Journal of Psychiatric Research 156 (2022) 261–267

Contents lists available at ScienceDirect

Journal of Psychiatric Research

journal homepage: www.elsevier.com/locate/jpsychires

Can machine learning identify childhood characteristics that predict future development of bipolar disorder a decade later?

Mai Uchida [a,b,1,*], Qasim Bukhari [c,1], Maura DiSalvo [a], Allison Green [a,d], Giulia Serra [e], Chloe Hutt Vater [a], Satrajit S. Ghosh [c,f], Stephen V. Faraone [g], John D.E. Gabrieli [c], Joseph Biederman [a,b]

[a] Clinical and Research Programs in Pediatric Psychopharmacology and Adult ADHD, Massachusetts General Hospital, Boston, MA, USA
[b] Department of Psychiatry, Harvard Medical School, Boston, MA, USA
[c] Department of Brain and Cognitive Sciences and McGovern Institute for Brain Research, Massachusetts Institute of Technology, Cambridge, MA, USA
[d] Department of Psychological and Brain Sciences, Indiana University, Bloomington, IN, USA
[e] Department of Neuroscience, Child Neuropsychiatry Unit, I.R.C.C.S. Children Hospital Bambino Gesù, Rome, Italy
[f] Department of Otolaryngology Head and Neck Surgery, Harvard Medical School, USA
[g] Departments of Psychiatry and of Neuroscience and Physiology, SUNY Upstate Medical University, Syracuse, NY, USA

ARTICLE INFO

Keywords:
Machine learning
Bipolar disorder
Mood disorders
Pediatric bipolar disorder

ABSTRACT

Early identification of bipolar disorder may provide appropriate support and treatment, however there is no current evidence for statistically predicting whether a child will develop bipolar disorder. Machine learning methods offer an opportunity for developing empirically-based predictors of bipolar disorder. This study examined whether bipolar disorder can be predicted using clinical data and machine learning algorithms. 492 children, ages 6–18 at baseline, were recruited from longitudinal case-control family studies. Participants were assessed at baseline, then followed-up after 10 years. In addition to sociodemographic data, children were assessed with psychometric scales, structured diagnostic interviews, and cognitive and social functioning assessments. Using the Balanced Random Forest algorithm, we examined whether the diagnostic outcome of full or subsyndromal bipolar disorder could be predicted from baseline data. 45 children (10%) developed bipolar disorder at follow-up. The model predicted subsequent bipolar disorder with 75% sensitivity, 76% specificity, and an Area Under the Receiver Operating Characteristics of 75%. Predictors best differentiating between children who did or did not develop bipolar disorder were the Child Behavioral Checklist Externalizing and Internalizing behaviors, the Child Behavioral Checklist Total t-score, problematic school functions indexed through the Child Behavioral Checklist School Competence scale, and the Child Behavioral Checklist Anxiety/Depression and Aggression scales. Our study provides the first quantitative model to predict bipolar disorder. Longitudinal prediction may help clinicians assess children with emergent psychopathology for future risk of bipolar disorder, an area of clinical and scientific importance. Machine learning algorithms could be implemented to alert clinicians to risk for bipolar disorder.

1. Introduction

Pediatric Bipolar Disorder (BP disorder) is a prevalent and morbid disorder estimated to affect at least 2% of youth (Van Meter et al., 2011). Individuals with BP disorder often present subsyndromal symptoms of mood dysregulation during their childhood that eventually develop into a full diagnosis. The full syndromatic diagnosis of BP disorder is associated with increased risks of suicide, substance use disorders, hospitalization, and social dysfunctions for the patients and their family (De Crescenzo et al., 2017; Faedda et al., 1995; Serra et al., 2017). Although longitudinal studies have found the prognosis of early-onset mood disorders to be unfavorable, research has also shown there are effective treatments and therapies that could significantly alleviate the patients' and their families' struggles from the diagnoses (DelBello et al., 2022; Pavuluri et al., 2005; West et al., 2014). Thus, early identification of the risks and interventions for early symptoms of pediatric mood

* Corresponding author. Massachusetts General Hospital, Warren 624, Boston, MA, 02114, USA.
E-mail address: muchida@partners.org (M. Uchida).
[1] Co-First Authors.

https://doi.org/10.1016/j.jpsychires.2022.09.051
Received 24 June 2022; Received in revised form 26 August 2022; Accepted 24 September 2022
Available online 13 October 2022

disorders is crucial. However, uncertainties remain on how to best predict the development of BP disorder in youth with emergent psychopathology referred to clinical practice (Faedda et al., 1995; Leverich et al., 2007).

In addition, the accurate identification of a developing bipolar disorder is extremely difficult in clinical practice. Since symptoms such as increased activity and impulsivity overlap with Attention Deficit Hyperactivity Disorder, many children are identified as having ADHD prior to them receiving the diagnosis of bipolar disorder. While ADHD could coincide with bipolar disorder, pharmacotherapy using stimulant medications could worsen the mood of children with bipolar disorder. Similarly, some children present with irritability, sadness or anxiety prior to having manic or hypomanic episodes and receive the diagnosis of major depressive disorder instead. The antidepressant medications could further worsen the agitation and activity levels of a child with underlying bipolar disorder due to activation or manic switches. These frequent scenarios in current psychiatric practice further emphasizes the importance of any assistance in prediction of bipolar diagnosis in the future.

Correlational studies have suggested early onset and severe mood symptoms, family history of BP disorder, and severe emotional dysregulation are associated with future development of BP disorder (Uchida et al., 2015a, 2015b). A review examining predictors of adulthood BP disorder reported that cyclothymic features, recurrent depression, anxiety disorders, psychotic symptoms, and family history of BP disorder predicted the development of BP disorder (Faedda et al., 2019). While these studies have aided clinicians in identifying risks for BP disorder, they uncovered group level risk factors and were not specific enough to help inform caregivers about such a prognosis in individuals. Further, these studies have been limited by not testing the generalization of findings from one sample of children to other, independent samples of children and this precludes evidence for generalization to larger populations of children. While the risk factors have been identified, there exists no way of predicting which individuals would develop BP disorder 10 years into the future.

