Stanford Children's Health | Lucile Packard Children's Hospital Stanford

Challenges of Deploying and Validating an AI Tool into Medical Practice

Safwan S. Halabi MD

Clinical Associate Professor

Department of Radiology

March 19, 2019

GPU TECHNOLOGY CONFERENCE

Disclosures

Advisor

Bunker Hill

Interfierce (CMO)

DNAFeed

Board Member, Society of Imaging Informatics in Medicine

Member, RSNA Informatics Committee

Chair, Data Science Standards Subcommittee

Motivations

Diagnostic errors play a role in up to 10% of patient deaths

21 percent of adults report having personally experienced a medical error

4% of radiology interpretations contain clinically significant errors

Improving Diagnosis in Health Care. National Academy of Medicine. Washington, DC: The National Academies Press, 2015.
Americans' Experiences with Medical Errors and Views on Patient Safety. Chicago, IL: University of Chicago and IHI/NPSF, 2017.
Waite S, Scott J, Gale B, Fuchs T, Kolla S, Reede D. Interpretive Error in Radiology. *Am J Roentgenol*. 2016:1-11
Berlin L. Accuracy of Diagnostic Procedures: Has It Improved Over the Past Five Decades? *Am J Roentgenol*. 2007;188(5):1173-1178.

Motivations

Empower radiologists to provide high level diagnostic interpretation in setting of increased volume and limited resources

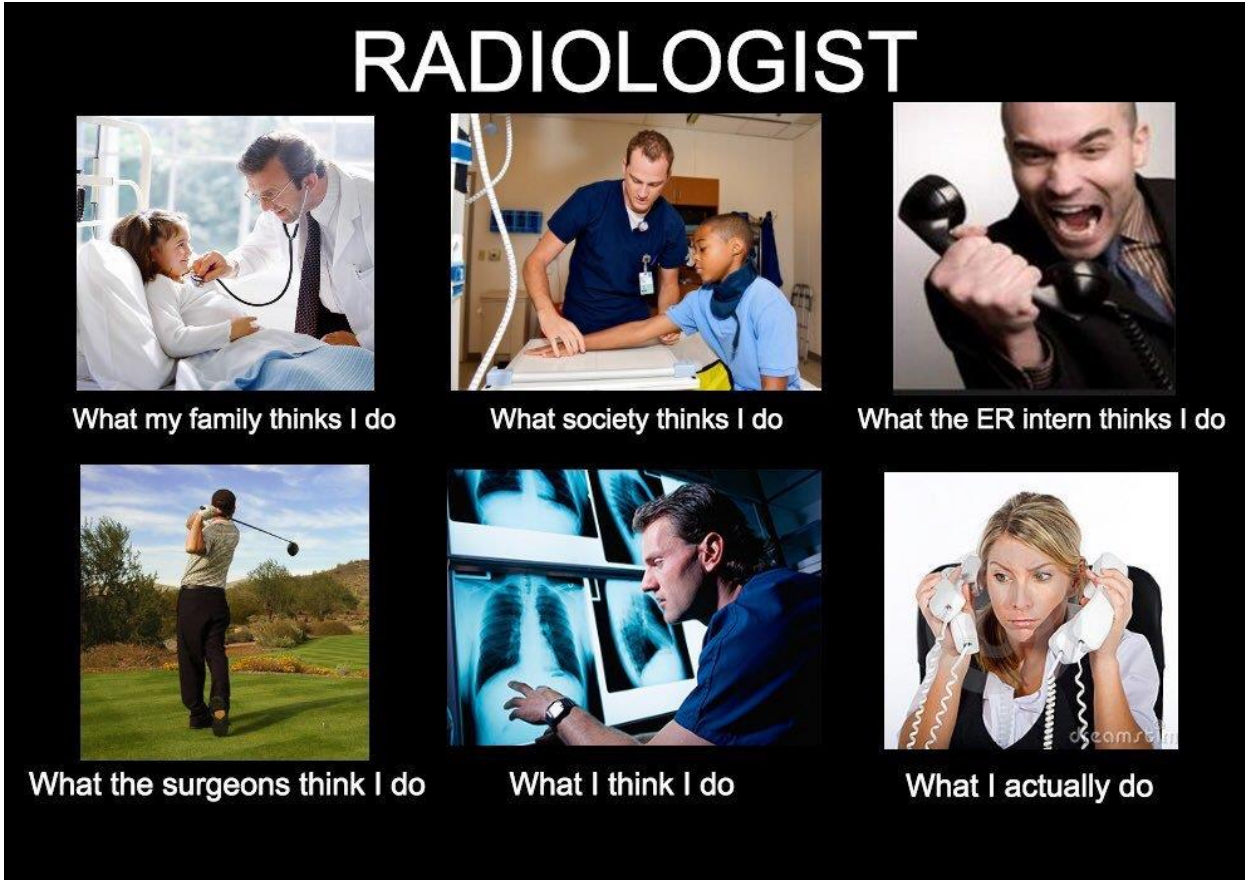
NOT to replace clinicians and radiologists

Radiologist disagreement

- Disagreement with colleagues – 25% of the time
- Disagreement with <u>themselves</u> – 30% of the time

Abujudeh, HH, Boland, GW, Kaewalai, R, et al. Abdominal and Pelvic Computed Tomography (CT) Interpretation: discrepancy rates among experienced radiologists. Eur Radiol.2010;20(8): 1952-7

What do radiologists do?


RADIOLOGIST
What my family thinks I do
What society thinks I do
What the ER intern thinks I do
What the surgeons think I do
What I think I do
What I actually do

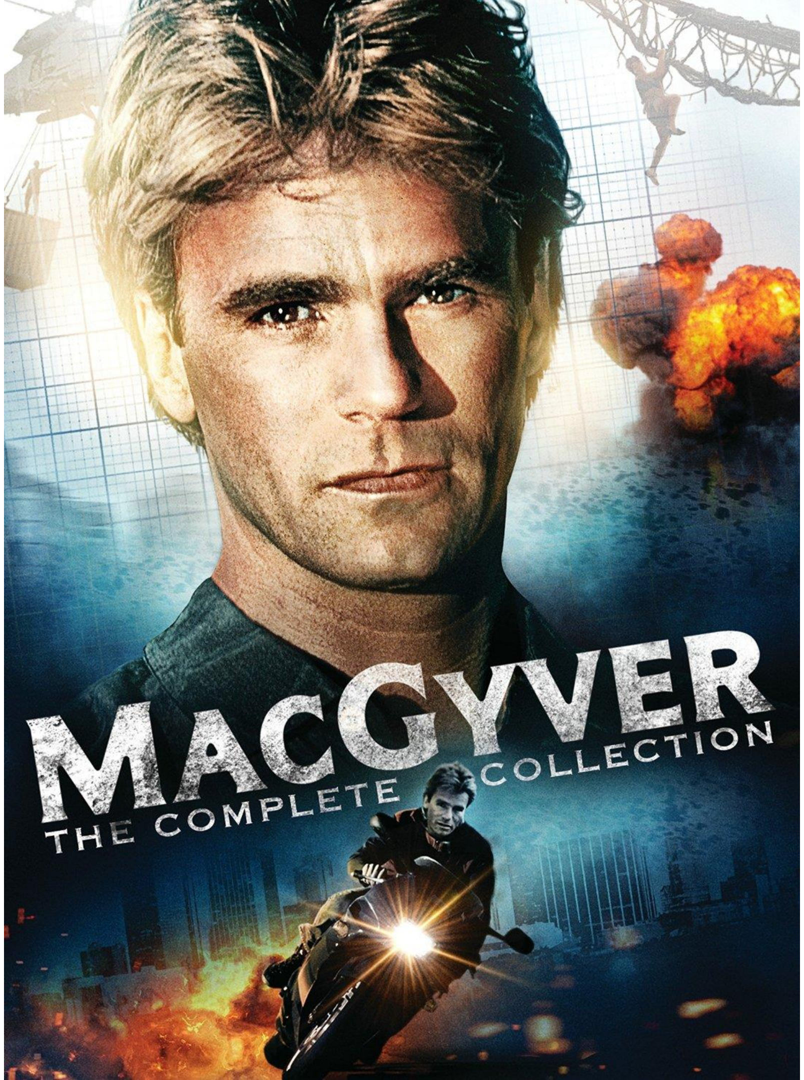
Acting as an expert consultant to your referring physician (the doctor who sent you to the radiology department or clinic for testing) by aiding him or her in choosing the proper examination, interpreting the resulting medical images, and using test results to direct your care

Treating diseases by means of radiation (radiation oncology) or minimally invasive, image-guided therapeutic intervention (interventional radiology)

Correlating medical image findings with other examinations and tests

Recommending further appropriate examinations or treatments when necessary and conferring with referring physicians

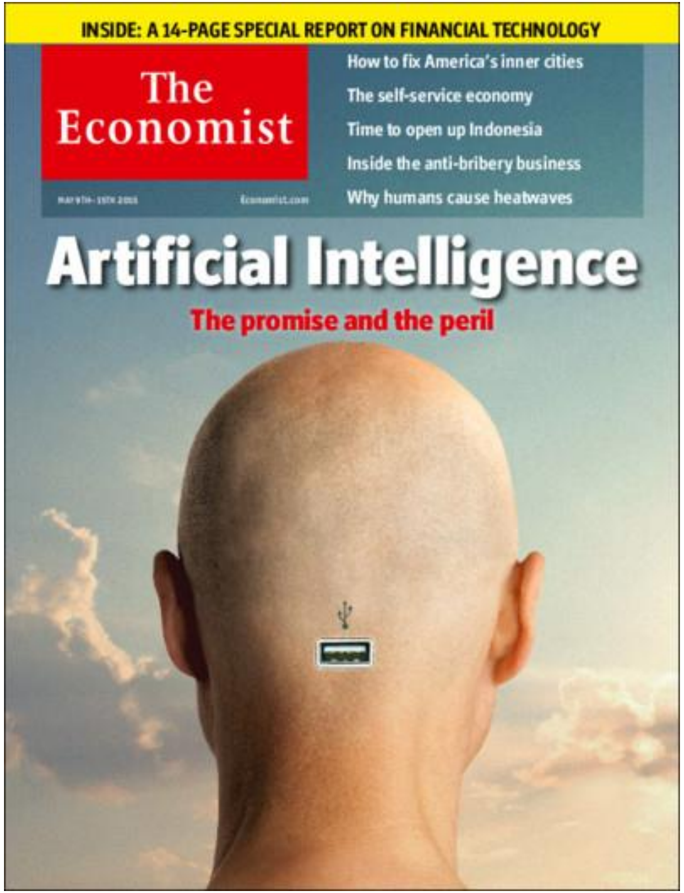
Directing radiologic technologists (personnel who operate the equipment) in the proper performance of quality exams


MacGyver
CBS

MacGyver
THE COMPLETE COLLECTION

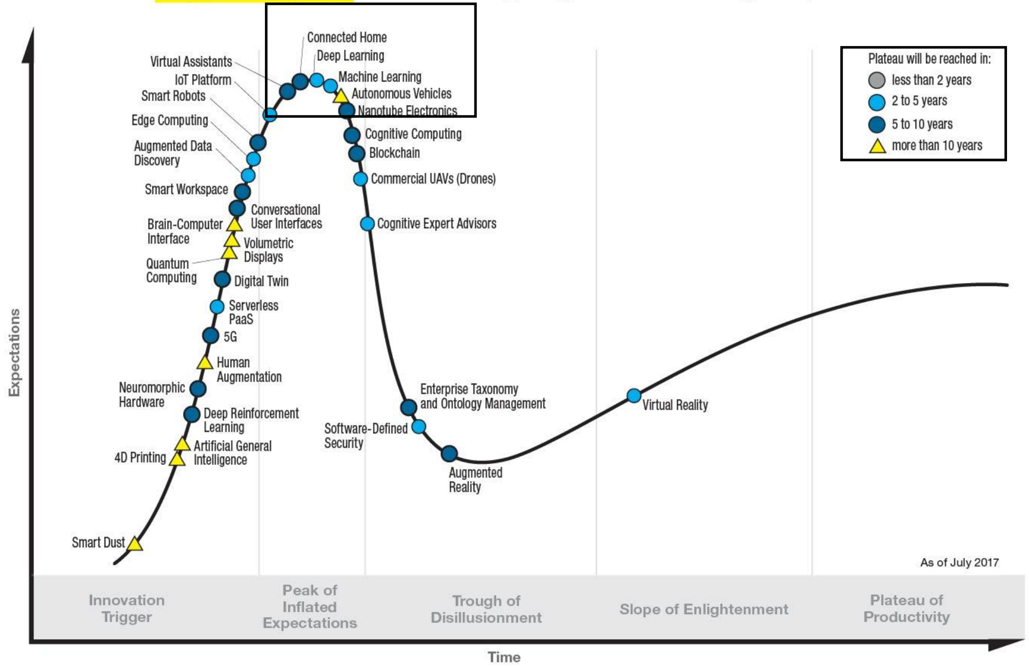

swiss

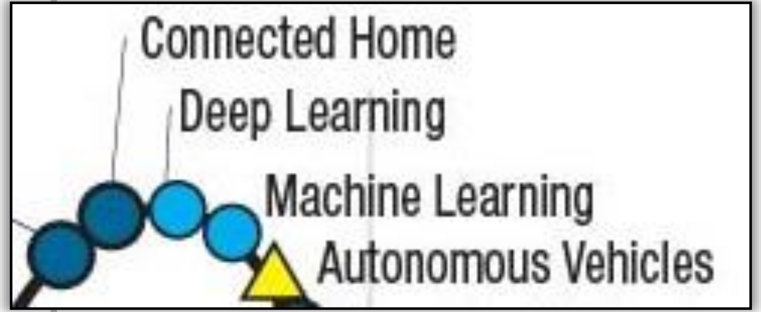
What is AI and Why All the Hype?

