RADIOLOGISTS PRONE TO MISTAKES BECAUSE OF AI

BY SIMON SPICHAK

Brain CT scan.
Photo source: Freepik; author: DC Studio

AI was supposed to help radiologists screen for cancer more accurately. Instead AI is making them doubt their own expertise and inadvertently harming their patients.

Imagine you're screened for cancer at a state-of-the-art institute using the latest and most powerful imaging techniques. You've read about the clinic online, rather than just getting a radiologists' opinion, the diagnosis is made all the more certain as with the help of artificial intelligence (AI). Even though the radiologist doesn't pick up anything strange on the scan, the AI alerts them of a suspicious-looking part of the scan — convincing the radiologist to order further testing. You might feel relieved that the AI was able to spot something the radiologist couldn't - but then it turns out the AI is wrong.

Research recently published in the Radiological Society of North America's peer-reviewed scientific journal Radiology suggests that relying on AI to help make these clinical decisions sways radiologists across different levels of expertise toward making the wrong call, leading toward expensive and unnecessary testing.

Worse at assessing mammograms

The study was conducted by a group of researchers and radiologists in Germany and the Netherlands, co-led by Thomas Dratsch and Xue Chen.

The researchers presented 27 radiologists with 50 mammograms, using the Breast Imaging Reporting and Data System (BI-RADS) assessment in conjunction with an AI assistant.

Rather than having two radiologists look at the image, this AI can act

as the second pair of eyes. But, when the AI assistant provided an incorrect BI-RADS assessment, radiologists were much more likely to second-guess themselves, dropping their assessment accuracy from 80 percent to less than 20 percent.

Because these radiologists considered the AI assistant as a trusted source, they were more likely to defer to its judgment, even when it was wrong. Even the most experienced radiologists in the study were prone to this cognitive error, called automation bias.

Accept a false conclusion?

"Even if the radiologists have excellent medical training, and they are looking at an image, and they know by their reasoning that this is a negative exam, but the AI flags it as positive that they have this introduced doubt," Jordan Perchik, MD, assistant professor at the department of diagnostic radiology at the University of Alabama at Birmingham, who was not involved in this study, tells Digest in a personal interview.

"They might trust the algorithm over their clinical reasoning and accept a false conclusion over their own true conclusion."

Computer-aided detection for decades

As AI is rolled out across more settings in radiology, clinicians are in a race to figure out the best ways to integrate these tools safely to minimize harm to patients. "We've had computer-aided detection and mammography for decades now, and that was marketed as something that would be this big leap forward in patient safety and diagnostic accuracy and physician workflow," Perchik says. "But it turns out that this particular tool had an overall very low accuracy and very high level of false positives."

What's the harm?

A false positive test disrupts the workflow, making the clinician waste extra time looking at the patient's case. The clinician has to rush through the rest of their cases, leading to expensive, unnecessary testing and anxiety for the patient.

"The problem goes beyond how we clinicians use these AI assistants; there are also built-in algorithmic biases that can embed racial biases and lead to poorer health for women and non-white patients."

AI underdiagnoses women, low socio-economic status and non-white patients

A landmark 2021 study in Nature Medicine led by Laleh Seyyed-Kalantari at the University of Toronto found that AI used to read chest X-rays would underdiagnose Black and Hispanic patients, as well as women and those with a lower socioeconomic background, compared to white patients.

Judy Wawira Gichoya, MD, MS, assistant professor in the Department of Radiology and Imaging Sciences at Emory University, who studies algorithmic fairness and biases built into these AI tools, tells Digest why this should be a key concern.

Gichoya explains that the AI model underdiagnosed the patients who already had the worst health outcomes and least access to care. These outcomes were encoded in the data scientists used to train the AI, so it learned a shortcut. Since these patients were less likely to access timely care even with a positive X-ray, the model would consider them healthy.

In an international setting, this suggests that AI bias could make health outcomes worst for the groups of people who have least access to healthcare. "What we need to do is come up with better frameworks for real-world monitoring," Gichoya says. Rather than focusing all the energy on developing the algorithm, clinicians must ensure that the outcomes are fair — meaning patients aren't underdiagnosed because of their race, sex, or socioeconomic status.

