Gac Sanit. 2026;40(S4):122

Gaceta Sanitaria

https://www.gacetasanitaria.org

239 - USE OF LANGUAGE-BASED ARTIFICIAL INTELLIGENCE (AI) MODELS FOR THE SCREENING OF DEPRESSION: A SYSTEMATIC REVIEW

J. Arias de la Torre, R. Dahl, J. Alonso, L. Botella-Juan, V. Martín, G. Martínez-Alés, G. Vilagut, A. Dregan, J.M. Valderas, et al.

King's College London; Universidad de León; CIBERESP; Hospital del Mar Medical Research Institute (IMIM); Universitat Pompeu Fabra University; Mount Sinai Hospital; Harvard TH Chan School of Public Health; National University of Singapore.

Resumen

Background/Objectives: Depressive disorders are among the most prevalent and burdensome health problems worldwide, with increasing prevalence and impact over the last decades. Advances in Artificial Intelligence (AI) models based on text and speech data have facilitated the generalisation of their use. The implementation of these models to screen for depressive disorders could potentially improve the effectiveness and efficiency of current methods, such as questionnaires or clinical interviews. In this context, we aimed to synthesise the evidence about the use of text- and speech-based AI tools for the screening and diagnosis of depression in adults.

Methods: A systematic review following the PRISMA guidelines was performed. Five databases (PubMed, Scopus, Embase, MEDLINE and APA PsycInfo) were systematically searched until August 2025 to identify relevant studies, and their quality, risk of bias, and the applicability of prediction models or algorithms used in the studies found were assessed using the updated version of the Prediction model Risk Of Bias Assessment Tool (PROBAST+AI). Data about the studies, the population in which they are focused, the datasets that they used, the models, their performance (F1 values), and their validation were extracted and narratively synthesised. The protocol of the review was registered in PROSPERO (CRD420251121073).

Results: A total of 1,114 studies were identified from which 10 were finally included in the review (n = 3,001 individuals), 6 of them focused on USA population. Six studies analysed text data using transformer models, three analysed speech data from audio through convolutional and recurrent neural networks, and one focused on both modalities. The performance of the text- and speech-based models for the screening of depression was similar to the performance of standard questionnaires and cut-off scores e.g., the 9-item version of the Patient Health Questionnaire (PHQ-9) with a cut-off point of 10+, with F1 values ranging between 0.73 and 0.90. Only four studies externally validated their findings and just two reported specific classification errors.

Conclusions/Recommendations: While further research is required, particularly using representative data from real-world settings, the evidence identified shows that language-based AI models can be used to support the screening of depressive disorders and, consequently, to prevent them and reduce their prevalence, burden, and impact.

Financiación: CIBERESP.