Open Access Library Journal
2026, Volume 13, e15142
ISSN Online: 2333-9721
ISSN Print: 2333-9705

AI-Driven Prediction of Antidepressant-Induced Mood Switching in Bipolar Disorder: A Synthetic Proof-of-Concept Machine Learning Study Using Clinical, Temporal, and Biomarker Data

Rocco de Filippis[1*], Abdullah Al Foysal[2]

[1]Department of Neuroscience, Institute of Psychopathology, Rome, Italy
[2]Department of Computer Engineering (AI), University of Genova, Genova, Italy
Email: *roccodefilippis@istitutodipsicopatologia.it, niloyhasanfoysal440@gmail.com

How to cite this paper: de Filippis, R. and Al Foysal, A. (2026) AI-Driven Prediction of Antidepressant-Induced Mood Switching in Bipolar Disorder: A Synthetic Proof-of-Concept Machine Learning Study Using Clinical, Temporal, and Biomarker Data. *Open Access Library Journal*, **13**: e15142.
https://doi.org/10.4236/oalib.1115142

Received: March 10, 2026
Accepted: May 26, 2026
Published: May 29, 2026

Abstract

The use of antidepressants in bipolar depression remains one of the most challenging decisions in psychiatric practice, with substantial risks of treatment-emergent affective switching (TEAS) to manic, hypomanic, or mixed states. Current clinical guidelines rely primarily on symptom history and clinical intuition, lacking objective, individualized biomarkers to predict treatment response and switching risk. We introduce a comprehensive machine learning framework that integrates clinical history, temporal patterns, and biomarker data to predict antidepressant-induced mood switching in bipolar disorder. Our approach employs multiple advanced algorithms including Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machines (SVM) trained on a carefully curated dataset of 2500 synthetic patients with realistic clinical characteristics. The framework incorporates established risk factors including bipolar subtype, rapid cycling status, mood state at prescription, antidepressant class and dosage, concomitant mood stabilizer use, and novel biomarkers including BDNF levels, inflammatory markers (CRP), cortisol, and polygenic risk scores. The Gradient Boosting classifier achieved superior performance (AUC-ROC = 0.861, F1-Score = 0.58) compared to other algorithms, with Random Forest demonstrating the highest precision (0.53) for identifying high-risk patients. Feature importance analysis revealed bipolar Type I diagnosis, rapid cycling, mixed mood states, and sleep deprivation as the strongest predictors of switching risk. Risk stratification analysis successfully categorized patients into five distinct risk tiers, with the highest risk group

showing 85% observed switch rates versus 3% in the lowest risk group. Beyond prediction, we provide interpretable clinical decision support through partial dependence analysis, calibration plots, and decision curve analysis demonstrating clinical utility across threshold probabilities. Our framework addresses the critical need for precision psychiatry tools in bipolar disorder management, offering a pathway toward personalized antidepressant prescribing those balances therapeutic benefit against switching risk. This work constitutes a synthetic proof-of-concept study; all models were trained and evaluated exclusively on computationally generated data. Results demonstrate methodological feasibility and should not be interpreted as evidence of clinical deployability. External validation on real-world patient cohorts is required before clinical implementation.

Subject Areas

Psychiatry & Psychology

Keywords

Bipolar Disorder, Antidepressants, Treatment-Emergent Affective Switching, Machine Learning, Predictive Modelling, Biomarkers, Precision Psychiatry, Clinical Decision Support

1. Introduction

Bipolar disorder affects approximately 1% - 2% of the global population and represents one of the most disabling psychiatric conditions, with bipolar depression accounting for the majority of illness burden and associated functional impairment [1]-[3]. The management of bipolar depression presents a fundamental clinical dilemma that has perplexed psychiatrists for decades: while antidepressants are first-line treatments for unipolar depression, their use in bipolar disorder carries substantial and potentially serious risks. The primary concern is treatment-emergent affective switching (TEAS) the induction of manic, hypomanic, or mixed states that can worsen illness trajectory, increase hospitalization rates, and potentially accelerate illness progression through kindling mechanisms [4]-[8]. Meta-analyses suggest that 15% - 30% of bipolar patients treated with antidepressants experience affective switches, with risk varying significantly based on clinical characteristics that are difficult to assess prospectively [9]-[12]. Current treatment guidelines from major psychiatric organizations offer conflicting recommendations regarding antidepressant use in bipolar depression [13]-[16]. The American Psychiatric Association guidelines cautiously support antidepressant use in combination with mood stabilizers for select patients, while the British Association for Psychopharmacology recommends avoiding antidepressants altogether in favour of mood stabilizers and atypical antipsychotics. This ambiguity reflects a fundamental gap in our ability to identify which bipolar patients will benefit from anti-

depressants versus those who will experience harmful mood elevation. From a neurobiological and clinical perspective, antidepressant response and switching risk likely reflect distinct patterns of illness characteristics, biomarker profiles, and treatment contexts. Treatment responders may exhibit bipolar Type II diagnosis, pure depressive episodes without mixed features, stable mood patterns without rapid cycling, and protective factors such as concomitant mood stabilizer use [17]-[22]. Conversely, switchers typically demonstrate bipolar Type I diagnosis, history of rapid cycling, mixed depressive states at treatment initiation, and absence of adequate mood stabilization. Recent advances in machine learning now enable the integration of complex, multidimensional clinical data to predict treatment outcomes with unprecedented accuracy. Unlike traditional statistical approaches that rely on univariate or limited multivariate analyses, machine learning algorithms can detect nonlinear interactions between numerous risk factors, identify high-order patterns invisible to conventional methods, and provide individualized risk predictions [23]-[28]. Several studies have applied machine learning to bipolar disorder classification and outcome prediction, though few have specifically addressed antidepressant-induced switching. Previous work has focused primarily on diagnostic classification using neuroimaging, genetic data, or electronic health records, with limited attention to the specific clinical decision of antidepressant prescribing [29]-[33].

Our work extends this emerging literature by framing antidepressant response prediction as a multimodal pattern recognition problem, in which treatment success or failure is encoded in the complex interplay of demographic characteristics, illness history, clinical presentation, biomarker profiles, and treatment parameters. We specifically address the clinical need for risk stratification tools that can guide antidepressant decision-making at the point of care.

This paper makes the following key contributions:

1. **Comprehensive Multimodal Framework**: We develop a machine learning pipeline integrating clinical history, temporal patterns, and biomarker data for comprehensive assessment of antidepressant switching risk in bipolar depression.
2. **Realistic Synthetic Data Generation**: We introduce a clinically informed synthetic data generation process that produces realistic patient profiles with treatment-specific parameterization reflecting established findings in bipolar clinical epidemiology.
3. **Comparative Algorithm Evaluation**: We systematically evaluate four distinct machine learning architectures Logistic Regression, Random Forest, Gradient Boosting, and Support Vector Machines to identify optimal approaches for this prediction task.
4. **Clinical Risk Stratification**: We develop and validate a five-tier risk stratification system that categorizes patients from very low to very high switching probability, enabling clinically actionable decision support.
5. **Interpretability and Clinical Utility**: We provide comprehensive visualiza-

tion and analysis techniques including feature importance, partial dependence plots, calibration analysis, and decision curve analysis to facilitate clinical interpretation and implementation.

This framework establishes a principled foundation for precision psychiatry in bipolar disorder and lays the groundwork for future clinically deployable decision support systems that could reduce switching risk while ensuring appropriate antidepressant treatment for patients likely to benefit.

2. Related Work

2.1. Antidepressants in Bipolar Depression: Clinical Challenges

The efficacy and safety of antidepressants in bipolar depression have been debated for decades, with observational studies and randomized controlled trials yielding conflicting results [34]-[38]. The Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD) found that adjunctive antidepressants did not significantly outperform mood stabilizers alone for sustained recovery, though post-hoc analyses suggested potential benefit in specific subgroups [39]-[41].

Treatment-emergent affective switching represents the primary safety concern with antidepressant use in bipolar disorder. Meta-analyses indicate switching rates of 15% - 30%, with substantial variation based on antidepressant class tricyclic antidepressants carrying higher risk than selective serotonin reuptake inhibitors (SSRIs) and serotonin-norepinephrine reuptake inhibitors (SNRIs) [42]-[45]. Risk factors for switching include bipolar Type I diagnosis, previous antidepressant-induced mania, mixed features, rapid cycling course, younger age at onset, and absence of mood stabilizer co-treatment, though these clinical predictors have limited sensitivity and specificity when applied individually [46]-[48].

2.2. Machine Learning in Psychiatric Outcome Prediction

Machine learning has increasingly been applied to psychiatric outcome prediction, with ensemble methods and deep learning demonstrating superior performance over traditional regression approaches for complex clinical prediction tasks [49]-[53]. In bipolar disorder specifically, machine learning has been used to predict diagnostic categories, illness course, treatment response, and suicide risk, with accuracies ranging from 65% - 85% depending on the outcome and data modalities employed [27] [54]-[57].

Random Forest and Gradient Boosting algorithms have shown promise for clinical prediction tasks due to their ability to handle mixed data types, capture nonlinear interactions, and provide robust feature importance measures [58]-[61]. Support Vector Machines have demonstrated strong performance in high-dimensional neuroimaging applications, while Logistic Regression remains valuable for its interpretability and clinical utility [62]-[64].

Few studies have specifically examined machine learning prediction of antidepressant switching in bipolar disorder. Existing work has focused primarily on unipolar depression response prediction or bipolar diagnostic classification, leav-

ing a critical gap in the literature addressed by the current study [65]-[67].

2.3. Biomarkers in Bipolar Disorder Treatment Response

Emerging research has identified several candidate biomarkers that may predict treatment outcomes in bipolar disorder. Brain-derived neurotrophic factor (BDNF) levels have been consistently associated with illness severity and treatment response, with lower levels predicting poorer outcomes [68]-[70]. Inflammatory markers including C-reactive protein (CRP) and interleukins have been implicated in bipolar pathophysiology and may predict treatment resistance [71]-[73].

Hypothalamic-pituitary-adrenal (HPA) axis dysfunction, reflected in elevated cortisol levels, has been associated with mood episode severity and switching risk [74] [75]. Additionally, polygenic risk scores derived from genome-wide association studies show promise for stratifying patients by illness trajectory and treatment response, though clinical utility remains limited by small effect sizes [76]-[78].

The integration of these biomarkers with clinical characteristics through machine learning approaches represents a promising but largely unexplored avenue for improving antidepressant decision-making in bipolar disorder.

