Authors' preprint (last submitted version)
Published in Eur Neuropsychopharmacol. 2020 May:34:28-38.
doi: 10.1016/j.euroneuro.2020.03.008.

Predicting differential diagnosis between bipolar and unipolar depression with multiple kernel learning on multimodal structural neuroimaging

Vai Benedetta[1,2,3], Parenti Lorenzo[1], Bollettini Irene[1], Cara Cristina[1], Verga Chiara[1], Melloni Elisa[1], Mazza Elena[1,2], Poletti Sara[1,2], Colombo Cristina[1,2], Benedetti Francesco[1,2]

[1] Psychiatry and Clinical Psychobiology, Division of Neuroscience, IRCCS Ospedale San Raffaele, Milano, Italy
[2] University Vita-Salute San Raffaele, Milano, Italy.
[3] Fondazione Centro San Raffaele, Milano, Italy

Corresponding author:
Dr. Benedetta Vai, Ph.D.
Division od Neuroscience
Istituto Scientifico Ospedale San Raffaele
San Raffaele Turro
Via Stamira d'Ancona 20
Milano
Italy
Tel +39/02/26433156
Fax +39/02/26433265
E-mail vai.benedetta@hsr.it

Abstract

One of the greatest challenges in providing early effective treatment in mood disorders is the early differential diagnosis between major depression (MDD) and bipolar disorder (BD). A remarkable need exists to identify reliable biomarkers for these disorders. We integrate structural neuroimaging techniques (i.e. Tract-based Spatial Statistics, TBSS, and Voxel-based morphometry) in a multiple kernel learning procedure in order to define a predictive function of BD against MDD diagnosis in a sample of 148 patients. We achieved a balanced accuracy of 73.65% with a sensitivity for BD of 74.32% and specificity for MDD of 72.97%. Mass-univariates analyses showed reduced grey matter volume in right hippocampus, amygdala, parahippocampal, fusiform gyrus, insula, rolandic and frontal operculum and cerebellum, in BD compared to MDD. Volumes in these regions and in anterior cingulate cortex were also reduced in BD compared to healthy controls (n=74). TBSS analyses revealed widespread significant effects of diagnosis on fractional anisotropy, axial, radial, and mean diffusivity in several white matter tracts, suggesting disruption of white matter microstructure in depressed patients compared to healthy controls, with worse pattern for MDD. To best of our knowledge, this is the first study combining grey matter and diffusion tensor imaging in predicting BD and MDD diagnosis. Our results prompt brain quantitative biomarkers and multiple kernel learning as promising tool for personalized treatment in mood disorders.

Keyword: bipolar disorder, machine learning, grey matter, white matter, biomarker, multiple kernel learning.

1 Introduction

The burden of mood disorders (major depression, MDD; bipolar disorder, BD) is growing despite the availability of new interventions and reducing it requires major shifts in research, clinical practice, and public health by incorporating multidisciplinary intervention models (Wittchen, 2012). One of the greatest challenges in providing effective interventions for these disorders is the differential diagnosis between MDD and BD, which often interferes with the selection of the optimal treatment: due to overlapping psychopathology, ~ 60% of depressed BD patients are initially misdiagnosed as MDD and wait on average 5–10 years for a proper diagnosis. Depression is the predominant abnormal mood state in BD (Kupka et al., 2007), and although distinct clinical features are associated with BD, including more frequent depressive episodes, psychotic symptoms, a family history of BD and poor response to pharmacological treatments, the MDD/BD discrimination capability is largely affected by i) the comparable symptom profiles of depressive episodes and ii) the lower prevalence of hypo/manic episodes during BD illness (Hirschfeld et al., 2003; Judd et al., 2003). Misdiagnosing of BD as MDD has severe consequences in terms of inadequate pharmacological therapy, leads to poor prognosis, increased switches to mania and greater health care costs (Goodwin, 2012; Hirschfeld et al., 2003; Judd et al., 2003). A remarkable need exists to identify reliable biomarkers that can help a more rapid diagnosis of BD and successful treatment selection.

Over the last decades, advanced neuroimaging techniques have been increasingly employed in studying the neural underpinning of mood disorders. Average-group differences in structural Magnetic Resonance highlighted more widespread white matter (WM) abnormalities in BD than MD affecting corpus callosum and in cingulum (Benedetti et al., 2011a; Cardoso de Almeida and Phillips, 2013; Matsuoka et al., 2017; Repple et al., 2017) paralleled by spatially disease-specific patterns of grey matter (GM) volume especially involving corticolimbic circuitry such as anterior cingulate cortex, prefrontal cortex and hippocampus (Matsuo et al., 2019; Niida et al., 2019).

In recent years, we observed a growing interest in machine learning procedures in order to provide automatic and objective predictive models. This is particularly remarkable when the clinical diagnosis of the considered illness remains uncertain (Schrouff et al., 2013). A very sensitive procedure in neuroimaging is provided by multivariate pattern recognition analyses and methods based on kernel. Kernel functions allow summarizing

original data (i.e.voxels) in pairwise similarity measures among subjects, with advantages in terms of computational efficiency. Furthermore, together with regularization constraints, kernel functions allow to face two relevant issues in neuroimaging: the ill-conditioned problem (dimensionality of data is very high compared to the number of observations) and overfitting (Schrouff et al., 2013; Shawe-Taylor and Cristianini, 2004). Previous studies implementing machine learning procedures on structural imaging classified BD against MDD subjects on GM measures with an accuracy of 73.1% (Rubin-Falcone et al., 2018), 92.1% (combined with resting state data) (Jie et al., 2015) , 62.5% (Fung et al., 2015), 75.9% (Redlich et al., 2014), 75.0% (in youth) (MacMaster et al., 2014), 62.86% (Sacchet et al., 2015) and 70% (combined with rest) (Rive et al., 2016), and on WM microstructure with an accuracy of 62.7% (Deng et al., 2018). Despite promising, these studies have not ever combined different GM and WM brain imaging modalities, and their results are potentially biased by small sample size (n per group <50 in 7 out of 8).

The aim of this study is to integrate cutting-edge structural neuroimaging voxel-wise techniques (i.e. Tract-based Spatial Statistics, TBSS (Smith et al., 2006) and Voxel-based morphometry, VBM (Ashburner and Friston, 2000)) in a kernel based machine learning procedure in order to define a predictive function of BD diagnosis against MDD in a sample of 148 patients. To combine heterogenous data (TBSS and VBM) and to enhance the interpretability of the decision function and improve performances (Lanckriet et al., 2004), we implemented Multiple Kernel Learning (MKL) algorithm where optimal combination of each kernel is learned, considering the interdependent information of the features. MKL have been previously implemented in psychiatry in order to classify patients with first-episode psychosis or schizophrenia vs healthy controls (HC) by neuroimaging data (Castro et al., 2011; Squarcina et al., 2017).

Conventional mass-univariate analyses are also performed including a sample of HC in order to explore specific anatomical differences among groups.

Implementing quantitative biological biomarkers in machine learning routine will help to translate the cutting-edge scientific knowledge into tools for personalized treatment, strengthening the impact of biomarkers research into clinical practice.

2 Experimental procedure

2.1 Participants and Data collection

We studied 222 participants (74 HC, 74 patients with a current diagnosis of MDD and 74 with BD-I, SCID-I, DSM-IV TR). MDD and BD patients were comparable for age, and sex and had a major depressive episode without psychotic features. Current severity of depression was rated on the Hamilton Depression Rating Scale (HDRS) (Hamilton, 1960). All subjects were recruited in parallel and scanned between 2.30 pm and 6 pm. Exclusion criteria for all subjects were: major medical and neurological disorders, history of drug or alcohol abuse or dependency; recruited MDD had no family history of BD in first-degree relatives in order to reduce possible confounding effect related to shared endophenotypes; HC had no previous history of psychiatric disorders or first-degree familiarity; patients had no comorbidity on axis I. The study was performed in the framework of two national grants (Italian Ministry of Health) focused on improving treatment and prognosis of mood disorders unveiling their neural and biological underpinnings. After a complete description of the study, written informed consent was obtained. All procedures contributing to this work comply with the ethical standards of the relevant national and institutional committees on human experimentation and with the Helsinki Declaration of 1975, as revised in 2008.

