Article

Machine learning prediction of estimated risk for bipolar disorders using hippocampal subfield and amygdala nuclei volumes

Fabian Huth, BSc[1], Leonardo Tozzi, PhD[2], Michael Marxen, PhD[1], Philipp Riedel, MD[1], Kyra Bröckel, MSc[1], Prof. Julia Martini[1], Christina Berndt, MSc[1], Cathrin Sauer[1], Christoph Vogelbacher,PhD[3,4,6], Prof. Andreas Jansen[3,5,6], Prof. Tilo Kircher[5,6], Prof. Irina Falkenberg[5,6], Florian Thomas-Odenthal[5,6], Prof. Martin Lambert[7], Vivien Kraft, MSc[7], Gregor Leicht, PhD[7], Prof. Christoph Mulert[6,7,8], Prof. Andreas J. Fallgatter[9], Prof. Thomas Ethofer[9], Anne Rau, PhD[9], Karolina Leopold, PhD[10], Prof. Andreas Bechdolf[10], Prof. Andreas Reif[11], Silke Matura, PhD[11], Silvia Biere[11], Prof. Felix Bermpohl[12], Jana Fiebig[12], Prof. Thomas Stamm[12,13], Prof. Christoph U. Correll[14,15,16], Prof. Georg Juckel[17], Vera Flasbeck, PhD[17], Philipp Ritter, PhD[1], Prof. Michael Bauer[1], Prof. Andrea Pfennig[1] and Pavol Mikolas, PhD[1]

[1]Department of Psychiatry and Psychotherapy, Carl Gustav Carus University Hospital, Technische Universität Dresden, Germany
[2]Department of Psychiatry and Behavioral Sciences, Stanford University School of Medicine, Stanford, CA, USA
[3]Core-Facility Brainimaging, Faculty of Medicine, University of Marburg, Germany
[4]Philipps-University Marburg, Translational Clinical Psychology, Marburg, Germany
[5]Department of Psychiatry and Psychotherapy, University of Marburg, Germany
[6]Center for Mind, Brain and Behavior (CMBB), University of Marburg and Justus Liebig University Giessen, Germany
[7]Department of Psychiatry and Psychotherapy, University Medical Center Hamburg-Eppendorf, Hamburg, Germany
[8]Centre for Psychiatry, Justus-Liebig University Giessen, Germany
[9]Department of Psychiatry, Tuebingen Center for Mental Health, University of Tuebingen, Tuebingen, Germany
[10]Department of Psychiatry, Psychotherapy and Psychosomatic Medicine, Vivantes Hospital Am Urban and Vivantes Hospital Im Friedrichshain, Charité-Universitätsmedizin Berlin, Germany
[11]Goethe University Frankfurt, University Hospital, Department of Psychiatry, Psychosomatic Medicine and Psychotherapy, Germany
[12]Department of Psychiatry and Psychotherapy, Charité Campus Mitte, Charité University Medicine, Berlin, Germany
[13]Department of Clinical Psychiatry and Psychotherapy, Brandenburg Medical School Theodor Fontane, Neuruppin, Germany
[14]Department of Child and Adolescent Psychiatry, Charité Universitätsmedizin Berlin, Berlin, Germany
[15]Department of Psychiatry, Northwell Health, The Zucker Hillside Hospital, Glen Oaks, NY, USA
[16]Department of Psychiatry and Molecular Medicine, Donald and Barbara Zucker School of Medicine at Hofstra/Northwell, Hempstead, NY, USA
[17]Department of Psychiatry, Psychotherapy and Preventive Medicine, LWL University Hospital, Ruhr-University, Bochum, Germany
*Correspondence: pavol.mikolas@uniklinikum-dresden.de; Tel.: +49 351 458-2662

Abstract: The pathophysiology of bipolar disorder (BD) remains mostly unclear. Yet, a valid biomarker is necessary to improve early detection of this serious disorder. Patients with manifest BD display reduced volumes of the hippocampal subfields and amygdala nuclei. In this pre-registered analysis, we used structural MRI (*N*=271, 7 sites), to compare volumes of hippocampus, amygdala, and their subfields/nuclei between help-seeking subjects divided in risk groups for BD as estimated by BPSS-P, BARS and EPI*bipolar*. We performed between-group comparisons using linear mixed effects models for all three risk assessment tools. Additionally, we aimed to differentiate the risk groups using linear support vector machine. We found no significant volume differences between the risk groups for all limbic structures during the main analysis. However, the SVM could still classify subjects at risk according to BPSS-P criteria with a balanced accuracy of 66.90% (95% *CI* 59.2– 74.6) for 10-fold cross-validation and 61.9% (95% *CI* 52.0– 71.9) for leave-one-site-out. Structural alterations of hippocampus and amygdala may not be as pronounced in young people at risk, nonetheless, machine learning can predict estimated risk for BD above chance. This suggests that neural changes may not merely be a consequence of BD and may have prognostic clinical value.

Keywords: bipolar risk, hippocampal subfields, amygdala nuclei, MRI, machine learning

1. Introduction

Bipolar disorders (BD) are serious recurrent, chronic mental disorders starting in early adulthood (18.4 – 20 years) **[1]**, which contribute to 6.8% (4.9–9.1) disability-adjusted life years lost due to mental disorders **[2]**. A longer duration of untreated illness leads to more depressive and manic episodes and more suicidal behavior **[3]**. Ten to twenty percent of patients suffering from BD commit suicide throughout their disease course **[4]**. Hence, early detection and treatment is crucial.

Detection of BD risk prior to diagnosis has been increasingly subject to scientific research **[5]**. One established approach is to study genetic risk at the individual level, i.e., BD diagnoses among first degree relatives **[6]**. The transition rates of first-degree relatives have been estimated at 4.2 – 22.4 % **[7–9]**. Another approach is to study a broader range of clinical risk factors in help-seeking populations using risk assessment tools. For instance, the prognostic accuracies of two clinical interviews have been investigated: The Bipolar at Risk States Revised (BARS; Harrell's $C = 0.777$) and the Semistructured Interview of Bipolar At Risk States (SIBARS; Harrell's $C = 0.742$) **[10]**. Besides SIBARS, there are two other established BD risk assessment tools which we used here: The Early Phase Inventory for Bipolar Disorders (EPI-*bipolar* **[11]**), a semi-structured interview consisting of a broad range of literature-derived risk factors for BD and the Bipolar Prodrome Symptom Scale—Prospective (BPSS-P **[12]**), a semi-structured interview assessing three different scales based on DSM IV criteria. Besides interviews, using biomarkers, such as structural neuroimaging, might also yield a good prognostic accuracy and thus lead to an earlier detection of psychiatric disorders and better clinical outcomes **[13,14]**.

During the search for such biomarkers, several structural abnormalities have already been identified in BD **[15]**. These include reduced cortical thickness in frontal, parietal and temporal regions **[16]**, as well as smaller subcortical structures (hippocampus, thalamus, and amygdala) **[17]**. Using more fine-grained parcellation methods it became possible to study subcortical structures and their subregions. One region of interest (ROI) has been investigated thoroughly in a currently published mega-analysis with 4698 subjects: The authors found significantly smaller volumes of the hippocampus and some of its subfields (whole hippocampus, GC-ML-DG, CA4, CA3, CA1, subiculum, presubiculum, molecular layer HP, HATA, and hippocampal tail) but not others (parasubiculum, fimbria, and hippocampal fissure) in BD patients compared to healthy controls (HC) **[18]**. Another study found some subfields to be smaller in both BD and schizophrenia (SZ) patients, compared to HC (bilateral CA2/3, CA4/dentate gyrus, subiculum and right CA1), whereas presubiculum volumes were smaller only in SZ **[19]**. In an investigation of different psychotic disorders, smaller subfield volumes in psychotic BD could be found only in the bilateral CA2/3, the left presubiculum and the right CA4/DG, compared to HC **[20]**.

Another valuable ROI for early detection of BD, with less BD-specific scientific literature available to date, is the amygdala and its nuclei. The amygdala is involved in emotion regulation via connections with the medial prefrontal and the orbitofrontal cortex, inferior frontal gyrus, hypothalamus, and the ventromedial striatum **[21,22]**. Its nuclei show widespread, but differentially organized connections. A large-scale multicentric study in 1710 BD patients and 2594 HC revealed smaller total amygdala volumes only for BD I patients **[17]**. More specifically, Barth et al. **[23]** found a decreased volume of the basal nucleus, accessory basal nucleus, anterior amygdaloid area, and cortico-amygdaloid transition area in BD I, but only a smaller volume of the basal nucleus and the cortico-amygdaloid transition area in BD II, both compared to HC. On the other hand, Bielau et al. **[24]** found no significant differences for the total amygdala volume in a post-mortem study between BD patients and HC. Thus, reduced volumes of the segmented nuclei of the amygdala, rather than total volume, might be considered a potential risk factor for BD. Although structural alterations of hippocampus and amygdala in manifest BD have been described, they have not been analyzed in individuals at risk. It is not understood, if the reduction in volume of the above-mentioned structures represents a risk factor for development of the disease, or rather its long-term consequence. The hippocampus is one of the most plastic brain areas: for example, in one study its volume was larger in long-term meditators compared to controls in many subfields, even up to 15% **[25]**. Interestingly, BD patients treated with mood stabilizers (i.e., taking lithium over 24 months) have larger amygdala and hippocampus volumes than untreated BD patients and cannot be differentiated from HC **[26]**. Indeed, both structures may react sensitively to environmental factors not only via neuroproliferation, but also via neurogenesis with comparable cell turnover rates **[27]**. An analysis of hippocampal subfields and amygdala nuclei in individuals at risk for BD might help us to understand the role of these brain alterations throughout the course of the disease.

While group comparisons using null-hypothesis testing can show mean structural abnormalities, they often cannot be used to draw inferences about individuals **[28]**. On the other hand, multivariate machine learning (ML) approaches, have a higher sensitivity and can potentially improve individual inferences required for clinical diagnostics and prognostics. We have already investigated the use of ML in combination with regional cortical thickness and surface area values as well as subcortical structural volumes (Mikolas et al., in press) **[29]**. Here, we classified subjects at risk of developing BD according to the BPSS-P criteria and achieved a balanced accuracy of 63.1 % using a 10-fold cross-validation.

In this pre-registered analysis of the same sample of help-seeking individuals with risk factors for BD, we first analyzed volume differences of hippocampal subfields and amygdala nuclei using statistical between-group comparisons. As the main analysis, we used a linear support vector machine (SVM) to classify subjects in different BD risk groups. SVMs are a widely used algorithm in medical research **[30]**. The risk stratification into risk states/syndromes for BD was performed using three assessment tools: BARS **[10]**, EPI*bipolar* **[11]** and BPSS-P **[12]**.

2. Materials and Methods

2.1 Pre-registration

The study was pre-registered via the open science framework (https://osf.io/xz9vt).

2.2 Sample

The data was collected as part of the Early-BipoLife project **[31,32]**. Early-BipoLife is a multi-centric, naturalistic, prospective-longitudinal observational cohort study of adolescents and young adults (age 15 – 35 years) at risk for BD. It was performed at ten German university centers and teaching hospitals with early detection facilities. For this analysis we only used the data acquired at baseline. MRI acquisitions were performed on seven of the ten study sites: Berlin, Bochum, Frankfurt, Hamburg, Dresden, Marburg, Tübingen. Of the total *N* = 1229 recruited adolescents and young adults included in this study, *N* = 313 opted to receive an MRI. For a detailed look on final sample sizes, see *2.6 Quality control and data exclusion*.

