Machine Learning for Medical Imaging: The Imaging Lifecycle

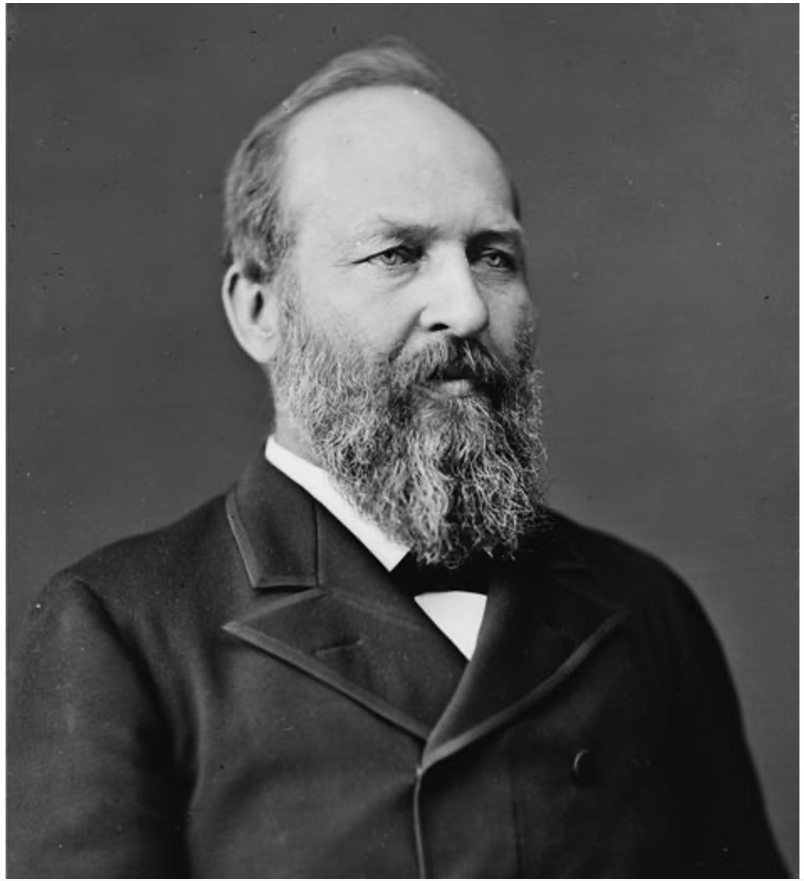
Brian E Chapman, PhD

September 2022

Machine Learning and Domain Expertise (Jarrahi, Memariani, and Guha 2023)

Imaging and the transformation of medicine

July 1881 US president Games Garfiedl shot. Died 80 days later]



March 1981 U.S. president Ronald Reagan shot. Died 23 years later.

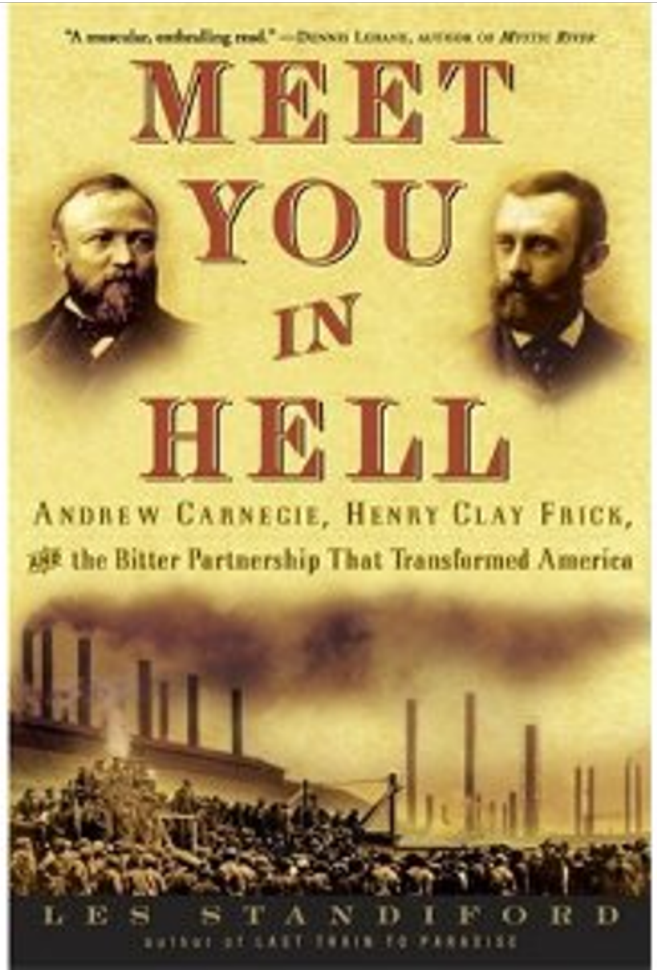
What happened in between to make the difference?

- Poll Everywhere


Hand mit Ringen
PHYSIK. INSTITUT
DER UNIVERSITÄT
WÜRZBURG.

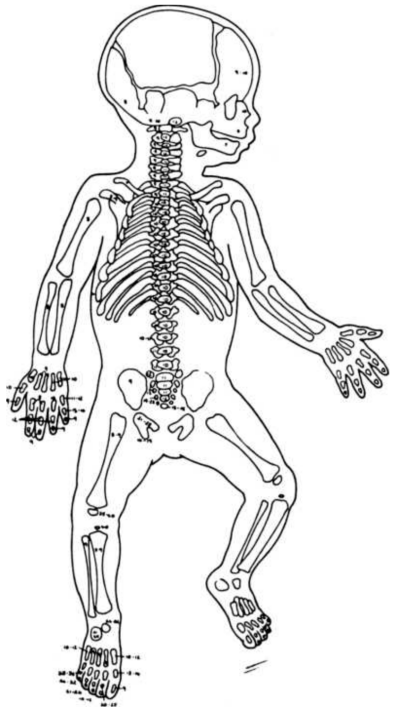
The price of success

"Monumental achievements come with monumental costs." (*Meet You In Hell*)



Imaging gone wild

"by 1930 many obstetricians suggested routine prenatal pelvic X-ray examinations. The development of the rabbit test in 1931 to verify pregnancy diminished the use of X-rays from the protocol of early confirmation. Obstetricians continued to use them, however, to estimate mate due dates and check on fetal condition. (Kevles 1997)



46. Diagram of a fetus (1917). (J. Hess, "The Diagnosis of the Age of the Fetus by Use of Roentgenograms," *American Journal of Diseases of Children* 14 [1917]: 397–423.) Courtesy of the American Medical Association.

Is this a good idea?

Ordering (or not) a medical imaging procedure

- "What is the action I'm going to take based on the results of this image?"
  - Obstrical ultrasound
- What costs and risks are being incurred by performing this imaging examination?

Screening for Breast Cancer

- I'm a woman with no family history of breast cancer and no symptoms.

Screening for Breast Cancer

- I'm a woman with no family history of breast cancer and no symptoms.
- Should I get a screening mammogram at

Screening for Breast Cancer

- I'm a woman with no family history of breast cancer and no symptoms.
- Should I get a screening mammogram at
  - 40?

Screening for Breast Cancer

- I'm a woman with no family history of breast cancer and no symptoms.
- Should I get a screening mammogram at
  - 40?
  - 60?

Screening for Breast Cancer

- I'm a woman with no family history of breast cancer and no symptoms.
- Should I get a screening mammogram at
  - 40?
  - 60?
  - 80?

Screening for Breast Cancer

- I'm a woman with no family history of breast cancer and no symptoms.
- Should I get a screening mammogram at
  - 40?
  - 60?
  - 80?
- What affects this decision?

