ARTICLE OPEN

Check for updates

Using deep learning to predict internalizing problems from brain structure in youth

Marlee M. Vandewouw[1,2]✉, Bilal Syed[3], Noah Barnett[4], Alfredo Arias[5], Elizabeth Kelley[6,7], Jessica Jones[6,7], Muhammad Ayub[6,8], Alana Iaboni[1], Paul D. Arnold[9,10], Jennifer Crosbie[11,12], Russell J. Schachar[12,13], Margot J. Taylor[11,14,15,16], Jason P. Lerch[11,17,18], Evdokia Anagnostou[1,11] and Azadeh Kushki[1,2]

Internalizing problems (e.g., anxiety and depression) are associated with a wide range of adverse outcomes. While some predictors of internalizing problems are known (e.g., their frequent co-occurrence with neurodevelopmental (ND) conditions), the biological markers of internalizing problems are not well understood. Here, we used deep learning, a powerful tool for identifying complex and multi-dimensional brain-behaviour relationships, to predict cross-sectional and worsening longitudinal trajectories of internalizing problems. Data were extracted from four large-scale datasets: the Adolescent Brain Cognitive Development study, the Healthy Brain Network, the Human Connectome Project Development study, and the Province of Ontario Neurodevelopmental network. We developed deep learning models that used measures of brain structure (thickness, surface area, and volume) to (a) predict clinically significant internalizing problems cross-sectionally ($N = 14,523$); and (b) predict subsequent worsening trajectories (using the reliable change index) of internalizing problems ($N = 10,540$) longitudinally. A stratified cross-validation scheme was used to tune, train, and test the models, which were evaluated using the area under the receiving operating characteristic curve (AUC). The cross-sectional model performed well across the sample, reaching an AUC of 0.80 [95% CI: 0.71, 0.88]. For the longitudinal model, while performance was sub-optimal for predicting worsening trajectories in a sample of the general population (AUC = 0.66 [0.65, 0.67]), good performance was achieved in a small, external test set of primarily ND conditions (AUC = 0.80 [0.78, 0.81]), as well as across all ND conditions (AUC = 0.73 [0.70, 0.76]). Deep learning with features of brain structure is a promising avenue for biomarkers of internalizing problems, particularly for individuals who have a higher likelihood of experiencing difficulties.

Translational Psychiatry (2025)15:326 ; https://doi.org/10.1038/s41398-025-03565-3

INTRODUCTION

Mental health problems in children and youth are widespread, with 11% estimated to have a mental health disorder [1]. Internalizing problems, specifically, which include feelings of anxiety, depression, and social withdrawal, are associated with profoundly negative outcomes, including poorer quality of life and worse long-term social and economic outcomes [2, 3]. This highlights the need for biological markers that help us understand who is most at risk for developing problems in this domain, taking a step towards improving our ability to deliver proactive care that can effectively minimize adverse outcomes. However, these markers have thus far remained elusive.

Neurobiological markers are a promising approach to bridging the gap between wide ranges of genetic and environmental variation and subsequent manifestations of internalizing problems. Measures of brain structure, such as cortical thickness, cortical surface area, and cortical and subcortical volume, are particularly appealing in this context due to their high test-retest reliability [4]. Recent large-scale studies in community samples have shown that internalizing symptoms in youth are associated with altered cortical thickness and volume in frontal, limbic, and temporal regions [5–7], although some studies have reported these alterations are not specific to internalizing symptoms [8, 9]. While these structural markers reveal associations with internalizing symptoms, they offer limited insight into individual-level risk; neuroimaging-based predictive models offer a promising avenue for translating these patterns into tools that can inform

[1]Autism Research Centre, Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, ON, Canada. [2]Institute of Biomedical Engineering, University of Toronto, Toronto, Canada. [3]Mouse Imaging Centre, The Hospital for Sick Children, Toronto, Canada. [4]Schulich School of Business, York University, Toronto, Canada. [5]Participant Advisory Committee, Province of Ontario Neurodevelopmental Network, Toronto, Canada. [6]Department of Psychiatry, Queen's University, Kingston, Canada. [7]Department of Psychology, Queen's University, Kingston, Canada. [8]Division of Psychiatry, University College London, London, UK. [9]The Mathison Centre for Mental Health & Education, Hotchkiss Brain Institute, University of Calgary, Calgary, Canada. [10]Department of Psychiatry and Medical Genetics, Cumming School of Medicine, University of Calgary, Calgary, Canada. [11]Program in Neurosciences & Mental Health, The Hospital for Sick Children, Toronto, Canada. [12]Department of Psychiatry, University of Toronto, Toronto, Canada. [13]Department of Psychiatry, The Hospital for Sick Children, Toronto, Canada. [14]Department of Diagnostic & Interventional Radiology, The Hospital for Sick Children, Toronto, Canada. [15]Department of Psychology, University of Toronto, Toronto, Canada. [16]Department of Medical Imaging, University of Toronto, Toronto, Canada. [17]Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, UK. [18]Department of Medical Biophysics, University of Toronto, Toronto, Canada. Emiritus: Russell J. Schachar. ✉email: mvandewouw@hollandbloorview.ca

Received: 20 January 2025 Revised: 2 July 2025 Accepted: 18 August 2025
Published online: 29 August 2025

personalized mental health care [10]. However, investigations into whether measures of brain structure can predict internalizing problems have been mixed [11–16]. For example, some have reported that children with lower cortical thickness experienced steeper declines in internalizing symptoms over time [11], while others have reported no predictive effects [13–15].

Thus far, this work has largely relied on methodological approaches that may be insufficient for characterizing the multi-dimensional and complex brain-behaviour associations related to mental health. For example, regression has typically been used, relying on manual feature selection, restricting brain-behaviour relationships to being linear, and modeling each brain region independently. This calls for complementary strategies which add depth to the existing work, advancing our understanding of the brain-behaviour associations of internalizing problems. Artificial intelligence, in particular deep learning, is increasingly recognized as a valuable tool for predicting mental health outcomes [17]. However, work in this space has primarily focused on predicting diagnostic categories [18], which overlooks individualized neurobiological variation that is critical for personalized mental health care. While some studies have begun using this approach to predict mental health symptoms from neurobiology [19, 20], it has yet to be applied to predict internalizing problems both cross-sectionally and longitudinally using measures of brain structure.

Here, we used deep learning to predict internalizing problems in youth from neurobiological measures. We drew from four multinational, independently collected datasets of children and youth from Canada and the United States. We first used cross-sectional measures of brain structure to predict the presence of clinically significant internalizing problems ($N = 14{,}523$), to provide insight into the underlying brain features beyond what can be revealed by traditional approaches. Next, we used longitudinal data ($N = 10{,}540$) to examine whether brain structure can predict subsequent worsening trajectories of internalizing problems. The cohorts include both neurotypical (NT) children and youth and those with neurodevelopmental (ND) conditions. Given the higher prevalence of internalizing problems in the ND compared to NT population [21], we also provide measures of model performance stratified by the presence or absence of any ND diagnosis alongside overall performance.