Machine learning approaches can help develop empirically driven childhood predictors of future onset of BP disorder. By aggregating large numbers of sociodemographic and clinical predictors, machine learning empirical models can produce high-quality predictions (Elshawi et al., 2019). For example, using a machine learning model trained on multisite data from the STAR*D consortium, prediction of remission from Major Depressive Disorder (MDD) after a 12-week course of Citalopram therapy was achieved with accuracy of 64.6%, which is a clinically meaningful level of accuracy (Chekroud et al., 2016). Neuroimaging studies predicting transition from mild cognitive impairment to Alzheimer's disease showed an average predictive accuracy above 70% (Arbabshirani et al., 2017). Another study examining the development of psychotic disorders also showed similar predictive rates using clinical measures (Mechelli et al., 2017). While machine learning methods are starting to be used in predicting the prognosis of various psychiatric disorders, to the best of our knowledge, it is yet to be used to predict the future development of BP disorder in children and adolescents with emergent psychopathology.

The main aim of the present study was to assess whether machine learning can help predict the presence of BP disorder in an individual without initial BP disorder a decade later based on presenting childhood sociodemographic and clinical characteristics. To this end, we applied a machine learning approach to analyze data from a unique and large longitudinal sample of children and adolescents of both sexes who were followed for 10 years from childhood to young adulthood with repeated comprehensive assessment batteries. We tested rigorously for generalizability by developing predictive models on a subset of the sample and then testing the accuracy of the models on an independent subset of the sample. We examined the sensitivity and specificity of childhood variables in predicting BP disorder in young adult years using a machine learning algorithm. To the best of our knowledge, this represents the first study using machine learning algorithms for this purpose in pediatric psychiatry.

2. Methods

2.1. Sample

The sample was derived from two identically designed longitudinal case-control family studies of psychiatrically and pediatrically referred youth of both sexes, ages 6–18 years at baseline, and their first-degree biological relatives (parents and siblings) (Biederman et al., 2006, 2010, 2011). The original study recruited equal numbers of boys and girls with and without attention deficit hyperactivity disorder (ADHD) as probands; equal numbers of boys with ADHD (ADHD boys proband), boys without ADHD (control boys proband), girls with ADHD (ADHD girls proband) and girls without ADHD (control girls proband). Their siblings were included without the restrictions of diagnosis, gender or age. Potential participants were excluded if they had major sensorimotor handicaps, psychosis, autism, inadequate command of the English language, or a Full-Scale Intelligence Quotient (IQ) less than 80. Parents and adult offspring provided written informed consent to participate, and parents provided consent for offspring under the age of 18. Children and adolescents provided written assent to participate. Participants from the Boys ADHD study were assessed at baseline, after 4 years, and after 10 years (Biederman et al., 2010). Participants from the Girls ADHD study were assessed at baseline, after 5 years, and after 11 years (Biederman et al., 2011). The human research committee at Massachusetts General Hospital approved this study.

For the current study, we included only participants who had both a baseline evaluation and a diagnostic evaluation 10 years later (N = 780). We excluded 1) probands and siblings who had a positive BP-I disorder diagnosis at baseline, and 2) the ADHD probands. Included were 1) all siblings (siblings of controls and ADHD probands) and the control probands who did not have BP-I disorder at the baseline assessment and 2) children and adolescents had at least 70% of the scales at the examined timeframes. From those participants, scales were included only if they had been completed by at least 70% of participants. For any missing data within a scale, we used the 'most frequent' imputer strategy, but this was necessary for only a single participant for a single scale.

The final sample consisted of 492 children and adolescents, 52% male, ranging in age – at their first evaluation – from 6 to 19 years (μ = 11), after excluding fifteen participants who already had BP-I disorder (Table 1). In this sample, 45 participants (10%) developed BP-I disorder by their 10-year follow-up.

2.2. Assessment procedures

Psychiatric assessments of participants older than 18 years relied on the Structured Clinical Interview for the DSM-IV (SCID) (First et al., 1997) supplemented with modules from the Schedule for Affective Disorder and Schizophrenia for Children (K-SADS-E)(Orvaschel, 1994) to assess childhood diagnoses. Children and adolescents were assessed

Table 1
Baseline demographic and clinical characteristics of the participants included in the analysis.

Characteristic	Included Participants (N = 492)
	Mean ± SD
Age at baseline	11.1 ± 3.2
Socioeconomic status	1.6 ± 0.8
Global Assessment of Functioning	67.8 ± 9.8
CBCL Total Problems T-score	45.1 ± 12.9
Full scale IQ	112.4 ± 12.0
	N (%)
Male	253 (52)

with the K-SADS-E completed with the parents. For youth older than 12 years, direct interviews were also conducted. For these double interviews, we combined data from direct and indirect interviews by considering a diagnostic criterion positive if it was endorsed in either interview. All diagnostic assessments were conducted by highly selective, highly trained, and closely supervised raters. Raters were blind to the ascertainment source of the families (ADHD or Controls). To assess the reliability of our overall diagnostic procedures, we computed kappa coefficients of agreement by having experienced, blinded, board-certified child and adult psychiatrists and licensed experienced clinical psychologists diagnose subjects from audiotaped interviews made by the assessment staff. Based on 500 assessments from interviews of children and adults, the median kappa coefficient was 0.98. Socioeconomic status (SES) was measured using the 5-point Hollingshead scale (Hollingshead, 1975).

