Development and Psychopathology (2026), 1–11
doi:10.1017/S0954579426101977

CAMBRIDGE UNIVERSITY PRESS

Regular Article

A machine learning approach to predicting high-risk irritability trajectories across the transition to adolescence

Leslie S. Jordan[1,2], Annika J. Quam[1], Alyssa J. Parker[1], Jillian Lee Wiggins[3,4] and Lea R. Dougherty[1,2]

[1]Department of Psychology, University of Maryland, College Park, MD, USA; [2]Neuroscience and Cognitive Science Program, University of Maryland, College Park, MD, USA; [3]Joint Doctoral Program in Clinical Psychology, San Diego State University/University of California, San Diego, CA, USA and [4]Department of Psychology, San Diego State University, San Diego, CA, USA

Abstract

Youth irritability, characterized as proneness to anger/frustration, is associated with negative mental health outcomes. This study identified irritability trajectories across the transition to early adolescence and applied machine learning to identify preadolescent predictors of irritability trajectories. Data are from the Adolescent Brain Cognitive Development (ABCD) study, a large longitudinal study that tracks youth development from preadolescence ($n = 11{,}862$, ages 8.92–11.08 years, $M = 9.91$, $SD = 0.62$). The baseline sample was 47.83% female ($n = 5{,}680$), 63.35% White, 15.74% Black, 12.07% Biracial or Multiracial, 2.32% Asian, and 20.30% Hispanic. Using parent-report youth irritability scores across four years, latent class growth analysis (LCGA) identified four distinct irritability trajectories: low stable ($n = 8{,}692$, 73.28%), high decreasing ($n = 1{,}083$, 9.13%), low increasing ($n = 1{,}449$, 12.22%), and high chronic ($n = 638$, 5.38%). Machine learning approaches generated predictive models for irritability trajectories, including baseline youth psychopathology, physical health, neurocognitive abilities, psychosocial environment, demographics, parent psychopathology, and brain structure metrics. The models demonstrated strong performance in detecting probability of being in the high chronic and low stable trajectories but poorer performance in classifying the irritability trajectories characterized by change. The most prominent predictors included youth externalizing and internalizing problems, parent psychopathology, and sleep problems. Findings provide insight into preadolescent predictors of irritability across a critical developmental period.

Keywords: adolescence; Adolescent Brain Cognitive Development Study; developmental trajectories; irritability; machine learning; preadolescence

(Received 17 December 2025 UTC; revised 12 August 2026 UTC; accepted 12 August 2026 UTC)

Adolescence is a crucial developmental turning point, characterized by increases in psychological disorders, including generalized anxiety, social anxiety, eating, trauma-related, and depressive disorders (Solmi et al., 2022). This increased severity and onset of psychopathology coincides with dramatic psychological and neurological maturation. Nevertheless, the same developmental sensitivity that amplifies the harmful effects of risk factors also creates a window of opportunity to foster resilience and positive wellbeing that persist into adulthood. One major challenge is identifying youth at high-risk before they reach adolescence to facilitate the timeliness of preventive interventions. Prior research indicates that one of the earliest antecedents of psychopathology is elevated irritability characterized by frequent temper outbursts and chronic angry mood (Sorcher et al., 2022; Wakschlag et al., 2024). Indeed, irritability is one of the most common reasons that youth are referred for psychiatric evaluation (e.g., Evans et al., 2023). Irritability is most common in early childhood and typically declines; however, adolescence is associated with an increase in the duration of irritable mood, as well as in mental health disorders characterized by irritable mood (Copeland et al., 2015). Furthermore, the heterogeneity in developmental trajectories and multifinality of this transdiagnostic symptom exacerbate the challenges of identifying youth at high-risk (Riglin et al., 2019). There is a critical need to develop predictive models that can determine which youth are at greatest risk for high and chronic levels of irritability across critical, transitional developmental periods.

Prior work has observed four or five youth irritability trajectories in longitudinal community samples and has reported an association between trajectories characterized by persistently elevated levels of irritability and poorer mental health outcomes (Orri et al., 2019; Riglin et al., 2019; Wiggins et al., 2014; Yu et al., 2023b). Across these studies, the majority of children exhibit normative, persistently low irritability levels throughout development, while a smaller percentage of youth experience higher levels of irritability either persistently throughout or at some point during the assessed developmental period. For example, in a sample of 1,393 children from ages 6 to 12 using teacher reports of irritability, four trajectories were identified and included low

Corresponding author: Lea R. Dougherty; Email: ldougher@umd.edu

Cite this article: Jordan, L. S., Quam, A. J., Parker, A. J., Wiggins, J. L., & Dougherty, L. R. (2026). A machine learning approach to predicting high-risk irritability trajectories across the transition to adolescence. *Development and Psychopathology*, 1–11, https://doi.org/10.1017/S0954579426101977

© The Author(s), 2026. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0954579426101977 Published online by Cambridge University Press

(74.7%), persistent (5.0%), rising (13.0%), and declining (7.4%) (Orri et al., 2019). Furthermore, Wiggins et al. (2014) identified five trajectories based on parent reports of child irritability from age 3 to 9: low-decreasing (60.8%), medium-decreasing (21.0%), high-steady (10.6%), initially very high-decreasing (5.4%), and high-increasing (2.2%). Interestingly, Yu et al. (2023b) identified an important inflection point in preadolescence (around age 9), after which irritability levels shifted to either high or low levels. Thus, it is critical to further investigate these changes in irritability from preadolescence into early adolescence.

Although previous research has built a foundation for our understanding of irritability trajectories in youth, critical gaps remain. First, adolescence is characterized by increased vulnerability to mental health disorders; thus, there is a need for research that closely examines the unfolding of irritability across the transition from preadolescence to adolescence with sufficiently frequent repeated assessments. Although a few studies have analyzed irritability trajectories that crossed over into the adolescent years, the time intervals between clinical assessments have often omitted ages that may be pivotal time points for the developmental course of irritability (Riglin et al., 2019; Yu et al., 2023b). Second, there is a need to predict the course of irritability to facilitate early intervention that can alter high-risk irritability trajectories. Few studies have investigated predictors of the course of irritability (Sorcher et al., 2023; Wiggins et al., 2014; Yu et al., 2023b), and no studies to our knowledge have applied machine learning techniques to develop predictive models for irritability trajectories. Machine learning provides an exploratory, computational tool to identify predictors across numerous domains. In turn, this approach can contribute to subsequent confirmatory research with a priori hypotheses and ultimately inform early detection of high-risk individuals.

Research investigating predictors for irritability is sparse, and the limited research has revealed predictors for youth irritability across a broad range of domains. Studies of irritability predictors have reported youth internalizing and externalizing problems (e.g. Leadbeater & Homel, 2015), sleep problems (e.g., Rubens et al., 2017), harsh parenting behaviors (Wiggins et al., 2014), lower parental education (Kessel et al., 2021; Wiggins et al., 2014), and parent personality and psychopathology (Dougherty et al., 2013; Sorcher et al., 2023; Wiggins et al., 2014) as predictors of later youth irritability; however, the majority of these studies examined predictors of later irritability but not the longitudinal course of irritability. Prior research has also interrogated both concurrent and longitudinal structural neural correlates of youth irritability with scrutiny of brain regions involved in reward processing, emotion regulation, and inhibitory control (e.g., Dennis et al., 2019). Given the paucity of research into predictors of the course of irritability, it is crucial to apply data-driven methods to examine predictors that have been previously identified in the youth irritability research while also investigating predictors from the broader youth psychopathology literature.

