Open Access Library Journal
2026, Volume 13, e15673
ISSN Online: 2333-9721
ISSN Print: 2333-9705

Simulation-Based Bayesian Modelling of Antidepressant Discontinuation Risk in Bipolar Disorder: Integrating Side-Effect Profiles, Therapeutic Alliance Scores, and Synthetic Prescription Records

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) Simulation-Based Bayesian Modelling of Antidepressant Discontinuation Risk in Bipolar Disorder: Integrating Side-Effect Profiles, Therapeutic Alliance Scores, and Synthetic Prescription Records. *Open Access Library Journal*, **13**: e15673.
https://doi.org/10.4236/oalib.1115673

Received: June 22, 2026
Accepted: August 23, 2026
Published: August 26, 2026

Abstract

Antidepressant discontinuation in bipolar disorder is a clinically important contributor to treatment failure. Side-effect burden, therapeutic alliance, and prescription-adherence behaviour may jointly influence unplanned cessation. This proof-of-concept study evaluates whether those domains can support discontinuation-risk modelling in a fully synthetic cohort; it does not analyse real patient records or provide clinical validation. We generated a synthetic cohort of N = 900 antidepressant-exposed patients with bipolar disorder and a simulated 90-day discontinuation prevalence of 35.0%. Five feature domains represented side effects, therapeutic alliance, synthetic prescription records, patient-reported outcomes, and clinical variables. Random Forest, XGBoost, and LightGBM were combined through five-fold out-of-fold stacking with a logistic-regression meta-learner. A MAP-regularised Bayesian logistic regression served as the principal probabilistic reference. SHAP analyses and domain-specific visualisations were used for interpretation. BIC, WAIC, and Bayes-factor calculations for non-likelihood models were retained only as exploratory approximations. Bayesian logistic regression achieved the highest AUC (0.939) and lowest Brier score (0.095). The stacked ensemble achieved AUC = 0.913 (95% CI: 0.860 - 0.956), F1 = 0.816 (CI: 0.722 - 0.891), sensitivity = 0.817, specificity = 0.896, and Brier score = 0.109. Therefore, the ensemble was not superior in discrimination or calibration; its value lies in modelling non-linear interactions and supporting complementary interpretation. Within the simulator, prior gap count, refill adherence, patient-clinician concordance, medication-harm beliefs, and prior discontinuation history were the dominant SHAP predictors. These results demon-

strate methodological feasibility in synthetic data only. The feature rankings and apparent clinical patterns may partly reflect the data-generation rules and require external validation using real dispensing, clinical, and therapeutic alliance data before any decision-support use.

Subject Areas

Psychiatry & Psychology

Keywords

Antidepressant Discontinuation, Bipolar Disorder, Therapeutic Alliance, Side-Effect Profile, Prescription Adherence, Bayesian Model, SHAP, Working Alliance Inventory, Real-World Data, Clinical Decision Support

1. Introduction

The pharmacological treatment of bipolar disorder is undermined by a paradox: the medications prescribed to stabilise mood are frequently discontinued by the very patients they are intended to help, not through clinical recommendation but through unilateral patient decision, the "quiet quitters" who stop taking their antidepressant without informing their prescribing clinician [1]. This phenomenon is both ubiquitous and systematically undercounted. Real-world pharmacy dispensing data consistently show antidepressant medication possession ratios well below those reported in clinical trial populations, with 30% - 40% of BD patients discontinuing within three months and 60% - 70% within twelve months [2]. The clinical consequences are substantial: discontinuation is a leading cause of depressive relapse, pharmacological escalation, and the accumulating treatment burden that characterises chronic bipolar depression.

Three distinct but interacting driver domains have been identified in qualitative research: side-effect intolerance (the most cited reason in patient surveys), poor therapeutic alliance (the quality of the working relationship between patient and clinician), and structural prescription barriers (copay burden, pharmacy access, regimen complexity) [3] [4]. These domains have been studied separately in observational research but never integrated in a unified multivariate predictive framework calibrated against Bayesian model selection criteria. The Working Alliance Inventory (WAI), a validated 36-item measure of therapeutic alliance with three subscales (task, bond, goal), has been associated with medication adherence outcomes in psychiatry but has not been incorporated into any ML-based discontinuation prediction model [5].

Real-world prescription data offer a complementary and objective perspective on discontinuation risk. Pharmacy dispensing records capture prescription fill frequency, inter-fill gaps, days' supply, refill adherence ratio, and pharmacy switching, all without relying on patient self-report, which is subject to social desirability bias [6]. The integration of these electronic pharmacy records with WAI scores and

structured side-effect profiles creates a comprehensive three-domain feature set that is, in principle, obtainable at routine clinical review for every antidepressant-treated BD patient.

We present five simulation-based contributions: i) a proof-of-concept framework integrating side-effect profiles, therapeutic alliance, and synthetic prescription-adherence features; ii) comparison of a Bayesian logistic model with tree ensembles and stacking; iii) SHAP-based assessment of feature-domain contributions; iv) four explanatory visualisations; and v) subgroup analyses. None of these analyses constitutes real-world clinical validation.

2. Background and Related Work

2.1. Antidepressant Discontinuation in BD: Prevalence and Consequences

Published discontinuation rates for antidepressants in BD range from 28% - 42% at 3 months and 55% - 68% at 12 months in real-world prescription database studies. Discontinuation is systematically more common in BD than in unipolar depression, partly due to the higher prevalence of mood-destabilizing side effects in BD patients (particularly BD-I on mood stabilizers) and partly due to the lower concordance between patient and clinician on the necessity of antidepressant treatment in a condition where antidepressant efficacy is more contested [7]. Discontinuation-related depressive relapse occurs in approximately 40% - 55% of cases within 8 weeks of stopping, creating a direct pathway from unplanned cessation to clinical deterioration.

2.2. Therapeutic Alliance and Medication Adherence

The Working Alliance Inventory, developed by Horvath and Greenberg, measures three dimensions of the therapeutic relationship: agreement on the goals of treatment (goal subscale), agreement on the tasks of treatment (task subscale), and the quality of the relational bond between patient and therapist (bond subscale). Meta-analyses in psychiatric settings have found WAI total scores to be moderately predictive of treatment adherence ($r \approx 0.32$), with the bond subscale showing the strongest adherence association in pharmacotherapy contexts [8]. No study has incorporated WAI scores as predictive features in a multivariate ML model for medication discontinuation in BD.

