

Contrastive learning enhances the links between functional signatures and antidepressant treatment

Badong Chen

chenbd@mail.xjtu.edu.cn

Xi'an Jiaotong University

Kaizhong Zheng

Xi'an Jiaotong University

Xinhu Zheng

The Hong Kong University of Science and Technology

Hongyu Shi

Xi'an Jiaotong University

Huaning Wang

Fourth Military Medical University

Bao-Juan Li

Xijing Hospital of Air Force Military Medical University

Article

Keywords:

Posted Date: April 18th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4230487/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

Contrastive learning enhances the links between functional signatures and antidepressant treatment

Kaizhong Zheng^{1#}, Xinhua Zheng^{2#}, Hongyu Shi^{1#}, Huaning Wang³, DIRECT consortium[†], Baojuan Li^{4*}, Badong Chen^{1*}

#These authors contributed equally to this work.

*Corresponding authors

Author affiliations:

1 National Key Laboratory of Human-Machine Hybrid Augmented Intelligence, National Engineering Research Center for Visual Information and Applications, and Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China

2 Intelligent Transportation Thrust, The Hong Kong University of Science and Technology (Guangzhou), Guangzhou, China

3 Xijing Hospital, Fourth Military Medical University, Xi'an, Shaanxi, 710032, China

4 School of Biomedical Engineering, Fourth Military Medical University, Xi'an, Shaanxi, 710032, China

Correspondence to:

Baojuan Li

Full address: School of Biomedical Engineering, Fourth Military Medical University, Xi'an, Shaanxi, 710032, China

Email: libjuan@163.com

Badong Chen

Full address: National Key Laboratory of Human-Machine Hybrid Augmented Intelligence, National Engineering Research Center for Visual Information and Applications, and Institute of Artificial Intelligence and Robotics, Xi'an Jiaotong University, Xi'an, Shaanxi, 710049, China.

Email: chenbd@mail.xjtu.edu.cn

†DIRECT consortium: Chao-Gan Yan, Ph.D., Xiao Chen, Ph.D., Li-Ping Cao, M.D., Wei Chen, M.D.¹¹, Yu-Qi Cheng, M.D., Yi-Ru Fang, M.D., Qi-Yong Gong, M.D., Wen-Bin Guo, M.D., Li Kuang, M.D., Bao-Juan Li, M.D., Tao Li, M.D., Yan-Song Liu, M.D., Zhe-Ning Liu, M.D., Jian-Ping Lu, M.D., Qing-Hua Luo, M.D., Hua-Qing Meng, M.D., Dai-Hui Peng, M.D., Jiang Qiu, Ph.D., Yue-Di Shen, M.D., Tian-Mei Si, M.D., Yan-Qing Tang, M.D., Chuan-Yue Wang, M.D., Fei Wang, M.D., Hua-Ning Wang, M.D., Kai Wang, M.D., Xiang Wang, M.D., Ying Wang, M.D., Xiao-Ping Wu, M.D., Chun-Ming Xie, M.D., Guang-Rong Xie, M.D., Peng Xie, M.D., Xiu-Feng Xu, M.D., Hong Yang, M.D., Jian Yang, M.D., Shu-Qiao Yao, M.D., Yong-Qiang Yu, M.D., Yong-Gui Yuan, M.D., Ke-Rang Zhang, M.D., Wei Zhang, M.D., Zhi-Jun Zhang, M.D., Jun-Juan Zhu, M.D., Xi-Nian Zuo, Ph.D., Jing-Ping Zhao, M.D., Yu-Feng Zang, M.D.

Abstract

Major depressive disorder (MDD) is highly heterogeneous in terms of responses to treatment, which hinders the improvement in treatment effectiveness and outcomes for MDD. Identifying MDD subtypes associated with treatment responses could inform interventions and facilitate personalized treatment. Here, we sought to identify reproducible MDD subtypes characterized by distinct neurofunctional (i.e., neuroimaging) patterns to delineate heterogeneity in MDD and explored the relationship between subtypes and antidepressant treatment response. We used contrastive variational autoencoders (CVAEs) to identify two distinct MDD subtypes with the REST-meta-MDD II dataset (1660 MDD participants, 1340 HCs). Subtype 1 exhibited increased functional activity in occipital, parietal, temporal, and frontal areas, while subtype 2 showed decreased functional activity in these areas. The number and patterns of MDD subtypes were validated in a further large multi-center dataset (1276 MDD participants, 1104 HCs). Notably, patients with subtype 1 could be considered the "treatment-sensitive" group, with a response rate of over 50% to all antidepressants and a better response to repetitive transcranial magnetic stimulation (rTMS) compared to patients with subtype 2. In contrast, patients with subtype 2 could be characterized as the "treatment-resistant" group, with a response rate of below 50% for most medications. The ensuing MDD-specific features from CVAEs may serve as a neuroimaging biomarker for predicting treatment outcomes for both medication and rTMS treatments. Our study shows that contrastive learning can be used to establish the predictive validity of functional brain signatures — in terms of responses to

antidepressant treatment — offering potential new targets for optimizing treatment strategies for treatment-resistant depression, and further lay a path toward higher treatment outcomes.

Introduction

Major depressive disorder (MDD) is a severe mental illness that affects more than 320 million people worldwide, posing a significant global health and economic burden ^{1,2}. As the leading cause of disability globally, MDD has significantly impacts on daily life, work, and relationships ^{3,4}. However, patients diagnosed with MDD demonstrate significant heterogeneity in symptoms, illness trajectories, and responses to treatment ¹. Such heterogeneity may contribute to the low initial treatment remission rate (30%) and the limited advantage of antidepressant medication over placebo (Cohen's *d* of ~0.3) ⁵⁻⁷. However, the neural mechanism of treatment-resistant depression remains uncertain. In addition, the current approach to treating MDD in clinical practice involves a sequential, trial-and-error method that can take several months to find the most effective antidepressant ⁸. This approach may lead to a delay in treatment and an increase in economic burden for patients. Unfortunately, no biomarker has demonstrated sufficient sensitivity and specificity to serve as a reliable prognostic tool for MDD.

With advancements in neuroimaging technology and machine learning, researchers can use neuroimaging biomarkers such as brain connectivity and morphology to identify potential neurobiological subtypes and redefine the clinical diagnosis of MDD, thus revealing the heterogeneity of MDD ^{9,10}. As summarized in *Supplementary*

Material, Tab. S1, 11 studies have investigated MDD subtypes based on structural magnetic resonance imaging (MRI), functional connectivity (fMRI) measures, and fractional anisotropy measures. These data-driven subtyping analyses as identified homogeneous subgroups of MDD patients, also known as subtypes, who exhibit unique psychopathology and demonstrate similar patterns of brain dysfunction ¹¹. The ultimate objective is to identify neuroimaging subtypes that can assess clinical heterogeneity and assist in selecting the most suitable treatment strategy (medication or repetitive transcranial magnetic stimulation (rTMS)) for each subtype, including identifying treatment-resistant cases at an earlier stage ¹².

In order to successfully achieve the ultimate objective, we must overcome three critical challenges. **First, it is difficult to ensure that the subtype results are solely caused by the underlying nature of MDD and are not influenced by confounding variables.** As is well known, the structure and function of the brain can vary due to various genetic and environmental factors ¹³, and neuroimaging data can also differ among individuals due to systematic differences between scanners and scanning sites, etc. ¹⁴. It can be challenging to identify MDD-specific signatures amidst a large number of irrelevant factors. Current methods used to address these issues have proven to be inadequate. **Second, there is limited research on the relationship between specific subtypes and treatment outcomes, and there is a lack of appropriate treatment strategies (such as different medications or rTMS) for each subtype.** In an earlier study¹¹, hierarchical clustering was utilized to identify four MDD subtypes that were found to be predictive of treatment responses to rTMS. In addition, a recent study used

heterogeneity through discriminative analysis (HYDRA) to identify two MDD subtypes and found patients with subtype 2 and subtype 1 showed different placebo effects¹⁵. These initial attempts have shown that it is possible to establish a link between subtypes and treatment responses, however, there is a lack of systematic research on various antidepressant treatment, such as different medications. **Third, the subtypes identified in independent cohorts have not yet been externally validated.**

Contrastive learning is a generic procedure in unsupervised or self-supervised learning, where the main idea is to compare (semantically) similar (positive) and dissimilar (negative) data pairs, encouraging the representations of similar pairs to be close, and those of dissimilar pairs to be more orthogonal^{16,17}. Recently, this concept has been applied in the field of neuroimaging to study the underlying neuroanatomical variations within autism¹⁸. The authors utilized contrastive variational autoencoders (CVAEs)^{19,20} to separate neuroanatomical variations specific to autism spectrum disorder (ASD). Unlike other unsupervised learning methods, this approach also allows for the inclusion of a healthy control group as a reference, enabling the separation of brain features into disease-specific and shared features. This prevents clustering based on disease-irrelevant features. In this study, we employed contrastive learning to obtain MDD-specific features, which we believe can capture the underlying nature of MDD, without being affected by confounding variables. The MDD subtypes identified through the use of MDD-specific features are reliable, providing a strong foundation for establishing a link between subtypes and treatment outcomes.

Here, we developed modified CVAEs to identify reproducible subtypes across two

large multisite datasets and examined the relationship between distinct subtypes and treatment response (medication and rTMS). First, we used CVAEs to disentangle MDD-specific features and shared features and asked whether the MDD-specific features were correlated with clinical symptoms of MDD (i.e., to establish the construct validity of MDD-specific features, in relation to clinical presentation). Here, the inputs of CVAEs are the fractional amplitude of low-frequency fluctuations (fALFF) rather than functional connectivity measures or structural MRI. The frequency range of fALFF, typically between 0.01-0.1 Hz, has been demonstrated to be a consistent indicator of spontaneous neuronal activity at rest ²¹ and has a moderate to substantial test-retest reliability ²². This highlights its validity as a measure for detecting individual differences in regional brain function ²³. Recently, it has been applied in the subtyping analysis ²⁴. Second, we applied cluster analysis to the MDD-specific features to identify MDD subtypes. Third, we repeated the preceding analyses using an independent dataset, to test generalization. Fourth, we compared the clinical outcomes after medication and rTMS treatment between different subtypes of patients. Finally, we used MDD-specific features to predict medication and rTMS treatment responses (i.e., to establish the predictive validity, in relation to treatment response).

Results

CVAE reveals two optimal subtypes. We used the second release of the REST-meta-MDD project from the Depression Imaging REsearch ConsorTium (DIRECT) functional magnetic resonance imaging (fMRI) dataset (REST-meta-MDD II ²⁵, 1660

MDD participants, 1340 HCs) to train a CVAE (see Methods) that disentangled MDD-specific and shared features. According to previous findings of a CVAE study focused on autism spectrum disorder (ASD) ¹⁸, ASD-specific and shared features could reveal clinically relevant individual variations. We hypothesized that features specific to Major Depressive Disorder (MDD) would exhibit correlations with clinical variations in MDD, while shared features would be unrelated to clinical changes in MDD, highlighting a substantial distinction between the two. To test this hypothesis, we employed representational similarity analysis (RSA ^{26,27}, see Methods) to evaluate the association between the functional features generated by the CVAE and individual variations in clinical characteristics among MDD patients (Fig. 1a). These characteristics included sex, age, education, first episode, medication status, treatment responsiveness, family history, as well as scores on the 17-item Hamilton Depression Rating Scale (HAMD-17) and Hamilton Anxiety Scale (HAMA). RSA results demonstrated significantly reduced clinical characteristics including age ($t = 31.65$, $p < 0.0001$, false discovery rate (FDR)-correction), education ($t = 5.762$, $p < 0.0001$, FDR-correction), first episode ($t = 14.72$, $p < 0.0001$, FDR-correction), medication status ($t = 124.4$, $p < 0.0001$, FDR-correction), treatment responsive ($t = 28.35$, $p < 0.0001$, FDR-correction), HAMD-17 ($t = 2.936$, $p = 0.0088$, FDR-correction) in shared features compared with MDD-specific features.

