



Commentary

Deep Learning: A New Horizon for Personalized Treatment of Depression?

David Benrimoh, MDCM¹, Sonia Israel², Robert Fratila³, Kelly Perlman²
MJM 2017 16(3)

Abstract

Globally, depression affects over 300 million people at any given time and is the leading cause of disability. While different patients may benefit more from different therapies, there is no principled way for clinicians to predict individual patient responses or side effect profiles. Deep learning is a form of machine learning based on artificial neural networks and might be useful for generating a predictive model that could aid in clinical decision making. Such a model's primary outcomes would be to help clinicians select the most effective treatment plans and mitigate adverse side effects, which would allow doctors to provide greater personalized care to a larger number of patients. In this commentary, we discuss the need for the personalization of depression treatment and how a deep learning model could be used to construct a clinical decision aid.

◆ Deep learning, mental health, psychiatry, depression, personalized medicine, precision medicine

¹Faculty of Medicine, McGill University, Montréal, Canada.

²Faculty of Science, McGill University, Montréal, Canada.

³Department of Computer Science and Biology, McGill University, Montréal, Canada.

Corresponding Author: David Benrimoh, email david.benrimoh@mail.mcgill.ca.

Introduction

Major Depression is a debilitating health condition that affects 11.1% of people over the course of their lives and is responsible for the majority of Disability Adjusted Life Years (DALYs) lost globally [1,2]. At any one time, over 322 million people around the world

struggle with depression [3]. Depression causes significant suffering and entails high treatment and social costs [4,5]. While seeking professional help for depression is indeed a step towards recovery, mental health professionals and the patients they work with

face another challenge: selecting the best treatment. Though a range of effective medications, therapies, and other treatments do exist, these are not equivalently effective for all patients. Some patients can spend years finding the right choice or combination from the dozens of available medications, multiple psychotherapies, and several neurostimulation techniques (e.g. repetitive Transcranial Magnetic Stimulation (rTMS), transcranial direct current stimulation (tDCS), and deep brain stimulation (DBS)). Currently, most patients and their physicians have little option but to undergo an educated “guess and check” approach to finding the right treatment. In the large Sequenced Treatment Alternatives to Relieve Depression (STAR*D) study, only about one third of patients improved after the first step of treatment in the trial, with decreasing response rates after further steps [6]. The decision about which treatment to try is one with significant consequences. In addition, different patients develop varying side effects to the same medication in an unpredictable manner, which further complicates treatment choice [7].

As such, a research objective in this domain should be to develop an evidence-based approach to rapidly select the most effective treatment for any given patient, as early on in their clinical course as possible, whilst minimizing side effects that lead to reduced quality of life or poor treatment adherence. Existing guidelines do categorize the large array of treatment options into first-, second-, and third-line treatments [7]. However, doctors lack a systematic, evidence-based tool for mental health conditions that helps clinicians choose treatments in a way that is personalized to each given patient [7,8,9,10]. Indeed, the most important study of comparative treatment efficacy to date, a meta-analysis of twenty-one antidepressants by Cipriani et al. [11], was unable to make any clear recommendations regarding the personalization of treatment. In this commentary article, the possibility of using deep learning as a solution to this problem will be discussed.

Deep Learning: A Possible Solution?

A clinical decision tool is a potential solution to the aforementioned problem. The tool would synthesize both existing and newly-recorded data with the aim of producing treatment recommendations to optimize symptom remission while mitigating adverse side effects. This can be framed as a classification problem that could be solved through a machine learning approach. Machine learning is the use of algorithms to learn patterns in often large datasets, sometimes in

order to make predictions about new pieces of data (i.e. learning about previous weather patterns to make predictions about future weather). These approaches can be as simple as classical linear or logistic regression, or as complex as multi-layered artificial neural networks, known as deep learning (DL) [12]. Artificial neural networks (ANN) have existed for decades in the form of the archetypal multilayer perceptron [13]. Due to limited computing capabilities and already superior performance by standard statistical methods, research into artificial networks slowed for decades [14]. However, recent advances [14] have led to an explosion of applications using DL [15]. DL works by passing information through several layers of weighted artificial neurons [12], producing increasingly abstract representations of relationships between variables in the original dataset [16]. The problem that comes with deeper networks is that while they can capture more complex relationships between features, the gradient-based training procedure’s error signal can either attenuate or magnify in a manner that can impede training. Superficial layers either learn very slowly, if at all (i.e., vanishing gradients), or much too quickly when compared to ‘deeper’ layers (i.e. exploding gradients). Several approaches, beyond the scope of this paper, have been developed to solve this problem.