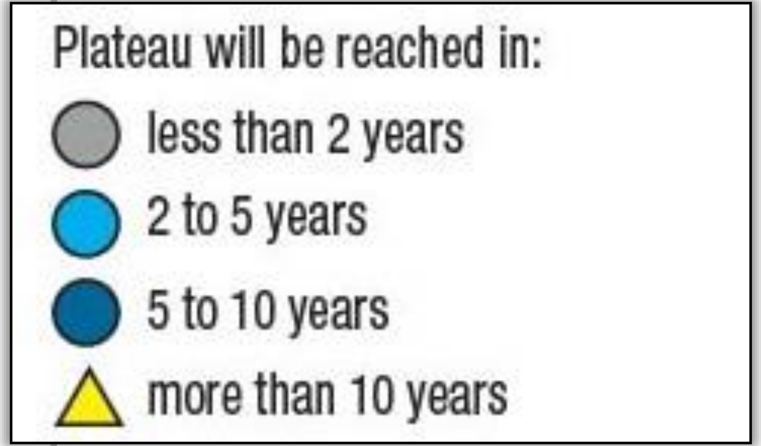

The Economist
Inside China's Ministry of Truth
Trump in trouble
Who are the Niger Delta Avengers?
The flaws in executive pay
Motorcycles that fly
March of the machines
A SPECIAL REPORT ON ARTIFICIAL INTELLIGENCE


INSIDE: A 14-PAGE SPECIAL REPORT ON FINANCIAL TECHNOLOGY
The Economist
How to fix America's inner cities
The self-service economy
Time to open up Indonesia
Inside the anti-bribery business
Why humans cause heatwaves
Artificial Intelligence
The promise and the peril

Gartner Hype Cycle for Emerging Technologies, 2017

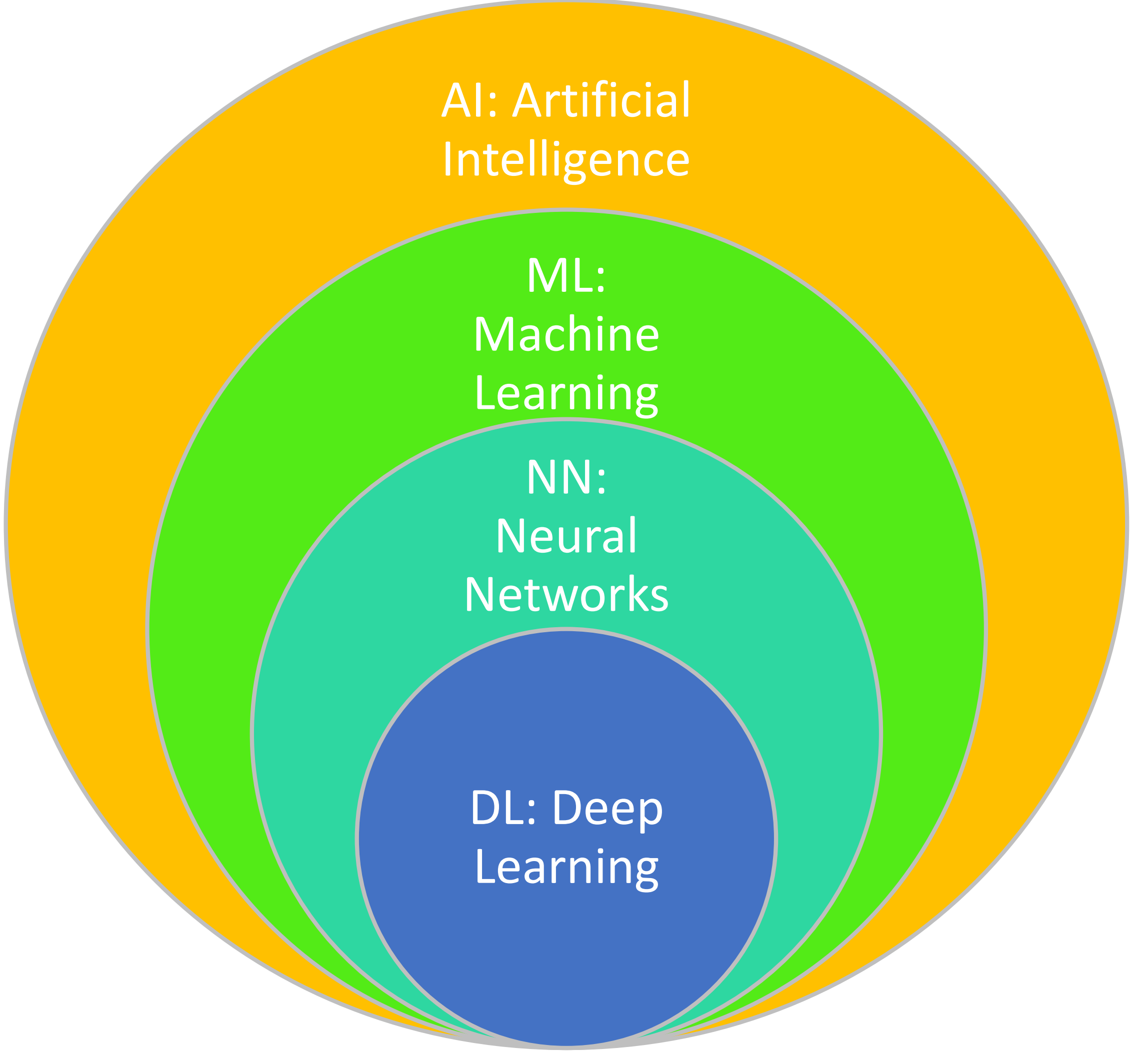






gartner.com/SmarterWithGartner

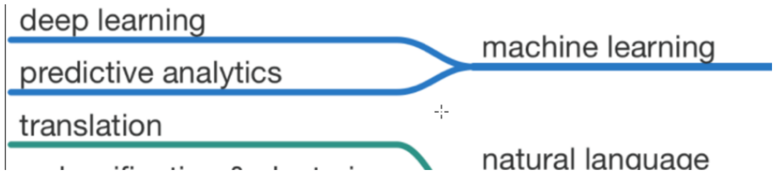
Source: Gartner (July 2017)



Definitions

- AI: When computers do things that make humans seem intelligent
- ML: Rapid automatic construction of algorithms from data
- NN: Powerful form of machine learning
- DL: Neural networks with many layers

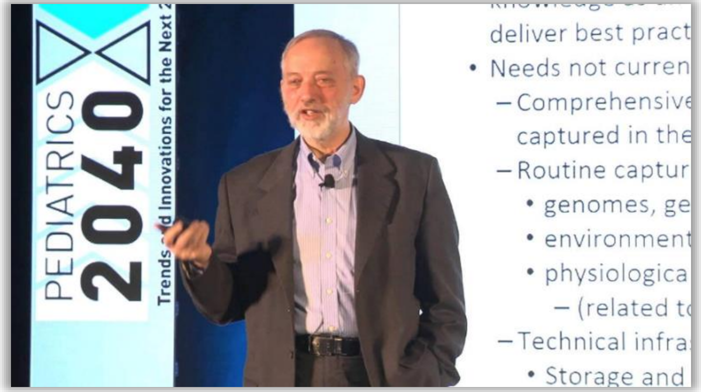
Deep Learning

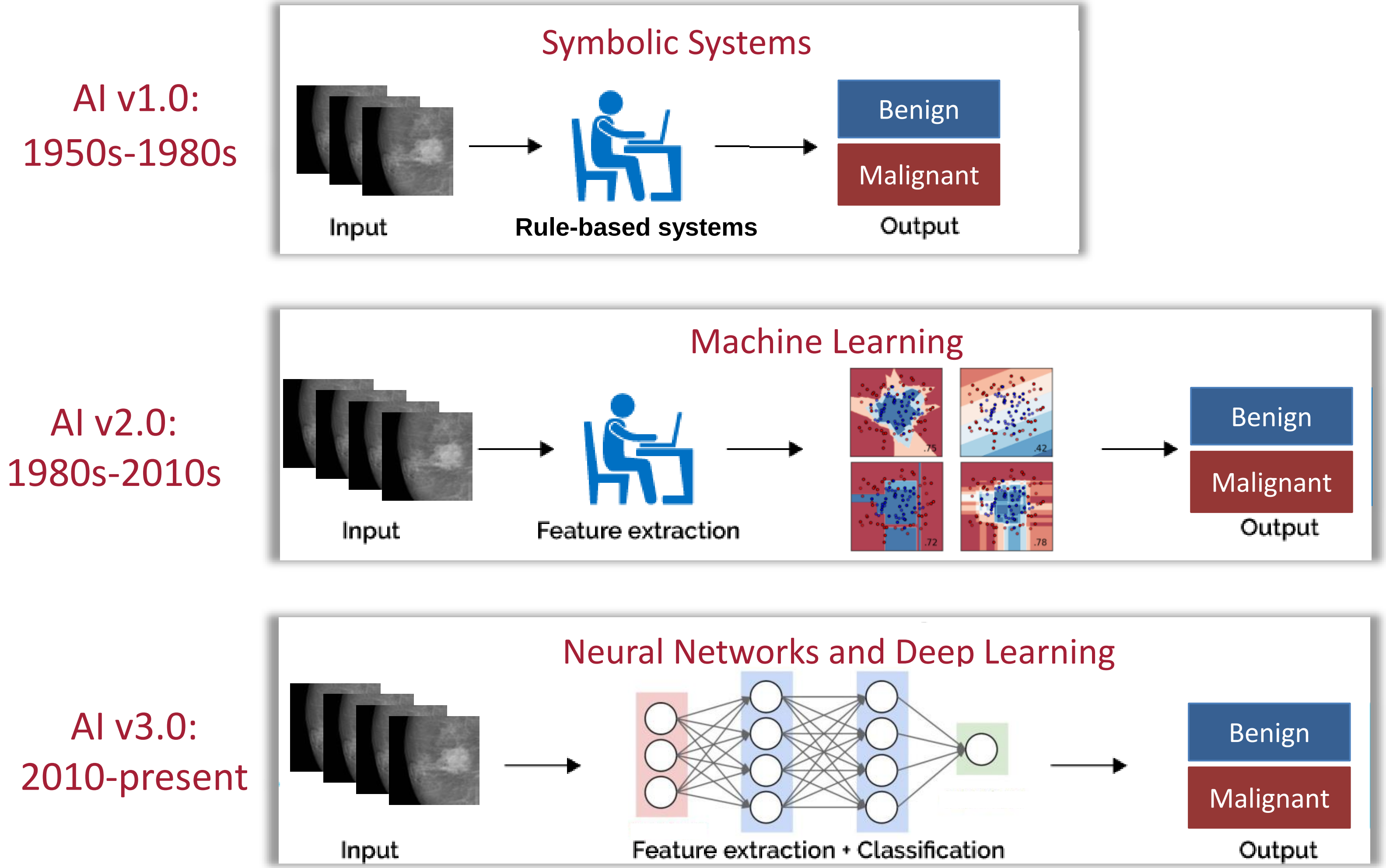


Ability for machines to **autonomously** mimic human thought patterns through artificial neural networks composed of cascading layers of information

“In the 1970s, an AI system that worked for one patient was worth a masters degree; if it worked for three patients, it was a PhD. Now, it's different.”

--Pete Szolovits, #Peds2040, Jan 2016



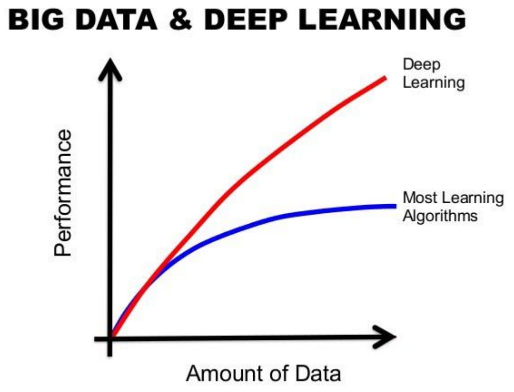


AI v1.0:
1950s-1980s
Symbolic Systems
Input
Rule-based systems
Benign
Malignant
Output
AI v2.0:
1980s-2010s
Machine Learning
Input
Feature extraction
.75
.42
.72
.78
Benign
Malignant
Output
AI v3.0:
2010-present
Neural Networks and Deep Learning
Input
Feature extraction + Classification
Benign
Malignant
Output

Augmented Intelligence

- Systems that are design to **enhance** human capabilities
  - Contrasted with Artificial Intelligence, which is intended to replicate or replace human intelligence
- In healthcare (HC), a more appropriate term is 'augmented intelligence,' reflecting the **enhanced** capabilities of human clinical decision making when coupled with these computational methods and systems

