


Diagnostic classification of unipolar depression based on resting-state functional connectivity MRI: effects of generalization to a diverse sample

Benedikt Sundermann¹  · Stephan Feder¹ · Heike Wersching² · Anja Teuber² · Wolfram Schwindt¹ · Harald Kugel¹ · Walter Heindel¹ · Volker Arolt³ · Klaus Berger² · Bettina Pfeiderer¹

Received: 14 November 2016 / Accepted: 23 December 2016 / Published online: 31 December 2016
© Springer-Verlag Wien 2016

Abstract In small, selected samples, an approach combining resting-state functional connectivity MRI and multivariate pattern analysis has been able to successfully classify patients diagnosed with unipolar depression. Purposes of this investigation were to assess the generalizability of this approach to a large clinically more realistic sample and secondarily to assess the replicability of previously reported methodological feasibility in a more homogeneous subgroup with pronounced depressive symptoms. Two independent subsets were drawn from the depression and control cohorts of the BiDirect study, each with 180 patients with and 180 controls without depression. Functional connectivity either among regions covering the gray matter or selected regions with known alterations in depression was assessed by resting-state fMRI. Support vector machines with and without automated feature selection were used to train classifiers differentiating between individual patients and controls in the entire first subset as well as in the subgroup. Model parameters were explored systematically. The second independent subset was used for validation of successful

models. Classification accuracies in the large, heterogeneous sample ranged from 45.0 to 56.1% (chance level 50.0%). In the subgroup with higher depression severity, three out of 90 models performed significantly above chance (60.8–61.7% at independent validation). In conclusion, common classification methods previously successful in small homogeneous depression samples do not immediately translate to a more realistic population. Future research to develop diagnostic classification approaches in depression should focus on more specific clinical questions and consider heterogeneity, including symptom severity as an important factor.

Keywords fMRI · Functional connectivity · Classification · MVPA · Depression

Introduction

There are currently ambitious but still preclinical efforts to complement the diagnostic spectrum for individual patients with unipolar depression, particularly the clinical entity major depressive disorder (MDD), by imaging biomarkers. These neurobiological markers are aimed at amending therapeutic decision making by capturing biological information which cannot be assessed by clinical interviews (Atluri et al. 2013; Phillips et al. 2015; Schneider and Prvulovic 2013). Particularly, diagnostic models based on multivariate pattern analyses (MVPA) of functional magnetic resonance imaging (fMRI) have been proposed as potentially powerful clinical tools for mental disorders (Arbabshirani et al. 2017; Castellanos et al. 2013; Haller et al. 2014; Klöppel et al. 2012; Lui et al. 2016; Orru et al. 2012; Patel et al. 2016; Sundermann et al. 2014a; Wolfers et al. 2015).

Electronic supplementary material The online version of this article (doi:10.1007/s00702-016-1673-8) contains supplementary material, which is available to authorized users.

✉ Benedikt Sundermann
benedikt.sundermann@uni-muenster.de

¹ Department of Clinical Radiology, University Hospital Münster, Albert-Schweitzer-Campus 1, Gebäude A1, 48149 Münster, Germany

² Institute of Epidemiology and Social Medicine, University of Münster, Münster, Germany

³ Department of Psychiatry, University Hospital Münster, Münster, Germany

FMRI data acquisition at rest has attracted particular interest of clinical researchers, because it seems to promise robust information on functionally relevant neural networks with simple setups and little demand for patient cooperation (Barkhof et al. 2014; Castellanos et al. 2013; Smith et al. 2009; Sundermann et al. 2014a; Zhang and Raichle 2010). Most approaches to analyze data obtained by this so-called resting-state fMRI (rs-fMRI) are based on analyses of functional connectivity (FC), the correlation of spontaneous activity in remote brain areas (Margulies et al. 2010; van den Heuvel and Hulshoff Pol 2010). Rs-fMRI has been used to characterize abnormal FC (rs-fcMRI) or spontaneous local activity in MDD. Major findings are increased spontaneous activity in cortical midline structures related to self-referential cognition and altered interactions of these regions with lateral cortical areas (Hamilton et al. 2015; Kaiser et al. 2015; Sundermann et al. 2014b). They have been interpreted as a correlate of a reduced top-down inhibition of cortical midline and limbic regions reflecting increased ruminative brooding (Hamilton et al. 2015; Marchetti et al. 2012; Nejad et al. 2013). Rs-fMRI, therefore, seems to provide information on important aspects of MDD etiopathogenesis and symptomatology (Kupfer et al. 2012).

MVPA conceptually overlaps with multivariate classification, pattern recognition, or predictive analyses. Here, mainly, instances of supervised learning (a subfield of machine learning) are summarized as MVPA. They facilitate the automated generation of decision rules based on previous experience, i.e., labeled training data (Alpaydin 2010; James et al. 2013). MVPA integrates information from multiple sources (for example, brain regions) with the aim to increase discriminative power compared to conventional univariate analyses of intrinsically noisy fMRI data. MVPA, particularly nonlinear techniques, additionally exploits complex relationships among individual features. Information captured by MVPA, therefore, exceeds and fundamentally differs from that assessed by standard univariate analyses (Arbabshirani et al. 2017; Pereira et al. 2009; Sundermann et al. 2014a). Popular and powerful classification tools for such analyses are support vector machines (SVM). In SVMs, subjects/imaging data sets are represented as points in a multidimensional feature space and the diagnostic problem can be operationalized as defining a hyperplane (i.e., a decision boundary) which best distinguishes between two groups of subjects. The classifier is trained using the ‘kernel trick’ by maximizing the margin of separation between groups based on the examples closest to this hyperplane (Alpaydin 2010; James et al. 2013; Orru et al. 2012; Pereira et al. 2009; Vapnik 2000). Different kernels and model parameters (such as C in SVMs) can be chosen to influence model complexity. These properties determine a model’s complexity and thus

its performance given a tradeoff between over- and underfitting. If a model is overfitted it perfectly conforms to the training samples while losing generalization ability for new samples. On the other hand, if a model is underfitted, it is too simple to capture the information essential for the diagnostic decision of interest. Consequently, parameters have to be optimized to generate successful diagnostic models that generalize to new clinical data (Alpaydin 2010; Arbabshirani et al. 2017; James et al. 2013; Orru et al. 2012; Vapnik 2000). The actual classification algorithm can be combined with different methods for dimensionality reduction and feature selection (FS) to prepare input data with the aim to improve diagnostic accuracy (Mwangi et al. 2014; Pereira et al. 2009; Sundermann et al. 2014a).

Groundbreaking work in the field of diagnostic MVPA in depression has been reported by Craddock et al. in 2009: The authors used FS and linear SVMs based on pairwise FC. They reached diagnostic accuracies of 83.33% (hold-out validation) (Craddock et al. 2009). Usually, MVPA models to differentiate patients from controls were trained on rs-fcMRI data of small, selected MDD samples and explicitly healthy subjects under ideal research conditions: these subjects were mostly young and had high levels of current depressive symptoms. However, other factors of clinical heterogeneity (e.g., antidepressant medication or any clinical comorbidity) were either excluded or not explicitly reported (Cao et al. 2014; Guo et al. 2014; Lord et al. 2012; Ma et al. 2013; Qin et al. 2015; Ramasubbu et al. 2016; Zeng et al. 2012; Zeng et al. 2014). One recent study reported successful diagnostic performance even for medication-free remitted MDD patients while even more strictly constraining sources of heterogeneity (Bhaumik et al. 2016). In contrast to that, Ramasubbu et al. observed above-change accuracies in a subgroup with the highest symptom severity only and still being in a range not deemed clinically meaningful (Ramasubbu et al. 2016). Another study included patients with schizophrenia as a second control group (Yu et al. 2013). For a detailed overview of sample characteristics and classification methods in previous studies on diagnostic MVPA of rs-fMRI data in depression, see Table 1.

Small samples as well as a high diversity of tested computational models in the field carry an inherent bias to a publication of false-positive results (Button et al. 2013). Diagnostic classification studies typically use cross validation (CV) to partially alleviate problems coming along with small samples. CV is an established method to assess the generalizability of classifiers to new data. It makes efficient use of data sets by repartitioning them into test- and training sets multiple times (Pereira et al. 2009). Nevertheless, rigorous confirmation of findings with independent data (Ioannidis 2005; Sundermann et al. 2014a)

Table 1 Sample characteristics, methods, and main classification results of the previous studies reporting diagnostic applications of rs-fcMRI and MVPA in unipolar depression