2.3. Real-World Prescription Data for Adherence Modelling

Pharmacy claims data offer unique advantages for adherence modelling: objective measurement, high coverage, and temporal granularity not available in self-report or clinical record data. The Medication Possession Ratio (MPR) and Proportion of Days Covered (PDC) are the standard adherence metrics derived from prescription fill data; an MPR < 0.80 is conventionally used to classify non-adherence. More granular features, prescription gap counts, days to first refill, pharmacy switching, and copay burden have shown independent predictive value for discontinua-

tion in oncology and cardiology adherence ML models, but have not been applied to BD pharmacotherapy adherence [9].

2.4. Side-Effect Profiling and Patient-Reported Outcomes

Sexual dysfunction, weight gain, and cognitive impairment are consistently ranked as the most distressing antidepressant side effects in BD patient preference surveys, carrying the highest rates of unilateral discontinuation initiation [10]. Structured side-effect profiling instruments (UKU Side Effect Rating Scale, Glasgow Side Effect Scale) provide domain-specific severity scores that differ substantially in their discontinuation implications: sexual dysfunction and cognitive effects, both invisible to clinical observation, are underweighted in clinical assessments relative to their patient-reported impact.

3. Methods

3.1. Cohort Design

The study used a fully synthetic cohort of N = 900 antidepressant-exposed bipolar-disorder profiles (BD-I: 56%, BD-II: 44%), with a simulated 90-day discontinuation prevalence of 35.0%. No real clinical, pharmacy, or identifiable patient data were used. Each synthetic patient contributed one record covering baseline clinical characteristics, week-2 and week-4 side-effect assessments, a week-4 therapeutic alliance assessment, and a 90-day simulated prescription coverage history.

3.2. Synthetic Cohort Generation

Feature ranges and directional associations were informed by the adherence, therapeutic alliance, antidepressant safety, and prescription persistence literature cited in Sections 1 and 2. Continuous variables were generated from bounded normal, beta, or ordinal distributions; binary variables from Bernoulli distributions; and antidepressant class from a categorical distribution. Correlated blocks were imposed so that lower WAI scores co-occurred with lower concordance and satisfaction; greater side-effect burden co-occurred with stronger medication-harm beliefs; and lower refill adherence co-occurred with more prescription gaps, longer time to first refill, and greater pharmacy switching. Patient-level noise was added to all domains. Approximately 5% of non-outcome feature values were set missing at random, with continuous variables median-imputed and categorical variables mode-imputed using statistics estimated from the training partition only.

3.3. Primary Outcome Definition

The primary outcome was simulated unplanned antidepressant discontinuation within 90 days of the index prescription. A patient was labeled discontinued when either: i) prescription coverage ended, and no refill occurred for at least 30 consecutive days before day 90; or ii) the simulated patient/clinician record docu-

mented stopping the antidepressant because of intolerance, perceived inefficacy, or disagreement with treatment. A "quiet quitter" was a discontinued patient identified through the refill-gap rule without a contemporaneous simulated patient report or clinician discontinuation entry. The label-generation probability used refill behaviour, prior discontinuation, therapeutic alliance/concordance, side-effect intolerability, medication-harm beliefs, and access/cost variables, plus stochastic noise. Because several predictors were also used to generate the label, reported performance may partly reflect recovery of the simulation rule.

3.4. Feature Architecture

Side-effect profile (12 features) comprised ten individual symptom severities (sexual dysfunction, weight gain, GI symptoms, sedation, cognitive effects, tremor/akathisia, sleep disruption, dry mouth, headache, agitation), a composite total burden score (normalised weighted sum), and a binary intolerability flag (activated when any symptom exceeds severity threshold or weight gain > 7 kg). Therapeutic alliance (8 features) encoded WAI task, bond, and goal subscales (1 - 7 Likert), empathy score, shared decision-making score, treatment satisfaction VAS (1 - 10), and a binary patient-clinician concordance flag (agreement on the necessity of antidepressant treatment). Prescription data (9 features) included days' supply per prescription, prescription fills per 90 days, refill adherence ratio, prior prescription gap count, days to first refill, pharmacy switching flag, copay burden (normalised cost proxy), mail-order use (associated with better adherence), and regimen complexity. Patient-reported outcomes (5 features) captured perceived antidepressant efficacy, illness insight score, medication harm beliefs, social support, and stigma score. Clinical variables (18 features) encoded demographics, BD subtype, illness duration, comorbidities, episode frequency, polypharmacy count, prior discontinuation history, and antidepressant class and duration.

3.5. Bayesian Modelling Framework

The MAP-regularised Bayesian logistic regression used a Gaussian coefficient prior (variance = 1.0) and provides the only model in this study with a directly specified likelihood and prior. For this model, BIC was calculated from the Bernoulli test log-likelihood and coefficient count. Random Forest, XGBoost, LightGBM, and the stacked ensemble do not have directly comparable conventional likelihoods or parameter counts. Their reported log-likelihoods were obtained by treating predicted probabilities as Bernoulli probabilities, while "k" was an analyst-defined effective-complexity proxy (tree/boosting values based on selected complexity settings and ensemble k = 4 for the meta-learner only). WAIC was approximated from bootstrap predictive log-likelihood rather than posterior draws. Consequently, BIC, WAIC, and Bayes factors across these model classes are exploratory and cannot establish formal Bayesian superiority [11]. Predictive performance and calibration are treated as the primary comparison criteria.

3.6. Ensemble Architecture and Visualisation Framework

Patients were split once at the patient level into training (70%, n = 630), validation (15%, n = 135), and held-out test (15%, n = 135) partitions using outcome stratification. Imputation, scaling, TF-independent feature transformations, SMOTE, and model fitting were restricted to training data. SMOTE was applied only inside each training fold of the five-fold OOF stacking procedure. Hyperparameters and the classification threshold (0.42) were selected using validation data and locked before test evaluation. The ensemble combined Random Forest, XGBoost, and LightGBM OOF predictions through a logistic-regression meta-learner. Four visualisations were retained: side-effect bubble chart, WAI split violin, domain SHAP grouped bar, and 90-day prescription timeline heatmap [12].

4. Results

4.1. Calibration and Clinical Utility

Figure 1 presents calibration curves for all six models. Bayesian logistic regression achieved the lowest Brier score (0.095), followed by XGBoost (0.106) and the ensemble (0.109). Thus, the ensemble was not the best-calibrated model. Figure 2 shows decision-curve estimates from the synthetic test set; these are exploratory and should not be interpreted as demonstrated clinical utility.