T1-weighted and Diffusion Tensor Imaging (DTI) were acquired on a 3.0 Tesla scanner (Gyroscan Intera, Philips, Netherlands) employing a 8 channels SENSE head coil using a T-1-weighted MPRAGE sequences (TR 25.00 ms, TE 4.6 ms, field of view FOV=230 mm, matrix=256×256, in-plane resolution 0.9×0.9 mm, yielding 220 transversal slices with a thickness of 0.8 mm), and using SE EPI (TR/TE=9000/58 ms, FoV (mm) 232(ap), 126 (fh), 240.00 (rl); acquisition matrix = 112×85; voxel acquisition 2.14×2.73×2.3; 55 contiguous, with in-plane voxel size 1.88×1.88 mm; SENSE acceleration factor=2; 1 b0 and 35 non-collinear directions of the diffusion gradients; b value=900 s/mm2). Fat saturation was performed to avoid chemical shift artifacts.

On the same occasion and using the same magnet 22 Turbo Spin Echo (TSE), T2 axial slices (TR =3000 ms; TE =85 ms; flip angle =90°; turbo factor 15; 5-mm- thick, axial slices with a 512×512 matrix and a 230×230 mm2 field of view) were acquired to rule out brain lesions.

2.2 Neuroimaging analyses

T1-weighted images were processed for VBM by using Computational Anatomy Toolbox (CAT12) (Gaser and Dahnke) for SPM12 (www.fil.ac.uk/spm/) in Matlab R2016b. This included segmentation into GM, WM and cerebrospinal fluid, bias regularization, non-linear modulation, and normalization to MNI space using

DARTEL (Ashburner, 2007) to a 1.5 mm isotropic MNI template. The homogeneity of grey matter images was checked using the covariance structure, as implemented in CAT12, and highlighted no outliers (SD >3). Each individual GM images were smoothed with an 8-mm full width at half maximum Gaussian filter. Measures of Total Intracranial Volume (TIV) were obtained.

Tensor calculations were done using the using the FMRIB Software Library (FSL 5.0) (Smith et al., 2006; Smith et al., 2004; Woolrich et al., 2009). Images were corrected for the effects of eddy currents and head motion (FLIRT, FMRIB's Linear Registration Tool) (Jenkinson et al., 2002; Jenkinson and Smith, 2001), and segmented (BET, Brain Extraction Tool software) (Smith, 2002); then, a diffusion tensor model was fitted at each voxel to each subject's fractional anisotropy (FA) image using FDT (FMRIB's Diffusion Tool) (Behrens et al., 2003), and FA maps were calculated. Next, the data of all the subjects were merged into a common 4D image and skeletonized, as used in TBSS (Smith et al., 2006), in order to focus on the centers of all fiber bundles that are common to the participants (the most compact WM skeleton). All maps were nonlinearly co-registered to the FMRIB58-FA template and normalized to the MNI space. A threshold of FA > 0.2 was applied to the skeleton to include only major fiber bundles. Analogous processes allowed to generate axial (AD), radial (RD) and mean diffusivity (MD) images.

We investigated the effect of diagnosis on regional GM volumes and DTI measures in the whole brain with univariate statistical analyses, entering diagnosis as categorical predictor, and age, sex, medication load, number of previous episodes, and TIV (only for GM analysis) as nuisance covariates. For medication load we used previously developed criteria and categorized each medication into low-dose or high-dose groupings, scored as 0 (no medication), 1 (low dosage), or 2 (high dosage). We then combined all individual medication scores for each medication category in each individual participant to obtain a single composite score (Sackeim, 2001). This procedure was already used in order to differentiate MDD and BD patients on WM microstructure (Benedetti et al., 2011a). A detailed description of the medication used is reported in supplementary materials (SM1). Post-hoc pairwise Family Wise Error (FWE) corrected comparisons were performed. VBM analyses were FWE corrected for multiple comparisons ($p<0.05$) at cluster level (peak $p<0.001$ uncorrected). For DTI measures, voxel-wise statistical analyses across-subjects were performed on the resulting skeletonized images using the FSL Randomise program, based on non-parametric permutation inference, within the framework of the GLM design setup. Threshold-Free Cluster Enhancement (TFCE)

(Smith and Nichols, 2009) to estimate voxel-wise levels of significance calculated with standard permutation (n=10000) testing and corrected for multiple comparisons (p_{FWE}<0.05) in a minimum cluster size of k=100 (Nichols and Holmes, 2002). We also explored the interactions between nuisance covariates (i.e. age, sex, medication load, and number of previous episodes) with diagnosis in affecting VBM and DTI measures. No significant effects were found.

2.3 Multiple Kernel Learning

We performed multivariate pattern analyses by using Multiple Kernel learning (MKL) for binary classification as implemented in PRoNTo software version 2.1 (http://www.mlnl.cs.ucl.ac.uk/pronto). In order to differentiate MDD and BD linear kernels (i.e. dot products) were built separately for the five feature sets: whole brain smoothed GM volume images (VBM) and skeletonized DTI measures brain maps (MD, RD, AD, FA). In MKL heterogenous kernels are linearly combined (Lanckriet et al., 2004): each kernel allow to define a decision boundaries based on support vector machines (SVM, (Rakotomamonjy, 2008)), these are weighted in order to define a global decision boundary. However, the regularization constraints entered in order to avoid overfitting can lead some kernels to have a null contribution, this could not be achieved when kernel are concatenated together in single SVM (Schrouff et al., 2016), ignoring the relative contribution of each modality. MKL include a weighting phase of the different modalities (in contrast with the SVM approach) (Donini et al., 2019). Furthermore, MKL instead of a single SVM was showed to enhance the interpretability of the decision function and improve performances (Lanckriet et al., 2004).

The five kernels for VBM, FA, MD, RD, and AD were mean centered and normalized before classification. Confound effects for age, sex, medication load, number of previous episodes, and TIV were removed from the kernel using a residual forming matrix framework (Chu et al., 2011; Mourão- Miranda et al., 2012). In supervised pattern recognition analysis, the predictive function is usually defined during a training phase where the algorithm learns patterns from the provided data in order to predict a label or a target. Then, during a test phase, the algorithm is used to predict outcome in an independent dataset. Predictions obtained are then compared to the true labels, providing measures of algorithm accuracy. In our study we performed 10-folds nested cross-validations on subjects per group to optimize the model's hyperparameter (i.e. soft-margin C ranged 0.01, 0.1, 1, 10, 100, 1000, inner loop) and to compute model performance (i.e. outer loop). Nested

cross-validation schemes have been showed to provided more reliable estimate than other validation procedure such as leave-one-out (Varoquaux, 2018). In the outer loop, 10-folds nested cross-validation MKL was trained on 90% of the sample and used to predict diagnosis on the remaining 10%, testing the related accuracy. This procedure was repeated recursively for 10 times allowing considering each subject once for testing. Hyperparameter was tuned applying the same procedure in the training sample (i.e. inner loop): the parameter with the highest performance (balanced accuracy, BA) is then applied in the outer loop (Schrouff et al., 2013). Predictions obtained in the test set allow to BA, computed as the average of the class accuracies (corresponding to the sensitivity for BD and specificity for MDD), positive and negative predictive values and area under the receiver operator curve (AUC). Each class accuracy was calculated averaging classification results across the 10 folds of cross validation (Schrouff et al., 2018). We also estimated the statistical significance ($p<0.05$) of accuracies by using 5000 permutations of the labels during training phase. MKL analyses classifying HC and mood disorders are included in supplementary materials (SM2).

3 Results

Our final sample included 74 HC, 74 MDD and 74 BD. No significant differences were found for age, sex, years of education, HDRS and medication load comparing BD and MDD; number of previous mood episodes and duration of illness (years) were significantly higher in BD (table 1). HC were younger, and with a higher education and number of males compared to both MDD and BD.

For VBM analyses, a significant main effect of diagnosis (fig. 2, table in supplementary materials SM1) was found on large clusters which cover right hippocampus, amygdala, parahippocampal, fusiform and lingual gyrus, precuneus, thalamus, putamen and orbitofrontal cortex (MNI 42 4 -22 , k= 4760, $p_{FWE} < 0.001$), cerebellar vermis and crus (MNI 4 -75 -28, k= 819, $p_{FWE} = 0.01$), rolandic and frontal operculum and insula (MNI 52 4 0, k= 1147, $p_{FWE} = 0.002$), anterior and middle cingulate cortex (MNI 10 38 14, k= 1880, $p_{FWE} < 0.001$). Post-hoc tests revealed that for all significant cluster BD patients showed reduced volume compared to HC, whereas no significant differences were found between MDD and HC. Finally, MDD showed higher GM volume compared to BD in all clusters excepted for anterior and middle cingulate cortex, pointing out intermediate pattern of GM reduction between HC and BD.

