The Classification of Conduct Disorder Using a Biopsychosocial Model and Machine Learning Method

Lena Chan

Department of Psychology, Yale University

PSYC 499: Senior Essay

Dr. Arielle Baskin-Sommers

April 30, 2021

Author Note

This thesis is submitted in partial fulfillment of the Psychology Bachelor of Science degree. This work was advised by Arielle Baskin-Sommers.

Abstract

Conduct disorder (CD), a psychiatric diagnosis describing aggressive and delinquent behavior among children and adolescents, is influenced and maintained by distinct social, psychological, and biological risk factors. However, no studies have assessed the extent to which biopsychosocial factors collectively influence and predict CD onset and severity. The present study tests the biopsychosocial etiological model and identifies the most influential risk factors for its onset by training machine learning classifiers that predict CD status among 4667 youth from the Adolescent Brain Cognitive Development Study. We constructed four feed-forward neural nets to classify a subject as meeting or not meeting the diagnostic criteria for CD based on their exposure to (a) social risk factors, (b) biological risk factors, (c) psychological risk factors, and (d) factors across all three risk domains. The classifier trained *across* risk domains yielded the highest accuracy of 80.8% and outperformed the classifiers trained *within* individual domains. Furthermore, comorbid symptoms of attention-deficit/hyperactive disorder (ADHD) and oppositional defiant disorder (ODD), the experience of physical harm and criticism within the family, and reduced cortical-subcortical connectivity were the strongest predictors of CD in this study. Together, these findings reinforce the efficacy of the biopsychosocial model and suggest that researchers must investigate the mental health, family environment, and neural topography of at-risk youth to adequately characterize CD.

Keywords: conduct disorder; biopsychosocial; comorbidity; neighborhood; family; neural topology; graph analysis; machine learning; neural net; classifier

Conduct disorder (CD) is a psychiatric diagnosis that describes severe antisocial, aggressive, and deceitful behavior among children and adolescents (American Psychiatric Association, 2013). Youth with CD display widespread social and neurocognitive deficits, such as adverse family relationships, academic underachievement, impaired executive functioning, and emotion dysregulation (Cappadocia et al., 2009; Frick, 2006; Frick & Dickens, 2006; Moffitt, 1993; Ogilvie et al., 2011). CD presents a high risk for developing other mental health conditions, including substance abuse, major depression, and antisocial personality disorder (Colman et al., 2009). The violent and disruptive behaviors that characterize the disorder pose serious physical and emotional harm to the parents, teachers, and peers surrounding the youth (Frick & Dickens, 2006; Lytton, 1990). Furthermore, CD burdens the criminal justice system, with approximately 40% of youth who offend meeting for its diagnosis (Dadds et al., 2005; Vermeiren, 2003). Given the severe impact of CD on youth and society, it is essential that researchers investigate the individual and environmental factors underlying its development.

Brief Review of Social, Psychological and Biological Factors Related to CD

Decades of research highlight that social, psychological, and biological risk factors influence the onset and maintenance of CD. Social factors, including living in poverty, exposure to neighborhood violence, inconsistent and harsh discipline from caregivers, and interactions with antisocial peers, predict the disorder (Frick et al., 1992; Loeber et al., 2000; Murray & Farrington, 2010). Psychological factors, such as impulsivity, substance misuse, difficult temperament, neuropsychological deficits, and psychiatric comorbidity with attention-deficit/hyperactive disorder (ADHD) and oppositional defiant disorder (ODD), are strongly associated with CD diagnosis (Azeredo et al., 2018; Biederman et al., 1991; Dodge et al., 1990). Finally, biological factors relating to neural abnormalities reliably characterize the disorder

(Baker et al., 2015; Blair et al., 2014; Moffitt, 1993). Youth with CD show several structural and functional differences in the amygdala, hippocampus, caudate, orbital frontal cortex, anterior cingulate cortex, superior temporal gyrus, prefrontal cortex, insula, and fusiform gyrus (Baker et al., 2015; Blair, 2013; Fairchild et al., 2019). Additionally, affected youth display atypical neural communication between various neural structures (e.g., reduced connectivity in emotional cortical-subcortical networks, low-level perceptual networks, and high-order cognitive networks and heightened connectivity in the frontoparietal network; Aghajani et al., 2016; Cohn et al., 2015; Finger et al., 2011; Lu et al., 2015; Zhou et al., 2016). Recent advances in graph theory, a higher-level connectivity analysis that examines the organization of neural networks and the brain as a whole, highlights that CD is related to functional deficits in communication efficiency and information segregation and structural deficits in node distribution (Jiang et al., 2016; Lu et al., 2017). More specifically, Tillem et al. (2021) identified greater global clustering of nodes and fewer cortical-subcortical connections among CD subjects. These widespread neural abnormalities alongside the familial, interpersonal, and cognitive difficulties experienced by affected youth highlight the importance of assessing a biopsychosocial model for CD development (Borsboom, 2017; Borsboom & Cramer 2013; Frick & Dickens, 2006; Molenaar & Campbell, 2009).

Unfortunately, the research on CD often examines the effects of only a few factors at a time (Bolón-Canedo et al., 2015; Dwyer et al., 2018; Ioannidis, 2016). To date, no studies have assessed the collective influence of biological, psychological, and social risk factors on CD. As a result, existing research may not account for the heterogeneous systems and interacting factors that influence CD onset and expression (Borsboom, 2017; Borsboom & Cramer 2013; Ioannidis, 2016; Molenaar & Campbell, 2009). Furthermore, the development of predictive models for CD

is extremely limited (De Brito et al., 2009; Yarkoni & Westfall, 2017). As a result, the extent to which characteristics and conditions such as neighborhood disadvantage, negative parenting styles, and disrupted neural connectivity reliably predict CD status or identify high-risk subjects is largely unexplored (Molenaar & Campbell, 2009; Yarkoni & Westfall, 2017).

Machine Learning as a Method to Test the Biopsychosocial Model of CD

Substantial research has systematically identified factors important for understanding CD. However, a paucity of empirical work has examined how these factors collectively influence the disorder. Machine learning is a novel technique that can assess high-dimensional relationships between risk factors and CD. This method enables models to automatically improve through experience, recognize complex trends, and conduct evidence-based decision-making (Jordan & Michell, 2015). Machine learning algorithms generate abstract, lower-level representations of data by conducting non-linear transformations that selectively turn off noise and preserve meaningful signals (LeCun et al., 2015; Shafiei et al., 2020; Zhang, 2018a). As a result, this technique is especially suited for analyzing linear and non-linear relationships between predictors and outcomes, identifying salient predictors, and suppressing non-relevant predictors (Belloni et al., 2014; Svozil et al., 1997). An additional advantage of machine learning is that the approach does not require assumptions regarding the distribution of the data or interactions between variables (Arnold et al., 2020; Jha et al., 2017).

Classifiers are specific types of machine learning models that extract trends and generate individual predictions from high-dimensional data (Shafiei et al., 2020). These models can assess whether group-level differences—regardless of effect size or statistical significance—translate into reliable individual differences (Pauli et al., 2020). Classifiers also can learn key features, such as unique neural signatures or combinations of risk factors, that identify subjects at high

risk for certain disorders (Koutsouleris et al., 2009). In the future, these models may act as decision aids for unclear diagnostic circumstances (Davatzikos et al., 2008; Dwyer et al., 2018; Redlich et al., 2014). Thus, machine learning assistance for clinical psychology has the potential to improve the prevention, diagnosis, and treatment of mental health conditions like CD.

A common criticism of machine learning methods is that they are "black box" algorithms that predict outcomes without interpreting the features most relevant to their analyses (Belloni et al., 2014). However, developers recently have built algorithms that ameliorate this shortcoming by determining the relative importance of individual predictors when classifying out-of-sample outcomes (Olden & Jackson, 2002; Shafiei et al., 2020). Furthermore, experimenters can compare a classifier's predictive performance when different sets of features are included to determine relative significance across risk factor domains (Yarkoni & Westfall, 2017). In the clinical context, these new interpretable qualities allow researchers to assess the collective and individual influence of biopsychosocial risk factors on psychological outcomes.

Recent applications of machine learning to clinical psychology predict psychiatric diagnoses using high-dimensional neuroimaging data (Arbabshirani et al. 2017; Dwyer et al., 2018). Primarily focused on separating individuals with Alzheimer's disease, depression, schizophrenia, and ADHD from healthy controls, these analyses achieve greater than 75% accuracy in their classifications and demonstrate machine learning's promising contributions to psychology (Arbabshirani et al. 2017; Davatzikos et al., 2008; Fu et al., 2008; Kambeitz et al., 2015; Kambeitz et al., 2017; Kim et al., 2016; Kuang & He, 2014; Schnack & Kahn, 2016). However, to date, only six published studies have applied machine learning to CD classification. Researchers have predicted CD diagnoses using gray matter volume data with 76% to 85% accuracy (Zhang et al., 2018b; Zhang et al., 2019; Zhang et al., 2020a). Tor et al. (2021)

developed classifiers from resting-state electroencephalogram signals that differentiated subjects with ADHD, CD, and comorbid ADHD + CD, while Zhang et al. (2020b) trained a support-vector machine from resting-state graph metrics that correctly identified CD groups 94.44% of the time. Finally, Pauli et al. (2020) separated subjects with CD from healthy controls based on negative and positive parenting characteristics, demonstrating 69% to 75% accuracy. These studies suggest that abnormal brain structure, disrupted neural activity, and harsh parenting strategies independently predict CD among youth.

However, this prediction-based research is limited to individual domains of risk factors. No researchers have developed classifiers for CD built on all three biological, psychological, and social domains nor compared the predictive power of classifiers built on subsets of these domains. Furthermore, many of these studies lack meaningful analyses of the individual risk factors most relevant in classification tasks. The purpose of the present study is to address these gaps in the literature by training four machine learning models to classify CD diagnoses based on neighborhood and family measures, brain topology, comorbid psychopathology, and the combination of these features. Furthermore, we intend to determine the relative importance of each feature during prediction and identify the "riskiest" factors across all four models.

The Present Study

We trained four feed-forward neural nets[1] (FNN) to predict if a subject meets the criteria for CD based on (a) neighborhood and family measures, (b) global and local connectivity graph metrics, (c) comorbid ADHD and ODD symptomatology, and (d) the combination of all three risk factor domains. We included graph metrics to account for widespread connectivity

[1] Despite their name, neural nets are not exclusive to neuroscientific variables. The algorithm adopts this title because its nodes or units of information processing are loosely modeled on the neurons in a biological brain. Neural nets can process both neural and non-neural data.

anomalies and build upon previous findings, which suggest that differences in global and local organization partially account for CD-specific deficits in neurocognitive functioning and emotional processing (Tillem et al., 2021). All data was collected from the Adolescent Brain Cognitive Development Study (ABCD Study), a longitudinal multi-site experiment following the brain development and health of over 10,000 eight to 11-year-olds (Casey et al., 2018). For each model, we calculated the accuracy of CD classification for unseen subjects and determined the relative importance of individual risk factors. These analyses test the biopsychosocial model for CD by comparing the accuracy rates of classifiers built within or across risk factor domains. Furthermore, insight into the classifiers' most predictive factors may shed light on causal theories of development and improve our ability to identify children at greatest risk for CD.

Methods

Participants

The ABCD Study is designed to include a diverse population that reflects the demographics of the United States (Garavan et al., 2018). Participants consisted of children between eight and 11 who completed the baseline session of the ABCD Study. Researchers gathered signed informed consent from parents and written assent from children prior to the study (Garavan et al., 2018). Given the large number of families with multiple children participating in the ABCD Study, siblings were overrepresented in the sample (Iacono et al., 2018). To control for any family-related effects, one child from each family was randomly selected for the current analysis (see Table 1 for a summary of demographics).

Table 1. Demographics Summary

Variable	*n*	Mean	Std. Dev.	Min	Max	Correlations 1	2[a]	3[b]	4
1. Age	4667	9.51	.51	8.00	11.00	—	.01	-.03*	-.05*
2. Sex[a]	4667						—	-.02	.08*
Male	2302								
Female	2365								
3. Race[b]	4667							—	.08*
White	2483								
Black	687								
Hispanic	977								
Asian	98								
Other	470								
4. CD Symptomatology	4667	.29	.79	.00	10.00				—
CD Diagnosis	204								
No CD Diagnosis	4463								

* Significant at alpha = .05 under the Spearman correlation test.

[a] Spearman correlations were used to examine the effect of Sex (dichotomously-coded).

[b] Spearman correlations were used to examine the effect of Race (dichotomously-coded, white vs. non-white).

Measures

CD Diagnosis

ABCD Study researchers administer the Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime version (K-SADS-PL) to measure CD symptomatology and diagnosis according to the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition's (DSM-5) criteria (Kaufman et al., 2013). Consistent with the prevalence rate of approximately 4.4% for young adolescents in the United States, 4.37% (n = 204) of the sample was diagnosed with CD (Merikangas et al., 2010).

Neighborhood and Family

The ABCD Study's Parental Monitoring Survey (PMQ) gathers youth-reported measures of parental monitoring and supervision, and the Family Environment Scale-Family Conflict Subscale (FES) obtains parent-reported measures of family dynamics, cohesion, expressiveness, and conflict (see Table 2 for descriptions of all environmental measures). Additionally, the severity of neighborhood crime is assessed with the ABCD Study's Neighborhood Safety/Crime Survey (NSC) administered to subjects' parents.

