



## Original Investigation | Psychiatry

# Developing an Electroencephalography-Based Model for Predicting Response to Antidepressant Medication

Benjamin Schwartzmann, MSc; Prabhjot Dhami, PhD; Rudolf Uher, MD, PhD; Raymond W. Lam, MD; Benicio N. Frey, MD, PhD; Roumen Milev, MD, PhD; Daniel J. Müller, MD, PhD; Pierre Blier, MD, PhD; Claudio N. Soares, MD, PhD; Sagar V. Parikh, MD; Gustavo Turecki, MD, PhD; Jane A. Foster, PhD; Susan Rotzinger, PhD; Sidney H. Kennedy, MD; Faranak Farzan, PhD

## Abstract

**IMPORTANCE** Untreated depression is a growing public health concern, with patients often facing a prolonged trial-and-error process in search of effective treatment. Developing a predictive model for treatment response in clinical practice remains challenging.

**OBJECTIVE** To establish a model based on electroencephalography (EEG) to predict response to 2 distinct selective serotonin reuptake inhibitor (SSRI) medications.

**DESIGN, SETTING, AND PARTICIPANTS** This prognostic study developed a predictive model using EEG data collected between 2011 and 2017 from 2 independent cohorts of participants with depression: 1 from the first Canadian Biomarker Integration Network in Depression (CAN-BIND) group and the other from the Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care (EMBARC) consortium. Eligible participants included those aged 18 to 65 years who had a diagnosis of major depressive disorder. Data were analyzed from January to December 2022.

**EXPOSURES** In an open-label trial, CAN-BIND participants received an 8-week treatment regimen of escitalopram treatment (10-20 mg), and EMBARC participants were randomized in a double-blind trial to receive an 8-week sertraline (50-200 mg) treatment or placebo treatment.

**MAIN OUTCOMES AND MEASURES** The model's performance was estimated using balanced accuracy, specificity, and sensitivity metrics. The model used data from the CAN-BIND cohort for internal validation, and data from the treatment group of the EMBARC cohort for external validation. At week 8, response to treatment was defined as a 50% or greater reduction in the primary, clinician-rated scale of depression severity.

**RESULTS** The CAN-BIND cohort included 125 participants (mean [SD] age, 36.4 [13.0] years; 78 [62.4%] women), and the EMBARC sertraline treatment group included 105 participants (mean [SD] age, 38.4 [13.8] years; 72 [68.6%] women). The model achieved a balanced accuracy of 64.2% (95% CI, 55.8%-72.6%), sensitivity of 66.1% (95% CI, 53.7%-78.5%), and specificity of 62.3% (95% CI, 50.1%-73.8%) during internal validation with CAN-BIND. During external validation with EMBARC, the model achieved a balanced accuracy of 63.7% (95% CI, 54.5%-72.8%), sensitivity of 58.8% (95% CI, 45.3%-72.3%), and specificity of 68.5% (95% CI, 56.1%-80.9%). Additionally, the balanced accuracy for the EMBARC placebo group (118 participants) was 48.7% (95% CI, 39.3%-58.0%), the sensitivity was 50.0% (95% CI, 35.2%-64.8%), and the specificity was 47.3% (95% CI, 35.9%-58.7%), suggesting the model's specificity in predicting SSRIs treatment response.

(continued)

## Key Points

**Question** Can a robust and generalizable electroencephalography-based model be developed to predict whether a patient with depression will benefit from specific medication treatments?

**Findings** In this prognostic study, a model based on electroencephalography predicted response to escitalopram for a cohort of 125 patients with an accuracy of 64.2%. It was validated in an independent sample of 105 patients, in which the model demonstrated an accuracy of 63.7% in predicting response to sertraline.

**Meaning** These findings suggest that the use of electroencephalography can provide a reliable method for predicting response to specific antidepressant medications and may help match patients with depression to optimized treatment.

## + Supplemental content

Author affiliations and article information are listed at the end of this article.

**Open Access.** This is an open access article distributed under the terms of the CC-BY License.

Abstract (continued)

**CONCLUSIONS AND RELEVANCE** In this prognostic study, an EEG-based model was developed and validated in 2 independent cohorts. The model showed promising accuracy in predicting treatment response to 2 distinct SSRIs, suggesting potential applications for personalized depression treatment.

JAMA Network Open. 2023;6(9):e2336094. doi:10.1001/jamanetworkopen.2023.36094

## Introduction

Major depressive disorder (MDD) is a leading cause of disability worldwide, affecting more than 300 million people.<sup>1</sup> Antidepressant medications are prescribed as a first-line treatment, but remission rates after the first and second trials range from 35% to 50%<sup>2,3</sup> and decline further with subsequent trials.<sup>2</sup> Given that each trial may take up to 12 weeks, predicting medication response could greatly reduce illness duration and associated negative outcomes.

In the past 2 decades, many studies have assessed the utility of electroencephalography (EEG) for identifying biological predictors of medication treatment outcomes.<sup>4-10</sup> However, EEG predictors identified in these studies are not robust enough to be used in clinical practice. For instance, 1 study<sup>4</sup> found lower theta activity in responders to several selective serotonin reuptake inhibitor (SSRI) medications, whereas another study<sup>5</sup> showed the opposite in responders to the SSRI fluoxetine. Responders to different SSRIs were also found to exhibit distinct patterns in alpha power.<sup>6,7</sup> Failure to replicate across studies can be attributed to small sample size and the complexity and heterogeneity of depression. Testing individual EEG features in small, underpowered studies has hindered the development of EEG-based predictive models for clinical use.<sup>11,12</sup> An alternative approach involves applying machine learning techniques on larger samples to potentially capture a more reliable neurobiological signature of response to antidepressants.<sup>13-15</sup> However, despite some promising results, several challenges remain in implementing this approach.

The main challenge is the lack of generalizability of predictive models. Inadequate validation methods can lead to overfitting, especially with a small sample size.<sup>16</sup> To avoid overfitting, proper internal and/or external validation is needed. Although internal validation is common due to scarce EEG data in depression studies,<sup>13,15</sup> external validation is crucial from a clinical standpoint to establish model generalizability.<sup>17</sup> However, to our knowledge, no study has rigorously used external validation with EEG-based models for predicting antidepressant response.<sup>18</sup> Although some recent attempts have been made to validate such models with small independent cohorts of patients receiving different antidepressant modalities,<sup>14</sup> independent replication is yet to be established.

Other challenges relate to the choice of EEG technology. Existing models have relied on high-density EEG devices with 64 or more channels, requiring extensive training and expertise to ensure data quality. Yet, for clinical viability, EEG signal must be easily recorded and robust to experimental setup noise. Developing models using features derived from portable EEG devices with fewer channels is an important step toward clinical EEG application. Efforts should be made to use fully automated pipelines for data cleaning and feature extraction, enabling nonexperts to use and process EEG data consistently across patients and sites, making them more suitable for clinical use.