Why use DL? This method has the tendency to overfit, or to produce solutions that fit the training data but fail to generalize to other data [17]. It also suffers from the black box problem: results provided by DL networks are difficult for humans to interpret, which is not as much of a problem with other approaches like decision trees. We will discuss possible solutions to this problem below. However, besides having surpassed other methods in a variety of tasks [12], DL has two distinct advantages. Firstly, it remains robust in the face of noisy or incomplete datasets [18,19], which are common in psychiatry. Secondly, the failure of simple models to explain or predict psychiatric phenomena speaks accurately to a complexity [20] of which DL may be able to capture through its increasingly abstract data representations [5].

Generally, to train deep networks large datasets are required. For prediction of response to depression treatment, these datasets could include information such as socio-demographic factors, symptom profiles, and previous response to treatment, as well as genetic, metabolic, endocrine, immunological, and neuroimaging data. Moreover, the training datasets must include valid outcome measures. Therefore, clinical trials and treatment research studies are of

significant interest. Another potential source of data are clinics that employ measurement-based patient care or structure clinical data collection in a manner which facilitates machine learning for outcome prediction. Despite its established efficacy, measurement-based care is not routinely implemented in psychiatry, thereby limiting the amount of available data [21]. Other clinical data (i.e., extant medical records) may not be optimal data for training because they are often incomplete and lack the same rigorous outcome measures as controlled trial data or data collected for machine learning applications.

Collecting sufficient data for DL is challenging. Producing data de novo that is conducive to machine learning requires extensive clinical partnerships and years of collection, a route being explored by the Alphabet-owned company Verily [22]. However, large, and well-designed clinical trials can be prohibitively expensive. For example, the STAR*D trial cost 35 million dollars (USD) and enrolled 4041 patients – a sizeable number, but below the tens of thousands of patients required for a machine learning application to make reliable predictions for a patient’s response to all available treatments [23]. For datasets large enough for machine learning, this leaves data collected by industry and researchers, which is often difficult to access, though recent open data initiatives such as the National Institutes of Mental Health Data Archive have facilitated this. The fact that little standardization exists between datasets means that data pooling for the purposes of machine learning remains a challenge [24].

treatment response when fed into a DL system as input features [25].

Implementation of Artificial Intelligence (AI) Technology

In order to capture clinical response patterns in a population of individuals with complex medical and personal histories and biological profiles, a predictive model must be trained on complex and heterogeneous data. Having a dataset of sufficient complexity is necessary to help the model generalize to real life clinical populations-- using unrepresentative data would be a source of bias [26]. One DL approach to such a dataset would be to use a feed-forward stacked denoising autoencoder (SDA) [5], which consists of a series of context-learning layers. The general architecture of the autoencoder is composed of the encoder (e.g., $E(x)$) and decoder (e.g. $D(x)$) and is defined as follows: the dataset, x , gets mapped to a learned latent space $E(x)$. Using the decoder, $D(E(x))$ is then mapped back to the original x . The features learned in $E(x)$ are a subset of the original, meaning that the ANN constructs representations of the interactions between input features. Pairing an SDA with a discriminative model (such as another ANN) allows for greater depth (i.e., numbers of layers) over standard feedforward neural networks, resulting in higher-order representations of data and an appreciation of more complex relationships between variables. It should be noted that recent developments in other approaches, such as generative adversarial networks (GANs), are not necessarily suited to our purpose. This is because they are designed to create artificial data samples that remain representative of the training data. Training on artificially generated patient data might skew further analyses.

