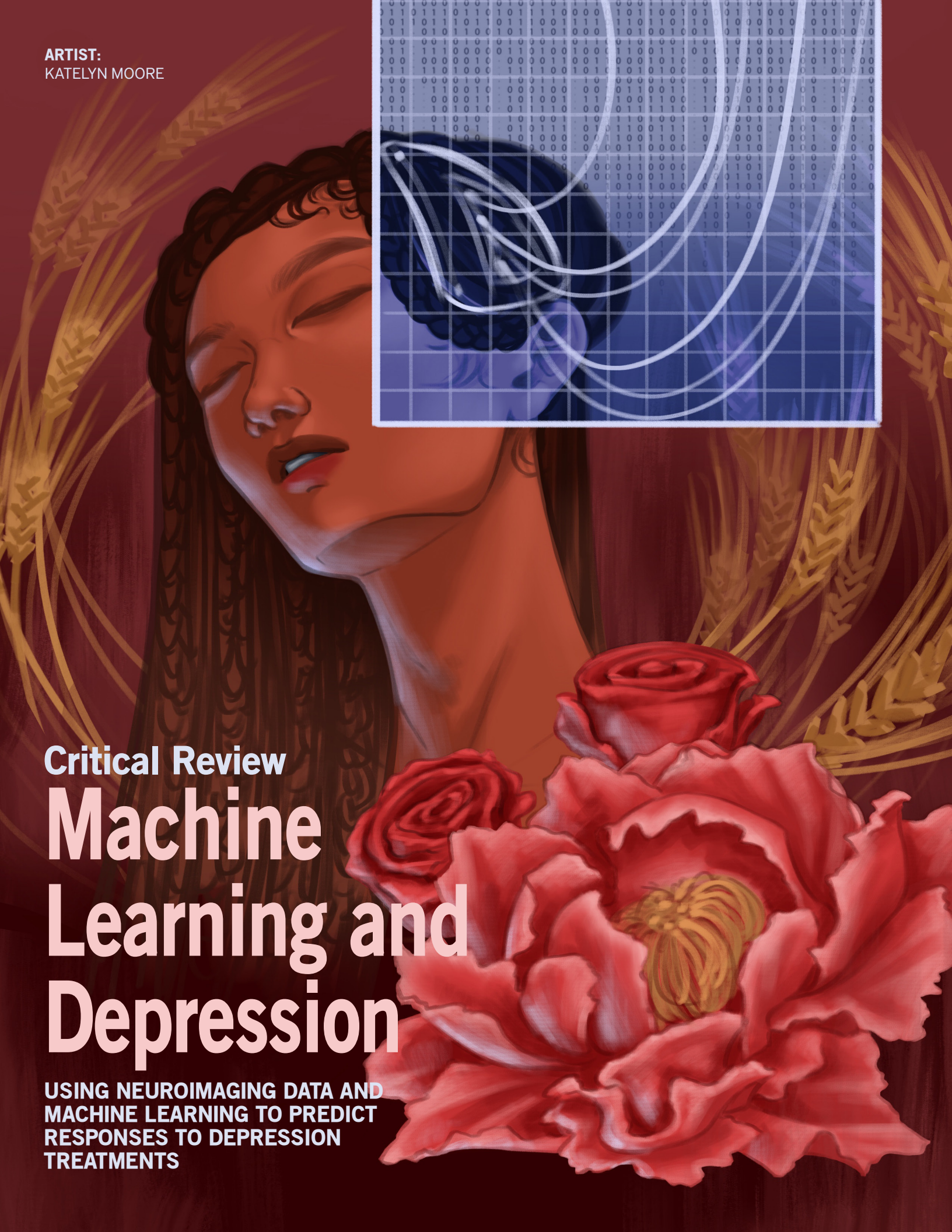
**ARTIST:**
KATELYN MOORE



Critical Review

# Machine Learning and Depression

## USING NEUROIMAGING DATA AND MACHINE LEARNING TO PREDICT RESPONSES TO DEPRESSION TREATMENTS

## ABSTRACT

Despite the high prevalence of major depressive disorder (MDD), there is a lack of tools for predicting individual patient responses to specific MDD treatments. However, a growing body of literature has been describing the use of machine learning (ML) to improve MDD treatment by using neuroimaging data to generate a model capable of predicting said treatment responses. Studies follow a general ML pipeline, though exact methodologies for sampling, treatment, and imaging vary. Overall, predictions using ML are relatively successful, with reasonable accuracy during cross-validation. However, generalizability of these algorithms has not yet been demonstrated and, at this stage, studies largely serve as "proof-of-concept", with many practical issues that still need to be addressed prior to clinical implementation. This review aims to discuss the potential benefits and limitations of ML in predicting patient responses to MDD treatment.