How does screening mammography work? [(Dyro and Dyro 2004)]

How does breast biopsy work? (Bick et al. 2020)

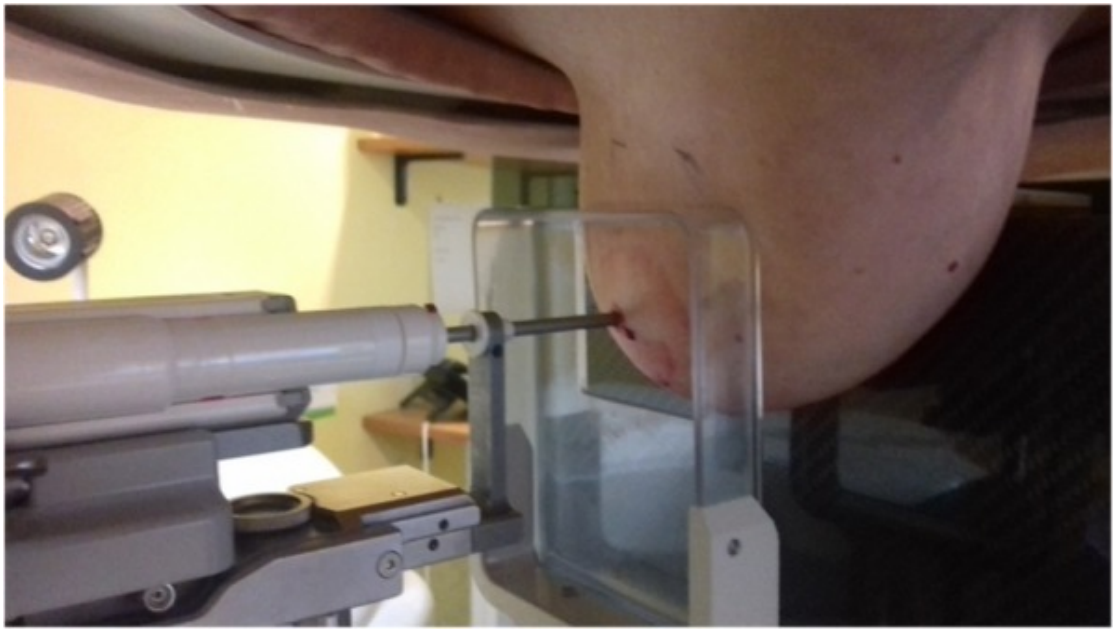
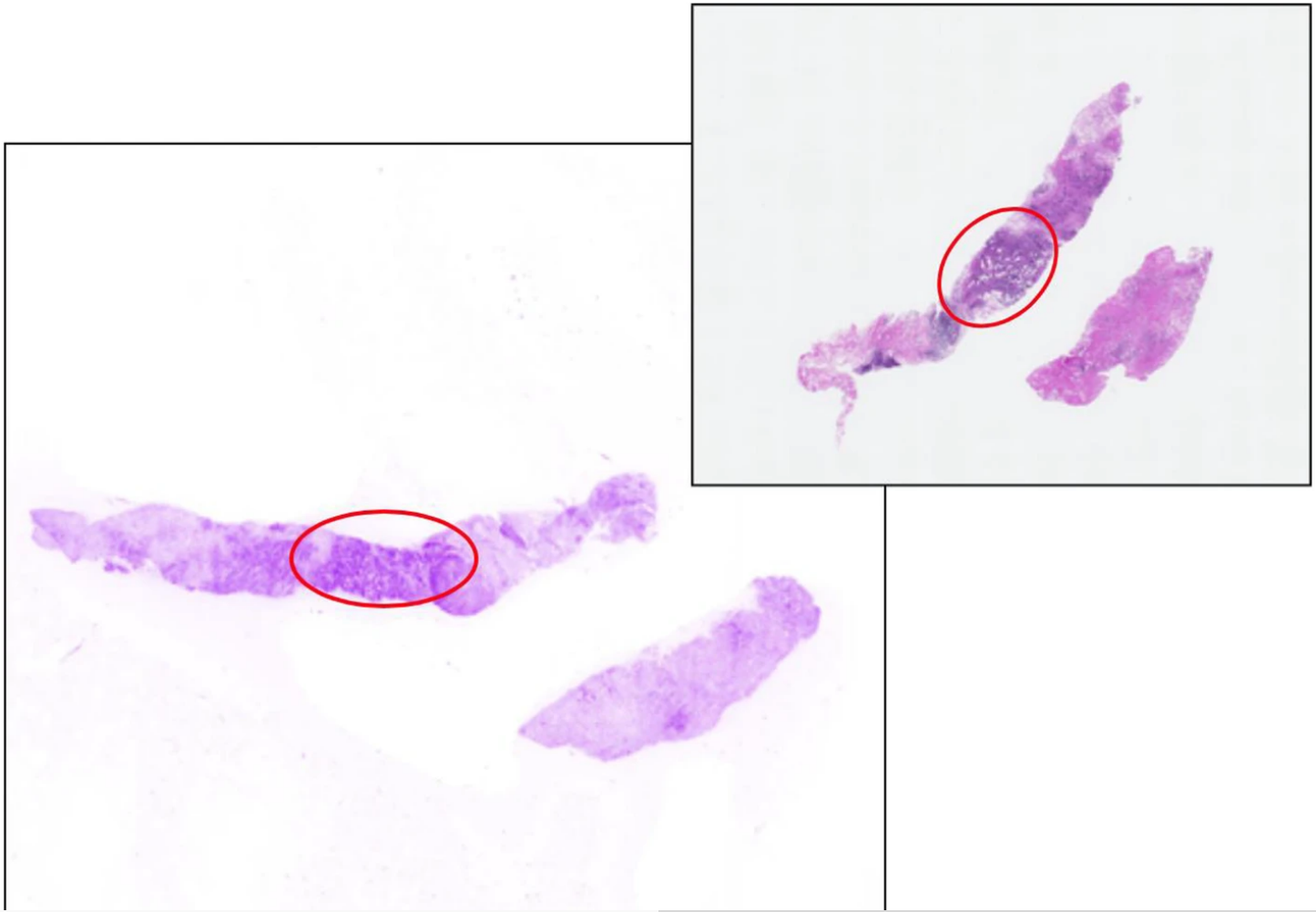


Figure 6: breast biopsy

Pathology interpretation (Elfgen et al. 2019)



Mammography Question #1

- For a woman aged 60 years, what is the positive predictive value of a positive screening mammogram for actually being positive for cancer?
- PollEv.com/brianchapman270

Mammography Question #1 Answer

- 9.8%

Mammography Question #2

- For a woman aged 40 years, what is the positive predictive value of a positive screening mammogram for actually being positive for cancer?
- PollEv.com/brianchapman270

Mammography Question #2 Answer

- 1.3%

Mammography Question #3

- I'm a 30-year-old woman with large, dense breasts and a family history of breast cancer who has tested positive for a BRCA1 mutation.

Mammography Question #3

- I'm a 30-year-old woman with large, dense breasts and a family history of breast cancer who has tested positive for a BRCA1 mutation.
- Should I get a mammogram?

Mammography Question #3

- I'm a 30-year-old woman with large, dense breasts and a family history of breast cancer who has tested positive for a BRCA1 mutation.
- Should I get a mammogram?
- Why or why not?

Mammography Question #3

- I'm a 30-year-old woman with large, dense breasts and a family history of breast cancer who has tested positive for a BRCA1 mutation.
- Should I get a mammogram?
- Why or why not?
- Should I get an MRI or US instead?

Mammography Question #3

- I'm a 30-year-old woman with large, dense breasts and a family history of breast cancer who has tested positive for a BRCA1 mutation.
- Should I get a mammogram?
- Why or why not?
- Should I get an MRI or US instead?
- How can we go from population statistics to personalized decisions?

What are possible roles for ML?

- Helping the GP?
- Helping the radiologist?
- Helping the consumer?
- Helping the pathologist?

Consumer decision-support

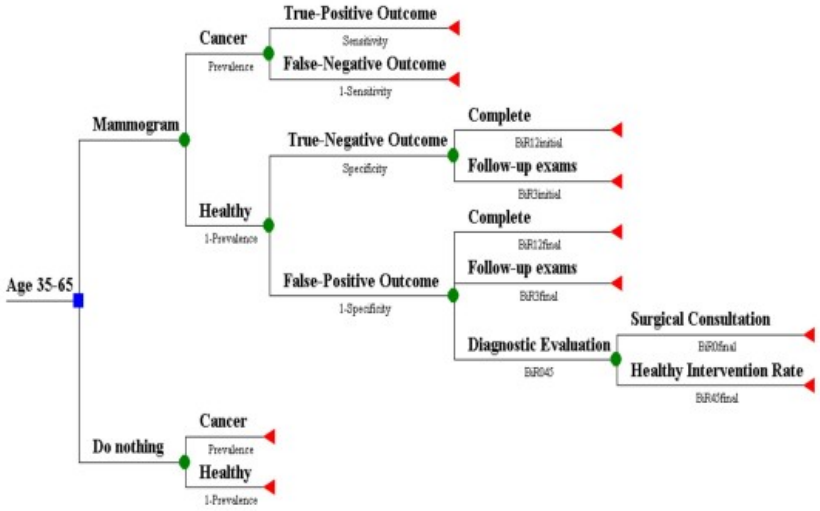


Figure 8: Decision model for baseline screening mammography used to predict primary and secondary performance measures.