The training technique for an SDA involves two steps: unsupervised learning followed by supervised learning. Unsupervised learning provides the network with contextual information about data by allowing it to autonomously explore latent groupings and dependencies - this trains the auto-encoder layers mentioned above. Supervised learning asks the network to take inputs, run them through the previously trained autoencoder layers, and maximize a specific output target - a process optimized by the network’s underlying knowledge of the dataset acquired during unsupervised learning [5]. Unsupervised learning may be effectively applied to unstructured data such as electronic medical records. Doing so might then improve supervised learning on more structured data with rigorous outcome measurement, like clinical trial data. In this way, both structured clinical data and



First there is a clinical interaction between the patient and the physician, and a diagnosis of depression is made by the physician. Then the physician would input the data they have collected in to the model (1). The model would use AI trained on large sets of patient data (2) and output a list of treatment suggestions ranked by likely efficacy (3). The physician could choose to use one of these suggestions (4) and prescribe it for the patient (5). The patient would then return for follow up, leading to another clinical interaction and the clinician could use the model again in cases of non-response or non-remission until response or remission is achieved. All of these interactions would provide new data and new questions for research. (Credit for figure design: Nykan Mirchi)

While there is ongoing effort to identify single biomarkers that could reliably and accurately predict response to treatment, none have emerged [25]. A more fruitful strategy may be to define a panel of biomarkers and clinical questions, which, together, might provide more reliable and accurate predictions of

unstructured electronic medical records may be processed and used to improve outcome prediction.

Patient outcomes would be assessed in terms of validated depression rating scales. Standard validation assessments could be used to critically examine all aspects of the classifier, including its vulnerabilities. These methods include computation of the sensitivity, specificity, positive predictive value, negative predictive value, accuracy, and the receiver operating curve (ROC) [27] along with the area under the ROC (AUC). To ensure model validity, k-fold cross validation could also be implemented [28]. As well, dropout should be implemented during training to prevent overfitting and to improve generalizability. Dropout refers to the practice of occasionally excluding artificial neurons during training to prevent over-reliance on certain nodes [29]. Clinicians require interpretability in their decision-making tools. Interpretation of DL models has been traditionally difficult due to their increasingly abstract internal representations of information. Failing to provide easily interpretable results of the model's decisions may compromise clinician trust and technology adoption. Nevertheless, certain tools for interpretability do exist and continue to improve as DL is applied towards the clinical domain [30]. For example, the authors of [31] used a DL model to "teach" a gradient boosted tree model to predict mortality in an ICU, which produced a model that benefits from the power of a DL model and the interpretability of a gradient boosted tree model.

Several other interpretability methods exist, including validating the final model input features by relating them to existing literature and using receptive field analysis on all layers in the network to gain a sense of low-level feature groupings [32,33]. Moreover, saliency map generation may elucidate features from the input data sample that most significantly contribute to network prediction [34]. Such clusters may correspond to similarities between certain patient types extracted via t-Distributed Stochastic Neighbor Embedding (t-SNE) [35]. Some of the feature clusters identified may even spawn novel research avenues. Interpretability tools such as these can produce reports explaining the most salient features in the making of a given decision. This provides a level of detail familiar to mental health clinicians, similar to analyses looking at risk factors for different conditions which cannot always definitively explain causal links.

Targeting response to treatment in depression gives rise to several challenges. For example, there is significant debate about the nosology of depression.

Anxiety is often comorbid with depression but is not part of the diagnostic criteria for the disorder [36]. To avoid the pitfalls associated with diagnostic validity when selecting the features to be used in a machine learning model, it is possible to adopt a "dimensional approach" [37] - that is, by focusing on symptoms and other patient features independently of the actual diagnosis (assuming, of course, that non-psychiatric causes of symptoms such as fatigue (e.g. hypothyroidism or anemia) have been ruled out or are not suspected). Another consideration is the temporally sensitive nature of response to treatment - that is, it may not be possible to predict treatment response solely from patient features at baseline, but also from observations shortly following treatment initiation. This is especially relevant to psychotherapy, where the strength of the relationship with the therapist - and not the individual patient or treatment features [38] - is most predictive of outcome. In this case, two machine learning models could be used. The first could predict whether the patient is likely to respond to psychotherapy based on data from patients who had a good relationship with their therapist (i.e. the 'best case' scenario). A few weeks into treatment, a second model would evaluate the actual patient-therapist relationship amongst other factors that predict response.