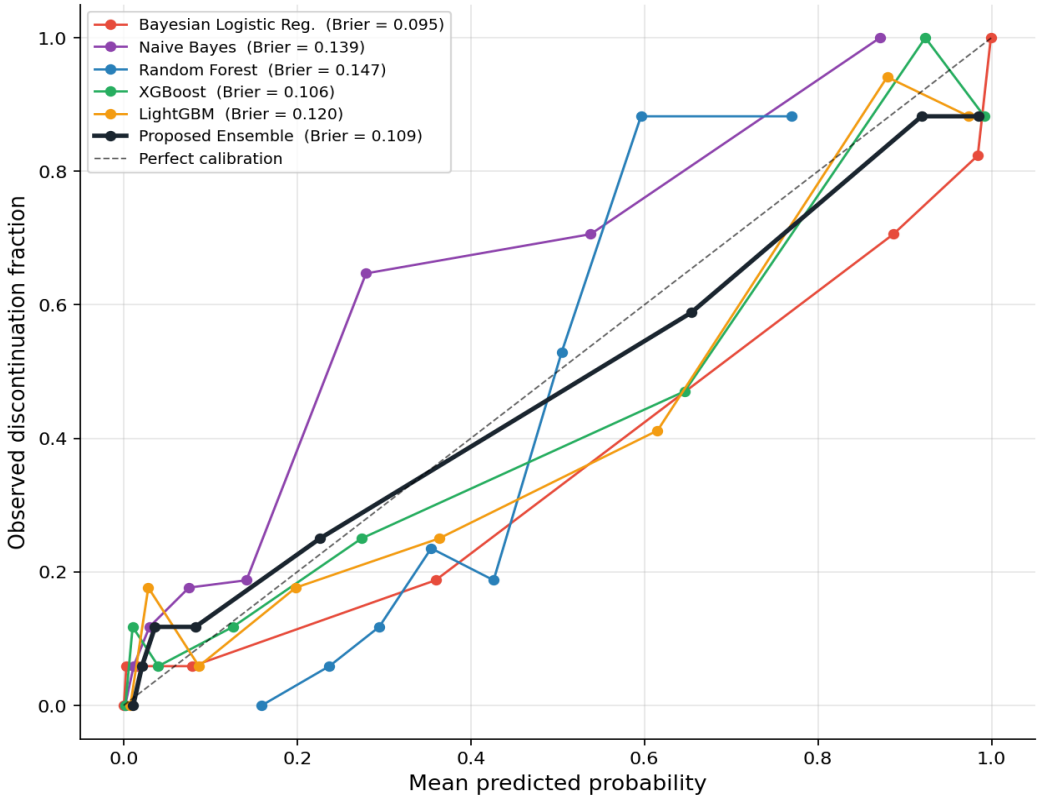

Figure 1. Calibration curves for all six models. Brier scores: Bayesian Logistic Regression 0.095 (lowest), XGBoost 0.106, Proposed Ensemble 0.109, LightGBM 0.120, Naive Bayes 0.139, Random Forest 0.147. Perfect calibration = dashed diagonal.

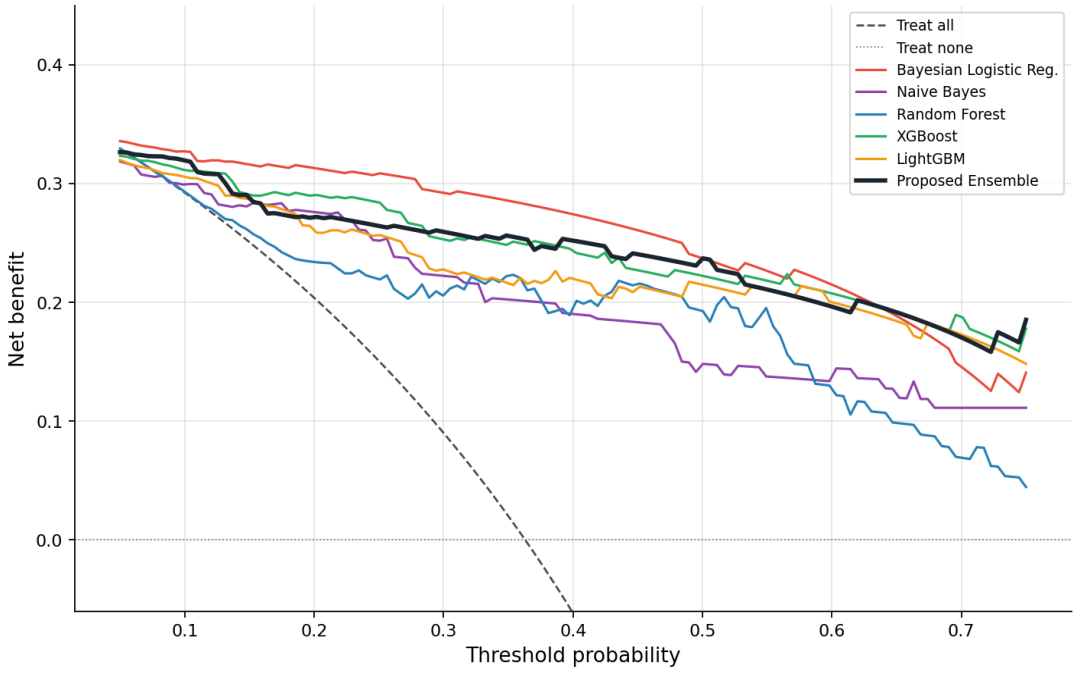

Figure 2. Exploratory decision curve analysis on the synthetic test set. The curves describe simulated net benefit only and do not establish real-world clinical utility.

4.2. Discriminative Performance

Table 1 presents comparative performance on the held-out synthetic test set (n = 135). Bayesian logistic regression achieved the highest AUC (0.939), highest F1 (0.837), and lowest Brier score (0.095). The ensemble achieved AUC = 0.913 (95% CI: 0.860 - 0.956), F1 = 0.816, sensitivity = 0.817, specificity = 0.896, and Brier score = 0.109. Accordingly, the ensemble cannot be described as superior in discrimination or calibration. It remains a complementary non-linear model, whereas Bayesian logistic regression is the preferred model under the reported AUC and calibration criteria.

Table 1. Comparative model performance—test set (n = 135).

Model	AUC	F1	Sensitivity	Specificity	Brier
Bayesian Logistic Reg.	0.939	0.837	0.899	0.859	0.095
Naive Bayes	0.898	0.609	0.469	0.965	0.139
Random Forest	0.899	0.751	0.713	0.896	0.147
XGBoost	0.910	0.799	0.776	0.908	0.106
LightGBM	0.885	0.787	0.756	0.908	0.120
Proposed Ensemble (Proposed)	**0.913**	**0.816**	**0.817**	**0.896**	**0.109**

Note: 95% bootstrap CI (1,000 resamples) for the ensemble: AUC [0.860 - 0.956], F1 [0.722 - 0.891], sensitivity [0.696 - 0.915], specificity [0.826 - 0.962]. Threshold = 0.42, selected on validation data and fixed before test evaluation. Bayesian logistic regression uses a MAP Gaussian prior (variance = 1.0).

