Radiologists outperformed AI in identifying lung diseases on chest X-ray
In a study of more than 2,000 chest X-rays, radiologists outperformed AI in accurately identifying the presence and absence of three common lung diseases, according to a study published in Radiology, a journal of the Radiological Society of North America (RSNA).

"Chest radiography is a common diagnostic tool, but significant training and experience is required to interpret exams correctly," said lead researcher Louis L. Plesner, M.D., resident radiologist and Ph.D. fellow in the Department of Radiology at Herlev and Gentofte Hospital in Copenhagen, Denmark.

While commercially available and FDA-approved AI tools are available to assist radiologists, Dr. Plesner said the clinical use of deep-learning-based AI tools for radiological diagnosis is in its infancy.
"While AI tools are increasingly being approved for use in radiological departments, there is an unmet need to further test them in real-life clinical scenarios," Dr. Plesner said. "AI tools can assist radiologists in interpreting chest X-rays, but their real-life diagnostic accuracy remains unclear."

Dr. Plesner and a team of researchers compared the performance of four commercially available AI tools with a pool of 72 radiologists in interpreting 2,040 consecutive adult chest X-rays taken over a two-year period at four Danish hospitals in 2020. The median age of the patient group was 72 years. Of the sample chest X-rays, 669 (32.8%) had at least one target finding.

The chest X-rays were assessed for three common findings: airspace disease (a chest X-ray pattern, for example, caused by pneumonia or lung edema), pneumothorax (collapsed lung) and pleural effusion (a buildup of water around the lungs).

AI tools achieved sensitivity rates ranging from 72 to 91% for airspace disease, 63 to 90% for pneumothorax, and 62 to 95% for pleural effusion.

"The AI tools showed moderate to a high sensitivity comparable to radiologists for detecting airspace disease, pneumothorax and pleural effusion on chest X-rays," he said. "However, they produced more false-positive results (predicting disease when none was present) than the radiologists, and their performance decreased when multiple findings were present and for smaller targets."

For pneumothorax, positive predictive values—the probability that patients with a positive screening test truly have the disease—for the AI systems ranged between 56 and 86%, compared to 96% for the radiologists.
"AI performed worst at identifying airspace disease, with positive predictive values ranging between 40 and 50%," Dr. Plesner said. "In this difficult and elderly patient sample, the AI predicted airspace disease where none was present five to six out of 10 times. You cannot have an AI system working on its own at that rate."

According to Dr. Plesner, the goal of radiologists is to balance the ability of finding and excluding disease, avoiding both significant overlooked diseases and overdiagnosis.

"AI systems seem very good at finding disease, but they aren't as good as radiologists at identifying the absence of disease especially when the chest X-rays are complex" he said. "Too many false-positive diagnoses would result in unnecessary imaging, radiation exposure and increased costs."

Dr. Plesner said most studies generally tend to evaluate the ability of AI to determine the presence or absence of a single disease, which is a much easier task than real-life scenarios where patients often present with multiple diseases.

"In many prior studies claiming AI superiority over radiologists, the radiologists reviewed only the image without access to the patient's clinical history and previous imaging studies," he said.

"In everyday practice, a radiologist's interpretation of an imaging exam is a synthesis of these three data points. We speculate that the next generation of AI tools could become significantly more powerful if capable of this synthesis as well, but no such systems exist yet."

"Our study demonstrates that radiologists generally outperform AI in real-life scenarios where there is a wide variety of patients," he said. "While an AI system is effective at identifying normal chest X-rays, AI
should not be autonomous for making diagnoses."

Dr. Plesner noted that these AI tools could boost radiologists' confidence in their diagnoses by providing a second look at chest X-rays.


Provided by Radiological Society of North America