Decoding Preadolescent Anxiety with Machine Learning Insights from Neuroimaging Data

Olimpia Carrioli

Dr. Deanna Greene[1], Dr. Alessandra Camassa[2], Dr. Jason G. Fleischer[1], Dr. Terrence Sejnowski[2]

[1]*UC San Diego Department of Cognitive Science*

[2]*Salk Institute for Biological Studies*

Introduction

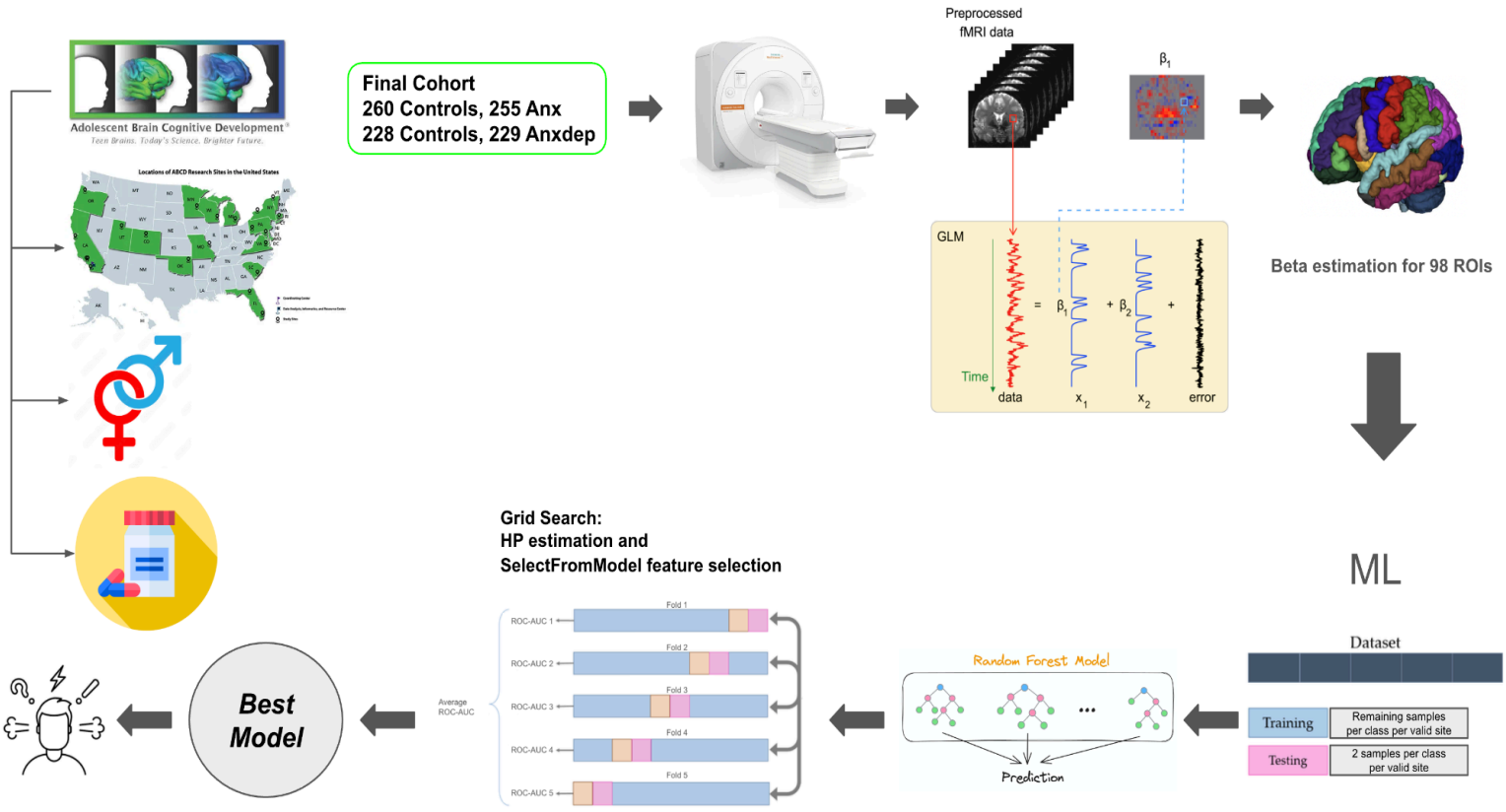

Figure 1. Project Pipeline
The ABCD dataset includes substantial imbalance across collection site, sex, and diagnostic groups (top left). To address this, we constructed a subset of the baseline cohort stratified by site and scanner type (therefore by diagnostic group) and with overall sex balance. This yielded final cohorts of 260 Controls matched to 255 Anxiety cases, and 228 Controls matched to 229 Anx/Dep cases. fMRI data from 22 sites were used to estimate beta activation coefficients for task contrasts via a general linear modeling (GLM), resulting in activation coefficients (β) features across 98 ROIs, segmented according to the Desikan-Killani atlas. For Anxiety classification, medication usage (ADHD medication, antidepressants, antipsychotics, anticonvulsants, and mood stabilizers) was included as binary features. For Anx/Dep classification, medication, sex, and site were included as features. Site-stratified data splitting held out two samples per class per valid site (≥2 samples per class per site) for testing, with remaining subjects used for training. A Random Forest classifier was trained using grid search with 5-fold cross-validation for hyperparameter tuning and feature selection via SelectFromModel. The best model is fit on the full training set and selected features. Final model performance was evaluated on the held-out test set for both Anxiety and Anx/Dep prediction tasks.

Anxiety disorders represent a growing public health concern among children and adolescents. In the United States, over one in three adolescents (31.9%) meets diagnostic criteria for an anxiety disorder by age 18 (Merikangas et al., 2010), and multiple studies indicate that prevalence is rising. Data from the National Survey of Children's Health, for example, show a 61% increase in anxiety prevalence between 2016 and 2023 (Health Resources and Services

Administration, Maternal and Child Health Bureau, 2024). These conditions often emerge in preadolescence, disrupting social relationships and academic performance, and they predispose those affected to the risk of long-term psychiatric impairment. Early diagnosis is critical, as timely intervention is known to shorten recovery time and improve outcomes later in life (McGorry & Mei, 2018).

Despite their acknowledged importance, psychiatric diagnoses are limited by both clinical and methodological challenges. Instruments such as the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-V) and the Achenbach System of Empirically Based Assessment (ASEBA) are the field standard, but their utility in neurobiological research is limited. DSM diagnoses are categorical and threshold-based, often collapsing diverse symptom profiles under the same diagnosis (Kotov et al., 2017). Additionally, the DSM and ASEBA scales rely on parent-report data, which may underrepresent internalizing symptoms that are less observable (De Los Reyes & Kazdin, 2005). Perhaps most critically, these diagnostic systems lack a direct mapping onto the neural mechanisms they aim to characterize, limiting their value for developing brain-based diagnostic tools (Pickersgill, 2014). Given these limitations, there is a growing need for approaches that can directly link behavioral symptoms to brain function.

In response, researchers have turned to neuroimaging and supervised machine learning (ML) as tools for understanding and predicting psychiatric conditions. Advances in functional Magnetic Resonance Imaging (fMRI) and computational modeling make it possible to examine distributed patterns of brain activity and their relationship to symptoms and diagnosis. In adults, ML models have shown moderate success—often achieving classification accuracies in the 70–85% range—in identifying anxiety, depression, and related disorders vs. control populations from structural and functional neuroimaging data (Portugal et al., 2019; Mousavian et al., 2021).

Similar work in adolescents has begun to explore whether these methods can be adapted for early diagnosis (Sawalha et al., 2021; Chavanne et al., 2022). However, realizing this potential requires overcoming key methodological challenges.