The recruitment process comprised three different pathways with the following inclusion criteria: In the first pathway (*N* = 123), subjects consulted an early recognition center and had a presence of at least one BD risk factor (family BD history, (sub)threshold affective symptomatology/depressive syndrome, hypomanic/mood swings, disturbances of circadian rhythm/sleep, and other clinical hints). The second pathway (*N* = 146) comprised in- or outpatients with a depressive syndrome (major depressive disorder, dysthymic disorder, cyclothymic disorder, minor depressive disorder, recurrent brief depressive disorder, adjustment disorder with depressed mood, depressive disorder not otherwise specified). The third pathway (*N* = 44) comprised in- or outpatients with a clinically confirmed ADHD diagnosis. ADHD patients were included since the combined subtype of ADHD, with both inattentive and hyperactive / impulsive symptoms was previously associated with BD **[11]**. For complete inclusion and exclusion criteria see *Appendix A*.

The age criterion was based on available studies on age of onset and time to diagnosis. About 75% of individuals with BD I would develop the disorder before the age of 42 years **[33]** and about 70 % would develop any BD by the age of 21 **[1]**. On the other hand, in the health care system of Germany it takes 12.4 years on average to establish the diagnosis since the onset of first symptoms **[34]**. Failure to establish the correct diagnoses occurs mainly due to the predominance of depressive symptoms, as well as unrecognized hypomania **[1,12,35]**. Thus, we extended the age criterion to 35 years to include older individuals with unrecognized BD.

Informed consent was obtained from all subjects involved in the study. Additionally, parents of adolescents gave their informed consent about their children's participation. The study was approved by the Ethics Committee of the Medical Faculty of the Technische Universität Dresden (No: EK290082014, 05.02.2015), as well as local ethics committees at each study site.

2.3 Risk assessment instruments

We assessed the risk state for BD with three independent assessment tools. The Semistructured Interview for Bipolar at Risk States (SIBARS **[10]**) is an assessment tool developed for young people aged 15–35 covering five different subscales: Subthreshold mania, depression, cyclothymic features, genetic risk and mood swings. While cyclothymic features and genetic risk are binary categories, the three other scales are measured with both a severity and a frequency score. The interview items for these subscales were adapted from established rating scales, decisions about inclusion were made by clinical expertise from the authors **[10]**. Experienced clinicians had a training how to apply the instrument and discussed the scoring under supervision. The instrument achieved a prognostic accuracy of Harrell's $C = 0.742$ in a sample of subjects with a high risk for psychosis **[10]**.

The Early Phase Inventory for Bipolar Disorders (EPI*bipolar* **[11]**) is a semi-structured interview, developed based on a systematic literature review. It covers a broad range of risk factors for BD. To quantify the BD risk, all risk factors are weighted and classified into either primary symptoms (such as genetic risk, increasing cyclothymia dynamic and prodromal (hypo-)manic symptoms) or secondary symptoms (such as specific changes in sleep and circadian rhythm, substance use, (suspected) diagnosis of ADHD, impairment of psychosocial functioning, manifest or earlier affective disorder other than BD, or fearfulness/anxiety. Finally, subjects are categorized into four different risk groups: No risk at present, risk status (at least one secondary symptom accompanied by specific changes in sleep and circadian rhythm), high risk status (one primary and at least one secondary symptom, and ultra-high risk (more than one primary symptom). In later analyses, the high risk and ultra-high-risk groups were fused, since the high-risk group contained a disproportionally low number of subjects (3.2%) **[36]**. Therefore, we used the EPI*bipolar* version with three risk categories: no risk, low-risk and high-risk across this manuscript. The Bipolar Prodrome Symptom Scale—Prospective BPSS-P **[12]** is a semi-structured interview assessing three scales: Mania (ten questions), depression (twelve questions) and a general symptom index (nine questions). Each question is rated between absent (zero) and extreme (six). The symptoms and scales are based on DSM-IV criteria and other established rating scales. Internal consistency was Cronbach's α= 0.87 for mania, $\alpha = 0.89$ for depression and $\alpha = 0.74$ for the general symptom index **[12]**. Also, it showed good inter-rater reliability (*ICC* = 0.939 for the total score) as well as convergent validity to comparable scales **[12]**.

2.4 MRI acquisition and preprocessing

All seven utilized MRI scanners had a field strength of 3 Tesla. Six of the seven were manufactured by Siemens and one scanner in Bochum by Philips. For a detailed explanation of hardware, software and scanner protocols please see the study protocol by Vogelbacher et al. **[37]**.

T1-weighted structural scans were preprocessed using Freesurfer 7.1.1. **[38–41]** running on the Centre for Information Services and High-Performance Computing (ZIH) by TU Dresden (https://tu-dresden.de/zih/hochleistungsrechnen). The ZIH offers a virtual environment specifically designed for parallel, data-intensive applications, holding about 60,000 cores. Here, we wrote a bash shell prompting parallel processing within Freesurfer 7.1.1 for all our subjects in a virtual Linux environment (see *Appendix B*). Since the ZIH did not hold the required MATLAB runtime necessary for the Freesurfer pipeline segmentation of hippocampal subfields and nuclei of the amygdala, we had to run this part at a remote computer. Here, we only could install Freesurfer 7.2.0 using Windows Subsystem for Linux (WSL 2). However, the automated segmentation pipeline is the same for all versions of Freesurfer 7 or higher. Thus, this change of software version did not have any effect on our processed data. The segmentation process is a widely used algorithm computing probabilistic atlases to localize the independent nuclei and subfields for each subject **[40–42]**. These statistical atlases were obtained from ultra-high resolution, ex vivo MRI scans. After a first manual segmentation and a manual annotation of adjacent areas using T1-weighted MRI scans, the developers of the software built a computational atlas using Bayesian inference **[40]**. Each node of this computational atlas contains probabilistic information about its assignment to each relevant subregion **[43]**. The segmentation of a previous unfamiliar dataset is then solved as "[…] a Bayesian inference problem of maximizing the probability of the segmentation—given the atlas and the input image" **[44]**. Summarized, we decided to use the Freesurfer segmentation pipeline to A) enhance reproducibility to other BD study groups, for instance ENIGMA and B) due to its good reliability metrics, such as an overall high test-retest reliability, with an *ICC* larger than 0.5 and most studies reporting an *ICC* around 0.9 **[44]**. For a detailed explanation on quality control, see below.

2.5 Measured variables

We assessed the risk for BD using three independent tools. Results were binary variables no risk vs. risk for BARS and BPSS-P and three risk groups for EPI*bipolar*. Further, we estimated the volumes of hippocampal subfields and nuclei of the amygdala with the above explained pipeline. The output consists of three main outputs, twelve hippocampal subfields and nine nuclei of the amygdala. For an overview, see Table 1. For an overview of graphical representation of the parcellations we recommend Tesli et al. (2020) **[45]**. For the medical history, we collected data on psychiatric medication at baseline (yes/no for five classes of medication: antidepressants, antipsychotics, mood stabilizers, anxiolytics and hypnotics, psychostimulants), as well as smoking status (never smoked / current smoker / past smoker). Data on present and lifetime use of cannabis was collected, since it previously was associated with a smaller hippocampus **[46]** (no use / <1 x month / ~1 x month / 2-9 x month / ≥10x/month). Early life adversities are typically associated with alterations of hippocampal subfield volumes **[47–50]**. We estimated it using the total score of the Child Trauma Questionnaire (CTQ) questionnaire, which consists of 25 five-point Likert scale items with five subscales on different traumatic experiences and three extra items on trivialization **[51]**.

Table 1. Output volumes of the segmented subcortical structures.

Main output	Hippocampal subfields	Nuclei of the amygdala
Hippocampus (total volume)	Hippocampal tail	Lateral nucleus
Amygdala (total volume)	Subiculum	Basal nucleus
Intracranial volume (ICV)	Hippocampal fissure	Central nucleus
	Presubiculum	Medial nucleus
	Parasubiculum	Cortical nucleus
	Molecular Layer	Accessory basal nucleus
	Granule cell layer of the dentate gyrus (GC ML DG)	Paralaminar nucleus
	CA1	Corticoamygdaloid transition area
	CA2/3	Anterior amygdaloid area
	CA4	
	Fimbria	
	Hippocampal amygdala transition area (HATA)	

[1] Note. All parcelled volumes are given bilaterally.

2.6 Quality control and data exclusion

For quality assurance, the MRI images were analyzed using the MRIQC tool [52]. For this dataset a visual inspection by two authors was performed. A set of *N* = 23 subjects was excluded from further analysis due to strong movement, ghosting or fold-over artifacts, which gives us remaining *N* = 290. The specific MRI quality control protocol is described elsewhere [37].

After preprocessing the T1 images, we performed a standardized quality control of the segmentation of hippocampal subfields and nuclei of the amygdala according to the established protocols of the ENIGMA working group (http://enigma.ini.usc.edu/protocols/imaging-protocols). Thus, we ran an RStudio script (Rstudio Team, 2022.07.2, Boston, MA) computing outliers and a ranking analysis for the subfield and nuclei volumes. After generating an HTML file containing several pictures for each subject and hemisphere in three different planes, two investigators performed visual quality control independently. The focus was on the subjects statistically identified as outliers. Here, we discarded *N* = 4 subjects: one due to failed preprocessing and three because of major segmentation errors. Other reasons for exclusion were failed preprocessing (*N* = 4), drop-out (*N*=4), no longer wished to participate (*N*=2), baseline data incomplete (*N*=1) and missing data in the three respective risk assessment tools (*N* = 11 for BPSS-P, *N* = 13 for BARS and *N*= 4 for EPI*bipolar*). This resulted in final sample sizes of *N* = 264 (BPSS-P), *N* = 262 (BARS) and *N* = 271 (EPI*bipolar*). For the covariates, we imputed the sample mean (continuous variables) or mode (discrete variables) using IBM SPSS Statistics (Version 28.0.1.1, Armonk, NY) since both LME and SVM cannot handle missing data. This was the case for smoking status (*N* = 2), cannabis use lifetime (*N* = 3), CTQ score (*N* = 13) and psychiatric medication (*N* = 4). The covariates gender, age, cannabis use present, education and study site were available for all subjects.

2.7 Statistical Analysis

We compared demographic and clinical variables between the respective risk score groups (low vs. high risk for BPSS-P and BARS, no risk vs. low-risk vs. high-risk for EPI*bipolar*) using $\chi 2$ analysis for categorical data and t-tests or Wilcoxon rank-sum tests for continuous data (depending on whether the data are normally distributed or not). Then, we tested if the volumes of the whole hippocampus and the whole amygdala differed significantly between high and low BD risk scores. To achieve this, we calculated six independent models: One for each combination of the three risk assessment tools with hippocampus and amygdala, respectively. To achieve reproducibility, we followed the statistical analysis pipeline from a previously published paper on hippocampal subfields in BD patients versus HC, where they also used a linear mixed effects model **[16]**. For the three LME models predicting the hippocampus, we included its total volume as a dependent variable and BD risk score (binary factor for BPSS-P and BARS and three-category factor for EPI*bipolar*) as a fixed effect variable. Sex, age, sex*age (factors, fixed), scanner site (factor, random), CTQ total (continuous), cannabis use present and lifetime (continuous), ICV (continuous) and medication (binary) were included as covariates. Sex, age and its interaction were included, because the volume of the hippocampus follows a non-linear age-related trend **[53]**, its subfields differ between men and women even after adjusting for total hippocampal volume and brain size **[54]** and the sex differences in age-related volume changes are not so clear yet. The same procedure was followed for the amygdala, where its total volume was used as a dependent variable. We further followed the previously published study and averaged the volumes of both hemispheres, since there was no theoretical consideration of potential lateral differences and to reduce number of tests **[18]**. For each significant model, we performed post-hoc LME models with above mentioned covariates and variables of interest using each of the twelve subfields (in case of hippocampus) or each of the nine nuclei (in the case of amygdala). To achieve a balance between reducing false-positive results and gaining an acceptable test power, we corrected for multiple testing using the false discovery rate (FDR by Benjamini and Hochberg **[55]**). We considered FDR $q < 0.05$ as significant. All LME models were calculated using IBM SPSS Statistics (Version 28.0.1.1, Armonk, NY), while the FDR correction was performed with a MATLAB script (R2022a, Natick, Massachusetts), see *Appendix C*. Since the three risk assessment tools differ, we decided to correct for each tool independently. Thus, we corrected for two p-values during the main analysis (total hippocampus and amygdala volumes) and 21 *p*-values during the explorative analysis (twelve hippocampal subfields and nine nuclei of the amygdala volumes) for each tool respectively.

