**Deep learning for the prediction of treatment response in depression**

Letizia Squarcina1\*, Filippo Maria Villa2\*, Maria Nobile2, Enrico Grisan3,4, Paolo Brambilla1,5

1 Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy; Department of Neurosciences and Mental Health

 2Scientific Institute, IRCCS E. Medea, Developmental Psychopathology Unit, Bosisio Parini, Lecco, Italy

3Department of Information Engineering, University of Padova, Padova, Italy

4School of Engineering, London South Bank University, London, UK.

5Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Milan, Italy.

\*The two authors contributed equally to the study

Corresponding author:

Letizia Squarcina, Department of Pathophysiology and Transplantation, University of Milan, Via Festa del Perdono, 7, 20122 Milan, Italy (email: letizia.squarcina@gmail.com, letizia.squarcina@unimi.it)

**Abstract**

**Background**

Mood disorders are characterized by heterogeneity in severity, symptoms and treatment response. The possibility of selecting the correct therapy on the basis of patient-specific biomarker may be a considerable step towards personalized psychiatry. Machine learning methods are gaining increasing popularity in the medical field. Once trained, the possibility to consider single patients in the analyses instead of whole groups makes them particularly appealing to investigate treatment response. Deep learning, a branch of machine learning, lately gained attention, due to its effectiveness in dealing with large neuroimaging data and to integrate them with clinical, molecular or -omics biomarkers.

**Methods**

In this mini-review, we summarize studies that use deep learning methods to predict response to treatment in depression. We performed a bibliographic search on PUBMED, Google Scholar and Web of Science using the terms “psychiatry”, “mood disorder”, “depression”, “treatment”, “deep learning”, “neural networks”. Only studies considering patients’ datasets are considered.

**Results**

Eight studies met the inclusion criteria. Accuracies in prediction of response to therapy were considerably high in all studies, but results may be not easy to interpret.

**Limitations**

The major limitation for the current studies is the small sample size, which constitutes an issue for machine learning methods

**Conclusions**

Deep learning shows promising results in terms of prediction of treatment response, often outperforming regression methods and reaching accuracies of around 80%. This could be of great help towards personalized medicine. However, more efforts are needed in terms of increasing datasets size and improved interpretability of results.

Mood disorders, as bipolar disorder or depression, present themselves with a great heterogeneity in symptoms severity, as well as in treatment response. In fact, patients with the same categorical diagnosis may respond differently to the same treatment. Indeed, a particular therapy is usually selected following guidelines which are defined on the basis of results obtained and validated at the population level, thus targeted to a generic, ideal patient (Perna et al., 2017). Moreover, a timely diagnosis and accurate treatment are both of paramount importance: it has been demonstrated that the quality of life of patients with depression, in terms of functional impairments, is directly related to the success of the first round of treatment (Trivedi et al., 2013). To date, the selection of a specific treatment is done using a trial-and-error procedure (Trivedi et al., 2006), which does not entirely take advantage of the great number of different and available therapies. Thus, it is clear how an objective method to identify the most suitable therapeutic intervention would be extremely beneficial. Therefore, the identification of biomarkers able to predict patients’ response to treatment with a high accuracy would be of great help in the clinical practice, paving the way towards personalized medicine, where the treatment is tailored to the specific patient (Snyderman et al., 2012). Many efforts have been devoted to link the efficacy of antidepressant to biomarkers of different types. Genetic and epigenetic characteristics have been investigated (Lin and Chen, 2008; Lisoway et al., 2018) both in genome-wide association studies (GWAS) aiming at identifying genetic loci affecting response to antidepressants, and in single nucleotide polymorphisms (SNPs) to be used as biomarkers (Kautzky et al., 2015; Maciukiewicz et al., 2017). Clinical variables (Chekroud et al., 2016; Iniesta et al., 2016; Ermers et al., 2020 ) as well as neuroimaging biomarkers (Patel et al., 2015; Fonseka et al., 2018), proved to be effective in explaining some variability in remission or resistance to treatment in bipolar patients. However, all the metanalysis conducted on the performance of different biomarkers in predicting patients’ response to treatment agree that the results are heterogeneous and at times inconsistent (Fonseka et al., 2018, Lisoway et al., 2018), possibly due to differences in methodologies, variations across biotypes and limited sample size.