Diagnosis of mood disorder had widespread significant effects in reducing FA (p_{FWE}=0.0041) and AD (p_{FWE}=0.0025), and increasing RD (p_{FWE}=0.0021) and MD (p_{FWE}=0.0102) in several WM tracts encompassing forceps minor and corpus callosum, cingulum bundle, uncinate fasciculus, anterior thalamic radiation, inferior and superior longitudinal fasciculus, inferior fronto-occipital fasciculus, and corticospinal tract (Table SM2, Fig. 2). Post-hoc pairwise comparisons (t-tests) showed significant spread higher FA and lower RD, in HC compared to both patients with BD and MDD. In turn, the latter showed also spread lower AD and higher MD compared to HC. These effects were not found in BD. MDD also showed lower FA and lower AD than patients with BD in uncinate, cingulum bundle, inferior fronto-occipital, inferior and superior longitudinal fasciculus, forceps minor and major, corticospinal tract and anterior thalamic radiation; with higher RD in uncinate, inferior fronto-occipital, forceps minor, and cingulum bundle (see detailed list in Table SM3).

The MKL produced a BA of 73.65% (p= 0.001), BD sensitivity of 74.32% (p=0.016), MDD specificity of 72.97% (p=0.024) and the AUC was 79% (Fig. 3, Tab. 2, figure SM1). Class predictive value was 73.33% for BD and 73.97% for MDD. Thus, model performance significantly exceeds the one expected by randomly guessing the labels, suggesting that the algorithm has successfully learned a predictive function (Schrouff et al., 2016). Margin C=1 obtained best performance (Frequency of selection 60%), followed by C=10 (20%) and C=1000 (20%). All five modalities contribute to MKL global decision: VBM showed the highest contribution (58% of total weight) followed by FA (24%), AD (11%), MD (4%), RD (3%) (additional SVMs, performed for each brain modality separately, can be found in table SM4). Finally, in order to explore a possible effect of nuisance covariates in biasing the MKL performance, we modelled logistic regressions exploring their main effects and their interactions with diagnosis in predicting classification outcome (MDD vs BD). No significant effects were found.

4 Discussion

In this study, we aimed to distinguish MDD and BD by using MKL on multimodal structural neuroimaging data of GM matter volume and WM integrity correcting for age, sex, number of previous episodes, medication load and TIV. Our predictive function was able to detect BD patients effectively affected with a significant accuracy of 74.32%, with a positive predictive value of 73.33%, which indicates the probability that subjects

with a positive BD screening test is truly affected. The accuracy of MDD was 72.97%, indicating the ability to correctly identify affected MDD, with a predictive value of 73.97%.

The MKL reached good significant accuracies (>70%, $p<0.05$) prompting multimodal structural neuroimaging as a reliable candidate biomarker for the differential diagnosis between BD and MDD. Despite previous studies, implementing machine learning procedures for classifying differential diagnosis on structural neuroimaging also achieved high accuracies (> 70%), the present one is the first study in mood disorders which included relevant nuisance covariates in pattern recognition such as age, sex, number of previous episodes and medication load, which may confound and bias the accuracy estimation. This is particularly relevant when considering previous evidence of the widespread effect of drug used in depression, lithium, multiple recurrent episodes and chronicity on considered brain structural measures (Hibar et al., 2018; McDonald, 2015; Schmaal et al., 2017). Furthermore, previous studies never combined multimodal structural imaging of GM volumes and WM microstructure. DTI measures emerged as candidate biomarkers of mood disorders and our results suggest that multi-modal neuroimaging techniques may provide reliable predictive models for differentiating mood disorders. In our model, VBM-based kernel contributed to 58% of total weight. Notably, the global contribution of DTI measures (42%) confirmed the role of WM integrity in differentiating the two diagnostic groups. DTI measures reflect myelination and WM integrity, coherence of WM tracts, axonal structural organization in the core WM skeleton (Schmierer et al., 2007). Though speculative, a tentative explanation of the greater contribution of FA (24%), and AD (11%), in respect to MD (4%), and RD (3%), in discriminating patients, suggest a major role for neuronal structures in respect to myelin sheaths and integrity and the microtubular structure of the axon (FA, AD), given that RD is mostly associated with the integrity and thickness of myelin sheaths (Song et al., 2002).

Furthermore, available literature results might be biased by the small or unbalanced sample sizes (i.e. BD=26, MDD=26 (Rubin-Falcone et al., 2018), BD=21, MDD=25 (Jie et al., 2015); BD=16, MDD=19 (Fung et al., 2015), BD=14, MDD=32 (MacMaster et al., 2014), BD=40, MDD=57 (Sacchet et al., 2015), BD=36, MDD=45 (Rive et al., 2016), BD=31, MDD=36 (Deng et al., 2018)), and by the implemented cross-validation procedure (Gao et al., 2018). The nested cross validation, used in our study, provides more stable performance compared to Leave-One-Out procedure, implemented in most of previous studies, which may lead to weak generalization and high overfitting (Elisseeff and Pontil, 2003; Gao et al., 2018).

Mass-univariate analyses revealed reduced GM volumes in BD compared to MDD in hippocampus, fusiform and lingual gyrus, amygdala, putamen, parahippocampal gyrus, cerebellar vermis and crus, rolandic and frontal operculum, insula and temporal pole. Volumes in these regions and in anterior cingulate cortex were also reduced in BD compared to HC. All these regions exert a crucial role in affective processing and mood regulation and volume and functional abnormalities in anterior cingulate, amygdala, hippocampus, insula, temporal cortex, thalamus, striatum and cerebellum have been widely replicated in both MDD and BD compared to HC and involved in the pathophysiology of the disorders (Drevets et al., 2008; Vai et al., 2019), also related to antidepressant response and current symptomatology (Fonseka et al., 2018; Frodl, 2017; Godlewska et al., 2018).

Few studies directly compared the disorders for GM volumes: lower grey matter volume in anterior cingulate cortex in BD compared to MDD was confirmed in a previous study (Cai et al., 2015), despite also contrasting results (Redlich et al., 2014), reduced volume in hippocampus and amygdala was also previously detected in BD patients compared to MDD (Cao et al., 2017; Redlich et al., 2014). Notably, our results are in contrast with a recent meta-analyses which pointed out smaller GM volumes in MDD compared to BD in the hippocampus, fusiform gyrus, parahippocampal gyrus, cerebellar vermis and inferior parietal lobule, whereas shared reduction in the two disorders compared to HC were detected in anterior cingulate cortex and insula. However, relevant methodological issues could contribute in explaining contrasting results. Despite meta-regressions pointed out no significant effect of mood state, pharmacological treatments in BD and MDD, authors cannot completely rule out possible confounding effects of these variables. MDD and BD samples were different for these variables (e.g. 94% of MDD and only 28% for BD experienced depression) and several original research papers, included in the meta-analysis, do not provide information in order to effectively meta-regress their effect over the entire sample. In our study, we included only depressed patients for both MDD and BD and medication load, number of previous episodes and effect of sex and age were directly regressed out from the analyses. This is particularly relevant by considering the neuroprotective effect of lithium, and drugs used in treating depression (http://www.nbn2.org/) in counteracting volume loss in BD (Hafeman et al., 2012; Hajek et al., 2012). Finally, studies directly comparing the two conditions may provide more accurate results that those based on effect size (Wise et al., 2017).

DTI analyses revealed a significant effect of diagnosis on FA, AD, RD and MD in several WM tracts encompassing forceps minor and corpus callosum, cingulum bundle, uncinate fasciculus, anterior thalamic radiation, inferior and superior longitudinal fasciculus, inferior fronto-occipital fasciculus, and corticospinal tract. DTI indirectly provides measures of the cellular density, integrity, and directionality of WM tracts. Compared to HC, depressed patients, both MDD and BD, presented lower FA and higher RD, in most of WM tracts. Lower FA may reflect a widespread reduced structural integrity, regional myelination levels, axonal density and diameter, while higher RD may reflect possible alterations in the myelin integrity as demyelination or dysmyelination processes in BD and MDD compared to HC. Notably, MDD also showed a widespread higher AD and MD compared to HC, not highlighted in BD. AD is associated to fiber diameter or organization: lower values may reflect axonal damage or loss or fragmentation specific for MDD patients and emerging as specific biomarker for the disorder.