Furthermore, we used K-means clustering methods to identify clusters of patients with MDD based on MDD-specific features, selecting the optimal number of clusters using the average distortion (Fig. 1b). 745 patients with MDD were assigned to subtype

1, while 915 patients with MDD were assigned to subtype 2. In subtype 1, fALFF was significantly increased in middle occipital areas, parietal areas, middle temporal areas and inferior frontal areas compared with HCs (Fig. 1c). The converse was found in subtype 2: fALFF was significantly decreased in middle occipital areas, parietal areas, middle and superior temporal areas and frontal areas compared to HCs (Fig. 1c).

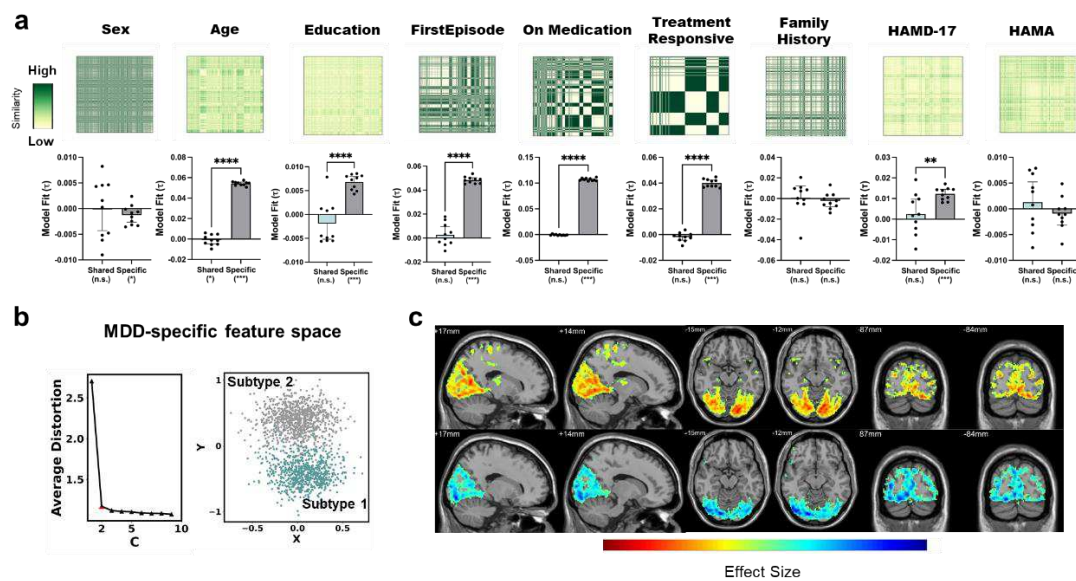


Fig. 1 Contrastive Variational Autoencoders (CVAEs) identify two MDD subtypes on the REST-meta-MDD II. **a** Similarity of neurofunctional matrices is compared to the similarity based on different participant characteristics. The MDD-specific features are more effective in capturing variables associated with MDD-related variation compared to shared features. **b** Optimal clustering. MDD-specific features are best captured by two clusters. Scatterplots show individual subjects' fALFF data from subtype 1 (green) and subtype 2 (gray) groups projected onto uniform manifold approximation and projection dimensions computed from the MDD-specific features. **c** False discovery rate (FDR)-corrected voxel-wise comparison of fALFF volume differences between subtype 1 (top row) and subtype 2 (bottom row) and controls are

displayed on transverse, sagittal, and coronal sections. The color bar illustrates the degree of group differences (t -statistic) between distinct MDD subtypes and healthy control participants. $**p < 0.01$, $****p < 0.0001$.

Generalization to an independent dataset. To ensure generalization, we repeated the analyses using the first release of the REST-meta-MDD project from the DIRECT fMRI dataset (REST-meta-MDD I, 1276 MDD participants, 1104 HCs). As expected, significantly reduced clinical characteristics including age ($t = 12.49$, $p < 0.0001$, FDR-correction), education ($t = 3.302$, $p = 0.004$, FDR-correction), first episode ($t = 11.90$, $p < 0.0001$, FDR-correction), medication status ($t = 55.49$, $p < 0.0001$, FDR-correction), HAM-D-17 ($t = 5.881$, $p < 0.0001$, FDR-correction) in shared features compared with MDD-specific features (Fig. 2a). These findings were in accordance with those observed in the REST-meta-MDD II dataset. Furthermore, we consistently identified two MDD subtypes (465 patients with subtype1 and 811 patients with subtype2) in the REST-meta-MDD I dataset (Fig. 2b). We observed a significant increase in fALFF in middle occipital areas, middle temporal areas, and inferior parietal areas in subtype 1 compared with HCs and a significant decrease in fALFF in frontal areas, parietal areas, middle temporal areas and middle occipital areas, aligning with the patterns identified for subtype 1 and subtype 2 in the REST-meta-MDD II dataset (Fig. 2c). However, despite the high similarity in patterns of subtype 1 and subtype 2 across both datasets, there are also notable differences. For instance, in REST-meta-MDD I, subtype 1 showed decreased fALFF in the cerebellum areas, which was not observed in REST-meta-MDD II.

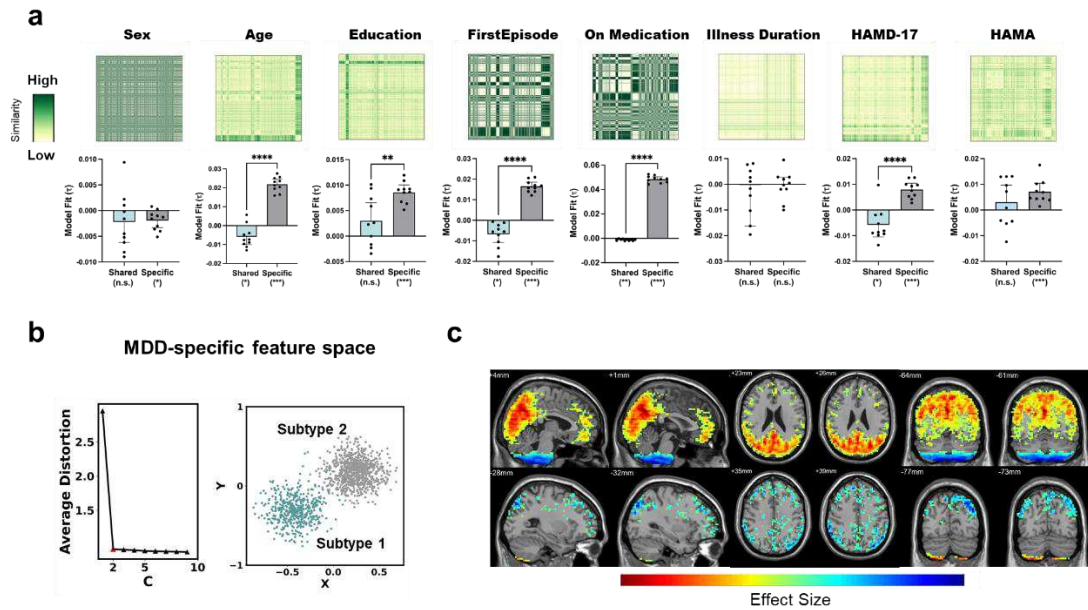


Fig. 2 The generalization of subtyping analyses in an independent dataset (REST-meta-MDD I). **a** Similarity of neurofunctional matrices is compared to the similarity based on different participant characteristics. The MDD-specific features are more effective in capturing variables associated with MDD-related variation compared to shared features. **b** Optimal clustering. MDD-specific features are best captured by two clusters. Scatterplots show individual subjects' fALFF data from subtype 1 (green) and subtype 2 (gray) groups projected onto uniform manifold approximation and projection dimensions computed from the MDD-specific features. **c** False discovery rate (FDR)-corrected voxel-wise comparison of fALFF volume differences between subtype 1 (top row) and subtype 2 (bottom row) and controls are displayed on transverse, sagittal, and coronal sections. The color bar illustrates the degree of group differences (t -statistic) between distinct MDD subtypes and healthy control participants. $**p < 0.01$, $****p < 0.0001$.

Clinical symptoms across subtypes. To investigate the relationship between clinical

symptoms and distinct subtypes, we compared HAMD total scores and each item of HAMD using linear mixed models (LMMs) between subtype 1 and subtype 2 on the REST-meta-MDD II. Significantly reduced agitation symptom (HAMD9, $t = -2.4578$, $p = 0.0142$, $d = -0.1131$, un-corrected) and significantly enhanced weight loss symptom (HAMD16, $t = 2.3776$, $p = 0.0176$, $d = 0.2416$, un-corrected) were observed in subtype 1 compared with subtype 2 (Tab. 1).

Interaction between different subtypes and antidepressant response. Through the comparison of clinical outcomes after antidepressant and rTMS treatment between different subtypes of patients, we could gain valuable insights into individual patient responses to the treatment and provide important information for personalized therapy. We initially compared the response rate to antidepressant treatment between subtype 1 and subtype 2. Patients with subtype 1 was characterized by superior efficacy across all medication (response rate of subtype 1 = 70.33%, response rate of subtype 2 = 44.44%; $\chi^2 = 13.79$, $p < 0.001$, FDR-correction), Medication-free (response rate of subtype 1 = 76.47%, response rate of subtype 2 = 54.26%; $\chi^2 = 10.64$, $p = 0.0017$, FDR-correction), SSRI (Selective Serotonin Reuptake Inhibitor) medication (response rate of subtype 1 = 50%, response rate of subtype 2 = 34.21%; $\chi^2 = 5.255$, $p = 0.0219$, FDR-correction) and SNRI (Serotonin and Norepinephrine Reuptake Inhibitor) medication (response rate of subtype 1 = 100%, response rate of subtype 2 = 30%; $\chi^2 = 107.7$, $p < 0.0001$, FDR-correction) (Fig. 3a). For specific medications, patients with subtype 1 exhibited superior efficacy compared to patients with subtype 2 with SSRI Citalopram (response rate of subtype 1 = 50%, response rate of subtype 2 = 9.52%; χ^2

= 38.1, $p < 0.0001$, FDR-correction), SSRI Escitalopram (response rate of subtype 1 = 75%, response rate of subtype 2 = 28.57%; $\chi^2 = 42.39$, $p < 0.0001$, FDR-correction), SSRI Paroxetine (response rate of subtype 1 = 100%, response rate of subtype 2 = 42.11%; $\chi^2 = 81.69$, $p < 0.0001$, FDR-correction), SNRI Venlafaxine (response rate of subtype 1 = 66.67%, response rate of subtype 2 = 28.57%; $\chi^2 = 28.93$, $p < 0.0001$, FDR-correction), while patients with subtype 2 showed better treatment efficacy following AAD Mirtazapine (response rate of subtype 1 = 50%, response rate of subtype 2 = 69.23%; $\chi^2 = 7.49$, $p = 0.0074$, FDR-correction) and SSRI Sertraline (response rate of subtype 1 = 0%, response rate of subtype 2 = 33.33%; $\chi^2 = 39.52$, $p < 0.0001$, FDR-correction) intervention (Fig. 3b).

