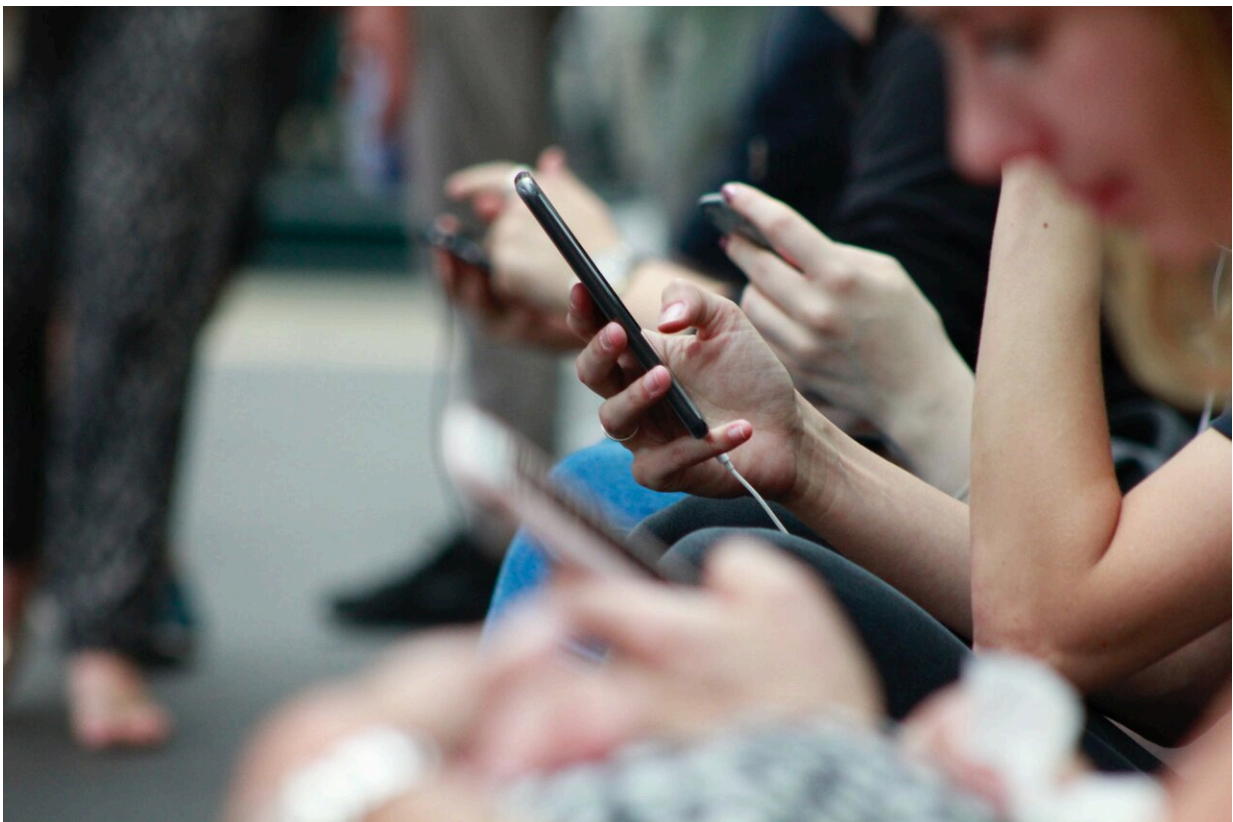


# Scientists create model to predict depression and anxiety using artificial intelligence and social media

April 10 2023, by Luciana Constantino

---



Credit: Unsplash/CC0 Public Domain

Researchers at the University of São Paulo (USP) in Brazil are using artificial intelligence (AI) and Twitter, one of the world's largest social

media platforms, to try to create anxiety and depression prediction models that could in future provide signs of these disorders before clinical diagnosis.

The study is reported in an article published in the journal *Language Resources and Evaluation*.

Construction of a database, called SetembroBR, was the first step in the study. The name is a reference to Yellow September, an annual suicide awareness and prevention campaign, and also to the fact that [data collection](#) for the study began one day in September.

The second step is still in progress but has provided some preliminary findings, such as the possibility of detecting whether a person is likely to develop depression solely on the basis of their social media friends and followers, without taking their own posts into account.