Machine learning is a field of artificial intelligence which employs algorithms which automatically learn models based on examples, and is particularly suited to the idea of predicting the response to a treatment based on certain characteristics of patients, and in particular to the aim of single-patient predictions (Perna et al., 2017). Machine learning techniques have been applied to this aim in patients with depression, with good performance, obtaining accuracies around 75-80% in the prediction of response to treatment and therapeutic outcome (Pigoni et al., 2019; Lee at al., 2018).

In the latest years, deep learning, a particular branch of machine learning, is becoming increasingly popular as it can efficiently exploit neuroimaging data and integrate non-imaging biomarkers. Its main principle is the use of neural networks (NN): many hidden layers, of increasing levels of abstraction, are employed to learn a hierarchical representation of the data. The use of deep learning methods is proving to be promising in the field of psychiatry, and deep learning algorithms are considered one of the most promising machine learning techniques, even if results are not easy to interpret (Zhang et al., 2018), and they are often criticised as “black box” models. Nonetheless, visualization methods have been proposed which may help in the interpretability of results, by highlighting regions driving the deep-learning decision, using sensitivity and saliency analysis (Simonyan et al, 2013; Zeiler and Fergus, 2014; Selvaraju et al., 2017; Shrikumar et al., 2017; Fong and Vedaldi, 2017, 2019; Taha et al., 2007).

A bibliographic search on PUBMED, Web of Science and Google Scholar was performed in July 2020 using the search terms “psychiatry”, “mood disorder”, “depression”, “treatment”, “deep learning”, “neural networks”. We considered only studies taking into account patients’ datasets, which considered clinical, genetic or medical information. Eight studies, which have been summarized in Table 1, met these inclusion criteria.

Different approaches were used, all with the aim of identifying biomarkers which may allow the prediction of response to treatment. In one of the first works on deep learning, Serretti and Smeraldi (2004) used a model called multilayer perceptron (MLP), in which each unit performs a biased weighted sum of its inputs (in this case, the polymorphisms: transcriptional control region upstream of the 5HTT coding sequence (SERTPR) and in the Tryptophan Hydroxylase genotypes) layered in a feedforward topology, with one hidden layer, to obtain the response status as an output. The MLP was trained with a back-propagation algorithm obtaining a ROC area of 0.73, outperforming linear or radial basis function networks, which resulted in a ROC area of 0.69 and 0.66, respectively. Later, the group of Serretti used the MLP for the prediction of response using clinical variables (2006), considering the Hamilton-D (HAM-D) scale for depressive symptoms as output to predict. They obtained a moderate correlation of the predicted and observed response of 0.46, that although suggestive of the possibility of including clinical variables and NN in response prediction, is somehow far from allowing a personalized medicine approach. Interestingly, the network provided to the investigators also the most relevant variables used for the prediction, which were number of episodes, side effects, delusional features, baseline HAM-D, length of current episode, lithium augmentation, current medical condition and personality disorders, which then may be of particular interest for the clinical practice. Along the same line of investigation, Mehltretter and colleagues (2019) used clinical variables as predictors, merging two publicly available datasets of patients with depression, obtaining a dataset of over 3000 individuals. After a feature selection procedure, which reduced the number of features from over 200 to 17, they used a NN to predict remission rate, with an area under the curve (AUC) of almost 0.7. Interestingly, they used the same NN to predict the most suitable treatment for a test set of 200 subjects, obtaining an improvement of remission rate both for hypothetical cases, where they compared the predicted remission rate to the population real remission rate, and for real cases, where they considered the actual remission rate of patients treated with the same drug the NN predicted to have the best performance. These results highlight once more how deep learning could have a profound impact on clinical practice and on personalized medicine. However, it has to be noted that this is a retrospective analysis which, albeit conservative as stated by the authors, does not really answer the question if patient would actually benefit from the therapy predicted by the NN. A prospective study would be needed to verify the usefulness of this method. The same group later (Mehltretter et al., 2020) applied five different DNN models to various combinations of the two datasets, identifying the most important features in the prediction of response or non-response to treatment. Importantly, they used ML-identified features to do a post-hoc analyses using classical statistics, so to improve the interpretability of the deep learning model, which may help elucidate causal models and ultimately the selection of a specific treatment for individual patients.

