

Editorial

Artificial Intelligence and Bipolar Disorder: Applications of Machine Learning Models for Diagnosis, Treatment, and Outcome Prediction

Francesco Bartoli^{1,*}, Daniele Cavaleri¹, Cristina Crocamo¹¹School of Medicine and Surgery, University of Milano-Bicocca, 20900 Monza, Italy*Correspondence: francesco.bartoli@unimib.it (Francesco Bartoli)

Submitted: 2 July 2025 Revised: 29 July 2025 Accepted: 4 August 2025 Published: 13 October 2025

Keywords: artificial intelligence; bipolar disorder; machine learning; mental health; predictive learning models

Artificial Intelligence (AI) and its subfields have the potential to transform medical practice and healthcare delivery by addressing the complexities of clinical decision-making [1]. Specifically, machine learning (ML), including deep learning, is a powerful tool that leverages advanced statistical methods and computer-science techniques to analyze large datasets and identify patterns that often elude traditional statistical approaches [2]. ML may be particularly useful in psychiatry, a discipline based on the assessments of diagnostic criteria, by enhancing personalized clinical decision-making [3]. This may be particularly relevant for bipolar disorder (BD); the complex clinical presentation and management of BD [4–6] can benefit from the potential held by ML. ML techniques may integrate specific information on individual clinical features with other characteristics across different sources of data to make personalized predictions and subsequent treatment decisions [3]. Although still in the earliest stages, research has already begun to show how ML methods are effectively combining heterogeneous data stemming from genetics, electrophysiology, neuroimaging, biomarkers, speech, social media, and mobile health analyses, to improve diagnostic accuracy, identify clinical subtypes of BD, characterize drug-response profiles, and predict illness trajectories.

1. Enhancing Diagnostic Accuracy

Differentiating BD from other mental disorders remains a significant clinical challenge [7]. Indeed, BD is often misdiagnosed as major depressive disorder at onset, leading to delays in optimal treatments and poorer clinical outcomes, partly due to the inappropriate use of antidepressant monotherapy [8]. To address this issue, the implementation of ML may support the integration of clinical data to improve diagnostic accuracy. A recent systematic review and meta-analysis [9], based on findings from 18 studies, analyzed 28 ML models to show a pooled sensitivity and specificity of 0.84 and 0.82, respectively, in distinguishing BD from major depressive disorder, demonstrating the strong discriminative potential of these models. Similarly, in another systematic review including 81 studies, ML showed a high degree of accuracy in distinguishing

BD from other mental disorders, even though a high risk of publication bias was estimated [10].

2. Personalizing Treatment

Beyond diagnosis, ML may be helpful in guiding clinical decision-making in BD. Indeed, predictive models can assist in stratifying patients based on their expected response to different mood stabilizers, antipsychotics, or non-pharmacological interventions, potentially enabling individualized treatment plans. Although research in this area is still limited, recent studies have offered promising preliminary evidence. For instance, interview-based clinical data showed that the response to lithium treatment was predictable, with clinical features such as the characteristics of clinical course, age, age at onset, and sociodemographic features emerging as particularly informative [11]. Consistently, ML models, incorporating polygenic risk scores and clinical factors, appeared effective in identifying patients who are most likely to respond to lithium treatment [12]. However, findings have not been uniformly positive. For instance, ML models applied to electronic health records in the United Kingdom failed to distinguish between lithium and olanzapine responders with BD [13].

3. Predicting Clinical Outcomes

The prediction of BD clinical outcomes is likely to be the main strength of ML approaches, as this prediction is critical for an effective management of BD. ML models have demonstrated encouraging performance in this area. A systematic review [14] including 18 studies and over 30,000 participants found that ML models, based on both neuroimaging and clinical data, could predict relapses, hospitalizations, and suicide, with generally acceptable (though heterogeneous) performance metrics across studies. That review also identified key clinical predictors of negative outcomes, including early onset, BD-I subtype, comorbid substance use, and circadian-rhythm disruptions, as well as neuroimaging markers involving frontolimbic connectivity and corticostriatal-circuit abnormalities. Moreover, speech markers, identified using both natural language and signal processing from audio data streams of people with



BD, have been used to train supervised learning models to assess the feasibility of detection of depressive and manic features [15]. That supports the utility of ML in predicting mood relapses, thereby opening promising perspectives for its integration into digital tools for ecological momentary assessment in psychiatric care [15]. Finally, ML models were also developed and tested to predict mortality. A recent national-register-based cohort study showed good performance in both 2-year and 10-year mortality prediction in both Sweden ($n = 31,013$, followed-up 2006–2021) and Finland ($n = 13,956$, followed-up 1996–2018) [16].

