

Is this AI output reliable?

Developing trustworthy AI for medical imaging

Abhinav K. Jha, PhD
Department of Biomedical Engineering
Mallinckrodt Institute of Radiology
Chair: SNMMI AI Task Force

Artificial intelligence (AI) is showing significant promise in multiple areas!



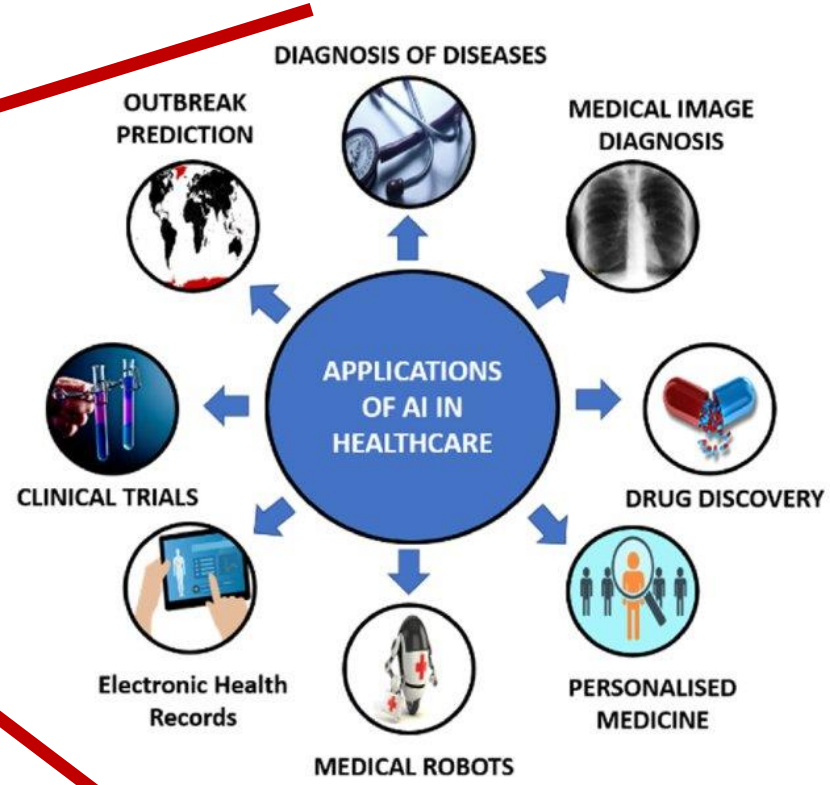


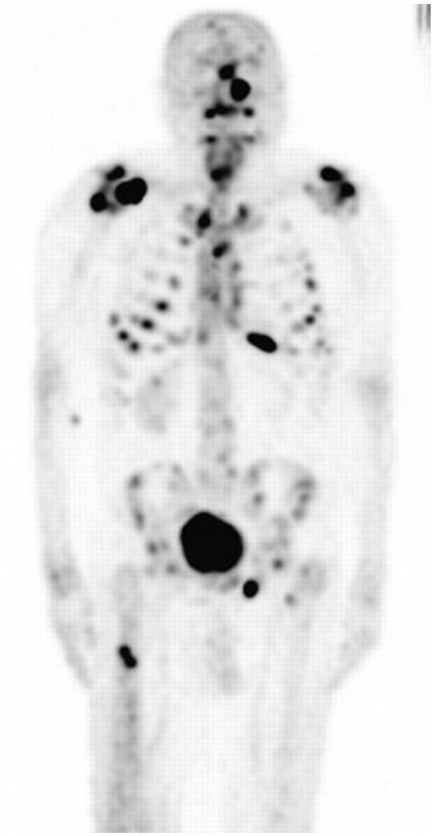
*Image source: https://www.javatpoint.com/application-of-ai

*Source: Pandya et al, Sensors, 2021

The broad application in this talk: Medical Imaging







The technique of creating images of the interior of a body for clinical analysis

Nuclear-medicine imaging: PET and SPECT



Single-photon emission computed tomography (SPECT)



Positron emission tomography (PET)

Imaging modalities that are widely used in multiple clinical applications

The role of computation in nuclear-medicine imaging



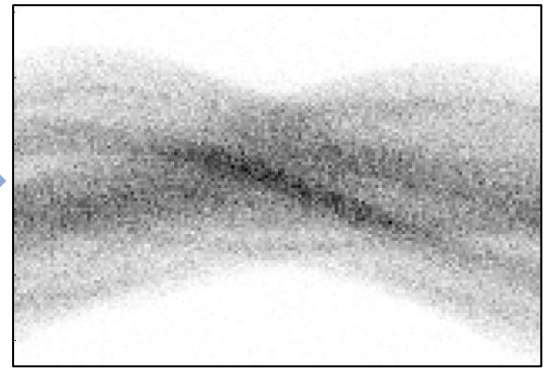


Image reconstruction

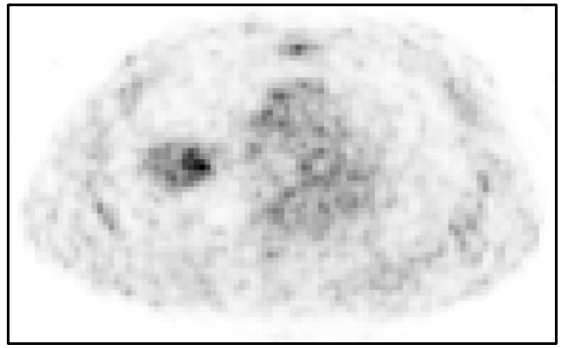


Image enhancement

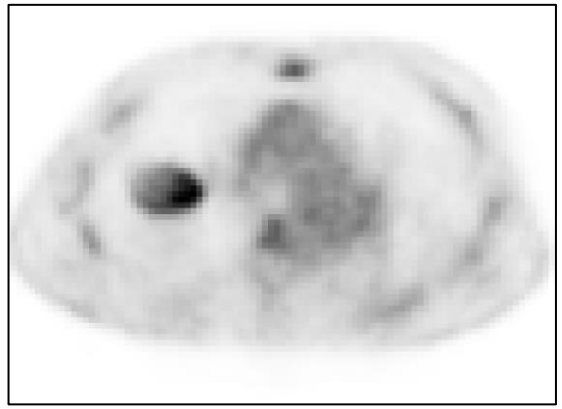
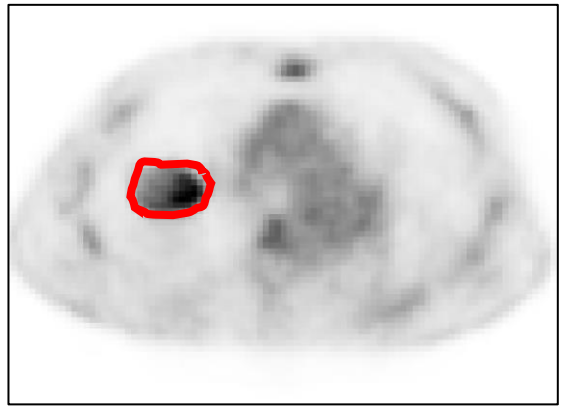
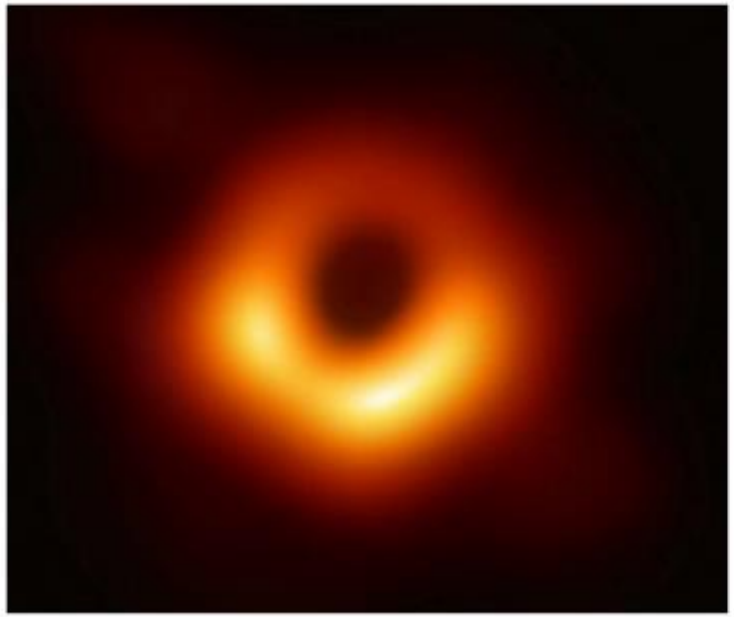


Image segmentation



Parameter quantification and clinical decision making

The role of computation in nuclear-medicine imaging



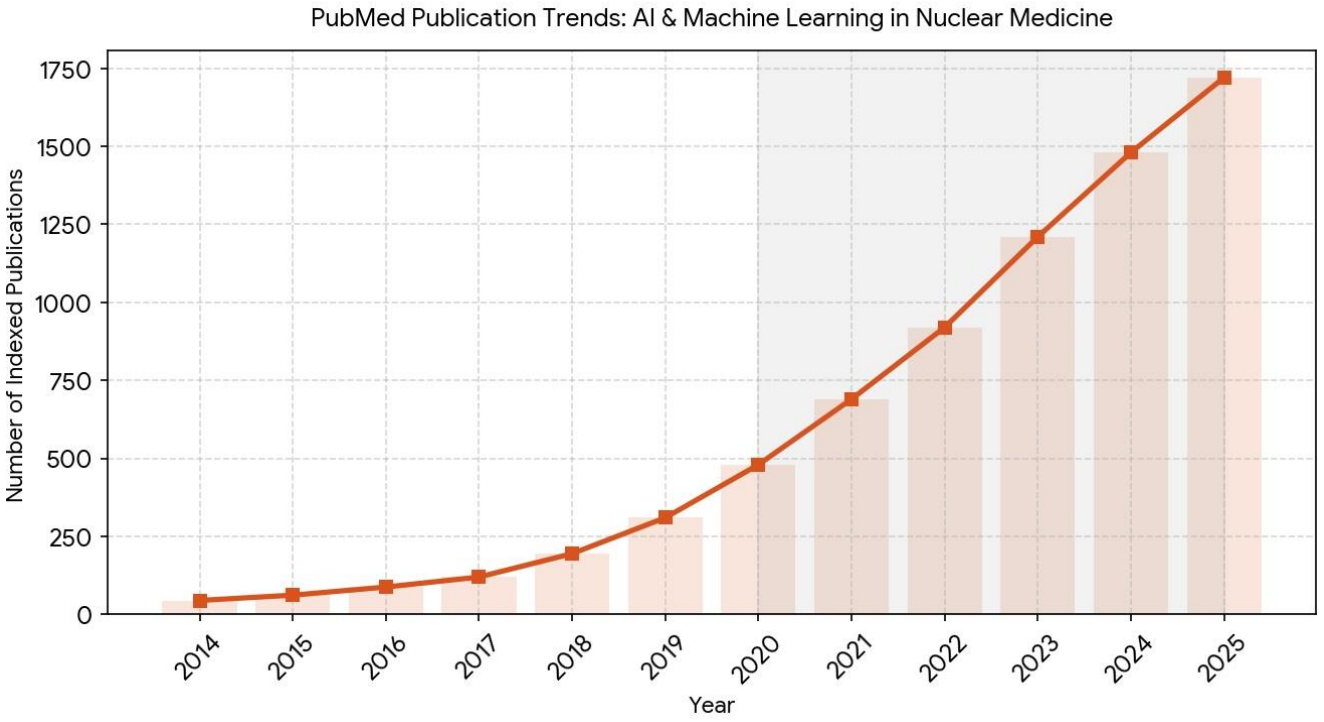
Black hole: Heart of the M87!



Image of a human heart with SPECT!

*Image source: Fortney

AI algorithms are showing strong promise in multiple aspects of nuclear medicine



News | **Videos** | February 26, 2025

Can Deep Learning Provide a CT-Less Alternative for Attenuation Compensation with SPECT MPI?

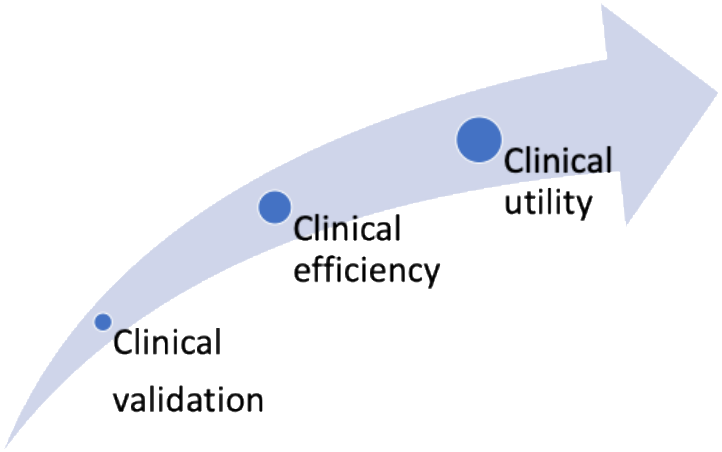
Author(s) *Jeff Hall*

CLINICAL NEWS | **MOLECULAR IMAGING**

AI cuts PET scan time in Alzheimer's disease

How do we translate these tools to clinical practice?

