Data 146: Foundations for CPH

Case Study of AI and Radiology

Irene Y. Chen

“We should stop training radiologists now. It’s just completely obvious that within five years, deep learning is going to do better than radiologists.”

- Geoff Hinton (2016)

It’s 2025. Why are there still radiologists?

A bit about Irene

- Assistant Professor, UC Berkeley and UCSF
- EECS PhD MIT, GEMS Certificate from Harvard-MIT Health Sciences Technology program
- Postdoc, Microsoft Research
- Data Scientist, Dropbox
- AB/SM, Harvard
- Enumerator, US Census Bureau
- Hobbies: fiction reading, pickleball

University of California San Francisco

Microsoft Research

Dropbox

United States® Census Bureau

Reading List

I like to share what I'm reading to keep me motivated. For

Books I've especially loved are marked with an asterik.

2025

1. The Anthropocene Reviewed, John Green*
2. Make Something Wonderful, Steve Jobs*
3. Welcome to the Monkey House, Kurt Vonnegut*
4. all about love, bell hooks
5. Dubliners, James Joyce*
6. All Fours, Miranda July*
7. Orbital, Samantha Harvey
8. The Body Keeps the Score, Bessel Van Der Kolk*
9. The Power Broker, Robert A. Caro*
10. Intermezzo, Sally Rooney
11. Martyr!, Kaveh Akbar*
12. Home Fire, Kamila Shamsie*
13. Goddess Complex, Sanjena Sathian*

A little about you

Course Logistics

- Live Syllabus linked on Bcourses
- Virtual classes:
  - Tues Nov 25 (Thanksgiving week)
  - Tues Dec 2 (NeurIPS)
  - Thurs Dec 4 (NeurIPS)
- Office Hours Wed 9-10am in Warren 120 Suite
- More details about final project later this week

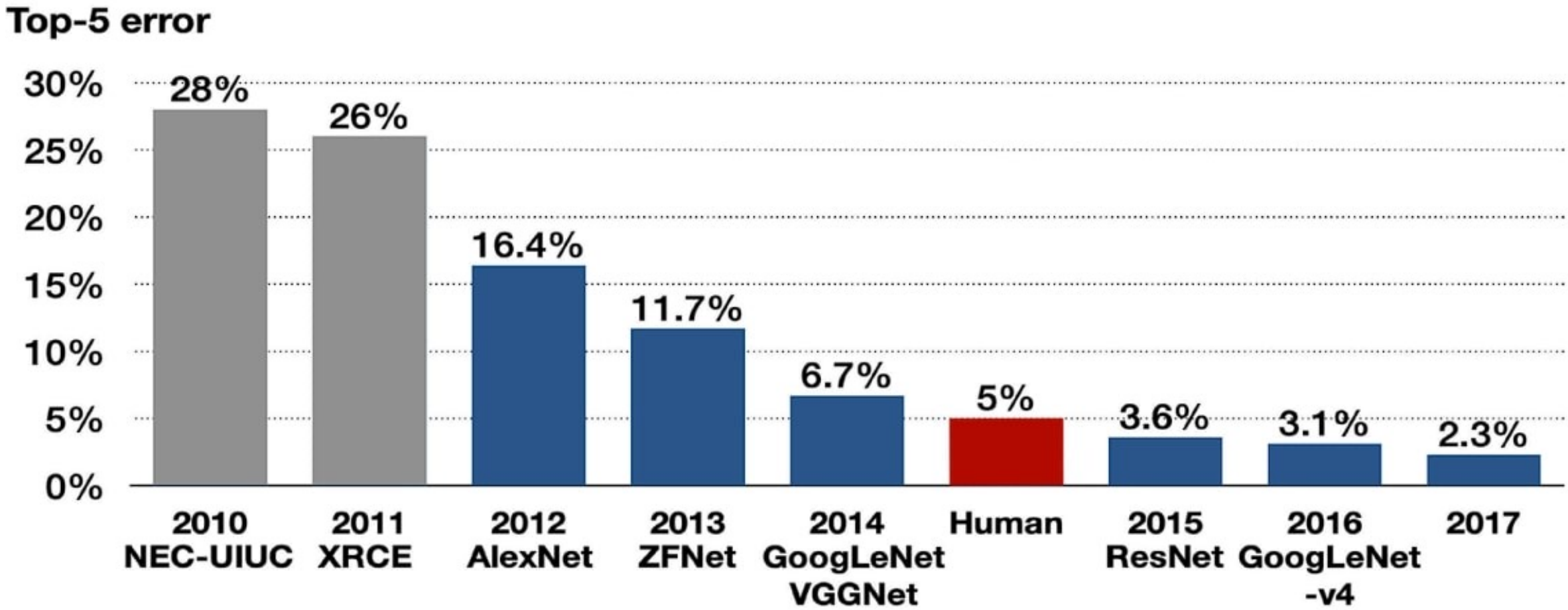
Outline

- **Course logistics** (5 mins)
- **Geoff Hinton on radiologists + discussion** (10 mins)
- **Introduction on AI for Radiology** (35 mins)

Learning Objective: Introduce second half of Data 146, understand challenges to AI for radiology

“We should stop training radiologists now. It’s just completely obvious that within five years, deep learning is going to do better than radiologists.”