One of the most pressing issues in the ML application to psychiatry is the problem of overfitting. High-dimensional brain data coupled with small samples can lead models to learn noise or dataset-specific artifacts instead of meaningful brain patterns. This problem is particularly relevant in psychiatric applications, where effect sizes are often small and labels may be uncertain or imprecise. Common pitfalls that lead to overfitting include performing feature selection on the full dataset or failing to use held-out test sets. In these cases, model performance may appear strong but does not reflect true predictive power. As several reviews have emphasized (Pulini et al., 2019; Anderson et al., 2019; Kassraian-Fard et al., 2016), even subtle forms of information leakage can significantly inflate reported accuracy. These risks are further increased when confounds such as site variability or scanner-specific artifacts correlate with diagnosis, creating spurious correlation which nonetheless can be learned. With these challenges in mind, we developed a machine learning pipeline designed to mitigate overfitting and aid interpretability in classifying individuals with anxiety from healthy controls using brain activation data.

Project Overview

The aim of this project is to determine whether patterns of brain activation during a working memory task could distinguish preadolescents with anxiety from healthy controls. As shown in Figure 1, this was done via a machine learning pipeline applied to task-based fMRI data from the Adolescent Brain Cognitive Development (ABCD) Study. This dataset presents

considerable heterogeneity in collection sites, scanner types, sex distribution, and medication intake, sources of potential confounds that were addressed to mitigate the risk of overfitting.

The ABCD dataset encompasses several neuroimaging data types; here we focus on beta activation coefficients extracted from the Emotional N-Back (EN-Back) task—a paradigm that probes working memory and emotional processing. We used these neural features, along with indicators of medication usage, to train a Random Forest classifier to predict anxiety status. Throughout the pipeline, particular attention was given to minimize methodological and demographic bias, including stratification by collection site and diagnostic status, as well as statistical validation using permutation testing. This framework was designed not only to evaluate classification performance, which can lead to new and improved diagnostic systems, but also to explore whether specific brain regions serve as early neural markers of anxiety. A parallel analysis was also conducted for Anxious-Depressed classification using the ASEBA scale, with results discussed in the appendix.

Methods

Participants

This study used baseline data from the Adolescent Brain Cognitive Development Study, a longitudinal research effort that maps biological, cognitive, and behavioral development from childhood to adulthood. Data collection occurs across 22 institutions in the United States, and the baseline cohort encompasses approximately 11,000 participants aged 9 and 10 years (Casey et al., 2018). Mental health is assessed annually through diagnostic interviews, with psychiatric conditions identified using both the Achenbach System of Empirically Based Assessment (ASEBA) and Diagnostic and Statistic Manual of Mental Disorders, Fifth Edition (DSM-V).

In this project, we primarily focus on the prediction of Anxiety Disorder based on the Child Behavioral Checklist (CBCL) DSM-oriented scale. In this context, Anxiety refers to a cluster of parent-reported behaviors and symptoms that reflect anxiety psychopathologies as

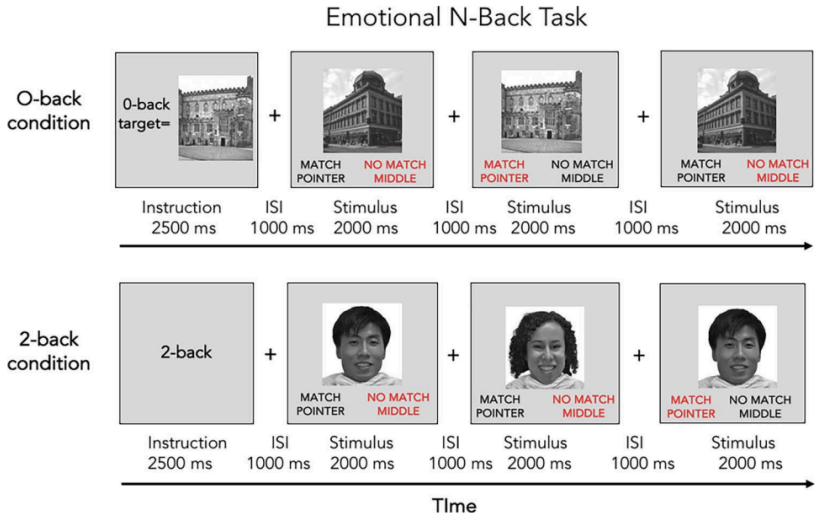

Figure 2. ABCD implementation of the Emotional N-back task (adapted from Casey et al., 2018).
This figure illustrates the specific timing and structure used in the ABCD Study's version of the EN-Back task. Each block begins with a 2500 ms instruction screen, followed by trials that include a 1000 ms inter-stimulus interval (ISI) and a 2000 ms stimulus presentation. Blocks alternate between emotional faces and places to probe working memory performance under emotional and neutral conditions.

defined by the DSM-V. Such conditions are united by excessive worry, restlessness, irritability, sleep disturbances, symptoms which cause distress and impairments in one's social and private lives. Anxious Depression classification using the Child Behavioral Checklist (CBCL) questionnaire based on ASEBA scales was also conducted but not retained for analysis; this outcome is examined in the Appendix section. We used baseline data as a first step, with the intention of extending the analysis to later time points in future work.

Task Paradigm and Neuroimaging Acquisition

Neuroimaging procedures were standardized across sites and involve an approximately 30-minute prescan interview, during which MRI contraindications are assessed and participants

practice task procedures. This was followed by two scan sessions of approximately one hour each (for further details, see Casey et al., 2018).

Among the multiple behavioral tasks included in the ABCD neuroimaging battery, this project focuses on the Emotional N-Back task, designed to measure emotional processing and working memory. In this task, participants were presented with faces displaying various emotional expressions or images of places. They were tasked with identifying whether the current stimulus matches either a predefined target (0-back condition) or the stimulus presented two trials earlier (2-back condition). The emotional stimuli include faces expressing negative, neutral, or positive emotions (Casey et al. 2018). A summary of the EN-Back task design is presented in Figure 2.

fMRI Preprocessing and Feature Extraction

Beta activation coefficients for each task condition were extracted using a Generalized Linear Model (GLM) applied to motion-corrected, distortion-corrected, and spatially normalized fMRI scans (Casey at al., 2018). These beta coefficients, which quantify brain activation during specific task conditions, served as the primary features for classification. Preprocessed beta values were provided directly by the ABCD consortium (Data Release 4.0; Yang & Jernigan, n.d.) . Since each subject completed two runs of the EN-Back task during the scanning session, beta coefficients for each condition were averaged across runs to generate a single feature set per subject.

Cohort Selection and Confounding Mitigation

To ensure robustness and minimize confounding effects, we applied a series of preprocessing steps and sampling strategies during subject selection. Below, we describe the distributions observed in the data and the steps taken to address them. Importantly, control