Insights from the 2026 SNMMI AI Summit: Clinical adoption of AI



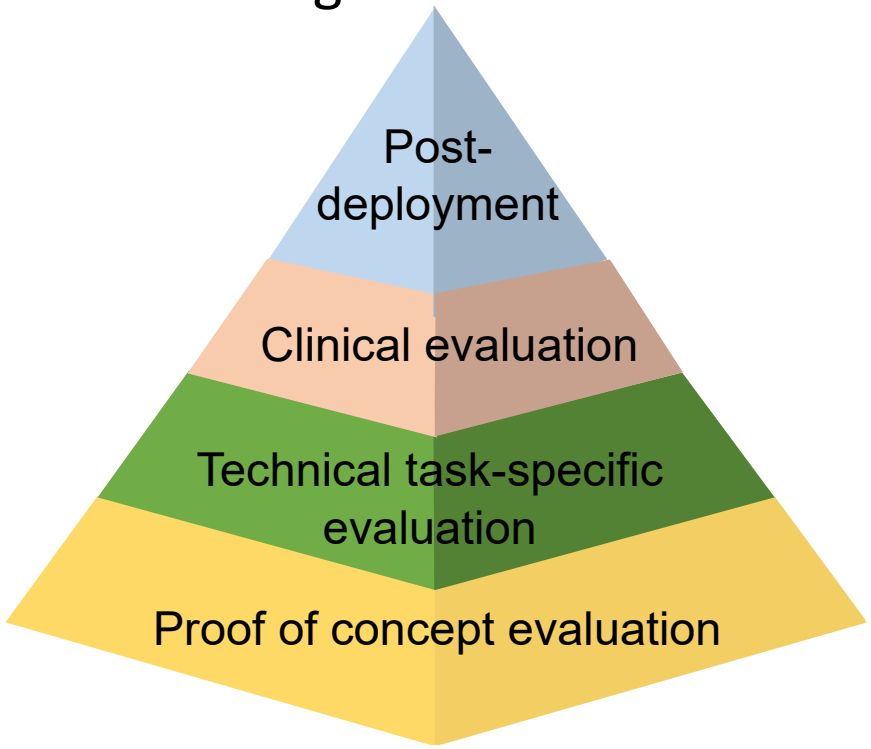
A theme emerging from the summit discussions

A key requirement for successful AI adoption is clinically relevant validation.

How do we achieve this goal?

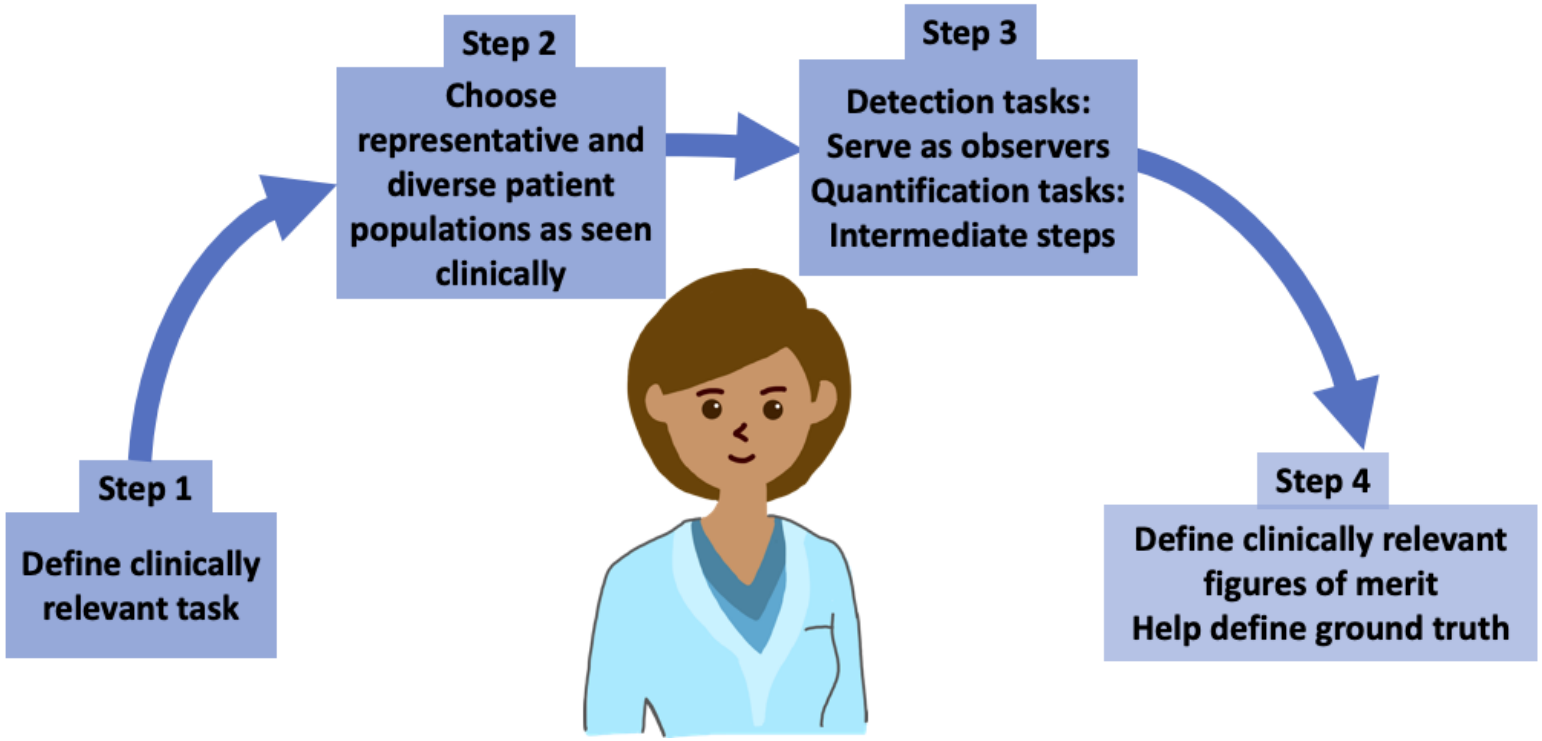
Guidelines and mechanisms have been developed for AI validation

RELAINCE guidelines provide a mechanism for performing rigorous validation



*Jha et al, J. Nuc. Med. 2023

Physicians have an important role in validation of AI algorithms

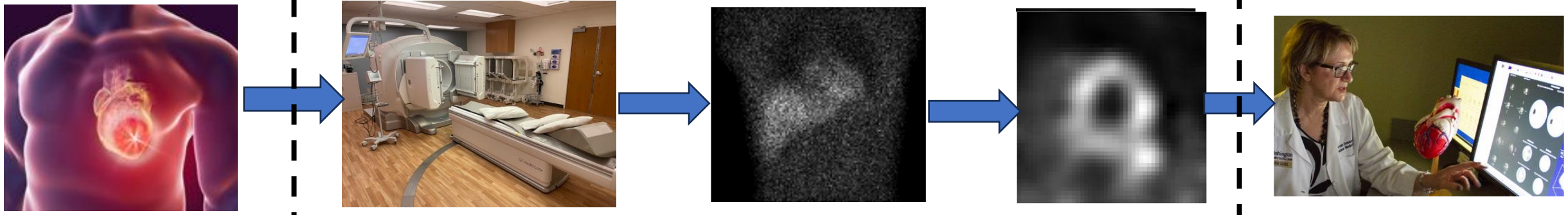


*Jha et al, PET Clinics, 2021

How do we implement these guidelines?

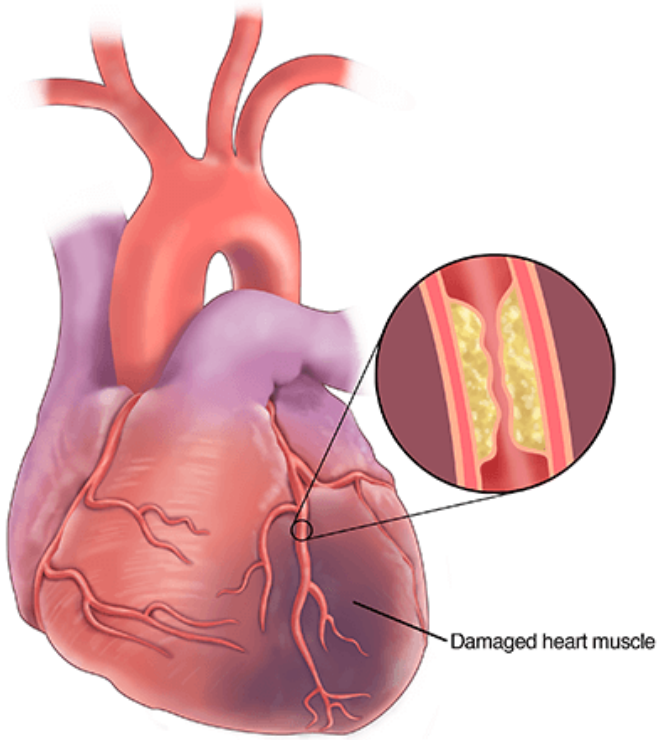
Clinical application for this talk: SPECT Myocardial perfusion imaging (MPI)

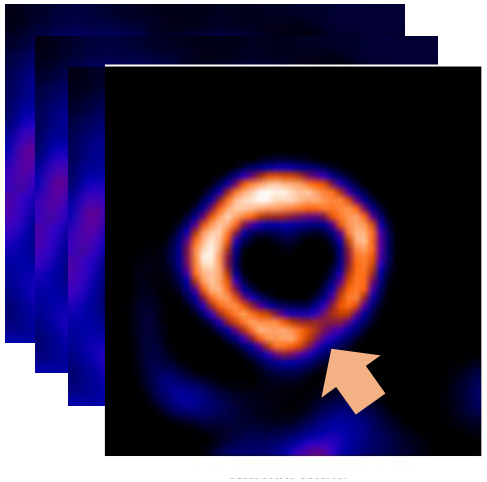
Cost effective and widely used modality for the diagnosis and evaluation of coronary artery disease (~9 million scans conducted in the USA alone)



Focus of this talk

Clinical application: Cardiac SPECT



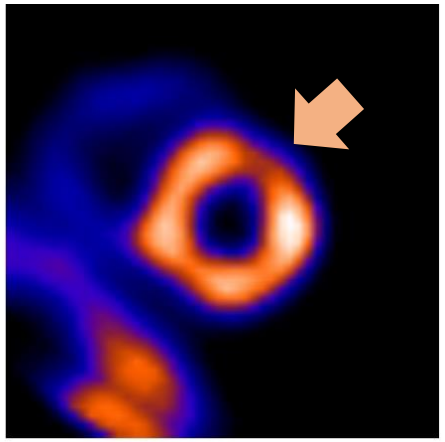


Area of reduced blood flow (defect)

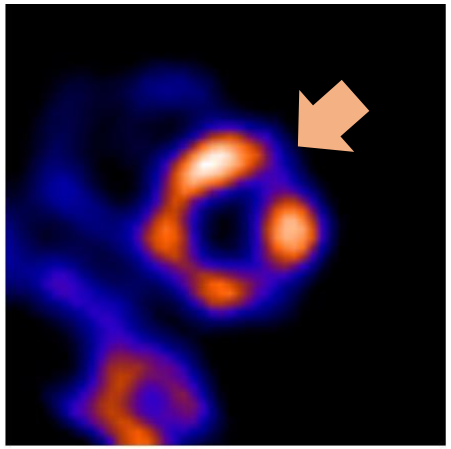
*Source: upmc.com

Challenges: SPECT MPI acquisition

- Requires administering substantial amounts of radiotracer dosage
- Long scan duration: 12 – 20 minutes