The detected widespread effect encompasses WM tracts that exert a crucial role in mood regulation and cognition connecting the lateral and medial areas of the bilateral frontal lobes, temporal cortex and limbic regions, such as amygdala and hippocampus. WM abnormalities in these tracts have been extensively associated to mood disorder (Ambrosi et al., 2013; Benedetti et al., 2011b; Canales-Rodriguez et al., 2014; Emsell et al., 2013; Helm et al., 2018; Jiang et al., 2017; Maller et al., 2014; Vai et al., 2019; Zanetti et al., 2009), and related symptomatology such suicidality (Matsuo et al., 2010). These results support that the widespread reduced WM integrity and structural connectivity may identify biomarkers for mood disorders: WM tracts exert a crucial role in connecting brain regions deeply involved in defining emotional experience and mood regulation. By directly comparing the two disorders, MDD had widespread lower FA, lower AD and higher RD in uncinate, inferior fronto-occipital, inferior and superior longitudinal fasciculus, forceps minor, anterior thalamic radiation, and cingulum bundle. These results are line with our machine learning outcome, which highlighted FA and AD as most relevant DTI measures in order to differentiate the two disorders.

From a general perspective, our results suggest a wider disruption of WM microstructure in MDD patients, which encompass cortical and subcortical tracts in frontal, parietal, ventral, temporal, and occipital lobes compared to BD. Our results are in contrast with previous evidence: a recent meta-analysis which showed a reduction in FA in the left posterior cingulum in BD compared to MDD (Wise et al., 2016), in line with single

studies which reported wider disrupted WM microstructure in BD (Benedetti et al., 2011a; Matsuoka et al., 2017; Repple et al., 2017). Despite our analyses were corrected for medication load, which specifically weights lithium assumption, we cannot completely rule out non-linear neurotropic effect of lithium on WM tracts in BD patients. Lithium has been related to neuroprotective and neurotrophic effects, which counteract the detrimental influences of BD on WM structure (Benedetti et al., 2013; Gildengers et al., 2015). Future studies in drug naïve patients should be performed. Finally, a word of caution is needed in the interpretation of the neurobiological basis of DTI measures: crossing fibers or different pathological processes (e.g. demyelination axon injury) may lead to similar pattern of DTI measured and may also change during course of illness. These limitations, however, do not bias the main finding of an effect on WM microstructure.

The lack of an independent validation set (i.e. external validation) identifies a limit, not allowing to control for possible biases associated with the specific samples. Nevertheless, the nested cross-validation scheme has been suggested to provide an accurate insight of the model generalization performance (Refaeilzadeh et al., 2009). Furthermore, patients were not drug-naive, and the drug treatments administered during life-time could have influenced MRI measures. The recruitment took place in a single center and in a single ethnic group, thus raising the possibility of population stratifications limiting the generalizability of the findings. The sample size, despite higher than previous studies can be considered relatively small, also by considering that a multimodal neuroimaging approach have been applied. In the next years, the algorithm will be refined including an independent sample of multicenter neuroimaging acquisitions, further increasing the sample size, and variability. A further experimental step can be to test or to refine the algorithm in larger datasets as collected in worldwide consortium, such as Enhancing Neuro Imaging Genetics Through Meta Analysis Consortium, ENIGMA. This will provide highly realistic and useful diagnostic functions (Schrouff et al., 2013). Furthermore, despite we achieved a significant accuracy higher than 70%, and that MDD patients with BD familiarity were excluded, we cannot completely rule out a possible initial mislabelling for MDD.

Moreover, by considering that our BD subjects were already diagnosed at the time of the scanning, in order to assess the effective prognostic ability of the algorithm, it should be further longitudinally validated in MDD patients clinically followed until their switch into BD. Finally, effects of covariates of no interest were directly removed in the kernel space, rather than in feature one, using a residual forming matrix. This is a more

computationally efficient approach compared to removing confounding variables at voxel level, however, this does not accounts for possible multivariate effect of confounds.

Future studies can include other kind of features: resting state, corticolimbic functional connectivity and other biological data, such as inflammatory markers and genes, that have been previously proposed as candidate biomarkers for the differential diagnosis in mood disorders (Bai et al., 2015; Menezes et al., 2018; Vai et al., 2019) and may remarkably improve the accuracy. This is particularly relevant for the genetic vulnerability: genetic influence explain 60–85% of risk of BD (Barnett and Smoller, 2009; McGuffin et al., 2003), despite it is partially shared with MDD, which notably has a lower heritability (h^2=31-50 % (Hamet and Tremblay, 2005; Jansen et al., 2016)), ~71% of the genetic variance of mania is not shared among the disorders (Barnett and Smoller, 2009; McGuffin et al., 2003). Several Single Nucleotide Polymorphisms, identified by the Psychiatric Genomics Consortium (see http://pgc.unc.edu) in genome-wide association studies (GWAS) for MDD and BD, can identify an interesting feature for machine learning procedures by also considering the new cost-effective high-density microarray procedures.

Genetic, immune markers and neuroimaging are key biological tools in the precision medicine that could define new algorithms to reliably predict BD diagnosis in depressed patients.

To the best of our knowledge this is the first study combining GM and diffusion tensor imaging in predicting the differential diagnosis between depressed BD and MDD patients. Our results prompt the quantitative biomarkers and multiple kernel learning routine as promising tool for personalized treatment in mood disorder, strengthening the impact of neurobiological research into clinical practice.

References

Ambrosi, E., Rossi-Espagnet, M.C., Kotzalidis, G.D., Comparelli, A., Del Casale, A., Carducci, F., Romano, A., Manfredi, G., Tatarelli, R., Bozzao, A., Girardi, P., 2013. Structural brain alterations in bipolar disorder II: a combined voxel-based morphometry (VBM) and diffusion tensor imaging (DTI) study. J Affect Disord 150, 610-615.

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.

Bai, Y.M., Su, T.P., Li, C.T., Tsai, S.J., Chen, M.H., Tu, P.C., Chiou, W.F., 2015. Comparison of pro-inflammatory cytokines among patients with bipolar disorder and unipolar depression and normal controls. Bipolar Disord 17, 269-277.

Barnett, J.H., Smoller, J.W., 2009. The genetics of bipolar disorder. Neuroscience 164, 331-343.

Behrens, T.E., Woolrich, M.W., Jenkinson, M., Johansen-Berg, H., Nunes, R.G., Clare, S., Matthews, P.M., Brady, J.M., Smith, S.M., 2003. Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn. Reson. Med. 50, 1077-1088.

Benedetti, F., Absinta, M., Rocca, M.A., Radaelli, D., Poletti, S., Bernasconi, A., Dallaspezia, S., Pagani, E., Falini, A., Copetti, M., Colombo, C., Comi, G., Smeraldi, E., Filippi, M., 2011a. Tract-specific white matter structural disruption in patients with bipolar disorder. Bipolar Disord 13, 414-424.

Benedetti, F., Bollettini, I., Barberi, I., Radaelli, D., Poletti, S., Locatelli, C., Pirovano, A., Lorenzi, C., Falini, A., Colombo, C., 2013. Lithium and GSK3-β promoter gene variants influence white matter microstructure in bipolar disorder. Neuropsychopharmacology 38, 313.

Benedetti, F., Yeh, P.H., Bellani, M., Radaelli, D., Nicoletti, M.A., Poletti, S., Falini, A., Dallaspezia, S., Colombo, C., Scotti, G., Smeraldi, E., Soares, J.C., Brambilla, P., 2011b. Disruption of white matter integrity in bipolar depression as a possible structural marker of illness. Biol Psychiatry 69, 309-317.

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.

Canales-Rodriguez, E.J., Pomarol-Clotet, E., Radua, J., Sarro, S., Alonso-Lana, S., Del Mar Bonnin, C., Goikolea, J.M., Maristany, T., Garcia-Alvarez, R., Vieta, E., McKenna, P., Salvador, R., 2014. Structural abnormalities in bipolar euthymia: a multicontrast molecular diffusion imaging study. Biol Psychiatry 76, 239-248.

Cao, B., Passos, I.C., Mwangi, B., Amaral-Silva, H., Tannous, J., Wu, M.J., Zunta-Soares, G.B., Soares, J.C., 2017. Hippocampal subfield volumes in mood disorders. Mol Psychiatry 22, 1352-1358.

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., Martínez-Ramón, M., Pearlson, G., Sui, J., Calhoun, V.D., 2011. Characterization of groups using composite kernels and multi-source fMRI analysis data: application to schizophrenia. Neuroimage 58, 526-536.

Chu, C., Ni, Y., Tan, G., Saunders, C.J., Ashburner, J., 2011. Kernel regression for fMRI pattern prediction. NeuroImage 56, 662-673.

Deng, F., Wang, Y., Huang, H., Niu, M., Zhong, S., Zhao, L., Qi, Z., Wu, X., Sun, Y., Niu, C., 2018. Abnormal segments of right uncinate fasciculus and left anterior thalamic radiation in major and bipolar depression. Progress in Neuro-Psychopharmacology and Biological Psychiatry 81, 340-349.