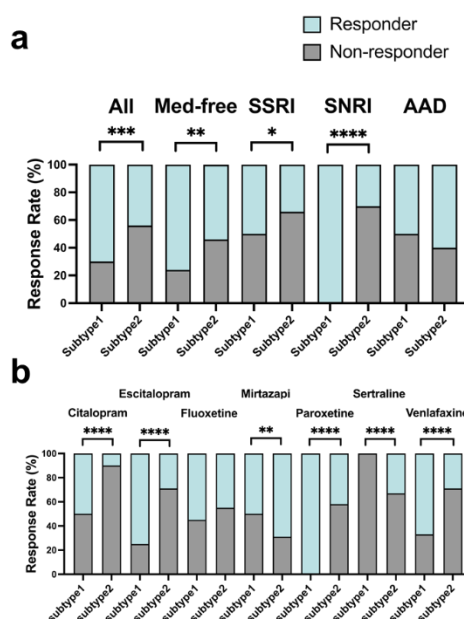


Fig. 3 Response rate across the subtypes and different medication types on the REST-meta-MDD II. a Comparison of response rate for all medication, medication-free, SSRI medication, SNRI medication and AAD medication between both subtypes. **b** Comparison of response rate for specific medications including Citalopram, Escitalopram, Fluoxetine, Mirtazapine, Paroxetine, Sertraline and Venlafaxine between

both subtypes. Chi-square tests were used to compare the medication outcomes (responder rate vs. non-responder rate) between different subtypes. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$, $****p < 0.0001$.

Interaction between different subtypes and treatment outcome for rTMS. We further compared the differences in 17-item Hamilton Depression Rating Scale (HAM-D-17), Montgomery-Asberg Depression Rating Scale (MADRS), and Beck Depression Inventory-II (BDI-II) scores between subtype 1 and subtype 2 after undergoing rTMS treatment for 5 days, 15 days, and 30 days on another independent dataset (SAINT). As shown in Fig. 4, patients with subtype 1 showed significantly enhanced HAM-D-17 changes after undergoing rTMS treatment for 5 days (subtype 1: 20.5 ± 3.8 , subtype2: 16.6 ± 5.1 ; $t = 2.807$, $p = 0.0247$, $d = 0.829$, FDR-correction) and 30 days (subtype 1: 24.2 ± 5.2 , subtype2: 19.1 ± 2.1 ; $t = 3.808$, $p = 0.009$, $d = 1.256$, FDR-correction) compared with patients with subtype 2. Furthermore, significantly increased MADRS changes after undergoing rTMS treatment for 30 days (subtype 1: 29.2 ± 4.3 , subtype2: 25.1 ± 2.9 ; $t = 3.108$, $p = 0.018$, $d = 1.07$, FDR-correction) were observed in patients with subtype 1 compared to patients with subtype 2. Similarly, significantly increased BDI-II changes after undergoing rTMS treatment for 15 days (subtype 1: 19.8 ± 7.0 , subtype 2: 11.8 ± 13.4 ; $t = 3.061$, $p = 0.018$, $d = 0.718$, FDR-correction) and 30 days (subtype 1: 23.6 ± 7.6 , subtype 2: 17.6 ± 12.7 ; $t = 2.539$, $p = 0.036$, $d = 0.557$, FDR-correction) were observed in patients with subtype 1 compared to patients with subtype 2.

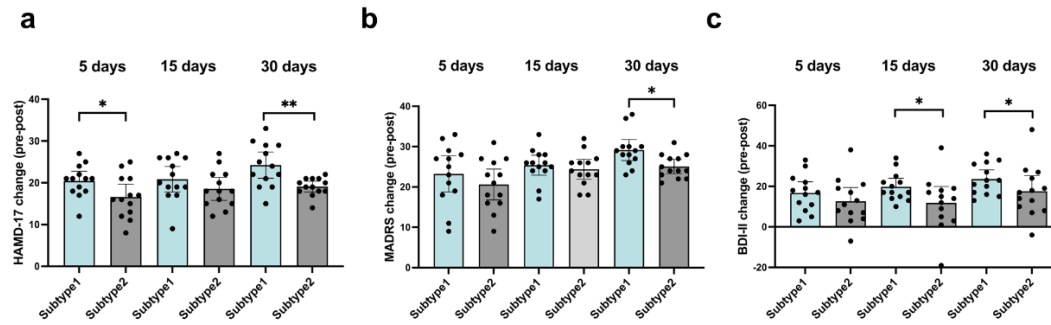


Fig. 4 Depressive symptoms across the subtypes and all three rTMS treatment groups. Difference in changes in HAMD-17, MADRS and BDI-II across two subtypes and three different treatment groups following 5 days rTMS treatment, 15 days rTMS treatment and 30 days rTMS treatment. The data is presented using a bar plot with mean values and 95th percentile error bars. The asterisks (*) indicate significant differences between the two subgroups using a two-sided Linear Mixed-Effects Model (LMM) with a p -value of less than 0.05. $*p < 0.05$, $**p < 0.01$.

Predicting outcomes of medication and rTMS through using MDD-Specific features. Considering the robust and reproducible properties of MDD-specific features, we attempted to use the MDD-specific features as a neuroimaging biomarker to predict treatment outcomes of medication and rTMS. To this end, we trained logistic regression (LR) classifiers to predict the diagnostic status of each individual (224 responders for all medications versus 174 non-responders for all medications; 41 responders for SSRI medication versus 65 non-responders for SSRI medication; 155 responders for medication-free versus 75 non-responders for medication-free). Figure 5a shows sensitivity-specificity and precision-recall curves based on model predictions for the all medications, SSRI medication and medication-free, respectively. The implementation

of 100 repetitions of 10-fold cross-validation yielded diagnostic status predictions with an accuracy of over 90% (All medication: accuracy, 0.90 ± 0.01 , AUC, 0.90 ± 0.05 , specificity, 0.89 ± 0.01 , MCC, 0.80 ± 0.01 , permutation test-validated using 1,000 permutations, $p < 0.001$; SSRI medication: accuracy, 0.99 ± 0.01 , AUC, 0.99 ± 0.01 , specificity, 0.99 ± 0.01 , MCC, 0.98 ± 0.01 , permutation test-validated using 1,000 permutations, $p < 0.001$; Medication-free: accuracy, 0.96 ± 0.01 , AUC, 0.96 ± 0.04 , specificity, 0.96 ± 0.01 , MCC, 0.98 ± 0.01 , permutation test-validated using 1,000 permutations, $p < 0.001$).

Furthermore, we further developed support vector regression (SVR) model to predict the changes in HAMD-17, MADRS, and BDI-II scores following 5, 15, and 30 days of rTMS treatment. For the HAMD-17 score, MDD-specific features were significantly predictive of the treatment outcome when performing 100 repetitions of 5-fold cross-validation (Fig. 5b; 5 days: Pearson's $r = 0.80 \pm 0.01$, MAE = 3.20 ± 0.01 , RMSE = 3.90 ± 0.01 , permutation test-validated using 1,000 permutations, $p < 0.001$; 15 days: Pearson's $r = 0.82 \pm 0.01$, MAE = 3.28 ± 0.01 , RMSE = 3.97 ± 0.01 , permutation test-validated using 1,000 permutations, $p < 0.001$; 30 days: Pearson's $r = 0.86 \pm 0.01$, MAE = 2.86 ± 0.01 , RMSE = 3.75 ± 0.02 , permutation test-validated using 1,000 permutations, $p < 0.001$). The results for MADRS and BDI-II can be found in Tab. 1. Our predictive model showed excellent performance and was not impacted by treatment duration.

In addition, our findings reveal a significant correlation between the dimensionality of MDD-specific features and diagnostic and regression results. Optimal results were

achieved at a dimensionality of 512, while a dimensionality of only 32 greatly reduced prediction accuracy. This indicates that the dimensionality of MDD-specific features plays a crucial role in treatment outcomes, with each feature representing a distinct interindividual variance.

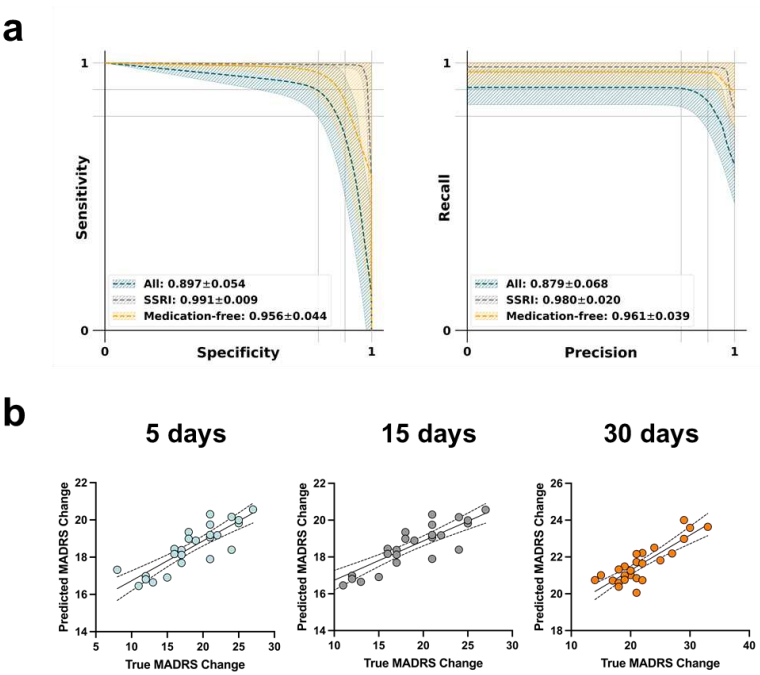


Fig. 5 MDD-specific features disentangled from CVAE characterize antidepressant response. **a** Sensitivity-specificity and precision-recall curves were generated based on the model's predictions of responder versus non-responder for all medications, SSRI medications, and medication-free individuals. **b** A 100×5 stratified cross-validation prediction of HAMD-17 change following 5, 15, and 30 days of rTMS treatment ($n = 26$).

Tab. 1 A 100×5 stratified cross-validation prediction of HAMD-17, MADRS and BDI-II changes following 5, 15, and 30 days of rTMS treatment. MAE, Mean Absolute Error; RMSE, Root Mean Squared Error; PCC, Pearson's Correlation

Coefficient; p , p -value of permutation test-validated using 1,000 permutations.

Measures	HAMD-17			MADRS			BDI-II		
	5 days	15 days	30 days	5 days	15 days	30 days	5 days	15 days	30 days
MAE	3.20(0.01)	3.28(0.01)	2.86(0.01)	4.81(0.02)	2.65(0.01)	2.75(0.01)	6.55(0.05)	6.63(0.04)	6.84(0.05)
RMSE	3.90(0.02)	3.97(0.01)	3.75(0.02)	5.68(0.04)	3.44(0.01)	3.57(0.01)	8.35(0.05)	9.75(0.04)	8.78(0.08)
PCC	0.80(0.01)	0.82(0.01)	0.86(0.01)	0.76(0.01)	0.66(0.01)	0.71(0.01)	0.84(0.01)	0.73(0.01)	0.83(0.01)
p	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001

Discussion

In this study, we developed modified CVAEs for MDD subtyping and investigated the sensitivity of the ensuing subtyping to antidepressant treatments among various subtypes of patients. CVAE identified two MDD subtypes. Patients with subtype 1 showed significantly increased fALFF in middle occipital areas, parietal areas, middle temporal areas and inferior frontal areas compared with HCs, while patients with subtype 2 demonstrated significantly decreased fALFF in middle occipital areas, parietal areas, middle and superior temporal areas and frontal areas compared to HCs. The number and patterns of subtypes were validated on an independent large multicenter dataset. Patients with subtype 1 could be considered as the "treatment-sensitive" group, as the response rate to all antidepressants was above 50% and the effectiveness of rTMS was better than that of patients with subtype 2. In contrast, patients with subtype 2 could be classified as the "treatment-resistant" group, as the response rate only to Atypical Antidepressant (AAD) medication was 60%, while the response rate to other medications was below 50%. Finally, we found that the MDD-specific features separated by CVAE can be used as a neuroimaging biomarker for predicting treatment outcomes for both medication and rTMS treatments.