Diagnostic workup

Figure 9: First x-ray

Figure 10: my hand

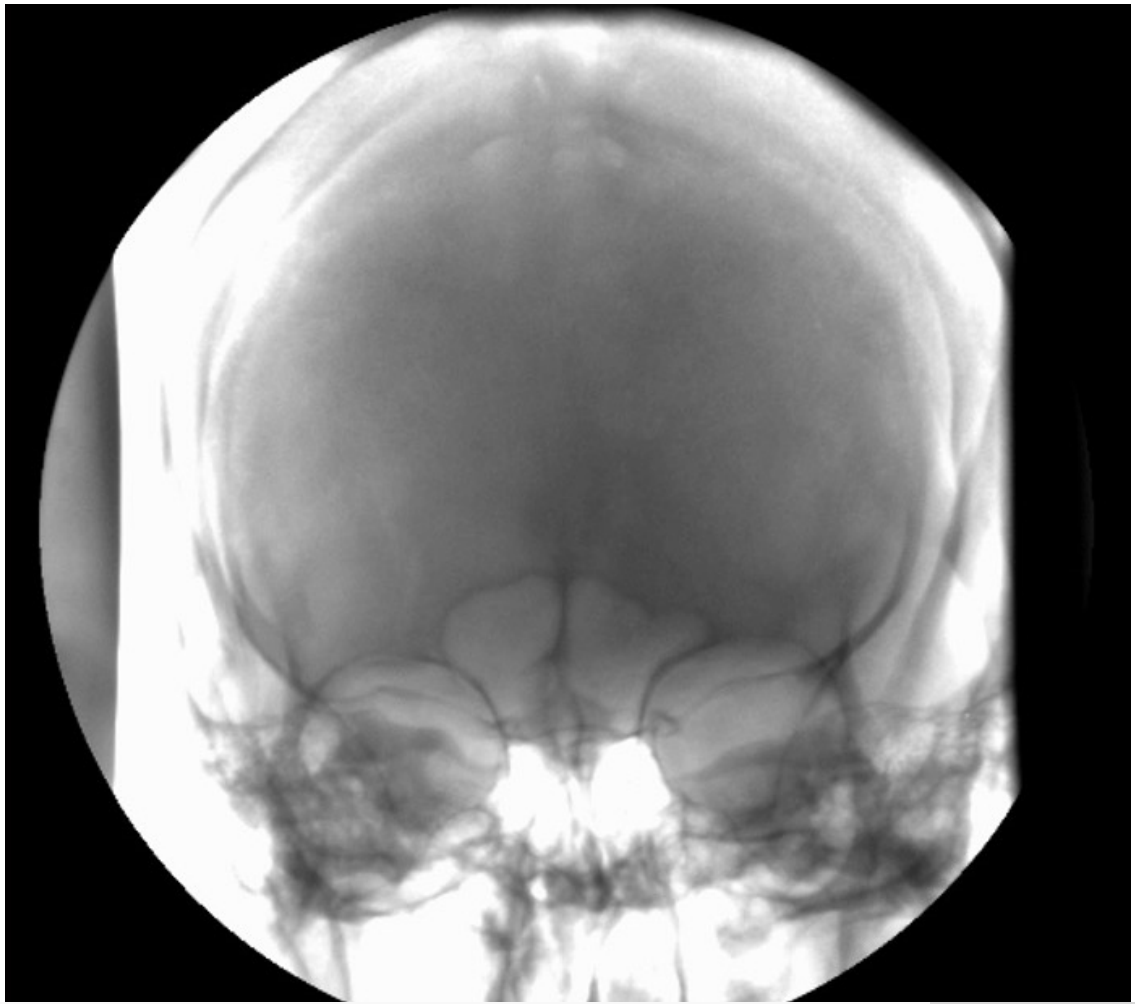
Diagnostic workup

Figure 11: Something is wrong?

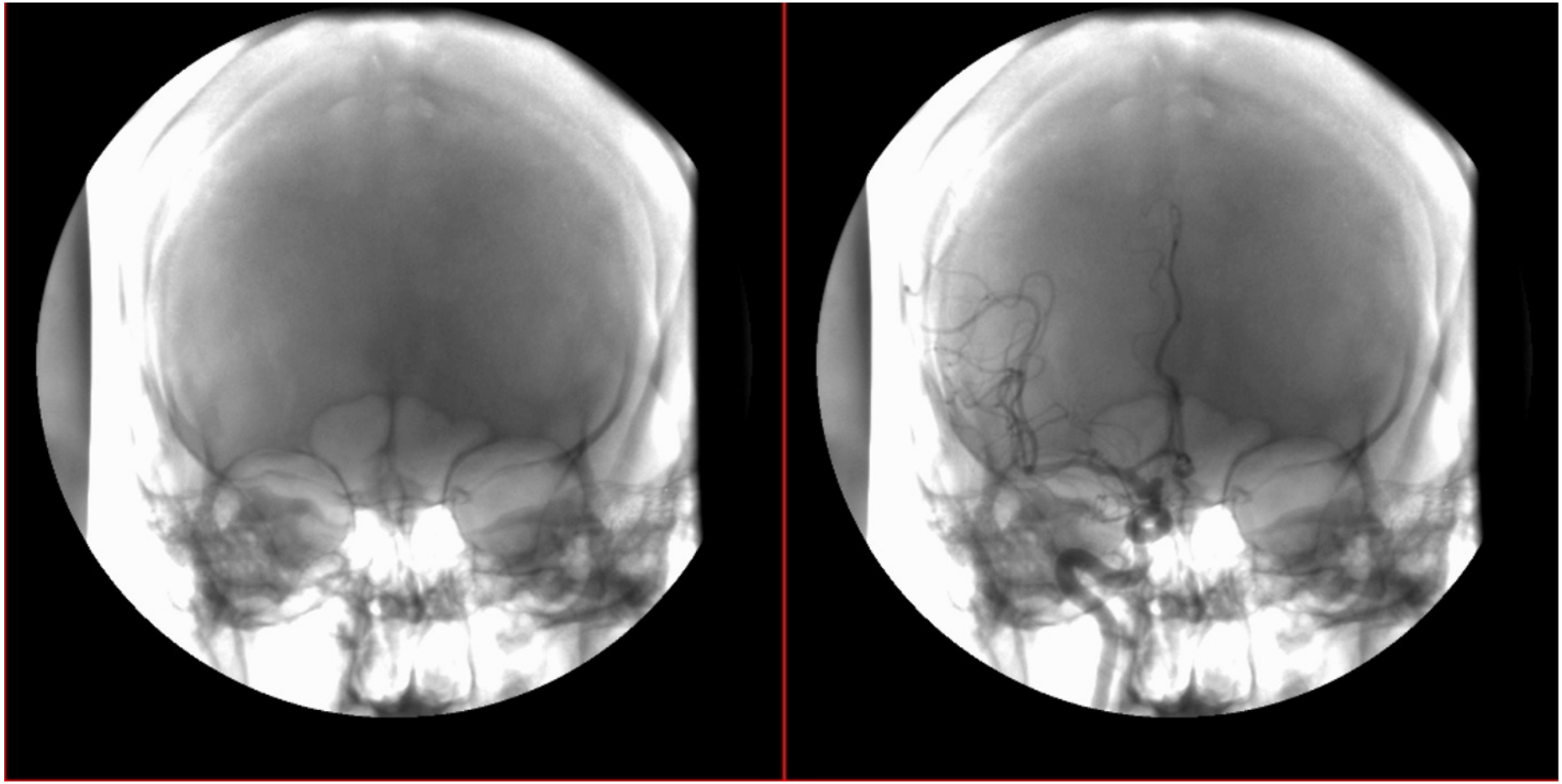
How do I pick my energy probe?

I want to visualize blood vessels?