Challenge #1: Dataset



- Collection of data
- Text and/or images

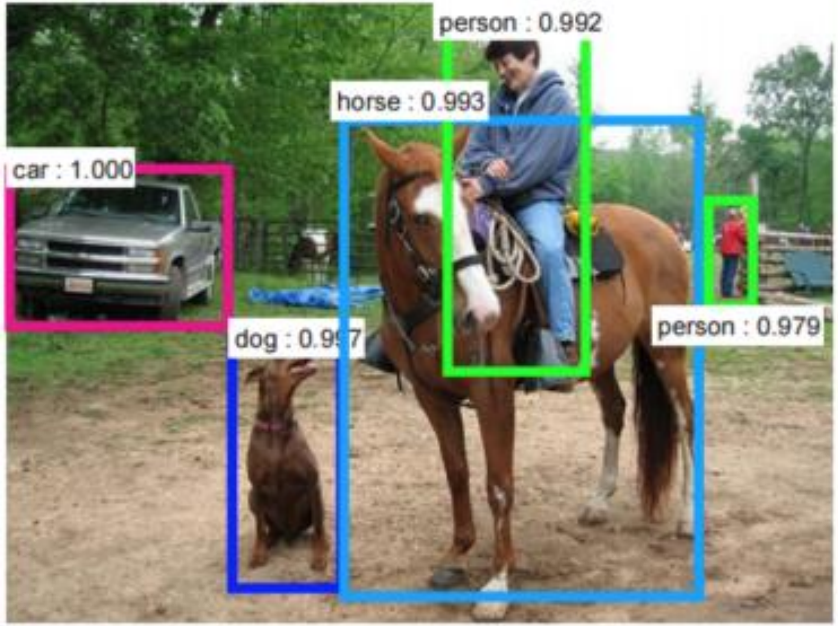
Data Challenges

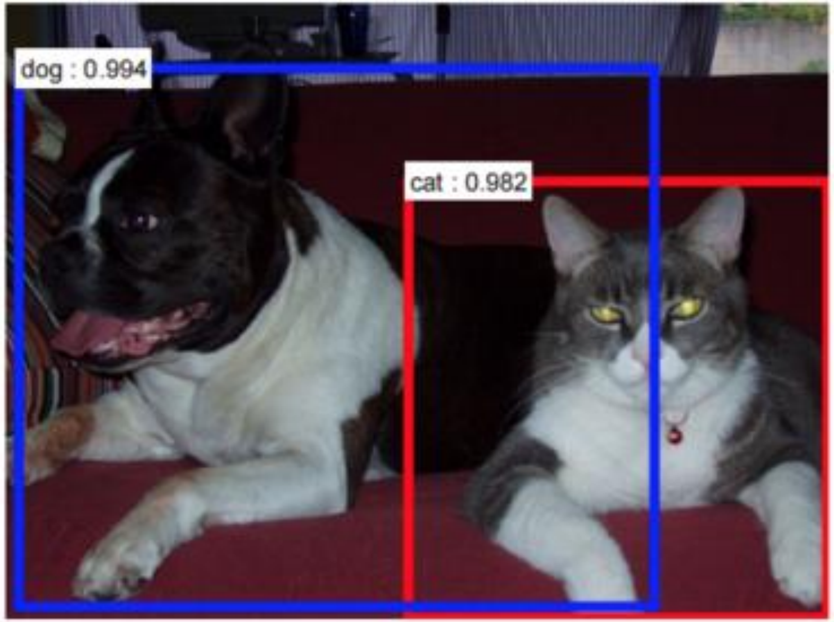
- Do I have enough?
- Balanced?
- Representative?
- Annotated/labeled?
- De-identified?
  - Metadata
  - Facial scrubbing
  - Burned in data
- Sharing rights?

Attribute Name	Tag	Action	Comments
Station Name	(0008,1010)	Removed	Their values are only relevant to the equipment
Device Serial Number	(0018,1000)		
Institution Name	(0008,0080)	Removed	Their values are not normally relevant for research on image processing or aided diagnosis algorithms
Institution Address	(0008,0081)		
Referring Physician's Name	(0008,0090)		
Referring Physician's Address	(0008,0092)		
Referring Physician's Telephone Numbers	(0008,0094)		

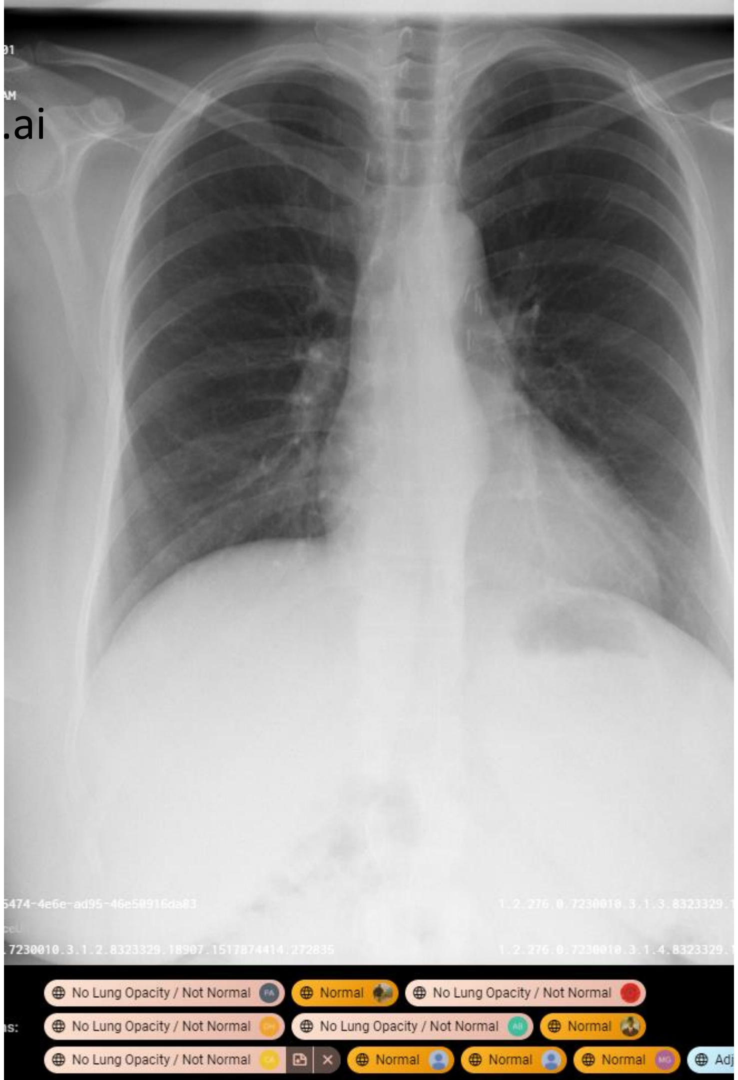
Attribute Name	Tag	Action	Comments
Series Description	(0008,103E)		algorithms
Protocol Name	(0018,1030)		
Patient's Sex	(0010,0040)	Unchanged	Attributes that may be relevant for research algorithms
Patient's Size	(0010,1020)		
Patient's Weight	(0010,1030)		
Requested Procedure Description	(0032,1060)	Unchanged	Their values are important for image processing algorithms
Scheduled Procedure Step Description	(0040,0007)		
Performed Procedure Step Description	(0040,0254)		

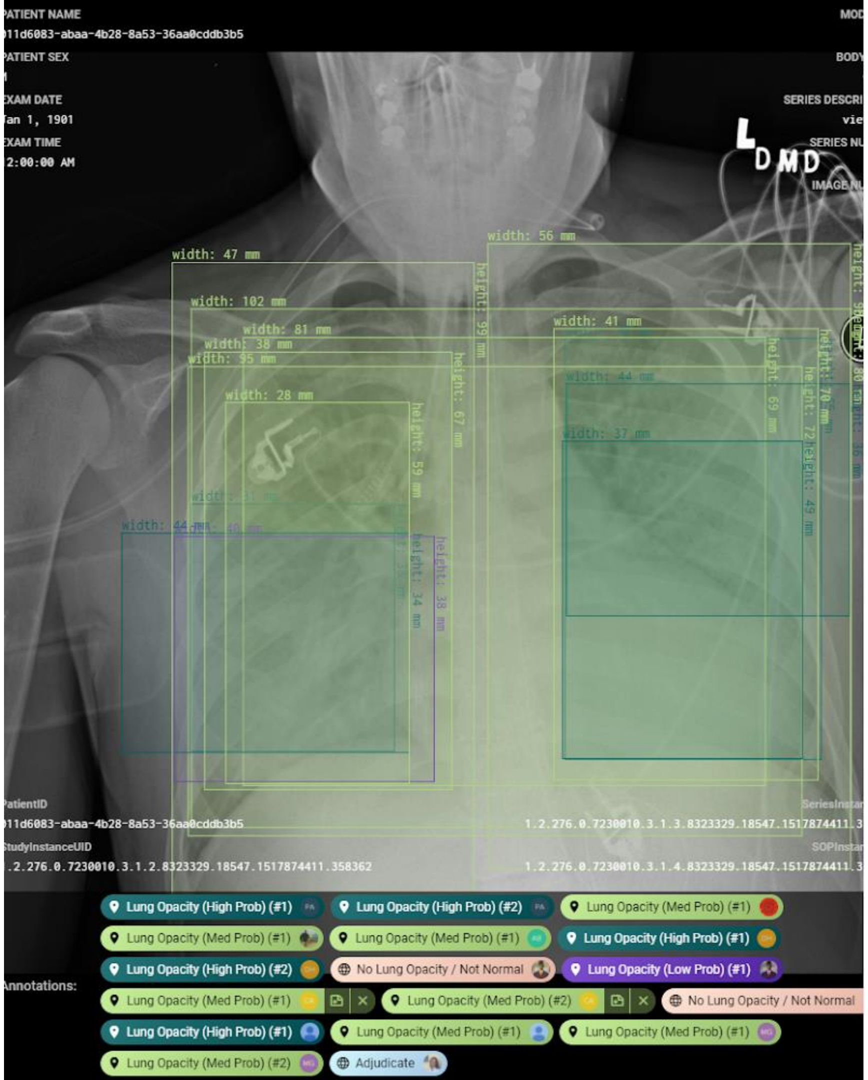
Challenge #2: Annotation



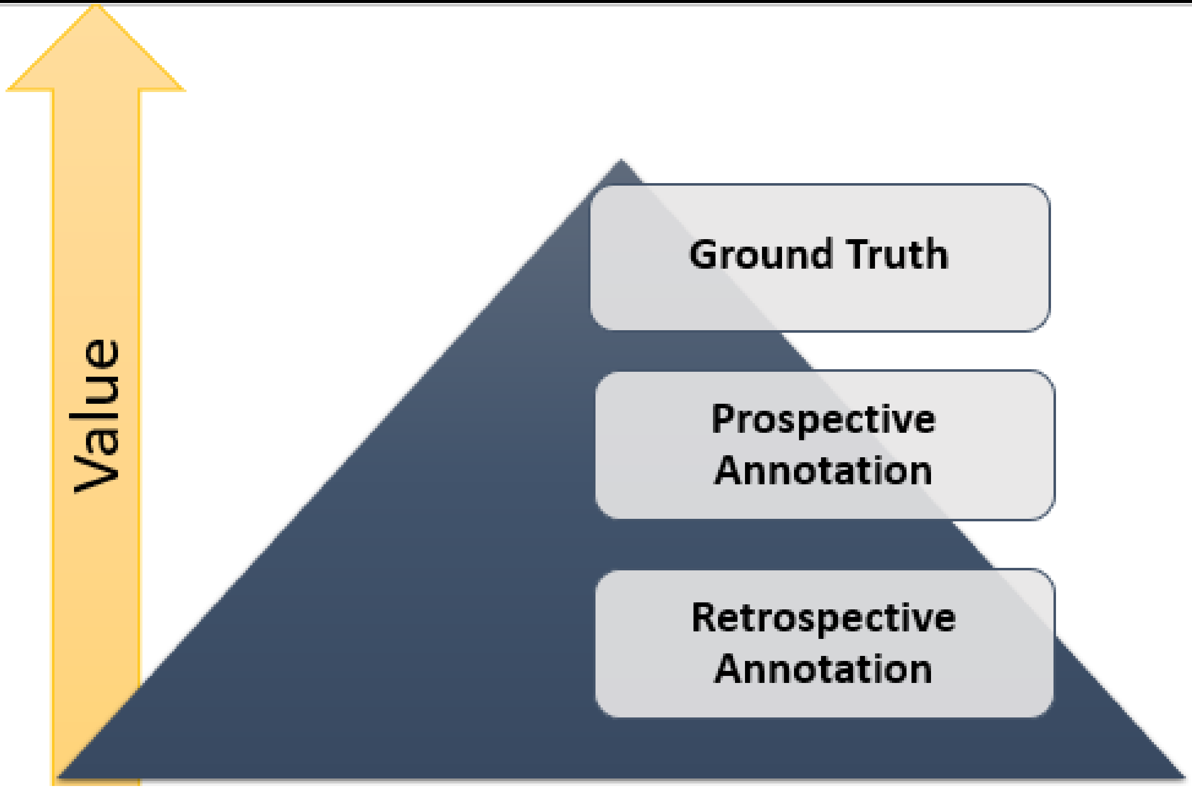


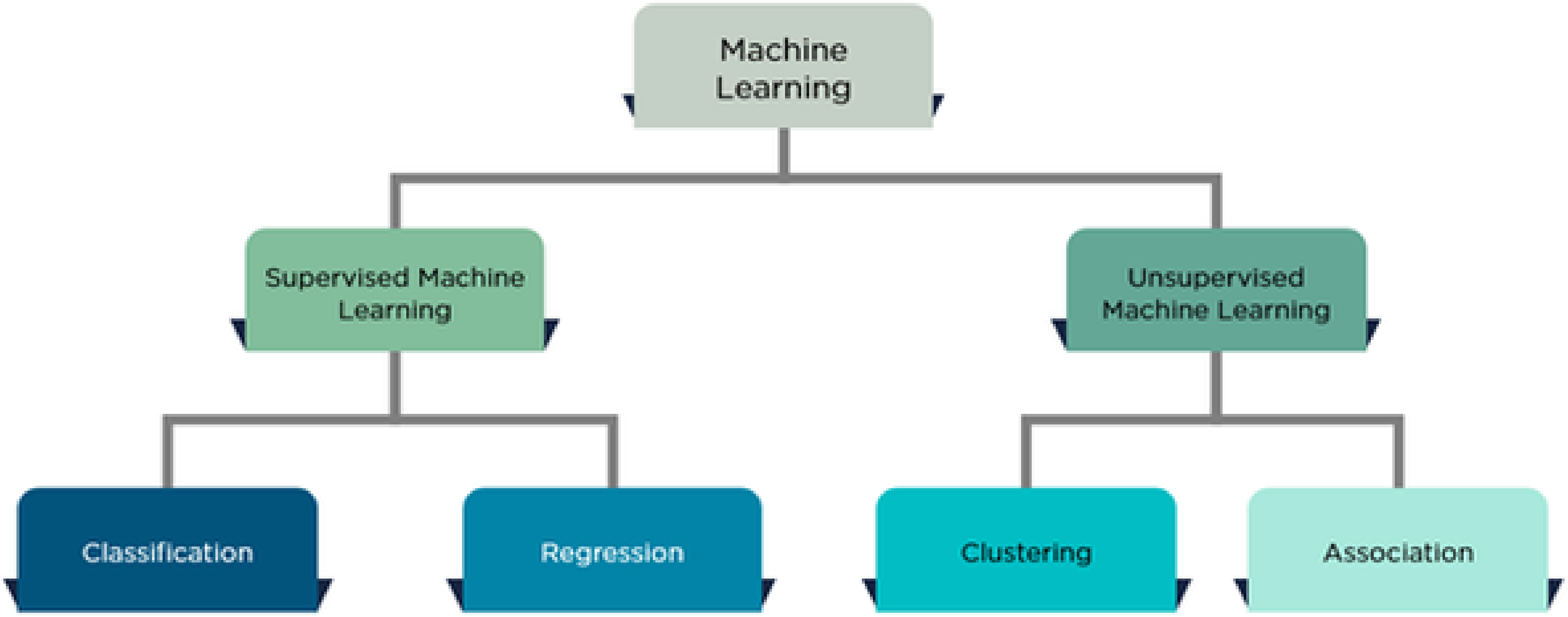
MD.ai

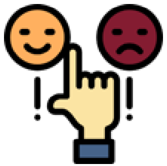

No Lung Opacity / Not Normal
Normal
No Lung Opacity / Not Normal
No Lung Opacity / Not Normal
No Lung Opacity / Not Normal
Normal
No Lung Opacity / Not Normal
Normal
Normal
Normal