Author and year	Depressed patients ^a	Controls ^a	Methods	Results
Craddock et al. (2009) ^a	<p><i>n</i> = 20, 43.2 ± 10.8 years, 60% female</p> <p><i>Diagnosis</i> current major depressive episode (DSM-IV)</p> <p><i>Depression severity</i> HAM-D-17: 23.7 ± 1.6</p> <p><i>Depression duration</i> not reported</p> <p><i>Comorbidities</i> none reported, exclusion of comorbid psychiatric disorders</p> <p><i>Medication</i> none</p> <p><i>Structural MRI findings</i> not reported</p>	<p>Healthy subjects</p> <p><i>n</i> = 20, 28.9 ± 7.2 years, 60% female</p> <p><i>Definition</i> without neurological/psychiatric condition</p> <p><i>Comorbidities</i> none reported</p> <p><i>Medication</i> none</p> <p><i>Structural MRI findings</i> not reported</p>	<p><i>Features</i> functional connectivity of prespecified ROIs</p> <p><i>Feature selection</i> different methods</p> <p><i>Classifier</i> linear SVM</p>	<p><i>Overall accuracy</i> (CV) 62.5–95.0%</p> <p><i>Sensitivity</i> (CV) not reported</p> <p><i>Specificity</i> (CV) not reported</p> <p><i>Overall accuracy</i> (IV) 16.7–83.3%</p> <p><i>Sensitivity</i> (IV) not reported</p> <p><i>Specificity</i> (IV) not reported</p>
Lord et al. (2012)	<p><i>n</i> = 21, 37.9 ± 11.4 years, 38% female</p> <p><i>Diagnosis</i> acute episode of MDD (ICD-10)</p> <p><i>Depression severity</i> HAM-D-17: 15.8 ± 4.8</p> <p><i>Depression duration</i> mean disease onset approximately 6 years prior to the study, exact number of depressive episodes is not available for all subjects, duration of episodes between one and twelve month (mean 4.8)</p> <p><i>Comorbidities</i> none reported, exclusion of atypical forms of depression and any additional psychiatric disorders</p> <p><i>Medication</i> none</p> <p><i>Structural MRI findings</i> not reported</p>	<p>Healthy subjects</p> <p><i>n</i> = 22, 34.6 ± 6.2 years, 41% female</p> <p><i>Definition</i> without neurological/psychiatric condition</p> <p><i>Comorbidities</i> none reported</p> <p><i>Medication</i> none</p> <p><i>Structural MRI findings</i> not reported</p>	<p><i>Features</i> graph-theoretical measures (whole brain)</p> <p><i>Feature selection</i> MRMR</p> <p><i>Classifier</i> linear SVM</p>	<p><i>Overall accuracy</i> (CV) up to 99.3%</p> <p><i>Sensitivity</i> (CV) up to 99.3%</p> <p><i>Specificity</i> (CV) up to 99.3%</p> <p><i>Overall accuracy</i> (IV) not reported</p> <p><i>Sensitivity</i> (IV) not reported</p> <p><i>Specificity</i> (IV) not reported</p>
Zeng et al. (2012), Ma et al. (2013), Zeng et al. (2014) and Qin et al. (2015) ^b	<p><i>n</i> = 24, 31.8 ± 11.0 years, 67% female</p> <p><i>Diagnosis</i> acute episode of MDD (DSM-IV)</p> <p><i>Depression severity</i> HAM-D-17: 26.4 ± 5.2</p> <p><i>Depression duration</i> age of illness onset (years) 28.7 ± 10.9, number of previous episodes 1.6 ± 0.8, duration of current episode (months) 5.3 ± 6.3</p> <p><i>Comorbidities</i> exclusion of major psychiatric or neurological illnesses other than depression</p> <p><i>Medication</i> none</p> <p><i>Structural MRI findings</i> not reported</p>	<p>Healthy subjects</p> <p><i>n</i> = 29, 33.6 ± 10.3 years, 69% female</p> <p><i>Definition</i> without neurological/psychiatric condition</p> <p><i>Comorbidities</i> none reported</p> <p><i>Medication</i> none</p> <p><i>Structural MRI findings</i> not reported</p>	<p><i>Features</i> different methods</p> <p><i>Feature selection</i> different methods</p> <p><i>Classifier</i> different methods</p>	<p><i>Overall accuracy</i> (CV) 69.8–96.2%</p> <p><i>Sensitivity</i> (CV) not reported</p> <p><i>Specificity</i> (CV) not reported</p> <p><i>Overall accuracy</i> (IV) not reported</p> <p><i>Sensitivity</i> (IV) not reported</p> <p><i>Specificity</i> (IV) not reported</p>

Table 1 continued

Author and year	Depressed patients*	Controls*	Methods	Results
Yu et al. (2013)	<p>$n = 19$, 26.7 ± 7.6 years, 42% female</p> <p>Diagnosis acute MDD (DSM-IV)</p> <p>Depression severity HAM-D-17: 25.4 ± 6.3</p> <p>Depression duration not reported</p> <p>Comorbidities exclusion of psychiatric or neurological illnesses other than depression or schizophrenia</p> <p>Medication not reported</p> <p>Structural MRI findings not reported</p>	<p>Healthy subjects</p> <p>$n = 38$, 24.4 ± 4.5 years, 29% female</p> <p>Definition without neurological/psychiatric condition</p> <p>Comorbidities none reported</p> <p>Medication not reported</p> <p>Structural MRI findings not reported</p> <p>Schizophrenia</p> <p>$n = 32$, 24.0 ± 5.7 years, 22% female</p> <p>Diagnosis DSM-IV</p> <p>Comorbidities exclusion of other psychiatric or neurological illnesses</p> <p>Medication not reported</p> <p>Structural MRI findings not reported</p>	<p>Features functional connectivity of ROIs (whole brain)</p> <p>Feature selection intrinsic discriminative analysis or principal component analysis</p> <p>Classifier linear SVM (multiclass)</p>	<p>Overall accuracy (CV) up to 80.9%</p> <p>Sensitivity (CV) not reported</p> <p>Specificity (CV) not reported</p> <p>Overall accuracy (IV) not reported</p> <p>Sensitivity (IV) not reported</p> <p>Specificity (IV) not reported</p>
Guo et al. (2014)	<p>$n = 36$, 28.4 ± 9.7 years, 61% female</p> <p>Diagnosis first onset of acute MDD (DSM-IV)</p> <p>Depression severity HAM-D-24: 23.1 ± 13.4</p> <p>Depression duration not reported</p> <p>Comorbidities none reported, exclusion of psychiatric or neurological disorders</p> <p>Medication none</p> <p>Structural MRI findings not reported</p>	<p>Healthy subjects</p> <p>$n = 27$, 26.8 ± 9.6 years, 56% female</p> <p>Definition without neurological/psychiatric condition</p> <p>Comorbidities none reported</p> <p>Medication none</p> <p>Structural MRI findings not reported</p>	<p>Features graph-theoretical measures (whole brain)</p> <p>Feature selection univariate filter</p> <p>Classifier neural network</p>	<p>Overall accuracy (CV): up to 90.5%</p> <p>Sensitivity (CV) not reported</p> <p>Specificity (CV) not reported</p> <p>Overall accuracy (IV) not reported</p> <p>Sensitivity (IV): not reported</p> <p>Specificity (IV) not reported</p>
Cao et al. (2014)	<p>$n = 39$, 28.0 ± 7.5 years, 59% female</p> <p>Diagnosis MDD patients</p> <p>Depression severity HAM-D-17: 25.0 ± 5.0</p> <p>Depression duration 23.3 ± 37.8 months</p> <p>Comorbidities exclusion of neurological/serious physical illness</p> <p>Medication not reported</p> <p>Structural MRI findings not reported</p>	<p>Healthy subjects</p> <p>$n = 37$, 28.2 ± 6.5 years, 69% female</p> <p>Definition without neurological/psychiatric condition</p> <p>Comorbidities none reported</p> <p>Medication none</p> <p>Structural MRI findings not reported</p>	<p>Features functional connectivity of ROIs (whole brain)</p> <p>Feature selection probabilistic density function or univariate filter</p> <p>Classifier nonlinear SVM</p>	<p>Overall accuracy (CV) up to 84.2%</p> <p>Sensitivity (CV) not reported</p> <p>Specificity (CV) not reported</p> <p>Overall accuracy (IV) not reported</p> <p>Sensitivity (IV) not reported</p> <p>Specificity (IV) not reported</p>

Table 1 continued

Author and year	Depressed patients*	Controls*	Methods	Results
Bhaumik et al. (2016)	<p>$n = 38, 21.0 \pm 1.5$ years, 76% female</p> <p>Diagnosis remitted MDD patients</p> <p>Depression severity HAM-D-17: 2.4 ± 3.0</p> <p>Depression duration 1.2 ± 1.3 episodes</p> <p>Comorbidities exclusion of neurological/serious physical illness</p> <p>Medication none</p> <p>Structural MRI findings not reported</p>	<p>Healthy subjects</p> <p>$n = 29, 21.0 \pm 1.6$ years, 55% female</p> <p>Definition subjects and their first degree relatives without axis I or II psychiatric disorders</p> <p>Comorbidities none reported</p> <p>Medication none</p> <p>Structural MRI findings not reported</p>	<p>Features functional connectivity of prespecified ROIs</p> <p>Feature selection different methods</p> <p>Classifier linear SVM</p>	<p>Overall accuracy (CV) up to 76.1%</p> <p>Sensitivity (CV) up to 81.5%</p> <p>Specificity (CV) up to 68.9%</p> <p>Overall accuracy (IV) up to 77.8%</p> <p>Sensitivity (IV) up to 80.0%</p> <p>Specificity (IV) up to 75.0%</p>
Ramasubbu et al. (2016)	<p>$n = 45, 37 \pm 11$ years, 64% female (subdivided into groups according to symptom severity)</p> <p>Diagnosis acute episode of MDD (DSM-IV)</p> <p>Depression severity HAM-D-17: 22 ± 4</p> <p>Depression duration 12 ± 8 years overall disease duration, 59 ± 66 months duration of current episode</p> <p>Comorbidities exclusion of major psychiatric disorders other than depression; exclusion of neurological or medical disorders; exclusion of treatment-resistant MDD patients</p> <p>Medication none</p> <p>Structural MRI findings not reported</p>	<p>Healthy subjects</p> <p>$n = 19, 33 \pm 10$ years, 58% female</p> <p>Definition subjects and family history without axis I psychiatric disorders</p> <p>Comorbidities not reported</p> <p>Medication not reported</p> <p>Structural MRI findings not reported</p>	<p>Features different representations of functional connectivity</p> <p>Feature selection different methods</p> <p>Classifier linear SVM</p>	<p>Overall accuracy (CV) 49–66% (only significant in group with most severe symptoms)</p> <p>Sensitivity (CV) 8–59%</p> <p>Specificity (CV) 59–89%</p> <p>Overall accuracy (IV) not reported</p> <p>Sensitivity (IV) not reported</p> <p>Specificity (IV) not reported</p>

CV cross validation, HAM-D-17 Hamilton Rating Scale for Depression, IV independent validation, MRRM maximum redundancy minimum relevance feature selection, ROI region of interest, SVM support vector machine

* Measures represent mean \pm SD if not stated differently

^a Training sample characteristics. This study involved a second validation sample

^b Grouped because of identical sample characteristics and overlapping authors

and sufficiently large samples (Arbabshirani et al. 2017) is crucial to assess whether these desirable methods can be translated from well-controlled research environments to routine clinical diagnostic applications in more heterogeneous patient populations.