To address these challenges, in this prognostic study we adapted the model developed by the Canadian Biomarker Integration Network in Depression (CAN-BIND)<sup>13</sup> group to make it more suitable for a clinical environment and assessed its generalizability in an independent sample. Our adapted model was built using data from the CAN-BIND1 study,<sup>19,20</sup> where participants received open-label escitalopram treatment across 6 clinical sites. Resting-state, eyes-closed EEG data were collected in 125 participants with high-density EEG devices. Zhdanov et al<sup>13</sup> used features derived from 58 common electrodes across sites at baseline and week 2 to develop the previous model. In our study, only baseline data were used for clinical relevance, and a selection of 32 electrodes was made to emulate a setup closer to those used in a clinical environment with feasibility in mind. We extracted

relative power density and multiscale entropy features and used a fully nested cross-validation on the CAN-BIND data set to obtain an unbiased estimation of the model's performance. We aimed to achieve a higher accuracy compared with the mean accuracy of 56% reported in the 3 adequate-quality studies that were identified in a 2021 meta-analysis<sup>18</sup> and focused on predicting response to medication treatments. Furthermore, we used an independent data set from the Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care (EMBARC) for Depression consortium,<sup>21</sup> where 105 participants were randomized to receive sertraline treatment and 118 participants were randomized to receive placebo treatment across 4 clinical sites. We hypothesized that the predictive accuracy with external validation in the sertraline-treated EMBARC sample would be comparable to the predictive accuracy obtained with internal validation in the CAN-BIND sample.

## Methods

In this prognostic study, we developed and validated a multivariate model utilizing EEG data to predict response to SSRIs, adhering to the Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) reporting guideline.<sup>22</sup> The CAN-BIND<sup>19,20</sup> and EMBARC<sup>21</sup> study protocols were approved by research ethics boards of participating institutions and complied with the Declaration of Helsinki.<sup>23</sup> All participants provided written informed consent.

### Participants Sample

Only participants with baseline EEG data and who completed all assessments at week 8 were included in this study. In CAN-BIND,<sup>19,20</sup> 6 Canadian sites recruited participants aged 18 to 60 years who met the *Diagnostic and Statistical Manual of Mental Disorders* (Fourth Edition) criteria for a major depressive episode of MDD. Participants received escitalopram during the first 8 weeks in an open-label design. The primary outcome was the Montgomery-Åsberg Depression Rating Scale (MADRS), and participants were identified as responders on the basis of a reduction in MADRS score of 50% or more from baseline to week 8.

In EMBARC,<sup>21</sup> 4 US sites recruited participants aged 18 to 65 years with MDD according to the *Diagnostic and Statistical Manual of Mental Disorders* (Fourth Edition). Participants were randomized to receive sertraline or placebo during the first 8 weeks using a double-blind design. The primary outcome measure was the Hamilton Depression Rating Scale (HDRS-17), and participants with a reduction in their HDRS-17 score of 50% or more from baseline to week 8 were classified as responders. More details about CAN-BIND and EMBARC are given in eAppendix 1 in [Supplement 1](#).

### EEG Data Recording and Processing

At baseline, EEG activity was recorded during the eyes-closed resting condition for participants at 4 sites in CAN-BIND<sup>19,20</sup> and at all sites in EMBARC.<sup>21</sup> Because EEG data were collected at multiple sites, data sets were standardized into a common site-independent format. A customized, fully automated cleaning pipeline was then used to remove artifacts from the data sets (more details in eAppendix 2, eTable 1, eAppendix 3, and eAppendix 4 in [Supplement 1](#)).

### EEG Features

#### Spectral Features

Power spectral density (absolute power) was estimated for each channel from 1 Hz to 50 Hz using the Welch method with 2-second nonoverlapping windows. Relative power was then obtained in 5 frequency bands—delta, theta, alpha, beta, and gamma—by calculating the ratio of the absolute power within a frequency band to the total sum of absolute power (1-50 Hz).

### Multiscale Entropy Features

According to methods described in prior studies,<sup>13,24,25</sup> multiscale entropy (MSE) was estimated for each channel on a scale of 1 to 70 using 30-second nonoverlapping windows. The following parameters were used for the computation: embedding dimension ( $m = 2$ ) and similarity tolerance ( $r = 0.15$ ). The MSE was then averaged within 3 timescale bands: fine timescales (1-20), medium timescales (21-35), and coarse timescales (36-70).

### EEG Features Reduction

To reduce the number of EEG features, the 32 electrodes were grouped into 14 brain regions, and features were averaged within each region. Asymmetry features were also computed by dividing features in the left hemisphere by features in the right hemisphere for 5 pairs of brain regions. This led to a total of 152 features. See more details on feature reduction in eAppendix 5 in [Supplement 1](#) and information on electrode mapping in eFigure 1 in [Supplement 1](#).

### Classifier Construction

For the current study, a nonparametric, supervised method called the  $k$ -nearest neighbors algorithm (kNN) was used. The kNN algorithm was chosen due to its capacity to find complex patterns, while making few assumptions regarding the data. With EEG signals exhibiting complex behavior with nonlinear dynamic properties, the kNN algorithm has been widely used in EEG research.<sup>26-29</sup> This algorithm has only a few hyperparameters to be optimized (eg, number of neighbors  $k$  and distance metrics), which simplifies the training process. Finally, the kNN algorithm performs well in low-dimensional EEG data sets.<sup>26</sup>

### Assessments of Generalizability

#### Metrics

Balanced accuracy, sensitivity, and specificity were used to assess model predictive performance. The 95% CIs for these metrics were calculated using the normal approximation method or the Clopper-Pearson exact method, wherever appropriate. Responder status was considered as a positive outcome, and nonresponder as a negative outcome.

#### Internal Validation With CAN-BIND

A kNN classifier was trained using a nested leave-one-out cross-validation with the CAN-BIND data set.<sup>30</sup> In this approach, no final model is built because a different set of optimal hyperparameters might be found in each fold. However, this procedure is essential to obtain an unbiased estimate of the model's performance when no independent data set is available.<sup>16</sup> Additionally, to reduce concerns related to overfitting, we conducted nested  $k$ -fold cross-validation with different values of  $k$  (more details in eAppendix 6 and eAppendix 7 in [Supplement 1](#)).<sup>31</sup>

#### External Validation With EMBARC

A kNN classifier was trained using a traditional leave-one-out cross-validation with the CAN-BIND data set. This process yielded a unique set of optimal hyperparameters, resulting in a final model capable of predicting new data on the basis of its  $k$ -nearest neighbors in the training set. The model's generalizability was tested using the independent EMBARC sertraline group data set. Additionally, its ability to specifically predict response to SSRIs was evaluated by assessing its performance on the EMBARC placebo group data set.