2.2.1. Child behavior checklist (CBCL)

The parent of each participant completed the 1991 version of the CBCL for ages 4–18 years. The CBCL queries the parent about the child's behavior in the past six months and aggregates this data into behavioral problem T scores (Achenbach, 1991). A computer program calculates the T scores for each scale. Raw scores are converted to sex- and age-standardized scores (T scores having a mean of 50 and standard deviation (SD) of 10). A minimum T score of 50 is assigned to scores that fall at midpoint percentiles of $\leq$50 on the syndrome scales to permit comparison of standardized scores across scales. T Scores above 70 (2SD) indicate clinical disorder. Clinical subscales include Anxious/-Depressed, Withdrawn/Depressed, Somatic Complains, Social Problems, Thought Problems, Attention Problems, Rule-Breaking Behavior, and Aggressive Behavior. Composite scales include Internalizing Problems, Externalizing Problems, and Total Problems. Competence scales include Activities, School, Social, and Total Competence. In addition, we included the Emotion Dysregulation Profile (AAA score); the aggregate T score of the Anxious/Depressed, Attention and Aggressive Behaviors subscales. While CBCL is a parent-reported scale, studies have documented cross-rater agreement among parent-report and self-report (Althoff et al., 2010; Huang, 2017; Rescorla et al., 2017).

2.2.2. Social Adjustment Inventory for Children and Adolescents (SAICA)

The parent of each participant also completed the Social Adjustment Inventory for Children and Adolescents (SAICA) (Kathoor et al., 1987). The SAICA is a semi-structured interview to assess social functioning. It examines the following domains: activities, peer relations, family relations, and academic performance.

2.2.3. Cognitive assessments

Cognitive ability was measured using the Wechsler Intelligence Scale for Children Revised Version (WISC-R) or Third Edition (WISC-III) for subjects younger than 17 years of age and the Wechsler Adult Intelligence Scale Third Edition (WAIS-III) for subjects 17 years of age or older (Wechsler, 1974, 1991, 2011). The WISC-R/WISC-III and WAIS-III are individually administered tests of intelligence that generate a Full-Scale IQ score and scores in the domains of verbal comprehension, visual spatial abilities, fluid reasoning, working memory, and processing speed.

2.3. Machine learning methods

We used the Random Forest algorithm (Chen et al., 2004; Kam, 1995) which is a decision tree-based machine learning algorithm that has been shown to perform well with complex data sets that have many features (Chen et al., 2020; Khalilia et al., 2011). The Random Forest algorithm uses bootstrap replicas thereby increasing the variance and chooses optimal cut-points in order to split nodes. For this specific study, the dataset is imbalanced with only 10% of the individuals having a BP-I diagnosis at 10 years. Therefore, we used the Balanced Random Forest (Chen et al., 2004) from the imblearn python library (version 0.8.0), which provides a mechanism to address the sample class bias during training by providing each tree a balanced bootstrap sample using an under-sampling strategy. We used 1000 trees and the default settings for their features split and maximum depth.

We then used stratified shuffle split, which is a repeated sampling cross-validation method to assess the accuracy of the model (sensitivity and specificity) for unseen data and to ascertain a distribution of model performance, while removing order effects. We created multiple training and testing pairs by sampling 80% of these data as a training set and using the remaining 20% as a test set of unseen, independent patients. In each pair, there was no overlap in samples between the training and test set. Each training set was used to train the model and the test set was used to evaluate model performance. Using stratification ensures that training and test sets have a similar percentage of each target class (BP+ and BP-). The distribution of performance on the test set of these models provides a better estimate of our confidence in the accuracy of the model, and the shape of the performance distribution can provide clues to sampling biases in the data. The algorithm's performance in predicting BP-I disorder in subjects using information collected at baseline (10 years prior) was evaluated using commonly employed parameters, such as accuracy, sensitivity, specificity, F1 score, area under the receiver operating characteristic curve (ROC-AUC) and the precision recall curve.

We also examined the feature importance map created during repeated sampling cross-validation to determine which features most accounted for the model's predictive accuracy. We estimated the median of the feature importance maps weighted by the ROC-AUC in each of the cross-validation loop. Although we selected the Random Forest algorithm *a priori* due to its usefulness with similar data sets, we also tested performance from an alternative algorithm, the balanced Bagging Classifier (Maclin and Opitz, 1997). The balanced Bagging Classifier performed less well (Supplementary Table) than the Random Classifier, and we report the findings from the Random Forest algorithm.

3. Results

3.1. Sensitivity and specificity of the model in predicting final BPD status

The model computes the probability that a child will develop BP-I disorder. When using a probability of 0.5 or greater to predict the onset of BP-I disorder, the model accurately predicted the development of BP-I disorder with a median sensitivity of 75% and median specificity of 76%. The area under the receiver operating characteristic curve (ROC-AUC) was with median of 75% and F1 score was with median of 79% (Table 2). Our model has a false positive rate of 21.6% and a false negative rate of 3.1%.

The average precision-recall curve of all bootstrapped iteration is presented in Fig. 1. This curve plots the positive predictive power against the sensitivity for each possible cut point on the model's output probability. The ROC-AUC curve is presented in Fig. 2.

Table 2
Performance evaluation using repeated sampling cross-validation.

	Median	Standard Deviation
Accuracy	75.69% [66.66 – 82.63]	3.29
Sensitivity	75.19% [66.14 - 81.88]	3.76
Specificity	76.47% [47.05 - 100]	10.79
ROC-AUC	75.26% [74.40 – 77.6]	0.66
F1 score	79.35% [72.48 - 85.17]	2.5

Table 2 shows the median and standard deviation values of accuracy, sensitivity, specificity, ROC-AUC and F1 score of our machine learning model over several repeated sampling cross validation iterations. All the performance evaluation metrics are close to each other representing balanced results.