4. Challenges and Future Directions

In view of all this, tools containing embedded ML algorithms are likely to markedly enhance decision-making in the clinical management of BD. However, methodological and practical challenges remain. Some validity concerns may be raised regarding the possible relationship between sample size and reported metrics in ML models, that somehow diverge from the expectations set by the theory of learning curves (i.e., performance typically improves or remains stable with increasing sample size) [17]. That suggests that additional factors, potentially including data quality or distribution, may have shaped observed outcomes. Therefore, further studies are needed to validate and replicate findings across different large datasets. In addition, transparent and prespecified analytic protocols are essential to minimize publication and selective-reporting bias. Moreover, the publication of studies on ML models showing low-accuracy results should be encouraged [17]. Finally, possible biases exist, including those related to missing data, misclassification, and measurement error [18]. From a clinical-utility standpoint, the complexity and limited interpretability of several ML models, which appear as “black boxes” providing limited insight into decision-making, may hinder their adoption into routine clinical practice [19]. In response to this, explainable AI (XAI), streamlining of model architecture and the implementing of post-hoc explanations, is gaining increasing attention [20]. Also, inadequate technical infrastructure and unresolved ethical issues may limit the effective adaptation to the clinical contexts in which ML algorithms are intended to be deployed [21]. Risks related to bias in algorithmic decision-making, overreliance on ML outputs, and data-privacy concerns require the adoption of core principles such as beneficence, non-maleficence, autonomy, and justice, especially because of the unique vulnerabilities of psychiatric populations [22,23].

Notwithstanding the aforementioned problems, ML might be the much sought-after evidence-based, human-centered method that finally gives the clinical characterization of BD a methodological boost and a pragmatic meaning. Future research should prioritize XAI models, promote open and reproducible methods, and foster international collaborations to analyze large, representative

datasets that may unlock the full transformative potential of ML in the management of BD. Longitudinal predictive ML models should also integrate clinical, neurobiological, genetic, behavioural, and digital phenotyping data to capture the evolving illness trajectories. Challenging the “one-size-fits-all accuracy-interpretability trade-off”, robust, standardized frameworks for ML validation, including out-of-sample assessment across diverse populations, should be established. These must encompass the entire modeling pipeline (from data preprocessing and feature selection, through model training and evaluation, to deployment), as each stage plays a critical role in ensuring its reliability, generalizability, and safe implementation in psychiatric practice.

Author Contributions

FB: Conceptualization, Writing—original draft. DC: Conceptualization, Writing—review & editing; CC: Conceptualization, Writing—review & editing. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

Not applicable.

Acknowledgment

Not applicable.

Funding

This research received no external funding.

Conflict of Interest

Francesco Bartoli serves as one of Editors-in-Chief and also as the Guest Editor of this journal. We declare that Francesco Bartoli was not involved in the editorial processing of this article. Full responsibility for the editorial process for this article was delegated to Wei Zheng.

Declaration of AI and AI-Assisted Technologies in the Writing Process

The authors used ChatGPT-4.5 to check the language accuracy and improve the readability of some sentences during the drafting of this article. All suggested changes were further reviewed and edited by the authors as suggested by the ICMJE guidelines.

References

- [1] Bajwa J, Munir U, Nori A, Williams B. Artificial intelligence in healthcare: transforming the practice of medicine. *Future Healthcare Journal*. 2021; 8: e188–e194. <https://doi.org/10.7861/fhj.2021-0095>.
- [2] Zhou Z, Wu TC, Wang B, Wang H, Tu XM, Feng C. Machine learning methods in psychiatry: a brief introduction. *General*