Potential Impact

Should a predictive model withstand clinical validation, it would be among the first personalized medicine tools in mental health specifically designed for use by clinicians. This solution has the potential to reduce the disappointingly high rate of failure to reach remission, as seen in the STAR*D trial [6]. It is difficult to ascertain the potential reduction of treatment failure rate prior to clinical testing. However, a recent machine learning approach using random forests [39] has attained some success in identifying patients who have attempted suicide (AUC = 0.84, precision = 0.79, recall = 0.95) using electronic medical records. The authors of [39] use a technique called bootstrap optimism to reduce overfitting and find that their predictions became more accurate closer in time to the actual attempt. Although this model was only tested on previously collected data, the finding that these retroactive predictions became more accurate closer to the attempt suggests that the model was able to capture the patients' temporal evolution. Given the similar nature of datasets used to predict suicide or response to depression treatment, we might reasonably expect a 60-80% accuracy in predicting the most effective treatment using DL (see [40] for a

previous approach to this using another machine learning technique). This accuracy would represent a significant improvement over the one-third success rate reported in Step 1 of STAR*D [6]. For more complex or chronic cases, we might expect a lower success rate while still aiming to surpass STAR*D Step 2 and 3 remission rates (25% and 12-20%, respectively) [6]. These are evidently rough predictions and will need to be continually revised.

Feature Reduction

Given existing clinical psychiatry data, there are likely to be less training samples than typical machine learning datasets [6,10]. Model training might prove problematic if there are too many predictors. Manually inputting many predictors would be time-consuming and cumbersome for clinicians. However, many of the features in a patient's file or a clinical trial are not actually expected to influence the response to antidepressants, making it possible to narrow down the number of features fed into the model. This feature reduction could be approached via several routes. One method would be to ask a panel of experts to review the features in the dataset and to select only the most important predictors. This list of predictors can be compared with the results of an exhaustive literature review. One can further decrease predictors by running many iterations of a model with certain input features omitted, to see which combination of predictors provides the most accurate results while being efficient enough with respect to time that clinicians will be able to incorporate it into their busy practices.

Clinical Validation

Clinical validation is key to ensure the safety and efficacy of a predictive model in the clinical setting. A predictive model should be subject to rigorous testing, including: an open-label clinical trial to establish safety and effectiveness and a randomized control trial, to evaluate efficacy by comparing the predictive model to a usual practice control group and to a group using a model loaded with static, non-personalized suggestions derived from current clinical guidelines; for an example of these guidelines, see [7]. It is important to compare both the static model and the 'practice as usual' groups to the AI-powered model group in order to assess the effect of clinicians using guideline-centered, measurement-based care, which, on its own may improve quality of care and patient outcomes.

Physicians are not familiar with using AI technologies in their practice. Building physician and patient trust will be critical to the success of any clinical decision tool. For this reason, physicians, as well as

patient representatives, should be involved in the design of such a product. Considering that a clinical decision tool must be incorporated into the medical workflow, the ultimate utility of these applications will depend in large part on them being user-friendly. Use of companion patient self-report applications might also help to reduce the time spent by the clinician inputting patient information into the predictive model while providing rich data. Most importantly, however, a clinical decision tool should be conceptualized as a tool used to compliment or augment physician capabilities, and not as a means to supplant or replace clinical judgment. As such, it is critical that the clinician actively engage with the application. Active clinician and patient engagement, coupled with a seamless integration into clinician workflow, will bring clinical medicine into the age of big data and AI-augmented care.

Funding

The authors received no funding for this research.

Disclosure

All authors are co-founders and shareholders of Aifred Health, the company engaging in the research described here.