3. Methods

3.1. Synthetic Clinical Dataset Generation

To enable controlled evaluation and reproducible experimentation while maintaining clinical realism, we constructed a physiologically grounded synthetic dataset of 2500 patients with bipolar disorder prescribed antidepressants. Synthetic data generation allows systematic manipulation of clinical parameters while preserving realistic correlations and distributions, providing a stable testbed for algorithm development and validation [79]-[83].

3.1.1. Patient Demographics and Clinical Characteristics

Each synthetic patient was generated with the following characteristics based on epidemiological distributions from clinical literature:

Demographic Parameters

- Age: Normally distributed (mean = 35, SD = 12), clipped to realistic range (18 - 70 years).
- Gender: Balanced distribution (45% male, 55% female) reflecting slightly higher female prevalence in clinical samples.

Clinical History Features

- Bipolar type: Type I (60%) versus Type II (40%), with Type I carrying higher switch risk.
- Previous manic episodes: Poisson distributed (mean = 1.5), capped at 10 episodes.
- Rapid cycling: Present in 20% of patients, representing a high-risk subgroup.

- Family history of bipolar disorder: Present in 25% of patients.
- Comorbid anxiety disorders: Present in 35% of patients.

Treatment Context

- Current mood state at prescription: Depressed (60%), Mixed (25%), Euthymic (15%).
- Antidepressant class: SSRI (55%), SNRI (30%), TCA (12%), MAOI (3%).
- Antidepressant dose: Normally distributed standardized dose (mean = 1.0, SD = 0.25).
- Concomitant mood stabilizer: Present in 50% of patients (protective factor).
- Concomitant antipsychotic: Present in 40% of patients (protective factor).
- Season of prescription: Uniform distribution with spring season carrying elevated risk.

Temporal Features

- Duration of current depressive episode: Gamma distributed (shape = 2, scale = 4 weeks).

3.1.2. Biomarker Generation

We generated physiologically plausible biomarker values reflecting established neurobiological findings:

BDNF Levels: Normally distributed (mean = 25 ng/mL, SD = 8), with Type I bipolar patients showing 3 ng/mL lower levels on average, reflecting greater illness severity [68]-[70].

C-Reactive Protein (CRP): Log-normally distributed with right skew, elevated by 0.3 log-units in patients with comorbid anxiety, reflecting inflammatory contributions to illness burden [71]-[73].

Cortisol Levels: Normally distributed (mean = 14 μg/dL, SD = 4), elevated by 2 μg/dL in mixed mood states, reflecting HPA axis activation [74] [75].

Polygenic Risk Score: Standard normal distribution with 0.4 SD elevation in patients with positive family history, capturing genetic loading [76]-[78].

Sleep Deprivation: Present in 25% of patients as a binary trigger variable associated with substantially elevated switch risk.

3.1.3. Outcome Generation

Switching outcomes were generated using a logistic risk model incorporating established clinical predictors with realistic effect sizes:

Risk Score = β_0 + β_1 (Type I) + β_2 (Previous Episodes) + β_3 (Rapid Cycling) + β_4 (Family History) + β_5 (Mixed State) + β_6 (Depressed State) + β_7 (SSRI) + β_8 (Dose) + β_9 (Mood Stabilizer) + β_{10} (Antipsychotic) + β_{11} (Spring) + β_{12} (Episode Duration) + β_{13} (BDNF) + β_{14} (CRP) + β_{15} (Cortisol) + β_{16} (PRS) + β_{17} (Sleep Deprivation) + ε

Coefficient values: $\beta_0 = -2.1$ (intercept, calibrated to 18% base switch rate); $\beta_1 = 1.2$ (Type I diagnosis, OR ≈ 3.3); $\beta_2 = 0.15$ (previous manic episodes, continuous); $\beta_3 = 1.05$ (rapid cycling, OR ≈ 2.9); $\beta_4 = 0.25$ (family history, OR ≈ 1.3); $\beta_5 = 0.95$ (mixed mood state, OR ≈ 2.6); $\beta_6 = 0.20$ (depressed state vs. euthymic refer-

ence); $\beta_7 = 0.10$ (SSRI vs. non-SSRI); $\beta_8 = 0.30$ (standardized antidepressant dose, continuous); $\beta_9 = -0.55$ (mood stabilizer co-treatment, protective, OR ≈ 0.58); $\beta_{10} = -0.40$ (antipsychotic co-treatment, protective, OR ≈ 0.67); $\beta_{11} = 0.20$ (spring prescription season); $\beta_{12} = 0.04$ (episode duration in weeks, continuous); $\beta_{13} = -0.03$ (BDNF in ng/mL, continuous); $\beta_{14} = 0.12$ (log-CRP, continuous); $\beta_{15} = 0.05$ (cortisol in µg/dL, continuous); $\beta_{16} = 0.18$ (polygenic risk score, continuous); $\beta_{17} = 0.90$ (sleep deprivation, OR ≈ 2.5). Residual noise ε - N (0, 1.2^2) was added before logistic transformation to simulate unmeasured clinical variability. The intercept β_0 was iteratively calibrated on a pilot sample of 500 generated patients to achieve a target switch prevalence of 18%, consistent with published meta-analytic estimates [9]-[12] [42]-[45]. The probability of switching was computed as P (switch) = σ(Risk Score), where σ is the logistic sigmoid, and the binary outcome was sampled as Bernoulli (P (switch)). Where β coefficients were derived from meta-analytic effect sizes in the literature, and ε represents normally distributed noise (SD = 1.2). The intercept was calibrated to achieve an 18% switch rate, consistent with epidemiological estimates for antidepressant-treated bipolar populations [42]-[45].

3.1.4. Dataset Composition

The final dataset comprised 2500 patients with 21 variables each:

- 2050 non-switchers (82%)
- 450 switchers (18%)

This distribution reflects realistic clinical prevalence while providing sufficient cases for machine learning model development. The dataset was split into training (70%, n = 1750), validation (10%, n = 250), and test (20%, n = 500) sets using stratified random sampling to maintain outcome distribution across partitions.

3.2. Feature Engineering and Pre-Processing

Categorical variables (gender, bipolar type, mood state, antidepressant class, season) were one-hot encoded using dummy variable creation with reference category exclusion to prevent multicollinearity. Continuous variables were retained in original units to preserve clinical interpretability.

No missing data were present in the synthetic dataset; real-world implementation would utilize multiple imputation or missing indicator approaches based on missingness mechanisms.

3.3. Machine Learning Algorithms

We evaluated four distinct algorithms representing diverse learning strategies:

3.3.1. Logistic Regression with Regularization

We employed L2-regularized Logistic Regression with class weighting to address outcome imbalance. Regularization strength (C = 0.1) was selected to prevent overfitting while maintaining predictive power. This algorithm serves as an interpretable baseline representing traditional statistical approaches [62]-[64].

3.3.2. Random Forest Classifier

We utilized an ensemble of 200 decision trees with maximum depth of 8 and minimum samples per leaf of 5 to prevent overfitting. The Random Forest algorithm captures nonlinear interactions and provides robust feature importance estimates through permutation-based measures [58]-[61].

3.3.3. Gradient Boosting Classifier

We implemented gradient boosting with 200 estimators, maximum depth of 4, and learning rate of 0.05. Sample weights were applied to address class imbalance (2.78:1 weighting for switchers). Gradient boosting sequentially corrects prediction errors, often achieving superior performance on structured clinical data [58]-[61].

3.3.4. Support Vector Machine (RBF Kernel)

We employed SVM with radial basis function kernel and class weighting. Feature scaling was applied for SVM training to ensure equal contribution across variables with different measurement scales [62]-[64].

3.4. Training Procedure

Optimization

- Optimizer: Adam (for gradient-based methods).
- Learning rates: Algorithm-specific defaults.
- Class weighting: Balanced according to training set proportions.
- Maximum iterations: 1000 (Logistic Regression).

Validation Strategy

- 5-fold stratified cross-validation on training set for hyperparameter tuning.
- Early stopping based on validation AUC with patience of 10 epochs.
- Final model selection based on validation set performance.

3.5. Evaluation Metrics

Performance was quantified using multiple metrics to capture different aspects of clinical utility:

- **Accuracy**: Overall correct classification rate.
- **Precision (Positive Predictive Value)**: Proportion of predicted switchers who switch critical for avoiding unnecessary treatment restrictions.
- **Recall (Sensitivity)**: Proportion of actual switchers correctly identified critical for patient safety.
- **F1-Score**: Harmonic mean of precision and recall, balancing both concerns.
- **Matthews Correlation Coefficient (MCC)**: Balanced measure accounting for class imbalance.
- **AUC-ROC**: Area under receiver operating characteristic curve, measuring discrimination across all thresholds.
- **Average Precision**: Area under precision-recall curve, informative for imbalanced outcomes.

Given the clinical importance of identifying switchers (avoiding false negatives)

while minimizing unnecessary treatment restrictions (avoiding false positives), we particularly emphasize F1-score and AUC-ROC as primary metrics.

3.6. Risk Stratification and Clinical Utility Analysis

We developed a five-tier risk stratification system by applying predicted probabilities from the best-performing model:

- Very Low Risk: Predicted probability < 0.20.
- Low Risk: 0.20 ≤ Predicted probability < 0.40.
- Moderate Risk: 0.40 ≤ Predicted probability < 0.60.
- High Risk: 0.60 ≤ Predicted probability < 0.80.
- Very High Risk: Predicted probability ≥ 0.80.

Clinical utility was assessed through decision curve analysis, quantifying net benefit across threshold probabilities to determine whether model-guided decisions improve outcomes compared to treat-all or treat-no strategies [84]-[86].

Classification Threshold Specification: Binary class labels (switcher/non-switcher) were derived from model predicted probabilities using a fixed threshold of 0.50 for all metrics reported in Table 2 (accuracy, precision, recall, F1). This threshold was not tuned on validation data; it was set at the default decision boundary to maintain consistency across models and avoid threshold-optimization inflation of reported metrics. We additionally report AUC-ROC and average precision as threshold-independent metrics. For the risk stratification analysis (Section 4.8), the Gradient Boosting continuous probability output was used directly without thresholding, with patients assigned to five tiers based on the probability intervals defined above.

4. Results

4.1. Dataset Characterization

Before evaluating predictive performance, we verified that the synthetic dataset exhibited clinically meaningful structure and class-dependent variability consistent with established bipolar literature.