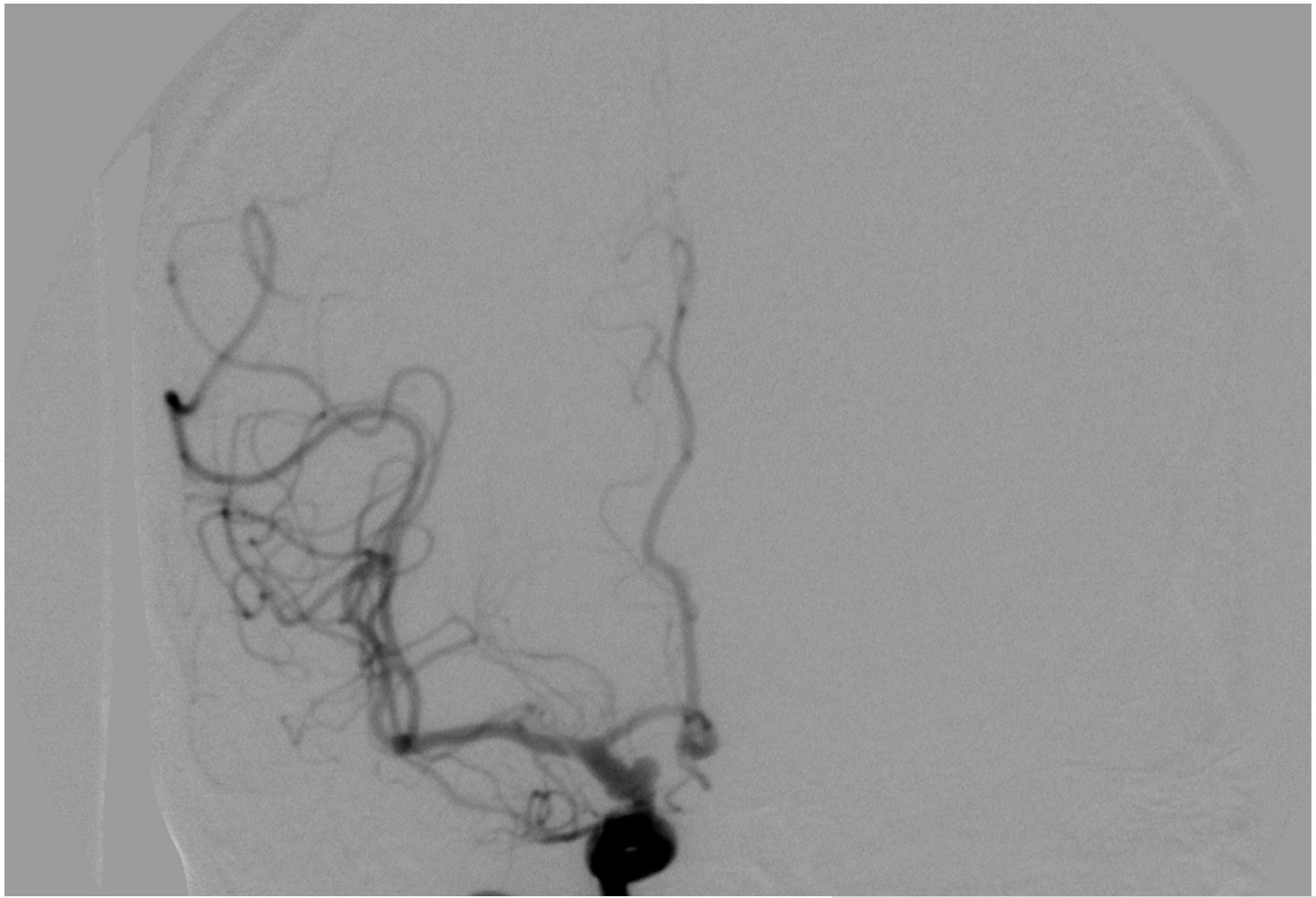
X-ray visualization of blood vessels?



X-ray visualization of blood vessels?



X-ray visualization of blood vessels?



PET Imaging

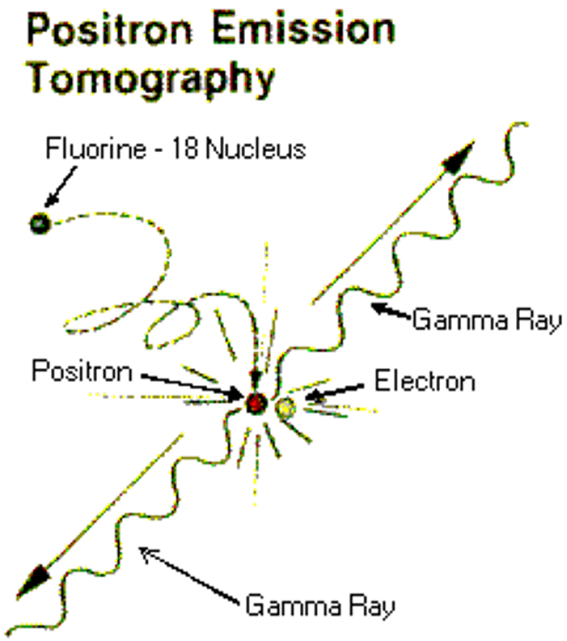


Figure 13: Annihilation

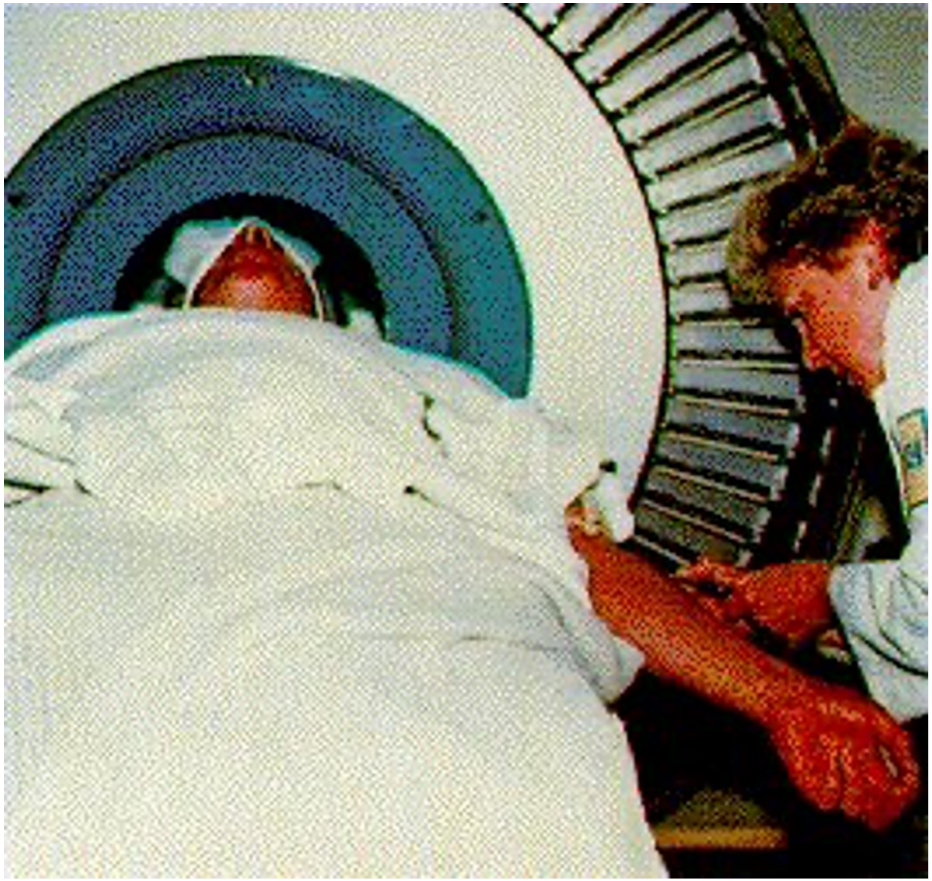


Figure 14: PET Scanner

Proportion of low-value examinations (Kjelle et al. 2022)

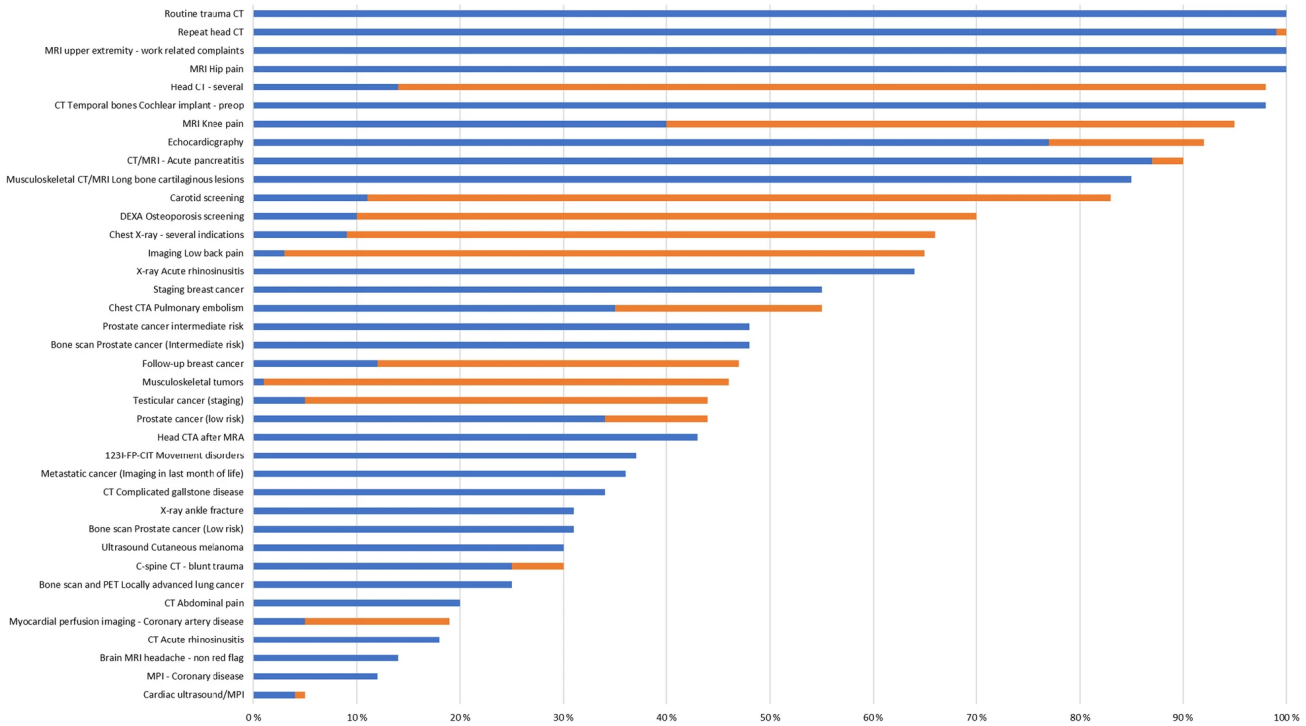


Figure 15: The blue bar represents the minimum rate and the combined blue and orange bar represents the maximum inappropriate rate

Appropriate vs low-value images?

