



Figure 1. Overview of the project timeline and workflow across seven sequential steps. Step 01 corresponds to the retrospective phase: aggregation and preprocessing of existing datasets from ConLiGen and PGBD (see Sections 2.1-2.2). Steps 02-07 comprise the prospective phase. Step 02 initiates prospective recruitment of 100 lithium-naïve patients with bipolar disorder (BD) (Section 2.2). Step 03 involves multi-omics and clinical data collection (Section 2.3). Steps 04-06 describe model development: training a multilayer feed-forward neural network (MFNN) using prospective data (Step 04; Section 2.6), validation with a dedicated internal subset (Step 05; Sections 2.6-2.7), and final testing on a held-out dataset (Step 06; Section 2.7). Step 07 captures longitudinal follow-up at 6-month intervals for up to 36 months, supporting outcome refinement and model updating (Section 2.8). Collectively, these steps follow the two-phase structure detailed in Section 2.1: retrospective pre-training followed by prospective fine-tuning and validation.

Step	Description	Target N	Key Outcome	Tools/Considerations
1. Retrospective Data Collection (Phase I)	Aggregate and preprocess existing lithium response datasets (e.g., ConLiGen, PGBD) to extract SNPs, available RNA-seq/methylation data and Alda scores. Format features for initial model input.	retrospective datasets	Foundation for pre-training and feature selection	Alda scale (≥ 7 = responder), data normalization, SMOTE, dimensionality reduction (LASSO, PCA)
2. Prospective Cohort Recruitment (Phase II)	Enroll 100 lithium-naïve BD patients via partner clinics. Screen for inclusion/exclusion and obtain consent for clinical and molecular data collection.	100	Consent-based, diverse cohort for multi-omic study	Batch control, standardized protocols, consistent sample timing, ethics/REB clearance
3. Multi-Omics Data Collection (Phase II)	Collect blood for RNA-seq and DNA methylation profiling. Store serum for future metabolomics. Gather standardized clinical data (e.g., YMRS, MADRS, CGI-BP).	100 (multi-omics)	High-resolution multi-modal input dataset	Biobanking, SOPs, QC, morning sampling to reduce circadian noise
4. Model Training (Phase II)	Train a multilayer feed-forward neural network (MFNN) using multi-omics and clinical data from 50 patients. Address class imbalance.	50 (training subset)	Trained model with internal cross-validation	SMOTE, class weighting, 10-fold CV, dropout, embedded feature selection (LASSO/PCA/autoencoders)
5. Validation (Phase II)	Optimize architecture and tune hyperparameters using a dedicated validation set. Assess feature importance.	25 (validation subset)	Refined model with interpretability checks	Early stopping, SHAP explanations, performance tuning
6. Testing (Phase II)	Evaluate final model performance on an unseen hold-out set.	25 (test subset)	Generalizability assessment; benchmark metrics	AUROC, accuracy, F1, precision/recall, calibration curve, confusion matrix
7. Longitudinal Follow-Up (Phase II)	Conduct 6-month follow-ups for up to 36 months. Track response durability, refine labels, and support model retraining.	100 (full cohort)	Outcome validation, label refinement, and model retraining	Imputation, survival analysis, sensitivity testing, model updating feedback loop

Table 1. Sequential project steps from retrospective dataset processing to real-world follow-up. Steps 1-7 align with the core workflow described in the methodology (Sections 2.1-2.8). Step 1 involves retrospective data aggregation from the International Consortium on Lithium Genetics (ConLiGen) and the Pharmacogenomics of Bipolar Disorder (PGBD) study for pre-training. Steps 2-6 represent the prospective pipeline: recruitment of 100 lithium-naïve patients with bipolar disorder (BD), multi-omics and clinical data collection, training of a multilayer feed-forward neural network (MFNN), model validation, and testing. Step 7 includes longitudinal follow-up to refine outcome labels and improve model performance. Training/validation/test subsets derived from prospective cohort of N=100.

[Abbreviations: BD - bipolar disorder; SNP - single-nucleotide polymorphism; RNA-seq - RNA sequencing; MFNN - multilayer feed-forward neural network; CGI-BP - Clinical Global Impressions Scale-Bipolar Version; YMRS - Young Mania Rating Scale; MADRS - Montgomery-Åsberg Depression Rating Scale; SHAP - SHapley Additive exPlanations; PCA - principal component analysis; LASSO - least absolute shrinkage and selection operator; SMOTE - Synthetic Minority Oversampling Technique; CV - cross-validation; SOP - standard operating procedure; QC - quality control; REB - Research Ethics Board.]