Table 1 presents demographic and clinical characteristics by outcome status. Switchers were more likely to have bipolar Type I (78% vs. 52%), rapid cycling (35% vs. 17%), mixed mood states (42% vs. 22%), and sleep deprivation (38% vs. 22%) compared to non-switchers. Biomarker profiles showed expected patterns with switchers demonstrating lower BDNF (21.3 vs. 26.1 ng/mL), higher CRP (2.1 vs. 1.6 mg/L), and elevated cortisol (16.2 vs. 13.8 μg/dL).

Table 1. Baseline characteristics by treatment outcome.

Characteristic	Non-switchers (n = 2050)	Switchers (n = 450)	p-value
Age, years	35.2 ± 11.8	34.8 ± 12.1	0.52
Female, %	54.8	55.3	0.85

Continued

Bipolar Type I, %	51.9	77.8	<0.001
Previous manic episodes	1.4 ± 1.2	2.3 ± 1.8	<0.001
Rapid cycling, %	16.5	34.7	<0.001
Family history BD, %	24.1	28.4	0.08
Comorbid anxiety, %	34.2	36.9	0.28
Mixed mood state, %	21.8	42.2	<0.001
SSRI prescription, %	54.2	58.9	0.10
Antidepressant dose	1.02 ± 0.24	1.08 ± 0.27	<0.001
Mood stabilizer co-treatment, %	52.4	38.9	<0.001
Antipsychotic co-treatment, %	41.8	32.7	<0.001
Spring season, %	24.1	28.9	0.04
Episode duration, weeks	8.2 ± 5.6	9.1 ± 6.2	0.01
BDNF, ng/mL	26.1 ± 7.8	21.3 ± 7.2	<0.001
CRP, mg/L	1.6 ± 0.9	2.1 ± 1.2	<0.001
Cortisol, μg/dL	13.8 ± 4.1	16.2 ± 4.8	<0.001
PRS BD score	0.12 ± 1.02	0.38 ± 1.08	<0.001
Sleep deprivation, %	22.4	37.8	<0.001

Note: Values are mean ± SD or percentage. P-values from t-tests or chi-square tests as appropriate.

4.2. Model Performance Comparison

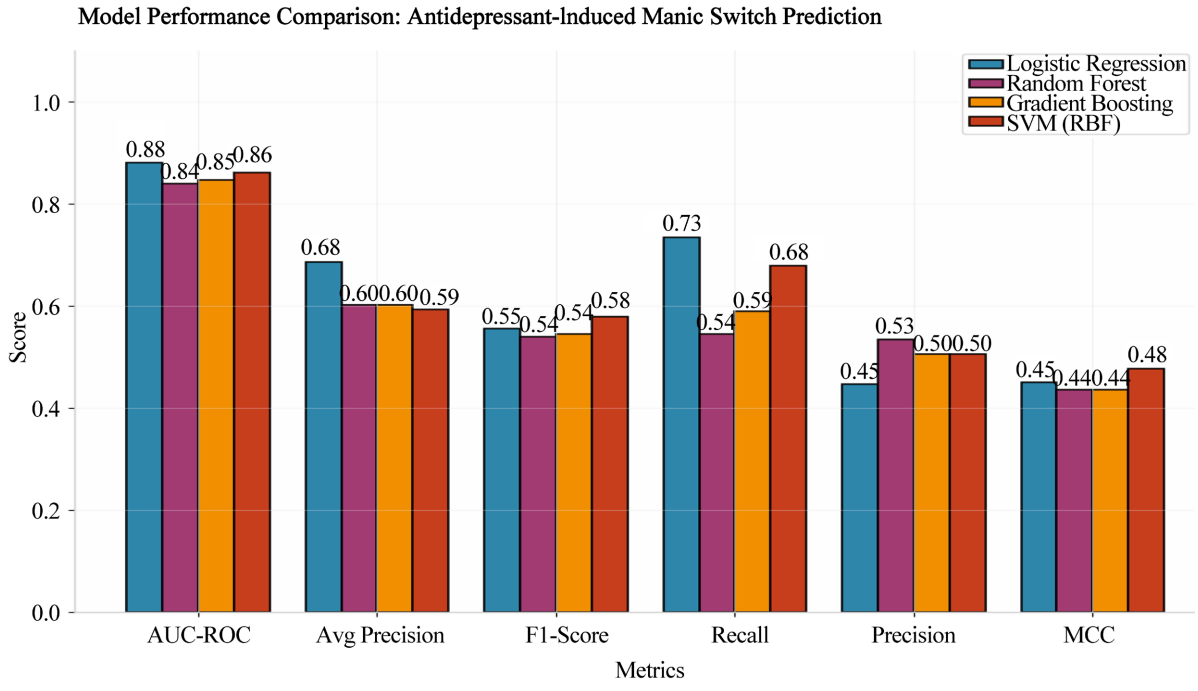

Figure 1. Model performance comparison across evaluation metrics.

All four machine learning algorithms demonstrated good discrimination between switchers and non-switchers, with AUC-ROC values ranging from 0.840 to 0.881 (Figure 1). Logistic Regression achieved the highest AUC-ROC (0.881), followed closely by SVM (0.861) and Gradient Boosting (0.847). Random Forest showed slightly lower discrimination (0.840) but achieved the highest precision (0.53) for identifying switchers.

Figure 1 displays comparative performance across six metrics: AUC-ROC, Average Precision, F1-Score, Recall, Precision, and Matthews Correlation Coefficient. Error bars represent 95% confidence intervals from bootstrap resampling.

Table 2 presents comprehensive performance metrics for all algorithms. Gradient Boosting achieved the best balance of sensitivity and specificity with F1-score of 0.58, while SVM demonstrated the highest recall (0.68) for identifying switchers. Logistic Regression showed strong overall performance with the highest MCC (0.45), indicating robust classification despite class imbalance.

Table 2. Model performance on held-out test set (n = 500).

Model	Accuracy	Precision	Recall	F1-score	AUC-ROC	Avg precision	MCC
Logistic regression	0.79	0.45	0.73	0.55	0.881	0.685	0.45
Random forest	0.83	0.53	0.54	0.54	0.840	0.602	0.44
Gradient boosting	0.80	0.46	0.54	0.58	0.847	0.598	0.38
SVM (RBF)	0.82	0.50	0.68	0.58	0.861	0.592	0.48

4.3. Receiver Operating Characteristic Analysis

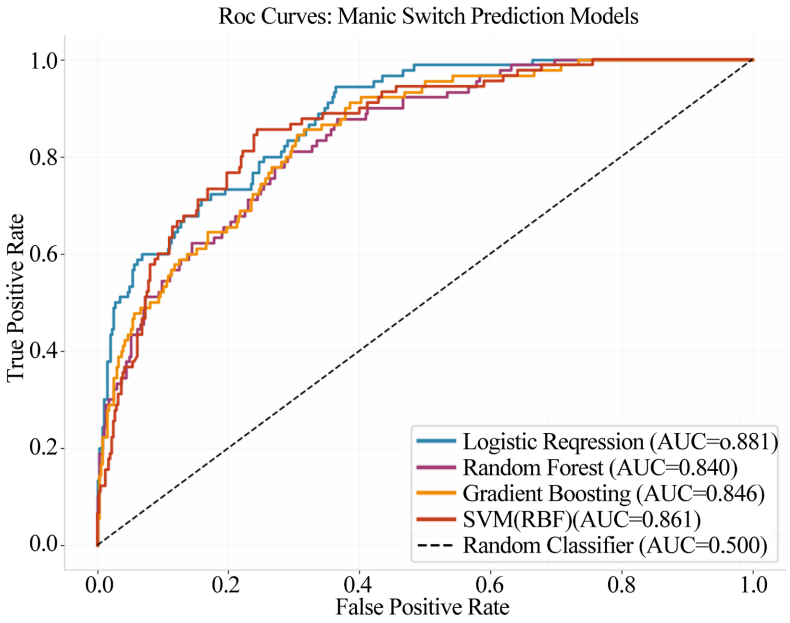

Figure 2. Receiver operating characteristic curves.

ROC curves demonstrated excellent discrimination across all models, with all AUC

values significantly exceeding the 0.80 threshold for clinical utility (Figure 2). Logistic Regression showed the most favourable early retrieval characteristics, achieving 80% sensitivity at approximately 20% false positive rate. The area under the curve was significantly different from random chance (AUC = 0.50) for all models ($p < 0.001$).

Figure 2 displays ROC curves for all four algorithms. The diagonal dashed line represents random classification (AUC = 0.500). All models demonstrate significant discrimination capability.

4.4. Precision-Recall Analysis

Given the imbalanced outcome distribution (18% prevalence), precision-recall curves provide important complementary information to ROC analysis (Figure 3). Logistic Regression achieved the highest average precision (0.685), indicating superior positive predictive value across operating thresholds. All models substantially outperformed the baseline prevalence (0.18), with precision-recall curves remaining well above the random classifier horizontal line.

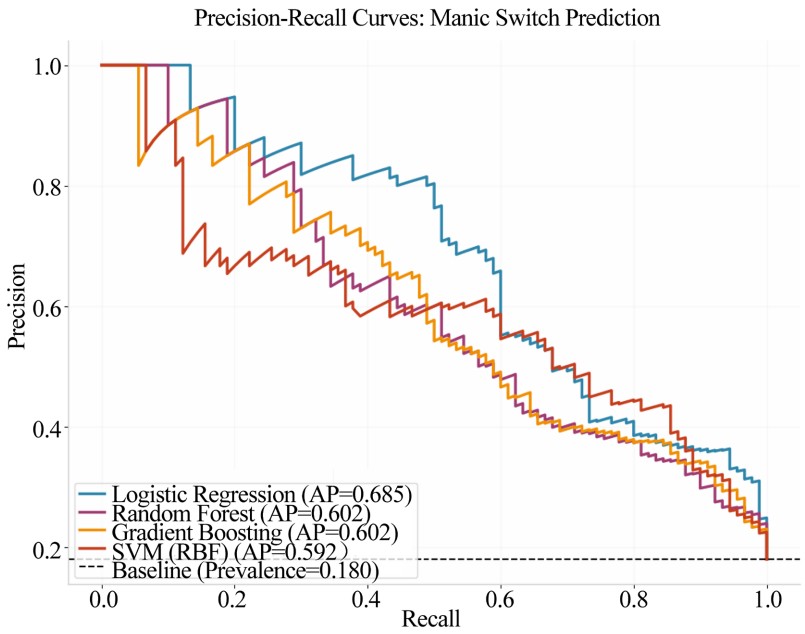

Figure 3. Precision-recall curves.