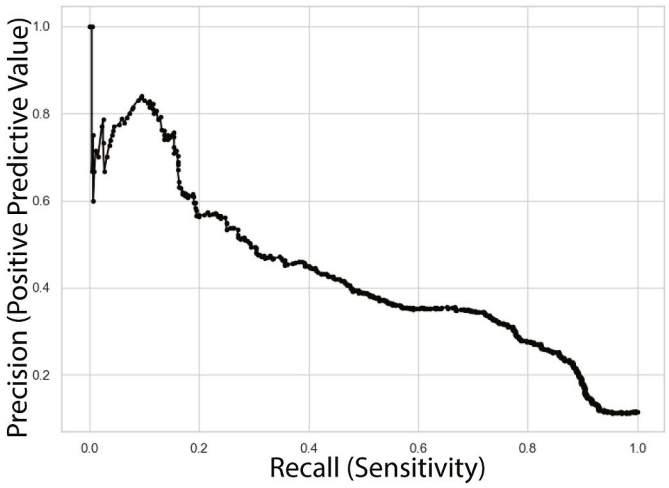

Fig. 1. Precision-Recall curve of bipolar disorder prediction for several cross-validation iterations. The figure shows the median precision-recall of all iterations of the cross-validation.

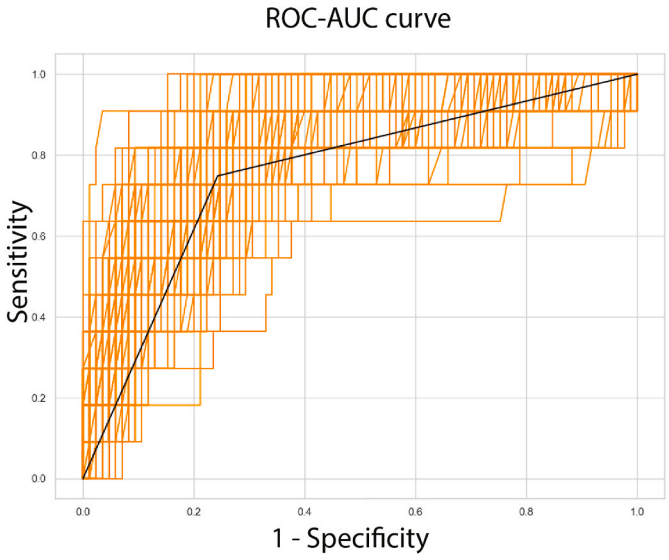

Fig. 2. ROC-AUC curve of bipolar disorder prediction for several cross-validation iterations. The figure shows the median ROC-AUC of all iterations of the cross-validation.

3.2. Important features identified to be predictive of BP-I disorder

To determine which childhood features contributed most to differentiating between children and adolescents who developed BP-I disorder and those who did not, we analyzed the feature importance map. The top 7 features were CBCL Total t-score, CBLC Externalizing t-score, CBCL-AAA score, CBCL Internalizing t-score, CBCL School Competence t-score, CBCL Anxious/Depressed t-score and CBCL Aggressive t-score (Table 3). It should be noted that for nonlinear models such as Random Forest, performance is a function of the combinations of features, not that of an isolated feature even if they are most salient. We only show the top 7 features because the remaining features were less strong and all similar to one another in feature importance.

4. Discussion

The main aim of the present study was to examine whether machine learning could predict the future development of BP-I disorder by using

Table 3
Top 7 important features identified from classification.

Features at Baseline	Importance values
CBCL Total t-score	0.065617
CBLC Externalizing t-score	0.062014
CBCL-AAA score	0.043109
CBCL Internalizing t-score	0.042678
CBCL School Competence t-score	0.039680
CBCL Anxious/Depressed t-score	0.037959
CBCL Aggressive t-score	0.034370

Table 3 shows the top 7 important features. Only the top 7 features are selected because there is a drop in importance values after the top 7. All the top 7 features contain CBCL related scores.

childhood clinical characteristics. Our model predicted the existence of BP-I disorder 10 years into the future with a sensitivity of 75% and a specificity of 76%. This is the first evidence that future diagnosis of BP-I disorder can be predicted significantly better than chance at an individual level. A limitation of machine learning methods with many features is that they are a sort of "black box" between inputs (features) and outputs (diagnostic status). For that reason, we used methods that identified which inputs had the strongest effects. The top seven features were all from the CBCL: CBCL Total t-score, CBCL Externalizing t-score, CBCL-AAA score, CBCL Internalizing t-score, CBCL School Competence t-score, CBCL Anxious/Depressed t-score and CBCL Aggressive t-score. All the other 56 features were less predictive and similar to one another in strength of prediction. Although identification of relative feature strength opens up the black box and can be related to recognizable measures, it is important to note that the strength of machine learning typically derives from its use of many combined features rather than the strength of a few features. While variables such as the CBCL total t-scores, AAA scores and externalizing t-scores that showed strength in predicting the development of BP-I disorder here are variables that do not specifically predict BP-I disorder in clinical practice, machine learning uses many combined features to learn and adapt by analyzing and drawing inferences from patterns in data that allow personalized predictions. This method is different from traditional variable-specific predictions and could aid clinical practice in the future.

Our study suggests that machine learning could aid clinicians in individual prognosis of the development of the diagnosis of BP-I disorder in the 10-year follow up period from clinical characteristics presented in childhood. As an example, such a model could alert clinicians and caregivers to improve monitoring over a 10-year period. This is especially important since frequently BP disorder is either preceded by or mistaken with depression or ADHD diagnosis, and if the pharmacological treatment for depression or ADHD is given to BP disorder patients, it could make their condition worse. Therefore, it is critically important to assist the clinicians towards better treatment selection in patients with high risk of developing BP disorder and our model is the first prognostic model to successfully predict the development of BP disorder.