CVAE identifies two reproducible subtypes. We identified 2 MDD subtypes, which is in line with previous studies. According to Tab. S1, out of the 11 studies, nine identified two MDD subtypes. Moreover, in addition to MRI studies, Zhang et al.²⁸ also used high-density resting-state electroencephalography to identify two MDD subtypes and discovered unique functional connectivity patterns in the frontoparietal control network and default mode network (DMN). In major psychiatric disorders such as schizophrenia (SZ), bipolar disorder (BD) and MDD, two subtypes were also identified using whole-brain ALFF²⁴. Based on multimodal imaging data and different data-driven methods, these findings indicate that two MDD subtypes are best aligned with the data distribution and the pathophysiology of MDD.

To ensure the reproducibility of the findings, we repeated the data-driven analyses using an independent dataset and found two MDD subtypes again. In recent years, there has been a lot of criticism surrounding the reproducibility of neuroimaging studies²⁹. The primary contributing factors to this issue are the typically small sample sizes used in neuroimaging studies, which can result in low sensitivity and reliability^{30,31}, as well as a lack of external validation. By utilizing two large multi-site fMRI datasets with over 5000 participants, this present study ensures the reproducibility and robustness of the results. Additionally, it expands our understanding of neurophysiological MDD subtypes.

The response to treatment varied between different subtypes. We further examined

the relationship between distinct subtypes and treatment response. Patients with subtype 1 showed a response rate of over 50% to all medication treatments, with an overall response rate of 70.27%. However, patients with subtype 2 were generally unresponsive to most medications, with only AAD showing an effectiveness of 60%. These results align with a recent study that used a semi-supervised machine-learning method called heterogeneity through discriminative analysis (HYDRA) to identify 2 MDD subtypes ¹⁵. In that study, subtype 1 patients showed a significant improvement in depressive symptoms after receiving treatment with selective serotonin reuptake inhibitor (SSRI) medication, while subtype 2 patients did not. Our study expands upon this research by delving deeper into the effects of different medication types, including Serotonin and Norepinephrine Reuptake Inhibitor (SNRI) and Atypical Antidepressant (AAD), as well as specific medications such as Citalopram, Escitalopram, Paroxetine, Venlafaxine, Mirtazapine and Sertraline. Our findings provide a more comprehensive understanding of the subtypes and their responses to a variety of medications. This knowledge can also aid doctors in creating effective treatment plans. For instance, patients classified as subtype 1 may respond well to conventional medications like SSRI and SNRI, while those classified as subtype 2 may benefit from alternative medications such as Mirtazapine. If medication is not effective for patients with subtype 2, other treatments like rTMS can be explored. This can be particularly helpful in reducing the duration of treatment and cost burden for patients. Interestingly, despite not receiving medication, patients identified as subtype 1 still had a response rate of 70.27%, whereas those identified as subtype 2 only had a response rate of 45.73%. This discovery could

potentially disclose the underlying reasons for the slightly higher effectiveness of antidepressants compared to placebos ⁷, as well as provide insight into potential self-healing mechanisms for MDD ³².

Moreover, significantly greater treatment responses to rTMS treatment were observed in patients with subtype 1 compared with patients with subtype 2. This finding is in line with a prior study that used hierarchical clustering to identify 4 MDD subtypes, where subtype 1 also showed significant improvement in HAMD scores compared to the other subtypes¹¹. We further investigated whether this difference exists longitudinally and does not change over time. Specifically, we compared the changes in HAMD-17, MADRS, and BDI-II scores between subtype 1 and subtype 2 after undergoing rTMS treatment for 5 days, 15 days, and 30 days. We found that the superiority of subtype 1 in responding to rTMS treatment remained consistent over time.

MDD-Specific features disentangled from CVAEs characterize antidepressant response. CVAEs could disentangle MDD-specific features and shared features using contrastive learning. MDD-specific features were significantly correlated with clinical variations in MDD compared with shared features. More importantly, the clinical property of MDD-specific features could be generalized to another independent dataset. These results indicate that MDD-Specific features disentangled from CVAEs are necessarily related to the underlying nature of MDD. This generalized characteristic enables its use in diagnostic settings, where a model trained on prior cases can be utilized to analyze data from new individuals. Thus, we further utilized MDD-specific

features to predict treatment outcomes. The results showed that MDD-specific features could effectively differentiate between responders and non-responders for medication treatment, and could also accurately predict changes in clinical indicators (HAMD-17, MADRS and BDI-II) after rTMS treatment. This finding suggests that MDD-specific features have the potential to serve as neuroimaging biomarkers and aid clinicians in early identification of participants with Treatment-Resistant Depression (TRD)³³, informing individualized treatment decisions.

Limitations. The limitations of this study should be considered. First, external validation is needed to further investigate the differences in treatment response between different subtypes and the potential of MDD-specific features as neuroimaging biomarkers for predicting the treatment outcomes. Second, our sample size for treatment data was small, with only 26 participants used to study the effects of rTMS treatment. Finally, additional research is needed to determine the applicability of our method for other mental disorders.

Competing interests

All authors declare that they have no competing interests.

Funding

This work was supported by the National Natural Science Foundation of China (62088102, U21A20485, 62311540022).

Data availability

Deidentified and anonymized data were contributed from studies approved by local Institutional Review Boards. All study subjects provided written informed consent at their local institution. Data of REST-meta-MDD I are available at: <http://rfmri.org/REST-meta-MDD>. Data of REST-meta-MDD II and SAINT used during the current study are available from the corresponding author upon reasonable request.

Code availability

Codes used during the current study are available at GitHub repositior (<https://github.com/ZKZ-Brain/Contrastive-Learning-for-MDD-subtyping>).

Methods

Participants. Data of REST-meta-MDD I and REST-meta-MDD II collected by REST-meta-MDD Project from the Depression Imaging REsearch ConsorTium (DIRECT) are currently the largest resting-state fMRI datasets for MDD. In this study, we used a sample of ($N = 3000$, 1660 MDDs/1340 HCs) on the REST-meta-MDD II and a sample of ($N = 2380$, 1276 MDDs/1104 HCs) on the REST-meta-MDD I. Demographic data for REST-meta-MDD I and REST-meta-MDD II are presented in Tab. 1 and 2, respectively.

SAINT dataset is collected by the Department of Psychiatry at the First Affiliated

Hospital, Air Force Military University. All participants received the Stanford Accelerated Intelligent Neuromodulation Therapy (SAINT ³⁴). In this study, we used a sample of ($N = 26$, 26 pre-treatment MDDs and corresponding post-treatment MDDs) on the SAINT. Demographic data for SAINT are presented in Tab. 3. More details are described in Supplementary Material.

Tab. 1 Demographic and clinical variables for MDD and healthy control participants on the REST-meta-MDD II.

	Healthy controls	MDD participants	MDD subtype1	MDD subtype2
Sample size	1340	1660	915	745
Age (yr)	32.0	32.9	28.9	37.1
Age range (yr)	9-67	11-77	11-67	12-77
Sex				
Male (number, percentage)	570(42.5)	599 (36.0)	323 (35.3)	276 (37.0)
Female (number, Percentage)	770 (57.5)	1057 (63.6)	588 (64.2)	469 (63.0)
Years of education	13.6 (12.9)	12.0 (13.8)	12.4 (12.8)	11.5 (14.7)
HAMD-17				
Total	—	22.4 (62.9)	22.2 (56.0)	22.8 (76.7)
HAMD1	—	2.5 (0.9)	2.7 (0.7)	2.4 (1.0)
HAMD2	—	1.2 (1.0)	1.2 (1.0)	1.0 (0.9)
HAMD3	—	1.3 (1.5)	1.5 (1.5)	1.2 (1.4)
HAMD4	—	1.3 (0.7)	1.3 (0.7)	1.3 (0.7)
HAMD5	—	1.2 (0.6)	1.2 (0.6)	1.2 (0.7)
HAMD6	—	1.1 (0.7)	1.1 (0.7)	1.1 (0.8)
HAMD7	—	2.3 (1.0)	2.4 (0.9)	2.2 (1.1)
HAMD8	—	1.1 (0.8)	1.0 (0.8)	1.1 (0.8)
HAMD9	—	0.7 (0.7)	0.7 (0.6)	0.8 (0.8)

HAMD10	—	1.8 (0.9)	1.8 (0.9)	1.7 (1.0)
HAMD11	—	1.5 (1.0)	1.5 (0.9)	1.5 (1.0)
HAMD12	—	0.8 (0.5)	0.8 (0.4)	0.8 (0.6)
HAMD13	—	1.1 (0.6)	1.1 (0.6)	1.0 (0.7)
HAMD14	—	0.6 (0.6)	0.6 (0.6)	0.6 (0.7)
HAMD15	—	0.8 (0.8)	0.8 (0.8)	0.8 (0.8)
HAMD16	—	0.6 (0.5)	0.6 (0.6)	0.5 (0.5)
HAMD17	—	0.4 (0.4)	0.5 (0.4)	0.4 (0.4)
HAMA	—	18.1 (68.5)	18.1 (63.5)	18.1 (74.3)
Illness Duration (mo)	—	42.3 (4201.6)	28.4 (2298.0)	56.6 (5766.5)
First episode/recurrent MDD	—	1034/456	641/159	393/297
Family History	—	97	41	56
Prospective treatment sample				
Total (number of participants)				
Treatment responsive	—	224/174	128/54	96/120
Medication-free	—	155/75	104/32	51/43
SSRI	—	41/65	15/15	26/50
SNRI	—	8/14	2/0	6/14
AAD	—	10/8	4/4	6/4
Citalopram	—	4/21	2/2	2/19
Escitalopram	—	9/16	3/1	6/15
Fluoxetine	—	16/17	6/5	10/12
Mirtazapine	—	13/8	4/4	9/4
Paroxetine	—	11/11	3/0	8/11
Sertraline	—	3/13	0/7	3/6
Venlafaxine	—	8/16	2/1	6/15

Tab. 2 Demographic and clinical variables for MDD and normal control participants on the REST-meta-MDD I.

	Healthy controls	MDD participants	MDD subtype1	MDD subtype2
Sample size	1104	1276	811	465
Age (yr)	36.2	36.2	35.6	37.3
Age range (yr)	14-80	12-82	12-80	14-82
Sex				
Male (number, percentage)	462 (41.8%)	463 (36.3%)	301 (37.1%)	162 (34.9%)
Female (number, Percentage)	642 (58.2%)	813 (63.7%)	510 (62.9%)	303 (65.1%)
Years of education	12.3 (24.8)	11.2 (17.7)	11.3 (16.8)	11.1 (19.3)
HAMD-17	—	20.5 (58.7)	20.6 (57.2)	20.3 (61.3)
HAMA	—	19.0 (81.2)	19.0 (78.1)	19.0 (90.3)
Illness Duration (mo)	—	36.5 (3562.2)	37.5 (3751.4)	34.5 (3185.9)
First episode/recurrent MDD	—	538/282	396/199	142/83
Medication (on/off)	—	408/447	210/339	198/108

Tab. 3 Demographic and clinical variables for MDD participants on the SAINT.

