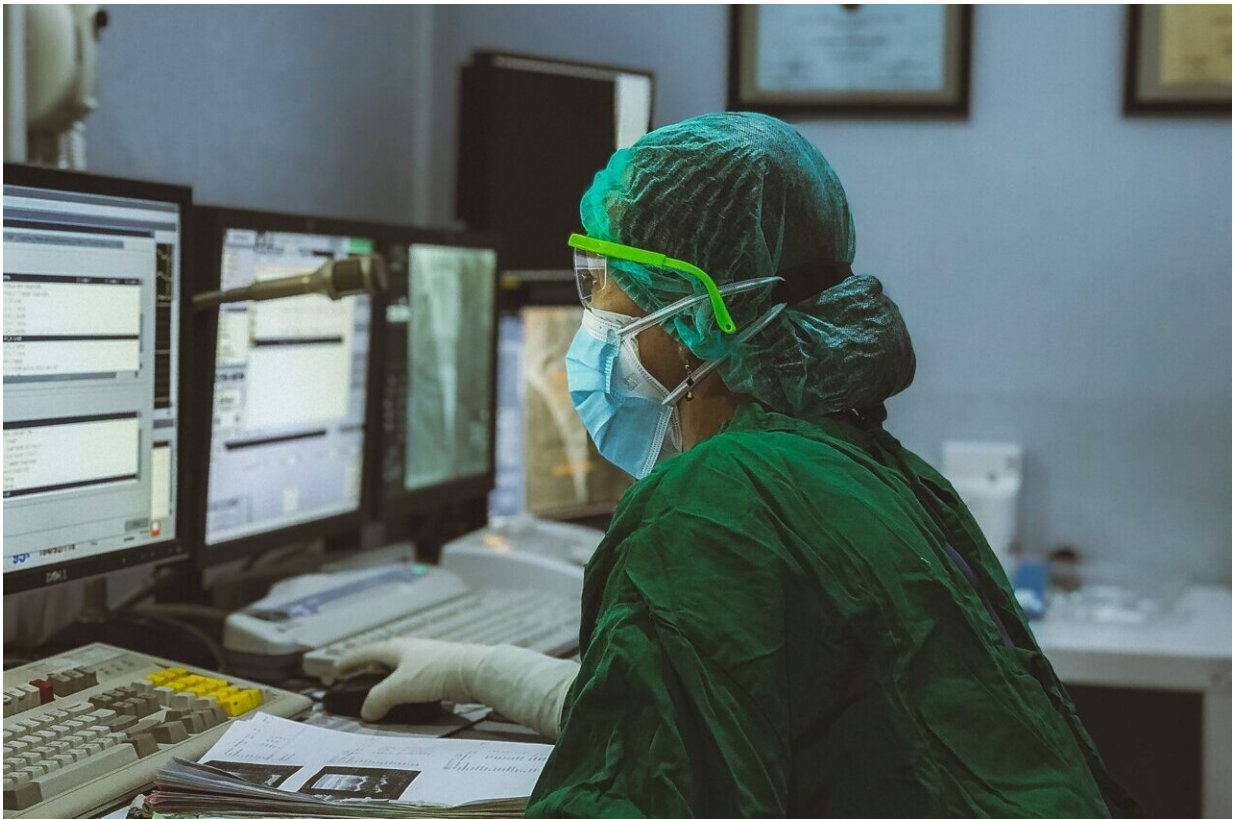# Hidden biases in medical data could compromise AI approaches to healthcare

February 2 2022, by Steve Nadis

Credit: Unsplash/CC0 Public Domain

While working toward her dissertation in computer science at MIT, Marzyeh Ghassemi wrote several papers on how machine-learning techniques from artificial intelligence could be applied to clinical data in

order to predict patient outcomes. "It wasn't until the end of my Ph.D. work that one of my committee members asked: "Did you ever check to see how well your model worked across different groups of people?'"

That question was eye-opening for Ghassemi, who had previously assessed the performance of models in aggregate, across all patients. Upon a closer look, she saw that models often worked differently—specifically worse—for populations including Black women, a revelation that took her by surprise. "I hadn't made the connection beforehand that [health disparities](#) would translate directly to model disparities," she says. "And given that I am a visible minority woman-identifying computer scientist at MIT, I am reasonably certain that many others weren't aware of this either."