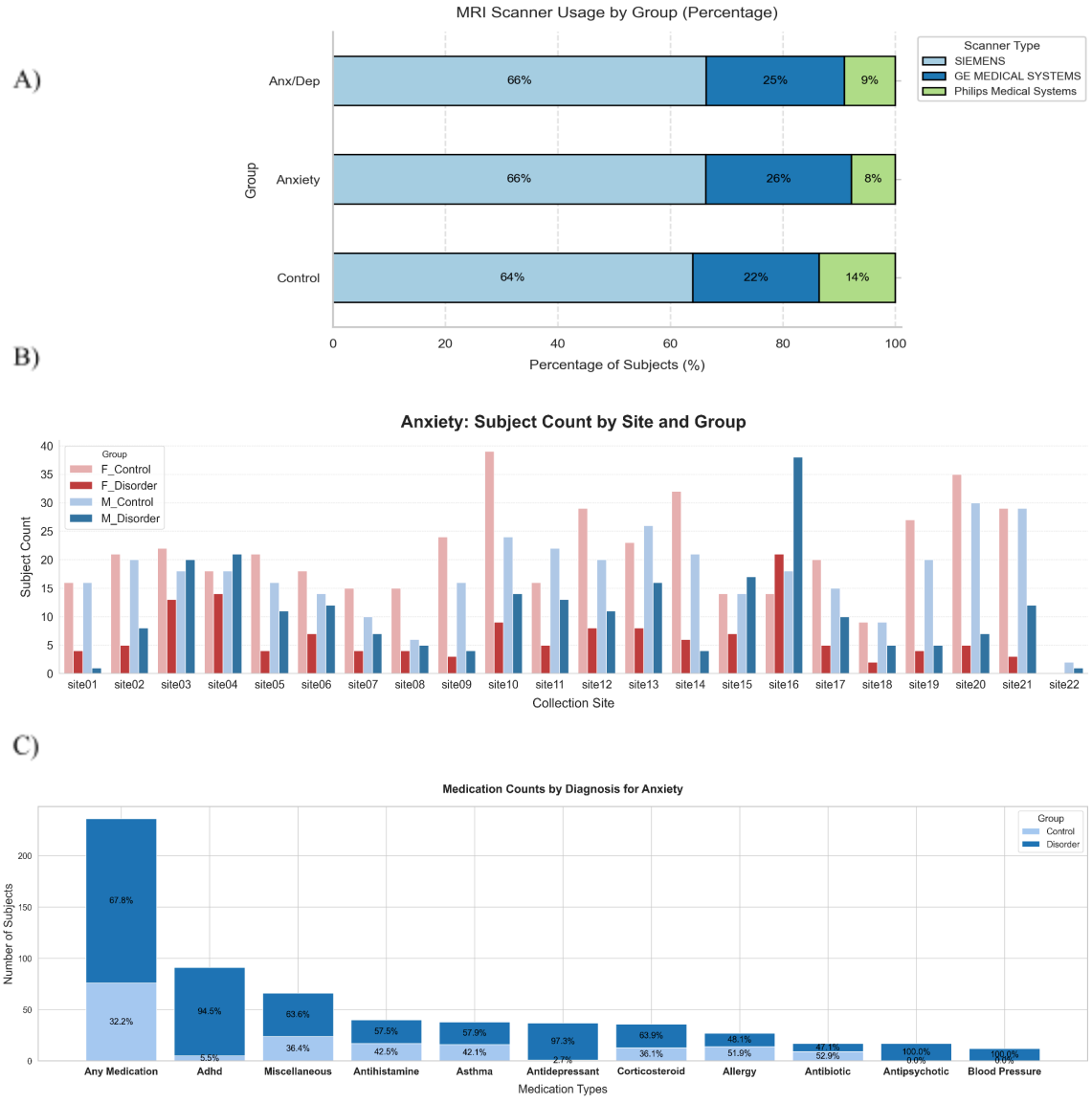

Figure 3.Analysis of ABCD dataset showing subject distribution according to scanner type, collection site coupled with sex and disorder status, and medication intake.

A) Horizontal bar plot summarizing the percentage of subjects scanned on Siemens, GE, and Philips MRI systems within each diagnostic group (Control, Anxiety, Anx/Dep); the percentage for each vendor is labelled on the bar, and the brand is color-coded.

B) For the anxiety diagnosis group and the global control cohort, vertical grouped bars give subject counts per collection site, further split by sex and diagnostic status (female control, female disorder, male control, male disorder). Red shades indicate females; blue shades indicate males. Refer to the appendix for an analogous Anx/Dep plot.

C) Stacked bar plot displaying medication intake counts for the anxiety diagnosis group and the global control cohort. The bar height corresponds to the raw count across the full ABCD baseline cohort, with percentages reflecting the anxiety and control proportions. Group type is color-coded. Refer to the appendix for an analogous Anx/Dep plot.

participants were selected to be free from *any* psychiatric diagnosis from the ASEBA and DSM-oriented scales (all scores ≤ 50) not free from anxiety, to maximize the difference between distributions. Additionally, anxiety participants were identified according to the clinical threshold of 69 points on the CBCL scale.

Scanner Type Distribution and Correction

As shown in Figure 3A, the majority of participants were scanned using Siemens scanners across all groups (66% for Anxiety, 66% for Anxious-Depressed, and 64% for Controls), but a notable fraction was scanned using GE Medical Systems (22–26%) and Philips Medical Systems (8–14%). Since scanner type is shown to have a significant impact on fMRI data (Marek et al. 2016), we ensured that the diagnostic groups were balanced within each scanner type, thereby mitigating scanner-related confounds while retaining maximum generalizability. Specifically, the final group included 159 anxiety participants and 161 controls scanned on Siemens machines, 78 anxiety participants and 80 controls scanned on GE Medical System machines, and 18 anxiety participants and 19 controls scanned on Philips machines.

Site Distribution and Correction

Subject distribution across the 22 ABCD collection sites also varied quite a bit (Figure 3B). To control for site effects, we stratified our subject selection by site. Specifically, for each site, we selected an equal number of participants with and without Anxiety. The number of participants we selected could vary site-by-site depending on how many subjects were available, but within each site, the diagnostic groups were balanced. The test set is composed of two participants per site per diagnostic group—meaning two with anxiety and two controls—ensuring a balanced representation across sites and diagnoses. Additionally, we

balanced the overall sample by sex, regardless of site or diagnosis, to address potential sex-related confounds.

Medication Usage

Medication usage differed between groups, as shown in Figure 3C. Participants with a psychiatric diagnosis were much more likely to report taking medication, especially ADHD medications, antidepressants, and corticosteroids. We tested several strategies to address this, including an analysis that excluded all medicated participants. However, this approach resulted in a substantial reduction in sample size and a concomitant increase in overfitting and decrease in test set performance. Therefore this approach was set aside in favor of the one described below. Instead, we decided to keep medicated participants in the analysis and to include medication usage as an explicit feature. The specific features included were use of ADHD medications, antipsychotic medications, antidepressants, and anticonvulsants. These features were one-hot encoded into a drug feature vector, and 10x importance weighting was applied by to account for the larger number of activation coefficient features in the dataset. This allowed us to control for medication effects without throwing away a large part of the data. Upon the aforementioned corrections, the final cohort consists of 515 (260 controls and 255 anxiety patients).

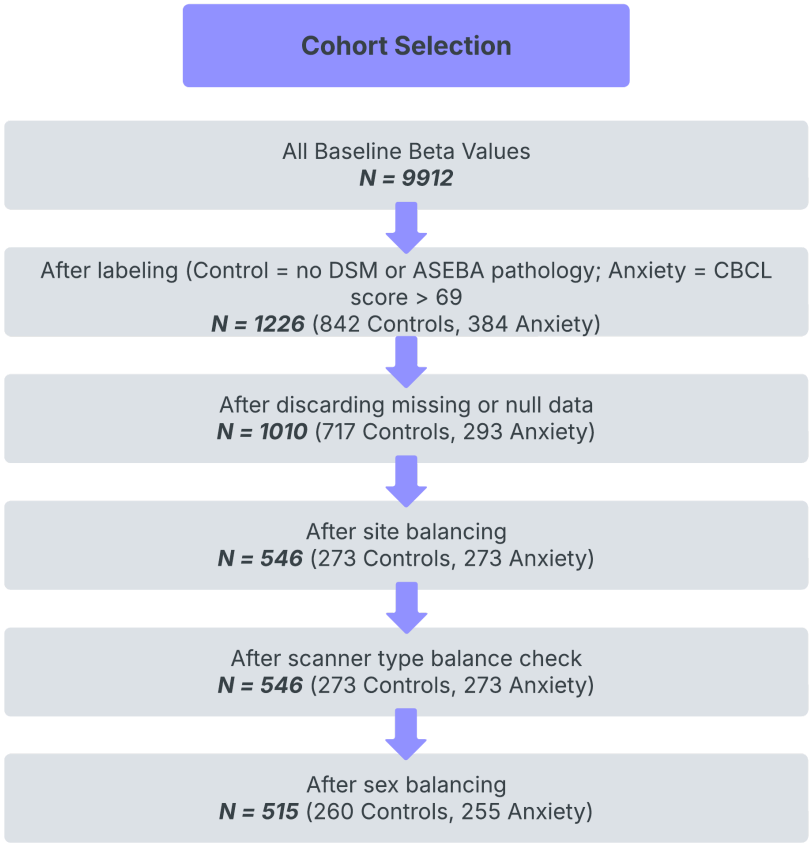

Figure 4. Cohort Selection Pipeline
Flowchart showing the sequential preprocessing steps applied to ABCD baseline beta data (N = 9912) to construct the final anxiety classification cohort. Labeling was based on CBCL-derived DSM-5 t-scores, with anxiety cases defined as scoring >69 and controls restricted to participants without any DSM or ASEBA pathologies. Subjects were removed due to missing data, then balanced within collection site. A scanner-type balance check was performed to ensure diagnostic groups were proportionally distributed across Siemens, GE, and Philips machines. Final sex balancing was applied globally, yielding a cohort of 515 participants (260 controls, 255 anxiety cases).