Lin and colleagues (2018) compared three different neural networks applied to a sample of patients with depression treated with antidepressants, with a varying number of hidden layers, from 1 to 3. The input were genetic variables, in particular various SNPs identified from the literature as related to treatment response or remission. They obtained the best performance in predicting response to antidepressants using 2 and 3 hidden layers, where the AUC reached the value of 0.8 and outperformed logistic regression.

Additionally, brain functionality and morphology measures have been used as features for the prediction of treatment response. In particular, Erzugel and colleagues (2015) investigated the possibility of predicting the response to repetitive transcranial magnetic stimulation (rTMS) in patients affected by major depression. They tried different approaches and obtained cross-validated accuracies over 85% with all models, demonstrating that deep learning could be a valuable tool for selecting candidates for rTMS, on the basis of the electroencephalography signal acquired prior to treatment. Brain functionality, at rest and during a task involving reward mechanisms, has been considered by Nguyen and colleagues (2019): they evaluated 800 distinct NN models on 37 subjects, with different combinations of functional magnetic resonance imaging (fMRI) ROIs and architectures, and identified the 10 most important ROIs, which included frontal cortex, amygdala, cingulate cortex and striatum. Brain morphology was taken into account in a NN regression model (Chang et al., 2019), together with DNA methylation, genetic variants and demographic features in a multimodal approach. Deep learning considerably outperformed all other machine learning approaches the authors considered as baseline methods, showing an accuracy of 0.85 in predicting response or remission, demonstrating the potentiality of NN in the prediction of treatment response, although in a small cohort. In particular, they compared NN with support vector regression, ridge regression, gradient boosting, MLP, K-nearest neighbors regression, and random forest regressor, obtaining consistently better results with NN in predicting both remission and outcome in terms of HAM-D scores.

In conclusion, deep learning is a powerful approach for the automatic analysis of medical data, and it proved particularly suited towards the development of a true personalized medicine. The use of neural networks enables the exploitation of large datasets and heterogeneous sources: this is particularly appealing, given the steady increase of information which is acquired during clinical practice in the last years. Despite the promising results, with AUC values consistently over 0.8 and results firmly outperforming regression or other machine learning methodologies, it has to be considered that most studies suffer from a small sample size, which constitutes an issue for machine learning and in particular for deep learning methods, which benefit from large datasets Moreover, deep learning methods are characterized by a substantially higher number of parameters with respect to other machine learning techniques as, for example, support vector machines, increasing the risk of overfitting the data. This could be avoided on the one side by, e.g., using regularization and data augmentation methods (Dvornek et al, Farruque et al., 2020, Mousavian et al., 2019, Nguyen et al., 2020), or on the other side by consistently increasing the number of samples used in the analysis. In addition, recent advances in generative-adversarial networks (GAN) allow the generation of realistic simulated data that are used to enhance the generalization performance of deep learning. In any case, studies employing larger populations are needed to confirm the results and provide to the clinicians the confidence to complement their evaluation with a deep-learning evaluation. Nevertheless, the considered studies demonstrate that the field of deep learning is promising for the prediction of treatment outcome, leading ultimately to personalized medicine in psychiatry.

**Funding:**

PB was partially supported by grants from the Italian Ministry of Health (RF-2016-02364582).

**References**

Chang B, Choi Y, Jeon M, et al. ARPNet: Antidepressant Response Prediction Network for Major Depressive Disorder. Genes (Basel). 2019;10(11):907. Published 2019 Nov 7. doi:10.3390/genes10110907

Chekroud AM, Zotti RJ, Shehzad Z, Gueorguieva R, Johnson MK, Trivedi MH, et al. Cross-trial prediction of treatment outcome in depression: a machine learning approach. Lancet Psychiatry (2016) 3:243–50. doi: 10.1016/S2215-0366(15)00471-X

Dvornek NC, Ventola P, Duncan JS. COMBINING PHENOTYPIC AND RESTING-STATE FMRI DATA FOR AUTISM CLASSIFICATION WITH RECURRENT NEURAL NETWORKS. Proc IEEE Int Symp Biomed Imaging. 2018 Apr;2018:725-728. doi: 10.1109/ISBI.2018.8363676. Epub 2018 May 24. PMID: 30288208; PMCID: PMC6166875.