### Feature Importance

To help in the interpretation of the model, permutation feature importance was used to identify the most impactful features. This procedure involved randomly permuting the values of each feature and measuring the resulting effect on the model's performance. To ensure robustness, we repeated this procedure 100 times and computed the mean reduction in balanced accuracy as a measure of

feature importance. We applied this approach to both internal validation with CAN-BIND and external validation with EMBARC (more details in eAppendix 8 in Supplement 1).

Statistical Analysis

In this study, a machine learning approach was used instead of traditional statistical analysis. The ability of the kNN classifier to predict treatment outcomes from EEG data served as a statistical measure of difference between responders and nonresponders. All analyses were performed using MATLAB statistical software version R2020b (MathWorks) from January to December 2022. Statistical significance was set at 2-sided  $P < .05$ .

Results

Participant Characteristics

We analyzed data from 125 participants in CAN-BIND (mean [SD] age, 36.4 [13.0] years; 78 [62.4%] women) who had a mean (SD) MADRS score of 30.0 (5.8) at baseline (Table 1). We also analyzed data from 223 participants in EMBARC with 105 participants in sertraline group (mean [SD] age, 38.4 [13.8] years; 72 [68.6%] women) (Table 2) and 118 participants in the placebo group (mean [SD] age 37.3 [13.1] years; 73 [61.8%] women). Sertraline-treated participants had a mean (SD) HDRS score of 18.5 (4.4) at baseline.

No differences were observed in age (36.4 years [95% CI, 34.1-38.7 years] vs 38.4 years [95% CI, 35.8-41.0 years];  $P = .26$ ) or sex (230 participants;  $\chi^2_1 = 0.96$ ;  $P = .33$ ) between participants in CAN-BIND and sertraline-treated participants in EMBARC.

No differences were observed in age or sex between participants in CAN-BIND and sertraline-treated participants in EMBARC. However, converting HDRS scores to MADRS scores on the basis of

Table 1. Demographics and Clinical Characteristics in the Canadian Biomarker Integration Network in Depression Cohort

Variable	Electroencephalography recording site					Responders, No. (%) (n = 56)	Nonresponders, No. (%) (n = 69)	P value
	CAMH (n = 7)	QNS (n = 18)	TGH (n = 46)	UBC (n = 54)	Total (N = 125)			
Age, mean (SD), y	30.4 (13.0)	42.7 (14.4)	35.8 (13.0)	35.6 (12.0)	36.4 (13.0)	36.1 (13.2)	36.7 (12.8)	.81
Sex, No. (%)								
Female	7 (100.0)	9 (50.0)	27 (58.7)	35 (64.8)	78 (62.4)	37 (66.1)	41 (59.4)	.45
Male	0	9 (50.0)	19 (41.3)	19 (35.2)	47 (37.6)	19 (33.9)	28 (40.6)	
MADRS wk 0 score, mean (SD)	28.0 (5.2)	30.0 (4.7)	32.4 (5.5)	28.3 (5.8)	30.0 (5.8)	29.4 (5.8)	30.5 (5.8)	.30
MADRS wk 8 score, mean (SD)	15.7 (6.0)	18.1 (10.2)	19.5 (11.9)	14.2 (9.2)	16.8 (10.5)	7.8 (5.0)	24.2 (7.6)	<.001
Change in MADRS score, mean (SD), %	44.5 (17.0)	39.2 (35.9)	39.6 (32.6)	49.8 (32.9)	44.2 (32.7)	73.5 (16.0)	20.5 (21.5)	<.001

Abbreviations: CAMH, Centre for Addiction and Mental Health; MADRS, Montgomery-Åsberg Depression Rating Scale; QNS, Queen's University; TGH, Toronto General Hospital; UBC, University of British Columbia.

Table 2. Demographic and Clinical Characteristics in the Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care Sertraline Group

Variable	Electroencephalography recording site				Total (N = 105)	Responders (n = 51)	Nonresponders (n = 54)	P value
	CU (n = 34)	MGH (n = 18)	TX (n = 36)	UM (n = 17)				
Age, mean (SD), y	35.0 (11.9)	36.1 (15.1)	43.8 (13.5)	36.2 (14.4)	38.4 (13.8)	38.7 (13.1)	38.2 (14.5)	.86
Sex, No. (%)								
Female	24 (70.6)	10 (55.6)	27 (75.0)	11 (64.7)	72 (68.6)	35 (68.6)	37 (68.5)	.99
Male	10 (29.4)	8 (44.4)	9 (25.0)	6 (35.3)	33 (31.4)	16 (31.4)	17 (31.5)	
HDRS wk 0 score, mean (SD)	17.4 (4.3)	19.3 (4.1)	19.3 (4.9)	18.4 (3.9)	18.5 (4.4)	19.4 (4.2)	17.8 (4.6)	.07
HDRS wk 8 score, mean (SD)	7.4 (6.1)	12.6 (5.6)	12.4 (6.2)	10.9 (6.4)	10.6 (6.5)	5.3 (6.5)	15.5 (3.0)	<.001
Change in HDRS score, mean (SD), %	54.4 (39.4)	31.9 (33.4)	32.0 (38.3)	39.2 (35.4)	40.4 (38.3)	72.7 (38.3)	10.0 (27.4)	<.001

Abbreviations: CU, Columbia University; HDRS, Hamilton Depression Rating Scale; MGH, Massachusetts General Hospital; TX, University of Texas Southwestern Medical Center; UM, University of Michigan, Ann Arbor.

item response theory–derived conversion tables<sup>32,33</sup> revealed that sertraline-treated participants in EMBARC had significantly lower symptoms severity at baseline (mean [SD] HDRS converted to MADRS score, 25.5 [5.5]; 95% CI, 24.5-26.6) than participants in CAN-BIND (mean [SD] MADRS score, 30.0 [5.8]; 95% CI, 29.0-31.0; *P* < .01). See eAppendix 9, eFigure 2, and eTables 2-5 in [Supplement 1](#) for more information on the CAN-BIND and EMBARC participants.

Classifier Construction and Performance Estimation

Internal Validation With CAN-BIND

The kNN algorithm with nested leave-one-out cross-validation predicted response with a balanced accuracy of 64.2% (95% CI, 55.8%-72.6%), a sensitivity of 66.1% (95% CI, 53.7%-78.5%), and a specificity of 62.3% (95% CI, 50.1%-73.8%) in CAN-BIND (**Table 3**). In comparison, the nested 10-fold, cross-validation led to a balanced accuracy of 61.7% (95% CI, 52.1%-69.3%), a sensitivity of 62.3% (95% CI, 49.6%-75.0%), and a specificity of 59.1% (95% CI, 47.5%-70.7%). Results for other values of *k* in nested *k*-fold cross-validation can be found in eAppendix 10 and eFigure 4 in [Supplement 1](#).