Donini, M., Monteiro, J.M., Pontil, M., Hahn, T., Fallgatter, A.J., Shawe-Taylor, J., Mourao-Miranda, J., Initiative, A.s.D.N., 2019. Combining heterogeneous data sources for neuroimaging based diagnosis: re-weighting and selecting what is important. NeuroImage 195, 215-231.

Drevets, W.C., Price, J.L., Furey, M.L., 2008. Brain structural and functional abnormalities in mood disorders: implications for neurocircuitry models of depression. Brain Struct Funct 213, 93-118.
Elisseeff, A., Pontil, M., 2003. Leave-one-out error and stability of learning algorithms with applications. NATO science series sub series iii computer and systems sciences 190, 111-130.
Emsell, L., Leemans, A., Langan, C., Van Hecke, W., Barker, G.J., McCarthy, P., Jeurissen, B., Sijbers, J., Sunaert, S., Cannon, D.M., McDonald, C., 2013. Limbic and callosal white matter changes in euthymic bipolar I disorder: an advanced diffusion magnetic resonance imaging tractography study. Biol Psychiatry 73, 194-201.
Fonseka, T.M., MacQueen, G.M., Kennedy, S.H., 2018. Neuroimaging biomarkers as predictors of treatment outcome in Major Depressive Disorder. Journal of Affective Disorders 233, 21-35.
Frodl, T., 2017. Recent advances in predicting responses to antidepressant treatment. F1000Research 6.
Fung, G., Deng, Y., Zhao, Q., Li, Z., Qu, M., Li, K., Zeng, Y.-w., Jin, Z., Ma, Y.-t., Yu, X., 2015. Distinguishing bipolar and major depressive disorders by brain structural morphometry: a pilot study. BMC psychiatry 15, 298.
Gao, S., Calhoun, V.D., Sui, J., 2018. Machine learning in major depression: From classification to treatment outcome prediction. CNS Neurosci Ther 24, 1037-1052.
Gaser, C., Dahnke, R., CAT-a computational anatomy toolbox for the analysis of structural MRI data.
Gildengers, A.G., Butters, M.A., Aizenstein, H.J., Marron, M.M., Emanuel, J., Anderson, S.J., Weissfeld, L.A., Becker, J.T., Lopez, O.L., Mulsant, B.H., 2015. Longer lithium exposure is associated with better white matter integrity in older adults with bipolar disorder. Bipolar disorders 17, 248-256.
Godlewska, B.R., Browning, M., Norbury, R., Igoumenou, A., Cowen, P.J., Harmer, C.J., 2018. Predicting Treatment Response in Depression: The Role of Anterior Cingulate Cortex. International Journal of Neuropsychopharmacology 21, 988-996.
Goodwin, G.M., 2012. Bipolar depression and treatment with antidepressants. The British Journal of Psychiatry 200, 5-6.
Hafeman, D.M., Chang, K.K.D., Garrett, A.S., Sanders, E.M., Phillips, M.L., 2012. Effects of medication on neuroimaging findings in bipolar disorder: an updated review. Bipolar Disorders 14, 375-410.
Hajek, T., Kopecek, M., Hoschl, C., Alda, M., 2012. Smaller hippocampal volumes in patients with bipolar disorder are masked by exposure to lithium: a meta-analysis. Journal of Psychiatry & Neuroscience 37, 333-343.
Hamet, P., Tremblay, J., 2005. Genetics and genomics of depression. Metabolism 54, 10-15.
Hamilton, M., 1960. A rating scale for depression. J. Neurol. Neurosurg. Psychiatry 23, 56-62.
Helm, K., Viol, K., Weiger, T.M., Tass, P.A., Grefkes, C., del Monte, D., Schiepek, G., 2018. Neuronal connectivity in major depressive disorder: a systematic review. Neuropsychiatric Disease and Treatment 14, 2715-2737.
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., 2018. Cortical abnormalities in bipolar disorder: an MRI analysis of 6503 individuals from the ENIGMA Bipolar Disorder Working Group. Mol Psychiatry 23, 932-942.
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.
Jansen, R., Penninx, B.W.J.H., Madar, V., Xia, K., Milaneschi, Y., Hottenga, J.J., Hammerschlag, A.R., Beekman, A., van der Wee, N., Smit, J.H., Brooks, A.I., Tischfield, J., Posthuma, D., Schoevers, R., van Grootheest, G., Willemsen, G., de Geus, E.J., Boomsma, D.I., Wright, F.A., Zou, F., Sun, W., Sullivan, P.F., 2016. Gene expression in major depressive disorder. Mol Psychiatr 21, 339-347.
Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved optimization for the robust and accurate linear registration and motion correction of brain images. Neuroimage 17, 825-841.
Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. Med Image Anal 5, 143-156.
Jiang, J., Zhao, Y.-J., Hu, X.-Y., Du, M.-Y., Chen, Z.-Q., Wu, M., Li, K.-M., Zhu, H.-Y., Kumar, P., Gong, Q.-Y., 2017. Microstructural brain abnormalities in medication-free patients with major depressive disorder: a systematic review and meta-analysis of diffusion tensor imaging. Journal of psychiatry & neuroscience: JPN 42, 150.
Jie, N.-F., Zhu, M.-H., Ma, X.-Y., Osuch, E.A., Wammes, M., Théberge, J., Li, H.-D., Zhang, Y., Jiang, T.-Z., Sui, J., 2015. Discriminating bipolar disorder from major depression based on SVM-FoBa: efficient feature selection with multimodal brain imaging data. IEEE transactions on autonomous mental development 7, 320-331.
Judd, L.L., Akiskal, H.S., Schettler, P.J., Coryell, W., Endicott, J., Maser, J.D., Solomon, D.A., Leon, A.C., Keller, M.B., 2003. A prospective investigation of the natural history of the long-term weekly symptomatic status of bipolar II disorder. Arch Gen Psychiatry 60, 261-269.
Kupka, R.W., Altshuler, L.L., Nolen, W.A., Suppes, T., Luckenbaugh, D.A., Leverich, G.S., Frye, M.A., Keck, P.E., Jr., McElroy, S.L., Grunze, H., Post, R.M., 2007. Three times more days depressed than manic or hypomanic in both bipolar I and bipolar II disorder. Bipolar Disord. 9, 531-535.
Lanckriet, G.R., Cristianini, N., Bartlett, P., Ghaoui, L.E., Jordan, M.I., 2004. Learning the kernel matrix with semidefinite programming. Journal of Machine learning research 5, 27-72.
MacMaster, F.P., Carrey, N., Langevin, L.M., Jaworska, N., Crawford, S., 2014. Disorder-specific volumetric brain difference in adolescent major depressive disorder and bipolar depression. Brain imaging and behavior 8, 119-127.
Maller, J.J., Thaveenthiran, P., Thomson, R.H., McQueen, S., Fitzgerald, P.B., 2014. Volumetric, cortical thickness and white matter integrity alterations in bipolar disorder type I and II. J Affect Disord 169, 118-127.
Matsuo, K., Harada, K., Fujita, Y., Okamoto, Y., Ota, M., Narita, H., Mwangi, B., Gutierrez, C.A., Okada, G., Takamura, M., Yamagata, H., Kusumi, I., Kunugi, H., Inoue, T., Soares, J.C., Yamawaki, S., Watanabe, Y., 2019. Distinctive Neuroanatomical Substrates for Depression in Bipolar Disorder versus Major Depressive Disorder. Cereb Cortex 29, 202-214.
Matsuo, K., Nielsen, N., Nicoletti, M.A., Hatch, J.P., Monkul, E.S., Watanabe, Y., Zunta-Soares, G.B., Nery, F.G., Soares, J.C., 2010. Anterior genu corpus callosum and impulsivity in suicidal patients with bipolar disorder. Neuroscience Letters 469, 75-80.
Matsuoka, K., Yasuno, F., Kishimoto, T., Yamamoto, A., Kiuchi, K., Kosaka, J., Nagatsuka, K., Iida, H., Kudo, T., 2017. Microstructural Differences in the Corpus Callosum in Patients With Bipolar Disorder and Major Depressive Disorder. J Clin Psychiatry 78, 99-104.