The purposes of this multi-step investigation were replication, clinical translation, and model optimization: The main goal was to clarify if the approach to combine rs-fcMRI with a common MVPA approach to diagnose MDD is feasible in a diverse population as a prerequisite for generalization to routine depression care. Thereby, we wanted to identify suitable modelling parameters and FS techniques. SVMs, a well-established classification method, were applied to rs-fcMRI data from a large cohort acquired under clinically realistic conditions. The MDD sample was heterogeneous regarding symptom severity, comorbidity, and therapy status. A non-depressed population sample served as control group. In addition, we aimed to corroborate the feasibility of this approach in a more homogeneous subgroup of patients with more distinct depressive symptoms.

Methods

First, we give a brief overview of our data analysis strategy: For the main analysis, two sub-samples—both comprising patients with MDD and controls—were drawn from participants in a cohort study. Estimates of functional connectivity derived from rs-fMRI data were used as features to train computational models to identify individual MDD patients. All this modelling was accomplished in the form of planned exploratory analyses with widely varying technical parameters in the first sub-sample. The second sub-sample was reserved to validate potentially successful models identified at that exploratory stage. In additional analyses, we evaluated modelling performance in a more selected subgroup and explored the univariate information content of the features used.

Subjects: the BiDirect study sample

This investigation is based on baseline data from the BiDirect study (Hermesdorf et al. 2016; Teismann et al. 2014; Teuber et al. 2017). BiDirect primarily aims to disentangle the bidirectional relationship between depression and subclinical arteriosclerosis. The particular analysis reported here uses data from the depression cohort (current or recent episode) and the population control cohort (invited via population register). The acquired data comprised clinical, psychological, and neuropsychometric testing. Structural and functional [wakeful rest, emotional faces paradigm (Dannowski et al. 2009)] imaging was also performed. Psychometric assessment included the Center

for Epidemiological Studies Depression Scale (CES-D) (Radloff 1977), Hamilton Rating Scale for Depression (HAM-D-17) (Hamilton 1960), and the Mini International Neuropsychiatric Interview (MINI) (Ackenheil et al. 1999). Clinical data included information on medical history and current medication. These drugs were classified according to the Anatomical Therapeutic Classification Systems (ATC, http://www.whooc.no/atc_ddd_index/). A comprehensive study protocol detailing recruitment and data acquisition of BiDirect has been published (Teismann et al. 2014). Sub-sample selection for this methodological work, including demographical and clinical characteristics, will be detailed in the following.

Resting-state fMRI data acquisition and feature extraction

MRI data were acquired in a setting equivalent to clinical appointments: The same 3 Tesla scanner is also used for clinical examinations, and data acquisition was accomplished by clinical personnel. Rs-fMRI data were acquired using T2*-weighted echo planar imaging at the end of the MRI protocol [see the methods supplement (Online Resource 1) for further details on the rs-fMRI protocol]. 1378 technically complete rs-fMRI data sets were obtained. Structural scans were subjected to neuroradiological reporting.

We followed a standard approach for preprocessing and analysis of individual rs-fMRI data sets implemented in the Data Processing Assistant for Resting-State fMRI (DPARSF 2.3) (Chao-Gan and Yu-Feng 2010), REST 1.8 (Song et al. 2011) and SPM8 (<http://www.fil.ion.ucl.ac.uk/spm/>) [see the methods supplement (Online Resource 1) for further details].

Signal time courses were extracted from two sets of ROIs: (1) peak coordinates ($n = 38$) derived from a meta-analysis on rs-fMRI in depression (Sundermann et al. 2014b) as center coordinates for spherical ROIs (5 mm radius). They represent prior knowledge on altered spontaneous activity in MDD, a common approach in MVPA of MRI data (Chu et al. 2012; Schrouff et al. 2013). (2) Another set consisted of 200 ROIs spanning the whole gray matter. They were derived using spatially constrained spectral clustering of rs-fMRI data with the aim to better comprehend the functional architecture of the human brain compared to conventional atlases based on surface anatomy (Craddock et al. 2012). They are distributed with DPARSF (Chao-Gan and Yu-Feng 2010). FC was determined by calculating pairwise correlation (Pearson's r) for all ROIs within each of the two sets separately. Subsequently, correlation coefficients were z-transformed. FC analyses resulted in 703 unique features for the meta-analytically defined ROIs and 19,900 unique features for the whole-brain parcellation.

Table 2 Demographical and medical characteristics of subjects in subset 1 used for exploratory model comparison

	MDD (<i>n</i> = 180)	Controls (<i>n</i> = 180)	Statistical significance ^a
<i>General demographical characteristics</i>			
Age (mean ± SD)	50.8 ± 7.1 years	51.2 ± 7.6 years	
Sex (female:male)	86:94	86:94	
<i>Depression-related characteristics</i>			
ICD F32 (<i>n</i>)	90 (50.0%)		
ICD F33 (<i>n</i>)	90 (50.0%)		
Disease duration ^b (median)	49.5 months		
Current HAM-D-17 score (mean ± SD)	13.2 ± 6.4		
Current CES-D score (mean ± SD)	26.0 ± 12.0		
Current antidepressant medication (<i>n</i>)	149 (82.8%)		
ATC N06AA—non-selective monoamine reuptake inhibitors (<i>n</i>)	20 (11.1%)		
ATC N06AB—selective serotonin reuptake inhibitors (<i>n</i>)	40 (22.2%)		
ATC N06AF—monoamine oxidase inhibitors, non-selective (<i>n</i>)	1 (0.6%)		
ATC N06AG—monoamine oxidase A inhibitors (<i>n</i>)	0 (0.0%)		
ATC N06AX—other antidepressants (<i>n</i>)	111 (61.7%)		
History of anxiety (<i>n</i>)	57 (31.7%)		
History of psychotic symptoms (<i>n</i>)	3 (1.7%)		
History of addiction (<i>n</i>)	10 (5.6%)		
History of anorexia (<i>n</i>)	3 (1.7%)		
<i>Physical (Co-)morbidity^c</i>			
History of cardiovascular disease (<i>n</i>)	85 (47.2%)	44 (24.4%)	<i>p</i> < 0.001 ^e
History of respiratory disease (<i>n</i>)	28 (15.6%)	20 (11.1%)	
History of endocrine and metabolic disorders (<i>n</i>)	46 (25.6%)	30 (16.7%)	<i>p</i> = 0.039 ^e
History of history of cancer (<i>n</i>)	15 (8.3%)	9 (5.0%)	
History of movement disorders (<i>n</i>)	4 (2.2%)	1 (0.6%)	
History of rheumatism or chronic pain (<i>n</i>)	44 (24.4%)	20 (11.1%)	<i>p</i> = 0.001 ^e
<i>Medication other than antidepressive</i>			
ATC A—alimentary tract and metabolism (<i>n</i>)	41 (22.8%)	14 (7.8%)	<i>p</i> < 0.001 ^e
ATC B—blood and blood forming organs (<i>n</i>)	6 (3.3%)	7 (3.9%)	
ATC C—cardiovascular system (<i>n</i>)	73 (40.6%)	43 (23.9%)	<i>p</i> < 0.001 ^e
ATC D—dermatologicals (<i>n</i>)	4 (2.2%)	2 (1.1%)	
ATC G—genito-urinary system and sex hormones (<i>n</i>)	13 (7.2%)	14 (7.8%)	
ATC H—systemic hormonal preparations, excluding sex hormones and insulins (<i>n</i>)	28 (15.6%)	25 (13.9%)	
ATC J—anti-infectives for systemic use (<i>n</i>)	1 (0.6%)	1 (0.6%)	
ATC L—antineoplastic and immunomodulating agents (<i>n</i>)	0 (0.0%)	1 (0.6%)	
ATC M—musculo-skeletal system (<i>n</i>)	21 (11.7%)	15 (8.3%)	
ATC N01—anesthetics (<i>n</i>)	0 (0.0%)	0 (0.0%)	
ATC N02—analgesics (<i>n</i>)	21 (11.7%)	15 (8.3%)	
ATC N03—antiepileptics (<i>n</i>)	9 (5.0%)	0 (0.0%)	
ATC N04—anti-Parkinson drugs (<i>n</i>)	2 (1.1%)	0 (0.0%)	
ATC N05—psycholeptics ^d (<i>n</i>)	98 (54.4%)	0 (0.0%)	<i>p</i> < 0.001 ^e
ATC N06—psychoanaleptics excl. antidepressants (see above) (<i>n</i>)	1 (0.6%)	0 (0.0%)	
ATC N07—other nervous system drugs (<i>n</i>)	0 (0.0%)	0 (0.0%)	
ATC P—antiparasitic products, insecticides and repellents (<i>n</i>)	0 (0.0%)	0 (0.0%)	
ATC R—respiratory system (<i>n</i>)	7 (3.9%)	16 (8.9%)	
ATC S—sensory organs (<i>n</i>)	5 (2.8%)	8 (4.4%)	
ATC V—various (<i>n</i>)	0 (0.0%)	0 (0.0%)	

^a Only significant results are reported, only tested for significance in cases when more than ten subjects were in any group