Machine learning, a diverse set of data-driven algorithms that discover patterns in large datasets, has the potential to facilitate our identification of youth at higher risk for long-term maladjustment by generating predictive models (Galatzer-Levy et al., 2018). Large community samples such as the Adolescent Brain Cognitive Development study (ABCD; Barch et al., 2018), a large national sample of preadolescents assessed for 10 years, offer the opportunity to apply machine learning methods to investigate irritability and mental health during critical developmental years (Xiang et al., 2022).

Our study aims to leverage the ABCD dataset to identify irritability trajectories and to develop predictive models for these irritability trajectories using a wide range of preadolescent predictors. We examine irritability trajectories across the transition from preadolescence (ages 9–10) to early adolescence (ages 13–14) using latent class growth analysis. We apply several different supervised machine learning algorithms to develop predictive models for irritability trajectories. Results for the predictive models describe model performance and identify important preadolescent predictors of irritability trajectories. Our full set of candidate predictors reflects a comprehensive biopsychosocial model of youth irritability and mental health across development, spanning multiple domains that have the potential to influence the developmental course of irritability, including youth psychopathology, youth physical health, neurocognitive abilities, psychosocial environment, parent psychopathology, and youth neurobiology. Based on prior research in community samples, we expect to observe that the majority of youth will exhibit consistently low levels of irritability, while a subset of youth will show persistently high or changing levels of irritability. In accordance with the data-driven nature of machine learning, we take an exploratory approach to the evaluation of model performance and the identification of top predictors without specific a priori hypotheses.

Method

Participants

Data are from the ABCD Study, a 10-year longitudinal study that tracks youth development in a large, national sample starting from preadolescence (Volkow et al., 2018). The baseline sample consisted of 11,862 9–10-year-old youth residing in the United States. The current study analyzed data from the ABCD Data Release 5.1 (10.15154/z563-zd24), which included the full, preadolescent sample at baseline, 1-year, 2-year, and 3-year follow-ups, and partial sample at the 4-year follow-up ($n = 4,678$). The outcome of interest is developmental trajectories of parent-reported youth irritability symptoms based on the Child Behavior Checklist (CBCL) irritability scale scores.

Youth irritability

The CBCL is a parent-report mental health measure that assesses the frequency and severity of a child's emotional and behavioral problems (Achenbach, 2009). Prior research has quantified an irritability dimension by summing three CBCL items that evaluate the extent to which their child can be described as "stubborn, sullen or irritable," exhibiting "sudden changes in mood or feelings," and having "temper tantrums or hot temper" (Roberson-Nay et al., 2015; Stringaris et al., 2012). Each item is rated as absent (0), occurs sometimes (1), or occurs often (2) over the past six months, resulting in an irritability index ranging from 0 to 6 with higher scores indicating higher levels of irritability. Prior exploratory and confirmatory factor analyses support the psychometric properties of this scale to measure the latent construct of irritability (Stringaris et al., 2012; Wiggins et al., 2014).

Baseline predictors

We used both the small literature on predictors and correlates of youth irritability and the broader literature on predictors of internalizing and externalizing symptomatology, particularly given the transdiagnostic nature of irritability, to inform the selection of the candidate predictors (Hill et al., 2025; Kessel et al., 2021). Specifically, we used baseline variables that spanned the wide range

https://doi.org/10.1017/S0954579426101977 Published online by Cambridge University Press

of ABCD emotional, behavioral, cognitive, and contextual domains, including youth's mental health, physical health, neurocognitive abilities, demographics, culture and environment, parent mental health, and structural neurobiological measures (Barch et al., 2018; Casey et al., 2018; Zucker et al., 2018). The breadth of candidate predictors along with data-driven methods allowed us to expand on the limited research on predictors of youth irritability with the aim of guiding theoretical models of irritability. The current study used baseline variables from Release 4.0 which contained the full sample at baseline. For the full list of selected predictors refer to Tables 1 and 2.

Data analysis plan

Developmental trajectories

To identify differential irritability trajectories from ages 9 to 14, we implemented a series of latent class growth analyses ranging from two to five classes, with growth factors fixed to zero in Mplus version 8.9 (Muthén & Muthén 2017). In addition to testing models based on linear trajectories, we repeated the series of models with two to five classes while estimating quadratic components for each class. To account for the hierarchical structure within the ABCD data, the models were stratified by site and clustered by family. The models classified participants by growth trajectory group based on their caregiver-rated CBCL irritability scores across the baseline, 1-year, 2-year, 3-year, and 4-year follow-up assessments. The models were estimated using full information maximum likelihood with all available data. Each model was evaluated based on information criteria model-fit indices, confidence in classification, likelihood ratio tests, and interpretability. Lower values of information criteria model-fit indices including the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC), and the Sample Size Adjusted BIC (SSABIC) indicate better fit (Nyund-Gibson & Choi, 2018). Entropy is a measure of confidence in classification with values greater than 0.8 indicating good classification (Nyund-Gibson & Choi, 2018). Likelihood-based tests, such as the Vuong-Lo-Mendell-Rubin adjusted likelihood ratio test (VLMR-LRT) indicates whether the current model provides a statistically significant improvement over the previous model with $k-1$ classes. A significant p-value on a likelihood-based test indicates that the current model offers statistically significant improvement over the previous model, whereas a non-significant p-value suggests maintaining the model with $k-1$.

Machine learning procedures

Each machine learning problem is unique and, therefore, development of accurate predictive models requires assessment of different algorithms. Applying multiple machine learning methods provides the opportunity to compare performance metrics and to examine the most important predictors for different approaches. Accordingly, we trained four different multiclass supervised machine learning models to classify youth irritability trajectories. A standardized machine learning pipeline was designed to maintain consistency across models in feature selection, data preprocessing, and adjusting for class imbalance. The full set of candidate predictors was reduced using Boruta, a data-driven procedure based on the Random Forest algorithm that identifies relevant predictors, thereby minimizing the risk of overfitting in the final models (Jiang et al., 2020; Kursa & Rudnicki, 2010; Kursa et al., 2010). Data preprocessing included normalization of numeric predictors, dummy coding of categorical predictors, and Bagged Tree Imputation (BTI) for missing data imputation (Joranov et al., 2018). Synthetic Minority Oversampling Technique (SMOTE) was used to address class imbalance by generating synthetic cases for minority classes using the nearest neighbors algorithm (KNN) (Chawla et al., 2002). Model training employed 10-fold cross-validation for hyperparameter tuning applied to each of the four machine learning models (Elastic Net, KNN, Random Forest, and XGBoost). The four machine learning models that we selected are classic supervised learning models which use different techniques to balance the tradeoff between prediction accuracy and overfitting. Elastic net is a regularized regression technique that combines LASSO and ridge regression by optimizing two hyperparameters (Whelan et al., 2014; Zou & Hastie, 2005). KNN assigns classification using k nearest neighbors (Cunningham & Delany, 2021). Random forest is an ensemble method that generates multiple models which are averaged together to produce predictions (Jiang et al., 2020). XGBoost is another ensemble method, yet this method builds multiple models sequentially, thus improving on each previous model (Chen & Guestrin, 2016).

All participants with structural MRI data were included in the machine learning sample ($n = 11{,}034$). The data were divided into training ($n = 8{,}826$, 80%) and test sets ($n = 2{,}208$, 20%), stratified by irritability trajectory.