doi: 10.35493/medu.40.18


**YIFAN WANG**
Bachelor of Health Sciences (Honours),
Class of 2024, McMaster University
wang66@mcmaster.ca


## INTRODUCTION

Major depressive disorder (MDD) is a highly prevalent disorder characterized by depressed mood, diminished interests, impaired cognitive function, disturbed sleep, and changes in appetite causing clinically significant distress or impairment.[1] Affecting 1 in 6 adults, it is estimated by the Global Burden of Disease Consortium to be the fourth leading contributor to global disease burden in individuals aged 10 to 24, and sixth in those aged 25 to 49.[2]

Although methods such as psychotherapy, electroconvulsive therapy (ECT), and pharmacotherapy are commonly used to treat depression, these approaches are effective in only 30-50% of patients.[3] This is partially due to the broad and heterogeneous nature of depression diagnoses: the DSM-5 does not break down MDD into more narrowly defined disease entities with specific biologies.[1] This impedes the personalization of treatment on a patient-specific level.[4] Instead, effective treatment is dependent on long-term interactions, where clinicians begin with recommendations based on broader symptom classifications before personalizing the treatment over time through trial and error.[3] Although this approach may eventually prove effective for patients, it prolongs disease complications and consumes significantly more resources compared to targeted approaches.[2] This review aims to investigate the potential of the application of machine learning to neuroimaging data to predict individualized responses to MDD treatment, while addressing specific and systemic limitations of current research.

## MACHINE LEARNING

Machine learning (ML) is a branch of artificial intelligence that uses an algorithmic and data-based approach to develop machines to perform tasks without explicit programming. In ML models, machines are "trained" using a bottom-up approach, where they are given examples from which they learn, automatically improving as further experience is gained to develop a more generalizable algorithm.[5] This means that ML can develop models capable of novel and generalizable predictions; this ability has led ML to gain significant traction in recent decades, with applications ranging from self-driving cars to medical diagnoses.[5]

When it comes to MDD treatment, ML provides two key advantages. First, it allows for predictions at the level of the individual rather than solely the identification of gross differences on a group level.[6] As such, it has high translational potential to a clinical setting. Second, its multivariate nature makes it more sensitive to subtle, spatially-distributed brain alterations.[6] This enables it to detect patterns in massively multivariate data, such as magnetic resonance imaging (MRI), that are far too complicated for humans to interpret. These patterns can be used to predict whether a specific treatment will be successful in decreasing depressive symptoms for a given patient. Models commonly focus on predicting the success of pharmacotherapy and ECT, while significantly less research has applied ML predictions to cognitive behavioral therapy, despite it being a common treatment for MDD.[7]

The machine learning pipeline typically begins with data preprocessing, which prepares and refines the raw data to make it more suitable for the machine learning model. This generally entails the alignment and normalization of image data, and the filtering-out of noise. However, there is significant variation in the tools used to perform these steps; some studies have also included further preprocessing steps, such as feature selection, where only features that are expected to be meaningful for prediction are included in the model.[8] Although these differences are not largely significant, they highlight the lack of a standardized and validated approach to preprocessing imaging data that may be necessary before clinical application is possible. Following preprocessing, models are trained using training data to build an algorithm to make predictions about the success of the drug or intervention. One method of analyzing neuroimaging data is an algorithm called a support vector machine (SVM).[6] SVMs seek to define a "hyperplane" in high-dimensional space, which is a decision boundary that separates data into discrete categories, namely whether a depression treatment is successful or not (see Fig. 1). It should be noted that different studies vary in their criteria for a "successful" intervention, with some using symptom reduction measured on the Hamilton Depression Rating Scale (HRSD) and others defining success as complete remission.[8,9] While this makes it more difficult to compare studies, these different definitions of success allow for greater application over multiple contexts, depending on the history and severity of MDD a patient experiences. Once a model is trained, it can be used to make predictions on new data.