Our model also has the strength that the features used for prediction can be captured in a cost-effective manner, because the top seven features can be measured through a parent reported questionnaire and no additional clinician time is required. Another strength is the high quality of our labeled data with well validated clinical scales and structured diagnostic interviews with high inter-rater reliability. The finding that severe forms of emotional dysregulation as indexed through the aggregate t-score of CBCL Attention, Aggression and Anxiety/Depression scales (CBCL-AAA, or CBCL Emotion Dysregulation Profile) represents a childhood predictor for the future development of BP disorder is consistent with the literature (Biederman et al., 2012). For example, CBCL-AAA scores above 195 can efficiently identify children with a structured interview derived cross-sectional diagnosis of pediatric BP-I disorder with high accuracy (Yule et al., 2019).

Our current study also found that not only were the aggregate CBCL t-scores predictive of BP-I disorder, but that the individual elevation of

the CBCL anxiety/depression, aggression and attention scores represented additional risk for future development of BP-I disorder in the youth. The high comorbidity between BP disorder and anxiety disorders as well as ADHD has been well documented (Biederman et al., 2013; Dineen Wagner, 2006; Wingo and Ghaemi, 2007). Also depression is part of the BP disorder mood presentation along with manic, hypomanic, and mixed mood states. Increased irritability that could involve aggression is also part of the criteria of manic episodes. As such, it is not surprising that these scales that measure elevation of anxious and depressed moods, aggressive behaviors and inattention could be associated with future development of BP-I disorder.

Likewise, the finding that childhood school behavior problems and rule breaking behaviors were predictive of the future development of BP-I disorder are consistent with the literature. School behavior problems are present in 83% of children with depression who eventually developed bipolar disorder vs. 59% in those who did not (Biederman et al., 2009). Other studies have also found that youth who developed BP disorder over time struggled with higher levels of school behavioral problems when compared with youth who developed unipolar depression (Wozniak et al., 2004). Rule breaking behaviors are also strongly associated with the development of BP disorder in youth (Tseng et al., 2015). There is a strong bidirectional overlap between BP disorder and Conduct Disorder, a disorder that its symptomatology is highly correlated with the CBCL Rule Breaking scores (Biederman et al., 1999, 2003).

While BP disorder is a known heritable disease confirmed by numerous genetic and family studies (Gordovez and McMahon, 2020; Muller and Muller, 2016; Ramos et al., 2019), family history of BP disorder had a low feature importance score similar to many other features. The reason for these results needs further investigation, but it could be due to family history being specific but not sensitive in prediction of future outcomes. Thus, it is likely that family history of BP disorder did not enter the model because it was redundant with, and weaker than, other predictors. Our study has several limitations. Our analysis was done in a 10-year longitudinal sample that originally were recruited based on the presence or absence of ADHD and was predominantly Caucasian. Thus, our results may not generalize to other ethnic groups of community samples. Our study focused on BP-I disorder but did not include other subtypes of bipolar disorders such as BP-II or cyclothymic disorders. Additionally, our study included patients from studies with different study assessment time points (boys: 10 years vs. girls: 11 years), which could have potentially biased the results since the patients from the boys had more time to develop bipolar disorder. However, there was only a half a year difference in average follow-up time between the boys and girls studies and the impact of this difference in follow-up time was likely minimal. Another limitation is the possibility that we used a specific machine learning model. It is possible that other machine learning models could perform better or worse than the Balanced Random Forest classifier. In the medical literature, however, Random Forest classifiers have attained state-of-the-art performance. For example, in a 10-year longitudinal aiming to predict the development of hypertension, Random Decision Forests outperformed five other machine learning models for predicting future hypertension (Elshawi et al., 2019).

When using such predictive models to inform practice, a clinician must consider the impact of missed predictions about individuals who go on to develop BP (false negatives; 3.1%) and the pressure on system resources for those who are monitored but do not develop BP (false positives; 21.6%). Given the generally low rate of conversion to BP disorder, it may be better to tolerate a high false-positive rate for the sake of monitoring a few more children who need such monitoring rather than having a high-false negative rate and overlooking children who should have been monitored. Improving accuracy may require deeper phenotyping, larger sample sizes, and more extensive evaluation of different machine learning models. Future research can also show whether other measures can improve prediction accuracy, such as genetics or neuroimaging; neuroimaging has shown some promise in predicting longitudinal progression of mood disorders or symptoms (Hirshfeld-Becker et al., 2019; Whitfield-Gabrieli et al., 2020).

While our model offers the first evidence that BP disorder can be statistically predicted 10 years before clinical onset, we need improved accuracy in predicting the development of BP disorder in order to better inform clinical practice. However, the current model, even with its overidentification of false positives, may have value in identifying children and adolescents who warrant additional attention by alerting clinicians treating the youth with emergent symptoms of mood disorders as to their future risk of BP disorder.

Financial disclosure statement

This work was partially supported by the Abdul Latif Jameel Clinic for Machine Learning in Health (J-Clinic), the MGH Pediatric Psychopharmacology Fund, and the Poitras Center for Psychiatric Disorders Research at the McGovern Institute for Brain Research at MIT. The supporters had no role in the design, analysis, interpretation, or publication of this study.

Author statement

Mai Uchida, MD: Conceptualization, Writing- Original Draft Preparation, Writing- Reviewing & Editing. **Qasim Bukhari, PhD:** Conceptualization, Writing- Original Draft Preparation, Writing- Reviewing & Editing. **Maura DiSalvo, MPH:** Conceptualization, Formal Analysis, Writing- Original Draft Preparation, Writing- Reviewing & Editing. **Allison Green, BA:** Conceptualization, Writing- Original Draft Preparation. **Giulia Serra, MD:** Conceptualization, Writing- Original Draft Preparation. **Chloe Hutt Vater, BA:** Writing- Reviewing & Editing, Resources. **Satrajit S. Ghosh, PhD:** Conceptualization, Writing- Original Draft Preparation. **Stephen V. Faraone, PhD:** Conceptualization, Writing- Original Draft Preparation. **John D. E. Gabrieli, PhD:** Conceptualization, Writing- Original Draft Preparation, Writing- Reviewing & Editing. **Joseph Biederman, MD:** Conceptualization, Writing- Original Draft Preparation, Writing- Reviewing & Editing. All authors gave their final approval of the version of the article to be published. All authors are responsible for the reported research and have approved the manuscript as submitted.