METHODS

Participants

Data from children and adolescents were extracted from four independent datasets: the Adolescent Brain Cognitive Development (ABCD [22]) study (Release 5.1), the Healthy Brain Network (HBN [23]; Release 10), the Human Connectome Project Development (HCP-D [24]) study (Release 2.0), and the Province of Ontario Neurodevelopmental (POND) network (exported January 2024). The ABCD and POND datasets contain longitudinal data, while HBN and HCP-D provide only cross-sectional data. Participants were eligible for inclusion based on the availability of T1-weighted structural magnetic resonance images (MRIs) and the Child Behaviour Checklist (CBCL [25]), resulting in at least one datapoint from 14,950 participants (ABCD: 11,796, HBN: 1,959, HCP-D: 498, POND: 697). Flowcharts of the study samples are presented in Supplemental Fig. 1.

We acknowledge that individuals have different language preferences for referring to neurodevelopmental disorders/conditions, and we honour such perspectives. Given that in this study we defined group membership according to the diagnostic criteria in the DSM-5, we have used this nomenclature in the manuscript, and in partnership with our family and youth advisories (represented by NB and AA).

Ethics approval and consent to participate

The Holland Bloorview Kids Rehabilitation Hospital's research ethics board approved the current study (#1854). The recruitment strategies, assent/consent procedures, institutional review board approvals, and data usage agreements for each of the datasets are provided in the Supplementary material; informed consent was obtained from all participants. All methods were performed in accordance with the relevant guidelines and regulations.

Data

Biological sex, race/ethnicity, annual household income, and highest household education were collected as part of all four study protocols (see ***Sociodemographic assessments*** in the Supplementary material).

Internalizing problems were measured using the CBCL [25], a caregiver-reported assessment of problem behaviour in children and adolescents. The assessment derives syndromic scales corresponding to behavioural problems, including anxious/depressed, withdrawn/depressed, and somatic complaints, which are used to calculate the higher-order construct of internalizing problems. Raw scores were converted to age- and sex-normed standardized *T*-scores, with higher scores indicating greater problems. We used binarized outcomes rather than continuous symptom scores to reduce the influence of measurement error and focus on clinically meaningful thresholds. While continuous scores can vary within a range that may not reflect true differences in severity, binary outcomes based on validated cut-offs help mitigate this issue. Additionally, binary models yield probabilistic outputs, allowing flexible interpretation and alignment with prior work in this area. For the cross-sectional analysis, the provided threshold (*T*-score ≥64) was used to categorize individuals as having clinically significant internalizing problems. For the longitudinal analysis, the reliable change index (RCI [26]) was used to categorize individuals as having worsening (RCI > 1.96) or non-worsening (RCI ≤ 1.96) trajectories of internalizing problems (see the ***Reliable change index*** section in the Supplementary material).

T1-weighted images were acquired as part of all four study protocols (***Imaging protocols*** in the Supplementary material). The FreeSurfer image analysis suite [27] was used to perform cortical reconstruction and volume segmentation of the structural MRIs (see the ***Imaging preprocessing*** section of the Supplementary material), which provided measures of cortical thickness, surface area, and volume according to the Desikan-Killiany parcellation [28], alongside volumes of subcortical structures, ventricles, white matter, brainstem, cerebellum, and whole-brain. Processed FreeSurfer outputs were provided by ABCD and HCP-D, while the pipeline was run in-house for POND and HBN. Datapoints with poor data quality were excluded from all subsequent analyses (see Supplementary material, section ***Quality control***).

After quality control, for the cross-sectional analysis, one datapoint per participant was selected, resulting in a sample size of 14,523 children and youth (ABCD: 11,632, HBN: 1,785, HCP-D: 498, POND: 608). For the longitudinal analysis, partial data were available for a third timepoint in the ABCD study, which was leveraged to maximize the total number of included datapoints, resulting in 10,532 datapoints (ABCD: 10,491, POND: 49) from 8,103 participants (ABCD: 8,062, POND: 49). See the ***Datapoint selection*** section in the Supplementary material for further information on how datapoints were selected for both analyses, and the distribution of age at both timepoints is visualized in Supplemental Fig. 2.

Deep learning models

The primary goal of the cross-sectional deep learning model was to establish the features of brain structure that contribute to clinically significant internalizing problems. To do so, we built a binary classification model, a type of model that predicts one of two possible outcomes (here, clinically significant versus non-significant internalizing problems), using features of brain structure as input. We then examined the contribution of each feature to the model's prediction. The model was tuned, trained, and tested using data aggregated across all four datasets. The model architecture was a multilayer perceptron, consisting of an input layer, blocks of hidden layers, and an output layer (see ***Deep learning model architecture*** in the Supplementary material for further details). Input features included the measures of brain structure ($N = 241$) along with age and sex, given their association with internalizing problems [29]. Note that we did not include ND diagnosis as an input feature, to allow neurobiological patterns to emerge without imposing diagnostic boundaries, consistent with findings of transdiagnostic, data-driven studies that have highlighted a misalignment between diagnostic labels and neurobiological features frameworks [30–34]. We also ran the model only including age and sex as predictors, to evaluate the added predictive power of neuroanatomy. All features were *z*-scored prior to building the model. A nested 5-fold stratified cross-validation scheme was used to tune, train, and test the model (Supplemental Figure 3). For each iteration, the dataset was split into a training set (80%), which was used for hyperparameter tuning and model training, and 20% was withheld for testing purposes. Prediction scores (probabilities of the positive class) and feature importances (Shapley Additive exPlanations [35] (SHAP) values) were obtained for the test set. To ensure generalizability, this was performed across five folds.

The primary goal of the longitudinal deep learning model was to predict whether an individual would experience worsening internalizing problems from baseline measures of brain structure. The model architecture and tuning, training, and testing procedure was identical to the cross-sectional model. Predictors included baseline age, baseline sex, baseline internalizing problems, and between-timepoint age differences as features alongside the baseline measures of brain structure. Given the sample size imbalance between ABCD and POND, tuning, training, and testing were performed using the ABCD dataset, holding out POND as an external testing set.

Performance evaluation

The primary measure used to evaluate model performance was the area under the receiving operating characteristic (ROC) curve (AUC). The secondary measure was accuracy, computed as the prediction score cutoff that maximized the average recall of both classes. Outcome measures were computed across all samples, as well as stratified by the presence or absence of any ND diagnosis. Feature importances (SHAP values) were used to measure each feature's contribution to the model's prediction.

Statistical analysis

AUCs and accuracies are reported across folds using the mean and 95% confidence intervals (CIs). Bias with respect to sociodemographic factors (sex, age, race/ethnicity, annual household income, and highest household education) was assessed by examining differences in AUC. Overall feature importances were computed across folds by taking the mean of the magnitudes of the SHAP values across all datapoints.

RESULTS

Samples

Data from 14,523 children and youth between 5 – 21 years of age were included for the cross-sectional analysis; participant demographics are presented in Table 1. Across all four datasets, 1967 [14%] participants had clinically significant internalizing problems, and internalizing problems were more prevalent in individuals with an ND diagnosis (31%) compared to those without (8%).

For the longitudinal analysis, 10,540 datapoints were included (Table 2). Participants were aged 10.3 [9.6, 10.9] years at baseline (range: 7-15 years), and 12.4 [11.6, 13.2] at follow-up (range: 10-19 years). Of these participants, 2029 [19%] had worsening changes in internalizing problems between the baseline and follow-up timepoints. The prevalence rates of worsening trajectories in individuals with (17%) and without (20%) an ND were similar.

Table 1. Participant demographics for the cross-sectional analysis.