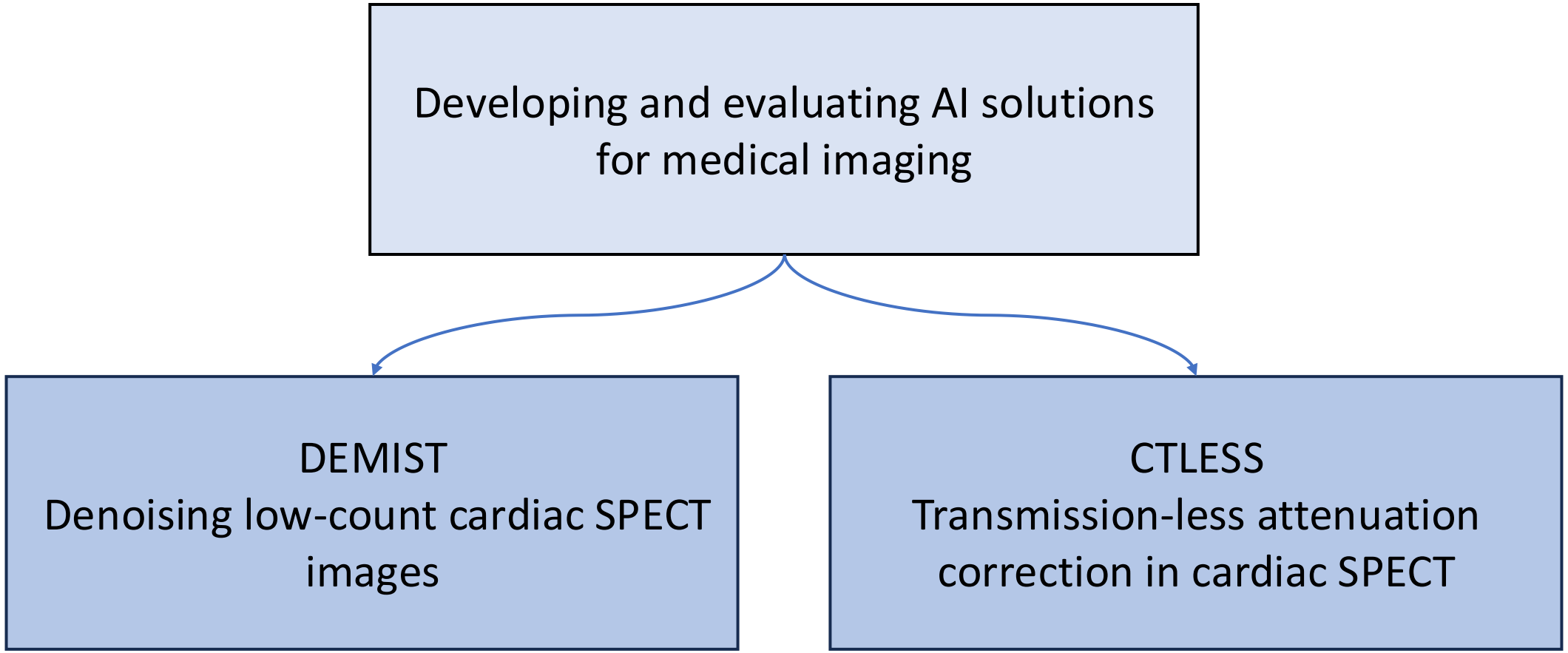
Normal-dose image



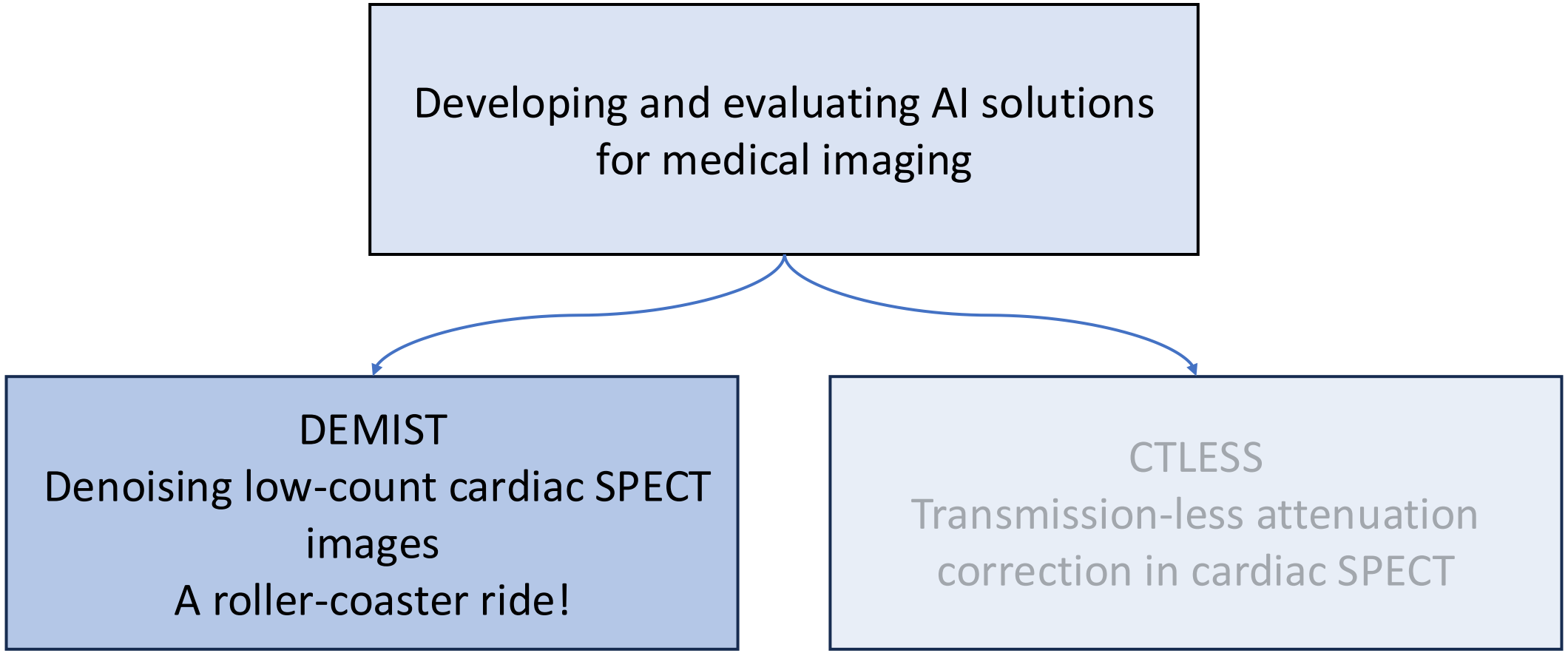
Low-dose image

- For attenuation compensation, additional X-ray CT scan typically acquired
- Limited image quality for patients with large anatomies

How AI can help address these challenges



How AI can help address these challenges

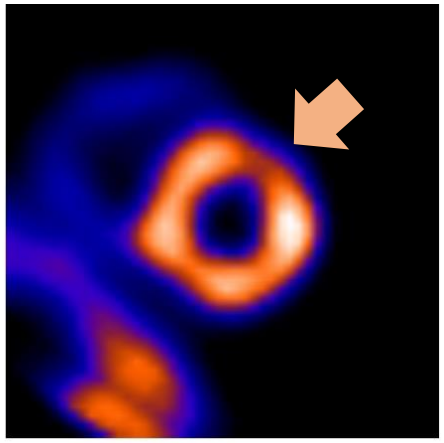


Tie your seat belts and get ready for the ride!

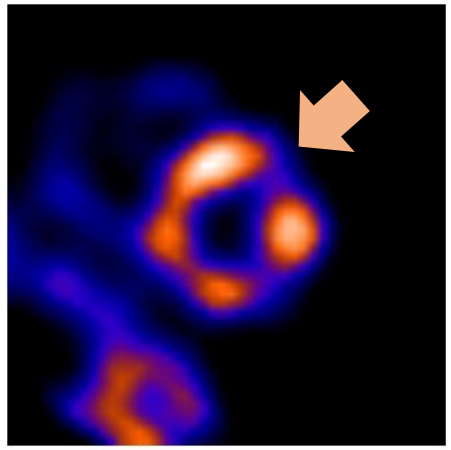


The problem

- Cardiac SPECT procedures require patients to be:
  - Administered substantial radiation dose (~48 mCi)
  - Scanned for a long time on the scanner
- Important need to reduce radiation dose/acquisition time for patients, but that impacts image quality

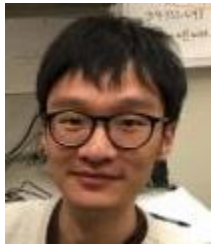


Normal-dose image



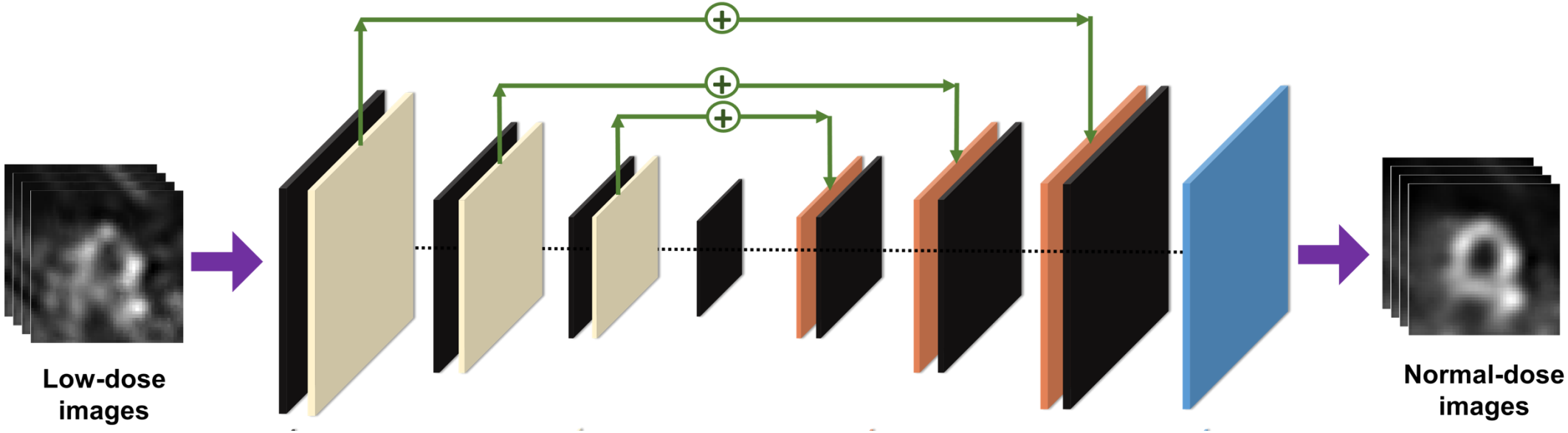
Low-dose image

Our initial efforts on developing a deep-learning denoising technique



Zitong Yu

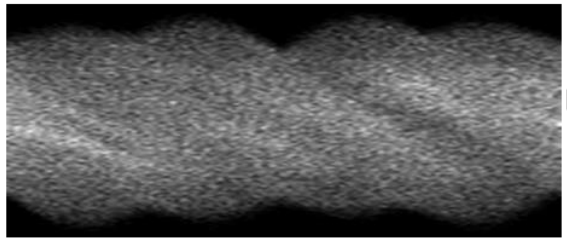
Explored a commonly used deep denoising technique



This algorithm minimized the mean square error between the normal dose and predicted normal dose images

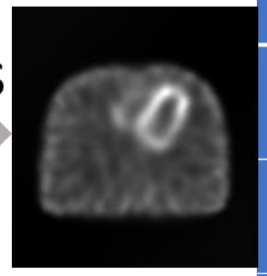
Very promising initial results!

Normal-dose sinogram

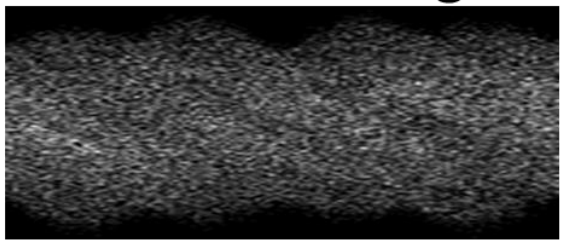


Pure-physics

Normal-dose image

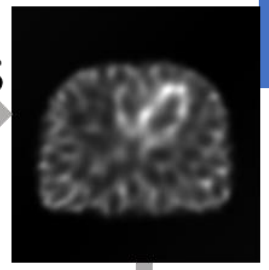


Low-dose sinogram



Pure-physics

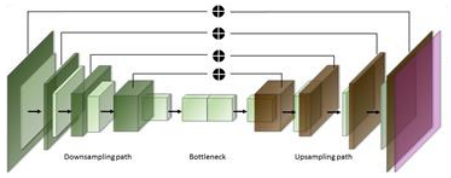
Low-dose image

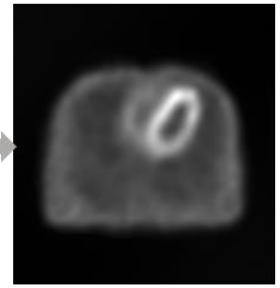


Conventional metrics

	Root mean square error	Structural similarity index	Peak signal-to-noise ratio
AI-based	0.045 (0.045,0.046)	0.87 (0.87,0.87)	27.25 (27.13,27.35)
Pure physics-based	0.047 (0.047,0.048)	0.85 (0.85,0.85)	26.80 (26.69,26.90)
Inference	**AI better ($p < 0.01$)**	**AI better ($p < 0.01$)**	**AI better ($p < 0.01$)**

CNN-based denoiser



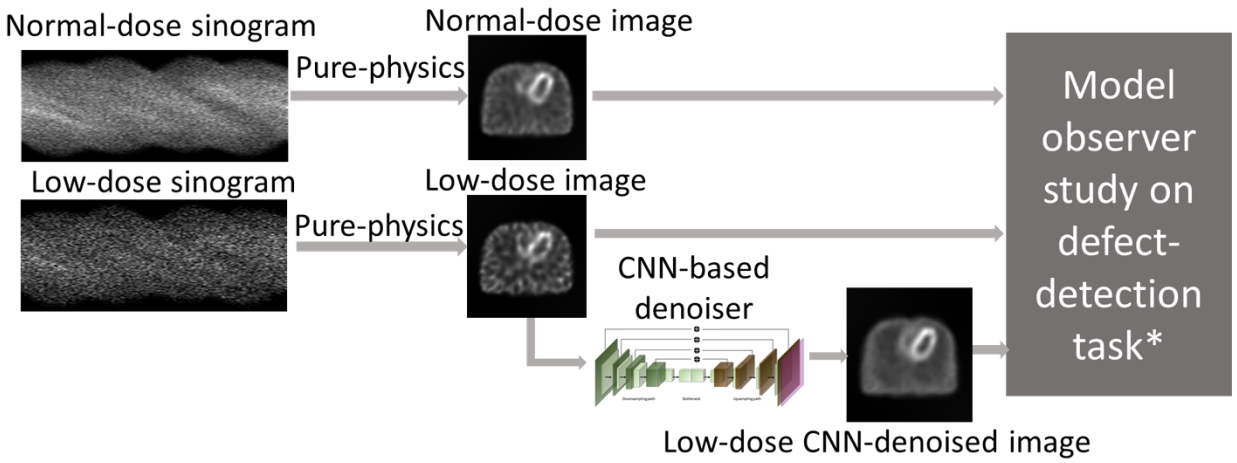


Low-dose CNN-denoised image

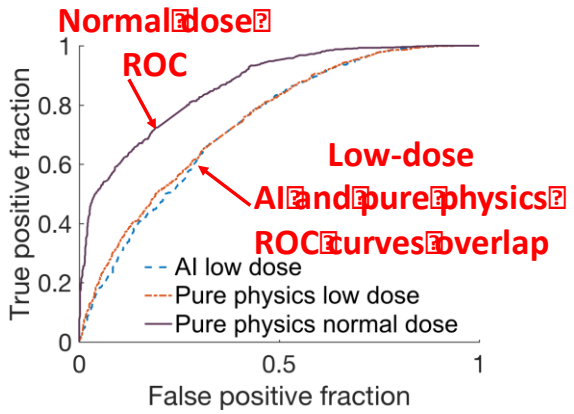
Wow!
We can reduce radiation dose by 10 times!