External Validation With EMBARC

After training the model with data from CAN-BIND, we assessed its ability to generalize to new data by testing it with data from the EMBARC sertraline-treated group (eAppendix 11 and eTable 8 in [Supplement 1](#)). The model achieved a balanced accuracy of 63.7% (95% CI, 54.5%-72.8%), a sensitivity of 58.8% (95% CI, 45.3%-72.3%), and specificity of 68.5% (95% CI, 56.1%-80.9%) (**Table 4**). Additionally, the model was tested with data from the EMBARC placebo group, but did not perform better than chance, yielding a balanced accuracy of 48.7% (95% CI, 39.3%-58.0%), a sensitivity of 50.0% (95% CI, 35.2%-64.8%), and specificity of 47.3% (95% CI, 35.9%-58.7%) (eTable 6 in [Supplement 1](#)). Furthermore, a significant difference was observed between the predictions in EMBARC sertraline and placebo groups (223 participants;  $\chi^2_1 = 5.41$ ; *P* = .02) (eTable 7 in [Supplement 1](#)). To assess the association of baseline symptoms severity differences between CAN-BIND and EMBARC participants with generalizability, we also conducted an exploratory analysis by excluding EMBARC participants with low baseline HDRS scores (eTable 9 in [Supplement 1](#)). More information regarding symptoms severity and generalizability can be found in eAppendix 12 and eFigure 5 in [Supplement 1](#).

Table 3. Internal Validation With Canadian Biomarker Integration Network in Depression Data					
Measure	Percentage (95% CI)				
	CAMH participants (n = 7)	QNS participants (n = 18)	TGH participants (n = 46)	UBC participants (n = 54)	Total participants (N = 125)
Sensitivity	66.7 (9.4-99.2)	100.0 (59.0-100.0)	65.0 (40.8-84.6)	57.7 (36.9-76.7)	66.1 (53.7-78.5)
Specificity	75.0 (19.4-99.4)	63.6 (30.8-89.1)	65.4 (44.3-82.8)	57.1 (37.2-75.5)	62.3 (50.1-73.8)
Balanced accuracy	70.8 (14.4-99.3)	81.8 (44.9-94.6)	65.2 (42.6-83.7)	57.4 (31.1-76.1)	64.2 (55.8-72.6)

Abbreviations: CAMH, Centre for Addiction and Mental Health; QNS, Queen's University; TGH, Toronto General Hospital; UBC, University of British Columbia.

Table 4. External Validation With Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care Sertraline Group Data					
Measure	Percentage (95% CI)				
	CU participants (n = 34)	MGH participants (n = 18)	TX participants (n = 36)	UM participants (n = 17)	Total participants (N = 105)
Sensitivity	61.9 (38.4-81.9)	80.0 (28.4-99.5)	52.9 (27.8-77.0)	50.0 (15.7-84.3)	58.8 (45.3-72.3)
Specificity	69.2 (38.6-90.9)	84.6 (54.6-98.1)	63.2 (38.4-83.7)	55.6 (21.2-86.3)	68.5 (56.1-80.9)
Balanced accuracy	65.6 (38.5-86.3)	82.3 (41.5-98.8)	58.1 (33.1-80.4)	52.8 (18.5-85.3)	63.7 (54.5-72.8)

Abbreviations: CU, Columbia University; MGH, Massachusetts General Hospital; TX, University of Texas Southwestern Medical Center; UM, University of Michigan Ann Arbor.



## Feature Importance

The process of randomly permuting the values of each of the 152 features resulted in a reduction in balanced accuracy ranging from 0.6% to 3.5% in internal validation with CAN-BIND, and from -0.2% to 3.7% in external validation with the EMBARC sertraline group (**Figure**). No clear structured pattern of informative features was identified when comparing internal and external validations. However, the asymmetry of power and multiscale entropy, especially in frontal, central, and parietal regions of the brain, provided the more robust features on the model's performance in both validations. Feature importance for EMBARC placebo group can be found in eAppendix 13 and eFigure 6 in [Supplement 1](#).

