[1]Polytechnic University of Catalonia (UPC); [2]Hospital Clínic of Barcelona, Bipolar and Depressive Disorders Unit, IDIBAPS, CIBERSAM, University of Barcelona, Barcelona, Spain and [3]Imaging of Mood- and Anxiety-Related Disorders (IMARD) group, IDIBAPS, Barcelona, Spain., Barcelona, Spain
*Corresponding author.
doi: 10.1192/j.eurpsy.2026.10415

Introduction: Cognitive distortions are systematic patterns of biased thinking strongly associated with a wide range of mental health disorders. However, most research on their automatic detection has focused on English-language corpora and relied heavily on opaque, black-box models, limiting clinical interpretability and cross-linguistic generalization.
Objectives: This study addresses a critical gap by developing and evaluating interpretable machine learning methods for detecting and classifying cognitive distortions in Spanish-language text.
Methods: A synthetic corpus of 1,509 paired distorted and alternative thoughts was generated by clinical psychologists. Using this dataset, models were trained and evaluated across three tasks: (1) binary detection of distorted thoughts, (2) detection in linguistically complex cases, and (3) multi-class classification into ten distortion categories.
Results: A Support Vector Machine with TF-IDF bigram features achieved strong performance in binary detection (F1 = 0.9287). For multi-class classification, a Logistic Regression model with part-of-speech features attained the highest accuracy (F1 = 0.8436). Extending the setup to 11 classes—including non-distorted thoughts—yielded comparable performance (F1 = 0.7782) but highlighted challenges in generalizing to the non-distorted category. Feature analyses revealed close alignment with clinical theory, identifying key lexical indicators such as absolutist adverbs and first-person language.
Conclusions: These findings demonstrate that classic, interpretable NLP approaches can provide accurate and transparent insights into distorted cognition. The study establishes a robust baseline for Spanish-language psychological text analysis and contributes to the development of automated, clinically meaningful tools for mental health support.

Disclosure of Interest: None Declared

EPP212

A pipeline for automatic bipolar disorder analysis through anonymization and LLM-as-a-judge

A. Tremopoulos[1], M. De Prisco[2], A. Mas-Musons[2], C. Valenzuela-Pascual[2], M. Korniyenko[2], V. Oliva[2], G. Fico[2], J. Raduà[3], E. Vieta[2], G. Anmella[2]*, D. Hidalgo-Mazzei[2] and C. Escolano[1]

[1]Polytechnic University of Catalonia (UPC); [2]Hospital Clínic of Barcelona, Bipolar and Depressive Disorders Unit, IDIBAPS, CIBERSAM, University of Barcelona, Barcelona, Spain and [3]Imaging of Mood- and Anxiety-Related Disorders (IMARD) group, IDIBAPS, Barcelona, Spain., Barcelona, Spain
*Corresponding author.
doi: 10.1192/j.eurpsy.2026.10414

Introduction: Bipolar Disorder (BD) is a psychiatric condition characterized by marked shifts in mood, energy, and thought processes, manifested in episodes of mania, depression, and euthymia. Despite advances in digital psychiatry, few studies have systematically assessed the potential of Large Language Models (LLMs) to analyze real clinical interviews in BD, particularly in multiple languages.
Objectives: This work presents a pipeline designed to (1) anonymize real psychiatric interviews while preserving clinical information, and (2) evaluate the capabilities of an open-source LLM, Qwen2-Audio, to classify affective states and their severity in BD.
Methods: A dataset of 41 audio recordings in Spanish and Catalan was used, covering different illness phases. A complete anonymization pipeline was developed to remove sensitive information from transcripts while extracting audio embeddings from the LLM's internal representation. These embeddings were explored in latent space to assess separability of affective states. A structured chain-of-thought (CoT) prompting strategy was then implemented, asking the model symptom-based clinical questions to emulate psychiatric reasoning. Multiple input modalities (audio, text, audio+text, and full transcripts) and prompt strategies were compared across sessions.
Results: The anonymization pipeline effectively protected patient confidentiality while enabling the use of real clinical data. Audio embeddings revealed meaningful differences between affective states within the same patient. Prompt design strongly influenced performance, with CoT prompting reducing hallucinations and improving consistency. Audio input achieved the best performance for depression (recall = 88.41%), full-text input was most effective for mania detection (F1 = 59%), while euthymia remained challenging (max F1 = 46.2%). Severity predictions were more accurate for higher severity levels (e.g., moderate depression F1 ≈ 70%, severe mania F1 = 53.2%).
Conclusions: This work demonstrates the feasibility of combining anonymization pipelines, real psychiatric data, and open-source multimodal LLMs for BD analysis. By jointly evaluating depression, mania, and euthymia, and by applying structured CoT reasoning, the study fills an important gap in the literature and provides a reproducible, privacy-preserving framework for future research in digital psychiatry.

Disclosure of Interest: None Declared

EPP213

Automated speech analysis in depressive disorder: Enhancing diagnosis and monitoring

A. Neuberg[1], M. De Prisco[2], A. Mas-Musons[2], C. Valenzuela-Pascual[2], M. Korniyenko[2], V. Oliva[2], G. Fico[2], J. Raduà[3], E. Vieta[2], D. Hidalgo-Mazzei[2], C. Escolano[1] and G. Anmella[2]*

[1]Polytechnic University of Catalonia (UPC); [2]Hospital Clínic of Barcelona, Bipolar and Depressive Disorders Unit, IDIBAPS, CIBERSAM, University of Barcelona, Barcelona, Spain and [3]Imaging of Mood- and Anxiety-Related Disorders (IMARD) group, IDIBAPS, Barcelona, Spain., Barcelona, Spain
*Corresponding author.
doi: 10.1192/j.eurpsy.2026.10413

Introduction: Depression is one of the most prevalent mental disorders and is associated with measurable alterations in speech. Automatic speech analysis offers a promising, non-invasive approach, yet progress is hindered by small, imbalanced datasets and the limited interpretability of many models.