References

1. World Health Organization. Depression and Other Common Mental Disorders: Global Health Estimates. Geneva: World Health Organization; 2017
2. Bromet, E., Andrade, L. H., Hwang, I., Sampson, N. A., Alonso, J., de Girolamo, G., ... Kessler, R. C. (2011). Cross-national epidemiology of DSM-IV major depressive episode. *BMC Medicine*, 9, 90. <https://doi.org/10.1186/1741-7015-9-90>
3. Depression. (n.d.). Retrieved August 16, 2017, from <http://www.who.int/mediacentre/factsheets/fs369/en/>
4. Greenberg PE, Fournier AA, Sisitsky T, Pike CT, Kessler RC. The economic burden of adults with major depressive disorder in the United States (2005 and 2010). *The Journal of clinical psychiatry*. 2015;76(2):155-62.
5. Vincent P, Larochelle H, Lajoie I, Bengio Y, Manzagol PA. Stacked Denoising Autoencoders: Learning Useful Representations in a Deep Network with a Local Denoising Criterion. *Journal of Machine Learning Research* 11 (2010) 3371-3408
6. Rush, A. J., Fava, M., Wisniewski, S. R., Lavori, P. W., Trivedi, M. H., Sackeim, H. A., . . . for the, S. D. I. G. (2004). Sequenced treatment alternatives to relieve depression (STAR*D): rationale and design. *Controlled Clinical Trials*, 25(1), 119-142. doi:10.1016/S0197-2456(03)00112-0
7. Kennedy, S. H., Lam, R. W., Parikh, S. V., MacQueen, G. M., Milev, R. V., Ravindran, A. V., & the, C. D. W. G. (2016). Canadian Network for Mood and Anxiety Treatments (CANMAT) 2016 Clinical Guidelines for the Management of Adults with Major Depressive Disorder: Introduction and Methods. *Canadian Journal of*