2.8 Machine learning analysis

To increase reproducibility and comparability with a previous study of structural MRI in BD by the ENIGMA consortium, we used a linear support vector machine (SVM) **[56]**. An advantage of a linear SVM is, that its coefficients can be interpreted as relative measures of feature importance, which allows for explainable models **[56]**. A linear SVM has one single hyperparameter C which determines whether the algorithm allows more misclassifications to avoid overfitting the training data (lower C) or less misclassifications to enhance the accuracy (higher C) **[57]**. We optimized C using nested cross-validation and a grid search method using the following values for C: 1e-5, 1e-4, 1e-3, 1e-2, 1e-1, 1, 10 and 100. As the subjects completed three risk assessment tools, we trained three independent SVMs. We used the volumes of 18 amygdala nuclei (right and left), 24 volumes of hippocampal subfields (right and left) and total volumes of amygdala and hippocampus (right and left), i.e., total of 44 features. To evaluate the SVM performance, we used 10-fold cross-validation, i.e., we divided the sample into ten independent subsets of similar size and trained the model ten times, using all the data of the subsamples except for one, which was left out as validation set **[58]**. To evaluate the generalizability across study sites, we additionally performed a leave-one-site-out cross-validation, training the classifier on the data of six sites and testing on the data from one left-out site in each fold. To account for the imbalanced class distribution within the data we A) kept the class ratio in all folds approximately the same (i.e., stratified cross-validation) B) we used random oversampling of the minority class **[59]** in the training set, so that the class ratios in each fold was balanced. To preserve the train/test separation, we standardized the features separately in the training and testing sets by removing the mean and scaling to unit variance. We evaluated the performance using standard metrics for imbalanced datasets (balanced accuracy, Cohen's kappa, sensitivity, specificity). We report the performance results as the average value on all folds. Similarly, we computed 95% confidence intervals based on the performance results on all folds. We performed binary classifications for each risk assessment instrument separately between subjects who fulfilled vs did not fulfill any risk criterion. For EPI*bipolar*, we pooled the low-risk and the high-risk groups to provide for a binary outcome. As 50 % equals random classification, we considered a classification significant, if the mean classification accuracy across folds minus lower confidence interval was higher than 50%. Additionally, we compared the correctly and incorrectly classified subjects to identify factors potentially influencing the achieved accuracy using $\chi 2$ for medication (yes/no), recruitment pathway, smoking status, cannabis use present / lifetime, study site and MRI scanner used; Mann-Whitney U test for early life stress (CTQ) as CTQ was not normally distributed (Kolmogorov-Smirnov $p < 0.001$). Of note, the recruitment pathways did not automatically represent the distributions of diagnoses ADHD and depression in the sample. Therefore, we also tested for ADHD diagnosis separately. As depressive symptoms were an inherent criterion of all three risk instruments, this comparison was not possible for depression.

3. Results

3.1. Demographics

See Table 2 and Table 3 for detailed information on sociodemographic and clinical variables among groups. The subjects who fulfilled the BD risk syndrome according to BPSS-P (Table 2) did not differ from those who did not fulfill any risk syndrome in any of the demographic variables. The subjects who fulfilled the BD risk syndrome according to BARS (Table 2) were more likely to take medication ($\chi2 = 5.018$, $p = 0.025$), to smoke ($\chi2 = 6.529$, $p = 0.038$) and to have entered the study via the depression recruitment pathway, but less likely via the ADHD pathway ($\chi2 = 8.823$, $p = 0.012$) than those who did not fulfill any risk syndrome. The subjects categorized to the low-risk and high-risk groups according to EPI*bipolar* (Table 3) both were more likely to take medication ($\chi2 = 8.077$, $p = 0.018$) and to have entered the study via the depression recruitment pathway, but less likely via the ADHD recruitment pathway ($\chi2 = 23.707$, $p < 0.001$), than those who did not fulfill any risk syndrome.

Table 2. Socio-demographic characteristics for BPSS-P and BARS.

Risk assessment instrument	**BPSS-P (*N* = 264)**			**BARS (*N* = 262)**		
Risk criterion fulfilled	**No**	Yes	Test	No	Yes	Test
N (%)	205 (77.7)	59 (22.3)	n/a	74 (28.2)	188 (71.8)	
Female (%)	93 (45.4)	34 (57.6)	$\chi2 = 2.759$, $p = .097$	35 (47.3)	91 (48.4)	$\chi2 = .026$, $p = .872$
Age (SD)	24.88 (4.2)	24.54 (4.7)	$t = -.532$, $df = 262$, $p = .595$	24.39 (3.7)	25.03 (4.6)	$t = 1.075$, $df = 260$, $p = .283$
Education high school (%)	165 (80.5)	41 (69.5)	$\chi2 = 3.232$, $p = .072$	62 (83.8)	142 (75.5)	$\chi2 = 2.098$, $p = .148$
Recruitment pathway						
Early recognition (%)	91 (44.4)	20 (33.9)	$\chi2 = 2.076$, $p = .354$	35 (47.3)	77 (41.0)	
Depression (%)	87 (42.4)	30 (50.8)		23 (31.1)	91 (48.4)	**$\chi2 = 8.823$, $p = .012*$**
ADHD (%)	27 (13.2)	9 (15.3)		16 (21.6)	20 (10.6)	
Psychiatric Medication						
Yes (%)	111 (54.1)	39 (66.1)	$\chi2 = 2.669$, $p = .102$	34 (45.9)	115 (61.2)	**$\chi2 = 5.018$, $p = .025*$**
Substance Use						
Smoking status						
• Never smoked (%)	97 (47.3)	20 (33.9)		42 (56.8)	74 (39.4)	
• Current smoker (%)	94 (45.9)	31 (52.5)	$\chi2 = 4.784$, $p = .091$	27 (36.5)	95 (50.5)	**$\chi2 = 6.529$, $p = .038*$**
• Past smoker (%)	14 (6.8)	8 (13.6)		5 (6.8)	19 (10.1)	
Cannabis present						
• No use (%)	147 (71.7)	45 (76.3)		61 (82.4)	129 (68.6)	
• < 1x/month (%)	17 (8.3)	3 (5.1)		4 (5.4)		
• ~ 1x/month (%)	12 (5.9)	2 (3.4)	$\chi2 = 3.836$, $p = .429$	3 (4.1)	16 (8.5)	$\chi2 = 5.350$, $p = .253$
• 2-9x/month (%)	15 (7.3)	2 (3.4)		3 (4.1)	11 (5.9)	
• ≥10x/month (%)						

	14 (6.8)	7 (11.9)		3 (4.1)	14 (7.4) 18 (9.6)	
Cannabis lifetime						
• No use (%)	84 (41.0)	23 (39.0)		38 (51.4)	68 (36.2)	
• < 1x/month (%)	46 (22.4)	11 (18.6)		16 (21.6)	40 (21.3)	
• ~1x/month (%)	9 (4.4)	2 (3.4)	*χ*2 = 1.977, *p* = .740	2 (2.7)	9 (4.8)	*χ*2 = 6.532, *p* = .163
• 2-9x/month (%)	24 (11.7)	6 (10.2)		7 (9.5)	24 (12.8)	
• ≥10x/month (%)	42 (20.5)	17 (28.8)		11 (14.9)	47 (25.0)	

p* ≤ 0.05; ** *p* ≤ 0.01; * *p* ≤ 0.001.

Table 3. Socio-demographic characteristics for *EPIbipolar*.

Risk assessment instrument	**EPIb*ipolar* (*N* = 271)**			
Risk criterion fulfilled	**No-risk**	**Low-risk**	High-risk	Test
N (%)	30 (11.1)	136 (50.2)	105 (38.7)	n/a
Female (%)	10 (33.3)	62 (45.6)	57 (54.3)	*χ*2 = 4.550, *p* = .103
Age (SD)	24.13 (3.03)	25.40 (4.61)	25.02 (4.34)	*F* = .570, *df* = 2, *p* = .566
Education high school (%)	24 (80.0)	107 (78.7)	81 (77.1)	*χ*2 = .144, *p* = .931
Recruitment pathway				
Early recognition (%)	14 (46.7)	50 (36.8)	51 (48.6)	***χ*2 = 23.707, *p* < .001****
Depression (%)	5 (16.7)	72 (52.9)	43 (41.0)	
ADHD (%)	11 (36.7)	14 (10.3)	11 (10.5)	
Psychiatric Medication				
Yes (%)	11 (36.7)	87 (64.0)	57 (54.3)	***χ*2 = 8.077, *p* = .018***
Substance Use				
Smoking status				
• Never smoked (%)	16 (53.3)	64 (47.1)	38 (36.2)	
• Current smoker (%)	9 (30.0)	64 (47.1)	56 (53.3)	*χ*2 = 8.771, *p* = .067
• Past smoker (%)	5 (16.7)	8 (5.9)	11 (10.5)	
Cannabis present				
• No use (%)	25 (83.3)	96 (70.6)	76 (72.4)	
• < 1x/month (%)	1 (3.3)	11 (8.1)	10 (9.5)	
• ~ 1x/month (%)	0 (0.0)	8 (5.9)	6 (5.7)	*χ*2 = 4.647, *p* = .795
• 2-9x/month (%)	2 (6.7)	8 (5.9)	7 (6.7)	
• ≥10x/month (%)	2 (6.7)	13 (9.6)	6 (5.7)	
Cannabis lifetime				
• No use (%)	14 (46.7)	52 (38.2)	44 (41.9)	
• < 1x/month (%)	8 (26.7)	29 (21.3)	21 (20.0)	
• ~1x/month (%)	0 (0.0)	6 (4.4)	5 (4.8)	*χ*2 = 3.173, *p* = .923

• 2-9x/month (%)	3 (10.0)	17 (12.5)	11 (10.5)
• ≥10x/month (%)	5 (16.7)	32 (23.5)	24 (22.9)

Note. *$p \leq 0.05$; ** $p \leq 0.01$; *** $p \leq 0.001$.

3.2. Statistical Analysis

There were no significant results for all LME models in the main analysis after correcting for multiple comparisons. Using BPSS-P, subjects did not show structural alterations in the total hippocampus $F(1, 251.458) = 0.088$, $p = 0.872$, nor in the total amygdala $F(1, 257.812) = .026$, $p = 0.872$. If categorized by EPI*bipolar* in three risk groups, there was neither a volume difference for the total hippocampus $F(2, 260.590) = 1.104$, $p = 0.333$, nor for the total amygdala $F(2, 262.751) = 3.107$, $p = 0.092$. Similarly, BARS categorization did not yield to different structural volumes between risk groups for the total hippocampus $F(1, 253.214) = 0.541$, $p = 0.670$ nor for the total amygdala $F(1, 255.927) = 0.182$, $p = 0.670$. For a view on descriptive volume differences between all groups, see the boxplot diagram *Appendix D*.

Since there was no significant difference in the main analysis, we did not perform any post-hoc tests for the total volumes and did not analyze the subfields and nuclei as part of our confirmatory hypothesis. To have a closer look on potential differences between single subfields and nuclei, we did perform LME models as an explicitly explorative analysis. Here, only the medial nucleus showed a significantly higher volume in subjects fulfilling the BPSS-P risk criterion (F (1, 231.662) = 13.706, $p < 0.05$). For a detailed view on all explorative LME results, see *Appendix E*.