Table 2. Descriptions of Environmental Measures

Metric	Definition
Neighborhood Crime	1-5 interval response to "My neighborhood is safe from crime." (1 indicates "Strongly disagree"; 5 indicates "Strongly agree")
Family Fighting	0/1 binary response to "We fight a lot in our family/Peleamos mucho en nuestra familia." (0 indicates "No"; 1 indicates "Yes")

Family Anger	0/1 binary response to “Family members rarely become openly angry/Los miembros de la familia raramente se enojan abiertamente.” (0 indicates “No”; 1 indicates “Yes”)
Family Throwing Objects	0/1 binary response to “Family members sometimes get so angry they throw things/Los miembros de la familia algunas veces se enojan tanto que avientan cosas.” (0 indicates “No”; 1 indicates “Yes”)
Family Temper	0/1 binary response to “Family members hardly ever lose their tempers/Los miembros de la familia dificilmente pierden su temperamento.” (0 indicates “No”; 1 indicates “Yes”)
Family Criticism	0/1 binary response to “Family members often criticize each other/Los miembros de la familia con frecuencia se critican unos a otros.” (0 indicates “No”; 1 indicates “Yes”)
Family Hitting	0/1 binary response to “Family members sometimes hit each other/Los miembros de la familia algunas veces se golpean unos a otros.” (0 indicates “No”; 1 indicates “Yes”)
Family Disagreement	0/1 binary response to “If there is a disagreement in our family, we try hard to smooth things over and keep the peace/Si hay un desacuerdo en nuestra familia, hacemos todo lo posible por resolverlo y conservar la paz.” (0 indicates “No”; 1 indicates “Yes”)
Family Competition	0/1 binary response to “Family members often try to one-up or outdo each other/Los miembros de la familia con frecuencia tratan de superar a los demás.” (0 indicates “No”; 1 indicates “Yes”)
Family Raising Voices	0/1 binary response to “In our family, we believe you don't ever get anywhere by raising your voice/En nuestra familia, creemos que no se llega a nada levantando la voz.” (0 indicates “No”; 1 indicates “Yes”)
Spanish-Speaking	0/1 binary response to “¿Español? (1, Sí).” (0 indicates “No”; 1 indicates “Yes”)

Monitoring Youth Location	1-5 interval response to “How often do your parents/guardians know where you are?” (1 indicates “Never”; 5 indicates “Always”)
Monitoring Youth Interactions	1-5 interval response to “How often do your parents know who you are with when you are not at school and away from home?” (1 indicates “Never”; 5 indicates “Always”)
Youth/Parent Communication	1-5 interval response to “If you are at home when your parents or guardians are not, how often do you know how to get in touch with them?” (1 indicates “Never”; 5 indicates “Always”)
Monitoring Youth Plans	1-5 interval response to “How often do you talk to your mom/dad or guardian about your plans for the coming day, such as your plans about what will happen at school or what you are going to do with friends?” (1 indicates “Never”; 5 indicates “Always”)
Dinner Frequency	1-5 interval response to “In an average week, how many times do you and your parents/guardians eat dinner together?” (1 indicates “Never”; 5 indicates “Always”)

ADHD and ODD Symptomatology

ADHD and ODD symptomatology is measured by the DSM-5 ADHD and ODD criteria of the K-SADS-PL. ADHD is a mental health disorder defined by highly hyperactive and impulsive behaviors, while ODD is a childhood disorder characterized by a persistent pattern of hostility and defiance against authority figures (Biederman et al., 1991; Loeber et al., 2000).

Graph Metrics

The ABCD Study acquires 15-20 minutes of resting-state fMRI data from each participant. Subjects are instructed to remain still and gaze at a central fixation cross during data collection. Imaging parameters are uniform across all 21 data collection sites and scanner models (Casey et al., 2018; Hagler et al., 2019).

Graph metrics of resting-state fMRI connectivity were calculated by Tillem et al. (2021) through unweighted, undirected graph analysis. Tillem et al. (2021) constructed a 14x14 connectivity matrix from (a) cortical-cortical connectivity data between the major resting-state networks (default, dorsal attention, frontoparietal, salience, ventral attention, cingulo-opercular, cingulo-parietal, visual, auditory, retrosplenial-temporal, sensorimotor-hand, sensorimotor-mouth, and "other" networks) provided by Hagler et al. (2019) and (b) cortical-subcortical connectivity data across the subcortical structures provided by Tillem et al. (2019). Tillem et al. (2021) converted these matrices into graphs with 14 nodes, representing 13 cortical networks and one subcortical node. From these graphs, they extracted global metrics that describe topology across the brain and local metrics that measure node-level organization (see Table 3 for a description of the graph metrics).

Graph metrics such as efficiency and path length measure the integration of information in the brain (Rubinov & Sporns, 2010; Stam & Reijneveld, 2007). Additionally, clustering coefficients quantify neural segregation or local specialization for certain tasks, while degree and betweenness centrality identify crucial, highly connected areas within different brain networks (Fallani et al., 2014).

Table 3. Descriptions of Graph Metrics (adapted from Tillem et al., 2021)

Metric	Definition
Maximum Degree	The number of connections between the largest hub (the node with the largest number of connections) and other nodes in the brain. High maximum degree indicates effective integration of information between nodes.
Mean Eccentricity	The longest path length from one node to any other node in the network.

Maximum Betweenness Centrality (BC)	The number of shortest paths passing through a specific node. Graphs with high maximum BC have more information traveling through a single, centrally located hub, allowing for efficient communication and effective information integration.
Diameter	The longest possible path in a network. High diameter indicates a spread-out network.
Average Shortest Path	The average shortest path length across all pairs of nodes. Low average shortest path indicates greater efficiency.
Global Clustering	The fraction of nodes that form triangular connections (i.e., the fraction of nodes whose neighbors are also interconnected with each other) at the network level. High global clustering indicates greater functional segregation, efficient local connectivity, and robustness to disruption.
Local (Mean) Clustering	The fraction of nodes that form triangular connections (i.e., the fraction of nodes whose neighbors are also interconnected with each other) at the node level, then averaged across all nodes in the network. High local clustering indicates greater functional segregation, efficient local connectivity, and robustness to disruption.
Global Efficiency	A metric related to the average inverse shortest path length across an entire graph. Graphs with high global efficiency allow information to travel through fewer connections to get from a node to any other node in the network, increasing the efficiency of neural communication.
Local (Mean) Efficiency	A metric related to the inverse shortest path length of a specific node within a smaller neighborhood. Nodes with high mean efficiency allow information to travel through fewer connections to get to other nodes in that neighborhood, increasing the efficiency of neural communication.

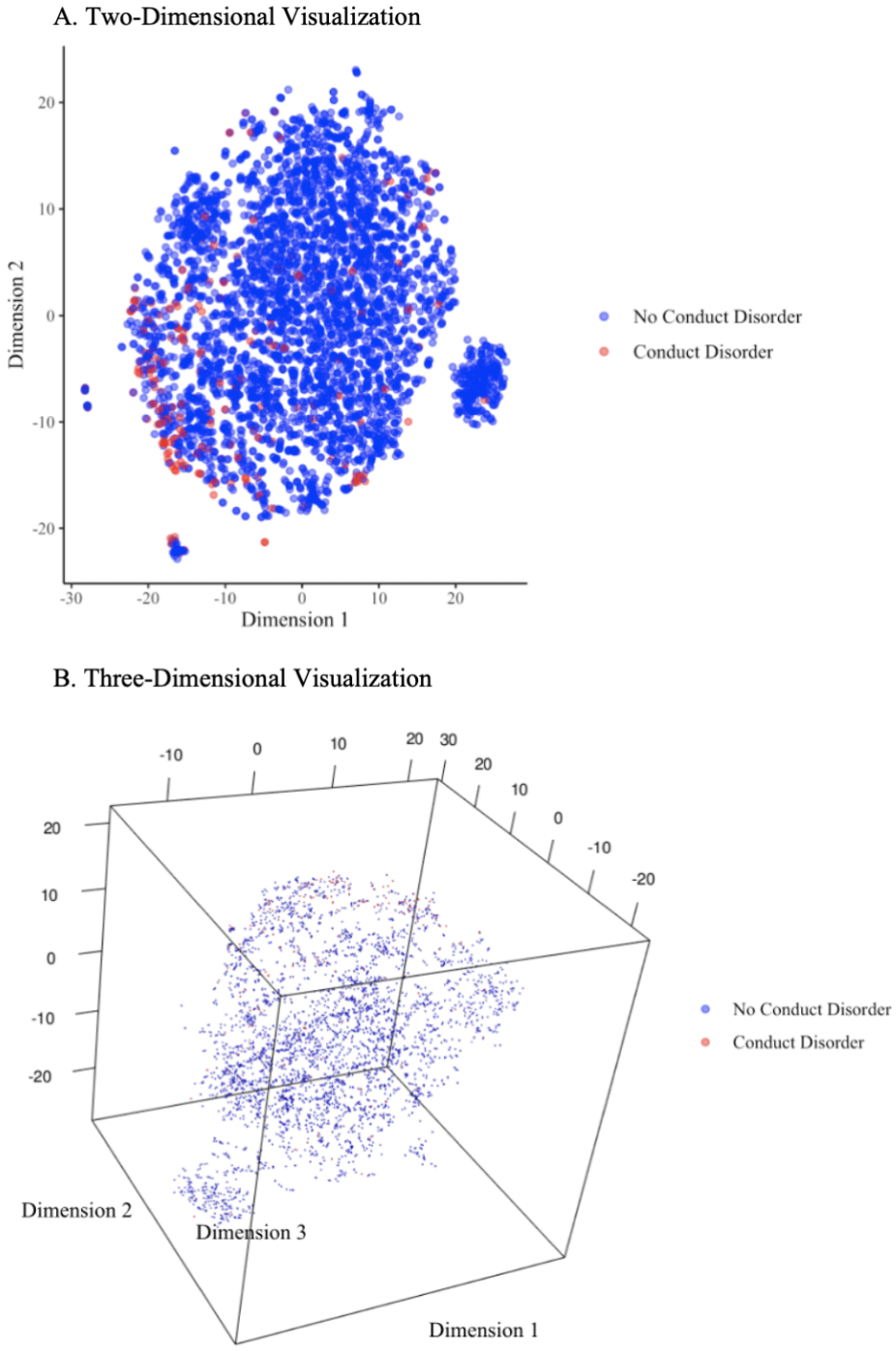
Subcortical Degree	The number of connections between subcortical structures and cortical networks. Nodes with high subcortical degree have greater connections and may act as more of a hub in the global flow of information.
Subcortical Betweenness Centrality (BC)	The number of shortest paths passing through a subcortical node in the global flow of information. Nodes with high subcortical BC have more information passing through them (i.e., are more central).
Subcortical Local Clustering	The number of triangular connections associated with subcortical structures. High subcortical local clustering indicates greater functional segregation, efficient local connectivity, and robustness to disruption in subcortical structures.
Subcortical Local Efficiency	The efficiency of communication between subcortical structures and other nodes in their immediate neighborhood.

Procedure

The classification of subjects as meeting or not meeting the criteria for CD requires that machine learning models identify highly non-linear patterns and develop meaningful decision rules. To illustrate the complexity of this task, we plotted a 36-dimensional representation of our sample for the 36 biopsychosocial risk factors used in later analysis and colored each subject by their CD diagnosis. Because it is impossible to visualize statistical relationships in higher than three dimensions, we reduced the data from 36 to two and three dimensions using the T-Distributed Stochastic Neighbor Embedding (tSNE) algorithm. tSNE is a non-linear dimensionality reduction technique that preserves the shape of high-dimensional data in lower dimensions for visualization purposes (Donaldson, 2016). The extensive overlap and lack of

clear separation between subjects with CD and healthy controls point to the classification task's immense difficulty (see Figure 1).

Figure 1. Visualization of the Classification Task



Note. Rows A and B visualize the classification task in two and three dimensions respectively. In reality, FNN classification occurs at the 36-dimensional level.

Using the nnet package in the programming language R, we built four FNNs to learn patterns within this high-dimensional data and classify the presence or absence of CD (Ripley & Venables, 2021). FNNs are machine learning models that develop decision rules and identify key features from labeled training data to classify unlabeled testing data (Zhang et al., 2018a). When designing FNNs, researchers must specify the network's architecture, activation function, and training algorithm (Jha et al., 2017).

Architecture

FNN architecture is defined by the number of layers within the network and the number of neurons or nodes within each layer. Information moves forward from the input neurons, through the hidden neurons, and to the output neurons. At each layer, the model learns complex patterns and relationships by conducting linear and non-linear transformations on the data.

For all four classifiers, we included participants from the initial sample of 4667 subjects if data on their (a) CD diagnosis, (b) age, sex, race, and research collection site, and (c) model-specific measures were available.

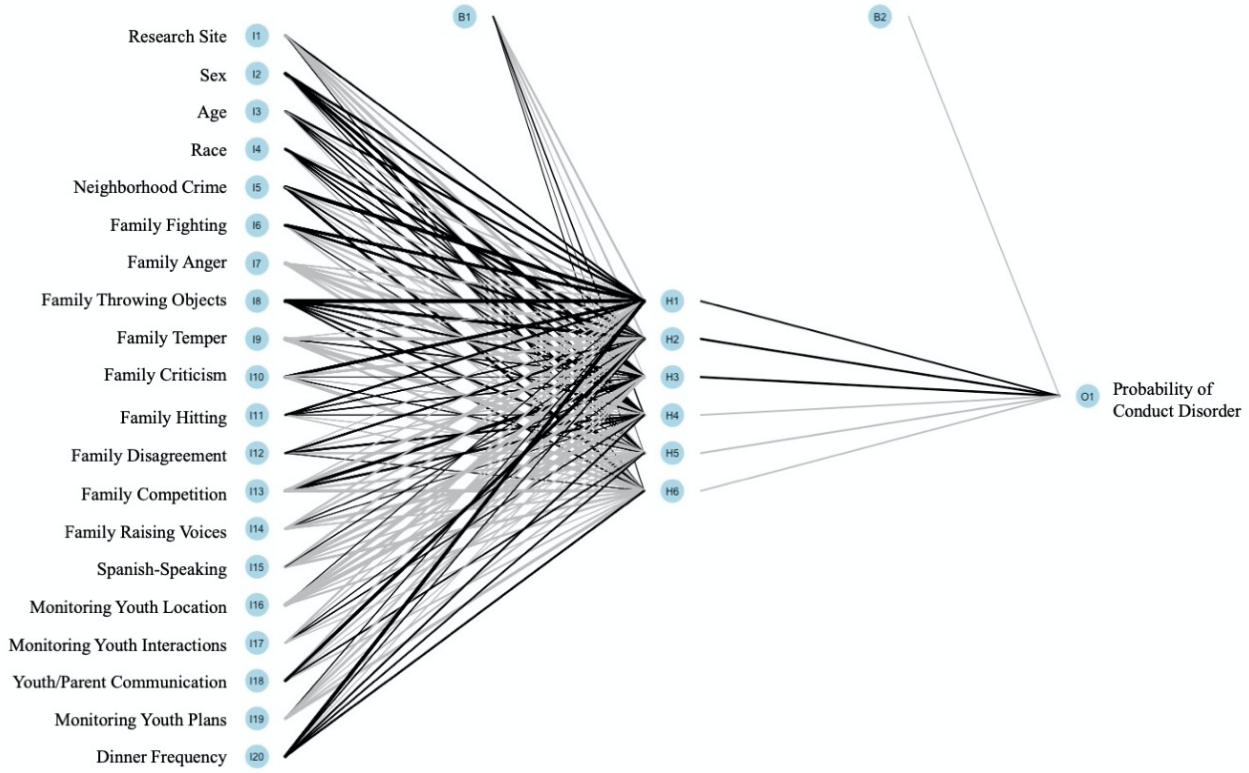
Model 1

To assess the predictive ability of neighborhood and family risk factors, model 1 was trained on NSC, FES, and PMQ responses. Demographic information on the research collection site, sex (dichotomously coded, male vs. female), race (dichotomously coded, white vs. non-white), and age were included to control for between-subject variability. Subjects with missing entries for NSC, FES, or PMQ were excluded from the original sample of 4667, resulting in a final sample of 4387 for model 1 training and testing.

