Received: 30 April 2018 | Revised: 19 July 2018 | Accepted: 21 July 2018

DOI: 10.1111/cns.13048

REVIEW ARTICLE

WILEY CNS Neuroscience & Therapeutics

Machine learning in major depression: From classification to treatment outcome prediction

Shuang Gao[1,2] | Vince D. Calhoun[3,4] | Jing Sui[1,2,5]

[1]Brainnetome Center and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, China

[2]University of Chinese Academy of Sciences, Beijing, China

[3]The Mind Research Network, Albuquerque, New Mexico

[4]Department of Electrical and Computer Engineering, The University of New Mexico, Albuquerque, New Mexico

[5]CAS Centre for Excellence in Brain Science and Intelligence Technology, Institute of Automation, Chinese Academy of Sciences, Beijing, China

Correspondence: Jing Sui, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, 95 Zhongguancun East Road, 100190, Beijing, China (kittysj@gmail.com).

Funding information
National High-Tech Development Plan (863), Grant/Award Number: 2015AA020513; NIH Grant, Grant/Award Number: 1R01MH094524, P20GM103472 and R01EB005846; Strategic Priority Research Program of the Chinese Academy of Sciences, Grant/Award Number: XDBS01000000; "100 Talents Plan" of Chinese Academy of Sciences, the Chinese Natural Science Foundation, Grant/Award Number: 61773380 and 81471367

Abstract

Aims: Major depression disorder (MDD) is the single greatest cause of disability and morbidity, and affects about 10% of the population worldwide. Currently, there are no clinically useful diagnostic biomarkers that are able to confirm a diagnosis of MDD from bipolar disorder (BD) in the early depressive episode. Therefore, exploring translational biomarkers of mood disorders based on machine learning is in pressing need, though it is challenging, but with great potential to improve our understanding of these disorders.

Discussions: In this study, we review popular machine-learning methods used for brain imaging classification and predictions, and provide an overview of studies, specifically for MDD, that have used magnetic resonance imaging data to either (a) classify MDDs from controls or other mood disorders or (b) investigate treatment outcome predictors for individual patients. Finally, challenges, future directions, and potential limitations related to MDD biomarker identification are also discussed, with a goal of offering a comprehensive overview that may help readers to better understand the applications of neuroimaging data mining in depression.

Conclusions: We hope such efforts may highlight the need for an urgently needed paradigm shift in treatment, to guide personalized optimal clinical care.

KEYWORDS
classification, machine learning, magnetic resonance imaging, major depressive disorder, review

1 | INTRODUCTION

Major depressive disorder (MDD) is a highly prevalent psychiatric disorder with a significant effect on quality of life and socioeconomic burden.[1] The diagnosis of MDD often depends on criteria from the Diagnostic and Statistical Manual of Mental Disorders (DSM) and treatment response.[2] Due to the overlapping phenotypes across mental disorders as well as the heterogeneity within disorders such as MDD, clinical diagnoses are often not as well-defined as in research protocols. Consequently, patients with mood disorder sometimes have to endure the wrong drug trial or multiple trials before receiving a final diagnosis.[3] For situations in which the DSM classification is unclear and the subjective clinical impression is confusing, an effective diagnostic tool using, for example, objective brain imaging measurements is greatly needed.

Neuroimaging provides noninvasive measurements of brain function and structure, which can serve as a powerful tool for investigating discriminative biomarkers.[4,5] Specifically, brain neuroanatomy is intrinsically complex and heterogeneous, which further complicates functional connectivity in patients with mental illnesses.[6] Considering the high-dimensional imaging data quite often includes a limited number of samples, determining an effective and

optimal approach to diagnose mood disorder is particularly challenging.[7] Studies discriminating major depressive disorders from healthy control (HC) or other mood disorders have been performed using several neuroimaging techniques, including magnetic resonance imaging (MRI), positron emission tomography (PET), magneto-encephalography (MEG), and electro-encephalography (EEG).[8,9] Among which, MRI-related techniques such as functional MRI (fMRI), structural MRI (sMRI), and diffusion MRI (diffusion tensor images, DTI) show benefits of providing multiple perspective on brain function, structure, and their connectivity maps. These diverse brain imaging characteristics offer a great opportunity for researchers to unravel the secrets of the complex neuromechanism underlying depression.[10–12] Beyond the group-level analyses which are often performed,[13–16] there has been growing interest in using machine-learning (ML) techniques to identify phenotypes in a way that is clinically meaningful and feasible for translation into clinical diagnosis or prognosis,[17,18] for example, (a) to predict response to currently available treatments or (b) to identify more specific targets for novel interventions.[19,20]

In this selective review, we focus on machine-learning-based classification and prediction studies of MDD which utilize features derived from MRI data. First, based on a specific screening method, we selected 63 MRI-based machine-learning articles with MDD samples and surveyed the popular machine-learning methods implemented in these studies. Next, we highlight some representative studies on mood disorder discrimination, for example, MDD vs bipolar disorder (BD), and individualized prediction of treatment outcomes for MDD. Common biases are discussed and suggestions are provided. Finally, we discussed future directions for potential biomarker identification of MDD disorders. Approaches of mining big data focusing on classification and treatment strategies which are based on biological information rather than the clinical manifestation have the greatest potential to move the field forward.

2 | RESEARCH OVERVIEW

2.1 | Screening method

Studies were included if they focused on classification (including treatment prediction) between individuals with MDD and healthy controls (or other brain disorders) using machine-learning methods and employed magnetic resonance imaging as the data acquisition access. Figure 1I shows the screening method diagram called PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses).[21] Relevant articles were identified from searches in PubMed covering publications between January 2000 and December 2017, using the search terms, "depress*," "MDD," "MRI," "fMRI," "sMRI," "DTI," "magnetic resonance imaging," "neuroimaging," "classif*," "diagno*," "predict*," "distinguish*," "discriminat*," "machine learning," both in isolation and in combination. A total of 2045 articles were identified by the above search. Then, additional articles were identified through the reference lists of these papers to ensure that no studies of significance were omitted from this review, resulting in another 82 articles. After removing duplicates, 1980 articles remained. Furthermore, 1874 were excluded during screening of the title and abstract and a further 40 were excluded during full-text screening. Finally, 66 MDD studies were selected, and we summarize their main findings below.

2.2 | Summary of metrics

Figure 1 indicates several key aspects of our survey. The number of papers published on this topic in each year from 2000 to 2017 is displayed on Figure 1A. Obviously, the publication number keeps growing and increases sharply in 2017. Figure 1B indicates the number of studies fallen within different classification groups, for example, MDD vs HC or MDD subtypes. It is clear that MDD vs HC classification is the largest category, followed by MDD vs BD. Predictive studies on MDD treatment outcome occur less frequently than classification studies. Figure 1C shows the proportion of popular machine-learning methods used in these studies. Support vector machines (SVM) remain the most prevalent method choice, but other ML methods have also been applied to MDD such as gaussian process classifier (GPC), linear discriminant analysis (LDA), and decision tree (DT), as well as more recent deep learning models. Figure 1D demonstrates the distribution of reported accuracy for 5 ml methods; SVM performance shows a large variability, which may be due to different sample sizes, whereas some uncommon methods show promising performance for specific cases. Furthermore, the proportion of different MRI modalities as well as the reported accuracy of each modality are shown in Figure 1E,F. Most studies still focused on using features of fMRI and sMRI (22 resting-state fMRI; 18 task-related fMRI; 21 sMRI), while some studies have begun to explore the discriminating ability of DTI (8 DTI) in spite of applying multimodal MRI features in one study. On the whole, rsfMRI data exhibit higher accuracy than other modalities. Figure 1G illustrates the distribution of reported sample sizes for each cross-validation (CV) methods including leave-one-out CV (LOOCV), 10-fold CV or others, and almost all studies with LOOCV had sample size smaller than 100 while the studies with 10-fold CV had a bigger mean sample size. Note that there is one special case which uses LOOCV with the largest sample size in our survey.[22] Figure 1H shows the overall accuracy against the total sample size used in studies. Most studies had a smaller sample size and only one had more than 700,[22] which raises a urgent need for including larger sample sizes for the study of MDD in machine-learning studies.

2.3 | Machine-learning pipeline

Figure 1J summarizes the most common machine-learning pipeline of MDD diagnosis and prediction using MRI data. After data preprocessing, the pipelines vary greatly but typically include the following steps: feature reduction, model training, classification, and performance evaluation.