3.3. Machine learning analysis

Detailed information on outcomes of the SVM for all three risk assessment tools can be found in Table 4. No significant predictions could be found for EPI*bipolar* and BARS. However, using the 10-fold cross-validation for BPSS-P, the SVM could significantly classify subjects at risk for BD with a balanced accuracy of 66.90% (95% *CI* 59.2– 74.6), a Cohen's kappa of 0.275 (95% *CI* 0.149-0.401), a sensitivity of 63% (95% *CI* 49.7-76.3) and a specificity of 63.0% (95% *CI* 49.7-76.3). There were no significant differences between subjects categorized at risk and those not at risk in terms of age ($df = 262$, $t = 1.545$, $p = 0.417$), sex ($df = 1$, $\chi2 = 0.148$, $p = 0.7$), medication ($df = 1$, $\chi2 = 1.182$, $p = 0.669$), recruitment pathway ($df = 2$, $\chi2 = 2.215$, $p = 0.33$), first degree relatives ($df = 1$, $\chi2 = 0.66$, $p = 0.797$), ADHD diagnosis ($df = 1$, $\chi2 = 0.673$, $p = 0.412$), early life stress (Mann–Whitney $U = 7309.5$, $p = 0.79$), smoking status ($df = 2$, $\chi2 = 2.575$, $p = 0.276$), cannabis use present (Fisher-Freeman-Halton's exact test $p = 0.972$), cannabis use lifetime (Fisher-Freeman-Halton's exact test $p = 0.730$), and site (Fisher-Freeman-Halton's exact test $p = 0.071$). The falsely and correctly classified subjects differed in scanner type (Fisher-Freeman-Halton's exact test $p = 0.032$). Using the leave-one-site out validation approach for BPSS-P, subjects could also be classified significantly above chance level with a balanced accuracy of 61.9% (95% *CI* 52.0– 71.9), a Cohen's kappa of 0.197 (95% *CI* 0.033-0.361), a sensitivity of 45.0% (95% *CI* 17.7-72.2) and a specificity of 78.9% (95% *CI* 60.2-97.7). There were no significant differences between subjects categorized at risk and those not at risk in terms of age ($df = 262$, $t = 0.522$, $p = 0.602$), sex ($df = 1$, $\chi2 = 0.1167$, $p = 0.28$), medication ($df = 1$, $\chi2 = 0.051$, $p = 0.822$), first degree relatives ($df = 1$, $\chi2 = 1.625$, $p = 0.202$), ADHD diagnosis ($df = 1$, $\chi2 = 0.338$, $p = 0.561$), early life stress (Mann–Whitney $U = 7001.0$, $p = 0.375$), cannabis use present (Fisher-Freeman-Halton's exact test $p = 0.883$), cannabis use lifetime (Fisher-Freeman-Halton's exact test $p = 0.736$), site (Fisher-Freeman-Halton's exact test $p = 0.08$) and scanner type (Fisher-Freeman-Halton's exact test $p = 0.615$). The falsely and correctly classified subjects ($N = 83$) differed in recruitment pathway ($df = 2$, $\chi2 = 6.838$, $p = 0.033$) and smoking status ($df = 2$, $\chi2 = 5.953$, $p = 0.51$). More specifically, the subjects recruited via the early recognition pathway were more frequently falsely classified as high-risk (35.2% versus 24.1% of all falsely classified) whereas the subjects recruited via the depression and ADHD pathways were more frequently falsely classified as no-risk (55.2% versus 46.3% and 20.7% versus 18.5% respectively). The non-smokers were more frequently classified as high risk (48.1% versus 27.6 of all falsely classified) whereas the current smokers and past-smokers were more frequently classified as no-risk (51.7% versus 40.7% and 20.7% versus 11.1 % respectively). Past or actual smoking was more frequent among the falsely classified subjects recruited via depression and ADHD pathways (23.7% early recognition, 68.1% depression, 72.2% ADHD). The distributions of model weights (i.e., contributions of individual features) among both 10-fold and leave-one-site-out approaches remained consistent (see Figure 1).

Table 4. Performance metrics of the linear SVM classification for all three risk assessment tools.

	Cohen's kappa (%)			Balanced accuracy (%)			Sensitivity (%)			Specificity (%)		
		95% CI			95% CI			95% CI			95% CI	
		lower	**upper**		lower	upper		lower	upper		lower	upper
BPSS-P												
10-fold	.275	.149	.401	66.9	59.2	74.6	63.0	49.7	76.3	63.0	49.7	76.3
Leave-one-site-out	.197	.033	.361	61.9	52.0	71.9	45.0	17.7	72.2	78.9	60.2	97.7
BARS												
10-fold	-.001	-.132	.129	49.2	41.9	56.5	56.2	44.2	68.3	42.1	35.7	48.5
Leave-one-site-out	-.000	-.103	.103	48.2	40.1	56.2	53.8	41.9	65.7	42.5	30.7	54.4
EPI*bipolar*												
10-fold	-.049	-.143	.045	45.0	35.3	54.8	66.8	58.0	75.5	23.3	7.2	39.4
Leave-one-site-out	-.027	-.203	.148	46.2	35.6	56.8	62.4	44.4	80.5	30.0	15.2	44.8

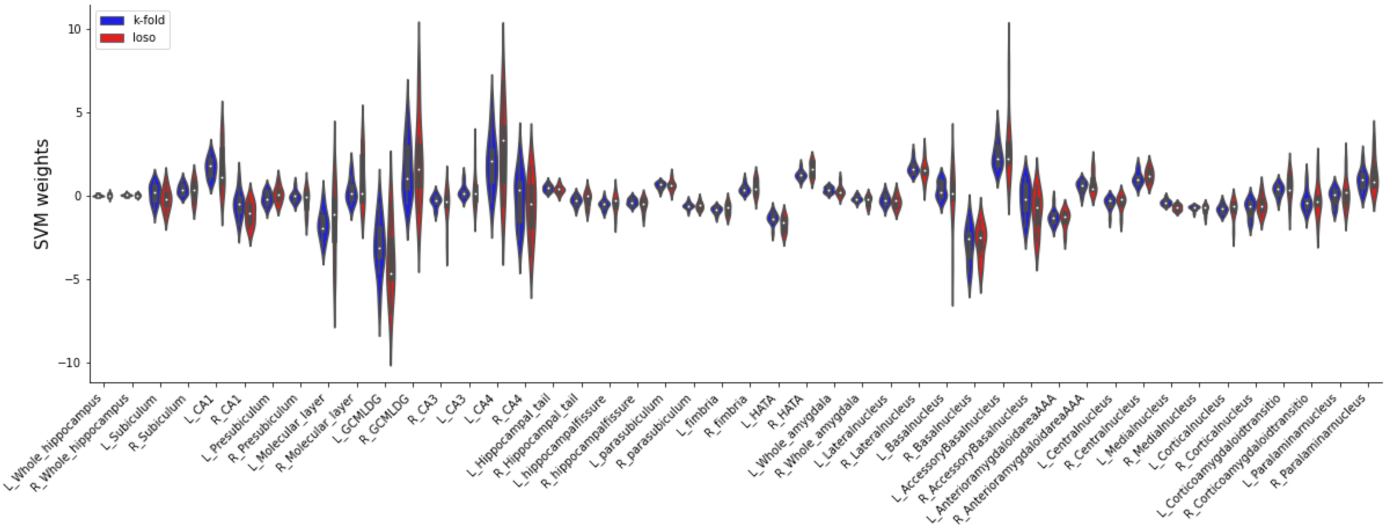


Figure 1. Comparison of the contribution of individual features to the SVM classification between 10-fold and leave-one-site-out classifications. The violin plots represent the SVM weights for each single feature across all folds. The similarity of both patterns suggests, that the SVM relied on similar structural patterns in both cross-validation methods.

4. Discussion

To our best knowledge, this is the first study investigating volume differences for the hippocampus, amygdala and their subfields/nuclei in help-seeking individuals fulfilling a risk criterion for BD. We found no significant structural differences for hippocampus and amygdala between subjects with no compared to at risk state for BD, using conventional statistics. However, we could classify the individuals fulfilling versus not fulfilling the risk criterion according to BPSS-P using machine learning (i.e., a SVM approach) with a moderate performance. The exploratory analyses revealed a higher volume of the medial nucleus of the amygdala in the subjects fulfilling the BPSS-P risk criterion. The SVM classification of risk criteria according to BARS and EPI*bipolar* was not significant.

Unlike some studies in manifest BD **[18,26,60]** we could not identify significant volume differences in hippocampus, amygdala and/or their subfields/nuclei in subjects fulfilling a risk criterion according to state-of-the art risk assessment tools during confirmatory hypothesis testing. This is in accordance with studies of subjects with genetic risk for BD (i.e., first degree relatives), which have not identified any differences in hippocampal or amygdala volumes compared to HC **[6,60,61]**. This might imply, that individuals at risk for BD may not show structural abnormalities in those limbic regions prior to diagnosis and that the volume reduction of hippocampus and its subfields found in manifest disorder may not be a causal mechanism for BD pathogenesis, since it seems to not happen prior to BD onset.

In contrast with group comparisons using conventional statistics, an SVM could classify the subjects fulfilling the BPSS-P risk criterion with a balanced accuracy of 66.9%, which remained significant in a leave-one-site-out cross-validation. Unlike null hypothesis testing, machine learning techniques might be more appropriate to identify discrete, multivariate differences, which are more representative of psychiatric disorders **[62]**. The classification performance using hippocampal subfields and nuclei of the amygdala was similar to our previous study using regional cortical thickness, cortical surface areas and volumes of subcortical structures (63.1 and 56.2% 10-fold versus leave-one-site-out) (Mikolas et al., in press) **[29]**. The classification of patients with manifest BD versus HC by the ENIGMA consortium performed similarly (65.23% and 58.67% k-fold versus leave-one-site-out) **[56]**. At least five other studies aimed to classify first-degree relatives, healthy subjects and/or other diagnoses and achieved accuracies of 59.7 up to 83.21% **[63]**. However, those studies mostly used small sample sizes and most importantly, did not perform a leave-one-site-out validation, which might favor higher accuracies **[13,62,64]**. Thus, our results suggest, that although not being detected by conventional statistic methods, structural abnormalities in limbic regions are already present in the at-risk state. These may increase after transition to the full-blown disorder and later during the time course of BD development **[17]**. Although the studies in adults with BD suggest volume reductions of amygdala and hippocampus, our exploratory analyses revealed a higher volume of the medial nucleus of the amygdala in our sample of help-seeking participants at risk. The amygdala is a dynamic region which develops continuously until the early adulthood **[65]**. The direction and exact dynamics of structural changes yet need to be understood.

Although demographic confounders seem to not have influenced the classification using the pooled sample (i.e., the 10-fold method), recruitment pathway and smoking status might have influenced the leave-one-site-out classification. Interestingly, subjects recruited via the depression and ADHD pathways were more likely to smoke or have smoked in the past, which suggests a shared effect. However, there were subjects with depression and ADHD also in the group entering the study via the early recognition pathway, which suggests that smoking status might have been the true source of this effect. Indeed, longitudinal evidence showed, that smoking might lead to reduced hippocampal volumes **[66]**. Another confounder was the type of scanner. Seven subjects were scanned on a different scanner type (see Methods), out of which five were falsely classified. Future clinical diagnostic tools based on machine learning should include clinical data, in particular smoking habits. Samples should include more subjects scanned using different scanner types, to allow for better generalization across different scanners.