The architecture of model 1 consists of 20 input neurons to represent 20 social risk factors (see Figure 2). As standard with binary classification problems, one output neuron calculated the probability that a subject meets the criteria for CD based on their input features.

The number of hidden neurons chosen by the researcher influences the fit and generalizability of the model. An insufficient number of hidden neurons may prevent the model from learning detailed patterns and representations, thus lowering accuracy. Conversely, an excessive number of hidden neurons may encourage overfitting to the training set and decrease test performance (Vujičić et al., 2016). The standard approach to address this issue is to select fewer hidden neurons than input neurons but more hidden neurons than output neurons (Huang & Babri, 1998; Sheela & Deepa, 2013; Vujičić et al., 2016). To determine this number, we ran simulations of possible architectures abiding by this rule and selected a FNN with six hidden neurons for model 1 because it optimized test accuracy.

Figure 2. Network Architecture for Model 1

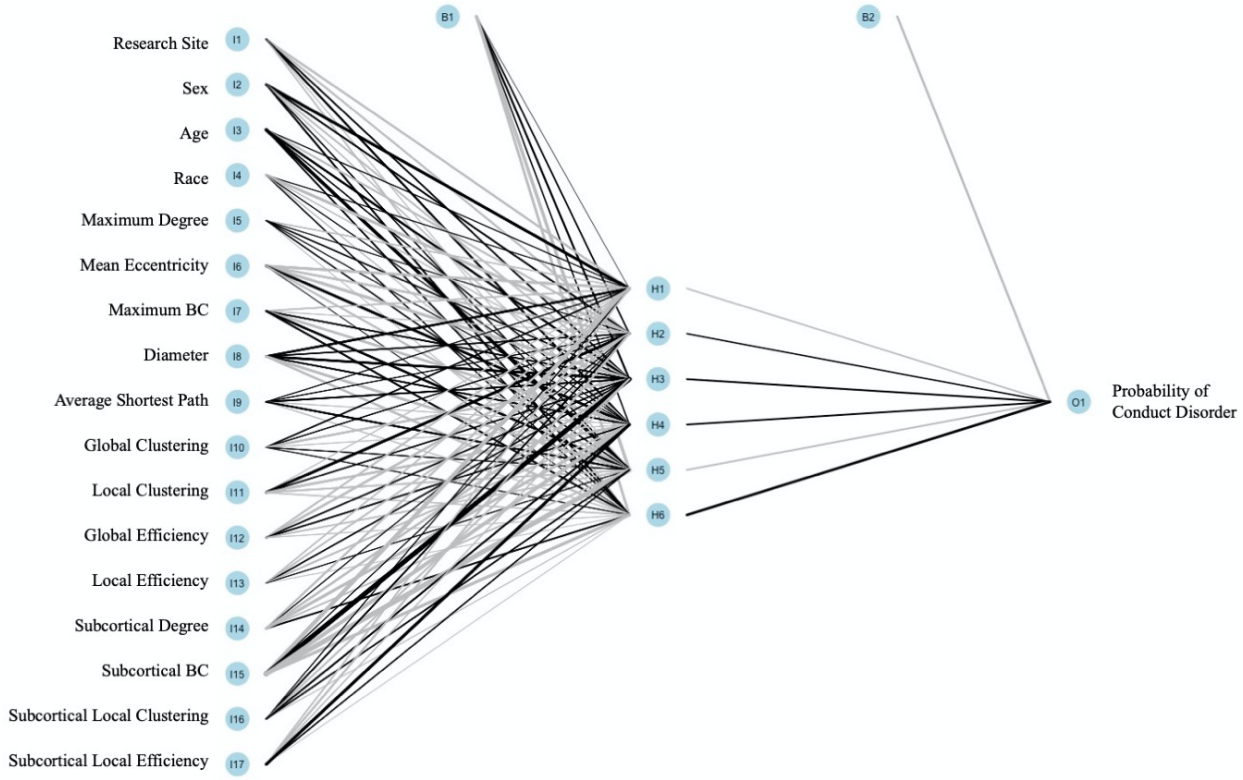


Note. Model 1 consists of 20 input neurons, six hidden neurons, and one output neuron. Line thickness is proportional to the magnitude of each weight and bias term. Black lines indicate positive parameters and grey lines indicate negative parameters.

Model 2

To assess the predictive ability of resting-state brain topography, model 2 was trained on global and local graph metrics. Like with model 1, demographic controls were included (see Figure 3). The model was trained on all 4667 subjects. Using the same methodology as above, we determined the optimal architecture of model 2 to contain 17 input neurons, six hidden neurons, and one output neuron.

Figure 3. Network Architecture for Model 2

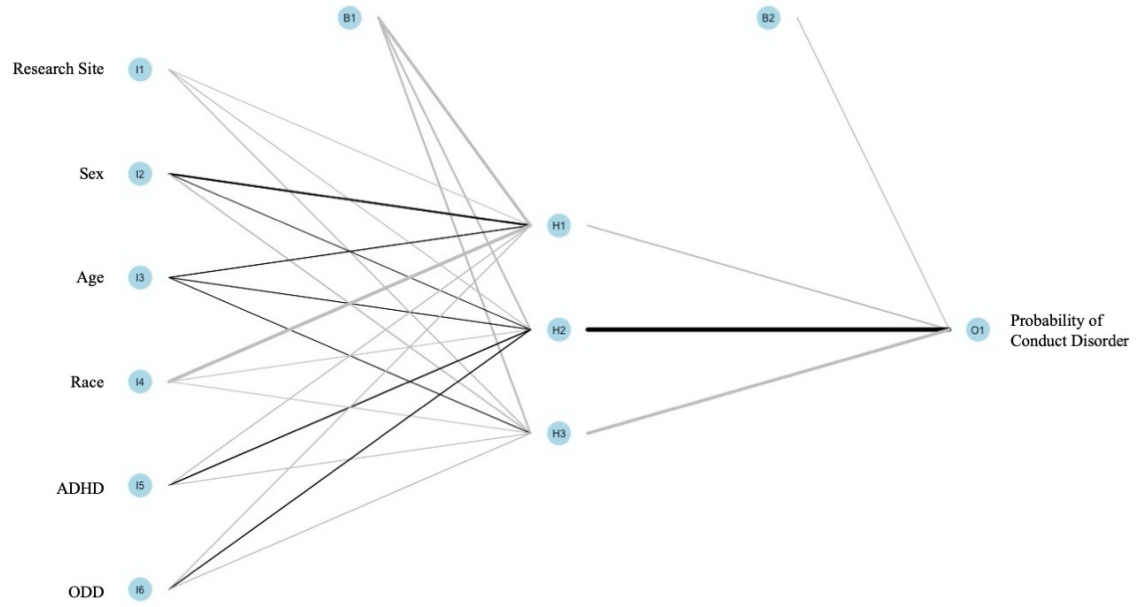


Note. Model 2 consists of 17 input neurons, six hidden neurons, and one output neuron. Line thickness is proportional to the magnitude of each weight and bias term. Black lines indicate positive parameters and grey lines indicate negative parameters.

Model 3

To assess the predictive ability of comorbid psychopathology, model 3 classifies CD status based on ADHD and ODD symptomatology measured by the K-SADS-PL. Demographic features were also included for model training (see Figure 4). We removed subjects with missing values, yielding a sample of 4400. After running multiple simulations to determine the optimal architecture, we specified six input neurons, three hidden neurons, and one output neuron for model 3.

Figure 4. Network Architecture of Model 3

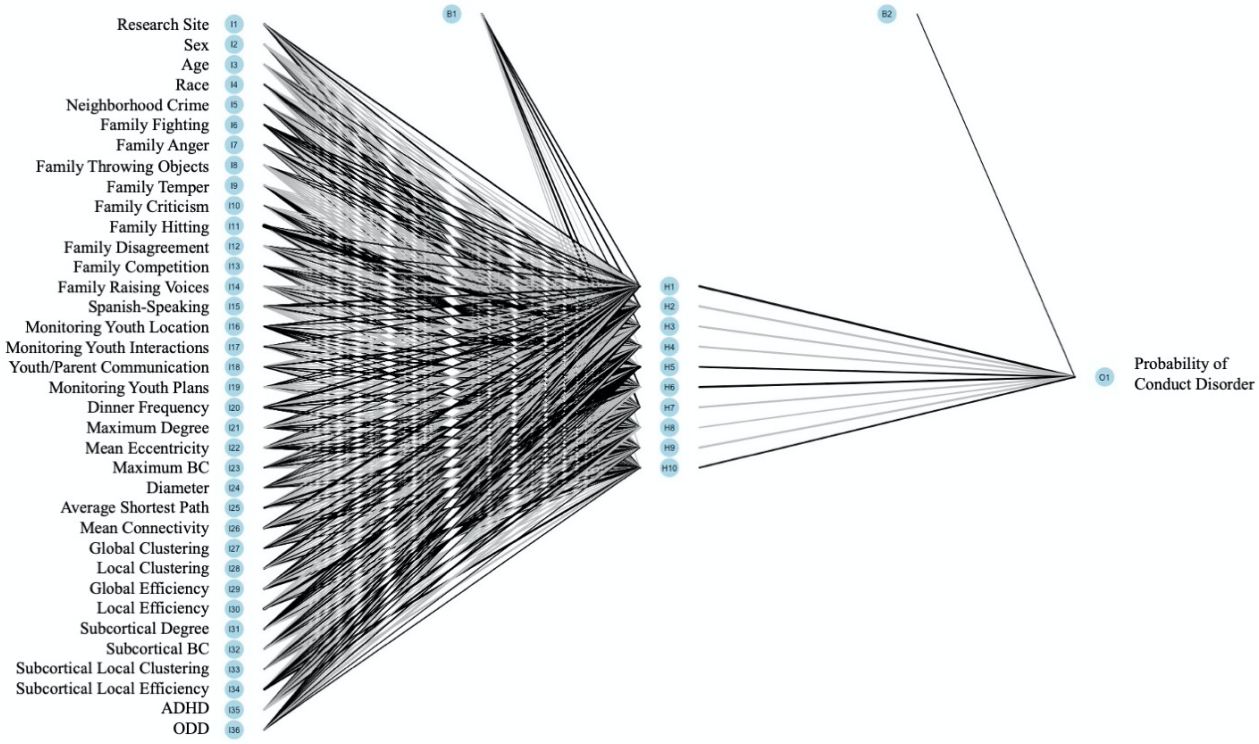


Note. Model 3 consists of six input neurons, three hidden neurons, and one output neuron. Line thickness is proportional to the magnitude of each weight and bias term. Black lines indicate positive parameters and grey lines indicate negative parameters.

Model 4

Finally, to test the collective influence of these biopsychosocial risk factors, model 4 predicts CD diagnosis based on neighborhood and family measures, global and local brain topography, psychiatric comorbidities, and demographic features combined (see Figure 5). We removed subjects with missing values, yielding a sample of 4387. This model consists of 36 input neurons, 10 hidden neurons, and one output neuron.

Figure 5. Network Architecture for Model 4



Note. Model 4 consists of 36 input neurons, 10 hidden neurons, and one output neuron. Line thickness is proportional to the magnitude of each weight and bias term. Black lines indicate positive parameters and grey lines indicate negative parameters.

Balancing Datasets

Classifiers encounter issues with model fitting when datasets are unbalanced (i.e., contain one group that is disproportionately smaller than the other). If left uncorrected, the model may ignore or discount the minority group during classification (Chawla et al., 2002; Qiao & Liu, 2009). Approximately 4.4% to 4.6% of subjects in our four model samples were diagnosed with CD. Thus, because typically developing subjects vastly outnumbered subjects with CD, the four datasets in the present study were unbalanced.

A standard approach to address unbalance is to undersample the larger class and oversample the smaller class until they are equal in size (Chawla et al., 2002). The ROSE package in R achieves oversampling through a smoothed bootstrapping method that draws

artificial samples from the feature space around the minority class. This algorithm preserves the distribution of data within each class and is preferred over oversampling processes that merely duplicate observations (Menardi & Torelli, 2014). In addition to oversampling the minority class, ROSE randomly undersamples the majority class without replacement (Lunardon et al., 2015). We used this method to oversample subjects with CD and undersample typically developing subjects until each dataset was balanced.

Gaussian Noise and Regularization

Overfitting occurs when models are closely trained to a limited set of data points and fail to flexibly adjust to new patterns or behaviors in unseen populations (Whelan & Garavan, 2014; Yarkoni & Westfall, 2017). For each classifier, 67% of the dataset was randomly selected for model training and 33% for model testing. We adopted the standard approach of adding Gaussian noise to the training set to prevent the model from memorizing patterns or trends in the data (Bishop, 1995). In addition, all models included a regularization decay term that penalized weight and bias parameters for overfitting to the training set (Ying, 2019).

Training

Each classifier was trained to predict whether an individual subject meets the diagnostic criteria for CD based on their exposure to risk factors. All predictors were standardized before training to improve model fitting. Risk factor inputs were then transformed by a series of weight and bias terms and summed to calculate hidden layer neurons. These sums then passed through a non-linear activation function that activated neurons above a certain threshold and deactivated neurons below that threshold.

The neurons of the hidden layer were calculated as follows (Jha et al., 2017):

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

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

We adopted the standard approach for binary classification models of defining f_H as the sigmoid activation function (Jurafsky & Martin, 2020). This activation function transforms inputs into values between 0 and 1 to amplify meaningful signals and suppress noise. The calculation of the sigmoid function f_H is displayed below (Jha et al., 2017):

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

The classifiers then conducted linear and non-linear transformations on the hidden neurons y_j to calculate values in the output layer. Output nodes were calculated as follows (Jha et al., 2017):

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

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

The output of the model is the probability that a given subject meets the criteria for CD. We adopted the standard classification threshold of .5 to convert probability outputs into binary classes of "CD" or "no CD." As a result, all outputs of probability ≥ .5 were classified as meeting for CD diagnosis, and all outputs of probability < .5 were classified as not meeting for CD diagnosis (Zou et al., 2016).

We trained weight and bias parameters with the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm, an iterative method for solving non-linear optimization problems (Ripley & Venables, 2021). The BFGS algorithm randomly assumes initial estimates for the optimal weight and bias terms and iteratively adjusts these values until they maximize classification accuracy and minimize a specified loss function (Dai, 2002). This approach processes random batches of data at a time, providing stability during model training and reliable convergence (Le et al., 2011).

As typical with binary classification models, the cross-entropy loss L_{CE} was used to optimize parameters (Mohri et al., 2018). L_{CE} measures the dissimilarity between the actual class y and the predicted class $\hat{y}$. By selecting weight and bias terms that minimize L_{CE}, the BFGS algorithm maximizes the model's classification accuracy. L_{CE} is calculated as follows (Jurafsky & Martin, 2020):

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

The BFGS algorithm conducts backpropagation to calculate the gradient of the loss function with respect to weight and bias parameters. Each gradient is computed one layer at a time, iterating backward from the last layer to avoid redundant calculations of intermediate terms. This method increases the efficiency of model training and updates parameters until L_{CE} converges at its minimum (Ciaburro & Venkateswaran, 2017).