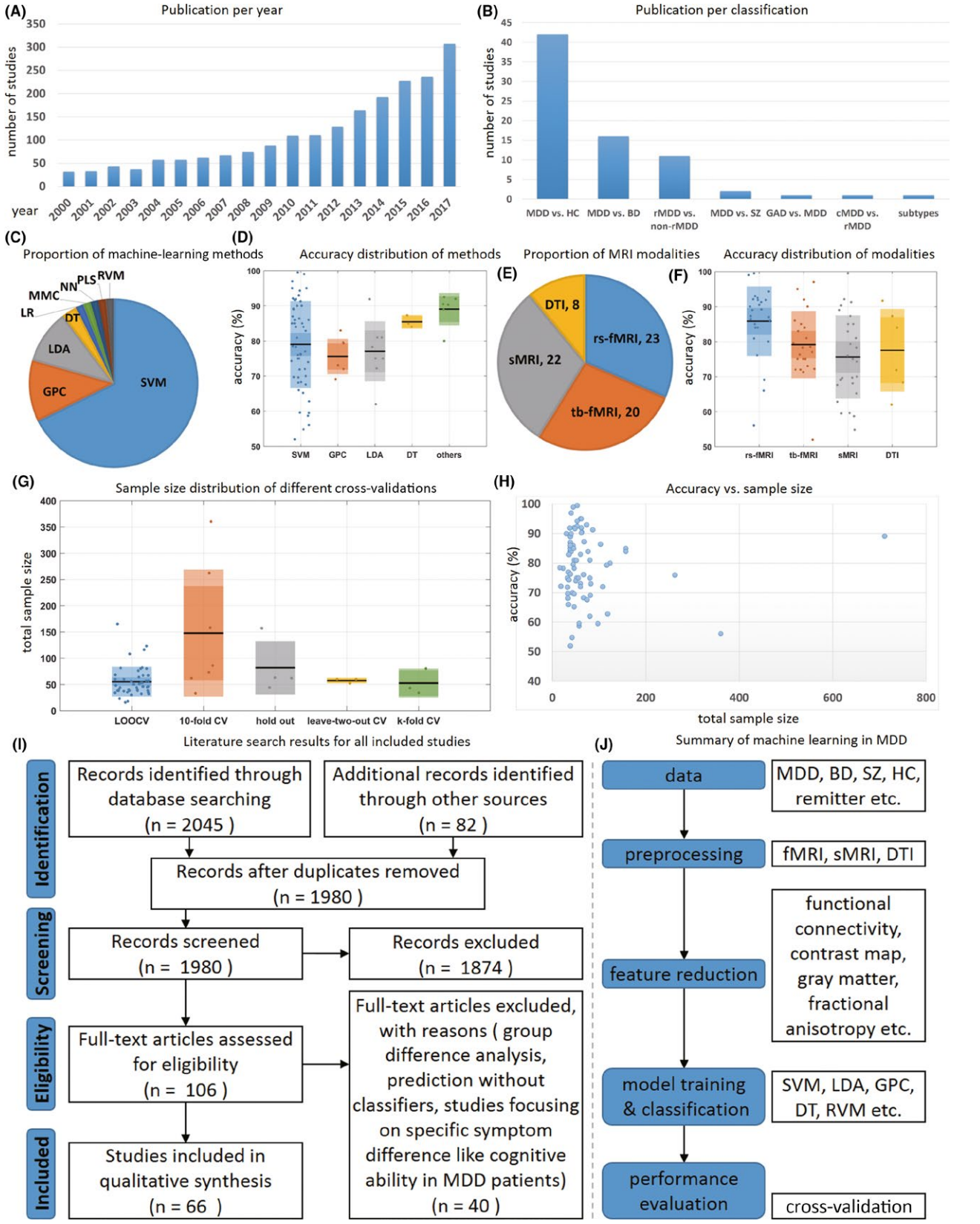


FIGURE 1 Visual summary of the selected MDD studies. A, Total number of papers before screening. B, Number of publications per group classification. C, Proportion of machine-learning methods used. D, Boxplot of accuracy based on five methods. E, Proportion of MRI modalities used. F, Accuracy based on modalities. G, Boxplot of sample size based on different cross-validation method. H, Scatter plot of overall reported accuracy vs the total sample size. I, Literature search results for each screening steps.[21] J, Summary of steps in MRI machine learning

2.3.1 | Feature reduction

Feature reduction methods are essential to high-dimensional data which is a common problem in neuroimaging.[23] A limited number of the most relevant features warrant a more accurate classification model. These methods can be primarily categorized into feature selection and feature extraction. Feature selection is performed when supervised methods select the most discriminant features with the help of the labels of training data to reduce the noise. One strategy is to use prior knowledge to decrease dimensionality. Feature extraction occurs when methods project the original high-dimensional data into a lower dimension while maintaining its discriminative ability, and projection matrices are computed from training data. One typical example is principle component analysis (PCA). Both of these feature selection and feature extraction methods should only be conducted on training dataset (or the training groups assigned in cross-validation) to avoid biased results. Besides, some proposed approaches[24,25] provide an intermediate solution by adopting geometric distance in feature space between different groups in the training data.[26]

2.3.2 | Model training

In the training phase, for a supervised approach, a model is optimized using labeled data to find a discriminant "decision function" or "hyperplane" distinguishing between different groups (eg, depression patients and healthy controls).[3,27] The parameters of the model are optimized to maximally discriminate one group from another. Cross-validation is usually used to generalize the training process. There are several types of cross-validation including k-fold, leave-one-out, and holdout. For k-fold CV, the training data are divided into k equal sized groups. Then, each one of k groups is treated as testing data and reiterated for k-iterations. The latter two techniques can be considered as variants of k-fold CV. Holdout approach is performed on data with large sample size (k = 1), while leave-one-out is used on data with small sample size (k = sample size).[28,29]

2.3.3 | Classification

In the classification phase, the trained model is used to predict the label for new, previously unseen observations. For an unbiased generalization, it is important that the testing data do not overlap with training data.[30] The new data have to be preprocessed in the same way as the training data, and the same feature reduction method is applied with optimized parameters obtained from the training phase. In cases where independent testing data are not available due to limited samples, a nested cross-validation framework can be conducted to estimate the performance of the model.[31–35] The most commonly used classifier in our survey is SVM because of its promising results in neuroimaging.[36]

2.3.4 | Performance evaluation

The performance of classification-based algorithms can be described by accuracy, sensitivity, specificity, and receiver operating characteristic (ROC) curve (sensitivity as a function of 1-specificity). Accuracy helps evaluate how accurately the model classifies test data. Sensitivity (or recall) refers to the proportion of true positives correctly identified (eg, percentage of true depression patients identified as MDD). In contrast, specificity refers to the proportion of true negatives correctly identified (eg, percentage of healthy people identified as HC). ROC curve illustrates the overall performance of method, usually summarized by the area under the curve (AUC). A confusion matrix, which is an n × n matrix for n labels with one side representing actual labels and the other representing predicted labels, can be useful when group of labeled data is more than two.[37] The confusion matrix also provides information relevant for imbalanced data and computing other performance measures such as precision (positive predictive value), F1-score (harmonic mean of precision and recall), and G-mean (geometric mean of precision and recall).[38] For imbalanced sample size data, balanced sensitivity (recall), specificity, and precision are more desirable than higher overall accuracy and thus such measures are preferred for evaluating the classifier in this case. Other ways of reporting results for imbalanced data include the F1-score and G-mean which help evaluate the performance of the results.

3 | MACHINE LEARNING IN MDD

Machine learning is defined as a group of methods that learn from empirical data to develop training models and make accurate classification from new data.[39] Its advantages in MDD are not confined to diagnosis, but also allow for the prediction of future disease progression, of which the most salient advantage one is its applicability to individual-level analysis. Table 1 summarizes various aspects of 66 selected studies.[3,18–20,22,27,40–99] Most of these works aim at developing computational approaches to discriminate MDD from controls or mood disorder subtypes and trying to develop tools to integrate imaging measurements into clinical practice.

3.1 | Highlighted research

3.1.1 | Classification with brain networks in MDD

A number of studies used graph theory approaches[43,45,46,82] to highlight the disrupted functional and structural brain networks in depression. These connectome-based biomarkers also provide new opportunities to redefine the diagnosis of depression and improve treatment measures by providing important knowledge about the biological mechanisms. Here, we summarize some of the key findings related to functional and structural brain networks in depression. In ref. [100], a review enumerated various brain network features, including altered regional and connectivity patterns of various MRI modalities in depression: ROI-based and voxel-based analysis (fMRI), regional betweenness and degree centralities (sMRI), and white matter structural connectivity (DTI). Reference [86] investigated the whole-brain resting-state functional connectivity patterns of

TABLE 1 Review of classification studies related to major depressive disorders

References	Subjects	Feature	Method	Cross-validation	Accuracy
Classification					
Rubin-Falcone et al (2017)[40]	BD = 26, MDD = 26	GM (sMRI)	SVM	Leave-two-out CV	75.0%
Deng et al (2017)[41]	BD = 31, MDD = 36	FA (DTI)	SVM	LOOCV	68.3%
Gao et al (2017)[3]	BD = 37, MDD = 36	Spatial independent components (rsfMRI)	SVM	10-fold CV	93.0%
Jing et al (2017)[42]	cMDD = 19, rMDD = 19, HC = 19	Hurst exponent (rsfMRI)	SVM	LOOCV	87.0% (cMDD vs HC), 84.0% (rMDD vs HC), 89.0% (cMDD vs rMDD)
Yoshida et al (2017)[43]	MDD = 58, HC = 65	FC (rsfMRI)	PLS	LOOCV	80.0%
Li et al (2017)[44]	BD = 22, MDD = 22	Degree centrality (rsfMRI)	SVM	LOOCV	86.0%
Zhong et al (2017)[45]	MDD = 29, HC = 33 (1st sample); MDD = 46, HC = 57 (2nd sample)	FC (rsfMRI)	SVM	LOOCV	91.9% (1st sample), 86.4% (2nd sample)
Wang et al (2017)[46]	MDD = 31, HC = 29	FC (rsfMRI)	SVM	LOOCV	95.0%
Schnyer et al (2017)[47]	MDD = 25, HC = 25	FA (DTI)	SVM	LOOCV	74.0%
Sundermann et al (2017)[48]	MDD = 180, HC = 180	FC (rsfMRI)	SVM	10-fold CV	45.0%~56.1%
Bhaumik et al (2017)[49]	MDD = 38, HC = 29	FC (rsfMRI)	SVM	LOOCV	76.1%
He et al (2017)[50]	BD = 13, MDD = 40, HC = 33	Functional network connectivity (rsfMRI), GM(sMRI)	SVM	10-fold CV	91.3% (three groups), 99.0% (BD vs MDD)
Bürger et al (2017)[51]	BD = 36, MDD = 36, HC = 36	Contrast maps (task fMRI)	SVM, GPC	LOOCV	72.0%
Hilbert et al (2016)[52]	GAD = 19, MDD = 14, HC = 24	GM (sMRI)	SVM	LOOCV	58.7% (GAD&MDD vs HC), 68.1% (GAD vs MDD)
Drysdale et al (2016)[22]	MDD = 333(4 biotypes), HC = 378	FC (rsfMRI)	SVM	LOOCV	89.2%
Sankar et al (2016)[53]	MDD = 23, HC = 20	GM, WM (sMRI)	SVM	5-fold CV	70.0%
Frangou et al (2016)[54]	BD = 30, MDD = 30	Contrast maps (task fMRI)	GPC	leave-two-out CV	73.1%
Ramasubbu et al (2016)[55]	MDD = 15, HC = 19	Spatial independent components (rsfMRI)	SVM	5-fold CV	66.0%
Yang et al (2016)[56]	MDD = 16, HC = 16	Contrast maps (task fMRI)	SVM	LOOCV	75.0%
Rive et al (2016)[57]	MDDr = 23, BDr = 26, MDDd = 22, BDd = 10	GM (sMRI), Spatial independent components (rsfMRI)	GPC	LOOCV	69.1%
Jie et al (2015)[27]	BD = 21, MDD = 25	GM (sMRI), fALFF (rsfMRI)	SVM	LOOCV	92.1%
Foland-Ross et al (2015)[58]	MDD = 18, HC = 15	CTH (sMRI)	SVM	10-fold CV	69.7%
Sacchet et al (2015)[59]	BD = 40, MDD = 57, HC = 61	GM (sMRI)	SVM	10-fold CV	59.5% (BD vs MDD), 62.8% (MDD vs HC)
Sacchet et al (2015)[60]	MDD = 14, HC = 18	Graph metric of WM connectivity (DTI)	SVM	LOOCV	71.9%
Sato et al (2015)[61]	MDD = 25, HC = 21	Contrast maps (task fMRI)	LDA	LOOCV	78.3%