- How is appropriate defined?

Appropriate vs low-value images?

- How is appropriate defined?
  - Professional consensus that this is appropriate

Appropriate vs low-value images?

- How is appropriate defined?
  - Professional consensus that this is appropriate
- How is value defined?

Appropriate vs low-value images?

- How is appropriate defined?
  - Professional consensus that this is appropriate
- How is value defined?
  - Imaging changes therapy

Appropriate vs low-value images?

- How is appropriate defined?
  - Professional consensus that this is appropriate
- How is value defined?
  - Imaging changes therapy
  - Information is obtained at minimal risk to patient

Example: Back pain

- “Me: Doctor, my lower back has hurt for years. I think I should get some imaging done.”
- “Doctor: ???”

Example: Back pain

The most commonly reported low-value procedures in musculoskeletal imaging was for low back pain. Ten studies demonstrated that X-ray, CT and MRI have a low impact on the treatment of patients without red flags, and 58.7% of MRI scans were negative, imaging for pain in the rest of the spine was also shown as low-value….

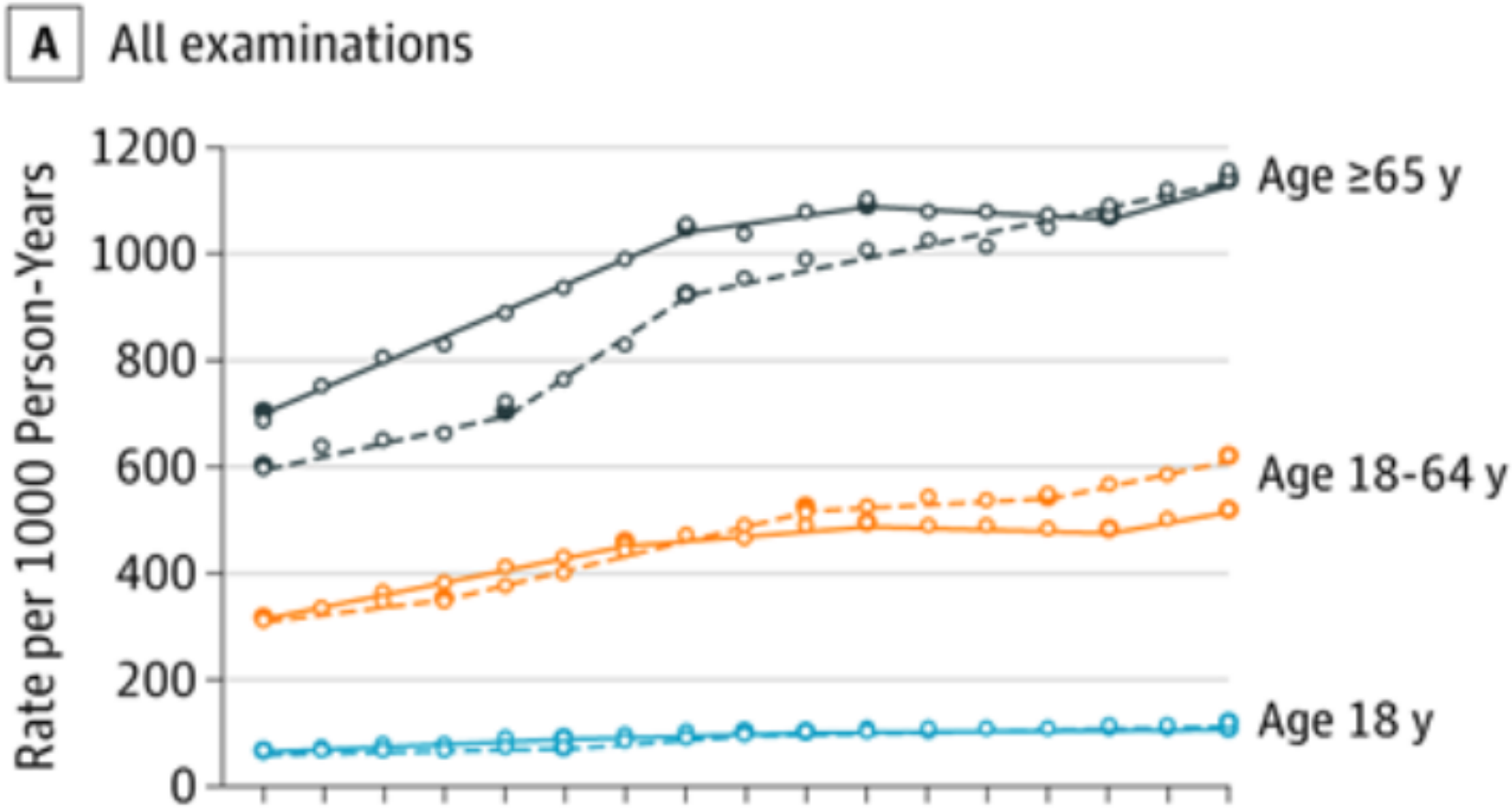
Example: Back pain

In addition, ***change in management were only seen in $< 1\%$ of routine post-op X-rays*** *after cervical (c)- or lumbar (l)-spine fusion. Another study found that even though* ***93% of the referrals for lumbar MRI were appropriate according to guidelines, only 13% of the scans showed actionable findings****. (Kjelle et al. 2022)*

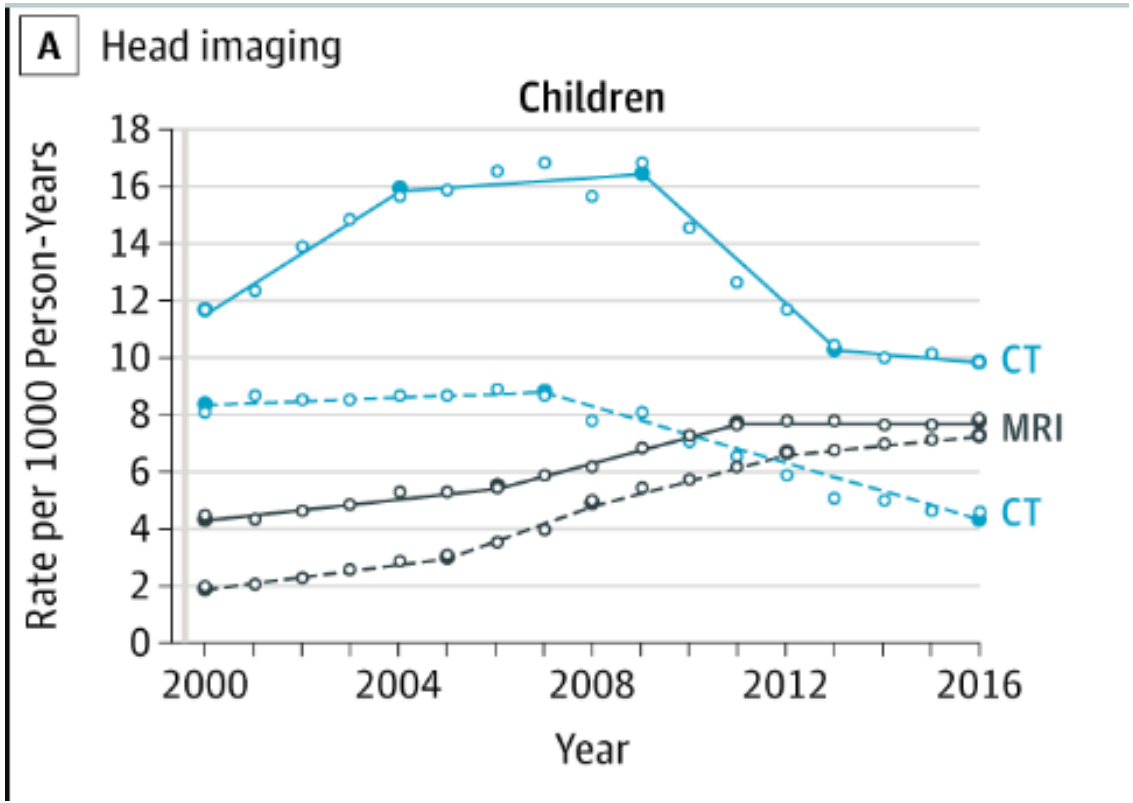
What drives utilization?