Machine Learning

We used a supervised machine learning approach to classify participants based on their diagnostic status. The final model used for classification was a Random Forest (RF) classifier, selected after extensive model comparison and hyperparameter tuning. Before settling on Random Forests, which consistently achieved moderate to high performance, we tested several classification algorithms: Linear Support Vector Machines (SVM) with various kernel types, Linear Discriminant Analysis (LDA), K-Nearest Neighbors (KNN), Multi-Layer Perceptron

(MLP), AdaBoost, and Naive Bayes. Alongside different classifiers, we explored multiple feature selection strategies, such as Recursive Feature Elimination (RFE), SelectKBest, Principal Component Analysis (PCA), and permutation testing.

The final sample (see figure 4, N=515) was divided into a test set (2 controls and 2 anxiety patients per site) and training set (the remaining samples). Hyperparameter tuning for the Random Forest Classifier was performed through a grid search procedure on the training set using 5-fold cross validation. The parameters tested included the maximum depth (10 or 20), maximum features considered at each split ('sqrt' or 'log2'), minimum samples required to split an internal node (2 or 5), and the splitting criterion ('gini' or 'entropy'). The best-performing model, selected based on Area Under the Receiving Operating Curve (ROC-AUC) cross-validated performance, had 100 estimators, a maximum depth of 10, used the 'sqrt' strategy for maximum features, required a minimum of 5 samples to split a node, and used 'entropy' as the splitting criterion. All other Random Forest parameters were left at their default settings.

Feature selection was incorporated into the machine learning pipeline using a "Select From Model" approach based on the feature importances output by the Random Forest classifier. Specifically, feature selection was conducted within the grid search with 5-fold cross validation on the training set. Standardization of the beta values was performed prior to model fitting, using z-scoring (StandardScaler) to ensure that the features were on a comparable scale. The best model was retrained on the entire dataset and the selected features after hyperparameter tuning and feature selection.

Software and Environment

All analyses were implemented in Python 3.11.2. The core libraries used include scikit-learn (1.5.2) for model training, pandas (2.2.3) and NumPy (2.0.2) for data handling, and matplotlib (3.9.3) and seaborn (0.13.2) for visualization. Statistical testing was done via SciPy (1.14.1) and statsmodels (0.14.4). Neuroimaging data manipulation relied on nilearn (0.11.1) and nibabel (5.3.2). All analyses were conducted on a 64-bit Unix-based system with 32 CPU threads and 188 GB of RAM. The full codebase is available at *github.com/holliwood8/olimpias-honors-project*.

Permutation Testing

To evaluate the statistical significance of our classification results and the robustness of the identified features, we performed a Monte Carlo permutation test. Specifically, we conducted 1,000 Monte Carlo simulations in which the diagnostic labels (Anxiety vs. Control) were randomly shuffled while preserving the overall distribution of classes across sites. This was done to avoid biasing the site distribution and the associated risk of overfitting. For each iteration, we retrained the full machine learning pipeline and recorded the test accuracy, the test ROC-AUC, and the feature importances. This procedure generated null distributions for each evaluation metric, allowing us to estimate thresholds for significance at the 95th percentile. Based on these null distributions, we determined that a test accuracy greater than 0.59, a test ROC-AUC greater than 0.61, and feature importance values greater than 0.024 can be considered statistically significant at $p < 0.05$.

Results

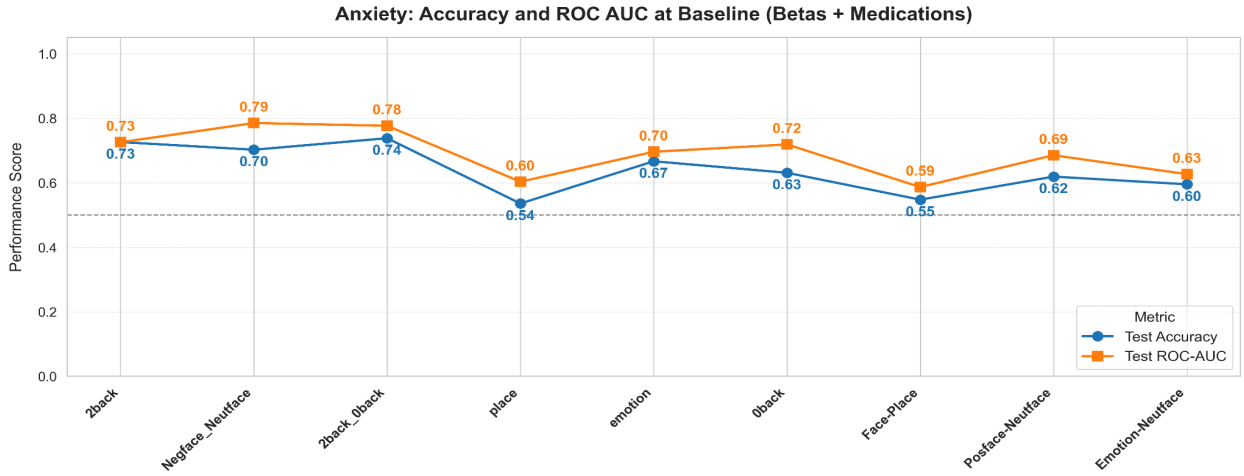

Figure 5. Test set performance metrics for anxiety classification.
Line plot showing test set accuracy (blue) and area under the ROC curve (ROC-AUC; orange) for binary classification of anxiety disorder. Features include beta activation coefficients grouped by task condition, along with medication intake (ADHD medication, antidepressants, antipsychotics, anticonvulsants, and mood stabilizers). Chance-level performance (0.5) is marked by the dashed horizontal line. See appendix for corresponding results on Anx/Dep classification.

2back_0back Contrast Yields Highest Classification Performance

We evaluated the performance of anxiety classification at baseline by combining beta coefficients with medication features across nine task conditions. The final test-set performance, summarized in Figure 5, is reported in terms of both Accuracy and ROC-AUC metrics. These metrics varied substantially across conditions. Overall, the best performance was observed for the 2back_0back contrast, with a test accuracy of 0.74 and a test ROC AUC of 0.78. This condition is designed to highlight differences in brain activity under increased working memory load. Intermediate performances were observed for the 2back and Negface_Neutface contrasts, with test accuracies of 0.73 and 0.70 and ROC AUC values of 0.73 and 0.79, respectively. The 0back and emotion conditions also showed moderate performance, with accuracies of 0.63 and 0.67 and ROC AUC values of 0.72 and 0.70, respectively. Lower performances were noted for the place and Face-Place contrasts, which yielded accuracies of 0.54 and 0.55 and ROC AUC values of 0.60 and 0.59, respectively (these metrics should not be considered statistically

different from chance based on the Monte Carlo simulations described above). Similar results were observed for the Posface-Neutface and Emotion-Neutface contrasts, with accuracies of 0.62 and 0.60 and ROC AUCs of 0.69 and 0.63.

