Journal of Affective Disorders 227 (2018) 498–505

Contents lists available at ScienceDirect

Journal of Affective Disorders

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

Research paper

Pattern recognition of magnetic resonance imaging-based gray matter volume measurements classifies bipolar disorder and major depressive disorder

Harry Rubin-Falcone[a,b,*], Francesca Zanderigo[a,b], Binod Thapa-Chhetry[b], Martin Lan[a,b], Jeffrey M. Miller[a,b], M. Elizabeth Sublette[a,b], Maria A. Oquendo[c], David J. Hellerstein[a,d], Patrick J. McGrath[a,d], Johnathan W. Stewart[a,d], J. John Mann[a,b,e]

[a] Department of Psychiatry, Columbia University, New York, NY, USA
[b] Molecular Imaging and Neuropathology Division, New York State Psychiatric Institute, New York, NY, USA
[c] Now at Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, USA
[d] Depression Evaluation Service, Division of Clinical Therapeutics, New York State Psychiatric Institute, New York, NY, USA
[e] Department of Radiology, Columbia University, New York, NY, USA

ARTICLE INFO

Keywords:
MDD
Depression
Bipolar
Gray-matter
MRI
SVM

ABSTRACT

Background: Bipolar Disorder (BD) cannot be reliably distinguished from Major Depressive Disorder (MDD) until the first manic or hypomanic episode. Consequently, many patients with BD are treated with antidepressants without mood stabilizers, a strategy that is often ineffective and carries a risk of inducing a manic episode. We previously reported reduced cortical thickness in right precuneus, right caudal middle-frontal cortex and left inferior parietal cortex in BD compared with MDD.

Methods: This study extends our previous work by performing individual level classification of BD or MDD in an expanded, currently unmedicated, cohort using gray matter volume (GMV) based on Magnetic Resonance Imaging and a Support Vector Machine. All patients were in a Major Depressive Episode and a leave-two-out analysis was performed.

Results: Nineteen out of 26 BD subjects and 20 out of 26 MDD subjects were correctly identified, for a combined accuracy of 75%. The three brain regions contributing to the classification were higher GMV in bilateral supramarginal gyrus and occipital cortex indicating MDD, and higher GMV in right dorsolateral prefrontal cortex indicating BD.

Limitations: This analysis included scans performed with two different headcoils and scan sequences, which limited the interpretability of results in an independent cohort analysis.

Conclusions: Our results add to previously published data which suggest that regional gray matter volume should be investigated further as a clinical diagnostic tool to predict BD before the appearance of a manic or hypomanic episode.

1. Introduction

Bipolar disorder (BD) is a leading cause of disability worldwide (Mathers et al., 2008) and is associated with significant risk of suicide and other medical morbidity and mortality (Baldessarini et al., 2010; Schneider et al., 2001). BD is often misdiagnosed as Major Depressive Disorder (MDD) for several reasons: the presentation of major depressive episodes (MDEs) in these conditions is very similar; hypomanic or manic symptoms may occur years after initial depressive symptoms (Culpepper, 2014); and patients with BD may under-report severity of manic/hypomanic symptoms. One study found that the mean time interval between first mood episode and a correct diagnosis of BD is greater than 10 years (Lish et al., 1994). Failure to diagnose BD can lead to prescription of antidepressants without mood stabilizers, which may trigger manic symptoms, increasing the risk of social and occupational impairment (Nasrallah, 2015). Some data suggest that usual antidepressants also may not be as effective at treating the bipolar depression (Sachs et al., 2007). It would therefore be clinically valuable to

* Correspondence to: Division of Molecular Imaging and Neuropathology Columbia Psychiatry and New York State Psychiatric Institute, 1051, Riverside Drive #42, New York, NY 10032, USA.
E-mail address: falcone@nyspi.columbia.edu (H. Rubin-Falcone).

https://doi.org/10.1016/j.jad.2017.11.043
Received 6 September 2017; Received in revised form 6 November 2017; Accepted 11 November 2017
Available online 13 November 2017
0165-0327/

identify biomarkers that improve diagnostic discrimination of BD from MDD prior to the first manic episode.

Noninvasive structural neuroimaging using Magnetic Resonance Imaging (MRI) is routinely available in clinical settings, and is less expensive, more reliable, and less burdensome to patients compared to other imaging modalities, such as Positron Emission Tomography or functional MRI (fMRI). Many studies have used structural MRI (sMRI) to identify morphological abnormalities in BD (including type 1, type 2 and not otherwise specified (NOS) patients) (Hanford et al., 2016b) and MDD (Zhang et al., 2016). Larger gray matter volume (GMV) was found in MDD compared to healthy volunteers (HV) in thalamus, cuneus, and superior frontal gyrus, while smaller GMV was found in insula and middle frontal gyrus (Peng et al., 2016b). Studies in BD patients have found lower cortical thickness values in BD compared to HV in frontal and limbic areas (Eker et al., 2014; Oertel-Knochel et al., 2015). MRI scans obtained before the first episode in individuals at high risk (HR) of developing mood disorders reveal thinner cortex in HR participants who went on to develop MDD compared to those who did not (Papmeyer et al., 2015). Children of BD parents have been found to have reduced cortical thickness compared to healthy controls, even when asymptomatic (Hanford et al., 2016c).

Potential structural brain differences between BD and MDD populations also have been investigated. We found less cortical thickness in right precuneus, right caudal middle-frontal cortex and left inferior parietal cortex in BD compared with MDD (Lan et al., 2014). Others found smaller GMV in BD compared with MDD in right anterior frontal gyrus and middle cingulate gyri (Cai et al., 2015). A recent meta-analysis found that MDDs have smaller GMV than BD when compared to HV in dorsolateral prefrontal cortex and hippocampus, although MDD and BD patients similarly exhibited lower GMV values compared to HV in medial prefrontal cortex, cingulate, and insula (Wise et al., 2016). Other analyses have found comparably lower/thinner cortical thickness in both groups compared to HV in temporal lobe (Niu et al., 2017). MDD, BD, and schizophrenia patients were found to share lower GMV values compared to HV in 88% of regions where the groups differed in a 4-group ANCOVA, although BD and schizophrenia patients showed white matter integrity alterations not seen in MDD (Miao Chang, 2017). A limitation of past studies comparing MDD and BD has been the inclusion of medicated patients, and of subjects in different mood states. Lithium is reported to increase gray matter density (Giakoumatos et al., 2015), and antidepressants confer neuroprotective effects in cell culture and in mouse models (Hunsberger et al., 2009), and promote neurogenesis and process extension in humans (Boldrini et al., 2009), which could mitigate gray matter loss.

In recent years, machine learning (an application of computer science that allows programs to learn from data sets and perform tasks such as pattern classification without direct user input), has been applied to structural brain imaging data in Alzheimer's disease (Zeifman et al., 2015), schizophrenia (Castro et al., 2014), obsessive compulsive disorder (Hoexter et al., 2013) and autism spectrum disorder (Zhou et al., 2014). It has also been applied to structural brain imaging data to identify MDD patients and to predict which patients would be treatment responders with an accuracy of 89% (Patel et al., 2016), and to predict the onset of MDD in healthy adolescents with an accuracy of 70% (Foland-Ross et al., 2015). Machine learning has also been used to classify Bipolar Type I from Bipolar Type II with up to 94% accuracy using Diffusion Tensor Imaging (DTI) data combined with neuropsychiatric measurements (Wu et al., 2016), and to classify pediatric BD compared to healthy volunteers with up to 79% accuracy (Mwangi et al., 2015).