PATIENT NAME
EXAM DATE
Jan 1, 1901
EXAM TIME
2:00:00 AM
width: 47 mm
width: 56 mm
width: 102 mm
width: 81 mm
width: 38 mm
width: 41 mm
width: 28 mm
width: 37 mm
PatientID
StudyInstanceUID
Lung Opacity (High Prob) (#1)
Lung Opacity (High Prob) (#2)
Lung Opacity (Med Prob) (#1)
Lung Opacity (Med Prob) (#1)
Lung Opacity (Med Prob) (#1)
Lung Opacity (High Prob) (#1)
Lung Opacity (High Prob) (#2)
No Lung Opacity / Not Normal
Lung Opacity (Low Prob) (#1)
Annotations:
Lung Opacity (Med Prob) (#1)
Lung Opacity (Med Prob) (#2)
No Lung Opacity / Not Normal
Lung Opacity (High Prob) (#1)
Lung Opacity (Med Prob) (#1)
Lung Opacity (Med Prob) (#1)
Lung Opacity (Med Prob) (#2)
Adjudicate

Imaging Annotation Value




Machine Learning
Supervised Machine Learning
Unsupervised Machine Learning
Classification
Regression
Clustering
Association



Classification

Used when the output is categorical like 'YES' or 'NO'

Algorithms used

- Decision Tree
- Naïve Bayes
- Random Forest
- Logistic regression
- KNN

Clustering

Used when the data needs to be organized to find patterns in the case of 'product recommendation'

Regression

Used when a value needs to be predicted like the '**stock prices**'

Algorithms used

- Linear Regression

Classification Models

- Logistic Regression
- Decision Tree
- Random Forest
- Support Vector Machine
- Gradient-Boosted Tree
- Multilayer Perceptron
- Naive Bayes

Algorithms

A set of rules or instructions given to an AI, neural network, or other machine to help it **learn on its own**

Clustering, classification, regression, and recommendations

Logistic Regression

Chest radiographs labeled for presence of pneumonia

If greater the 50% of labels or labelers consider image contains pneumonia, then model considers that image positive for pneumonia

Knee MRI Classifier

- Dataset:

  1400 knee MRI

  3 series

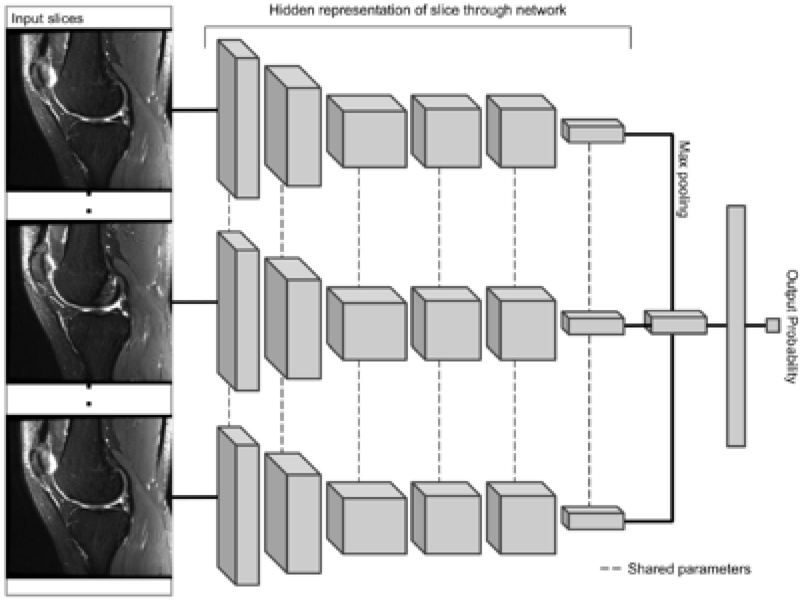
- Labels:

  (1) normal/abnormal

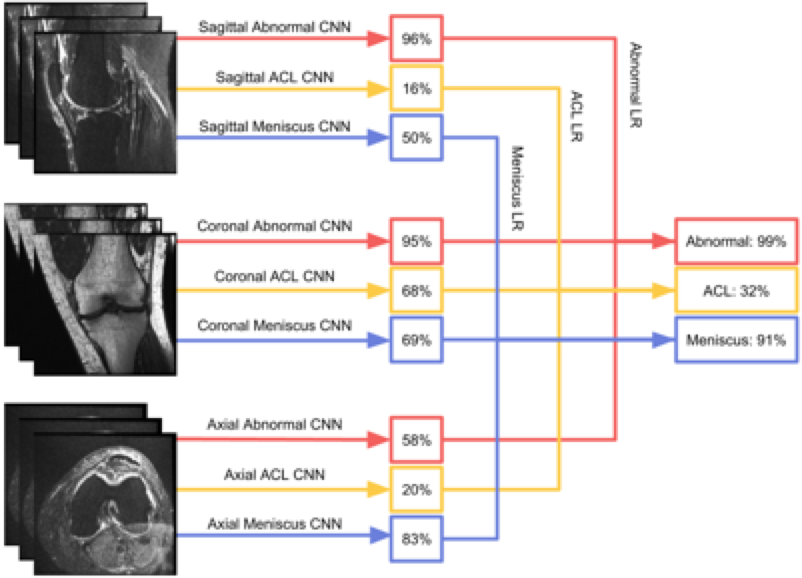
  (2) ACL tear

  (3) Meniscus tear

Architecture



Logistic Regression



Knee MRI Deep Learning Classifier

Label	AUC
Abnormal	.94
ACL Tear	.97
Meniscal Tear	.85

MURA

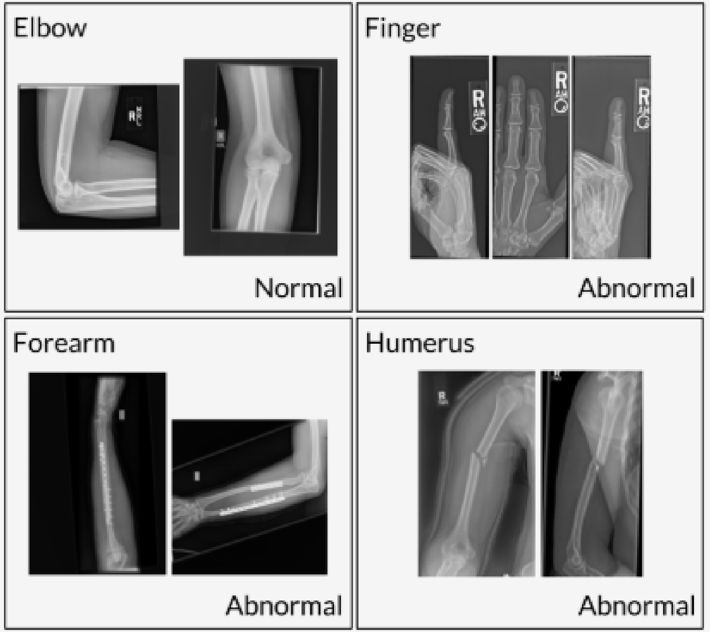
Bone X-Ray Deep Learning Competition

How did we collect MURA?

MURA is a dataset of musculoskeletal radiographs consisting of 14,863 studies from 12,173 patients, with a total of 40,561 multi-view radiographic images. Each belongs to one of seven standard upper extremity radiographic study types: elbow, finger, forearm, hand, humerus, shoulder, and wrist. Each study was manually labeled as normal or abnormal by board-certified radiologists from the Stanford Hospital at the time of clinical radiographic interpretation in the diagnostic radiology environment between 2001 and 2012.

Test Set Collection

To evaluate models and get a robust estimate of radiologist performance, we collected additional labels from six board-certified Stanford radiologists on the test set, consisting of 207 musculoskeletal studies. The radiologists individually retrospectively reviewed and labeled each study in the test set as a DICOM file as normal or abnormal in the clinical reading room environment using the PACS system. The radiologists have 8.83 years of experience on average ranging from 2 to 25 years. We randomly chose 3 of these radiologists to create a gold standard, defined as the majority vote of labels of the radiologists.



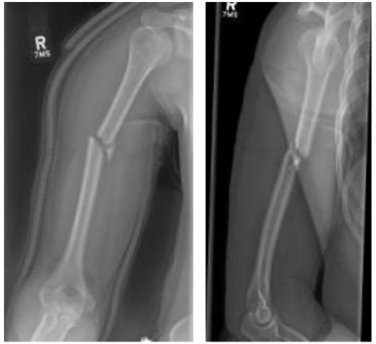
https://stanfordmlgroup.github.io/competitions/mura/

Prospective Labels

1.5M exams labeled prospectively
@ Stanford Radiology

MURA

40k prospectively labeled MSK X-rays
released in 2018 for data challenge



Abnormal

Normal

MURA Dataset: Towards Radiologist-Level Abnormality Detection in Musculoskeletal Radiographs

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

MU RA

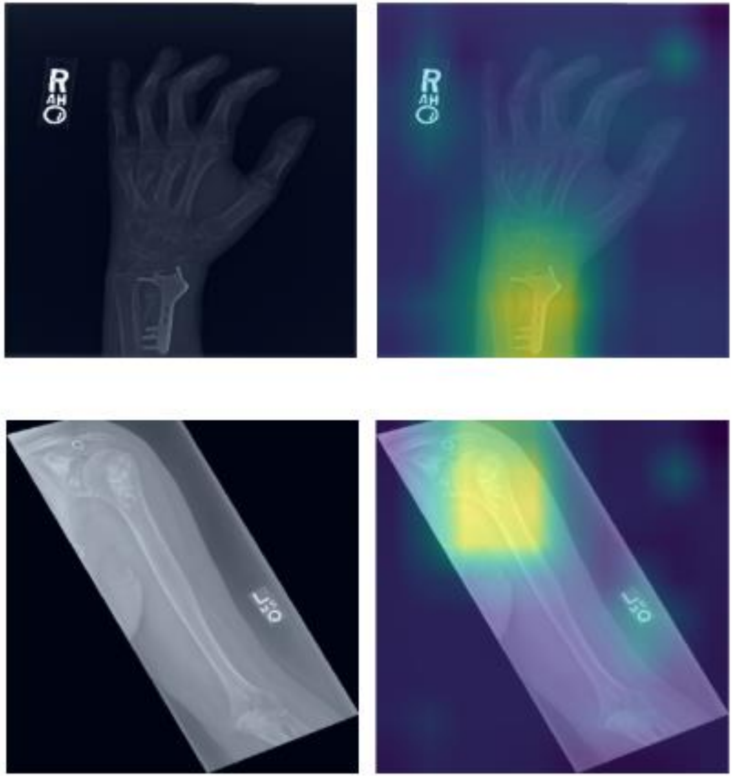
Bone X-Ray Deep Learning Competition

How does our baseline do?

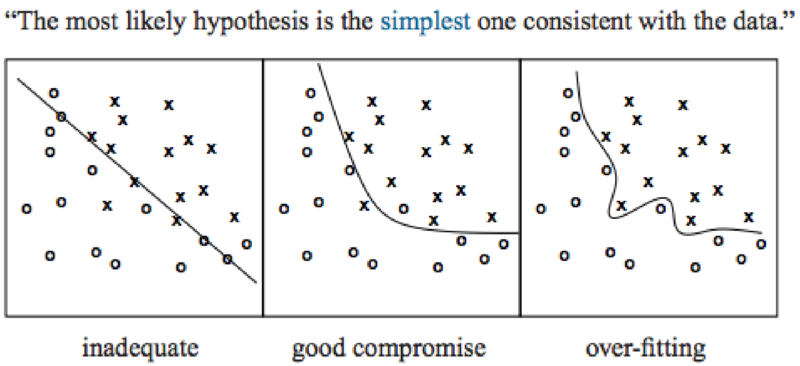
We evaluated our baseline on the Cohen's kappa statistic, which expresses the agreement of the model with the gold standard. Baseline performance is comparable to radiologist performance in detecting abnormalities on finger studies and equivalent on wrist studies. However, baseline performance is lower than best radiologist performance in detecting abnormalities on elbow, forearm, hand, humerus, shoulder studies, and overall, indicating that the task is a good challenge for future research.