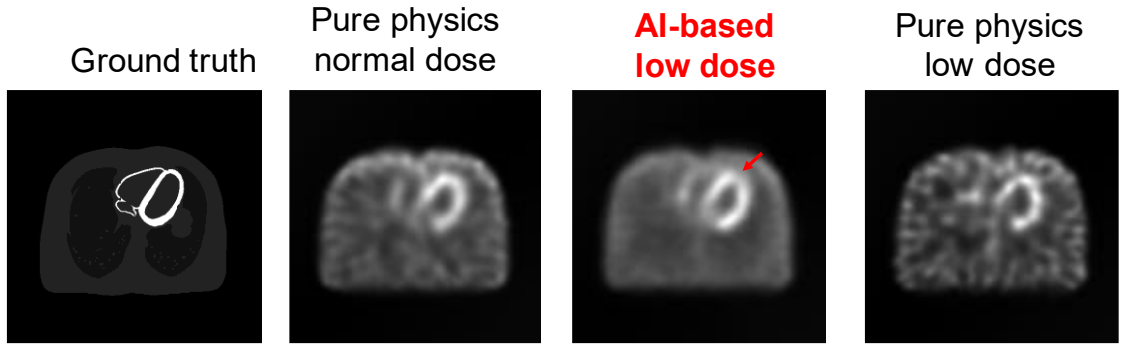


What about performance on clinical task?



	Conventional metrics			Task-specific metric
	Root mean square error	Structural similarity index	Peak signal-to-noise ratio	Area under the ROC curve (AUC)
AI-based	0.045 (0.045,0.046)	0.87 (0.87,0.87)	27.25 (27.13,27.35)	0.74 (0.71,0.76)
Pure physics-based	0.047 (0.047,0.048)	0.85 (0.85,0.85)	26.80 (26.69,26.90)	0.74 (0.72,0.77)
Inference	**AI better ($p < 0.01$)**	**AI better ($p < 0.01$)**	**AI better ($p < 0.01$)**	**AI gives similar performance as physics!**





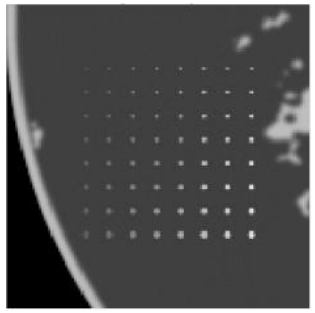
Even though the deep denoising algorithm was creating pretty images, the images were not useful!

*Yu et al, J. Nuc. Med. Supp 2020

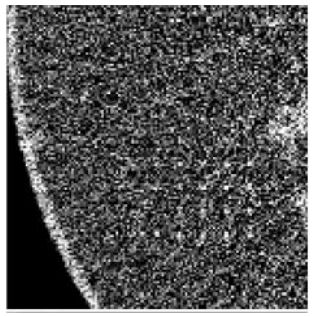
What about other clinical applications?

- Similar findings were observed in other imaging modalities[1-3].

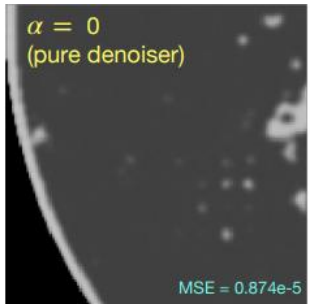
Ground truth



Low dose



Denoised



Denoising washed out the small low-contrast signals[3]

- And in other applications of AI in nuclear medicine, including image segmentation[4]

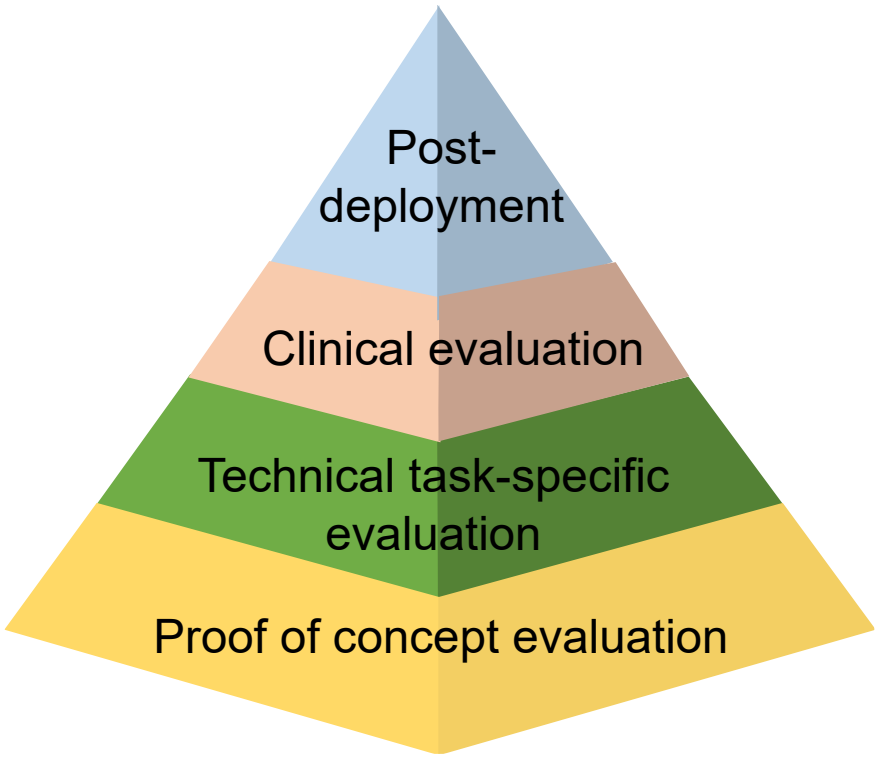
Important need for evaluation of AI algorithms on clinical tasks

[1]Badal et al., *SPIE Med. Imaging* 2019
[2]Li et al., *IEEE Trans. Med. Imaging* 2021
[3]Ongie et al., *SPIE Med. Imaging* 2021
[4] Liu et al, J. Nuc. Med. 2024

Led evaluation team of SNMMI AI taskforce that recommended the RELAINCE guidelines



Team consisted of nuclear medicine physicists, computational imaging scientists, physicians, statisticians, representatives from the industry and from regulatory agencies



THE STATE OF THE ART

Nuclear Medicine and Artificial Intelligence: Best Practices for Evaluation (the RELAINCE Guidelines)

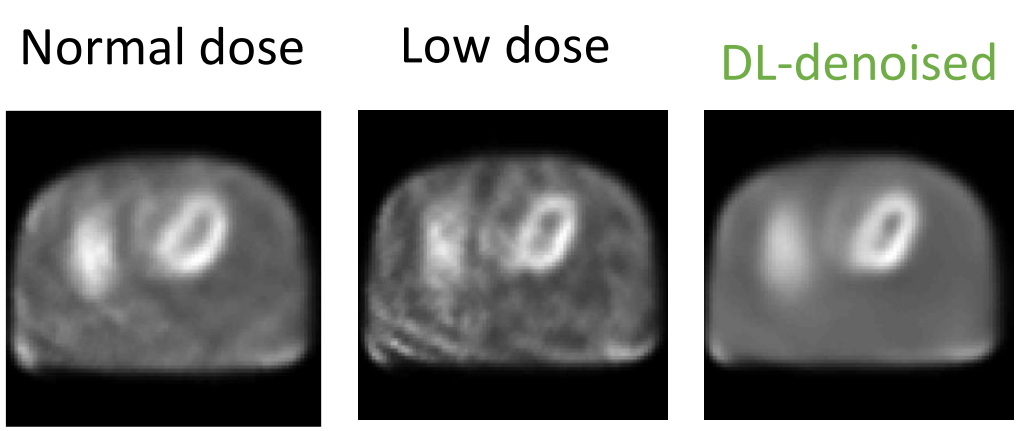
Abhinav K. Jha[1], Tyler J. Bradshaw[2], Irène Buvat[3], Mathieu Hatt[4], Prabhat KC[5], Chi Liu[6], Nancy F. Obuchowski[7], Babak Saboury[8], Piotr J. Slomka[9], John J. Sunderland[10], Richard L. Wahl[11], Zitong Yu[12], Sven Zuehlsdorff[13], Arman Rahmim[14], and Ronald Boellaard[15]

Does this mean deep learning cannot help reduce radiation dose?

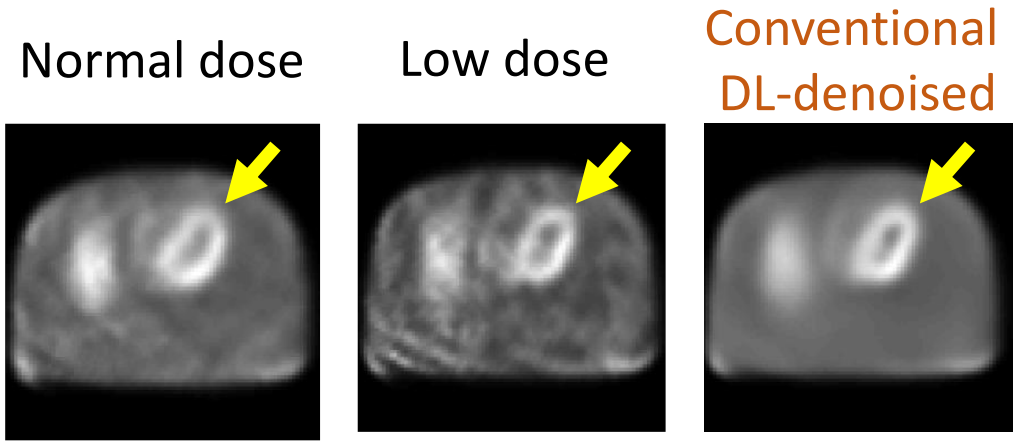
Or can we advance deep learning to help?



Why was deep learning not improving observer performance?



Denoising makes the images less noisy



Denoising reduces defect contrast

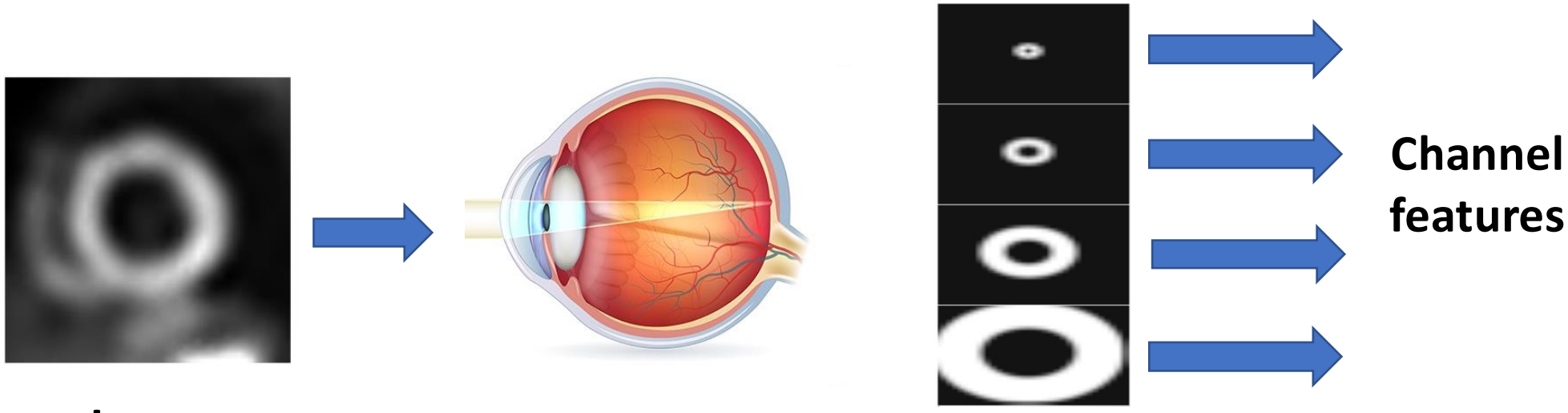
No overall improvement!