Our work adds to a growing body of literature attempting to distinguish BD and MDD (Cardoso de Almeida and Phillips, 2013) by applying machine learning methods to cortical thickness or voxel-based morphometry (VBM) data. Previous studies have classified MDD and BD subjects with accuracies of 92.1% (Jie et al., 2015), 74.3% (Fung et al., 2015), 79.3% (Redlich et al., 2014), and 69.1% (Rive et al., 2016).

We sought to use brain-wide VBM gray matter volume measurements and machine learning, specifically Support Vector Machine (SVM), to discriminate MDD and BD patients in a cohort of 26 MDD and 26 BD patients. Participants were in a MDE and were antidepressant and mood-stabilizer-medication-free for at least 2 weeks at the time of scanning. Medications like SSRIs and lithium can produce neuron process extension, angiogenesis, and neurogenesis, and can enlarge gray matter (Giakoumatos et al., 2015; Boldrini et al., 2009). Next to nothing is known about the offset of such trophic effects of psychotropic medications and given the burden of untreated major depression, a 2-week drug-free period was considered reasonable.

2. Methods and materials

2.1. Participants

Twenty-six patients with DSM-IV BD (15 with Bipolar Type-I, 11 with Bipolar Type-II), and 26 age/sex-matched patients with DSM-IV MDD were included in this study. Patients were initially assessed by experienced masters or doctoral level psychologists using the Structured Clinical Interview for Axis I disorders (SCID-I/P) (First et al., 1995), followed by psychiatric interview with an experienced research psychiatrist. Final diagnoses were made by consensus in a conference where research psychiatrists and psychologists reviewed all available information. All patients also had a current 17-item Hamilton Depression Rating Scale (HDRS-17) score of 16 or greater (Hamilton, 1960) or a Quick Inventory of Depressive Symptomatology-Self Rated version (QIDS-SR) score $\geq$16 (Rush et al., 2003). Patient ages ranged from 20 to 60 years. Patients taking psychotropic medication at the time of enrollment (n = 1/26 MDD, 1/26 BD), and whose depression was not responding adequately, were tapered off medications and completed a washout of $\geq$2 weeks prior to scanning. No participants were receiving fluoxetine or depot antipsychotics at the time of enrollment. Short-acting benzodiazepines were permitted in small doses. The Young Mania Rating Scale (YMRS) (Young et al., 1978) was used to assess manic symptoms. Four subjects did not have HAM-17 scores, so for these subjects Montgomery–Åsberg Depression Rating Scale (MADRS) scores (Montgomery and Asberg, 1979) were converted to HAM-17 scores (using the ID-QUIDS conversion table http://www.ids-qids.org/index2.html#table 5) in order to evaluate an effect of depression severity on SVM classification. For all participants, exclusion criteria included (i) medical condition that may affect brain integrity assessed by physical examination, medical history, review of systems, and screening laboratory tests, including blood dyscrasias, lymphomas, hypersplenism, endocrinopathies, renal failure or chronic obstructive lung disease, multi-system autoimmune disorders, autonomic neuropathies, peripheral vascular disease, or malignancy, (ii) positive urine toxicology screen, (iii) positive pregnancy test or planned pregnancy, and (iv) alcohol or substance use disorder within the previous six months. This is a secondary analysis of pooled MRI data from patients recruited between May 2007 and April 2016. Twenty out of 52 subjects were from as-yet unpublished studies and the rest were included in earlier reports (Chhetry et al., 2016; Gray et al., 2013; Lan et al., 2013; Miller et al., 2013; Parsey et al., 2010). The Institutional Review Board of the New York State Psychiatric Institute approved all studies from which subjects were gathered, and all participants gave written informed consent.

A total pool of 59 BD and 58 MDD patients had potentially usable data for this analysis. Although all subjects were imaged on the same scanner, two different pulse sequences and head coils were employed in this sample, as described below. As both of these factors can affect GMV measurements (Focke et al., 2011) and were found to impact GMV measurements in our sample (see Supplemental materials S1), subjects were matched on both pulse sequence and head coil to prevent them from influencing the classifier's performance. Since GMV measurements are also affected by sex and age (Peng et al., 2016a), subjects were

further matched on age and sex as follows: for each BD patient, the closest-in-age (with up to a 10 year difference) MDD patient of same sex and with data from the same head coil and scan sequence was selected as its counterpart, resulting in a sample of 26 matched pairs. The remaining subjects (who lacked head coil and sequence-matched counterparts) were used as a second independent cohort to test the classifier in an alternate analysis.

2.2. Image acquisition

Images were acquired using a GE Signa HDx 3 T scanner. Seven subjects in each group were scanned with a 32-channel head coil, while the remaining subjects were scanned with an 8-channel head coil. High-resolution structural 3D T1-weighted magnetization prepared ultrafast spoiled gradient echo (IR-FSPGR) images were acquired in 1 mm slices with repetition times (TR) between 5.0 and 7.5 msec, echo times (TE) between 2.4 and 3.0 msec, flip angles of 9°, acquisition matrix of 256 × 256, and 1 mm^3 voxel size. Twelve subjects with the 8-channel head coil were scanned with TI = 500 and 168 axial slices; all remaining subjects were scanned with TI = 900 and 174 sagittal slices. Subjects were matched on head coil and scan sequence as described above. All scans were visually inspected to check for artifacts or structural abnormalities (none were excluded). All 33 alternate BD subjects were scanned with the 32-channel head coil and TI = 900, all 32 alternate MDD subjects were scanned with the 8-channel coil, with 21 scanned with the TI = 500 sequence and the remaining 11 with the TI = 900 sequence.

2.3. MRI processing and gray matter volume measurement

T1-weighted images were preprocessed and analyzed with Statistical Parametric Mapping 8 (SPM8) software package (www.fil.ion.ucl.ac.uk/spm; Wellcome Department of Imaging Neuroscience, London, UK) using VBM8 toolbox (http://dbm.neuro.uni-jena.de/vbm/) (Ashburner and Friston, 2000; Mechelli et al., 2005) and Matlab 2012a (www.mathworks.com/). Images were bias corrected, segmented, and spatially normalized to standard Montreal Neurological Institute (MNI) space at a voxel size of 1.5 × 1.5 × 1.5 mm^3 using 12-parameter affine linear transformation and diffeomorphic anatomical registration through exponentiated lie algebra (DARTEL) (Ashburner, 2007; Ashburner and Friston, 2009). Segmented gray matter images were multiplied by the measure of warped and unwarped structures derived from the nonlinear step of the spatial normalization. The modulated gray matter images were smoothed with an isotropic Gaussian kernel of 8 mm full width at half maximum (FWHM). Alternate smoothing kernels were examined (see Supplemental materials S2).

2.4. Statistical analysis

SVM, a type of multivariate pattern analysis, was chosen as the machine learning algorithm to classify MDD and BD. SVM classifies data by maximizing the distance between classes in a high-dimensional space (Cherkassky, 1997) and is highly suitable for analysis of high-dimension imaging data sets. SVM has been applied to structural brain imaging data as a diagnostic tool and predictor of illness course and treatment outcome in various pathologies including MDD, BD, schizophrenia and others (Orru et al., 2012).