Model Validation

Machine learning models are advantageous because they minimize overfitting and maximize prediction accuracy on unseen observations (Belloni et al., 2014; Dwyer et al., 2018). To achieve this, standard validation techniques such as the holdout method set aside around 33% of the data for testing the model's performance (Kim, 2009). Comparisons between in-sample

and out-of-sample predictions allow experimenters to check against overfitting and adjust their models accordingly; if the classifier demonstrates significantly higher accuracy on seen training data compared to unseen test data, there is strong evidence of overfitting. Thus, by maximizing test accuracy, experimenters can guard against overfitting and improve generalizability (Domingos, 2012; Yarkoni & Westfall, 2017). Consistent with standard model validation practices, all predictions and performance measures discussed in the Results section were calculated on the out-of-sample testing set.

Results

Confusion Matrix

The confusion matrix evaluates model efficacy by displaying correct and incorrect classifications (see Table 4). Confusion matrices for binary classification depict true positive (TP), true negative (TN), false positive (FP), and false negative (FN) predictions (Jha et al., 2017).

Table 4. Sample Confusion Matrix with Annotations

		Predicted Class	
		Control	Patient
True Class	Control	TN	FP
	Patient	FN	TP

We compared each classifier's predictions against the known diagnostic status of subjects to construct confusion matrices (see Tables 5-8). For the present study, TP and TN predictions refer to subjects with CD that are correctly classified as having CD and typically developing subjects that are correctly classified as not having CD respectively. FN predictions incorrectly

categorize subjects who have CD as not having CD, and FP predictions incorrectly categorize typically developing subjects as having CD.

Table 5. Confusion Matrix for Model 1

		Predicted Class	
		No CD	CD
True Class	No CD	535	226
	CD	269	418

Table 6. Confusion Matrix for Model 2

		Predicted Class	
		No CD	CD
True Class	No CD	463	344
	CD	312	422

Table 7. Confusion Matrix for Model 3

		Predicted Class	
		No CD	CD
True Class	No CD	642	121
	CD	256	433

Table 8. Confusion Matrix for Model 4

		Predicted Class	
		No CD	CD
True Class	No CD	607	154
	CD	124	563

Confusion matrices were used to calculate performance measures that normalize the number of TP and TN predictions by the model's sample size, allowing researchers to compare classification accuracy across different approaches.

Performance Measures

Accuracy is a performance measure that quantifies the overall precision of a classifier. This metric describes the proportion of TP and TN predictions among all evaluated cases (Jha et al., 2017):

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Sensitivity, or the true positive rate, is the proportion of subjects with CD that are correctly classified with the CD label, and specificity, or the true negative rate, is the proportion of typically developing subjects that are correctly classified as not having CD. These performance measures are calculated with the following formulas (Jha et al., 2017):

$$Sensitivity = true\ positive\ rate = \frac{TP}{TP + FN}$$

$$Specificity = true\ negative\ rate = \frac{TN}{TN + FP}$$

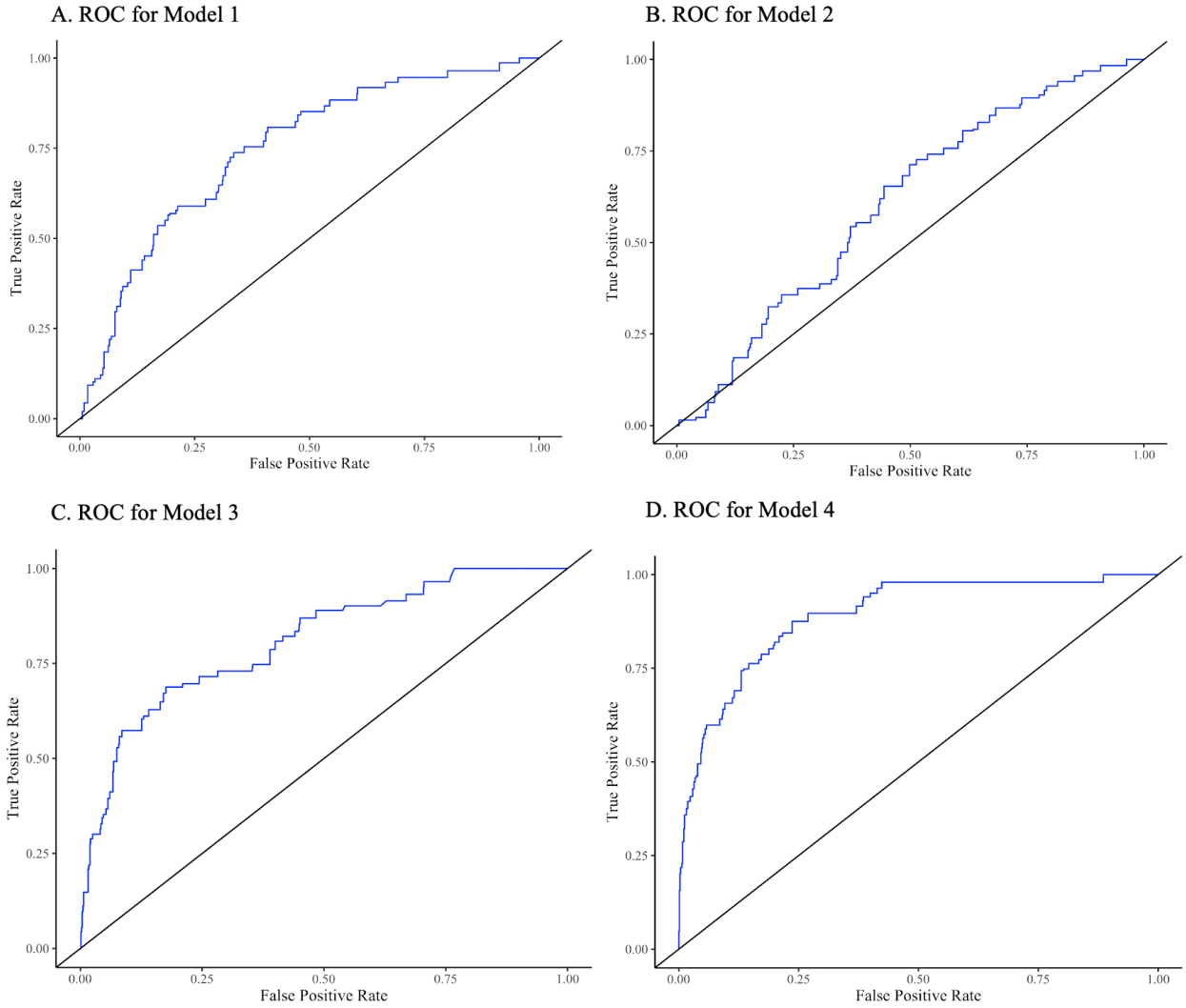
The strongest classifiers find an optimal balance between sensitivity and specificity. A model with high sensitivity may lack clinical relevance if it demonstrates low specificity and is biased towards TP and FP classifications. Conversely, a classifier may yield high specificity at the cost of low sensitivity. The receiver operating characteristic (ROC) curve allows researchers to assess and optimize this balance by plotting sensitivity against 1 − specificity, also termed the false positive rate, at all possible classification thresholds (Mandrekar, 2010).

$$1 - Specificity = false\ positive\ rate = 1 - \frac{TN}{TN + FP}$$

The coordinate (0,1) in the upper left corner represents a perfect classification with 100% sensitivity and specificity. The diagonal line depicts random predictions with 50% sensitivity and specificity that fail to meaningfully discriminate between classes. Thus, ROC curves that closely approach the left corner represent models that optimize true positive and true negative rates and maximize overall accuracy (Mandrekar, 2010).

The ROC curve for model 4 best approaches the (0,1) point of perfect prediction, suggesting that this classifier maximizes sensitivity and specificity. Model 3, model 1, and finally model 2 follow model 4 in terms of ROC curve performance. Across the four plots, the ROC curve for model 2 more closely mirrors the line of no discrimination, indicating that this classifier yields lower overall accuracy (see Figure 6).

Figure 6. ROC curves



Note. The ROC curves display the tradeoff between specificity and sensitivity at various classification thresholds. The ROC curve for model 4 in figure 6D optimally balances specificity and sensitivity by maximizing true positive rates and minimizing false positive rates. Model 4 is followed by model 3 in figure 6C, model 1 in figure 6A, and finally model 2 in figure 6B in order of optimal ROC curve performance.

The area under the ROC curve (AUC) is a metric that quantifies the balance between sensitivity and specificity and the overall diagnostic accuracy of the model (Mandrekar, 2010). The AUC is calculated via standard integration and adopts values between 0 and 1, where 0 indicates a perfectly inaccurate classifier and 1 reflects a perfectly accurate classifier. An AUC

of .5 represents a ROC curve that falls on the diagonal line and displays no discriminatory ability (Mandrekar, 2010).

Model 4 demonstrated the highest accuracy of 80.8%, followed by model 3 (74.04%), model 1 (65.81%), and model 2 (57.43%). In other words, model 4 classified the greatest proportion of subjects with CD as having CD and typically developing subjects as not having CD (see Table 9).

Model 4 also demonstrated the largest AUC of .8908, followed by model 3 (.8085), model 1 (.7471), and model 2 (.6077). These measures suggest that model 4 achieved an optimal balance between sensitivity and specificity and maximized overall prediction accuracy (see Table 9).

Finally, model 4 yielded the highest sensitivity of 81.95%, while model 3 demonstrated the highest specificity of 84.14%. These findings indicate that model 4 is the best classifier for detecting CD among subjects with CD, and model 3 is the best classifier for determining the absence of CD among typically developing subjects. Following model 3, model 1 produced a relatively high specificity rate of 70.3% but a lower sensitivity rate of 60.84%. In contrast, model 2 displayed the lowest sensitivity and specificity rates of 57.49% and 57.37% respectively (see Table 9).

Table 9. Performance Measures

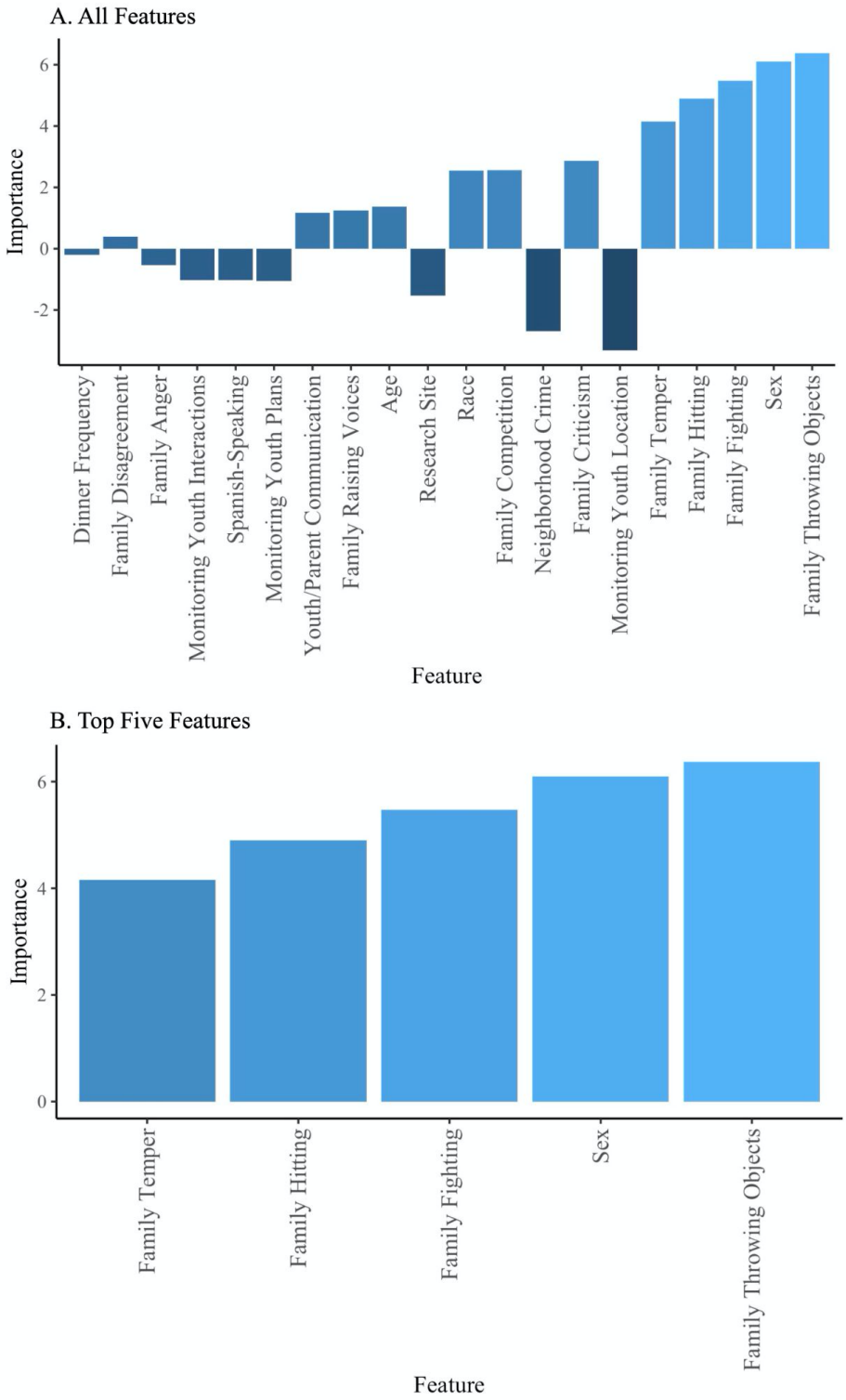
Model	Accuracy	Sensitivity	Specificity	AUC
Model 1	65.81%	60.84%	70.3%	0.7471
Model 2	57.43%	57.49%	57.37%	0.6077
Model 3	74.04%	84.14%	74.04%	0.8085
Model 4	80.8%	81.95%	79.76%	0.8908

Feature Importance

Feature importance describes the risk factors most influential in a classifier's decision-making process. Garson's algorithm in R identifies the relative importance of explanatory variables on the classification task by deconstructing and assessing weighted connections associated with each predictor (Garson, 1991). This technique displays both the sign and magnitude of the feature's influence on CD diagnosis.

In model 1, family members throwing items, fighting, and hitting each other are the environmental risk factors with the greatest feature importance. These predictors are positively associated with CD diagnosis. Male sex also predicts CD within this model (see Figure 7).