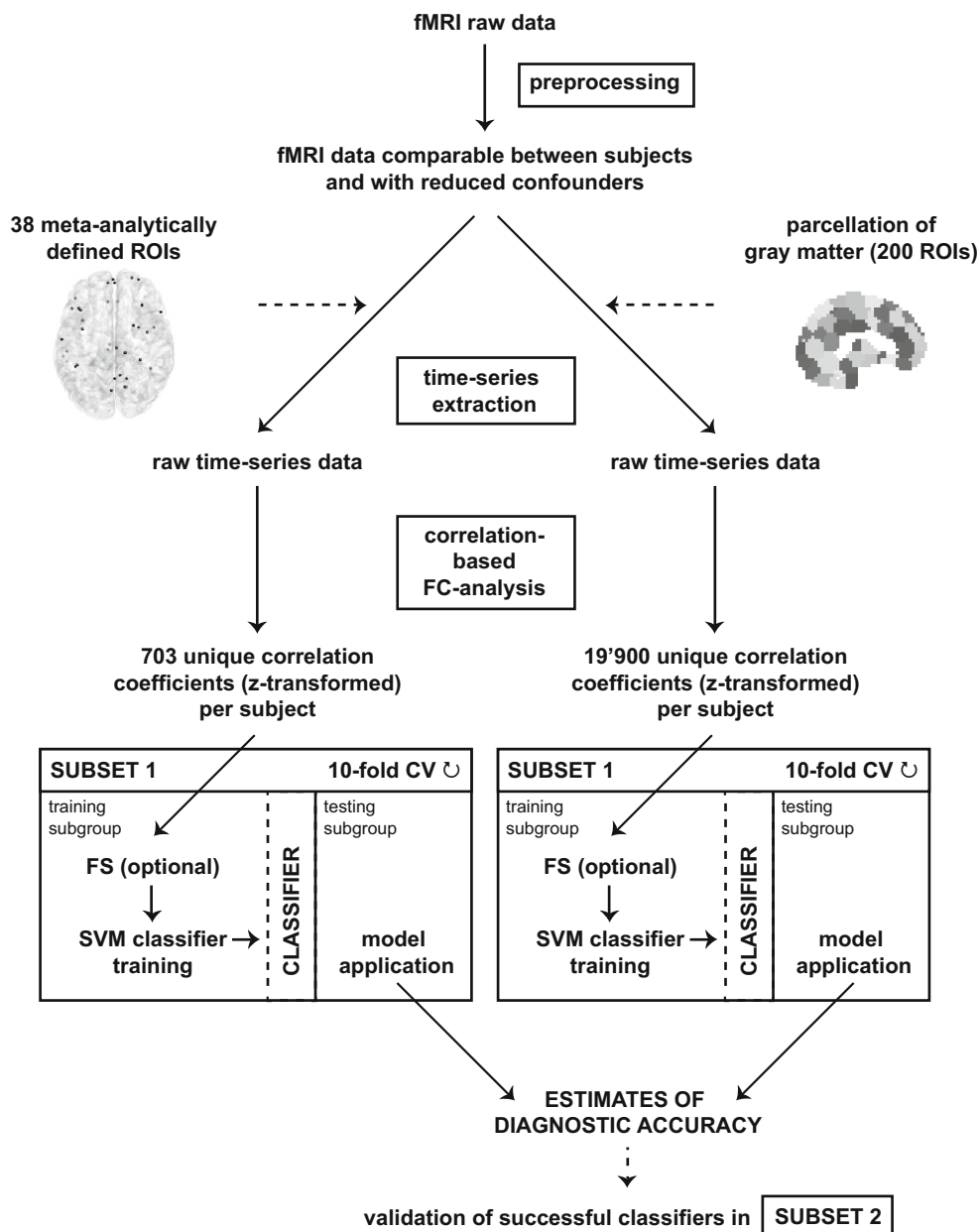
^b Since initial medical diagnosis

^c Self-report data (summary of prespecified items)

^d Including anxiolytic and antipsychotic medication

^e Chi²-test

Fig. 2 Overview of the general data analysis pipeline for generating FC-based classifiers



independent hold-out data set to validate potentially successful models with a high level of evidence.

See Fig. 1 and the methods supplement (Online Resource 1) for further details on sub-sample selection.

Training and performance estimation of diagnostic MVPA models

Subsequent MVPA used the z-transformed correlation coefficients as features to train diagnostic models for this two-class problem (MDD vs. controls). Analyses were implemented using RapidMiner (open source edition, version 5.3.013, RapidMiner GmbH, Dortmund, North Rhine-

Westphalia, Germany, <http://github.com/rapidminer/>) (Mierswa et al. 2006; Schowe 2011). Given the comparably large data set available, we opted for a hierarchical approach to identification and validation of potential diagnostic models. In a first step, a wide range of models was explored in subset 1. As pointed out previously, the independent subset 2 was kept for validation and potentially further refinement of models. An overview of the MVPA approach is presented in Fig. 2. In-depth information on our MVPA approach is presented in the methods supplement (Online Resource 1).

We explored the diagnostic accuracy, sensitivity, and specificity of soft-margin SVM models [C-SVC from LIBSVM (Chang and Lin 2011)]. Linear kernels are the

preferential choice in diagnostic SVM models of neuroimaging data given the typically higher number of features than patients (Orri et al. 2012), have been the most common choice in MDD (see Table 1), and were consequently evaluated here. We additionally explored non-linear radial basis function (RBF) kernels to facilitate a higher model complexity. Model parameters were varied systematically over a wide range spanning potential under- and overfitting to identify optimal settings. Models were tested either with one of four different FS techniques or without any FS: a t test-based filter, a filter based on linear SVM weights, recursive feature elimination (SVM-RFE) (Guyon et al. 2002), and minimum redundancy maximum relevance (MRMR) FS (Ding and Peng 2005). Classification accuracies were estimated using cross validation (CV) (Pereira et al. 2009).

In total, 210 different settings were assessed in the exploratory analysis in subset 1.

Subgroup analysis in patients with more severe current depressive symptoms

To further explore the technical feasibility of different models as well as the potential influence of the severity of depressive symptoms at the time of MR scanning and to allow for comparison with pilot studies in severely depressed patients, we investigated the diagnostic accuracies in a subgroup of subset 1. This subgroup comprised one-third of MDD patients ($n = 60$) with the most severe depressive symptoms according to HAM-D-17 and their respective matched controls ($n = 60$). The mean HAM-D-17-score in this MDD subgroup was $20.2 \pm \text{SD } 2.9$. For further subject characteristics, see supplementary Table 3 (Online Resource 2). Exploratory analyses comprised the same modelling and validation approaches as in the main analysis. As an exception, only moderate C values were tested as very low or very high C values in the main analysis did either not change results or led to biased classifiers with a high tendency towards a single diagnostic label (see results). 90 different settings were assessed.

Parameter sets that achieved an overall cross-validated accuracy of at least 60% were further validated in subset 2. We, therefore, drew a comparable subgroup of the most severely depressed MDD patients ($n = 60$, mean HAM-D-17 score: $20.9 \pm \text{SD } 3.3$) and their controls ($n = 60$) from the independent subset 2 (see supplementary Table 6 in Online Resource 2). Parameters derived from successful models in the cross-validated exploratory subgroup analyses in subset 1 were used to train classifiers based on the entire subgroup of subset 1. Generalizability of these hypothetically most powerful models was assessed by applying the resulting decision rules to subjects in the

subgroup of subset 2. These hypotheses about overall accuracies generated in the more severely depressed subgroup of subset 1 were tested using a one-tailed binomial test and corrected for multiple comparisons by controlling the false-discovery rate (FDR) (Benjamini and Hochberg 1995).

Post hoc estimation of feature set information content by univariate analyses

As a final step, we retrospectively assessed how much information about the clinical diagnosis of unipolar depression the FC-based feature sets inherited by means of more classical univariate analyses. We, therefore, conducted two-sided two-sample t tests in the entire subset 1 and the subgroup of subset 1 with pronounced depressive symptoms both for FC coefficients based on the meta-analytical ROIs and the whole-brain parcellation. To assess information content while controlling for multiple comparisons, we then estimated the proportion π_1 of univariate results that truly followed the alternative hypothesis of a group difference of patients and controls using Matlab. This method is based on the assumption of a uniform distribution of p values that follow the null hypothesis (Storey and Tibshirani 2003) and was introduced in the framework of the positive false-discovery rate for high-dimensional data sets (Storey and Tibshirani 2003; Storey 2002).

Results

Models based on features supported by a priori knowledge

In the explorative model, identification step of the main analysis, SVM models based on all pairwise connections of 38 meta-analytically defined ROIs, reached overall diagnostic accuracies from 47.5% (linear kernel, $C = 0.1$) to 53.6% (linear kernel, $C = 0.001$) in the explorative analysis in the entire subset 1. Sensitivities ranged from 45.0 to 97.2% and specificities from 1.7 to 53.9% as some models assigned nearly all subjects to one single group. FS did not improve performance with overall accuracies ranging from 47.0% (linear kernel, $C = 0.1$, FS based on SVM weights) to 53.3% (linear kernel, $C = \{0.001, 0.01, 0.1\}$, FS based on t tests) with sensitivities from 45.0 to 98.33% and specificities from 3.3 to 51.7%. To summarize, results were distributed closely around the chance level of 50% accuracy with a typical tradeoff of sensitivities and specificities. For detailed results, see supplementary Table 1 (Online Resource 2).