Statistical analysis was performed using the Library for Support Vector Machines (LIBSVM) software package (Chang, 2011) and Matlab 2012a (www.mathworks.com/). Brain-wide GMV analysis was performed, with no *a priori* regions of interest (ROIs) selected. A leave-two-subjects-out analysis was used to test the classifier: for 26 iterations, one scan and its matched counterpart in the other group were excluded from the sample, and a support vector classifier was trained on the remaining subjects. Each subject was excluded with its counterpart so that the training group was always balanced in terms of scan sequence and head coil. Each classifier was generated using C-SVC with a linear kernel (selected because of the high dimensionality of the data). A default cost parameter of 1 was used. The algorithm was then used to classify the subject pair that was excluded during the training. Accuracy, specificity and sensitivity were calculated across the 26 trials as percentage of subjects classified correctly in the full sample, percentage of MDD subjects correctly identified, and percentage of BD subjects correctly identified, respectively, with consensus diagnosis as the gold standard. A permutation test (Ojala and Garriga, 2010) was used to calculate significance: the group labels were shuffled (while keeping pairs intact) and the leave-two-out analysis was repeated on the randomized data for 1000 trials. The p-value was calculated as the number of trials where the accuracy was equal to or higher than the value obtained in the real test divided by 1000.

Regions that contributed most to the classification were selected as clusters where all voxels had weights greater than 0.0019 or lower than −0.0019 with an extent of 50 or more voxels (cutoff chosen as the average magnitude of weight plus 3 standard deviations). Contributing regions were identified using the FSL Harvard-Oxford atlases (Desikan et al., 2006).

The 65 subjects who lacked head coil and sequence-matched counterparts were used to test a classifier that was trained on the full 26 matched pairs sample.

In order to examine the effect of clinical variables on the classifier, the leave-two-out analysis was repeated using HAM-17 scores, existence of past substance abuse, number of prior depressive episodes, and age of onset as additional features.

2.5. Alternative analysis

Our primary outcome measure derived from the sMRI data was VBM, because it allows examination of both cortical and subcortical structures, and because it provides gray matter characteristics additional to those identified in cortical thickness analyses (Hutton et al., 2009). Fifteen out of 26 BD subjects and 11 out of 26 MDD subjects in this analysis had also been included in our earlier work with structural brain imaging in BD and MDD populations (Lan et al., 2014). Since this earlier work was performed using cortical thickness as measured in Freesurfer (http://surfer.nmr.mgh.harvard.edu), we repeated SVM analyses in this cohort using cortical thickness data estimated by Freesurfer 5.1.0 with the same methods as our previous publication (Dale et al., 1999; Fischl and Dale, 2000). Each subject's cortical thickness was mapped onto a standard average subject's surface. The resulting maps were smoothed with a Gaussian filter with a FWHM of 8 mm. The entire cortical surface was then used in a SVM analysis as described above. Four subjects in each group had to be excluded from this analysis due to poor performance of the Freesurfer software, including under-labeling of occipital regions and poor conformity to anatomy.

3. Results

3.1. Sample

MDD and BD groups were matched on age (MDD: mean age = 37.5 ± 10.4 years; BD: mean age = 37.5 ± 9.7 years) and sex (each group: 41% male), and did not differ in intracranial volume, or prior exposure to antidepressants. MDD and BD subjects did differ on a few clinical variables, largely consistent with each pathology: BD patients had higher HAM-17 scores, more prior depressive episodes, an earlier age of onset, and a greater incidence of remitted substance abuse. As discussed below, these variables did not appear to affect classifier performance. Clinical and demographic data, along with statistical comparisons between groups, are summarized in Table 1.

Table 1
Clinical and demographic information.

	MDD (N = 26)	BD (N = 26)	p-value (2-tailed *t*-test)
Age	37.5 ± 9.72	37.49 ± 10.38	N/A
Intracranial Volume	1384.63 ± 132.89	1383.24 ± 161.02	0.97
Depression Severity (HDRS − 17 or converted MADRS)	17.36 ± 3.84	21.43 ± 4.92	0.002
Length of Current Episode (days)	249.65 ± 309.03	99.22 ± 192.23	0.09
Number of Previous Episodes	2.88 ± 4.43	10.39 ± 8.48	< 0.001
Age of Onset	24.88 ± 7.99	16.67 ± 6.51	< 0.001
Years of Education	16.04 ± 2.63	15.58 ± 2.37	0.51
Categorical Variables	**N (%)**		**p-value (Fisher's exact)**
Male	11 (42.31)	11 (42.31)	N/A
Prior Exposure to AntiDepressants	14 (53.85)	18 (69.23)	0.39
Co-morbid Anxiety Disorder	10 (38.46)	13 (50)	0.58
Relatives with Major Depression	9 (34.62)	12 (46.15)	0.57
Remitted Substance Abuse	2 (7.69)	10 (38.46)	0.02
Race/Ethnicity			
Asian	4 (15.38)	2 (7.69)	
Hispanic	4 (15.38)	2 (7.69)	
African American	4 (15.38)	5 (19.23)	
Caucasian	13 (50)	16 (61.54)	
> 1 Race	1 (3.85)	1 (3.85)	

3.2. Main analysis

In a leave-two-out analysis, the SVM accurately classified 75.0% of subjects. It generated 73.1% sensitivity for BD diagnosis, correctly identifying 19 out of 26 BD subjects, and 76.9% specificity, correctly identifying 20 out of 26 MDD patients. Fig. 1 shows classification values for each subject and summarizes the classifier's accuracy. A receiver operator curve had an area under the curve (AUC) of 0.682, and in our permutation test, p = 0.002. Voxel-by-voxel SVM weights, which represent the contribution of each voxel to the classification, are reported in Fig. 2. Regions that contributed most to the classification included bilateral supramarginal gyrus, occipital pole, and right lateral occipital cortex (MDD > BD) and right dorsolateral superior frontalgyrus (BD > MDD) and are reported in Table 2.

A classifier trained with all 26 matched-pairs and tested on the second cohort was able to correctly classify 19 out of 33 BD subjects (57% sensitivity) and 25 out of 32 MDD subjects (78% specificity), for an overall accuracy of 68%. A permutation test (randomly shuffling all labels) yielded p = 0.004.

Although MDD and BD groups differed in number of previous depressive episodes, age of onset, HAM-17 scores, and number of subjects with past substance abuse, none of these variables differed between misclassified and correctly classified subjects in either group, nor did they differ between all patients classified as MDD and all patients classified as BD (2 sample *t*-test; p > 0.1 for all comparisons). The number of subjects with comorbid anxiety disorders and the number of patients with medication exposure also did not differ between any of these groups. This means that during classification, subjects were evenly distributed between groups in terms of all variables, despite the difference in the training samples. Bipolar subtype did not appear to impact classification as well: of the 7 BD subjects who were misclassified as MDD, 4 were BD type I and 3 were BD type II.

The classifier that used clinical variables as additional features accurately classified 20 out of 26 MDD subjects and 20 out of 26 BD subjects, for a combined accuracy of 77%.

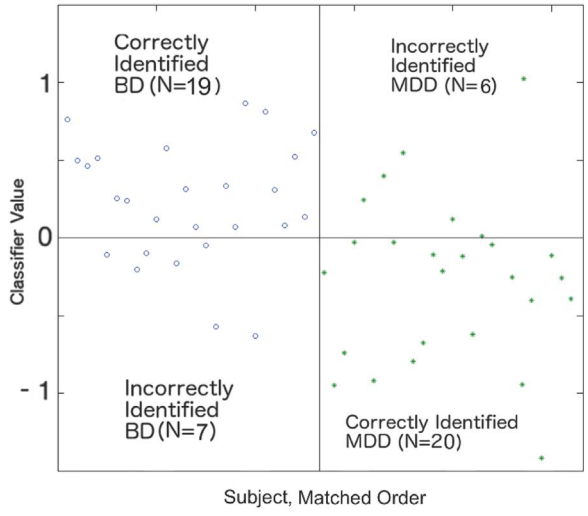

Fig. 1. Confusion matrix. Subjects are arbitrarily ordered along the X-axis with BD on the left and MDD on the right. The Y-axis shows the values that the classifier assigned to each subject in the leave-two-out analysis, with positive values representing a BD classification and negative values a MDD classification.

