



Jan 09, 2026

Version 2

AI at the Bedside of Psychiatry: Comparative Meta-analysis of Imaging vs Non-imaging Machine-Learning Models for Differentiating Bipolar vs Unipolar Depression at First Episode V.2

DOI

<https://dx.doi.org/10.17504/protocols.io.4r3l21w5pg1y/v2>

Andrei Dăescu^{1,2}

¹Doctoral School Department, "Victor Babes" University of Medicine and Pharmacy, 300041 Timisoara, Romania;

²Neurosciences Department, Discipline of Psychiatry, "Victor Babes" University of Medicine and Pharmacy, 300041 Timisoara, Romania



andreidaescu

Create & collaborate more with a free account

Edit and publish protocols, collaborate in communities, share insights through comments, and track progress with run records.

Create free account

OPEN  ACCESS



DOI: <https://dx.doi.org/10.17504/protocols.io.4r3l21w5pg1y/v2>

Protocol Citation: Andrei Dăescu 2026. AI at the Bedside of Psychiatry: Comparative Meta-analysis of Imaging vs Non-imaging Machine-Learning Models for Differentiating Bipolar vs Unipolar Depression at First Episode. **protocols.io**

<https://dx.doi.org/10.17504/protocols.io.4r3l21w5pg1y/v2> Version created by [andreidaescu](#)

License: This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: In development

We are still developing and optimizing this protocol

Created: January 09, 2026

Last Modified: January 09, 2026

Protocol Integer ID: 238291

Keywords: bipolar vs unipolar depression, bipolar disorder, episode bipolar disorder, differentiating bipolar, unipolar major depressive disorder, unipolar depression, ai at the bedside, diagnostic accuracy framing, comparing diagnostic performance, proteomic biomarker, diagnostic performance across modality, psychiatry, imaging, analysis of imaging, ai consideration, depression, biomarker

Abstract

First-episode bipolar disorder (BD) is frequently misclassified as unipolar major depressive disorder (MDD). Artificial intelligence/machine-learning (AI/ML) models trained on imaging (e.g., MRI, EEG) and non-imaging (e.g., EHR/clinical, voice/NLP, actigraphy, blood/proteomic biomarkers) data show promise for early differentiation. This protocol describes a PRISMA-aligned systematic review and meta-analysis comparing diagnostic performance across modalities. We include primary human studies (2014–2025) that report test-set discrimination (primary outcome: AUC; secondary: sensitivity/specificity at a specified threshold and calibration metrics). Risk of bias will be appraised with QUADAS-2 (diagnostic accuracy framing) and PROBAST/AI considerations (prediction-model analysis). Random-effects synthesis will use Hartung–Knapp adjustment with REML for τ^2 . A priori, we contrast imaging vs non-imaging models and run sensitivity analyses excluding high-risk-of-bias or potential data-leakage studies.

Search snapshot: identified 158 records, removed 39 duplicates, screened 119, included 17 (qualitative), of which 6 were meta-ready.



Version Information:

- 1 Version: v1.1
Date: 05/01/2026
Corresponding author: ana-maria.daescu@umft.ro

Materials and Equipment:

- 2 Reference manager (e.g., Mendeley/Zotero) for import and de-duplication
Spreadsheet/REDCap/Google Sheet for screening and data extraction
R (≥ 4.2) with packages: metafor, meta, robvis, DiagrammeR

Procedure Steps:

- 3 **Step 1: Objectives and Eligibility**
Objective. To estimate the diagnostic performance of AI/ML models distinguishing BD vs MDD at first episode, and compare imaging vs non-imaging modalities.
Inclusion. Original human studies; first-episode mood participants with separable BD vs MDD groups; supervised ML classifier (e.g., SVM, RF, XGBoost, penalized LR, CNN); imaging or non-imaging inputs; test-set performance (AUC; CI/SE preferred) and BD/MDD class counts; internal hold-out/k-fold or external validation.
Exclusion. Schizophrenia/FEP cohorts without separable BD vs MDD; prognosis/treatment-response/service-evaluation tasks; non-ML analyses; not first-episode or not clearly separable; MDD vs healthy controls only; reviews/editorials/protocols/abstracts without analyzable data; insufficient reporting (e.g., no test-set AUC/CI/SE or missing class counts).
Clarification. (applied at full-text stage): Studies enrolling "recent-onset" or prevalent cases were included only if first-episode participants were clearly separable; otherwise, such studies were excluded from quantitative synthesis to preserve a strict first-episode-at-presentation definition.
- 4 **Step 2: Information Sources and Search Strategy**
Databases: PubMed, Scopus, Europe PMC, Semantic Scholar, OpenAlex, The Lens, medRxiv, ClinicalTrials.gov, Web of Science (coverage 2014 – 3 Nov 2025).
Core Boolean (adapted per interface/field tags):
("bipolar disorder") AND ("major depressive disorder" OR MDD OR unipolar) AND ("first episode" OR "first-episode") AND ("machine learning" OR "artificial intelligence" OR "deep learning" OR "neural network" OR SVM OR "support vector" OR "random forest" OR XGBoost OR classification).
Controlled vocabulary (e.g., MeSH) combined with Title/Abstract terms where available; free-text/topic platforms used stepwise narrowing (disorder → first episode → AI/ML).
For low-yield sources (e.g., Semantic Scholar, medRxiv, ClinicalTrials.gov), the fully

combined query was executed directly. No language or study-type filters at search. Export RIS/CSV and complete the search strings template verbatim.

