Interpretable Machine Learning Models for Precision Psychiatry

Scott Fleming[1*], Adina Fischer[2*], Bailey Holt-Gosselin[2], Leanne Williams[2]
[1]Biomedical Informatics Training Program, Stanford University
[2]Department of Psychiatry and Behavioral Sciences, Stanford University

PanLab
Neuroscience is Personal

Introduction / Related Work

- Major Depressive Disorder (MDD) is the leading cause of disability across the globe.[1] In 2017, the World Health Organization (WHO) estimated that 322 million people (5% of world population) suffer from MDD.[2]
- To-date response to antidepressant medication is unpredictable, with a 30% remission rate after 12 weeks of treatment and 30–40% failure to have an adequate response even after several trials of medication (or psychotherapy) after one year of treatment.[3]
- Unsuccessful treatment leads to chronic and recurrent depression, resulting in functional impairment and increased risk of suicide.[4, 5, 6]
- Machine learning methods may help identify unique patient signatures that optimize treatment selection for MDD. Previous work has demonstrated an accuracy of 59.6% in predicting antidepressant response.[7]
- We developed a machine learning model to:
  1. Accurately predict whether a patient will respond to first-line antidepressants;
  2. Provide patient-level model interpretability.

Methods

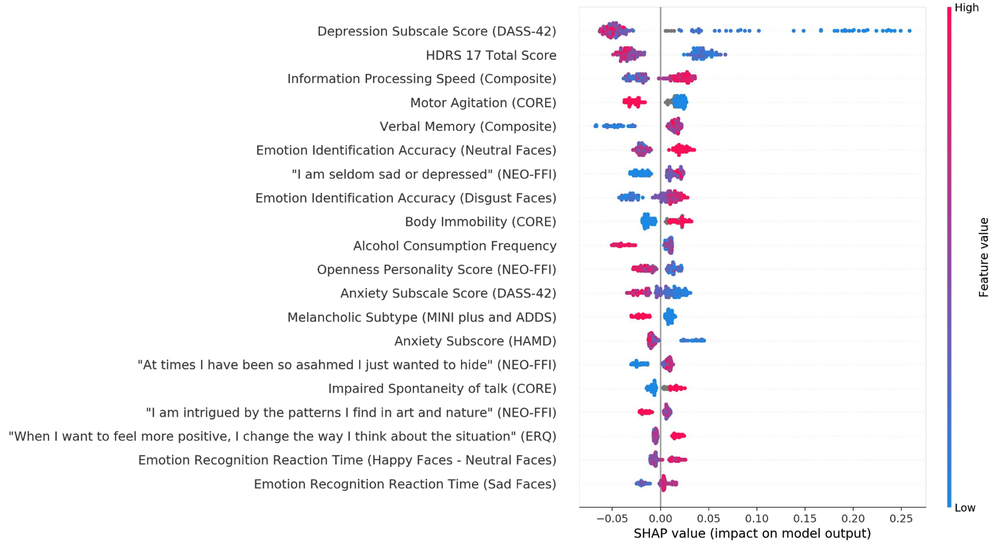
- Data was obtained from the International Study to Predict Optimized Treatment in Depression (iSPOT-D), a longitudinal controlled trial of 1,008 patients with MDD who were randomly assigned to one of three first-line antidepressants (Escitalopram, Venlafaxine, or Sertraline).[8]
- Clinical, neuropsychological, and behavioral information was collected at baseline and after 8 weeks of treatment, including:
  - Demographic information
  - Clinical information [Hamilton Depression Rating Scale (HDRS-17); Core Rating Scale (CORE); Depression Anxiety and Stress Scale (DASS); MINI Neuropsychiatric Interview (MIN); etc.]
  - Measures of social and occupational functioning (SOFAS), quality of life (WHO-QOL), resilience (BRISC), personality (NEO-FFI Personality Inventory), and stress (Early Life Stress Questionnaire; ELS)
  - Neuropsychological measures of cognition (e.g. working memory, information processing speed, verbal memory, attention) and emotion (Emotion Regulation Questionairre; ERQ and Emotion Recognition Task)
- Antidepressant response was defined as ≥ 50% symptom reduction and remission was defined as score of ≤ 7 on HDRS-17.[9]
- 712 subjects who completed the study were included in the analysis. These subjects were split into 80/20 train (n_{train} = 557) and test (n_{test} = 155) sets.
- State-of-the art gradient boosting algorithm, XGBoost[10], using 5-fold cross-validation and Tree of Parzen-based Bayesian Optimization[11] were applied to the training set.
- Shapley values were used for model interpretation, both for the model as a whole and for individual participants.[12]

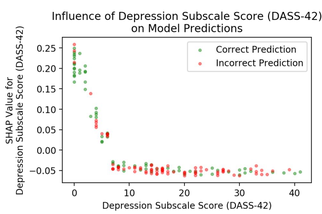
Results

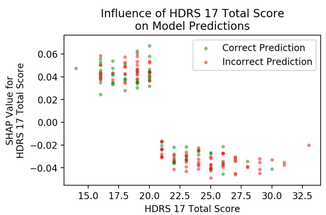
- Our model achieved an accuracy of 68.4%, substantially better than the majority vote baseline classifier which achieved an accuracy of 54%.