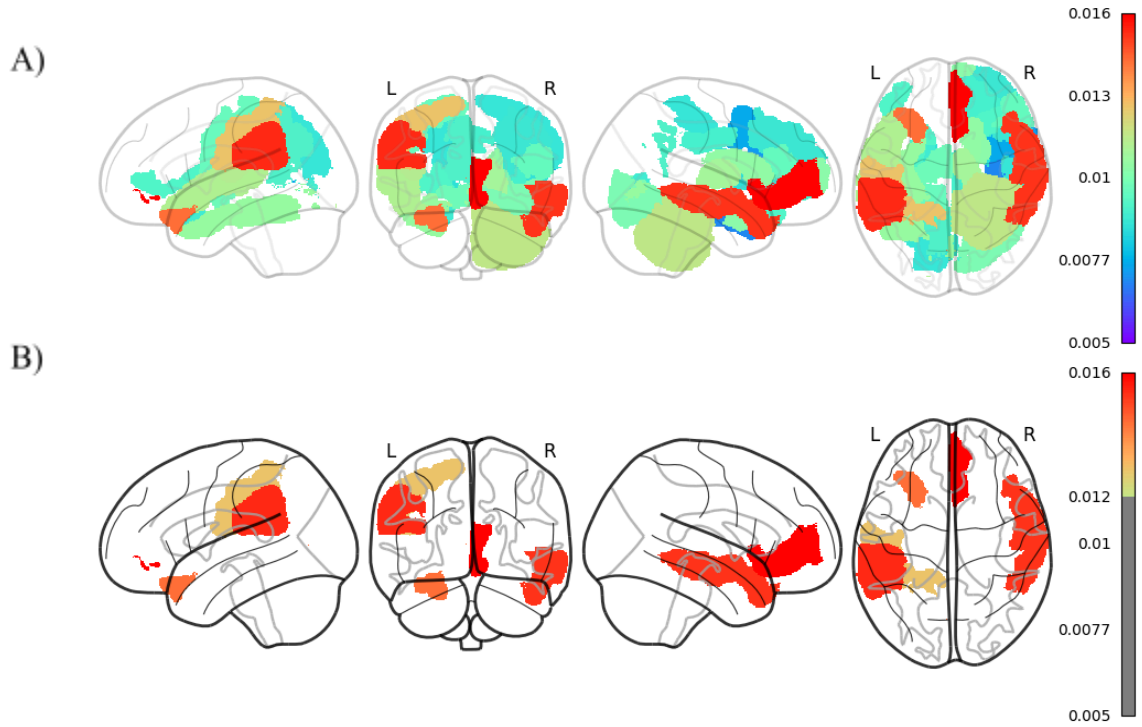

Figure 6. Feature importance brain map for anxiety classification using beta values from the 2-back vs. 0-back contrast and including medication as a feature.

A) The unthresholded map displays all 33 selected ROIs, with color intensity reflecting feature importance: red indicates high importance; purple indicates lower importance, as shown on the accompanying scale.

B) The thresholded map highlights only ROIs with importance values in the 95th percentile, computed after excluding medication features from the distribution.

ADHD Medication Use Emerges as the Most Important Feature

As previously mentioned, this analysis used 'SelectFromModel' for feature selection. Figure 6 shows the selected features shaded according to their assigned importance. In both panels, red shading corresponds to higher feature importance. In total, 35 features were retained following feature selection, including 33 brain activation coefficients and 2 medication features

(ADHD and antidepressant use). All selected neural features are included in figure 6A, while only those within the 95th percentile of importance are shown in the figure 6B. Importantly, this percentile is calculated on the activation coefficients only, with the medication features excluded from the thresholding step. Among all selected features, ADHD medication use obtained the highest importance overall.

Frontal and Temporal Regions Show High Predictive Value for Anxiety

Five brain regions emerged from the thresholding step: the right medial orbitofrontal cortex, the left supramarginal gyrus, the right middle temporal gyrus, the left temporal pole, and the left postcentral gyrus (Figure 6B). These regions span frontal, temporal, and somatosensory areas, and may reflect processes relevant to emotion regulation, semantic memory, and sensory integration—domains often implicated in anxiety. The threshold was applied to mean feature importances obtained from the Random Forest Classifier using the SelectFromModel algorithm. As noted, ADHD medication use and antidepressant use were also retained in the full feature set but are not shown in Figure 6.

Discussion

This study investigated whether brain activation during a working memory task could be used to distinguish preadolescents with anxiety from those without. We found that the model performed above chance across all task conditions, with the best results coming from the most cognitively demanding ones. A small set of brain regions, especially those linked to emotion regulation, cognitive control, and bodily awareness, stood out as important for the classification. Interestingly, ADHD and antidepressant medications' intake was also a strong predictor, highlighting both the clinical relevance and potential limitations of including treatment status in neuroimaging-based models.

To better understand what drove the model's predictions, we next consider the influence of medication use, differences across task conditions, and the functional roles of the selected brain regions.

Advantages and Limitations of Medication Intake as a Predictive Feature

Including medication use as a feature considerably improved classification performance, suggesting that medication status carries clinically relevant information. It is possible that participants prescribed psychiatric medications—especially psychoactive ones like antidepressants or ADHD medications—tend to have more severe or persistent symptoms, which could produce more distinct neural activation patterns. These types of drugs are also known to alter brain structure and function in ways that might enhance group separability (Rubia et al., 2014; Frodl et al., 2011; Delaveau et al., 2011). We cannot distinguish between these possibilities in the current study.

That said, this finding may reflect a methodological vulnerability. The model could be relying too heavily on medication status, identifying participants based on treatment rather than core features of anxiety. This raises concerns about whether the classification is driven more by treatment-related signals than by underlying biomarkers of anxiety itself, especially since medicated participants may differ from unmedicated ones in ways unrelated to diagnosis, such as treatment history, comorbidities, or parental intervention. However, this pattern aligns with dimensional models of psychopathology, like the Research Domain Criteria (RDoC) framework (Cuthbert & Insel, 2013), which argues that brain–behavior relationships become more detectable as symptom severity increases. In that sense, medication use may be a proxy for

chronicity or symptoms severity, and its predictive power, while not ideal from a Machine Learning standpoint, still carries clinical relevance.

Classification Performance by Task Condition

Looking at performance across EN-Back conditions, the strongest classification results came from the 2back_0back contrast (0.74 accuracy, 0.78 ROC-AUC) and the 2back condition (0.73 accuracy, 0.73 ROC-AUC), both of which involve high working memory load. This fits with a long line of research suggesting that anxiety interferes with cognitive control under pressure (Moldawsky & Moldawsky, 1952; Rashkis & Welsh, 1946). Moran (2016) found that this working memory impairment scales with anxiety severity, particularly in dynamic tasks like the one considered here. Other studies (Shackman et al., 2006; Vytal et al., 2012) have pointed to prefrontal dysfunction as a likely mechanism at the base of the anxiety and working memory interaction, especially under cognitive stress. The fact that 0back performed notably worse (0.63 accuracy, 0.72 ROC-AUC) supports the idea that anxiety-related differences become more pronounced when the system is under heightened demand. However, it is also possible that the improved classification in high-load conditions reflects greater overall neural engagement or signal variability, rather than something specific to anxiety-related impairment.

Emotion-based conditions did not perform as well as expected. For instance, Emotion-Neutral Face contrast (0.60 accuracy, 0.63 ROC-AUC) and Positive Face-Neutral Face contrast (0.62 accuracy, 0.69 ROC-AUC) had weaker results, even though anxiety is commonly linked to altered emotion processing. One possible explanation is that emotional faces evoke similar neural responses in both anxious and non-anxious participants, reducing between-group separability. This idea is supported by Chaarani et al. (2021), who showed that emotional contrasts in the ABCD EN-Back task produced weaker and less reproducible activation than

high-load working memory conditions. It's also worth considering that emotion dysregulation may not be uniquely characteristic of anxiety: Ladouceur et al. (2005) found that emotional interference during working memory tasks was more pronounced in youth with depression or comorbid conditions. Thus, emotion-based contrasts might be less specific to anxiety, and more broadly reflective of affective disturbance across diagnoses.

The place-based conditions—place and place versus face—had the worst performance overall. These tasks are low in both emotional salience and cognitive demand, which may limit their ability to engage brain systems typically altered in anxiety. However, another possibility is that these conditions simply lack psychological relevance for the participants. Unlike faces or affectively charged stimuli, place images may not tap into the threat-monitoring or social-evaluative processes that are central to anxiety. Without capturing attention or evoking internal response, these trials may produce noisy or shallow activation patterns. While behavioral engagement metrics are not included, prior findings suggest that spatial contrasts often underperform in differentiating internalizing symptoms, particularly in youth populations (Chaarani et al., 2021; Bachmann et al., 2024; Nord et al., 2027).

