**Prediction of estimated risk for bipolar disorder using machine learning and structural MRI features**

**Supplemental materials**

**Supplementary note 1:** Inclusion and exclusion criteria for the three recruitment pathways.

**Supplementary note 2:** The list of 20 selected features for secondary analysis according to Hibar et al.

**Supplementary note 3:** Analysis of the diagnoses ADHD and depression on the classification.

**Supplementary note 4:** Secondary analysis of EPI*bipolar* risk

**Supplementary table 1:** Overview of the risk assessment tools and corresponding sample sizes

**Supplementary table 2.** Breakdown of sample sizes and participants fulfilling and not fulfilling the risk criterion per study site.

**Supplementary table 3.** Ranking of features according to their SVM coefficients.

**Supplementary note 1:** Inclusion and exclusion criteria for the three recruitment pathways.

Inclusion criteria

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:

• 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

Our inclusion criteria regarding age (15-35 years), which were based on available studies on age of onset and time to diagnosis. A study by Kessler et al. (2005) reported that about 75% of individuals with bipolar I disorder would develop the disorder until the age of 42 years. Other studies report, that approximately 70 % of all individuals would develop bipolar disorder by the age of 21 (Merikangas et al., 2011; Lambert et al., 2013). However, the time-to-diagnosis for bipolar disorder is typically long. Within current psychiatric services in Germany, it takes in average 12.4 years from the appearance of first symptoms to establish the diagnosis (Pfennig et al., 2011). Due to the predominance of depressive symptoms, as well as difficulties to recognize hypomania, the most typical false diagnosis is unipolar depression (Merikangas et al., 2011; Lambert et al., 2013; Correll et al., 2014). For these reasons, we can assume that some older individuals in our sample might have an unrecognized bipolar disorder and therefore decided to extend the age range up to 35 years.

**Supplementary note 2:** The list of 20 selected features for secondary analysis according to Hibar et al. (2018).

lh\_parsopercularis\_thickness

lh\_fusiform\_thickness

lh\_rostralmiddlefrontal\_thickness

lh\_parstriangularis\_thickness

rh\_fusiform\_thicknes

lh\_caudalmiddlefrontal\_thickness

lh\_inferiorparietal\_thickness

rh\_rostralmiddlefrontal\_thickness

rh\_inferiorparietal\_thickness

rh\_superiorfrontal\_thickness

lh\_supramarginal\_thickness

lh\_middletemporal\_thickness

lh\_inferiortemporal\_thickness

rh\_parsopercularis\_thickness

lh\_parsorbitalis\_thickness

rh\_parsorbitalis\_thickness

lh\_superiorfrontal\_thickness

rh\_parstriangularis\_thickness

rh\_medialorbitofrontal\_thickness

rh\_middletemporal\_thickness

**Supplementary note 3: Analysis of the diagnoses ADHD and depression on the classification.**

In the Results – primary analysis section, we report post-hoc tests of the correctly and incorrectly classified subjects, in order to evaluate the potential confounders on the classification. Here, we compared the proportion of participants from the three recruitment pathways (early recognition / depression / ADHD) between the correctly and incorrectly classified subjects. However, for ADHD, there were 19 participants both in the early recognition and depression recruitment pathways. Taking only the ADHD diagnosis irrespective of recruitment pathway into account, there was no difference between the correctly and incorrectly classified subjects (df = 1, Chi2 = 0.016,*p* = 0.898). For the lifetime or present major depression, we did not perform post-hoc tests, as depression and specific depressive symptoms belong to criteria of the risk syndromes according to BPSS-P (similarly to BARS as well as EPI*bipolar*).

Additionally, we performed the classification after discarding participants recruited via ADHD and depression pathways with following results: without ADHD (N = 202) Cohen's kappa 0.054 (95% CI -0.139-0.247), balanced accuracy 53.3 % (95% CI 42.5-64.1), without depression (N = 152) Cohen's kappa 0.071 (95% CI -0.089-0.231), balanced accuracy 55.4 % (95% CI 45.3-65.4).