	MDD participants	MDD subtype1	MDD subtype2
Sample size	26	11	15
Age (yr)	27.8	29.0	26.9
Age range (yr)	18-54	18-54	18-54
Sex			
Male (number, percentage)	6 (23.0%)	3 (27.2%)	3 (20.0%)
Female (number, Percentage)	20 (77.0%)	8 (72.7%)	12 (80.0%)
Years of education	12.2 (7.1)	12.0 (12.0)	12.4 (4.0)
Baseline			
HAMD-17	27.9 (17.0)	27.0 (17.2)	28.6 (16.8)
MADRS	36.6 (23.6)	36.0 (15.6)	37.0 (30.6)
BDI-II	36.0 (82.0)	37.8 (91.2)	34.7 (77.1)
5 days after treatment			
HAMD-17	9.4 (31.8)	6.9 (16.9)	11.2 (36.5)

MADRS	14.7(57.1)	12.1 (18.3)	16.5 (79.9)
BDI-II	21.3 (128.6)	21.2 (165.8)	21.5 (111.3)
15 days after treatment			
HAMD-17	8.2 (24.3)	7.8 (20.0)	8.5 (28.8)
MADRS	11.7 (37.9)	11.3 (27.6)	12.0 (47.9)
BDI-II	20.2 (124.9)	19.0 (157.2)	21.1 (108.7)
30 days after treatment			
HAMD-17	6.3(20.8)	5.0(8.8)	7.2 (28.7)
MADRS	9.5(34.0)	7.7(16.4)	10.7 (44.9)
BDI-II	15.4(67.9)	15.1(86.1)	15.7 (59.5)

Preprocessing. Standard preprocessing of the REST-meta-MDD I and REST-meta-MDD II were done using the Data Processing Assistant for Resting-State fMRI (DPARF, <http://rfmri.org/DPARF>). Functional fMRI data of SAINT were preprocessed using the Graph Theoretical Network Analysis (GRETNA, <https://www.nitrc.org/projects/gretna>). Preprocessing steps encompassed discarding the initial 10 volumes, slice-timing correction, head motion correction, space normalization, and temporal bandpass filtering (0.01-0.1 Hz). The Friston 24-parameter model was employed to regress out head motion effects and linear regression was applied to eliminate the influences of white matter, cerebrospinal fluid signals, and linear trends.

Calculation of fALFF. We utilized DPARF to calculate the fractional amplitude of low-frequency fluctuations (fALFF). In particular, the filtered time series of a voxel underwent a fast Fourier transform process to convert it to the frequency domain. Subsequently, the ALFF value for the voxel was determined as the average square root of the power spectrum. To improve specificity and sensitivity, the fALFF value was

calculated as the ratio of the power in the specific frequency band (0.01 -0.1 Hz) to that of the entire detected frequency range. This approach aimed to suppress nonspecific signals in the resting-state fMRI data.

Architecture of contrastive variational autoencoder (CVAE). The CVAE model utilized an adapted architecture derived from a recent study¹⁸. The overall framework of CVAE is illustrated in Figure 6a. The inputs consisted of fALFF volumes of MDDs $M \in \mathbb{R}^{64 \times 64 \times 64}$ and HCs $H \in \mathbb{R}^{64 \times 64 \times 64}$, both resampled to a resolution of $64 \times 64 \times 64$ in their native space. The CVAE consists of three modules: an encoder q_{Φ_Z} is used to learn MDD-specific features $Z_M \in \mathbb{R}^{32 \times 1}$; an encoder q_{Φ_S} is used to learn shared features $S_M \in \mathbb{R}^{32 \times 1}$ and $S_H \in \mathbb{R}^{32 \times 1}$ and a decoder p_θ is able to reconstruct fALFF images of MDD $\hat{M} \in \mathbb{R}^{64 \times 64 \times 64}$ and HCs $\hat{H} \in \mathbb{R}^{64 \times 64 \times 64}$ from the latent 64-dimensional features. Here, q_{Φ_Z} and q_{Φ_S} are convolutional layers, while p_θ is two deconvolutional layers. For reconstructing fALFF of MDD participants, the latent 64-dimensional representation was a concatenation of the MDD-specific features and shared features (Z_M and S_M). For reconstructing fALFF of HC participants, a concatenation of the shared features (S_H) and 32 zeros were fed into p_θ .

In addition, we forced the independence between MDD-specific features Z_M and shared features S_M through minimizing the mutual information between Z_M and S_M . In general, the estimation of mutual information is commonly done by using the mutual information neural estimator (MINE), which requires the training of an auxiliary neural network f . However, the use of f can lead to instability in the estimation of mutual information between Z_M and S_M , and in some cases, even result in negative values in

the REST-meta-MDD II (see *Supplementary Material*, Fig. S1a). To address this issue, we used Hilbert-Schmidt independence criterion (HSIC) to calculate the mutual information between Z_M and S_M . HSIC is defined as the Hilbert-Schmidt norm of the cross-covariance operator between distributions in Reproducing Kernel Hilbert Space (RKHS). This allows us to directly estimate the mutual information between Z_M and S_M from data without the need for discrete probability density function (PDF) estimation or any auxiliary neural networks. The formulation of HSIC is:

$$\text{HSIC}(Z_M, S_M) = (m-1)^{-2} \text{tr}(K_{Z_M} H K_{S_M} H), \quad (1)$$

where m is number of batch size, $\text{tr}(\cdot)$ is the trace operator, $K_{Z_M} \in \mathbb{R}^{m \times m}$ and $K_{S_M} \in \mathbb{R}^{m \times m}$ are kernel matrices with $K_{Z_M}^{i,j} = k_{Z_M}(z_M^i, z_M^j)$ and $K_{S_M}^{i,j} = k_{S_M}(s_M^i, s_M^j)$, respectively and $H \in \mathbb{R}^{m \times m}$ is the centering matrix $H = I_m l_m l_m^T$. Here, we used Gaussian kernel $k(x, y) \sim \exp(-\frac{1}{2}\|x - y\|^2/\sigma^2)$ to estimate kernel matrices. In *Supplementary Material* Fig. S1 a-b, we found that the use of HSIC mutual information significantly improved the stability of training compared to MINE.

The objective of our modified CVAE can be defined:

$$\begin{aligned} L_{\text{CVAE}} = & \mathbb{E}[\|M - \hat{M}\|_F] + \mathbb{E}[\|H - \hat{H}\|_F] - \alpha \cdot \left(\mathbb{E}[D_{KL}[q_{\Phi_Z}(Z_M | M) \parallel p_\theta(Z_M)]] \right. \\ & \left. + \mathbb{E}[D_{KL}[q_{\Phi_S}(S_M | M) \parallel p_\theta(S_M)]] \right. \\ & \left. + \mathbb{E}[D_{KL}[q_{\Phi_S}(S_H | H) \parallel p_\theta(S_H)]] \right) \\ & + \beta \cdot \text{HSIC}(Z_M, S_M), \end{aligned} \quad (2)$$

where $\mathbb{E}$ is expectation, D_{KL} is KL divergence and α and β are hyperparameters.

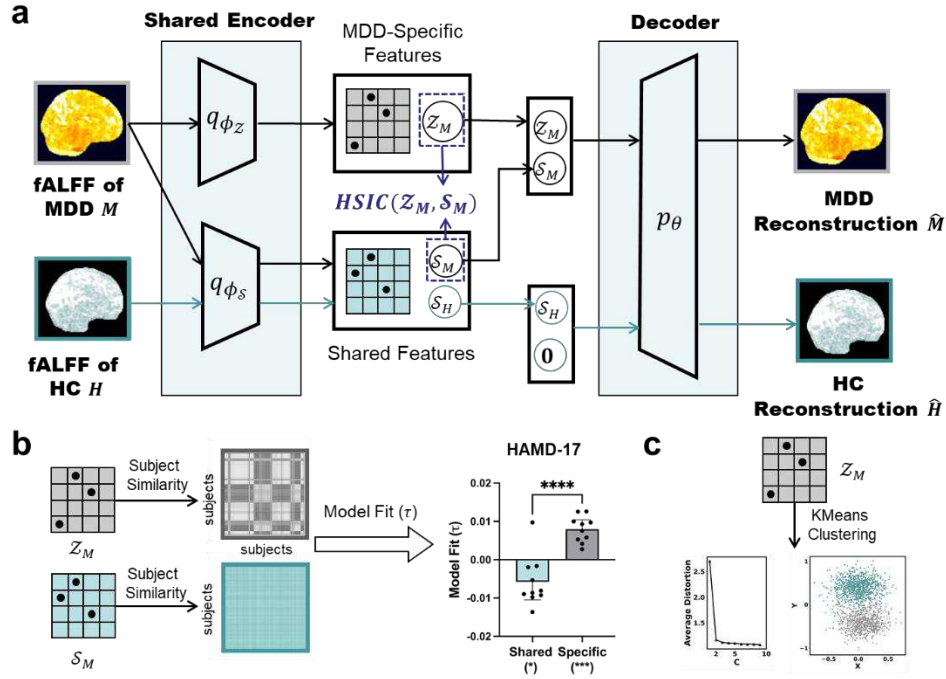


Fig. 6 MDD subtyping analysis pipeline. **a** Architecture for CVAEs. The fALFF of MDD was fed into two sets of shared encoders q_{ϕ_Z} and q_{ϕ_S} , which outputted the MDD-specific features Z_M and shared features S_M respectively. While the fALFF of HC was fed into q_{ϕ_S} , which outputted shared features S_H . The latent variables Z_M and S_M for the MDD were concatenated and fed into a decoder network, while only the (zero-padded) latent variable S_H was fed into the decoder network. The decoder generated reconstructions of the fALFF of MDD and HC. **b** Representational Similarity Analysis (RSA). MDD-specific Z_M and shared features S_M were used to generate a dissimilarity matrix to represent the pairwise dissimilarity between participants. Then the dissimilarity matrix was compared to the dissimilarity matrix for each individual characteristic using the Kendall rank correlation coefficient (Kendall τ). **c** MDD-specific features Z_M were used as input for the K-means clustering method to identify MDD subtypes.

Optimization and Implementation Details. The CVAE model was implemented using PyTorch ³⁵, and the Adam optimizer ³⁶ was used with an initial learning rate of 0.001. The model was trained for 15000 epochs with a batch size of 32. For the REST-meta-MDD I and REST-meta-MDD II datasets, we trained a new model using their respective data. In the SAINT dataset, due to a lack of a healthy control group, we trained the CVAE model using the healthy participants from the REST-meta-MDD II dataset and the pre-treatment participants from the same dataset. Hyper-parameters and parameters of CVAE for the REST-meta-MDD I and REST-meta-MDD II dataset are shown in *Supplementary material* Tab. S2-5.