Altogether, the performance patterns observed across conditions align well with Attentional Control Theory (Eysenck et al., 2005), which proposes that anxiety impairs executive efficiency, especially under cognitively demanding conditions. Under low-load conditions, anxious participants may compensate more effectively, but as task difficulty increases, top-down control becomes more fragile. Nonetheless, other potential confounding factors such as comorbidity in the anxiety group or engagement levels were not explicitly accounted for, and it's possible that other unmeasured factors are influencing these results.

Brain Features in the Top 5% of Importance Factor, and Their Relationship to Anxiety

The selected brain features offer further insight. The top 5% regions according to feature importance—right medial orbitofrontal cortex, left supramarginal gyrus (SM), right middle temporal gyrus (MTG), left temporal pole, and left postcentral gyrus (S1)—span prefrontal, temporal, and somatosensory areas. The medial orbitofrontal cortex supports affective valuation and reward-based decision-making, a process often disrupted in anxiety. The left supramarginal gyrus, more traditionally associated with phonological processing and verbal working memory, is less commonly linked to anxiety in the literature. However, its involvement may reflect differences in language-based internal processing or verbal rehearsal strategies in anxious individuals, especially under cognitive load. These findings align with theories emphasizing disrupted executive control in anxious youth (Xie et al., 2021).

The middle temporal gyrus, associated with semantic processing and social cognition, has also been linked to adolescent social anxiety. For example, Wang et al. (2021) found that increased gray matter volume in the right MTG was associated with higher social anxiety, a relationship that was mediated by emotional intelligence. Golde et al. (2023) showed that MTG activation, along with the temporal pole, increases with stress and is tied to maladaptive emotional behaviors like avoidance. The temporal pole itself, which helps integrate social context with emotional meaning, has shown promise as a biomarker of pediatric anxiety: Sawalha et al. (2021) reported that activation in this region alone could classify anxiety children with over 80% accuracy, although this study was done in a small sample.

Lastly, the left postcentral gyrus, part of the primary somatosensory cortex, is traditionally involved in processing tactile and bodily input. Yet recent work suggests it may also support emotional awareness and interoceptive regulation through its connections to the insula and the amygdala. Kropf et al. (2019) highlight this role in their review, proposing that the

primary somatosensory cortex contributes to emotional experience by representing internal bodily states. Its inclusion here might reflect a somatic dimension of anxiety that is especially salient in younger populations.

Of course, these interpretations remain tentative. I did not employ model interpretability tools beyond feature importance, and it is possible that some of the selected features reflect noise or indirect associations rather than true signal. Moreover, the stability of these features across different subsamples or in longitudinal designs is uncertain. Nevertheless, the observed overlap with prior findings is encouraging and suggests that the model may be capturing meaningful neurobiological patterns relevant to anxiety.

Limitations

While this study achieves promising classification performances and highlights potential biomarkers of anxiety in preadolescents, several limitations must be acknowledged. First, the use of beta activation coefficients, while standard in task-fMRI analysis, has notable limitations. These coefficients condense complex neural dynamics into averaged responses across a condition, potentially overlooking meaningful temporal dynamics or distributed information. They are also affected by violations of key GLM assumptions—like autocorrelation and HRF mis-modeling—as well as by scanner noise, task variability, and preprocessing steps, all of which can distort the beta estimates and make them less reliable for classification or diagnosis (Monti, 2011).

Second, although including medication use as a feature helped account for treatment-related variance, it complicates interpretability. As noted earlier, the model may be capturing medication-related brain changes or underlying symptom severity rather than features specific to anxiety.

Third, diagnostic labels were based on a parent-reported questionnaires: the Child Behavioral Checklist. While widely used, these assessments are subject to bias and may not reliably reflect internalizing symptoms, which are often less visible to caregivers (Drotar et al., 1995; De Los Reyes & Kazdin, 2005). In addition, both the DSM-V and ASEBA frameworks have been criticized for their rigid categorical structure, lack of neurobiological foundation, and arbitrary diagnostic thresholds, as noted in the introduction. As a result, the training labels may lack the accuracy required which is at the core of supervised Machine Learning.

Fourth, the young age of participants represents a challenge per se. Preadolescents tend to be more restless, and thus show high head motion measures and attention fluctuations during scanning, both of which can undermine signal quality and introduce noise into the beta estimates (Engelhardt et al., 2017; Frew et al., 2022).

Finally, comorbid psychiatric conditions were not excluded in the anxiety and anxious-depressed group, though they were excluded in the control group. While this choice reflects clinical reality and supports generalizability, it limits the ability to isolate features that are specific to anxiety, given the overlapping neural correlates found with other internalizing disorders such as depression.

Future Developments

In light of the limitations discussed above, there are several ways to build on this work. First, future studies should aim to disentangle the effects of medication intake from anxiety itself. As discussed, including medication as a feature helped control for treatment-related variance, but may have introduced a confound. Stratifying by medication status or tracking participants longitudinally could clarify whether the model is detecting diagnosis-specific patterns or simply treatment history.

Second, improving the quality of the target labels could make a meaningful difference. Since the current labels came from parent-report, combining them with child self-reports or clinician administered assessments might better capture internalizing symptoms and reduce misclassification.

Future work could also explore more detailed neural features. As mentioned above, beta coefficients provide a valuable summary but may miss important details. Emerging approaches like Multi-Voxel-Pattern-Analysis, which leverages distributed neural activity, might reveal additional differences between groups (Mahmoudi et al. 2012).

It would also be useful to test the model's stability. Replicating these findings on future ABCD releases would help assess generalizability. Additionally, comparing performance across diagnostic groups could also test whether the selected features are specific to anxiety or are common to other internalizing disorders.

Lastly, future studies should consider moving beyond binary classification and instead use regression models to predict continuous measures of symptom severity. This would be more consistent with dimensional frameworks of psychopathology and could help capture meaningful variation in symptoms severity that binary labels miss. Using structured assessments like the CBCL to generate continuous symptom scores may improve the model's sensitivity and offer a more nuanced view of how brain activation relates to internalizing symptoms.

Conclusion

This Honors Thesis represents a first step toward a data-driven approach to diagnosing anxiety in preadolescents. The abundance of neurobiological data and the maturity of the machine learning field offer a real opportunity to move toward faster, more accurate, and more biologically meaningful psychiatric diagnoses—advances that could ultimately improve

outcomes. At the same time, it's important to be realistic about the limitations of classical machine learning, especially its tendency to overfit in noisy, high-dimensional datasets. The quality of the underlying data matters just as much as the modeling approach, and future work should continue to prioritize careful design, confound control, and meaningful validation.

Appendix

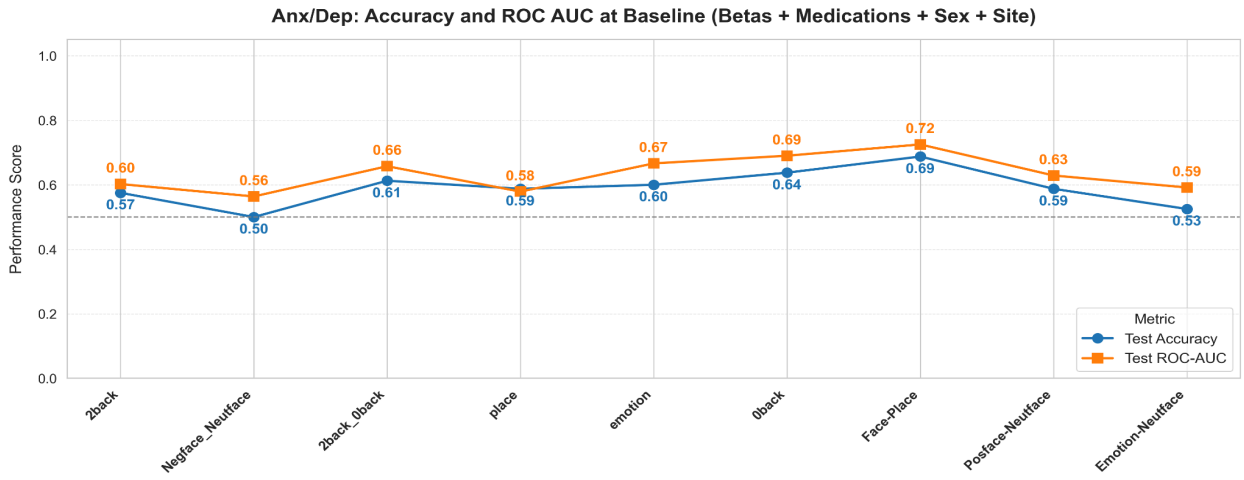

Figure 7. Test set performance metrics for anxious depressed classification.
Line plot showing test set accuracy (blue) and area under the ROC curve (ROC-AUC; orange) for binary classification of anxious depressed disorder. Features include beta activation coefficients grouped by task condition, along with medication intake (ADHD medication, antidepressants, antipsychotics, anticonvulsants, and mood stabilizers), biological sex, and collection site. Chance-level performance (0.5) is marked by the dashed horizontal line.