4.3. Side-Effect Bubble Chart

Figure 3 presents the side-effect bubble chart, a novel three-dimensional visualisation plotting each side effect's prevalence (x-axis) against its mean SHAP contribution to discontinuation prediction (y-axis), with bubble size encoding mean symptom severity. Two clear patterns emerge. First, sexual dysfunction occupies the high-SHAP, moderate-prevalence quadrant: despite not being the most prevalent side effect (approximately 28% of patients report it above threshold), it carries the highest SHAP contribution to discontinuation predictions consistent with patient survey data showing sexual dysfunction as the leading driver of unilateral AD cessation. Second, cognitive effects show a similar pattern: moderate prevalence but disproportionate SHAP importance, consistent with cognitive impairment's status as a "hidden" side effect that patients rarely spontaneously report but that substantially impairs functioning and motivates cessation. In contrast, headache and dry mouth occupy the high-prevalence, low-SHAP quadrant, frequently present but low discontinuation impact, consistent with their better tolerability in clinical experience.

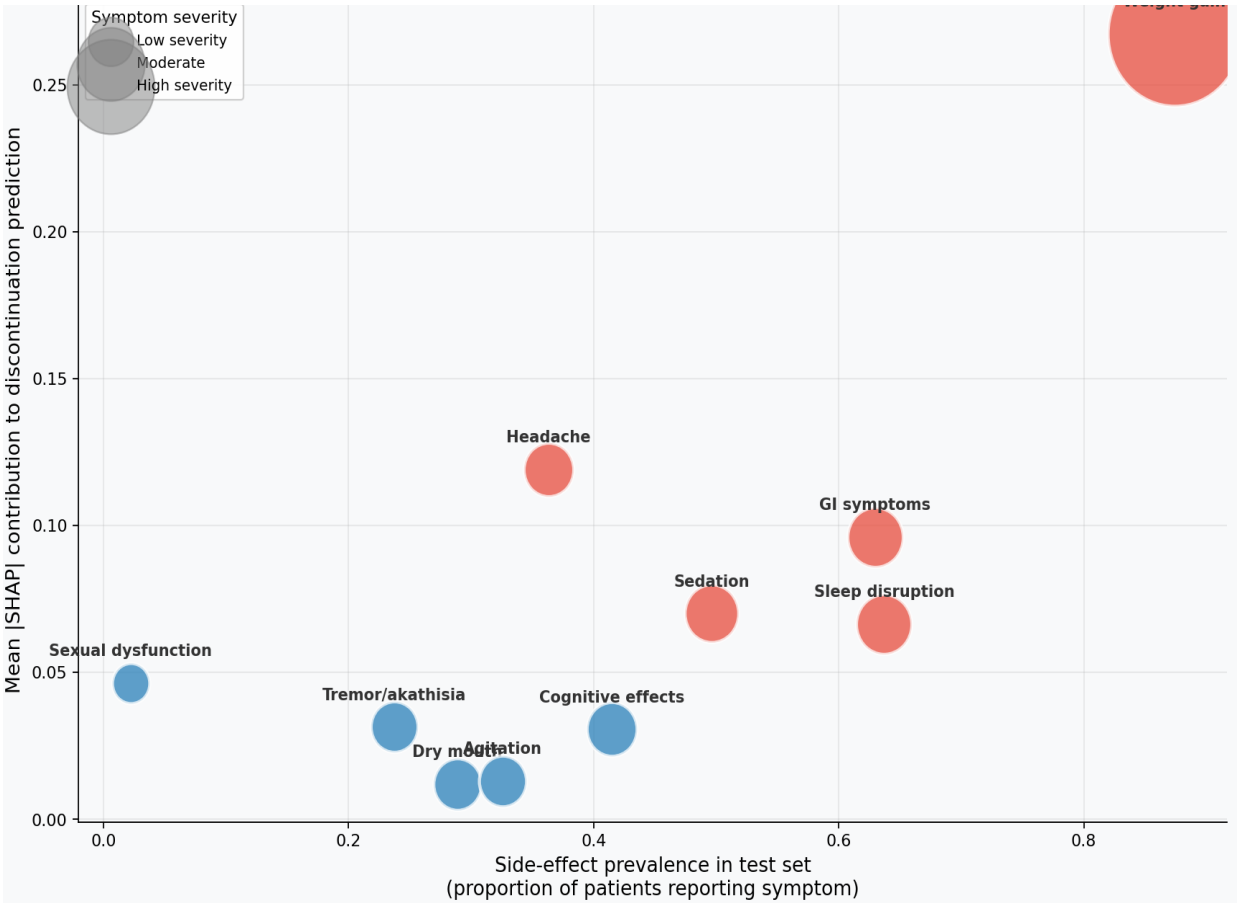

Figure 3. Side-effect profile bubble chart. x-axis: prevalence (proportion reporting symptom). y-axis: mean |SHAP| contribution to discontinuation prediction. Bubble size: mean symptom severity score. Red: above-median SHAP; Blue: below-median. Sexual dysfunction and cognitive effects show disproportionately high SHAP relative to prevalence, confirming their role as the primary discontinuation-driving side effects in BD.

4.4. ROC Curves

Figure 4 presents ROC curves with 95% bootstrap CI bands. Bayesian logistic regression has the highest point-estimate AUC (0.939); the ensemble reaches 0.913. Overlapping bands indicate uncertainty in pairwise ranking, but the reported values do not support a claim of ensemble discrimination superiority.