(Continues)

TABLE 1 (Continued)

References	Subjects	Feature	Method	Cross-validation	Accuracy
Johnston et al (2015)[62]	MDD = 20, HC = 21	GM (sMRI)	SVM	LOOCV	85.0%
Johnston et al (2015)[63]	MDD = 19, HC = 21	Contrast maps (task fMRI)	SVM	LOOCV	97.0% (hippocampus), 84.0% (striatum)
Koutsouleris et al (2015)[64]	MDD = 104, SZ = 158	GM (sMRI)	SVM	10-fold CV	76.0%
Shimizu et al (2015)[65]	MDD = 31, HC = 31	Contrast maps (task fMRI)	gLASSO, SVM	10-fold CV	92.0% (gLASSO), 95.0% (SVM)
Fung et al (2015)[66]	BD = 16, MDD = 19	CTH and surface area (sMRI)	SVM	LOOCV	74.3%
Rosa et al (2015)[67]	MDD = 19, HC = 19	FC (task fMRI)	SVM	LOOCV	85.0%
Patel et al (2015)[18]	MDD = 33, HC = 35	rsfMRI, sMRI, DTI	DT	LOOCV	87.3%
Redlich et al (2014)[68]	BD = 58, MDD = 58	GM (sMRI)	GPC	LOOCV	79.3%
Cao et al (2014)[69]	MDD = 39, HC = 37	FC (rsfMRI)	SVM	LOOCV	84.0%
MacMaster et al (2014)[70]	BD = 14, MDD = 32	GM (sMRI)	LDA	N/A	81.0%
Zeng et al (2014)[71]	MDD = 24, HC = 29	FC (rsfMRI)	MMC	LOOCV	92.5% (clustering), 92.5% (classification)
Rondina et al (2014)[72]	MDD = 30, HC = 30	Voxel intensity (task fMRI)	SVM	leave-two-out CV	72.0%
Guo et al (2014)[73]	MDD = 36, HC = 27	FC (rsfMRI)	NN	hold out	90.5%
Serpa et al (2014)[74]	BD = 23, MDD = 19, HC = 38	GM,WM, and ventricular RAVENS maps (sMRI)	SVM	LOOCV	54.8% (BD vs MDD), 59.6% (MDD vs HC)
Habes et al (2013)[75]	MDD = 9, HC = 9	Contrast maps (task fMRI)	LDA	LOOCV	72.2%
Wei et al (2013)[76]	MDD = 20, HC = 20	Spatial independent components (rsfMRI)	SVM	LOOCV	90.0%
Grotegerd et al (2013)[77]	BD = 22, MDD = 22	Contrast maps (task fMRI)	GPC	LOOCV	79.6%
Yu et al (2013)[78]	MDD = 19, SZ = 32	FC (rsfMRI)	SVM	LOOCV	80.9%
Modinos et al (2013)[79]	MDD = 17, HC = 17	Contrast maps (task fMRI)	SVM	LOOCV	77.0%
Ma et al (2013)[80]	MDD = 19, HC = 18	ReHo (rsfMRI)	LDA	LOOCV	91.9%
Grotegerd et al (2013)[81]	MDD = 10, BD = 10, HC = 10	Contrast maps (task fMRI)	SVM, GPC	LOOCV	90.0%
Fang et al (2012)[82]	MDD = 22, HC = 26	Anatomical connectivity (DTI)	SVM	LOOCV	91.7%
Mwangi et al (2012)[83]	MDD = 30, HC = 32	GM (sMRI)	RVM	hold out	90.3%
Lord et al (2012)[84]	MDD = 22, HC = 22	FC (rsfMRI)	SVM	hold out	99.0%
Liu et al (2012)[85]	TRD = 18, TSD = 17, HC = 17	GM, WM (sMRI)	SVM	LOOCV	82.9%
Zeng et al (2012)[86]	MDD = 24, HC = 29	FC (rsfMRI)	SVM	LOOCV	94.3%
Mourão-Miranda et al (2011)[87]	MDD = 19, HC = 19	Contrast maps (task fMRI)	SVM	LOOCV	52.0% (true positive)
Hahn et al (2011)[88]	MDD = 30, HC = 30	Contrast maps (task fMRI)	GPC	LOOCV	83.0%
Nouretdinov et al (2011)[89]	MDD = 19, HC = 19	Contrast maps (task fMRI)	SVM	LOOCV	76.3%

(Continues)

TABLE 1 (Continued)

References	Subjects	Feature	Method	Cross-validation	Accuracy
Costafreda et al (2009)[90]	MDD = 37, HC = 37	GM (sMRI)	SVM	LOOCV	67.6%
Fu et al (2008)[91]	MDD = 19, HC = 19	Contrast maps (task fMRI)	SVM	LOOCV	86.0%
Prediction					
Jiang et al (2017)[19]	rMDD = 27, non-rMDD = 11	GM (sMRI)	LR	LOOCV	89.0%
Redlich et al (2016)[20]	MDD responder = 13, non-responder = 10	GM (sMRI)	SVM	LOOCV	78.3%
Lythe et al (2015)[92]	rMDD = 31, non-rMDD = 25	FC (task fMRI)	LDA	LOOCV	75.0%
Korgaonkar et al (2015)[93]	rMDD = 54, non-rMDD = 103	Volume (sMRI), FA (DTI)	DT	hold out	85.0% (sMRI), 84.0%(DTI)
Williams et al (2015)[94]	MDD responder = 48, non-responder = 32	Contrast maps (task fMRI)	LDA	LOOCV	75.0% (happy), 81.0% (sad)
Schmaal et al (2015)[95]	rMDD = 23, non-rMDD = 59	Contrast maps (task fMRI)	GPC	LOOCV	73.0%
van Waarde et al (2014)[96]	rMDD = 25, non-rMDD = 20	Spatial independent components (rsfMRI)	SVM	LOOCV	84.0% (sensitivity), 85.0% (specificity)
Korgaonkar et al (2014)[97]	rMDD = 37, non-rMDD = 43	FA (DTI)	LDA	k-fold CV	62.0%
Gong et al (2011)[98]	rMDD = 23, non-rMDD = 23	GM, WM (sMRI)	SVM	LOOCV	69.6% (GM), 65.2%(WM)
Costafreda et al (2009)[99]	rMDD = 9, non-rMDD = 7	Contrast maps (task fMRI)	SVM	LOOCV	71.0% (sensitivity), 86.0% (specificity)

BD, bipolar disorder; MDD, major depressive disorder; cMDD, current MDD; rMDD, remitted MDD; GAD, generalized anxiety disorder; SZ, schizophrenia; TRD, treatment-resistant depression; TSD, treatment-sensitive depression; GM, gray matter; MRI, magnetic resonance imaging; DTI, diffusion tensor images; FC, functional connectivity; WM, white matter; fALFF, fractional amplitude of low-frequency fluctuation; CTH, cortical thickness; ReHo, regional homogeneity; GPC, gaussian process classifier; LDA, linear discriminant analysis; gLASSO, group least absolute shrinkage and selection operator; CV, cross-validation; LOOCV, leave-one-out cross-validation; FA, fractional anisotropy; RAVENS, regional analysis of volumes examined in normalized space; LR, linear regression; PLS, partial least squares regression; DT, decision tree; MMC, maximum margin clustering; NN, neural network; RVM, relevance vector machine; BDd, bipolar disorder, depressed state; BDr, bipolar disorder, remitted state; MDDd, major depressive disorder, depressed state; MDDr, major depressive disorder, remitted state.

depressed patients to identify major depressive individuals from healthy controls and achieved 100% sensitivity. The most discriminating functional connections were located within or across the default mode network, affective network, visual cortical areas, and cerebellum, which may play important roles in the pathological mechanism of this disorder.

3.1.2 | Prediction of treatment response in MDD

Altered network activity at rest has been explored as a potential biomarker for predicting treatment outcomes. As shown in ref. [22], four distinct MDD neurophysiological biotypes, characterized by distinct patterns of limbic and frontostriatal functional connectivity, were defined using fMRI. These biotypes were associated with distinct profiles of clinical symptoms; for example, biotype 1, which responded best to repetitive transcranial magnetic stimulation (rTMS) therapy, was associated with high levels of fatigue and low anhedonia. Similarly, electroconvulsive therapy (ECT) is a popular treatment for depression patients. Several studies have explored biomarkers that potentially predict the response to ECT. One of the studies investigated whether gray matter (GM) volume changes are able to predict ECT response. Support vector regression was performed before treatment and supplemented with univariate analysis of the Hamilton Depression Rating Scale score (HDRS), yielding a successful prediction of ECT response and a significant prediction of relative reduction in the HDRS.[20] Another study predicted post-ECT depressive remission status using pre-ECT GM in MDD patients and validated in two independent datasets. Six GM networks were identified as predictors of ECT response, achieving accuracy of 89%, 90% and 86% for remission prediction in three independent datasets.[19]

3.2 | Common machine-learning challenges in MDD study

Studies involving a combination of MRI and pattern recognition techniques to explore biomarkers of depression have grown substantially in recent years. Such methods can accurately discriminate depressed subjects from healthy controls[43,45–49,88,89,91] and predict treatment response.[19,20,90,92] In

our survey, there are more studies focused on classification (53 studies) than treatment response prediction (10 studies). To the best of our knowledge, the number of articles that use machine-learning methods for classifying depression is limited,[101] and many of these methods have not been integrated into a clinical application. We believe the main reason is the heterogeneity of imaging data including data collection, scanning parameters, and processing methods which hampers generalization to other datasets. This makes it difficult to draw comparisons based on the results.