Declaration of competing interest

Dr. Mai Uchida is partially supported by a K award, grant number 1K23MH122667-01. Dr. Uchida also provided a one-time consultation to the Moderna scientific advisory board.

Allison Green received funding from the Eunice Kennedy Shriver National Institute of Child Health & Human Development of the National Institutes of Health under award number T32HD007475.

In the past year, **Dr. Faraone** received income, potential income, travel expenses continuing education support and/or research support from Aardvark, Akili, Genomind, Ironshore, KemPharm/Corium, Noven, Ondosis, Otsuka, Rhodes, Supernus, Takeda, Tris and Vallon. With his institution, he has US patent US20130217707 A1 for the use of sodium-hydrogen exchange inhibitors in the treatment of ADHD. In previous years, he received support from: Alcobra, Arbor, Aveksham, CogCubed, Eli Lilly, Enzymotec, Impact, Janssen, Lundbeck/Takeda, McNeil, NeuroLifeSciences, Neurovance, Novartis, Pfizer, Shire, and Sunovion. He also receives royalties from books published by Guilford Press: *Straight Talk about Your Child's Mental Health*; Oxford University Press: *Schizophrenia: The Facts;* and Elsevier: *ADHD: Non-Pharmacologic Interventions.* He is also Program Director of www.adhdinadults.com. Dr. Faraone is supported by the European Union's Horizon 2020 research and innovation programme under grant agreement No 965381; NIMH grants U01AR076092-01A1, 1R21MH1264940, R01MH116037; Oregon Health and Science University, Otsuka Pharmaceuticals, Noven

Pharmaceuticals Incorporated, and Supernus Pharmaceutical Company.

Dr. Joseph Biederman is currently receiving research support from the following sources: AACAP, Feinstein Institute for Medical Research, Genentech, Headspace Inc., NIDA, Pfizer Pharmaceuticals, Roche TCRC Inc., Sunovion Pharmaceuticals Inc., Takeda/Shire Pharmaceuticals Inc., Tris, and NIH. Dr. Biederman and his program have received royalties from a copyrighted rating scale used for ADHD diagnoses, paid by Biomarin, Bracket Global, Cogstate, Ingenix, Medavent Prophase, Shire/Takeda, Sunovion, and Theravance; these royalties were paid to the Department of Psychiatry at MGH. Through Partners Healthcare Innovation, Dr. Biederman has a partnership with MEMOTEXT to commercialize a digital health intervention to improve adherence in ADHD. Through MGH corporate licensing, Dr. Biederman has a US Patent (#14/027,676) for a non-stimulant treatment for ADHD, a US Patent (#10,245,271 B2) on a treatment of impaired cognitive flexibility, and a patent pending (#61/233,686) on a method to prevent stimulant abuse. In 2022: Dr. Biederman received honoraria from the MGH Psychiatry Academy for tuition-funded CME courses. In 2021: Dr. Biederman received an honorarium for a scientific presentation from Multi-Health Systems, and a one-time consultation for Cowen Healthcare Investments. He received honoraria from AACAP, the American Psychiatric Nurses Association, BIAL - Portela & C[a]. S.A. (Portugal), Medscape Education, and MGH Psychiatry Academy for tuition-funded CME courses. In 2020: Dr. Biederman received an honorarium for a scientific presentation from Tris and from the Institute of Integrated Sciences – INI (Brazil), and research support from the Food & Drug Administration. He received honoraria from Medlearning Inc, NYU, and MGH Psychiatry Academy for tuition-funded CME courses. In 2019, Dr. Biederman was a consultant for Akili, Avekshan, Jazz Pharma, and Shire/Takeda. He received research support from Lundbeck AS and Neurocentria Inc. Through MGH CTNI, he participated in a scientific advisory board for Supernus. He received honoraria from the MGH Psychiatry Academy for tuition-funded CME courses.

Dr. Qasim Bukhari, Ms. Maura DiSalvo, Dr. Guilla Serra, Ms. Chloe Hutt Vater, Dr. Satrajit Ghosh, and Dr. John Gabrieli do not have any financial relationships to disclose.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jpsychires.2022.09.051.

References

Achenbach, T.M., 1991. Manual for the Child Behavior Checklist/4-18 and the 1991 Profile. University of Vermont, Department of Psychiatry, Burlington, VT.

Althoff, R.R., Rettew, D.C., Ayer, L.A., Hudziak, J.J., 2010. Cross-informant agreement of the dysregulation profile of the child behavior checklist. Psychiatr. Res. 178 (3), 550–555.

Arbabshirani, M.R., Plis, S., Sui, J., Calhoun, V.D., 2017. Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. Neuroimage 145 (Pt B), 137–165.

Biederman, J., Faraone, S.V., Chu, M.P., Wozniak, J., 1999. Further evidence of a bidirectional overlap between juvenile mania and conduct disorder in children. J. Am. Acad. Child Adolesc. Psychiatry 38 (4), 468–476.

Biederman, J., Faraone, S.V., Petty, C., Martelon, M., Woodworth, K.Y., Wozniak, J., 2013. Further evidence that pediatric-onset bipolar disorder comorbid with ADHD represents a distinct subtype: results from a large controlled family study. J. Psychiatr. Res. 47 (1), 15–22.

Biederman, J., Mick, E., Wozniak, J., Monuteaux, M.C., Galdo, M., Faraone, S.V., 2003. Can a subtype of conduct disorder linked to bipolar disorder be identified? Integration of findings from the Massachusetts General Hospital Pediatric Psychopharmacology Research Program. Biol. Psychiatr. 53 (11), 952–960.