- Geoff Hinton (2016)

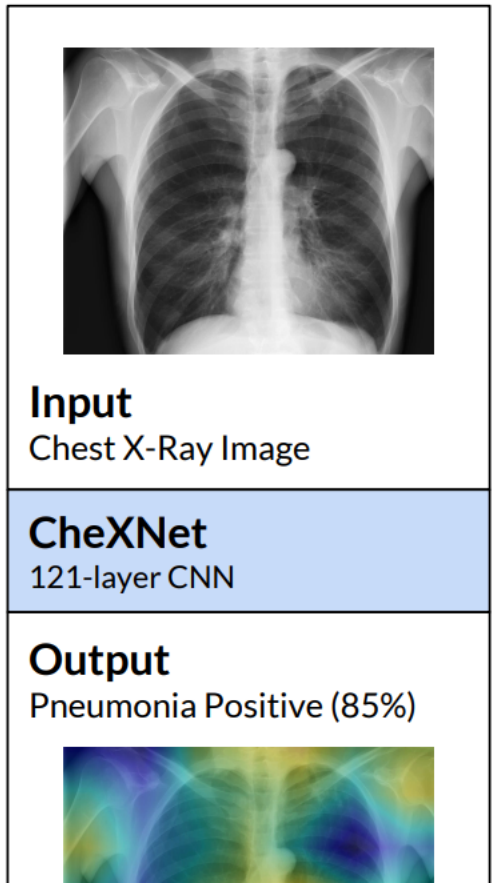

Top-5 error
30%
25%
20%
15%
10%
5%
0%
28%
26%
16.4%
11.7%
6.7%
5%
3.6%
3.1%
2.3%
2010
NEC-UIUC
2011
XRCE
2012
AlexNet
2013
ZFNet
2014
GoogLeNet
VGGNet
Human
2015
ResNet
2016
GoogLeNet
-v4
2017

CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

Pranav Rajpurkar [*1] **Jeremy Irvin** [*1] **Kaylie Zhu** [1] **Brandon Yang** [1] **Hershel Mehta** [1]
Tony Duan [1] **Daisy Ding** [1] **Aarti Bagul** [1] **Robyn L. Ball** [2] **Curtis Langlotz** [3] **Katie Shpanskaya** [3]
Matthew P. Lungren [3] **Andrew Y. Ng** [1]

Abstract

We develop an algorithm that can detect pneumonia from chest X-rays at a level exceeding practicing radiologists. Our algorithm, CheXNet, is a 121-layer convolutional neural network trained on ChestX-ray14, currently the largest publicly available chest X-ray dataset, containing over 100,000 frontal-view X-ray images with 14 diseases. Four practicing academic radiologists annotate a test set, on which we compare the performance of CheXNet to that of radiologists. We find that CheXNet exceeds average radiologist performance on the F1 metric. We extend CheXNet to detect all 14 diseases in ChestX-ray14 and achieve state of the art results on all 14 diseases.

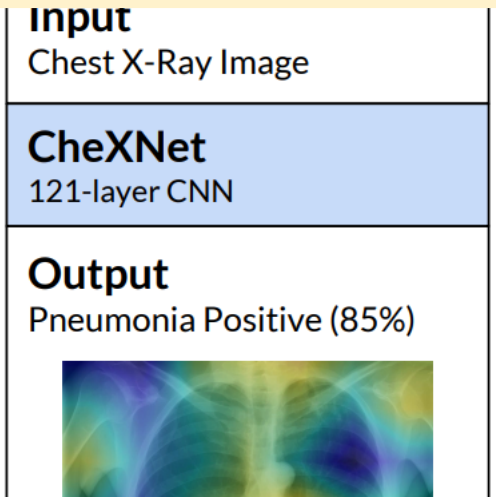


CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

Pranav Rajpurkar [*1] Jeremy Irvin [*1] Kaylie Zhu [1] Brandon Yang [1] Hershel Mehta [1]
Tony Duan [1] Daisy Ding [1] Aarti Bagul [1] Robyn L. Ball [2] Curtis Langlotz [3] Katie Shpanskaya [3]
Matthew P. Lungren [3] Andrew Y. Ng [1]

"CheXNet exceeds average radiology performance on F1 metric"

practicing academic radiologists annotate a test set, on which we compare the performance of CheXNet to that of radiologists. We find that CheXNet exceeds average radiologist performance on the F1 metric. We extend CheXNet to detect all 14 diseases in ChestX-ray14 and achieve state of the art results on all 14 diseases.



At Mayo Clinic, how has the size of the radiology team changed since 2016?

A. Decrease by more than 10%

B. Decrease by 5%

C. About the same

D. Increase by 5%

E. Increase by more than 10%

At Mayo Clinic, how has the size of the radiology team changed since 2016?