- Psychiatry. *Revue Canadienne de Psychiatrie*, 61(9), 506-509. doi:10.1177/0706743716659061
8. Leuchter, A. F., Cook, I. A., Hunter, A. M., & Korb, A. S. (2009). A new paradigm for the prediction of antidepressant treatment response. *Dialogues Clin Neurosci*, 11(4), 435-446.
9. Young, J. J., Silber, T., Bruno, D., Galatzer-Levy, I. R., Pomara, N., & Marmar, C. R. (2016). Is there Progress? An Overview of Selecting Biomarker Candidates for Major Depressive Disorder. *Frontiers in Psychiatry*, 7, 72. <http://doi.org/10.3389/fpsy.2016.00072>
10. Labermaier, C., Masana, M., & Müller, M. B. (2013). Biomarkers Predicting Antidepressant Treatment Response: How Can We Advance the Field? *Disease markers*, 35(1), 23-31. doi:10.1155/2013/984845
11. Cipriani, A., Furukawa, T. A., Salanti, G., Chaimani, A., Atkinson, L. Z., Ogawa, Y., ... Geddes, J. R. (2018). 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*, 0(0). [https://doi.org/10.1016/S0140-6736\(17\)32802-7](https://doi.org/10.1016/S0140-6736(17)32802-7)
12. LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444. <https://doi.org/10.1038/nature14539>
13. Bengio, Y., & LeCun, Y. (2007). Scaling learning algorithms towards AI. *Large-scale kernel machines*, 34(5).
14. Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7), 1527-1554.
15. SABCS 2016: IBM Watson for Oncology Platform Shows High Degree of Concordance With Physician Recommendations - The ASCO Post. (n.d.). Retrieved November 16, 2017, from <http://www.ascopost.com/News/44214>
16. Bengio, Y., Courville, A., & Vincent, P. (2012). Representation Learning: A Review and New Perspectives. arXiv:1206.5538 [Cs]. Retrieved from <http://arxiv.org/abs/1206.5538>
17. Deep Learning. Retrieved March 18, 2018, from <http://www.deeplearningbook.org/>
18. Rolnick, D., Veit, A., Belongie, S., & Shavit, N. (2017). Deep Learning is Robust to Massive Label Noise. arXiv:1705.10694 [Cs]. Retrieved from <http://arxiv.org/abs/1705.10694>
19. Lasko, T. A., Denny, J. C., & Levy, M. A. (2013). Computational Phenotype Discovery Using Unsupervised Feature Learning over Noisy, Sparse, and Irregular Clinical Data. *PLOS ONE*, 8(6), e66341. <https://doi.org/10.1371/journal.pone.0066341>
20. Huys, Q. J. M., Maia, T. V., & Frank, M. J. (2016). Computational psychiatry as a bridge from neuroscience to clinical applications. *Nature Neuroscience*, 19(3), 404-413. <https://doi.org/10.1038/nn.4238>
21. Waldrop J, McGuinness TM. Measurement-Based Care in Psychiatry. *J Psychosoc Nurs Ment Health Serv*. 2017 Nov 1;55(11):30-5.
22. Verily Life Sciences [Internet]. [cited 2018 Mar 18]. Available from: <https://verily.com/>
23. Pigott, H. E. (2015). The STAR*D Trial: It Is Time to Reexamine the Clinical Beliefs That Guide the Treatment of Major Depression. *Canadian Journal of Psychiatry. Revue Canadienne de Psychiatrie*, 60(1), 9-13.
24. Wilkinson, M. D., Dumontier, M., Aalbersberg, Ij. J., Appleton, G., Axton, M., Baak, A., ... Mons, B. (2016, March 15). The FAIR Guiding Principles for scientific data management and stewardship [Comments and Opinion]. <https://doi.org/10.1038/sdata.2016.18>
25. Brand, S. J., Mller, M., & Harvey, B. H. (2015). A Review of Biomarkers in Mood and Psychotic Disorders: A Dissection of Clinical vs. Preclinical Correlates. *Current Neuropharmacology*, 13(3), 324-368. <http://doi.org/10.2174/1570159X13666150307004545>
26. Knight, W. (2017) Google's AI chief says forget Elon Musk's killer robots, and worry about bias in AI systems instead. Retrieved March 18, 2018, from <https://www.technologyreview.com/s/608986/forget-killer-robotsbias-is-the-real-ai-danger/>
27. Hajian-Tilaki K. Receiver Operating Characteristic (ROC) Curve Analysis for Medical Diagnostic Test Evaluation. *Caspian Journal of Internal Medicine*. 2013;4(2):627-35.
28. Fushiki T. Estimation of prediction error by using K-fold cross-validation. *Statistics and Computing*. 2011;21(2):137-46.
29. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of machine learning*. 15(Jun):1929-1958
30. Sha, Y., & Wang, M. D. (2017). Interpretable Predictions of Clinical Outcomes with An Attention-based Recurrent Neural Network. In *Proceedings of the 8th ACM International Conference on Bioinformatics, Computational Biology, and Health Informatics* (pp. 233-240). New York, NY, USA: ACM. <https://doi.org/10.1145/3107411.3107445>
31. Che, Z., Purushotham, S., Khemani, R., & Liu, Y. (2017). Interpretable Deep Models for ICU Outcome Prediction. *AMIA Annual Symposium Proceedings, 2016*, 371-380.
32. Günter Klambauer TU, Andreas Mayr, Sepp Hochreiter. Self-Normalizing Neural Networks. arXiv. 2017.
33. Haffner YLLBYBP. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 1998;86(11):2278-324.
34. Karen Simonyan AV, Andrew Zisserman. Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. arXiv. 2013.
35. Maaten, L., Hinton, G. (2008). Visualizing Data using t-SNE. *JMLR* 9(Nov):2579--2605.
36. American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders (5th ed.)*. Arlington, VA: American Psychiatric Publishing.
37. Heinz, A., Schlagenhaut, F., Beck, A., & Wackerhagen, C. (2016). Dimensional psychiatry: mental disorders as dysfunctions of basic learning mechanisms. *Journal of Neural Transmission (Vienna, Austria: 1996)*, 123(8), 809-821. <https://doi.org/10.1007/s00702-016-1561-2>
38. Ardito, R. B., & Rabellino, D. (2011). Therapeutic Alliance and Outcome of Psychotherapy: Historical Excursus, Measurements, and Prospects for Research. *Frontiers in Psychology*, 2. <https://doi.org/10.3389/fpsyg.2011.00270>
39. Walsh, C. G., Ribeiro, J. D., & Franklin, J. C. (2017). Predicting Risk of Suicide Attempts Over Time Through Machine Learning. *Clinical Psychological Science*, 5(3), 457-469. <https://doi.org/10.1177/2167702617691560>
40. Chekroud, A. M., Zotti, R. J., Shehzad, Z., Gueorguieva, R., Johnson, M. K., Trivedi, M. H., ... Corlett, P. R. (2016). Cross-trial prediction of treatment outcome in depression: a machine learning approach. *The Lancet. Psychiatry*, 3(3), 243-250. [https://doi.org/10.1016/S2215-0366\(15\)00471-X](https://doi.org/10.1016/S2215-0366(15)00471-X)