Medical Xpress
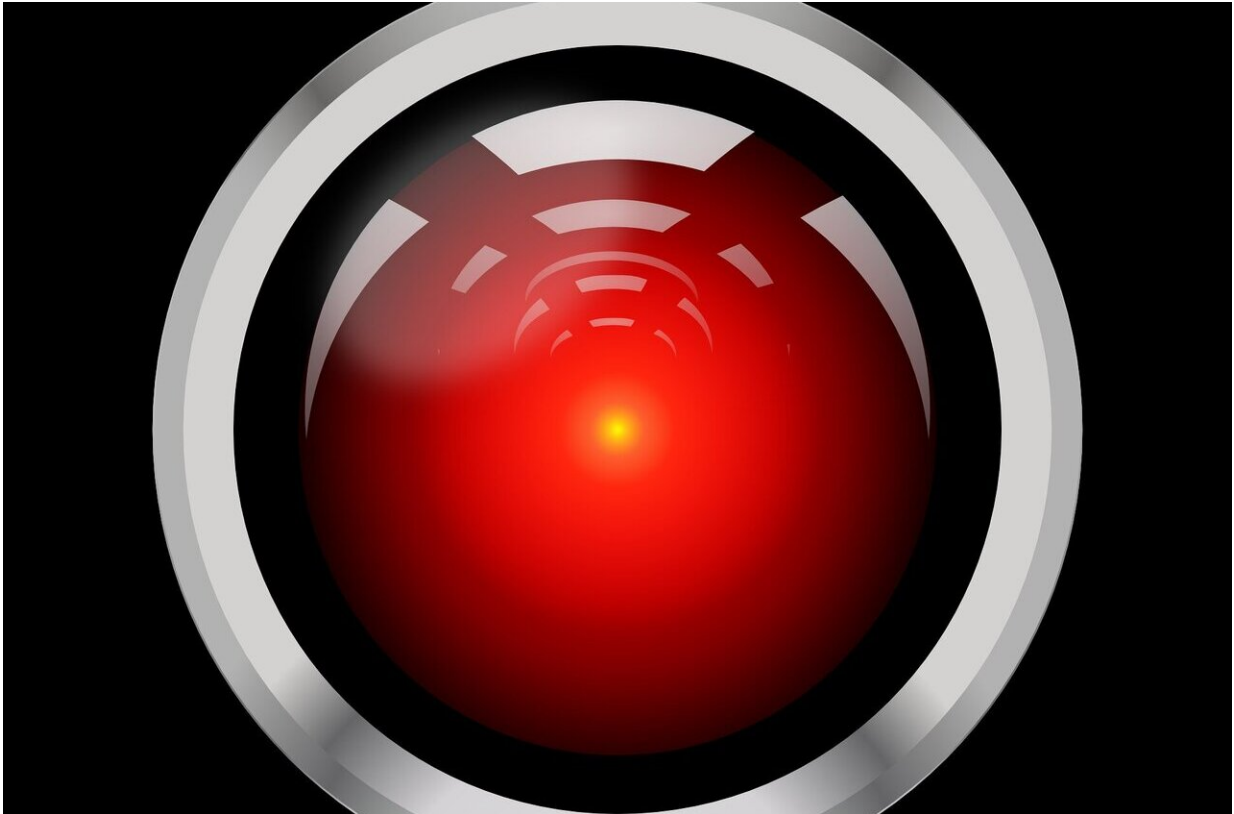
The geographic bias in medical AI tools

September 23 2020, by Shana Lynch

Credit: Pixabay/CC0 Public Domain

Just a few decades ago, scientists didn't think much about diversity when studying new medications. Most clinical trials enrolled mainly white men living near urban research institutes, with the assumption that any findings would apply equally to the rest of the country. Later research demonstrated that assumption to be false; examples accumulated of

medications that were later determined to be less effective or caused more side effects in populations that were underrepresented in the initial study.

To address these inequities, federal requirements for participation in medical research were broadened in the 1990s, and clinical trials now attempt to enroll diverse populations from the onset of the study.

But we are now at risk of repeating these same mistakes as we develop new technologies, such as AI. Researchers from Stanford University examined clinical applications of machine learning to find that most algorithms are trained on datasets from patients in only three geographic areas, and that the majority of states have no represented patients whatsoever.

"AI algorithms should mirror the community," says Amit Kaushal, an attending physician at VA Palo Alto Hospital and Stanford adjunct professor of bioengineering. "If we're building AI-based tools for patients across the United States, as a field, we can't have the data to train these tools all coming from the same handful of places."

Kaushal, along with Russ Altman, a Stanford professor of bioengineering, genetics, medicine, and biomedical data science, and Curt Langlotz, a professor of radiology and biomedical informatics research, examined five years of peer-reviewed articles that trained a deep-learning algorithm for a diagnostic task intended to assist with patient care. Among U.S. studies where geographic origin could be characterized, they found the majority (71%) used patient data from California, Massachusetts, or New York to train the algorithms. Some 60% solely relied on these three locales. Thirty-four states were not represented at all, while the other 13 states contributed limited data.

The research didn't expose bad outcomes from AI trained on the

geographies, but raised questions about the validity of the algorithms for patients in other areas. "We need to understand the impact of these biases and whether considerable investments should be made to remove them," says Altman, associate director of the Stanford Institute for Human-Centered Artificial Intelligence.

"Geography correlates to a zillion things relative to health," Altman says. "It correlates to lifestyle and what you eat and the diet you are exposed to; it can correlate to weather exposure and other exposures depending on if you live in an area with fracking or high EPA levels of toxic chemicals—all of that is correlated with geography."

If these datasets were used for an [algorithm](#) to diagnose patients across the United States, "you could be doing actual harm to the people not included in the sample."

Limited data also means limited vision. "The data you have available impacts the problems you can study in the first place," Kaushal says. "If I only have access to data from California, Massachusetts, and New York, I can build algorithms to help people in those places. But problems that are more common in other geographies won't even be on my radar."

The takeaways from this study: Larger and more diverse datasets are needed for the development of innovative AI algorithms. "Stanford has led the way in making diagnostic datasets freely available for science—more than any other center by far," says Langlotz, director of the Stanford Center for Artificial Intelligence in Medicine and Imaging. "But it's expensive and it's not enough. Resources are needed to help centers across the country contribute to more diverse training datasets."

The public also should be skeptical when medical AI systems are developed from narrow training datasets. And regulators must scrutinize the training methods for these new machine learning systems.

"Medicine has been down this road before—early clinical trials didn't think much about gender, racial, or geographic diversity and we are still working to address that oversight," Kaushal says. "As AI is set to enter clinical medicine, we shouldn't have to wait 30, 40 years to make all the same mistakes and fix them again. We should see where this is headed and address it upfront."

Provided by Stanford University