Compared to our previous study using regional cortical thickness, cortical surface areas and volumes of unparcelled cortical structures (Mikolas et al., in press) **[29]**, balanced accuracies were not much different numerically when using hippocampal subfields and amygdala nuclei as features for the SVM. The latter approach even seems to be more prone to confounders. On the other hand, the benefit of these features should separately be evaluated in future studies using longitudinal data and models should be trained to recognize subjects who transitioned to BD. Whereas the SVM classification based on BPSS-P risk assessment was above chance, the predictions using EPI*bipolar* and BARS failed, which was a similar pattern to Mikolas et al. (in press) **[29]**. BPSS-P assigned noticeably less subjects to the risk group, compared to both other instruments (22.3% for BPSS-P, 71.8% for BARS, 88.9% low-risk and high-risk pooled for EPI*bipolar*). The more conservative classification approach using BPSS-P seems to go along with a higher sensitivity in detection of volume differences in limbic regions. Thus, structural abnormalities in these regions might only be traceable for a specific subgroup with a higher risk state. Also, subjects categorized at risk by EPI*bipolar* and BARS were more often diagnosed with unipolar depression comparing to BPSS-P. This could have led to a higher heterogeneity in the high-risk sample of EPI*bipolar* and BARS, thus impeding the classification sensitivity for BD solely.

The concept of the at-risk state for BD is in development. Studies showed transition rates of 8 and 25% using different criteria with a follow-up length from 1 to 21 years compared to a cumulative lifetime incidence between 1.5- to 2% in the general population **[5]**. Although subjects fulfilling the risk criterion might benefit from psychiatric treatment options based on the symptoms they display, better prediction rates are necessary to aid specific clinical decisions, such as initiating a mood stabilizer. Structural MRI combined with longitudinal follow-up may improve prediction rates for conversion to manifest disease, especially through studies with a sufficient conversion ratio in subjects at risk. In this study we had no control group in form of HC to which the risk population could have been compared. Since this was a naturalistic study carried out at university clinics and the recruitment pathways comprised ADHD diagnosis, depression diagnosis or people contacting early recognition centers, the sample was not representative for the general population. However, since people with several psychiatric disorders seem to show structural abnormalities, such as reduced volumes of hippocamps or amygdala, the statistical difference of these subcortical volumes might be underestimated **[68,69]**.

5. Conclusions

Univariate testing did not reveal significant volume differences in hippocampal subfields and amygdala nuclei in subjects fulfilling risk criteria for developing BD. On the other hand, machine learning differentiated between subjects fulfilling versus not fulfilling the BPSS-P risk criteria with a moderate performance. Parcellation of subcortical structures might identify patients who are at risk of developing BD with similar performance to cortical features. The specific benefit of using hippocampal subfields and amygdala nuclei should be evaluated in models using multimodal features and longitudinal data on conversion to BD. However, we found that a ML approach can be more sensitive in early illness detection than classic null-hypothesis testing.

Author Contributions: Conceptualization, Fabian Huth, Leonardo Tozzi, Michael Marxen, Philipp Riedel, Andreas Jansen and Pavol Mikolas; Data curation, Kyra Bröckel, Christina Berndt and Christoph Vogelbacher; Funding acquisition, Philipp Ritter, Michael Bauer and Andrea Pfennig; Investigation, Fabian Huth, Cathrin Sauer and Pavol Mikolas; Methodology, Fabian Huth, Michael Marxen, Philipp Riedel, Andreas Jansen and Pavol Mikolas; Supervision, Philipp Ritter, Michael Bauer and Andrea Pfennig; Writing – original draft, Fabian Huth; Writing – review & editing, Leonardo Tozzi, Michael Marxen, Philipp Riedel, Kyra Bröckel, Julia Martini, Christina Berndt, Christoph Vogelbacher, Tilo Kircher, Irina Falkenberg, Florian Odenthal, Martin Lambert, Vivien Kraft, Gregor Leicht, Christoph Mulert, Andreas Fallgatter, Thomas Ethofer, Anne Rau, Karolina Leopold, Andreas Bechdolf, Andreas Reif, Silke Matura, Silvia Biere, Felix Bermpohl, Jana Fiebig, Thomas Stamm, Christoph U. Correll, Georg Juckel, Vera Flasbeck, Philipp Ritter, Michael Bauer, Andrea Pfennig and Pavol Mikolas. All authors have read and agreed to the published version of the manuscript.

Funding: Early-BipoLife is funded by the Federal Ministry of Education and Research (BMBF, grant numbers: 01EE1404A, 01EE1404E and 01EE1404F). MM was supported by the Deutsche Forschungsgemeinschaft (DFG grant Nos. 178833530 [SFB 940] and 402170461 [TRR 265].

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Medical Faculty of the Technische Universität Dresden (No: EK290082014, 05.02.2015), as well as local ethics committees at each study site.

Informed Consent Statement: We obtained a written informed consent after comprehensive information about study aims and procedures. Additionally, parents of adolescents gave their informed consent about their children's participation.

Data Availability Statement: Data available upon reasonable request.

Acknowledgments: The author gratefully acknowledges the GWK support for funding this project by providing computing time through the Center for Information Services and HPC (ZIH) at TU Dresden and the Early-Bipolife consortium for collecting and providing the data.

Conflicts of Interest: K. Leopold has been a consultant and/or advisor to or has received honoraria from: Janssen/J&J, Lundbeck, Otsuka, Recordati and ROVI. She has received grant support from Janssen/J&J and Otsuka. All other authors declared no conflict of interest.

Appendix A. Inclusion and exclusion criteria for the three recruitment pathways.

Inclusion criteria for each pathway:

1. Youth and young adults consulting early recognition centres/facilities:

- Age: 15 to 35 years
- Consultation of an early recognition centre/facility
- Presence of at least one of the proposed risk factors for bipolar disorder: Family history of bipolar disorder, (sub)threshold affective symptomatology/depressive syndrome, hypomanic/mood swings, disturbances of circadian rhythm/sleep other clinical hints

2. Young individuals with diagnosed depression:

- Age: 15 to 35 years*
- In- or outpatients with a depressive syndrome in the context of: Major depressive disorder, dysthymic disorder, cyclothymic disorder, minor depressive disorder, recurrent brief depressive disorder, adjustment disorder with depressed mood, depressive disorder Not Otherwise Specified (NOS)

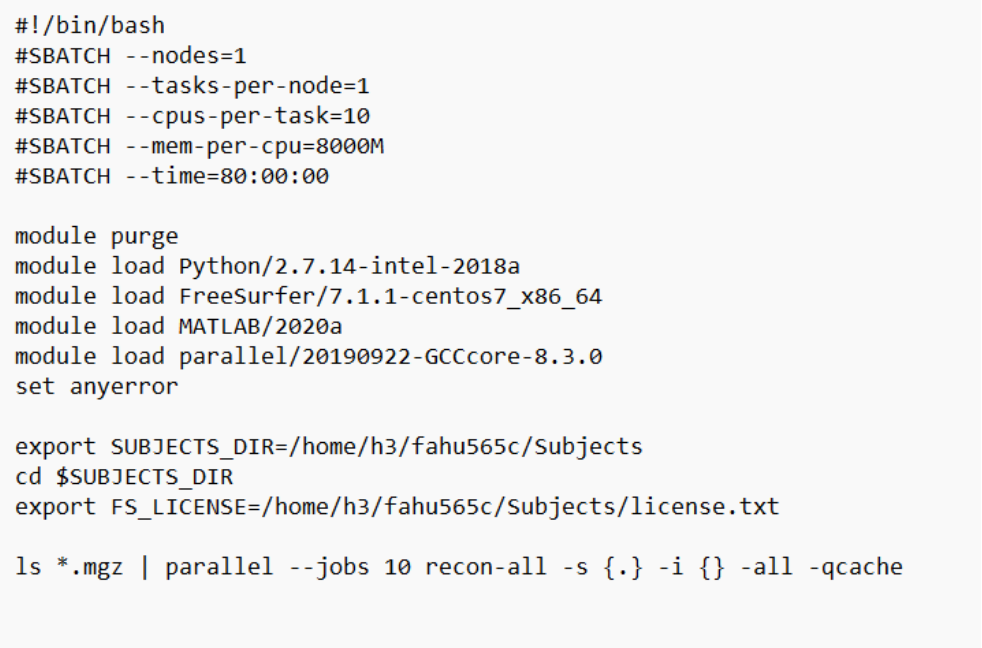
3. Patients with ADHD:

- Age: 15 to 35 years
- In- or outpatients with a clinically confirmed ADHD diagnosis

Exclusion criteria for all three pathways:

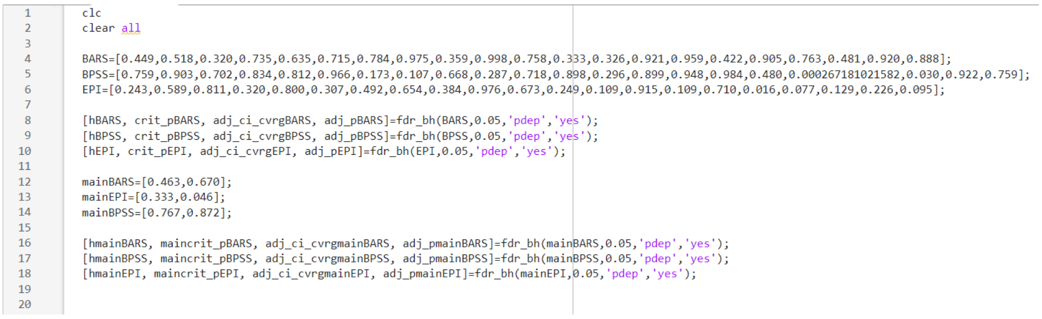
- Diagnosis of bipolar disorder, schizoaffective disorder, schizophrenia
- Diagnosis of anxiety, obsessive–compulsive or substance dependence disorder that fully explains the whole symptomatology
- Limited ability to comprehend the study
- Implied expressed negative declaration of intent to participate in the study by a minor and
- Acute suicidality

Appendix B. Bash shell script in a virtual Linux environment at the high-performance computing (ZIH) prompting a parallel preprocessing of the subcortical segmentation procedure.



