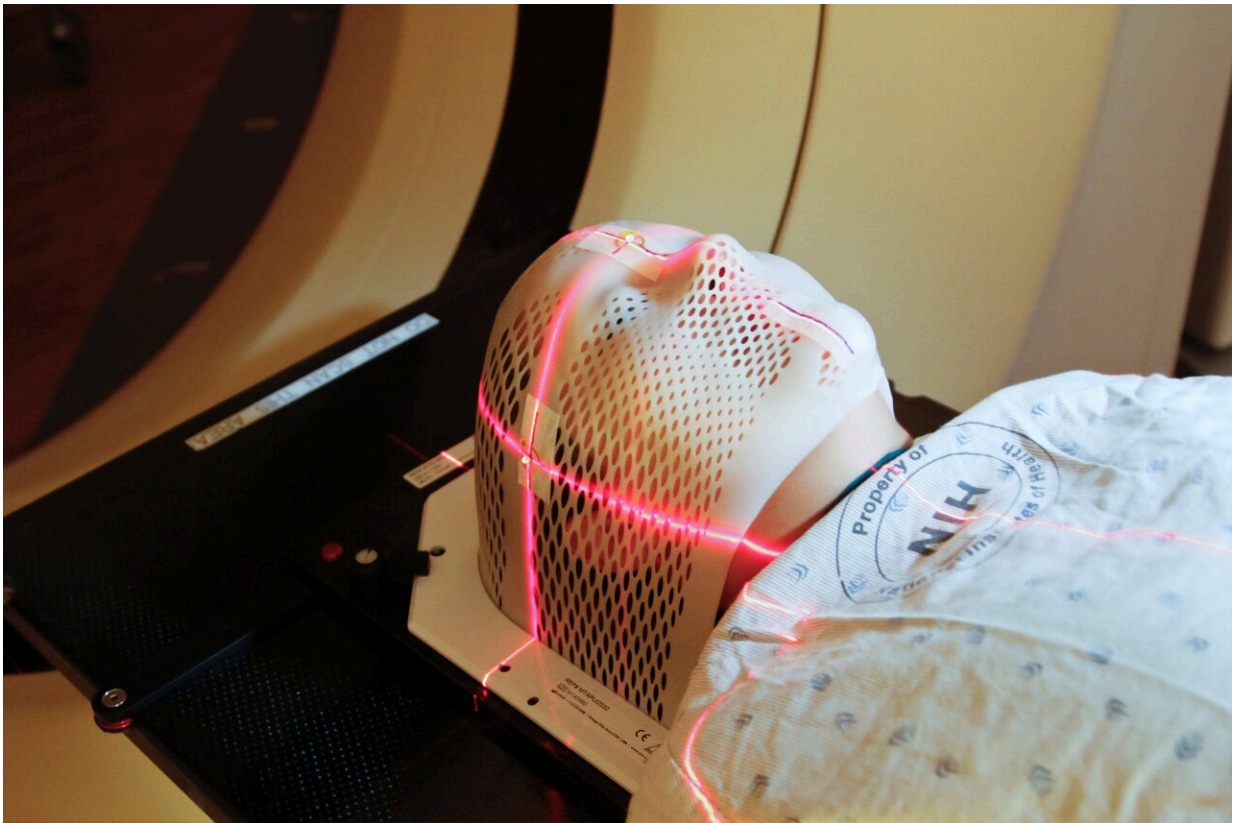


Evaluation of AI for medical imaging: A key requirement for clinical translation

September 13 2022, by Beth Miller



Credit: Unsplash/CC0 Public Domain

Artificial intelligence (AI) is showing significant promise in medical imaging. To translate this promise to reality requires rigorous evaluation of these algorithms.

