Linguistic red flags from Facebook posts can predict future depression diagnoses

October 15 2018

Credit: CC0 Public Domain

In any given year, depression affects more than 6 percent of the adult population in the United States—some 16 million people—but fewer than half receive the treatment they need. What if an algorithm could scan social media and point to linguistic red flags of the disease before a

formal medical diagnosis had been made?

New research from the University of Pennsylvania and Stony Brook University published in the *Proceedings of the National Academy of Sciences* shows this is now more plausible than ever. Analyzing social media data shared by consenting users across the months leading up to a depression diagnosis, the researchers found their algorithm could accurately predict future depression. Indicators of the condition included mentions of hostility and loneliness, words like "tears" and "feelings," and use of more first-person pronouns like "I" and "me."

"What people write in social media and online captures an aspect of life that's very hard in medicine and research to access otherwise," says H. Andrew Schwartz, senior paper author and a principal investigator of the World Well-Being Project (WWBP). "It's a dimension that's relatively untapped compared to biophysical markers of disease. Considering conditions such as depression, anxiety, and PTSD, for example, you find more signals in the way people express themselves digitally."

For six years, the WWBP, based in Penn's Positive Psychology Center and Stony Brook's Human Language Analysis Lab, has been studying how the words people use reflect inner feelings and contentedness. In 2014, Johannes Eichstaedt, WWBP founding research scientist, started to wonder whether it was possible for social media to predict mental health outcomes, particularly for depression.

"Social media data contain markers akin to the genome," Eichstaedt explains. "With surprisingly similar methods to those used in genomics, we can comb social media data to find these markers. Depression appears to be something quite detectable in this way; it really changes people's use of social media in a way that something like skin disease or diabetes doesn't."

Eichstaedt and Schwartz teamed with colleagues Robert J. Smith, Raina Merchant, David Asch, and Lyle Ungar from the Penn Medicine Center for Digital Health for this study. Rather than do what previous studies had done—recruit participants who self-reported depression—the researchers identified data from people consenting to share Facebook statuses and electronic medical-record information, and then analyzed the statuses using machine-learning techniques to distinguish those with a formal depression diagnosis.

"This is early work from our Social Mediome Registry from the Penn Medicine Center for Digital Health," Merchant says, "which joins social media with data from health records. For this project, all individuals are consented, no data is collected from their network, the data is anonymized, and the strictest levels of privacy and security are adhered to."

Nearly 1,200 people consented to provide both digital archives. Of these, just 114 people had a diagnosis of depression in their medical records. The researchers then matched every person with a diagnosis of depression with five who did not have such a diagnosis, to act as a control, for a total sample of 683 people (excluding one for insufficient words within status updates). The idea was to create as realistic a scenario as possible to train and test the researchers' algorithm.

"This is a really hard problem," Eichstaedt says. "If 683 people present to the hospital and 15 percent of them are depressed, would our algorithm be able to predict which ones? If the algorithm says no one was depressed, it would be 85 percent accurate."

To build the algorithm, Eichstaedt, Smith, and colleagues looked back at 524,292 Facebook updates from the years leading to diagnosis for each individual with depression and for the same time span for the control. They determined the most frequently used words and phrases and then

modeled 200 topics to suss out what they called "depression-associated language markers." Finally, they compared in what manner and how frequently depressed versus control participants used such phrasing.

They learned that these markers comprised emotional, cognitive, and interpersonal processes such as hostility and loneliness, sadness and rumination, and that they could predict future depression as early as three months before first documentation of the illness in a medical record.

"There's a perception that using social media is not good for one's mental health," Schwartz says, "but it may turn out to be an important tool for diagnosing, monitoring, and eventually treating it. Here, we've shown that it can be used with clinical records, a step toward improving mental health with social media."

Eichstaedt sees long-term potential in using these data as a form of unobtrusive screening. "The hope is that one day, these screening systems can be integrated into systems of care," he says. "This tool raises yellow flags; eventually the hope is that you could directly funnel people it identifies into scalable treatment modalities."

Despite some limitations to the study, including its strictly urban sample, and limitations in the field itself—not every depression diagnosis in a medical record meets the gold standard that structured clinical interviews provide, for example—the findings offer a potential new way to uncover and get help for those suffering from depression.

  More information: Johannes C. Eichstaedt el al., "Facebook language predicts depression in medical records," *PNAS* (2018). www.pnas.org/cgi/doi/10.1073/pnas.1802331115

Provided by University of Pennsylvania