RSNA Press Release

Artificial Intelligence Shows Potential for Triaging Chest X-rays

Released: January 22, 2019

At A Glance

- Researchers trained an artificial intelligence system to interpret and prioritize abnormal chest X-rays with critical findings.
- Chest X-rays account for 40 percent of all diagnostic imaging worldwide and create significant backlogs at health care facilities.
- The AI system distinguished abnormal from normal X-rays with high accuracy.

OAK BROOK, Ill. — An artificial intelligence (AI) system can interpret and prioritize abnormal chest X-rays with critical findings, potentially reducing the backlog of exams and bringing urgently needed care to patients more quickly, according to a study appearing in the journal *Radiology*.

download full-size photo

Giovanni Montana, Ph.D.

Chest X-rays account for 40 percent of all diagnostic imaging worldwide. The number of exams can create significant backlogs at health care facilities. In the U.K. there are an estimated 330,000 X-rays at any given time that have been waiting more than 30 days for a report.

"Currently there are no systematic and automated ways to triage chest X-rays and bring those with critical and urgent findings to the top of the reporting pile," said study co-author Giovanni Montana, Ph.D., formerly of King's College London in London and currently at the University of Warwick in Coventry, England.

Deep learning (DL), a type of AI capable of being trained to recognize subtle patterns in medical images, has been proposed as an automated means to reduce this backlog and identify exams that merit immediate attention, particularly in publicly-funded health care systems.

For the study, Professor Montana and colleagues used 470,388 adult chest X-rays to develop an AI system that could identify key findings. The images had been stripped of any identifying information to protect patient privacy. The radiologic reports were pre-processed using Natural Language Processing (NLP), an important algorithm of the AI system that extracts labels from written text. For each X-ray, the researchers' in-house system required a list of labels indicating which specific abnormalities were visible on the image.

"The NLP goes well beyond pattern matching," Dr. Montana said. "It uses AI techniques to infer the structure of each written sentence; for instance, it identifies the presence of clinical findings and body locations and their relationships. The development of the NLP system for labeling chest X-rays at scale was a critical milestone in our study."

The NLP analyzed the radiologic report to prioritize each image as critical, urgent, non-urgent or normal. An AI system for computer vision was then trained using labeled X-ray images to predict the clinical priority from appearances only. The researchers tested the system's performance for prioritization in a simulation using an independent set of 15,887 images.

The AI system distinguished abnormal from normal chest X-rays with high accuracy. Simulations showed that critical findings received an expert radiologist opinion in 2.7 days, on average, with the AI approach—significantly sooner than the 11.2-day average for actual practice.

"The initial results reported here are exciting as they demonstrate that an AI system can be successfully trained using a very large database of routinely acquired radiologic data," Dr. Montana said. "With further clinical validation, this technology is expected to reduce a radiologist's workload by a significant amount by detecting all the normal exams so more time can be spent on those requiring more attention."

The researchers plan to expand their research to a much larger sample size and deploy more complex algorithms for better performance. Future research goals include a multi-center study to prospectively assess the performance of the triaging software.

"A major milestone for this research will consist in the automated generation of sentences describing the radiologic abnormalities seen in the images," Dr. Montana said. "This seems an achievable objective given the current AI technology."

"Automated Triaging of Adult Chest Radiographs with Deep Artificial Neural Networks." Collaborating with Dr. Montana were Mauro Annarumma, M.Sc., Samuel J. Withey, M.Res., Robert J. Bakewell, M.Res., Emanuele Pesce, M.Sc., and Vicky Goh, M.D., FRCR.