**Supplementary note 4:** Secondary analysis of EPI*bipolar* risk

We analyzed EPIbipolar in more detail in two exploratory analyses. First, we performed a classification using all three groups – no-risk, low-risk and high-risk (N = 32 / 104 / 137 respectively) achieving a balanced accuracy of 34.7 % (95% CI 31.1- 38.3) (i.e. chance level for a three category outcome) in the 10-fold crossvalidation. Second, we removed the low-risk group from the sample, retaining the no-risk and high-risk groups exclusively (N = 136), and used the left pars opercularis thickness as a single feature, achieving a balanced accuracy of 60.9 % (95% CI 48.4- 73.4) in the 10-fold crossvalidation and 55.5 % (95% CI 42.4- 68.6) in the leave-one-site-out crossvalidation (i.e. chance level for a binary outcome considering the CI). We also performed the classification while grouping the 'no risk' and 'low risk' groups vs 'high risk' group with following results: balanced accuracy 48.3 % (CI 40.9-55.6%), Cohen's kappa -0.032 (CI -0.171-0.107).

**Supplementary table 1:** Overview of the risk assessment tools and corresponding sample sizes (adapted according to Bröckel et al. (*In submission*) and Mikolas et al. (2021)).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Instrument (N)** | **Risk states** | **N (% Sample)** | **Validation** | **Note** |
| BPSS-P (276) | Attenuated mania symptom syndrome (AMSS) | 54 (19.6) | Good internal consistency, convergent validity and inter-rater reliability (Correll *et al.*, 2014) | Semi-structured interview based on the DSM-5 criteria for bipolar disorder and major depressive disorder (Correll *et al.*, 2014) |
| Genetic mania risk and deterioration syndrome (GMRDS) | 2 (0.7) |
| BARS (264) | Sub-threshold mania, assessed by BPSS-P | 26 (9.8) | BARS criteria had an adequate prognostic accuracy (Harrell's C = 0.742) and clinical utility (Fusar-Poli *et al.*, 2018) | Extension of the BAR criteria (2 additional symptom domains) (Fusar-Poli *et al.*, 2018) |
| Sub-threshold depression, assessed by BPSS-FP or SCID *and* cyclothymic features | 149 (56.4) |
| Sub-threshold depression plus genetic risk | 13 (4.9) |
| Mixed symptoms, assessed by BPSS-P | 3 (1.1) |
| Mood swings, assessed by EPI*bipolar* | 118 (44.7) |
| EPI*bipolar* (273) | No-risk | 32 (11.6) | No longitudinal (ongoing study)  Includes and integrates items from validated tools (BPSS-P, BAR) as well as genetic risk | Semi-structured interview  Integrates risk factors based on a systematic review of literature (Leopold *et al.*, 2012) |
| Low-risk | 137 (49.6) |
| High-risk | 104 (37.7) |

**Supplementary table 2.** Breakdown of sample sizes and participants fulfilling and not fulfilling the risk criterion per study site.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Site | | | | | | | Total |
| Berlin | Bochum | Dresden | Frankfurt | Hamburg | Marburg | Tübingen |
| BPSS-P criterion | not fullfilled | 57 | 7 | 31 | 21 | 21 | 67 | 16 | 220 |
| fullfilled | 6 | 1 | 5 | 20 | 13 | 7 | 4 | 56 |
| Total | | 63 | 8 | 36 | 41 | 34 | 74 | 20 | 276 |