- Psychiatry. 2020; 33: e100171. <https://doi.org/10.1136/gpsych-2019-100171>.
- [3] Chekroud AM, Bondar J, Delgadillo J, Doherty G, Wasil A, Fokkema M, *et al.* The promise of machine learning in predicting treatment outcomes in psychiatry. *World Psychiatry*. 2021; 20: 154–170. <https://doi.org/10.1002/wps.20882>.
 - [4] Bartoli F, Malhi GS, Carrà G. Combining predominant polarity and affective spectrum concepts in bipolar disorder: towards a novel theoretical and clinical perspective. *International Journal of Bipolar Disorders*. 2024; 12: 14. <https://doi.org/10.1186/s40345-024-00336-9>.
 - [5] Bourin M. The Diagnosis of Bipolar Disorders: A Major Public Health Issue. *Alpha Psychiatry*. 2024; 25: 750–751. <https://doi.org/10.5152/alphapsychiatry.2024.241801>.
 - [6] Oliva V, Fico G, De Prisco M, Gonda X, Rosa AR, Vieta E. Bipolar disorders: an update on critical aspects. *The Lancet Regional Health. Europe*. 2024; 48: 101135. <https://doi.org/10.1016/j.lanepe.2024.101135>.
 - [7] Hirschfeld RM. Differential diagnosis of bipolar disorder and major depressive disorder. *Journal of Affective Disorders*. 2014; 169: S12–S16. [https://doi.org/10.1016/S0165-0327\(14\)70004-7](https://doi.org/10.1016/S0165-0327(14)70004-7).
 - [8] Gitlin MJ. Antidepressants in bipolar depression: an enduring controversy. *International Journal of Bipolar Disorders*. 2018; 6: 25. <https://doi.org/10.1186/s40345-018-0133-9>.
 - [9] Pan Y, Wang P, Xue B, Liu Y, Shen X, Wang S, *et al.* Machine learning for the diagnosis accuracy of bipolar disorder: a systematic review and meta-analysis. *Frontiers in Psychiatry*. 2025; 15: 1515549. <https://doi.org/10.3389/fpsyt.2024.1515549>.
 - [10] Colombo F, Calesella F, Mazza MG, Melloni EMT, Morelli MJ, Scotti GM, *et al.* Machine learning approaches for prediction of bipolar disorder based on biological, clinical and neuropsychological markers: A systematic review and meta-analysis. *Neuroscience and Biobehavioral Reviews*. 2022; 135: 104552. <https://doi.org/10.1016/j.neubiorev.2022.104552>.
 - [11] Nunes A, Ardaur R, Berghöfer A, Bocchetta A, Chillotti C, Deiana V, *et al.* Prediction of lithium response using clinical data. *Acta Psychiatrica Scandinavica*. 2020; 141: 131–141. <https://doi.org/10.1111/acps.13122>.
 - [12] Cearns M, Amare AT, Schubert KO, Thalamuthu A, Frank J, Streit F, *et al.* Using polygenic scores and clinical data for bipolar disorder patient stratification and lithium response prediction: machine learning approach. *The British Journal of Psychiatry: the Journal of Mental Science*. 2022; 220: 219–228. <https://doi.org/10.1192/bjp.2022.28>.
 - [13] Hayes JF, Ben Abdesslem F, Eloranta S, Osborn DPJ, Boman M. Predicting maintenance lithium response for bipolar disorder from electronic health records-a retrospective study. *PeerJ*. 2024; 12: e17841. <https://doi.org/10.7717/peerj.17841>.
 - [14] Amanollahi M, Jameie M, Looha MA, A Basti F, Cattarinussi G, Moghaddam HS, *et al.* Machine learning applied to the prediction of relapse, hospitalization, and suicide in bipolar disorder using neuroimaging and clinical data: A systematic review. *Journal of Affective Disorders*. 2024; 361: 778–797. <https://doi.org/10.1016/j.jad.2024.06.061>.
 - [15] Crocamo C, Cioni RM, Canestro A, Nasti C, Palpella D, Piacenti S, *et al.* Acoustic and Natural Language Markers for Bipolar Disorder: A Pilot, mHealth Cross-Sectional Study. *JMIR Formative Research*. 2025; 9: e65555. <https://doi.org/10.2196/65555>.
 - [16] Lieslehto J, Tiihonen J, Lähteenvuo M, Kautzky A, Akhtar A, Ármannsdóttir B, *et al.* Machine learning-based mortality risk assessment in first-episode bipolar disorder: a transdiagnostic external validation study. *EClinicalMedicine*. 2025; 81: 103108. <https://doi.org/10.1016/j.eclinm.2025.103108>.
 - [17] Saidi P, Dasarathy G, Berisha V. Unraveling overoptimism and publication bias in ML-driven science. *Patterns (New York, N.Y.)*. 2025; 6: 101185. <https://doi.org/10.1016/j.patter.2025.101185>.
 - [18] Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential Biases in Machine Learning Algorithms Using Electronic Health Record Data. *JAMA Internal Medicine*. 2018; 178: 1544–1547. <https://doi.org/10.1001/jamainternmed.2018.3763>.
 - [19] Hofweber T, Walker RL. Machine Learning in Health Care: Ethical Considerations Tied to Privacy, Interpretability, and Bias. *North Carolina Medical Journal*. 2024; 85: 240–245. <https://doi.org/10.18043/001c.120562>.
 - [20] Han H, Liu X. The challenges of explainable AI in biomedical data science. *BMC Bioinformatics*. 2022; 22: 443. <https://doi.org/10.1186/s12859-021-04368-1>.
 - [21] Rajkomar A, Dean J, Kohane I. Machine Learning in Medicine. *The New England Journal of Medicine*. 2019; 380: 1347–1358. <https://doi.org/10.1056/NEJMr1814259>.
 - [22] Fisher CE. The real ethical issues with AI for clinical psychiatry. *International Review of Psychiatry*. 2025; 37: 14–20. <https://doi.org/10.1080/09540261.2024.2376575>.
 - [23] Starke G, De Clercq E, Borgwardt S, Elger BS. Computing schizophrenia: ethical challenges for machine learning in psychiatry. *Psychological Medicine*. 2021; 51: 2515–2521. <https://doi.org/10.1017/S0033291720001683>.