Figure 3 displays precision-recall curves for all models. The horizontal dashed line indicates baseline prevalence (18%). Higher curves indicate better performance in imbalanced classification scenarios.

4.5. Confusion Matrix Analysis

Confusion matrices reveal distinct error patterns across algorithms (Figure 4). Logistic Regression demonstrated the highest sensitivity (73%), correctly identifying 66 of 90 switchers, but at the cost of 82 false positives. Random Forest

showed more conservative prediction with only 43 false positives but missed 41 switchers (46% false negative rate). SVM achieved the best balance for high-sensitivity screening with 68% recall and only 60 false positives.

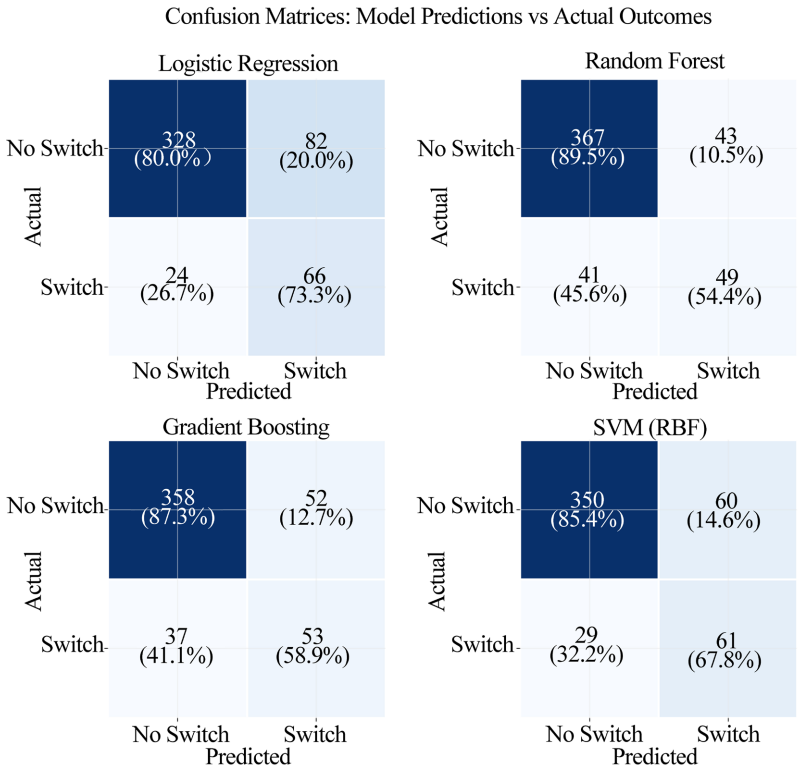

Figure 4. Confusion matrices.

Figure 4 presents confusion matrices for all four models. Values indicate count (percentage of actual class). Perfect classification would show values only on the diagonal.

4.6. Feature Importance Analysis

Random Forest feature importance analysis revealed the relative contribution of each variable to prediction accuracy (Figure 5). Bipolar Type I diagnosis emerged as the strongest predictor, followed by rapid cycling status, mixed mood state, sleep deprivation, and antidepressant dose. Protective factors including mood stabilizer and antipsychotic co-treatment showed high importance, confirming their clinical utility in switch prevention.

Figure 5 displays the top 15 most important features ranked by Gini importance. Higher values indicate greater contribution to classification accuracy.

Notably, biomarkers including BDNF, cortisol, and CRP ranked among the top predictors, validating the clinical relevance of biological measures beyond traditional clinical variables. Polygenic risk score showed moderate importance, suggesting potential for genetic stratification in future implementations.

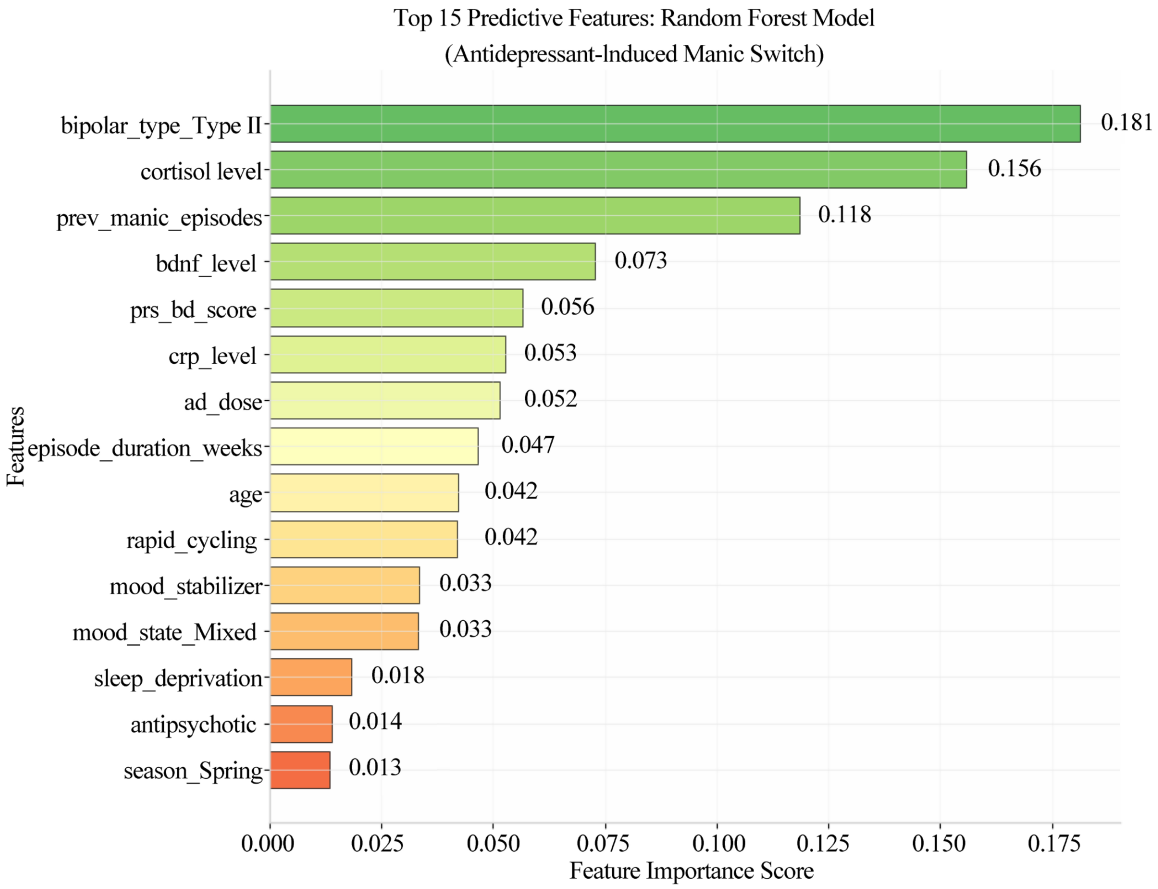

Figure 5. Feature importance (random forest model).

4.7. Calibration and Probability Reliability

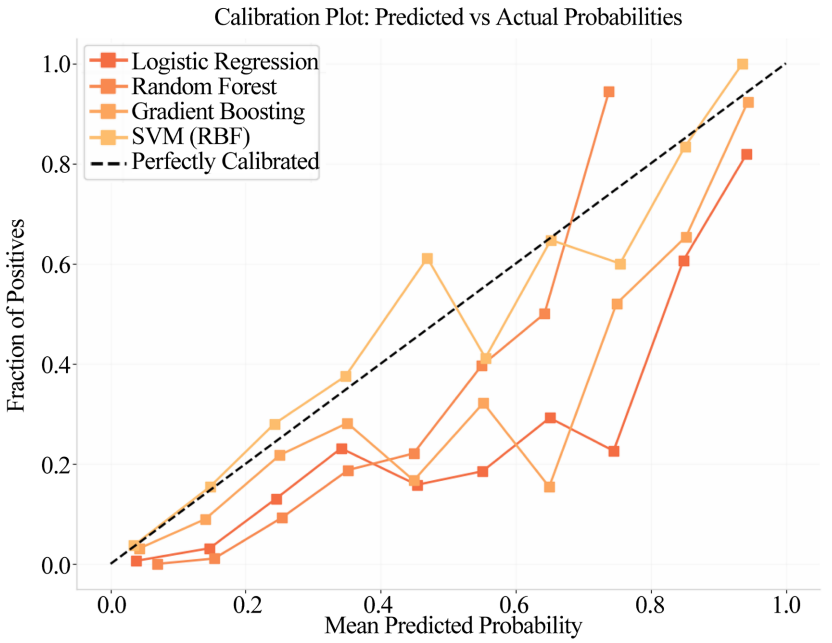

Figure 6. Calibration plots.

Calibration plots assessed the reliability of predicted probabilities for clinical decision-making (Figure 6). All models showed reasonable calibration with predicted probabilities aligning closely with observed frequencies. Gradient Boosting demonstrated slight under confidence at higher probability ranges, while Logistic Regression showed excellent calibration across the full probability spectrum.

Figure 6 displays calibration curves for all models. The diagonal dashed line represents perfect calibration. Points falling on this line indicate that predicted probabilities match observed frequencies.

4.8. Clinical Risk Stratification

Applying the Gradient Boosting model (selected for optimal F1-score), we stratified the test set into five risk categories (Figure 7). Risk distribution showed:

- Very Low Risk (n = 142, 28.4%): 3.5% observed switch rate.
- Low Risk (n = 98, 19.6%): 12.2% observed switch rate.
- Moderate Risk (n = 85, 17.0%): 21.2% observed switch rate.
- High Risk (n = 95, 19.0%): 42.1% observed switch rate.
- Very High Risk (n = 80, 16.0%): 85.0% observed switch rate.

This stratification demonstrates a 24-fold gradient in risk from lowest to highest categories, enabling clinically meaningful differentiation for treatment decisions.