For all observations in the test dataset, each model produced an estimated probability of class assignment to each of the four trajectories and assigned the final predicted class based on the highest estimated probability across all four trajectories. The current study provided a comprehensive overview of model performance, including measures of overall performance and performance by trajectory. Performance metrics included measures of the overall ratio of accurately classified cases to inaccurately classified cases (accuracy), the ability of the models to accurately detect true cases of a trajectory (true positive rate), the ability of the model to accurately detect only true cases of a trajectory (precision), and the harmonic mean of true positive rate and precision which quantifies the balance between true positive rate and precision (F1 Score) (Sokolova & Lapalme, 2009). The possible values range from 0 to 1, with 1 representing perfect performance and 0 representing worst possible performance. Overall performance measures were calculated using micro- and macro-averaging techniques. Micro-averaging gives each observation in the test dataset an equal influence on the performance metric, while macro-averaging gives each class (or trajectory) an equal influence on the performance metric. Performance metrics by trajectory were calculated by modeling the results as one-vs-rest binary classification with the trajectory of interest being considered the positive class and all other classes being considered the negative class, including data visualizations for ROC curves along with area under the curve for each trajectory (Mathew et al., 2023). This method of plotting one-vs-rest ROC curves and calculating AUC by trajectory provided insight into the model's ability to rank order estimated probability of class assignment to the trajectory of interest as opposed to any of the other three trajectories. Lastly, the rankings for the top ten variables are reported for the Elastic Net, Random Forest, and XGBoost algorithms. Variable importance results provided insight into the relative influence of each predictor on the final model's predictions (Bartoszuk & Gagolewski, 2020). The top ten important variables were obtained using the "vip" package in R (Greenwell & Boehmke, 2020). Machine learning analyses were conducted in R version 4.2.2.

https://doi.org/10.1017/S0954579426101977 Published online by Cambridge University Press

Table 1. Candidate behavioral predictors and feature selection results

Domain	Respondent	Measure	Scales	Boruta feature selection result
Mental health	Parent	Child Behavior Checklist (CBCL)	Depression CBCL DSM5 scale	Selected
			Anxiety disorder CBCL DSM5 scale	Selected
			Somatic problems CBCL DSM5 scale	Selected
			ADHD CBCL DSM5 scale	Selected
			Conduct CBCL DSM5 scale	Selected
	Youth	BISBAS	Behavioral inhibition scale	Not selected
			Reward responsiveness scale	Not selected
			Drive scale	Not selected
			Fun seeking scale	Not Selected
	Youth	Impulsivity (UPPS-P) scale for children short form	Lack of premeditation scale score	Selected
			Negative urgency scale score	Selected
			Sensation seeking scale score	Not selected
			Lack of perseverance scale score	Selected
			Positive urgency scale score	Selected
	Parent	KSADS	Depression diagnosis lifetime	Selected
			Anxiety diagnosis lifetime	Selected
			Oppositional defiant disorder lifetime	Selected
			Suicidal ideation lifetime	Not selected
			Suicide attempt lifetime	Not selected
			Nonsuicidal self injury lifetime	Not selected
	Youth	KSADS	Suicidal ideation lifetime	Not selected
			Suicide attempt lifetime	Not Selected
			Nonsuicidal self injury lifetime	Not selected
Family history/ parent mental health	Parent	Adult Self Report (ASR)	Depressive problems ASR DSM-5 scale	Selected
			Anxiety problems ASR DSM-5 scale	Selected
			Somatic problems ASR DSM-5 scale	Selected
			Avoidant personality problems ASR DSM-5 scale	Selected
			ADHD problems ASR DSM-5 scale	Selected
			Antisocial personality problems ASR DSM-5 scale	Selected
			Inattention ASR DSM-5 scale	Selected
			Hyperactivity–Impulsivity ASR DSM-5 scale	Selected
			ASR irritability index	Selected
	Parent	Family history summary scores	Either father or mother with alcohol problem	Not selected
			Either father or mother with drug problem	Not selected
Physical health	Parent	Youth sleep disturbance	Disorders of arousal or nightmares	Selected
			Disorders of initiating and maintaining sleep	Selected
			Disorders of excessive somnolence	Selected
			Sleep breathing disorders	Selected
			Sleep hyperhidrosis	Selected
			Sleep–wake transition disorders	Selected
	Youth	Youth risk behavior survey exercise physical activity	During the past 7 days, on how many days were you physically active for a total of at least 60 minutes per day?	Not selected
			On how many of the past 7 days did you do exercises to strengthen or tone your muscles, such as push-ups, sit-ups, or weightlifting?	Not selected
	Parent/ Youth	Average parent youth pubertal development scale	Parent puberty score (youth score if parent is missing)	Not selected
Neurocognition	Youth	NIH toolbox	Cognition fluid composite	Selected
			Crystallized composite	Selected
			Flanker inhibitory control and attention test	Not selected
			Dimensional change card sort	Selected
Demographic	Parent	Parent demographics survey	Youth sex	Not selected
			Youth age	Not selected
			Parent education	Selected
			Youth race	Selected
			Youth ethnicity	Not selected
	Parent	Residential history derived scores	Modified area deprivation index	Selected
	Youth	Youth diagnostic interview for DSM-5 background items	Youth sexuality	Not selected
			Youth transgender	Not selected

(*Continued*)

https://doi.org/10.1017/S0954579426101977 Published online by Cambridge University Press

Table 1. (*Continued*)

Domain	Respondent	Measure	Scales	Boruta feature selection result
Culture & environment	Parent	ABCD parent family environment scale	Family conflict subscale	Selected
	Youth	ABCD youth family environment scale	Family conflict subscale	Selected
	Parent	School risk and protective factors	Prosocial opportunity subscale Prosocial reward subscale General protective subscale School protective factors overall scale	Selected Selected Selected Selected
	Youth	Children's report of parental behavior inventory	Parental acceptance and warmth (Caregiver 1)	Selected
	Parent/ Youth	Adverse childhood experiences	Adverse Childhood Experiences (ACE) score	Selected

Table 2. Candidate structural MRI predictors and feature selection results

Brain region of interest	Boruta feature selection result
Bilateral Nucleus Accumbens	Selected
Bilateral Caudate	Selected
Bilateral Putamen	Selected
Bilateral Amygdala	Selected
Bilateral Superior Frontal Gyrus	Not selected
Bilateral Insular Gyrus and Central Sulcus of the Insula	Selected
Bilateral Short Insular Gyri	Not selected
Bilateral Anterior Segment of the Circular Sulcus of the Insula	Selected
Bilateral Inferior Segment of the Circular Sulcus of the Insula	Selected
Bilateral Superior Segment of the Circular Sulcus of the Insula	Not selected
Bilateral Opercular Part of the Inferior Frontal Gyrus	Not selected
Bilateral Orbital Part of the Interior Frontal Gyrus	Selected
Bilateral Triangular Part of the Inferior Frontal Gyrus	Not selected
Bilateral Middle Frontal Gyrus	Selected
Bilateral Anterior Part of the Cingulate Gyrus and Sulcus	Not selected
Bilateral Posterior-Dorsal Part of the Cingulate Gyrus	Selected

Results

Irritability trajectories

Prior to latent growth curve analysis, we first tested the longitudinal measurement invariance of the CBCL irritability scale over time. Specifically, we conducted a series of confirmatory factor analyses to test configural, metric, scalar, and strict invariance, allowing correlated residuals between identical items at adjacent timepoints (for results, see Supplementary Table 1). Global model fit was evaluated by the following criteria: CFI $\geq$ 0.90, TLI $\geq$ 0.90, RMSEA $\leq$ 0.08, SRMR $\leq$ 0.08 (Sathyanarayana & Mohanasundaram, 2024), although a CFI > 0.95 is preferred (Hu & Bentler, 1999). As shown in Supplementary Table 1, the configural, metric, and scalar models are all above the less conservative threshold of CFI $\geq$ 0.90. Longitudinal measurement invariance was assessed based on Δ CFI $\leq$ 0.010, Δ RMSEA $\leq$0.015, Δ SRMR $\leq$ 0.030 for metric invariance, and Δ SRMR $\leq$ 0.010 for scalar invariance (Chen, 2007; Cheung & Rensvold, 2002). Results supported longitudinal measurement invariance at the configural, metric, and scalar levels. Scalar invariance is considered sufficient support for further analysis (Tse et al., 2024). Thus, we proceeded with latent class growth analyses to identify youth irritability trajectories.

