A precision health approach to medication management in neurodevelopmental conditions: a model development and validation study using four international cohorts

Marlee M. Vandewouw[1,2,3], Kamran Niroomand[1], Harshit Bokadia[1], Sophia Lenz[1], Jesiqua Rapley[1], Alfredo Arias[4], Jennifer Crosbie[5,6], Elisabetta Trinari[7], Elizabeth Kelley[8], Robert Nicolson[9], Russell J. Schachar[5,6], Paul D. Arnold[10], Alana Iaboni[1], Jason P. Lerch[11,12,13], Melanie Penner[1], Danielle Baribeau[1, 5,6,14,15], Evdokia Anagnostou[1,14,16], Azadeh Kushki[1,17]

1. Autism Research Center, Bloorview Research Institute, Holland Bloorview Kids Rehabilitation Hospital, Toronto, Canada
2. Center for Precision Psychiatry, Massachusetts General Hospital, Boston, United States of America
3. Department of Psychiatry, Harvard Medical School, Boston, United States of America
4. Province of Ontario Neurodevelopmental (POND) Network, Toronto, Canada
5. Department of Psychiatry, The Hospital for Sick Children, Toronto, Canada
6. Department of Psychiatry, University of Toronto, Toronto, Canada
7. Faculty of Health Sciences, McMaster University, Hamilton, Canada
8. Departments of Psychology and Psychiatry, Queen's University, Kingston, Canada
9. Department of Psychiatry, University of Western Ontario, London, Canada
10. Department of Psychiatry, Cumming School of Medicine, University of Calgary, Calgary, Canada
11. Wellcome Centre for Integrative Neuroimaging, FMRIB, Nuffield Department of Clinical Neurosciences, University of Oxford, Oxford, United Kingdom
12. Department of Medical Biophysics, University of Toronto, Toronto, Canada
13. Program in Neurosciences & Mental Health, The Hospital for Sick Children, Toronto, Canada
14. Institute of Medical Sciences, University of Toronto, Toronto, Canada
15. Institute of Health Policy, Management, and Evaluation, Dalla Lana School of Public Health, University of Toronto, Toronto, Canada
16. Department of Paediatrics, University of Toronto, Toronto, Canada
17. Institute of Biomedical Engineering, The University of Toronto, Toronto, Canada

Corresponding author:
Dr. Azadeh Kushki
Bloorview Research Institute
Holland Bloorview Kids Rehabilitation Hospital
150 Kilgour Road Toronto, Ontario M4G 1R8
Tel: 416-425-6220, extension 3563

Keywords: artificial intelligence, medication management, neurodevelopmental conditions, autism

Abstract

Psychotropic medications are commonly used for children with neurodevelopmental conditions, but their effectiveness varies, making treatment selection challenging. This study developed artificial intelligence (AI) models to predict successful use of stimulants, anti-depressants, and anti-psychotics. Cross-sectional data from research cohorts (*N*=4,758) were used to predict medication use from Child Behaviour Checklist scores, validating the feasibility and generalizability of this approach. Longitudinal prediction of medication success was then evaluated using electronic medical records from the Psychopharmacology Program (PPP; *N*=312) at Holland Bloorview Kids Rehabilitation Hospital. Ensemble models achieved strong performance, indicated by high area under the receiving operating characteristic curve (median [IQR], stimulants: 0.84 [0.81,0.88], anti-depressants: 0.82 [0.77,0.87], anti-psychotics: 0.87 [0.83,0.91]). Findings demonstrate that AI can accurately learn expert prescribing patterns and predict treatment outcomes, supporting the potential of data-driven tools to guide personalized medication management for neurodevelopmental conditions and reduce the trial-and-error burden in clinical practice.

Introduction

Neurodevelopmental conditions, such as autism spectrum disorder (autism), intellectual disability (ID), and attention-deficit/hyperactivity disorder (ADHD), are highly heterogeneous and overlapping in their biology and phenotypic presentation[1,2], which includes hundreds of associated genes[3], and significant variability in neurobiological correlates[2]. Despite decades of research, no markers exist to indicate which biological pathway may be implicated for an individual child to guide individualized care[4,5], and current approaches to care provision remain reliant on broad diagnostic labels. This lack of precision is particularly challenging for pharmacological interventions in neurodevelopmental conditions. These interventions are highly prevalent with greater than 50% of children with neurodevelopmental conditions prescribed psychotropic medications to address behavioural and/or mental health concerns (e.g., anxiety, hyperactivity, aggressive or self-injurious behavior), and high rates of polypharmacy during childhood (~30%)[6].

Despite the prevalent use of psychotropic medications in neurodevelopmental conditions, there is tremendous variability in individual-level responses to psychotropic treatment for co-occurring mental health and behavioural condition[7], reflecting the heterogeneity of biological pathways impacted, diverse clinical presentations, as well as sociodemographic complexities that influence the acceptability and uptake of medications. This variability poses a significant challenge in determining the appropriate medication for an individual child, especially in cases where co-occurring conditions (e.g., autism, ADHD, anxiety) can interact and increase the complexity of presentation and response. In these cases, prescribing practices in the children and adolescents remain almost entirely off-label[6,8–12] and based on trial-and-error approaches that can expose children to multiple medication trials with the potential for severe and life-long adverse effects for some classes of medication[13]. This highlights the critical need for precision health tools that can provide medication recommendations based on an individual's unique presentation.

To fill this gap, we propose an artificial intelligence (AI) model that uses an individual's characteristics to predict the medication choice among the commonly prescribed classes in neurodevelopmental conditions: stimulants, anti-depressants, and

anti-psychotics. We first establish the feasibility of this approach by building models to predict medication usage using a research cohort of children and youth with neurodevelopmental conditions, using two international research cohorts for external validation. Then, to examine how this approaches generalizes to real-world clinical settings, we leveraged a patient cohort to predict continuation of a medication class after 3 months from electronic medical record (EMR) data of children and youth with neurodevelopmental conditions who visited a specialized psychopharmacology clinic.

Methods

Study design and cohorts

For the research cohorts, training and internal testing was conducted using data from 598 children and youth with neurodevelopmental conditions extracted from the Province of Ontario Neurodevelopmental (POND) network (Ontario, Canada). External testing was performed using two international cohorts: 1,764 children and youth with neurodevelopmental conditions from the Healthy Brain Network (HBN[14]; New York, United States), and 2,396 children and youth with neurodevelopmental conditions from the Adolescent Brain Cognitive Development study (ABCD[15]; United States).

For clinical validation, longitudinal EMRs from the Psychopharmacology Program (PPP) at Holland Bloorview Kids Rehabilitation Hospital (Toronto, Ontario, Canada) were used. PPP sees children with neurodevelopmental conditions who have tried at least one psychoactive medication without adequate success. Data were extracted from electronic medical records of 312 children who visited the PPP over a three-year period (2019 – 2022) and returned for a follow-up approximately 3 months later.

Participant demographics are summarized in

Table 1. Further details on the cohorts, including eligibility and consent procedures, and a comparison of the research cohorts can be found in the Supplementary Material. This study was approved by the Research Ethics Board at Holland Bloorview Kids Rehabilitation Hospital.

Table 1. Cohort demographics and medication outcomes.