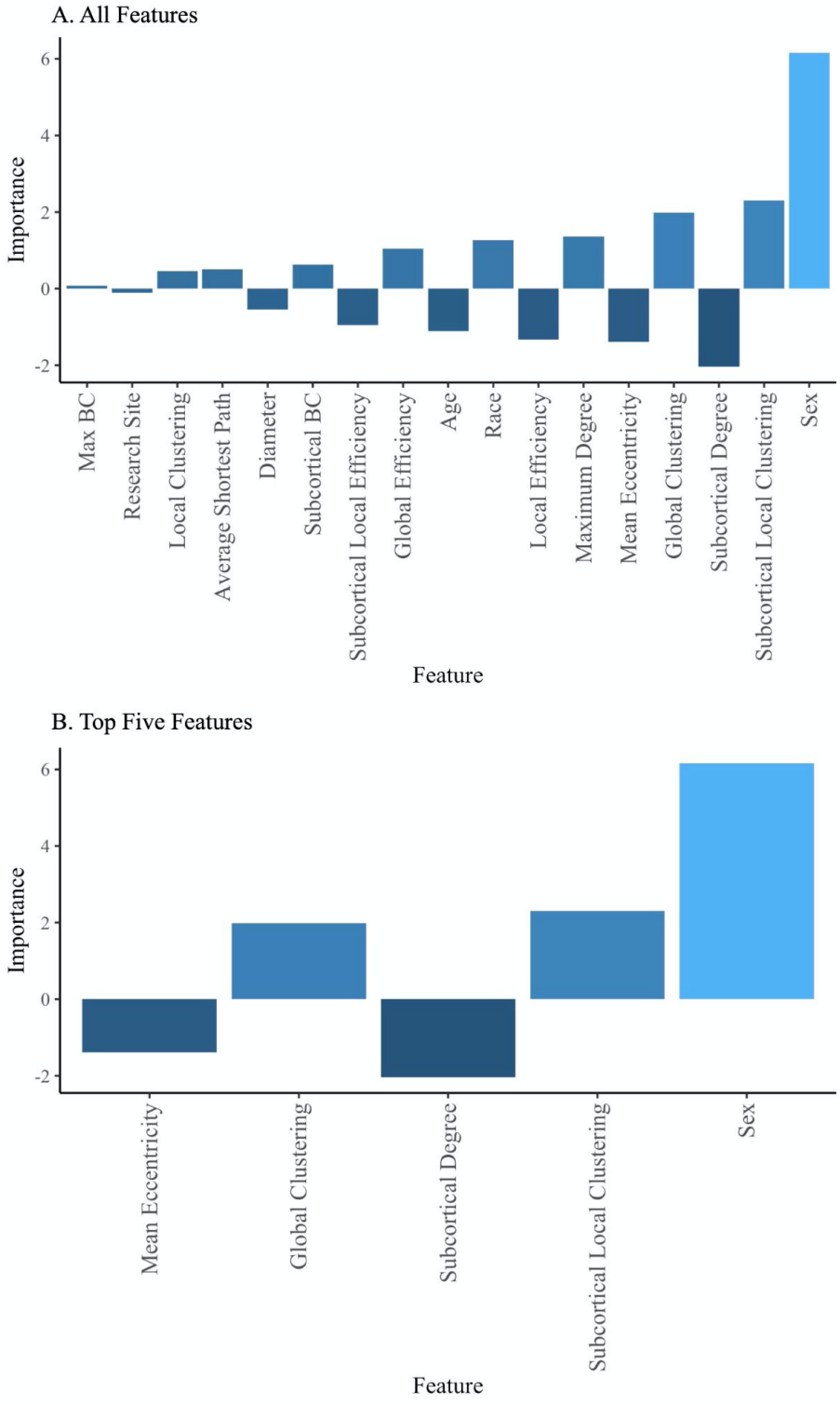
Figure 7. Feature Importance for Model 1



Note. Features are ordered from left to right by increasing absolute value of importance. Row A represents all features included in model 1. Row B highlights the five most important features in model 1.

The graph metrics with the greatest feature importance in model 2 are subcortical clustering, subcortical degree, and global clustering. Subcortical clustering and global clustering are positively associated with CD, suggesting that affected youth may experience greater functional segregation. In contrast, subcortical degree is negatively correlated with diagnosis, indicating that reduced cortical-subcortical connections characterize CD (see Figure 8).

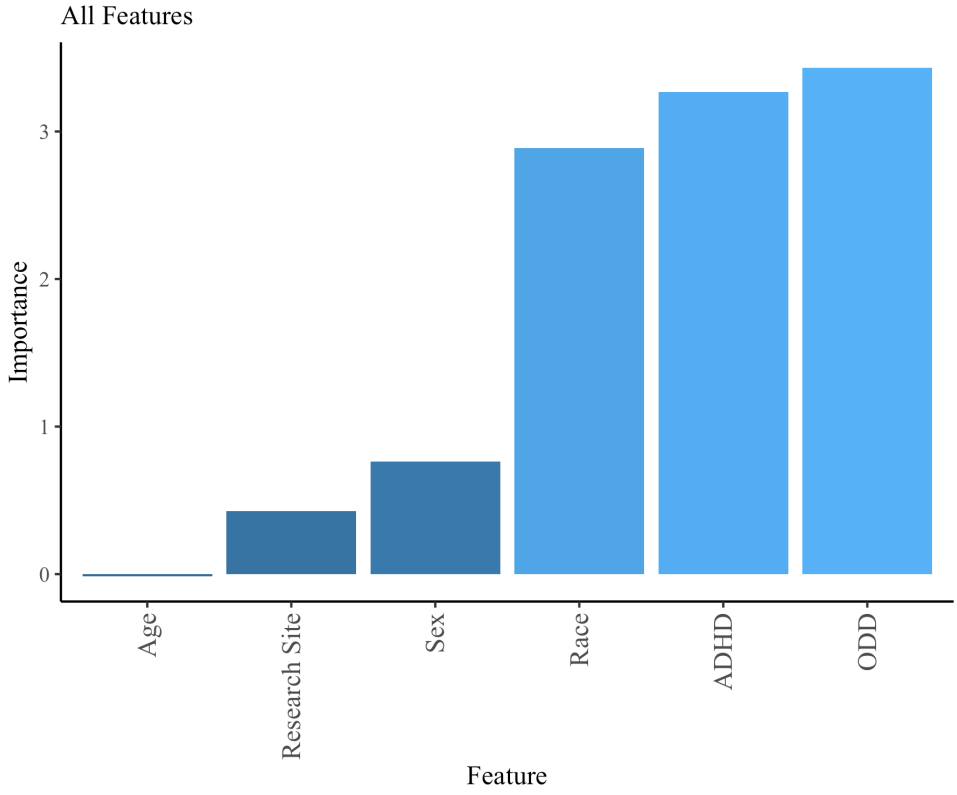
Figure 8. Feature Importance for Model 2



Note. Features are ordered from left to right by increasing absolute value of importance. Row A represents all features included in model 2. Row B highlights the five most important features in model 2.

In model 3, the severity of ADHD and ODD symptomatology strongly predicts CD among subjects (see Figure 9).

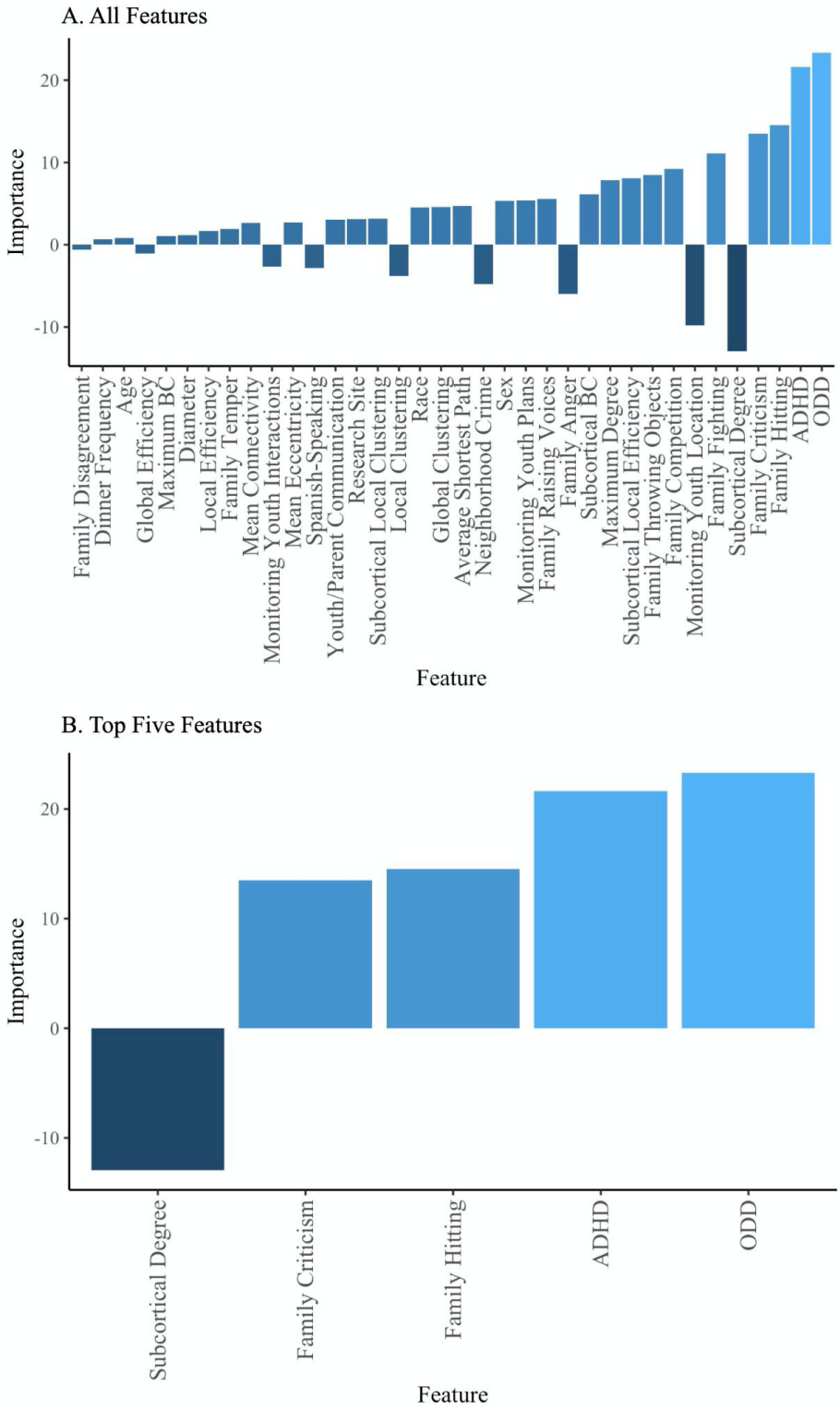
Figure 9. Feature Importance for Model 3



Note. Features are ordered from left to right by increasing absolute value of importance.

Finally, in model 4, the most important features for CD classification are comorbid ADHD and ODD symptoms, physical abuse, family criticism, and subcortical degree. Greater ADHD and ODD symptomatology, family violence, and family criticism all predict the disorder. In contrast, the number of cortical-subcortical connections is negatively correlated with CD (see Figure 10).

Figure 10. Feature Importance for Model 4



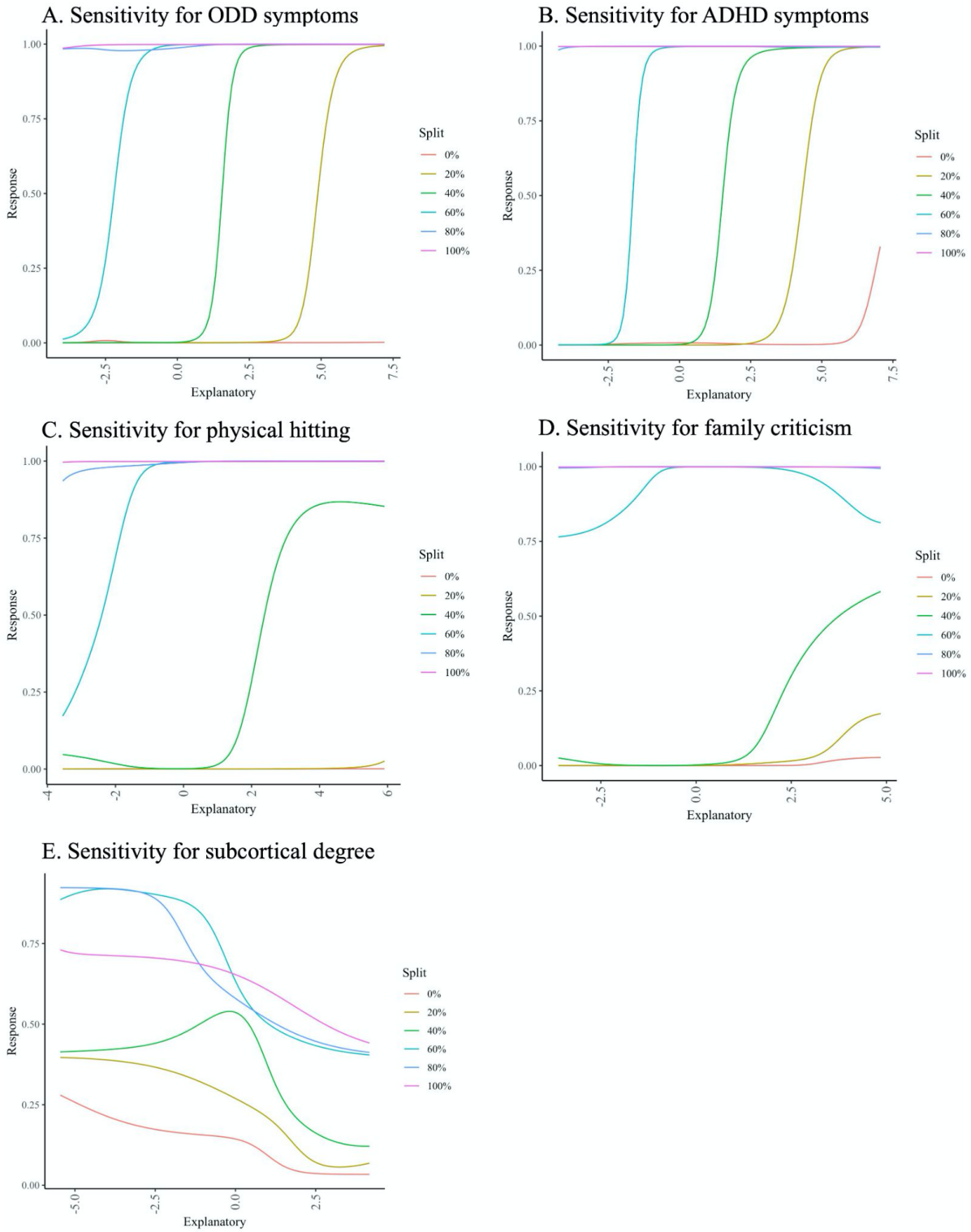
Note. Features are ordered from left to right by increasing absolute value of importance. Row A represents all features included in model 4. Row B highlights the five most important features in model 4.