Radiology is edited by David A. Bluemke, M.D., Ph.D., University of Wisconsin School of Medicine and Public Health, Madison, Wis., and owned and published by the Radiological Society of North America, Inc. (*http://radiology.rsna.org/*)

RSNA is an association of over 53,400 radiologists, radiation oncologists, medical physicists and related scientists, promoting excellence in patient care and health care delivery through education, research and technologic innovation. The Society is based in Oak Brook, Ill. (*RSNA.org*)

For patient-friendly information on chest X-rays, visit *RadiologyInfo.org*.
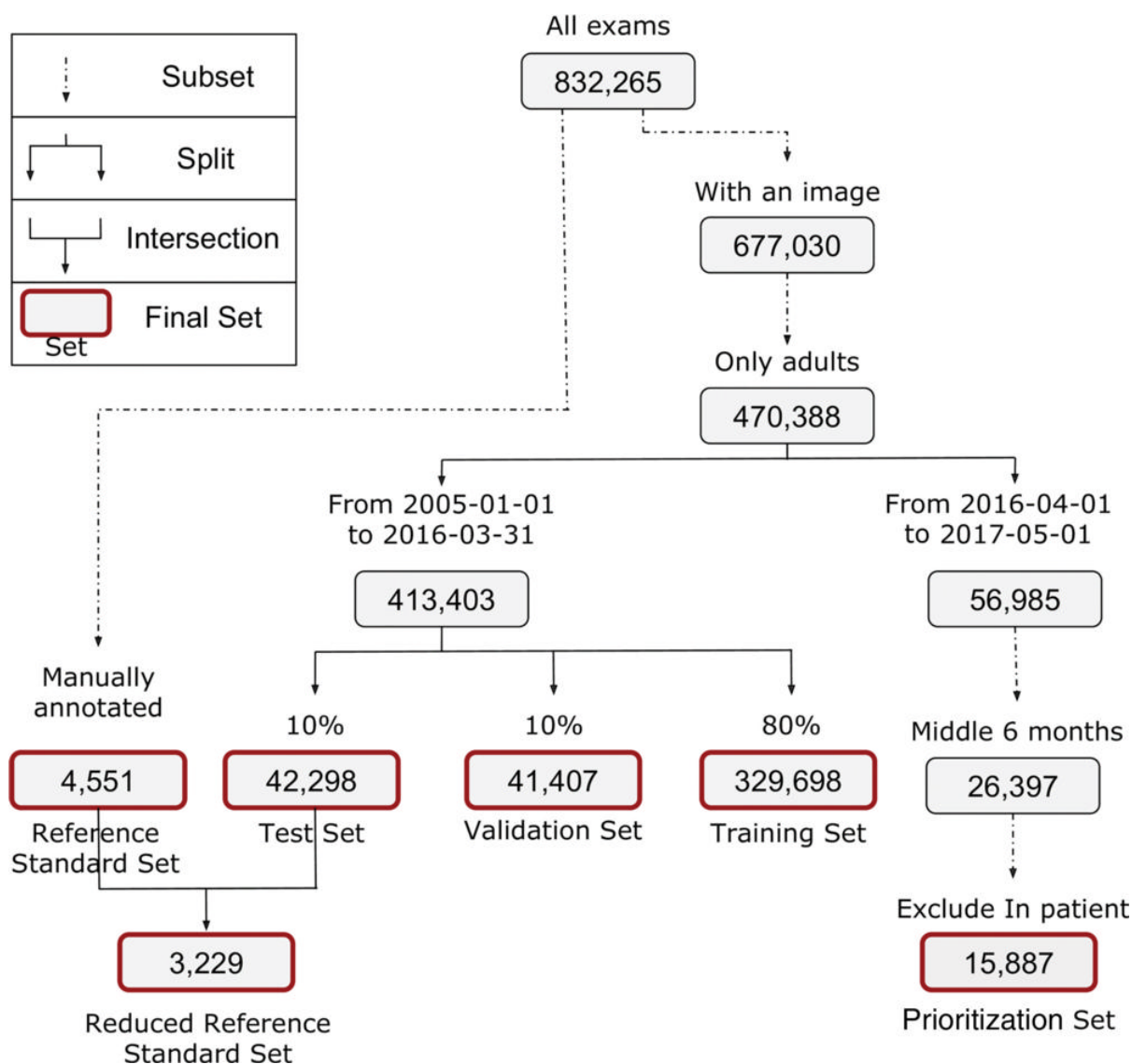
Images (JPG, TIF):