		Research cohorts			EMR cohort
		POND	**HBN**	**ABCD**	**PPP**
N		598	1,764	2,396	312
Age range (years)		5 – 19	5 – 17	8 – 11	4 – 17
Mean age [SD] (years)		11.4 [3.2]	10.1 [2.7]	9.9 [0.6]	10.4 [2.9]
Sex [%]	Male	424 [71]	1,213 [69]	1,523 [64]	262 [84]
	Female	174 [29]	551 [31]	873 [36]	50 [16]
Autism [%]		338 [57]	325 [18]	0 [0]	292 [94]
ADHD [%]		193 [32]	1,381 [78]	1,751 [73]	181 [58]
OCD [%]		67 [11]	90 [5]	1,048 [44]	16 [5]
Intellectual disability [%]	Yes	84 [14]	86 [5]	149 [6]	105 [34]
	No	514 [86]	1,678 [95]	2,247 [94]	135 [43]
	Unknown	0 [0]	0 [0]	0 [0]	72 [23]
Race/ ethnicity [%]	White	327 [55]	871 [49]	1,257 [52]	NA
	Minoritized	156 [26]	866 [49]	1,137 [47]	NA
	Not collected	115 [19]	27 [2]	2 [1]	312 [100]
Annual household income [%]	Low	86 [14]	307 [17]	770 [32]	NA
	Medium	98 [16]	345 [20]	608 [25]	NA
	High	170 [28]	786 [45]	798 [33]	NA
	Not collected	244 [41]	326 [18]	220 [9]	312 [100]
Outcome [%]	Stimulant	171 [29]	177 [10]	551 [23]	114 [37]
	Anti-depressant	96 [16]	72 [4]	142 [6]	82 [26]
	Anti-psychotic	52 [9]	26 [1]	43 [2]	123 [39]

EMR: electronic medical record; PPP: Psychopharmacology Program; POND: Province of Ontario Neurodevelopmental network; HBN: Healthy Brain Network; ABCD: Adolescent Brain Cognitive Development study; SD: standard deviation; ADHD: attention-deficit/hyperactivity disorder; OCD: obsessive-compulsive disorder. For POND, primary diagnosis is reported; for ABCD, HBN, and PPP, refer to supplemental.

Predictors and outcome

For the PPP cohort, the set of predictors were co-created with the clinicians at the PPP program to reflect features relevant to the medication classes considered, and the availability of variables in the electronic medical records at the first visit. These predictors were broadly grouped into five categories: demographics, clinical, family-related, child-related, and care-related. The predicted outcome was the medication decision at the follow-up visit (

Table 1).

For the research cohorts, demographics (age and sex), neurodevelopmental diagnoses (autism, ADHD, and OCD), IQ as a continuous score, and data from the Child Behaviour Checklist[16] were used as predictors. The predicted outcome was the medication class the child was on at the time of study (

Table 1). A full description of the predictors and outcome are presented in the Supplementary Material.

Model

AI models were built to predict the outcome from the set of predictors, separately for each medication class (Figure 1). For the research cohorts, the models were trained and internally tested using the POND dataset, while HBN and ABCD were withheld for external testing. By validating with two external datasets, we assessed the generalizability and robustness of the models, ensuring results are not specific to a single dataset and can be replicated across diverse populations. For the EMR cohort (PPP), only internal testing was performed.

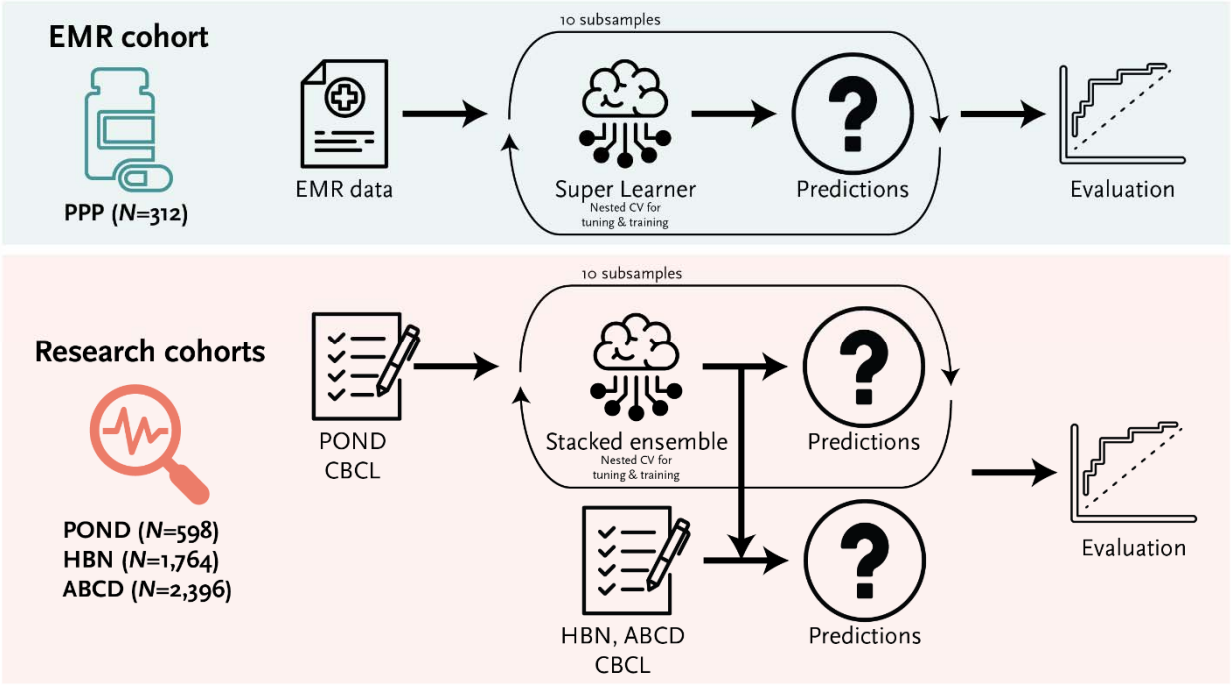

Figure 1. Study design.

To address class imbalance, we applied random under-sampling of the majority class to achieve approximately balanced 1:1 class ratios. This algorithm-agnostic approach reduces majority-class dominance and strengthens the minority-class learning signal without relying on synthetic data or class weighting. To limit dependence on any single balanced draw, we repeated the under-sampling for 10 independent iterations, noting that findings were stable across 5, 10, and 20 iterations.

For the AI models, we implemented a Super Learner approach[17], combining CatBoost gradient boosting[18] and deep neural networks as base learners, with logistic regression meta-learner trained on out-of-fold predictions to generate final predictions. We used nested cross-validation with inner 3-fold loops for hyperparameter optimization using Optuna[19] and outer 5-fold loops where base learners generated out-of-fold predictions for the logistic regression meta-learner. Further details are provided in the Supplementary material.

Performance evaluation

Model performance was assessed using the area under the receiving operating characteristic curve (AU-ROC), reporting medians and interquartile ranges (IQR) across folds and under-sampling iterations. Additional metrics (area under the precision-recall curve (AU-PR), sensitivity, and specificity) are provided in the Supplementary Material. To interpret which features were most important for predicting longitudinal medication success, Shapley Additive exPlanations (SHAP) values[20] were computed for the EMR (PPP) models. SHAP values were computed for each of the base models and combined by taking their weighted average using the coefficients from the logistic regression meta-learner.