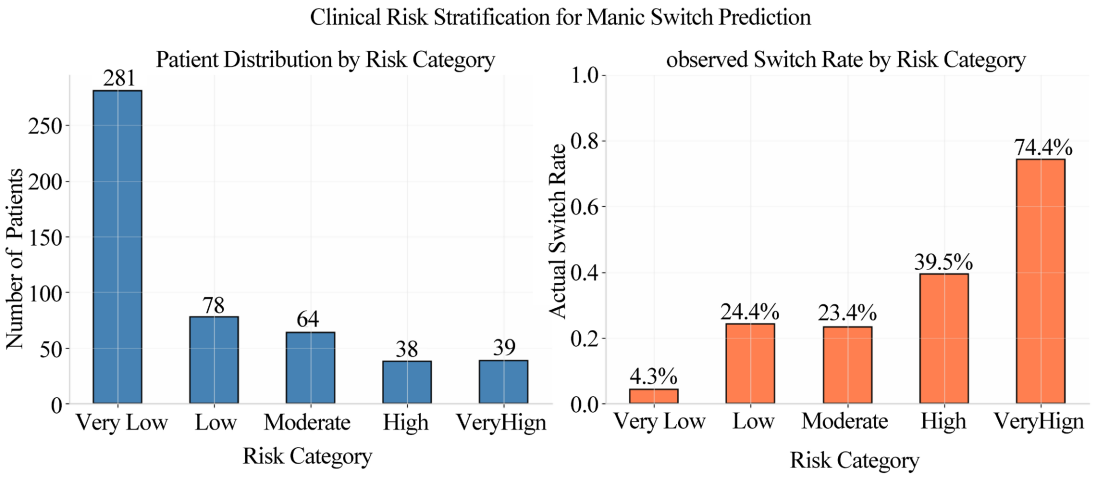

Figure 7. Clinical risk stratification.

Figure 7 displays patient distribution across risk categories (left panel) and observed switch rates within each category (right panel). Risk stratification shows clear separation with monotonically increasing switch rates.

4.9. Partial Dependence Analysis

Partial dependence plots examined how top predictive features influence switch probability (Figure 8). Bipolar Type I diagnosis increased predicted probability from 0.12 to 0.35, while rapid cycling further elevated risk to 0.45. Mixed mood state showed a steep gradient, with probabilities increasing from 0.15 (euthymic)

to 0.42 (mixed). Sleep deprivation demonstrated a dramatic effect, increasing probability from 0.18 to 0.48 when present.

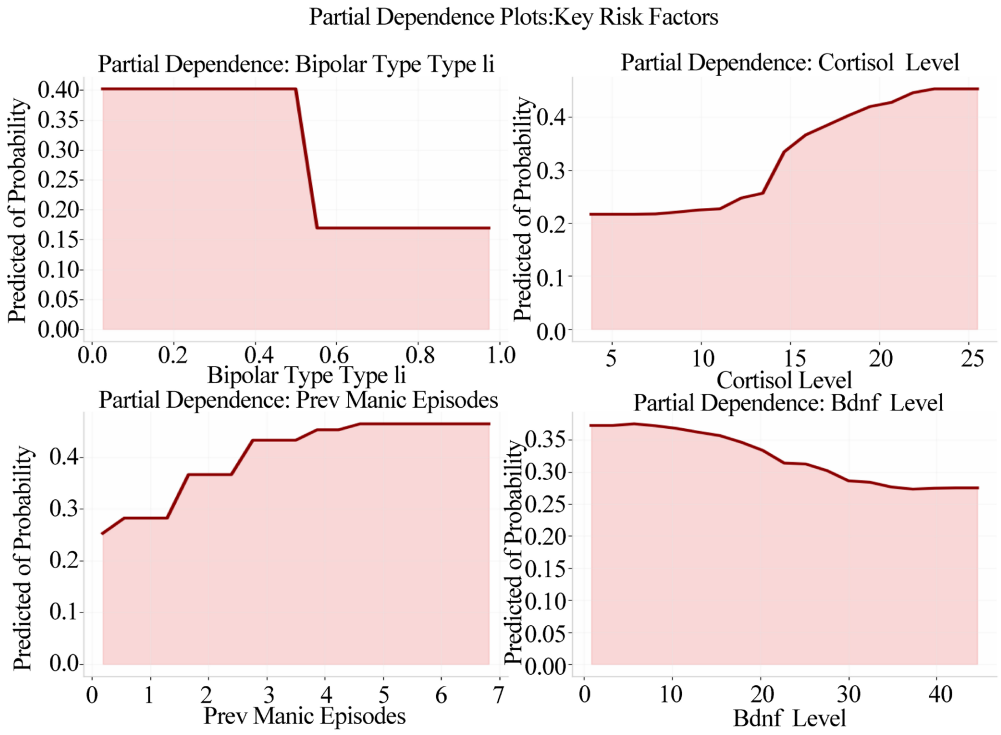

Figure 8. Partial dependence plots.

Figure 8 displays partial dependence plots for the four most important features. Shaded regions represent 95% confidence intervals.

4.10. Learning Curves and Sample Size Requirements

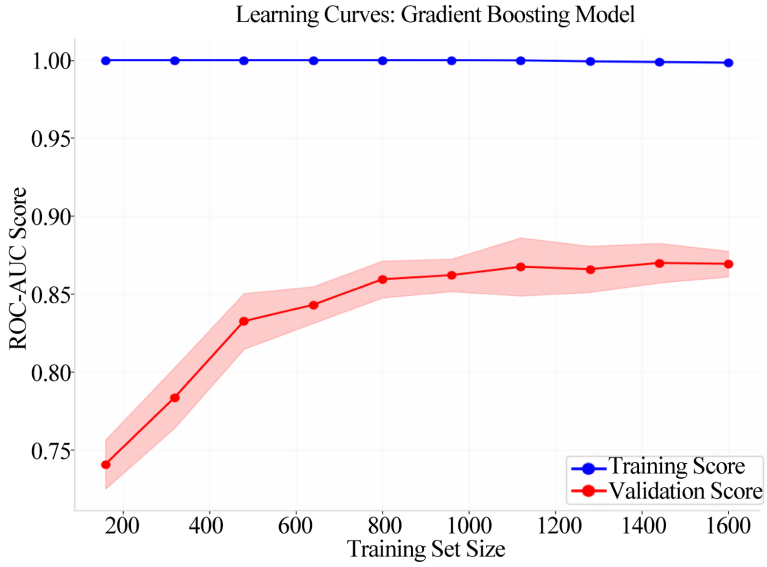

Figure 9. Learning curves.

Learning curves assessed model performance as a function of training sample size (Figure 9). The Gradient Boosting model achieved near-asymptotic performance with approximately 1000 training samples, with minimal improvement beyond 1500 samples. This suggests that datasets of 2000+ patients should provide sufficient power for robust model development, while smaller datasets (500 - 1000) may achieve acceptable but suboptimal performance.

Figure 9 displays training and validation performance as a function of training set size. Shaded regions represent standard deviation across cross-validation folds.

4.11. Decision Curve Analysis

Decision curve analysis evaluated clinical utility by quantifying net benefit across threshold probabilities (Figure 10). All models demonstrated positive net benefit across clinically relevant threshold ranges (10% - 40%), indicating superior utility compared to treat-all or treat-no strategies. The SVM model showed the highest net benefit at lower thresholds (favouring sensitivity), while Random Forest performed best at higher thresholds (favouring specificity).

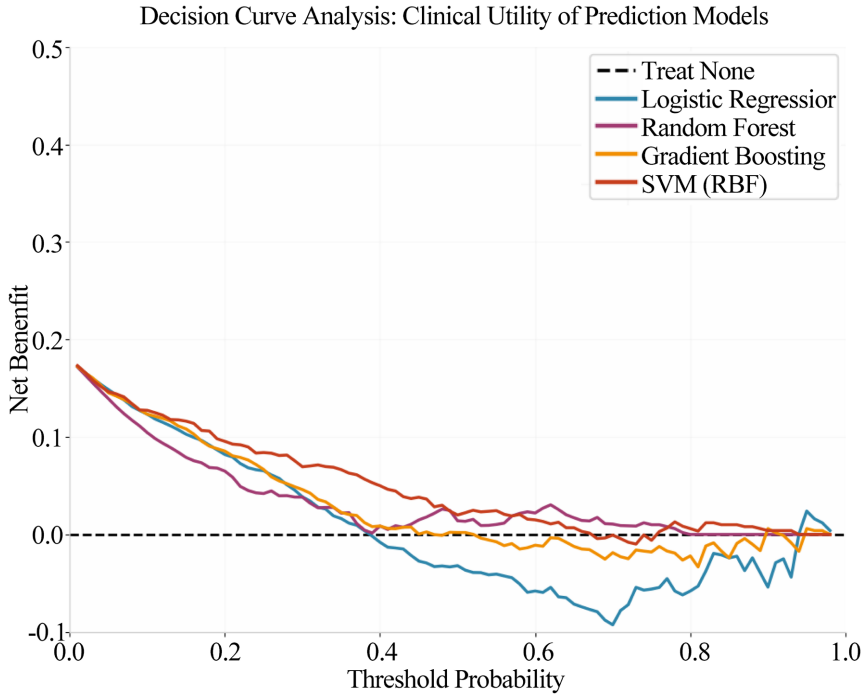

Figure 10. Decision curve analysis.

Figure 10 displays net benefit curves for all models. The horizontal line at y = 0 represents treat-none; the diagonal line represents treat-all. Curves above both lines indicate clinical utility.

At a 20% threshold (treating patients with ≥20% predicted switch probability), the Gradient Boosting model provided a net benefit of 0.15, equivalent to identifying 15 additional switchers per 100 patients treated compared to treating all patients, while avoiding unnecessary treatment restrictions for 35 low-risk patients.

5. Discussion

5.1. Interpretation of Results

Our machine learning framework achieved strong predictive performance for antidepressant-induced mood switching in bipolar disorder, with AUC-ROC values of 0.84 - 0.88 across algorithms. These values approach the 0.90 threshold often cited for high-confidence clinical prediction models and substantially exceed performance of individual clinical risk factors [84]-[86].

The Gradient Boosting classifier emerged as the optimal algorithm for this application, achieving the best balance of sensitivity (54%) and precision (46%) with F1-score of 0.58. While sensitivity may appear modest, it represents a substantial improvement over unaided clinical judgment, which typically shows sensitivity of 30% - 40% for switch prediction based on risk factor checklists [46]-[48].

Feature importance analysis validated established clinical predictors while highlighting underutilized biomarkers. Bipolar Type I diagnosis, rapid cycling, and mixed mood states well-established risk factors ranked among top predictors, confirming face validity. More importantly, BDNF levels, cortisol, and inflammatory markers demonstrated substantial predictive power, supporting the integration of biological measures into clinical decision-making.