Representational Similarity Analysis. To explore the type of information represented in the latent features, we employed representational similarity analysis (RSA) to investigate the potential correlation between MDD-specific features and individual variations in clinical characteristics among MDD participants, such as age, education, first episode, and scores on the HAM-D-17 and HAMA. The workflow of the RSA analysis is demonstrated in Fig. 6b. Initially, we calculated the pairwise dissimilarity between participants based on the CVAE functional features (MDD-specific and shared features) to generate a dissimilarity matrix. This process was repeated for each clinical characteristic. For ratio-valued measurements, such as age and HAM-D, the subject distance was calculated as the absolute Euclidean distance between their measurements (i.e., the difference in age or HAM-D score). As for categorical variables (such as sex or medication status), the subject distance was considered 0 if the categorical variables matched (e.g., same gender or medication status) and 1 otherwise. Finally, the CVAE

dissimilarity matrix was compared to the dissimilarity matrix for each individual characteristic using the Kendall rank correlation coefficient (Kendall τ). To ensure the reliability of the RSA analyses on various samples extracted from the CVAE's latent distributions, we randomly sampled the latent distribution for each MDD participant 10 times, which resulted in 10 different reconstructions of the dataset. RSA models then were fitted to each reconstruction separately. Please note that we deliberately limited the number of repetitions (10 times) to maintain a strict approach in these tests, as larger samples can result in significant effects even with minor differences.

Cluster Analysis. We used K-means clustering method to identify distinct MDD subtypes based on MDD-specific features. To determine the optimal number of clusters, we utilized the elbow plot method. We fitted a K-means clustering model with a range of clusters (1-10) and calculated the distortion as a measure of model fit. Distortion is defined as the average of the squared Euclidean distance from the respective cluster's centroid³⁷. Fig. 1c, Fig.2c show mean distortion for each number of clusters in REST-meta-MDD II and REST-meta-MDD I, respectively. The average distortion shows a slower decrease after $C = 2$.

Case-control analysis of fALFF. In order to obtain the functional patterns of each subtype, we conducted group comparisons of fALFF values using Statistical Parametric Mapping (SPM, <https://www.fil.ion.ucl.ac.uk/spm/software/spm12/>). For each voxel, a general linear model (GLM) was used to investigate the difference in fALFF between each subtype and HC. The GLM analyses included gender, age, and education as covariates, with the primary focus on the effect of group on fALFF. Significance was

determined by individual voxel $p < 0.05$ with FDR correction.

Predictive modeling of treatment responses using machine learning approaches.

Logistic regression classifiers³⁸ were used to distinguish between patients who responded positively to medication treatment and those who did not. In this study, REST-meta-MDD II encompassed 224 responders versus 174 non-responders for all medication, 41 responders versus 65 non-responders for SSRI medication and 155 responders versus 75 non-responders for medication-free. Specifically, to mitigate overfitting to noise in the data and prevent inflation of prediction performance, we utilized the recursive feature elimination method³⁹ for feature selection, reducing the dimensionality of MDD-specific features from 512 to 175. Then 175 dimensions of MDD-specific features were fed into logistic regression classifiers. In addition, we used the grid search function to select the optimal hyper-parameters (see *Supplementary materials* Tab. S6). Finally, we implemented a 100×10 stratified cross-validation to assess performance and accuracy, area under the curve (AUC), specificity, Matthews Correlation Coefficient (MCC)⁴⁰ were used as evaluating indicators. Demographic and clinical variables for responders and non-responders for all medication, medication-free and SSRI medication classification are shown in *Supplementary Material*, Tab. S7-9.

Furthermore, support vector regression (SVR⁴¹) was used to predict clinical improvements observed 5 days, 10 days, 30 days after rTMS treatment. Clinical improvements were quantified by subtracting the post-treatment scores from the pre-treatment scores. These clinical scales include the HAM-D-17⁴², MADRS and BDI-II scores. SelectKBest⁴³ was employed to reduce the dimension of MDD-specific features

from 512 to 175. Then 175 dimensions of MDD-specific features were fed into SVR model. In addition, we used the grid search function to select the optimal hyper-parameters (see Supplementary materials Tab. S10). A 100×5 stratified cross-validation to assess performance using the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE) and Pearson's correlation coefficient (PCC).

Statistics. Linear mixed models (LMM) were employed to compare the clinical variables (HAMD-17 total scores and each item of HAMD-17) and rTMS treatment outcomes (differences in HAMD-17, MADRS, and BDI-II after undergoing rTMS treatment for 5 days, 15 days, and 30 days) between distinct subtypes. The model was defined as $y \sim 1 + \text{Diagnosis} + \text{Age} + \text{Sex} + \text{Education} + \text{Motion} + (1 | \text{Site}) + (\text{Diagnosis} | \text{Site})$. This model provided t and p values for the fixed effect of Diagnosis. Cohen's d effect size was also calculated as $d = \frac{T(n_1+n_2)}{\sqrt{df}\sqrt{n_1n_2}}$ ⁴⁴. The diagnosis variable of LMM was defined as the subtype status. In addition, chi-square tests were used to compare the medication outcomes (responder rate vs. non-responder rate) between different subtypes.

To determine the statistical significance of the machine learning based treatment prediction results, a nonparametric permutation test was employed. In the medication task, we randomly shuffle the labels of medication treatment response across participants 1,000 times, then repeat the cross-validated classification procedure to obtain an accuracy distribution. The p value is defined as the proportion of cross-validated accuracy that is higher than those without permutation. Similarly, in the rTMS task, we randomly shuffle the observed HAMD-17, MADRS, and BDI-II score changes

across participants 1,000 times, then repeat the cross-validated classification procedure to obtain a distribution of Pearson's correlation coefficients. The p value is defined as the proportion of correlation coefficients that are higher than those without permutation.

References

1. Kessler, R.C. The costs of depression. *Psychiatric Clinics* **35**, 1-14 (2012).
2. Vos, T., *et al.* Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *The lancet* **396**, 1204-1222 (2020).
3. James, S.L., *et al.* Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *The Lancet* **392**, 1789-1858 (2018).
4. Vos, T., *et al.* Years lived with disability (YLDs) for 1160 sequelae of 289 diseases and injuries 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The lancet* **380**, 2163-2196 (2012).
5. Rush, A.J. The varied clinical presentations of major depressive disorder. *Journal of Clinical Psychiatry* **68**, 4 (2007).
6. Cipriani, A., *et al.* Comparative efficacy and acceptability of 21 antidepressant drugs for the acute treatment of adults with major depressive disorder: a systematic review and network meta-analysis. *The Lancet* **391**, 1357-1366 (2018).
7. Fournier, J.C., *et al.* Antidepressant drug effects and depression severity: a patient-level meta-analysis. *Jama* **303**, 47-53 (2010).
8. Tian, Y.E. Toward Reproducible, Generalizable, and Clinically Useful Neurophysiological Subtypes of Major Depressive Disorder. *Biological Psychiatry* **94**, e45-e47 (2023).
9. Wen, J., *et al.* Characterizing heterogeneity in neuroimaging, cognition, clinical symptoms, and genetics among patients with late-life depression. *JAMA psychiatry* **79**, 464-474 (2022).
10. Lalouis, P.A., *et al.* Neurobiologically based stratification of recent-onset depression and psychosis: identification of two distinct transdiagnostic phenotypes. *Biological psychiatry* **92**, 552-562 (2022).
11. Drysdale, A.T., *et al.* Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nature medicine* **23**, 28-38 (2017).
12. Fu, C.H., Fan, Y. & Davatzikos, C. Addressing heterogeneity (and homogeneity) in treatment mechanisms in depression and the potential to develop diagnostic and predictive biomarkers. *NeuroImage: Clinical* **24**, 101997 (2019).
13. Bouchard Jr, T.J. & McGue, M. Genetic and environmental influences on human psychological differences. *Journal of neurobiology* **54**, 4-45 (2003).

14. Yarkoni, T., Poldrack, R.A., Nichols, T.E., Van Essen, D.C. & Wager, T.D. Large-scale automated synthesis of human functional neuroimaging data. *Nature methods* **8**, 665-670 (2011).
15. Fu, C.H., *et al.* Neuroanatomical dimensions in medication-free individuals with major depressive disorder and treatment response to SSRI antidepressant medications or placebo. *Nature Mental Health*, 1-13 (2024).
16. Chuang, C.-Y., Robinson, J., Lin, Y.-C., Torralba, A. & Jegelka, S. Debaised contrastive learning. *Advances in neural information processing systems* **33**, 8765-8775 (2020).
17. Tian, Y., *et al.* What makes for good views for contrastive learning? *Advances in neural information processing systems* **33**, 6827-6839 (2020).
18. Aglinskias, A., Hartshorne, J.K. & Anzellotti, S. Contrastive machine learning reveals the structure of neuroanatomical variation within autism. *Science* **376**, 1070-1074 (2022).
19. Sevenson, K.A., Ghosh, S. & Ng, K. Unsupervised learning with contrastive latent variable models. in *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33 4862-4869 (2019).
20. Abid, A. & Zou, J. Contrastive variational autoencoder enhances salient features. *arXiv preprint arXiv:1902.04601* (2019).
21. Yu-Feng, Z., *et al.* Altered baseline brain activity in children with ADHD revealed by resting-state functional MRI. *Brain and Development* **29**, 83-91 (2007).
22. Zuo, X.-N. & Xing, X.-X. Test-retest reliabilities of resting-state FMRI measurements in human brain functional connectomics: a systems neuroscience perspective. *Neuroscience & Biobehavioral Reviews* **45**, 100-118 (2014).
23. Zuo, X.-N., Xu, T. & Milham, M.P. Harnessing reliability for neuroscience research. *Nature human behaviour* **3**, 768-771 (2019).
24. Chang, M., *et al.* Identifying and validating subtypes within major psychiatric disorders based on frontal-posterior functional imbalance via deep learning. *Molecular psychiatry* **26**, 2991-3002 (2021).
25. Chen, X., *et al.* The DIRECT consortium and the REST-meta-MDD project: towards neuroimaging biomarkers of major depressive disorder. *Psychoradiology* **2**, 32-42 (2022).
26. Nili, H., *et al.* A toolbox for representational similarity analysis. *PLoS computational biology* **10**, e1003553 (2014).
27. Kriegeskorte, N., Mur, M. & Bandettini, P.A. Representational similarity analysis-connecting the branches of systems neuroscience. *Frontiers in systems neuroscience* **2**, 249 (2008).
28. Zhang, Y., *et al.* Identification of psychiatric disorder subtypes from functional connectivity patterns in resting-state electroencephalography. *Nature biomedical engineering* **5**, 309-323 (2021).
29. Botvinik-Nezer, R. & Wager, T.D. Reproducibility in neuroimaging analysis: challenges and solutions. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging* **8**, 780-788 (2023).
30. Button, K.S., *et al.* Power failure: why small sample size undermines the reliability of neuroscience. *Nature reviews neuroscience* **14**, 365-376 (2013).