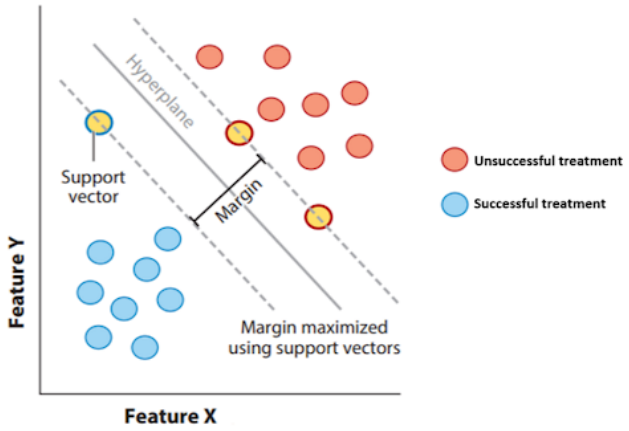



Figure 1: **A two-dimensional support vector machine.** SVM is an ML algorithm that finds a hyperplane in N-dimensional space (where each dimension corresponds to a feature of the dataset) that most successfully classifies cases, while also maximizing the margin between the data points closest to the hyperplane (support vectors).[3]

## PREDICTIVE PERFORMANCE

The performance of an ML model is commonly evaluated by first training the model using a training set, after which the generated algorithm is applied to a test set to analyze performance by comparing the prediction to the correct label. Specifically, most studies in this area use leave-one-out cross-validation, a method where a single participant from both responder and non-responder groups is excluded to use as a test set, while the model is trained on the remaining patients.[7] This process is repeated with different participants being excluded at each iteration, until every participant is excluded once. While this cross-validation approach produces high-variability and potentially biased estimations, other approaches are currently difficult to implement due to small sample sizes.[3]

Performance can be evaluated with a range of metrics, but the most commonly reported metrics are accuracy (the proportion of correctly predicted cases), sensitivity (the proportion of correctly predicted responders out of all responders), and specificity (the proportion of correctly predicted non-responders out of all non-responders). Model accuracy varies significantly between investigations, ranging from 62-89%.[10,11] This variability is partly due to objective model performance, but can also be attributed to different methodologies for sampling, treatment, imaging, and analysis. The average sensitivity and specificity was estimated to be 77% and 79% respectively in a 2021 meta-analysis conducted by Cohen et al., which demonstrates that models perform fairly successfully overall.[7] There were no significant differences resulting from imaging modality: structural and functional MRI studies yielded similar sensitivities and specificities. However, the predictive accuracy of ECT interventions has generally been found to be higher than that of pharmacotherapy.[7]

Studies have had widely varying results for brain regions with high predictive importance for MDD treatment. Costafreda et al. found that increased grey matter density in the cingulate gyrus (CgC) predicted an increased probability of clinical remission in response to fluoxetine.[8] These findings agree with an earlier functional MRI study by Marquand et al., which also identified the CgC as a biomarker for successful antidepressant response.[12] Grieve et al. conducted a subsequent study using different drugs, yet also found that the CgC predicted non-remitting patients.[13] This raises a potential issue: current algorithms could be predicting the overall success of antidepressant treatments in general, rather than being drug-specific. While such predictions may still be helpful, it detracts from their clinical utility where deciding between multiple potential drugs is often required. Many other regions of interest have been identified, such as the amygdala and the hippocampus, with some studies identifying as many as 25 regions of interest.[7,14] However, this varies significantly between publications, as regions of high predictive importance identified in some studies have been completely excluded from others.[11,15] Overall, there is no clear agreement in the literature that suggests a single region of interest as a potential biomarker.

