Interpretable Machine Learning Models for Precision Psychiatry

Scott Fleming1, Adina Fischer2, Bailey Holt-Gosselin2, Leanne Williams2
1Biomedical Informatics Training Program, Stanford University
2Department of Psychiatry and Behavioral Sciences, Stanford University

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.2
• 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]; MNI Neuropsychiatric Interview [MNI], etc.)
  • Measures of social and occupational functioning (SOFS), quality of life (WHO-QOL), resilience (BSRI), personality (NEO-FFI Personality Inventory), and stress (Early-Life Stress Questionnaire: ELS)
  • Neuropsychological measures of cognition (i.e. working memory, information processing speed, verbal memory, attention) and emotion (Emotion Regulation Questionnaire; ERQ and Emotion Recognition Task)
• Antidepressant response was defined as ≥ 50% symptom reduction and remission was defined as score ≤ 7 on HDRS-17.9
• 712 subjects who completed the study were included in the analysis. These subjects were split into 80/20 train (ntrain = 557) and test (n = 155) sets.
• State-of-the-art gradient boosting algorithm, XGBoost10, using 5-fold cross-validation and Tree of Partition-based Bayesian Optimization11 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%.

<table>
<thead>
<tr>
<th>Actual (Left) vs. Predicted (Top)</th>
<th>n</th>
<th>Label</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonresponder</td>
<td>Responders</td>
<td>Nonresponder</td>
<td>63</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>Responder</td>
<td>27</td>
<td>43</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average</td>
<td>0.68</td>
<td>0.68</td>
<td>0.68</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Depression Subscale Score (DASS-42) HDRS-17 Total Score
Information Processing Speed (Composite) Motor Agitation (CORE)
Verbal Memory (Composite)
Emotion Identification Accuracy (Neutral Faces)
"I am sad, sad or depressed" (MDD-PR)
Emotion Identification Accuracy (Happy Faces) Body Immobility (CORE)
Alcohol Consumption Frequency
Openness Personality Score (NEO-FFI-PR)
Anxiety Subscale Score (DASS-42)
Mental Health Subscale (MHS)
"I am interested in the patterns I find in art and nature" (MDD-PR)
"If I was tired then I would feel more positive, I change the way I think about the situation" (ERQ)
Emotion Recognition Reaction Time (Happy Faces - Neutral Faces)
Emotion Recognition Reaction Time (Sad Faces)

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

Conclusion

• Our model performed substantially better than the naive baseline of assuming 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.

Acknowledgements

I would like to thank the NIH for its support through a BD2K Fellowship (LM012409); Marvin Billings for his guidance in understanding the associated metadata; I would like to thank the NIH for its support through a BD2K Fellowship (LM012409); Marvin Billings for his guidance in understanding the associated metadata; and Leonardo Tozzi, Tali Ball, and Andrea Goldstein-Piekarski for their guidance in understanding the associated metadata.

References

[2] In 2017, the World Health Organization (WHO) estimated that 322 million people (5% of the world population) suffer from MDD.
[3] 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.
[4] Unsuccessful treatment leads to chronic and recurrent depression, resulting in functional impairment and increased risk of suicide.
[5] 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.
[6] 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.
[7] Our model achieved an accuracy of 68.4%, substantially better than the majority vote baseline classifier which achieved an accuracy of 54%.
[8] 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).
[9] 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]; MNI Neuropsychiatric Interview [MNI], etc.)
  • Measures of social and occupational functioning (SOFS), quality of life (WHO-QOL), resilience (BSRI), personality (NEO-FFI Personality Inventory), and stress (Early-Life Stress Questionnaire: ELS)
  • Neuropsychological measures of cognition (i.e. working memory, information processing speed, verbal memory, attention) and emotion (Emotion Regulation Questionnaire; ERQ and Emotion Recognition Task)
[10] Antidepressant response was defined as ≥ 50% symptom reduction and remission was defined as score ≤ 7 on HDRS-17.
[11] 712 subjects who completed the study were included in the analysis. These subjects were split into 80/20 train (ntrain = 557) and test (n = 155) sets.
[12] State-of-the-art gradient boosting algorithm, XGBoost, using 5-fold cross-validation and Tree of Partition-based Bayesian Optimization were applied to the training set.
[13] Shapley values were used for model interpretation, both for the model as a whole and for individual participants.