The issue: The AI algorithm was designed to improve visual fidelity, and not for the clinical task

Yu et al., *Med. Phys.* 2023

Addressing the issue: Design AI for clinical tasks

Using psychophysical studies, it has been seen that human visual system can be modeled as processing images through frequency selective channel[1,2].

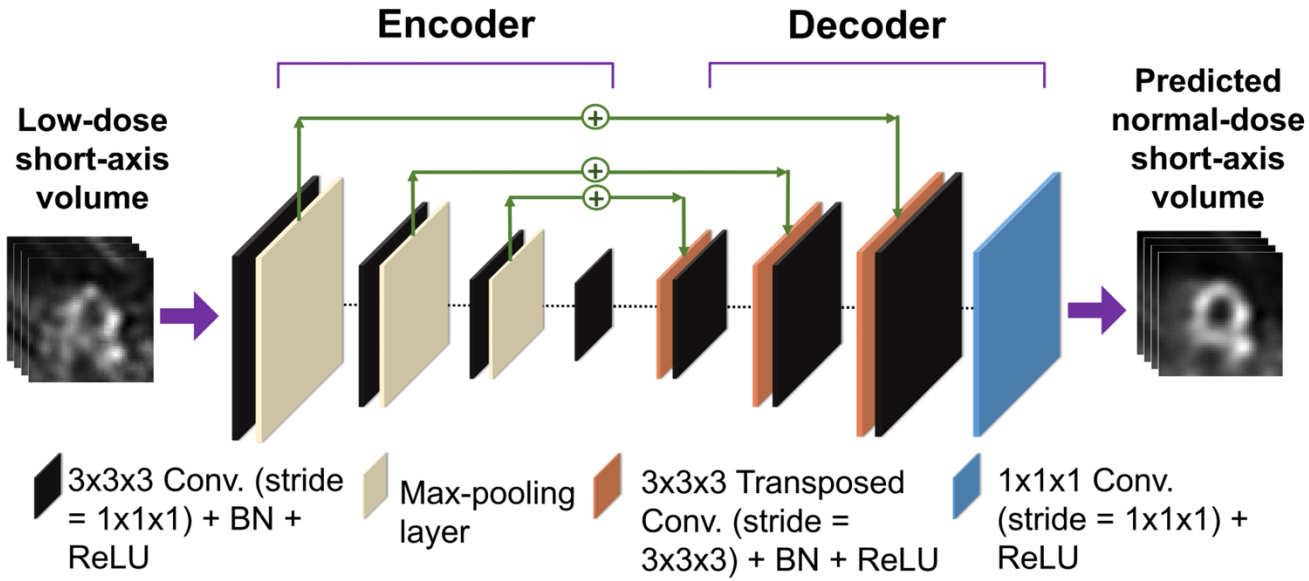


Can we design a deep learning algorithm that, while performing denoising, ensures that the features used by human observers are preserved?

A detection-task-specific deep denoising method (DEMIST)*



Ashequr Rahman, PhD



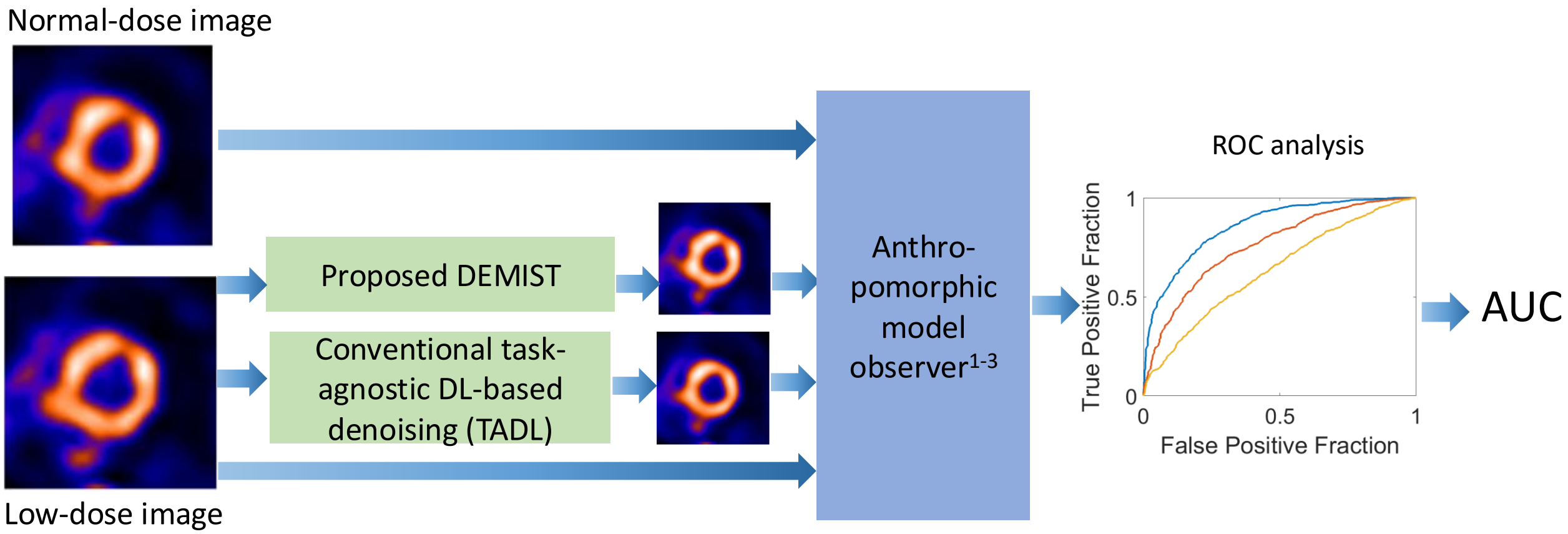
$$\text{Loss function} = \text{Fidelity loss} + \lambda * \text{Observer loss}$$

Minimize mean square error between normal-dose and predicted normal=dose image

Preserve features that assist in detection task: Minimize mean square error between channel features in normal and predicted normal-dose image

*Rahman et al, IEEE TRPMS April 2024

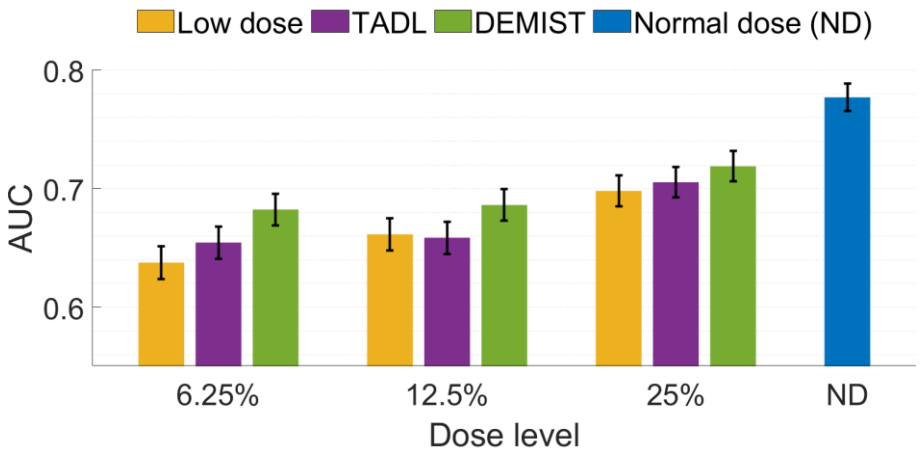
Evaluation on the clinical task of detecting myocardial perfusion defects



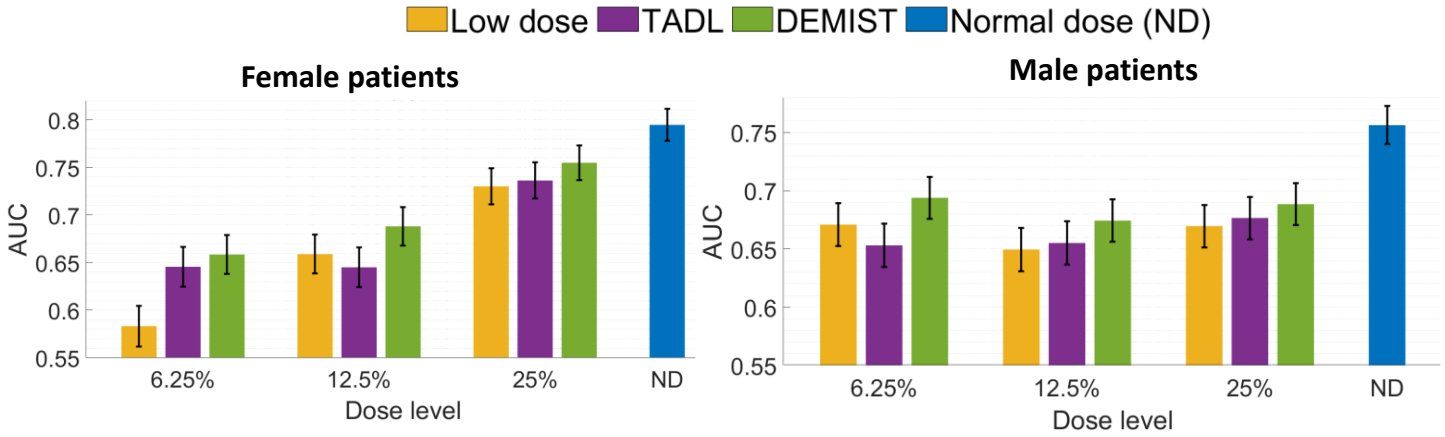
[1] Frey et al. IEEE TMI. 2002.
[2] Wollenweber et al. IEEE TNS. 1999
[3] Sankaran et al. JNM. 2002