Sensitivity Analysis

Sensitivity analysis allows researchers to visualize the linear and non-linear relationships between the classification task and predictors in the presence and absence of other predictors (Zhang et al., 2018a). We used Lek's profile method in R to conduct sensitivity analysis and illustrate how the probability of CD diagnosis and individual risk factors covary while holding other factors constant (specifically at their minimum, maximum, and 20th, 40th, 60th, and 80th quartiles).

We focused our sensitivity analysis on model 4 given its high accuracy rate and biopsychosocial assessment. When all other predictors are held at intermediate levels, higher values of ODD and ADHD symptom severity, family fighting, and family criticism and lower values of subcortical degree predict CD (see Figure 11). These relationships are consistent with the strongly positive feature importance of comorbid ADHD and ODD symptomatology, physical abuse, and criticism and negative feature importance of cortical-subcortical connectivity reported in model 4.

Figure 11. Sensitivity Plots



Note. The sensitivity plots in figures 11A-E display the relationship between model 4 predictions and the risk factors of ODD symptoms, ADHD symptoms, family fighting, family criticism, and subcortical degree respectively. The

explanatory variable denotes risk factor values, and the response variable denotes the probability that a subject is diagnosed with CD. The risk factors in figures 11A-D positively correlate with the likelihood of developing CD, while the risk factor in figure 11E negatively correlates with the likelihood of developing CD.

Discussion

The present study is the first to utilize different subsets of biopsychosocial risk factors to predict CD diagnoses among youth. Comparing accuracy rates across the various classifiers, we demonstrate that a multifactorial biopsychosocial model for CD yields greater predictive accuracy than domain-specific collections of risk factors. Furthermore, we identify comorbid ADHD and ODD symptoms, physical hitting, family criticism, and reduced cortical-subcortical connectivity as the "riskiest" factors for CD.

Reaffirming the Biopsychosocial Model

Our findings suggest that a biopsychosocial model is superior to neural, psychological, or environmental predictors alone in explaining and predicting CD onset. Model 4, built on all 36 biopsychosocial risk factors, reports the highest accuracy and AUC among the classifiers, supporting the cross-domain theory for CD development. Furthermore, the five most important features for CD classification in model 4 were the psychological risk factors of ADHD and ODD symptomatology, the environmental risk factors of physical hitting and family criticism, and the neural risk factor of subcortical degree. The involvement of all three domains in achieving accurate classification further reinforces the biopsychosocial model.

The high feature importance of ADHD and ODD symptomology in model 4 indicates that hyperactive and impulsive behaviors, hostility, and defiance against authority figures strongly predict CD diagnosis among youth (Turgay, 2005). In fact, the absolute value of feature importance for ODD symptoms (23.31) and ADHD symptoms (21.60) is almost twice that for subcortical degree (12.93). Considering that subcortical degree is the fifth most important factor

in model 4, this comparison suggests that the predictive power of ADHD and ODD is substantial. Furthermore, prior research identifies greater delinquency, aggression, and symptom severity as well as different treatment responses among subjects with comorbid CD + ADHD and CD + ODD compared to subjects with these disorders in isolation (Connor & Doerfler, 2008; Jensen et al., 2001). Given these associations, the present study demonstrates that ADHD and ODD comorbidity may indicate particularly high-risk subsets of youth in addition to strongly predicting CD. The high influence of ADHD and ODD also implies that these disorders share underlying risk factors or latent characteristics with CD (Biederman et al., 1991). Indeed, the association between ODD and CD is consistent with the common themes of rule-breaking and aggression that define their diagnoses (Loeber et al., 2000). Furthermore, the prediction of CD status based on ADHD symptomatology supports the hybrid theory for comorbid CD and ADHD development. Under the hybrid model, even if the risk profiles for CD and ADHD are somewhat distinct (e.g., CD is associated with psychosocial adversity and ADHD with developmental delays and neurocognitive deficits), the presence of risk factors for one disorder increases the likelihood of risk factors for the other (Schachar & Tannock, 1995). These findings propose that mutually influential etiological pathways may connect CD and ADHD.

The ranking of physical hitting and family criticism as the third and fourth most influential predictors in model 4 demonstrates the robust relationship between negative parenting, home dysfunction, and adverse child development. These results support extensive findings that physical abuse from caretakers greatly increases the risk for CD (Dodge & Pettit, 2003). Furthermore, the salience of the family criticism predictor is consistent with the theory that coercive parenting styles aggravate antisocial youth outcomes (McFadyen-Ketchum et al., 1996; Patterson, 1995). Under the coercion model, a parent harshly scolds or punishes their child

for misbehavior, the child reacts by screaming or rebelling, and the parent adopts a positive or neutral response to quell the child's aversive conduct. According to Patterson (1995), this coercive parenting style negatively reinforces disruptive and aggressive behavior and is a crucial life experience risk factor for later conduct problems. The present study suggests that family criticism may play a key role in promoting this coercive exchange between parent and child.

Although prior studies find strong correlations between community violence, social disorganization, and CD, the neighborhood crime feature yields only moderate predictive power (Loeber et al., 2000). A comparison across social risk factors reveals that physical hitting (14.51) and family criticism (13.47) demonstrate substantially higher absolute importance than neighborhood crime (4.76) when assessing for CD in model 4. Furthermore, as the 17th most predictive factor among 36 factors, neighborhood crime ranks below the median feature importance value. This midtier ranking initially implies that neighborhood crime only moderately influences the classification task and is less predictive of CD than physically and verbally harmful interactions between parent and child. However, given that the neighborhood crime predictor reflects parental assessments of both violent and non-violent crime, these results are not necessarily inconsistent with widespread findings that social disorder is a crucial risk factor for CD. First, parental reports of neighborhood crime may differ substantially from children's perceptions and may fail to measure social disorganization as experienced by affected youth. Second, since crime is only one aspect of the neighborhood environment, other unmeasured components such as community economic disadvantage, neighborhood violence, and involvement with neighborhood-based deviant peers may more directly define the social context and predict CD (Ingoldsby & Shaw, 2002). Consequently, further research is needed to

explore the relationship between CD and different aspects of social disorganization as reported by youth to re-assess the relative importance of neighborhood and family risk factors.

Finally, the negative association between subcortical degree and CD suggests that reduced connectivity between subcortical and cortical regions places youth at significant risk for diagnosis. This finding is consistent with well-established theories positing that reduced connectivity and communication between subcortical structures, such as the amygdala, hippocampus, and caudate, and the rest of the brain significantly hinder crucial neurocognitive processes (e.g., affective responding, memory, reward and punishment processing, decision-making, and violence inhibition; Baas et al., 2004; Blair, 1995; Eichenbaum, 2001; Knutson & Cooper, 2005; Tillem et al., 2019). Disrupted cortical-subcortical communication poses particularly insidious implications for youth with CD, who may express aggressive or impulsive behaviors even at the cost of others' physical and emotional wellbeing because affective processing and self-inhibitory mechanisms are disrupted. Interestingly, subcortical degree (12.93) demonstrates substantially higher absolute importance than global clustering (4.60). Although both graph metrics are relevant to the classification task, decreased cortical-subcortical connectivity is more predictive of CD than abnormalities in global clustering and other measures of functional segregation. Thus, reduced communication between subcortical structures, such as the amygdala, and cortical structures, such as the prefrontal cortex, appears especially relevant for the etiopathogenesis of CD and its related symptomatology.

The positive feature importance of global clustering for CD classification is initially surprising given that higher clustering coefficients are associated with optimal network organization and functioning (Achard et al., 2006). For example, the "small-world" network structure—a specific topology hypothesized to enhance robust and effective information

processing—is characterized by both high global clustering and high global efficiency (Liao et al., 2017). However, the greater feature importance of global clustering (4.60) relative to global efficiency (1.05) in the present student reinforces a recent hypothesis that heightened clustering without a corresponding increase in efficiency impairs general neurocognitive functioning for subjects with CD (Tillem et al., 2021). Higher clustering relative to efficiency may increase functional segregation to such a degree that it inhibits flexible communication and information integration between more distal nodes or networks (Tillem et al., 2021). The present study is consistent with this suggestion, and these collective findings demonstrate the relevance of further investigation into aberrant brain topology among youth with CD.

The risk factors for CD cut across various domains and levels of analysis. Youth with the greatest likelihood of developing CD face extensive physical and verbal conflict in the home and demonstrate hyperactivity, impulsivity, and hostility towards authority figures. At the neural level, these youth also experience abnormalities in cortical-subcortical connections that may interfere with emotional regulation and self-inhibitory mechanisms. Thus, current findings suggest that researchers should continue investigating the psychological well-being, family environment, and neural topography of at-risk youth to adequately understand, detect, and treat CD.

Domain-Specific Models

A comparison across domain-specific classifiers demonstrates that model 3—built on risk factors of psychological comorbidity alone—outperforms models 1 and 2. This result is consistent with the high feature importance of ADHD and ODD reported in model 4 and suggests that, in isolation, ADHD and ODD symptom severity is more predictive of CD than disruptions in brain topology or neighborhood and family-level risk factors.

Model 1 performed the next best among the classifiers built on domain-specific risk factors. Its accuracy rate of 65.81% demonstrates that neighborhood and family features are moderately predictive of CD on their own. This finding is slightly surprising given widespread suggestions that neighborhood crime, physical discipline, lack of parental supervision, and coercive parenting styles are the most consistent risk factors for later conduct problems (Dishion & Bullock, 2002; Dodge & Pettit, 2003; Lansford et al., 2002; Patterson, 1995). Nonetheless, the performance of model 1 indicates that environmental features are decent predictors of CD on their own and achieve even greater accuracy when assessed alongside psychological and biological risk factors. Furthermore, because the predictors in model 1 primarily measure child-parent interactions within the home, the inclusion of additional features such as exposure to community violence, antisocial peer influences, and the school environment may better represent social risk factors for CD and improve classification.

With a prediction accuracy of 57.43%, model 2 performed worse than not only the models in the present study but also the classifier developed by Zhang et al. (2020b). Yielding up to 94.44% accuracy, Zhang et al.'s (2020b) model separated subjects with CD from healthy controls based on 232 network property indices (e.g., clustering coefficient, average path length, average node degree, and measures of small-worldness) between highly correlated nodes within the brain. Because Zhang et al. (2020b) did not report the features most influential in their predictions, it is difficult to compare the neural profiles of CD classification observed in their study against our own. However, this discrepancy in performance accuracy indicates that indices between many node combinations, as performed in Zhang et al. (2020b), may offer more predictive power than indices between 13 cortical nodes and one subcortical node, as performed in the present study. This interpretation suggests that researchers may more precisely articulate

the neural abnormalities that define CD by extracting graph metrics across a greater number of nodes.

Limitations

The current study provides strong evidence that youth with CD face notable biological, psychological, and social risk exposures. However, these findings must be considered in light of several key limitations. First, feature importance quantifies the relative influence of predictors on CD classification based on other predictors in the model and cannot measure their absolute influence. Furthermore, the training of each classifier incorporates inherent variation as samples are randomly separated into training and testing sets and model parameters are randomly initialized before their optimization. As a result, performance measures and feature importance scores may vary slightly for each training of the same model. This quality of feature importance suggests that researchers should build numerous, diverse models on many combinations of risk factors to identify the features that consistently yield strong predictions.

Additionally, the present study does not evaluate interactions between risk factors during classification nor does it assess if risk factors mediate the relationship between neurocognitive difficulties, emotional dysfunction, antisocial conduct, or other behavioral outcomes that characterize CD. These two analyses may greatly benefit researchers by determining how unique combinations of key risk factors yield additive or diminishing effects on CD onset and differentially account for its heterogeneous behavioral expressions.

Conclusions and Future Directions

Machine learning classifiers are powerful tools for psychologists to determine salient risk factors, identify clinically relevant subgroups, and predict diagnoses and other psychiatric outcomes. These methods are especially relevant for CD research because they examine the

collective influence of high-dimensional biopsychosocial factors spanning the neighborhood, home, and child. Identifying comorbid ADHD and ODD symptoms, verbal and physical conflict, and reduced cortical-subcortical connectivity as highly predictive risk factors for CD, we demonstrate the potential for classifiers to test etiological theories. Future research should build upon the present study and also advance predictive modeling methods to determine if identified biopsychosocial factors meaningfully improve clinical outcomes for youth with CD. Specifically, the classification of high-risk youth before disorder onset may enhance prevention and early intervention strategies. Researchers can build classifiers that separate subjects into low, moderate, and high-risk groups or report a continuous risk score for CD development using biopsychosocial risk exposures at a baseline assessment. A comparison of these predictions against CD onset or symptom severity at later time points may illuminate the extent to which classifiers reliably can predict psychological outcomes before they occur. In the future, these longitudinal classification models may help establish machine-assisted risk detection processes that identify high-risk subjects prior to CD development.

Furthermore, predictive models may improve etiological theories and psychiatric treatment by assessing the childhood-onset and adolescent-onset subtypes of CD. According to Moffitt et al. (1996), youth diagnosed with CD during adolescence temporarily engage in aggressive behavior to meet survival needs, adhere to antisocial peer or gang influences, or explore typical rebellious tendencies. In contrast, youth who develop CD during childhood experience more severe, long-term genetic and neurocognitive deficits. This theory proposes two distinct profiles for youth with CD: (a) an adolescence-onset subgroup whose antisocial behavior only occurs during adolescence and (b) a childhood-onset or life-course-persistent subgroup whose antisocial behavior emerges early in childhood and persists into adulthood. Future work in

clinical psychology and machine learning can test Moffitt et al.'s (1996) hypothesis by assessing how classifiers built on various combinations of risk factors differentially predict outcomes in the childhood-onset or adolescent-onset groups. In doing so, researchers may develop nuanced profiles for these subgroups that promote targeted and effective interventions for youth with CD.

In addition, machine learning is especially suited for assessing massively multivariate data and engaging in image pattern recognition. As a result, classifiers can further explore the neural abnormalities introduced in the present study by analyzing highly information-rich neural data such as graph metrics between numerous node pairings or raw images of brain connectivity. These investigations may identify salient biological risk factors in addition to disrupted cortical-subcortical connectivity and clarify the unique neural profile for CD.

Given the immense burden of antisocial behavior on individuals, their families, and society at large, research on CD is crucial. Machine learning techniques that assess high-dimensional biopsychosocial variables, extract patterns from complex neural data, and robustly predict psychological outcomes are especially promising. Thus, the application of the biopsychosocial model and classification method to CD research may help alleviate burdens on the mental health and criminal justice system and improve outcomes for affected youth by refining etiological models and treatment.