The Anxious-Depressed (Anx/Dep) scale is one of the empirically derived syndrome scales from the Achenbach System of Empirically Based Assessment, specifically the Child Behavior Checklist for ages 6–18 (Achenbach 2018, Achenbach & Rescorla, 2001). It captures a combination of anxiety and depression symptoms such as nervousness, excessive worry, sadness, crying, and low self-worth which are based on caregiver ratings of the child's behavior over the previous six months. While commonly used in developmental research, it's important to note that the Anx/Dep scale does not correspond directly to any single DSM diagnosis. Instead, it reflects a broader measure of internalizing symptoms.

In this project, a secondary analysis was conducted using the Anx/Dep scale as the target label. The same pipeline was applied, with medication use, sex, and site included as features. However, as shown in the figure above, classification performance was modest. Accuracy scores hovered around chance across most task conditions, and although one condition

(NegFace–NeutFace) reached a ROC-AUC of 0.72, this still fell short of the more consistent results observed in the Anxiety classification.

There are a few likely reasons for this. First, the Anx/Dep scale blends two overlapping but distinct symptom domains, which may have introduced noise and reduced group separability. Second, while the CBCL is an important screening tool, it may not align cleanly with neurobiological patterns in the brain, especially when those patterns are subtle or diagnosis-specific. Finally, because the Anx/Dep group includes a wider range of internalizing symptom presentations, the underlying signal may be more diffuse, making it harder for the model to learn consistent patterns. In addition to these considerations specific to the Anx/Dep scale, the aforementioned limitations in beta activation coefficients, label accuracy, participants' age, and comorbidities remain valid.

For these reasons, and to keep the focus on a more clinically interpretable outcome, the Anx/Dep results were not retained in the main analysis. They are included here for completeness.

References

Achenbach, T. M. (2017). Achenbach System of Empirically Based Assessment (ASEBA). In J. Kreutzer, J. DeLuca, & B. Caplan (Eds.), *Encyclopedia of Clinical Neuropsychology* (pp. 1–7). Springer International Publishing. https://doi.org/10.1007/978-3-319-56782-2_1529-3

Anderson, A. N., King, J. B., & Anderson, J. S. (2019). Neuroimaging in Psychiatry and Neurodevelopment: Why the emperor has no clothes. *The British Journal of Radiology*, *92*(1101), 20180910. https://doi.org/10.1259/bjr.20180910

Bachmann, H. P., Japee, S., Merriam, E. P., & Liu, T. T. (2024). Emotion and anxiety interact to bias spatial attention. *Emotion*, *24*(4), 1109–1124. https://doi.org/10.1037/emo0001322

Bishop, S. J. (2009). Trait anxiety and impoverished prefrontal control of attention. *Nature Neuroscience*, *12*(1), 92–98. https://doi.org/10.1038/nn.2242

Casey, B. J., Cannonier, T., Conley, M. I., Cohen, A. O., Barch, D. M., Heitzeg, M. M., Soules, M. E., Teslovich, 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

Chaarani, B., Hahn, S., Allgaier, N., Adise, S., Owens, M. M., Juliano, A. C., Yuan, D. K., Loso, H., Ivanciu, A., Albaugh, M. D., Dumas, J., Mackey, S., Laurent, J., Ivanova, M., Hagler, D. J., Cornejo, M. D., Hatton, S., Agrawal, A., Aguinaldo, L., … Garavan, H. P. (2021).

Baseline brain function in the preadolescents of the ABCD Study. *Nature Neuroscience*, *24*(8), 1176–1186. https://doi.org/10.1038/s41593-021-00867-9

Chavanne, A. V., Paillère Martinot, M. L., Penttilä, J., Grimmer, Y., Conrod, P., Stringaris, A., Van Noort, B., Isensee, C., Becker, A., Banaschewski, T., Bokde, A. L. W., Desrivières, S., Flor, H., Grigis, A., Garavan, H., Gowland, P., Heinz, A., Brühl, R., Nees, F., … Whelan, R. (2023). Anxiety onset in adolescents: A machine-learning prediction. *Molecular Psychiatry*, *28*(2), 639–646. https://doi.org/10.1038/s41380-022-01840-z

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*, *11*, 126. https://doi.org/10.1186/1741-7015-11-126

De Los Reyes, A., & Kazdin, A. E. (2005). Informant Discrepancies in the Assessment of Childhood Psychopathology: A Critical Review, Theoretical Framework, and Recommendations for Further Study. *Psychological Bulletin*, *131*(4), 483–509. https://doi.org/10.1037/0033-2909.131.4.483

Delaveau, P., Jabourian, M., Lemogne, C., Guionnet, S., Bergouignan, L., & Fossati, P. (2011). Brain effects of antidepressants in major depression: A meta-analysis of emotional processing studies. *Journal of Affective Disorders*, *130*(1), 66–74. https://doi.org/10.1016/j.jad.2010.09.032

Drotar, D., Stein, R. E. K., & Perrin, E. C. (1995). Methodological issues in using the child behavior checklist and its related instruments in clinical child psychology research. *Journal of Clinical Child Psychology*, *24*(2), 184–192. https://doi.org/10.1207/s15374424jccp2402_6

Engelhardt, L. E., Roe, M. A., Juranek, J., DeMaster, D., Harden, K. P., Tucker-Drob, E. M., & Church, J. A. (2017). Children's head motion during fMRI tasks is heritable and stable over time. *Developmental Cognitive Neuroscience*, *25*, 58–68. https://doi.org/10.1016/j.dcn.2017.01.011

Eysenck, M. W., & and Calvo, M. G. (1992). Anxiety and Performance: The Processing Efficiency Theory. *Cognition and Emotion*, *6*(6), 409–434. https://doi.org/10.1080/02699939208409696

Eysenck, M. W., Payne, S., & Derakshan, N. (2005). Trait anxiety, visuospatial processing, and working memory. *Cognition and Emotion*, *19*(8), 1214–1228. https://doi.org/10.1080/02699930500260245

Feng, Y. X., Kiguchi, M., Ung, W. C., Dass, S. C., Mohd Hani, A. F., Tang, T. B., & Ho, E. T. W. (2021). Working Memory Performance under a Negative Affect Is More Susceptible to Higher Cognitive Workloads with Different Neural Haemodynamic Correlates. *Brain Sciences*, *11*(7), Article 7. https://doi.org/10.3390/brainsci11070935

Frew, S., Samara, A., Shearer, H., Eilbott, J., & Vanderwal, T. (2022). Getting the nod: Pediatric head motion in a transdiagnostic sample during movie- and resting-state fMRI. *PLOS ONE*, *17*(4), e0265112. https://doi.org/10.1371/journal.pone.0265112

Frodl, T., Scheuerecker, J., Schoepf, V., Linn, J., Koutsouleris, N., Bokde, A. L. W., Hampel, H., Möller, H.-J., Brückmann, H., Wiesmann, M., & Meisenzahl, E. (2011). Different effects of mirtazapine and venlafaxine on brain activation: An open randomized controlled fMRI

study. *The Journal of Clinical Psychiatry*, *72*(4), 448–457. https://doi.org/10.4088/JCP.09m05393blu

Golde, S., Gleich, T., Romund, L., Stippl, A., Pelz, P., Raufelder, D., Lorenz, R. C., & Beck, A. (2023). Adolescents' neural reactivity to acute psychosocial stress: Dysfunctional regulation habits are linked to temporal gyrus response. *Development and Psychopathology*, *35*(1), 332–344. https://doi.org/10.1017/S0954579421000572.