**Supplementary table 3.** Ranking of features according to their SVM coefficients.

|  |  |
| --- | --- |
| Feature | Mean SVM coefficient |
| lh\_precuneus\_thickness | 1.783976879 |
| rh\_parstriangularis\_thickness | 1.608533374 |
| rh\_fusiform\_thickness | 0.928239047 |
| lh\_transversetemporal\_thickness | 0.800646946 |
| rh\_superiorfrontal\_thickness | 0.789743825 |
| rh\_cuneus\_thickness | 0.713079441 |
| lh\_inferiorparietal\_thickness | 0.658000082 |
| lh\_parstriangularis\_thickness | 0.642315548 |
| lh\_caudalanteriorcingulate\_thickness | 0.631354016 |
| rh\_caudalanteriorcingulate\_thickness | 0.492421592 |
| rh\_parahippocampal\_thickness | 0.471491982 |
| rh\_insula\_thickness | 0.431770073 |
| rh\_frontalpole\_thickness | 0.420177869 |
| rh\_lateraloccipital\_thickness | 0.405688992 |
| lh\_frontalpole\_thickness | 0.395092914 |
| rh\_medialorbitofrontal\_thickness | 0.391455904 |
| rh\_rostralanteriorcingulate\_thickness | 0.382318411 |
| lh\_supramarginal\_thickness | 0.359961695 |
| rh\_supramarginal\_thickness | 3.40E-01 |
| lh\_inferiortemporal\_thickness | 0.292938256 |
| lh\_posteriorcingulate\_thickness | 0.267600265 |
| lh\_superiorparietal\_thickness | 0.240793864 |
| lh\_parahippocampal\_thickness | 0.232690747 |
| rh\_postcentral\_thickness | 0.218904162 |
| rh\_pericalcarine\_thickness | 0.167581752 |
| rh\_parsopercularis\_thickness | 0.156305279 |
| rh\_precentral\_thickness | 0.124376678 |
| lh\_temporalpole\_thickness | 0.118953415 |
| rh\_lateralorbitofrontal\_thickness | 0.03930299 |
| lh\_entorhinal\_thickness | 0.027801043 |
| rh\_superiortemporal\_thickness | 0.012469632 |
| lh\_bankssts\_thickness | 0.009063421 |
| rh\_bankssts\_thickness | 0.005956208 |
| rh\_transversetemporal\_thickness | -0.00309838 |
| lh\_superiorfrontal\_thickness | -0.03653334 |
| lh\_insula\_thickness | -0,04220697 |
| rh\_inferiortemporal\_thickness | -0.05566412 |
| lh\_superiortemporal\_thickness | -0.05851827 |
| rh\_caudalmiddlefrontal\_thickness | -0.07425086 |
| rh\_inferiorparietal\_thickness | -0.08364897 |
| lh\_pericalcarine\_thickness | -0.08678666 |
| lh\_precentral\_thickness | -0.09148176 |
| lh\_parsorbitalis\_thickness | -0.14161886 |
| lh\_middletemporal\_thickness | -0.226009 |
| lh\_isthmuscingulate\_thickness | -0.22893125 |
| lh\_cuneus\_thickness | -0.26492932 |
| rh\_temporalpole\_thickness | -0.27046636 |
| lh\_lateralorbitofrontal\_thickness | -0.28265985 |
| lh\_lateraloccipital\_thickness | -0.28658943 |
| lh\_medialorbitofrontal\_thickness | -0.29211138 |
| lh\_rostralanteriorcingulate\_thickness | -0.3898791 |
| rh\_entorhinal\_thickness | -0.39351071 |
| lh\_paracentral\_thickness | -0.47477732 |
| lh\_caudalmiddlefrontal\_thickness | -0.48499482 |
| lh\_fusiform\_thickness | -0.51077598 |
| rh\_isthmuscingulate\_thickness | -0.51830264 |
| rh\_parsorbitalis\_thickness | -0.55182755 |
| lh\_rostralmiddlefrontal\_thickness | -0.57001052 |
| rh\_precuneus\_thickness | -0.59037961 |
| rh\_lingual\_thickness | -0.59452749 |
| rh\_middletemporal\_thickness | -0.59767057 |
| lh\_lingual\_thickness | -0.60498745 |
| rh\_superiorparietal\_thickness | -0.66991654 |
| rh\_rostralmiddlefrontal\_thickness | -0.7187164 |
| lh\_parsopercularis\_thickness | -0.84061014 |
| rh\_posteriorcingulate\_thickness | -1.08693179 |
| lh\_postcentral\_thickness | -1.1165027 |
| rh\_paracentral\_thickness | -1.14023448 |