McDonald, C., 2015. Brain Structural Effects of Psychopharmacological Treatment in Bipolar Disorder. Curr Neuropharmacol 13, 445-457.
McGuffin, P., Rijsdijk, F., Andrew, M., Sham, P., Katz, R., Cardno, A., 2003. The heritability of bipolar affective disorder and the genetic relationship to unipolar depression. Arch Gen Psychiatry 60, 497-502.
Menezes, I.C., von Werne Baes, C., Lacchini, R., Juruena, M.F., 2018. Genetic biomarkers for differential diagnosis of major depressive disorder and bipolar disorder: A systematic and critical review. Behav Brain Res.
Mourão- Miranda, J., Almeida, J.R., Hassel, S., de Oliveira, L., Versace, A., Marquand, A.F., Sato, J.R., Brammer, M., Phillips, M.L., 2012. Pattern recognition analyses of brain activation elicited by happy and neutral faces in unipolar and bipolar depression. Bipolar disorders 14, 451-460.
Nichols, T.E., Holmes, A.P., 2002. Nonparametric permutation tests for functional neuroimaging: a primer with examples. Hum. Brain Mapp. 15, 1-25.
Niida, R., Yamagata, B., Matsuda, H., Niida, A., Uechi, A., Kito, S., Mimura, M., 2019. Regional brain volume reductions in major depressive disorder and bipolar disorder: An analysis by voxel-based morphometry. Int J Geriatr Psychiatry 34, 186-192.
Rakotomamonjy, A., 2008. SimpleMKL Toolbox. http://asi. insarouen. fr/enseignants/rakotom/code/mklindex. html.
Redlich, R., Almeida, J.R., 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.
Refaeilzadeh, P., Tang, L., Liu, H., 2009. Cross-validation. Encyclopedia of database systems, 532-538.
Repple, J., Meinert, S., Grotegerd, D., Kugel, H., Redlich, R., Dohm, K., Zaremba, D., Opel, N., Buerger, C., Forster, K., Nick, T., Arolt, V., Heindel, W., Deppe, M., Dannlowski, U., 2017. A voxel-based diffusion tensor imaging study in unipolar and bipolar depression. Bipolar Disord 19, 23-31.
Rive, M.M., Redlich, R., Schmaal, L., Marquand, A.F., Dannlowski, U., Grotegerd, D., Veltman, D.J., Schene, A.H., Ruhé, H.G., 2016. Distinguishing medication- free subjects with unipolar disorder from subjects with bipolar disorder: state matters. Bipolar disorders 18, 612-623.
Rubin-Falcone, H., Zanderigo, F., Thapa-Chhetry, B., Lan, M., Miller, J.M., Sublette, M.E., Oquendo, M.A., Hellerstein, D.J., McGrath, P.J., Stewart, J.W., Mann, J.J., 2018. Pattern recognition of magnetic resonance imaging-based gray matter volume measurements classifies bipolar disorder and major depressive disorder. J Affect Disord 227, 498-505.
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. Journal of psychiatric research 68, 91-98.
Sackeim, H.A., 2001. The definition and meaning of treatment-resistant depression. Journal of Clinical Psychiatry.
Schmaal, L., Hibar, D.P., Samann, P.G., Hall, G.B., Baune, B.T., Jahanshad, N., Cheung, J.W., van Erp, T.G.M., Bos, D., Ikram, M.A., Vernooij, M.W., Niessen, W.J., Tiemeier, H., Hofman, A., Wittfeld, K., Grabe, H.J., Janowitz, D., Bulow, R., Selonke, M., Volzke, H., Grotegerd, D., Dannlowski, U., Arolt, V., Opel, N., Heindel, W., Kugel, H., Hoehn, D., Czisch, M., Couvy-Duchesne, B., Renteria, M.E., Strike, L.T., Wright, M.J., Mills, N.T., de Zubicaray, G.I., McMahon, K.L., Medland, S.E., Martin, N.G., Gillespie, N.A., Goya-Maldonado, R., Gruber, O., Kramer, B., Hatton, S.N., Lagopoulos, J., Hickie, I.B., Frodl, T., Carballedo, A., Frey, E.M., van Velzen, L.S., Penninx, B., van Tol, M.J., van der Wee, N.J., Davey, C.G., Harrison, B.J., Mwangi, B., Cao, B., Soares, J.C., Veer, I.M., Walter, H., Schoepf, D., Zurowski, B., Konrad, C., Schramm, E., Normann, C., Schnell, K., Sacchet, M.D., Gotlib, I.H., MacQueen, G.M., Godlewska, B.R., Nickson, T., McIntosh, A.M., Papmeyer, M., Whalley, H.C., Hall, J., Sussmann, J.E., Li, M., Walter, M., Aftanas, L., Brack, I., Bokhan, N.A., Thompson, P.M., Veltman, D.J., 2017. Cortical abnormalities in adults

and adolescents with major depression based on brain scans from 20 cohorts worldwide in the ENIGMA Major Depressive Disorder Working Group. Mol Psychiatry 22, 900-909.
Schmierer, K., Wheeler-Kingshott, C.A., Boulby, P.A., Scaravilli, F., Altmann, D.R., Barker, G.J., Tofts, P.S., Miller, D.H., 2007. Diffusion tensor imaging of post mortem multiple sclerosis brain. Neuroimage 35, 467-477.
Schrouff, J., Monteiro, J.M., Portugal, L., Rosa, M.J., Phillips, C., Mourao-Miranda, J., 2018. Embedding Anatomical or Functional Knowledge in Whole-Brain Multiple Kernel Learning Models. Neuroinformatics 16, 117-143.
Schrouff, J., Mourao-Miranda, J., Phillips, C., Parvizi, J., 2016. Decoding intracranial EEG data with multiple kernel learning method. J Neurosci Methods 261, 19-28.
Schrouff, J., Rosa, M.J., Rondina, J.M., Marquand, A.F., Chu, C., Ashburner, J., Phillips, C., Richiardi, J., Mourao-Miranda, J., 2013. PRoNTo: pattern recognition for neuroimaging toolbox. Neuroinformatics 11, 319-337.
Shawe-Taylor, J., Cristianini, N., 2004. Kernel methods for pattern analysis. Cambridge university press.
Smith, S.M., 2002. Fast robust automated brain extraction. Hum Brain Mapp 17, 143-155.
Smith, S.M., Jenkinson, M., Johansen-Berg, H., Rueckert, D., Nichols, T.E., Mackay, C.E., Watkins, K.E., Ciccarelli, O., Cader, M.Z., Matthews, P.M., Behrens, T.E., 2006. Tract-based spatial statistics: voxelwise analysis of multi-subject diffusion data. Neuroimage 31, 1487-1505.
Smith, S.M., Jenkinson, M., Woolrich, M.W., Beckmann, C.F., Behrens, T.E., Johansen-Berg, H., Bannister, P.R., De Luca, M., Drobnjak, I., Flitney, D.E., 2004. Advances in functional and structural MR image analysis and implementation as FSL. Neuroimage 23, S208-S219.
Smith, S.M., Nichols, T.E., 2009. Threshold-free cluster enhancement: addressing problems of smoothing, threshold dependence and localisation in cluster inference. Neuroimage 44, 83-98.
Song, S.K., Sun, S.W., Ramsbottom, M.J., Chang, C., Russell, J., Cross, A.H., 2002. Dysmyelination revealed through MRI as increased radial (but unchanged axial) diffusion of water. Neuroimage 17, 1429-1436.
Squarcina, L., Castellani, U., Bellani, M., Perlini, C., Lasalvia, A., Dusi, N., Bonetto, C., Cristofalo, D., Tosato, S., Rambaldelli, G., 2017. Classification of first-episode psychosis in a large cohort of patients using support vector machine and multiple kernel learning techniques. NeuroImage 145, 238-245.
Vai, B., Bertocchi, C., Benedetti, F., 2019. Cortico-limbic connectivity as a possible biomarker for bipolar disorder: where are we now? Expert Rev Neurother 19, 159-172.
Varoquaux, G., 2018. Cross-validation failure: small sample sizes lead to large error bars. Neuroimage 180, 68-77.
Wise, T., Radua, J., Nortje, G., Cleare, A.J., Young, A.H., Arnone, D., 2016. Voxel-Based Meta-Analytical Evidence of Structural Disconnectivity in Major Depression and Bipolar Disorder. Biol Psychiatry 79, 293-302.
Wise, T., Radua, J., Via, E., Cardoner, N., Abe, O., Adams, T.M., Amico, F., Cheng, Y., Cole, J.H., Perico, C.D.M., Dickstein, D.P., Farrow, T.F.D., Frodl, T., Wagner, G., Gotlib, I.H., Gruber, O., Ham, B.J., Job, D.E., Kempton, M.J., Kim, M.J., Koolschijn, P.C.M.P., 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.A., Veltman, D.J., Young, A.H., Fu, C.H., Cleare, A.J., Arnone, D., 2017. Common and distinct patterns of grey-matter volume alteration in major depression and bipolar disorder: evidence from voxel-based meta-analysis. Molecular Psychiatry 22, 1455-1463.
Wittchen, H.U., 2012. The burden of mood disorders. Science 338, 15.
Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S.M., 2009. Bayesian analysis of neuroimaging data in FSL. Neuroimage 45, S173-S186.