Results: Observer-study evaluation

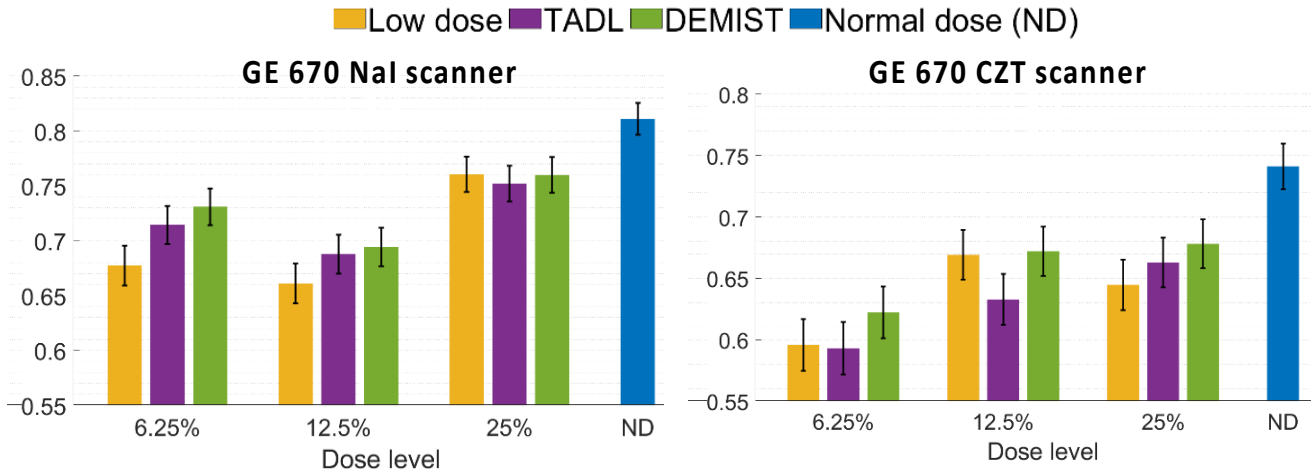
Entire test population



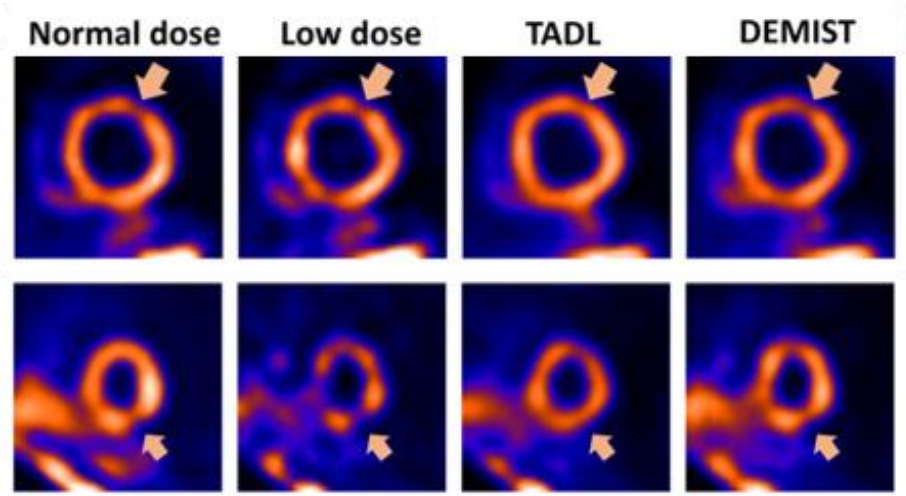
Stratified analysis: Patient sex



Stratified analysis: Different scanners



Representative results

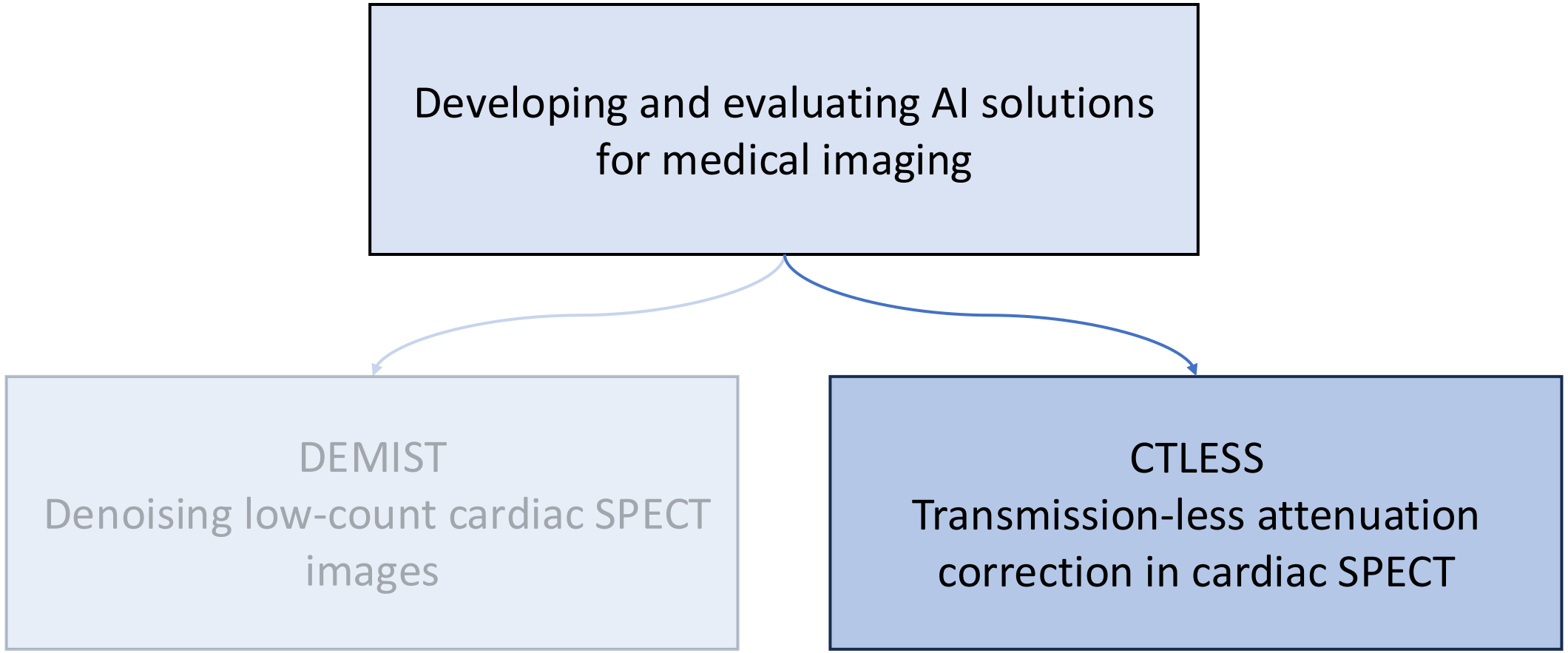


DEMIST significantly improved performance on defect detection task compared to low-dose

Key take aways

- Integrating clinical task, statistical decision theory and our understanding of the human visual system into a deep learning approach helped design AI solutions for medical imaging
- The resultant deep learning approach, DEMIST, yielded improved performance on the task of detecting cardiac defects
- AI-based dose reduction is promising, but further development and validation is needed to demonstrate performance comparable to current clinical protocols

How AI can help address these challenges



The challenge of attenuation in SPECT

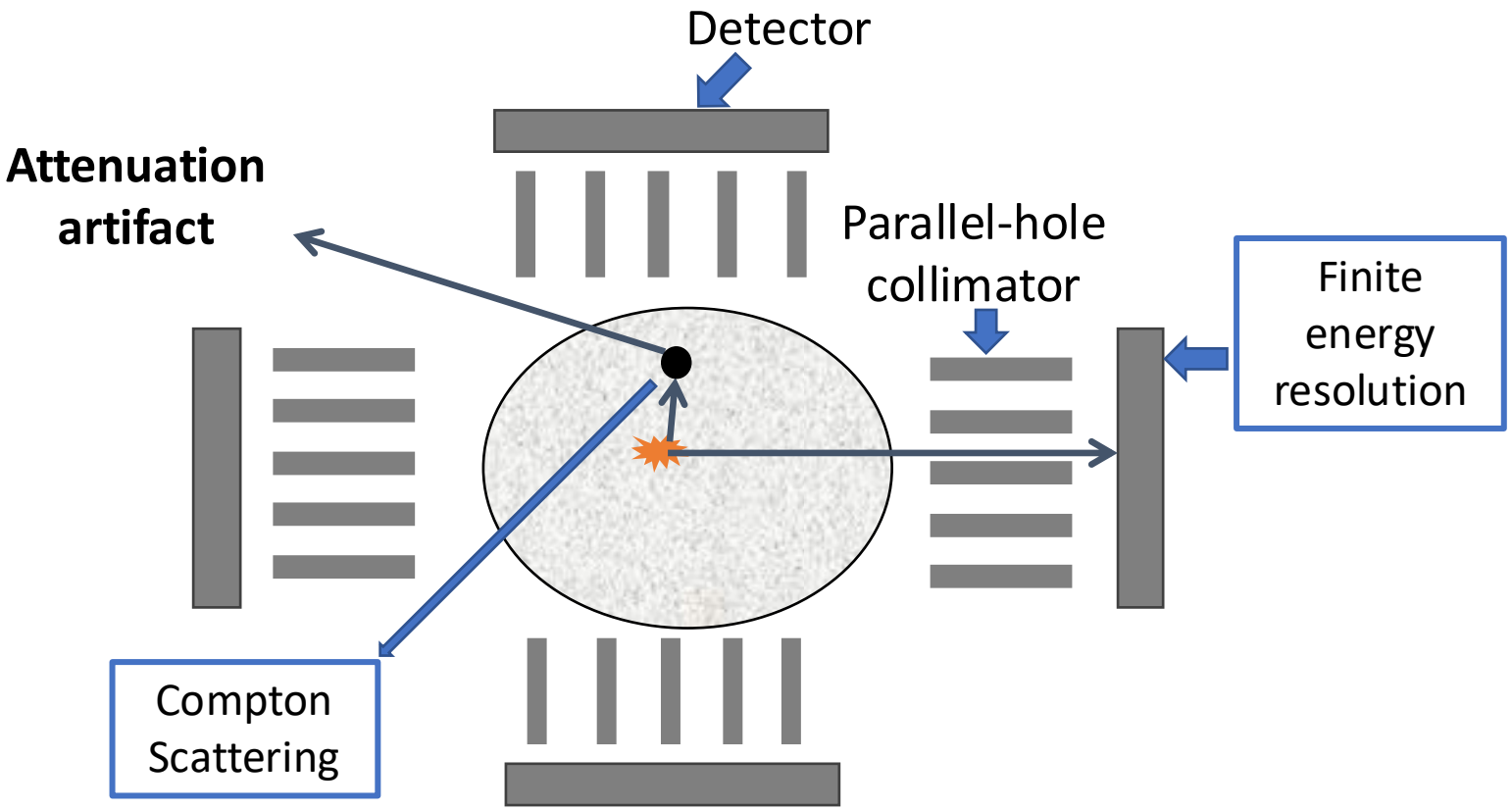
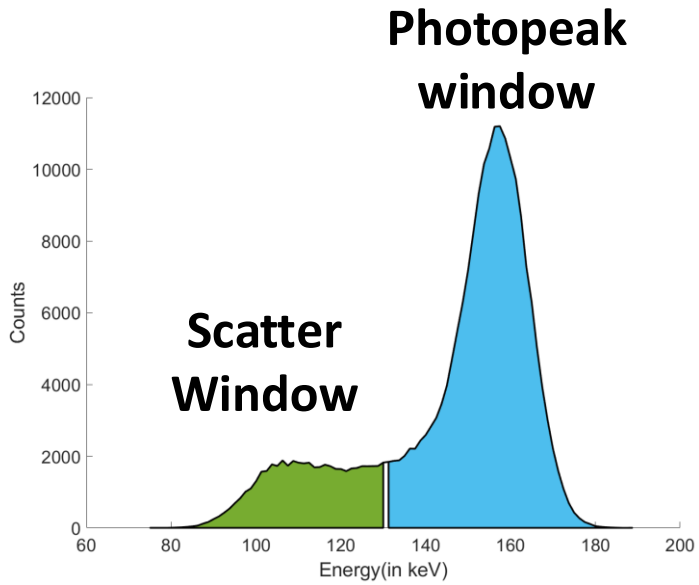


Fig. Attenuation and scatter in SPECT



Energy spectrum of a typical SPECT detector

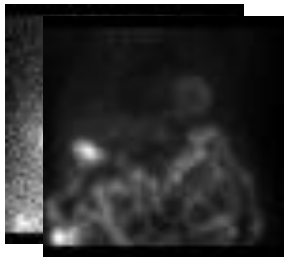
Attenuation: A major image-degrading artifact in SPECT

Attenuation compensation in SPECT

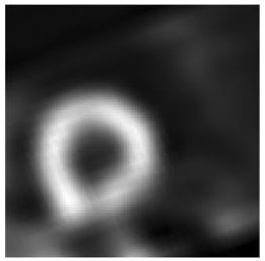
- Attenuation compensation (AC) is beneficial for visual interpretation tasks and pre-requisite for quantification tasks in SPECT.[1]
- Traditional AC methods require an additional X-ray CT scan:
  - Many SPECT systems do not have CT imaging capability.
  - Even when available, there are multiple disadvantages.
  - A staggering 75% of SPECT scans are performed without AC!







SPECT emission data

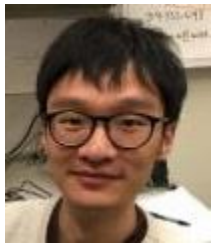


SPECT image with AC

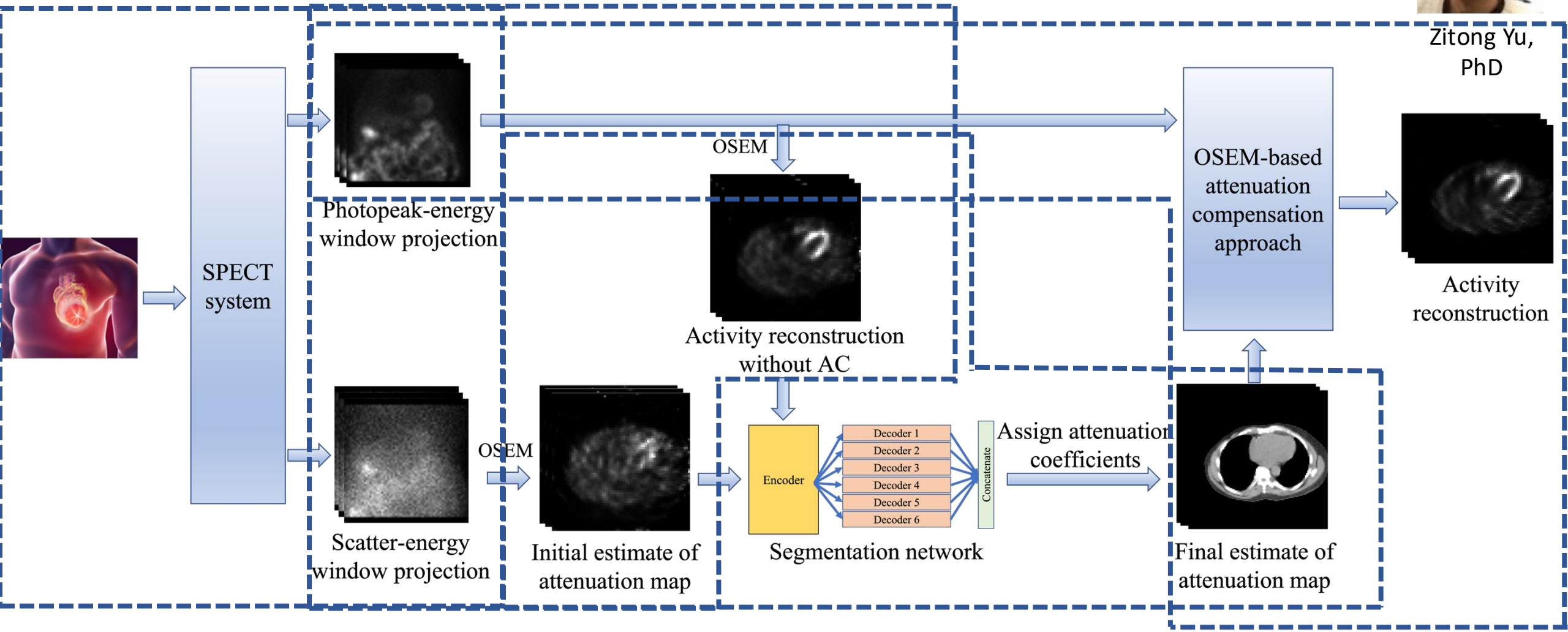
Can we perform AC using SPECT emission data only?