3.3. Alternative analyses

The cortical thickness analysis had an overall accuracy of 68.2%, with 63.6% sensitivity, 72.2% specificity, a ROC AUC of 0.448 and p = 0.027. Regions that contributed most to the classification using freesurfer data included right supramarginal gyrus, right caudal middle frontal, right parahippocampal gyrus and left insula. They are described in more detail in Supplemental materials S3.

4. Discussion

Using VBM, we were able to classify MDD and BD with comparable accuracy, which suggests that although MDD and BD have different structural brain changes, there is a similar degree of heterogeneity in structural pathology in BD compared with MDD, consistent with our finding that bipolar subtype did not appear to impact classification. The VBM-based classifier performed better than the cortical-thickness based classifier. While cortical thickness analyses measure only one aspect of gray matter, VBM data have been shown to provide mixed measures of gray matter characteristics including cortical thickness, surface area, and folding (Hutton et al., 2009). It is therefore possible that BD and MDD subjects differ in these additional measures. SPM and VBM have been found to have better segmentation accuracy and reliability than FreeSurfer (Eggert et al., 2012), and this might have impacted the accuracy of the classifier as well. Alternatively, the subcortical regions, which are not quantified in cortical thickness analyses, could have contributed to the greater accuracy observed with VBM. However, this is not supported by the regions identified to have the greatest impact on our classification (none of which were subcortical), and recent analyses

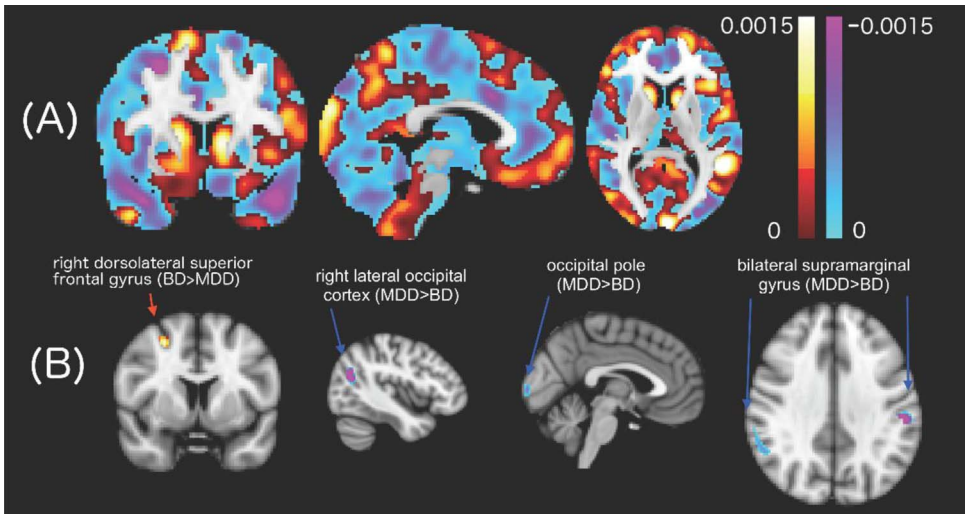

Fig. 2. Regions that contributed most to classification (a) SVM weights overlaid on a T1-weighted MRI image. Red-to-yellow regions represent voxels where GMV is higher for BDs than MDDs, and blue-to-cyan regions represent voxels where GMV is higher for MDDs than BDs. (b) Locations of clusters that contributed most to the classification (as defined by all voxel weights greater than 0.0019 or less than −0.0019 with extent greater than 50 voxels).

using only subcortical regions to classify MDD and BD have not performed as well as those that use cortex, with only 59.5% accuracy (Sacchet et al., 2015).

The clinical variables described in the results section did not differ between subjects classified as MDD and subjects classified as BD. If the SVM algorithm had been influenced by a clinical variable, this would have been evidenced by a difference in values for that variable in the classified groups. Since there was no significant difference of values between the SVM-classified groups for any variable (including those whose values differed in the actual diagnostic groups), there is no evidence that any of these variables impacted our analysis. Although the classifier performed slightly better when these variables were used as additional features compared to when they not used (77% vs. 75% accuracy), the only difference was one additional correctly classified subject. That the subjects were otherwise classified the same way with and without the use of the clinical variables as features is further evidence that the SVM was not heavily influenced by these variables.

Since our SVM analysis used a linear kernel, the regions that were most relevant in distinguishing MDD and BD can be interpreted as the regions with the most gray matter differences between the two pathologies in our sample (Ma and Guo, 2014).

The regions identified in our primary analysis did not overlap with the regions found to have lower cortical thickness in BD compared to MDD in our previous work (Lan et al., 2014), perhaps due to different approaches to age and sex: subjects were matched in this sample, resulting in a much smaller MDD group, while age and sex were included as covariates in our earlier work. The non-overlapping regions identified might also be explained by the different modalities used: right caudal middle frontal was found to have lower cortical thickness in BD in our previous work and here again was associated with MDD classification in our freesurfer-based cortical thickness analysis (see supplement) but not in our primary GMV analysis. Even though no regions in our main analysis overlapped with the BD vs MDD contrast in our earlier study, lower supramarginal gyrus GMV/cortical thickness was found to be associated with BD classification in this analysis using both modalities, and in our older work this same region was found to have less cortical thickness in BD, compared to healthy volunteers but not MDD subjects. This region has been previously shown to have gray matter deficits in BD (Saricicek et al., 2015). A recent large-sample study found brain-wide deficits in BD compared to healthy volunteers in regions including occipital and supramarginal cortex (Hibar et al., 2017), and GMV in right lateral occipital cortex has been found to be lower in BD compared to healthy volunteers but not in MDD compared to the same cohort of healthy volunteers (Niu et al., 2017). Supramarginal gyrus has been shown to be involved in verbal working memory (Deschamps et al., 2014) and empathy (Silani et al., 2013). It has been shown that euthymic BD patients have deficits in verbal working memory (McKenna et al., 2014), and in one study BD patients were found to have empathy deficits whereas MDD patients were comparable to normal controls in most tasks (Derntl et al., 2012). Taken together, these results point to a structural abnormality in BD that can be used to distinguish BD from MDD patients.

The only region in which higher GMV increased likelihood for BD classification was right dorsolateral superior frontal gyrus. A meta-analysis of 329 medication-naïve MDD subjects and 340 healthy volunteers found this region to have the most significantly lower GMV in MDD compared to healthy volunteers (Peng et al., 2016b). This region has been found to be activated during an introspection task (Goldberg et al., 2006), and MDD patients often struggle with introspection, becoming ruminative and preoccupied with feelings of worthlessness, rather than adapting realistic views of themselves (Nolen-Hoeksema, 1991). Our findings suggests that structural deficits in this region may be associated with MDD pathology but not BD pathology.

We were able to identify only one other published study that has used gray matter measurements and machine learning to classify MDD and BD in a medication-free sample, and that study reported 69.1% accuracy (Rive et al., 2016). That study was confined to emotion-regulation related regions, while our analysis examined GMV values brainwide. That our analysis had slightly higher accuracy and found peaks outside of emotion regulation related brain regions implies that

Table 2
Regions that contributed most to the discrimination of BD and MDD using SVM along with peak coordinates. Cutoff chosen arbitrarily at ± 0.0019 k > 50. All coordinates in MNI space (mm).