	Radiologist 1	Radiologist 2	Radiologist 3	Model
Elbow	0.850 (0.830, 0.871)	0.710 (0.674, 0.745)	0.719 (0.685, 0.752)	0.710 (0.674, 0.745)
Finger	0.304 (0.249, 0.358)	0.403 (0.339, 0.467)	0.410 (0.358, 0.463)	0.389 (0.332, 0.446)
Forearm	0.796 (0.772, 0.821)	0.802 (0.779, 0.825)	0.798 (0.774, 0.822)	0.737 (0.707, 0.766)
Hand	0.661 (0.623, 0.698)	0.927 (0.917, 0.937)	0.789 (0.762, 0.815)	0.851 (0.830, 0.871)
Humerus	0.867 (0.850, 0.883)	0.733 (0.703, 0.764)	0.933 (0.925, 0.942)	0.600 (0.558, 0.642)
Shoulder	0.864 (0.847, 0.881)	0.791 (0.765, 0.816)	0.864 (0.847, 0.881)	0.729 (0.697, 0.760)
Wrist	0.791 (0.766, 0.817)	0.931 (0.922, 0.940)	0.931 (0.922, 0.940)	0.931 (0.922, 0.940)
Overall	0.731 (0.726, 0.735)	0.763 (0.759, 0.767)	0.778 (0.774, 0.782)	0.705 (0.700, 0.710)

https://stanfordmlgroup.github.io/competitions/mura/



Challenge #3: Validation

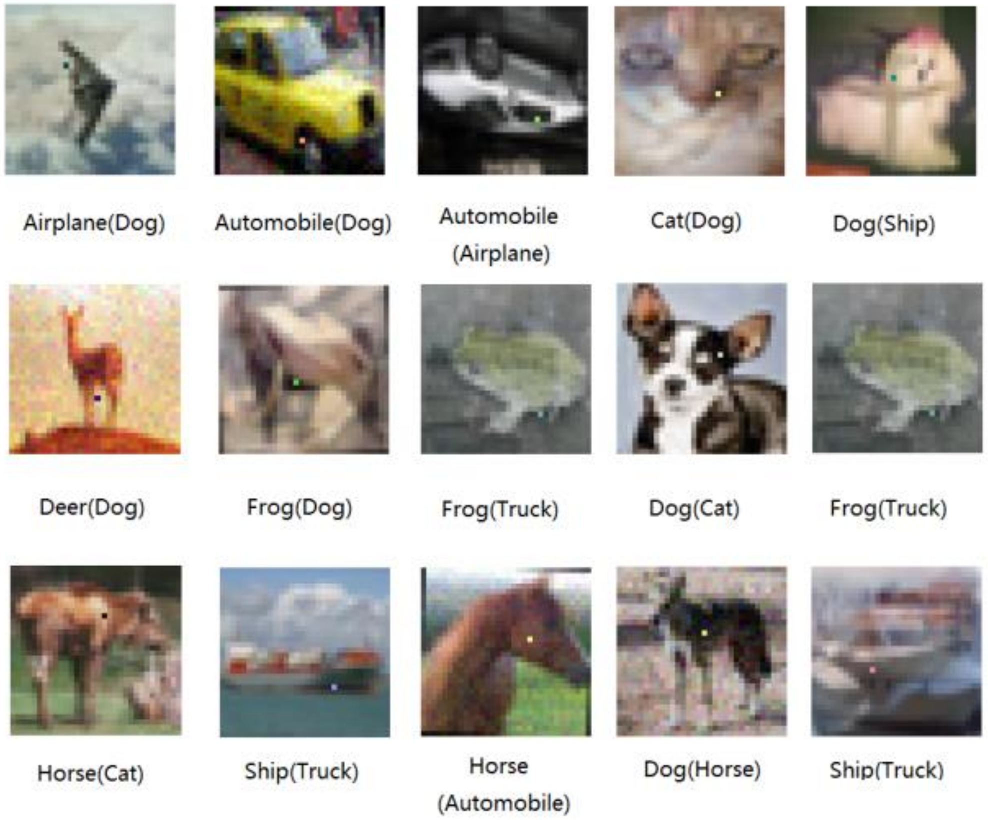


- Does the AI tool work in all scenarios?
  - Patient population
  - Imaging modalities
- Overfitting
  - The production of an analysis that corresponds too closely or exactly to a particular set of data, and may therefore fail to fit additional data or predict future observations reliably
  - Overfitting and underfitting can occur in machine learning, in particular

Machine learning security: These are not stop signs?

Eykholt et al. Robust Physical-World Attacks on Machine Learning Models.
arxiv.org/abs/1707.08945

Single Pixel Attacks



Su et al: https://arxiv.org/pdf/1710.08864.pdf

FDA News Release

FDA permits marketing of artificial intelligence algorithm for aiding providers in detecting wrist fractures

SHARE | TWEET | LINKEDIN | PIN IT | EMAIL | PRINT

For Immediate Release May 24, 2018

Summary FDA permits marketing of artificial intelligence algorithm for aiding providers in detecting wrist fractures

Low Bar for FDA Approval?

Manufacturer Imagen Technologies of New York City submitted to the FDA a study of 1000 radiographic images that evaluated the software's independent performance in detecting wrist fractures (OsteoDetect)

Study assessed how accurately the software indicated the location of fractures compared with reviews from 3 board-certified orthopedic hand surgeons

Also submitted a retrospective study in which 24 clinicians reviewed 200 patient cases

FDA

- FDA said both studies showed that sensitivity, specificity, and positive and negative predictive values in detecting wrist fractures improved when clinicians used the software
- Approved through the FDA's De Novo regulatory pathway for novel low- to moderate-risk devices

Imagen OsteoDetect is a type of computer-aided detection and diagnostic software that uses machine learning techniques to identify signs of distal radius fracture during reviews of posterior-anterior and medial-lateral x-ray images of the wrist

Software marks the location of a fracture on the image to aid clinicians with their diagnoses

Clinicians can use the software in a variety of settings, including primary care, emergency departments, urgent care centers, and for specialty care such as orthopedics

OsteoDetect is an adjunct tool

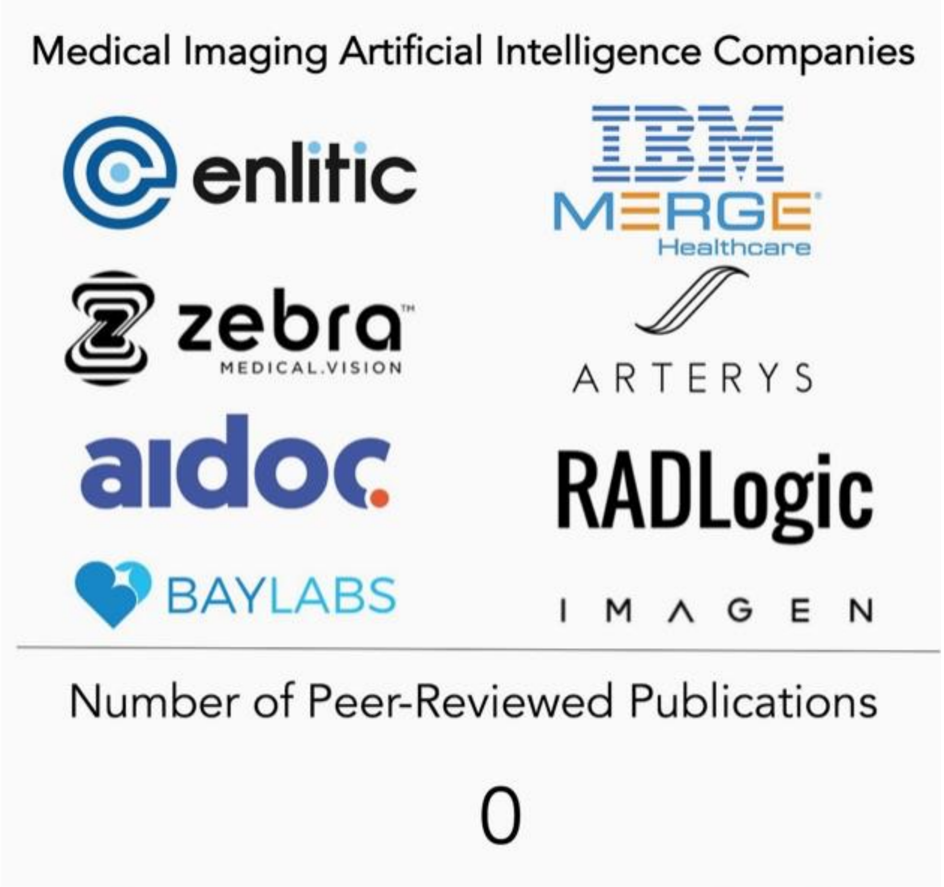
Not meant to replace clinicians' radiograph reviews or clinical judgment

Eric Topol
@EricTopol

physician-scientist, author, editor

La Jolla, CA

stsiweb.org

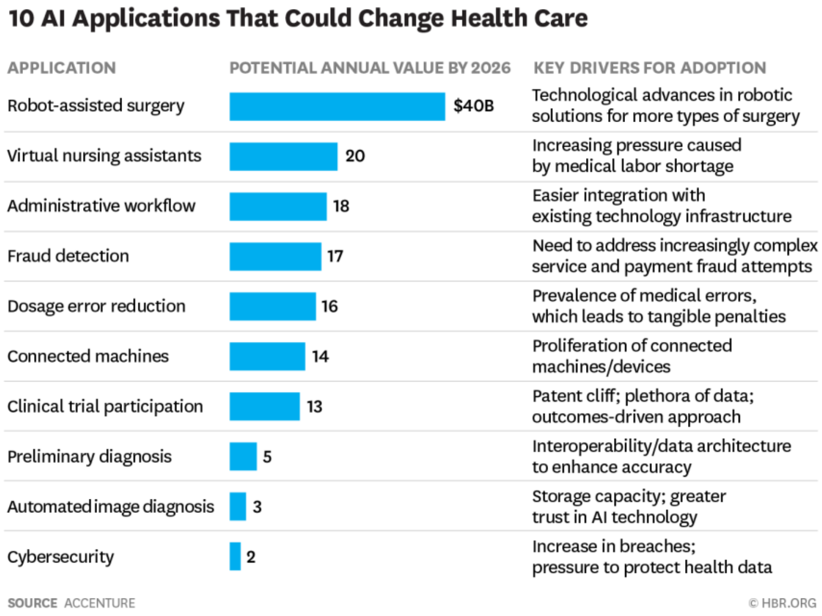


Eric Topol @EricTopol · 18 nov. 207 327

There's been a lot of talk about radiologists being replaced by machines. So I looked up the peer-reviewed publications of #AI companies
@enlitic @ZebraMedVision @baylabsinc @ArterysInc @radlogics
@IBMWatsonHealth

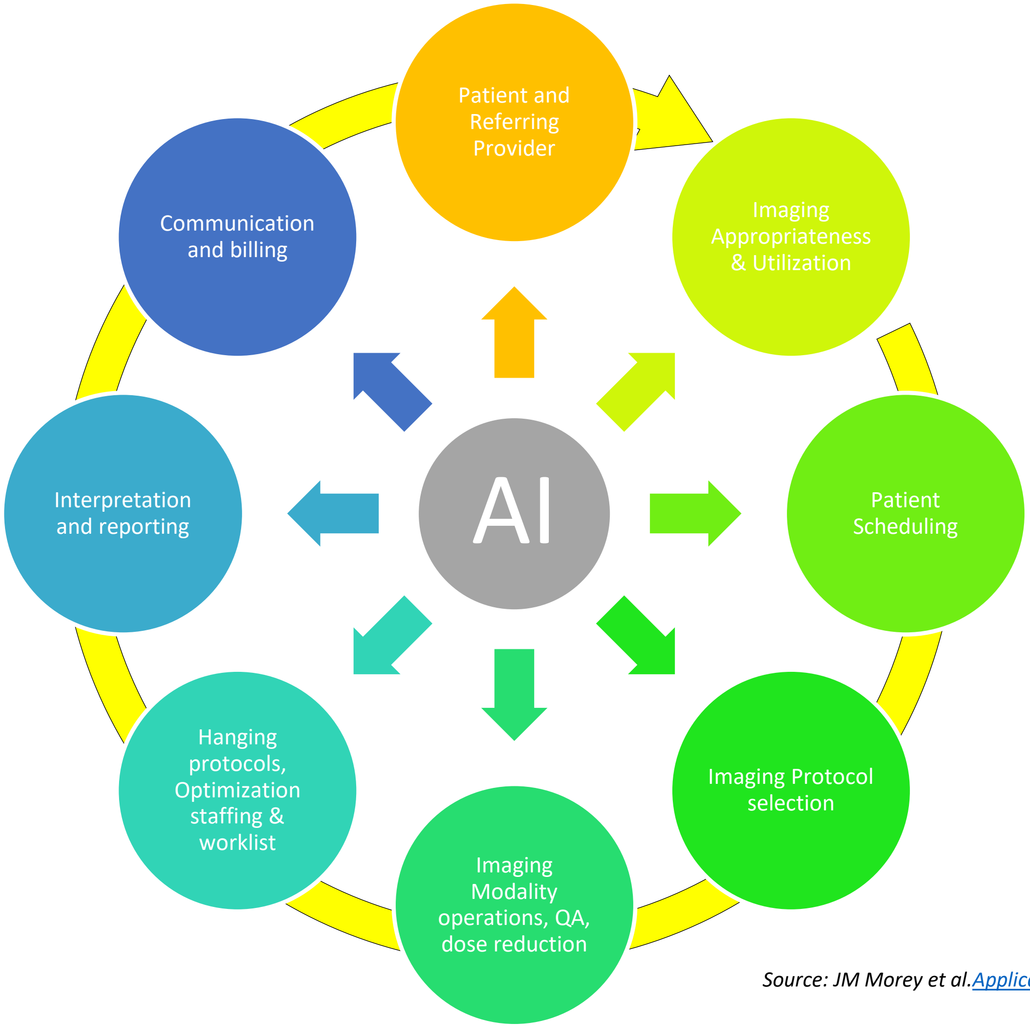
Greatest Potential of AI in HC

Making <u>back-end processes</u> more efficient



https://www.accenture.com/us-en/insight-artificial-intelligence-healthcare

AI Imaging Value Chain



Source: JM Morey et al.[*Applications of AI Beyond Image Interpretation, Springer 2018 – in press*]