Model-fit indices for the latent class growth analyses, including AIC, BIC, and SSABIC, indicated better fit as the number of classes increased. All models resulted in entropy > 0.8 indicating acceptable accuracy. Models with two, three, and four classes resulted in significant Lo-Mendell-Rubin Adjusted LRT Tests suggesting that the previous model should be rejected in favor of the model with more classes (Nyland-Gibson & Choi, 2018). However, the five-class model had a non-significant Lo-Mendell-Rubin Adjusted LRT Test indicating that the four-class model cannot be rejected in favor of the five-class model (See Table 3). We

https://doi.org/10.1017/S0954579426101977 Published online by Cambridge University Press

Table 3. Model-fit statistics for quadratic models of irritability trajectories

					Lo-Mendell-Rubin adjusted LRT test		
Number of classes	AIC	BIC	SSABIC	Entropy	Value	*p*	*n* (%) of smallest class
2	137,574.62	137,641.05	137,612.44	0.91	19,834.60	<.01	2,564 (21.62)
3	135,352.24	135,440.81	135,402.68	0.90	4,999.43	<.01	655 (5.52)
4	133,457.94	133,576.03	133,525.19	0.89	1,852.92	<.01	638 (5.38)
5	132,274.06	132,421.68	132,358.12	0.88	1,160.94	.15	291 (2.45)

Note: Model-fit indices include Akaike Information Criterion (AIC), Bayesian Information Criteria (BIC), and Sample Size Adjusted Bayesian Information Criterion (SSABIC).

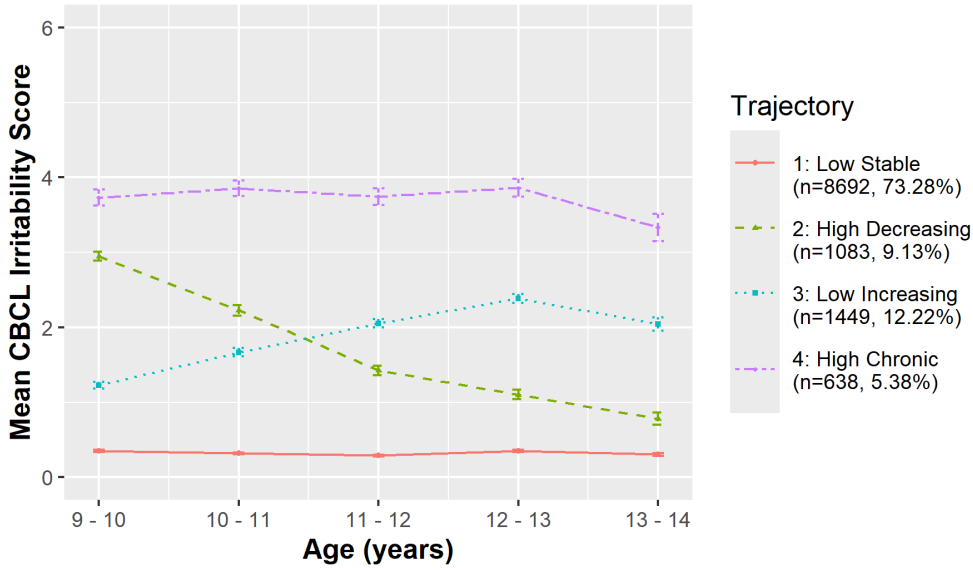


Figure 1. Irritability trajectories across the transition to adolescence in ABCD sample. *Note*: This figure shows mean CBCL irritability scores with standard error (95% confidence interval) for each irritability trajectory class across the transition from preadolescence to early adolescence.

observed the same pattern of model-fit indices, entropy, and Lo-Mendell-Rubin Adjusted LRT Test as the number of classes increased for both the linear and quadratic models. The quadratic classes produced smaller AIC, BIC, and SSABIC for each model in comparison to the linear models with the same number of classes, indicating better fit for the trajectories with quadratic components (See Supplementary Table 2). Moreover, quadratic components were significant for each of the four trajectories (See Supplementary Table 3). Thus, we selected the four-class latent growth curve model with quadratic components. The four-class model differentiated growth trajectories in irritability into classes that were characterized as (1) low stable ($n = 8{,}692$, 73.28%), (2) high decreasing ($n = 1{,}083$, 9.13%), (3) low increasing ($n = 1{,}449$, 12.22%), and (4) high chronic ($n = 638$, 5.38%). Model results are detailed in Supplementary Table 3 and illustrated in Figure 1.

Demographic information for each irritability trajectory group is presented in Supplementary Table 4. We conducted Rao-Scott design-corrected chi-square tests to evaluate whether trajectory classification was associated with sex or parental education and subsequently applied pairwise proportion tests with a Bonferroni correction. The irritability trajectory groups significantly differed by youth sex, χ^2 (3, $N = 11{,}862$) = 21.44, $p < .001$, and parental education, χ^2 (3, $N = 11{,}848$) = 34.27, $p < .001$. Relative to their male peers, female youth were disproportionately more likely to be classified in the low stable trajectory, χ^2 (1) = 11.62, $p < .01$. Conversely, male youth were more likely to be classified as either high decreasing, χ^2 (1) = 7.86, $p = .02$, or high chronic, χ^2 (1) = 11.78, $p < .01$, as compared to females. In comparison to youth whose parents did not have a 4-year college degree, youth having at least one parent who obtained a 4-year college degree or higher were more likely to be classified as following a low increasing trajectory, χ^2 (1) = 11.76, $p < .01$, whereas they were less likely to follow a high decreasing trajectory, χ^2 (1) = 10.96, $p < .01$, or a high chronic trajectory, χ^2 (1) = 14.07, $p < .01$.

Machine learning results

The proportions of the machine learning sample classified to each irritability trajectory were nearly identical to the proportions in the full sample: low stable ($n = 8{,}125$, 73.64%), high decreasing ($n = 988$, 8.95%), low increasing ($n = 1{,}339$, 12.14%), high chronic ($n = 582$, 5.27%). The Boruta feature selection procedure confirmed 51 predictors from the full set of 82 candidate features as potentially relevant. Refer to Tables 1 and 2.

Model performance varied by both algorithm and performance metric. Overall, the Elastic Net, Random Forest, and XGBoost algorithms achieved similar performance, while the KNN algorithm exhibited poorer performance. For example, Random Forest achieved the highest accuracy with a value of 0.74, which was slightly better than the accuracy achieved through the XGBoost algorithm (0.70) and through the Elastic Net algorithm (0.64), whereas the KNN algorithm resulted in the lowest accuracy with a

https://doi.org/10.1017/S0954579426101977 Published online by Cambridge University Press

Table 4. Performance metrics for machine learning models

	Elastic net	KNN	Random forest	XGBoost
Overall metrics				
Accuracy	0.64 [0.62, 0.66]	0.42 [0.39, 0.44]	0.74 [0.72, 0.76]	0.70 [0.68, 0.72]
Macro F1 score	0.42 [0.39, 0.44]	0.30 [0.28, 0.33]	0.43 [0.41, 0.46]	0.42 [0.39, 0.45]
Micro F1 score	0.64 [0.63, 0.66]	0.42 [0.39, 0.44]	0.74 [0.72, 0.76]	0.70 [0.68, 0.72]
True positive rate	0.46 [0.43, 0.49]	0.37 [0.34, 0.40]	0.42 [0.40, 0.45]	0.45 [0.42, 0.48]
Precision	0.40 [0.38, 0.42]	0.33 [0.32, 0.35]	0.46 [0.43, 0.50]	0.41 [0.38, 0.44]
Specificity	0.86 [0.85, 0.87]	0.81 [0.80, 0.82]	0.84 [0.83, 0.85]	0.85 [0.84, 0.86]
Average AUC				
Micro AUC	0.85 [0.84, 0.86]	0.71 [0.70, 0.72]	0.92 [0.91, 0.93]	0.89 [0.88, 0.90]
Macro AUC	0.78 [0.76, 0.80]	0.65 [0.63, 0.67]	0.82 [0.80, 0.83]	0.80 [0.78, 0.81]

Note: Values are performance metrics with 95% confidence intervals in brackets. True positive rate, precision, and specificity are macro scores. Area Under the Curve (AUC).

value of 0.42. A similar pattern emerged from the macro F1 scores, such that the Elastic Net, Random Forest, and XGBoost algorithms exhibited similar performance, whereas the KNN algorithm resulted in the lower macro F1 score. See Table 4 for details.