Table 3 Models with cross-validated accuracies of at least 60.0% and corresponding results of model validation in the hold-out data set

(a) Main analysis

Yielded no positive results

(b) Subgroup analysis [most severely depressed patients (n = 60) and their corresponding controls (n = 60)]

FS	Features (n)	C-SVC-classifier	CV in subset 1			Validation in subset 2			p(FDR)	ACC
			ACC (%)	SENS (%)	SPEC (%)	ACC (%)	SENS (%)	SPEC (%)		
<i>Models based on connectivity of meta-analytically defined regions</i>										
./.	703	RBF ($\gamma = 0.01$), $C = 10$	60.0	60.0	60.0	55.0	46.7	63.3	0.171	
t test ^a	141	lin, $C = 0.1$	65.0	63.3	66.7	57.5	45.0	70.0	0.087	
t test ^a	141	RBF ($\gamma = 0.01$), $C = 10$	61.7	60.0	63.3	55.8	45.0	66.7	0.140	
SVM ^b	141	RBF ($\gamma = 0.01$), $C = 1$	61.7	71.7	51.7	61.7	58.3	65.0	0.046*	
MRMR	141	lin, $C = 0.1$	61.7	63.3	60.0	60.0	46.7	73.3	0.059	
MRMR	141	RBF ($\gamma = 0.01$), $C = 10$	60.8	61.7	60.0	56.7	41.7	71.7	0.111	
SVM-RFE	141	RBF ($\gamma = 0.01$), $C = 0.1$	60.0	75.0	45.0	61.7	61.7	61.7	0.046*	
SVM-RFE	141	RBF ($\gamma = 0.01$), $C = 1$	60.8	68.3	53.3	60.8	58.3	63.3	0.048*	
<i>Models based on whole gray matter connectivity</i>										
SVM ^b	3980	RBF ($\gamma = 0.01$), $C = 1$	60.8	68.3	53.3	58.3	61.7	55.0	0.076	
SVM ^b	3980	RBF ($\gamma = 0.01$), $C = 10$	60.8	68.3	53.3	59.2	61.7	56.7	0.070	
MRMR	141	lin, $C = 0.1$	60.0	60.0	60.0	54.2	46.7	61.7	0.206	
SVM-RFE	3980	RBF ($\gamma = 0.01$), $C = 1$	60.0	66.7	53.3	57.5	61.7	53.3	0.087	
SVM-RFE	3980	RBF ($\gamma = 0.01$), $C = 10$	60.0	66.7	53.3	58.3	63.3	53.3	0.076	

CV cross validation, FDR false-discovery rate, FS feature selection, ACC overall accuracy, SENS sensitivity, SPEC specificity, SVC support vector classification, MRMR maximum redundancy minimum relevance feature selection, SVM-RFE recursive feature elimination by support vector machines, lin linear kernel, RBF radial basis function kernel

* Models with overall accuracies significantly above chance in the validation set, corrected (false-discovery rate)

^a t test-based weighting with weight-based feature selection

^b SVM-based weighting with weight-based feature selection

Models based on whole-brain connectivity

Models based on the entire gray matter parcellation in the model identification step of the main analysis reached overall diagnostic accuracies from 45.0% (RBF kernel, $\gamma = 0.01$, $C = \{10, 100, 1000\}$) to 52.8% (linear kernel, $C = \{0.01, 0.1, 1, 10, 100, 1000\}$) in the explorative analysis in the entire subset 1. Sensitivities ranged from 0.0 to 88.3% and specificities from 9.4 to 100.0%. Some instances of automated feature selection yielded slightly better performance in this case with overall accuracies ranging from 48.1% (RBF kernel, $\gamma = 1$, $C = \{0.001, 0.01, 0.1\}$, MRMR FS) to 56.1% (linear kernel, $C = 0.1$, MRMR FS) with sensitivities from 51.1 to 83.9% and specificities from 12.8 to 55.0%. Thus, these exploratory results were slightly shifted with regard to chance level. For detailed results, see supplementary Table 2 (Online Resource 2).

With the range of diagnostic performances observed, no model generation technique tested in the entire subset 1

reached clinically relevant accuracies in the exploratory analysis. We, therefore, refrained from further model validation in subset 2 (Table 3a).

Subgroup analysis in patients with higher current depressive symptoms

Models explored in the subgroup of most depressed patients in subset 1 to assess the general feasibility of this approach achieved diagnostic accuracies from 40.8 to 65.0%. 13 models reached cross-validated accuracies of at least 60.0%. For detailed results, see supplementary Tables 4 and 5 (Online Resource 2).

These 13 models were further assessed by validation in an independent non-overlapping subgroup with the most severely depressed patients from subset 2 (hypothesis tests). Three of these modelling approaches performed significantly above chance ($p < 0.05$, FDR), and further six models reached a formal trend to statistical significance ($p < 0.1$, FDR). Detailed validation results are presented in Table 3b.

Post hoc estimation of feature set information content by univariate analyses

The following results calculated with a conventional approach serve as a surrogate of the information content of features. The estimated proportion π_1 of pairwise connections between meta-analytically defined ROIs which contain group information of diagnostic interest in the univariate analyses was 9.05% in the entire subset 1 and 16.38% in the subgroup with more severely depressed patients and matched controls. Corresponding π_1 of features based on the whole-brain analysis was 19.00% in the entire subset 1 and 26.18% in the subgroup analysis. p value histograms illustrating this deviation of univariate effects from a uniform distribution under the null hypothesis are shown in Online Resource 3.

Discussion

This study in a diverse unipolar depression and population control sample fails to confirm reports (Cao et al. 2014; Craddock et al. 2009; Guo et al. 2014; Lord et al. 2012; Ma et al. 2013; Qin et al. 2015; Yu et al. 2013; Zeng et al. 2012; Zeng et al. 2014) that the combination of MVPA and rs-fMRI facilitates the clinically reliable identification of individual MDD patients. This is particularly noteworthy, since the current investigation follows methodological principles (combinations of correlation-based FC analyses and SVMs) that have yielded particularly promising results in pilot studies not only in depression but have been commonly used in pilot studies in diverse mental disorders in recent years (Arbabshirani et al. 2017; Orru et al. 2012; Sundermann et al. 2014a).

Sample characteristics

Pilot studies on the combination of rs-fMRI and MVPA in MDD (Cao et al. 2014; Craddock et al. 2009; Guo et al. 2014; Lord et al. 2012; Ma et al. 2013; Qin et al. 2015; Yu et al. 2013; Zeng et al. 2012; Zeng et al. 2014) as well as most work on diagnostic MVPA of fMRI data so far (Sundermann et al. 2014a) have adopted control groups of mostly young explicitly healthy subjects. In contrast, this analysis features controls from the general population (Teismann et al. 2014). Some recruitment bias can also be expected in general population samples like this (Heun et al. 1997). Even though subjects with signs of depression among controls as well as disorders potentially mimicking MDD in the depression cohort were excluded and data sets were balanced for potential general confounders, we aimed at keeping sample heterogeneity at its original level. Thus, compared to the

above-detailed previous work, there is substantial heterogeneity in our data regarding age-related physical and mental comorbidity, medication in both groups, and levels of depressive symptoms as well as disease duration in the MDD group (Table 2). There is initial evidence that in addition to increasing the level of heterogeneity antidepressant medication may obscure typical FC alterations in depression beyond correlates of symptom reduction (McCabe and Mishor 2011). Thus, current antidepressant medication may limit the diagnostic power of MVPA (Qin et al. 2015). Moreover, the specific recruitment strategy for BiDirect (Teismann et al. 2014) and the matching procedure essential for unbiased estimates of classifier performance resulted in a comparatively older sample lacking female predominance compared to typical disease-onset MDD populations (Andrade et al. 2003). There is evidence that particularly age has a major effect on brain structure and function (Douaud et al. 2014) including measures typically derived from rs-fcMRI (Damoiseaux et al. 2008; Dosenbach et al. 2010). In addition, effects of cardiovascular disorders on depression pathogenesis and vice versa are expected to be more prevalent than in younger depression samples. This reflects the core objective of the underlying BiDirect study (Teismann et al. 2014).

Despite its inherent limitations we believe that the data set used here represents clinical populations far better than homogenous MDD samples and explicitly healthy control groups in the previous studies and is, therefore, beneficial to rigorously assess the transferability of current MVPA approaches to routine care.

Significant positive results in the subgroup analysis with more severely depressed patients further support the assumption that sample heterogeneity is an important determinant of the ineffectiveness of this approach in this sample. Moreover, this indicates that these rs-fcMRI models in depression may be particularly sensitive to the current depressive state. This finding is in line with a recently reported failure to achieve above-change diagnostic accuracies in subgroups other than one comprising patients with the highest current symptom severity (Ramasubbu et al. 2016). That study is, however, limited by its small sample size.

It remains unclear if these models are also capable of sufficiently representing longer term traits (Bhaumik et al. 2016; Graham et al. 2013; Qin et al. 2015), which may be more important for clinical decision making in this population. Nevertheless, results from the subgroup analysis confirm the general feasibility of this computational approach based on rs-fcMRI. However, even most successful models in the subgroup analysis did not reach sufficient diagnostic accuracies for clinical use.

Methodological aspects

We used two conceptually different approaches to extract relevant features from the original data. Compared to voxel-based FC analyses, our ROI-based approaches limit redundancy. Avoiding excessive numbers of features is supposed to improve the power of MVPA (Mwangi et al. 2014; Pereira et al. 2009). We have compared two different sets of ROIs: One set of regions optimized specificity by relying on prior knowledge about the effects of diagnostic interest (Sundermann et al. 2014b). Despite the potential advantage of a strictly hypothesis-driven approach, important discriminative features may be missed. We, therefore, additionally extracted features based on the entire gray matter (Craddock et al. 2012). The univariate post hoc analysis of information content confirms that both ROI definitions lead to informative feature sets.

First, SVM models were trained and tested based on the whole feature set of pairwise connections. In addition, we applied different automated FS methods to select most discriminative features. Model parameters were systematically varied over a wide range at the exploratory stage of analyses as such parameter optimization is mandatory to avoid under- or overfitting and to improve the generalizability of our results to comparable, yet still versatile MVPA approaches in MDD (Arbabshirani et al. 2017; Pereira et al. 2009). Please also note that if no successful model is identified in a wide search like this, poor classification performance can be assumed to be rather caused by the fact the underlying data does not convey sufficient information for the diagnostic question of interest to be captured by this commonly used family of computational models.

Successful modelling approaches in the subgroup analysis of patient with a preeminent depressive state were dominated by nonlinear models. This indicates a high complexity of the underlying information to be captured, i.e., beyond a pure summation of univariate information. As an observation—not directly amenable to statistical analyses—the feature set based on the previous knowledge about FC alterations in MDD led to superior models compared to features representing whole gray matter connectivity. However, in this setting, all successful SVM models required further FS, preferably based on SVMs themselves. In contrast to findings in the subgroup analysis, whole-brain features with powerful FS (MRMR) yielded slightly better diagnostic accuracies compared to whole-brain models with other types of FS or literature-informed features in the main analysis in the original more heterogeneous data set. However, as stated previously, the generality of this result is limited by its explorative nature as no approach reached sufficient accuracies in this diverse sample.