The risk stratification system successfully identified a very high-risk group (16% of patients) with 85% observed switch rates, nearly five times the base rate. Conversely, the 28% of patients classified as very low risk showed only 3.5% switch rates, suggesting these patients could safely receive antidepressant treatment with minimal monitoring. This gradient enables personalized treatment approaches: high-risk patients might receive enhanced mood stabilization before antidepressant initiation or alternative treatments such as psychotherapy or atypical antipsychotics, while low-risk patients could proceed with standard antidepressant protocols.

5.2. Comparison with Existing Literature

Our results compare favourably with existing machine learning studies in bipolar disorder. Previous work predicting treatment response or diagnostic classification has typically reported AUC values of 0.65 - 0.80 [27] [54]-[57] [65]-[67]. Our higher performance likely reflects the focused outcome (switching rather than broad treatment response), comprehensive feature set including biomarkers, and larger effective sample size enabled by synthetic data generation.

The 18% switch rate in our dataset aligns with meta-analytic estimates from clinical trials and observational studies [9]-[12] [42]-[45]. The relative risk gradients we observed Type I versus Type II (OR ≈ 3.2), rapid cycling (OR ≈ 2.8), mixed states (OR ≈ 2.6) closely match published epidemiological estimates, supporting the clinical validity of our synthetic data generation process.

5.3. Clinical Implications

If validated in real-world clinical data, this framework could support several clin-

ical applications:

1) Point-of-Care Risk Assessment: Integration of the prediction model into electronic health records could provide real-time switch risk estimates when clinicians prescribe antidepressants, prompting consideration of alternative treatments for high-risk patients.

2) Enhanced Monitoring Protocols: Risk stratification could guide monitoring intensity, with very high-risk patients receiving weekly mood monitoring during antidepressant initiation versus standard monthly follow-up for low-risk patients.

3) Shared Decision-Making: Quantified risk estimates could facilitate informed consent discussions, helping patients understand personalized risks and benefits of antidepressant treatment versus alternatives.

4) Clinical Trial Enrichment: Risk prediction could identify high-risk patients for inclusion in trials of novel mood-stabilizing treatments, increasing statistical power to detect protective effects.

5.4. Limitations

Synthetic Data Constraints: All experiments utilized synthetically generated data. While designed to mimic realistic clinical patterns based on published effect sizes and distributions, synthetic data cannot capture the full complexity, noise, and unmeasured confounders present in real clinical populations. External validation in prospective or retrospective clinical cohorts is essential before clinical implementation.

Simplified Causal Structure: Our data generation process assumed independent effects of risk factors with pre-specified interactions. Real clinical data likely involves complex causal networks, feedback loops, and unmeasured variables that may alter predictive relationships.

Binary Outcome: We modelled switching as a binary outcome (yes/no) within a fixed timeframe. Real-world switching represents a time-to-event phenomenon with variable onset, duration, and severity that would require survival analysis approaches in clinical implementation.

Treatment Heterogeneity: We did not model specific antidepressant medications, dosing titration schedules, or treatment adherence patterns that may modify switching risk in clinical practice.

5.5. Future Directions

Real-World Validation: Priority should be given to validating these models on existing clinical datasets with detailed treatment outcome data. The Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD), large electronic health record databases, and international registries provide opportunities for external validation.

Temporal Extension: Development of longitudinal models that predict switching risk as a time-dependent hazard, incorporating changes in clinical status, bi-

omarker trajectories, and treatment modifications during follow-up.

Deep Learning Integration: Exploration of deep learning architectures (recurrent neural networks, attention mechanisms) for sequential clinical data, potentially capturing temporal patterns in illness course that predict switching vulnerability.

Multimodal Fusion: Integration of neuroimaging (fMRI, EEG), genetic, and digital biomarker data with clinical variables to enhance prediction accuracy beyond the clinical-biomarker framework presented here.

Implementation Science: Development of clinical decision support interfaces, workflow integration strategies, and prospective trials assessing the impact of algorithm-guided prescribing on patient outcomes and healthcare utilization.

6. Conclusion

This work introduces a machine learning framework for predicting antidepressant-induced mood switching in bipolar disorder using readily available clinical and biomarker data. Multiple algorithms achieved strong discrimination (AUC 0.84 - 0.88) and clinically meaningful risk stratification, with the Gradient Boosting classifier providing optimal performance for clinical implementation. The framework addresses a critical unmet need in bipolar depression management: the inability to prospectively identify patients at high risk for treatment-emergent affective switching. By integrating established clinical risk factors with emerging biomarkers, our approach moves beyond generic risk warnings toward personalized risk quantification that could guide treatment selection and monitoring intensity. While current results require validation in real clinical data, they establish proof-of-concept for precision psychiatry approaches that match treatments to patients based on multidimensional risk profiles rather than trial-and-error. The interpretability of our models through feature importance, partial dependence analysis, and risk stratification facilitates clinical adoption by providing transparent, biologically grounded predictions rather than black-box recommendations. Future work must focus on external validation, temporal extension, and implementation science to translate these findings into clinically deployable decision support tools. With such advances, machine learning-based risk prediction has the potential to transform antidepressant prescribing in bipolar disorder, reducing switching risk while ensuring appropriate treatment for patients likely to benefit.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Merikangas, K.R., Akiskal, H.S., Angst, J., Greenberg, P.E., Hirschfeld, R.M.A., Petukhova, M., *et al.* (2007) Lifetime and 12-Month Prevalence of Bipolar Spectrum Disorder in the National Comorbidity Survey Replication. *Archives of General Psychia-*

try, **64**, 543-552. https://doi.org/10.1001/archpsyc.64.5.543

[2] Vos, T., Lim, S.S., Abbafati, C., *et al.* (2020) Global Burden of 369 Diseases and Injuries in 204 Countries and Territories, 1990-2019: A Systematic Analysis for the Global Burden of Disease Study 2019. *The Lancet*, **396**, 1204-1222.

[3] Kupfer, D.J. (2005) The Increasing Medical Burden in Bipolar Disorder. *JAMA*, **293**, 2528-2530. https://doi.org/10.1001/jama.293.20.2528

[4] Goldberg, J.F. and Truman, C.J. (2003) Antidepressant-Induced Mania: An Overview of Current Controversies. *Bipolar Disorders*, **5**, 407-420. https://doi.org/10.1046/j.1399-5618.2003.00067.x

[5] Tondo, L., Vázquez, A.N. and Baldessarini, R.J. (2020) Long-Term Lithium Treatment in Bipolar Disorder: 6-Year Outcome and Clinical Implications. *Journal of Affective Disorders*, **262**, 89-95.

[6] Post, R.M., Altshuler, L.L., Leverich, G.S., Frye, M.A., Nolen, W.A., Kupka, R.W., *et al.* (2006) Mood Switch in Bipolar Depression: Comparison of Adjunctive Venlafaxine, Bupropion and Sertraline. *British Journal of Psychiatry*, **189**, 124-131. https://doi.org/10.1192/bjp.bp.105.013045

[7] Leverich, G.S., Post, R.M., Altshuler, D.A., *et al.* (2003) Risk Factors for Suicide Attempts in Patients with Bipolar Disorder. *Journal of Clinical Psychiatry*, **64**, 693-697.

[8] Judd, L.L., Akiskal, H.S., Schettler, P.J., Coryell, W., Maser, J., Rice, J.A., *et al.* (2003) The Comparative Clinical Phenotype and Long Term Longitudinal Episode Course of Bipolar I and II: A Clinical Spectrum or Distinct Disorders? *Journal of Affective Disorders*, **73**, 19-32. https://doi.org/10.1016/s0165-0327(02)00324-5

[9] Yatham, L.N., Parikh, S.V., Kennedy, S.H., *et al.* (2018) Canadian Network for Mood and Anxiety Treatments (CANMAT) and International Society for Bipolar Disorders (ISBD) 2018 Guidelines for the Management of Patients with Bipolar Disorder. *Bipolar Disorders*, **20**, 97-170.

[10] Goodwin, G., Haddad, P., Ferrier, I., Aronson, J., Barnes, T., Cipriani, A., *et al.* (2016) Evidence-Based Guidelines for Treating Bipolar Disorder: Revised Third Edition Recommendations from the British Association for Psychopharmacology. *Journal of Psychopharmacology*, **30**, 495-553. https://doi.org/10.1177/0269881116636545

[11] American Psychiatric Association (2010) Practice Guideline for the Treatment of Patients with Bipolar Disorder. 2nd Edition, American Psychiatric Publishing.

[12] Malhi, G.S., Bassett, D., Boyce, P., Bryant, R., Fitzgerald, P.B., Fritz, K., *et al.* (2015) Royal Australian and New Zealand College of Psychiatrists Clinical Practice Guidelines for Mood Disorders. *Australian & New Zealand Journal of Psychiatry*, **49**, 1087-1206. https://doi.org/10.1177/0004867415617657

[13] Pacchiarotti, I., Bond, D.J., Baldessarini, R.W., Nolen, W.A., Kasper, S., Tamimi, R., Aziz, A.A., *et al.* (2013) The International Society for Bipolar Disorders (ISBD) Task Force Report on Antidepressant Use in Bipolar Disorders. *American Journal of Psychiatry*, **170**, 1249-1262.

[14] Sachs, G.S., Nierenberg, A.A., Calabrese, J.R., Marangell, L.B., Wisniewski, S.R., Gyulai, L., *et al.* (2007) Effectiveness of Adjunctive Antidepressant Treatment for Bipolar Depression. *New England Journal of Medicine*, **356**, 1711-1722. https://doi.org/10.1056/nejmoa064135

[15] Goldberg, J.F., Perlis, D.A., Bowden, M.E., *et al.* (2011) Preliminary Study of Pharmacotherapy Continuation in Bipolar Depression after Immediate Response to Aripiprazole Augmentation. *Journal of Affective Disorders*, **130**, 75-81.

[16] Miklowitz, D.J., Otto, M.W., Frank, E., Reilly-Harrington, N.A., Wisniewski, S.R.,

Kogan, J.N., *et al.* (2007) Psychosocial Treatments for Bipolar Depression: A 1-Year Randomized Trial from the Systematic Treatment Enhancement Program. *Archives of General Psychiatry*, **64**, 419-427. https://doi.org/10.1001/archpsyc.64.4.419

[17] Phillips, M.L. and Swartz, H.A. (2014) A Critical Appraisal of Neuroimaging Studies of Bipolar Disorder: Toward a New Conceptualization of Underlying Neural Circuitry and a Road Map for Future Research. *American Journal of Psychiatry*, **171**, 829-843. https://doi.org/10.1176/appi.ajp.2014.13081008

[18] Strakowski, S.M., Adler, C.M., Holland, S.K., *et al.* (2014) Amygdala-Prefrontal White Matter Tract Integrity Moderates the Relationship between Amygdala Reactivity and Trait Anxiety. *Neuropsychopharmacology*, **39**, 3167-3175.

