Advancement of artificial intelligence opens health data privacy to attack

December 24 2018, by John Hickey

Credit: CC0 Public Domain

Advances in artificial intelligence have created new threats to the privacy of health data, a new UC Berkeley study shows.

The study, led by professor Anil Aswani of the Industrial Engineering & Operations Research Department (IEOR) in the College of Engineering and his team, suggests current laws and regulations are nowhere near

sufficient to keep an individual's health status private in the face of AI development. The research was released today on *JAMA Network Open*.

In the work, which was funded in part by UC Berkeley's Center for Long-Term Cybersecurity, Aswani shows that by using artificial intelligence, it is possible to identify individuals by learning daily patterns in step data (like that collected by activity trackers, smartwatches and smartphones) and correlating it to demographic data. The mining of two years' worth of data covering more than 15,000 Americans led to the conclusion that the privacy standards associated with 1996's HIPAA (Health Insurance Portability and Accountability Act) legislation need to be revisited and reworked.

"We wanted to use NHANES (the National Health and Nutrition Examination Survey) to look at privacy questions because this data is representative of the diverse population in the U.S.," Aswani says. "The results point out a major problem. If you strip all the identifying information, it doesn't protect you as much as you'd think. Someone else can come back and put it all back together if they have the right kind of information."

"In principle, you could imagine Facebook gathering step data from the app on your smartphone, then buying health care data from another company and matching the two," he explains. "Now they would have health care data that's matched to names, and they could either start selling advertising based on that or they could sell the data to others."

Aswani makes it clear that the problem isn't with the devices, but with how the information the devices capture can be misused and potentially sold on the open market.

"I'm not saying we should abandon these devices," he says. "But we need to be very careful about how we are using this data. We need to protect

the information. If we can do that, it's a net positive."

Though the study specifically looked at step data, Aswani says the results suggest a broader threat to the privacy of health data. "HIPAA regulations make your health care private, but they don't cover as much as you think," he says. "Many groups, like tech companies, are not covered by HIPAA, and only very specific pieces of information are not allowed to be shared by current HIPAA rules. There are companies buying health data. It's supposed to be anonymous data, but their whole business model is to find a way to attach names to this data and sell it."

Aswani says he is worried that as advances in AI make it easier for companies to gain access to health data, the temptation for companies to use it in illegal or unethical ways will increase. Employers, mortgage lenders, credit card companies and others could potentially use AI to discriminate based on pregnancy or disability status, for instance.

"Ideally, what I'd like to see from this are new regulations or rules that protect health data," he says. "But there is actually a big push to even weaken the regulations right now. For instance, the rule-making group for HIPAA has requested comments on increasing data sharing. The risk is that if people are not aware of what's happening, the rules we have will be weakened. And the fact is the risks of us losing control of our privacy when it comes to health care are actually increasing and not decreasing."

More information: Liangyuan Na et al. Feasibility of Reidentifying Individuals in Large National Physical Activity Data Sets From Which Protected Health Information Has Been Removed With Use of Machine Learning, *JAMA Network Open* (2018). DOI: 10.1001/jamanetworkopen.2018.6040