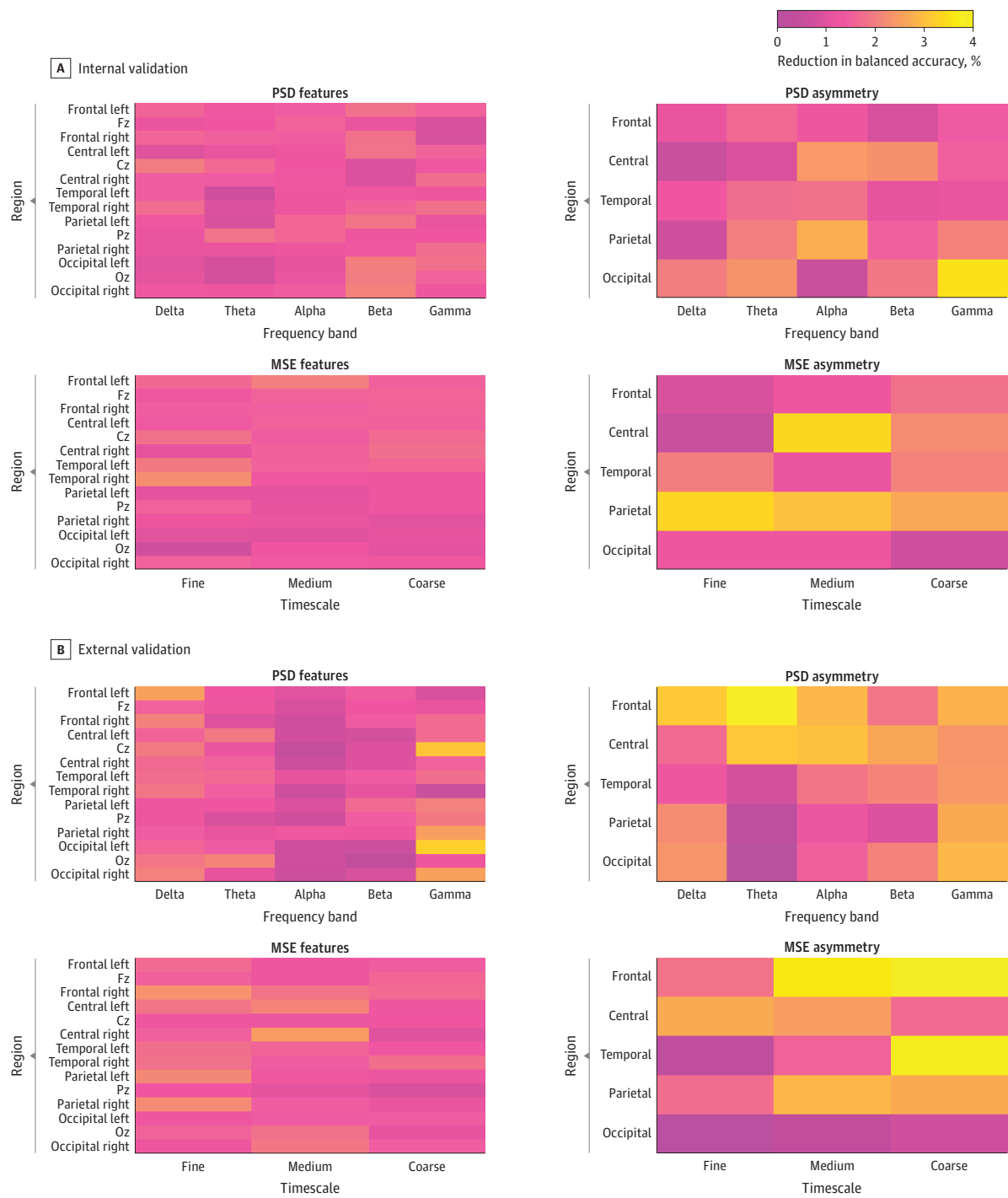
## Discussion

In this prognostic study, we described an EEG-based model that predicts outcomes of treatment with SSRI antidepressants using 2 independent data sets. Our model achieved 64.2% balanced accuracy using internal validation with CAN-BIND, and 63.7% accuracy using external validation with EMBARC. This finding is of clinical importance because it provides a substantial improvement over trial-and-error methods. This observation is notably accentuated by the initial low response rates in the CAN-BIND and EMBARC samples. Accurately predicting treatment response in almost two-thirds of cases could lead to improved patient outcomes, reducing ineffective treatments and health care burden. Prior studies that reported higher accuracy with EEG-based models<sup>13,34-36</sup> may have been overoptimistic due to limited sample sizes and overfitting in their validation procedures. In fact, Sajjadi et al<sup>18</sup> found that of 54 studies predicting outcomes to medication treatment, only 8 met the criteria for adequate quality (ie, defined as having a minimum sample of 100 participants and an adequate validation method), and none of these studies used EEG for prediction. Moreover, the mean accuracy across these 8 adequate-quality studies was 63%, with prediction accuracies of 69% for treatment resistance, 60% for remission, and 56% for treatment response.<sup>18</sup> Here, our study demonstrated that a model using EEG predictors for treatment response could perform as well as, if not better than, models using other types of predictors, such as genetic, clinical, or demographic data.

Furthermore, our EEG-based model presents substantial advantages for clinical practice. First, using objective measures such as EEG data to predict treatment response is preferable to relying on subjective variables in clinical setting. Objective biomarkers have been suggested to reduce the self-stigma associated with seeking help and reporting symptoms on clinical questionnaires.<sup>37</sup> Second, our model here generalized across 2 cohorts of patients who received different SSRIs (escitalopram and sertraline). For clinical practice in the near future, it will be probably more feasible to have a model that predicts response to a certain class of antidepressant (ie, SSRIs), than a predictive model for every single existing antidepressant medication, which would require extensive time and effort to collect sufficient EEG data. Furthermore, the model did not predict response to placebo treatment, suggesting its specificity for predicting outcomes with SSRIs and not just symptom improvement. Third, our model used only 32 electrodes, making it compatible with data recorded with portable EEG devices. Compared with research-grade EEG devices with 64 or more channels, these devices are more suitable for clinical settings with limited resources and time, prioritizing cost-effectiveness, ease of use, patient comfort, and signal quality over spatial resolution. Finally, the kNN algorithm used here is relatively simple and does not require complex training because it relies on observable data similarities to generate accurate predictions.

Regarding feature importance, the results suggest that power and multiscale entropy features from different regions are necessary to accurately predict treatment response. Interestingly, it was observed that asymmetry features in frontal, central, and parietal regions had the most impact on the model's performance in both validations. This finding is consistent with previous studies that reported differences in EEG asymmetry features between responders and nonresponders to medications.<sup>6,7,38,39</sup> These results support the potential usefulness of asymmetry measures in predicting treatment response and may contribute to the development of more accurate prediction

Figure. Feature Importance in Internal and External Validations



The figure displays the feature importance results obtained in internal validation with the Canadian Biomarker Integration Network in Depression (CAN-BIND) cohort (A) and external validation with the Establishing Moderators and Biosignatures of Antidepressant Response for Clinical Care (EMBARC) sertraline-treated group (B), where each matrix corresponds to the importance of power spectral density (PSD), PSD asymmetry, multiscale entropy (MSE), and MSE asymmetry features. The depicted features in pink colors indicate a low level of importance for the model, whereas the features represented in yellow colors indicate a stronger level of importance. Feature importance was evaluated on the basis of the reduction in balanced accuracy in internal validation (A) and external validation (B). To enhance the clarity of the visualization, features exhibiting an increase in balanced accuracy were adjusted to 0%, indicating a lack of valuable information for predictive purposes (low level of importance). Cz indicates central zone; Fz, frontal zone; and Oz, occipital zone.



models. Furthermore, the impact of individual features on the predictive model varied across the internal and external validation sets, indicating that a combination of features, rather than individual features, may be more relevant for accurate prediction. Overall, these findings highlight the importance of considering multiple features and their interactions when predicting treatment response to SSRIs.

An interesting point of this study was the association of symptom severity with model generalizability. The results suggest that the model might be better suited with participants experiencing moderate or severe depression. These findings highlight the potential need for distinct predictive models based on baseline symptom severity, driven by the inherent heterogeneity of depression.