31. Chen, X., Lu, B. & Yan, C.G. Reproducibility of R - fMRI metrics on the impact of different strategies for multiple comparison correction and sample sizes. *Human brain mapping* **39**, 300-318 (2018).
32. Xin, Z. & Li, S. Analysis of self-healing of depression by helping others in adolescents from the perspective of constructivism. *Frontiers in Psychiatry* **14**, 1201923 (2023).
33. Kim, E., Bagalman, J. & Goetzel, R. Cost burden of treatment resistance in patients with depression. *Am J Manag Care* **16**, 370-377 (2010).
34. Williams, N.R., *et al.* High-dose spaced theta-burst TMS as a rapid-acting antidepressant in highly refractory depression. *Brain* **141**, e18-e18 (2018).
35. Imambi, S., Prakash, K.B. & Kanagachidambaresan, G. PyTorch. *Programming with TensorFlow: Solution for Edge Computing Applications*, 87-104 (2021).
36. Zhang, Z. Improved adam optimizer for deep neural networks. in *2018 IEEE/ACM 26th international symposium on quality of service (IWQoS)* 1-2 (Ieee, 2018).
37. Daviran, M., Ghezelbash, R., Niknezhad, M., Maghsoudi, A. & Ghaeminejad, H. Hybridizing K-means clustering algorithm with harmony search and artificial bee colony optimizers for intelligence mineral prospectivity mapping. *Earth Science Informatics* **16**, 2143-2165 (2023).
38. Nick, T.G. & Campbell, K.M. Logistic regression. *Topics in biostatistics*, 273-301 (2007).
39. Chen, X.-w. & Jeong, J.C. Enhanced recursive feature elimination. in *Sixth international conference on machine learning and applications (ICMLA 2007)* 429-435 (IEEE, 2007).
40. Chicco, D., Tötsch, N. & Jurman, G. The Matthews correlation coefficient (MCC) is more reliable than balanced accuracy, bookmaker informedness, and markedness in two-class confusion matrix evaluation. *BioData mining* **14**, 1-22 (2021).
41. Awad, M., Khanna, R., Awad, M. & Khanna, R. Support vector regression. *Efficient learning machines: Theories, concepts, and applications for engineers and system designers*, 67-80 (2015).
42. Ballesteros, J., *et al.* Sensitivity to change, discriminative performance, and cutoff criteria to define remission for embedded short scales of the Hamilton depression rating scale (HAM-D). *Journal of affective disorders* **102**, 93-99 (2007).
43. Ayyanar, M., Jeganathan, S., Parthasarathy, S., Jayaraman, V. & Lakshminarayanan, A.R. Predicting the Cardiac Diseases using SelectKBest Method Equipped Light Gradient Boosting Machine. in *2022 6th International conference on trends in electronics and informatics (ICOEI)* 117-122 (IEEE, 2022).
44. Rosenthal, R. & Rosnow, R.L. *Essentials of behavioral research: Methods and data analysis*, (2008).
45. Sun, X., *et al.* Mapping neurophysiological subtypes of major depressive disorder using normative models of the functional connectome. *Biological Psychiatry* **94**, 936-947 (2023).
46. Zheng, K., *et al.* Redefining and subtyping of major depression based on brain functional connectivity signatures with high generalization: an ensemble hybrid framework. *bioRxiv*, 2022.2006.2007.495073 (2022).
47. Wang, Y., *et al.* Data-driven clustering differentiates subtypes of major depressive disorder with distinct brain connectivity and symptom features. *The British Journal of Psychiatry* **219**, 606-613 (2021).

48. Liang, S., *et al.* Biotypes of major depressive disorder: Neuroimaging evidence from resting-state default mode network patterns. *NeuroImage: Clinical* **28**, 102514 (2020).
49. Feder, S., *et al.* Sample heterogeneity in unipolar depression as assessed by functional connectivity analyses is dominated by general disease effects. *Journal of affective disorders* **222**, 79-87 (2017).
50. Price, R.B., *et al.* Parsing heterogeneity in the brain connectivity of depressed and healthy adults during positive mood. *Biological psychiatry* **81**, 347-357 (2017).
51. Price, R.B., Gates, K., Kraynak, T.E., Thase, M.E. & Siegle, G.J. Data-driven subgroups in depression derived from directed functional connectivity paths at rest. *Neuropsychopharmacology* **42**, 2623-2632 (2017).
52. Cheng, Y., *et al.* Delineation of early and later adult onset depression by diffusion tensor imaging. *PloS one* **9**, e112307 (2014).

Supplementary Materials

SAINT Dataset

Trial registration. The dataset was approved by the Ethics Committee of the First Affiliated Hospital, Fourth Military Medical University, and was registered with clinicaltrial.gov (identifier: NCT 04653337).

Treatment. Each treatment was administered using a Black Dolphin Navigation Robot system (SmarPhin S-50, Solide Brain Control Medical Technology Co., Ltd., Xi'an, China). In a precise experimental protocol, three consecutive sessions of intermittent theta-burst stimulation (iTBS) were delivered at 90% of the resting motor threshold (RMT) within a 9 minute and 52 second interval. This regimen was repeated for a total of ten iTBS sessions, with a 50-minute interval between each session. Patients received this treatment daily for five consecutive days, resulting in a cumulative total of 90,000 pulses throughout the entire duration of treatment. For ethical and safety considerations, venlafaxine (75mg/d) or duloxetine (30mg/d) were prescribed at the start of treatment. Additionally, dexzopiclone or zolpidem was administered to improve sleep quality for individuals with severe insomnia.

Inclusion Criteria: Participants included in the study were (i) between 18 and 60 years old; (ii) diagnosed with unipolar MDD according to the criteria of the Diagnostic and Statistical Manual of Mental Disorders, Fifth Edition (DSM-5) as assessed by the Mini-Neuropsychiatric Interview (MINI); (iii) right-handed; (iv) with a score > 17 on the HAMD-17; (v) with a score ≥ 6 on the Beck Scale for Suicidal Ideation-Chinese Version

(BSI-CV); (vi) showed normal results on physical examination and electroencephalography.

Exclusion criteria: Participants who (i) had received antidepressant treatment within two months prior to the study; (ii) had any other current or past psychiatric axis-I or axis-II disorders; (iii) had severe physical illnesses; (iv) experienced psychotic symptoms, alcohol or drug abuse; (v) had a history of neurological disorders including seizures, cerebral trauma, or MRI evidence of structural brain abnormalities; (vi) had contraindications to MRI and rTMS, such as metallic implants in the body, cardiac pacemakers, or claustrophobia; (vii) were experiencing acute suicidal or self-injury behavior requiring immediate intervention; (viii) were pregnant, lactating, or planning a pregnancy, were excluded from the study.

Image Acquisition: The parameters for resting-state fMRI with eyes closed were as follows: slices = 35, repetition time = 2000 ms, echo time = 30 ms, slice thickness = 4 mm, matrix size = 64 × 64, field of view = 224 × 224 mm², and flip angle = 90°.

Supplementary Tab. S1 A summary of MDD subtyping studies based on MRI.

Study	Sample Size (MDD)	Subtyping based on	Method	Number of subtypes	Subtype descriptions
Cynthia et al. ¹⁵ , 2024	685	T1-weighted MRI image	HYDRA	2	1. preserved gray and white matter widespread subtle reductions in gray and white matter
Sun et al. ⁴⁵ , 2023	1148	Functional connectivity	normative model	2	1. positive deviations in the default mode, limbic, and subcortical areas and negative deviations in the sensorimotor and attention areas.

					2. a moderate but converse deviation pattern.
Wen et al. 9, 2022	501	T1-weighted MRI image	HYDRA	2	1. preserved brain anatomy without WM disruptions 3. widespread brain atrophy and WM integrity disruptions
Zheng et al. ⁴⁶ , 2022	848	Functional connectivity	Graph neural networks and clustering by fast search and find of density peaks (CFDP)	3	1. subcortical-limbic, 2. DMN-FPN 4. VN-SMN
Wang et al. ⁴⁷ , 2021	115	Functional connectivity	Regularized CCA and K-means	2	1. hyperconnectivity within the ventral attention network 2. hypoconnectivity in the subcortical and dorsal attention networks
Liang et al. ⁴⁸ , 2020	690	Functional connectivity	K-means and PCA	2	1. increase DMN connectivity 2. decrease DMN connectivity
Drysdale et al. ¹¹ , 2017	220	Functional connectivity and HAMD-17	Hierarchical clustering	4	1. Reduced connectivity in frontoamygdala 2. Low anhedonia-related connectivity 3. Hyperconnectivity in thalamic and frontostriatal networks 4. Large, reduced connectivity in frontoamygdala, hyperconnectivity in thalamic and frontostriatal networks
Feder et al. ⁴⁹ , 2017	159	Functional connectivity	Hierarchical clustering and K-means	2	1. Higher connectivity 2. Lower connectivity
Price et al. ⁵⁰ , 2017	68	Functional connectivity	Group Iterative Multiple Model Estimation	2	1. Fewer specific connections. 2. greater overall connectivity
Price et al. ⁵¹ , 2017	80	Effective connectivity	Group Iterative Multiple Model Estimation	2	1. Directed pathway linking two DMN hubs 2. Lacking within-DMN connectivity
Cheng et al. ⁵² , 2014	61	Fractional anisotropy	Classification analysis	2	1: Stronger WM connectivity 3. 2. WM microstructural deficits

Supplementary Tab. S2 Architecture and parameters of CVAE on the REST-meta-

MDD II dataset.

Layer	Output Shape	# Param	Activation	Bias	Input
Shared encoder					
Input Layer	(-1,1,64,64,64)	-	-	-	-
Cov3D-1	(-1,1,31,31,31)	28	ReLU	False	Input Layer
Cov3D-2	(-1,1,15,15,15)	28	ReLU	False	Cov3D-1
Flatten	(-1,3375)	0	-	False	Cov3D-2
Linear	(-1,64)	216064	ReLU	True	Flatten
S-mu	(-1,32)	2080	-	True	Linear
S-log_var	(-1,32)	2080	-	True	Linear
Specific encoder					
Input Layer	(-1,1,64,64,64)	-	-	-	-
Cov3D-1	(-1,1,31,31,31)	28	ReLU	False	Input Layer
Cov3D-2	(-1,1,15,15,15)	28	ReLU	False	Cov3D-1
Flatten	(-1,3375)	0	-	False	Cov3D-2
Linear	(-1,64)	216064	ReLU	True	Flatten
Z-mu	(-1,32)	2080	-	True	Linear
Z-log_var	(-1,32)	2080	-	True	Linear
Decoder					
Linear-1	(-1,64)	4160	LeakyReLU	True	[S _M , Z _M] or [S _H , 0]
Linear-2	(-1,3375)	219375	LeakyReLU	True	Linear-1
Reshape	(-1,1,15,15,15)	0	-	False	Linear-2
CovTranspose 3D-1	(-1,1,31,31,31)	28	LeakyReLU	False	Reshape
CovTranspose 3D-2	(-1,1,64,64,64)	28	LeakyReLU	False	CovTranspose 3D-1
Total params: 1,108,022					

Linear: Fully connected layer; Cov3D: Three-dimensional convolution layer; CovTranspose 3D: Three-dimensional transposed convolution layer; ReLU: Rectified Linear Unit; LeakyReLU: Leaky Rectified Linear Unit; S-mu/S-log_var: Fully connected layer outputting the mean / logarithm of the variance of shared features S; Z-mu/Z-log_var: Fully connected layer outputting the mean / logarithm of the variance of MDD-specific features Z.

Supplementary Tab. S3 Hyperparameters of CVAE on the REST-meta-MDD II dataset.