AI in Radiology: Current State

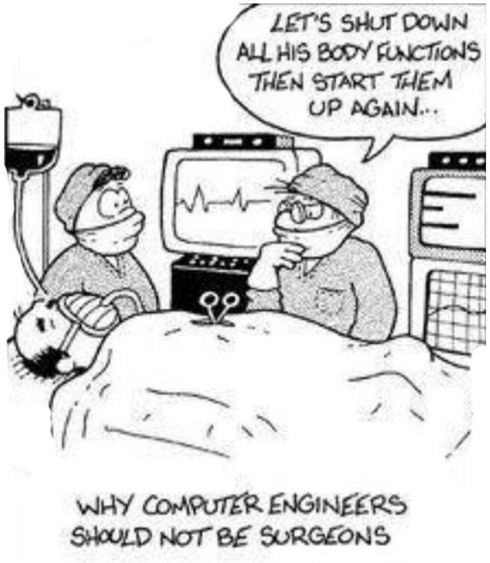
- **Individual** AI software developers are currently working with **individual** radiologists at single institutions to create AI algorithms that are focused on **targeted** interpretive needs
- Developers are using a **single** institution's prior imaging data for training and testing the algorithms, and the algorithm output is **specifically** tailored to that site's perspective of the clinical workflow
- Will models be generalizable to widespread clinical practices?
- How will model be integrated into clinical workflows across a variety of practice settings?

https://www.radiologybusiness.com/topics/artificial-intelligence/advancing-ai-algorithms-clinical-practice-how-can-radiology-lead-way

Advancing AI Algorithms for Radiology

- *"Ensuring that algorithms can be integrated into radiologists' clinical workflow is of paramount importance because if the AI tool is not readily available to the end users in their workflow, adoption in clinical practice will be less likely to occur."*
  (B. Allen, K. Dreyer)
- Interoperability between all systems is prerequisite
- Radiologists have to chose the best model for implementing AI
  - How to activate AI analysis and for what purpose
  - How to incorporate image analysis results in their reports

M. Walter, Radiology Business, May 07, 2018
B. Allen, JACR, DOI: https://doi.org/10.1016/j.jacr.2018.02.032



Implementing AI in Radiology

- Developers of AI algorithms do not always have a strong medical background or understanding of physician workflow
- Lack of well curated and diverse datasets
- "You have to have validated data sets to train [the algorithms], and so the use cases now are just being driven by data availability, not by cases that people care about. No one cares about bone age" (Paul Chang MD)

Implementing AI in Radiology: Challenges

- Heterogeneity of data
- Heterogeneity of workflow
- Determination of ground truth
- Validation of AI models at different institutions
- FDA approval of AI models for clinical use

Implementing AI: 3 Possible scenarios

1. AI on demand
2. Automated image analysis
3. Discrepancy management

P. Lakhani, NIBIB AI in Medical Imaging Workshop, Aug 23, 2018
P. Lakhani et al. JACR https://doi.org/10.1016/j.jacr.2017.09.044

Scenario 1

1. **AI on demand**
   - For a single image or series of images
   - PACS ➔radiologist ➔ AI server ➔ PACS, RIS, EHR
   - Radiologist would be in control of asking relevant AI interpretations
   - Requires manual step

Scenario 2

2. Automated AI image analysis

- Exams automatically sent to AI server (before reading)
- modality ➔ AI server ➔ PACS ➔radiologist ➔ RIS, EHR
- Helps to prioritizing reading order -> reduce TAT
- Radiologist views AI findings <u>before final report is made</u>
- Radiologist is able to ensure accuracy

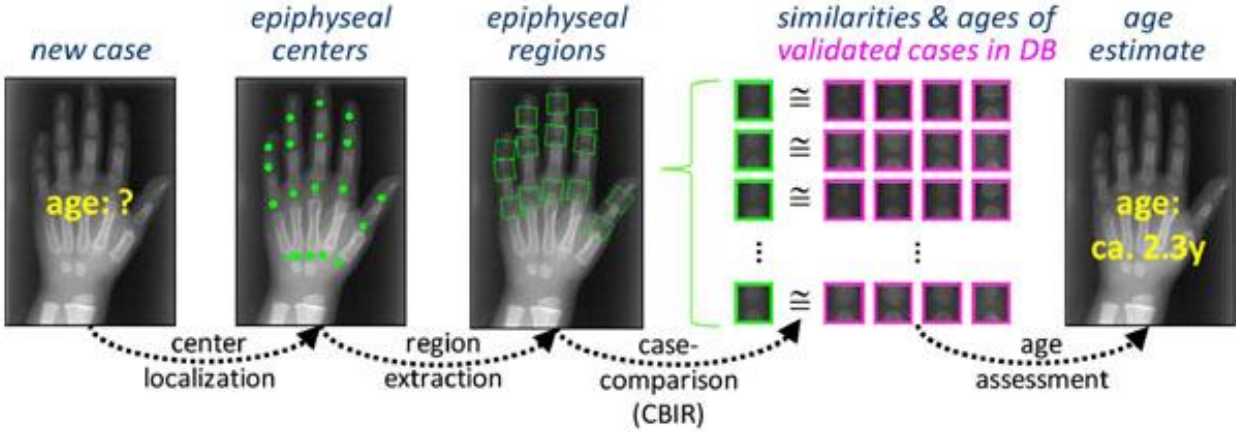
Scenario 3

3. Discrepancy management

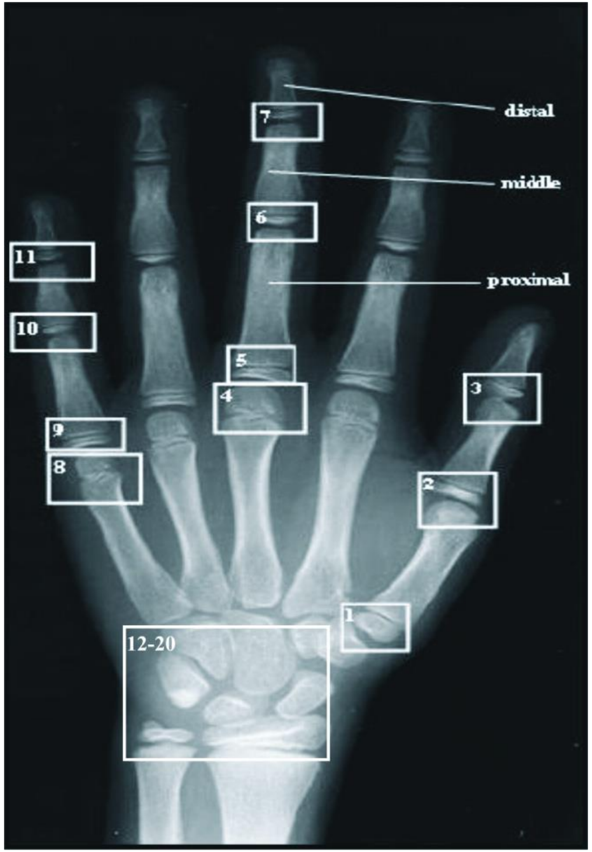
- As in 2. but results are automatically routed to RIS or EHR
- Requires discrepancy management
- AI -> preliminary -> RIS/EHR -> staff radiologist -> final
- Accurate AI needed (highly sens and spec), high confidence
- Fastest TAT although potential risk
- Might increase calls to radiology reading room
- Might have medicolegal consequences

Source: P. Lakhani, NIBIB AI in Medical Imaging Workshop, Aug 23, 2018

Bone Age The Old Way

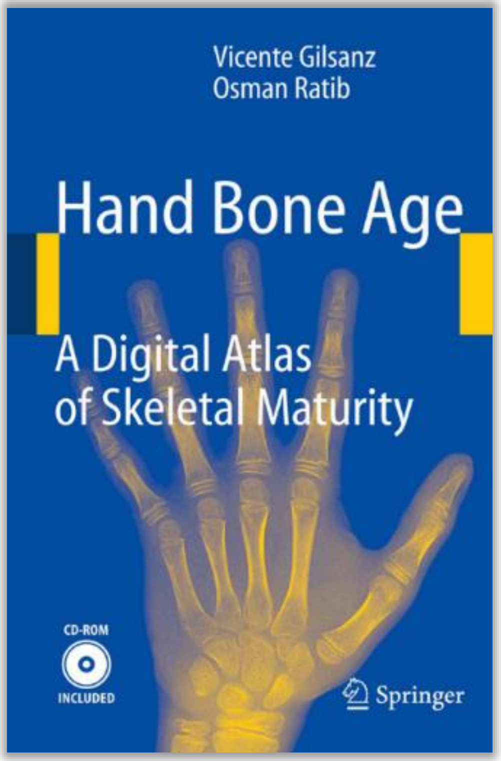


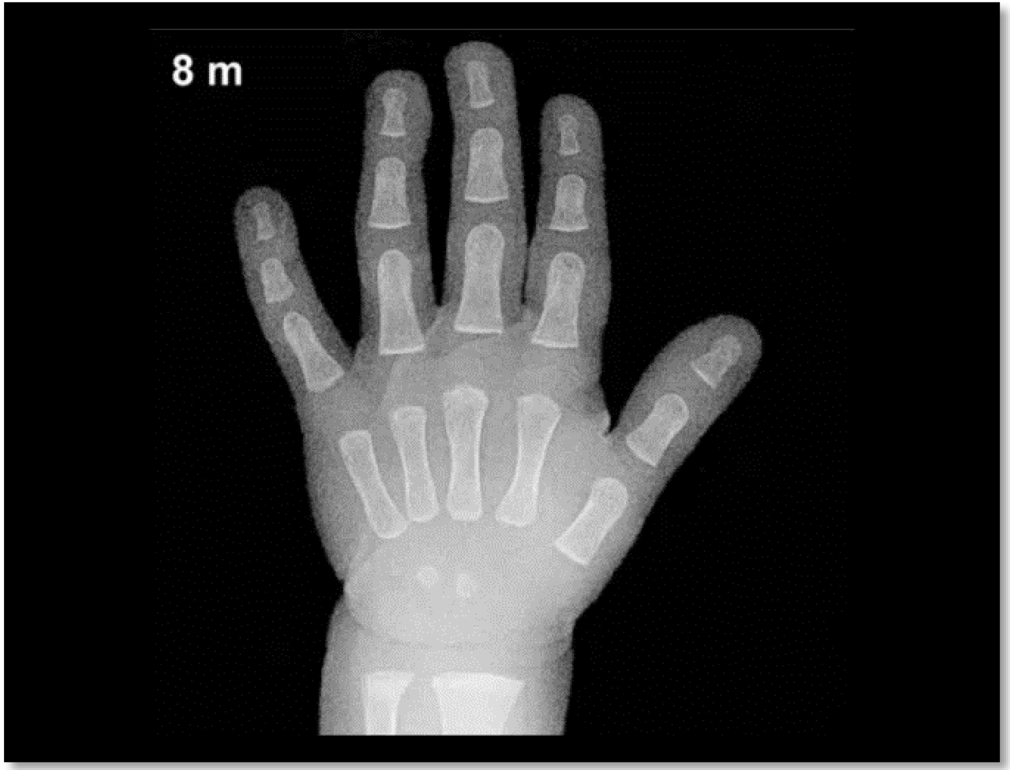
A Depeursinge et al, Open Medical Informatics Journal 11:2017



V Rai et al. Journal of Clinical and Diagnostic Research 8(9): 2014

Measuring Delayed Growth

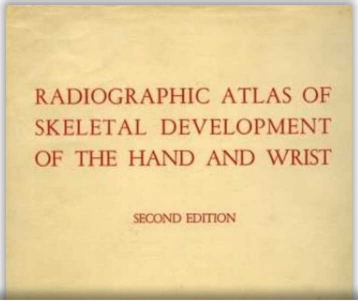




Radiology

Performance of a Deep-Learning Neural Network Model in Assessing Skeletal Maturity on Pediatric Hand Radiographs[1]

David B. Larson, MD, MBA
Matthew C. Chen, MS
Matthew P. Lungren, MD, MPH
Safwan S. Halabi, MD
Nicholas V. Stence, MD
Curtis P. Langlotz, MD, PhD



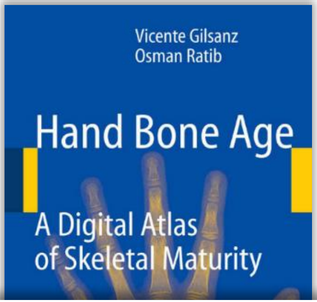


Table 2

Summary Statistics of Paired Interobserver Difference between Bone Age Estimate of Each Reviewer and Mean of the Other Three Human Reviewers' Estimates, Compared with That of Model

Variable	Clinical Report	Reviewer 1	Reviewer 2	Reviewer 3	Mean
MAD					
Reviewer	0.65	0.55	0.53	0.69	0.61
Model	0.51	0.53	0.53	0.53	0.52
P value (paired *t* test)	<.01	.50	.99	<.01	

Note.—Unless otherwise noted, data are expressed as years. The authors of the clinical report were treated collectively as a single reviewer.