A. Decrease by more than 10%

B. Decrease by 5%

C. About the same

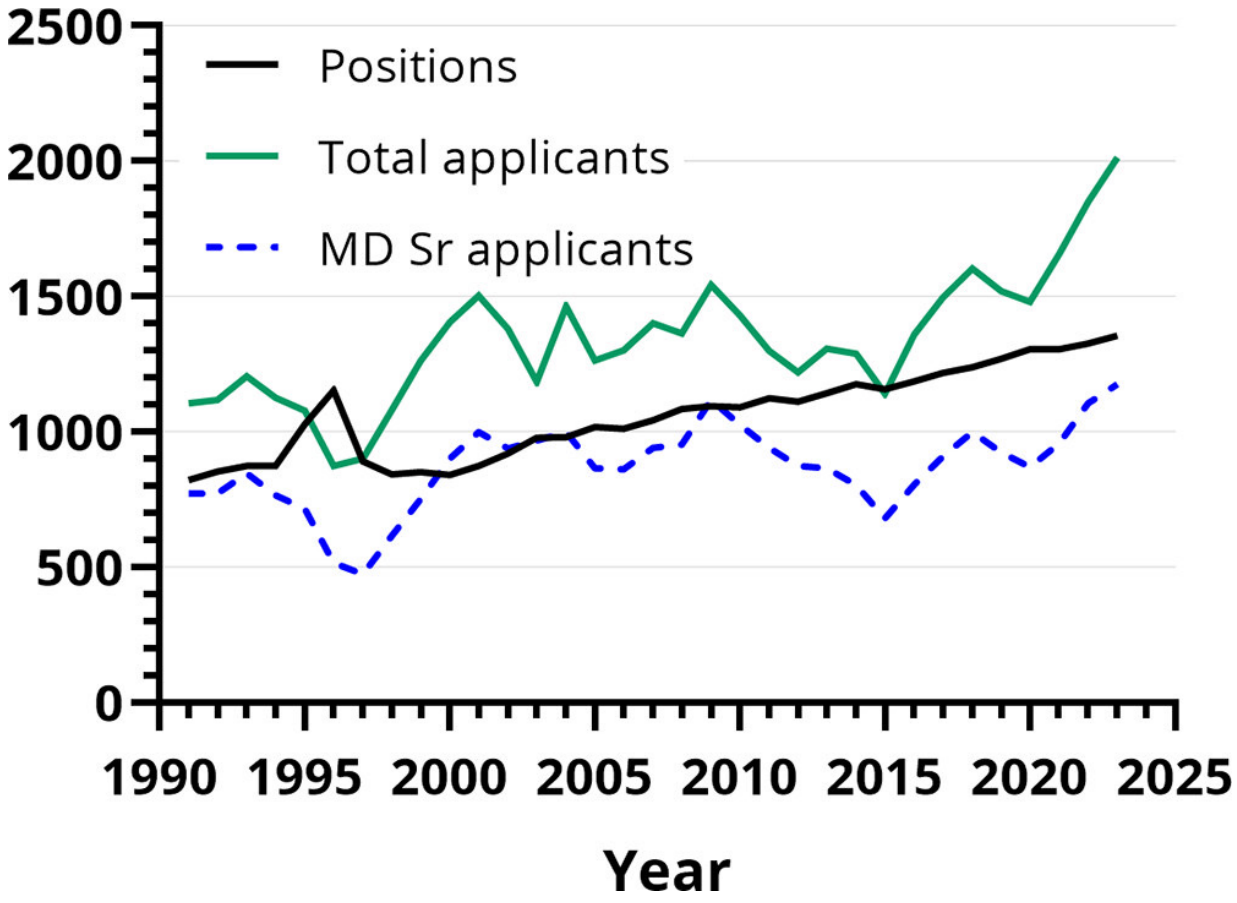
D. Increase by 5%

E. Increase by more than 10% (55% increase)

Radiology Today

- Second-highest-paid medical specialty in US with average income of $520,000
- US diagnostic radiology residency programs offered a record 1,208 positions across all radiology specialties
  - 4% increase from 2024
  - Vacancy rates are at all-time highs

Radiology Today



Deng and Moy, The U.S. Radiology Residency Match: Update and Multidecade Trends, *Radiology* 2023.