Figure 1. Flowchart shows different data sets used for training, learning, and testing. Approximately 8 percent of radiographs were critical, 40 percent urgent, 26 percent nonurgent, and 26 percent normal across the training, test, and validation data sets.
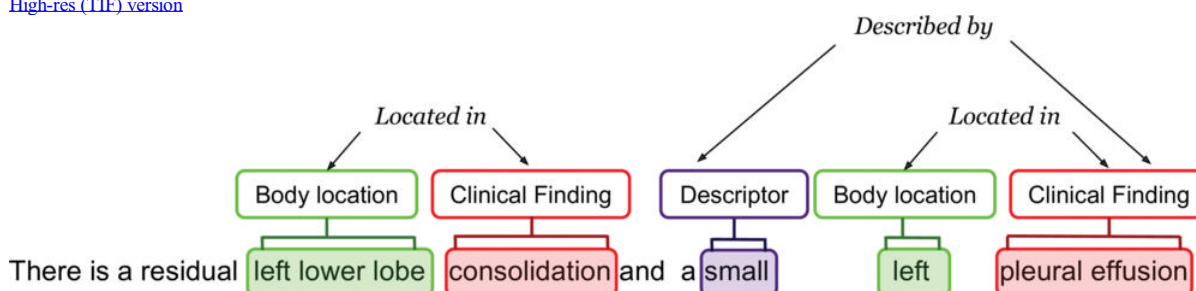
High-res (TIF) version

Figure 2. Example of radiologic report annotated by natural language processing system. "Entities" are highlighted with different colors, one for each semantic class. Arrows represent relationships between entities. Final annotation extracted by the rule-based system was "airspace opacification; pleural effusion/abnormality."
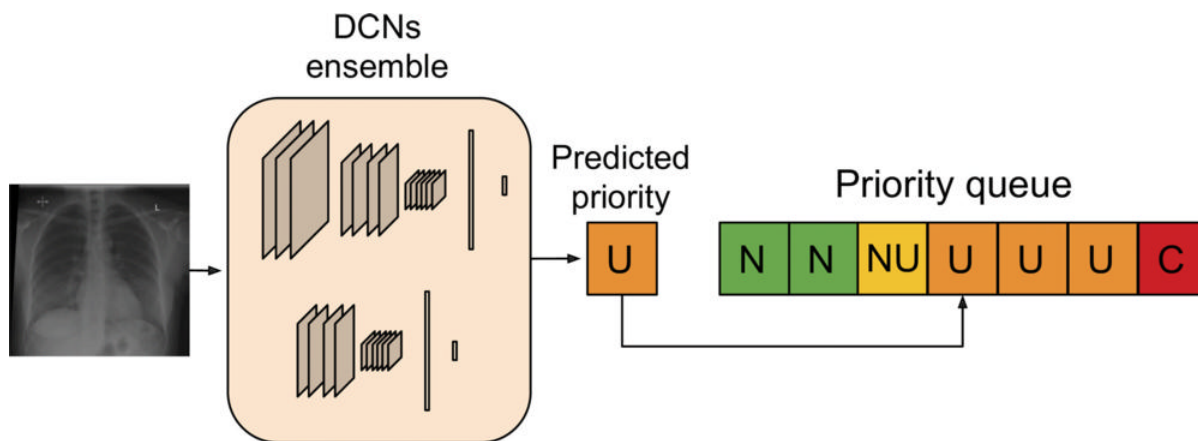
High-res (TIF) version