The demographic parity and equalized odds ratios[21,22] were used to evaluate whether the model predictions were independent of sensitive group status for sex, intellectual disability (ID), race/ethnicity, and annual household income. The metric was computed at the default probability threshold (0.50) across folds, under-sampling iterations, and datasets for the research cohorts; the 80% rule was used to determine any biases[23].

Results

Model performance

The research cohorts demonstrated the feasibility of AI predicting treatment status cross-sectionally for each medication class (Table 2 and Figure 2), with internal testing (POND) demonstrating good-to-excellent AU-ROC that was significantly higher than chance for stimulants (median [IQR]: 0.75 [0.73, 0.80]), anti-depressants (0.83 [0.78, 0.87]), and anti-psychotics (0.79 [0.72, 0.86]). The performance of the external

testing sets (HBN and ABCD) indicated good generalizability, with the AU-ROCs slightly reduced but still in the good-to-excellent range.

Table 2. Model performance.

	Research cohorts			**EMR cohort**
	Internal testing	**External testing**		**Internal testing**
Medan AU-ROC [IQR]	**POND**	**HBN**	**ABCD**	**PPP**
Stimulant	.75 [.73, .80]	.70 [.67, .71]	.70 [.70, .71]	.84 [.81, .88]
Anti-depressant	.83 [.78, .87]	.75 [.74, .77]	.67 [.65, .69]	.82 [.77, .87]
Anti-psychotic	.79 [.72, .86]	.79 [.75, .81]	.81 [.79, .83]	.87 [.83, .91]

EMR: electronic medical record; PPP: Psychopharmacology Program; POND: Province of Ontario Neurodevelopmental network; HBN: Healthy Brain Network; ABCD: Adolescent Brain Cognitive Development study; CI: confidence interval; AU-ROC: area under the receiving operating characteristic curve; IQR: interquartile range

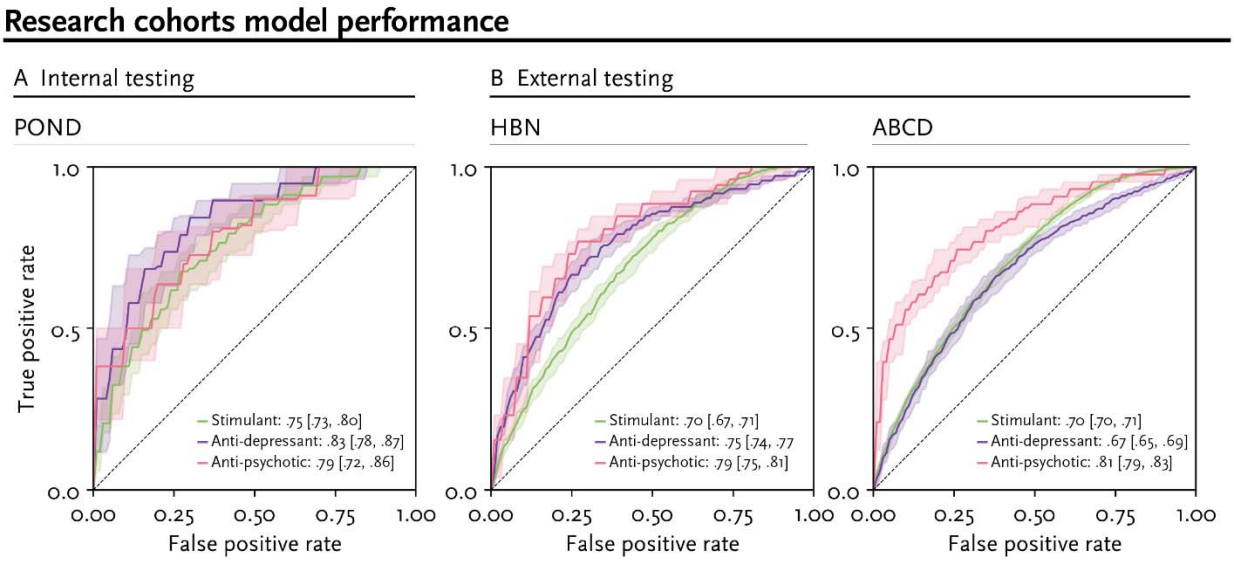

Figure 2. ROC curves for the research cohorts, for the internal testing (A; POND) and external testing (B; HBN and ABCD) sets for each of the stimulant classes (green: stimulants, purple: anti-depressants, pink: anti-psychotics); AU-ROCs are also presented.
POND: Province of Ontario Neurodevelopmental network; HBN: Healthy Brain Network; ABCD: Adolescent Brain Cognitive Development study; ROC: receiving operating characteristic curve; AU-ROC: area under the ROC

For the EMR cohort (PPP; Table 2 and Figure 3A), the AU-ROC across the cross-validation folds was excellent for anti-psychotics (0.87 [0.83, 0.91]), stimulants (0.84 [0.81, 0.88]) and anti-depressants (0.82 [0.77, 0.87]).