What happened?
Where did Geoff go wrong?
(Partner discussion)

Main reasons

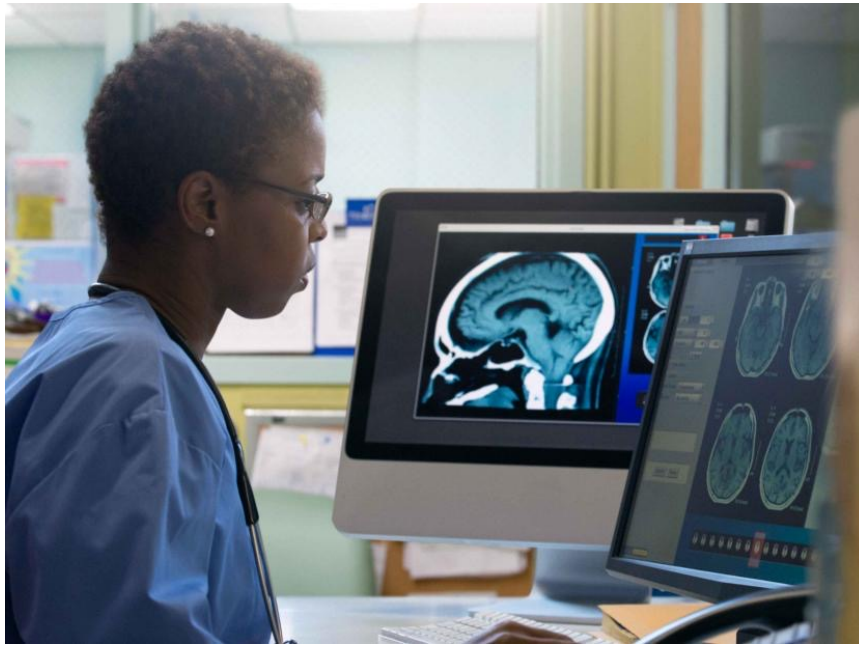
1. Misunderstanding of radiology job specifications
2. Benchmarks didn’t show true performance
3. Implementation and regulatory blockers

What is radiology?

- Field was started when Wilhelm Rontgen discovered x-rays in 1895
- Until 1990s, radiologists used film to examine x-rays, CT scans, and MRI images
- By mid 2000s, hospitals had shifted to digital imagining, including DR and CR
- Radiologists were the first to be digitized!

What is radiology?

- In the US, radiologists undergo **13-15** years of post-HS training:
  - 4 years undergrad
  - 4 years medical school
  - 1 year internship (internal medicine or surgery)
  - 4 years in diagnostic radiology
  - 1-2 years fellowship in a subspecialty
- Responsibilities include diagnosis, consult on treatment and procedures, and perform procedures



What is AI for radiology?

What is AI for radiology?

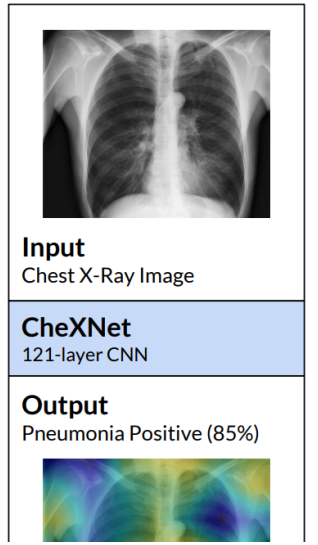
- CheXNet: 121-layer CNN
- Trained on ChestX-ray14 dataset with 100k frontal-view images labeled for 14 diseases
- Labels derived from NLP algorithms of radiology reports (text written by radiologists)
- For labeling pneumonia:
  - **CheXNet F1: 0.435**
  - **Stanford radiologists F1: 0.387**

CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

Pranav Rajpurkar [*1] Jeremy Irvin [*1] Kaylie Zhu [1] Brandon Yang [1] Hershel Mehta [1]
Tony Duan [1] Daisy Ding [1] Aarti Bagul [1] Robyn L. Ball [2] Curtis Langlotz [3] Katie Shpanskaya [3]
Matthew P. Lungren [3] Andrew Y. Ng [1]

Abstract

We develop an algorithm that can detect pneumonia from chest X-rays at a level exceeding practicing radiologists. Our algorithm, CheXNet, is a 121-layer convolutional neural network trained on ChestX-ray14, currently the largest publicly available chest X-ray dataset, containing over 100,000 frontal-view X-ray images with 14 diseases. Four practicing academic radiologists annotate a test set, on which we compare the performance of CheXNet to that of radiologists. We find that CheXNet exceeds average radiologist performance on the F1 metric. We extend CheXNet to detect all 14 diseases in ChestX-ray14 and achieve state of the art results on all 14 diseases.



What is AI for radiology?

- CNN trained on **128k retinal fundus photographs**
- Graded by 3-7 board-certified ophthalmologists
- Binary classification of referable diabetic retinopathy
- **AUC of 0.993** with 98.1% specificity and 90.3% sensitivity

Home | JAMA | Vol. 316, No. 22

Original Investigation | Innovations in Health Care Delivery FREE

Development and Validation of a Deep Learning Algorithm for Detection of Diabetic Retinopathy in Retinal Fundus Photographs