1. Garcia et al., JNC, 2007 2. Garcia et al., JNC, 2007
2. Hagio et al., Eur J Nucl Med Mol Imaging, 2022

CTLESS: A scatter-window projection and deep learning-based transmission-less AC method[1]



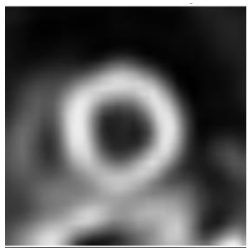
Zitong Yu, PhD



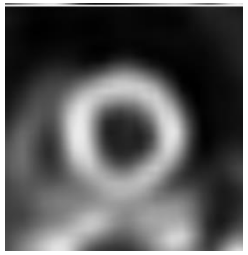
[1]Yu et al., IEEE Trans. Med. Imag., 2024

CTLESS was demonstrating promise in anthropomorphic model observer studies*

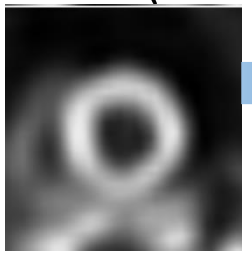
CT-based AC (CTAC)



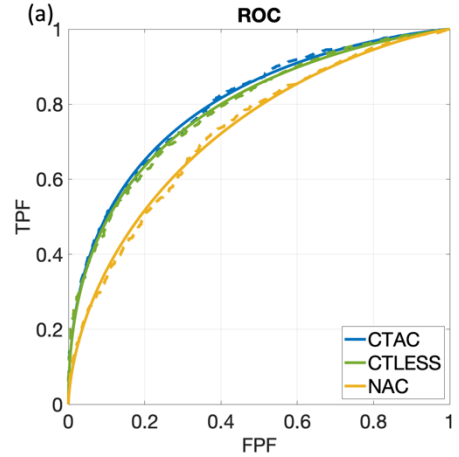
CTLESS

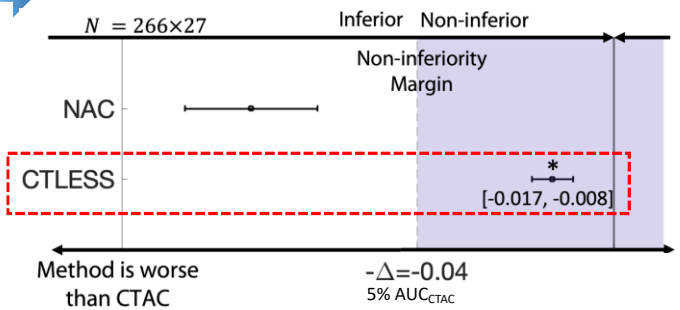


No AC (NAC)



Anthro-pomorphic model observer studies





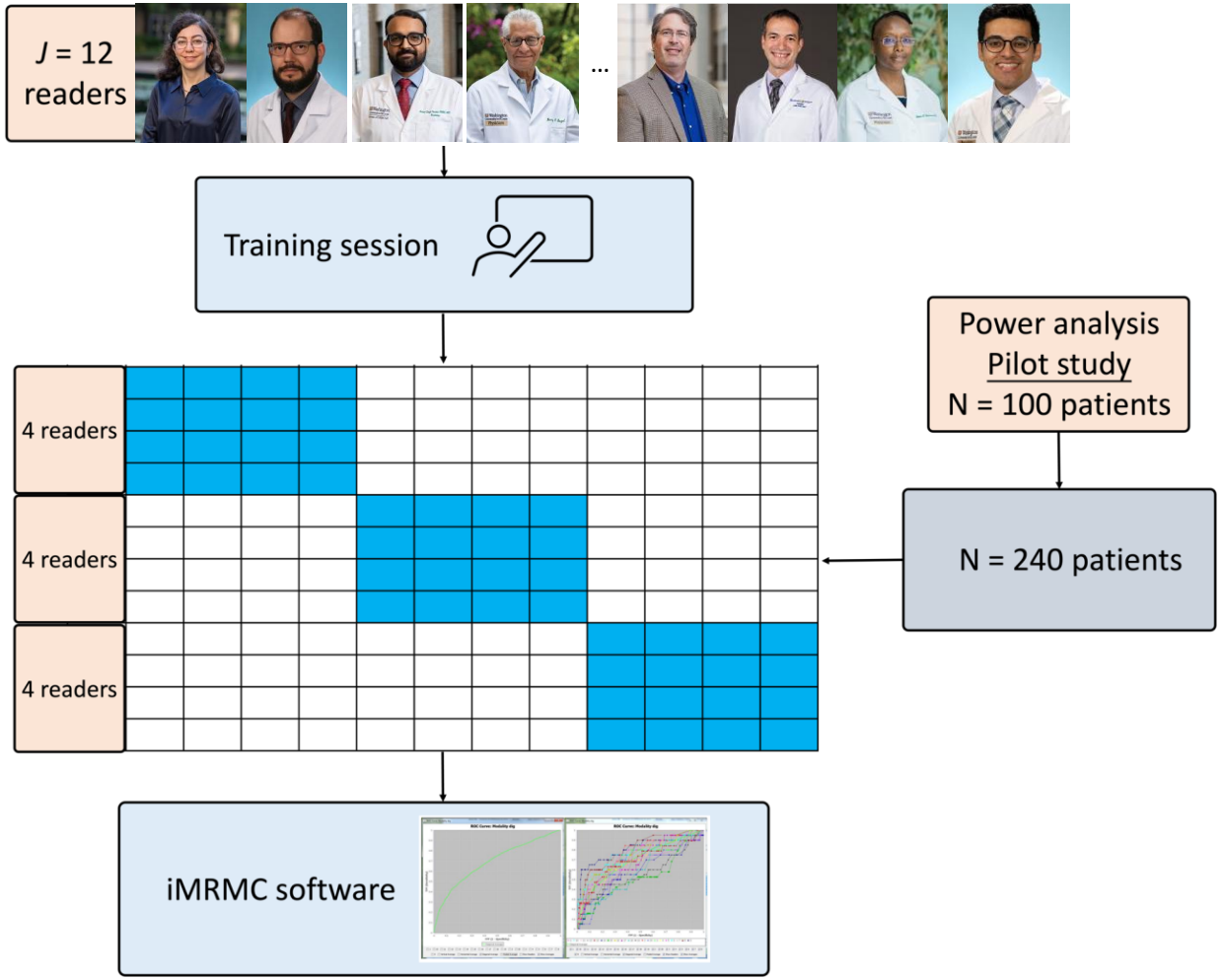
Motivated human-observer studies



CTLESS yielded non-inferior performance compared to standard of care CTAC method and outperformed no AC method

*Yu et al., IEEE Trans. Med. Imag., 2024

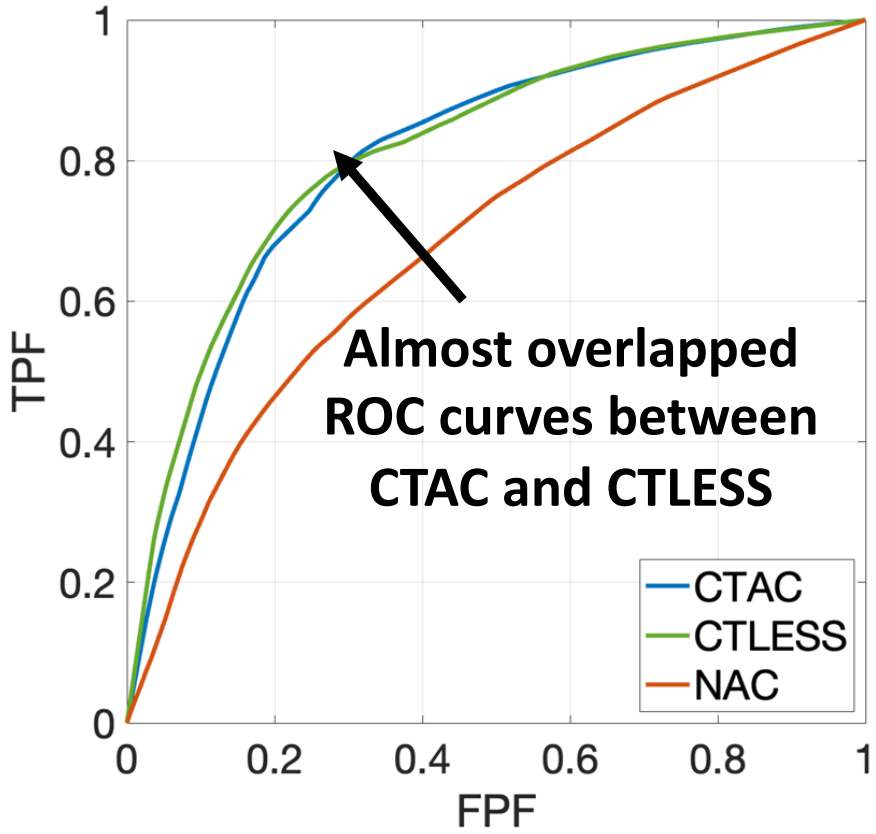
Evaluation of CTLESS with human-observer study on the task of detecting perfusion defects*

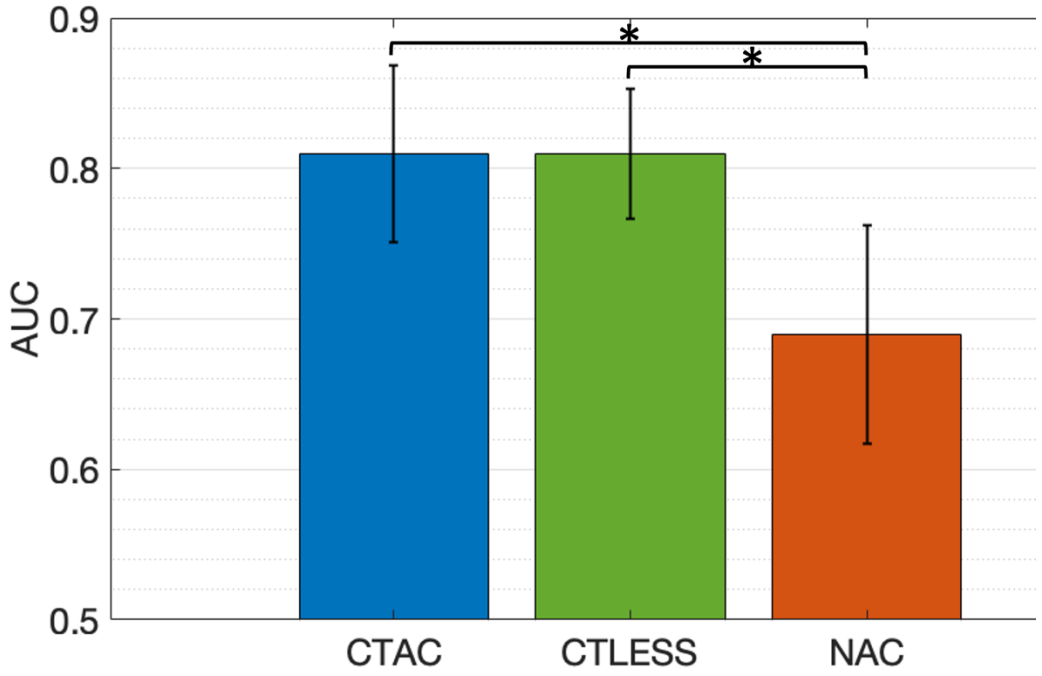


- *J* = 12 expert readers
- Synthetic inserted perfusion defects
- Carefully designed training session
- Power analysis based on a pilot study
- Images from *N* = 240 patients
- Split-plot design
- Reading results analysis using iMRMC

*Yu et al, in review

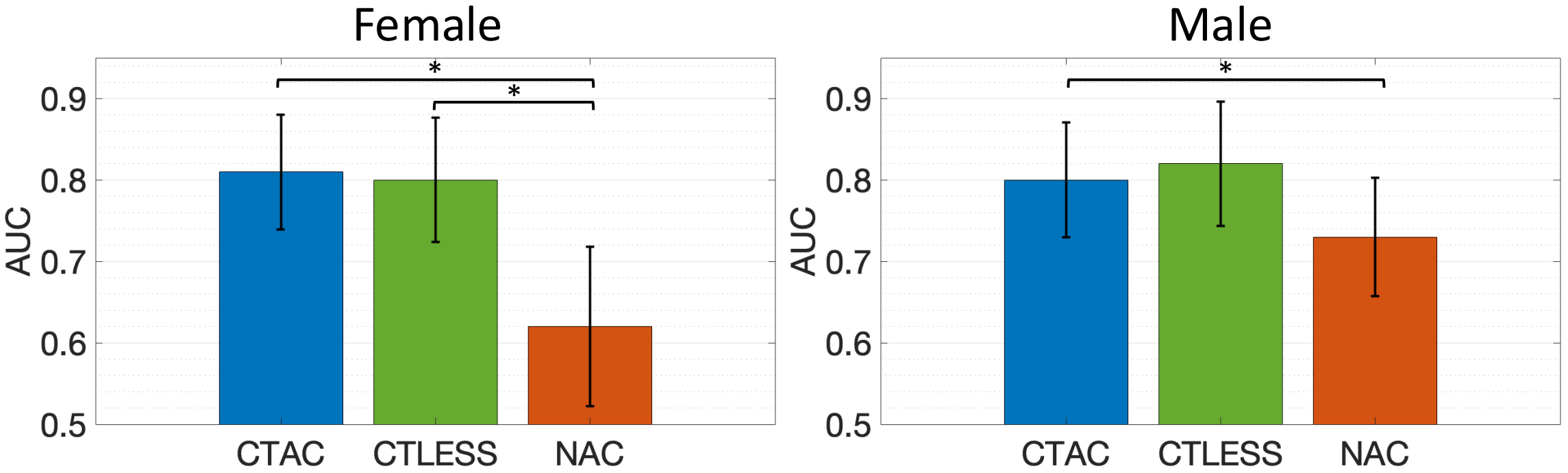
Human-observer study: ROC plots and AUC values





The ROC curve obtained by CTLESS almost overlapped that obtained by the CTAC.
AUC obtained by CTLESS was similar to CTAC and significantly outperformed NAC.