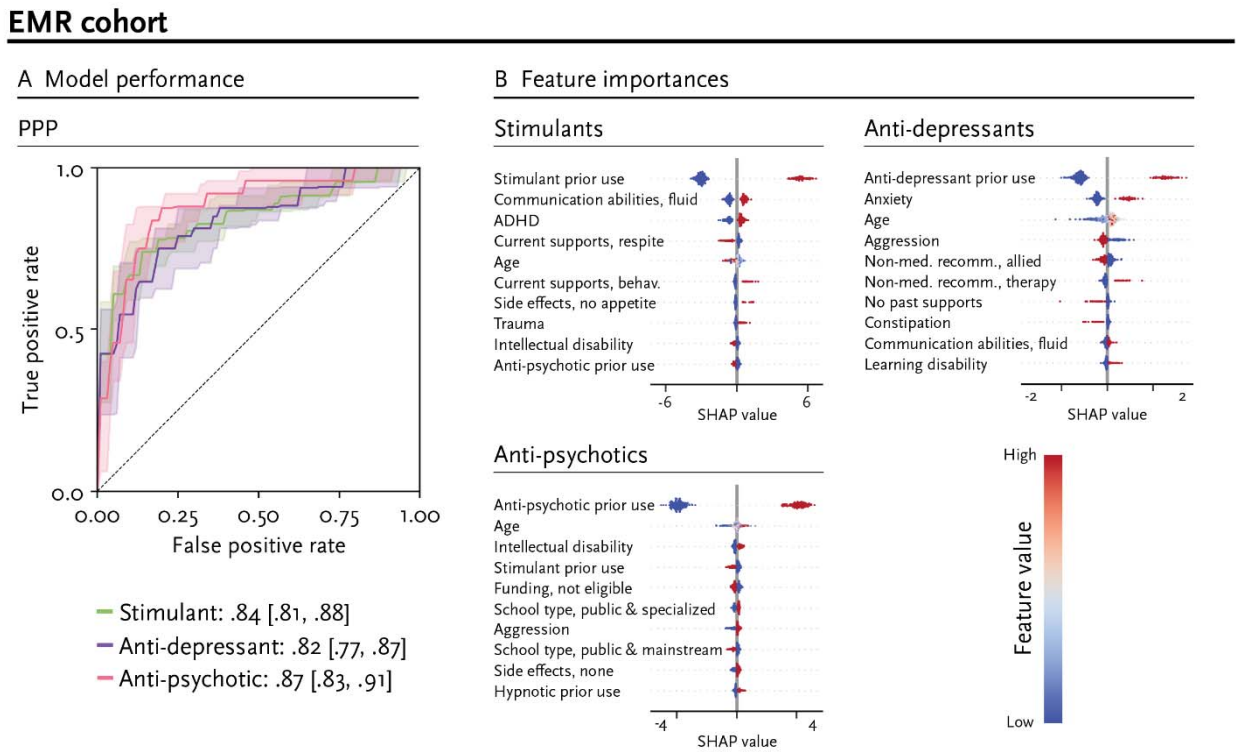

Figure 3. (A) ROC curves for the EMR cohort (PPP) for each of the stimulant classes (green: stimulants, purple: anti-depressants, pink: anti-psychotics); AU-ROCs are also presented. (B) Feature importance for the EMR cohort (PPP) assessed using SHAP values. The SHAP values are plotted for each participant for the 10 most important features for each medication class, sorted by importance (mean absolute SHAP value) from top to bottom. Colours represent the feature value (red for high values, blue for low values), indicating whether higher or lower values contribute positively or negatively to prediction scores.

EMR: electronic medical record; PPP: Psychopharmacology Program; ROC: receiving operating characteristic curve; AU-ROC: area under the ROC; SHAP: Shapley Additive exPlanations; ADHD: attention-deficit/hyperactivity disorder

Feature importance

SHAP values were computed for the EMR cohort (PPP), to identify which features were most important for predicting medication choice (Figure 3B). Prior use of a medication class was the most important feature across all four classes. In particular, prior use of a medication within the same class, increased the likelihood of a medication choice in that class. Age was also a salient feature: younger, middle, and older participants were more likely to be recommended stimulants, anti-psychotics, and anti-depressants, respectively. Other important features included ADHD and communication abilities for stimulants, anxiety and aggression for anti-depressants, and intellectual disability and prior prescription of stimulants for anti-psychotics.

Bias

The demographic parity and equalized odds ratios were computed to evaluate the bias of the models (Table 3), where values exceeding 0.80 indicate bias according to the four-fifths rule[24]. For the research cohorts, parity was mixed. Individuals who were female, and who did not have an intellectual disability (ID) were more likely to be recommended a stimulant (male, female: 39%, 56%; ID, no-ID: 37%, 52%). Individuals who were male, did not have an ID, who were white, and who were of high income class were more likely to be recommended an anti-depressant (male, female: 40%, 27%; ID, no-ID: 23%, 32%; white, minority: 38%, 22%; low, middle, high income: 31%, 25%, 35%). On the other hand, those with an ID, and those with a lower annual household income were more likely to be recommended anti-psychotics (ID, no-ID: 54%, 31%; low, middle, high income: 45%, 22%, 30%). These observations extended to the EMR cohort, with the exception of sex, with males being more likely to be recommended stimulants (male, female: 45%, 36%) and females being more likely to be recommended anti-depressants (male, female: 44%, 55%). This was in line with trends in the data.

Table 3. Demographic parity ratio (DPR) and equalized odds ratio (EOR) for sensitive groups.

		Research cohorts		**EMR cohort**	
		DPR	**EOR**	**DPR**	**EOR**
Sex	Stimulant	0.70	0.70	0.79	0.49
	Anti-depressant	0.69	0.68	0.79	0.72
	Anti-psychotic	0.81	0.83	0.91	0.81
Intellectual disability	Stimulant	0.72	0.63	0.45	0.29
	Anti-depressant	0.71	0.65	0.62	0.41
	Anti-psychotic	0.57	0.74	0.62	0.59
Race/ethnicity	Stimulant	0.95	0.94	–	
	Anti-depressant	0.57	0.71	–	
	Anti-psychotic	0.83	0.90	–	
Annual household income	Stimulant	0.94	0.85	–	
	Anti-depressant	0.72	0.77	–	
	Anti-psychotic	0.50	0.56	–	

EMR: electronic medical record

Discussion

Medication management for children with neurodevelopmental conditions remains a significant challenge, particularly in community settings where subspecialty expertise may not be available. To address this gap, this study aimed to develop and evaluate an AI algorithm to predict medication choices in neurodevelopmental conditions. Ultimately, this type of tool could be used as a decision-aid to support medication management where specialized expertise is not available.

We showed that AI can learn expert clinician medication decisions with high accuracy. We also demonstrated that the results generalize across independent, international datasets, which is notable given the drastically different characteristics of these cohorts, in terms of demographics, diagnosis, phenotype, and region-specific access to service and supports (see Supplemental Table 1 and Supplemental Table 2). While the accuracy in these research cohorts is lower compared to the EMR model, this is likely due to the cross-sectional nature of the prediction in these cohorts. For example, use of stimulant medications to target ADHD behaviours confounds the measures of ADHD traits when considered cross-sectionally, in turn impacting predictive performance.

Overall, these results support the feasibility of using an AI-based decision aid tool. This can be particularly advantageous for community practitioners, where subspecialty expertise required for medication management is often not available. Ultimately, an AI-based decision aid can bring precision to current prescribing practices in the community which may expose children to multiple medication trials that may have severe and life-long adverse effects[25]. Prospective trials are needed to further validate this type of algorithms in clinical settings.

Explainability and interpretability are critical features of responsible AI systems, contributing to enhancing trust and transparency with both clinicians and patients[26]. To promote explainability, we characterized the relative contributions of clinical and demographic features to the AI algorithm's medication recommendations. For the EMR cohort, these contributions were closely aligned with indications for each medication class (e.g., stimulants for ADHD, SSRIs for anxiety). Interestingly, the top contributors

for all classes were prior use of a medication in that class. This finding reflects the fact that medication choices often remain within the same class of medication but entail changes in the medication type and dosing within that class (e.g., amphetamines, methylphenidate within the stimulant class). Future work will focus on adding this type of resolution to the AI predictions. Other contributing features to medication decision reflected sociodemographic and environmental factors that may contribute to clinical decision making beyond primary indications. For example, school environment and funding eligibility were significant contributors to the decision to recommend an antipsychotic. These finding highlight the complexity of the clinical and sociodemographic factors that contribute to value-laden decisions around medication management for children with neurodevelopmental conditions.

Another aspect of responsible AI is potential algorithmic biases that may disproportionately disadvantage subgroups of individuals. Our analysis revealed several biases including increased recommendation of antipsychotics to children who had intellectual disability and were from lower income households. On the other hand, anti-depressants recommendations were increased for children who did not have an intellectual disability, and who were white. It is important to note that these biases are not related to any specific feature of the AI algorithm, but mirror previously reported biases in prescription patterns which exist across the system[27]. For example, decreased access to early diagnosis and behavioural interventions for children from lower income household may contribute to a higher likelihood of crisis situations that necessitate prescription of antipsychotic medications. Similarly, it is well established that depression can be challenging to detect in individuals with an intellectual disability[28]. While computational approaches to bias mitigation can be applied to equalize predicted prescription rates across groups, these will not solve the underlying system-level inequities that impact children with marginalized identities.