## Limitations

This study has limitations regarding predictive accuracy. EEG-based models are not necessarily intended for stand-alone use in clinical practice, and future studies may enhance accuracy by incorporating other data sources. Recent research<sup>40</sup> combining clinical, molecular, and magnetic resonance imaging data in the CAN-BIND sample showed improved treatment outcome prediction, but their models using baseline data achieved a mean balanced accuracy of 57%, which is still inferior to the accuracy obtained with EEG-alone here. Future studies could explore combining multimodal predictors with EEG to further improve predictive accuracy. It is also important to consider the theoretical ceiling in predicting treatment response in depression due to uncertainties associated with the response labels. For instance, Zhdanov and colleagues<sup>13</sup> estimated the maximal achievable classification accuracy to be approximately 86%. Furthermore, EEG data can be sensitive to noise, artifacts, and other sources of variability, and the estimated 86% accuracy could be an upper limit, potentially attenuated in a clinical setting. Nonetheless, the achieved 64.2% balanced accuracy obtained in both internal and external validation suggests robustness of our EEG-based predictive model for treatment response in the presence of potential clinical variability and measurement errors. Other limitations include medication-free participants at baseline and exclusion of those with missing data. Future studies could use more diverse data sets including patients who are already taking an antidepressant or those for whom their first treatment failed to enhance the model's generalizability and ensure its applicability to a wider range of patients. Additionally, validating the findings in larger cohorts through future research or follow-up studies would be valuable for further refining the predictive model's performance. Furthermore, the study focused only on eyes-closed resting-state EEG databased on previous evidence.<sup>13</sup> However, Wu et al<sup>14</sup> found eyes-open condition to be more predictive of sertraline response. Future research may wish to further explore the value of both types of EEG data to improve predictive accuracy.

## Conclusions

We developed an EEG-based model with potential for clinical use to predict response to SSRIs, taking into account the practical application of EEG technology in a clinical setting. This study is the first, to our knowledge, to present replicable evidence of such a model and be validated in a large, independent cohort. These findings represent a substantial advancement toward using EEG in future clinical practice and supporting its potential to match patients with MDD to optimized treatment.

## ARTICLE INFORMATION

**Accepted for Publication:** August 19, 2023.

**Published:** September 28, 2023. doi:[10.1001/jamanetworkopen.2023.36094](https://doi.org/10.1001/jamanetworkopen.2023.36094)

**Open Access:** This is an open access article distributed under the terms of the [CC-BY License](https://creativecommons.org/licenses/by/4.0/). © 2023 Schwartzmann B et al. *JAMA Network Open*.

**Corresponding Author:** Faranak Farzan, PhD, School of Mechatronic Systems Engineering, Simon Fraser University, 250-13450 102 Avenue, Surrey, BC V3T 0A3, Canada ([ffarzan@sfu.ca](mailto:ffarzan@sfu.ca)).

**Author Affiliations:** eBrain Lab, School of Mechatronic Systems Engineering, Simon Fraser University, Surrey, British Columbia, Canada (Schwartzmann, Dhami, Farzan); Department of Psychiatry, University of Toronto, Toronto, Ontario, Canada (Dhami, Müller, Rotzinger, Kennedy, Farzan); Centre for Addiction and Mental Health, Toronto, Ontario, Canada (Dhami, Müller, Farzan); Department of Psychiatry, Dalhousie University, Halifax, Nova Scotia, Canada (Uher); Department of Psychiatry, University of British Columbia, Vancouver, British Columbia, Canada (Lam); Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario, Canada (Frey); Mood Disorders Program and Women's Health Concerns Clinic, St. Joseph's Healthcare, Hamilton, Ontario, Canada (Frey, Rotzinger); Department of Psychiatry, Queen's University, Providence Care, Kingston, Ontario, Canada (Milev, Soares); Department of Psychology, Queen's University, Providence Care, Kingston, Ontario, Canada (Milev, Soares); Mood Disorders Research Unit, University of Ottawa Institute of Mental Health Research, Ottawa, Ontario, Canada (Blier); University of Michigan Depression Center, Ann Arbor (Parikh); Douglas Mental Health University Institute, Department of Psychiatry, McGill University, Montreal, Quebec, Canada (Turecki); Department of Psychiatry and Behavioural Neurosciences, McMaster University, Hamilton, Ontario, Canada (Foster); Center for Depression Research and Clinical Care, University of Texas Southwestern Medical Center, Dallas (Foster); Unity Health Toronto, Toronto, Ontario, Canada (Kennedy).

**Author Contributions:** Mr Schwartzmann had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

*Concept and design:* Uher, Frey, Milev, Blier, Soares, Foster, Rotzinger, Kennedy, Farzan.

*Acquisition, analysis, or interpretation of data:* Schwartzmann, Dhami, Uher, Lam, Frey, Milev, Mueller, Soares, Parikh, Turecki, Farzan.

*Drafting of the manuscript:* Schwartzmann, Dhami, Soares, Farzan.

*Critical review of the manuscript for important intellectual content:* All authors.

*Statistical analysis:* Schwartzmann, Dhami, Uher.

*Obtained funding:* Lam, Milev, Soares, Foster, Rotzinger, Kennedy, Farzan.

*Administrative, technical, or material support:* Lam, Frey, Milev, Mueller, Soares, Parikh, Foster, Rotzinger.

*Supervision:* Dhami, Uher, Frey, Milev, Kennedy, Farzan.

**Conflict of Interest Disclosures:** Dr Lam reported receiving grants from Janssen; personal fees from Carnot, Lundbeck, AbbVie, Bausch, Medscape, Otsuka, Neurotorium, and Pfizer/Viatris; and holding a copyright patent for the Lam Employment Absence and Productivity Scale outside the submitted work. Dr Milev reported receiving grants from, CANBIND, Lallemand, and Nybyiota; and personal fees from AbbVie, Janssen, and Otsuka, outside the submitted work. Dr Soares reported receiving grants from Eisai Inc, Clairvoyant Therapeutics; and personal fees from Otsuka and Ontario Health outside the submitted work. Dr Parikh reported receiving grants from Sage, Merck, Aifred, and Janssen; personal fees from Sage, Myriad Research, Aifred, Janssen, Otsuka, Mensante, and Takeda outside the submitted work. Dr Foster reported receiving personal fees from Rothman, Benson, Hedges Inc, Takeda, WebMD, Alphasights, and Microbial Resource Management Health Newfoundland and Labrador outside the submitted work. Dr Rotzinger holding a patent for Teneurin C-terminal associated peptides (TCAP) and methods and uses thereof. Dr Kennedy reported receiving personal fees from Abbvie, Boehringer-Ingelheim, Janssen, Lundbeck, Merck, Neurotorium, Otsuka, Pfizer, Sunovion, and Servier; grants from Abbott, Brain Canada, Janssen, Otsuka, and Strategy for Patient-Oriented Research. Dr Farzan reported receiving grants from the Michael Smith Foundation for Health Research during the conduct of the study. No other disclosures were reported.

**Funding/Support:** CAN-BIND is an Integrated Discovery Program that is carried out in partnership with and receives financial support from the Ontario Brain Institute, an independent nonprofit corporation, funded partially by the Ontario government. Additional funding was provided by the Canadian Institutes of Health Research (CIHR), Lundbeck, and Servier. Funding and/or in-kind support was also provided by the investigators' universities and academic institutions. All study medications are independently purchased at wholesale market values. Some data and/or research tools used in the preparation of this manuscript were obtained from the National Institute of Mental Health Data Archive (NDA). NDA is a collaborative informatics system created by the National Institutes of Health to provide a national resource to support and accelerate research in mental health (dataset identifier: 10.15154/1528397).

