Predicting patients at high risk for non-response may improve clinical outcomes, for when antidepressants are used to treat depression, only a little over half respond. Nie et al.'s 2018 paper, a recent paper, and the basis for this replication

Around 7 prior studies have made incremental progress in these predictions. To truly gauge clinical applicability, machine learning models must be externally validated to investigate the how different predictors affect prediction performance.

Methods: We replicated and adapted a prior study’s methodology to predict initial antidepressant response and to achieve remission, and externally validated these models with the clinical features found in both datasets. We then evaluated additional models to investigate the how different parameters affect prediction performance.

Objectives: Antidepressant monotherapy remains the first-line treatment for major depressive disorder. However, 40-60% of patients will not initially respond. Predicting a patient’s treatment outcome based on clinical symptoms and episode features represents an exciting application of modern machine learning. We sought to independently replicate recent work predicting antidepressant outcome using the STAR*D dataset, and then externally validate these models using new data from CAN-BIND.

Background

Data and Data Processing:

- Data came from CAN-BIND (n=4046) and CAN-BIND-1 (n=324).
- The STAR*D data was different than that used in Nie et al., predicting both QIDS-SR and QIDS-P responses.
- Data was processed using a pipeline modeled after feature selection, to improve model performance.
- Predicted the prior feature selection methods, including an elastic-net to find top 30 features with largest weights, and their other method using k-means clustering and chi-squared scores.

Results:

- We successfully replicated a prior study using machine learning to predict antidepressant outcomes, supporting the validity of their methods.
- We then externally validated the models on a new dataset, finding similar performance when predicting if a patient will achieve remission, but reduced performance if predicting antidepressant response. These results motivate future work to investigate the generalizability of this finding, as well as other efforts to improve prediction performance.

Conclusion: Prior work has suggested that replication and external validation may have an important role in driving acceptance and applicability of machine learning methods in psychiatry. We successfully replicate prior work predicting antidepressant treatment outcomes using clinical data. We then externally validate these models on a new dataset, finding similar performance when predicting if a patient will achieve remission, but reduced performance if predicting antidepressant response. These results motivate future work to investigate the generalizability of this finding, as well as other efforts to improve prediction performance.

References


4. Elton et al. "Predicting treatment outcomes in depression.