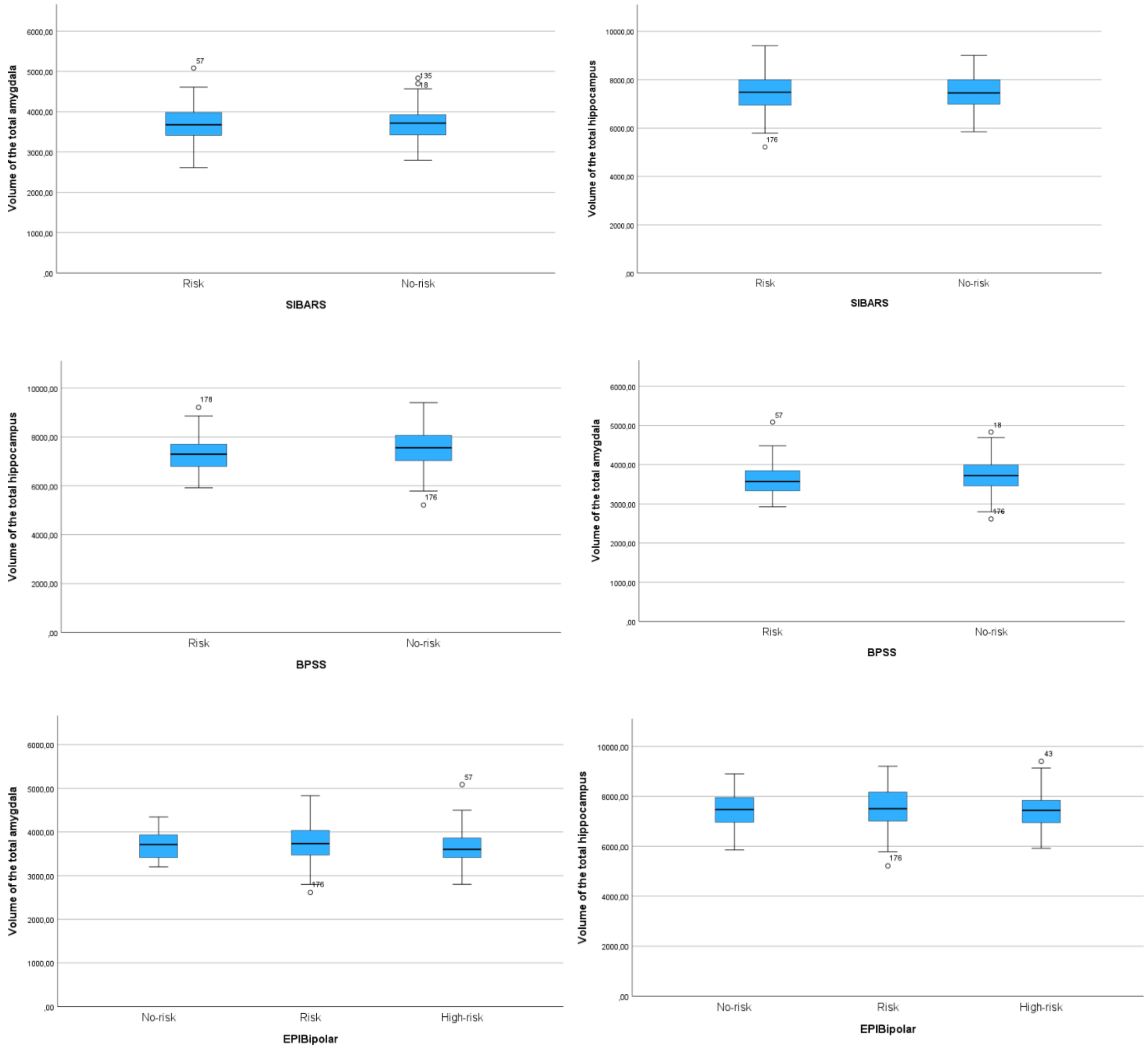
```
#!/bin/bash
#SBATCH --nodes=1
#SBATCH --tasks-per-node=1
#SBATCH --cpus-per-task=10
#SBATCH --mem-per-cpu=8000M
#SBATCH --time=80:00:00

module purge
module load Python/2.7.14-intel-2018a
module load FreeSurfer/7.1.1-centos7_x86_64
module load MATLAB/2020a
module load parallel/20190922-GCCcore-8.3.0
set anyerror

export SUBJECTS_DIR=/home/h3/fahu565c/Subjects
cd $SUBJECTS_DIR
export FS_LICENSE=/home/h3/fahu565c/Subjects/license.txt

ls *.mgz | parallel --jobs 10 recon-all -s {.} -i {} -all -qcache
```

Appendix C. MATLAB script for calculating the FDR of the explorative analysis.



```
clc
clear all

BARS=[0.449,0.518,0.320,0.735,0.635,0.715,0.784,0.975,0.359,0.998,0.758,0.333,0.326,0.921,0.959,0.422,0.905,0.763,0.481,0.920,0.888];
BPSS=[0.759,0.903,0.702,0.834,0.812,0.966,0.173,0.107,0.668,0.287,0.718,0.898,0.296,0.899,0.948,0.984,0.480,0.000267181021582,0.030,0.922,0.759];
EPI=[0.243,0.589,0.811,0.320,0.800,0.307,0.492,0.654,0.384,0.976,0.673,0.249,0.109,0.915,0.109,0.710,0.016,0.077,0.129,0.226,0.095];

[hBARS, crit_pBARS, adj_ci_cvrgBARS, adj_pBARS]=fdr_bh(BARS,0.05,'pdep','yes');
[hBPSS, crit_pBPSS, adj_ci_cvrgBPSS, adj_pBPSS]=fdr_bh(BPSS,0.05,'pdep','yes');
[hEPI, crit_pEPI, adj_ci_cvrgEPI, adj_pEPI]=fdr_bh(EPI,0.05,'pdep','yes');

mainBARS=[0.463,0.670];
mainEPI=[0.333,0.046];
mainBPSS=[0.767,0.872];

[hmainBARS, maincrit_pBARS, adj_ci_cvrgmainBARS, adj_pmainBARS]=fdr_bh(mainBARS,0.05,'pdep','yes');
[hmainBPSS, maincrit_pBPSS, adj_ci_cvrgmainBPSS, adj_pmainBPSS]=fdr_bh(mainBPSS,0.05,'pdep','yes');
[hmainEPI, maincrit_pEPI, adj_ci_cvrgmainEPI, adj_pmainEPI]=fdr_bh(mainEPI,0.05,'pdep','yes');
```

Appendix D. Boxplot diagrams comparing the volumes of the total hippocampus and of the total amygdala for all three risk assessment tools.



Appendix E. LME results for the explorative analysis of all twelve hippocampal subfields and nine nuclei of the amygdala for all three risk assessment tools.

	BPSS-P	BARS	EPI*bipolar*
Hipoocampal subfields			
CA1	*F*(1, 258.392) = 0.015, *p* = 0. 984	*F*(1, 253.719) = 0.418, *p* = 0.998	*F*(2, 261.014) = 1.423 , *p* = 0.581
CA2/3	*F*(1, 257.737) = 0.057, *p* = 0. 984	*F*(1, 253.306) = 0.226, p = 0.998	*F*(2, 260.560) = 0.531, *p* = 0.877
CA4	*F*(1, 256.403) = 0.002, *p* = 0. 984	*F*(1, 252.530) = 0.133, *p* = 0.998	*F*(2, 260.047) = 0.210, *p* = 0.896
Molecular layer	*F*(1, 257.835) = 0.016, *p* = 0. 984	*F*(1, 253.161) = 0.942, *p* = 0.998	*F*(2, 260.538) = 1.143, *p* = 0.611
GC ML DG	*F*(1, 256.400) = 0.044, *p* = 0. 984	*F*(1, 252.529) = 0.115, *p* = 0.998	*F*(2, 260.057) = 0.224, *p* = 0.896
Hippocampal tail	*F*(1, 257.433) = 1.866, *p* = 0. 908	*F*(1, 255.414) = 0.075, *p* = 0.998	*F*(2, 262.303) = 1.186, *p* = 0.611
Subiculum	*F*(1, 258.919) = 0.094, *p* = 0. 984	*F*(1, 253.869) = 0.574, *p* = 0.998	*F*(2, 261.239) = 0.712, *p* = 0.795
Presubiculum	*F*(1, 258.126) = 0.146, *p* = 0. 984	*F*(1, 252.960) = 0.994, *p* = 0.998	*F*(2, 260.245) = 0.425, *p* = 0.877
Parasubiculum	*F*(1, 252.592) = 0.131, *p* = 0. 984	*F*(1, 254.835) = 0.095, *p* = 0.998	*F*(2, 260.532) = 0.962, *p* = 0.672
Fimbria	*F*(1, 238.399) = 0.184, *p* = 0. 984	*F*(1, 256.843) = 0.844, *p* = 0.998	*F*(2, 262.917) = 0.024, *p* = 0.976
Hippocampal fissure	*F*(1, 258.786) = 2.610, *p* = 0. 749	*F*(1, 254.065) = 0.001, *p* = 0.998	*F*(2, 261.224) = 0.397, *p* = 0.877

HATA	$F(1, 259) = 1.138$, $p = 0. 984$	$F(1, 254.478) = 0.000$, $p = 0.998$	$F(2, 261.619) = 1.397$, $p = 0.581$
Nuclei of the amygdala			
Cortical ncl.	$F(1, 247.747) = 4.766$, $p = 0. 315$	$F(1, 256.286) = 0.498$, $p = 0.998$	$F(2, 262.740) = 2.239$, $p = 0.452$
Medial ncl.	**$F(1, 231.662) = 13.706$, $p = 0.0056$****	$F(1, 256.062) = 0.091$, $p = 0.998$	$F(2, 262.792) = 0.089$, $p = 0.961$
Basal ncl.	$F(1, 257.337) = 0.016$, $p = 0. 984$	$F(1, 255.524) = 0.010$, $p = 0.998$	$F(2, 262.493) = 2.233$, $p = 0.452$
Central ncl.	$F(1, 259) = 0.500$, $p = 0. 984$	$F(1, 257) = 0.014$, $p = 0.998$	$F(2, 265) = 0.342$, $p = 0.877$
Lateral ncl.	$F(1, 252.575) = 1.096$, $p = 0. 984$	$F(1, 255.551) = 0.970$, $p = 0.998$	$F(2, 262.244) = 4.214$, $p = 0.336$
Accessory basal ncl.	$F(1, 249.097) = 0.004$, $p = 0. 984$	$F(1, 256.320) = 0.003$, $p = 0.998$	$F(2, 263.123) = 2.586$, $p = 0.452$
AAA	$F(1, 233.022) = 0.000$, $p = 0. 984$	$F(1, 256.675) = 0.648$, $p = 0.998$	$F(2, 263.346) = 2.066$, $p = 0.452$
Corticoamygdaloid transition area	$F(1, 247.829) = 0.009$, $p = 0. 984$	$F(1, 255.937) = 0.010$, $p = 0.998$	$F(2, 262.786) = 1.497$, $p = 0.581$
Paralaminar ncl.	$F(1, 256.680) = 0.095$, $p = 0. 984$	$F(1, 255.558) = 0.020$, $p = 0.998$	$F(2, 262.455) = 2.375$, $p = 0.452$

$^*p \leq 0.05$; $^{**}\ p \leq 0.01$; $^{***}\ p \leq 0.001$

References

1. Merikangas, K.R.; Jin, R.; He, J.-P.; Kessler, R.C.; Lee, S.; Sampson, N.A.; Viana, M.C.; Andrade, L.H.; Hu, C.; Karam, E.G.; et al. Prevalence and Correlates of Bipolar Spectrum Disorder in the World Mental Health Survey Initiative. *Arch Gen Psychiatry* **2011**, *68*, 241–251, doi:10.1001/archgenpsychiatry.2011.12.
2. He, H.; Hu, C.; Ren, Z.; Bai, L.; Gao, F.; Lyu, J. Trends in the Incidence and DALYs of Bipolar Disorder at Global, Regional, and National Levels: Results from the Global Burden of Disease Study 2017. *Journal of Psychiatric Research* **2020**, *125*, 96–105, doi:10.1016/j.jpsychires.2020.03.015.
3. Drancourt, N.; Etain, B.; Lajnef, M.; Henry, C.; Raust, A.; Cochet, B.; Mathieu, F.; Gard, S.; MBailara, K.; Zanouy, L. Duration of Untreated Bipolar Disorder: Missed Opportunities on the Long Road to Optimal Treatment. *Acta Psychiatrica Scandinavica* **2013**, *127*, 136–144.
4. Müller-Oerlinghausen, B.; Berghöfer, A.; Bauer, M. Bipolar Disorder. *The Lancet* **2002**, *359*, 241–247.
5. Keramatian, K.; Chakrabarty, T.; Saraf, G.; Yatham, L. Transitioning to Bipolar Disorder: A Systematic Review of Prospective High-Risk Studies. *Current Opinion in Psychiatry* **2021**, *Publish Ahead of Print*, doi:10.1097/YCO.0000000000000762.
6. Hajek, T.; Cullis, J.; Novak, T.; Kopecek, M.; Blagdon, R.; Propper, L.; Stopkova, P.; Duffy, A.; Hoschl, C.; Uher, R.; et al. Brain Structural Signature of Familial Predisposition for Bipolar Disorder: Replicable Evidence For Involvement of the Right Inferior Frontal Gyrus. *Biological Psychiatry* **2013**, *73*, 144–152, doi:10.1016/j.biopsych.2012.06.015.
7. Kerner, B. Genetics of Bipolar Disorder. *Appl Clin Genet* **2014**, *7*, 33–42, doi:10.2147/TACG.S39297.
8. Hafeman, D.M.; Merranko, J.; Goldstein, T.R.; Axelson, D.; Goldstein, B.I.; Monk, K.; Hickey, M.B.; Sakolsky, D.; Diler, R.; Iyengar, S.; et al. Assessment of a Person-Level Risk Calculator to Predict New-Onset Bipolar Spectrum Disorder in Youth at Familial Risk. *JAMA Psychiatry* **2017**, *74*, 841, doi:10.1001/jamapsychiatry.2017.1763.
9. Post, R.M.; Altshuler, L.L.; Kupka, R.; McElroy, S.L.; Frye, M.A.; Rowe, M.; Grunze, H.; Suppes, T.; Keck, P.E.; Leverich, G.S.; et al. Multigenerational Transmission of Liability to Psychiatric Illness in Offspring of Parents with Bipolar Disorder. *Bipolar Disord* **2018**, *20*, 432–440, doi:10.1111/bdi.12668.
10. Fusar-Poli, P.; De Micheli, A.; Rocchetti, M.; Cappucciati, M.; Ramella-Cravaro, V.; Rutigliano, G.; Bonoldi, I.; McGuire, P.; Falkenberg, I. Semistructured Interview for Bipolar at Risk States (SIBARS). *Psychiatry Research* **2018**, *264*, 302–309.
11. Leopold, K.; Ritter, P.; Correll, C.U.; Marx, C.; Özgürdal, S.; Juckel, G.; Bauer, M.; Pfennig, A. Risk Constellations Prior to the Development of Bipolar Disorders: Rationale of a New Risk Assessment Tool. *Journal of affective disorders* **2012**, *136*, 1000–1010.