	Region	Peak Weight	Cluster size (2 mm voxels)	MNI coordinates (mm)
$MDD_{GMV} > BD_{GMV}$	Right lateral occipital cortex	−0.00374	155	42, −66, 22
	Right supramarginal gyrus	−0.00313	105	50, −28, 36
	Left supramarginal gyrus	−0.00219	113	−58, −46, 32
	Medial occipital pole	−0.0022	50	2, −98, 4
$BD_{GMV} > MDD_{GMV}$	Right dorsolateral superior frontal gyrus	0.0025	55	26, 8, 50

BD and MDD patients have brainwide structural differences. We were able to obtain accuracies comparable to other studies (Fung et al., 2015; Redlich et al., 2014; Rive et al., 2016), which had accuracies of 69–79%, even though they included currently medicated subjects. The only study that was able to achieve significantly higher accuracy (92.1%) included resting state fMRI data in addition to sMRI (Jie et al., 2015).

4.1. Limitations

Although we tested our classifier in an independent sample, this sample contained MDD and BD subjects scanned with different headcoils. Because the training sample was completely balanced in terms of headcoil, this should not have impacted the accuracy of the classifier. However, since MDD subjects were 88% 8-channel scans and BD subjects were 68% 32-channel scans, it would be impossible to control for headcoil without regressing out the differences between the two groups, which prevented us from performing an expanded leave-two-out analysis or using alternate cohort groupings. It also makes interpretation of this analysis more difficult, which is why the leave-two-analysis was discussed in more detail above. Further validation with independent datasets from other groups and scanners is a crucial next step.

Our primary analysis was permutation-based in order to examine the properties of a classifier that was trained on as many subjects as possible in our relatively small matched sample. Since this approach is not as convincing as a two-group analysis, a classifier was also trained on 13 of the matched subject groups and tested on the remaining subjects, and it performed with 74% accuracy. Our primary analysis also involved matching subjects in order to minimize confounding factors. However, when this matching was disrupted (by training the classifier on all but one subject at a time and testing it on that subject), the accuracy lowered to 64%, showing that scan parameters and age and sex need to be controlled for in these analyses.

A predictive accuracy of 68% compares favorably to rates of clinical diagnoses in BD, which have been found to be as low as 31% for first diagnoses (Hirschfeld et al., 2003). The diagnostic accuracy reported in other studies, using neuropsychological indices, was higher than that reported here (Wu et al., 2016), although those studies also require independent validation. Another limitation of the current study is that while our diagnostic evaluation was extremely rigorous, it is possible that participants diagnosed as MDD in any cross-sectional study would later develop BD.

Participants were unmedicated at the time of scan. However, a considerable proportion of BD and MDD subjects were previously exposed to antidepressant medication (54% of MDD and 69% of BD subjects). GMV may be altered by medications such as lithium and antipsychotics (Abramovic et al., 2016) and past medication use might have affected results. In addition, subjects had been living with these illnesses for some time and both duration of untreated depression (Sheline et al., 2003) and lifetime number of manic episodes (Ekman et al., 2010) are associated with lower GMV. Thus, the duration of illnesses theoretically could have affected the results. However, prior exposure to anti-depressants and number of prior episodes, as discussed above, did not appear to have any impact on our results.

In high-dimensional analyses, over-fitting is a potential limitation, but since a reasonably low cost function of 1 was selected, this should not have been an issue in our analysis.

4.2. Future directions

This study needs further replication in another sample of mood disorder subjects. An ideal sample in which to test this algorithm would be medication-naïve subjects presenting with a first major depressive episode with ambiguous diagnosis, followed by careful prospective clinical follow-up. The machine-learning algorithm using the sMRI data could be used to predict future conversion to BD. A pragmatic difficulty in such a study would be the low anticipated conversion rate to BD, requiring a very large sample size and a long study time period to obtain a sufficient number of converters. In addition, high retention rates for rigorous longitudinal clinical follow-up would be required. Alternatively, HR offspring of individuals with MDD or BD would be expected to have a higher conversion rate. A previous study has found gray matter reductions in HR offspring compared with healthy offspring, which were more significant for symptomatic than asymptomatic participants, using both cortical thickness (Hanford et al., 2016c) and gray matter volume (Hanford et al., 2016a).

Combining several imaging modalities and cognitive assessments could improve the performance of such classifiers and shed light on the biological differences between these two mood disorders. A large-sample meta-analysis found that resting state fMRI and DTI data were better able to classify MDD and HV than T1 data alone, pointing to the utility of these modalities (Kambeitz et al., 2017). It would also be interesting to examine feature reduction techniques and other machine learning approaches, such as neural networks and gaussian process classifiers, to explore the robustness of these findings.

After validation of this method in a clinical setting, it could be used to help health professionals diagnose patients who present ambiguous symptoms with either MDD or BD. Patients with BD presenting with an initial depressive episode cannot be distinguished from MDD until the first sign of mania or hypomania. This diagnosis could be considered based on an MRI performed relatively quickly (T1 scan takes ~7 min) and classification can be performed in under a minute once the classifier has been trained. The result can then be used to help make an earlier diagnosis of BD and thereby guide treatment choices. The interval between the first mood disorder episode and a clinical diagnosis of BD has been estimated to be about ten years. Accuracy of the classifier can be improved as the training data set grows.

This work distinguishing MDD and BD may be extended to the prediction of treatment outcome. Preliminary work has already been conducted to use machine learning to predict the likelihood of a future suicide attempt in patients with mood disorders (Passos et al., 2016). Structural and other types of brain imaging and machine learning may help classify subtypes of mood disorders and other psychopathologies, providing an adjunctive approach to clinical diagnostic systems.

5. Conclusions

Using only sMRI brain images, we were able to classify BD and unipolar MDD subjects with 75% accuracy, supporting the claim that the two pathologies exhibit differences in brain structure. Similar techniques may one day be used clinically to prevent misdiagnoses, particularly in first-episode medication-naive depressed patients.

Acknowledgements

Dr. R. Todd Ogden was consulted about statistical analyses used.

Contributors

This work used data from projects supported by grants from the National Institute of Mental Health (J.J.M.: P50 MH090964-04, J.J.M.: R01 MH40695-22, J.J.M.: R01 MH108032-03, M.E.S: R21 MH096255-02, M.E.S: K-08 MH079033, M.O: R01 MH48514, J.M.M.: NIMH K-08 MH085061, ML: K23 MH105688) and an Independent Investigator Research Grant from Sunovion Pharmaceuticals.

Role of the funding source

The funding sources were not involved in study design, in the collection, analysis and interpretation of data, in the writing of the report, or in the decision to submit the article for publication.

Conflicts of interest

Dr. Miller's family previously owned stock in Johnson & Johnson, unrelated to the current manuscript.

Drs. Oquendo and Mann receives royalties for the commercial use of the Columbia Suicide Severity Rating Scale. Dr. Oquendo's family owns stock in Bristol Myers Squibb.

Dr. Lan received travel reimbursement to scientific conferences from Sunovion Pharmaceuticals.

Drs. Zanderigo and Sublette, and Mr. Rubin-Falcone have no conflict of interest.

Dr. Hellerstein has received research grant support from Pfizer, Inc., and Eli Lilly, unrelated to the current manuscript.

Dr. McGrath has received research grant support from Naurex (now Allergan), Sunovion, the NIMH, and the State of New York.

Appendix A. Supplementary material

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.jad.2017.11.043.