Varun Gulshan, PhD[1]; Lily Peng, MD, PhD[1]; Marc Coram, PhD[1] ; et al

Author Affiliations | Article Information

RELATED ARTICLES FIGURES SUPPLEMENTAL CONTENT

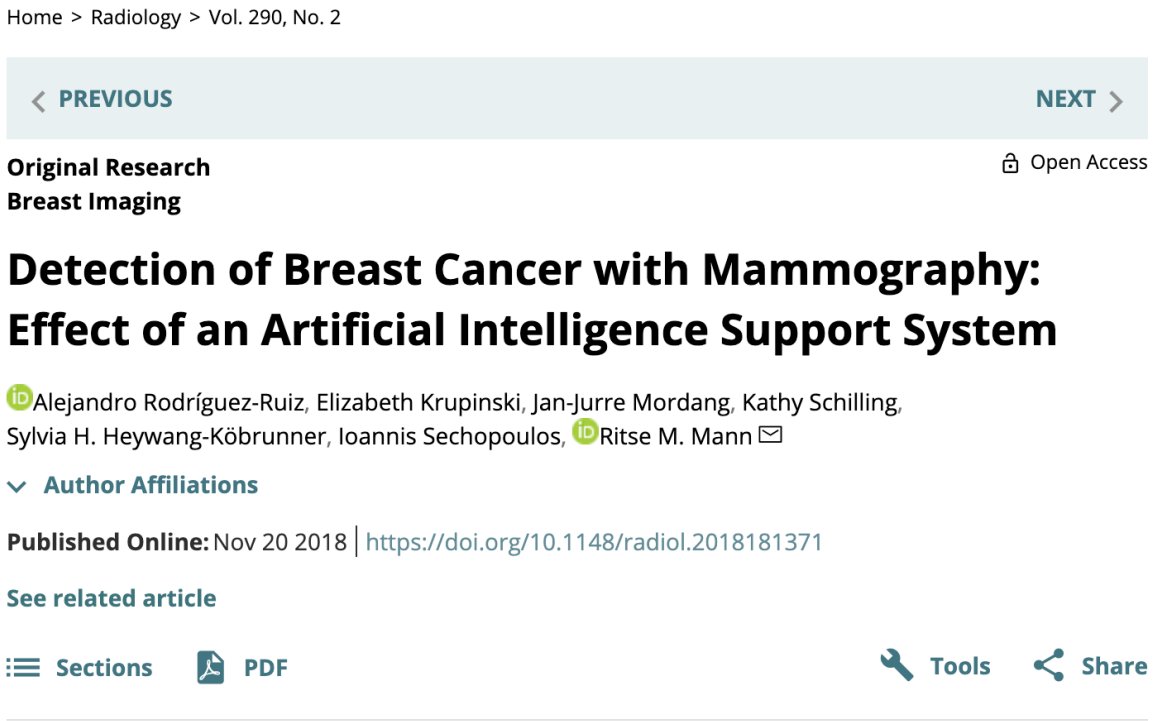
Key Points

Question How does the performance of an automated deep learning algorithm compare with manual grading by ophthalmologists for identifying diabetic retinopathy in retinal fundus photographs?

Finding In 2 validation sets of 9963 images and 1748 images, at the operating point selected for high specificity, the algorithm had 90.3% and 87.0% sensitivity and 98.1% and 98.5% specificity for detecting referable diabetic retinopathy, defined as moderate or worse diabetic retinopathy or referable macular edema by the majority decision of a panel of at least 7 US board-certified ophthalmologists. At the operating point selected for high sensitivity, the algorithm had 97.5% and 96.1% sensitivity and 93.4% and 93.9% specificity in the 2 validation sets.

What is AI for radiology?

- CNN trained on 9000 mammograms with cancer and 180k without
- **Cancer detection AUC:**
  - **0.87 (unaided)**
  - **0.89 (AI-supported)**
- Lesion detection and risk scoring; evaluation by 14 radiologists on an evaluate set of 240 examinations
- Commercial AI system (MammoScreen)



Home > Radiology > Vol. 290, No. 2

PREVIOUS NEXT

Original Research
Breast Imaging

Open Access

Detection of Breast Cancer with Mammography: Effect of an Artificial Intelligence Support System

Alejandro Rodríguez-Ruiz, Elizabeth Krupinski, Jan-Jurre Mordang, Kathy Schilling, Sylvia H. Heywang-Köbrunner, Ioannis Sechopoulos, Ritse M. Mann

Author Affiliations

Published Online: Nov 20 2018 | https://doi.org/10.1148/radiol.2018181371

See related article

Sections PDF Tools Share

777 FDA-approved radiology AIs (out of 1016 total)

777 FDA-approved radiology AIs (out of 1016 total)

aidoc

(Acute critical findings triage platform)

Lunit

(Breast and chest cancer screening)

qure.ai

(Emergency head CT, chest screening)

annalise

(Comprehensive chest X-ray, head CT)

iz.ai

(Stroke care coordination and triage)