Erguzel TT, Ozekes S, Gultekin S, Tarhan N, Hizli Sayar G, Bayram A. Neural Network Based Response Prediction of rTMS in Major Depressive Disorder Using QEEG Cordance. Psychiatry Investig. 2015;12(1):61-65. doi:10.4306/pi.2015.12.1.61

Ermers NJ, Hagoort K, Scheepers FE. The Predictive Validity of Machine Learning Models in the Classification and Treatment of Major Depressive Disorder: State of the Art and Future Directions. Front Psychiatry. 2020;11:472. Published 2020 May 25. doi:10.3389/fpsyt.2020.00472

Farruque N., Zaiane O., Goebel R. (2020) Augmenting Semantic Representation of Depressive Language: From Forums to Microblogs. In: Brefeld U., Fromont E., Hotho A., Knobbe A., Maathuis M., Robardet C. (eds) Machine Learning and Knowledge Discovery in Databases. ECML PKDD 2019. Lecture Notes in Computer Science, vol 11908. Springer, Cham. https://doi.org/10.1007/978-3-030-46133-1\_22

Fong R, Patrick M, Vedaldi A. Understanding deep networks via extremal perturbations and smooth masks. Proceedings of the IEEE International Conference on Computer Vision. (2019): 2950–2958.

Fong RC and Vedaldi A. Interpretable explanations of black boxes by meaningful perturbation. Proceedings of the IEEE International Conference on Computer Vision (207): 3429–3437.

Fonseka, T.M.; MacQueen, G.M.; Kennedy, S.H. Neuroimaging biomarkers as predictors of treatment outcome in Major Depressive Disorder. J. Affect. Disord. 2018, 233, 21–35.

Iniesta R, Malki K, Maier W, Rietschel M, Mors O, Hauser J, et al. Combining clinical variables to optimize prediction of antidepressant treatment outcomes. J Psychiatr Res. (2016) 78:94–102. doi: 10.1016/j.jpsychires.2016.03.016

Kautzky A, Baldinger P, Souery D, et al. The combined effect of genetic polymorphisms and clinical parameters on treatment outcome in treatment-resistant depression. Eur Neuropsychopharmacol. 2015;25(4):441-453. doi:10.1016/j.euroneuro.2015.01.001

Lin E, Chen PS. Pharmacogenomics with antidepressants in the STAR\*D study. Pharmacogenomics. 2008;9(7):935-946. doi:10.2217/14622416.9.7.935

Lin E, Kuo PH, Liu YL, Yu YW, Yang AC, Tsai SJ. A Deep Learning Approach for Predicting Antidepressant Response in Major Depression Using Clinical and Genetic Biomarkers. Front Psychiatry. 2018;9:290. Published 2018 Jul 6. doi:10.3389/fpsyt.2018.00290

Lisoway, A.J.; Zai, C.C.; Tiwari, A.K.; Kennedy, J.L. DNA methylation and clinical response to antidepressant medication in major depressive disorder: A review and recommendations. Neurosci. Lett. 2018, 669, 14–23

Maciukiewicz M, Marshe VS, Hauschild AC, et al. GWAS-based machine learning approach to predict duloxetine response in major depressive disorder. J Psychiatr Res. 2018;99:62-68. doi:10.1016/j.jpsychires.2017.12.009

Mehltretter J, Rollins C, Benrimoh D, Fratila R, Perlman K, Israel S, Miresco M, Wakid M, Turecki G. Analysis of Features Selected by a Deep Learning Model for Differential Treatment Selection in Depression. F rontiers in Artificial Intelligence 2020; 2:21. doi: 10.3389/frai.2019.00031

Mehltretter, J., Fratila, R., Benrimoh, D., Kapelner, A., Perlman, K., Snook, E., et al. (2019). Differential treatment benefit prediction for treatment selection in depression: a deep learning analysis of STAR\*D and CO-MED data. BioRxiv [Preprint]. doi: 10.1101/679779

Mousavian M, Chen J and Greening S. Depression Detection Using Feature Extraction and Deep Learning from sMRI Images. 2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA), Boca Raton, FL, USA, 2019, pp. 1731-1736, doi: 10.1109/ICMLA.2019.00281.