“Trends in Use of Medical Imaging in US Health Care Systems and in Ontario, Canada, 2000-2016” (Smith-Bindman et al. 2019)

"Trends in Use of Medical Imaging in US Health Care Systems and in Ontario, Canada, 2000-2016" [Smith-Bindman et al. (2019)]



"Trends in Use of Medical Imaging in US Health Care Systems and in Ontario, Canada, 2000-2016" [Smith-Bindman et al. (2019)]



Example: Low-value imaging in pediatrics

Overview of imaging identified as low-value in pediatrics sorted by body system

Low value for whom?

- low value for whom?

Low value for whom?

- low value for whom?
- Is negative information necessarily low value?

Low value for whom?

- low value for whom?
- Is negative information necessarily low value?
- Even if therapy hasn't changed, haven't we reduced uncertainty which has value?

Summary of low-value imaging

- ML has at least two possible roles

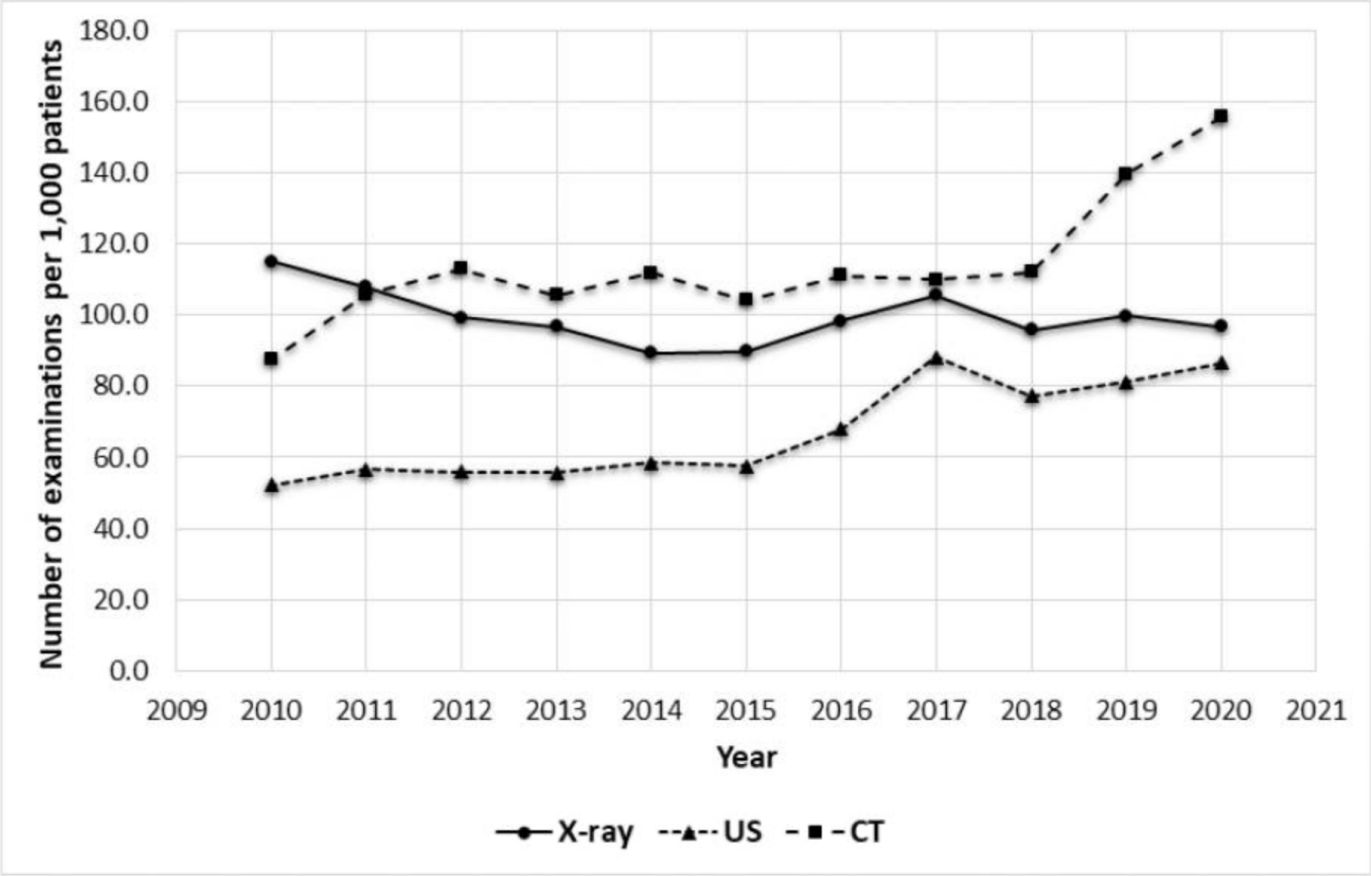
Summary of low-value imaging

- ML has at least two possible roles
  - Help ordering physicians order better

Summary of low-value imaging

- ML has at least two possible roles
  - Help ordering physicians order better
  - Help radiologists interpret images with low positivity rates

Change in utilization (Winder et al. 2021)



Normal/Abnormal or more?

- When is classification (normal/abnormal) sufficient?

Normal/Abnormal or more?

- When is classification (normal/abnormal) sufficient?
- Negative study?

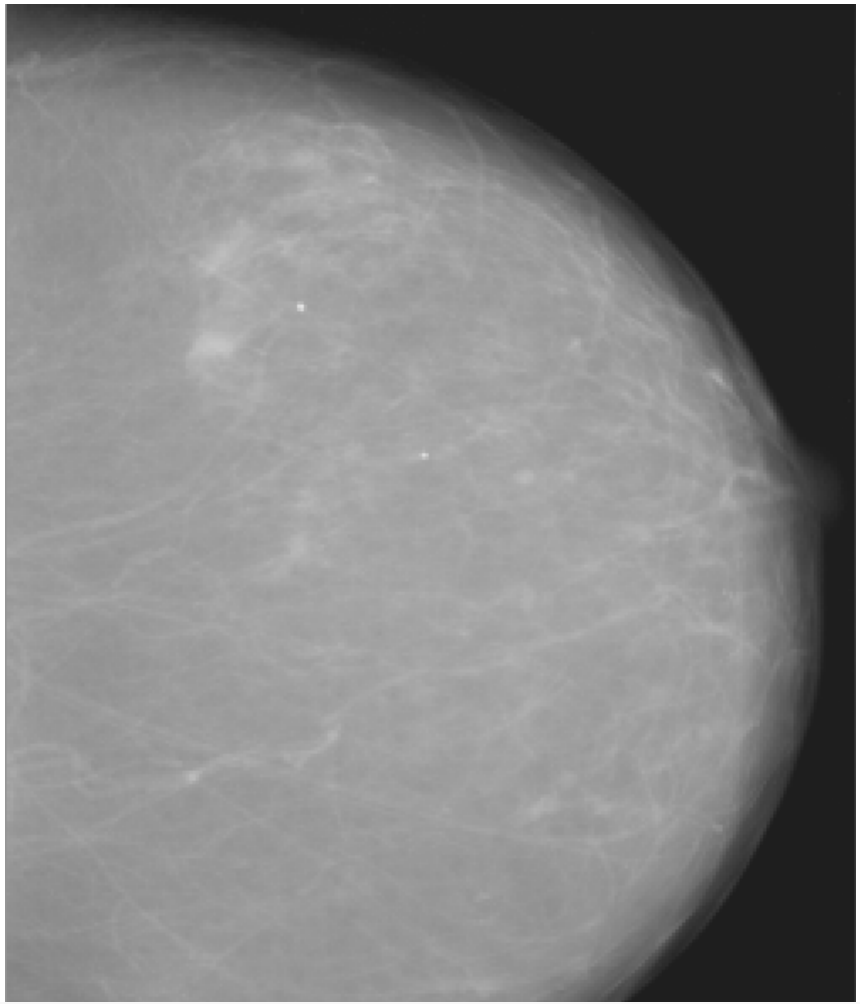
Normal/Abnormal or more?

- When is classification (normal/abnormal) sufficient?
- Negative study?
- Positive CXR for pneumonia?