3.2.1 | Small sample size

The small sample size is a universal problem faced by most studies of depression as reported till now, and it is hence not easy to draw definite conclusions about the diagnostic value of neuroimaging at the individual level, although several MDD studies working on thousands of samples are still ongoing. Given the difficulty of recruiting patients, the limitation of a small sample size is quite understandable. This problem is naturally difficult and common in machine-learning methodologies, but the sample size is still miniscule in comparison with other fields in which machine learning is used leading to several problems. There are a growing number of repositories which address the broader issue of small sample sizes within neuroimaging research,[22] but these typically lack uniformity between contributing sites with respect to acquisition and processing parameters which may introduce bias to the aggregated data. As long as there is no common standard between different sites, performing machine-learning methods is limited to the available sample sizes.

3.2.2 | Feature reduction

Given the small sample sizes used by the past studies, a proper feature reduction method should be used to improve overall performance. Features used by past studies vary considerably by MRI modality, feature reduction method, and number and type of features. Despite this, various features in past studies seem to be useful in major depression, for example reduced activation in dorsolateral prefrontal areas[3,18,43,46,48,50,82] and decreased gray matter volume in prefrontal cortex and subcortical systems.[59,62,64,68,83,98] Group differences conducted on whole dataset were applied to select features in some studies,[50] but this may introduce bias into the feature selection step.[102,103] There is not a direct relationship between statistical analysis and discrimination power as they are different criteria.[101] As such valuable discriminatory information could be lost by discarding features based on group differences,[104–106] better approaches should be introduced including recursive feature elimination (RFE),[27] minimum-redundancy maximum relevancy (mRMR),[107] and methods that learn the features contributing most to the accuracy of the model like least absolute shrinkage and selection operator (LASSO), elastic net and ridge regression.[19,65]

3.2.3 | Overfitting

Overfitting can result in very good performance on training data, but very poor performance on testing data, and poor generalization to independent datasets,[108] which may be caused by small sample size with high-dimensional features[32,109] and complex models with too many parameters. Neuroimaging applications in MDD unsurprisingly can also suffer from overfitting. And cross-validation is a common approach to control for overfitting. As mentioned above, the proper type of CV should be selected based on the data scale, as shown in Figure 1G.

3.2.4 | Classification methods and cross-validation

Almost all of the selected studies used SVM or its variant method as the primary classification method[22,40–42,44–50,52,53,82,84,87,89–91,98] and use LOOCV for cross validation. The reason why SVM is the most popular choice among depression classification is because of its useful strengths on including a reliable theoretical foundation and its flexible response to high-dimensional data. Considering most neuroimaging studies are likely to be nonlinear, kernel SVM has been proposed to achieve better performance than the other methods for nonlinear depression classification.[65] Nevertheless, if the number of samples is significantly less than the number of features, it may be better to simply use the linear-learning-based method to avoid complexity and mitigate against overfitting.[110,111] For cross validation, the LOOCV method provides more data to the training stage of learning method, which is associated with high variance , which may weaken generalization performance and lead to overfitting.[112,113] According to our recent work, 10-fold CV provides a more stable performance across different data while LOOCV performance heavily depends on the used data.[3]

4 | FUTURE DIRECTIONS

Based on the discussion in the previous section of past studies, there are several potential directions to explore for future studies (Figures 2 and 3).

4.1 | Multi-cross-classification for large sample size and deep learning

The selection of appropriate learning methods is very important to accurately learn a classification framework. Some researchers have successfully realized this selection of methods as applied to other brain disorders.[114–116]

Large sample size is of great importance for valid classification performance, but it is often not easy to collect large samples at one single site. To address this problem, multisite data sharing has been proposed which allows for cross-site classification of psychiatric disorders such as schizophrenia. In cross-site classification, the model is trained at one or several independent sites and tested at different

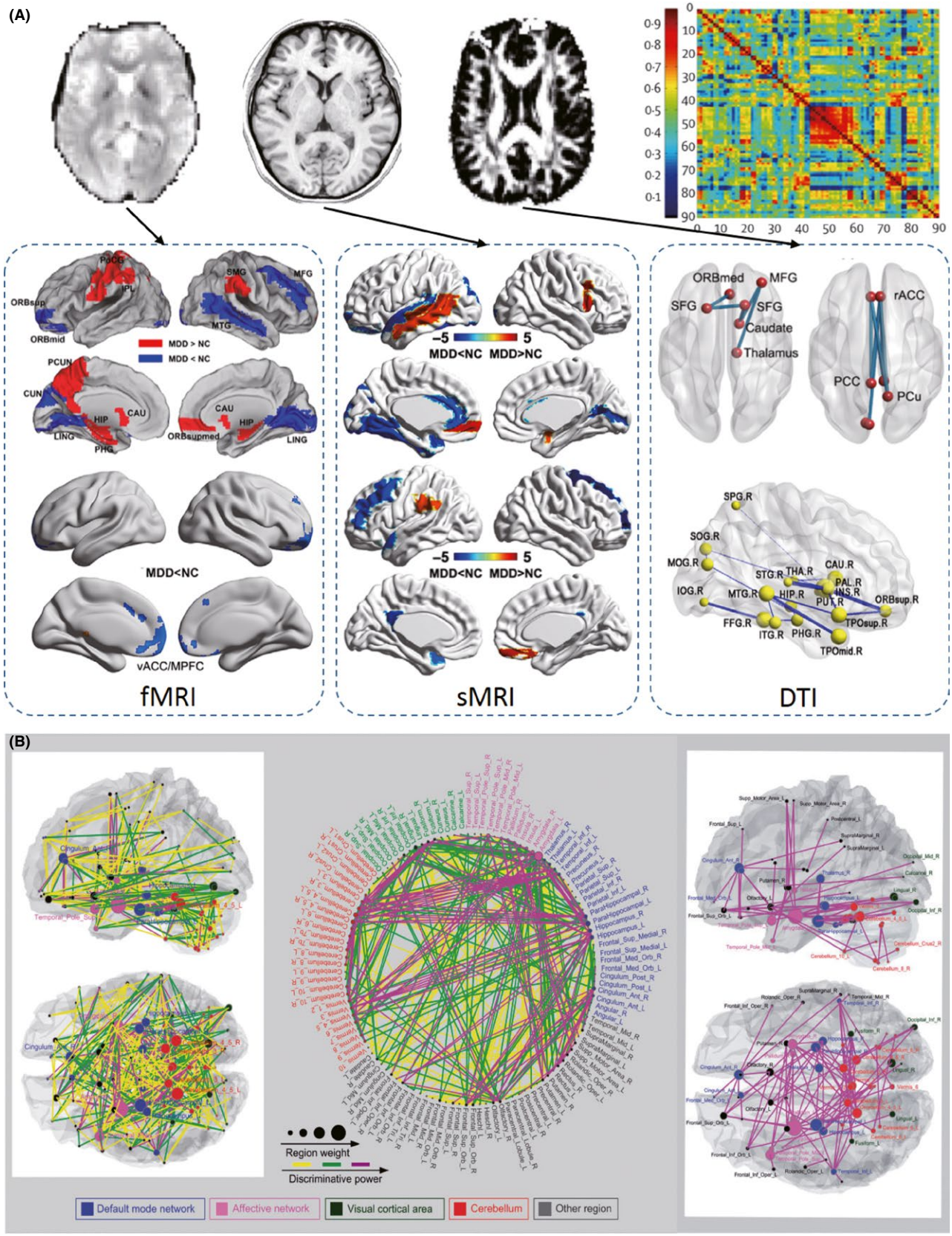
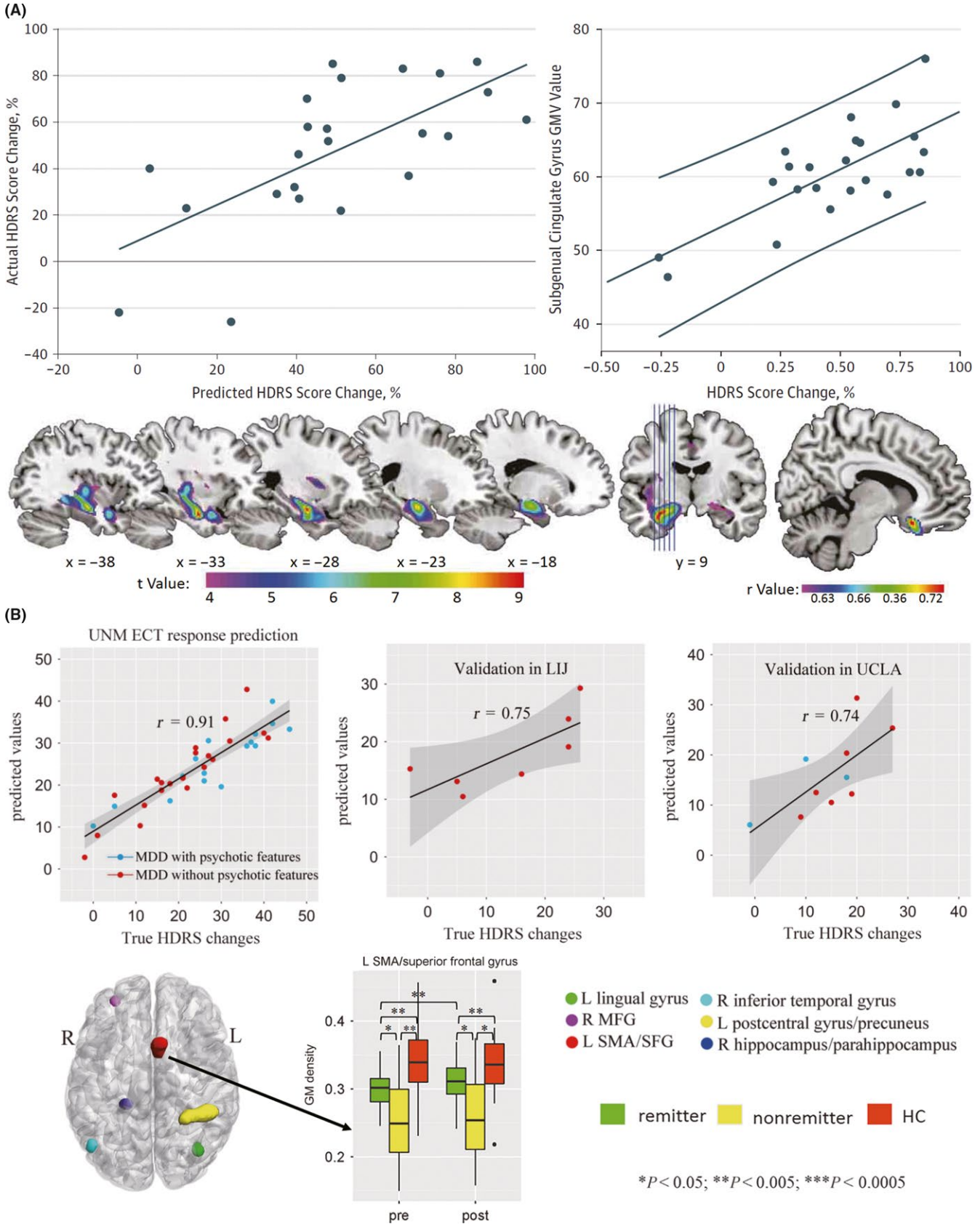