References

Abramovic, L., Boks, M.P., Vreeker, A., Bouter, D.C., Kruiper, C., Verkooijen, S., van Bergen, A.H., Ophoff, R.A., Kahn, R.S., van Haren, N.E., 2016. The association of antipsychotic medication and lithium with brain measures in patients with bipolar disorder. Eur. Neuropsychopharmacol.: J. Eur. Coll. Neuropsychopharmacol.

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. NeuroImage 38, 95–113.

Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry – the methods. NeuroImage 11, 805–821.

Ashburner, J., Friston, K.J., 2009. Computing average shaped tissue probability templates. NeuroImage 45, 333–341.

Baldessarini, R.J., Vieta, E., Calabrese, J.R., Tohen, M., Bowden, C.L., 2010. Bipolar depression: overview and commentary. Harv. Rev. Psychiatry 18, 143–157.

Boldrini, M., Underwood, M.D., Hen, R., Rosoklija, G.B., Dwork, A.J., John Mann, J., Arango, V., 2009. Antidepressants increase neural progenitor cells in the human hippocampus. Neuropsychopharmacol.: Off. Publ. Am. Coll. Neuropsychopharmacol. 34, 2376–2389.

Cai, Y., Liu, J., Zhang, L., Liao, M., Zhang, Y., Wang, L., Peng, H., He, Z., Li, Z., Li, W., Lu, S., Ding, Y., Li, L., 2015. Grey matter volume abnormalities in patients with bipolar I depressive disorder and unipolar depressive disorder: a voxel-based morphometry study. Neurosci. Bull. 31, 4–12.

Cardoso de Almeida, J.R., Phillips, M.L., 2013. Distinguishing between unipolar depression and bipolar depression: current and future clinical and neuroimaging perspectives. Biol. Psychiatry 73, 111–118.

Castro, E., Gupta, C.N., Martinez-Ramon, M., Calhoun, V.D., Arbabshirani, M.R., Turner, J., 2014. Identification of patterns of gray matter abnormalities in schizophrenia using source-based morphometry and bagging. Conference proceedings: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference 2014. pp. 1513–1516.

Chang, C., Lin, C.J., 2011. LIBSVM: a library for support vector machines. ACM Trans. Intell. Syst. Technol. 2, 27.

Cherkassky, V., 1997. The nature of statistical learning theory. IEEE Trans. Neural Netw. Publ. IEEE Neural Netw. Counc. 8, 1564.

Chhetry, B.T., Hezghia, A., Miller, J.M., Lee, S., Rubin-Falcone, H., Cooper, T.B., Oquendo, M.A., Mann, J.J., Sublette, M.E., 2016. Omega-3 polyunsaturated fatty acid supplementation and white matter changes in major depression. J. Psychiatr. Res. 75, 65–74.

Culpepper, L., 2014. Misdiagnosis of bipolar depression in primary care practices. J. Clin. Psychiatry 75, e05.

Dale, A.M., Fischl, B., Sereno, M.I., 1999. Cortical surface-based analysis. I. Segmentation and surface reconstruction. NeuroImage 9, 179–194.

Derntl, B., Seidel, E.M., Schneider, F., Habel, U., 2012. How specific are emotional deficits? A comparison of empathic abilities in schizophrenia, bipolar and depressed patients. Schizophr. Res. 142, 58–64.

Deschamps, I., Baum, S.R., Gracco, V.L., 2014. On the role of the supramarginal gyrus in phonological processing and verbal working memory: evidence from rTMS studies. Neuropsychologia 53, 39–46.

Desikan, R.S., Segonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31, 968–980.

Eggert, L.D., Sommer, J., Jansen, A., Kircher, T., Konrad, C., 2012. Accuracy and reliability of automated gray matter segmentation pathways on real and simulated structural magnetic resonance images of the human brain. PloS One 7, e45081.

Eker, C., Simsek, F., Yilmazer, E.E., Kitis, O., Cinar, C., Eker, O.D., Coburn, K., Gonul, A.S., 2014. Brain regions associated with risk and resistance for bipolar I disorder: a voxel-based MRI study of patients with bipolar disorder and their healthy siblings. Bipolar Disord. 16, 249–261.

Ekman, C.J., Lind, J., Ryden, E., Ingvar, M., Landen, M., 2010. Manic episodes are associated with grey matter volume reduction - a voxel-based morphometry brain analysis. Acta Psychiatr. Scand. 122, 507–515.

First, M., Spitzer, R., Gibbon, M., Williams, J., 1995. Structured Clinical Interview for DSM-IV Axis I Disorders (SCID-I/P, Version2.0). Biometrics Research Dept., New York State Psychiatric Institute, New York.

Fischl, B., Dale, A.M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. Proc. Natl. Acad. Sci. USA 97, 11050–11055.

Focke, N.K., Helms, G., Kaspar, S., Diederich, C., Toth, V., Dechent, P., Mohr, A., Paulus, W., 2011. Multi-site voxel-based morphometry–not quite there yet. NeuroImage 56, 1164–1170.

Foland-Ross, L.C., Sacchet, M.D., Prasad, G., Gilbert, B., Thompson, P.M., Gotlib, I.H., 2015. Cortical thickness predicts the first onset of major depression in adolescence. International journal of developmental neuroscience: the official journal of the International Society for. Dev. Neurosci. 46, 125–131.

Fung, G., Deng, Y., Zhao, Q., Li, Z., Qu, M., Li, K., Zeng, Y.W., Jin, Z., Ma, Y.T., Yu, X., Wang, Z.R., Shum, D.H., Chan, R.C., 2015. Distinguishing bipolar and major depressive disorders by brain structural morphometry: a pilot study. BMC Psychiatry 15, 298.

Giakoumatos, C.I., Nanda, P., Mathew, I.T., Tandon, N., Shah, J., Bishop, J.R., Clementz, B.A., Pearlson, G.D., Sweeney, J.A., Tamminga, C.A., Keshavan, M.S., 2015. Effects of lithium on cortical thickness and hippocampal subfield volumes in psychotic bipolar disorder. J. Psychiatr. Res. 61, 180–187.

Goldberg, I.I., Harel, M., Malach, R., 2006. When the brain loses its self: prefrontal inactivation during sensorimotor processing. Neuron 50, 329–339.

Gray, N.A., Milak, M.S., DeLorenzo, C., Ogden, R.T., Huang, Y.Y., Mann, J.J., Parsey, R.V., 2013. Antidepressant treatment reduces serotonin-1A autoreceptor binding in major depressive disorder. Biol. Psychiatry 74, 26–31.

Hamilton, M., 1960. A rating scale for depression. J. Neurol. Neurosurg. Psych. 23, 56–62.

Hanford, L.C., Hall, G.B., Minuzzi, L., Sassi, R.B., 2016a. Gray matter volumes in symptomatic and asymptomatic offspring of parents diagnosed with bipolar disorder. Eur. Child Adolesc. Psychiatry 25, 959–967.

Hanford, L.C., Nazarov, A., Hall, G.B., Sassi, R.B., 2016b. Cortical thickness in bipolar disorder: a systematic review. Bipolar Disord. 18, 4–18.

Hanford, L.C., Sassi, R.B., Minuzzi, L., Hall, G.B., 2016c. Cortical thickness in symptomatic and asymptomatic bipolar offspring. Psychiatry Res. 251, 26–33.