		***N* [%]**			
		ABCD	**HBN**	**HCP-D**	**POND**
N		11,632	1785	608	498
ND diagnosis[a]		1650 [14]	1310 [74]	0 [0]	408 [67]
Internalizing problems[b]	Total sample	1129 [10]	560 [31]	30 [6]	248 [41]
	No ND diagnosis	691 [7]	130 [28]	30 [6]	54 [27]
	ND diagnosis	435 [26]	428 [33]	–	194 [48]
Age range (years)		8 – 15	5 – 21	6 – 17	5 – 19
Median age (years) [IQR]		10.3 [9.5, 11.4]	10.1 [8.2, 13.1]	13.0 [10.2, 14.9]	11.7 [9.7, 14.4]
Sex	Male	6051 [52]	1143 [64]	229 [46]	430 [71]
	Female	5581 [48]	642 [36]	269 [54]	178 [29]
Race and ethnicity[c]	Hispanic/Latino	2368 [20]	301 [17]	40 [8]	23 [4]
	Non-Hispanic Asian	245 [2]	50 [3]	23 [5]	29 [5]
	Non-Hispanic Black	1719 [15]	191 [11]	33 [7]	8 [1]
	Non-Hispanic White	6077 [52]	750 [42]	308 [62]	293 [48]
	Multi-Racial/Other	1222 [11]	275 [15]	89 [18]	86 [14]
	Unknown/Not reported	1 [0]	218 [12]	5 [1]	169 [28]
Annual household income[d]	<$50,000	3002 [26]	287 [16]	61 [12]	81 [13]
	$50,000-$99,999	2924 [25]	322 [18]	105 [21]	83 [14]
	≥$100,000	4705 [40]	719 [40]	298 [60]	197 [32]
	Unknown/Not reported	1001 [9]	457 [26]	34 [7]	247 [41]
Highest household education[e]	Below high school	429 [4]	45 [3]	6 [1]	2 [0]
	High school or equivalent	1930 [17]	285 [16]	49 [10]	41 [7]
	Undergraduate	3213 [28]	570 [32]	182 [37]	236 [39]
	Graduate	2849 [24]	854 [48]	261 [52]	121 [20]
	Unknown/Not reported	3211 [28]	31 [2]	0 [0]	208 [34]

ABCD adolescent brain cognitive development study, *HBN* healthy brain network, *HCP-D* human connectome project development study, *POND* province of Ontario neurodevelopmental network, *IQR* interquartile range.
[a]According to the DSM-5 categories; data was not available for 519 (4%) of participants.
[b]Determined using a clinically significant cutoff of ≥64.
[c]Reported according to recommendations of the ABCD study [48].
[d]In local currency (CAD or USD).
[e]High school or equivalent includes: high school graduate, General Education Development (GED) diploma or equivalent, less than one year of college credit/post-secondary education (or less than 10 classes), or one year or more of college credit without a degree; Graduate includes: master's degree, professional School degree, or doctoral degree.

Table 2. Participant demographics for the longitudinal analysis.

		N [%]	
		ABCD	**POND**
N	Unique participants	8062	49
	Datapoints	10,491	49
ND diagnosis[a]		1522 [15]	32 [65]
Worsening internalizing problems[b]	Total sample	2019 [19]	10 [20]
	No ND diagnosis	1756 [20]	5 [29]
	ND diagnosis	259 [17]	5 [16]
Age range (years)	Baseline	8 – 13	7 – 15
	Follow-up	10 – 15	10 – 19
Median age (years) [IQR]	Baseline	10.3 [9.6, 10.9]	11.7 [9.8, 12.8]
	Follow-up	12.3 [11.6, 13.3]	15.2 [12.7, 16.8]
Sex	Male	5584 [53]	35 [71]
	Female	4907 [47]	14 [29]
Race and ethnicity[c]	Hispanic/Latino	2038 [19]	1 [2]
	Non-Hispanic Asian	206 [2]	0 [0]
	Non-Hispanic Black	1312 [13]	0 [0]
	Non-Hispanic White	5850 [56]	33 [67]
	Multi-Racial/Other	1085 [10]	9 [18]
	Unknown/Not reported	0 [0]	6 [12]
Annual household income[d]	<$50,000	2616 [25]	10 [20]
	$50,000-$99,999	2886 [28]	12 [24]
	≥$100,000	4222 [32]	18 [37]
	Unknown/Not reported	767 [7]	9 [18]
Highest household education[e]	Below high school	355 [3]	0 [0]
	High school or equivalent	1649 [16]	5 [10]
	Undergraduate	3128 [30]	25 [51]
	Graduate	2799 [27]	14 [29]
	Unknown/Not reported	2560 [24]	5 [10]

ABCD adolescent brain cognitive development study, *POND* province of ontario neurodevelopmental network, *IQR* interquartile range.
[a]Basline data, according to the DSM-5 categories; data was not available for 0.2% of samples.
[b]Determined using the reliable change index (RCI).
[c]Baseline data, reported according to recommendations of the ABCD study [48].
[d]Baseline data in local currency (CAD or USD).
[e]Baseline data; High school or equivalent includes: high school graduate, General Education Development (GED) diploma or equivalent, less than one year of college credit/post-secondary education (or less than 10 classes), or one year or more of college credit without a degree; Graduate includes: master's degree, professional School degree, or doctoral degree.

Predicting cross-sectional internalizing problems

We first examined the features of brain structure that contribute to internalizing symptoms by predicting clinically significant (CBCL *T*-score ≥64) versus non-significant (CBCL *T*-score <64) problems cross-sectionally. The ROC curves across the entire sample and disaggregated by ND diagnosis are shown in Fig. 1A. Across the cross-validation folds, the overall AUC was 0.80 (95% CI: [0.71, 0.88]), while the AUC for individuals with and without a diagnosis of an ND was 0.71 [0.60, 0.82] and 0.78 [0.68, 0.89], respectively. At the optimal prediction score cut-off (0.49 [0.46, 0.52]), an accuracy of 77% [68, 86] was achieved (ND: 62% [50, 75]; no ND: 82% [74, 90]). When only age and sex were included as predictors, AUC was reduced to 0.58 [0.56, 0.59].

AUCs were also compared amongst sociodemographic groups (age, sex, race/ethnicity, annual household income, and highest household education) to evaluate fairness (Fig. 1B). No significant differences were observed (Table 3), indicating that our model performed comparably for the different sociodemographic groups.

Feature importances, which capture the relative extent to which each feature influences the final prediction, were calculated using the magnitude of SHAP values (Fig. 1C and D). The most important features were sex (with females having a negative impact on prediction scores), the area and thickness of the right frontal pole, the areas of the left temporal pole, left rostral middle frontal, and right medial orbitofrontal gyri, and the volumes of the left medial orbitofrontal gyrus, left pars triangularis, right pericalcarine cortex, and posterior corpus callosum.