Zanetti, M.V., Jackowski, M.P., Versace, A., Almeida, J.R., Hassel, S., Duran, F.L., Busatto, G.F., Kupfer, D.J., Phillips, M.L., 2009. State-dependent microstructural white matter changes in bipolar I depression. Eur. Arch. Psychiatry Clin. Neurosci. 259, 316-328.

Figures

Figure 1

Mass-Univariate analyses for Grey Matter volume.

Main effect of Diagnosis in anterior cingulate cortex, Rolandic operculum/insula, Cerebellar Vermis and Fusiform gyrus/Hippocampus. Whole brain analyses corrected at cluster $p_{FWE}<0.05$. Whiskers indicate 90% confidence intervals. BD, bipolar disorder, MDD, major depressive disorder, HC, Healthy controls.

Figure 2

Mass-Univariate analyses for Diffusion Tensor Imaging.

White matter tracts where diagnosis had significant main effects on (A) Fractional Anisotropy, (B) Radial Diffusivity, (C) Mean Diffusivity and (D) Mean Diffusivity. Data showed are extracted at peak level. Analyses were corrected for multiple comparisons ($p_{FWE}<0.05$) in a minimum cluster size of k=100 using Threshold-Free Cluster Enhancement (TFCE).

Figure 3

Multiple kernel learning classification.

A) Smoothed density version of histogram plot of function values; B) Prediction plots per fold, decision threshold is displayed by a vertical line; C) Receiver Operator Curve, Areas Under the Curve=0.79.

Table 1. Clinical and demographic characteristics. The sample is divided by diagnosis (mean ± standard deviation and statistics).
* p<0.05; Abbreviations: BD, Bipolar Disorder; HC, healthy controls; HDRS-21, Hamilton Depression Rating Scale; MDD, major depressive disorder; F, Female; M, Males.

	MDD	BD	HC	T /F / χ^2	p	Significant Post-hoc
Sex	F53 M21	F55 M19	F39 M35	9.182	0.01*	F: HC < MD; HC < BD
Age	49.62±9.78	47.26±9.39	36.38±12.47	34.06	< 0.001*	BD > HC MDD> HC
Education - Years	12.82±4.2	11.5±4	16.08 ± 2.61	29.61	< 0.001*	HC > BD; HC>MDD
Duration of illness - years	13.61±9.78	17.35±10.12	/	2.24	0.03*	
HDRS-21	24.96±5.23	23.74±9.46	/	0.83	0.4	
Medication Load	4.11±1.86	4.2±2	/	0.298	0.766	
N. Episodes	4.45±4.12	9.82±9.63	/	4.418	<0.001*	

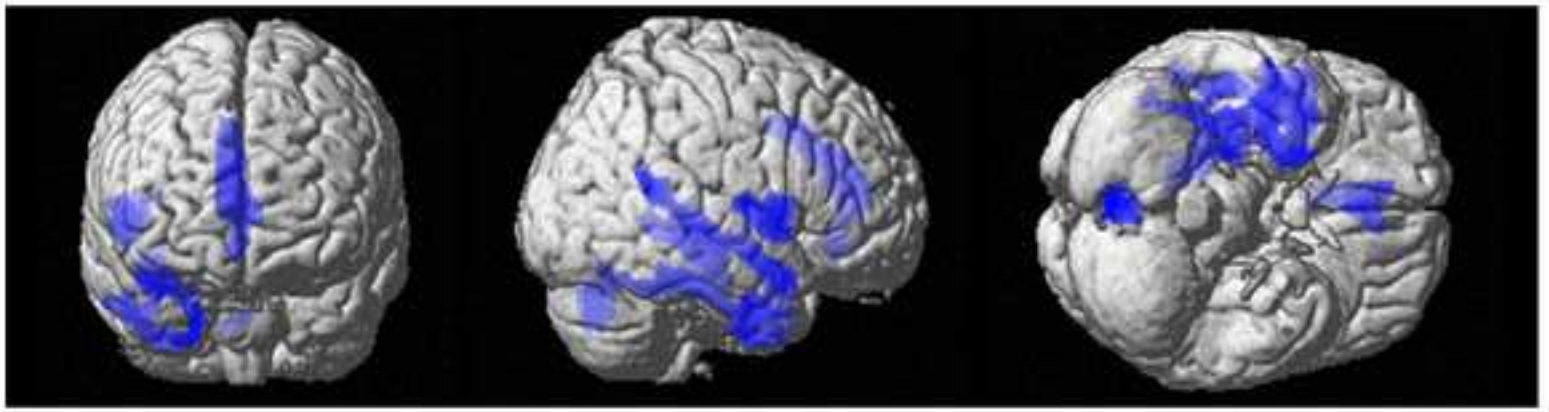
Table 2 Confusion matrix.

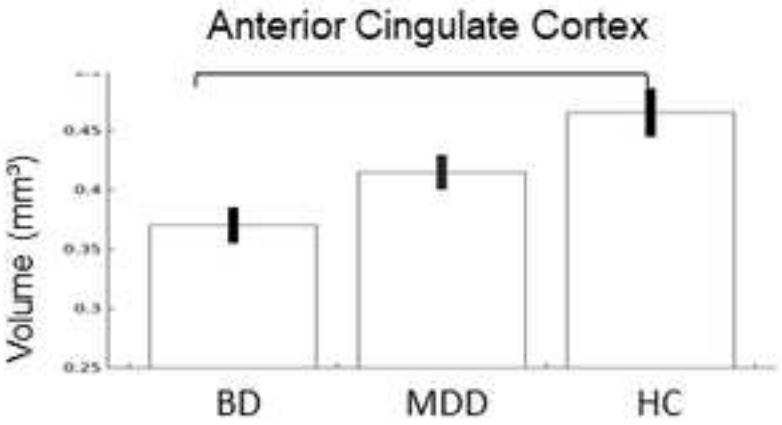
		PREDICTED LABEL	
		BD	MDD
TRUE LABEL	BD	55	20
	MDD	19	54

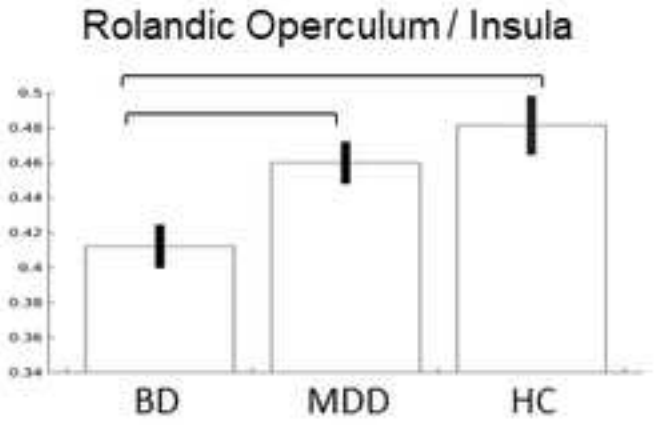
Multiple kernel learning predicted diagnosis from multimodal structural neuroimaging.

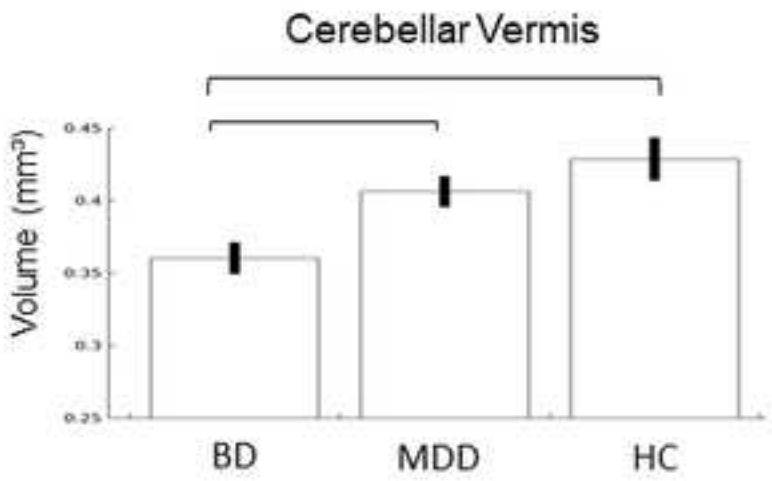
Abbreviations: BD, Bipolar Disorder; MDD, major depressive disorder.