References

Achard, S., Salvador, R., Whitcher, B., Suckling, J., & Bullmore, E. D. (2006). A resilient, low-frequency, small-world human brain functional network with highly connected association cortical hubs. *Journal of Neuroscience, 26*(1), 63-72. https://doi.org/10.1523/jneurosci.3874-05.2006

Aghajani, M., Colins, O. F., Klapwijk, E. T., Veer, I. M., Andershed, H., Popma, A., van der Wee, N. J., & Vermeiren, R. R. J. M. (2016). Dissociable relations between amygdala subregional networks and psychopathy trait dimensions in conduct-disordered juvenile offenders. *Human Brain Mapping, 37*(11), 4017-4033. https://doi.org/10.1002/hbm.23292

American Psychiatric Association. (2013). *Diagnostic and Statistical Manual of Mental Disorders (DSM-5®)* (5th ed.). https://doi.org/10.1176/appi.books.9780890425596

Arbabshirani, M. R., Plis, S., Sui, J., & Calhoun, V. D. (2017). Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *NeuroImage, 145*, 137-165. https://doi.org/10.1016/j.neuroimage.2016.02.079

Arnold, K., Davies, V., de Kamps, M., Tennant, P., Mbotwa, J., & Gilthorpe, M. (2020). Reflections on modern methods: Generalised linear models for prognosis and intervention–theory, practice, and implications for machine learning. *International Journal of Epidemiology, 46*(6), 2074–2082. https://doi.org/10.1093/ije/dyaa049

Azeredo, A., Moreira, D., & Barbosa, F. (2018). ADHD, CD, and ODD: Systematic review of genetic and environmental risk factors. *Research in Developmental Disabilities, 82*, 10-19. https://doi.org/10.1016/j.ridd.2017.12.010

Baas, D., Aleman, A., & Kahn, R. S. (2004). Lateralization of amygdala activation: A systematic review of functional neuroimaging studies. *Brain Research Reviews, 45*(2), 96-103. https://doi.org/10.1016/j.brainresrev.2004.02.004

Baker, R. H., Clanton, R. L., Rogers, J. C., & De Brito, S. A. (2015). Neuroimaging findings in disruptive behavior disorders. *CNS Spectrums, 20*(4), 369-381. https://doi.org/10.1017/s1092852914000789

Belloni, A., Chernozhukov, V., & Hansen, C. (2014). High-dimensional methods and inference on structural and treatment effects. *Journal of Economic Perspectives, 28*(2), 29-50. doi: 10.1257/jep.28.2.29

Biederman, J., Newcorn, J., & Sprich, S. (1991). Comorbidity of attention deficit hyperactivity disorder with conduct, depressive, anxiety, and other disorders. *The American Journal of Psychiatry, 148*(5), 564–577. https://doi.org/10.1176/ajp.148.5.564

Bishop, C. M. (1995). Training with noise is equivalent to Tikhonov regularization. *Neural computation, 7*(1)*,* 108-116. https://doi.org/10.1162/neco.1995.7.1.108

Blair, R. J. R. (1995). A cognitive developmental approach to morality: Investigating the psychopath. *Cognition, 57*, 1–29. doi:10.1016/0010-0277(95)00676-p

Blair, R. J. R. (2013). The neurobiology of psychopathic traits in youths. *Nature Reviews Neuroscience, 14*(11), 786-799. https://doi.org/10.1038/nrn3577

Blair, R. J. R., Leibenluft, E., & Pine, D. S. (2014). Conduct disorder and callous–unemotional traits in youth. *New England Journal of Medicine, 371*(23), 2207-2216. doi: 10.1056/NEJMra1315612

Bolón-Canedo, V., Sánchez-Maroño, N., & Alonso-Betanzos, A. (2015). Recent advances and emerging challenges of feature selection in the context of big data. *Knowledge-Based Systems, 86,* 33-45. https://doi.org/10.1016/j.knosys.2015.05.014

Borsboom, D. (2017). A network theory of mental disorders. *World Psychiatry, 16*(1), 5-13. https://doi.org/10.1002/wps.20375

Borsboom, D., & Cramer, A. O. (2013). Network analysis: An integrative approach to the structure of psychopathology. *Annual Review of Clinical Psychology, 9*, 91-121. https://doi.org/10.1146/annurev-clinpsy-050212-185608

Cappadocia, M. C., Desrocher, M., Pepler, D., & Schroeder, J. H. (2009). Contextualizing the neurobiology of conduct disorder in an emotion dysregulation framework. *Clinical Psychology Review, 29*(6), 506-518. https://doi.org/10.1016/j.cpr.2009.06.001

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E., Teslovichm T., Dellarco, D. V., Garavan, H., Orr, C. A., Wager, T. D., Banich, M. T., Speer, N. K., Sutherland, M. T., Riedel, M. C., Dick, A. S., Bjork, J. M., Thomas, K. 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

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

Ciaburro, G. & Venkateswaran, B. (2017). *Neural Networks with R*. Packt.

Cohn, M. D., Pape, L. E., Schmaal, L., van den Brink, W., van Wingen, G., Vermeiren, R. R., Doreleijers, T. A. H., Veltman, D. J., & Popma, A. (2015). Differential relations between

juvenile psychopathic traits and resting state network connectivity. *Human Brain Mapping, 36*(6), 2396-2405. https://doi.org/10.1002/hbm.22779

Colman, I., Murray, J., Abbott, R. A., Maughan, B., Kuh, D., Croudace, T. J., & Jones, P. B. (2009). Outcomes of conduct problems in adolescence: 40 year follow-up of national cohort. *BMJ, 338*(7688), 208-211. https://doi.org/10.1136/bmj.a2981

Connor, D. F., & Doerfler, L. A. (2008). ADHD with comorbid oppositional defiant disorder or conduct disorder: Discrete or nondistinct disruptive behavior disorders? *Journal of Attention Disorders, 12*(2), 126-134. https://doi.org/10.1177/1087054707308486

Dadds, M. R., Fraser, J., Frost, A., & Hawes, D. J. (2005). Disentangling the underlying dimensions of psychopathy and conduct problems in childhood: A community study. *Journal of Consulting and Clinical Psychology, 73*(3), 400–410. https://doi.org/10.1037/0022-006X.73.3.400

Dai, Y. H. (2002). Convergence properties of the BFGS algorithm. *SIAM Journal on Optimization, 13*(3), 693-701. https://doi.org/10.1137/S1052623401383455

Davatzikos, C., Fan, Y., Wu, X., Shen, D., & Resnick, S. M. (2008). Detection of prodromal Alzheimer's disease via pattern classification of magnetic resonance imaging. *Neurobiology of Aging, 29*(4), 514-523. https://doi.org/10.1016/j.neurobiolaging.2006.11.010

De Brito, S. A., Mechelli, A., Wilke, M., Laurens, K. R., Jones, A. P., Barker, G. J., Hodgins, S., & Viding, E. (2009). Size matters: Increased grey matter in boys with conduct problems and callous–unemotional traits. *Brain, 132*(4), 843-852. doi:10.1017/S0954579420000279

Dishion, T. J., & Bullock, B. M. (2002). Parenting and adolescent problem behavior: An ecological analysis of the nurturance hypothesis. In J. G. Borkowski, S. L. Ramey, & M. Bristol-Power (Eds.), *Monographs in parenting. Parenting and the child's world: Influences on academic, intellectual, and social-emotional development* (p. 231–249). Lawrence Erlbaum Associates Publishers.

Dodge, K. A., & Pettit, G. S. (2003). A biopsychosocial model of the development of chronic conduct problems in adolescence. *Developmental Psychology, 39*(2), 349–371. https://doi.org/10.1037/0012-1649.39.2.349

Dodge, K. A., Price, J. M., Bachorowski, J. A., & Newman, J. P. (1990). Hostile attributional biases in severely aggressive adolescents. *Journal of Abnormal Psychology, 99*(4), 385–392. https://doi.org/10.1037/0021-843X.99.4.385

Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM, 55*(10), 78-87. doi:10.1145/2347736.2347755

Donaldson, J. (2016, July 15). *Package 'tsne'*. The Comprehensive R Archive Network. https://cran.r-project.org/web/packages/tsne/tsne.pdf

Dwyer, D. B., Falkai, P., & Koutsouleris, N. (2018). Machine learning approaches for clinical psychology and psychiatry. *Annual Review of Clinical Psychology, 14*, 91-118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

Eichenbaum, H. (2001). The hippocampus and declarative memory: Cognitive mechanisms and neural codes. *Behavioural Brain Research, 127*, 199–207. doi:10.1016/s0166-4328(01)00365-5

Fairchild, G., Hawes, D. J., Frick, P. J., Copeland, W. E., Odgers, C. L., Franke, B., Freitag, C. M., & De Brito, S. A. (2019). Conduct disorder. *Nature Reviews Disease Primers, 5*(1), 1-25. https://doi.org/10.1038/s41572-019-0095-y

Fallani, F. D. V., Richiardi, J., Chavez, M., & Achard, S. (2014). Graph analysis of functional brain networks: Practical issues in translational neuroscience. *Philosophical Transactions of the Royal Society B, 369*(1653), Article 20130521. https://doi.org/10.1098/rstb.2013.0521

Finger, E. C., Marsh, A. A., Blair, K. S., Reid, M. E., Sims, C., Ng, P., Pine, D. S., & Blair, R. J. R. (2011). Disrupted reinforcement signaling in the orbitofrontal cortex and caudate in youths with conduct disorder or oppositional defiant disorder and a high level of psychopathic traits. *American Journal of Psychiatry, 168*(2), 152-162. https://doi.org/10.1176/appi.ajp.2010.10010129

Frick, P. J. (2006). Developmental pathways to conduct disorder. *Child and Adolescent Psychiatric Clinics, 15*(2), 311-331. https://doi.org/10.1016/j.chc.2005.11.003

Frick, P. J., & Dickens, C. (2006). Current perspectives on conduct disorder. *Current Psychiatry Reports, 8*(1), 59-72. doi: 10.1007/s11920-006-0082-3

Frick, P. J., Lahey, B. B., Loeber, R., Stouthamer-Loeber, M., Christ, M. A. G., & Hanson, K. (1992). Familial risk factors to oppositional defiant disorder and conduct disorder: Parental psychopathology and maternal parenting. *Journal of Consulting and Clinical Psychology, 60*(1), 49–55. https://doi.org/10.1037/0022-006X.60.1.49

Fu, C. H., Mourao-Miranda, J., Costafreda, S. G., Khanna, A., Marquand, A. F., Williams, S. C., & Brammer, M. J. (2008). Pattern classification of sad facial processing: Toward the

development of neurobiological markers in depression. *Biological Psychiatry, 63*(7), 656-662. https://doi.org/10.1016/j.biopsych.2007.08.020

Garavan, H., Bartsch, H., Conway, K., Decastro, A., Goldstein, R. Z., Heeringa, S., Jernigan, T., Potter, A., Thompson, W., & Zahs, D. (2018). Recruiting the ABCD sample: Design considerations and procedures. *Developmental Cognitive Neuroscience, 32*, 16-22. https://doi.org/10.1016/j.dcn.2018.04.004

Garson, G. D. (1991). Interpreting neural-network connection weights. *AI Expert, 6*(4), 46-51. https://link.gale.com/apps/doc/A10581496/ITOF?u=29002&sid=ITOF&xid=eb0cece7

Hagler, D. J., Hatton, S., Cornejo, M. D., Makowski, C., Fair, D. A., Dick, A., Sutherland, M. T., Casey, B. J., Barch, D. M, Harms, M. P., Watts, R., Bjork, J. M., Garavan, H. P., Hilmer, L., Pung, C. J., Sicat, C. S., Kuperman, J., Hauke, B., Xue, F., ... Dale, A. M. (2019). Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. *NeuroImage, 202*, Article 116091. https://doi.org/10.1016/j.neuroimage.2019.116091

Huang, G. B., & Babri, H. A. (1998). Upper bounds on the number of hidden neurons in feedforward networks with arbitrary bounded nonlinear activation functions. *IEEE Transactions on Neural Networks, 9*(1), 224-229. doi: 10.1109/72.655045

Iacono, W. G., Heath, A. C., Hewitt, J. K., Neale, M. C., Banich, M. T., Luciana, M. M., Madden, P. A., Barch, D. M., & Bjork, J. M. (2018). The utility of twins in developmental cognitive neuroscience research: How twins strengthen the ABCD research design. *Developmental Cognitive Neuroscience, 32,* 30-42. https://doi.org/10.1016/j.dcn.2017.09.001

Ingoldsby, E. M., & Shaw, D. S. (2002). Neighborhood contextual factors and early-starting antisocial pathways. *Clinical Child and Family Psychology Review, 5*, 21-55. https://doi.org/10.1023/A:1014521724498

Ioannidis, J. P. (2016). Why Most Clinical Research Is Not Useful. *Plos Medicine, 13*(6), Article e1002049. https://doi.org/10.1371/journal.pmed.1002049

Jensen, P. S., Hinshaw, S. P., Kraemer, H. C., Lenora, N., Newcorn, J. H., Abikoff, H. B., March, J. S., Arnold, L. E., Cantwell, D. P., Conners, C. K., Elliott, G. R., Greenhill, L. L., Hechtman, L., Hoza, B., Pelham, W. E., Severe, J. B., Swanson, J. M., Wells, K. C., Wigal, T., & Vitiello, B. (2001). ADHD comorbidity findings from the MTA study: Comparing comorbid subgroups. *Journal of the American Academy of Child & Adolescent Psychiatry, 40*(2), 147-158. https://doi.org/10.1097/00004583-200102000-00009

Jha, D., Kim, J. I., & Kwon, G. R. (2017). Diagnosis of Alzheimer's disease using dual-tree complex wavelet transform, PCA, and feed-forward neural network. *Journal of Healthcare Engineering*, *2017*, Article 9060124. https://doi.org/10.1155/2017/9060124

Jiang, Y., Liu, W., Ming, Q., Gao, Y., Ma, R., Zhang, X., Situ. W., Wang, X., Yao, S., & Huang, B. (2016). Disrupted topological patterns of large-scale network in conduct disorder. *Scientific Reports, 6,* 1-10. https://doi.org/10.1038/srep37053

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science, 349*(6245), 255-260. doi: 10.1126/science.aaa8415

Jurafsky, D. & Martin, J. H. (2020). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall.

Kambeitz, J., Cabral, C., Sacchet, M. D., Gotlib, I. H., Zahn, R., Serpa, M. H., Walter, M., Falkai, P., & Koutsouleris, N. (2017). Detecting neuroimaging biomarkers for depression: A meta-analysis of multivariate pattern recognition studies. *Biological Psychiatry, 82*(5), 330-338. https://doi.org/10.1016/j.biopsych.2016.10.028

Kambeitz, J., Kambeitz-Ilankovic, L., Leucht, S., Wood, S., Davatzikos, C., Malchow, B., Falkai, P., & Koutsouleris, N. (2015). Detecting neuroimaging biomarkers for schizophrenia: A meta-analysis of multivariate pattern recognition studies. *Neuropsychopharmacology, 40*(7), 1742-1751. doi: 10.1038/npp.2015.22

Kaufman, J., Birmaher, B., Axelson, D., Perepletchikova, F., Brent, D., & Ryan, N. (2013). *Kiddie Schedule for Affective Disorders and Schizophrenia Present and Lifetime Version 2013 (K-SADS-PL)*. Western Psychiatric Institute and Yale University.