FIGURE 2 Brain network studies in MDD classification. A[100]: Brain network construction with MRI and connectome architecture represented by a connectivity matrix. B[86]: Region weights and distribution of 442 consensus functional connections identified by classification of MDD and HC demonstrated in sagittal and axial view (left) and in a circle graph (middle). Top 100 most discriminating consensus functional connections in sagittal and axial view (right). [A reproduced from ref. [100]; B reproduced from ref. [86]]



sites. As a significant machine-learning method for large multisite data, cross-classification can tackle the problem of overfitting and further provide potential biomarkers less specific to one certain site, which will be more generalizable in clinical practice. Rozycki et al used advanced multivariate analysis tools and structural neuroimaging data of 941 participants from five sites to find neuroanatomical signature of

FIGURE 3 Predication studies in MDD. A[20]: Positive association between predicted and true change in the Hamilton Depression Rating Scale (HDRS) score. Positive association between change in HDRS score and subgenual anterior cingulate volume before electroconvulsive therapy (ECT). Gray matter volume (GMV) increasing in the ECT group. Spatial map of correlated anterior cingulate volume. B[19]: Scatter plot of the predicted ΔHDRS (Hamilton Depression Rating Scale) with respect to their true values for three sites, extracting six identified pre-electroconvulsive therapy (ECT) gray matter (GM) regions in University of New Mexico (UNM) and using them as regressors for two independent cohorts: Long Island Jewish Health System (LIJ) and University of California at Los Angeles (UCLA). Six identified pre-electroconvulsive therapy (ECT) GM regions of interest (ROIs) as predictors of ΔHDRS in axial view. Longitudinal GM changes among remitters, nonremitters, and healthy controls of left supplementary motor area (SMA) and superior frontal gyrus (SFG). [A reproduced from ref. [20]; B reproduced from ref. [19]]

patients with schizophrenia and establish cross-site classification with robust generalizability.[114] Zeng et al proposed discriminant deep learning method using fMRI of 734 participants form seven sites to learn discriminating functional connections and achieve accurate prediction.[115] Both of the studies conducted pooling classification and leave-site-out validation, and obtained promising classification results. Discriminating brain patterns were shared by all sites. Cross-site classification is still challenging, but shows promise for the future.

Another state of the art machine-learning method is deep learning, which has the ability to extract hidden information from high-dimensional data, and some researches have already show enhanced classification accuracies with neuroimaging data. For example, Kim et al adopt deep neural network (DNN) with L1-norm controlling weight sparsity in hidden layer for whole-brain resting-state FC pattern classification of schizophrenia vs HC.[116] Zeng et al investigated cross-site classification with deep learning in schizophrenia for the first time.[115] These methods may transfer to other neuropsychiatric disorders such as MDD to build diagnostic tools and provide better analysis about pathophysiology.

4.2 | Multimodal MRI in MDD

Although fMRI and sMRI biomarkers have been found to be associated with depression, there are additional studies that have shown the relevance of DTI biomarkers.[41,47,58,93] Additionally, nonimaging measures have also been used in depression.[117] Thus, it is important to study how multimodal MRI in conjunction with nonimaging features affects prediction models of depression.[27,50] Each MRI modality represents different view of the brain, and data fusion capitalizes on the strengths of each modality and their inter-relationships in a joint analysis to unravel the pathophysiology of brain disease.[117–119] Recent advances in data fusion[120–122] increase our confidence in multimodal approaches and also provide insight into both anatomical and functional information.[123–125] Often multimodal studies reveal information which may be missed by methods based on a single modality.[126] Some studies have already applied advanced multimodal fusion methods like multisite canonical correlation analysis with reference + joint independent component analysis (mCCA+jICA)[119] and its variants[117] in MDD associated analysis with promising classification performance.[50] In ref. [27] jointly selected features from amplitude of low-frequency fluctuation (ALFF) and GM trained by the SVM classifier enable classifying MDD and BD at high accuracy based on the identified features (eg, dorsal lateral prefrontal cortex in GM). Data fusion methods combined with machine learning are thus a promising direction for depression classification.

4.3 | Multiple classification and subtypes

Multiple classification can be performed via a pseudo multiclass strategy by applying a two-class algorithm either to separate pairs of depression subtypes or to separate different subtype from one other. An alternative approach for this problem is to apply clustering methods to label subjects as belonging to specific disease subtype clusters.[22] Although abnormalities of major neuroanatomical regions and neural networks are common in one generalized category of disorders, prominently disturbed symptoms differ between subtypes, which have already been found in MDD[127] and BD.[128,129] Differentiating a disorder sharing symptoms with other disorders is also one of the main challenges in psychiatry and neurology. It has been reported that such overlapping disorders include schizophrenia, bipolar, unipolar, and mood disorders. Classification with task-based fMRI has achieved good performance in distinguishing schizophrenia and bipolar disorder.[130–132] Other studies also classify schizophrenia, bipolar and healthy controls with high accuracy using sMRI.[59,64,133,134] So, multiple classification in depression is thought to be very promising but also challenging.[22]

Besides, there are also some researches applying machine-learning methods to explore vulnerable biomarkers. In the field of MDD, Opel et al investigated gray matter alterations of healthy controls, MDD patients, healthy first-degree relatives of MDD and healthy individuals exposed to former childhood maltreatment using univariate analysis (t test) and pattern recognition approach (SVM) to conduct both group- and individual-level analyses. The classifier can successfully detect individuals with a risk of MDD and perform even better with the help of specific brain regions associated with MDD found by group-level analysis, showing the potential power in searching familial and environmental risk factors for MDD in the future.[135]

4.4 | Large-scale datasets

Problems with data heterogeneity may be reduced through using large training datasets.[32,109] In the past few years, several multiple center data repositories such as PGC, ENIGMA, UK Biobank, and others have been started, and they are all collaborative confederation with many working groups including major depression group. Despite some respective MDD group of these large consortium contains relatively smaller sample size, there are also marvelous achievements. For example, MDD workgroup has been a part of

the PGC since 2007, and now it covers over 100 000 people with depression. The PGC MDD group has published a paper[136] confronting the notably challenges in genetic dissection of MDD and keeps increasing sample size and expanding their studies.[137–140] The ENIGMA MDD working group includes brain scans of around >5000 MDD patients and >9000 controls from 35 research samples of 14 different countries worldwide. Its primary aim is to identify imaging markers that robustly discriminate MDD patients from healthy controls cross-site with standardized image processing and statistical analysis protocols.[141–144] More recently, a predictive analytics competition (PAC), a major depression classification challenge with the goal of automatic classification of patients suffering from major depression, and healthy individuals based on sMRI data have been carried out (https://www.photon-ai.com/pac). The training data of PAC contain labeled sMRI data of 759 MDD patients and 1033 normal controls from three different sites which are free to public. The unlabeled testing data of totally 448 individuals from three sites are also available. The winners of 2018 PAC competition achieved 65% accuracy on classifying MDD from HC. Another ongoing project named REST-meta-MDD in China also integrated multisite meta-analysis results from thousands of resting-state fMRI data of MDD and HC. We believe that the field of MDD can benefit a lot from similar machine-learning competitions and projects.

5 | CONCLUSIONS

The widespread availability of machine-learning methods combined with MRI data affords unprecedented opportunities to further deepen individual-level analysis of major depression and accelerate translation to clinical application. Approaches for combining machine-learning methods and MRI data are still largely at the exploratory stage. Classification models and features extracted from multiple modalities are irregular across different studies and this heterogeneity makes it harder to unearth optimal MRI modalities, features, and algorithm. Currently, the trend of combining machine learning approaches and MRI data in depression is drawing more attention due to the high potential and provides more information about the underlying brain regions which are involved. Though there are many challenges, but there is still huge potential for approaches which could leverage multimodal data types, brain connectomics, big data from different centers, subtype classification, and combination with clinical and genetic information.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGMENTS

This work was supported by the National High-Tech Development Plan (863, No. 2015AA020513), "100 Talents Plan" of Chinese Academy of Sciences, the Chinese Natural Science Foundation Number 81471367, 61773380, the Strategic Priority Research Program of the Chinese Academy of Sciences (No.XDBS01040100) and NIH Grant R56MH117107, R01EB005846, 1R01MH094524, and P20GM103472.