The results of this study should be interpreted considering several limitations. Firstly, for the longitudinal prediction, we defined outcome as the medication choice at the follow-up visit. Treatment success in neurodevelopmental conditions is a complex, subjective, and value-laden construct. The use of medication choice was co-created

with clinicians as an outcome as it reflects the aggregate effects of the multitude of factors contributing to a clinical decision to not stop a medication (e.g., ineffectiveness, side-effects, acceptability or other factors). These children also visited the specialized clinic after at least one psychoactive medication was tried without success, and thus this may introduce biases when the models are used to make a first-line medication decision. Although gender identity and race/ethnicity were not included as predictors due to incomplete capture in the primary EMR cohort, their omission may contribute to model misspecification bias. These variables remain important contextual factors for understanding medication access and prescribing patterns. Future work should incorporate reliably captured measures and evaluate model performance across gender and racial/ethnic groups, alongside fairness assessments, updating the model as needed to mitigate potential inequities. The models were trained using data exclusively from North America; future work should investigate their translation to other global populations to account for potential differences in healthcare systems, prescribing practices, and sociocultural influences on medication usage. Finally, this study exclusively focused on predicting medication classes; these classes are broad, and future work should examine within-class predictability.

In summary, the present study focused on AI model development and evaluation, demonstrating the feasibility of this approach for medication recommendation for children with neurodevelopmental conditions. To enable effective clinical implementation of this type of AI-tool, future studies are needed to evaluate the algorithm in a prospective design at the point of care as well as acceptability, and barriers and facilitators to uptake.

Declaration of interests

EA has received research funding from Roche and Anavex, consulting fees from Roche, Quadrant Therapeutics, Ono, and Impel Pharmaceuticals, in-kind support from AMO Pharma and CRA-Simons Foundation, royalties from APPI and Springer, an editorial honorarium from Wiley, and has a patent for holly™. DB has received research funding from MapLight Therapeutics. MP has received consulting fee from Addis & Associates/Roche, and MedCounsel. AK has a patent for holly™ with royalties paid from Awake Labs and has received consulting fees from DNAStack and Shaftesbury. The remaining authors have no conflicts to report.

Acknowledgements

Funding was provided by the Canadian Institutes of Health Research (Operating Grant #527447). Acknowledgements for each of the research cohorts (POND, HBN, and ABCD) used in this study are provided below.

POND: This research was conducted with the support of the Ontario Brain Institute (POND, PIs: Dr. Evdokia Anagnostou and Dr. Jason P. 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.

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.

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

Data and code availability statement

Data from four cohorts were used in this study, each governed by their own data sharing restrictions; full details can be found in the Supplementary Material. Code used in this manuscript is available at GitHub (https://github.com/marlvan/med_prediction).

Supplementary material

PPP

Dataset description

The clinical cohort used for this study was longitudinal electronic medical records from the Psychopharmacology Program (PPP) at Holland Bloorview Kids Rehabilitation Hospital (Toronto, Canada). The PPP is a program with provincial catchment in Ontario that provides medication management for children and youth with neurodevelopmental conditions. Children were eligible for this study if (1) their first visit was between 2019-2022, (2) they were at least four years of age at their first visit, and (3) they returned to the clinic for a follow-up visit. As the data was extracted from electronic medical records, no consent was obtained. Data from PPP is not publicly available due to ethics restrictions. Clinical data from all longitudinal clinical visits were extracted into a RedCap database from semi-structured medical record templates by trained chart reviewers. Two chart reviewers co-reviewed each medical record in sets of five and reviewed discrepancies until 85% agreement on all items was achieved. Reviewers proceeded independently after that, and any uncertainties were discussed with the study team.

Predictors

For PPP, the predictors were grouped into five categories: demographics, clinical, family-related, child-related, and care-related. Categorical variables with unknown or undocumented values were dummy-coded with ‘Unknown’ as the implicit reference category. This retains all participants and prevents missing data from being conflated with a substantive level. Missingness ranged between 0 and 28% across predictors, and 65% of participants had at least one missing predictor.

Demographic variables included age, sex (male or female), and region (urban (Toronto, General Toronto Area, or other Ontario city) or rural (rural or northern community)). Due to missing data, gender and race/ethnicity could not included as predictors. To capture socioeconomic status, whether a parent was receiving social assistance or means-tested disability support as a source of income was included as a predictor.

Clinical variables included neurodevelopmental diagnoses (attention-deficit/hyperactivity disorder (ADHD), autism, obsessive-compulsive disorder (OCD), intellectual disability, learning disorder, other psychiatric condition), mental health symptoms (anxiety, depression, irritability), and chronic conditions (epilepsy, cerebral palsy, asthma, migraines, gastroesophageal reflux disease, persisting constipation, obesity, and other).

Family-related factors included a family history of psychiatric or developmental diagnoses in a first degree relative, and family reports of conflict between family members in the household, caregiver burnout, and financial barriers to care, as reported by families and documented by the treating clinician in the child's chart.

Child-related factors included school type (public, mainstream classroom, private, mainstream classroom, public, specialized classroom, private, specialized classroom, homeschool, or other), school attendance (full days, part days or withdrawal, frequent calls home for pick-up, frequent school refusal or frequent absences, not attending or demitted from school, or other), communication abilities (speaks fluidly in sentences, speaks a few words or short phrases, non-speaking and uses communication technique, non-speaking and no other formal communication form, and other). History of trauma (children's aid involvement, witnessing violence, experiencing violence, physical abuse, sexual abuse, or neglect, other significant stressor as perceived by the child, or other stressor), police involvement, hospitalization, emergency department visits, intensive crisis service involvement due to behavioural or mental health concerns, aggression, self-injury, and suicidal thoughts were also included.

Care-related factors included receiving government funding for care (specifically, the Ontario Assistance for Children with Severe Disabilities Program), which was coded as "yes", "has not applied", "not eligible", or "other". Data on current, past, and recommended supports were collected using the following categories: (1) behavioural therapy, Applied Behaviour Analysis (ABA), or parent coaching by a behaviour therapist, (2) child therapy (e.g., Cognitive Behavioural Therapy (CBT) or group therapy), (3) case management, (4) allied health, (5) respite, (6) other, or (7) none. Prior use of five commonly prescribed medication classes (stimulants, non-stimulants, hypnotics, anti-depressants, and anti-psychotics) was determined as whether a child

was taking a medication at the time of the first visit to the clinic. The behaviour that was targeted for treatment at the first visit was collected using the following categories: physical aggression, swearing/verbal threats, elopement, self-injury, verbal refusal, repetitive behaviours, property destruction, tantrums/meltdowns, and/or other. Data on side effects reported at the first visit were categorized as: headache, stomach-ache, constipation, nausea/vomiting, weight gain, weight loss, irritability, drowsy/fatigue/sedation, insomnia, appetite decrease, appetite increase, other, and/or none. Whether adherence problems for medications were discussed at the first visit was also included as a predictor.

Outcome

The predicted outcome was the medication class the child was on at the follow-up visit.

Bias

Bias was assessed on sex (male versus female) and intellectual disability (yes versus no).