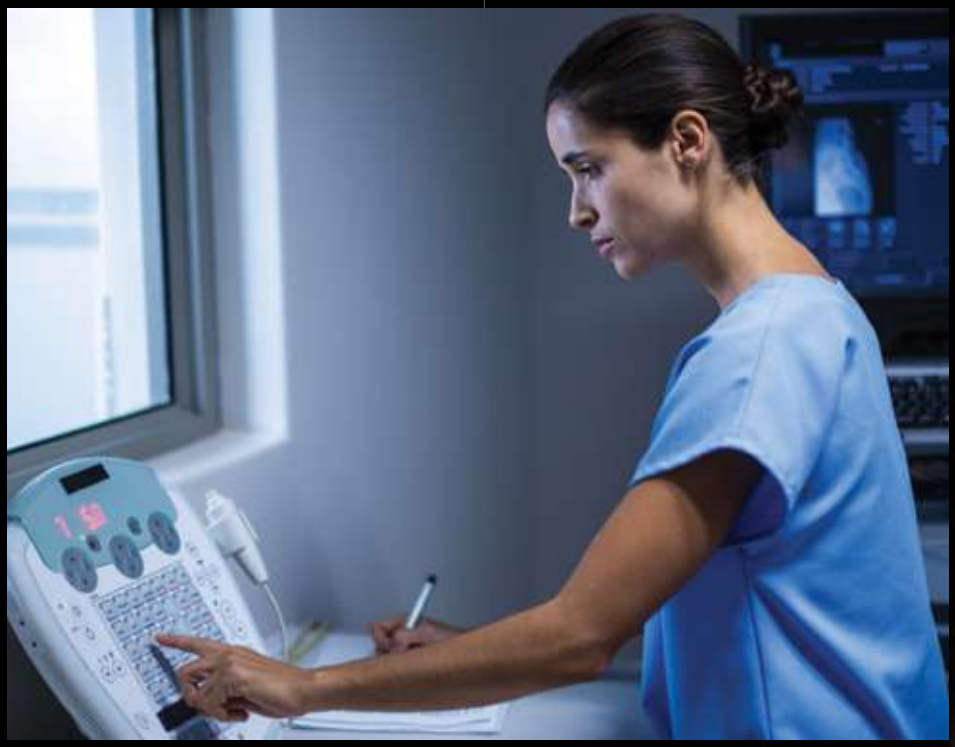
Judy Wawira Gichoya, MD, MS, an assistant professor of interventional radiology and informatics at the department of radiology and imaging sciences in Emory University, cautions that AI and author automated systems could worsen outcomes for underdiagnosed populations due to built in racial and algorithmic biases. Credit: Emory University

INTEGRATING AI INTO HEALTHCARE

BY SIMON SPICHAK

J. Raymond Geis, MD, a senior scientist at the American College of Radiology's Data Science Institute, believes that healthcare systems aren't yet ready for widespread adoption of these tools.

"Today, almost all clinical AI tools in medical imaging are installed and used on an ad hoc basis, with no or minimal formal monitoring or quality controls," Geis says. "To me, this is the greatest challenge: what skills do we need, what processes do we need, and how do we start to build the framework to run cascades of ensembles of clinical AI tools that receive and deliver information among themselves and humans."



Radiologist conducting a scan.
Photo source: Freepik; author: @wawebreakmedia_micro

Problems for patients may arise when an AI introduces a false negative. "For something like mammography, where giving the wrong answer means a patient has undiagnosed cancer that has potentially major consequences, the need to incorporate human+machine intelligence accurately is extremely important," Geis continuous. "Like anything, overreliance on computer programs when we don't know how well they work is not safe, and it is dangerous to trust those systems."

Using AI responsibly

Learning to use and integrate AI responsibly into certain parts of the radiology workflow is a far cry from the disruptive game changer that prominent computer scientists like Geoffrey Hinton believed AI would become. In 2016, at the Machine Learning and Market for Intelligence Conference in Hinton, he boldly proclaimed that these algorithms will replace radiologists altogether, comparing them to "the coyote already over the edge of the cliff who hasn't yet looked down." He also proclaimed it made no sense to continue training radiologists.

