

# Identifying Predictors of Therapy Responsiveness from Meta-Emotion Beliefs and Cognitive Flexibility Using ML

Musli. Salleh Sahimi<sup>1</sup>, Noorhayati. Saharuddin<sup>2</sup>, Ahmad. Ibrahim Ahmad<sup>2\*</sup>, Nora. Syed Alwi<sup>3</sup>


<sup>1</sup> Department of Psychiatry, Faculty of Medicine, National University of Malaysia, Kuala Lumpur, Malaysia

<sup>2</sup> Department of Educational Psychology and Counseling, Faculty of Education, Universiti Malaya, 50603 Kuala Lumpur, Malaysia


<sup>3</sup> Department of Psychology, Faculty of Education, UCSI University, Kuala Lumpur, Malaysia


\* Corresponding author email address: ah.ibrahimahmad@um.edu.my

## Editor

Asoke Kumar Saha  
Professor Department of  
Psychology, Jagannath University,  
Dhaka, Bangladesh  
drasoke@psychology.jnu.ac.bd

## Reviewers

**Reviewer 1:** Saeid Motevalli  
Associate Professor, Department of Psychology, Faculty of Education, UCSI  
University, Kuala Lumpur, Malaysia. Email: Saeid@ucsiuniversity.edu.my

**Reviewer 2:** Yoko Wong  
Faculty of Humanities, Department of Psychology, Nanyang Technological  
University, Nanyang, Singapore. Email: yokowong@ntu.edu.sg

## 1. Round 1

### 1.1. Reviewer 1

Reviewer:

The conceptualization of meta-emotion as secondary emotional and cognitive responses to one's own primary emotions is compelling and well-framed, with useful clinical examples (e.g., viewing sadness as unacceptable, anger as moral failure). Nonetheless, the manuscript should clarify how these beliefs are operationalized in the study: which specific instruments are used, which subscales enter the models, and whether meta-emotion is treated as a unidimensional or multidimensional construct. This clarification is crucial both for interpretability of the machine learning results and for allowing replication or extension of the work by other researchers.

The manuscript would be strengthened by a clearer description of the sample characteristics and clinical context in which the study was conducted, including diagnostic composition, treatment modality (e.g., CBT, integrative, pharmacotherapy adjunct), number of sessions, setting (e.g., outpatient clinic, university center), and cultural background of participants. These details are essential for evaluating the generalizability of the findings and for understanding whether the predictive model is tied to a specific therapeutic framework or might extend to other modalities and populations.

From an interpretive standpoint, the authors should consider integrating model interpretability tools (e.g., feature importance estimates, SHAP values, partial dependence plots) and explicitly report them, so that the field can learn not only that meta-emotion beliefs and cognitive flexibility predict therapy responsiveness, but how specific facets of these constructs contribute to prediction. This would bridge the gap between black-box prediction and clinically meaningful insight, and it would make the study far more useful for therapists seeking to translate the results into assessment or case formulation practices.

The discussion section would benefit from a more nuanced consideration of ethical and practical implications of deploying ML-based prediction of therapy responsiveness in clinical settings. Issues such as algorithmic bias, potential stigmatization of individuals labeled as “low responders,” the risk of prematurely excluding patients from certain treatments, and the need for human clinical judgment to remain central should be explicitly addressed. Articulating safeguards and recommendations for responsible use would significantly enhance the rigor and maturity of the manuscript’s contribution.

Authors revised and uploaded the document.

## 1.2. Reviewer 2

Reviewer:

The decision to examine cognitive flexibility as a potential moderator or co-determinant of therapy responsiveness is theoretically sound, particularly in the context of psychological interventions that require patients to adopt new perspectives and revise core beliefs. At the same time, the manuscript should describe more explicitly how cognitive flexibility is defined and measured (e.g., self-report vs. performance-based tasks, specific scales, psychometric properties in the current sample). It would also be valuable to explain whether the authors conceptualize flexibility as a stable trait, a dynamic capacity, or both, and how this conceptualization aligns with the timing of assessment relative to therapy.

A major strength of the paper is its argument that traditional linear statistical methods are limited in capturing complex, potentially nonlinear interactions among transdiagnostic processes, and that machine learning provides a more appropriate modeling framework. However, for a methodological contribution, the manuscript needs to provide considerably more detail about the analytical pipeline: which ML algorithms were used, how hyperparameters were tuned, how training/validation/test splits or cross-validation were handled, what metrics were prioritized (e.g., AUC, balanced accuracy, sensitivity/specificity), and how class imbalance (if any) was addressed. Without this information, readers cannot adequately evaluate the robustness or clinical utility of the proposed predictive models.

The concept of “therapy responsiveness” is central to the manuscript, yet in its current form the operational definition remains unclear. The authors should specify precisely how response was defined (e.g., percentage reduction on a symptom scale, achievement of remission, reliable and clinically significant change, therapist rating, or some composite index), at what time point(s) it was assessed, and whether responsiveness was dichotomized or treated as a continuous outcome in the ML models. Additionally, it would strengthen the paper to justify this definition with reference to prior outcome literature and to discuss how different operationalizations might affect the results.

While the introduction convincingly situates meta-emotion beliefs and cognitive flexibility within a broader transdiagnostic framework, it would benefit from a more explicit discussion of alternative or competing predictors that have been examined in the psychotherapy outcome literature (e.g., therapeutic alliance, baseline symptom severity, comorbidity, interpersonal functioning) and how the current predictors are expected to add incremental value beyond these established factors. Including such a comparison, at least conceptually, would enhance the paper’s contribution and help position the proposed model within the broader landscape of precision mental health approaches.

Authors revised and uploaded the document.

## 2. Revised

Editor's decision after revisions: Accepted.

Editor in Chief's decision: Accepted.