Kim, J. H. (2009). Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. *Computational Statistics & Data Analysis, 53*(11), 3735-3745. https://doi.org/10.1016/j.csda.2009.04.009

Kim, J., Calhoun, V. D., Shim, E., & Lee, J. H. (2016). Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: Evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. *NeuroImage, 124*, 127-146. https://doi.org/10.1016/j.neuroimage.2015.05.018

Knutson, B., & Cooper, J. C. (2005). Functional magnetic resonance imaging of reward prediction. *Current Opinion in Neurology, 18*(4), 411-417. https://doi.org/10.1097/01.wco.0000173463.24758.f6

Koutsouleris, N., Meisenzahl, E. M., Davatzikos, C., Bottlender, R., Frodl, T., Scheuerecker, J., Schmitt, G., Zetzsche, T., Decker, P., Reiser, M., Möller, H., & Gaser, C. (2009). Use of neuroanatomical pattern classification to identify subjects in at-risk mental states of psychosis and predict disease transition. *Archives of General Psychiatry, 66*(7), 700-712. doi:10.1001/archgenpsychiatry.2009.62

Kuang, D., & He, L. (2014). Classification on ADHD with deep learning. In J. E. Guerrero (Ed.), *Proceedings of the 2014 International Conference on Cloud Computing and Big Data* (pp. 27-32). The Institute of Electrical and Electronics Engineers

Lansford, J. E., Dodge, K. A., Pettit, G. S., Bates, J. E., Crozier, J., & Kaplow, J. (2002). A 12-year prospective study of the long-term effects of early child physical maltreatment on psychological, behavioral, and academic problems in adolescence. *Archives of Pediatrics and Adolescent Medicine, 156*(8), 824-830. doi:10.1001/archpedi.156.8.824

Le, Q. V., Ngiam, J., Coates, A., Lahiri, A., Prochnow, B., & Ng, A. Y. (2011). On optimization methods for deep learning. In L. Getoor & T. Scheffer (Eds.), *Proceedings of the 28 th International Conference on Machine Learning* (pp. 265–272). Omnipress

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature, 521*, 436-444. https://doi.org/10.1038/nature14539

Liao, X., Vasilakos, A. V., & He, Y. (2017). Small-world human brain networks: Perspectives and challenges. *Neuroscience & Biobehavioral Reviews, 77*, 286-300. https://doi.org/10.1016/j.neubiorev.2017.03.018

Loeber, R., Burke, J. D., Lahey, B. B., Winters, A., & Zera, M. (2000). Oppositional defiant and conduct disorder: A review of the past 10 years, part I. *Journal of the American Academy*

of Child & Adolescent Psychiatry, 39(12), 1468-1484. https://doi.org/10.1097/00004583-200012000-00007

Lu, F. M., Zhou, J. S., Zhang, J., Wang, X. P., & Yuan, Z. (2017). Disrupted small-world brain network topology in pure conduct disorder. *Oncotarget, 8*(39), 65506–65524. https://doi.org/10.18632/oncotarget.19098

Lu, F. M., Zhou, J. S., Zhang, J., Xiang, Y. T., Zhang, J., Liu, Q., Wang, X., & Yuan, Z. (2015). Functional connectivity estimated from resting-state fMRI reveals selective alterations in male adolescents with pure conduct disorder. *PLoS One, 10*(12), Article e0145668. https://doi.org/10.1371/journal.pone.0145668

Lunardon, N., Menardi, G., & Torelli, N. (2015, February 19). *Package 'ROSE'*. The Comprehensive R Archive Network. https://cran.r-project.org/web/packages/ROSE/ROSE.pdf

Lytton, H. (1990). Child and parent effects in boys' conduct disorder: A reinterpretation. *Developmental Psychology, 26*(5), 683-697. doi: 10.1037/0012-1649.26.5.683

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology, 5*(9), 1315-1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

McFadyen-Ketchum, S. A., Bates, J. E., Dodge, K. A., & Pettit, G. S. (1996). Patterns of change in early childhood aggressive-disruptive behavior: Gender differences in predictions from early coercive and affectionate mother-child interactions. *Child Development, 67*(5), 2417-2433. https://doi.org/10.1111/j.1467-8624.1996.tb01865.x

Menardi, G., & Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data Mining and Knowledge Discovery, 28*(1), 92-122. doi: 10.1007/s10618-012-0295-5

Merikangas, K. R., He, J. P., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., Benjet, C., Georgiades, K., & Swendsen, J. (2010). Lifetime prevalence of mental disorders in US adolescents: Results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *Journal of the American Academy of Child & Adolescent Psychiatry, 49*(10), 980-989. https://doi.org/10.1016/j.jaac.2010.05.017

Moffitt, T. E. (1993). The neuropsychology of conduct disorder. *Development and Psychopathology, 5*(1-2), 135-151. https://doi.org/10.1017/S0954579400004302

Moffitt, T. E., Caspi, A., Dickson, N., Silva, P., & Stanton, W. (1996). Childhood-onset versus adolescent-onset antisocial conduct problems in males: Natural history from ages 3 to 18 years. *Development and Psychopathology, 8*(2), 399-424. https://doi.org/10.1017/S0954579400007161

Mohri, M., Rostamizadeh, A., & Talwalkar, A. (2018). *Foundations of Machine Learning* (2nd ed.). MIT Press.

Molenaar, P. C., & Campbell, C. G. (2009). The new person-specific paradigm in psychology. *Current Directions in Psychological Science, 18*(2), 112-117. https://doi.org/10.1111/j.1467-8721.2009.01619.x

Murray, J., & Farrington, D. P. (2010). Risk factors for conduct disorder and delinquency: Key findings from longitudinal studies. *The Canadian Journal of Psychiatry, 55*(10), 633-642. doi: 10.1177/070674371005501003

Ogilvie, J. M., Stewart, A. L., Chan, R. C., & Shum, D. H. (2011). Neuropsychological measures of executive function and antisocial behavior: A meta-analysis. *Criminology, 49*(4), 1063-1107. https://doi.org/10.1111/j.1745-9125.2011.00252.x

Olden, J. D., & Jackson, D. A. (2002). Illuminating the "black box": A randomization approach for understanding variable contributions in artificial neural networks. *Ecological Modelling, 154*(1-2), 135-150. https://doi.org/10.1016/S0304-3800(02)00064-9

Patterson, G. R. (1995). Coercion as a basis for early age of onset for arrest. In J. McCord (Ed.), *Coercion and punishment in long-term perspectives* (p. 81–105). Cambridge University Press. https://doi.org/10.1017/CBO9780511527906.005

Pauli, R., Tino, P., Rogers, J.C., Baker, R., Clanton, R., Birch, P., Brown, A., Daniel, G., Ferreira, L., Grisley, L., Kohls, G., Baumann, S., Bernhard, A., Martinelli, A., Ackermann, A., Lazaratou, H., Tsiakoulia, F., Bali, P., Oldenhof, H., ... , De Brito, S.A. (2020). Positive and negative parenting in conduct disorder with high versus low levels of callous-unemotional traits. *Development and Psychopathology*, 1-12. https://doi.org/10.1017/S0954579420000279

Qiao, X., & Liu, Y. (2009). Adaptive weighted learning for unbalanced multicategory classification. *Biometrics, 65*(1), 159-168. https://doi.org/10.1111/j.1541-0420.2008.01017.x

Redlich, R., Almeida, J. R., Grotegerd, D., Opel, N., Kugel, H., Heindel, W., Arolt, V., Phillips, M. L., & Dannlowski, U. (2014). Brain morphometric biomarkers distinguishing unipolar and bipolar depression: A voxel-based morphometry–pattern classification approach. *JAMA Psychiatry, 71*(11), 1222-1230. doi:10.1001/jamapsychiatry.2014.1100

Ripley, B. & Venables, W. (2021, January 24). *Package 'nnet'*. The Comprehensive R Archive Network. https://cran.r-project.org/web/packages/nnet/nnet.pdf

Rubinov, M., & Sporns, O. (2010). Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage, 52*(3), 1059-1069. https://doi.org/10.1016/j.neuroimage.2009.10.003

Schachar, R., & Tannock, R. (1995). Test of four hypotheses for the comorbidity of attention-deficit hyperactivity disorder and conduct disorder. *Journal of the American Academy of Child & Adolescent Psychiatry, 34*(5), 639-648. https://doi.org/10.1097/00004583-199505000-00016

Schnack, H. G., & Kahn, R. S. (2016). Detecting neuroimaging biomarkers for psychiatric disorders: Sample size matters. *Frontiers in Psychiatry, 7*, Article 50. https://doi.org/10.3389/fpsyt.2016.00050

Shafiei, S. B., Lone, Z., Elsayed, A. S., Hussein, A. A., & Guru, K. A. (2020). Identifying mental health status using deep neural network trained by visual metrics. *Translational Psychiatry, 10*(1), 1-8. https://doi.org/10.1038/s41398-020-01117-5.

Sheela, K. G., & Deepa, S. N. (2013). Review on methods to fix number of hidden neurons in neural networks. *Mathematical Problems in Engineering*, *2013*, Article 425740. https://doi.org/10.1155/2013/425740

Stam, C. J., & Reijneveld, J. C. (2007). Graph theoretical analysis of complex networks in the brain. *Nonlinear Biomedical Physics, 1*, Article 3. https://doi.org/10.1186/1753-4631-1-3

Svozil, D., Kvasnicka, V., & Pospichal, J. (1997). Introduction to multi-layer feed-forward neural networks. *Chemometrics and Intelligent Laboratory Systems, 39*(1), 43-62. https://doi.org/10.1016/S0169-7439(97)00061-0

Tillem, S., Conley, M. I., & Baskin-Sommers, A. (2021). Conduct disorder symptomatology is associated with an altered functional connectome in a large national youth sample. *Development and Psychopathology*, 1-12. https://doi.org/10.1017/S0954579421000237

Tillem, S., Harenski, K., Harenski, C., Decety, J., Kosson, D., Kiehl, K. A., & Baskin-Sommers, A. (2019). Psychopathy is associated with shifts in the organization of neural networks in a large incarcerated male sample. *NeuroImage: Clinical, 24*, Article 102083. https://doi.org/10.1016/j.nicl.2019.102083

Tor, H. T., Ooi, C. P., Lim-Ashworth, N. S., Wei, J. K. E., Jahmunah, V., Oh, S. L., Acharya, U. R., Shuen, D., & Fung, D. S. S. (2021). Automated detection of conduct disorder and attention deficit hyperactivity disorder using decomposition and nonlinear techniques with EEG signals. *Computer Methods and Programs in Biomedicine, 200*, Article 105941. https://doi.org/10.1016/j.cmpb.2021.105941

Turgay, A. (2005). Treatment of comorbidity in conduct disorder with attention-deficit hyperactivity disorder (ADHD). *Essential Psychopharmacology, 6*(5), 277-290. https://europepmc.org/article/med/16222912

Vermeiren, R. (2003). Psychopathology and delinquency in adolescents: A descriptive and developmental perspective. *Clinical Psychology Review, 23*(2), 277-318. https://doi.org/10.1016/S0272-7358(02)00227-1

Vujicic, T., Matijevic, T., Ljucovic, J., Balota, A., & Sevarac, Z. (2016). Comparative analysis of methods for determining number of hidden neurons in artificial neural network. In T. Hunjak, V. Kirinić, & M. Konecki (Eds.), *Central European Conference on Information and Intelligent Systems* (pp. 219-223). Faculty of Organization and Informatics Varazdin.

Whelan, R., & Garavan, H. (2014). When optimism hurts: Inflated predictions in psychiatric neuroimaging. *Biological Psychiatry, 75*(9), 746-748. https://doi.org/10.1016/j.biopsych.2013.05.014

Yarkoni, T., & Westfall, J. (2017). Choosing prediction over explanation in psychology: Lessons from machine learning. *Perspectives on Psychological Science, 12*(6), 1100-1122. https://doi.org/10.1177/1745691617693393

Ying, X. (2019). An overview of overfitting and its solutions. *Journal of Physics: Conference Series*. *1168*(2), 1–6. doi:10.1088/1742-6596/1168/2/022022.

Zhang, J., Cao, W., Wang, M., Wang, N., Yao, S., & Huang, B. (2019). Multivoxel pattern analysis of structural MRI in children and adolescents with conduct disorder. *Brain Imaging and Behavior, 13*(5), 1273-1280. https://doi.org/10.1007/s11682-018-9953-6

Zhang, J., Li, X., Li, Y., Wang, M., Huang, B., Yao, S., & Shen, L. (2020a). Three dimensional convolutional neural network-based classification of conduct disorder with structural MRI. *Brain Imaging and Behavior, 14*, 2333-2340. https://doi.org/10.1007/s11682-019-00186-5

Zhang, J., Liu, W., Zhang, J., Wu, Q., Gao, Y., Jiang, Y., Gao, J., Yao, S., & Huang, B. (2018b). Distinguishing adolescents with conduct disorder from typically developing youngsters based on pattern classification of brain structural MRI. *Frontiers in Human Neuroscience*, *12*, Article 152. https://doi.org/10.3389/fnhum.2018.00152

Zhang, J., Liu, Y., Luo, R., Du, Z., Lu, F., Yuan, Z., Zhou, J., & Li, S. (2020b). Classification of pure conduct disorder from healthy controls based on indices of brain networks during resting state. *Medical & Biological Engineering & Computing, 58*(9), 2071-2082. https://doi.org/10.1007/s11517-020-02215-8

Zhang, Z., Beck, M. W., Winkler, D. A., Huang, B., Sibanda, W., & Goyal, H. (2018a). Opening the black box of neural networks: Methods for interpreting neural network models in clinical applications. *Annals of Translational Medicine, 6*(11). doi: 10.21037/atm.2018.05.32

Zhou, J., Yao, N., Fairchild, G., Cao, X., Zhang, Y., Xiang, Y. T., Zhang, L., & Wang, X. (2016). Disrupted default mode network connectivity in male adolescents with conduct disorder. *Brain Imaging and Behavior, 10*(4), 995-1003. https://doi.org/10.1007/s11682-015-9465-6

Zou, Q., Xie, S., Lin, Z., Wu, M., & Ju, Y. (2016). Finding the best classification threshold in imbalanced classification. *Big Data Research, 5*, 2-8. https://doi.org/10.1016/j.bdr.2015.12.001