AI matched, outperformed radiologists in screening X-rays for certain diseases

21 November 2018, by Hanae Armitage

Credit: CC0 Public Domain

In a matter of seconds, a new algorithm read chest X-rays for 14 pathologies, performing as well as radiologists in most cases, a Stanford-led study says.

A new artificial intelligence algorithm can reliably screen chest X-rays for more than a dozen types of disease, and it does so in less time than it takes to read this sentence, according to a new study led by Stanford University researchers.

The algorithm, dubbed CheXNeXt, is the first to simultaneously evaluate X-rays for a multitude of possible maladies and return results that are consistent with the readings of radiologists, the study says.

Scientists trained the algorithm to detect 14 different pathologies: For 10 diseases, the algorithm performed just as well as radiologists; for three, it underperformed compared with radiologists; and for one, the algorithm outdid the experts.

"Usually, we see AI algorithms that can detect a

brain hemorrhage or a wrist fracture—a very narrow scope for single-use cases," said Matthew Lungren, MD, MPH, assistant professor of radiology. "But here we're talking about 14 different pathologies analyzed simultaneously, and it's all through one algorithm."

The goal, Lungren said, is to eventually leverage these algorithms to reliably and quickly scan a wide range of image-based medical exams for signs of disease without the backup of professional radiologists. And while that may sound disconcerting, the technology could eventually serve as high-quality digital "consultations" to resource-deprived regions of the world that wouldn't otherwise have access to a radiologist's expertise. Likewise, there's an important role for AI in fully developed health care systems too, Lungren added. Algorithms like CheXNeXt could one day expedite care, empowering primary care doctors to make informed decisions about X-ray diagnostics faster, without having to wait for a radiologist.

"We're seeking opportunities to get our algorithm trained and validated in a variety of settings to explore both its strengths and blind spots," said graduate student Pranav Rajpurkar. "The algorithm has evaluated over 100,000 X-rays so far, but now we want to know how well it would do if we showed it a million X-rays—and not just from one hospital, but from hospitals around the world."

A paper detailing the findings of the study was published online Nov. 20 in *PLOS Medicine*. Lungren and Andrew Ng, Ph.D., adjunct professor of computer science at Stanford, share senior authorship. Rajpurkar and fellow graduate student Jeremy Irvin are the lead authors.

Practice makes perfect

Lungren and Ng's diagnostic algorithm has been in development for more than a year. It builds on their work on a previous iteration of the technology that

could outperform radiologists when diagnosing pneumonia from a chest X-ray. Now, they've boosted the abilities of the algorithm to flag 14 ailments, including masses, enlarged hearts and collapsed lungs. For 11 of the 14 pathologies, the algorithm made diagnoses with the accuracy of radiologists or better.

Back in the summer of 2017, the National Institutes of Health released a set of hundreds of thousands of X-rays. Since then, there's been a mad dash for computer scientists and radiologists working in artificial intelligence to deliver the best possible algorithm for chest X-ray diagnostics.

The scientists used about 112,000 X-rays to train the algorithm. A panel of three radiologists then reviewed a different set of 420 X-rays, one by one, for the 14 pathologies. Their conclusions served as a "ground truth"— a diagnosis that experts agree is the most accurate assessment—for each scan. This set would eventually be used to test how well the algorithm had learned the telltale signs of disease in an X-ray. It also allowed the team of researchers to see how well the algorithm performed compared to the radiologists.

"We treated the algorithm like it was a student; the NIH data set was the material we used to teach the student, and the 420 images were like the final exam," Lungren said. To further evaluate the performance of the algorithm compared with human experts, the scientists asked an additional nine radiologists from multiple institutions to also take the same "final exam."

"That's another factor that elevates this research," Lungren said. "We weren't just comparing this against other algorithms out there; we were comparing this model against practicing radiologists."

What's more, to read all 420 X-rays, the radiologists took about three hours on average, while the algorithm scanned and diagnosed all pathologies in about 90 seconds.

Next stop: the clinic

Now, Lungren said, his team is working on a

subsequent version of CheXNeXt that will bring the researchers even closer to in-clinic testing. The algorithm isn't ready for that just yet, but Lungren hopes that it will eventually help expedite the X-ray-reading process for doctors diagnosing urgent care or emergency patients who come in with a cough.

"I could see this working in a few ways. The algorithm could triage the X-rays, sorting them into prioritized categories for doctors to review, like normal, abnormal or emergent," Lungren said. Or the algorithm could sit bedside with primary care doctors for on-demand consultation, he said. In this case, Lungren said, the algorithm could step in to help confirm or cast doubt on a diagnosis. For example, if a patient's physical exam and lab results were consistent with pneumonia, and the algorithm diagnosed pneumonia on the patient's X-ray, then that's a pretty high-confidence diagnosis and the physician could provide care right away for the condition. Importantly, in this scenario, there would be no need to wait for a radiologist. But if the algorithm came up with a different diagnosis, the primary care doctor could take a closer look at the X-ray or consult with a radiologist to make the final call.

"We should be building AI algorithms to be as good or better than the gold standard of human, expert physicians. Now, I'm not expecting AI to replace radiologists any time soon, but we are not truly pushing the limits of this technology if we're just aiming to enhance existing radiologist workflows," Lungren said. "Instead, we need to be thinking about how far we can push these AI models to improve the lives of patients anywhere in the world."

Provided by Stanford University Medical Center