Parameter	Training Value	Range	Description
Epoch	15000	5000,10000,15000	Iteration epoch
Learning_rate	0.001	[0.001:0.200]	Learning rate
Input_size	64, 64, 64	-	Dimension of input fALFF data
Batch_size	32	16,32,64	Batch size

Latent dimension	32	32,64,128,512	Dimension of Z/S
Kernel_size	3	[2:5]	Kernel size of cov3D && CovTranspose3D
negative_slope	0.4	[0.1:0.8]	Negative slope of LeakyReLU
α	20000	[15000:25000]	Weight of disentangle loss
β	2	[0.1:2.0]	Weight of KL loss

Supplementary Tab. S4 Architecture and parameters of CVAE on the REST-meta-

MDD I dataset

Layer	Output Shape	# Param	Activation	Bias	Input
Shared encoder					
Input Layer	(-1,1,64,64,64)	-	-	-	-
Cov3D-1	(-1,1,31,31,31)	28	ReLU	False	Input Layer
Cov3D-2	(-1,1,15,15,15)	28	ReLU	False	Cov3D-1
Flatten	(-1,3375)	0	-	False	Cov3D-2
Linear	(-1,64)	216064	ReLU	True	Flatten
S-mu	(-1,32)	2080	-	True	Linear
S-log_var	(-1,32)	2080	-	True	Linear
Specific encoder					
Input Layer	(-1,1,64,64,64)	-	-	-	-
Cov3D-1	(-1,1,31,31,31)	28	ReLU	False	Input Layer
Cov3D-2	(-1,1,15,15,15)	28	ReLU	False	Cov3D-1
Flatten	(-1,3375)	0	-	False	Cov3D-2
Linear	(-1,64)	216064	ReLU	True	Flatten
Z-mu	(-1,32)	2080	-	True	Linear
Z-log_var	(-1,32)	2080	-	True	Linear
Decoder					
Linear-1	(-1,64)	4160	LeakyReLU(0.45)	True	[S _M , Z _M] or [S _H , 0]
Linear-2	(-1,3375)	219375	LeakyReLU(0.45)	True	Linear-1
Reshape	(-1,1,15,15,15)	0	-	False	Linear-2
CovTranspose 3D-1	(-1,1,31,31,31)	28	LeakyReLU(0.45)	False	Reshape
CovTranspose 3D-2	(-1,1,64,64,64)	28	LeakyReLU(0.45)	False	CovTranspose 3D-1
Total params: 1,108,022					

Linear: Fully connected layer; Cov3D: Three-dimensional convolution layer; CovTranspose 3D: Three-dimensional transposed convolution layer; ReLU: Rectified Linear Unit; LeakyReLU: Leaky Rectified Linear Unit; S-mu/S-log_var: Fully connected layer outputting the mean / logarithm of the variance of shared features S; Z-mu/Z-log_var: Fully connected layer outputting the mean / logarithm of the variance of MDD-specific features Z.

Supplementary Tab. S5 Hyperparameters of CVAE on the REST-meta-MDD I dataset.

Parameter	Training Value	Range	Help
Epoch	15000	5000,10000,15000	Iteration epoch
Learning_rate	0.001	[0.001:0.200]	Learning rate
Input_size	64, 64, 64	-	Dimension of input fALFF data
Batch_size	32	16,32,64	Batch size
Latent dimension	32 && 512	32,64,128,512	Dimension of Z/S
Kernel_size	3	[2:5]	Kernel size of cov3D && CovTranspose3D
negative_slope	0.4	[0.1:0.8]	Negative slope of LeakyReLU
α	20000	[15000:25000]	Weight of disentangle loss
β	2	[0.1:2.0]	Weight of KL loss

Supplementary Tab. S6 Hyperparameters of logistic regression model for medication treatment.

Parameter	Selection	Range	Description
C	1	[0.1:10]	Inverse of regularization strength in Logistic regression
solver	'liblinear'	-	Optimization algorithm to be used in Logistic regression
N	175	[10:500]	The number of features to select in Recursive Feature Elimination (RFE)
n_splits	10	-	Number of folds in Repeated K-Fold cross validator
n_repeats	100	50,100,1000	Number of times K-Fold cross-validator needs to be repeated

Supplementary Tab. S7 Demographic and clinical variables for responders and non-responders for all medication classification.

	Responsive	Non-Responsive
Sample size	224	174

Age (yr)	35.8	31.6
Age range (yr)	12-77	11-65
Sex		
Male (number, percentage)	72 (32.1%)	73 (41.9%)
Female (number, Percentage)	152 (67.8%)	101 (58.1%)
Years of education	12.5 (13.1)	12.7 (9.6)
HAMD-17	20.3 (21.1)	17.4 (90.3)
HAMA	22.0 (84.7)	14.5 (73.6)
Illness Duration (mo)	32.3 (3041.1)	67.1 (6357.2)
First episode/recurrent MDD	198/25	85/72

Supplementary Tab. S8 Demographic and clinical variables for responders and non-responders for medication-free classification.

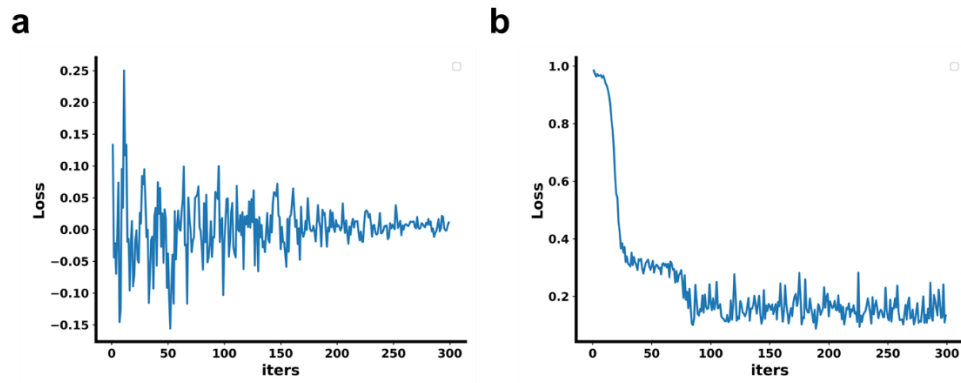
	Responsive	Non-Responsive
Sample size	155	75
Age (yr)	34.6	28.5
Age range (yr)	14-77	11-61
Sex		
Male (number, percentage)	52 (33.5%)	34 (45.3%)
Female (number, Percentage)	103 (66.5%)	41 (54.7%)
Years of education	12.7 (13.4)	12.7 (8.9)
HAMD-17	21.1 (14.1)	18.8 (115.1)
HAMA	23.4 (80.6)	13.8 (64.7)
Illness Duration (mo)	21.8 (2028.4)	44.2 (3670.1)
First episode/recurrent MDD	144/12	38/32

Supplementary Tab. S9 Demographic and clinical variables for responders and non-responders for SSRI medication classification.

	Responsive	Non-Responsive
Sample size	41	65
Age (yr)	40.2	31.5
Age range (yr)	19-77	13-65
Sex		
Male (number, percentage)	9 (21.9%)	26 (40.0%)
Female (number, Percentage)	32 (78.1%)	39 (60.0%)
Years of education	11.7 (11.7)	12.7 (10.8)
HAMD-17	18.5 (26.3)	17.4 (85.7)
HAMA	19.3 (67.8)	14.0(69.2)
Illness Duration (mo)	61.2 (5521.9)	83.9 (7819.1)
First episode/recurrent MDD	33/8	32/27

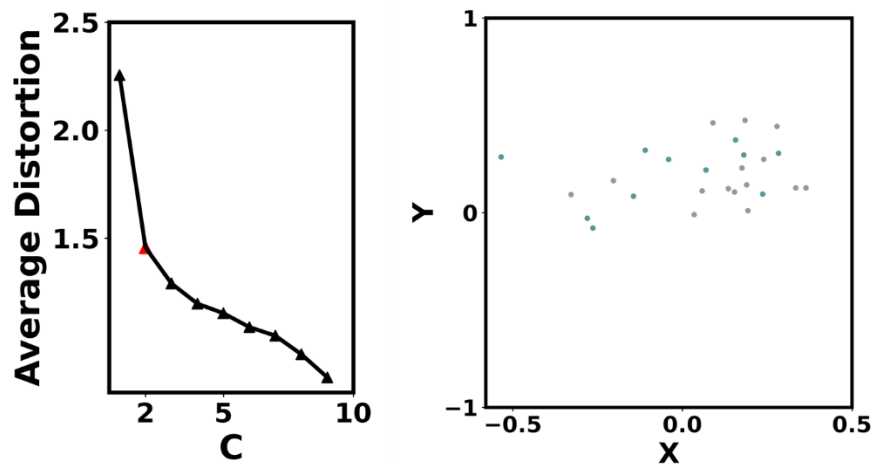
Supplementary Tab. S10 Hyperparameters of SVR model for rTMS treatment.

Parameter		Selection	Range	Description
C	HAMD-17-5	11	[0.1:15]	Inverse of regularization strength in support vector regression (SVR)
	HAMD-17-15	10		
	HAMD-17-30	10		
	MADRS-5	14		
	MADRS-15	8		
	MADRS-30	11		
	BDI-II-5	15		
	BDI-II-15	15		
	BDI-II-30	15		
Kernel		'rbf'	-	The kernel type to be used in SVR
K		175	[10:500]	Number of top K features to select in SelectKBest method
n_splits		5	-	Number of folds in Repeated K-Fold cross validator
n_repeats		100	50,100,1000	Number of times K-Fold cross-validator needs to be repeated

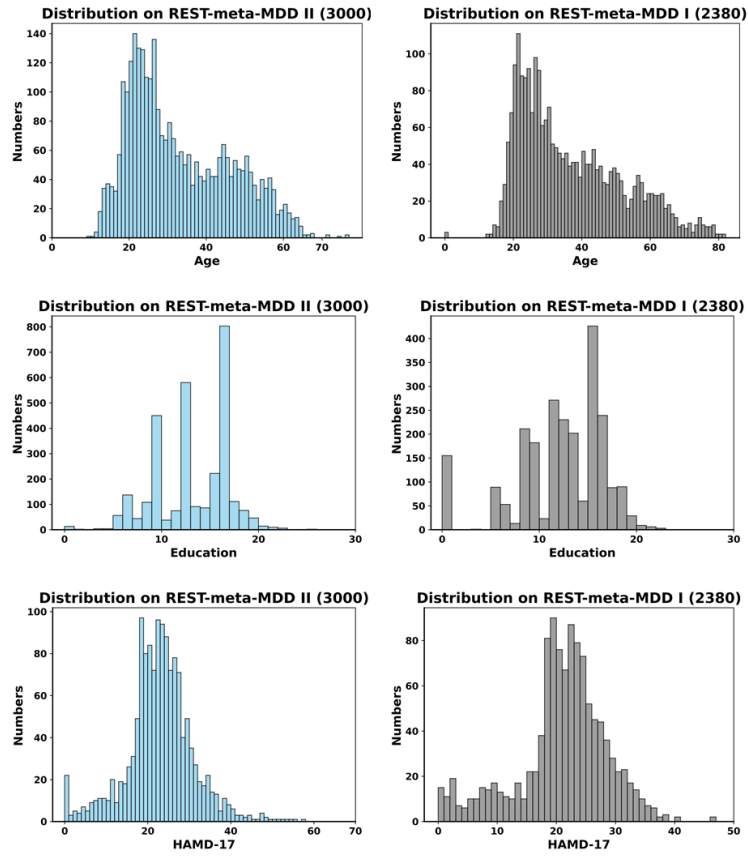


Supplementary Fig. S1 Comparison of decoupling loss using: a MINE and b HSIC.

The figure shows training decoupling loss difference for the first 300 batches.



Supplementary Fig. S2 Optimal clustering on the SAINT dataset. MDD-specific features are best captured by two clusters. Scatterplots show individual subjects' fALFF data from subtype1 (green) and subtype2 (gray) groups projected onto uniform manifold approximation and projection dimensions computed from the MDD-specific features.



Supplementary Fig. S3 Subject ages, educations and HAM-D-17 scores for REST-meta-MDD II and REST-meta-MDD I, respectively.