5 **Step 3: Selection Process**

Title/Abstract Screening. Two independent reviewers apply eligibility criteria to all records post-deduplication.

Full-Text Review. Candidate records undergo full-text eligibility assessment.

Discrepancies. Resolved by discussion or third-reviewer arbitration.

PRISMA Flowchart. Report counts (identified 158, duplicates 39, screened 119, qualitative 17, meta-analyzed 6) and list exclusion motifs beneath the relevant nodes (full-text: topic misalignment; added/interposed pathology; no AI/ML; not first-episode; wrong task; insufficient data; meta-exclusion: insufficient reporting—missing test-set AUC/CI/SE, missing class counts, unclear validation split/potential leakage, unstated threshold basis, absent calibration).

6 **Step 4: Data Extraction**

Use a piloted form to collect: author/year/setting; modality (imaging subtype; non-imaging category); model type and feature inputs; validation (k-fold/hold-out/external) and dataset split; test-set N and BD/MDD class counts; AUC with 95% CI or SE; threshold and basis (if reported); calibration (slope, intercept, Brier score); explicit data-leakage checks. Where CIs are missing but SE/raw numbers allow, compute them; otherwise include qualitatively only. Dual extraction with consensus reconciliation.

7 **Step 5: Risk of Bias Assessment**

Use QUADAS-2 (patient selection; index test; reference standard; flow/timing) with ML-specific signaling (blinding to reference; pre-specified threshold when relevant; no tuning on test data; leakage ruled out). Complement with PROBAST/AI considerations: Participants (target similarity); Predictors (definition/blinding/missingness); Outcome (BD vs MDD definition); Analysis (events-per-predictor, internal validation, overfitting control, independent test, calibration). Two independent raters; consensus resolution. Visualize with robvis (traffic-lights and domain barplots).

8 **Step 6: Data Synthesis and Analysis**

Qualitative. Tabulate study characteristics, modalities, model classes, validation design, and outcome reporting completeness.

Quantitative (Meta-analysis). Primary measure: AUC on an independent test set.

Compute SEs from CIs or via standard approximations when feasible. Pool AUCs under random-effects with Hartung–Knapp adjustment and REML estimation of τ^2 ; use generic inverse-variance weighting. Quantify heterogeneity with τ^2 and I^2 and report 95% prediction intervals for the overall meta-analysis and any subgroup or sensitivity analysis where at least three studies are pooled. Prespecified subgroup: imaging vs non-imaging. Sensitivity analyses: exclude high-risk-of-bias or potential-leakage studies; compare external vs internal validation. Explore small-study effects when k permits; interpret cautiously for diagnostic AUCs. Threshold metrics (sensitivity/specificity) summarized narratively unless thresholds are harmonized. Calibration summarized narratively and



meta-analyzed if consistently reported. Given the anticipated small number of eligible studies, all quantitative syntheses will be interpreted cautiously, with emphasis on effect estimates and uncertainty rather than formal hypothesis testing or heterogeneity statistics.

9 **Step 7: Documentation and PRISMA compliance**

Complete PRISMA 2020 flow diagram, detailing the identification, screening, eligibility and inclusion process. Any deviations from the registered protocol identified during full-text assessment or analysis will be documented explicitly in the final report, with justification.

Expected Time:

- 10
 - Database search and deduplication: 5-7 days
 - Screening and full-text retrieval: 3-5 days
 - Data extraction and validation: 7-10 days
 - Analysis and synthesis: 5-10 days

Notes

- 11
 - Discrepancies during screening or extraction are resolved by consensus or third-reviewer adjudication.
 - If multiple models per study are eligible, prioritize the prespecified primary model for the main analysis; alternate models are explored in sensitivity analysis.
 - Studies lacking test-set AUCs but otherwise relevant remain in the qualitative synthesis only.

Deviations note

- 12 During full-text assessment, one study (Tomasik et al., 2021) was excluded from quantitative synthesis because the cohort could not be confirmed as strictly first-episode at the point of clinical presentation and first-episode participants were not separable from recent-onset/prevalent cases. This exclusion was consistent with the prespecified eligibility criterion requiring clearly separable first-episode BD vs MDD groups and was undertaken to preserve conceptual alignment with the study objective.