## LIMITATIONS

While ML seems like a promising technology to assist with the treatment of MDD, it has many limitations that should be addressed. Firstly, training and testing data are often unrepresentative of the actual populations to which ML will be applied. Most current studies seek to obtain a pure estimate of the population mean without influence from other factors like comorbidities or medication effects, which often serve as exclusion criteria. In the general population, however, MDD cases are often comorbid with other psychiatric disorders.[16] As such, a model that displays high accuracy within a single study may not necessarily produce successful results in larger, more heterogeneous populations. In fact, the accuracy of these models trend downwards with increasing sample size, despite the fact that ML models generally improve with more data.[17] This inflation of accuracy may be due to overfitting, where small sample sizes can cause the model to capture dynamics that are specific to training data, but consequently do not generalize well to new data. As such, before clinical application is possible, the generalizability of ML in broader samples must be proven, especially in MDD-afflicted individuals with comorbid conditions that may complicate prediction. Dwyer and colleagues have proposed a validation hierarchy to achieve greater generalizability, where models can be gradually applied to more diverse selections of individuals, such as leave-site-out cross validation (where models are trained in one site and tested in another).[3]

Another significant limitation with current research is that investigators often use a classification approach. This approach considers the success of a treatment as a discrete variable, separating patients into responder and non-responder groups based on arbitrarily defined boundaries, such as a 50% reduction in HRSD score.[9] However, efficacy is a continuous variable, with varying degrees of symptom reduction: most non-ML studies of treatment efficacy do indeed report results continuously.[18] As such, a machine learning algorithm using a classification approach will be fundamentally constrained by being an inaccurate representation of the true nature of treatment. Therefore, it could be worth investigating a regression approach, where predictions are made on a continuous scale that estimates the degree of success of a treatment.

Finally, there are several challenges that must be addressed

before ML can be used in clinical settings. One issue is that models created by ML are often difficult to interpret, meaning that it is hard to understand how variables are combined to make predictions on both a computational and biological level. In the context of patient care, this lack of transparency could potentially undermine trust in this technology by both clinicians and patients, preventing adoption. However, improvement in this area reduces accuracy, given that the best performing models tend to be the least explainable, and vice versa.[19] ML will also likely create ethical issues of accountability: if an ML model makes an incorrect prediction, it is difficult to determine where culpability lies, creating further complications for existing legal and regulatory systems that address medical malpractice. Finally, there are a range of practical issues that also limit the usefulness of ML. These include a lack of clinician experience and training in using ML models, restricted access to neuroimaging, and the high time and computational demand required to implement this technology on a larger scale.

## CONCLUSION

The potential of ML to predict treatment response on an individual level could help rectify the current lack of targeted treatment methodologies and significantly improve patient care. However, many limitations still prevent clinical implementation. Future research should focus on improving generalizability, as successful validation across multiple sites or across different investigations would greatly improve the value of ML in real-world applications. Alongside this, researchers should also consider issues of how to make this technology readily available for clinician use, while enhancing transparency to encourage patient adoption.