Instead, AI algorithms are a new type of tool, unlike X-ray machines and other diagnostics. The software has built-in algorithmic biases, and the technology is prone to sway experienced radiologists to the wrong medical decision. The problems can be addressed by training medical school students and tracking patient outcomes.

J. Raymond Geis, MD, senior scientist at the American College of Radiology's Data Science Institute, says that overreliance on automated systems like AI can be dangerous without understanding the underlying limitations. Credit: Photo courtesy of J. Raymond Geis.

RESPONSIBLE AI 101: AUTOMATION AND ALGORITHMIC BIAS

BY SIMON SPICHAK

Dr. Jordan Perchik, MD, is assistant professor at the Department of Diagnostic Radiology at the University of Alabama at Birmingham. Since 2020, Perchik has been running a week-long AI Literacy course to make AI practical and understandable. He spoke with Digest about automation, algorithmic bias, and the challenges of integrating AI into clinical practice.

Patient undergoing a scan.
Photo source: Unsplash; author: Accuray

What exactly is automation bias?

Perchik:
When you have an algorithm deployed in clinical practice, someone who's using that algorithm, even if they have excellent medical training and know by their reasoning that this is a negative exam, if the AI flags it as positive, they have introduced doubt. They might trust the algorithm over their clinical reasoning and accept a false over their true conclusion.

How does that affect the workflow?

Perchik:
When AI would flag these things, they would add another step for the radiologist: instead of looking at a mammogram, seeing something negative and saying, Okay, this was a negative study. I can sign this off in two minutes. Now, they have to go back and scrutinize these areas marked as false positives by the AI. And the two-minute exam becomes a five-minute exam. And you also have this extra cognitive load and the feeling of doubt that, in addition to your diagnostic process, you're having to disprove the AI.

How big of a risk is automation bias?

Perchik:
As we bring in more applications, it will be complicated because the vast majority of radiologists in practice, even those using AI currently, have yet to have formal AI training and may not be aware of this bias. That puts us at risk.

What about algorithm bias?

Perchik:
That's an area that I am very passionate about and something that is not discussed enough. Many of these, like computer vision and image-related algorithms, are prone to minor imaging differences.

If an AI for pulmonary scans was trained entirely on patients from the northeast US, and now encounters cases from the southeast US, its performance worsens. The AI algorithm starts to see a lot of pulmonary nodules in the southeast and might begin to classify almost everyone down there as having cancer. It will not realize that the nodules are common and caused by endemic fungi.

What exactly does that mean for the patients?

Perchik:
These are the kinds of things that are within the spectrum of normal. If the algorithm is trained on one area, it doesn't see the other side of normal. This causes patient harm and cost thousands or tens of thousands of dollars in additional patient workup.

What kinds of diagnostic or screening applications is AI currently used for?

Perchik:
The AIs are used for flagging things like pulmonary nodules and breast masses. So, it's very limited right now. We had our AI boom about one or about two years ago, and we still are having, you know, dozens and dozens of new applications getting FDA clearance, but they're going for these same targets. High-volume, high-acuity studies for targeted cancer applications have not been as highly researched.

What should radiologists know about integrating AI? What do you tell new clinicians and trainees?

Perchik:
The most important thing is fundamental knowledge of how the AI works. This means understanding what type of data the AI uses and how well this was trained. Was it challenged with a lot of external data? Or was the AI developed in a single institution? Those kinds of single-institution tools are the most brittle and might perform worse if challenged with new types of data. Beyond that, knowing how to roll AI out in a regimented, meticulous way is important. This isn't like buying a new X-ray machine; it won't work right out of the box. You have to test it. You have to have feedback; you have to have a close relationship with the vendor. You have to be more meticulous, and you have to monitor its performance over time.

Do you have any final words of advice?

Perchik:
AI is not something that's going to replace your clinical judgment. It is something that can supplement your clinical judgment or help you to prioritize and work faster. It's an AI assistant. It's not an AI replacer.

"Clinicians need to understand the way that these AI algorithms work": Jordan Perchik, MD, assistant professor of abdominal imaging and informatics. Credit: University of Alabama at Birmingham.