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., Kramer, B., Overs, B., Hartberg, C.B., Abe, C., Dima, D., Grotegerd, D., Sprooten, E., Boen, E., Jimenez, E., Howells, F.M., Delvecchio, G., Temmingh, H., Starke, J., Almeida, J.R.C., Goikolea, J.M., Houenou, J., Beard, L.M., Rauer, L., Abramovic, L., Bonnin, M., Ponteduro, M.F., Keil, M., Rive, M.M., Yao, N., Yalin, N., Najt, P., Rosa, P.G., Redlich, R., Trost, S., Hagenaars, S., Fears, S.C., Alonso-Lana, S., van Erp, T.G.M., Nickson, T., Chaim-Avancini, T.M., Meier, T.B., Elvsashagen, T., Haukvik, U.K., Lee, W.H., Schene, A.H., Lloyd, A.J., Young, A.H., Nugent, A., Dale, A.M., Pfennig, A., McIntosh, A.M., Lafer, B., Baune, B.T., Ekman, C.J., Zarate, C.A., Bearden, C.E., Henry, C., Simhandl, C., McDonald, C., Bourne, C., Stein, D.J., Wolf, D.H., Cannon, D.M., Glahn, D.C., Veltman, D.J., Pomarol-Clotet, E., Vieta, E., Canales-Rodriguez, E.J., Nery, F.G., Duran, F.L.S., Busatto, G.F., Roberts, G., Pearlson, G.D., Goodwin, G.M., Kugel, H., Whalley, H.C., Ruhe, H.G., Soares, J.C., Fullerton, J.M., Rybakowski, J.K., Savitz, J., Chaim, K.T., Fatjo-Vilas, M., Soeiro-de-Souza, M.G., Boks, M.P., Zanetti, M.V., Otaduy, M.C.G., Schaufelberger, M.S., Alda, M., Ingvar, M., Phillips, M.L., Kempton, M.J., Bauer, M., Landen, M., Lawrence, N.S., van Haren, N.E.M., Horn, N.R., Freimer, N.B., Gruber, O., Schofield, P.R., Mitchell, P.B., Kahn, R.S., Lenroot, R., Machado-Vieira, R., Ophoff, R.A., Sarro, S., Frangou, S., Satterthwaite, T.D., Hajek, T., Dannlowski, U., Malt, U.F., Arolt, V., Gattaz, W.F., Drevets, W.C., Caseras, X., Agartz, I., Thompson, P.M., Andreassen, O.A., 2017. Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. Mol. Psychiatry.

Hirschfeld, R.M., Lewis, L., Vornik, L.A., 2003. Perceptions and impact of bipolar disorder: how far have we really come? Results of the national depressive and manic-depressive association 2000 survey of individuals with bipolar disorder. J. Clin. Psychiatry 64, 161–174.

Hoexter, M.Q., Miguel, E.C., Diniz, J.B., Shavitt, R.G., Busatto, G.F., Sato, J.R., 2013. Predicting obsessive-compulsive disorder severity combining neuroimaging and machine learning methods. J. Affect. Disord. 150, 1213–1216.

Hunsberger, J., Austin, D.R., Henter, I.D., Chen, G., 2009. The neurotrophic and neuroprotective effects of psychotropic agents. Dialog-. Clin. Neurosci. 11, 333–348.

Hutton, C., Draganski, B., Ashburner, J., Weiskopf, N., 2009. A comparison between voxel-based cortical thickness and voxel-based morphometry in normal aging. NeuroImage 48, 371–380.

Jie, N.F., Zhu, M.H., Ma, X.Y., Osuch, E.A., Wammes, M., Theberge, J., Li, H.D., Zhang, Y., Jiang, T.Z., Sui, J., Calhoun, V.D., 2015. Discriminating bipolar disorder from major depression based on SVM-FoBa: efficient feature selection with multimodal brain imaging data. IEEE Trans. Auton. Ment. Dev. 7, 320–331.

Kambeitz, J., Cabral, C., Sacchet, M.D., Gotlib, I.H., Zahn, R., Serpa, M.H., Walter, M., Falkai, P., Koutsouleris, N., 2017. Detecting neuroimaging biomarkers for depression: a meta-analysis of multivariate pattern recognition studies. Biol. Psychiatry 82, 330–338.

Lan, M.J., Hesselgrave, N., Ciarleglio, A., Ogden, R.T., Sullivan, G.M., Mann, J.J., Parsey, R.V., 2013. Higher pretreatment 5-HT1A receptor binding potential in bipolar disorder depression is associated with treatment remission: a naturalistic treatment pilot PET study. Synapse 67, 773–778.

Lan, M.J., Chhetry, B.T., Oquendo, M.A., Sublette, M.E., Sullivan, G., Mann, J.J., Parsey,

