When AI comes for your job: the radiology story

Robert M. Nishikawa, Ph.D.

Department of Radiology

University of Pittsburgh

Disclosures: Research contracts with: GE Healthcare, Hologic Inc., Koios Medical
Advisory committee: MaiData Corp., iCAD Medical
Royalties: Hologic, Inc.

So What, Who Cares

- Implementation and how the AI tool is actually used in the real-world are of paramount importance
- Will discuss these concepts from the perspective of AI in radiology
- AI is and will transform how radiologists do their jobs leading to better patient care
- Will begin with a brief history of AI in radiology

Sept. 29, 2016

The NEW ENGLAND JOURNAL of MEDICINE

Predicting the Future — Big Data, Machine Learning, and Clinical Medicine

Ziad Obermeyer, M.D., and Ezekiel J. Emanuel, M.D., Ph.D.

Machine learning will displace much of the work of radiologists.

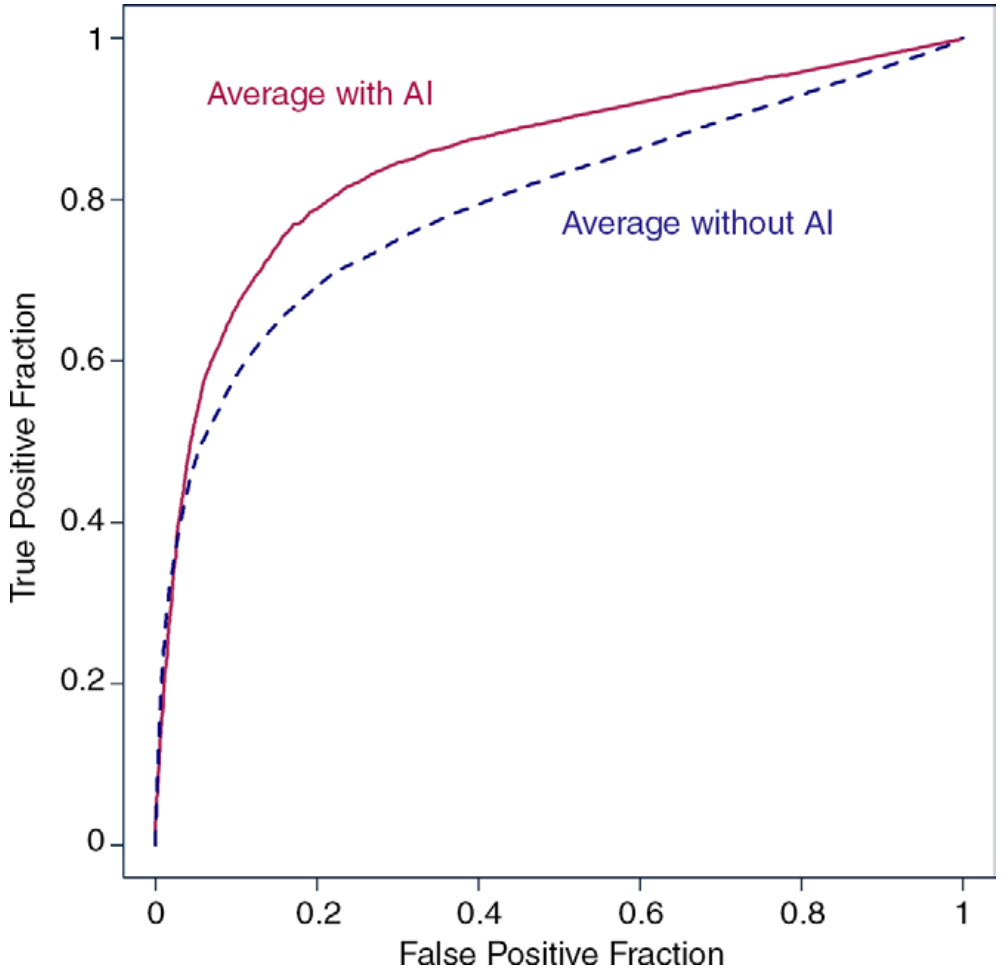
Emanuel suggested that computers may replace radiologists in the next 4-5 years (ACR 2016 keynote)

Incorporating AI into Daily Practice

- Current AI developments are scratching the surface of what radiologists do routinely
  - Radiologists do much more than just read images
- Most radiologists see AI as a tool to improve their jobs
  - Triaging
  - Lesion segmentation
- Gap in knowledge in understanding how a radiologist will use a tool in actual clinical practice
  - As oppose to how the tool has been designed to be used

AI algorithms testing/regulatory requirements

- Observer studies (simulated clinical reading not real world)
  - ~30 radiologists reading ~300 cases
  - Read once with AI tool and once without



There is no post-approval evaluation required
Real-world data

Radiologists are not using the CAD tool as intended

Study	Sensitivity			False Positive Rate (per 100 women screened)		
	Prior to CAD	With CAD	% Change	Prior to CAD	With CAD	% Change
Fenton (2007)	0.80	0.84	4.5%	10.1	13.2	30.7%
Fenton (2011)	0.80	0.81	1.8%	8.4	8.9	5.6%
Lehman (2015)	0.87	0.85	-2.3%	9.1	8.7	-4.4%

Nishikawa and Bae. Journal of the American College of Radiology. 2018; 15(1, Part A):49-52

- To avoid racial bias, need to over-collect cases from low representative populations

2019 U.S. Census Bureau Estimates		
Self-identified race	**Percent of population**	**1,000,000 Sample**
Non-Hispanic white	60.3%	600,000
Hispanic and Latino	18.5%	185,000
Black or African American	13.4%	134,000
Asian	5.9%	59,000
Two or more races	2.8%	28,000
Native Americans and Alaska Natives	1.3%	13,000
Native Hawaiians and Other Pacific Islanders	0.2%	2,000

Differences between Radiology and HS

- Radiologists politically powerful	- TSA agents are not powerful
- Deeply involved in AI development	- Involvement in AI development?
- Dependent component in chain: they consult with other types of MDs	- End of security chain
- Images are saved: medical record	- Images are not saved long-term
- Multi-modality imaging	- Single type of image to review
- Prevalence of disease is low	- Prevalence of target extremely low
- Radiologist work at own pace	- TSA agent make quick decisions
- Radiologist “don’t know” for any decision whether they are correct	- Agent may “know” TP and FP decisions
- Radiologists are “allowed” to make mistakes	- TSA agent cannot make a mistake that would lead to a catastrophic event

Baggage checking is most similar to mammographic screening

- Mammographic screening is only looking for breast cancer
- Multiple different targets indicating breast cancer
- Every mammogram looks different (i.e., no consistent landmarks)
- Prevalence is low
- FP rate is relatively high
- All women are recommended to be screened
  - very few known risk factors

AI and Screening Mammography

- AI will eventually read all mammograms
  - AI is equivalent to an expert breast radiologist
- First step is triaging
  - AI selects cases that do not need to read by a radiologist

AI and Baggage Screening

- If AI equivalent to an expert baggage screener can follow mammography example
  - Triaging and eventually all screened bags are read by AI only
    - Will still need hand checking by humans

How should AI/ATR be tested

- Need to compare performance of AI to humans
  - Collect a large number of screens by humans
  - Retrospective run AI on those cases and compare to human performance
    - Not just accuracy, but what types of cases are missed and are false positives
- If AI is equivalent to a good baggage screener could do triaging
- If you want to use AI to assist screener, then
  - Need to test screener performance with and without AI (step 1)
  - Need to collect data in real world