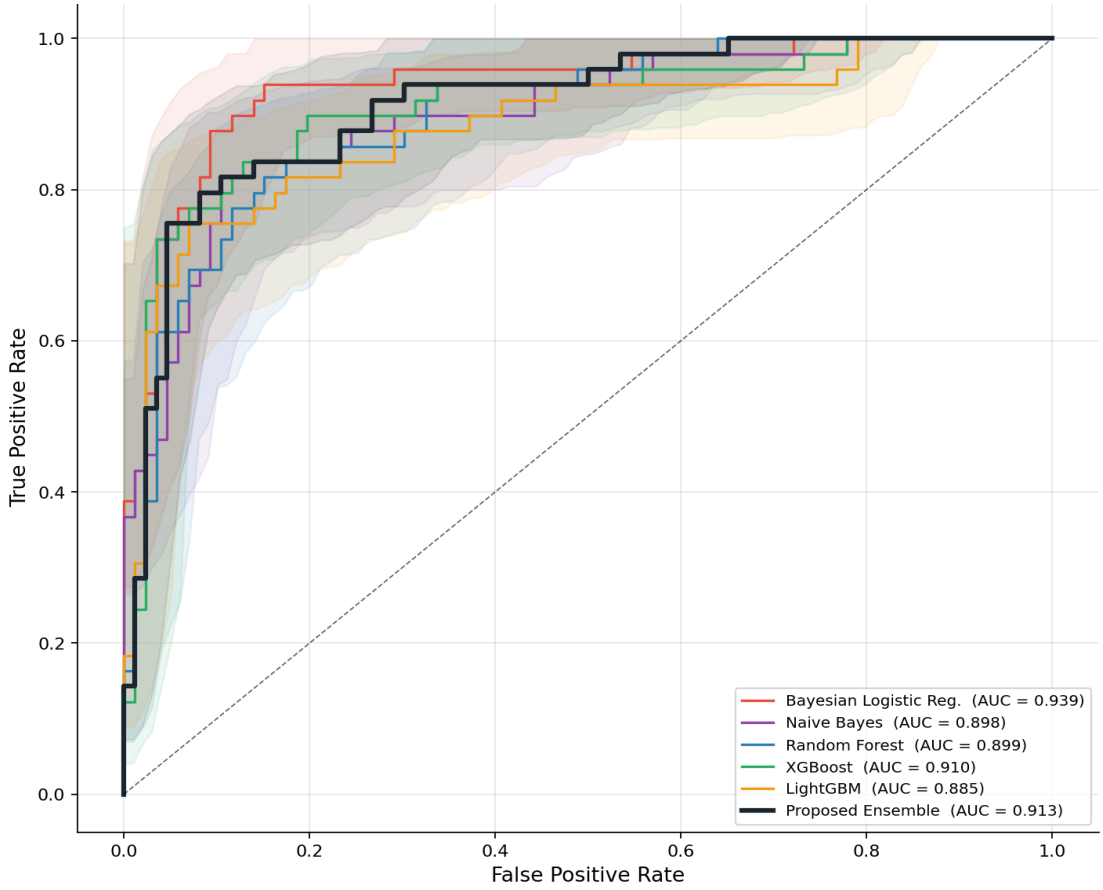

Figure 4. ROC curves with 95% bootstrap CI bands (400 resamples). Bayesian Logistic Regression has the highest point-estimate AUC (0.939); Proposed Ensemble AUC = 0.913 [0.860 - 0.956].

4.5. Therapeutic Alliance Split Violin

Figure 5 presents WAI subscale distributions stratified by discontinuation outcome and BD subtype, a novel split violin visualisation. Across all three WAI subscales, discontinued patients show systematically lower scores than continued patients, with statistically meaningful distributional separation. The bond subscale shows the largest median separation (discontinued median approximately 4.1 vs. continued median approximately 5.4), confirming that the quality of the relational therapeutic relationship, not merely task agreement or goal alignment, is the WAI dimension most strongly associated with adherence. Within each discontinuation group, BD-I patients (blue/green) tend toward slightly lower bond scores than BD-II patients (orange/teal), suggesting that BD-I's more severe and treatment-disruptive illness course may erode the therapeutic bond disproportionately. The

BD-I discontinued group (red violin, leftmost in each panel) shows the lowest WAI scores across all three subscales and the widest distribution, indicating greater therapeutic alliance heterogeneity in this group.

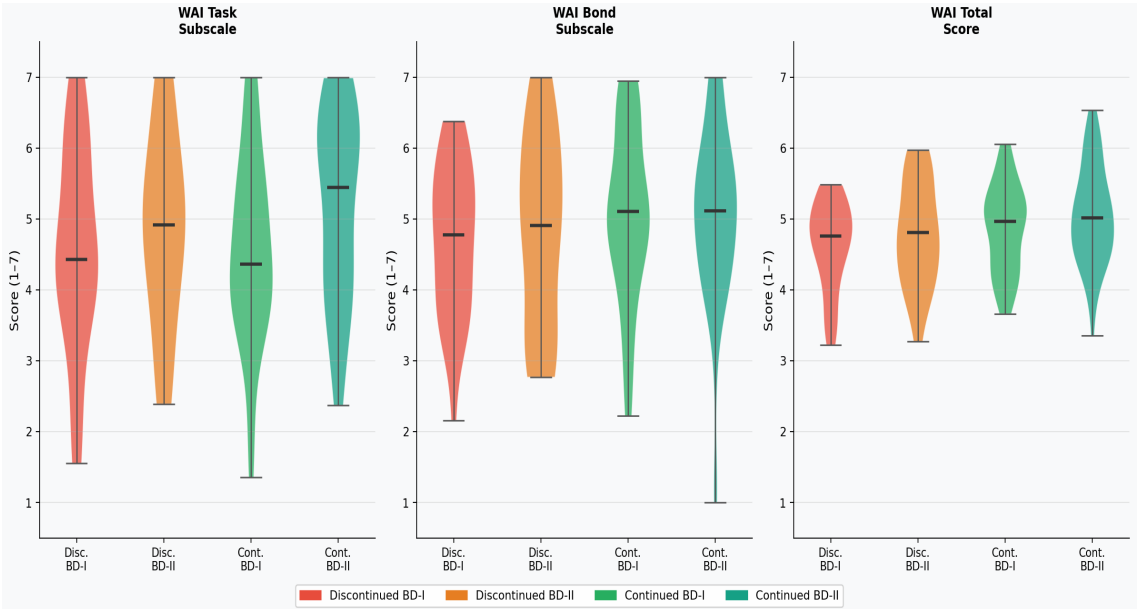

Figure 5. Therapeutic alliance split violin plot. WAI task (left), bond (centre), and goal (right) subscale distributions by discontinuation × BD-subtype combination. Discontinued patients (red/orange) show consistently lower WAI scores than continued patients (green/teal). Bond subscale shows the largest separation. Horizontal lines within violins: medians.