Nguyen KP, Chin Fatt C, Treacher A, Mellema C, Trivedi MH, Montillo A. Anatomically informed data augmentation for functional MRI with applications to deep learning. Proc. SPIE 11313, Medical Imaging 2020: Image Processing, 113130T (10 March 2020);

Nguyen KP, Fatt CC, Treacher A, Mellema C, Trivedi MH, Montillo A. Predicting Response to the Antidepressant Bupropion using Pretreatment fMRI. Predict Intell Med (2019). 2019;11843:53-62. doi:10.1007/978-3-030-32281-6\_6

Patel MJ, Andreescu C, Price JC, Edelman KL, Reynolds CF III, Aizenstein HJ. Machine learning approaches for integrating clinical and imaging features in late-life depression classification and response prediction. Int J Geriatr Psychiatry (2015) 30:1056–67. doi: 10.1002/gps.4262

Perna G, Grassi M, Caldirola D, Nemeroff CB. The revolution of personalized psychiatry: will technology make it happen sooner?. Psychol Med. 2018;48(5):705-713. doi:10.1017/S0033291717002859

Pigoni A, Delvecchio G, Madonna D, Bressi C, Soares J, Brambilla P. Can Machine Learning help us in dealing with treatment resistant depression? A review. J Affect Disord. 2019 Dec 1;259:21-26. doi: 10.1016/j.jad.2019.08.009. Epub 2019 Aug 13. PMID: 31437696.

Selvaraju RR, Cogswell R, Das A, Vedantam R, Parikh D, and Batra . Grad-cam: Visual explanations from deep networks via gradient-based localization. Proceedings of the IEEE international conference on computer vision (2017):618–626.

Serretti A, Zanardi R, Mandelli L, Smeraldi E, Colombo C. A neural network model for combining clinical predictors of antidepressant response in mood disorders. J Affect Disord. 2007;98(3):239-245. doi:10.1016/j.jad.2006.08.008

Shrikumar A, Greenside P, Kundaje A. Learning important features through propagating activation differences. arXiv preprint arXiv:1704.02685, 2017.

Simonyan K, Vedaldi A, and Zisserman A. Deep inside convolutional networks: Visualising image classification models and saliency maps. arXiv preprint arXiv:1312.6034, 2013.

Snyderman R. Personalized health care: from theory to practice. Biotechnol J. (2012) 7:973–9. doi: 10.1002/biot.201100297

Taha A, Yang X, Shrivastava A, Davis L. A generic visualization approach for convolutional neural networks. rXiv preprint arXiv:2007.09748, 2020.

Trivedi MH, Fava M, Marangell LB, Osser DN, Shelton RC. Use of treatment algorithms for depression. Prim Care Companion J Clin Psychiatry. 2006;8(5):291-298. doi:10.4088/pcc.v08n0506

Trivedi MH. Evaluating and monitoring treatment response in depression using measurement-based assessment and rating scales. J Clin Psychiatry. 2013;74(7):e14. doi:10.4088/JCP.12075tx2c

Zeiler MD and Fergus R. Visualizing and understanding convolutional networks. European conference on computer vision. Springer, 2014, pp. 818–833.

Zhang, Q., Zhu, S. Visual interpretability for deep learning: a survey. Frontiers Inf Technol Electronic Eng 19, 27–39 (2018). https://doi.org/10.1631/FITEE.1700808