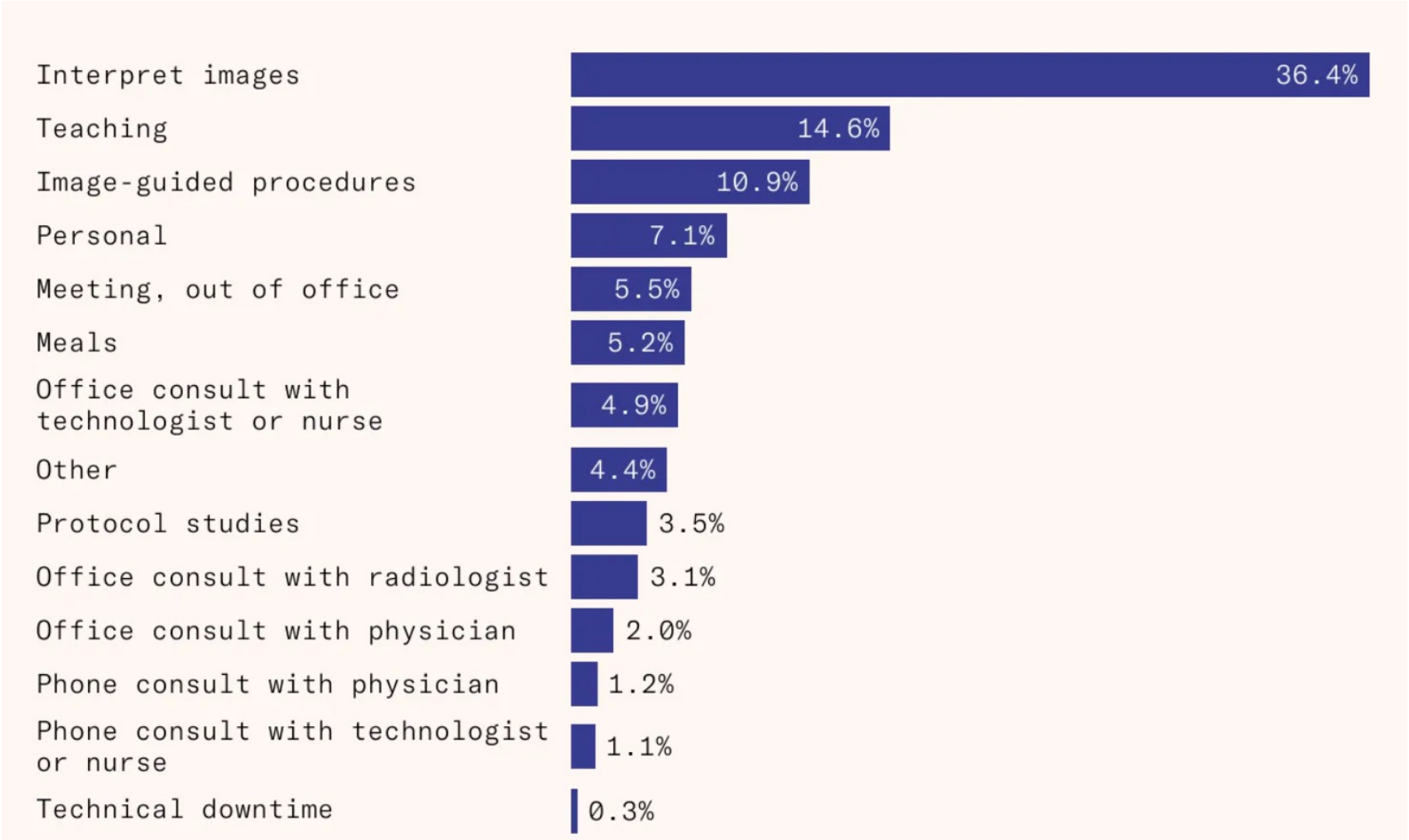
IDx
EyeDiagnosis.com

(Autonomous diabetic retinopathy screening)

Common themes

- Focus on detection/diagnosis
- Label intensive to get labels
- Comparison to clinicians with and without the tool

How much of radiologists' jobs are diagnostic reasoning?



Dhanoa et al, “The evolving role of the radiologist: the Vancouver workload utilization evaluation study”, *Journal of American College of Radiology* 2013

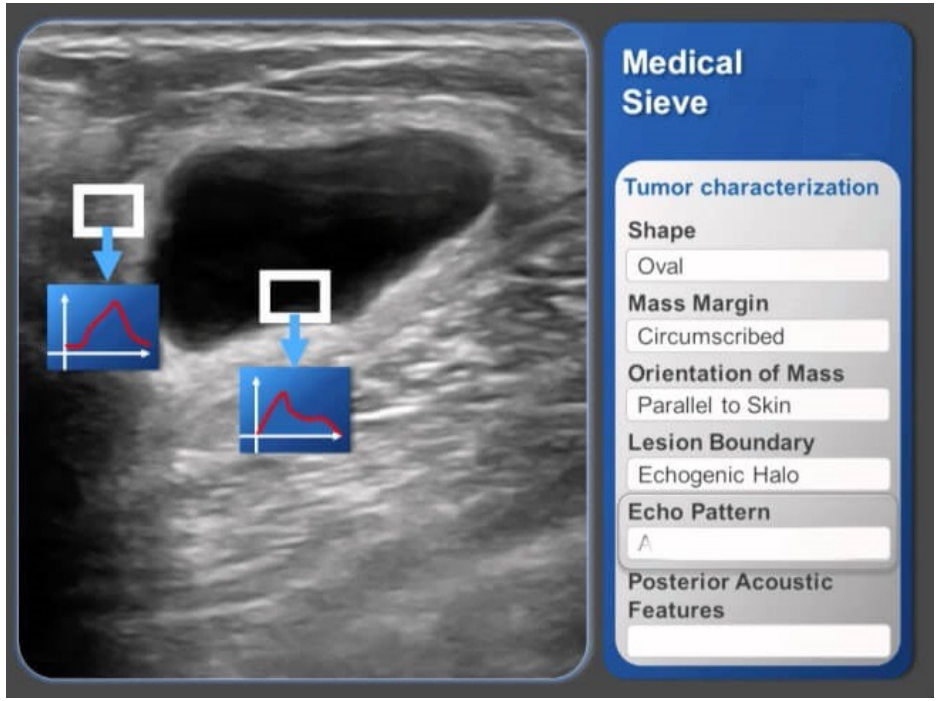
Do radiologists use AI?