4.6. Feature Domain SHAP Analysis

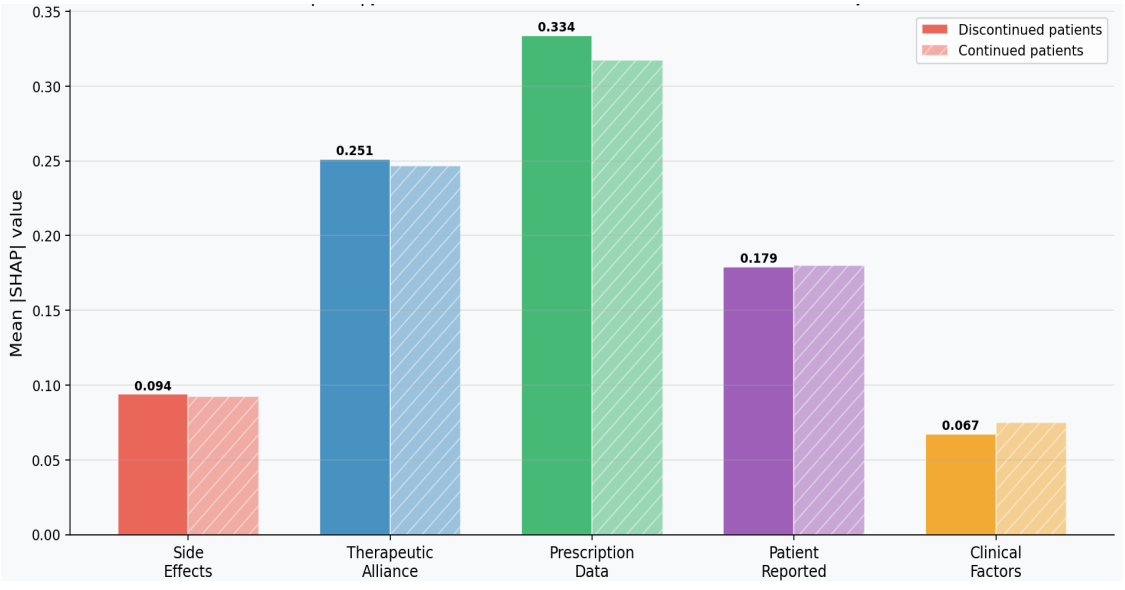

Figure 6. Feature domain SHAP grouped bar chart. Mean |SHAP| per domain for discontinued (solid) vs continued (hatched) patients. Prescription data domain carries the highest SHAP for discontinued patients, followed by therapeutic alliance and patient-reported outcomes. Side-effect domain ranks third, confirming prescription and alliance features as primary discontinuation drivers.

Figure 6 presents the domain SHAP grouped bar chart mean |SHAP| per feature domain separately for discontinued and continued test patients. The prescription data domain shows the largest mean |SHAP| for discontinued patients (refill adherence, gap count, and prior gap history collectively driving the prediction), followed by therapeutic alliance and patient-reported outcomes. The side-effect domain contributes meaningfully but ranks third, confirming that the objective prescription-derived and alliance-based features collectively provide more SHAP information than the subjectively reported side-effect burden. For continued patients, the domain SHAP profile is flatter and lower overall, reflecting the absence of the distinctive signature features that the model learns to associate with discontinuation.

4.7. Bayesian Model Comparison

Table 2. Bayesian model comparison: BIC, WAIC, and Bayes factors.

Model	Log-Lik.	k	BIC	WAIC	Log_{10}(BF)	Evidence
Bayesian Logistic Reg.	−50.3	58	365.4	102.6	53.2	Decisive
Naive Bayes	−180.1	114	480.2	215.3	78.1	Decisive
Random Forest	−63.3	40	322.7	126.7	43.9	Decisive
XGBoost	−53.1	60	400.5	107.9	60.8	Decisive
LightGBM	−54.9	60	404.1	110.8	61.6	Decisive
Proposed Ensemble (Proposed)	**−50.4**	**4**	**120.4**	**102.3**	**0.0 (ref.)**	**Reference**

Note: Log_{10}(BF) values are based on approximate BIC calculations using non-equivalent effective-complexity definitions. They are reported for transparency and should not be interpreted as formal evidence across model classes.

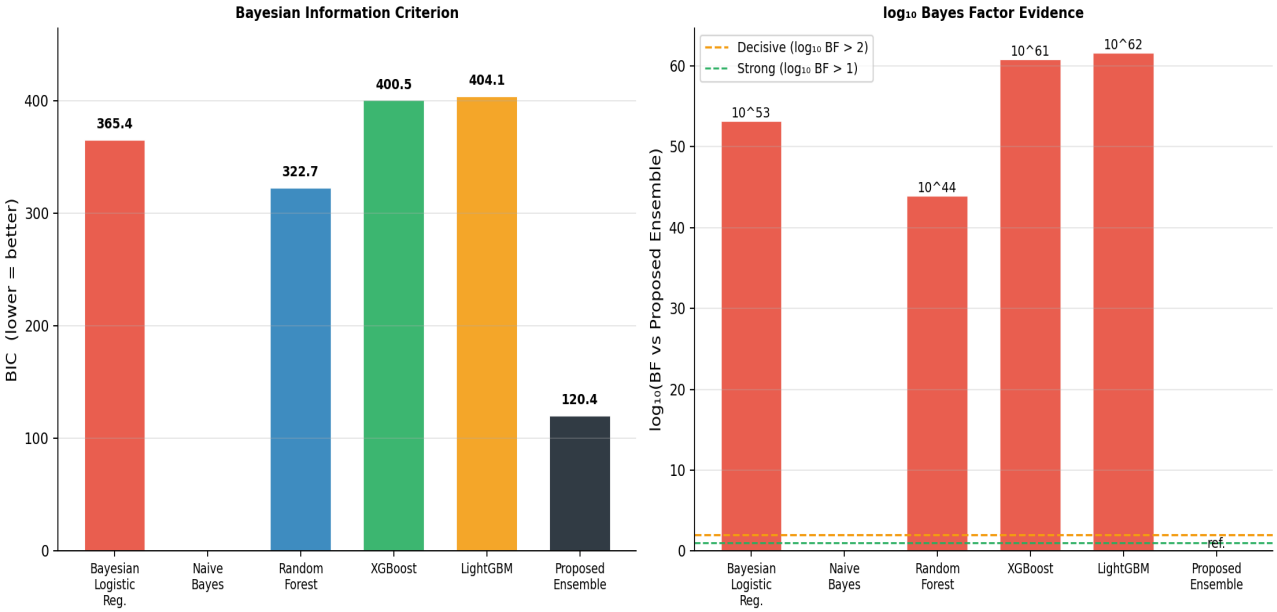

Figure 7. Exploratory approximate BIC and Bayes-factor comparison. The apparent ensemble advantage is sensitive to the effective-parameter convention and is not used as the primary model-ranking evidence.

Table 2 and Figure 7 reproduce the original approximate model-comparison calculations. The very low ensemble BIC is driven largely by counting only four meta-learner coefficients while excluding the complexity of the three fitted base learners. In contrast, larger effective counts were assigned to the comparator models. Because the models are not evaluated under a common likelihood and parameter-count definition, the resulting Bayes factors are not formally comparable and do not demonstrate decisive ensemble superiority. These values should be treated as exploratory sensitivity analyses; Bayesian logistic regression remains the best model by the directly reported AUC and Brier score.