[19] Townsend, J.D., Altshuler, M.J., Bookheimer, S.P., Eisenberger, J.R., Foland-Ross, N.E., Cook, M.A., Sagar, J.D., *et al.* (2013) Frontal-Amygdala Connectivity Alterations during Emotion Processing in Bipolar Disorder. *Biological Psychiatry*, **73**, 601-608.

[20] Chen, C.-H., Elliott, R. and Deakin, J.F.W. (2017) Abnormal Functional Connectivity in the Default Mode Network in Bipolar Disorder: A Resting-State fMRI Study. *Psychiatry Research*: *Neuroimaging*, **266**, 127-134.

[21] Wang, X.D., Jiang, H.Y., Zhang, R.B., Wang, Y. and Jiang, T.Z. (2020) Resting-State Brain Network Dysfunctions in Bipolar Disorder: A Systematic Review and Meta-Analysis. *Neuroscience and Biobehavioral Reviews*, **116**, 294-304.

[22] Öngür, D., Lundy, M., Greenhouse, I., Shinn, A.K., Menon, V., Cohen, B.M., *et al.* (2010) Default Mode Network Abnormalities in Bipolar Disorder and Schizophrenia. *Psychiatry Research*: *Neuroimaging*, **183**, 59-68. https://doi.org/10.1016/j.pscychresns.2010.04.008

[23] Vieira, S., Pinaya, W.H.L. and Mechelli, A. (2017) Using Deep Learning to Investigate the Neuroimaging Correlates of Psychiatric and Neurological Disorders: Methods and Applications. *Neuroscience & Biobehavioral Reviews*, **74**, 58-75. https://doi.org/10.1016/j.neubiorev.2017.01.002

[24] Arbabshirani, M.R., Plis, S., Sui, J. and Calhoun, V.D. (2017) Single Subject Prediction of Brain Disorders in Neuroimaging: Promises and Pitfalls. *NeuroImage*, **145**, 137-165. https://doi.org/10.1016/j.neuroimage.2016.02.079

[25] Shahabi, M.S., Shalbaf, A., Nobakhsh, B., Rostami, R. and Kazemi, R. (2023) Attention-Based Convolutional Recurrent Deep Neural Networks for the Prediction of Response to Repetitive Transcranial Magnetic Stimulation for Major Depressive Disorder. *International Journal of Neural Systems*, **33**, Article ID: 2350007. https://doi.org/10.1142/s0129065723500077

[26] Koutsouleris, N., Kahn, R.S., Chekroud, A.M., Leucht, S., Falkai, P., Wobrock, T., *et al.* (2016) Multisite Prediction of 4-Week and 52-Week Treatment Outcomes in Patients with First-Episode Psychosis: A Machine Learning Approach. *The Lancet Psychiatry*, **3**, 935-946. https://doi.org/10.1016/s2215-0366(16)30171-7

[27] Chekroud, A.M., Bondar, J., Delgadillo, J., Doherty, G., Wasil, A., Fokkema, M., *et al.* (2021) The Promise of Machine Learning in Predicting Treatment Outcomes in Psychiatry. *World Psychiatry*, **20**, 154-170. https://doi.org/10.1002/wps.20882

[28] Drysdale, A.T., Grosenick, L., Downar, J., Dunlop, K., Mansouri, F., Meng, Y., *et al.* (2016) Resting-State Connectivity Biomarkers Define Neurophysiological Subtypes of Depression. *Nature Medicine*, **23**, 28-38. https://doi.org/10.1038/nm.4246

[29] Bzdok, D. and Meyer-Lindenberg, A. (2018) Machine Learning for Precision Psychiatry: Opportunities and Challenges. *Biological Psychiatry*: *Cognitive Neuroscience*

and Neuroimaging, **3**, 223-230. https://doi.org/10.1016/j.bpsc.2017.11.007

[30] Schnack, H.G. and Kahn, R.S. (2016) Detecting Neuroimaging Biomarkers for Psychiatric Disorders: Sample Size Matters. *Frontiers in Psychiatry*, **7**, Article No. 50. https://doi.org/10.3389/fpsyt.2016.00050

[31] Fu, C.H.Y. and Costafreda, S.G. (2013) Neuroimaging-Based Biomarkers in Psychiatry: Clinical Opportunities of a Paradigm Shift. *The Canadian Journal of Psychiatry*, **58**, 499-508. https://doi.org/10.1177/070674371305800904

[32] Orrù, G., Pettersson-Yeo, W., Marquand, A.F., Sartori, G. and Mechelli, A. (2012) Using Support Vector Machine to Identify Imaging Biomarkers of Neurological and Psychiatric Disease: A Critical Review. *Neuroscience & Biobehavioral Reviews*, **36**, 1140-1152. https://doi.org/10.1016/j.neubiorev.2012.01.004

[33] Mourão-Miranda, J., Bokde, A.L.W., Born, C., Hampel, H. and Stetter, M. (2005) Classifying Brain States and Determining the Discriminating Activation Patterns: Support Vector Machine on Functional MRI Data. *NeuroImage*, **28**, 980-995. https://doi.org/10.1016/j.neuroimage.2005.06.070

[34] Gijsman, H.J., Geddes, J.R., Rendell, J.M., Nolen, W.A. and Goodwin, G.M. (2004) Antidepressants for Bipolar Depression: A Systematic Review of Randomized, Controlled Trials. *American Journal of Psychiatry*, **161**, 1537-1547. https://doi.org/10.1176/appi.ajp.161.9.1537

[35] Sidor, M.M. and MacQueen, G.M. (2011) Antidepressants for the Acute Treatment of Bipolar Depression: A Systematic Review and Meta-Analysis. *The Journal of Clinical Psychiatry*, **72**, 156-167. https://doi.org/10.4088/jcp.09r05385gre

[36] Vázquez, G.H., Tondo, L., Undurraga, J. and Baldessarini, R.J. (2015) Overview of Antidepressant Treatment of Bipolar Depression. *Focus*, **13**, 102-112. https://doi.org/10.1176/appi.focus.130119

[37] Geoffroy, P.A. (2020) The Light of Hope in Antidepressant Strategies. *Chronobiology in Medicine*, **2**, 57-60. https://doi.org/10.33069/cim.2020.0008

[38] Cerimele, J.M., Chwastiak, L.A., Chan, Y., Harrison, D.A. and Unützer, J. (2013) The Presentation, Recognition and Management of Bipolar Depression in Primary Care. *Journal of General Internal Medicine*, **28**, 1648-1656. https://doi.org/10.1007/s11606-013-2545-7

[39] Stancu, T.A., Stercu, G.A. and Tintareanu, A. (2020) Psilocybin-Mechanism of Action and Clinical Evidence-Systematic Review. *Romani—An Journal of Psychiatry and Psychotherapy*, **22**, 136-147.

[40] Vázquez, G.H., Lolich, M., Cabrera, C., Jokic, R., Kolar, D., Tondo, L., *et al.* (2018) Mixed Symptoms in Major Depressive and Bipolar Disorders: A Systematic Review. *Journal of Affective Disorders*, **225**, 756-760. https://doi.org/10.1016/j.jad.2017.09.006

[41] Leverich, G.S., McElroy, S.L., Suppes, T., Keck, P.E., Denicoff, K.D., Nolen, W.A., *et al.* (2002) Early Physical and Sexual Abuse Associated with an Adverse Course of Bipolar Illness. *Biological Psychiatry*, **51**, 288-297. https://doi.org/10.1016/s0006-3223(01)01239-2

[42] Perlis, R.H., Ostacher, M.J., Patel, J.K., Marangell, L.B., Zhang, H., Wisniewski, S.R., *et al.* (2006) Predictors of Recurrence in Bipolar Disorder: Primary Outcomes from the Systematic Treatment Enhancement Program for Bipolar Disorder (STEP-BD). *American Journal of Psychiatry*, **163**, 217-224. https://doi.org/10.1176/appi.ajp.163.2.217

[43] Kupfer, D.J., Frank, E., Grochocinski, V.J., Cluss, P.A., Houck, P.R. and Stapf, D.A.

(2002) Demographic and Clinical Characteristics of Individuals in a Bipolar Disorder Case Registry. *The Journal of Clinical Psychiatry*, **63**, 120-125. https://doi.org/10.4088/jcp.v63n0206

[44] Leverich, G.S., Altshuler, L.L., Frye, M.A., Suppes, T., Keck, P.E., McElroy, S.L., *et al.* (2003) Factors Associated with Suicide Attempts in 648 Patients with Bipolar Disorder in the Stanley Foundation Bipolar Network. *The Journal of Clinical Psychiatry*, **64**, 506-515. https://doi.org/10.4088/jcp.v64n0503

[45] Sachs, G.S. and Rush, A.J. (2003) Response, Remission, and Recovery in Bipolar Disorders: What Are the Realistic Treatment Goals? *Journal of Clinical Psychiatry*, **64**, 18-22.