Predicting worsening trajectories of internalizing problems

We then predicted whether an individual would experience worsening (RCI > 1.96) versus non-worsening (RCI ≤ 1.96) internalizing problems from baseline measures of brain structure. The ROC curves across the entire sample and disaggregated by ND diagnosis are shown in Fig. 2A. Across the cross-validation folds, the AUC was 0.66 (95% CI: [0.65, 0.67]) for ABCD, while the AUC for POND was 0.80 [0.78, 0.81]. Stratified by ND diagnosis across both datasets, the AUCs for individuals with and without an ND were 0.73 (95% CI: [0.70, 0.76]) and 0.65 (95% CI: [0.65, 0.67]), respectively. At the optimal prediction score cutoff (0.53 [0.50, 0.55]), an accuracy of 71% [65%, 77%] was achieved (ABCD: 63%

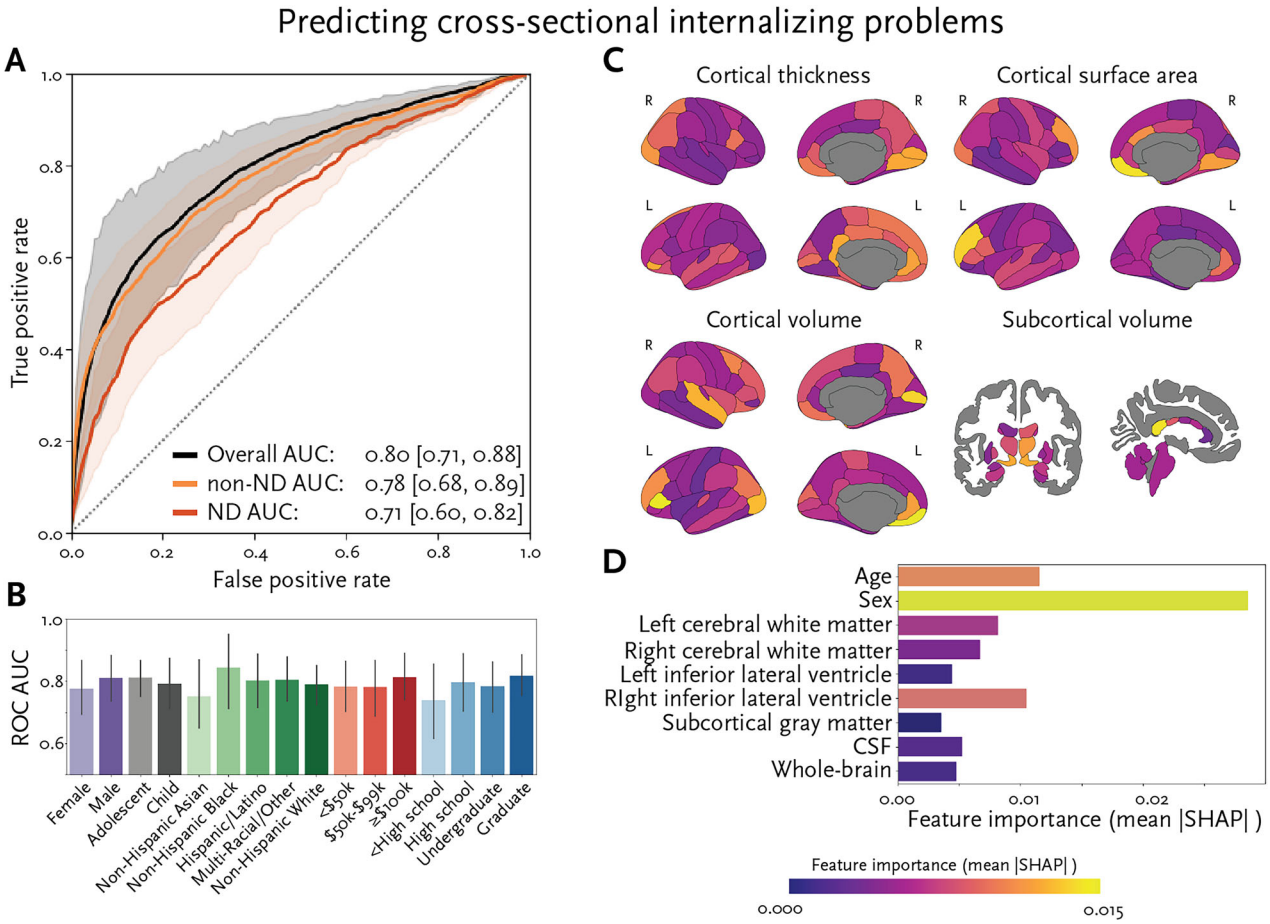

Fig. 1 Performance and feature importances for the model predicting cross-sectional internalizing problems. A The mean [95% CI] receiving operating characteristic curve and corresponding area under the curve (AUC) across all data (black) and stratified by individuals with (red) and without (orange) an ND diagnosis. **B** AUCs in different sociodemographic groups (age, sex, race/ethnicity, annual household income, and highest household education level). **C** Feature importances (mean absolute SHAP values) for the cortical and subcortical features used in the prediction model. **D** Feature importances (mean absolute SHAP values) for the sociodemographic and global brain features.

Table 3. Statistics comparing AUC amongst sociodemographic groups for the cross-sectional prediction model.

	F-statistic	*p*-value	η^2
ND	0.86	0.380	0.10
Sex	0.27	0.615	0.03
Age[a]	0.11	0.754	0.01
Race/ethnicity[b]	0.39	0.813	0.07
Annual household income[c]	0.14	0.873	0.02
Highest household education[d]	0.42	0.743	0.07

ND neurodevelopmental condition.
[a]Children (<13 years of age) versus adolescents (≥13 years of age).
[b]Non-Hispanic Asian, Non-Hispanic Black, Hispanic/Latino, Multi-Racial/Other, Non-Hispanic White.
[c] < $50,000, $50,000 - $99,999, ≥$100,000.
[d]<High school, High school, Undergraduate, Graduate.

[58, 69]; POND: 79% [73, 76]), and higher accuracy was achieved in those with (73% [70, 76]) an ND compared to those without (65% [64, 66]).

AUCs were also compared amongst sociodemographic groups (age, sex, race/ethnicity, annual household income, and highest household education) to evaluate fairness (Fig. 2B); a significant difference in sex was observed ($F(1,8) = 24.9$, $p = 0.001$, $\eta^2 = 0.76$), with males having higher AUC compared to females (Table 4).

Feature importances were calculated using the magnitude of SHAP values (Fig. 2C and D). The most important features were the baseline score of internalizing problems (with fewer problems having a positive impact on prediction scores), sex (with females having a positive impact), the thicknesses of the right entorhinal cortex and fusiform, the areas of the right temporal pole and left lingual gyrus, and the volumes of the left thalamus, right lingual gyrus, anterior corpus callosum, and brainstem. Feature importance values from the longitudinal analysis were not significantly correlated with those from the cross-sectional analysis ($R = 0.02$, $p = 0.816$).