**Role of the Funder/Sponsor:** The funders had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Disclaimer:** The opinions, results and conclusions are those of the authors and no endorsement by the Ontario Brain Institute is intended or should be inferred. This manuscript reflects the views of the authors and may not

reflect the opinions or views of the National Institutes of Health or of the submitters submitting original data to NDA.

**Data Sharing Statement:** See [Supplement 2](#).

## REFERENCES

1. World Health Organization. Depressive disorder (depression). March 31, 2023. Accessed August 25, 2023. <https://www.who.int/news-room/fact-sheets/detail/depression>
2. Rush AJ, Trivedi MH, Wisniewski SR, et al. Acute and longer-term outcomes in depressed outpatients requiring one or several treatment steps: a STAR\*D report. *Am J Psychiatry*. 2006;163(11):1905-1917. doi:10.1176/ajp.2006.163.11.1905
3. Trivedi MH, Rush AJ, Wisniewski SR, et al; STAR\*D Study Team. Evaluation of outcomes with citalopram for depression using measurement-based care in STAR\*D: implications for clinical practice. *Am J Psychiatry*. 2006;163(1):28-40. doi:10.1176/appi.ajp.163.1.28
4. Iosifescu DV, Greenwald S, Devlin P, et al. Frontal EEG predictors of treatment outcome in major depressive disorder. *Eur Neuropsychopharmacol*. 2009;19(11):772-777. doi:10.1016/j.euroneuro.2009.06.001
5. Korb AS, Hunter AM, Cook IA, Leuchter AF. Rostral anterior cingulate cortex theta current density and response to antidepressants and placebo in major depression. *Clin Neurophysiol*. 2009;120(7):1313-1319. doi:10.1016/j.clinph.2009.05.008
6. Baskaran A, Farzan F, Milev R, et al; CAN-BIND Investigators Team. The comparative effectiveness of electroencephalographic indices in predicting response to escitalopram therapy in depression: a pilot study. *J Affect Disord*. 2018;227:542-549. doi:10.1016/j.jad.2017.10.028
7. Bruder GE, Sedoruk JP, Stewart JW, McGrath PJ, Quitkin FM, Tenke CE. Electroencephalographic alpha measures predict therapeutic response to a selective serotonin reuptake inhibitor antidepressant: pre- and post-treatment findings. *Biol Psychiatry*. 2008;63(12):1171-1177. doi:10.1016/j.biopsych.2007.10.009
8. Knott V, Mahoney C, Kennedy S, Evans K. Pre-treatment EEG and its relationship to depression severity and paroxetine treatment outcome. *Pharmacopsychiatry*. 2000;33(6):201-205. doi:10.1055/s-2000-8356
9. Knott V, Mahoney C, Kennedy S, Evans K. EEG correlates of acute and chronic paroxetine treatment in depression. *J Affect Disord*. 2002;69(1-3):241-249. doi:10.1016/S0165-0327(01)00308-1
10. Jaworska N, Blondeau C, Tessier P, et al. Examining relations between alpha power as well as anterior cingulate cortex-localized theta activity and response to single or dual antidepressant pharmacotherapies. *J Psychopharmacol*. 2014;28(6):587-595. doi:10.1177/0269881114523862
11. Widge AS, Bilge MT, Montana R, et al. Electroencephalographic biomarkers for treatment response prediction in major depressive illness: a meta-analysis. *Am J Psychiatry*. 2019;176(1):44-56. doi:10.1176/appi.ajp.2018.17121358
12. Iosifescu DV. Are electroencephalogram-derived predictors of antidepressant efficacy closer to clinical usefulness? *JAMA Netw Open*. 2020;3(6):e207133. doi:10.1001/jamanetworkopen.2020.7133
13. Zhdanov A, Atluri S, Wong W, et al. Use of machine learning for predicting escitalopram treatment outcome from electroencephalography recordings in adult patients with depression. *JAMA Netw Open*. 2020;3(1):e1918377. doi:10.1001/jamanetworkopen.2019.18377
14. Wu W, Zhang Y, Jiang J, et al. An electroencephalographic signature predicts antidepressant response in major depression. *Nat Biotechnol*. 2020;38(4):439-447. doi:10.1038/s41587-019-0397-3
15. Rajpurkar P, Yang J, Dass N, et al. Evaluation of a machine learning model based on pretreatment symptoms and electroencephalographic features to predict outcomes of antidepressant treatment in adults with depression: a prespecified secondary analysis of a randomized clinical trial. *JAMA Netw Open*. 2020;3(6):e206653. doi:10.1001/jamanetworkopen.2020.6653
16. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. *PLoS One*. 2019;14(11):e0224365. doi:10.1371/journal.pone.0224365
17. Park SH, Han K. Methodologic guide for evaluating clinical performance and effect of artificial intelligence technology for medical diagnosis and prediction. *Radiology*. 2018;286(3):800-809. doi:10.1148/radiol.2017171920
18. Sajjadian M, Lam RW, Milev R, et al. Machine learning in the prediction of depression treatment outcomes: a systematic review and meta-analysis. *Psychol Med*. 2021;51(16):2742-2751. doi:10.1017/S0033291721003871