Separate calculations of true positive rate and precision for each trajectory revealed differences in performance across the set of algorithms. For example, XGBoost produced the highest true positive rate for the high chronic trajectory (0.56). On the other hand, Random Forest had the highest true positive rate for both the low stable trajectory (0.92) and the high decreasing trajectory (0.28). Finally, KNN generated the highest true positive rate for the low increasing trajectory (0.38). In terms of precision, the Random Forest algorithm had the highest precision for the three minority trajectories: high decreasing (0.28), low increasing (0.28), and high chronic (0.46), whereas the Elastic Net algorithm had the highest precision for the low stable trajectory (0.90). See Supplementary Table 5.

Visual inspection of the ROC Curves for each trajectory revealed similar patterns for the Elastic Net, Random Forest, and XGBoost algorithms. The ROC curves indicated that classification of the high chronic trajectory has the best tradeoff between true positive rate and false positive rate across the range of thresholds. The low stable trajectory had the next best tradeoff between true positive rate and false positive rate, while the two trajectories with change in irritability overtime were more difficult to accurately classify. The ROC curve for the low increasing trajectory was closest to the diagonal line for all algorithms indicating the poorer tradeoff between true positive rate and false positive rate. See Supplementary Table 5 and Supplementary Figure 1 for details.

Baseline predictors

The CBCL DSM-5 Conduct Scale was consistently ranked as the most important predictor by Elastic Net, Random Forest, and XGBoost. All three algorithms identified predictors from the youth psychopathology domain (CBCL Depression Scale, CBCL Anxiety Disorder Scale, and KSADS lifetime diagnosis of ODD) and predictors from the parent psychopathology domain (ASR irritability index and ASR antisocial index) among the top ten predictors. Four additional predictors were uniquely selected as top ten predictors by the Elastic Net algorithm: Negative Urgency Scale, Black race, Prosocial Reward Scale, and bilateral structural MRI volume for the circular sulcus of the insula. Three predictors were identified by both the Random Forest and XGBoost algorithms: CBCL Attention Deficit Hyperactivity Disorder (ADHD) Scale, Sleep–Wake Transition Disorders (SWTD) Score, and Disorders of Excessive Somnolence (DOES) Score. Lastly, the Random Forest and XGBoost algorithms each uniquely identified another predictor from the parent psychopathology domain. Specifically, Random Forest ranked the Adult Self Report (ASR) anxiety disorder score among the top ten important predictors, whereas XGBoost ranked the ASR depression score among the top ten important predictors. See Table 5 for details.

A supplementary analysis was conducted to evaluate model performance when only behavioral predictors were included. Results are reported in the Supplementary Materials and presented in Supplementary Tables 6–8 and Supplementary Figure 2.

Discussion

The current study identified distinct trajectories of youth irritability across the transition to early adolescence and top preadolescent predictors of the course of irritability using multiple machine learning methods. We identified four developmental trajectories for youth irritability assessed annually from preadolescence to early adolescence in the ABCD sample: low stable ($n = 8{,}692$, 73.28%), high decreasing ($n = 1{,}083$, 9.13%), low increasing ($n = 1{,}449$, 12.22%), and high chronic ($n = 638$, 5.38%). The four irritability trajectories illustrate differential longitudinal patterns in irritability severity as youth enter early adolescence, which is a sensitive developmental stage characterized by an increase in the onset of mental health disorders (Solmi et al., 2022). Specifically, the high chronic trajectory signals a high level of risk that persists from preadolescence into early adolescence. The low increasing trajectory indicates a moderate level of risk that exacerbates from preadolescence into early adolescence. Conversely, the high decreasing trajectory shows a mitigation of risk during the transition to early adolescence. Lastly, the low stable trajectory represents normative, low-risk levels of youth irritability throughout this developmental transition period. These irritability trajectories reflect previous findings in the irritability trajectory literature, which has consistently indicated that there is heterogeneity in the developmental course of irritability (e.g., Riglin et al., 2019; Wiggins et al., 2014; Yu et al., 2023b).

Three of the four supervised machine learning algorithms that we applied to predict youth irritability trajectories using a broad

https://doi.org/10.1017/S0954579426101977 Published online by Cambridge University Press

Table 5. Top ten predictors by algorithm for machine learning models

Rank	Elastic net	Random forest	XGBoost
1	CBCL DSM5 conduct Scale	CBCL DSM5 conduct scale	CBCL DSM5 conduct scale
2	ASR irritability index	CBCL DSM5 depression scale	CBCL DSM5 depression scale
3	CBCL DSM5 depression scale	CBCL DSM5 ADHD scale	CBCL DSM5 ADHD scale
4	KSADS ODD lifetime diagnosis	KSADS ODD lifetime diagnosis	KSADS ODD lifetime diagnosis
5	UPPS-P negative urgency	CBCL DSM5 anxiety disorder scale	CBCL DSM5 anxiety disorder scale
6	Race (Black)	ASR irritability index	ASR irritability index
7	CBCL DSM5 anxiety disorder scale	Disorders of Excessive Somnolence (DOES) score	Sleep–Wake Transition Disorders (SWTD) score
8	ASR antisocial scale	ASR anxiety disorder scale	ASR depression scale
9	Prosocial reward scale	Sleep–Wake Transition Disorders (SWTD) score	Disorders of Excessive Somnolence (DOES) score
10	Bilateral ROI volume (mm) of anterior segment of the circular sulcus of the insula	ASR antisocial scale	ASR antisocial scale

Note: Abbreviations in the top predictors include: Attention Deficit Hyperactivity Disorder (ADHD), Adult Self Report (ASR), Child Behavior Checklist (CBCL), Kiddie Schedule for Affective Disorders and Schizophrenia (KSADS), Oppositional Defiant Disorder (ODD), Region of Interest (ROI), and Urgency-Premeditation-Perseverance-Sensation Seeking-Positive Urgency (UPPS-P).

spectrum of baseline predictors achieved comparable performance, in contrast to substantially poorer predictive performance of the KNN algorithm. Further review showed variation in performance by trajectory. The algorithms were best at detecting the two trajectories that consistently fell on the extreme ends of the irritability range (low stable and high chronic). Specifically, the low stable trajectory had the highest true positive rate and precision across all algorithms indicating that the models could accurately identify true cases of this trajectory without a substantial excess of false positive predictions. The one-vs-rest measures (ROC curves and area under the curve) for the Elastic Net, Random Forest, and XGBoost algorithms illustrated a strength in differentiating the high chronic trajectory from the other three trajectories. On the other hand, the two irritability trajectories characterized by change in irritability over time (high decreasing and low increasing) were more challenging to accurately differentiate from the other trajectories for all algorithms using our set of preadolescent predictors. While the machine learning models were designed to differentiate between the four youth irritability trajectories, the strengths and weaknesses in performance inform cautious interpretation of the top predictors. Specifically, the top predictors likely provide more insight into identifying the high chronic and low stable trajectories with limited insight into identifying the low increasing and high decreasing trajectories.