As there were no successful models in the first exploratory stage of the main analysis, there was no model to be validated in a second stage. The hold-out data set was, therefore, only used in the subgroup analysis with most severely depressed patients. While overall performance was slightly poorer in the validation set of the subgroup analysis, there was generally a good agreement and most results were either significant or exhibited a formal trend towards significance after correcting for multiple comparisons. These findings are in line with earlier work in this field, for example, the seminal paper by Craddock et al. (2009).

Potential limitations and future directions

Reliable estimates of FC can be obtained using relatively short runs of data acquisition (Van Dijk et al. 2010). According to recent evidence prolonged acquisitions—even beyond clinically realistic timeframes—can, however, improve reliability depending on analysis techniques (Anderson et al. 2011, Birn et al. 2013). Determining the optimal scan duration for these particular analysis methods was not within the scope of this study. Therefore, it cannot be excluded that prolonged scanning may help reach sufficient diagnostic accuracies. All subjects were examined in a single center with a single MRI scanner. In a clinical context, it will be important that successful techniques can be adapted to different hardware.

The resting period in most subjects of this study followed an fMRI run with emotional faces, a popular paradigm in MDD research (Stuhrmann et al. 2011). Preceding tasks can influence quantitative estimates in rs-fcMRI (Pyka et al. 2013; Waites et al. 2005). During preparation of the training and validation sets, the percentage of subjects with this preceding task was balanced across MDD patients and controls by matching. This ensures that diagnostic decisions were not driven by the existence of this protocol variation. Though these preceding stimuli marginally limit the generalizability to other integrations of the resting period in imaging protocols, the expected diagnostic power has not been diminished.

FC analysis based on Pearson correlation is one of the most common approaches in rs-fcMRI (Margulies et al. 2010). Recently, sparse connectivity models have been proposed that were able to achieve higher diagnostic power than conventional correlation-based approaches (Rosa et al. 2015). Partial correlation and independent component analyses are further alternative strategies for extracting FC features (Margulies et al. 2010). Such approaches may be an important area of future research as long as computational expenses can be accommodated to clinical settings.

It has been proposed that clinically defined mental disorders, such as MDD, may in fact represent a diverse

spectrum of brain disorders with important differences in underlying disease mechanisms (Krishnan 2014). This potential neurobiological source of heterogeneity in clinically defined samples can be subject to future work that tries to optimize diagnostic strategies in mental disorders (Atluri et al. 2013). However, this also points toward the possibility that a clinical diagnosis of MDD is probably a weak reference (Hickie et al. 2013) for assessing new diagnostic tests. Yet, no other reference is currently available. The weaknesses of clinical signs and scores as diagnostic reference also apply to our group definition approach: In particular, the adoption of a range of surrogate parameters to achieve a high sensitivity for excluding potentially depressed subjects from the original control cohort reduces the spectrum of psychiatric comorbidity in the final samples. Moreover, information about mental and older age-related comorbidity is limited by the fact that these in part rely upon self-report data. These facts limit conclusions about the differential diagnostic ability of this classification approach regarding other mental conditions associated with depression.

The current investigation aims at distinguishing MDD patients from non-depressed controls. Using this principal diagnosis as the key diagnostic question is a straightforward approach to identify classification methods and parameters viable in MDD. It has been followed in a majority of such pilot studies (Patel et al. 2016; Sundermann et al. 2014a). However, such tools have recently been applied to more specific clinical questions, such as prognostics or differential diagnoses (Fu et al. 2013; Hahn et al. 2015; Lener and Iosifescu 2015; Patel et al. 2016; Phillips et al. 2015; Qin et al. 2015; Schmaal et al. 2014; Sundermann et al. 2014a; van Waarde et al. 2015). Prediction of therapy response in MDD by biomarkers, including fMRI, is currently investigated as part of prospective trials (Dunlop et al. 2012; Grieve et al. 2013; Kennedy et al. 2012; Trivedi et al. 2016; Williams et al. 2011). Beyond the fact that such diagnostic problems presumably involve more highly selected samples, they may in part rely on different neurobiological bases (Fu et al. 2013; Kupfer et al. 2012; Lener and Iosifescu 2015; Phillips et al. 2015). The finding that the combination of rs-fcMRI and MVPA does not generalize to a clinically more realistic population here does, therefore, not necessarily hold true for such more specific clinical questions. Hence, we propose that these treatment-related questions and adequate selection of patients to be examined (with disease severity as a crucial factor) are important lines of future research. Even if rs-fcMRI alone does not serve as a clinically reliable diagnostic biomarker, it may be included in MRI biomarkers in future studies which may rely on integrated multimodal instead of unimodal information (Calhoun and Sui 2016; Douaud et al. 2013; Wee et al. 2012). Another limitation is

that information about potential clinical confounders, such as comorbidity or medication, cannot be reasonably dealt with directly in diagnostic models using common techniques, such as SVMs. We, therefore, believe that another direction of future research should be the development of more sophisticated computational models (Li et al. 2011) which are to a lesser degree influenced by clinical heterogeneity.

We would like to conclude by saying that we are aware from extensive discussions with other scientists that these sobering results may be very disappointing and even discouraging for clinical and non-clinical scientists working in this interdisciplinary field. However, we would like to stress that this paper should not be considered as an objection against the work people have and will put into these kinds of methods. Rather, we would like this article to be understood as a note of caution to prevent a premature translation of “standard” methods in this emerging research area into clinical practice or large-scale prospective clinical trials for derivative medical products. In the context of this study, this statement particularly refers to the SVM family of classifiers. As discussed above (where directly related to our results), there are further new, potentially more sophisticated methods currently being tested, developed and to be developed. Further lines of research cover the whole range from improved data acquisition (Ugurbil et al. 2013) through improved pre-processing strategies (Murphy et al. 2013; Salimi-Khorshidi et al. 2014) to new, more complex classes of classification methods, including deep learning (Arbabshirani et al. 2017; Plis et al. 2014; Sarraf and Tofghi 2016) and elastic net approaches (Bowman et al. 2016; Mwangi et al. 2014; Schouten et al. 2016; Zou and Hastie 2005). Therefore, our hope is that these results will motivate researchers to improve such techniques for diagnostic classification.

Conclusions

Straightforward combinations of classification methods that are seemingly established based on the results of small homogenous samples do not translate to a diverse sample in a situation closer to daily life and thus a potential clinical application. The sample size in this investigation limits the risk of false-positive results compared to earlier work and is, therefore, suitable to assess the reproducibility of such findings. Results of a subgroup analysis show that this methodological approach is feasible yet still not clinically reliable even in patients with a preeminent depressive state. This indicates that such MVPA approaches need to take the heterogeneity of clinical populations (including symptom severity) into account. Presumably, they also need to focus

on more specific clinical questions, such as therapeutic outcomes as well as improvement of data acquisition and analysis techniques.

Acknowledgements The authors thank all study participants and the entire team of the BiDirect study, including collaborators in associated institutions. Thanks also go to Jens Bode for support with the execution of data analyses.

Compliance with ethical standards

Conflicts of interest The following authors declared additional financial relationships not directly related to this work: BP has received grants from EU European Social Fund and the German Ministry of Education and Research. VA is board member for Lundbeck, Otsuka, Servier and Tromsdorff and has received honoraria from Lundbeck, Otsuka and Servier.

Funding BiDirect is funded by a research Grant (FZK: 01ER0816) from the German Federal Ministry of Education and Research (BMBF). This analysis was additionally supported by BMBF Grant 01ER1205.

Ethical approval The study was approved by the ethics committee of the University of Münster and the Westphalian Chamber of Physicians in Münster. All procedures were done in accordance with the Helsinki Declaration.

Informed consent Written informed consent for participation in the study was obtained from all participants.