R.V., 2014. Cortical thickness differences between bipolar depression and major depressive disorder. Bipolar Disord. 16, 378–388.
Lish, J.D., Dime-Meenan, S., Whybrow, P.C., Price, R.A., Hirschfeld, R.M., 1994. The National Depressive and Manic-depressive Association (DMDA) survey of bipolar members. J. Affect. Disord. 31, 281–294.
Ma, Y., Guo, G., 2014. Support Vector Machines Applications. Springer Science & Business Media, Morgantown, WV.
Mathers, C., Fat, D.M., Boerma, J.T., Organization, W.H., 2008. The Global Burden of Disease: 2004 Update. World Health Organization.
McKenna, B.S., Sutherland, A.N., Legenkaya, A.P., Eyler, L.T., 2014. Abnormalities of brain response during encoding into verbal working memory among euthymic patients with bipolar disorder. Bipolar Disord. 16, 289–299.
Mechelli, A., Price, C.J., Friston, K.J., Ashburner, J., 2005. Voxel-based morphometry of the human brain: methods and applications. Curr. Med. Imaging Rev. 1, 105–113.
Miao Chang, F.Y.W., Kale Edmiston, E., Bai, Chuan, Zhou, Qian, Jiang, Xiaowei, Wei, Shengnan, Wei, Yange, Ye, Yuting, Huang, Haiyan, He, Yong, Xu, Ke, Tang, Yanqing, Wang, Fei, 2017. Neurobiological commonalities and distinctions among three major psychiatric diagnostic categories: a structural MRI study. Schizophr. Bull.
Miller, J.M., Hesselgrave, N., Ogden, R.T., Zanderigo, F., Oquendo, M.A., Mann, J.J., Parsey, R.V., 2013. Brain serotonin 1A receptor binding as a predictor of treatment outcome in major depressive disorder. Biol. Psychiatry 74, 760–767.
Montgomery, S.A., Asberg, M., 1979. A new depression scale designed to be sensitive to change. Br. J. Psychiatry.: J. Ment. Sci. 134, 382–389.
Mwangi, B., Wu, M.J., Bauer, I.E., Modi, H., Zeni, C.P., Zunta-Soares, G.B., Hasan, K.M., Soares, J.C., 2015. Predictive classification of pediatric bipolar disorder using atlas-based diffusion weighted imaging and support vector machines. Psychiatry Res. 234, 265–271.
Nasrallah, H.A., 2015. Consequences of misdiagnosis: inaccurate treatment and poor patient outcomes in bipolar disorder. J. Clin. Psychiatry 76, e1328.
Niu, M., Wang, Y., Jia, Y., Wang, J., Zhong, S., Lin, J., Sun, Y., Zhao, L., Liu, X., Huang, L., Huang, R., 2017. Common and specific abnormalities in cortical thickness in patients with major depressive and bipolar disorders. EBioMedicine 16, 162–171.
Nolen-Hoeksema, S., 1991. Responses to depression and their effects on the duration of depressive episodes. J. Abnorm Psychol. 100, 569–582.
Oertel-Knochel, V., Reuter, J., Reinke, B., Marbach, K., Feddern, R., Alves, G., Prvulovic, D., Linden, D.E., Knochel, C., 2015. Association between age of disease-onset, cognitive performance and cortical thickness in bipolar disorders. J. Affect. Disord. 174, 627–635.
Ojala, M., Garriga, G.C., 2010. Permutation tests for studying classifier performance. J. Mach. Learn Res. 11, 1833–1863.
Orru, G., Pettersson-Yeo, W., Marquand, A.F., Sartori, G., Mechelli, A., 2012. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: a critical review. Neurosci. Biobehav. Rev. 36, 1140–1152.
Papmeyer, M., Giles, S., Sussmann, J.E., Kielty, S., Stewart, T., Lawrie, S.M., Whalley, H.C., McIntosh, A.M., 2015. Cortical thickness in individuals at high familial risk of mood disorders as they develop major depressive disorder. Biol. Psychiatry 78, 58–66.
Parsey, R.V., Ogden, R.T., Miller, J.M., Tin, A., Hesselgrave, N., Goldstein, E., Mikhno, A., Milak, M., Zanderigo, F., Sullivan, G.M., Oquendo, M.A., Mann, J.J., 2010. Higher serotonin 1A binding in a second major depression cohort: modeling and reference region considerations. Biol. Psychiatry 68, 170–178.
Passos, I.C., Mwangi, B., Cao, B., Hamilton, J.E., Wu, M.J., Zhang, X.Y., Zunta-Soares, G.B., Quevedo, J., Kauer-Sant'Anna, M., Kapczinski, F., Soares, J.C., 2016. Identifying a clinical signature of suicidality among patients with mood disorders: a pilot study using a machine learning approach. J. Affect. Disord. 193, 109–116.
Patel, M.J., Khalaf, A., Aizenstein, H.J., 2016. Studying depression using imaging and machine learning methods. NeuroImage. Clin. 10, 115–123.
Peng, F., Wang, L., Geng, Z., Zhu, Q., Song, Z., 2016a. A cross-sectional voxel-based morphometric study of age- and sex-related changes in gray matter volume in the normal aging brain. J. Comput. Assist. Tomogr. 40, 307–315.
Peng, W., Chen, Z., Yin, L., Jia, Z., Gong, Q., 2016b. Essential brain structural alterations in major depressive disorder: a voxel-wise meta-analysis on first episode, medication-naive patients. J. Affect. Disord. 199, 114–123.
Redlich, R., Almeida, J.J., Grotegerd, D., Opel, N., Kugel, H., Heindel, W., Arolt, V., Phillips, M.L., Dannlowski, U., 2014. Brain morphometric biomarkers distinguishing unipolar and bipolar depression. A voxel-based morphometry-pattern classification approach. JAMA Psychiatry 71, 1222–1230.
Rive, M.M., Redlich, R., Schmaal, L., Marquand, A.F., Dannlowski, U., Grotegerd, D., Veltman, D.J., Schene, A.H., Ruhe, H.G., 2016. Distinguishing medication-free subjects with unipolar disorder from subjects with bipolar disorder: state matters. Bipolar Disord. 18, 612–623.
Rush, A.J., Trivedi, M.H., Ibrahim, H.M., Carmody, T.J., Arnow, B., Klein, D.N., Markowitz, J.C., Ninan, P.T., Kornstein, S., Manber, R., Thase, M.E., Kocsis, J.H., Keller, M.B., 2003. The 16-Item quick inventory of depressive symptomatology (QIDS), clinician rating (QIDS-C), and self-report (QIDS-SR): a psychometric evaluation in patients with chronic major depression. Biol. Psychiatry 54, 573–583.
Sacchet, M.D., Livermore, E.E., Iglesias, J.E., Glover, G.H., Gotlib, I.H., 2015. Subcortical volumes differentiate major depressive disorder, bipolar disorder, and remitted major depressive disorder. J. Psychiatr. Res. 68, 91–98.
Sachs, G.S., Nierenberg, A.A., Calabrese, J.R., Marangell, L.B., Wisniewski, S.R., Gyulai, L., Friedman, E.S., Bowden, C.L., Fossey, M.D., Ostacher, M.J., Ketter, T.A., Patel, J., Hauser, P., Rapport, D., Martinez, J.M., Allen, M.H., Miklowitz, D.J., Otto, M.W., Dennehy, E.B., Thase, M.E., 2007. Effectiveness of adjunctive antidepressant treatment for bipolar depression. N. Engl. J. Med. 356, 1711–1722.
Saricicek, A., Yalin, N., Hidiroglu, C., Cavusoglu, B., Tas, C., Ceylan, D., Zorlu, N., Ada, E., Tunca, Z., Ozerdem, A., 2015. Neuroanatomical correlates of genetic risk for bipolar disorder: a voxel-based morphometry study in bipolar type I patients and healthy first degree relatives. J. Affect. Disord. 186, 110–118.
Schneider, B., Muller, M.J., Philipp, M., 2001. Mortality in affective disorders. J. Affect. Disord. 65, 263–274.
Sheline, Y.I., Gado, M.H., Kraemer, H.C., 2003. Untreated depression and hippocampal volume loss. Am. J. Psychiatry 160, 1516–1518.
Silani, G., Lamm, C., Ruff, C.C., Singer, T., 2013. Right supramarginal gyrus is crucial to overcome emotional egocentricity bias in social judgments. J. Neurosci.: Off. J. Soc. Neurosci. 33, 15466–15476.
Wise, T., Radua, J., Via, E., Cardoner, N., Abe, O., Adams, T.M., Amico, F., Cheng, Y., Cole, J.H., de Azevedo Marques Perico, C., Dickstein, D.P., Farrow, T.F., Frodl, T., Wagner, G., Gotlib, I.H., Gruber, O., Ham, B.J., Job, D.E., Kempton, M.J., Kim, M.J., Koolschijn, P.C., Malhi, G.S., Mataix-Cols, D., McIntosh, A.M., Nugent, A.C., O'Brien, J.T., Pezzoli, S., Phillips, M.L., Sachdev, P.S., Salvadore, G., Selvaraj, S., Stanfield, A.C., Thomas, A.J., van Tol, M.J., van der Wee, N.J., Veltman, D.J., Young, A.H., Fu, C.H., Cleare, A.J., Arnone, D., 2016. Common and distinct patterns of grey-matter volume alteration in major depression and bipolar disorder: evidence from voxel-based meta-analysis. Mol. Psychiatry.
Wu, M.J., Mwangi, B., Bauer, I.E., Passos, I.C., Sanches, M., Zunta-Soares, G.B., Meyer, T.D., Hasan, K.M., Soares, J.C., 2016. Identification and individualized prediction of clinical phenotypes in bipolar disorders using neurocognitive data, neuroimaging scans and machine learning. NeuroImage.
Young, R.C., Biggs, J.T., Ziegler, V.E., Meyer, D.A., 1978. A rating scale for mania: reliability, validity and sensitivity. Br. J. Psychiatry.: J. Ment. Sci. 133, 429–435.
Zeifman, L.E., Eddy, W.F., Lopez, O.L., Kuller, L.H., Raji, C., Thompson, P.M., Becker, J.T., 2015. Voxel level survival analysis of grey matter volume and incident mild cognitive impairment or Alzheimer's disease. J. Alzheimer's Dis.: JAD 46, 167–178.
Zhang, H., Li, L., Wu, M., Chen, Z., Hu, X., Chen, Y., Zhu, H., Jia, Z., Gong, Q., 2016. Brain gray matter alterations in first episodes of depression: a meta-analysis of whole-brain studies. Neurosci. Biobehav. Rev. 60, 43–50.
Zhou, Y., Yu, F., Duong, T., 2014. Multiparametric MRI characterization and prediction in autism spectrum disorder using graph theory and machine learning. PloS One 9, e90405.