JR, Smallwood J, Spreng RN. The default network and self-generated thought: component processes, dynamic control, and clinical relevance. *Ann N Y Acad Sci.* (2014) 1316:29–52. doi: 10.1111/nyas.12360

78. Anticevic A, Cole MW, Murray JD, Corlett PR, Wang XJ, Krystal JH. The role of default network deactivation in cognition and disease. *Trends Cogn Sci.* (2012) 16:584–92. doi: 10.1016/j.tics.2012.10.008

79. Balleine BW, Delgado MR, Hikosaka O. The role of the dorsal striatum in reward and decision-making. *J Neurosci.* (2007) 27:8161–5. doi: 10.1523/JNEUROSCI.1554-07.2007

80. Baxter MG, Croxson PL. Facing the role of the amygdala in emotional information processing. *Proc Natl Acad Sci USA.* (2012) 109:21180–1. doi: 10.1073/pnas.1219167110

81. Baxter MG, Murray EA. The amygdala and reward. *Nat Rev Neurosci.* (2002) 3:563–73. doi: 10.1038/nrn875

82. Murray J, Farrington DP. Risk factors for conduct disorder and delinquency: key findings from longitudinal studies. *Can J Psychiatry.* (2010) 55:633–42. doi: 10.1177/070674371005500103

83. Waller R, Hawes SW, Byrd AL, Dick AS, Sutherland MT, Riedel MC, et al. Disruptive behavior problems, callous-unemotional traits, and regional gray matter volume in the adolescent brain and cognitive development study. *Biol Psychiatry Cogn Neurosci Neuroimaging.* (2020) 5:481–9. doi: 10.1016/j.bpsc.2020.01.002

84. Noordermeer SD, Luman M, Oosterlaan J. A systematic review and meta-analysis of neuroimaging in oppositional defiant disorder (ODD) and conduct disorder (CD) taking attention-deficit hyperactivity disorder (ADHD) into account. *Neuropsychol Rev.* (2016) 26:44–72. doi: 10.1007/s11065-015-9315-8

85. Baker RH, Clanton RL, Rogers JC, De Brito SA. Neuroimaging findings in disruptive behavior disorders. *CNS Spectr.* (2015) 20:369–81. doi: 10.1017/S1092852914000789

86. Rogers JC, De Brito SA. Cortical and subcortical gray matter volume in youths with conduct problems: a meta-analysis. *JAMA Psychiatry.* (2016) 73:64–72. doi: 10.1001/jamapsychiatry.2015.2423

87. Cope LM, Ermer E, Nyalakanti PK, Calhoun VD, Kiehl KA. Paralimbic gray matter reductions in incarcerated adolescent females with psychopathic traits. *J Abnorm Child Psychol.* (2014) 42:659–68. doi: 10.1007/s10802-013-9810-4

88. Ermer E, Cope LM, Nyalakanti PK, Calhoun VD, Kiehl KA. Aberrant paralimbic gray matter in criminal psychopathy. *J Abnorm Psychol.* (2012) 121:649–58. doi: 10.1037/a0026371

89. Ermer E, Cope LM, Nyalakanti PK, Calhoun VD, Kiehl KA. Aberrant paralimbic gray matter in incarcerated male adolescents with psychopathic traits. *J Am Acad Child Adolesc Psychiatry.* (2013) 52:94–103 e3. doi: 10.1016/j.jaac.2012.10.013

90. Kim WB, Cho JH. Encoding of contextual fear memory in hippocampal-amygdala circuit. *Nat Commun.* (2020) 11:1382. doi: 10.1038/s41467-020-15121-2