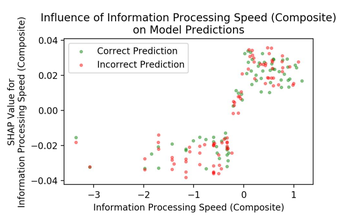
Actual (Left) vs. Predicted (Top)		
	Nonresponder	Responder
Nonresponder	63	22
Responder	27	43

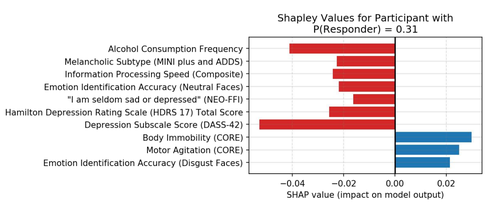
Label	Precision	Recall	F1	n
Nonresponder	0.70	0.74	0.72	85
Responder	0.66	0.61	0.64	70
Average	0.68	0.68	0.68	-

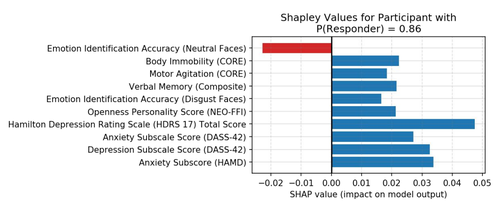




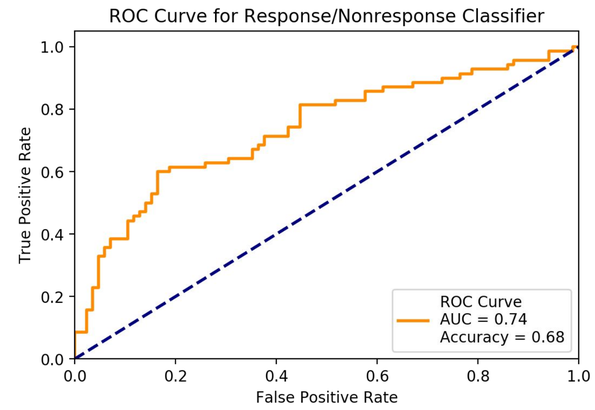








Results



Conclusion

- Our model performed substantially better than the naive baseline of assuming either everyone will respond (Acc. = 45.2%) or not respond (Acc. = 54.8%) and exceeds clinician-level performance.
- Our model (Acc. = 68.4%) surpassed the performance of current leading models in the field (Acc. = 59.6%).[7]
- Using Shapley values from coalitional game theory, we were able to provide patient-level model interpretability without sacrificing predictive accuracy.
- Interpretation suggests that patients with mild depression, better cognition (i.e. information processing and verbal memory) and emotion identification are most likely to respond to antidepressant treatment.

References

[1] World Health Organization et al. "Depression and other common mental disorders: global health estimates". In: (2017).
[2] MJ Friedrich. "Depression is the leading cause of disability around the world". In: Jama 317.15 (2017), pp. 1517–1517.
[3] Rush, A. John, et al. "Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR* D report." American Journal of Psychiatry 163.11 (2006): 1905-1917.
[4] Van Londen, L., et al. "Three-to 5-year prospective follow-up of outcome in major depression." Psychological Medicine 28.3 (1998): 731-735.
[5] Judd, Lewis L., et al. "Socioeconomic burden of subsyndromal depressive symptoms and major depression in a sample of the general population." (1996): 1411-1417.
[6] Dean Fergusson et al. "Association between suicide attempts and selective serotonin re- uptake inhibitors: systematic review of randomised controlled trials". In: Bmj 330.7488 (2005), p. 396.
[7] Adam Mourad Chekroud et al. "Cross-trial prediction of treatment outcome in depression: a machine learning approach". In: The Lancet Psychiatry 3.3 (2016), pp. 243–250.
[8] Leanne M Williams et al. "International Study to Predict Optimized Treatment for De- pression (iSPOT-D), a randomized clinical trial: rationale and protocol". In: Trials 12.1 (2011), p. 4.
[9] Leucht, Stefan, et al. "What does the HAMD mean?." Journal of affective disorders 148.2-3 (2013): 243-248.
[10] Chen, Tianqi, and Carlos Guestrin. "Xgboost: A scalable tree boosting system." Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining. ACM, 2016.
[11] Bergstra, J., Yamins, D., Cox, D. D. (2013) Making a Science of Model Search: Hyperparameter Optimization in Hundreds of Dimensions for Vision Architectures. To appear in Proc. of the 30th International Conference on Machine Learning (ICML 2013).
[12] Lundberg, Scott M., and Su-In Lee. "A unified approach to interpreting model predictions." Advances in Neural Information Processing Systems. 2017.

Acknowledgements

I would like to thank the NIH for its support through a BD2K Fellowship (LM012409); Leanne Williams and the Brain Resource for access to the data; and Leonardo Tozzi, Tali Ball, and Andrea Goldstein-Piekarski for their guidance in understanding the associated metadata.