https://doi.org/10.1148/radiol.2017170236

Saliency Maps

Figure 6

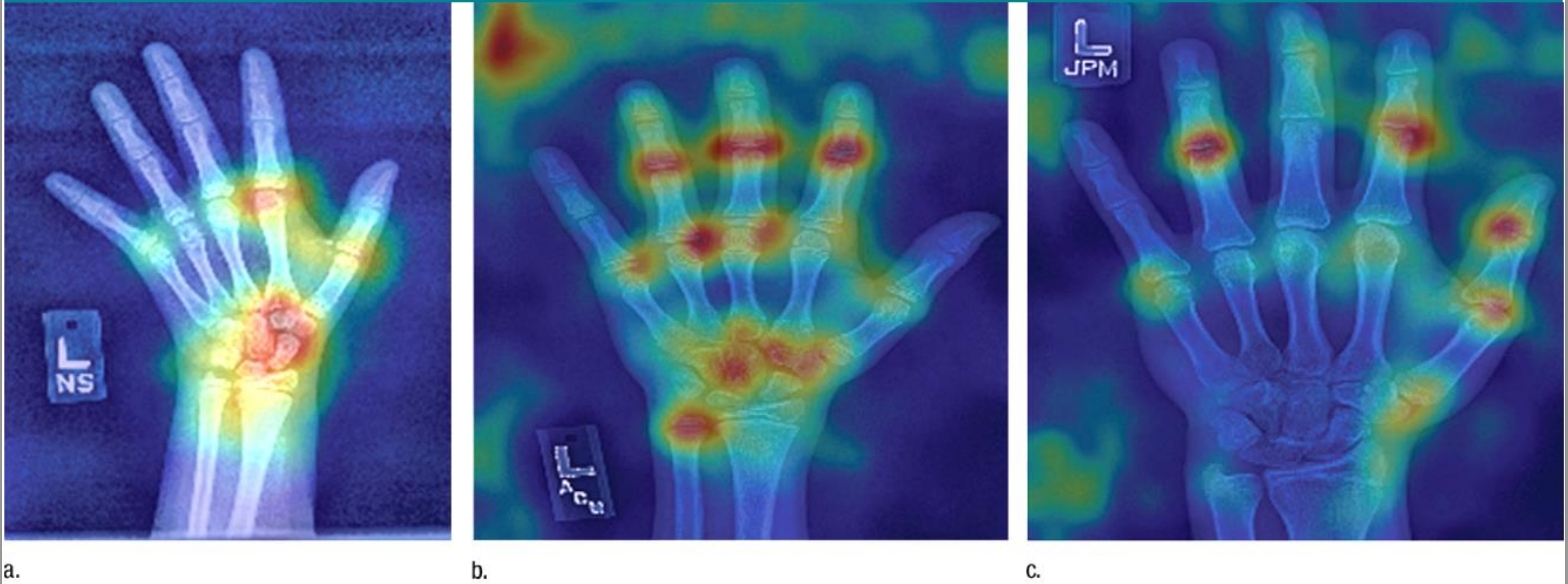


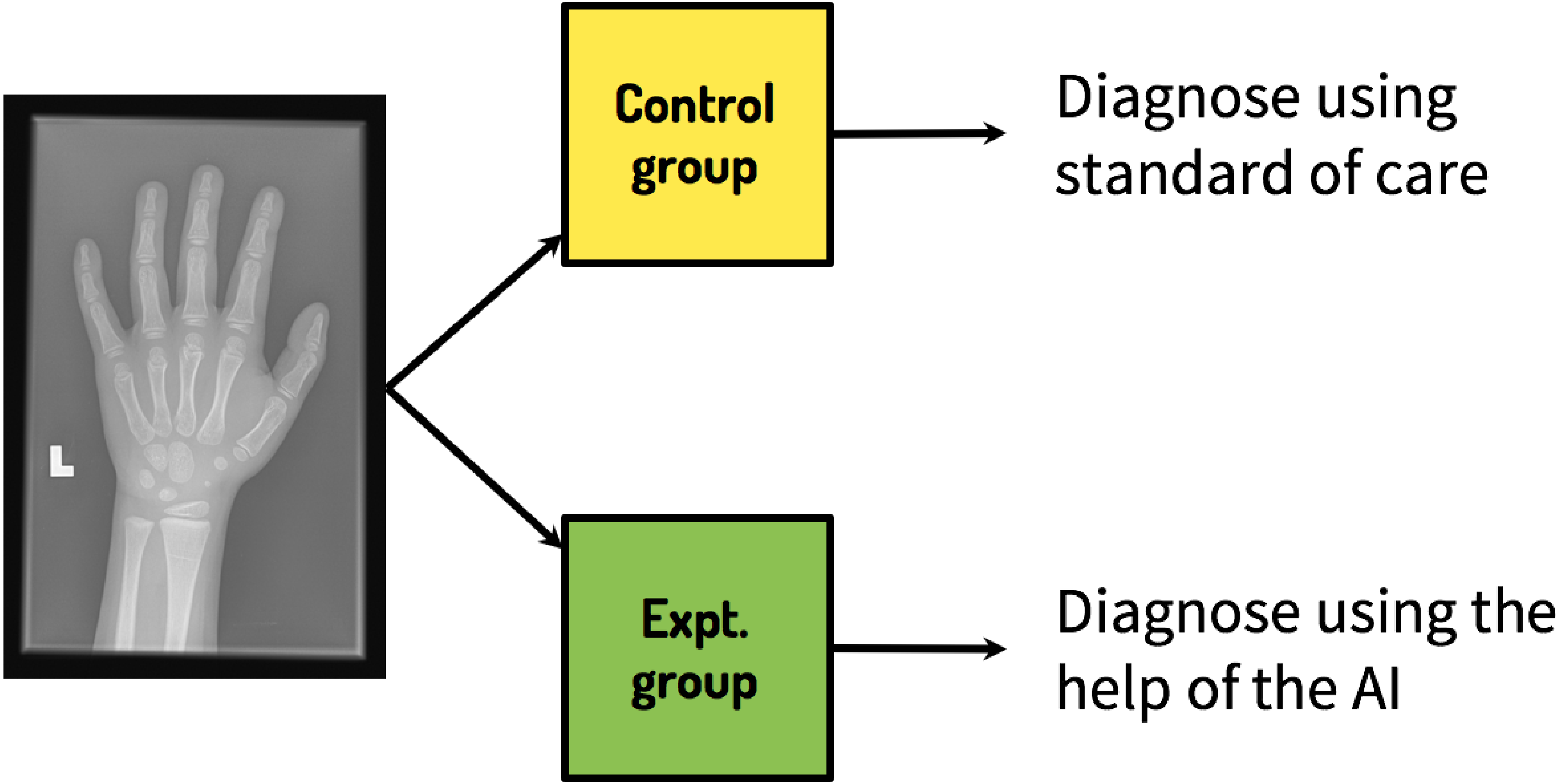
Figure 6: Original image with superimposed saliency map for sample hand radiographic images in three male patients age 4 years (**a**), 15 years (**b**), and 17 years (**c**).

Implementing BA Model Clinically

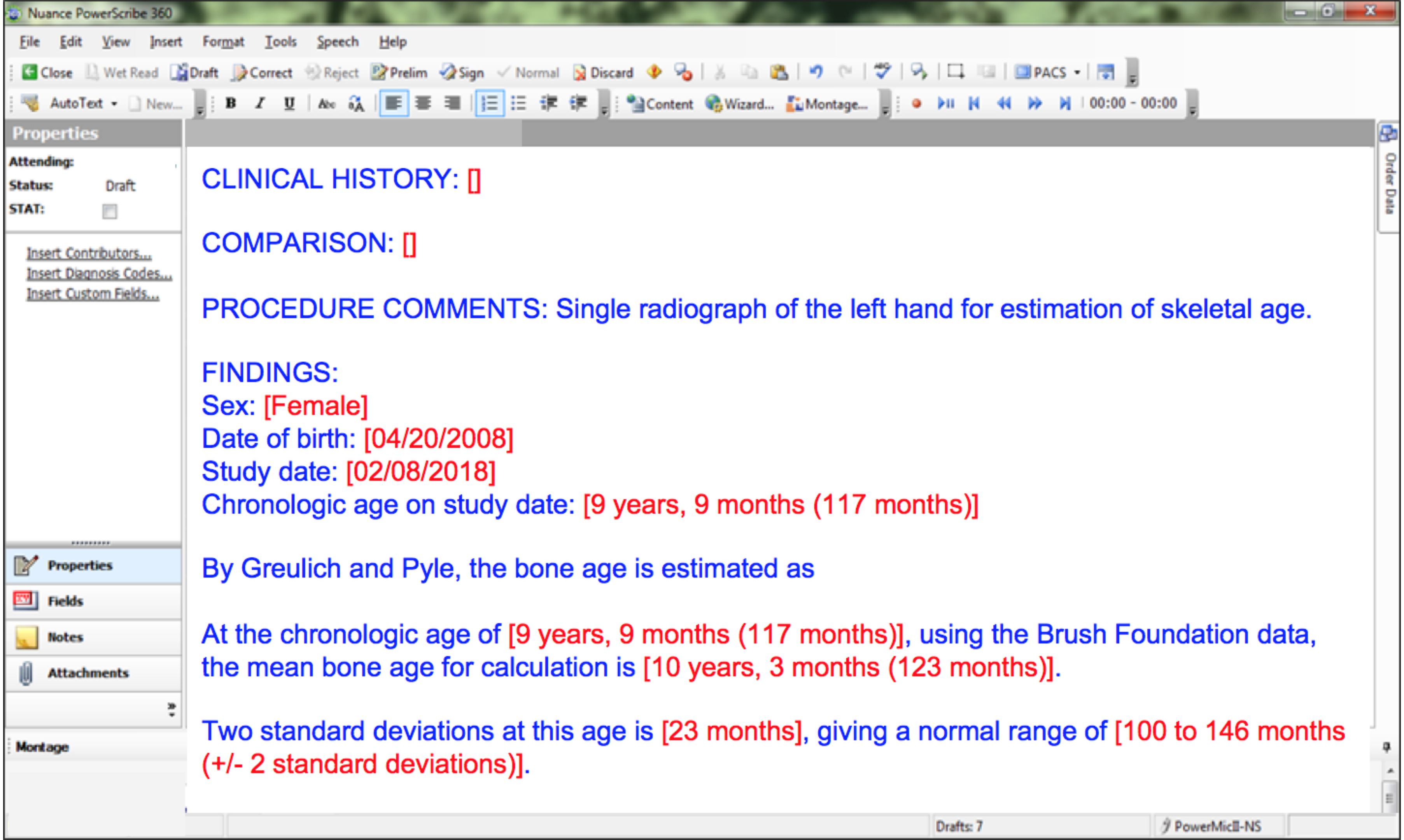
- Institutional Review Board (IRB)
- Data Use Agreement (DUA)
- Consent (Patient? Radiologist?)
- Interfaces
- Workflow
- AI Model

Validation of BA tool by Randomized Control Trial

How does exposing the prediction of the AI model to the attending radiologist prospectively affect diagnosis?



Control group
Diagnose using standard of care
Expt. group
Diagnose using the help of the AI


Nuance PowerScribe 360
File Edit View Insert Format Tools Speech Help
Close Wet Read Draft Correct Reject Prelim Sign Normal Discard PACS
AutoText New... Content Wizard... Montage... 00:00 - 00:00
Properties
Attending:
Status: Draft
STAT:
Insert Contributors...
Insert Diagnosis Codes...
Insert Custom Fields...
Properties
Fields
Notes
Attachments
Montage
Order Data
CLINICAL HISTORY: []
COMPARISON: []
PROCEDURE COMMENTS: Single radiograph of the left hand for estimation of skeletal age.
FINDINGS:
Sex: [Female]
Date of birth: [04/20/2008]
Study date: [02/08/2018]
Chronologic age on study date: [9 years, 9 months (117 months)]
By Greulich and Pyle, the bone age is estimated as
At the chronologic age of [9 years, 9 months (117 months)], using the Brush Foundation data, the mean bone age for calculation is [10 years, 3 months (123 months)].
Two standard deviations at this age is [23 months], giving a normal range of [100 to 146 months (+/- 2 standard deviations)].
Drafts: 7
PowerMicII-NS

Validation Design Scenarios

- Scenario 1: Popup window with recommendation and prediction?
- Scenario 2: Prepopulate report?
- Scenario 3: Automatically publish report?

Abbreviated Timeline of Implementing BA Model at Stanford Children's

10/16 - **Submitted DRA** for review
11/29 - Conference call with **DRA committee** (Lily from ISO, Annie from PO)
12/1 - Meeting with Dr. Halabi in OU; asked for intro to LPCH IS team
12/6 - Meeting with Marvin for **DICOM-SR**
12/8 - Follow-up meeting for DICOM-SR; Requested **firewall change**
12/22 - **DRA approved**
1/3 - **Firewall change approved**
1/9 - **IRB submitted**
1/29 - Modlink can receive my DICOM-SR messages, but cannot interpret them
2/23 - IRB approved
3/5 - Configured LPCH DICOM router to route new studies to the machine learning model
3/28 - Configured Modlink to receive DICOM-SR and tested in test environment; but we need to wait for new Nuance key (at this point, all technical integration work on our end is complete)
4/11 - Received Nuance key; required another firewall change for this key
4/26 - **Firewall change approved**
4/27 - **Change control and additional LPCH security review** for the first time
5/8 - **Security review form** submitted

Clinical Scenarios

- Quick question since you do a lot of bone age stuff. Patient JG 13y8m genetic female, transitioning to male and on hormone therapy. What is current practice in reporting in these cases? We are just going to report bone age for both genders. Thoughts?

Clinical Scenarios

- What BA reference should we use?
  - G&P
  - Snell
  - Tanner-Whitehouse
- Does BA model account for brachymetacarpia, dysplasia, malnutrition?
- Does BA model take into account demographics, clinical history, referring clinician practice?

Multi-Institutional Trial

KU MEDICAL CENTER
The University of Kansas

Boston Children's Hospital

Yale-New Haven Children's Hospital

Stanford Children's Health

240

The Children's Hospital of Philadelphia

Egleston Children's Hospital at Emory University

300

NYU Langone Health

450

MedStar Georgetown University Hospital

80

Cincinnati Children's

Key Recommendations

Goals to be accomplished for using AI in daily clinical practice

1. AI solutions should address a <u>significant clinical need</u>
2. Technology must perform <u>at least as well</u> as the existing standard approach
3. <u>Substantial clinical testing</u> must validate the new technology
4. New technology should provide <u>improvements</u> in patient outcomes, patient quality of life, practicality in use, and reduce medical costs
5. <u>COORDINATED APPROACH</u> between multiple stakeholders is needed

Coordinated Approach

- End users must first define the purpose (clinical use case)
- Developers must translate users' needs to program code
- Managers must coordinate resources and strategies to bring SW in workflow
- Companies must mass distribute the SW product and integrate it with existing infrastructure
- Policy experts and legal teams must ensure there are no legal/ethical barriers

Who are the Stakeholders?