Biederman, J., Monuteaux, M.C., Mick, E., Spencer, T., Wilens, T.E., Silva, J.M., Snyder, L.E., Faraone, S.V., 2006. Young adult outcome of attention deficit hyperactivity disorder: a controlled 10-year follow-up study. Psychol. Med. 36 (2), 167–179.

Biederman, J., Petty, C.R., Byrne, D., Wong, P., Wozniak, J., Faraone, S.V., 2009. Risk for switch from unipolar to bipolar disorder in youth with ADHD: a long term prospective controlled study. J. Affect. Disord. 119 (1–3), 16–21.

Biederman, J., Petty, C.R., Clarke, A., Lomedico, A., Faraone, S.V., 2011. Predictors of persistent ADHD: an 11-year follow-up study. J. Psychiatr. Res. 45 (2), 150–155.

Biederman, J., Petty, C.R., Day, H., Goldin, R.L., Spencer, T., Faraone, S.V., Surman, C.B., Wozniak, J., 2012. Severity of the aggression/anxiety-depression/attention child behavior checklist profile discriminates between different levels of deficits in emotional regulation in youth with attention-deficit hyperactivity disorder. J. Dev. Behav. Pediatr. 33 (3), 236–243.

Biederman, J., Petty, C.R., Evans, M., Small, J., Faraone, S.V., 2010. How persistent is ADHD? A controlled 10-year follow-up study of boys with ADHD. Psychiatr. Res. 177 (3), 299–304.

Chekroud, A.M., Zotti, R.J., Shehzad, Z., Gueorguieva, R., Johnson, M.K., Trivedi, M.H., Cannon, T.D., Krystal, J.H., Corlett, P.R., 2016. Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatr. 3 (3), 243–250.

Chen, C., Liaw, A., Breiman, L., 2004, 1-12. Using Random Forest to Learn Imbalanced Data, vol. 110. University of California, Berkeley, p. 24.

Chen, R.-C., Dewi, C., Huang, S.-W., Caraka, R.E., 2020. Selecting critical features for data classification based on machine learning methods. Journal of Big Data 7 (1), 1–26.

De Crescenzo, F., Serra, G., Maisto, F., Uchida, M., Woodworth, H., Casini, M.P., Baldessarini, R.J., Vicari, S., 2017. Suicide attempts in juvenile bipolar versus major depressive disorders: systematic review and meta-analysis. J. Am. Acad. Child Adolesc. Psychiatry 56 (10), 825–831 e823.

DelBello, M.P., Kadakia, A., Heller, V., Singh, R., Hagi, K., Nosaka, T., Loebel, A., 2022. Systematic review and network meta-analysis: efficacy and safety of second-generation antipsychotics in youths with bipolar depression. J. Am. Acad. Child Adolesc. Psychiatry 61 (2), 243–254.

Dineen Wagner, K., 2006. Bipolar disorder and comorbid anxiety disorders in children and adolescents. J. Clin. Psychiatr. 67 (Suppl. 1), 16–20.

Elshawi, R., Al-Mallah, M.H., Sakr, S., 2019. On the interpretability of machine learning-based model for predicting hypertension. BMC Med. Inf. Decis. Making 19 (1), 146.

Faedda, G.L., Baldessarini, R.J., Marangoni, C., Bechdolf, A., Berk, M., Birmaher, B., Conus, P., DelBello, M.P., Duffy, A.C., Hillegers, M.H.J., Pfennig, A., Post, R.M., Preisig, M., Ratheesh, A., Salvatore, P., Tohen, M., Vázquez, G.H., Vieta, E., Yatham, L.N., Youngstrom, E.A., Van Meter, A., Correll, C.U., 2019. An International Society of Bipolar Disorders task force report: precursors and prodromes of bipolar disorder. Bipolar Disord. 21 (8), 720–740.

Faedda, G.L., Baldessarini, R.J., Suppes, T., Tondo, L., Becker, I., Lipschitz, D.S., 1995. Pediatric-onset bipolar disorder: a neglected clinical and public health problem. Harv. Rev. Psychiatr. 3 (4), 171–195.

First, M., Spitzer, R., Gibbon, M., Williams, J., 1997. Structured Clinical Interview for DSM-IV Axis I Disorders. American Psychiatric Press, Washington, DC.

Gordovez, F.J.A., McMahon, F.J., 2020. The genetics of bipolar disorder. Mol. Psychiatr. 25 (3), 544–559.

Hirshfeld-Becker, D.R., Gabrieli, J.D.E., Shapero, B.G., Biederman, J., Whitfield-Gabrieli, S., Chai, X.J., 2019. Intrinsic functional brain connectivity predicts onset of major depression disorder in adolescence: a pilot study. Brain Connect. 9 (5), 388–398.

Hollingshead, A.B., 1975. Four Factor Index of Social Status. Yale Press, New Haven, CT.

Huang, C., 2017. Cross-Informant agreement on the child behavior checklist for youths: a meta-analysis. Psychol. Rep. 120 (6), 1096–1116.

Kam, H.T., 1995. Random Decision Forest. Proceedings of the 3rd international conference on document analysis and recognition, 278282. Montreal, Canada, August.

Kathoor, J., Gammon, G.D., Prusoff, B.A., Warner, V., 1987. The social adjustment inventory for children and adolescents (SAICA); testing of a new semi-structured interview (SAICA). J. Am. Acad. Child Adolesc. Psychiatr. 26 (6), 898–911.

Khalilia, M., Chakraborty, S., Popescu, M., 2011. Predicting disease risks from highly imbalanced data using random forest. BMC Med. Inf. Decis. Making 11 (1), 1–13.