DISCUSSION

This study examined the utility of using deep learning to predict mental health problems from measures of brain structure. Cross-sectionally, good performance (mean AUC = 0.80) was achieved across a large sample ($N = 14{,}523$) of children and youth, providing insight into the neurobiology underlying internalizing problems beyond what could be revealed by traditional regression. Longitudinally ($N = 10{,}540$), performance was sub-optimal for predicting worsening trajectories of internalizing problems in children and youth from the general population (mean AUC = 0.66). However, the model performed well for an external testing

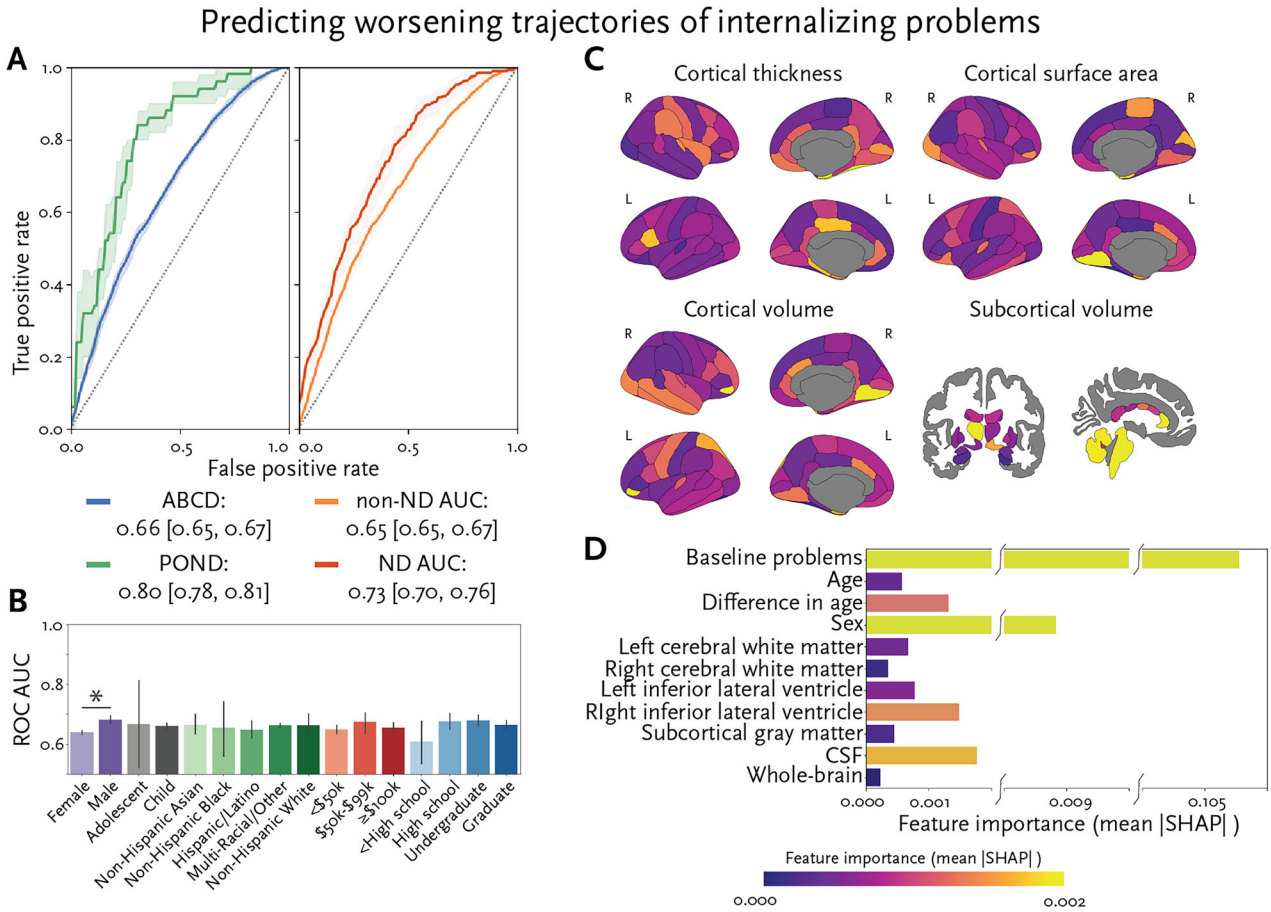

Fig. 2 Performance and feature importances for the model predicting worsening trajectories of internalizing problems. A The mean [95% CI] receiving operating characteristic curve and corresponding area under the curve (AUC) across all data from ABCD (left; blue) and POND (left; green) and stratified by individuals with (right; red) and without (right; orange) an ND diagnosis. **B** AUCs in different sociodemographic groups (age, sex, race/ethnicity, annual household income, and highest household education level). **C** Feature importances (mean absolute SHAP values) for the cortical and subcortical features used in the prediction model. **D** Feature importances (mean absolute SHAP values) for the sociodemographic and global brain features.

Table 4. Statistics comparing AUC amongst sociodemographic groups for the longitudinal prediction model.

	F-statistic	*p*-value	η^2
ND	26.7	0.001	0.77
Sex	24.9	0.001	0.76
Age[a]	0.01	0.943	0.00
Race/ethnicity[b]	0.07	0.990	0.01
Annual household income[c]	1.03	0.387	0.15
Highest household education[d]	2.10	0.141	0.28

ND neurodevelopmental condition.
[a]Children (<13 years of age) versus adolescents (≥13 years of age).
[b]Non-Hispanic Asian, Non-Hispanic Black, Hispanic/Latino, Multi-Racial/Other, Non-Hispanic White.
[c]< $50,000, $50,000 - $99,999, ≥$100,000.
[d]<High school, High school, Undergraduate, Graduate.

set composed primarily of youth with an ND (mean AUC = 0.80), and when only considering individuals across the whole sample with an ND diagnosis (mean AUC = 0.73). These results indicate that deep learning with measures of neurobiology could potentially improve our ability to detect which individuals, especially which neurodivergent individuals, are at risk of experiencing worsening trajectories of mental health, allowing for timely interventions to help mitigate the rising burden on youth, families, and healthcare systems around the world.

Deep learning was able to predict internalizing problems from cross-sectional measures of neurobiology with good performance (mean AUC = 0.80). While cross-sectional prediction does not have direct clinical relevance, it does allow for the comparison of the utility of using symptom measures, rather than diagnostic labels, as prediction targets. As the popularity of artificial intelligence has grown, there has been substantial work on its application to predicting ND conditions, such as autism and attention-deficit/hyperactivity disorder, from measures of brain structure. However, despite having distinct behaviour-based diagnostic criteria, these ND conditions are highly heterogeneous in neurobiology and phenotype, and overlap with each other, and even typical development [30, 31]. This has posed two main challenges for the existing literature. First, for clinical translation, the observed heterogeneity necessitates larger sample sizes to capture the full spectrum of variability. Yet, most studies have used small sample sizes [36]. While the diagnostic prediction accuracies of these underpowered studies have been high when predicting ND diagnosis versus typical development, accuracy was shown to decrease with sample size, with the best powered studies only achieving an accuracy of just over 60% [36]. Second, considering the heterogeneity and overlap, there is evidence that the diagnostic labels do not align with the underlying neurobiology [30, 31], which fundamentally limits work attempting to predict diagnosis from

neurobiology. Here, we have instead predicted mental health symptoms experienced across neurotypical and neurodivergent children and youth from measures of brain structure. Across this sample, we have achieved an accuracy of 77%, a substantial improvement over the existing literature, supporting the use of transdiagnostic symptom measures as prediction targets.

Comparatively, the deep learning model trained to predict worsening trajectories of internalizing problems did not perform well for a sample of children and youth representative of the general population (ABCD; mean AUC = 0.66). However, when applied to a small withheld external testing set composed primarily of children with an ND diagnosis (POND; $N = 49$), performance improved considerably (mean AUC = 0.80). While this external sample was small, this provides some indication that the model holds value for predicting worsening symptoms in individuals at high-risk of mental health symptoms. When examining performance across the individuals with an ND diagnosis across the entire sample, performance improved considerably (mean AUC = 0.73). This improvement cannot be explained by an increased representation of worsening changes in the ND group, as prevalence rates were comparable in individuals with (17%) and without (20%) an ND diagnosis. The performance difference was also not observed in the cross-sectional model, and the lack of significant difference in model performance across demographic variables (aside from sex, discussed later) suggests that sample characteristics alone are unlikely to explain this longitudinal-specific effect. Nonetheless, we acknowledge that unmeasured differences between the cross-sectional and longitudinal samples may have contributed to this pattern.

