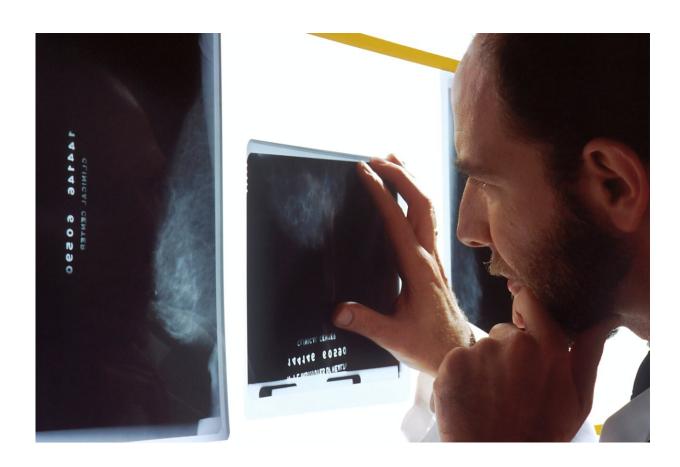# Does AI help or hurt human radiologists' performance? It depends on the doctor

March 19 2024

Credit: Unsplash/CC0 Public Domain

One of the most touted promises of medical artificial intelligence tools is their ability to augment human clinicians' performance by helping them interpret images such as X-rays and CT scans with greater precision to

make more accurate diagnoses.

But the benefits of using AI tools on image interpretation appear to vary from clinician to clinician, according to new research led by investigators at Harvard Medical School, working with colleagues at MIT and Stanford.

The study findings suggest that individual clinician differences shape the interaction between human and machine in critical ways that researchers do not yet fully understand. The analysis, published in *Nature Medicine*, is based on data from an earlier working paper by the same research group, released by the National Bureau of Economic Research.

In some instances, the research showed, use of AI can interfere with a radiologist's performance and interfere with the accuracy of their interpretation.

"We find that different radiologists, indeed, react differently to AI assistance—some are helped while others are hurt by it," said co-senior author Pranav Rajpurkar, assistant professor of biomedical informatics in the Blavatnik Institute at HMS.

"What this means is that we should not look at radiologists as a uniform population and consider just the 'average' effect of AI on their performance," he said. "To maximize benefits and minimize harm, we need to personalize assistive AI systems."

The findings underscore the importance of carefully calibrated implementation of AI into clinical practice, but they should in no way discourage the adoption of AI in radiologists' offices and clinics, the researchers said.

Instead, the results should signal the need to better understand how

humans and AI interact and to design carefully calibrated approaches that boost human performance rather than hurt it.

"Clinicians have different levels of expertise, experience, and decision-making styles, so ensuring that AI reflects this diversity is critical for targeted implementation," said Feiyang "Kathy" Yu, who conducted the work while at the Rajpurkar lab with co-first author on the paper with Alex Moehring at the MIT Sloan School of Management.

"Individual factors and variation would be key in ensuring that AI advances rather than interferes with performance and, ultimately, with diagnosis," Yu said.

AI tools affected different radiologists differently

While previous research has shown that AI assistants can indeed boost radiologists' diagnostic performance, these studies have looked at radiologists as a whole without accounting for variability from radiologist to radiologist.

In contrast, the new study looks at how individual clinician factors—area of specialty, years of practice, prior use of AI tools—come into play in human-AI collaboration.

The researchers examined how AI tools affected the performance of 140 radiologists on 15 X-ray diagnostic tasks—how reliably the radiologists were able to spot telltale features on an image and make an accurate diagnosis. The analysis involved 324 patient cases with 15 pathologies—abnormal conditions captured on X-rays of the chest.

To determine how AI affected doctors' ability to spot and correctly identify problems, the researchers used advanced computational methods that captured the magnitude of change in performance when using AI

and when not using it.

The effect of AI assistance was inconsistent and varied across radiologists, with the performance of some radiologists improving with AI and worsening in others.

AI tools influenced human performance unpredictably

AI's effects on human radiologists' performance varied in often surprising ways.

For instance, contrary to what the researchers expected, factors such as how many years of experience a radiologist had; whether they specialized in thoracic, or chest, radiology; and whether they'd used AI readers before did not reliably predict how an AI tool would affect a doctor's performance.

Another finding that challenged the prevailing wisdom: Clinicians who had low performance at baseline did not benefit consistently from AI assistance. Some benefited more, some less, and some none at all. Overall, however, lower-performing radiologists at baseline had lower performance with or without AI. The same was true among radiologists who performed better at baseline. They performed consistently well, overall, with or without AI.

Then came a not-so-surprising finding: More accurate AI tools boosted radiologists' performance, while poorly performing AI tools diminished the diagnostic accuracy of human clinicians.

While the analysis was not done in a way that allowed researchers to determine why this happened, the finding points to the importance of

testing and validating AI tool performance before clinical deployment, the researchers said. Such pre-testing could ensure that inferior AI doesn't interfere with human clinicians' performance, and therefore, patient care.

What do these findings mean for the future of AI in the clinic?

The researchers cautioned that their findings do not provide an explanation for why and how AI tools seem to affect performance across human clinicians differently, but note that understanding why would be critical to ensuring that AI radiology tools augment human performance rather than hurt it.

To that end, the team noted, AI developers should work with physicians who use their tools to understand and define the precise factors that come into play in the human-AI interaction.

The researchers added that the radiologist-AI interaction should be tested in experimental settings that mimic real-world scenarios and reflect the actual patient population for which the tools are designed.

Apart from improving the accuracy of the AI tools, it's also important to train radiologists to detect inaccurate AI predictions and to question an AI tool's diagnostic call, the research team said. To achieve that, AI developers should ensure that they design AI models that can "explain" their decisions.

"Our research reveals the nuanced and complex nature of machine-human interaction," said study co-senior author Nikhil Agarwal, professor of economics at MIT. "It highlights the need to understand the multitude of factors involved in this interplay and how they influence the

ultimate diagnosis and care of patients."

Additional authors included Oishi Banerjee at HMS and Tobias Salz at MIT, who was co-senior author on the paper.

More information: Heterogeneity and predictors of the effects of AI assistance on radiologists, *Nature Medicine* (2024). DOI: 10.1038/s41591-024-02850-w. www.nature.com/articles/s41591-024-02850-w

Provided by Harvard Medical School