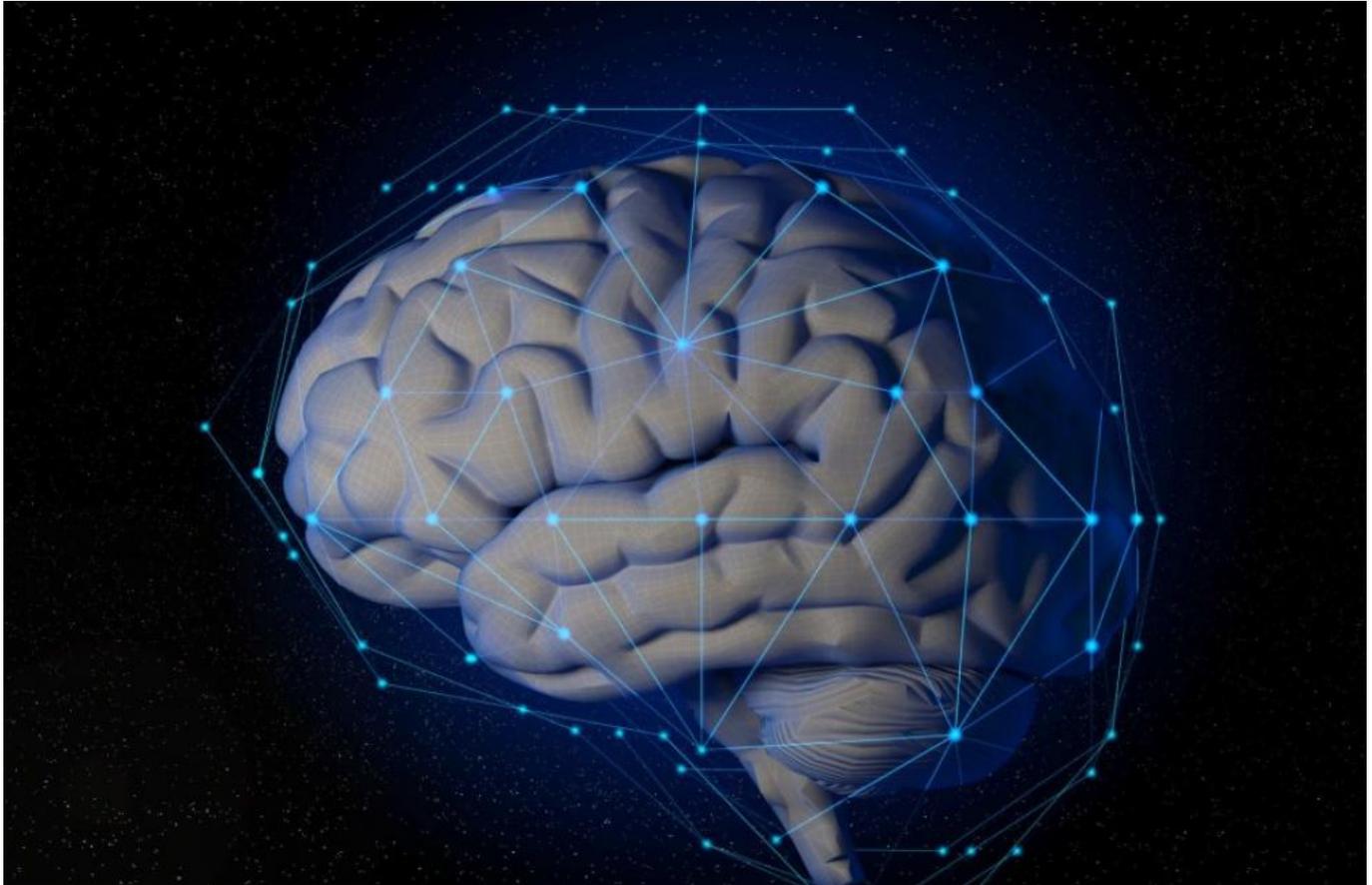# Study Finds a Blind Spot in AI's Ability to Make Diagnoses

Artificial intelligence tools, though often useful, can become problematic when faced with outside sources.