1. Otte C, Gold SM, Penninx BW, Pariante CM, Etkin A, Fava M, et al. Major depressive disorder. *Nat Rev Dis Primers*. 2016;2:16065. Available from: doi:10.1038/nrdp.2016.65.
2. Vos T, Lim SS, Abbafati C, Abbas KM, Abbasi M, Abbasifard M, 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. *Lancet*. 2020;396(10258):1204-22. Available from: doi:10.1016/S0140-6736(20)30925-9.
3. Dwyer DB, Falkai P, Koutsouleris N. Machine learning approaches for clinical psychology and psychiatry. *Annu Rev Clin Psychol*. 2018;14:91-118. Available from: doi:10.1146/annurev-clinpsy-032816-045037.
4. Khan A, Faucett J, Lichtenberg P, Kirsch I, Brown WA. A systematic review of comparative efficacy of treatments and controls for depression. *PLoS One*. 2012;7(7):e41778. Available from: doi:10.1371/journal.pone.0041778.
5. Jordan MI, Mitchell TM. Machine learning: Trends, perspectives, and prospects. *Science*. 2015;349(6245):255-60. Available from: doi:10.1126/science.aaa8415.
6. Orrù G, Pettersson-Yeo W, Marquand AF, Sartori G, Mechelli A. Using Support Vector Machine to identify imaging biomarkers of neurological and psychiatric disease: A critical review. *Neurosci Biobehav Rev*. 2012;36(4):1140-52. Available from: doi:10.1016/j.neubiorev.2012.01.004.
7. Cohen SE, Zantvoord JB, Wezenberg BN, Bockting CLH, van Wingen GA. Magnetic resonance imaging for individual prediction of treatment response in major depressive disorder: A systematic review and meta-analysis. *Transl Psychiatry*. 2021;11(1):168. Available from: doi:10.1038/s41398-021-01286-x.
8. Costafreda SG, Chu C, Ashburner J, Fu CHY. Prognostic and diagnostic potential of the structural neuroanatomy of depression. *PLoS One*. 2009;4(7):e6353. Available from: doi:10.1371/journal.pone.0006353.
9. Gong Q, Wu Q, Scarpazza C, Lui S, Jia Z, Marquand A, et al. Prognostic prediction of therapeutic response in depression using high-field MR imaging. *Neuroimage*. 2011;55(4):1497-503. Available from: doi:10.1016/j.neuroimage.2010.11.079.
10. Korgaonkar MS, Williams LM, Song YJ, Usherwood T, Grieve SM. Diffusion tensor imaging predictors of treatment outcomes in major depressive disorder. *Br J Psychiatry*. 2014;205(4):321-8. Available from: doi:10.1192/bjp.bp.113.140376.
11. Jiang R, Abbott CC, Jiang T, Du Y, Espinoza R, Narr KL, et al. SMRI biomarkers predict electroconvulsive treatment outcomes: Accuracy with independent data sets. *Neuropsychcopharmacology*. 2018;43(5):1078-87. Available from: doi:10.1038/npp.2017.165.
12. Marquand AF, Mourão-Miranda J, Brammer MJ, Cleare AJ, Fu CHY. Neuroanatomy of verbal working memory as a diagnostic biomarker for depression. *Neuroreport*. 2008;19(15):1507-11. Available from: doi:10.1097/WNR.0b013e328310425e.
13. Grieve SM, Korgaonkar MS, Gordon E, Williams LM, Rush AJ. Prediction of nonremission to antidepressant therapy using diffusion tensor imaging. *J Clin Psychiatry*. 2016;77(4):e436-43. Available from: doi:10.4088/JCP.14m09577.
14. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat Med*. 2017;23(1):28-38. Available from: doi:10.1038/nm.4246.
15. Redlich R, Nils O, Dominik G, Katharina D, Dario Z, Christian B, et al. Prediction of individual response to electroconvulsive therapy via machine learning on structural magnetic resonance imaging data. *JAMA Psychiatry*. 2016;73(6):557-64. Available from: doi:10.1001/jamapsychiatry.2016.0316.
16. Hasin DS, Sarvet AL, Meyers JL, Saha TD, Ruan WJ, Stohl M, et al. Epidemiology of adult DSM-5 major depressive disorder and its specifiers in the United States. *JAMA Psychiatry*. 2018;75(4):336-46. Available from: doi:10.1001/jamapsychiatry.2017.4602
17. Flint C, Cearns M, Opel N, Redlich R, Mehler DMA, Emden D, et al. Systematic misestimation of machine learning performance in neuroimaging studies of depression. *Neuropsychopharmacology*. 2021;46(8):1510-7. Available from: doi:10.1038/s41386-021-01020-7.
18. McGirr A, Berlim MT, Bond DJ, Fleck MP, Yatham LN, Lam RW. A systematic review and meta-analysis of randomized, double-blind, placebo-controlled trials of ketamine in the rapid treatment of major depressive episodes. *Psychol Med*. 2015;45(4):693-704. Available from: doi:10.1017/S0033291714001603.
19. Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med*. 2019;17(1):195. Available from: doi:10.1186/s12916-019-1426-2.



**REVIEWED BY: PEDRO L. BALLESTER**

Pedro L. Ballester has a bachelor's and master's degree in computer science. He is currently a neuroscience PhD candidate at McMaster University. His research is focused on artificial intelligence and mental health, particularly accelerated brain aging in mood and psychotic disorders. He is a trainee at the Canadian Biomarker Integration Network in Depression (CAN-BIND) searching for neuroimaging biomarkers of treatment response in depression.

**EDITED BY: SHANZEY ALI**