Table 1 Selection of studies investigating deep learning methods for the prediction of treatment response in depression.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Study** | **Sample** | **Treatment** | **Features** | **Data Preprocessing** | **Feature Selection** | **Selected features** | **Target output** | **Deep learning** | **Performance assessment** | **Results** |
| **Serretti, Smeraldi (2004)** | 217 depressed patients | Fluvoxamine | Polymorphism in the transcriptional control region upstream of the 5HTT coding sequence (SERTPR) and in the Tryptophan Hydroxylase (TPH) gene  |  |  |  | Response status | MLP (1 hidden layer, 7 nodes)Comparison with linear and RBF networks | Permutation test. | AUC 0.73Sensitivity 0.77Specificity 0.51Linear network:AUC 0.69,Specificity 0.56Sensitivity 0.67RBF network:AUC 0.66,Specificity 035Sensitivity 0.86 |
| **Serretti et al (2006)** | 145 subjects (MDD /BD: 111/34). | Fluvoxamine | 16 Clinical variables | Univariate analyses with Student t-test. Multivariate ANOVA and multiple regression. |  |  | slope of HAMD scores | MLP network (1 hidden layer, 13 nodesComparison with multiple regression analysis |  | Correlation of predicted versus observed response was 0.46Multiple regression analysis: p<0.01, explained variance 24.7% |
| **Lin et al. (2018)** | 455 MDD patients treated with SSRI | SSRI antidepressants | 20 SNPs6 clinical biomarkers | Plate-wise biases and inbreeding coefficient check.Multi-dimensional scaling analysis | Association of SNPs with treatment response investigated with logistic regression  | 10 top SNPs:ABCA13rs4917029 BNIP3rs9419139 CACNA1E rs704329EXOC4rs6978272 GRIN2B rs7954376 12 LHFPL3rs4352778NELL1rs2139423–NUAK1rs2956406 PREX1rs4810894 SLIT3rs139863958  | Response status | NN, 1-3 hidden layersComparison with logistic regression | 10-fold cross-validation | Best results with 16 biomarkers, 2 hidden layers:AUC 0.823Sensitivity 0.755;Specificity 0.692Logistic regression: AUC 0.817, Sensitivity 0.749, Specificity 0.701,  |
| **Mehltretter et al; (2019)** | CO-MED (Rush et al., 2011): 665 patients with non-psychotic depressionSTAR\*D (Trivedi et al., 2006): 2757 patients with depression | CO-MED: escitalopram and placebo, bupropion and escitalopram, or mirtazapine and venlafaxine. STAR∗D: citalopram | 213 features | Removal of features not in common | variance thersholding, recursive features elimination | 17 clinical variables | Remission | DNN (no information given on NN architecture) | 10 fold cross validations and test set of 200 subjects | AUC of 0.69Naive analysis: improvement of remission rate of over 30%.Conservative analysis: improvement of remission rate of 7.2% |
| **Mehltretter et al. (2020)** | CO-MEDSTAR\*D | As above | 213 features | As above | As above | 17 features common to STAR-D and CO-MED | Remission | DNN(no information given on NN architecture) | Test set: 750 subjects | AUC 0.69 |
| STAR∗D | As above | 213 features | As above | As above | 21 features in STAR-D | remission with citalopram | DNN(no information given on NN architecture) | Test set: 750 subjects | AUC 0.71 |
| CO-MED (citalopram only) | As above | 213 features | As above | As above | 14 features common to CO-MED and STAR-D | remission with citalopram | DNN(no information given on NN architecture) | Test set: 750 subjects | AUC 0.70 |
| CO-MED | As above | 213 features | As above | As above | 26 features in CO-MED | remission | DNN(no information given on NN architecture) | Test set: 750 subjects | AUC 0.80 |
| **Chang et al. (2019)** | 121 patients | 14 antidepressants | demographics, MRI, geneetic, epigenetic | imputation for missing variables | Selection of features identified by a literature survey followed by elastic net | 127 demographic information,20 MRI, 20 genetic variants, 20 methylation loci; | ΔHAMD | DNN, 2 hidden layersComparison with baseline models: SVR, Ridge Regressor, Gradient Boosting, MLP, K-Nearest Neighbors Regressor, and Random Forest Regressor | Cross validation | R2 0.55ARPNet improves R2 of 372.9%, 31.5%, 174.0%, 44.4%, and 34.1% over SVR, Ridge Regressor, Gradient Boosting, MLP Regressor, K-Nearest Neighbors Regressor, and Random Forest Regressor, respectively |
| Remission | DNN, 2 hidden layers | Cross validation | Accuracy 0.85Improvements of at least 40.0%, 74.4%, 86.0%, and 57.1% in Specificity, Precision, F1-score, and Accuracy, respectively, over the best baseline model. |
| **Nguyen et al. (2019)** | EMBARC trial: 37 patients | bupropion and placebo | fMRI | fMRI preprocessing pipeline, parcellation and region-based features |  |  | ΔHAMD | DNN, 2 hidden layers | Nested cross validation | R2 0.26 |
| Remission | DNN, 2 hidden layers | Nested cross validation | AUC 0.71 |

BD: bipolar disorder; MDD: major depressive disorder; MLP: Multi-layer perceptron; HAM-D: Hamilton – D; SSRI: selective serotonin reuptake inhibitors; NN: neural network; AUC: area under the curve; DNN: dense neural network; rTMS: repetitive transcranial magnetic stimulation; MRI: magnetic resonance imaging; fMRI: functional magnetic resonance imaging; SVR: support vector regression; ROI: region of interest;