To develop guidelines to evaluate [artificial intelligence](#) (AI) in nuclear-medicine imaging, Richard L. Wahl, MD, director of Mallinckrodt Institute of Radiology (MIR) at Washington University School of Medicine in St. Louis, then president of the Society of Nuclear Medicine and Molecular Imaging (SNMMI), established the SNMMI Artificial Intelligence Task Force, within which an evaluation team was formed. This team, comprised of computational imaging scientists, [physicians](#), [physicists](#), biostatisticians and representatives from industry and [regulatory agencies](#), was led by Abhinav Jha, assistant professor of biomedical engineering in the McKelvey School of Engineering and assistant professor of radiology at MIR.

The guidelines proposed by this team, known as the Recommendations for Evaluation of AI for Nuclear Medicine (RELAINCE) guidelines, will be published in the *Journal of Nuclear Medicine*.

"Artificial intelligence is showing tremendous promise in [medical imaging](#), specifically in nuclear medicine imaging, in a multitude of applications ranging from image generation, enhancement and analysis," Jha said. "We have seen a lot of research in this area, including multiple papers by our own group. However, for clinical translation of these algorithms, rigorous evaluation is needed."

"Lack of rigorous evaluation may have multiple adverse consequences, including reducing credibility of research findings, misdirection of future research, and, most importantly, yielding tools that are useless or even harmful to patients," Jha continued. "In our discussions, it emerged that there is an important need for [guidelines](#) to perform such evaluation."

As an example, there is significant research in developing AI algorithms for processing nuclear-medicine images acquired at low doses.

"Along this direction, in our own lab, we developed an algorithm to process low-dose cardiac SPECT images and were excited by the results because the resulting images visually looked great," Jha said. "But it's not whether they look great that matters—it is how they will do on the task that is required from the images, in this case, detecting cardiac defects. And on that task, the algorithm performed no better compared to the original low-dose images."

"We found that while the images looked good, in some cases, they removed the lesions, and in others, they introduced false lesions," he said.

None of those results is acceptable, showing the need to evaluate the algorithms based on the clinical task.

The task force proposed that all AI algorithms should be evaluated on clinical tasks, and the evaluations should yield a claim that defines the clinical task for which the algorithm was evaluated, demographics, imaging procedures, and process to extract task-specific information that was used in the evaluation study, and quantitative figures of merit to evaluate performance on clinical task.

"A proper and clear definition of a claim indicating the intended use and the validity of the AI algorithm is of utmost importance, and the claim should be substantiated with a proper and extensive evaluation of the AI method," said Ronald Boellaard, professor of radiology and nuclear medicine at Amsterdam University Medical Centers and senior author of the paper. "The claim should specifically indicate under which conditions and for which patients the algorithm can be used as well as any limitations or factors that could result in incorrect or less accurate performances."

Wahl, the Elizabeth E. Mallinckrodt Professor of Radiology and a co-

author, said the AI task force has generated several important papers which will help move AI methods forward from research to clinical practice in nuclear medicine.

"AI has the potential to spread expertise globally, but if not implemented properly, could spread inaccuracies," he said. "Thus, the work by Dr. Jha and colleagues on the RELAINCE criteria is so important to assuring valid population and task-specific AI methods are developed and deployed."

The task force proposed a framework with four classes to evaluate the algorithms for promise, technical task-specific efficacy, clinical decision making and post-deployment efficacy. The intent behind this framework is that it will guide AI developers conduct the evaluation study that provides evidence to support their intended claim. Additionally, the task force proposed best practices to evaluate the AI algorithms for each of the four classes.

"We want to make sure these algorithms are evaluated well so that they can assist well in clinical tasks, and hence the patients get the best treatment," Jha said. "We want AI to help, not hurt the patient."

More information: Abhinav K. Jha et al, Nuclear Medicine and Artificial Intelligence: Best Practices for Evaluation (the RELAINCE Guidelines), *Journal of Nuclear Medicine* (2022). [DOI: 10.2967/jnumed.121.263239](https://doi.org/10.2967/jnumed.121.263239)

Provided by Washington University School of Medicine in St. Louis

Citation: Evaluation of AI for medical imaging: A key requirement for clinical translation (2022, September 13) retrieved 24 June 2024 from <https://medicalxpress.com/news/2022-09-ai-medical->

[imaging-key-requirement.html](#)

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.