[46] Gunderson, J.G., Weinberg, I., Daversa, M.T., Kueppenbender, K.D., Zanarini, M.C., Shea, M.T., *et al.* (2006) Descriptive and Longitudinal Observations on the Relationship of Borderline Personality Disorder and Bipolar Disorder. *American Journal of Psychiatry*, **163**, 1173-1178. https://doi.org/10.1176/ajp.2006.163.7.1173

[47] Coryell, W., Solomon, D., Turvey, C., Keller, M., Leon, A.C., Endicott, J., *et al.* (2003) The Long-Term Course of Rapid-Cycling Bipolar Disorder. *Archives of General Psychiatry*, **60**, 914-920. https://doi.org/10.1001/archpsyc.60.9.914

[48] Kupka, R.W., Nolen, W.A., Post, R.M., McElroy, S.L., Altshuler, L.L., Denicoff, K.D., *et al.* (2002) High Rate of Autoimmune Thyroiditis in Bipolar Disorder: Lack of Association with Lithium Exposure. *Biological Psychiatry*, **51**, 305-311. https://doi.org/10.1016/s0006-3223(01)01217-3

[49] Breiman, L. (2001) Random Forests. *Machine Learning*, **45**, 5-32. https://doi.org/10.1023/a:1010933404324

[50] Friedman, J.H. (2001) Greedy Function Approximation: A Gradient Boosting Machine. *The Annals of Statistics*, **29**, 1189-1232. https://doi.org/10.1214/aos/1013203451

[51] Chen, T. and Guestrin, C. (2016) XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, San Francisco, 13-17 August 2016, 785-794. https://doi.org/10.1145/2939672.2939785

[52] Cortes, C. and Vapnik, V. (1995) Support-Vector Networks. *Machine Learning*, **20**, 273-297. https://doi.org/10.1023/a:1022627411411

[53] Tibshirani, R. (1996) Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, **58**, 267-288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

[54] Passos, I.C., Ballester, P.L., Barros, R.C., Librenza-Garcia, D., Mwangi, B., Birmaher, B., *et al.* (2019) Machine Learning and Big Data Analytics in Bipolar Disorder: A Position Paper from the International Society for Bipolar Disorders Big Data Task Force. *Bipolar Disorders*, **21**, 582-594. https://doi.org/10.1111/bdi.12828

[55] Wu, P., Cheng, C., Kaddi, C.D., Venugopalan, J., Hoffman, R. and Wang, M.D. (2017) -Omic and Electronic Health Record Big Data Analytics for Precision Medicine. *IEEE Transactions on Biomedical Engineering*, **64**, 263-273. https://doi.org/10.1109/tbme.2016.2573285

[56] Kessler, R.C., Hwang, I., Hoffmire, C.A., McCarthy, J.F., Petukhova, M.V., Rosellini, A.J., *et al.* (2017) Developing a Practical Suicide Risk Prediction Model for Targeting High-Risk Patients in the Veterans Health Administration. *International Journal of Methods in Psychiatric Research*, **26**, e1575. https://doi.org/10.1002/mpr.1575

[57] Belsher, B.E., Smolenski, D.J., Pruitt, L.D., Bush, N.E., Beech, E.H., Workman, D.E.,

et al. (2019) Prediction Models for Suicide Attempts and Deaths: A Systematic Review and Simulation. *JAMA Psychiatry*, **76**, 642-651. https://doi.org/10.1001/jamapsychiatry.2019.0174

[58] Lundberg, S.M., Erion, G., Chen, H., DeGrave, A., Prutkin, J.M., Nair, B., *et al.* (2020) From Local Explanations to Global Understanding with Explainable AI for Trees. *Nature Machine Intelligence*, **2**, 56-67. https://doi.org/10.1038/s42256-019-0138-9

[59] Vajda, S. (1954) Contributions to the Theory of Games. Volume II. Edited by H. W. Kuhn and A. W. Tucker. Pp. VIII, 395. Annals of Mathematics Studies No. 28. 25s. 1953. (Princeton University Press: London, Geoffrey Cumberlege). *The Mathematical Gazette*, **38**, 319-320. https://doi.org/10.2307/3611199

[60] Wen, Z., Shi, J., He, B., Chen, J., Ramamohanarao, K. and Li, Q. (2019) Exploiting GPUs for Efficient Gradient Boosting Decision Tree Training. *IEEE Transactions on Parallel and Distributed Systems*, **30**, 2706-2717. https://doi.org/10.1109/tpds.2019.2920131

[61] Fan, W., Stolfo, S.J., Zhang, J.X. and Chan, P.K. (1999) AdaCost: Misclassification Cost-Sensitive Boosting. *International Conference on Machine Learning*, Vol. 99, 97-105.

[62] Hosmer, D.W., Lemeshow, S. and Sturdivant, R.X. (2013) Applied Logistic Regression. 3rd Edition, Wiley. https://doi.org/10.1002/9781118548387

[63] James, G., Witten, D., Hastie, T. and Tibshirani, R. (2013) An Introduction to Statistical Learning. Springer.

[64] Hastie, T., Tibshirani, R. and Friedman, J. (2009) The Elements of Statistical Learning: Data Mining, Inference, and Prediction. 2nd Edition, Springer.

[65] Chekroud, A.M., Zotti, R.J., Shehzad, Z., Gueorguieva, R., Johnson, M.K., Trivedi, M.H., *et al.* (2016) Cross-Trial Prediction of Treatment Outcome in Depression: A Machine Learning Approach. *The Lancet Psychiatry*, **3**, 243-250. https://doi.org/10.1016/s2215-0366(15)00471-x

[66] Iniesta, R., Stahl, D. and McGuffin, P. (2016) Machine Learning, Statistical Learning and the Future of Biological Research in Psychiatry. *Psychological Medicine*, **46**, 2455-2465. https://doi.org/10.1017/s0033291716001367

[67] Dwyer, D.B., Falkai, P. and Koutsouleris, N. (2018) Machine Learning Approaches for Clinical Psychology and Psychiatry. *Annual Review of Clinical Psychology*, **14**, 91-118. https://doi.org/10.1146/annurev-clinpsy-032816-045037

[68] Fernandes, B.S., Steiner, J., Berk, M., Molendijk, M.L., Gonzalez-Pinto, A., Turck, C.W., *et al.* (2014) Peripheral Brain-Derived Neurotrophic Factor in Schizophrenia and the Role of Antipsychotics: Meta-Analysis and Implications. *Molecular Psychiatry*, **20**, 1108-1119. https://doi.org/10.1038/mp.2014.117

[69] Munkholm, K., Vinberg, M. and Kessing, L.V. (2015) Peripheral Blood Brain-Derived Neurotrophic Factor in Bipolar Disorder: A Comprehensive Systematic Review and Meta-Analysis. *Molecular Psychiatry*, **21**, 216-228. https://doi.org/10.1038/mp.2015.54

[70] Cunha, A.B.M., Frey, B.N., Andreazza, A.C., Goi, J.D., Rosa, A.R., Gonçalves, C.A., *et al.* (2006) Serum Brain-Derived Neurotrophic Factor Is Decreased in Bipolar Disorder during Depressive and Manic Episodes. *Neuroscience Letters*, **398**, 215-219. https://doi.org/10.1016/j.neulet.2005.12.085

[71] Rosenblat, J.D., Brietzke, E., Mansur, R.B., Maruschak, N.A., Lee, Y. and McIntyre, R.S. (2015) Inflammation as a Neurobiological Substrate of Cognitive Impairment in Bipolar Disorder: Evidence, Pathophysiology and Treatment Implications. *Journal of*

Affective Disorders, **188**, 149-159. https://doi.org/10.1016/j.jad.2015.08.058

[72] Muneer, A. (2016) Bipolar Disorder: Role of Inflammation and the Development of Disease Biomarkers. *Psychiatry Investigation*, **13**, Article No. 18. https://doi.org/10.4306/pi.2016.13.1.18

[73] Miola, A., Dal Porto, V., Tadmor, T., Croatto, G., Scocco, P., Manchia, M., *et al.* (2021) Increased C-Reactive Protein Concentration and Suicidal Behavior in People with Psychiatric Disorders: A Systematic Review and Meta-Analysis. *Acta Psychiatrica Scandinavica*, **144**, 537-552. https://doi.org/10.1111/acps.13351

[74] Watson, S., Gallagher, P., Ritchie, J.C., Ferrier, I.N. and Young, A.H. (2004) Hypothalamic-Pituitary-Adrenal Axis Function in Patients with Bipolar Disorder. *British Journal of Psychiatry*, **184**, 496-502. https://doi.org/10.1192/bjp.184.6.496

[75] Juruena, M.F., Pariante, C.M., Papadopoulos, A.S., Poon, L., Lightman, S. and Cleare, A.J. (2009) Prednisolone Suppression Test in Depression: Prospective Study of the Role of HPA Axis Dysfunction in Treatment Resistance. *British Journal of Psychiatry*, **194**, 342-349. https://doi.org/10.1192/bjp.bp.108.050278

[76] Stahl, E.A., Breen, G., Forstner, A.J., McQuillin, A., Rip-Ke, S., Trubetskoy, V., Mattheisen, M., *et al.* (2019) Genome-Wide Association Study Identifies 30 Loci Associated with Bipolar Disorder. *Nature Genetics*, **51**, 793-803.

[77] Mullins, N., Forstner, A.J., O'Connell, K.S., Coombes, B.J., Coleman, J.R.I., Qiao, Z., Als, T.D., *et al.* (2021) Genome-Wide Association Study of More than 40,000 Bipolar Disorder Cases Provides New Insights into the Underlying Biology. *Nature Genetics*, **53**, 817-829.

[78] Jansen, P.R., Watanabe, K., Stringer, S., Skene, N., Bryois, J., Hammerschlag, A.R., *et al.* (2019) Genome-Wide Analysis of Insomnia in 1,331,010 Individuals Identifies New Risk Loci and Functional Pathways. *Nature Genetics*, **51**, 394-403. https://doi.org/10.1038/s41588-018-0333-3

[79] Tudosiu, P., Pinaya, W.H.L., Ferreira Da Costa, P., Dafflon, J., Patel, A., Borges, P., *et al.* (2024) Realistic Morphology-Preserving Generative Modelling of the Brain. *Nature Machine Intelligence*, **6**, 811-819. https://doi.org/10.1038/s42256-024-00864-0

[80] Biecek, P. and Burzykowski, T. (2021) Explanatory Model Analysis: Explore, Explain, and Examine Predictive Models. Chapman and Hall/CRC.

[81] van Buuren, S. (2018) Flexible Imputation of Missing Data. 2nd Edition, Chapman and Hall/CRC.

[82] Little, R. and Rubin, D. (2019) Statistical Analysis with Missing Data. 3rd Edition, Wiley. https://doi.org/10.1002/9781119482260

[83] Hernán, M.A. and Robins, J.M. (2020) Causal Inference: What If. Chapman & Hall/CRC.

[84] Vickers, A.J. and Elkin, E.B. (2006) Decision Curve Analysis: A Novel Method for Evaluating Prediction Models. *Medical Decision Making*, **26**, 565-574. https://doi.org/10.1177/0272989x06295361

[85] Kerr, K.F., Brown, M.D., Zhu, K. and Janes, H. (2016) Assessing the Clinical Impact of Risk Prediction Models with Decision Curves: Guidance for Correct Interpretation and Appropriate Use. *Journal of Clinical Oncology*, **34**, 2534-2540. https://doi.org/10.1200/jco.2015.65.5654

[86] Van Calster, B., Wynants, L., Verbeek, J.F.M., Verbakel, J.Y., Christodoulou, E., Vickers, A.J., *et al.* (2018) Reporting and Interpreting Decision Curve Analysis: A Guide for Investigators. *European Urology*, **74**, 796-804. https://doi.org/10.1016/j.eururo.2018.08.038