- Of 1427 responses:
  - 33.5% of radiologists used AI in the clinic
    - 94.3% reported that AI performance was “inconsistent”
    - 5.7% that it “always worked”
    - 2% that it “never worked”
  - 66.5% not using AI
    - 80% saw “no benefit” in the technology
    - 1/3 couldn’t justify the purchase

Allen B, Agarwal S, Coombs L, Wald C, Dreyer K. 2020 ACR data science institute artificial intelligence survey. J Am Coll Radiol 2021

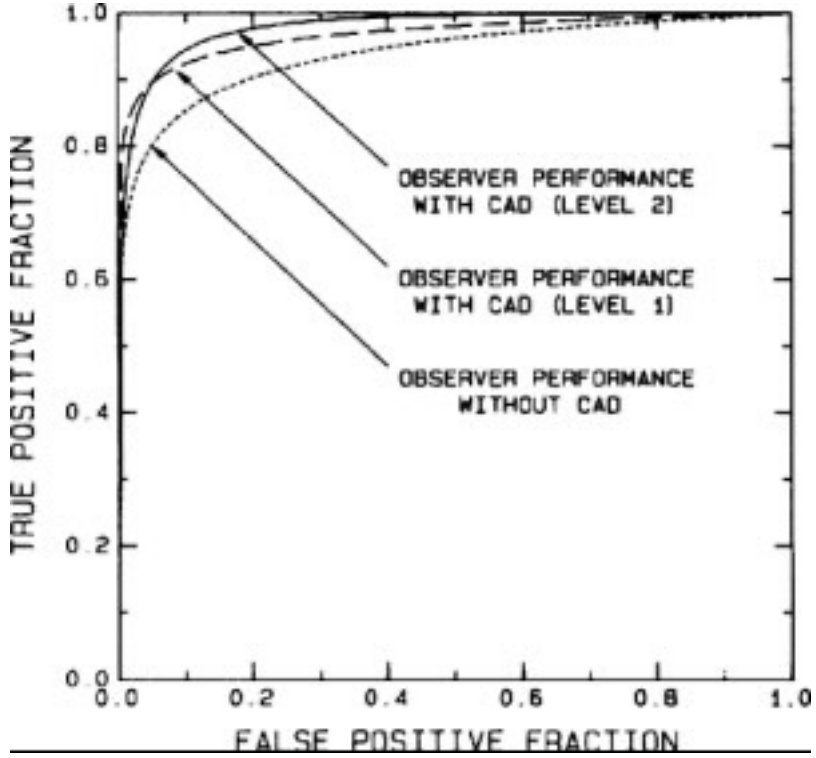
Computer-aided diagnosis

- Precursor to AI:
  1. Preprocessing
  2. Segmentation
  3. Structure analysis of region of interest
  4. Given features, apply favorite ML algorithms (e.g., Naïve Bayes, SVM)
- 1950s: flow charts
- By 1990s: used for screening mammography



Computer-aided diagnosis

- In 2001, Medicare paid $8 more to use CAD for a mammogram than without
- By 2010, more than 74% of mammograms were read by CAD
- CAD actually led to more false positives and no more real cancer
- In 2018, Medicare stopped reimbursing more for mammograms read with CAD



Rao VM, Levin DC, Parker L, Cavanaugh B, Frangos AJ, Sunshine JH. How widely is computer-aided detection used in screening and diagnostic mammography? Journal of the American College of Radiology. 2010;7(10):802-5.

But why did the ML results look so promising?