Figure 3. Artificial intelligence prioritization system: When a chest radiograph is acquired, the deep learning architecture (consisting of two different deep convolutional neural networks [DCNs] operating at different input sizes) processes the image in real time and predicts its priority level (urgent, in this example). Given the predicted priority, the image is then automatically inserted in a dynamic priority-based reporting queue. C = critical, N = normal, NU = nonurgent, U = urgent.
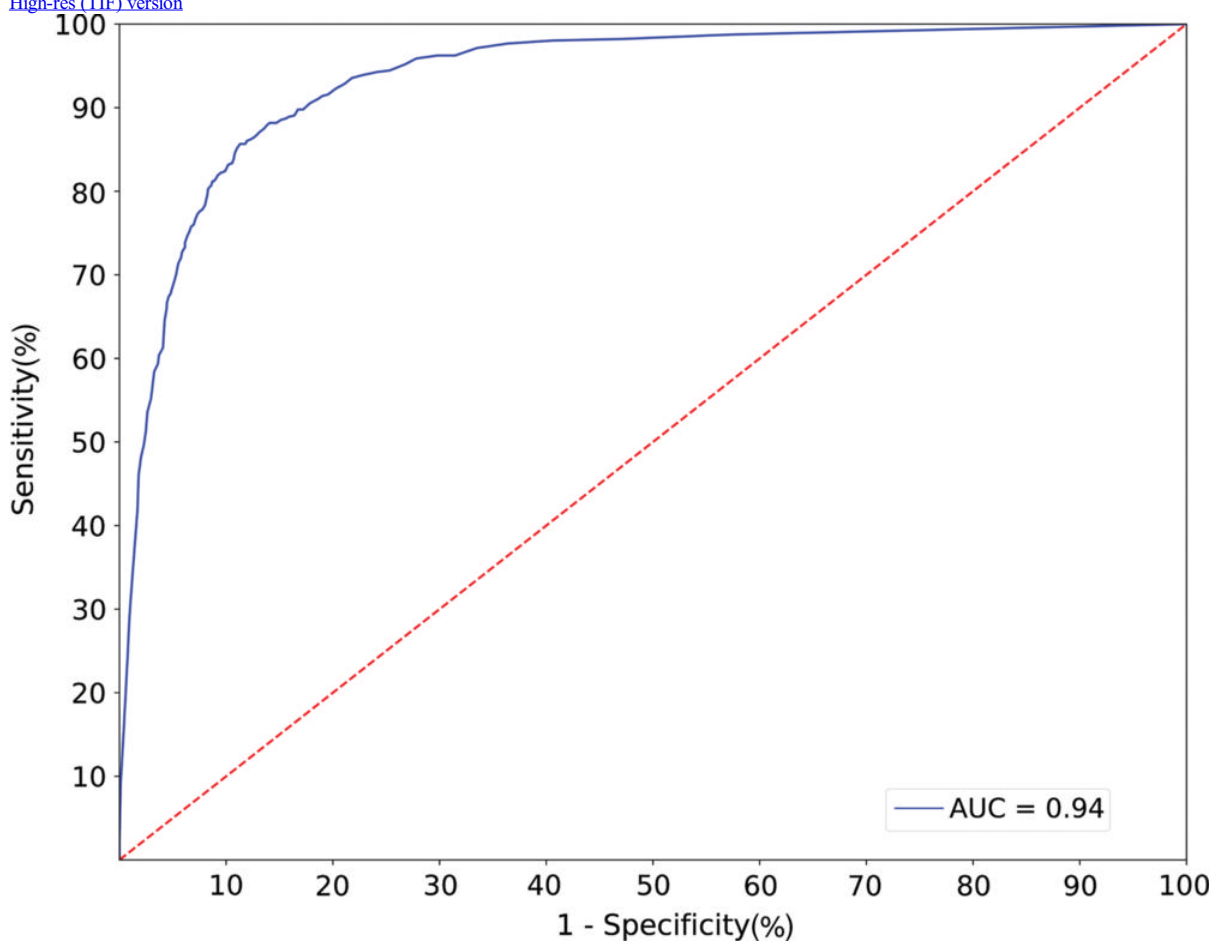
High-res (TIF) version

Figure 4. Receiver operating characteristic curve for normality prediction obtained by the artificial intelligence system. The system achieved an area under the receiver operating characteristic curve (AUC) of 0.94.
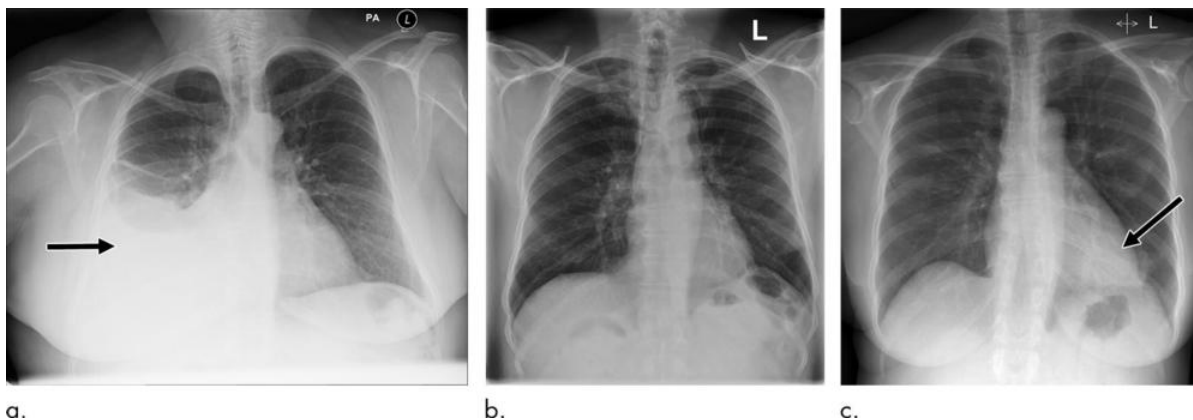
High-res (TIF) version