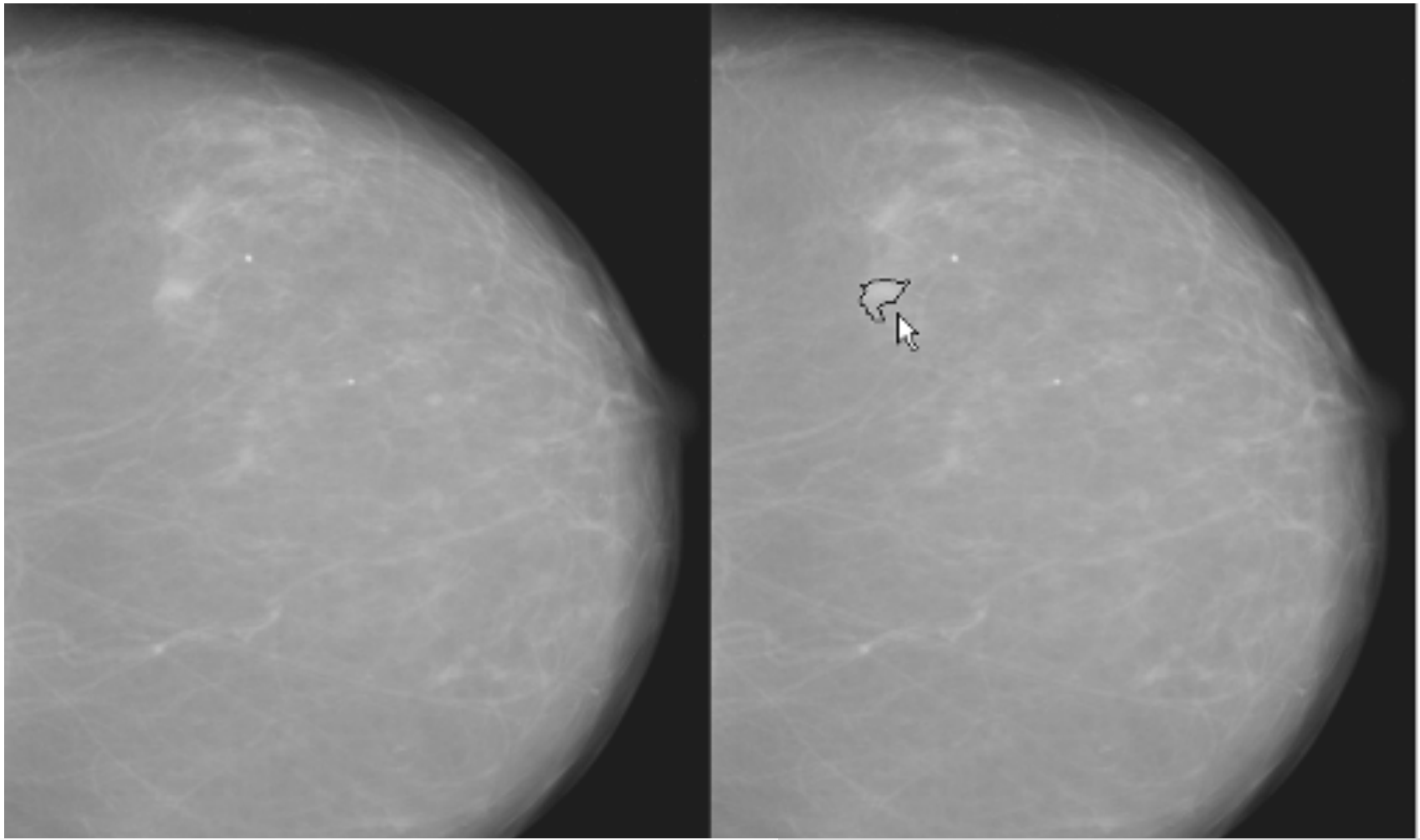
Normal/Abnormal or more?

- When is classification (normal/abnormal) sufficient?
- Negative study?
- Positive CXR for pneumonia?
- Positive CT for lung cancer (suspicious nodule)?

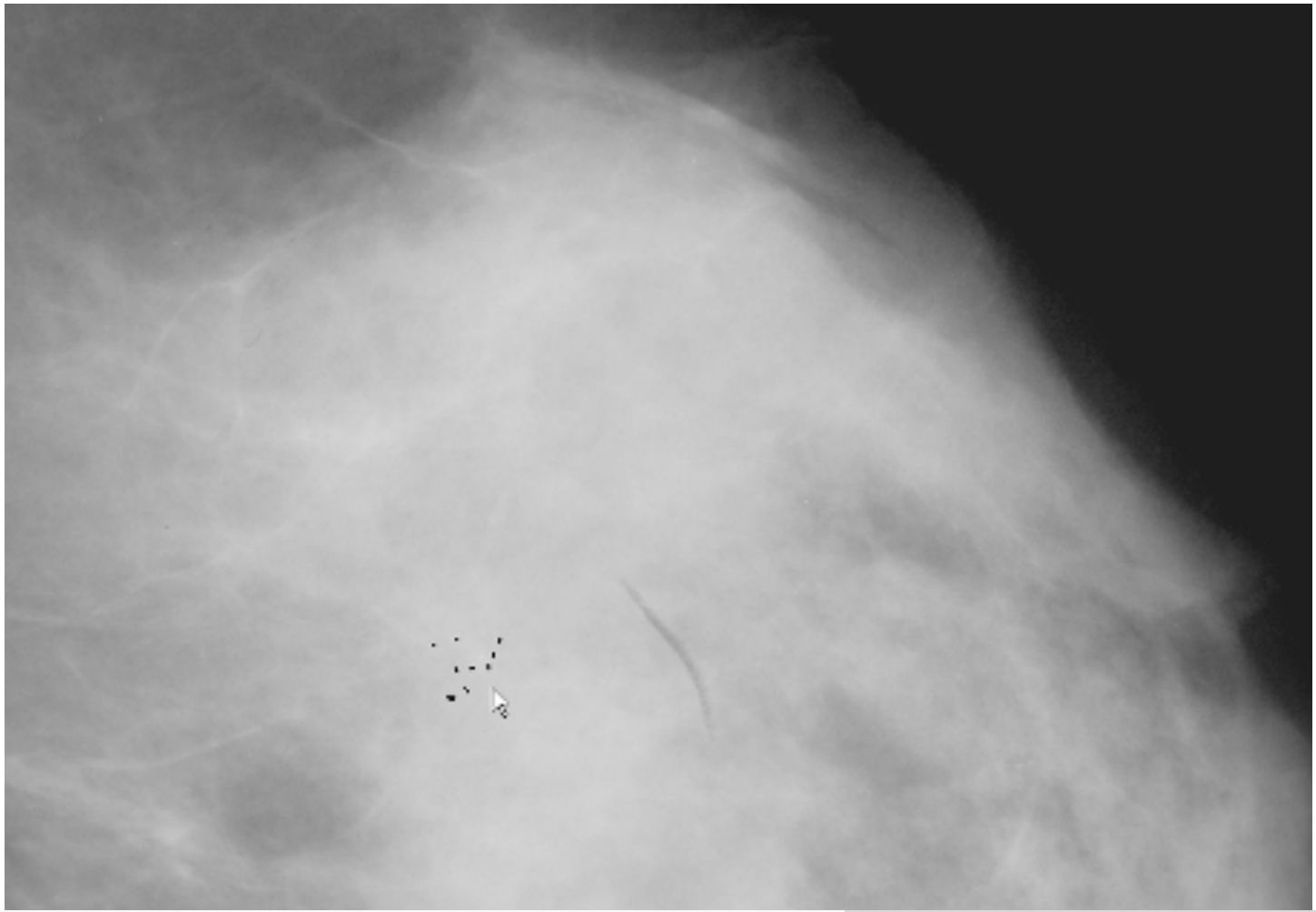
Example: Positive Mammogram



Example: Positive Mammogram



Example: Positive Mammogram #2



Normal/Abnormal or more?

- What is the confidence of our model?

Normal/Abnormal or more?

- What is the confidence of our model?
- Could I triage images based on my confidence that a study/image is negative?

ML Applications and Radiology Utilization (Naidich et al. 2020)

Editorial aside: Do ML applications match utilization?