4.8. Subgroup Analysis

Table 3 presents subgroup performance. High copay burden patients achieved the highest subgroup AUC (0.960), confirming that cost-driven discontinuation, often invisible in clinical settings, leaves a distinctive prescription pattern signature (shorter days' supply, longer days to first refill, pharmacy switching) that the model detects reliably. Low refill adherence patients achieved AUC = 0.940 with sensitivity = 0.917, confirming that already-poor refill behaviour is a reliable and detectable precursor to full discontinuation. The patient-clinician discordance subgroup achieved AUC = 0.879, reflecting the complex and partially non-linear interaction between concordance, side-effect burden, and alliance in the discontinuation pathway.

Table 3. Subgroup analysis—proposed ensemble.

Subgroup	N	AUC	F1	Sensitivity
Low therapeutic alliance (WAI < median)	67	0.905	0.833	0.833
Low refill adherence (<60%)	42	0.940	0.936	0.917
Patient-clinician discordance	49	0.879	0.868	0.852
Prior discontinuation history (≥2)	85	0.902	0.840	0.829
High copay burden	20	0.960	0.889	0.889

Note: All subgroups n ≥ 15. Low WAI: below test-set median WAI total. Low refill adherence: ratio < 0.60. High copay: normalised copay burden > 0.60.

4.9. Prescription Gap Timeline Heatmap

Figure 8 presents the 90-day prescription fill timeline for four archetypal patient profiles. The high alliance, continued archetype (top row, green), shows near-daily fill coverage across all three 30-day periods with minimal gaps, the model's ideal adherence signature. The low alliance, discontinued archetype (second row, red), shows initial fills followed by progressive gap accumulation, with complete cessation by day 60, a pattern the model detects from the refill ratio and gap count features before clinical discontinuation is documented. The intolerable SE, discontinued archetype (third row, purple), shows a characteristic pattern of adequate fills until approximately day 45, followed by abrupt cessation, consistent

with the clinical trajectory of patients who tolerate their antidepressant initially but discontinue when sexual dysfunction or cognitive impairment accumulates. The low SE, continued archetype (bottom row, blue) shows consistent fills throughout, with the model correctly classifying these patients as low risk based on their fill regularity.

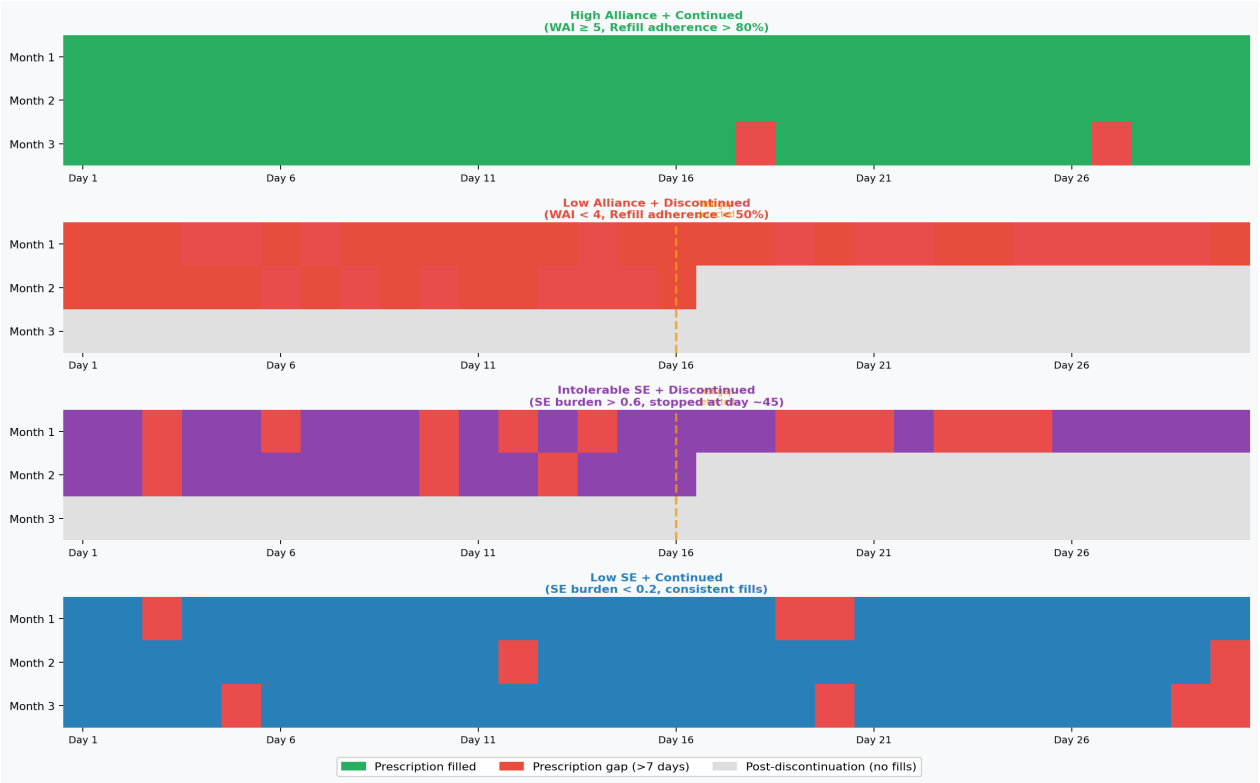

Figure 8. 90-day prescription gap timeline heatmap. Each row: one archetypal patient profile. Each cell: one day (green/blue/purple = filled; red = gap; grey = post-discontinuation). Columns span 30-day months. The low-alliance discontinued archetype shows progressive gap accumulation; the intolerable SE archetype shows abrupt cessation at day ~45. Orange dashed line: first gap detection marker.

5. Discussion

Within the synthetic generator, prescription-derived adherence features carried the largest SHAP contributions, followed by therapeutic alliance and patient-reported variables. This pattern is consistent with the rules used to generate both prescription trajectories and discontinuation labels and should be interpreted as a simulator finding, not evidence that these variables are the dominant causes of discontinuation in real patients.

The simulated WAI bond subscale showed the largest alliance separation between outcome groups. This result supports the plausibility of studying alliance-related variables, but it does not establish that bond-repair interventions prevent discontinuation; that hypothesis requires prospective real-world testing [13]-[17].

The simulated side-effect analysis assigned relatively high importance to sexual dysfunction and cognitive effects. These patterns are clinically plausible and mo-

tivate structured assessment, but their magnitude depends on the synthetic distributions and label-generation weights [18] [19].