Variable importance metrics revealed that the majority of the top predictors were from the youth psychopathology domain, particularly parent-report CBCL symptom scales. Notably, the CBCL DSM-5 Conduct Problems Scale was consistently identified as the most important predictor for irritability trajectory regardless of algorithm. We also observed CBCL ADHD, lifetime diagnosis of oppositional defiant disorder (ODD), and negative urgency (impulsivity) as top predictors, which is consistent with previous findings linking externalizing problems with youth irritability (e.g., Leadbeater & Homel, 2015). Internalizing symptoms on the CBCL (depression, anxiety disorder) also ranked among the top-ranked predictors for irritability trajectory, in accordance with prior research that found reciprocal effects between youth irritability and internalizing symptoms (Leadbeater & Homel, 2015; Savage et al., 2015). The literature strongly supports youth irritability as an antecedent to later psychopathology with some studies expanding on our knowledge with findings of bidirectional associations between externalizing and internalizing symptoms and youth irritability over time (Chad-Friedman et al., 2023). The current study indicates that preadolescent externalizing and internalizing symptoms predict the course of youth irritability, thus adding support for the less studied direction of this bidirectional association; nevertheless, our analyses cannot confirm the directionality, only temporal precedence of assessment.

Top predictors for youth irritability trajectory spanned domains beyond youth psychopathology, including parent psychopathology, youth physical health, demographics, and environment and culture. Among these domains, parental irritability was consistently the highest ranked, providing empirical evidence of familial liability for irritability specifically. The youth irritability literature indicates that irritability may be transmitted from parent to child through genetic risk factors (Roberson-Nay et al., 2015) and family environment (Sorcher et al., 2023; Wiggins et al., 2014). Parental antisocial symptoms and parental depression were also identified as top predictors. These top predictors align with prior research indicating that parent psychopathology can increase risk for youth irritability (Dougherty et al., 2013; Roberson-Nay et al., 2015; Sorcher et al., 2023; Wiggins et al., 2014; Yu et al., 2023a). Specifically, parental history of depression predicts youth irritability (Dougherty et al., 2013; Sorcher et al., 2023; Wiggins et al., 2014). Although to our knowledge no prior studies have found an association between parental antisocial behavior and youth irritability, a twin study, spanning preadolescence to late adolescence, reported shared genetic influence between irritability and antisocial behavior (Mikolajewski et al., 2017).

Multiple methods (Random Forest and XGBoost) highlighted the importance of two predictors from the youth sleep disturbance measure, specifically excessive somnolence or sleepiness and sleep–wake transition (i.e., difficulty falling or staying asleep). In particular, youth in the low stable trajectory tended to have lower sleep disturbance scores as compared to youth in any of the other three trajectories, with youth in the high chronic trajectory exhibiting the highest sleep disturbance scores. Prior research has found a bidirectional association between sleep duration and daily irritability such that shorter sleep duration was correlated with higher irritability symptoms in youth ages 8–18 (Meigs et al., 2024). The present findings highlight two features of sleep, somnolence and sleep–wake transition, that warrant further

https://doi.org/10.1017/S0954579426101977 Published online by Cambridge University Press

investigation into their association with the course of irritability during the transition to early adolescence.

In addition, one algorithm (Elastic Net) uniquely identified two other top predictors from the environment and demographic domains, respectively. These top predictors for the Elastic Net model included the Prosocial Reward Subscale from the School Risk and Protective Factors and race (Black or African American). Prosocial behavior has been linked to positive mental health outcomes (Oberle et al., 2023). Trajectories with lower preadolescent irritability (low stable and low increasing) had slightly higher prosocial reward, suggesting that this school-level variable may be a potential protective factor as well as a possible target for intervention. From the demographic domain, a higher proportion of Black or African American youth, as compared to White youth, were classified in the low stable trajectory, which is consistent with a previous sample (Yu et al., 2023b). The diversity of domains represented among the top predictors demonstrates the importance of a comprehensive clinical evaluation to identify youth at high risk.

The top ten important predictors were dominated by behavioral and clinical variables with only one algorithm selecting one structural MRI variable as the tenth-ranked predictor. Specifically, the Elastic Net algorithm indicated that structural volume of the anterior segment of the circular sulcus of the insula may be an important biomarker for irritability trajectories. Structural neuroimaging research has found a concurrent association between higher youth irritability (ages 9.1–13.9) and larger right insula volume (Dennis et al., 2019). As only one algorithm identified this region as a top predictor indicating that further research is needed. The paucity of structural MRI variables among the top predictors of the current models is consistent with prior work that has found that psychometric and behavioral variables outperform structural MRI variables in predicting adolescent mental health outcomes (Chavanne et al., 2023; Whelan et al., 2014). It is likely that common method variance increased the relevance of variables obtained from questionnaires, particularly parent-report measures, over the relevance of neurobiological variables because the irritability trajectories were ascertained from caregiver-reported CBCL.

The current report has several strengths. The use of a large, racially diverse sample enabled us to apply data-driven methods during a critical developmental transition to early adolescence. Moreover, the wide range of predictors that were used in the machine learning analyses provides avenues for future research and may increase the generalizability of our results. Despite these strengths, there are some noteworthy limitations that can inform future research directions. First, this study relied on annual reports of caregiver-rated youth irritability using the three-item CBCL irritability index. Future research could improve the measure of irritability by integrating youth report and multidimensional measures of irritability. Second, this study included annual assessments from preadolescence to early adolescence and future research is needed to extend the assessments into adolescence and early adulthood. Third, although we were careful not to include any items that overlap between the CBCL irritability scale and the other CBCL scales used as baseline predictors, we recognize that irritability is a symptom in several diagnoses including oppositional defiant, depressive, and anxiety disorders. Thus, it is possible that irritability symptoms occurring within these baseline diagnoses may influence our results; however, because only one diagnosis with irritability as a symptom made it to the top 10 predictors (lifetime ODD diagnosis), item overlap does not appear to be driving our findings. Nevertheless, future research should investigate whether a lifetime diagnosis of ODD with or without irritable mood differentially predicts the course of irritability or influences its rank as a predictor. Fourth, the machine learning models showed modest performance in predicting irritability trajectories. The limitations of the machine learning models bring to light new research questions for further investigation. Specifically, poorer classification of the low increasing and high decreasing trajectories suggests that the preadolescent predictors were insufficient at differentiating these trajectories from the others. Research that incorporates distal and proximal factors across development may improve predictive accuracy and discover important predictors for youth irritability trajectories. Additionally, future research could build binary machine learning models that discriminate between the high chronic and high decreasing and low stable and low increasing trajectories respectively. Furthermore, better true positive rate and precision for the low stable trajectory as compared to the other trajectories suggests more homogeneity in the low-risk group. In contrast, the high-risk groups may be characterized by more heterogeneity across risk factors. Future research could investigate whether there are different risk profiles within each of the high-risk trajectories. Finally, the data-driven approaches applied in this work offer advantages for analyzing large datasets and building predictive models; however, it is crucial to follow up with more targeted research to assess specific hypotheses and expand on this research.

In conclusion, this report expanded on the irritability trajectory literature by characterizing irritability trajectories across the developmental transition from preadolescence to early adolescence in a large, national sample. Subsequently, we applied a robust machine learning approach with multiple methods which enabled us to evaluate a broad set of predictors of the course of irritability during a crucial developmental period. Machine learning methods are data-driven, exploratory approaches designed for prediction rather than the identification of causal associations. However, examining the predictors that contribute most strongly to model performance can help researchers generate new hypotheses and identify variables that may be incorporated into developmental theory following additional confirmatory research (Van Lissa, 2023). Thus, our top predictors warrant further investigation to identify their role in the course of high-risk irritability. It is possible that some may serve as potential targets for intervention and prevention efforts to alter the course of high-risk irritability trajectories.