POND

Dataset description

The POND network is a research network spanning several institutions across Ontario, Canada that collects demographic, phenotypic, and biological data from children and youth with and without neurodevelopmental conditions. Participants were eligible for the current study if they received a diagnosis of autism, ADHD, or OCD, if they were between 5 and 19 years of age, and if diagnostic, medication usage, IQ, and the Child Behaviour Checklist (CBCL[16]) was available. Children and youth provided written informed consent if they were able; otherwise, they provided verbal assent, and their caregiver provided written consent. Data from POND is managed by the Ontario Brain Institute and is available through a controlled data release. All requests for access should be directed to the Brain-CODE platform (https://www.braincode.ca/content/controlled-data-releases#dr006).

Predictors

Predictors for POND included demographics (age and sex), diagnosis (autism, ADHD, or OCD), IQ, and the CBCL. Diagnoses were confirmed with expert clinical

judgement alongside the Autism Diagnostic Observation Schedule, 2 (ADOS-2[29]) and Autism Diagnostic Interview, Revised (ADI-R[30]) for autism, the Parent Interview for Child Symptoms (PI/orCS[31]) and/or the Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS[32]) for ADHD, and the Children's Yale-Brown Obsessive-Compulsive Scale (YBOCS[33]) for OCD. Full-scale IQ was collected using age-appropriate Wechsler[34–38] or Standford-Binet[39] scales. All the individual questions, syndromic scale *T*-scores (anxious/depressed, withdrawn/depressed, somatic complaints, social problems, thought problems, attention problems, rule-breaking behaviour, and aggressive behaviour), and summary *T*-scores (internalizing problems, externalizing problems, and total problems) from the CBCL were included as predictors.

Outcome

Caregivers were asked to provide a list of medication their child was taking. These medications were classified as stimulants, non-stimulants, anti-depressants, and anti-psychotics with assistance from clinicians. The predicted outcome was the medication class(es) the child was on.

Bias

Bias was assessed for sex, intellectual disability, race/ethnicity, and socioeconomic status. For sex, biological sex (male versus female) was used. Intellectual disability (yes versus no) was determined using the full-scale IQ cutoff of 70. For race/ethnicity, participants/caregivers were asked to select all groups that applied: Aboriginal, Arab, Black, Chinese, East Asian, Filipino, Japanese, Korean, Latin American/Hispanic, South Asian, Southeast Asian, West Asian, White, other; this data was used to categorize individuals as white or minoritized. For socioeconomic status, we used annual household income categorized as low (<$50,000), middle ($50,000-$99,999), and high (≥$100,000).

HBN

Dataset description

HBN is a mental health study of children and adolescents across the New York city (United States) area that has collected demographic, phenotypic, and biological data; families who had psychiatric concerns about their child were recruited for the study. Participants were eligible for the current study if they received a diagnosis of

autism, ADHD, or OCD, if they were between 5 and 19 years of age, and if diagnostic, medication usage, IQ, and the CBCL[16] was available. Participants at least 18 years of age provided written informed consent; otherwise, they provided written assent, and their caregiver provided written consent. Access to the data used in this study is protected by a Data Usage Agreement (DUA; see https://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network for more details), which was awarded to AK for this study.

Predictors

Aligning with POND, predictors for HBN included demographics (age and sex), diagnosis (autism, ADHD, and/or OCD), full-scale IQ, and the CBCL. For diagnosis, the consensus DSM-5[40] diagnoses provided by the HBN clinical team were used. Full-scale IQ was assessed using Wechsler scales[37,38] or the Kaufman Brief Intelligence Test[41].

Outcome

Caregivers were asked to provide a list of medication their child had taken the day of the visit. These medications were classified as stimulants, non-stimulants, anti-depressants, and anti-psychotics with assistance from clinicians. The predicted outcome was the medication class(es) the child was on.

Bias

Bias was assessed on sex, intellectual disability, race/ethnicity, and socioeconomic status. For sex, biological sex (male versus female) was used. Intellectual disability (yes versus no) was determined using the full-scale IQ cutoff of 70. For race/ethnicity, participants/caregivers indicated their ethnicity (not Hispanic or Latino, Hispanic or Latino, decline to specify, or unknown) and race (White/Caucasian, Black/African American, Hispanic, Asian, Indian, Native American Indian, American Indian/Alaskan Native, Native Hawaiian/Other Pacific Islander, two or more races, or other race), which were used to categorize individuals as white or minoritized. For socioeconomic status, we used annual household income categorized as low (<$50,000), middle ($50,000-$99,999), and high (≥$100,000).

ABCD

Dataset description

The ABCD study[42] is a longitudinal study of adolescents from the general population across the United States that has collected demographic, phenotypic, and biological data. For the purposes of this study, only the first time point (collected at approximately 9-11 years of age) was used. Participants were eligible for the current study if medication usage, IQ, the Kiddie Schedule for Affective Disorders and Schizophrenia (K-SADS), and the CBCL was collected. Caregivers provided written informed consent, and participants provided written assent. Access to the data used in this study is protected by an NIMH Data Archive DUA for the Adolescent Brain Cognitive Development (ABCD) permission group (see https://nda.nih.gov/user/dashboard/data_permissions.html for more details), which was awarded under DUA #15619 for this study.

Predictors

Aligning with POND, predictors for ABCD included demographics (age and sex), diagnosis (autism, ADHD, and/or OCD), IQ, and the CBCL. Diagnoses were assigned using a computerized version of the K-SADS[43] completed by parents and/or youth. IQ was assessed using the NIH Toolbox[44] Cognition Battery's age-corrected total cognition composite score, which shows a high degree of alignment with the Wechsler scales used in POND[45].

Outcome

Caregivers were asked to provide a list of medication their child had taken in the two weeks preceding the visit. These medications were classified as stimulants, non-stimulants, anti-depressants, and anti-psychotics with assistance from clinicians. The predicted outcome was the medication class(es) the child was on.

Bias

Bias will be assessed on sex, intellectual disability, race/ethnicity, and socioeconomic status. For sex, biological sex (male versus female) was used. Intellectual disability (yes versus no) was determined using the IQ cutoff of 70. For race/ethnicity, participants/caregivers indicated their race/ethnicity (Hispanic/Latino, non-Hispanic Asian, non-Hispanic Black, non-Hispanic White, and Multi-Racial/Other), which were used to categorize individuals as white or minoritized. For socioeconomic

status, we used annual household income categorized as low (<$50,000), middle ($50,000-$99,999), and high (≥$100,000).

AI models

We implemented a Super Learner approach[17] in Python (version 3.12), combining CatBoost gradient boosting[18] (version 1.2.7) and deep neural networks (Keras[46] (version 3.7.0) with TensorFlow[47] (version 2.18.0)) as base learners, with a logistic regression model trained on out-of-fold predictions to generate final predictions.

Hyperparameters for both base learners were optimized using Optuna's Tree-structured Parzen Estimator[19] with 100 trials and nested 3-fold cross-validation within each training fold, with both models optimizing AU-ROC. The deep neural network architecture was tuned across 1-3 hidden layers, 48-96 units per layer, and dropout rates of 0.20-0.40. Networks used ReLU activation and the Adam optimizer, with learning rates tuned between 3×10^{-4} to 3×10^{-3}. Training used a fixed batch size of 32 with early stopping (patience=12) monitoring AU-ROC over a maximum of 200 epochs. CatBoost models were configured with 50-400 iterations, learning rates of 0.01-0.30, tree depths of 3-6, and early stopping after 200 rounds without improvement.