12. Correll, C.U.; Olvet, D.M.; Auther, A.M.; Hauser, M.; Kishimoto, T.; Carrión, R.E.; Snyder, S.; Cornblatt, B.A. The Bipolar Prodrome Symptom Interview and Scale–Prospective (BPSS-P): Description and Validation in a Psychiatric Sample and Healthy Controls. *Bipolar disorders* **2014**, *16*, 505–522.

13. Dwyer, D.B.; Falkai, P.; Koutsouleris, N. Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annu Rev Clin Psychol* **2018**, *14*, 91–118, doi:10.1146/annurev-clinpsy-032816-045037.

14. Koutsouleris, N.; Dwyer, D.B.; Degenhardt, F.; Maj, C.; Urquijo-Castro, M.F.; Sanfelici, R.; Popovic, D.; Oeztuerk, O.; Haas, S.S.; Weiske, J.; et al. Multimodal Machine Learning Workflows for Prediction of Psychosis in Patients With Clinical High-Risk Syndromes and Recent-Onset Depression. *JAMA Psychiatry* **2021**, *78*, 195–209, doi:10.1001/jamapsychiatry.2020.3604.

15. Arnone, D.; Cavanagh, J.; Gerber, D.; Lawrie, S.M.; Ebmeier, K.P.; McIntosh, A.M. Magnetic Resonance Imaging Studies in Bipolar Disorder and Schizophrenia: Meta-Analysis. *The British Journal of Psychiatry* **2009**, *195*, 194–201.

16. Hibar, D.P.; Westlye, L.T.; Doan, N.T.; Jahanshad, N.; Cheung, J.W.; Ching, C.R.K.; Versace, A.; Bilderbeck, A.C.; Uhlmann, A.; Mwangi, B.; et al. Cortical Abnormalities in Bipolar Disorder: An MRI Analysis of 6503 Individuals from the ENIGMA Bipolar Disorder Working Group. *Mol Psychiatry* **2018**, *23*, 932–942, doi:10.1038/mp.2017.73.

17. Hibar, D.P.; Westlye, L.T.; van Erp, T.G.M.; Rasmussen, J.; Leonardo, C.D.; Faskowitz, J.; Haukvik, U.K.; Hartberg, C.B.; Doan, N.T.; Agartz, I.; et al. Subcortical Volumetric Abnormalities in Bipolar Disorder. *Mol Psychiatry* **2016**, *21*, 1710–1716, doi:10.1038/mp.2015.227.

18. Haukvik, U.K.; Gurholt, T.P.; Nerland, S.; Elvsåshagen, T.; Akudjedu, T.N.; Alda, M.; Alnæs, D.; Alonso-Lana, S.; Bauer, J.; Baune, B.T. In Vivo Hippocampal Subfield Volumes in Bipolar Disorder—A Mega-analysis from The Enhancing Neuro Imaging Genetics through Meta-Analysis Bipolar Disorder Working Group. *Human brain mapping* **2022**, *43*, 385–398.

19. Haukvik, U.K.; Westlye, L.T.; Mørch-Johnsen, L.; Jørgensen, K.N.; Lange, E.H.; Dale, A.M.; Melle, I.; Andreassen, O.A.; Agartz, I. In Vivo Hippocampal Subfield Volumes in Schizophrenia and Bipolar Disorder. *Biological psychiatry* **2015**, *77*, 581–588.

20. Mathew, I.; Gardin, T.M.; Tandon, N.; Eack, S.; Francis, A.N.; Seidman, L.J.; Clementz, B.; Pearlson, G.D.; Sweeney, J.A.; Tamminga, C.A. Medial Temporal Lobe Structures and Hippocampal Subfields in Psychotic Disorders: Findings from the Bipolar-Schizophrenia Network on Intermediate Phenotypes (B-SNIP) Study. *JAMA psychiatry* **2014**, *71*, 769–777.

21. Heller, A.S. Cortical-Subcortical Interactions in Depression: From Animal Models to Human Psychopathology. *Frontiers in Systems Neuroscience* **2016**, *10*.

22. Nikolenko, V.N.; Oganesyan, M.V.; Rizaeva, N.A.; Kudryashova, V.A.; Nikitina, A.T.; Pavliv, M.P.; Shchedrina, M.A.; Giller, D.B.; Bulygin, K.V.; Sinelnikov, M.Y. Amygdala: Neuroanatomical and Morphophysiological Features in Terms of Neurological and Neurodegenerative Diseases. *Brain Sciences* **2020**, *10*, 502, doi:10.3390/brainsci10080502.

23. Barth, C.; Nerland, S.; de Lange, A.-M.G.; Wortinger, L.A.; Hilland, E.; Andreassen, O.A.; Jørgensen, K.N.; Agartz, I. In Vivo Amygdala Nuclei Volumes in Schizophrenia and Bipolar Disorders. *Schizophr Bull* **2021**, *47*, 1431–1441, doi:10.1093/schbul/sbaa192.

24. Bielau, H.; Trübner, K.; Krell, D.; Agelink, M.W.; Bernstein, H.-G.; Stauch, R.; Mawrin, C.; Danos, P.; Gerhard, L.; Bogerts, B. Volume Deficits of Subcortical Nuclei in Mood Disorders. *European archives of psychiatry and clinical neuroscience* **2005**, *255*, 401–412.

25. Luders, E.; Thompson, P.M.; Kurth, F.; Hong, J.-Y.; Phillips, O.R.; Wang, Y.; Gutman, B.A.; Chou, Y.-Y.; Narr, K.L.; Toga, A.W. Global and Regional Alterations of Hippocampal Anatomy in Long-term Meditation Practitioners. *Human brain mapping* **2013**, *34*, 3369–3375.

26. Sani, G.; Simonetti, A.; Janiri, D.; Banaj, N.; Ambrosi, E.; De Rossi, P.; Ciullo, V.; Arciniegas, D.B.; Piras, F.; Spalletta, G. Association between Duration of Lithium Exposure and Hippocampus/Amygdala Volumes in Type I Bipolar Disorder. *Journal of Affective Disorders* **2018**, *232*, 341–348.

27. Roeder, S.S.; Burkardt, P.; Rost, F.; Rode, J.; Brusch, L.; Coras, R.; Englund, E.; Håkansson, K.; Possnert, G.; Salehpour, M.; et al. Evidence for Postnatal Neurogenesis in the Human Amygdala. *Commun Biol* **2022**, *5*, 1–8, doi:10.1038/s42003-022-03299-8.

28. Orru, G.; Pettersson-Yeo, W.; Marquand, A.F.; Sartori, G.; Mechelli, A. Using Support Vector Machine to Identify Imaging Biomarkers of Neurological and Psychiatric Disease: A Critical Review. *Neuroscience & Biobehavioral Reviews* **2012**, *36*, 1140–1152.

29. Mikolas, P.; Marxen, M.; Riedel, P.; Bröckel, K.; Martini, J.; Huth, F.; Berndt, C.; Vogelbacher, C.; Jansesn, A.; Kircher, T.; et al. *Prediction of Estimated Risk for Bipolar Disorder Using Machine Learning and Structural MRI Features*; In Review, 2022;

30. Garg, A.; Mago, V. Role of Machine Learning in Medical Research: A Survey. *Computer Science Review* **2021**, *40*, 100370, doi:10.1016/j.cosrev.2021.100370.

31. Pfennig, A.; Leopold, K.; Martini, J.; Boehme, A.; Lambert, M.; Stamm, T.; Bermpohl, F.; Reif, A.; Kittel-Schneider, S.; Juckel, G. Improving Early Recognition and Intervention in People at Increased Risk for the Development of Bipolar Disorder: Study Protocol of a Prospective-Longitudinal, Naturalistic Cohort Study (Early-BipoLife). *International Journal of Bipolar Disorders* **2020**, *8*, 1–14.

32. Ritter, P.S.; Bermpohl, F.; Gruber, O.; Hautzinger, M.; Jansen, A.; Juckel, G.; Kircher, T.; Lambert, M.; Mulert, C.; Pfennig, A. Aims and Structure of the German Research Consortium BipoLife for the Study of Bipolar Disorder. *International journal of bipolar disorders* **2016**, *4*, 1–9.

33. Kessler, R.C.; Berglund, P.; Demler, O.; Jin, R.; Merikangas, K.R.; Walters, E.E. Lifetime Prevalence and Age-of-Onset Distributions of DSM-IV Disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry* **2005**, *62*, 593, doi:10.1001/archpsyc.62.6.593.

34. Pfennig, A.; Jabs, B.; Pfeiffer, S.; Weikert, B.; Leopold, K.; Bauer, M. Health care service experiences of bipolar patients in Germany survey prior to the introduction of the S3 Guideline for diagnostics and treatment of bipolar disorders. *Nervenheilkunde* **2011**, *30*, 333–340, doi:10.1055/s-0038-1627819.

35. Lambert, M.; Bock, T.; Naber, D.; Löwe, B.; Schulte-Markwort, M.; Schäfer, I.; Gumz, A.; Degkwitz, P.; Schulte, B.; König, H.; et al. Die psychische Gesundheit von Kindern, Jugendlichen und jungen Erwachsenen – Teil 1: Häufigkeit, Störungspersistenz, Belastungsfaktoren, Service-Inanspruchnahme und Behandlungsverzögerung mit Konsequenzen. *Fortschr Neurol Psychiatr* **2013**, *81*, 614–627, doi:10.1055/s-0033-1355843.

36. Mikolas, P.; Bröckel, K.; Vogelbacher, C.; Müller, D.K.; Marxen, M.; Berndt, C.; Sauer, C.; Jung, S.; Fröhner, J.H.; Fallgatter, A.J.; et al. Individuals at Increased Risk for Development of Bipolar Disorder Display Structural Alterations Similar to People with Manifest Disease. *Transl Psychiatry* **2021**, *11*, 485, doi:10.1038/s41398-021-01598-y.

37. Vogelbacher, C.; Sommer, J.; Schuster, V.; Bopp, M.H.; Falkenberg, I.; Ritter, P.S.; Bermpohl, F.; Hindi Attar, C.; Rauer, L.; Einenkel, K.E. The German Research Consortium for the Study of Bipolar Disorder (BipoLife): A Magnetic Resonance Imaging Study Protocol. *International journal of bipolar disorders* **2021**, *9*, 1–15.