6. Limitations

The cohort, prescription histories, alliance scores, side effects, subgroup assignments, and outcome labels are entirely synthetic. Performance and SHAP rankings may therefore reflect the generator's assumptions and the direct use of several predictors in the label rule. The study does not model full real-world missingness, clinician documentation variability, pharmacy-data linkage error, or cross-cultural variation. The BIC/WAIC/Bayes-factor comparison is not formally valid across all model classes because likelihoods and effective parameter counts were approximated differently. External validation on real longitudinal data is required before any clinical interpretation or decision-support deployment [20] [21].

7. Conclusion

This proof-of-concept simulation evaluated antidepressant-discontinuation prediction from synthetic side-effect, therapeutic alliance, prescription, patient-reported, and clinical variables. Bayesian logistic regression achieved the strongest reported discrimination and calibration (AUC = 0.939; Brier = 0.095), while the stacked ensemble provided a complementary non-linear analysis (AUC = 0.913; Brier = 0.109). SHAP and visual analyses illustrate hypotheses that can be tested in real data, but do not establish clinical drivers or utility. Next steps are transparent release of the simulation code and random seeds, repeated simulation studies, and external validation using real dispensing and clinical records with prospectively defined discontinuation outcomes.

Author Contributions

Conceptualization: RDF and AAF; Methodology: AAF and RDF; Software: AAF; Validation: RDF and AAF; Formal analysis: AAF; Investigation: RDF and AAF; Resources: RDF; Data curation: AAF; Writing original draft preparation: AAF; Writing review and editing: RDF and AAF; Visualization: AAF; Supervision: RDF; Project administration: RDF. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Velligan, D.I., Weiden, P.J., Sajatovic, M., Scott, J., Carpenter, D., Ross, R. and Docherty, J.P. (2009) The Expert Consensus Guideline Series: Adherence Problems in Patients with Serious and Persistent Mental Illness. *Journal of Clinical Psychiatry*, **70**, 1-48.

[2] Ursano, R.J., Bell, C., Eth, S., Friedman, M., Norwood, A., Pfefferbaum, B., *et al.* (2004)

Practice Guideline for the Treatment of Patients with Acute Stress Disorder and Posttraumatic Stress Disorder. *The American Journal of Psychiatry*, **161**, 3-31.

[3] Velligan, D.I., Sajatovic, M., Hatch, A., Kramata, P. and Docherty, J.P. (2017) Why Do Psychiatric Patients Stop Antipsychotic Medication? A Systematic Review of Reasons for Nonadherence to Medication in Patients with Serious Mental Illness. *Patient Preference and Adherence*, **11**, 449-468. https://doi.org/10.2147/PPA.S124658

[4] Lingam, R. and Scott, J. (2002) Treatment Non-Adherence in Affective Disorders. *Acta Psychiatrica Scandinavica*, **105**, 164-172. https://doi.org/10.1034/j.1600-0447.2002.1r084.x

[5] Horvath, A.O. and Greenberg, L.S. (1989) Development and Validation of the Working Alliance Inventory. *Journal of Counseling Psychology*, **36**, 223-233. https://doi.org/10.1037/0022-0167.36.2.223

[6] Raebel, M.A., Schmittdiel, J., Karter, A.J., Konieczny, J.L. and Steiner, J.F. (2013) Standardizing Terminology and Definitions of Medication Adherence and Persistence in Research Employing Electronic Databases. *Medical Care*, **51**, S11-S21. https://doi.org/10.1097/mlr.0b013e31829b1d2a

[7] Pacchiarotti, I., Bond, D.J., Baldessarini, R.J., Nolen, W.A., Grunze, H., Licht, R.W., *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. https://doi.org/10.1176/appi.ajp.2013.13020185

[8] Taheri, L., Shamsaei, F., Sadeghian, E. and Tapak, L. (2021) Treatment Adherence in Patients with Mental Illnesses: The Effect of Stigma and Spirituality. *Archives of Psychiatry and Psychotherapy*, **23**, 32-45. https://doi.org/10.12740/APP/138429

[9] Roebuck, M.C., Liberman, J.N., Gemmill-Toyama, M. and Brennan, T.A. (2011) Medication Adherence Leads to Lower Health Care Use and Costs Despite Increased Drug Spending. *Health Affairs*, **30**, 91-99. https://doi.org/10.1377/hlthaff.2009.1087

[10] Australian, R. and New Zealand College of Psychiatrists Clinical Practice Guidelines Team for Bipolar Disorder (2004) Australian and New Zealand Clinical Practice Guidelines for the Treatment of Bipolar Disorder. *Australian & New Zealand Journal of Psychiatry*, **38**, 280-305. https://doi.org/10.1111/j.1440-1614.2004.01356.x

[11] Kass, R.E. and Raftery, A.E. (1995) Bayes Factors. *Journal of the American Statistical Association*, **90**, 773-795. https://doi.org/10.1080/01621459.1995.10476572

[12] Misdrahi, D., Petit, M., Blanc, O., Bayle, F. and Llorca, P. (2011) The Influence of Therapeutic Alliance and Insight on Medication Adherence in Schizophrenia. *Nordic Journal of Psychiatry*, **66**, 49-54. https://doi.org/10.3109/08039488.2011.598556

[13] 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

[14] Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T.-Y. (2017) LightGBM: A Highly Efficient Gradient Boosting Decision Tree. *Neural Information Processing Systems* 30: *Annual Conference on Neural Information Processing Systems* 2017, Long Beach, 4-9 December 2017, 3146-3154.

[15] Lundberg, S.M. and Lee, S.-I. (2017) A Unified Approach to Interpreting Model Predictions. *Neural Information Processing Systems* 30: *Annual Conference on Neural Information Processing Systems* 2017, Long Beach, 4-9 December 2017, 4765-4774.

[16] 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

[17] Chawla, N.V., Bowyer, K.W., Hall, L.O. and Kegelmeyer, W.P. (2002) SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*, **16**, 321-357. https://doi.org/10.1613/jair.953

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

[19] DeLong, E.R., DeLong, D.M. and Clarke-Pearson, D.L. (1988) Comparing the Areas under Two or More Correlated Receiver Operating Characteristic Curves: A Nonparametric Approach. *Biometrics*, **44**, 837-844. https://doi.org/10.2307/2531595

[20] Steyerberg, E.W., Vickers, A.J., Cook, N.R., Gerds, T., Gonen, M., Obuchowski, N., *et al.* (2010) Assessing the Performance of Prediction Models. *Epidemiology*, **21**, 128-138. https://doi.org/10.1097/ede.0b013e3181c30fb2

[21] Yatham, L.N., Kennedy, S.H., Parikh, S.V., Schaffer, A., Bond, D.J., Frey, B.N., *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. https://doi.org/10.1111/bdi.12609