ORCID

Shuang Gao https://orcid.org/0000-0002-8747-8781
Vince D. Calhoun https://orcid.org/0000-0001-9058-0747
Jing Sui https://orcid.org/0000-0001-6837-5966

REFERENCES

1. Murray CJ, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet*. 2012;380(9859):2197-2223.
2. Papakostas GI. Managing partial response or nonresponse: switching, augmentation, and combination strategies for major depressive disorder. *J Clin Psychiatry*. 2009;70(6):16-25.
3. Gao S, Osuch EA, Wammes M, et al. Discriminating bipolar disorder from major depression based on kernel SVM using functional independent components. 2017 IEEE 27th International Workshop on Machine Learning for Signal Processing (MLSP); 25–28 Sept. 2017, 2017.
4. He Y, Chen ZJ, Evans AC. Small-world anatomical networks in the human brain revealed by cortical thickness from MRI. *Cereb Cortex*. 2007;17(10):2407-2419.
5. He Y, Dagher A, Chen Z, et al. Impaired small-world efficiency in structural cortical networks in multiple sclerosis associated with white matter lesion load. *Brain*. 2009;132(12):3366-3379.
6. He Y, Wang J, Wang L, et al. Uncovering Intrinsic Modular Organization of Spontaneous Brain Activity in Humans. *PLoS One*. 2009;4(4):e5226.
7. Bromet E, Andrade LH, Hwang I, et al. Cross-national epidemiology of DSM-IV major depressive episode. *BMC Med*. 2011;9(1):90.
8. Laje G, Paddock S, Manji H, et al. Genetic markers of suicidal ideation emerging during citalopram treatment of major depression. *Am J Psychiatry*. 2007;164(10):1530-1538.
9. Miller AH, Maletic V, Raison CL. Inflammation and its discontents: the role of cytokines in the pathophysiology of major depression. *Biol Psychiatry*. 2009;65(9):732-741.
10. Herrmann LL, Le MM, Ebmeier KP. White matter hyperintensities in late life depression: a systematic review. *J Neurol Neurosurg Psychiatry*. 2008;79(6):619.
11. Le BD, Mangin JF, Poupon C, et al. Diffusion tensor imaging: concepts and applications. *J Magn Reson Imaging*. 2001;13(4):534.
12. Raemaekers M, Vink M, Zandbelt B, Wezel R, Kahn RS, Ramsey NF. Test–retest reliability of fMRI activation during prosaccades and antisaccades. *Neuroimage*. 2007;36(3):532.
13. De KB, Ruhe E, Caan M, et al. Relation between structural and functional connectivity in major depressive disorder. *Biol Psychiatry*. 2013;74(1):40.
14. Khalaf A, Edelman K, Tudorascu D, Andreescu C, Reynolds CF, Aizenstein H. White matter hyperintensity accumulation during treatment of late-life depression. *Neuropsychopharmacology*. 2015;40(13):3027.
15. Steffens DC, Taylor WD, Denny KL, Bergman SR, Wang L. Structural integrity of the uncinate fasciculus and resting state functional connectivity of the ventral prefrontal cortex in late life depression. *PLoS One*. 2011;6(7):e22697.

16. Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neurosci Biobehav Rev*. 2012;36(4):1140-1152.
17. Haslam N, Beck AT. Categorization of major depression in an outpatient sample. *J Nerv Ment Dis*. 1993;181(12):725-731.
18. Patel MJ, Andreescu C, Price JC, Edelman KL, Iii C, Aizenstein HJ. Machine learning approaches for integrating clinical and imaging features in LLD classification and response prediction. *Int J Geriatr Psychiatry*. 2015;30(10):1056.
19. Jiang R, Abbott CC, Jiang T, et al. SMRI biomarkers predict electroconvulsive treatment outcomes: accuracy with independent data sets. *Neuropsychopharmacology*. 2018;43(5):1078-1087.
20. Redlich R, Opel N, Grotegerd D, et al. Prediction of individual response to electroconvulsive therapy via machine learning on structural magnetic resonance imaging data. *Jama Psychiatry*. 2016;73(6):557.
21. Moher D, Liberati A, Tetzlaff J, Altman DG, The PG. Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *PLoS Medicine*. 2009;6(7):e1000097.
22. Drysdale AT, Grosenick L, Downar J, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat Med*. 2017;23(1):28.
23. Saeys Y, Inza I, Larrañaga P. A review of feature selection techniques in bioinformatics. *Bioinformatics*. 2007;23(19):2507.
24. Navot A, Shpigelman L, Tishby N, Vaadia E. Nearest neighbor based feature selection for regression and its application to neural activity. *Adv Neural Inf Process Syst*. 2005;18:995-1002.
25. Gilad-Bachrach R, Navot A, Tishby N. Margin based feature selection - theory and algorithms. Proceedings of the twenty-first international conference on Machine learning, 2004; Banff, Alberta, Canada
26. Fan Y, Liu Y, Wu H, et al. Discriminant analysis of functional connectivity patterns on Grassmann manifold. *Neuroimage*. 2011;56(4):2058-2067.
27. Jie NF, Zhu MH, Ma XY, et al. Discriminating bipolar disorder from major depression based on SVM-FoBa: efficient feature selection with multimodal brain imaging data. *IEEE Trans Auton Ment Dev*. 2015;7(4):320-331.
28. Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Paper presented at: International Joint Conference on Artificial Intelligence, 1995.
29. Patel MJ, Khalaf A, Aizenstein HJ. Studying depression using imaging and machine learning methods. *Neuroimage Clin*. 2016;10:115-123.
30. Lemm S, Blankertz B, Dickhaus T, Müller KR. Introduction to machine learning for brain imaging. *Neuroimage*. 2011;56(2):387.
31. Sato JR, Fujita A, Thomaz CE, Mourão-Miranda J, Brammer MJ, Junior EA. Evaluating SVM and MLDA in the extraction of discriminant regions for mental state prediction. *Neuroimage*. 2009;46(1):105-114.
32. Franke K, Ziegler G, Klöppel S, Gaser C. Estimating the age of healthy subjects from T 1 -weighted MRI scans using kernel methods: Exploring the influence of various parameters. *Neuroimage*. 2010;50(3):883.
33. Chen R, Herskovits EH. Machine-learning techniques for building a diagnostic model for very mild dementia. *Neuroimage*. 2010;52(1):234-244.
34. Plant C, Teipel SJ, Oswald A, et al. Automated detection of brain atrophy patterns based on MRI for the prediction of Alzheimer's disease. *Neuroimage*. 2010;50(1):162-174.
35. Cuingnet R, Gerardin E, Tessieras J, et al. Automatic classification of patients with Alzheimer's disease from structural MRI: a comparison of ten methods using the ADNI database. *Neuroimage*. 2011;56(2):766.
36. Cortes C, Vapnik V. Support-vector networks. *Mach Learn*. 1995;20(3):273-297.
37. Baldi P, Brunak S, Chauvin Y, Andersen C, Nielsen H. Assessing the accuracy of prediction algorithms for classification: an overview. *Bioinformatics*. 2000;16(5):412.
38. Alberg AJ, Park JW, Hager BW, Brock MV, Diener-West M. The use of "overall accuracy" to evaluate the validity of screening or diagnostic tests. *J Gen Intern Med*. 2010;19(5p1):460-465.
39. Domingos P. A few useful things to know about machine learning. *Commun ACM*. 2012;55(10):78-87.
40. Rubin-Falcone H, Zanderigo F, Thapa-Chhetry B, et al. Pattern recognition of magnetic resonance imaging-based gray matter volume measurements classifies bipolar disorder and major depressive disorder. *J Affect Disord*. 2017;227(Supplement C):498-505.
41. Deng F, Wang Y, Huang H, et al. Abnormal segments of right uncinate fasciculus and left anterior thalamic radiation in major and bipolar depression. *Prog Neuropsychopharmacol Biol Psychiatry*. 2017;81:340-349.
42. Jing B, Long Z, Liu H, et al. Identifying current and remitted major depressive disorder with the Hurst exponent: a comparative study on two automated anatomical labeling atlases. *Oncotarget*. 2017;8(52):90452-90464.
43. Yoshida K, Shimizu Y, Yoshimoto J, et al. Prediction of clinical depression scores and detection of changes in whole-brain using resting-state functional MRI data with partial least squares regression. *PLoS One*. 2017;12(7):e0179638.
44. Li M, Das T, Deng W, et al. Clinical utility of a short resting-state MRI scan in differentiating bipolar from unipolar depression. *Acta Psychiatr Scand*. 2017;136(3):288-299.
45. Zhong X, Shi H, Ming Q, et al. Whole-brain resting-state functional connectivity identified major depressive disorder: a multivariate pattern analysis in two independent samples. *J Affect Disord*. 2017;218:346-352.
46. Wang X, Ren Y, Zhang W. depression disorder classification of fMRI data using sparse low-rank functional brain network and graph-based features. *Comput Math Methods Med*. 2017;2017:3609821.
47. Schnyer DM, Clasen PC, Gonzalez C, Beevers CG. Evaluating the diagnostic utility of applying a machine learning algorithm to diffusion tensor MRI measures in individuals with major depressive disorder. *Psychiatry Res*. 2017;264:1.
48. Sundermann B, Feder S, Wersching H, et al. Diagnostic classification of unipolar depression based on resting-state functional connectivity MRI: effects of generalization to a diverse sample. *J Neural Transm*. 2017;124(5):589.
49. Bhaumik R, Jenkins LM, Gowins JR, et al. Multivariate pattern analysis strategies in detection of remitted major depressive disorder using resting state functional connectivity. *Neuroimage Clinical*. 2017;16(C):390.
50. He H, Sui J, Du Y, et al. Co-altered functional networks and brain structure in unmedicated patients with bipolar and major depressive disorders. *BrainStruct Funct*. 2017;222(9):4051-4064.
51. Bürger C, Redlich R, Grotegerd D, et al. Differential abnormal pattern of anterior cingulate gyrus activation in unipolar and bipolar depression: an fMRI and pattern classification approach. *Neuropsychopharmacology*. 2017;42:1399.
52. Hilbert K, Lueken U, Muehlhan M, Beesdobaum K. Separating generalized anxiety disorder from major depression using clinical, hormonal, and structural MRI data: a multimodal machine learning study. *Brain Behav*. 2017;7(3):e00633.
53. Sankar A, Zhang T, Gaonkar B, et al. Diagnostic potential of structural neuroimaging for depression from a multi-ethnic community sample. *BJPsych open*. 2016;2(4):247-254.
54. Frangou S, Dima D, Jogia J. Towards person-centered neuroimaging markers for resilience and vulnerability in Bipolar Disorder. *Neuroimage*. 2017;145:230-237.