Figure 5. Examples of correctly and incorrectly prioritized radiographs. (a) Radiograph was reported as showing large right pleural effusion (arrow). This was correctly prioritized as urgent. (b) Radiograph reported as showing "lucency at the left apex suspicious for pneumothorax." This was prioritized as normal. On review by three independent radiologists, the radiograph was unanimously considered to be normal. (c) Radiograph reported as showing consolidation projected behind heart (arrow). The finding was missed by the artificial intelligence system, and the study was incorrectly prioritized as normal.
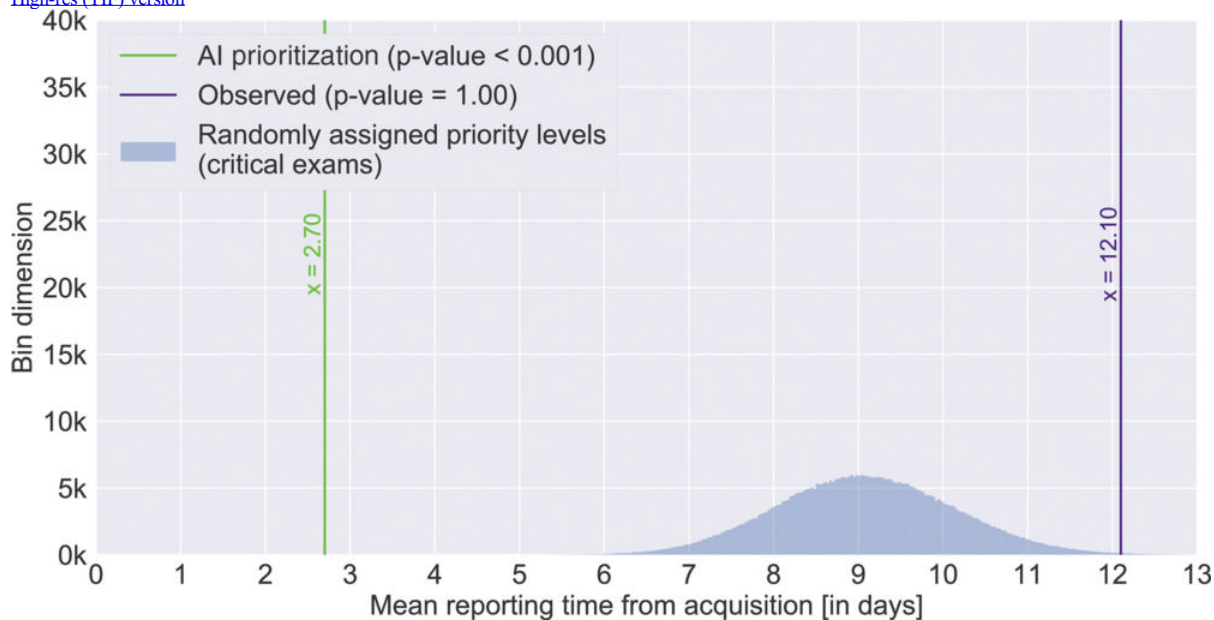
High-res (TIF) version

Figure 6. Mean reporting time from acquisition with AI prioritization system compared with observed mean for critical radiographs.

High-res (TIF) version

Resources:

Editorial
Study abstract