We hypothesize that biological differences associated with NDs more reliably predict subsequent mental health difficulties, potentially suggesting that mental health challenges may be part of the constellation of biological atypicality seen in NDs. There are ongoing debates about whether mental health difficulties in NDs represent shared or distinct mechanisms compared to the general population, with symptom overlap, measurement challenges, and diagnostic ambiguity obscuring the relationship between diagnosis and co-occurring mental health conditions [21, 37]. While our findings point to the neurobiological signatures of mental health trajectories being more distinguishable in NDs, it remains unclear whether these reflect shared mechanisms or unique pathways intrinsic to neurodevelopmental conditions. These complexities underscore the importance of developing tailored assessment approaches and clarifying underlying mechanisms to improve prediction and support personalized care. In addition, there may be more heterogeneity in the biological correlates of mental health trajectories in NDs; while feature variability can make initial learning challenging for a deep learning model, it can eventually lead to more general and robust performance. Finally, the increased prevalence of clinically significant internalizing problems in the NDs may be a contributing factor – worsening trajectories that extend from elevated baseline traits may have more distinct, or well-defined, neurobiological correlates compared to trajectories that arise from low baseline symptoms.

The cross-sectional model is also able to provide insight into the importance of individual neurobiological features for predicting internalizing problems. Nearly all of the most important features resided in the prefrontal cortex and included the medial orbitofrontal cortex, along with the temporal poles and corpus callosum. The role of these brain structures in internalizing disorders has been consistently documented [7, 38, 39], reflecting their roles in emotion processing, regulation, and awareness. The prefrontal cortex, in particular, has often been linked more strongly to anxiety over other domains of internalizing symptoms, such as depression [40–42], suggesting that our model may be particularly sensitive to this dimension, though this remains speculative. Furthermore, while the deep learning model relied on features already known to be associated with internalizing problems to make predictions, we have provided further insight into the multivariate complexities in these associations. For example, the importance of the lateral prefrontal cortex was specific to its surface area and volume, not thickness, and the patterns of importances throughout the brain were largely hemisphere dependent. Traditional approaches, like regression, typically search for linear brain-behaviour associations within each brain region and structural modality in isolation. The capability of deep learning models to learn from high-dimensional data allows these models to capture intricate patterns and relationships across many variables that might be challenging for traditional statistical methods. This helps advance our understanding of the neural correlates of mental health problems, which can be used to generate hypotheses for future research.

There was some overlap between features important for the cross-sectional and worsening models, for example the temporal pole and corpus callosum; however, the patterns of feature importances were strikingly different between the two models. This finding suggests that there are neuroanatomical patterns that influence internalizing problems over time, and these patterns are distinct from those that arise from the presence of clinically significant problems. In addition to brain features, baseline levels of internalizing problems were also important for predicting worsening trajectories of internalizing problems. Caution should be used in interpreting this finding through a clinical lens, given the limited range of the CBCL and the use of the reliable change index to identify worsening trajectories. Thus, this likely reflects the fact that individuals with very few baseline problems are more likely to exhibit worsening trajectories than individuals with problems already at the upper limit of the scale.

Our findings suggest that individual differences in brain structure may encode predictive information about both current and future internalizing symptoms. This indicates that structural features are not merely correlates of mental health status but may reflect underlying neurobiological vulnerabilities or compensatory mechanisms. For example, if characteristics of the prefrontal cortex predict current clinical status, this reinforces its established role in emotion processing, regulation, and awareness. Conversely, if other regions are more predictive of longitudinal symptom changes, this may point to roles in resilience, disorder progression, or neural plasticity. The absence of predictive value in certain regions also contributes to our understanding by delineating areas less involved in symptom emergence or stability. More broadly, these findings support the notion that the brain's structural architecture captures meaningful interindividual variability in mental health risk and trajectory, and they offer insight into regions that may be most relevant for early detection and intervention.

Sex was also one of the most important predictive features in both the cross-sectional and worsening models. In the cross-sectional model, the feature importances indicated that being male positively impacted the predicted probability of having internalizing problems, while being female had a negative impact. While internalizing problems have classically been reported as increased in females compared to males [43], more recent work has reported no sex differences in both neurotypical and neurodivergent populations [44]. Considering this work, we hypothesize that the positive predictive impact of males in the cross-sectional model may reflect the male preponderance in neurodivergence [29], which is in turn associated with increased internalizing problems [21]. On the other hand, in the longitudinal model, being female positively impacted the probability of predicted worsening internalizing problems, while being male had a negative impact. This is in line with evidence showing that both neurotypical and neurodivergent females are more likely to exhibit worsening longitudinal trajectories of internalizing problems across adolescence [45]. Together, the present study supports the importance of sex as an influencing factor on internalizing problems. Importantly, this study only considered

biological sex, and not gender. There is some evidence of differences in internalizing problems between gender identities (e.g., Herrmann et al. [46]), although work in this domain has been limited; there is also evidence that sex differences in internalizing problems may be caused, in part, by gender socialization, rather than biological sex [47]. While considering biological sex as a feature alongside neurobiology in the deep learning models, rather than regressing out its effects, is a step towards addressing its contribution to mental health; future work should consider measures of gender alongside measures of biological sex.

The present study provides data demonstrating that AI approaches using biological markers can be useful both at the cross-sectional level to uncover complex relationships between biology and phenotype that are not typically amenable to traditional statistical methods, as well as for predicting emerging clinically important phenotypes longitudinally. As such it is not an exhaustive search for brain-behaviour relationships during development in mental health. We only considered measures of brain structure and limited sociodemographic measures (age and sex) as predictors of cross-sectional and longitudinal internalizing problems. Other neurobiological measures, such as brain function and structural connectivity, alongside genetic, environmental, and broader socio-demographic variables, likely also hold predictive value. We also only examined internalizing problems as a prediction target, which are just one aspect of mental health; examining the utility of predicting sub-scores of internalizing problems (e.g., anxious/depressed, withdrawn/depressed, and somatic complaints), externalizing problems, or both problem domains, is an important next step. Furthermore, we measured internalizing problems using a caregiver report, which has been associated with response biases. Other limitations include that, due to the use of the ABCD dataset, the bulk of the sample's age is in the early adolescent years; this is particularly true for the longitudinal analysis, where two of the four datasets (HBN, HCP-D) cannot be used. Future work should include more younger children and older youth to ensure generalizability across development. We also did not consider the onset of the COVID-19 pandemic.