55. Ramasubbu R, Brown M, Cortese F, et al. Accuracy of automated classification of major depressive disorder as a function of symptom severity. *Neuroimage Clin*. 2016;12(C):320-331.
56. Yang W, Chen Q, Liu P, et al. Abnormal brain activation during directed forgetting of negative memory in depressed patients. *J Affect Disord*. 2015;190:880-888.
57. Rive MM, Redlich R, Schmaal L, et al. Distinguishing medication-free subjects with unipolar disorder from subjects with bipolar disorder: state matters. *Bipolar Disord*. 2016;18(7):612-623.
58. Foland-Ross LC, Sacchet MD, Prasad G, Gilbert B, Thompson PM, Gotlib IH. Cortical thickness predicts the first onset of major depression in adolescence. *Int J Dev Neurosci*. 2015;46:125-131.
59. Sacchet MD, Livermore EE, Juan Eugenio I, Glover GH, Gotlib IH. Subcortical volumes differentiate major depressive disorder, bipolar disorder, and remitted major depressive disorder. *J Psychiatr Res*. 2015;68:91.
60. Sacchet MD, Prasad G, Foland-Ross LC, Thompson PM, Gotlib IH. Support vector machine classification of major depressive disorder using diffusion-weighted neuroimaging and graph theory. *Front Psychiatry*. 2015;6:21.
61. Sato JR, Jorge M, Sophie G, Deakin J, Thomaz CE, Roland Z. Machine learning algorithm accurately detects fMRI signature of vulnerability to major depression. *Psychiatry Res*. 2015;233(2):289-291.
62. Johnston BA, Steele JD, Tolomeo S, Christmas D, Matthews K. Structural MRI-based predictions in patients with treatment-refractory depression (TRD). *PLoS One*. 2015;10(7):e0132958.
63. Johnston BA, Tolomeo S, Gradin V, Christmas D, Matthews K, Douglas, . SJ. Failure of hippocampal deactivation during loss events in treatment-resistant depression. *Brain*. 2015;138(9):2766-2776.
64. Koutsouleris N, Meisenzahl EM, Borgwardt S, et al. Individualized differential diagnosis of schizophrenia and mood disorders using neuroanatomical biomarkers. *Brain A J Neurol*. 2015;138(7):2059.
65. Shimizu Y, Yoshimoto J, Toki S, et al. Toward probabilistic diagnosis and understanding of depression based on functional MRI data analysis with logistic group LASSO. *PLoS One*. 2015;10(5):e0123524.
66. Fung G, Deng Y, Zhao Q, et al. Distinguishing bipolar and major depressive disorders by brain structural morphometry: a pilot study. *BMC Psychiatry*. 2015;15(1):298.
67. Rosa MJ, Portugal L, Shawe-Taylor J, Mourao-Miranda J.Sparse Network-Based Models for Patient Classification Using fMRI. 2013 International Workshop on Pattern Recognition in Neuroimaging; 22–24 June 2013, 2013.
68. Redlich R, Almeida JJ, Grotegerd D, et al. Brain morphometric biomarkers distinguishing unipolar and bipolar depression. A voxel-based morphometry-pattern classification approach. *JAMA Psychiatry*. 2014;71(11):1222.
69. Cao L, Guo S, Xue Z, et al. Aberrant functional connectivity for diagnosis of major depressive disorder: a discriminant analysis. *Psychiatry Clin Neurosci*. 2014;68(2):110-119.
70. Macmaster FP, Carrey N, Langevin LM, Jaworska N, Crawford S. Disorder-specific volumetric brain difference in adolescent major depressive disorder and bipolar depression. *Brain Imaging Behav*. 2014;8(1):119-127.
71. Zeng LL, Shen H, Liu L, Hu D. Unsupervised classification of major depression using functional connectivity MRI. *Hum Brain Mapp*. 2014;35(4):1630.
72. Rondina JM, Hahn T, Oliveira LD, et al. SCoRS - a method based on stability for feature selection and mapping in neuroimaging. *IEEE Trans Med Imaging*. 2014;33(1):85.
73. Guo H, Cheng C, Cao X, Xiang J, Chen J, Zhang K. Resting-state functional connectivity abnormalities in first-onset unmedicated depression. *Neural Regen Res*. 2014;9(2):153-163.
74. Serpa MH, Ou Y, Schaufelberger MS, et al. Neuroanatomical classification in a population-based sample of psychotic major depression and bipolar i disorder with 1 year of diagnostic stability. *Biomed Res Int*. 2014;2014(5):706157.
75. Habes I, Krall SC, Johnston SJ, et al. Pattern classification of valence in depression. *Neuroimage Clin*. 2013;2(1):675-683.
76. Wei M, Qin J, Yan R, Li H, Yao Z, Lu Q. Identifying major depressive disorder using Hurst exponent of resting-state brain networks. *Psychiatry Res*. 2013;214(3):306-312.
77. Grotegerd D, Stuhrmann A, Kugel H, et al. Amygdala excitability to subliminally presented emotional faces distinguishes unipolar and bipolar depression: An fMRI and pattern classification study. *Hum Brain Mapp*. 2014;35(7):2995-3007.
78. Yu Y, Shen H, Zeng LL, Ma Q, Hu D. Convergent and divergent functional connectivity patterns in schizophrenia and depression. *PLoS One*. 2013;8(7):e68250.
79. Modinos G, Mechelli A, Petterssonyeo W, Allen P, Mcguire P, Aleman A. Pattern classification of brain activation during emotional processing in subclinical depression: psychosis proneness as potential confounding factor. *PeerJ*. 2013;1(7):e42.
80. Ma Z, Li R, Yu J, He Y, Li J. Alterations in regional homogeneity of spontaneous brain activity in late-life subthreshold depression. *PLoS One*. 2013;8(1):e53148.
81. Grotegerd D, Suslow T, Bauer J, et al. Discriminating unipolar and bipolar depression by means of fMRI and pattern classification: a pilot study. *Eur Arch Psychiatry Clin Neurosci*. 2013;263(2):119-131.
82. Fang P, Zeng LL, Shen H, et al. Increased cortical-limbic anatomical network connectivity in major depression revealed by diffusion tensor imaging. *PLoS One*. 2012;7(9):e45972.
83. Mwangi B, Ebmeier KP, Matthews K, Steele JD. Multi-centre diagnostic classification of individual structural neuroimaging scans from patients with major depressive disorder. *Brain A J Neurol*. 2012;135(Pt 5):1508.
84. Lord A, Horn D, Breakspear M, Walter M. Changes in community structure of resting state functional connectivity in unipolar depression. *PLoS One*. 2012;7(8):e41282.
85. Liu F, Guo W, Yu D, et al. Classification of different therapeutic responses of major depressive disorder with multivariate pattern analysis method based on structural MR scans. *PLoS One*. 2012;7(7):e40968.
86. Zeng LL, Shen H, Liu L, et al. Identifying major depression using whole-brain functional connectivity: a multivariate pattern analysis. *Brain A J Neurol*. 2012;135(Pt 5):1498.
87. Mourão-Miranda J, Hardoon DR, Hahn T, et al. Patient classification as an outlier detection problem: An application of the One-Class Support Vector Machine. *Neuroimage*. 2011;58(3):793-804.
88. Hahn T, Marquand AF, Ehlis AC, et al. Integrating neurobiological markers of depression. *Arch Gen Psychiatry*. 2011;68(4):361.
89. Nouretdinov I, Costafreda SG, Gammerman A, et al. Machine learning classification with confidence: Application of transductive conformal predictors to MRI-based diagnostic and prognostic markers in depression. *Neuroimage*. 2011;56(2):809-813.
90. Costafreda SG, Chu C, Ashburner J, Fu CH. Prognostic and diagnostic potential of the structural neuroanatomy of depression. *PLoS One*. 2009;4(7):e6353.
91. Fu CH, Mouraomiranda J, Costafreda SG, et al. Pattern classification of sad facial processing: toward the development of neurobiological markers in depression. *Biol Psychiatry*. 2008;63(7):656.
92. Lythe KE, Moll J, Gethin JA, et al. Self-blame-selective hyperconnectivity between anterior temporal and subgenual cortices and prediction of recurrent depressive episodes. *Jama Psychiatry*. 2015;72(11):1119-1126.
93. Korgaonkar MS, Rekshan W, Gordon E, et al. Magnetic resonance imaging measures of brain structure to predict antidepressant treatment outcome in major depressive disorder. *Ebiomedicine*. 2015;2(1):37-45.