References

- Ackenheil M, Stotz-Ingenlath G, Dietz-Bauer R, Vossen A (1999) Mini International Neuropsychiatric Interview (German version 5.0.0, DSM-IV). Psychiatric University Clinic, Munich
- Alpaydin E (2010) Introduction to machine learning. The MIT Press, Cambridge
- Anderson JS, Ferguson MA, Lopez-Larson M, Yurgelun-Todd D (2011) Reproducibility of single-subject functional connectivity measurements. *AJNR Am J Neuroradiol* 32:548–555
- Andrade L, Caraveo-Anduaga JJ, Berglund P, Bijl RV, De Graaf R, Vollebergh W, Dragomirecka E, Kohn R, Keller M, Kessler RC, Kawakami N, Kilic C, Offord D, Ustun TB, Wittchen HU (2003) The epidemiology of major depressive episodes: results from the International Consortium of Psychiatric Epidemiology (ICPE) Surveys. *Int J Methods Psychiatr Res* 12:3–21
- Arbabshirani MR, Plis S, Sui J, Calhoun VD (2017) Single subject prediction of brain disorders in neuroimaging: promises and pitfalls. *Neuroimage* 145 (part B):137–165. doi:10.1016/j.neuroimage.2016.02.079
- Atluri G, Padmanabhan K, Fang G, Steinbach M, Petrella JR, Lim K, Macdonald A 3rd, Samatova NF, Doraiswamy PM, Kumar V (2013) Complex biomarker discovery in neuroimaging data: finding a needle in a haystack. *Neuroimage Clin* 3:123–131
- Barkhof F, Haller S, Rombouts SA (2014) Resting-state functional mr imaging: a new window to the brain. *Radiology* 272:29–49
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Stat Soc Ser B (Methodol)* 57:289–300
- Bhaumik R, Jenkins LM, Gowins JR, Jacobs RH, Barba A, Bhaumik DK, Langenecker SA (2016) Multivariate pattern analysis strategies in detection of remitted major depressive disorder using resting state functional connectivity. *Neuroimage Clin*. doi:10.1016/j.nicl.2016.02.018
- Birn RM, Molloy EK, Patriat R, Parker T, Meier TB, Kirk GR, Nair VA, Meyerand ME, Prabhakaran V (2013) The effect of scan length on the reliability of resting-state fMRI connectivity estimates. *Neuroimage* 83:550–558
- Bowman FD, Drake DF, Huddleston DE (2016) Multimodal imaging signatures of Parkinson’s disease. *Front Neurosci* 10:131
- Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafò MR (2013) Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci* 14:365–376
- Calhoun VD, Sui J (2016) Multimodal fusion of brain imaging data: A key to finding the missing link(s) in complex mental illness. *Biol Psychiatry Cogn Neurosci Neuroimaging* 1:230–244
- Cao L, Guo S, Xue Z, Hu Y, Liu H, Mwansisya TE, Pu W, Yang B, Liu C, Feng J, Chen EY, Liu Z (2014) Aberrant functional connectivity for diagnosis of major depressive disorder: a discriminant analysis. *Psychiatry Clin Neurosci* 68:110–119
- Castellanos FX, Di Martino A, Craddock RC, Mehta AD, Milham MP (2013) Clinical applications of the functional connectome. *Neuroimage* 80:527–540
- Chang C, Lin C (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2:27:1–27:27
- Chao-Gan Y, Yu-Feng Z (2010) DPARSF: a MATLAB toolbox for “pipeline” data analysis of resting-state fMRI. *Front Syst Neurosci* 4:13
- Chu C, Hsu AL, Chou KH, Bandettini P, Lin C, Alzheimer’s Disease Neuroimaging Initiative (2012) Does feature selection improve classification accuracy? Impact of sample size and feature selection on classification using anatomical magnetic resonance images. *Neuroimage* 60:59–70
- Craddock RC, James GA, Holtzheimer PE 3rd, Hu XP, Mayberg HS (2012) A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum Brain Mapp* 33:1914–1928
- Craddock RC, Holtzheimer PE 3rd, Hu XP, Mayberg HS (2009) Disease state prediction from resting state functional connectivity. *Magn Reson Med* 62:1619–1628
- Damoiseaux JS, Beckmann CF, Arigita EJ, Barkhof F, Scheltens P, Stam CJ, Smith SM, Rombouts SA (2008) Reduced resting-state brain activity in the “default network” in normal aging. *Cereb Cortex* 18:1856–1864
- Dannlowski U, Ohrmann P, Konrad C, Domschke K, Bauer J, Kugel H, Hohoff C, Schoning S, Kersting A, Baune BT, Mortensen LS, Arolt V, Zwieterlood P, Deckert J, Heindel W, Suslow T (2009) Reduced amygdala-prefrontal coupling in major depression: association with MAOA genotype and illness severity. *Int J Neuropsychopharmacol* 12:11–22
- Ding C, Peng H (2005) Minimum redundancy feature selection from microarray gene expression data. *J Bioinform Comput Biol* 3:185–205
- Dosenbach NU, Nardos B, Cohen AL, Fair DA, Power JD, Church JA, Nelson SM, Wig GS, Vogel AC, Lessov-Schlaggar CN, Barnes KA, Dubis JW, Feczko E, Coalson RS, Pruett JR Jr, Barch DM, Petersen SE, Schlaggar BL (2010) Prediction of individual brain maturity using fMRI. *Science* 329:1358–1361
- Douaud G, Groves AR, Tamnes CK, Westlye LT, Duff EP, Engvig A, Walhovd KB, James A, Gass A, Monsch AU, Matthews PM, Fjell AM, Smith SM, Johansen-Berg H (2014) A common brain network links development, aging, and vulnerability to disease. *Proc Natl Acad Sci USA* 111:17648–17653
- Douaud G, Menke RA, Gass A, Monsch AU, Rao A, Whitcher B, Zamboni G, Matthews PM, Sollberger M, Smith S (2013) Brain microstructure reveals early abnormalities more than two years

- prior to clinical progression from mild cognitive impairment to Alzheimer's disease. *J Neurosci* 33:2147–2155
- Dunlop BW, Binder EB, Cubells JF, Goodman MM, Kelley ME, Kinkead B, Kutner M, Nemeroff CB, Newport DJ, Owens MJ, Pace TW, Ritchie JC, Rivera VA, Westen D, Craighead WE, Mayberg HS (2012) Predictors of remission in depression to individual and combined treatments (PReDICT): study protocol for a randomized controlled trial. *Trials* 13:106
- Fu CH, Steiner H, Costafreda SG (2013) Predictive neural biomarkers of clinical response in depression: a meta-analysis of functional and structural neuroimaging studies of pharmacological and psychological therapies. *Neurobiol Dis* 52:75–83
- Graham J, Salimi-Khorshidi G, Hagan C, Walsh N, Goodyer I, Lennox B, Suckling J (2013) Meta-analytic evidence for neuroimaging models of depression: state or trait? *J Affect Disord* 151:423–431
- Grieve SM, Korgaonkar MS, Etkin A, Harris A, Koslow SH, Wisniewski S, Schatzberg AF, Nemeroff CB, Gordon E, Williams LM (2013) Brain imaging predictors and the international study to predict optimized treatment for depression: study protocol for a randomized controlled trial. *Trials* 14:224
- Guo H, Cheng C, Cao X, Xiang J, Chen J, Zhang K (2014) Resting-state functional connectivity abnormalities in first-onset unmedicated depression. *Neural Regen Res* 9:153–163
- Guyon I, Weston J, Barnhill S, Vapnik V (2002) Gene selection for cancer classification using support vector machines. *Mach Learn* 46:389–422
- Hahn T, Kircher T, Straube B, Wittchen HU, Konrad C, Strohle A, Wittmann A, Pfeleiderer B, Reif A, Arolt V, Lueken U (2015) Predicting treatment response to cognitive behavioral therapy in panic disorder with agoraphobia by integrating local neural information. *JAMA Psychiatry* 72(1):68–74
- Haller S, Lovblad KO, Giannakopoulos P, Van De Ville D (2014) Multivariate pattern recognition for diagnosis and prognosis in clinical neuroimaging: state of the art, current challenges and future trends. *Brain Topogr* 27:329–337
- Hamilton M (1960) A rating scale for depression. *J Neurol Neurosurg Psychiatr* 23:56–62
- Hamilton JP, Farmer M, Fogelman P, Gotlib IH (2015) Depressive rumination, the default-mode network, and the dark matter of clinical neuroscience. *Biol Psychiatry* 78:224–230
- Hermesdorf M, Sundermann B, Feder S, Schwindt W, Minnerup J, Arolt V, Berger K, Pfeleiderer B, Wersching H (2016) Major depressive disorder: findings of reduced homotopic connectivity and investigation of underlying structural mechanisms. *Hum Brain Mapp* 37:1209–1217
- Heun R, Hardt J, Muller H, Maier W (1997) Selection bias during recruitment of elderly subjects from the general population for psychiatric interviews. *Eur Arch Psychiatry Clin Neurosci* 247:87–92
- Hickie IB, Scott J, Hermens DF, Scott EM, Naismith SL, Guastella AJ, Glozier N, McGorry PD (2013) Clinical classification in mental health at the cross-roads: which direction next? *BMC Med* 11:125
- Ioannidis JP (2005) Why most published research findings are false. *PLoS Med* 2:e124
- James G, Witten D, Hastie T (2013) An introduction to statistical learning with applications in R. Springer, New York
- Kaiser RH, Andrews-Hanna JR, Wager TD, Pizzagalli DA (2015) Large-scale network dysfunction in major depressive disorder: a meta-analysis of resting-state functional connectivity. *JAMA Psychiatry* 72(6):603–611
- Kennedy SH, Downar J, Evans KR, Feilottter H, Lam RW, MacQueen GM, Milev R, Parikh SV, Rotzinger S, Soares C (2012) The Canadian biomarker integration network in depression (CAN-BIND): advances in response prediction. *Curr Pharm Des* 18:5976–5989
- Klöppel S, Abdulkadir A, Jack CR Jr, Koutsouleris N, Mourao-Miranda J, Vemuri P (2012) Diagnostic neuroimaging across diseases. *Neuroimage* 61(2):457–463
- Krishnan R (2014) Unipolar depression in adults: epidemiology, pathogenesis, and neurobiology. In: Solomon D (ed) *UpToDate*. Wolters Kluwer Health, Alphen aan de Rijn
- Kupfer DJ, Frank E, Phillips ML (2012) Major depressive disorder: new clinical, neurobiological, and treatment perspectives. *Lancet* 379:1045–1055
- Lener MS, Iosifescu DV (2015) In pursuit of neuroimaging biomarkers to guide treatment selection in major depressive disorder: a review of the literature. *Ann N Y Acad Sci* 1344:50–65
- Li L, Rakitsch B, Borgwardt K (2011) ccSVM: correcting support vector machines for confounding factors in biological data classification. *Bioinformatics* 27:i342–i348
- Lord A, Horn D, Breakspear M, Walter M (2012) Changes in community structure of resting state functional connectivity in unipolar depression. *PLoS One* 7:e41282
- Lui S, Zhou XJ, Sweeney JA, Gong Q (2016) Psychoradiology: the frontier of neuroimaging in psychiatry. *Radiology* 281:357–372
- Ma Q, Zeng LL, Shen H, Liu L, Hu D (2013) Altered cerebellar-cerebral resting-state functional connectivity reliably identifies major depressive disorder. *Brain Res* 1495:86–94
- Marchetti I, Koster EH, Sonuga-Barke EJ, De Raedt R (2012) The default mode network and recurrent depression: a neurobiological model of cognitive risk factors. *Neuropsychol Rev* 22:229–251
- Margulies DS, Böttger J, Long X, Lv Y, Kelly C, Schafer A, Goldhahn D, Abbushi A, Milham MP, Lohmann G, Villringer A (2010) Resting developments: a review of fMRI post-processing methodologies for spontaneous brain activity. *MAGMA* 23:289–307
- McCabe C, Mishor Z (2011) Antidepressant medications reduce subcortical-cortical resting-state functional connectivity in healthy volunteers. *Neuroimage* 57:1317–1323
- Mierswa I, Wurst M, Klinkenberg R, Scholz M, Euler T (2006) YALE: rapid prototyping for complex data mining tasks. In: *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp 935–940
- Murphy K, Birn RM, Bandettini PA (2013) Resting-state fMRI confounds and cleanup. *Neuroimage* 80:349–359
- Mwangi B, Tian TS, Soares JC (2014) A review of feature reduction techniques in neuroimaging. *Neuroinformatics* 12(2):229–244
- Nejad AB, Fossati P, Lemogne C (2013) Self-referential processing, rumination, and cortical midline structures in major depression. *Front Hum Neurosci* 7:666
- Orru G, Pettersson-Yeo W, Marquand AF, 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
- Patel MJ, Khalaf A, Aizenstein HJ (2016) Studying depression using imaging and machine learning methods. *NeuroImage Clin* 10:115–123
- Pereira F, Mitchell T, Botvinick M (2009) Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage* 45:S199–S209
- Phillips ML, Chase HW, Sheline YI, Etkin A, Almeida JR, Deckersbach T, Trivedi MH (2015) Identifying predictors, moderators, and mediators of antidepressant response in major depressive disorder: neuroimaging approaches. *Am J Psychiatry* 172:124–138
- Plis SM, Hjeltn DR, Salakhutdinov R, Allen EA, Bockholt HJ, Long JD, Johnson HJ, Paulsen JS, Turner JA, Calhoun VD (2014) Deep learning for neuroimaging: a validation study. *Front Neurosci* 8:229
- Pyka M, Hahn T, Heider D, Krug A, Sommer J, Kircher T, Jansen A (2013) Baseline activity predicts working memory load of preceding task condition. *Hum Brain Mapp* 34:3010–3022