Figure 1









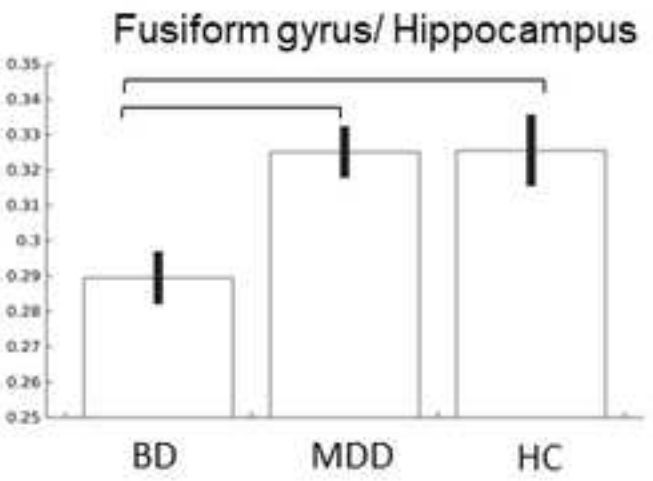


Figure 2

Click here to access/download;Figure(s);Figure_2.tif

Diapositiva2.TIF

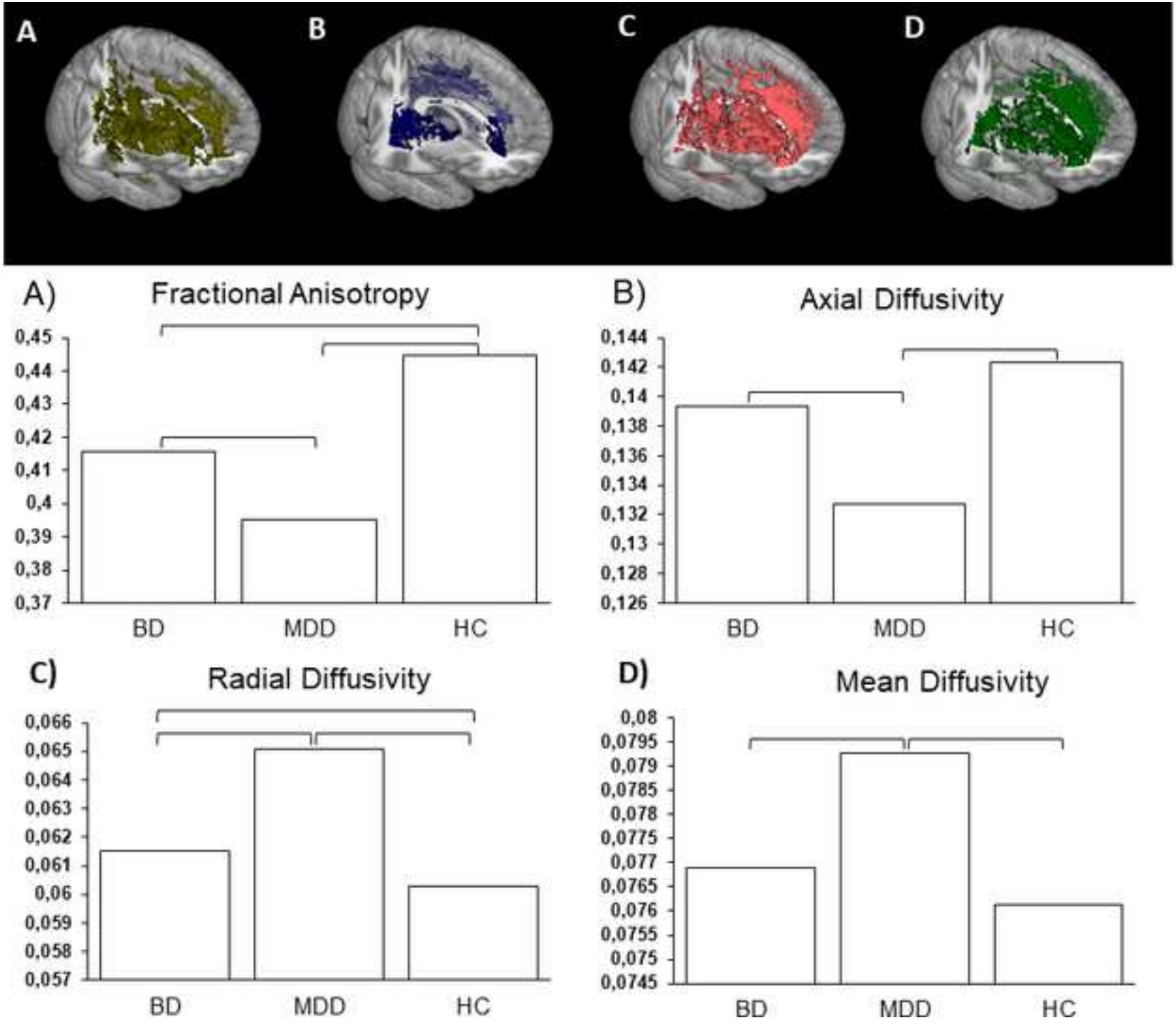
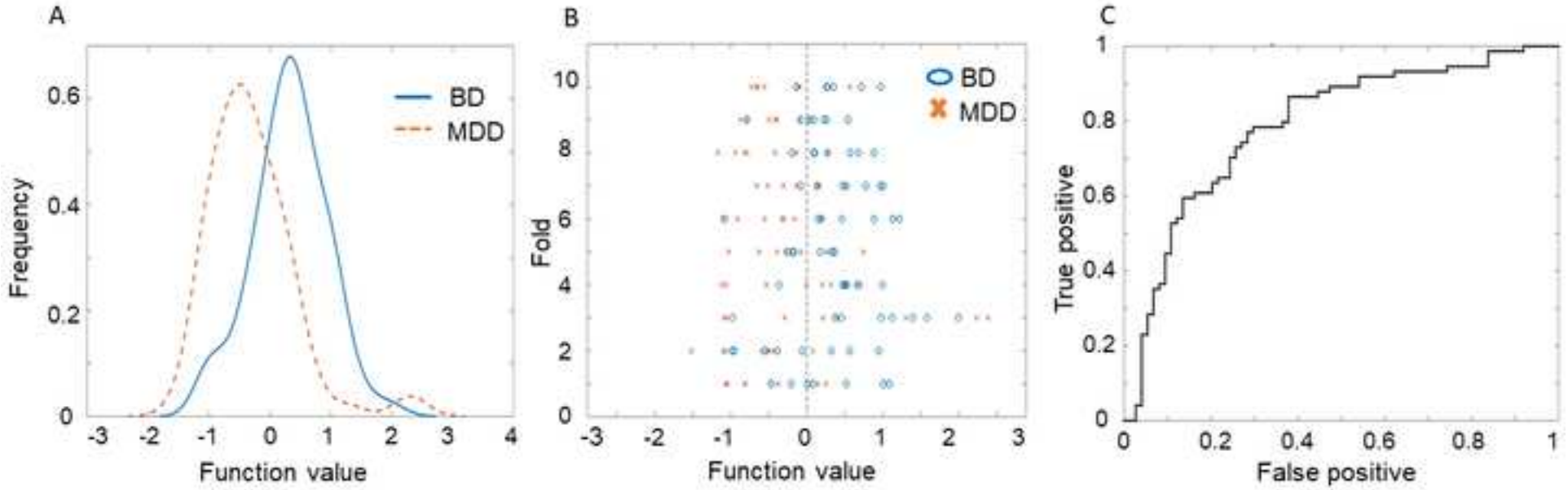


Figure 3



Role of Funding Source.

This work was supported by the Italian Ministry of Health (grants numbers RF-2011-02350980, RF-2011-02349921). BV's research activities are supported by Fondazione Umberto Veronesi. The Italian Ministry of Health and Fondazione Umberto Veronesi had no further role in study design; in the collection, analysis and interpretation of data; in the writing of the report; and in the decision to submit the paper for publication.

Contributors

Benedetta Vai and Francesco Benedetti designed the study and wrote the protocol. Benedetta Vai together with Lorenzo Parenti and Cristina Cara performed the machine learning analyses. Irene Bollettini, Poletti Sara, Elena Mazza, Elisa Melloni and Chiara Verga contributed in grey matter and white matter images preprocessing and mass-univariate group analyses. Cristina Colombo managed the subjects recruitment. All authors contributed to this paper and have approved the final manuscript

Conflict of Interest

All other authors declare that they have no conflicts of interest.

Acknowledgment

The authors thanks Silvio Conte for excellent technical help in MRI image acquisition.