HC Community

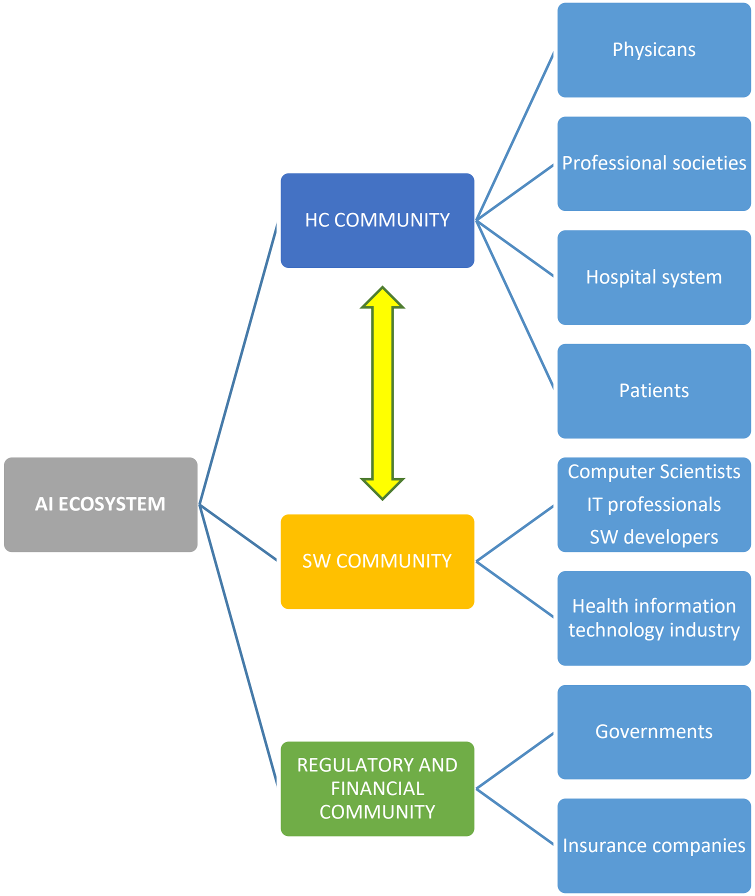
- Radiologists and residents/trainees
- Referring physicians and patients
- Medical professional societies
- Hospital systems, IT departments
- Academics and medical scientists

SW Community

- IT professionals, SW developers
- Health information technology (HIT) industry
- Academic IT professionals: engineers, computer scientists

Other Stakeholders

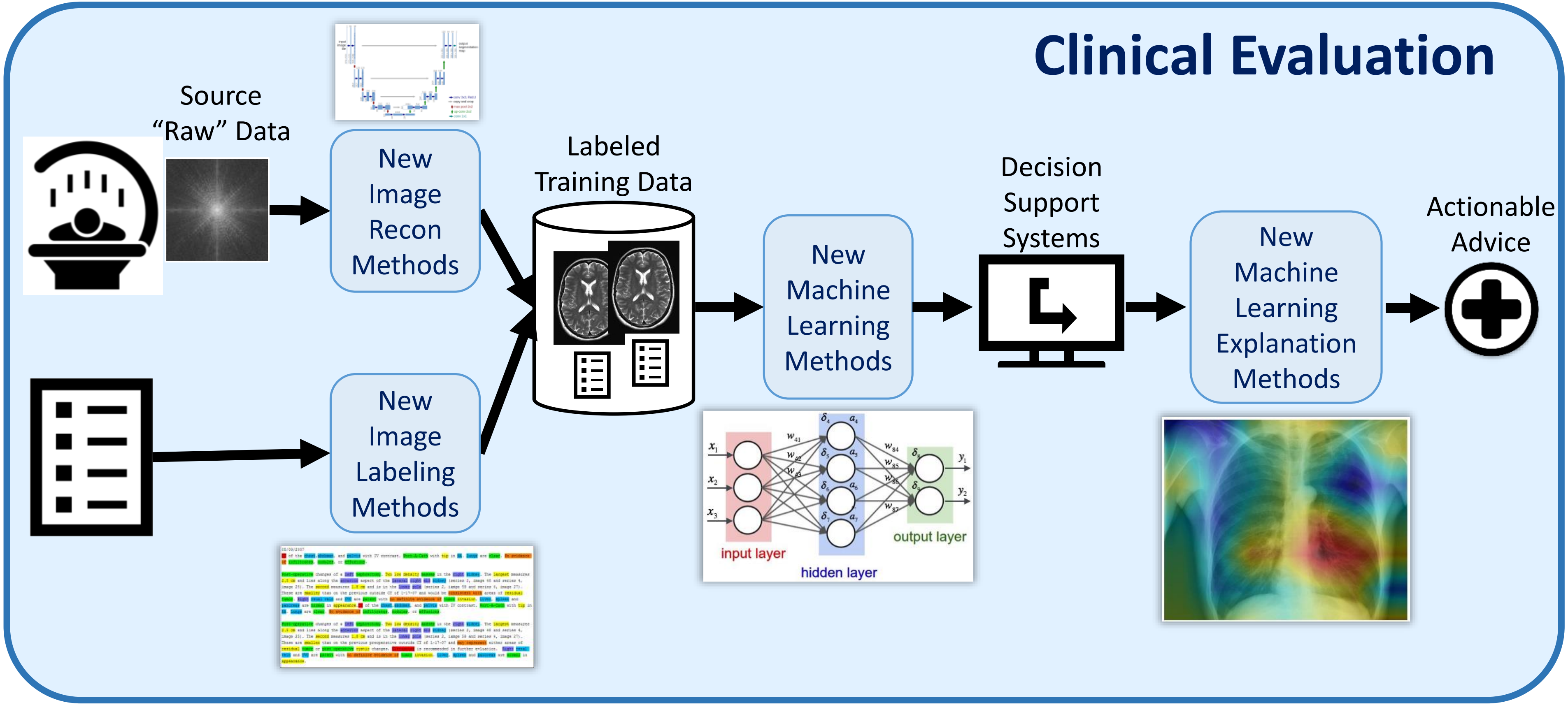
- Governments and insurance companies
  - Financing, reimbursement
  - Different payment models (public, hybrid)
  - Variable strategies for fostering AI software in general and for HC
- Regulatory agencies (FDA, CE)
- Patients


AI ECOSYSTEM
HC COMMUNITY
SW COMMUNITY
REGULATORY AND FINANCIAL COMMUNITY
Physicans
Professional societies
Hospital system
Patients
Computer Scientists
IT professionals
SW developers
Health information technology industry
Governments
Insurance companies

$ Financial Considerations

- Difficult to define a business plan for a narrow AI product that may solve one clinical question on one modality
- May be a pricing disparity between what customers will pay and the costs involved
- Who will pay? Insurance, patient, health system, radiology group, vendor?
- Who is in charge of AI model implementation? Vendor, hospital IS?
- What happens when the model fails or is not fully validated?

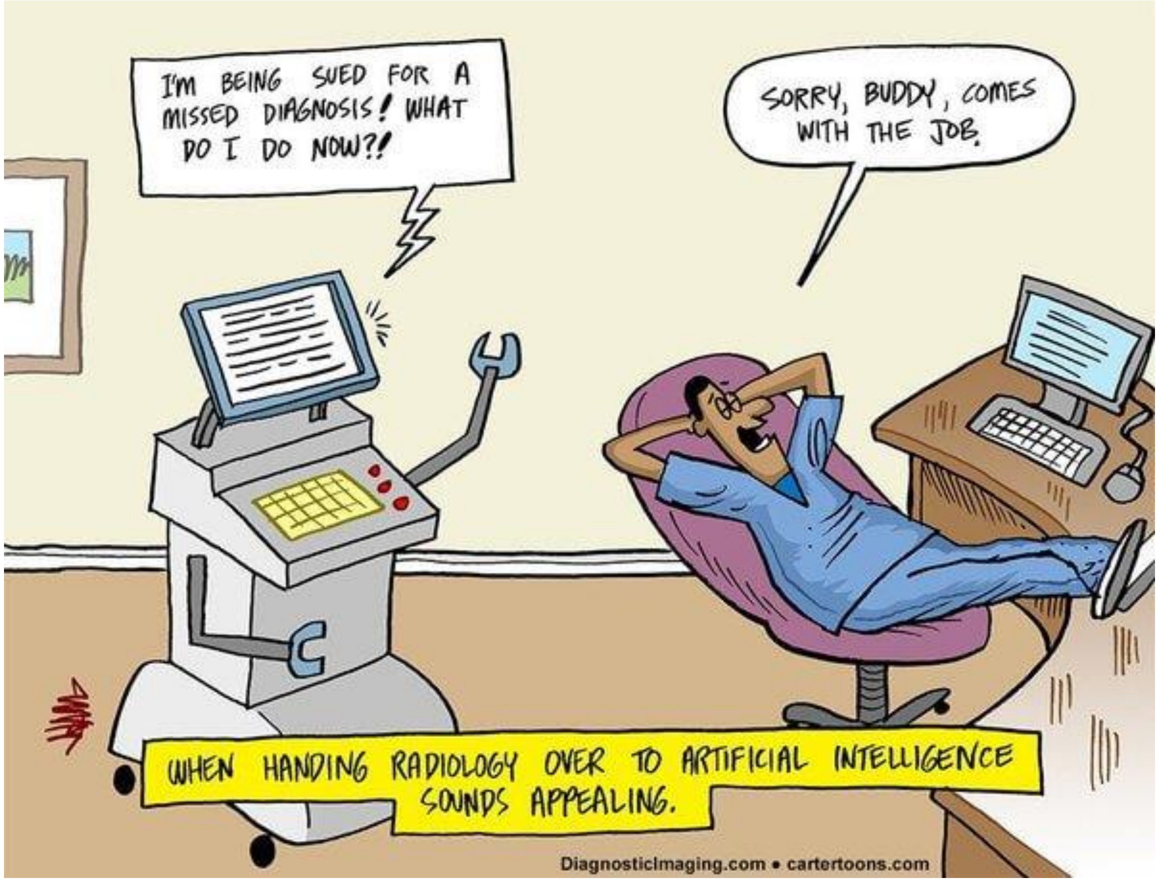
Technical Considerations



CT scan icon by Sergey Demushkin from the Noun Project

AI and the Radiologist

- How does the AI algorithm influence the performance of the radiologist?
- Does Radiologist + AI outperform just the Radiologist?
- What is considered the “ground truth”?
- How will the AI model be displayed?
- Will the AI model learn over time?


I'M BEING SUED FOR A MISSED DIAGNOSIS! WHAT DO I DO NOW?!
SORRY, BUDDY, COMES WITH THE JOB.
WHEN HANDING RADIOLOGY OVER TO ARTIFICIAL INTELLIGENCE SOUNDS APPEALING.
DiagnosticImaging.com • cartertoons.com

Building Radiology AI:
The Role of Professional Organizations

- Educate clinical users of AI algorithms
- Develop a robust technical workforce
- Convene collaborations: radiologists, scientists, industry
- Support development of AI use cases
- Assemble publicly-available training data sets
- Advocate for and provide research funding for AI
- Establish standards for AI data and algorithms
- Encourage balanced regulation of AI technology

RSNA ACR SIIM ESR

SPR The Society for Pediatric Radiology

Take Home Messages

- AI is a powerful tool with many applications that can help radiology practices today *beyond image interpretation*
- Integrating AI models holds promise for improving radiology practices and patient care
- More research needs to be done regarding the evaluation of AI in a clinical setting, including its impact on workflow and value of services
- No matter how AI is implemented in the workflow, the radiologists will have an important role in ensuring accuracy, safety and quality of the algorithms

Curt Langlotz
@curtlanglotz

Follow

Will #AI replace radiologists? The answer is NO. But rads who use #AI will replace rads who don't @RSNAInformatics @SIIM_Tweets

12:55 AM - 8 Feb 2017

86 Retweets **108** Likes

6 86 108

AIMI
CENTER FOR ARTIFICIAL INTELLIGENCE
IN MEDICAL IMAGING

AIMI.STANFORD.EDU
@STANFORDAIMI

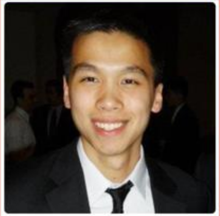
Members

Bao Do
Radiologist

Chris Beaulieu
Radiologist

Curt Langlotz
Professor of Radiology and Biomedical Informatics, Lab Director

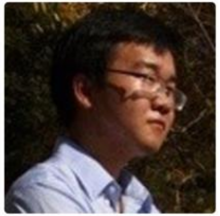
David Larson
Radiologist



Matt Chen
Graduate Student

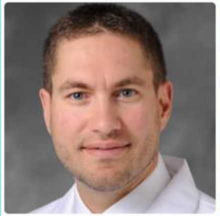
Mike Muelly
Clinical Instructor, Radiology

Matt Lungren
Assistant Professor of Radiology, Lab Faculty



Rui Shu
Graduate Student

Rusty Hoffman
Radiologist



Safwan Halabi
Radiologist

Vanessa Sochat
Research Software Engineer for Research Computing and Stanford Medicine

Nicholas Stence
Radiologist

boneage.stanford.edu

David Eng • 1st

Student at Stanford University

Stanford, California

Nishith Khandwala • 1st

Machine Learning Researcher at Stanford University

Stanford, California

Questions?

Safwan.Halabi@Stanford.edu
@SafwanHalabi