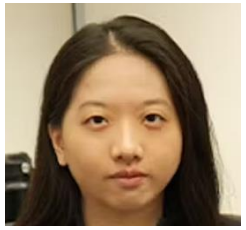
Human-observer study: Stratified analysis



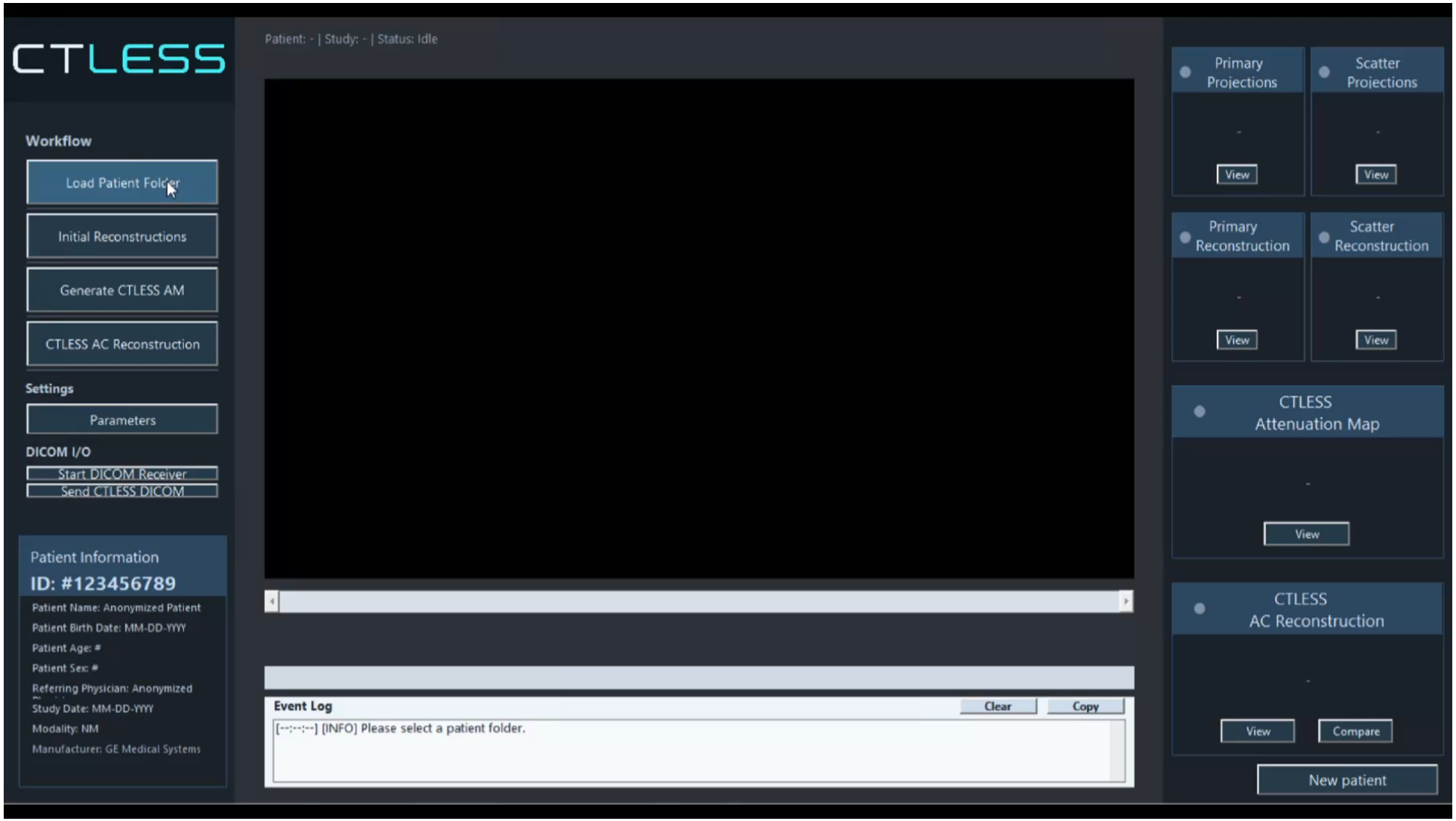
AUC values obtained by CTLESS were similar to CTAC in both males and females.

Proposed CTLESS method continued to yield similar performance as CTAC in human-observer studies! These results motivate further clinical validation of CTLESS.

An initial software implementation of CTLESS



Janice Tania, BS



Key takeaway

AI-based attenuation compensation methods are showing strong promise as alternatives to CT-based attenuation correction, motivating broader multicenter validation efforts.

Important to be mindful of ethics as we push boundaries with AI



Anya Plutynski, PhD

What do patients want?

- Uncertainty of AI algorithms should be accounted for when making clinical decisions
- Patient risk-value profiles should be considered when using AI for making clinical decisions

nature medicine

Explore content ⌄ About the journal ⌄ Publish with us ⌄

nature › nature medicine › comment › article

Comment | Published: 31 January 2022

Clinical decisions using AI must consider patient values

Jonathan Birch ✉, Kathleen A. Creel, Abhinav K. Jha & Anya Plutynski

Nature Medicine **28**, 229–232 (2022) | Cite this article

1038 Accesses | **30** Altmetric | Metrics

Built-in decision thresholds for AI diagnostics are ethically problematic, as patients may differ in their attitudes about the risk of false-positive and false-negative results, which will require that clinicians assess patient values.

Patients want AI in medicine, but to help physicians!

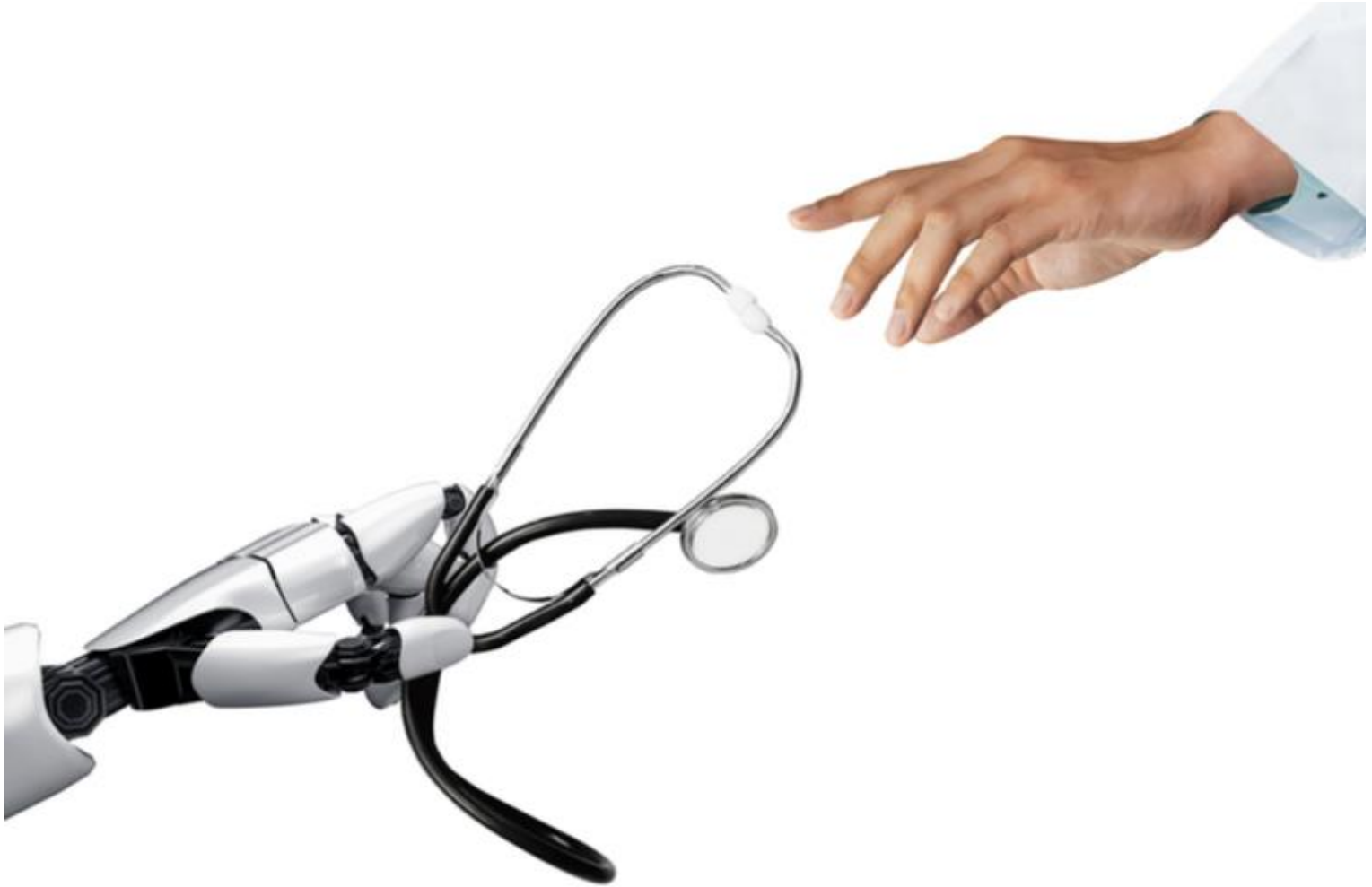
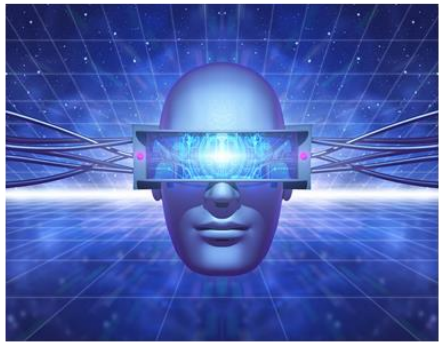


*Photo credit: Shutterstock

Conclusion

AI algorithms



+

Clinical task-based evaluation



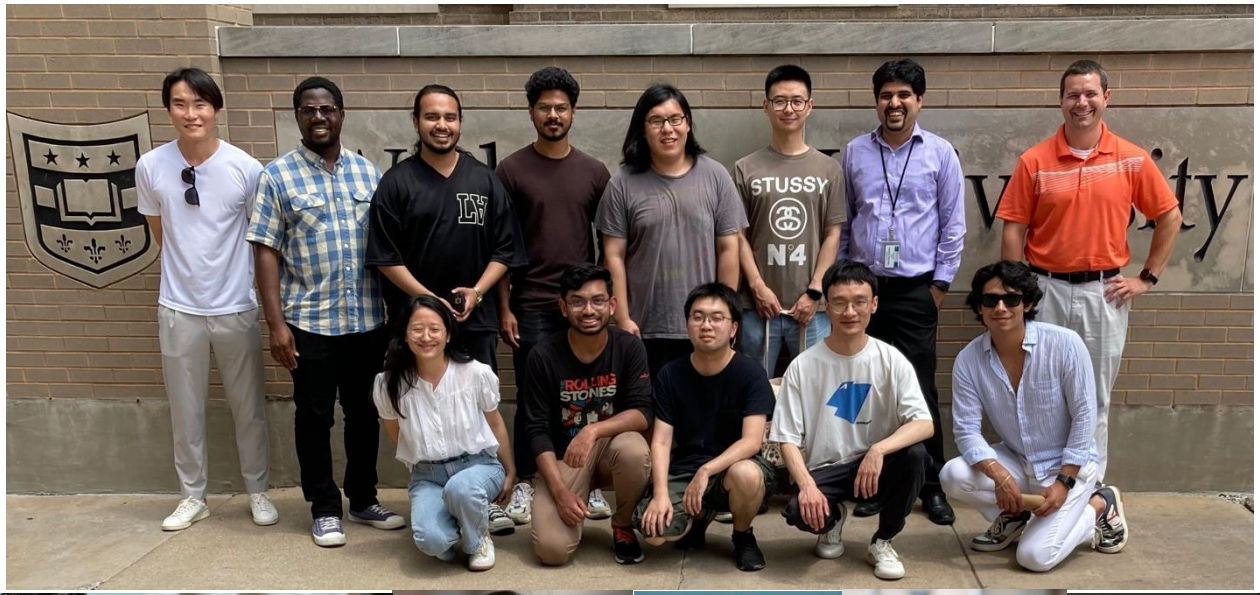
=

Transform medical imaging



Acknowledgements









National Institute of Biomedical Imaging and Bioengineering

R21 EB024647 (Trailblazer award)
R01 EB031051
R01 EB031962
R56 EB028287
R01 EB031051-S2
R01 EB031962-S3

NSF CAREER 2239707

SNMMI SOCIETY OF NUCLEAR MEDICINE & MOLECULAR IMAGING

WashU Here and Next AI for Health Seed Grant

Thank you!

Welcome questions!



jhalab.wustl.edu



a.jha@wustl.edu



jhalab_washu