The Super Learner consisted of a logistic regression model (L2 penalty = 1.0, liblinear solver) to combine predicted probabilities from both base learners. We employed nested cross-validation, with inner 3-fold loops for hyperparameter optimization and outer 5-fold loops in which base learners trained on 4 folds generated out-of-fold predictions for the logistic regression model. This design followed standard Super Learner methodology to prevent overfitting, avoid data leakage, and yield unbiased performance estimates.

Comparison of research cohorts

The demographic variables (age, sex, diagnosis, intellectual disability, race/ethnicity, and socioeconomic status) were compared amongst the research cohorts (Supplemental Table 1). Chi-squared tests were used for categorical variables, holding significance at $p<.050$ and reporting Cramer's *v* effect sizes. Shapiro-Wilks tests indicated that all continuous variables were non-normally distributed; as such, Kruskal-Wallis tests were used, holding significance at $p<.050$, and reporting eta-squared effect sizes.

Supplemental Table 1. Demographic comparison of the research cohorts.

Measure	*p*-value	Effect size	Post-hocs
Age	<.0001	.05	ABCD > POND > HBN
Sex	.0001	.06	Females: ABCD > POND, HBN
Autism	<.0001	.53	POND > HBN > ABCD
ADHD	<.0001	.32	HBN > ABCD > POND
OCD	<.0001	.43	ABCD > POND > HBN
Intellectual disability	<.0001	.11	POND > HBN, ABCD
Race/ethnicity	<.0001	.10	White: POND > HBN, ABCD
Annual household income	<.0001	.13	HBN > POND > ABCD

POND: Province of Ontario Neurodevelopmental network; HBN: Healthy Brain Network; ABCD: Adolescent Brain Cognitive Development study; ADHD: attention-deficit/hyperactivity disorder; OCD: obsessive-compulsive disorder

The CBCL summary *T*-scores were also compared amongst the datasets (Supplemental Table 33), using Kruskal-Wallis tests due to non-normality of all variables (Shapiro-Wilks tests), holding significance at $p<.050$, and reporting eta-squared effect sizes.

Supplemental Table 2. CBCL comparison of the research cohorts.

Measure	Median [IQR] ABCD	HBN	POND	*p*-value	Effect size	Post-hocs
Anxious/depressed	54 [50, 63]	55 [50, 64]	62 [53, 68]	<.0001	.03	POND > HBN, ABCD
Withdrawn/depressed	54 [60, 52]	56 [50, 64]	60 [54, 68]	<.0001	.04	POND > HBN > ABCD
Somatic complaints	57 [50, 64]	56 [50, 61]	57 [53, 64]	.0004	<.01	ABCD > POND, POND > HBN
Social problems	54 [51, 60]	58 [51, 64]	61 [54, 68]	<.0001	.06	POND > HBN > ABCD
Thought problems	58 [51, 64]	58 [51, 67]	66 [58, 71]	<.0001	.05	POND > HBN, ABCD
Attention problems	59 [53, 60]	64 [57, 69]	65 [59, 71]	<.0001	.05	POND > HBN > ABCD
Rule-Breaking behaviour	53 [50, 60]	53 [51, 60]	54 [51, 62]	.0008	<.01	ABCD > HBN, POND > ABCD
Aggressive behaviour	54 [50, 62]	57 [50, 66]	58 [51, 66]	<.0001	.02	POND > HBN > ABCD
Internalizing problems	56 [48, 64]	57 [50, 65]	63 [55, 69]	<.0001	.03	POND > HBN > ABCD
Externalizing problems	54 [46, 62]	56 [48, 65]	58 [50, 66]	<.0001	.01	POND > HBN > ABCD
Total problems	56 [50, 64]	60 [52, 68]	64 [56, 70]	<.0001	.04	POND > HBN > ABCD

CBCL: Child Behaviour Checklist; POND: Province of Ontario Neurodevelopmental network; HBN: Healthy Brain Network; ABCD: Adolescent Brain Cognitive Development study

Performance evaluation

Additional metrics are provided in the Supplemental Table 3, including the area under the precision–recall curve (AU-PR; where recall = sensitivity and precision = positive predictive value [PPV]), along with sensitivity and specificity.

Supplemental Table 33. Model performance.

	Research cohorts			Clinical cohort
	Internal testing	External testing		Internal testing
	POND	HBN	ABCD	PPP
Median AU-PR [IQR]				
Stimulant	.70 [.67, .72]	.68 [.68, .69]	.65 [.65, .66]	.86 [.86, .87]
Anti-depressant	.82 [.79, .84]	.75 [.74, .76]	.66 [.65, .67]	.84 [.82, .85]

Anti-psychotic	.81 [.78, .84]	.79 [.77, .81]	.81 [.80, .82]	.90 [.89, .92]
Median sensitivity [IQR]				
Stimulant	.72 [.70, .73]	.82 [.80, .83]	.77 [.76, .79]	.70 [.68, .72]
Anti-depressant	.73 [.70, .76]	.51 [.49, .52]	.44 [.41, .46]	.69 [.66, .72]
Anti-psychotic	.71 [.66, .76]	.60 [.57, .64]	.58 [.54, .61]	.83 [.68, .72]
Medan specificity [IQR]				
Stimulant	.63 [.61, .65]	.46 [.44, .47]	.53 [.51, .54]	.78 [.77, .80]
Anti-depressant	.76 [.73, .79]	.83 [.82, .85]	.76 [.73, .78]	.80 [.78, .83]
Anti-psychotic	.71 [.66, .76]	.80 [.77, .84]	.86 [.84, .89]	.90 [.89, .92]

PPP: Psychopharmacology Program; POND: Province of Ontario Neurodevelopmental network; HBN: Healthy Brain Network; ABCD: Adolescent Brain Cognitive Development study; CI: confidence interval; AU-ROC: area under the receiving operating characteristic curve

References

1. Astle, D. E., Holmes, J., Kievit, R. & Gathercole, S. E. Annual Research Review: The transdiagnostic revolution in neurodevelopmental disorders. *Journal of Child Psychology and Psychiatry* **n/a**, (2021).

2. Vandewouw, M. M. *et al.* Identifying Replicable Subgroups in Neurodevelopmental Conditions Using Resting-State Functional Magnetic Resonance Imaging Data. *JAMA Network Open* **6**, e232066 (2023).

3. Trost, B. *et al. Genomic Architecture of Autism Spectrum Disorder from Comprehensive Whole-Genome Sequence Annotation*. http://medrxiv.org/lookup/doi/10.1101/2022.05.05.22274031 (2022) doi:10.1101/2022.05.05.22274031.

4. Persico, A. M. *et al.* The pediatric psychopharmacology of autism spectrum disorder: A systematic review - Part I: The past and the present. *Prog Neuropsychopharmacol Biol Psychiatry* **110**, 110326 (2021).

5. Lord, C. *et al.* The Lancet Commission on the future of care and clinical research in autism. *Lancet* **399**, 271–334 (2022).

6. Feroe, A. G. *et al.* Medication Use in the Management of Comorbidities Among Individuals With Autism Spectrum Disorder From a Large Nationwide Insurance Database. *JAMA Pediatrics* **175**, 957–965 (2021).

7. Baribeau, D., Vorstman, J. & Anagnostou, E. Novel treatments in autism spectrum disorder. *Curr Opin Psychiatry* **35**, 101–110 (2022).