19. Kennedy SH, Lam RW, Rotzinger S, et al; CAN-BIND Investigator Team. Symptomatic and functional outcomes and early prediction of response to escitalopram monotherapy and sequential adjunctive aripiprazole therapy in patients with major depressive disorder: a CAN-BIND-1 report. *J Clin Psychiatry*. 2019;80(2):18m12202. doi:10.4088/JCP.18m12202
20. Lam RW, Milev R, Rotzinger S, et al; CAN-BIND Investigator Team. Discovering biomarkers for antidepressant response: protocol from the canadian biomarker integration network in depression (CAN-BIND) and clinical characteristics of the first patient cohort. *BMC Psychiatry*. 2016;16(1):105. doi:10.1186/s12888-016-0785-x
21. Trivedi MH, McGrath PJ, Fava M, et al. Establishing moderators and biosignatures of antidepressant response in clinical care (EMBARC): rationale and design. *J Psychiatr Res*. 2016;78:11-23. doi:10.1016/j.jpsychires.2016.03.001
22. Collins GS, Reitsma JB, Altman DG, Moons KGM. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD Statement. *BMC Med*. 2015;13(1):1-10. doi:10.1186/s12916-014-0241-z
23. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191-2194. doi:10.1001/jama.2013.281053
24. Costa M, Goldberger AL, Peng CK. Multiscale entropy analysis of biological signals. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2005;71(2 Pt 1):021906. doi:10.1103/PhysRevE.71.021906
25. Farzan F, Atluri S, Mei Y, et al. Brain temporal complexity in explaining the therapeutic and cognitive effects of seizure therapy. *Brain*. 2017;140(4):1011-1025. doi:10.1093/brain/awx030
26. Sha'abani MNAH, Fuad N, Jamal N, Ismail MF. kNN and SVM classification for EEG: a review. In: Nor Kasruddin Nasir A, Ashraf Ahmad M, Najib MS, eds. *ECCE2019: Proceedings of the 5th International Conference on Electrical, Control & Computer Engineering, Kuantan, Pahang, Malaysia, 29th July 2019*. Springer Singapore; 2020:555-565.
27. Hasanzadeh F, Mohebbi M, Rostami R. Prediction of rTMS treatment response in major depressive disorder using machine learning techniques and nonlinear features of EEG signal. *J Affect Disord*. 2019;256(May):132-142. doi:10.1016/j.jad.2019.05.070
28. Khosla A, Khandnor P, Chand T. Automated diagnosis of depression from EEG signals using traditional and deep learning approaches: a comparative analysis. *Biocybern Biomed Eng*. 2022;42(1):108-142. doi:10.1016/j.bbe.2021.12.005
29. Cai H, Qu Z, Li Z, Zhang Y, Hu X, Hu B. Feature-level fusion approaches based on multimodal EEG data for depression recognition. *Inf Fusion*. 2020;59:127-138. doi:10.1016/j.inffus.2020.01.008
30. Wong TT. Performance evaluation of classification algorithms by k-fold and leave-one-out cross validation. *Pattern Recognit*. 2015;48(9):2839-2846. doi:10.1016/j.patcog.2015.03.009
31. Varoquaux G. Cross-validation failure: small sample sizes lead to large error bars. *Neuroimage*. 2018;180(Pt A):68-77. doi:10.1016/j.neuroimage.2017.06.061
32. Uher R, Farmer A, Maier W, et al. Measuring depression: comparison and integration of three scales in the GENDEP study. *Psychol Med*. 2008;38(2):289-300. doi:10.1017/S0033291707001730
33. Leucht S, Fennema H, Engel RR, Kaspers-Janssen M, Szegedi A. Translating the HAM-D into the MADRS and vice versa with equipercentile linking. *J Affect Disord*. 2018;226:326-331. doi:10.1016/j.jad.2017.09.042
34. Jaworska N, de la Salle S, Ibrahim MH, Blier P, Knott V. Leveraging machine learning approaches for predicting antidepressant treatment response using electroencephalography (EEG) and clinical data. *Front Psychiatry*. 2019;9(JAN):768. doi:10.3389/fpsy.2018.00768
35. Khodayari-Rostamabad A, Reilly JP, Hasey GM, de Bruin H, Maccrimmon DJ. A machine learning approach using EEG data to predict response to SSRI treatment for major depressive disorder. *Clin Neurophysiol*. 2013;124(10):1975-1985. doi:10.1016/j.clinph.2013.04.010
36. Mumtaz W, Xia L, Mohd Yasin MA, Azhar Ali SS, Malik AS. A wavelet-based technique to predict treatment outcome for major depressive disorder. *PLoS One*. 2017;12(2):e0171409. doi:10.1371/journal.pone.0171409
37. Clement S, Schauman O, Graham T, et al. What is the impact of mental health-related stigma on help-seeking? a systematic review of quantitative and qualitative studies. *Psychol Med*. 2015;45(1):11-27. doi:10.1017/S0033291714000129
38. Jaworska N, Blier P, Fusee W, Knott V. A Power, a asymmetry and anterior cingulate cortex activity in depressed males and females. *J Psychiatr Res*. 2012;46(11):1483-1491. doi:10.1016/j.jpsychires.2012.08.003
39. Arns M, Bruder G, Hegerl U, et al. EEG alpha asymmetry as a gender-specific predictor of outcome to acute treatment with different antidepressant medications in the randomized iSPOT-D study. *Clin Neurophysiol*. 2016;127(1):509-519. doi:10.1016/j.clinph.2015.05.032

40. Sajjadian M, Uher R, Ho K, et al. Prediction of depression treatment outcome from multimodal data: a CAN-BIND-1 report. *Psychol Med*. Published online August 25, 2022. doi:[10.1017/S0033291722002124](https://doi.org/10.1017/S0033291722002124)

## SUPPLEMENT 1.

**eAppendix 1.** Participant Samples

**eAppendix 2.** EEG Data Recording

**eTable 1.** EEG Amplifier Settings Used in CAN-BIND and EMBARC

**eAppendix 3.** EEG Data Standardization

**eAppendix 4.** EEG Data Processing

**eAppendix 5.** EEG Features Reduction

**eFigure 1.** Division of the Brain Regions and Their Included Electrodes

**eAppendix 6.** Nested Leave-One-Out Cross-Validation

**eAppendix 7.** Nested K-Fold Cross-Validation

**eAppendix 8.** Feature Importance

**eAppendix 9.** Results: Participant Characteristics

**eFigure 2.** Flow of Participants in CAN-BIND

**eTable 2.** Demographics and Clinical Characteristics in CAN-BIND of Participants Who Completed Week 8

**eTable 3.** Demographics and Clinical Characteristics in CAN-BIND of Participants Who Completed Week 8 and Had EEG Recording

**eFigure 3.** Flow of Participants in EMBARC

**eTable 4.** Demographics and Clinical Characteristics in EMBARC Active Arm of Participants Who Completed Week 8

**eTable 5.** Demographics and Clinical Characteristics in EMBARC Active Arm of Participants Who Completed Week 8 and Had EEG Recording

**eAppendix 10.** Classifier Construction and Performance Estimation: Internal Validation with CAN-BIND

**eFigure 4.** Balanced Accuracy in Internal Validation with CAN-BIND

**eAppendix 11.** Classifier Construction and Performance Estimation: External Validation with EMBARC

**eTable 6.** External Validation With EMBARC Placebo Arm

**eTable 7.** Contingency Table With Correct and Incorrect Predictions in EMBARC Sertraline and Placebo Arms

**eAppendix 12.** Impact of Symptoms Severity on Generalizability

**eTable 8.** External Validation With EMBARC Sertraline Arm (Participants With HDRS  $\geq 16$ )

**eTable 9.** External Validation With EMBARC Placebo Arm (Participants With HDRS  $\geq 16$ )

**eFigure 5.** Impact of Symptoms Severity on Generalizability

**eAppendix 13.** Feature Importance

**eFigure 6.** Feature Importance in External Validation With EMBARC Placebo Arm

**eReferences**

## SUPPLEMENT 2.

**Data Sharing Statement**