Supplementary material. The supplementary material for this article can be found at https://doi.org/10.1017/S0954579426101977.

Data availability statement. Data for the current study were from the Adolescent Brain Cognitive Development (ABCD) Study. De-identified data from the ABCD Study are available to qualified researchers through the NIMH Data Archive (NDA). To request access, submit a Data Access Request (DAR) through the ABCD Study website (abcdstudy.org) in accordance with the instructions described at https://nda.nih.gov/abcd/request-acess. Code available upon request.

Acknowledgements. Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive DevelopmentSM (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9–10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039,

https://doi.org/10.1017/S0954579426101977 Published online by Cambridge University Press

U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators.

We acknowledge the support of the University of Maryland Baltimore, Institute for Clinical & Translational Research (ICTR) and the University of Maryland Strategic Partnership: MPowering the State (MPower), which provided training and educational stipend (LSJ).

Funding statement. Research reported in this publication was supported by the National Institute of Mental Health under award number R01MH122487 (MPIs: Wiggins, Dougherty).

Competing interests. The authors declare no competing interests to declare.

Pre-registration statement. The analytic plan was pre-registered (https://doi.org/10.17605/OSF.IO/GE3XC) prior to conducting this research [December 22, 2021]. Due to the data-driven nature of this study no a priori hypotheses were made. Given the comprehensive nature of the preregistration, the findings will be disseminated across multiple papers, of which this is one.

References

Achenbach, T. M. (2009). The Achenbach system of empirically based assessment (ASEBA): Development, findings, theory, and applications. University of Vermont, Research Center for Children, Youth, & Families.

Barch, D. M., Albaugh, M. D., Avenevoli, S., Chang, L., Clark, D. B., Glantz, M. D., & Sher, K. J. (2018). Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. *Developmental Cognitive Neuroscience*, *32*, 55–66. https://doi.org/10.1016/j.dcn.2017.10.010

Bartoszuk, B. M., & Gagolewski, M. (2020). Variable importance plots: An introduction to the vip package. https://doi.org/10.32614/rj-2020-013

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., & Dale, A. M. (2018). The adolescent brain cognitive development (ABCD) study: Imaging acquisition across 21 sites. *Developmental Cognitive Neuroscience*, *32*, 43–54. https://doi.org/10.1016/j.dcn.2018.03.001

Chad-Friedman, E., Galano, M. M., Lemay, E. P., Olino, T. M., Klein, D. N., & Dougherty, L. R.2023). Parsing between-and within-person effects: Longitudinal associations between irritability and internalizing and externalizing problems from early childhood through adolescence. *Development and Psychopathology*, *35*(3), 1371–1381. https://doi.org/10.1017/s0954579421001267

Chavanne, A. V., Paillère Martinot, M. L., Penttilä, J., Grimmer, Y., Conrod, P., Stringaris, A., & Artiges, E. (2023). Anxiety onset in adolescents: A machine-learning prediction. *Molecular Psychiatry*, *28*(2), 639–646. https://doi.org/10.1038/s41380-022-01840-z

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357. https://doi.org/10.1613/jair.953

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(3), 464–504. https://doi.org/10.1080/10705510701301834

Chen, T., & Guestrin, C. (2016, August). Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794). https://doi.org/10.1145/2939672.293978

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Copeland, W. E., Brotman, M. A., & Costello, E. J. (2015). Normative irritability in youth: Developmental findings from the Great Smoky Mountains study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(8), 635–642. https://doi.org/10.1016/j.jaac.2015.05.008

Cunningham, P., & Delany, S. J. (2021). K-nearest neighbour classifiers-a tutorial. *ACM Computing Surveys (CSUR)*, *54*(6), 1–25. https://doi.org/10.1145/3459665

Dennis, E. L., Humphreys, K. L., King, L. S., Thompson, P. M., & Gotlib, I. H. (2019). Irritability and brain volume in adolescents: Cross-sectional and longitudinal associations. *Social Cognitive & Affective Neuroscience*, *14*(7), 687–698. https://doi.org/10.1093/scan/nsz053

Dougherty, L. R., Smith, V. C., Bufferd, S. J., Stringaris, A., Leibenluft, E., Carlson, G. A., & Klein, D. N. (2013). Preschool irritability: Longitudinal associations with psychiatric disorders at age 6 and parental psychopathology. *Journal of the American Academy of Child & Adolescent Psychiatry*, *52*(12), 1304–1313. https://doi.org/10.1016/j.jaac.2013.09.007

Evans, S. C., Corteselli, K. A., Edelman, A., Scott, H., & Weisz, J. R. (2023). Is irritability a top problem in youth mental health care? A multi-informant, multi-method investigation. *Child Psychiatry & Human Development*, *54*(4), 1027–1041. https://doi.org/10.1007/s10578-021-01301-8

Galatzer-Levy, I. R., Huang, S. H., & Bonanno, G. A. (2018). Trajectories of resilience and dysfunction following potential trauma: A review and statistical evaluation. *Clinical Psychology Review*, *63*, 41–55. https://doi.org/10.1016/j.cpr.2018.05.008

Greenwell, B. M., & Boehmke, B. C. (2020). Variable importance plots—An introduction to the vip package. *The R Journal*, *12*(1), 343–366. https://doi.org/10.32614/RJ-2020-013

Hill, E. D., Kashyap, P., Raffanello, E., Wang, Y., Moffitt, T. E., Caspi, A., & Posner, J. (2025). Prediction of mental health risk in adolescents. *Nature Medicine*, *31*(6), 1840–1846. https://doi.org/10.1038/s41591-025-03560-7

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Jiang, T., Gradus, J. L., & Rosellini, A. J. (2020). Supervised machine learning: A brief primer. *Behavior Therapy*, *51*(5), 675–687. https://doi.org/10.1016/j.beth.2020.05.002

Jordanov, I., Petrov, N., & Petrozziello, A. (2018). Classifiers accuracy improvement based on missing data imputation. *Journal of Artificial Intelligence and Soft Computing Research*, *8*(1), 31–48. https://doi.org/10.1515/jaiscr-2018-0002

Kessel, E. M., Dougherty, L. R., Hubachek, S., Chad-Friedman, E., Olino, T., Carlson, G. A., & Klein, D. N. (2021). Early predictors of adolescent irritability. *Child and Adolescent Psychiatric Clinics*, *30*(3), 475–490. https://doi.org/10.1016/j.chc.2021.04.002

Kursa, M. B., Jankowski, A., & Rudnicki, W. R.2010). Boruta–A system for feature selection. *Fundamenta Informaticae*, *101*(4), 271–285. https://doi.org/10.3233/fi-2010-288

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of statistical software*, *36*, 1–13. https://doi.org/10.18637/jss.v036.i11

Leadbeater, B. J., & Homel, J. (2015). Irritable and defiant sub-dimensions of ODD: Their stability and prediction of internalizing symptoms and conduct problems from adolescence to young adulthood. *Journal of Abnormal Child Psychology*, *43*, 407–421. https://doi.org/10.1007/s10802-014-9908-3

Mathew, J., Kshirsagar, R., Abidin, D. Z., Griffin, J., Kanarachos, S., James, J., & Fitzpatrick, M. E. (2023). A comparison of machine learning methods to classify radioactive elements using prompt-gamma-ray neutron activation data. *Scientific Reports*, *13*(1), 9948. https://doi.org/10.1038/s41598-023-36832-8

Meigs, J. M., Kiderman, M., Kircanski, K., Cardinale, E. M., Pine, D. S., Leibenluft, E., & Naim, R. (2024). Sleepless nights, sour moods: Daily sleep-irritability links in a pediatric clinical sample. *Journal of Child Psychology & Psychiatry*, *65*(9), 1175–1183. https://doi.org/10.1111/jcpp.13959

Mikolajewski, A. J., Taylor, J., & Iacono, W. G. (2017). Oppositional defiant disorder dimensions: Genetic influences and risk for later psychopathology. *Journal of Child Psychology & Psychiatry*, *58*(6), 702–710. https://doi.org/10.1111/jcpp.12683

Muthén, L. K., & Muthén, B. O. (2017). Mplus Statistical Analysis with Latent Variables User's Guide (Version 8). Los Angeles, CA Authors.