In a paper published Jan. 14 in the journal *Patterns*, Ghassemi—who earned her doctorate in 2017 and is now an assistant professor in the Department of Electrical Engineering and Computer Science and the MIT Institute for Medical Engineering and Science (IMES)—and her coauthor, Elaine Okanyene Nsoesie of Boston University, offer a cautionary note about the prospects for AI in medicine. "If used carefully, this technology could improve performance in health care and potentially reduce inequities," Ghassemi says. "But if we're not actually careful, technology could worsen care."

It all comes down to data, given that the AI tools in question train themselves by processing and analyzing vast quantities of data. But the data they are given are produced by humans, who are fallible and whose judgments may be clouded by the fact that they interact differently with patients depending on their age, gender, and race, without even knowing it.

Furthermore, there is still great uncertainty about medical conditions themselves. "Doctors trained at the same medical school for 10 years

can, and often do, disagree about a patient's diagnosis," Ghassemi says. That's different from the applications where existing machine-learning algorithms excel—like object-recognition tasks—because practically everyone in the world will agree that a dog is, in fact, a dog.

Machine-learning algorithms have also fared well in mastering games like chess and Go, where both the rules and the "win conditions" are clearly defined. Physicians, however, don't always concur on the rules for treating patients, and even the win condition of being "healthy" is not widely agreed upon. "Doctors know what it means to be sick," Ghassemi explains, "and we have the most data for people when they are sickest. But we don't get much data from people when they are healthy because they're less likely to see doctors then."

Even mechanical devices can contribute to flawed data and disparities in treatment. Pulse oximeters, for example, which have been calibrated predominately on light-skinned individuals, do not accurately measure blood oxygen levels for people with darker skin. And these deficiencies are most acute when oxygen levels are low—precisely when accurate readings are most urgent. Similarly, women face increased risks during "metal-on-metal" hip replacements, Ghassemi and Nsoesie write, "due in part to anatomic differences that aren't taken into account in implant design." Facts like these could be buried within the data fed to computer models whose output will be undermined as a result.

Coming from computers, the product of [machine-learning algorithms](#) offers "the sheen of objectivity," according to Ghassemi. But that can be deceptive and dangerous, because it's harder to ferret out the faulty data supplied en masse to a computer than it is to discount the recommendations of a single possibly inept (and maybe even racist) doctor. "The problem is not machine learning itself," she insists. "It's people. Human caregivers generate bad data sometimes because they are not perfect."

Nevertheless, she still believes that machine learning can offer benefits in health care in terms of more efficient and fairer recommendations and practices. One key to realizing the promise of machine learning in health care is to improve the quality of data, which is no easy task. "Imagine if we could take data from doctors that have the best performance and share that with other doctors that have less training and experience," Ghassemi says. "We really need to collect this data and audit it."

The challenge here is that the collection of data is not incentivized or rewarded, she notes. "It's not easy to get a grant for that, or ask students to spend time on it. And data providers might say, "Why should I give my data out for free when I can sell it to a company for millions?" But researchers should be able to access data without having to deal with questions like: "What paper will I get my name on in exchange for giving you access to data that sits at my institution?"

"The only way to get better health care is to get better data," Ghassemi says, "and the only way to get better data is to incentivize its release."

It's not only a question of collecting data. There's also the matter of who will collect it and vet it. Ghassemi recommends assembling diverse groups of researchers—clinicians, statisticians, medical ethicists, and computer scientists—to first gather diverse patient data and then "focus on developing fair and equitable improvements in health care that can be deployed in not just one advanced medical setting, but in a wide range of medical settings."

The objective of the *Patterns* paper is not to discourage technologists from bringing their expertise in machine learning to the medical world, she says. "They just need to be cognizant of the gaps that appear in treatment and other complexities that ought to be considered before giving their stamp of approval to a particular computer model."

This story is republished courtesy of MIT News ([web.mit.edu/newsoffice/](#)), a popular site that covers news about MIT research, innovation and teaching.

Provided by Massachusetts Institute of Technology