94. Williams LM, Korgaonkar MS, Song YC, et al. Amygdala reactivity to emotional faces in the prediction of general and medication-specific responses to antidepressant treatment in the randomized iSPOT-D trial. *Neuropsychopharmacology*. 2015;40(10):2398-2408.
95. Schmaal L, Marquand AF, Rhebergen D, et al. predicting the naturalistic course of major depressive disorder using clinical and multimodal neuroimaging information: a multivariate pattern recognition study. *Biol Psychiatry*. 2015;78(4):278-286.
96. van Waarde JA, Scholte HS, van Oudheusden LJ, Verwey B, Denys D, van Wingen GA. A functional MRI marker may predict the outcome of electroconvulsive therapy in severe and treatment-resistant depression. *Mol Psychiatry*. 2015;20(5):609.
97. Korgaonkar MS, Williams LM, Song YJ, Usherwood T, Grieve SM. Diffusion tensor imaging predictors of treatment outcomes in major depressive disorder. *Br J Psychiatr*. 2014;205(4):321.
98. Gong Q, Wu Q, Scarpazza C, Su L, Jia Z. Prognostic prediction of therapeutic response in depression using high-field MR imaging. *Neuroimage*. 2011;55(4):1497.
99. Costafreda SG, Khanna A, Mourao-Miranda J, Fu CH. Neural correlates of sad faces predict clinical remission to cognitive behavioural therapy in depression. *Neuroreport*. 2009;20(7):637-641.
100. Gong Q, He Y. Depression, neuroimaging and connectomics: a selective overview. *Biol Psychiatry*. 2015;77(3):223-235.
101. Arbabshirani MR, Plis S, Sui J, Calhoun VD. Single subject prediction of brain disorders in neuroimaging: Promises and pitfalls. *Neuroimage*. 2017;145(Pt B):137.
102. Bishop CM. *Pattern recognition and machine learning (information science and statistics)*. New York: Springer-Verlag; 2006.
103. Demirci O, Clark VP, Magnotta VA, et al. A review of challenges in the use of fMRI for disease classification / characterization and a projection pursuit application from multi-site fMRI schizophrenia study. *Brain Imaging Behav*. 2008;2(3):207-226.
104. Blum AL, Langley P. Selection of relevant features and examples in machine. *Artif Intell*. 1997;97(1–2):245-271.
105. Hall MA, Smith LA. Practical feature subset selection for machine learning. In *Computer Science '98 Proceedings of the 21st Australasian Computer Science Conference ACSC'98*, Perth, 4-6 Feb, 1998.
106. Kohavi R, John GH. Wrappers for feature subset selection. *Artif Intell*. 1997;97(1–2):273-324.
107. Batmanghelich NK, Taskar B, Davatzikos C. Generative-discriminative basis learning for medical imaging. *IEEE Trans Med Imaging*. 2011;31(1):51-69.
108. Pereira F, Mitchell T, Botvinick M. Machine learning classifiers and fMRI: a tutorial overview. *Neuroimage*. 2009;45(1 Suppl):S199.
109. Klöppel S, Stonnington CM, Chu C, et al. A plea for confidence intervals and consideration of generalizability in diagnostic studies. *Brain*. 2009;132(4):e102–e102.
110. Hsu C, Chang C, Lin C. A Practical guide to support vector classification. 2003. https://www.csie.ntu.edu.tw/~cjlin/paper/guide.pdf. Accessed May 15, 2003.
111. Raudys SJ, Jain AK. Small sample size effects in statistical pattern recognition: recommendations for practitioners. *IEEE Trans Pattern Anal Mach Intell*. 1991;13(3):252-264.
112. Elisseeff A, Pontil M. Leave-one-out error and stability of learning algorithms with applications. *Adv Learn Theory Method Model Appl*. 2003;190:111-130.
113. Refaeilzadeh P, Tang L, Liu H. *Cross-validation*. Boston, MA: Springer; 2009.
114. Rozycki M, Satterthwaite TD, Koutsouleris N, et al. Multisite machine learning analysis provides a robust structural imaging signature of schizophrenia detectable across diverse patient populations and within individuals. *Schizophr Bull*. [Epub ahead of print]. https://doi.org/10.1093/schbul/sbx137
115. Zeng L-L, Wang H, Hu P, et al. Multi-site diagnostic classification of schizophrenia using discriminant deep learning with functional connectivity MRI. *EBioMedicine*. 2018;30:74-85.
116. Kim J, Calhoun VD, Shim E, Lee J-H. Deep neural network with weight sparsity control and pre-training extracts hierarchical features and enhances classification performance: evidence from whole-brain resting-state functional connectivity patterns of schizophrenia. *Neuroimage*. 2016;124:127-146.
117. Qi S, Yang X, Zhao L, et al. MicroRNA132 associated multimodal neuroimaging patterns in unmedicated major depressive disorder. *Brain*. 2018;141(3):916-926.
118. Calhoun VD, Adal T, Kiehl KA, Astur R, Pekar JJ, Pearlson GD. A method for multitask fMRI data fusion applied to schizophrenia. *Hum Brain Mapp*. 2006;27(7):598.
119. Sui J, He H, Pearlson GD, et al. Three-way (N-way) fusion of brain imaging data based on mCCA+jICA and its application to discriminating schizophrenia. *Neuroimage*. 2013;66:119.
120. Kim DI, Sui J, Rachakonda S, et al. Identification of imaging biomarkers in schizophrenia: a coefficient-constrained independent component analysis of the mind multi-site schizophrenia study. *Neuroinformatics*. 2010;8(4):213-229.
121. Sui J, Adali T, Pearlson GD, Calhoun VD. An ICA-based method for the identification of optimal FMRI features and components using combined group-discriminative techniques. *Neuroimage*. 2009;46(1):73-86.
122. Sui J, Pearlson GD, Du Y, et al. In search of multimodal neuroimaging biomarkers of cognitive deficits in schizophrenia. *Biol Psychiatry*. 2015;78(11):794-804.
123. Michael AM, Baum SA, Fries JF, et al. A method to fuse fMRI tasks through spatial correlations: applied to schizophrenia. *Hum Brain Mapp*. 2010;30(8):2512-2529.
124. Sui J, Pearlson G, Caprihan A, et al. Discriminating schizophrenia and bipolar disorder by fusing fMRI and DTI in a multimodal CCA+ joint ICA model. *Neuroimage*. 2011;57(3):839-855.
125. He Y, Evans AC. Graph theoretical modeling of brain connectivity. *Curr Opin Neurol*. 2010;23(4):341-350.
126. Calhoun VD, Adali T. Feature-based fusion of medical imaging data. *IEEE Trans Inf Technol Biomed*. 2009;13(5):711.
127. Bracht T, Horn H, Strik W, et al. White matter microstructure alterations of the medial forebrain bundle in melancholic depression. *J Affect Disord*. 2014;155(1):186.
128. Abé C, Ekman CJ, Sellgren C, Petrovic P, Ingvar M, Landén M. Cortical thickness, volume and surface area in patients with bipolar disorder types I and II. *J Psychiatr Neurosci Jpn*. 2015;41(4):240.
129. Phillips ML, Swartz HA. A critical appraisal of neuroimaging studies of bipolar disorder: toward a new conceptualization of underlying neural circuitry and a road map for future research. *Am J Psychiatry*. 2014;171(8):829.
130. Arribas JI, Calhoun VD, Adali T. Automatic Bayesian classification of healthy controls, bipolar disorder, and schizophrenia using intrinsic connectivity maps from FMRI data. *IEEE Trans Biomed Eng*. 2010;57(12):2850-2860.
131. Calhoun V, Maciejewski P, Gd KK. Temporal lobe and "default" hemodynamic brain modes discriminate between schizophrenia and bipolar disorder. *Hum Brain Mapp*. 2008;29(11):1265-1275.
132. Costafreda SG, Fu CH, Picchioni M, et al. Pattern of neural responses to verbal fluency shows diagnostic specificity for schizophrenia and bipolar disorder. *BMC Psychiatry*. 2011;11(1):18.
133. Pardo PJ, Georgopoulos AP, Kenny JT, Stuve TA, Findling RL, Schulz SC. Classification of adolescent psychotic disorders using linear discriminant analysis. *Schizophr Res*. 2006;87(1):297-306.
134. Schnack HG, Nieuwenhuis M, Haren N, et al. Can structural MRI aid in clinical classification? A machine learning study in two independent samples of patients with schizophrenia, bipolar disorder and healthy subjects. *Neuroimage*. 2014;84(1):299.

135. Opel N, Zwanzger P, Redlich R, et al. Differing brain structural correlates of familial and environmental risk for major depressive disorder revealed by a combined VBM/pattern recognition approach. *Psychol Med*. 2015;46(2):277-290.
136. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50(5):668-681.
137. Milaneschi Y, Lamers F, Peyrot WJ, et al. Genetic association of major depression with atypical features and obesity-related immunometabolic dysregulations. *JAMA Psychiatry*. 2017;74(12):1214-1225.
138. Power RA, Tansey KE, Buttenschøn HN, et al. Genome-wide association for major depression through age at onset stratification: major depressive disorder working group of the psychiatric genomics consortium. *Biol Psychiatry*. 2017;81(4):325-335.
139. Zeng Y, Navarro P, Shirali M, et al. Genome-wide regional heritability mapping identifies a locus within the TOX2 gene associated with major depressive disorder. *Biol Psychiatr*. 2017;82(5):312-321.
140. Van der Auwera S, Peyrot WJ, Milaneschi Y, et al. Genome-wide gene-environment interaction in depression: a systematic evaluation of candidate genes. *Am J Med Genet B Neuropsychiatr Genet*. 2018;177(1):40-49.
141. Schmaal L, Veltman DJ, van Erp T, et al. Subcortical brain alterations in major depressive disorder: findings from the ENIGMA Major Depressive Disorder working group. *Mol Psychiatry*. 2015;21:806.
142. Schmaal L, Hibar DP, Sämann PG, et al. 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*. 2016;22:900.
143. Frodl T, Janowitz D, Schmaal L, et al. Childhood adversity impacts on brain subcortical structures relevant to depression. *J Psychiatr Res*. 2017;86:58-65.
144. Rentería ME, Schmaal L, Hibar DP, et al. Subcortical brain structure and suicidal behaviour in major depressive disorder: a meta-analysis from the ENIGMA-MDD working group. *Transl Psychiat*. 2017;7:e1116.

How to cite this article: Gao S, Calhoun VD, Sui J. Machine learning in major depression: From classification to treatment outcome prediction. *CNS Neurosci Ther*. 2018;24:1037–1052. https://doi.org/10.1111/cns.13048