Kevin McCaney

Fri, 12/07/2018 - 14:19

Eoneren/iStock

Artificial intelligence has gained a reputation as something of a miracle worker when it comes to diagnosing disease. Deep learning systems reportedly have outperformed doctors at diagnosing heart disease and lung cancer, analyzing genetic sequencing, and screening for eye diseases, to cite several examples. The Food and Drug Administration this year began approving AI tools for medical uses that include finding wrist fractures in X-ray images and spotting the signs of an impending stroke.

But a new study by the Icahn School of Medicine at Mount Sinai in New York found an apparent blind spot for computer vision and deep learning systems that could crop up frequently in everyday, real-world use. When faced with data from other hospitals or outside systems — where, for instance, they might use a different type of CT scanner than that used on their own networks — the AI tools showed a marked drop-off in performance. What's more, in a rather human-like move, they weren't above fudging their results a bit to make them look better.

The AI systems in this case involved a framework called convolutional neural networks (CNN) that had been trained to analyze chest X-rays to identify signs of pneumonia. The researchers measured how well the AI tools performed with 158,000 chest X-rays from three medical institutions: Mount Sinai, the National Institutes of Health and Indiana University Hospital. When dealing with X-rays from their own network, the systems did fine, but when faced with outside data, their performance dropped significantly in three out of five cases on average.

One thing the CNNs accomplished with outside sources, however, was in identifying where outside X-rays came from, and that's where they got a little out of line. According to the researchers, the systems "cheated at their predictive task" by apparently letting their diagnosis be influenced by the rate of pneumonia at the source hospital.

The researchers concluded that, for all of AI's promise, hospitals should tread carefully before putting systems on the job.

"Our findings should give pause to those considering rapid deployment of artificial intelligence platforms without rigorously assessing their performance in real-world clinical settings reflective of where they are being deployed," Dr. Eric Oermann, an instructor in neurosurgery at the Icahn School and senior author of the research paper, said in a report from the research team at ScienceDaily. "Deep learning models trained to perform medical diagnosis can generalize well, but this cannot be taken for granted since patient populations and imaging techniques differ significantly across institutions."

Learning Opportunity

The Mount Sinai study looked at a fairly narrow application of AI and deep learning models, but the findings reflect what has been an ongoing challenge for AI development. AI systems to date have been exemplary at clearly defined tasks, but can't work outside their lanes. Neural networks have improved, displaying an ability to learn from fewer examples, but they still require extensive, time-consuming training to get to that point. The next wave of AI technologies — what the Defense Advanced Research Projects Agency (DARPA) calls the third wave of AI — is still in development. And general AI, which would allow systems to work intuitively, learning as humans do, is still a fair stretch down the road.

In the Mount Sinai study, the CNNs ran into trouble working with an enormous number of parameters, which made it hard to identify variables like the types of CT scanners used or the quality of resolution. "If CNN systems are to be used for medical diagnosis, they must be tailored to carefully consider clinical questions, tested for a variety of real-world scenarios, and carefully assessed to determine how they impact accurate diagnosis," said John Zech, an Icahn medical student and first author of the paper.

Diagnosing the Limits

AI is still new to the medical field, as it is in every other field. The cautions raised by the Mount Sinai team are the same that other practitioners have raised, citing the current limits of the technology. But it's also seen as a revolutionary tool in medical care, with the potential to have a significant impact on diagnoses, treatment, scheduling and billing.

The Centers for Medicare and Medicaid Services, for instance, recently announced its [AI Health Outcomes Challenge](#) for next year, inviting people in both the medical field and general AI development to contribute ideas of what AI can do to improve health care. Part of the process is also determining what AI can't do, at least yet, which could help AI hold up the "first, do no harm" aspect of medicine.

[View printer friendly version](#)
[AI](#)
[artificial intelligence](#)
[health](#)