38. Fischl, B.; Salat, D.H.; Kouwe, A.J.W. van der; Makris, N.; Ségonne, F.; Quinn, B.T.; Dale, A.M. Sequence-Independent Segmentation of Magnetic Resonance Images. *NeuroImage* **2004**, *23*, S69–S84, doi:DOI: 10.1016/j.neuroimage.2004.07.016.

39. Fischl, B.; Sereno, M.I.; Tootell, R.B.H.; Dale, A.M. High-Resolution Intersubject Averaging and a Coordinate System for the Cortical Surface. *Human Brain Mapping* **1999**, *8*, 272–284, doi:10.1002/(SICI)1097-0193(1999)8:4<272::AID-HBM10>3.0.CO;2-4.

40. Iglesias, J.E.; Augustinack, J.C.; Nguyen, K.; Player, C.M.; Player, A.; Wright, M.; Roy, N.; Frosch, M.P.; McKee, A.C.; Wald, L.L. A Computational Atlas of the Hippocampal Formation Using Ex Vivo, Ultra-High Resolution MRI: Application to Adaptive Segmentation of in Vivo MRI. *Neuroimage* **2015**, *115*, 117–137.

41. Saygin, Z.M.; Kliemann, D.; Iglesias, J.E.; van der Kouwe, A.J.; Boyd, E.; Reuter, M.; Stevens, A.; Van Leemput, K.; McKee, A.; Frosch, M.P. High-Resolution Magnetic Resonance Imaging Reveals Nuclei of the Human Amygdala: Manual Segmentation to Automatic Atlas. *Neuroimage* **2017**, *155*, 370–382.

42. Iglesias, J.E.; Van Leemput, K.; Augustinack, J.; Insausti, R.; Fischl, B.; Reuter, M.; Initiative, A.D.N. Bayesian Longitudinal Segmentation of Hippocampal Substructures in Brain MRI Using Subject-Specific Atlases. *Neuroimage* **2016**, *141*, 542–555.

43. Van Leemput, K.; Bakkour, A.; Benner, T.; Wiggins, G.; Wald, L.L.; Augustinack, J.; Dickerson, B.C.; Golland, P.; Fischl, B. Automated Segmentation of Hippocampal Subfields from Ultra-High Resolution in Vivo MRI. *Hippocampus* **2009**, *19*, 549–557, doi:10.1002/hipo.20615.

44. Sämann, P.G.; Iglesias, J.E.; Gutman, B.; Grotegerd, D.; Leenings, R.; Flint, C.; Dannlowski, U.; Clarke-Rubright, E.K.; Morey, R.A.; Erp, T.G.M.; et al. FREESURFER -based Segmentation of Hippocampal Subfields: A Review of Methods and Applications, with a Novel Quality Control Procedure for ENIGMA Studies and Other Collaborative Efforts. *Human Brain Mapping* **2022**, *43*, 207–233, doi:10.1002/hbm.25326.

45. Tesli, N.; van der Meer, D.; Rokicki, J.; Storvestre, G.; Røsæg, C.; Jensen, A.; Hjell, G.; Bell, C.; Fischer-Vieler, T.; Tesli, M.; et al. Hippocampal Subfield and Amygdala Nuclei Volumes in Schizophrenia Patients with a History of Violence. *Eur Arch Psychiatry Clin Neurosci* **2020**, *270*, 771–782, doi:10.1007/s00406-020-01098-y.

46. Yücel, M.; Lorenzetti, V.; Suo, C.; Zalesky, A.; Fornito, A.; Takagi, M.J.; Lubman, D.I.; Solowij, N. Hippocampal Harms, Protection and Recovery Following Regular Cannabis Use. *Transl Psychiatry* **2016**, *6*, e710, doi:10.1038/tp.2015.201.

47. Tozzi, L.; Garczarek, L.; Janowitz, D.; Stein, D.J.; Wittfeld, K.; Dobrowolny, H.; Lagopoulos, J.; Hatton, S.N.; Hickie, I.B.; Carballedo, A.; et al. Interactive Impact of Childhood Maltreatment, Depression, and Age on Cortical Brain Structure: Mega-Analytic Findings from a Large Multi-Site Cohort. *Psychol Med* **2020**, *50*, 1020–1031, doi:10.1017/S003329171900093X.

48. Mikolas, P.; Tozzi, L.; Doolin, K.; Farrell, C.; O'Keane, V.; Frodl, T. Effects of Early Life Adversity and FKBP5 Genotype on Hippocampal Subfields Volume in Major Depression. *Journal of Affective Disorders* **2019**, *252*, 152–159, doi:10.1016/j.jad.2019.04.054.

49. Twait, E.L.; Blom, K.; Koek, H.L.; Zwartbol, M.H.T.; Ghaznawi, R.; Hendrikse, J.; Gerritsen, L.; Geerlings, M.I.; UCC SMART Study Group Psychosocial Factors and Hippocampal Subfields: The Medea-7T Study. *Hum Brain Mapp* **2022**, doi:10.1002/hbm.26185.

50. Tozzi, L.; Farrell, C.; Booij, L.; Doolin, K.; Nemoda, Z.; Szyf, M.; Pomares, F.B.; Chiarella, J.; O'Keane, V.; Frodl, T. Epigenetic Changes of FKBP5 as a Link Connecting Genetic and Environmental Risk Factors with Structural and Functional Brain Changes in Major Depression. *Neuropsychopharmacology* **2018**, *43*, 1138–1145, doi:10.1038/npp.2017.290.

51. Klinitzke, G.; Romppel, M.; Häuser, W.; Brähler, E.; Glaesmer, H. The German Version of the Childhood Trauma Questionnaire (CTQ): Psychometric Characteristics in a Representative Sample of the General Population. *Psychotherapie, Psychosomatik, Medizinische Psychologie* **2011**, *62*, 47–51.

52. Esteban, O.; Birman, D.; Schaer, M.; Koyejo, O.O.; Poldrack, R.A.; Gorgolewski, K.J. MRIQC: Advancing the Automatic Prediction of Image Quality in MRI from Unseen Sites. *PloS one* **2017**, *12*, e0184661.

53. Fjell, A.M.; Westlye, L.T.; Grydeland, H.; Amlien, I.; Espeseth, T.; Reinvang, I.; Raz, N.; Holland, D.; Dale, A.M.; Walhovd, K.B. Critical Ages in the Life Course of the Adult Brain: Nonlinear Subcortical Aging. *Neurobiology of Aging* **2013**, *34*, 2239–2247, doi:10.1016/j.neurobiolaging.2013.04.006.

54. van Eijk, L.; Hansell, N.K.; Strike, L.T.; Couvy-Duchesne, B.; de Zubicaray, G.I.; Thompson, P.M.; McMahon, K.L.; Zietsch, B.P.; Wright, M.J. Region-Specific Sex Differences in the Hippocampus. *NeuroImage* **2020**, *215*, 116781, doi:10.1016/j.neuroimage.2020.116781.

55. Benjamini, Y.; Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal statistical society: series B (Methodological)* **1995**, *57*, 289–300.

56. Nunes; Nunes, A.; Schnack, H.G.; Ching, C.R.K.; Agartz, I.; Akudjedu, T.N.; Alda, M.; Alnæs, D.; Alonso-Lana, S.; Bauer, J.; et al. Using Structural MRI to Identify Bipolar Disorders – 13 Site Machine Learning Study in 3020 Individuals from the ENIGMA Bipolar Disorders Working Group. *Mol Psychiatry* **2020**, *25*, 2130–2143, doi:10.1038/s41380-018-0228-9.

57. Ching, C.R.K.; Hibar, D.P.; Gurholt, T.P.; Nunes, A.; Thomopoulos, S.I.; Abé, C.; Agartz, I.; Brouwer, R.M.; Cannon, D.M.; Zwarte, S.M.C.; et al. What We Learn about Bipolar Disorder from Large-scale Neuroimaging: Findings and Future Directions from the ENIGMA Bipolar Disorder Working Group. *Hum Brain Mapp* **2020**, hbm.25098, doi:10.1002/hbm.25098.

58. Lemm, S.; Blankertz, B.; Dickhaus, T.; Müller, K.-R. Introduction to Machine Learning for Brain Imaging. *NeuroImage* **2011**, *56*, 387–399, doi:10.1016/j.neuroimage.2010.11.004.

59. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-Sampling Technique. *jair* **2002**, *16*, 321–357, doi:10.1613/jair.953.

60. Blumberg, H.P.; Kaufman, J.; Martin, A.; Whiteman, R.; Zhang, J.H.; Gore, J.C.; Charney, D.S.; Krystal, J.H.; Peterson, B.S. Amygdala and Hippocampal Volumes in Adolescents and Adults with Bipolar Disorder. *Archives of general psychiatry* **2003**, *60*, 1201–1208.

61. Cattarinussi, G.; Di Giorgio, A.; Wolf, R.C.; Balestrieri, M.; Sambataro, F. Neural Signatures of the Risk for Bipolar Disorder: A Meta-analysis of Structural and Functional Neuroimaging Studies. *Bipolar Disord* **2019**, *21*, 215–227, doi:10.1111/bdi.12720.

62. Pereira, F.; Mitchell, T.; Botvinick, M. Machine Learning Classifiers and FMRI: A Tutorial Overview. *Neuroimage* **2009**, *45*, S199-209, doi:10.1016/j.neuroimage.2008.11.007.

63. Claude, L.; Houenou, J.; Duchesnay, E.; Favre, P. Will Machine Learning Applied to Neuroimaging in Bipolar Disorder Help the Clinician? A Critical Review and Methodological Suggestions. *Bipolar Disord* **2020**, *22*, 334–355, doi:10.1111/bdi.12895.

64. Nieuwenhuis, M.; van Haren, N.E.M.; Hulshoff Pol, H.E.; Cahn, W.; Kahn, R.S.; Schnack, H.G. Classification of Schizophrenia Patients and Healthy Controls from Structural MRI Scans in Two Large Independent Samples. *NeuroImage* **2012**, *61*, 606–612, doi:10.1016/j.neuroimage.2012.03.079.

65. Lupien, S.J.; Juster, R.-P.; Raymond, C.; Marin, M.-F. The Effects of Chronic Stress on the Human Brain: From Neurotoxicity, to Vulnerability, to Opportunity. *Front Neuroendocrinol* **2018**, *49*, 91–105, doi:10.1016/j.yfrne.2018.02.001.

66. Logtenberg, E.; Overbeek, M.F.; Pasman, J.A.; Abdellaoui, A.; Luijten, M.; van Holst, R.J.; Vink, J.M.; Denys, D.; Medland, S.E.; Verweij, K.J.H.; et al. Investigating the Causal Nature of the Relationship of Subcortical Brain Volume with Smoking and Alcohol Use. *Br J Psychiatry* **2022**, *221*, 377–385, doi:10.1192/bjp.2021.81.

67. Videbech, P. Hippocampal Volume and Depression: A Meta-Analysis of MRI Studies. *American Journal of Psychiatry* **2004**, *161*, 1957–1966, doi:10.1176/appi.ajp.161.11.1957.

68. Hayano, F.; Nakamura, M.; Asami, T.; Uehara, K.; Yoshida, T.; Roppongi, T.; Otsuka, T.; Inoue, T.; Hirayasu, Y. Smaller Amygdala Is Associated with Anxiety in Patients with Panic Disorder. *Psychiatry and clinical neurosciences* **2009**, *63*, 266–276.

69. Sala, M.; Perez, J.; Soloff, P.; Di Nemi, S.U.; Caverzasi, E.; Soares, J.C.; Brambilla, P. Stress and Hippocampal Abnormalities in Psychiatric Disorders. *European Neuropsychopharmacology* **2004**, *14*, 393–405.l