The database compiled by the group contains information relating to a corpus of texts (in Portuguese) and the network of connections involving 3,900 Twitter users who reported having been diagnosed with or treated for mental health problems before the survey. The corpus includes all public tweets posted by these users individually (without retweets), for a total of some 47 million of these short texts.

"First, we collected timelines manually, analyzing tweets by some 19,000 users, equivalent to the population of a village or small town. We then used two datasets, one for users who reported being diagnosed with a [mental health problem](#) and another selected at random for control purposes. We wanted to distinguish between people with depression and the general population," said Ivandre Paraboni, last author of the article and a professor at USP's School of Arts, Sciences and Humanities (EACH).

The study also collected tweets from friends and followers, in accordance with the observation that people with mental health problems tend to follow certain accounts, such as discussion forums, influencers and celebrities who publicly acknowledge their depression. "These people are attracted to each other. They have shared interests," said Paraboni, who is a researcher with the Center for Artificial Intelligence (C4AI).

Mental health disturbances, including depression and anxiety, are a growing global concern. The World Health Organization (WHO) estimated on the basis of 2021 data that 3.8% of the world population, or some 280 million people, were affected by depression.

WHO also estimated an increase of 25% in global prevalence of these mental health problems [during the COVID-19 pandemic](#). The tweets were collected for the study during this period.

In a recent survey by the Brazilian Health Ministry involving 784,000 participants, 11.3% said they had been diagnosed with depression. Most were women.

According to previous research, mental health problems are often reflected by the language used by the sufferers. This finding has led to a considerable number of studies involving natural language processing (NLP), with a focus on depression, anxiety and bipolar disorder, among others. However, most of these studies analyze texts in English and do not always match the profile of most Brazilians.

## Models

The researchers pre-processed the corpus to remove hashtags, URLs, emoticons and non-standard characters while maintaining the original texts. They then deployed [deep learning](#), an AI technique that teaches

computers to process data in a way inspired by the human brain, to create four text classifiers and [word embeddings](#) (context-dependent mathematical representations of relations between words) using models based on bidirectional encoder representations from transformers (BERT), a machine learning algorithm for NLP. These models correspond to a neural network that learns contexts and meanings by monitoring sequential data relationships, such as words in a sentence.

The training input consisted of a sample of 200 tweets selected at random from each user. The parameters were defined by executing cross-validation of the training data five times and calculating the average result.

The conclusion was that BERT performed best in terms of predicting depression and anxiety, with a statistically significant difference between it and LogReg, the next best option. Because the models analyzed sequences of words and complete sentences, it was possible to observe that people with depression, for example, tended to write about subjects connected to themselves, using verbs and phrases in the first person, as well as topics such as death, crisis and psychology.

"The signs of depression that can be detected during a visit to the doctor aren't necessarily the same as the ones that appear on social media," Paraboni said. "For example, use of the first-person singular pronouns I and me was very evident, and in psychology this is considered a classic sign of depression. We also observed frequent use of the heart emoji by depressive users. This is widely felt to be a symbol of affection and love, but maybe psychologists haven't yet characterized it as such."

All the collected texts were anonymized. "We published neither actual tweets nor users' names. We took care to ensure that the students involved in the project didn't have access to user data so as to protect people's identity," he said.

The researchers are now extending the database, refining their computational techniques and upgrading the models in order to see if they can produce a tool for future use in screening prospective sufferers from [mental health problems](#) and helping families and friends of young people at risk from [depression](#) and anxiety.

Brazil ranks third among the countries that most consume [social media](#) in the world, according to a Comscore survey published in early March, behind India and Indonesia but ahead of the United States, Mexico and Argentina. Its 131.5 million users are online for 46 hours a month on average. The most widely used platforms are YouTube, Facebook, Instagram, TikTok, Kwai and Twitter, which recently changed its rules and began charging for certain services.

**More information:** Wesley Ramos dos Santos et al, SetembroBR: a social media corpus for depression and anxiety disorder prediction, *Language Resources and Evaluation* (2023). [DOI: 10.1007/s10579-022-09633-0](#)

Provided by FAPESP

Citation: Scientists create model to predict depression and anxiety using artificial intelligence and social media (2023, April 10) retrieved 21 July 2023 from <https://medicalxpress.com/news/2023-04-scientists-depression-anxiety-artificial-intelligence.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--