Leverich, G.S., Post, R.M., Keck Jr., P.E., Altshuler, L.L., Frye, M.A., Kupka, R.W., Nolen, W.A., Suppes, T., McElroy, S.L., Grunze, H., Denicoff, K., Moravec, M.K., Luckenbaugh, D., 2007. The poor prognosis of childhood-onset bipolar disorder. J. Pediatr. 150 (5), 485–490.

Maclin, R., Opitz, D.W., 1997. An Empirical Evaluation of Bagging and Boosting. AAAI/IAAI.

Mechelli, A., Lin, A., Wood, S., McGorry, P., Amminger, P., Tognin, S., McGuire, P., Young, J., Nelson, B., Yung, A., 2017. Using clinical information to make individualized prognostic predictions in people at ultra high risk for psychosis. Schizophr. Res. 184, 32–38.

Muller, W.E., Muller, J.K., 2016. Basic data for bipolar disorders: genetics, neurobiology and pharmacology. Med. Monatsschr. Pharm. 39 (9), 371–376.

Orvaschel, H., 1994. Schedule for Affective Disorders and Schizophrenia for School-Age Children Epidemiologic Version, fifth ed. Nova Southeastern University, Center for Psychological Studies, Ft, Lauderdale.

Pavuluri, M.N., Birmaher, B., Naylor, M.W., 2005. Pediatric bipolar disorder: a review of the past 10 years. J. Am. Acad. Child Adolesc. Psychiatry 44 (9), 846–871.

Ramos, B.R., Librenza-Garcia, D., Zortea, F., Watts, D., Zeni, C.P., Tramontina, S., Passos, I.C., 2019. Clinical differences between patients with pediatric bipolar disorder with and without a parental history of bipolar disorder. Psychiatr. Res. 280, 112501.

Rescorla, L.A., Ewing, G., Ivanova, M.Y., Aebi, M., Bilenberg, N., Dieleman, G.C., Döpfner, M., Kajokiene, I., Leung, P.W., Plück, J., Steinhausen, H.C., Winkler Metzke, C., Zukauskiene, R., Verhulst, F.C., 2017. Parent-adolescent cross-informant agreement in clinically referred samples: findings from seven societies. J. Clin. Child Adolesc. Psychol. 46 (1), 74–87.

Serra, G., Uchida, M., Battaglia, C., Casini, M.P., De Chiara, L., Biederman, J., Vicari, S., Wozniak, J., 2017. Pediatric mania: the controversy between euphoria and irritability. Curr. Neuropharmacol. 15 (3), 386–393.

Tseng, W.L., Guyer, A.E., Briggs-Gowan, M.J., Axelson, D., Birmaher, B., Egger, H.L., Helm, J., Stowe, Z., Towbin, K.A., Wakschlag, L.S., Leibenluft, E., Brotman, M.A., 2015. Behavior and emotion modulation deficits in preschoolers at risk for bipolar disorder. Depress. Anxiety 32 (5), 325–334.

Uchida, M., Serra, G., Zayas, L., Kenworthy, T., Faraone, S.V., Biederman, J., 2015a. Can unipolar and bipolar pediatric major depression be differentiated from each other? A systematic review of cross-sectional studies examining differences in unipolar and bipolar depression. J. Affect. Disord. 176, 1–7.

Uchida, M., Serra, G., Zayas, L., Kenworthy, T., Hughes, B., Koster, A., Faraone, S.V., Biederman, J., 2015b. Can manic switches be predicted in pediatric major depression? A systematic literature review. J. Affect. Disord. 172, 300–306.

Van Meter, A.R., Moreira, A.L., Youngstrom, E.A., 2011. Meta-analysis of epidemiologic studies of pediatric bipolar disorder. J. Clin. Psychiatr. 72 (9), 1250–1256.

Wechsler, D., 1974. Manual for the Wechsler Intelligence Scale for Children-Revised. The Psychological Corporation, New York.

Wechsler, D., 1991. Manual for the Wechsler Intelligence Scale for Children, third ed. The Psychological Corporation, Harcourt Brace Jovanovich, Inc., San Antonio.

Wechsler, D., 2011. Wechsler Abbreviated Scale of Intelligence (WASI-II), second ed. NCS Pearson, Inc., Bloomington, MN.

West, A.E., Weinstein, S.M., Peters, A.T., Katz, A.C., Henry, D.B., Cruz, R.A., Pavuluri, M. N., 2014. Child- and family-focused cognitive-behavioral therapy for pediatric bipolar disorder: a randomized clinical trial. J. Am. Acad. Child Adolesc. Psychiatry 53 (11), 1168–1178, 1178.e1161.

Whitfield-Gabrieli, S., Wendelken, C., Nieto-Castanon, A., Bailey, S.K., Anteraper, S.A., Lee, Y.J., Chai, X.Q., Hirshfeld-Becker, D.R., Biederman, J., Cutting, L.E., Bunge, S. A., 2020. Association of intrinsic brain architecture with changes in attentional and mood symptoms during development. JAMA Psychiatr. 77 (4), 378–386.

Wingo, A.P., Ghaemi, S.N., 2007. A systematic review of rates and diagnostic validity of comorbid adult attention-deficit/hyperactivity disorder and bipolar disorder. J. Clin. Psychiatr. 68 (11), 1776–1784.

Wozniak, J., Spencer, T., Biederman, J., Kwon, A., Monuteaux, M., Rettew, J., Lail, K., 2004. The clinical characteristics of unipolar versus bipolar major depression in ADHD youth. J. Affect. Disord. 82, S59–S69.

Yule, A., Fitzgerald, M., Wilens, T., Wozniak, J., Woodworth, K.Y., Pulli, A., Uchida, M., Faraone, S.V., Biederman, J., 2019. Further evidence of the diagnostic utility of the child behavior checklist for identifying pediatric bipolar I disorder. Scand J Child Adolesc Psychiatr Psychol 7 (1), 29–36.