8. Murray, M. L. *et al.* Pharmacological treatments prescribed to people with autism spectrum disorder (ASD) in primary health care. *Psychopharmacology (Berl)* **231**, 1011–1021 (2014).

9. Lake, J. K., Weiss, J. A., Dergal, J. & Lunsky, Y. Child, parent, and service predictors of psychotropic polypharmacy among adolescents and young adults with an autism spectrum disorder. *J Child Adolesc Psychopharmacol* **24**, 486–493 (2014).

10. Spencer, D. *et al.* Psychotropic medication use and polypharmacy in children with autism spectrum disorders. *Pediatrics* **132**, 833–840 (2013).

11. Mellahn, O. J. *et al.* Understanding the Diversity of Pharmacotherapeutic Management of ADHD With Co-occurring Autism: An Australian Cross-Sectional Survey. *Front Psychiatry* **13**, 914668 (2022).

12. Ritter, C., Hewitt, K. & McMorris, C. A. Psychotropic Polypharmacy Among Children and Youth with Autism: A Systematic Review. *J Child Adolesc Psychopharmacol* **31**, 244–258 (2021).

13. Howes, O. D. *et al.* Autism spectrum disorder: Consensus guidelines on assessment, treatment and research from the British Association for Psychopharmacology. *J Psychopharmacol* **32**, 3–29 (2018).

14. Alexander, L. M. *et al.* An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Sci Data* **4**, 170181 (2017).

15. Barch, D. M. *et al.* Demographic, physical and mental health assessments in the adolescent brain and cognitive development study: Rationale and description. *Dev Cogn Neurosci* **32**, 55–66 (2018).

16. Achenbach, T. M. & Rescorla, L. A. *Manual for the ASEBA School-Age Forms and Profiles*. (University of Vermont Research Center for Children, Youth, and Families, Burlington, VT, 2001).

17. van der Laan, M. J., Polley, E. C. & Hubbard, A. E. Super learner. *Stat Appl Genet Mol Biol* **6**, Article25 (2007).

18. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. & Gulin, A. CatBoost: unbiased boosting with categorical features. in *Proceedings of the 32nd International Conference on Neural Information Processing Systems* 6639–6649 (Curran Associates Inc., Red Hook, NY, USA, 2018).

19. Xu, L., Liu, S., Lin, A., Su, Z. & Liang, D. Interpretable Prediction and Analysis of PVA Hydrogel Mechanical Behavior Using Machine Learning. *Gels* **11**, 550 (2025).

20. Lundberg, S. M. & Lee, S.-I. A unified approach to interpreting model predictions. in *Proceedings of the 31st International Conference on Neural Information Processing Systems* 4768–4777 (Curran Associates Inc., Red Hook, NY, USA, 2017).

21. Bird, S. *et al. Fairlearn: A Toolkit for Assessing and Improving Fairness in AI*. https://www.microsoft.com/en-us/research/publication/fairlearn-a-toolkit-for-assessing-and-improving-fairness-in-ai/ (2020).

22. Agarwal, A., Beygelzimer, A., Dudik, M., Langford, J. & Wallach, H. A Reductions Approach to Fair Classification. in *Proceedings of the 35th International Conference on Machine Learning* (eds Dy, J. & Krause, A.) vol. 80 60–69 (PMLR, 2018).

23. Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C. & Venkatasubramanian, S. Certifying and Removing Disparate Impact. in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* 259–268 (Association for Computing Machinery, New York, NY, USA, 2015). doi:10.1145/2783258.2783311.

24. Holzer, H. & Neumark, D. Assessing Affirmative Action. *Journal of Economic Literature* **38**, 483–568 (2000).

25. Howes, O. D. *et al.* Autism spectrum disorder: Consensus guidelines on assessment, treatment and research from the British Association for Psychopharmacology. *J Psychopharmacol* **32**, 3–29 (2018).

26. Cutillo, C. M. *et al.* Machine intelligence in healthcare-perspectives on trustworthiness, explainability, usability, and transparency. *NPJ Digit Med* **3**, 47 (2020).

27. Antoniou, T. *et al.* Geographic variation and sociodemographic correlates of prescription psychotropic drug use among children and youth in Ontario, Canada: a population-based study. *BMC Public Health* **23**, 85 (2023).

28. Buckley, N. *et al.* Prevalence estimates of mental health problems in children and adolescents with intellectual disability: A systematic review and meta-analysis. *Aust N Z J Psychiatry* **54**, 970–984 (2020).

29. Lord, C. *et al. Autism Diagnostic Observation Schedule: ADOS-2*. (Western Psychological Services, Los Angeles, CA, 2012).

30. Rutter, M., Le Couteur, A. & Lord, C. *Autism Diagnostic Interview Revised*. (Western Psychological Services, Los Angeles, CA, 2003).

31. Ickowicz, A. *et al.* The parent interview for child symptoms: a situation-specific clinical research interview for attention-deficit hyperactivity and related disorders. *Can J Psychiatry* **51**, 325–328 (2006).

32. Kaufman, J. *et al.* Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version (K-SADS-PL): Initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry* **36**, 980–988 (1997).

33. Scahill, L. *et al.* Children's Yale-Brown Obsessive Compulsive Scale: Reliability and validity. *J Am Acad Child Adolesc Psychiatry* **36**, 844–852 (1997).

34. Wechsler, D. *Wechsler Abbreviated Scale of Intelligence*. (The Psychological Corporation, San Antonio, TX, 1999).

35. Wechsler, D. *Wechsler Preschool and Primary Scales of Intelligence*. (The Psychological Corporation, San Antonio, TX, 2002).

36. Wechsler, D. *Wechsler Intelligence Scale for Children*. (The Psychological Corporation, San Antonio, TX, 2003).

37. Wechsler, D. *Wechsler Abbreviated Scale of Intelligence*. (NCS Pearson, San Antonio, TX, 2011).

38. Wechsler, D. *Wechsler Intelligence Scale for Children*. (Pearson, Bloomington, MN, 2014).

39. Roid, G. H. *Stanford-Binet Intelligence Scales*. (Riverside Publishing, Itasca, IL, 2003).

40. *Diagnostic and Statistical Manual of Mental Disorders*. (American Psychiatric Association, Arlington, TX, 2013).

41. Kaufman, A. S. & Kaufman, N. L. *Kaufman Brief Intelligence Test*. (Pearson).

42. Volkow, N. D. *et al.* The conception of the ABCD study: From substance use to a broad NIH collaboration. *Dev Cogn Neurosci* **32**, 4–7 (2018).

43. Townsend, L. *et al.* Development of Three Web-Based Computerized Versions of the Kiddie Schedule for Affective Disorders and Schizophrenia Child Psychiatric

Diagnostic Interview: Preliminary Validity Data. *J Am Acad Child Adolesc Psychiatry* **59**, 309–325 (2020).

44. Weintraub, S. *et al.* Cognition assessment using the NIH Toolbox. *Neurology* **80**, S54-64 (2013).

45. Heaton, R. K. *et al.* Reliability and validity of composite scores from the NIH Toolbox Cognition Battery in adults. *J Int Neuropsychol Soc* **20**, 588–598 (2014).

46. Chollet, F. & others. Keras. (2015).

47. Martín Abadi *et al.* TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems. (2015).