In conclusion, this study demonstrated that deep learning can be used to predict internalizing problems cross-sectionally, with performance exceeding that of comparable studies predicting ND diagnosis. We also showed that baseline measures of brain structure can predict worsening trajectories of internalizing problems, especially for children and youth with an ND. This suggests that neurobiology may offer a promising path towards identifying biomarkers of mental health trajectories in populations most vulnerable to adverse outcomes. Ultimately this type of prediction can support the delivery of proactive care aimed at preventing the escalation of mental health difficulties. While this study establishes the feasibility of this approach, future research is needed to evaluate this approach in prospective designs and real-world settings. Consideration must also be given to the availability and integration of neuroimaging within current clinical pathways. To ensure acceptability, equity, and clinical relevance, co-design with both clinicians and the neurodivergent community will be essential for translating these tools into practice. This work adds to the growing evidence that responsible applications of AI can advance personalized approaches to mental health care.

DATA AVAILABILITY

Access to the data used in this study is controlled by each of the datasets (ABCD, HBN, HCP-D, and POND). Details on the data usage agreements can be found in the Supplementary material.

CODE AVAILABILITY

Code is available on GitHub (https://github.com/marlvan/internalizing_prediction).

REFERENCES

1. Kieling C, Buchweitz C, Caye A, Silvani J, Ameis SH, Brunoni AR, et al. Worldwide prevalence and disability from mental disorders across childhood and adolescence: Evidence from the global burden of disease study. JAMA Psychiatry. 2024;81:347–56.
2. Vergunst F, Commisso M, Geoffroy MC, Temcheff C, Poirier M, Park J, et al. Association of childhood externalizing, internalizing, and comorbid symptoms with long-term economic and social outcomes. JAMA Netw. Open. 2023;6:e2249568–e2249568.
3. Stevanovic D. Impact of emotional and behavioral symptoms on quality of life in children and adolescents. Qual Life Res. 2013;22:333–7.
4. Cannon TD, Sun F, Mcewen SJ, Papademetris X, He G, van Erp TGM, et al. Reliability of neuroanatomical measurements in a multisite longitudinal study of youth at risk for psychosis. Hum Brain Mapp. 2014;35:2424–34.
5. Shen X, MacSweeney N, Chan SWY, Barbu MC, Adams MJ, Lawrie SM, et al. Brain structural associations with depression in a large early adolescent sample (the ABCD study®). EClinicalMedicine 2021; 42. https://doi.org/10.1016/j.eclinm.2021.101204.
6. Whittle S, Vijayakumar N, Simmons JG, Allen NB. Internalizing and externalizing symptoms are associated with different trajectories of cortical development during late childhood. J Am Acad Child Adolesc Psychiatry. 2020;59:177–85.
7. Schmaal L, Hibar DP, Sämann PG, Hall GB, Baune BT, Jahanshad N, et al. Cortical abnormalities in adults and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. Mol Psychiatry. 2016;22:900–9.
8. Durham EL, Jeong HJ, Moore TM, Dupont RM, Cardenas-Iniguez C, Cui Z, et al. Association of gray matter volumes with general and specific dimensions of psychopathology in children. Neuropsychopharmacology. 2021;46:1333–9.
9. Zhang Y, Xu B, Kim HH, Muetzel R, Delaney SW, Tiemeier H. Differences in cortical morphology and child internalizing or externalizing problems: Accounting for the co-occurrence. JCPP Adv. 2022;2:e12114.
10. Tejavibulya L, Rolison M, Gao S, Liang Q, Peterson H, Dadashkarimi J, et al. Predicting the future of neuroimaging predictive models in mental health. Mol Psychiatry. 2022;27:3129–37.
11. Romer AL, Ren B, Pizzagalli DA. Brain structure relations with psychopathology trajectories in the ABCD study. J Am Acad Child Adolesc Psychiatry. 2023;62:895–907.
12. Lipschutz R, Powers A, Minton ST, Stenson AF, Ely TD, Stevens JS, et al. Smaller hippocampal volume is associated with anxiety symptoms in high-risk Black youth. J. Mood Anxiety Disord. 2024;7:100065.
13. Andre QR, Geeraert BL, Lebel C. Brain structure and internalizing and externalizing behavior in typically developing children and adolescents. Brain Struct Funct. 2020;225:1369–78.
14. Muetzel RL, Blanken LME, van der Ende J, El Marroun H, Shaw P, Sudre G, et al. Tracking brain development and dimensional psychiatric symptoms in children: A longitudinal population-based neuroimaging study. Am J Psychiatry. 2018;175:54–62.
15. Jarvers I, Kandsperger S, Schleicher D, Ando A, Resch F, Koenig J, et al. The relationship between adolescents' externalizing and internalizing symptoms and brain development over a period of three years. Neuroimage Clin. 2022;36:103195.
16. Bos MGN, Peters S, Van De Kamp FC, Crone EA, Tamnes CK. Emerging depression in adolescence coincides with accelerated frontal cortical thinning. J Child Psychol Psychiatry. 2018;59:994–1002.
17. Chung J, Teo J. Mental health prediction using machine learning: taxonomy, applications, and challenges. Appl Computational Intell Soft Comput. 2022;2022:9970363.
18. Rashid B, Calhoun V. Towards a brain-based predictome of mental illness. Hum Brain Mapp. 2020;41:3468–535.
19. Chen J, Tam A, Kebets V, Orban C, Ooi LQR, Asplund CL, et al. Shared and unique brain network features predict cognitive, personality, and mental health scores in the ABCD study. Nat Commun. 2022;13:1–17.
20. Yeung HW, Stolicyn A, Buchanan CR, Tucker-Drob EM, Bastin ME, Luz S, et al. Predicting sex, age, general cognition and mental health with machine learning on brain structural connectomes. Hum Brain Mapp. 2023;44:1913–33.
21. Lai MC, Kassee C, Besney R, Bonato S, Hull L, Mandy W, et al. Prevalence of co-occurring mental health diagnoses in the autism population: a systematic review and meta-analysis. Lancet Psychiatry. 2019;6:819–29.
22. Volkow ND, Koob GF, Croyle RT, Bianchi DW, Gordon JA, Koroshetz WJ, et al. The conception of the ABCD study: From substance use to a broad NIH collaboration. Dev Cogn Neurosci. 2018;32:4–7.
23. Alexander LM, Escalera J, Ai L, Andreotti C, Febre K, Mangone A, et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. Sci Data. 2017;4:170181.
24. Somerville LH, Bookheimer SY, Buckner RL, Burgess GC, Curtiss SW, Dapretto M, et al. The Lifespan Human Connectome Project in Development: A large-scale study of brain connectivity development in 5–21 year olds. Neuroimage. 2018;183:456–68.
25. Achenbach TM, Rescorla LA. Manual for the ASEBA school-age forms & profiles: An integrated system of multi-informant assessment. Burlington, VT, VT: Aseba; 2001.