Health Resources and Services Administration, Maternal and Child Health Bureau. (2024). National Survey of Children's Health: Adolescent mental and behavioral health, 2023. U.S. Department of Health and Human Services. https://mchb.hrsa.gov/sites/default/files/mchb/data-research/nsch-data-brief-adolescent-mental-behavioral-health-2023.pdf

Kassraian-Fard, P., Matthis, C., Balsters, J. H., Maathuis, M. H., & Wenderoth, N. (2016). Promises, Pitfalls, and Basic Guidelines for Applying Machine Learning Classifiers to Psychiatric Imaging Data, with Autism as an Example. *Frontiers in Psychiatry*, *7*, 177. https://doi.org/10.3389/fpsyt.2016.00177

Kotov, R., Krueger, R. F., Watson, D., Achenbach, T. M., Althoff, R. R., Bagby, R. M., Brown, T. A., Carpenter, W. T., Caspi, A., Clark, L. A., Eaton, N. R., Forbes, M. K., Forbush, K. T., Goldberg, D., Hasin, D., Hyman, S. E., Ivanova, M. Y., Lynam, D. R., Markon, K., … Zimmerman, M. (2017). The Hierarchical Taxonomy of Psychopathology (HiTOP): A dimensional alternative to traditional nosologies. *Journal of Abnormal Psychology*, *126*(4), 454–477. https://doi.org/10.1037/abn0000258

Kropf, E., Syan, S. K., Minuzzi, L., & Frey, B. N. (2018). From anatomy to function: The role of the somatosensory cortex in emotional regulation. *Brazilian Journal of Psychiatry*, *41*, 261–269. https://doi.org/10.1590/1516-4446-2018-0183

Ladouceur, C. D., Dahl, R. E., Williamson, D. E., Birmaher, B., Ryan, N. D., & Casey, B. J. (2005). Altered Emotional Processing in Pediatric Anxiety, Depression, and Comorbid Anxiety-Depression. *Journal of Abnormal Child Psychology*, *33*(2), 165–177. https://doi.org/10.1007/s10802-005-1825-z

Mahmoudi, A., Takerkart, S., Regragui, F., Boussaoud, D., & Brovelli, A. (2012). Multivoxel Pattern Analysis for fMRI Data: A Review. *Computational and Mathematical Methods in Medicine*, *2012*, 961257. https://doi.org/10.1155/2012/961257

Marek, S., Tervo-Clemmens, B., Nielsen, A. N., Wheelock, M. D., Miller, R. L., Laumann, T. O., Earl, E., Foran, W. W., Cordova, M., Doyle, O., Perrone, A., Miranda-Dominguez, O., Feczko, E., Sturgeon, D., Graham, A., Hermosillo, R., Snider, K., Galassi, A., Nagel, B. J., … Dosenbach, N. U. F. (2019). Identifying reproducible individual differences in childhood functional brain networks: An ABCD study. *Developmental Cognitive Neuroscience*, *40*, 100706. https://doi.org/10.1016/j.dcn.2019.100706

McGorry, P. D., & Mei, C. (2018). Early intervention in youth mental health: Progress and future directions. *Evidence Based Mental Health*, *21*(4). https://doi.org/10.1136/ebmental-2018-300060

Merikangas, K. R., He, J., Burstein, M., Swanson, S. A., Avenevoli, S., Cui, L., Benjet, C., Georgiades, K., & Swendsen, J. (n.d.). Lifetime Prevalence of Mental Disorders in U.S.

Adolescents: Results from the National Comorbidity Survey Replication– Adolescent Supplement (NCS-A). *Adolescent Psychiatry*.

Moldawsky, S., & Moldawsky, P. C. (1952). Digit span as an anxiety indicator. *Journal of Consulting Psychology*, *16*(2), 115–118. https://doi.org/10.1037/h0059340

Monti, M. M. (2011). Statistical Analysis of fMRI Time-Series: A Critical Review of the GLM Approach. *Frontiers in Human Neuroscience*, *5*. https://doi.org/10.3389/fnhum.2011.00028

Moran, T. P. (2016). Anxiety and working memory capacity: A meta-analysis and narrative review. *Psychological Bulletin*, *142*(8), 831–864. https://doi.org/10.1037/bul0000051

Mousavian, M., Chen, J., Traylor, Z., & Greening, S. (2021). Depression detection from sMRI and rs-fMRI images using machine learning. *Journal of Intelligent Information Systems*, *57*(2), 395–418. https://doi.org/10.1007/s10844-021-00653-w

Nord, C. L., Gray, A., Charpentier, C. J., Robinson, O. J., & Roiser, J. P. (2017). Unreliability of putative fMRI biomarkers during emotional face processing. *NeuroImage*, *156*, 119–127. https://doi.org/10.1016/j.neuroimage.2017.05.024

Pickersgill, M. D. (2014). Debating DSM-5: Diagnosis and the sociology of critique. *Journal of Medical Ethics*, *40*(8), 521–525. https://doi.org/10.1136/medethics-2013-101762

Portugal, L. C. L., Schrouff, J., Stiffler, R., Bertocci, M., Bebko, G., Chase, H., Lockovitch, J., Aslam, H., Graur, S., Greenberg, T., Pereira, M., Oliveira, L., Phillips, M., & Mourão-Miranda, J. (2019). Predicting anxiety from wholebrain activity patterns to

emotional faces in young adults: A machine learning approach. *NeuroImage: Clinical*, *23*, 101813. https://doi.org/10.1016/j.nicl.2019.101813

Pulini, A. A., Kerr, W. T., Loo, S. K., & Lenartowicz, A. (2019). Classification Accuracy of Neuroimaging Biomarkers in Attention-Deficit/Hyperactivity Disorder: Effects of Sample Size and Circular Analysis. *Biological Psychiatry. Cognitive Neuroscience and Neuroimaging*, *4*(2), 108–120. https://doi.org/10.1016/j.bpsc.2018.06.003

Rashkis, H. A., & Welsh, G. S. (1946). Detection of anxiety by use of the Wechsler scale. *Journal of Clinical Psychology*, *2*(4), 354–357. https://doi.org/10.1002/1097-4679(194610)2:4<354::AID-JCLP2270020408>3.0.CO;2-B

Rubia, K., Alegria, A. A., Cubillo, A. I., Smith, A. B., Brammer, M. J., & Radua, J. (2014). Effects of Stimulants on Brain Function in Attention-Deficit/Hyperactivity Disorder: A Systematic Review and Meta-Analysis. *Biological Psychiatry*, *76*(8), 616–628. https://doi.org/10.1016/j.biopsych.2013.10.016

Sawalha, J., Yousefnezhad, M., Selvitella, A. M., Cao, B., Greenshaw, A. J., & Greiner, R. (2021). Predicting pediatric anxiety from the temporal pole using neural responses to emotional faces. *Scientific Reports*, *11*(1), 16723. https://doi.org/10.1038/s41598-021-95987-4

Wang, S., Zhao, Y., Wang, X., Yang, X., Cheng, B., Pan, N., Suo, X., & Gong, Q. (2021). Emotional intelligence mediates the association between middle temporal gyrus gray matter volume and social anxiety in late adolescence. *European Child & Adolescent Psychiatry*, *30*(12), 1857–1869. https://doi.org/10.1007/s00787-020-01651-z

Xie, S., Zhang, X., Cheng, W., & Yang, Z. (2021). Adolescent anxiety disorders and the developing brain: Comparing neuroimaging findings in adolescents and adults. *General Psychiatry*, *34*(4), e100411. https://doi.org/10.1136/gpsych-2020-100411

Yang, R. & Jernigan, Terry. (n.d.). *Adolescent Brain Cognitive Development Study (ABCD)—Annual Release 4.0* [Dataset]. NIMH Data Repositories. https://doi.org/10.15154/1523041