- Qin J, Shen H, Zeng LL, Jiang W, Liu L, Hu D (2015) Predicting clinical responses in major depression using intrinsic functional connectivity. *NeuroReport* 26:675–680
- Radloff LS (1977) The CES-D scale: a self-report depression scale for research in the general population. *Appl Psychol Meas* 1:385–401
- Ramasubbu R, Brown MR, Cortese F, Gaxiola I, Goodyear B, Greenshaw AJ, Dursun SM, Greiner R (2016) Accuracy of automated classification of major depressive disorder as a function of symptom severity. *Neuroimage Clin* 12:320–331
- Rosa MJ, Portugal L, Hahn T, Fallgatter AJ, Garrido MI, Shawe-Taylor J, Mourao-Miranda J (2015) Sparse network-based models for patient classification using fMRI. *Neuroimage* 105:493–506
- Salimi-Khorshidi G, Douaud G, Beckmann CF, Glasser MF, Griffanti L, Smith SM (2014) Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* 90:449–468
- Sarraf S, Tofghi G (2016) Classification of Alzheimer's disease using fMRI data and deep learning convolutional neural networks. *arXiv* 1603.08631
- Schmaal L, Marquand AF, Rhebergen D, van Tol M, Ruhe HG, van der Wee NJ, Veltman DJ, Penninx BWJH (2014) Predicting the naturalistic course of major depressive disorder using clinical and multimodal neuroimaging information: a multivariate pattern recognition study. *Biol Psychiatry* 78(4):278–286
- Schneider B, Prvulovic D (2013) Novel biomarkers in major depression. *Curr Opin Psychiatry* 26:47–53
- Schouten TM, Koini M, de Vos F, Seiler S, van der Grond J, Lechner A, Hafkemeijer A, Moller C, Schmidt R, de Rooij M, Rombouts SA (2016) Combining anatomical, diffusion, and resting state functional magnetic resonance imaging for individual classification of mild and moderate Alzheimer's disease. *Neuroimage Clin* 11:46–51
- Schowe B (2011) Feature selection for high-dimensional data with RapidMiner. In: *Proceedings of the 2nd RapidMiner Community Meeting and Conference (RCOMM 2011)*
- Schrouff J, Rosa MJ, Rondina JM, Marquand AF, Chu C, Ashburner J, Phillips C, Richiardi J, Mourao-Miranda J (2013) PRoNTo: pattern recognition for neuroimaging toolbox. *Neuroinformatics* 11:319–337
- Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, Mackay CE, Filippini N, Watkins KE, Toro R, Laird AR, Beckmann CF (2009) Correspondence of the brain's functional architecture during activation and rest. *Proc Natl Acad Sci USA* 106:13040–13045
- Song XW, Dong ZY, Long XY, Li SF, Zuo XN, Zhu CZ, He Y, Yan CG, Zang YF (2011) REST: a toolkit for resting-state functional magnetic resonance imaging data processing. *PLoS One* 6:e25031
- Storey JD (2002) A direct approach to false discovery rates. *J R Stat Soc Ser B (Statistical Methodology)* 64:479–498
- Storey JD, Tibshirani R (2003) Statistical significance for genome-wide studies. *Proc Natl Acad Sci USA* 100:9440–9445
- Stuhrmann A, Suslow T, Dannlowski U (2011) Facial emotion processing in major depression: a systematic review of neuroimaging findings. *Biol Mood Anxiety Disord* 1:10–5380-1-10
- Sundermann B, Herr D, Schwindt W, Pfeleiderer B (2014a) Multivariate classification of blood oxygen level-dependent FMRI data with diagnostic intention: a clinical perspective. *AJNR Am J Neuroradiol* 35:848–855
- Sundermann B, Olde Lütke Beverborg M, Pfeleiderer B (2014b) Toward literature-based feature selection for diagnostic classification: a meta-analysis of resting-state fMRI in depression. *Front Hum Neurosci* 8:692
- Teismann H, Wersching H, Nagel M, Arolt V, Heindel W, Baune BT, Wellmann J, Hense HW, Berger K (2014) Establishing the bidirectional relationship between depression and subclinical arteriosclerosis—rationale, design, and characteristics of the BiDirect Study. *BMC Psychiatry* 14:174
- Teuber A, Sundermann B, Kugel H, Schwindt W, Heindel W, Minnerup J, Dannlowski U, Berger K, Wersching H (2017) MR imaging of the brain in large cohort studies—feasibility report of the population- and patient-based BiDirect study. *Eur Radiol* 27:231–238
- Trivedi M, McGrath PJ, Fava M, Parsey RV, Kurian BT, Phillips ML, Oquendo M, Bruder G, Pizzagalli DA, Toups M, Cooper C, Adams P, Weyandt S, Morris DW, Grannemann BD, Ogden RT, Buckner R, McInnis M, Kramer HC, Petkova E, Carmody T, Weissmann MM (2016) Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): rationale and design. *J Psychiatr Res* 78:11–23. doi:10.1016/j.jpsychires.2016.03.001
- Ugurbil K, Xu J, Auerbach EJ, Moeller S, Vu AT, Duarte-Carvajalino JM, Lenglet C, Wu X, Schmitter S, Van de Moortele PF, Strupp J, Sapiro G, De Martino F, Wang D, Harel N, Garwood M, Chen L, Feinberg DA, Smith SM, Miller KL, Sotiropoulos SN, Jbabdi S, Andersson JL, Behrens TE, Glasser MF, Van Essen DC, Yacoub E, WU-Minn HCP Consortium (2013) Pushing spatial and temporal resolution for functional and diffusion MRI in the human connectome project. *Neuroimage* 80:80–104
- van den Heuvel MP, Hulshoff Pol HE (2010) Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur Neuropsychopharmacol* 20:519–534
- Van Dijk KR, Hedden T, Venkataraman A, Evans KC, Lazar SW, Buckner RL (2010) Intrinsic functional connectivity as a tool for human connectomics: theory, properties, and optimization. *J Neurophysiol* 103:297–321
- van Waarde JA, Scholte HS, van Oudheusden LJ, Verwey B, Denys D, van Wingen GA (2015) A functional MRI marker may predict the outcome of electroconvulsive therapy in severe and treatment-resistant depression. *Mol Psychiatry* 20(5):609–614
- Vapnik VN (2000) *The nature of statistical learning theory*, 2nd edn. Springer-Verlag, New York
- Waites AB, Stanislavsky A, Abbott DF, Jackson GD (2005) Effect of prior cognitive state on resting state networks measured with functional connectivity. *Hum Brain Mapp* 24:59–68
- Wee CY, Yap PT, Zhang D, Denny K, Browndyke JN, Potter GG, Welsh-Bohmer KA, Wang L, Shen D (2012) Identification of MCI individuals using structural and functional connectivity networks. *Neuroimage* 59:2045–2056
- Williams LM, Rush AJ, Koslow SH, Wisniewski SR, Cooper NJ, Nemeroff CB, Schatzberg AF, Gordon E (2011) International study to predict optimized treatment for depression (iSPOT-D), a randomized clinical trial: rationale and protocol. *Trials* 12:4
- Wolfers T, Buitelaar JK, Beckmann CF, Franke B, Marquand AF (2015) From estimating activation locality to predicting disorder: a review of pattern recognition for neuroimaging-based psychiatric diagnostics. *Neurosci Biobehav Rev* 57:328–349
- Yu Y, Shen H, Zeng LL, Ma Q, Hu D (2013) Convergent and divergent functional connectivity patterns in schizophrenia and depression. *PLoS One* 8:e68250
- Zeng LL, Shen H, Liu L, Wang L, Li B, Fang P, Zhou Z, Li Y, Hu D (2012) Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis. *Brain* 135:1498–1507
- Zeng LL, Shen H, Liu L, Hu D (2014) Unsupervised classification of major depression using functional connectivity MRI. *Hum Brain Mapp* 35:1630–1641
- Zhang D, Raichle ME (2010) Disease and the brain's dark energy. *Nat Rev Neurol* 6:15–28
- Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *J R Stat Soc Ser B (Statistical Methodology)* 67:301–320