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, *4*(4), 440. https://doi.org/10.1037/tps0000176

Oberle, E., Ji, X. R., & Molyneux, T. M. (2023). Pathways from prosocial behaviour to emotional health and academic achievement in early adolescence. *The Journal of Early Adolescence*, *43*(5), 632–653. https://doi.org/10.1177/02724316221113349

Orri, M., Galera, C., Turecki, G., Boivin, M., Tremblay, R. E., Geoffroy, M. C., & Côté, S. M. (2019). Pathways of association between childhood irritability and adolescent suicidality. *Journal of the American Academy of Child & Adolescent Psychiatry*, *58*(1), 99–107. https://doi.org/10.1016/j.jaac.2018.06.034

Riglin, L., Eyre, O., Thapar, A. K., Stringaris, A., Leibenluft, E., Pine, D. S., & Thapar, A. (2019). Identifying novel types of irritability using a developmental genetic approach. *American Journal of Psychiatry*, *176*(8), 635–642. https://doi.org/10.1176/appi.ajp.2019.18101134

Roberson-Nay, R., Leibenluft, E., Brotman, M. A., Myers, J., Larsson, H., Lichtenstein, P., & Kendler, K. S. (2015). Longitudinal stability of genetic and environmental influences on irritability: From childhood to young adulthood. *American Journal of Psychiatry*, *172*(7), 657–664. https://doi.org/10.1176/appi.ajp.2015.14040509

Rubens, S. L., Evans, S. C., Becker, S. P., Fite, P. J., & Tountas, A. M. (2017). Self-reported time in bed and sleep quality in association with internalizing and externalizing symptoms in school-age youth. *Child Psychiatry & Human Development*, *48*, 455–467. https://doi.org/10.1007/s10578-016-0672-1

Sathyanarayana, S., & Mohanasundaram, T. (2024). Fit indices in structural equation modeling and confirmatory factor analysis: Reporting guidelines. *Asian Journal of Economics, Business and Accounting*, *24*(7), 561–577. https://doi.org/10.9734/ajeba/2024/v24i71430

Savage, J., Verhulst, B., Copeland, W., Althoff, R. R., Lichtenstein, P., & Roberson-Nay, R. (2015). A genetically informed study of the longitudinal relation between irritability and anxious/depressed symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry*, *54*(5), 377–384. https://doi.org/10.1016/j.jaac.2015.02.010

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Solmi, M., Radua, J., Olivola, M., Croce, E., Soardo, L., Salazar de Pablo, G., & Fusar-Poli, P. (2022). Age at onset of mental disorders worldwide: Large-scale meta-analysis of 192 epidemiological studies. *Molecular Psychiatry*, *27*(1), 281–295. https://doi.org/10.1038/s41380-021-01161-7

Sorcher, L. K., Goldstein, B. L., Finsaas, M. C., Carlson, G. A., Klein, D. N., & Dougherty, L. R. (2022). Preschool irritability predicts adolescent psychopathology and functional impairment: A 12-year prospective study. *Journal of the American Academy of Child & Adolescent Psychiatry*, *61*(4), 554–564. https://doi.org/10.1016/j.jaac.2021.08.016

Sorcher, L. K., Mennies, R. J., Robeson, M., Seeley, J. R., Klein, D. N., Dougherty, L. R., & Olino, T. M. (2023). Offspring irritability: Associations with parental psychopathology and personality. *European Child & Adolescent Psychiatry*, *32*(9), 1691–1699. https://doi.org/10.1007/s00787-022-01985-w

Stringaris, A., Zavos, H., Leibenluft, E., Maughan, B., & Eley, T. C. (2012). Adolescent irritability: Phenotypic associations and genetic links with depressed mood. *American Journal of Psychiatry*, *169*(1), 47–54. https://doi.org/10.1176/appi.ajp.2011.10101549

Van Lissa, C. J. (2023). Developmental data science: How machine learning can advance theory formation in developmental psychology. *Infant and Child Development*, *32*(6), e2370. https://doi.org/10.1002/icd.2370

Volkow, N. D., Koob, G. F., Croyle, R. T., Bianchi, D. W., Gordon, J. A., Koroshetz, W. J., Pérez-Stable, E. J., Riley, W. T., Bloch, M. H., & Conway, K. (2018). The conception of the ABCD study: From substance use to a broad NIH collaboration. *Developmental Cognitive Neuroscience*, *32*, 4–7. https://doi.org/10.1016/j.dcn.2017.10.002

Wakschlag, L. S., MacNeill, L. A., Pool, L. R., Smith, J. D., Adam, H., Barch, D. M., & Allen, N. B. (2024). Predictive utility of irritability "in context": Proof-of-principle for an early childhood mental health risk calculator. *Journal of Clinical Child & Adolescent Psychology*, *53*(2), 231–245. https://doi.org/10.1080/15374416.2023.2188553

Whelan, R., Watts, R., Orr, C. A., Althoff, R. R., Artiges, E., Banaschewski, T., & Garavan, H. (2014). Neuropsychosocial profiles of current and future adolescent alcohol misusers. *Nature*, *512*(7513), 185–189. https://doi.org/10.1038/nature13402

Wiggins, J. L., Mitchell, C., Stringaris, A., & Leibenluft, E. (2014). Developmental trajectories of irritability and bidirectional associations with maternal depression. *Journal of the American Academy of Child & Adolescent Psychiatry*, *53*(11), 1191–1205. https://doi.org/10.1016/j.jaac.2014.08.005

Xiang, Q., Chen, K., Peng, L., Luo, J., Jiang, J., Chen, Y., & Zhou, X. (2022). Prediction of the trajectories of depressive symptoms among children in the adolescent brain cognitive development (ABCD) study using machine learning approach. *Journal of Affective Disorders*, *310*, 162–171. https://doi.org/10.1016/j.jad.2022.05.020

Yu, Q., Hernandez, B., Swineford, C., Walker, N., MacNeill, L., Zhang, Y., & Wiggins, J. L. (2023a). Toward an optimized assessment of adolescent psychopathology risk: Multilevel environmental profiles and child irritability as predictors. *JCPP Advances*, e12180. https://doi.org/10.1002/jcv2.12180

Yu, Q., Hodgdon, E. A., Kryza-Lacombe, M., Osuna, L., Bozzetto, L. E., Ciro, D., & Wiggins, J. L. (2023b). Roads diverged: Developmental trajectories of irritability from toddlerhood through adolescence. *Journal of the American Academy of Child & Adolescent Psychiatry*, *62*(4), 457–471. https://doi.org/10.1016/j.jaac.2022.07.849

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

Zucker, R. A., Gonzalez, R., Ewing, S. W. F., Paulus, M. P., Arroyo, J., Fuligni, A., & Wills, T. (2018). Assessment of culture and environment in the Adolescent Brain and Cognitive Development Study: Rationale, description of measures, and early data. *Developmental Cognitive Neuroscience*, *32*, 107–120. https://doi.org/10.1016/j.dcn.2018.03.004

https://doi.org/10.1017/S0954579426101977 Published online by Cambridge University Press