26. Jacobson NS, Truax P. Clinical significance: a statistical approach to defining meaningful change in psychotherapy research. J Consult Clin Psychol. 1991;59:12–19.
27. Fischl B. FreeSurfer. Neuroimage. 2012;62:774–81.
28. Desikan RS, Se F, Fischl B, Quinn BT, Dickerson BC, Blacker D, et al. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. Neuroimage. 2006;31:968–80.
29. Rapee RM, Oar EL, Johnco CJ, Forbes MK, Fardouly J, Magson NR, et al. Adolescent development and risk for the onset of social-emotional disorders: a review and conceptual model. Behav Res Ther. 2019;123:103501.
30. Kushki A, Anagnostou E, Hammill C, Duez P, Brian J, Iaboni A et al. Examining overlap and homogeneity in ASD, ADHD, and OCD: A data-driven, diagnosis-agnostic approach. Transl Psychiatry 2019; 9. https://doi.org/10.1038/s41398-019-0631-2.
31. Vandewouw MM, Brian J, Crosbie J, Schachar RJ, Iaboni A, Georgiades S, et al. Identifying replicable subgroups in neurodevelopmental conditions using resting-state functional magnetic resonance imaging data. JAMA Netw Open. 2023;6:e232066–e232066.
32. Morgado F, Vandewouw MM, Hammill C, Kelley E, Crosbie J, Schachar R, et al. Behaviour-correlated profiles of cerebellar-cerebral functional connectivity observed in independent neurodevelopmental disorder cohorts. Transl Psychiatry. 2024;14:1–11.
33. Kushki A, Cardy RE, Panahandeh S, Malihi M, Hammill C, Brian J, et al. Cross-diagnosis structural correlates of autistic-like social communication differences. Cereb Cortex. 2021;31:5067–76.
34. Sadat-Nejad Y, Vandewouw MM, Cardy R, Lerch J, Taylor MJ, Iaboni A, et al. Investigating heterogeneity across autism, ADHD, and typical development using measures of cortical thickness, surface area, cortical/subcortical volume, and structural covariance. Front Child Adolesc Psychiatry. 2023;2:1171337.
35. Lundberg SM, Allen PG, Lee S-I A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 2017; 30. https://github.com/slundberg/shap (accessed 26 Feb2024).
36. Bahathiq RA, Banjar H, Bamaga AK, Jarraya SK. Machine learning for autism spectrum disorder diagnosis using structural magnetic resonance imaging: Promising but challenging. Front Neuroinform. 2022;16:949926.
37. Lai MC. Mental health challenges faced by autistic people. Nat Hum Behav. 2023;7:1620–37.
38. Serra-Blasco M, Radua J, Soriano-Mas C, Gómez-Benlloch A, Porta-Casteràs D, Carulla-Roig M, et al. Structural brain correlates in major depression, anxiety disorders and post-traumatic stress disorder: a voxel-based morphometry meta-analysis. Neurosci Biobehav Rev. 2021;129:269–81.
39. Wu GR, Baeken C. Normative modeling analysis reveals corpus callosum volume changes in early and mid-to-late first episode major depression. J Affect Disord. 2023;340:10–16.
40. Peng W, Jia Z, Huang X, Lui S, Kuang W, Sweeney JA, et al. Brain structural abnormalities in emotional regulation and sensory processing regions associated with anxious depression. Prog Neuropsychopharmacol Biol Psychiatry 2019; 94. https://doi.org/10.1016/j.pnpbp.2019.109676.
41. Liang J, Yu Q, Liu Y, Qiu Y, Tang R, Yan L, et al. Gray matter abnormalities in patients with major depressive disorder and social anxiety disorder: a voxel-based meta-analysis. Brain Imaging Behav. 2023;17:749–63.
42. Liu X, Klugah-Brown B, Zhang R, Chen H, Zhang J, Becker B. Pathological fear, anxiety and negative affect exhibit distinct neurostructural signatures: evidence from psychiatric neuroimaging meta-analysis. Transl Psychiatry. 2022;12:1–19.
43. Zahn-Waxler C, Klimes-Dougan B, Slattery MJ. Internalizing problems of childhood and adolescence: prospects, pitfalls, and progress in understanding the development of anxiety and depression. Dev Psychopathol. 2000;12:443–66.
44. Mayes SD, Castagna PJ, Waschbusch DA. Sex differences in externalizing and internalizing symptoms in ADHD, autism, and general population samples. J Psychopathol Behav Assess. 2020;42:519–26.
45. Horwitz E, Vos M, De Bildt A, Greaves-Lord K, Rommelse N, Schoevers R, et al. Sex differences in the course of autistic and co-occurring psychopathological symptoms in adolescents with and without autism spectrum disorder. Autism. 2023;27:1716–29.
46. Herrmann L, Barkmann C, Bindt C, Fahrenkrug S, Breu F, Grebe J, et al. Binary and non-binary fender identities, internalizing problems, and treatment wishes among adolescents referred to a gender identity clinic in Germany. Arch Sex Behav. 2024;53:91–106.
47. Farhane-Medina NZ, Luque B, Tabernero C, Castillo-Mayén R, Factors associated with gender and sex differences in anxiety prevalence and comorbidity: a systematic review. Sci Prog 2022;105. https://doi.org/10.1177/00368504221135469/ASSET/IMAGES/LARGE/10.1177_00368504221135469-FIG1.JPEG.
48. Cardenas-Iniguez C, Gonzalez MR. Recommendations for the responsible use and communication of race and ethnicity in neuroimaging research. Nat Neurosci. 2024;27:615–28.

ACKNOWLEDGEMENTS

Funding for the current study was provided by the Canadian Institutes of Health Research and New Frontiers in Research Fund. ***ABCD***: Data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development[SM] (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA). This is a multisite, longitudinal study designed to recruit more than 10,000 children age 9-10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators. The ABCD data repository grows and changes over time. The ABCD data used in this report came from DOI: 10.15154/z563-zd24. DOIs can be found at https://nda.nih.gov/abcd. ***HBN***: This manuscript was prepared using a limited access dataset obtained from the Child Mind Institute Biobank, the Healthy Brain Network (HBN). This manuscript reflects the views of the authors and does not necessarily reflect the opinions or views of the Child Mind Institute. ***HCP-D***: Research reported in this publication was supported by the National Institute of Mental Health of the National Institutes of Health under Award Number U01MH109589 and by funds provided by the McDonnell Center for Systems Neuroscience at Washington University in St. Louis. The HCP-Development 2.0 Release data used in this report came from https://doi.org/10.15154/1520708. ***POND***: This research was conducted with the support of the Ontario Brain Institute (POND, PIs: Anagnostou/Lerch), an independent non-profit corporation, funded partially by the Ontario government. The opinions, results and conclusions are those of the authors and no endorsement by the Ontario Brain Institute is intended or should be inferred.

AUTHOR CONTRIBUTIONS

Conceptualization: MMV, AK; Data curation: MMV, BS; Formal analysis: MMV; Funding acquisition: MMV, JPL, EA, AK; Investigation: MMV; Methodology: MMV; Project administration: AK; Resources: EK, JJ, MA, PDA, JC, RJS, MJT, JPL, EA, AK; Software: MMV; Supervision: AK; Validation: MMV; Visualization: MMV; Writing – original draft: MMV; Writing – review & editing: MMV, BS, NB, AA, EK, JJ, MA, AI, PDA, JC, RJS, MJT, JPL, EA, AK.

COMPETING INTERESTS

AK has a patent for holly[TM] (formerly Anxiety Meter) with royalties paid from Awake Labs. AK has received consulting fees from DNAStack and Shaftesbury. EA has received grants from Roche and Anavex, served as a consultant to Roche, Quadrant Therapeutics, Ono, and Impel Pharmaceuticals, has received in-kind support from AMO Pharma and CRA-Simons Foundation, received royalties from APPI and Springer, received an editorial honorarium from Wiley, and has a patent for holly[TM] (formerly Anxiety Meter). The remaining authors have no potential conflicts of interest to report.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41398-025-03565-3.

Correspondence and requests for materials should be addressed to Marlee M. Vandewouw.