Next time

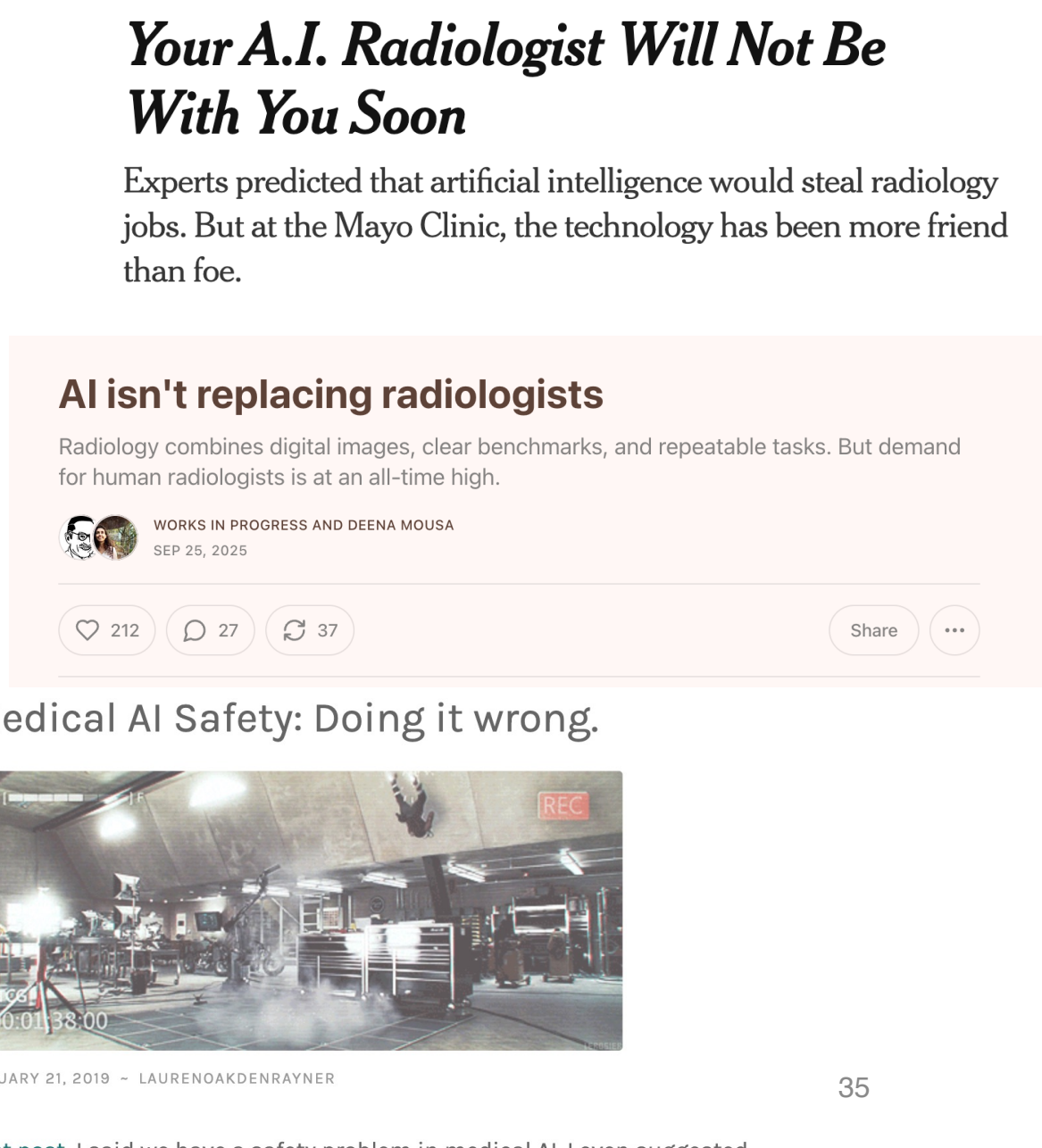
1. Misunderstanding of radiology job specifications
2. Benchmarks didn’t show true performance
3. Implementation and regulatory blockers

Themes for the rest of the class

- AI and the Workforce
- Health Datasets
- Measurement and Evaluation
- AI Policy and Regulation
- Interpretability
- Real-world Impact and Ethics

More reading

- Lohr, "Your A.I. Radiologist Will Not Be With You Soon", New York Times, May 2025
- Mousa, "AI Isn't Replacing Radiologists", Works in Progress Blog, Sept 2025.
- Oakden-Rayner, "Medical AI Safety: Doing it Wrong", Personal blog, Jan 2019



Your A.I. Radiologist Will Not Be With You Soon

Experts predicted that artificial intelligence would steal radiology jobs. But at the Mayo Clinic, the technology has been more friend than foe.

AI isn't replacing radiologists

Radiology combines digital images, clear benchmarks, and repeatable tasks. But demand for human radiologists is at an all-time high.

WORKS IN PROGRESS AND DEENA MOUSA
SEP 25, 2025

212 27 37 Share

Medical AI Safety: Doing it wrong.

JANUARY 21, 2019 ~ LAURENOAKDENRAYNER

Summary

- ✓ **Course logistics** (5 mins)
- ✓ **Geoff Hinton was wrong about radiologists** (10 mins)
- ✓ **Intro to AI for radiology + radiology do more than diagnostic imaging** (35 mins)

How can we make Data 146 better for you?

Next Class: AI and radiology benchmarks and regulation