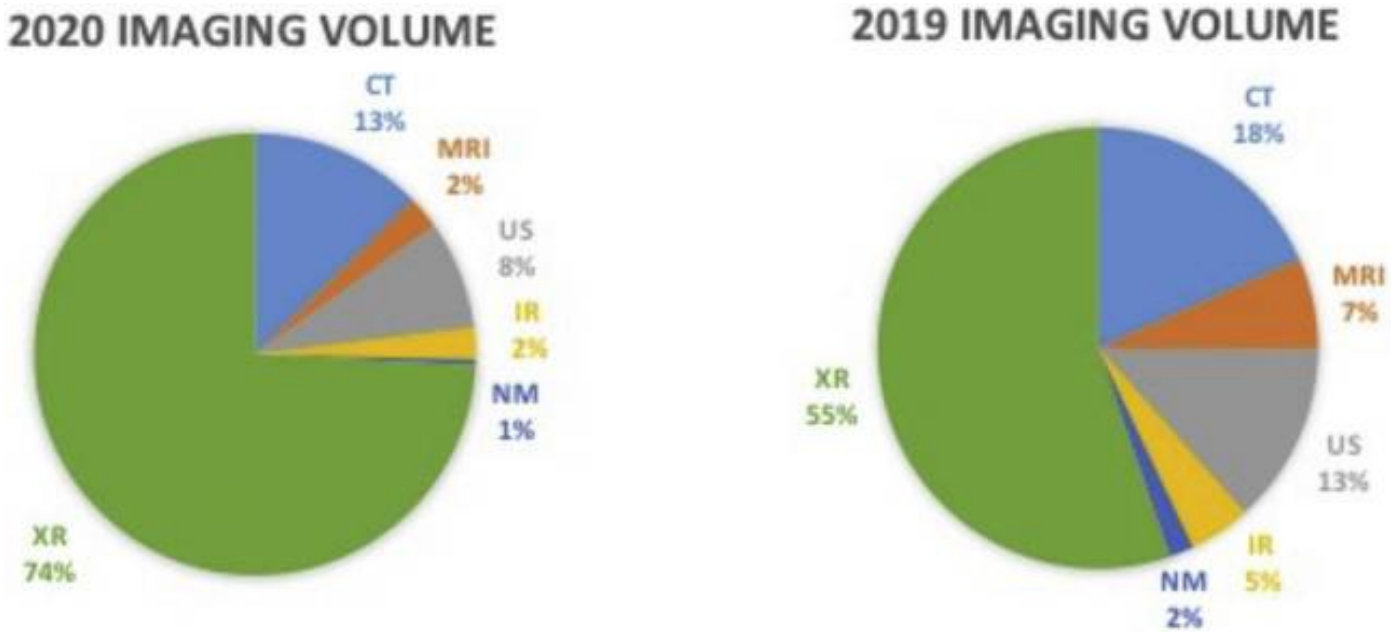


Figure 16: Modality utilization

Why do people create these applications?

Classification vs Object Identification

Qualitative vs Quantitative information

Carl Jaffe, MD:

"No one in clinical [drug] trials takes radiology seriously." [1]

- The inability to quantitatively monitor therapy
- The inability to validate findings
- Lack of transparency and data sharing between institutions
- Failure to integrate clinical information in the image assessment
- The unreliability of site interpretations of the imaging studies

Could ML help with any of these?

[1](CaBIG, Dec. 2005)

Quantification

- "A radiologist with a ruler is a radiologist in trouble." (anonymous 20th century radiologist)
- "Classification is necessary. But unless you can progress form classification to mathematics, your reasoning will not take you very far." (Alfred North Whitehead)

Example: Is it bigger?

Figure 17: kidney 2020

Figure 18: Kidney 2021

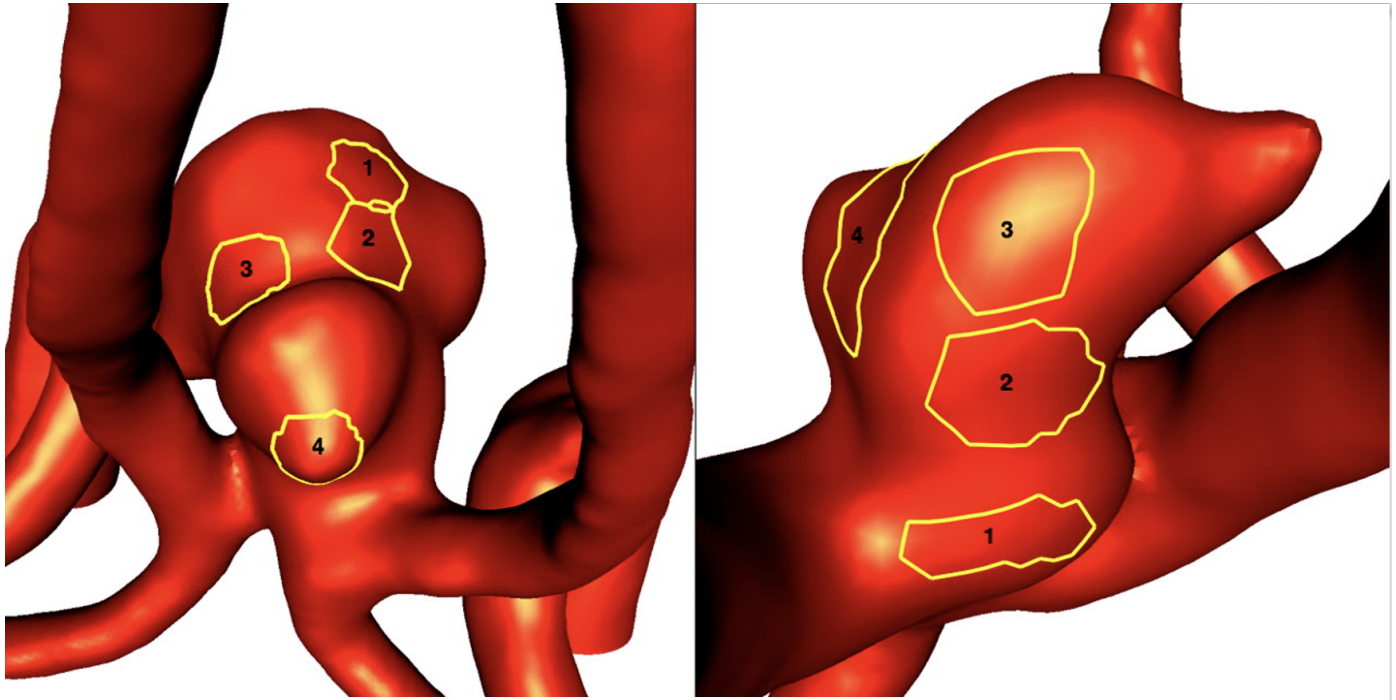
Example: Is it smaller?

Example: Assessing aneurysm risk

- Treat or don't treat? (Space shuttle story)

Figure 19: Cerebral aneurysm

Example: Assessing aneurysm risk



Castro, et al, AJNR November 2006 27: 2061-2068

Example: Risk of developing aneurysm

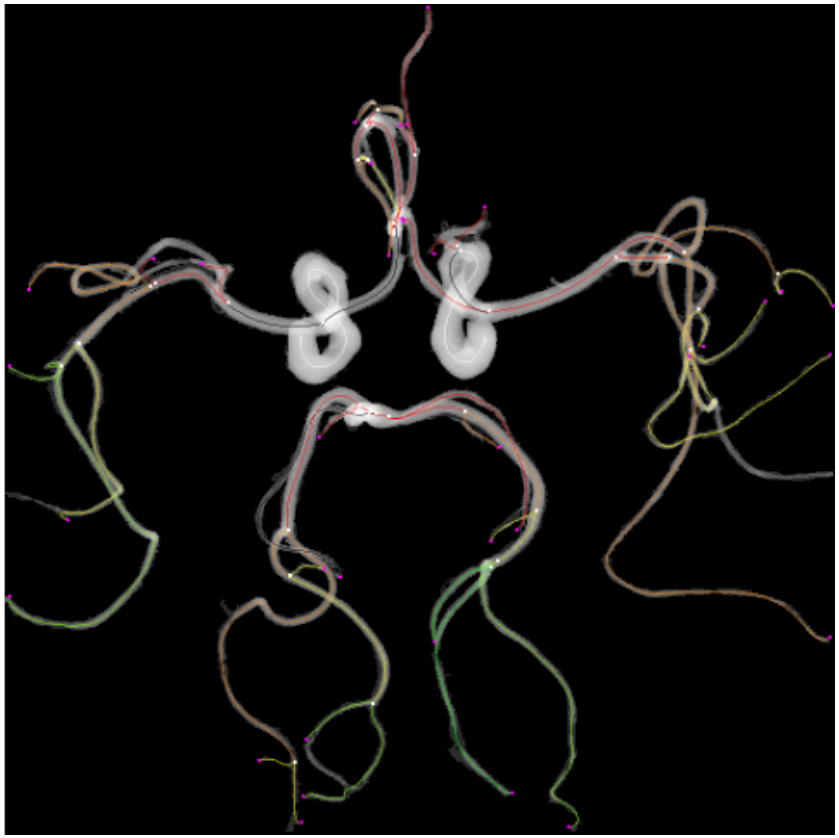


Figure 20: Semantically labeled Circle-of-Willis

Validation: Example

Figure 21: CT of liver

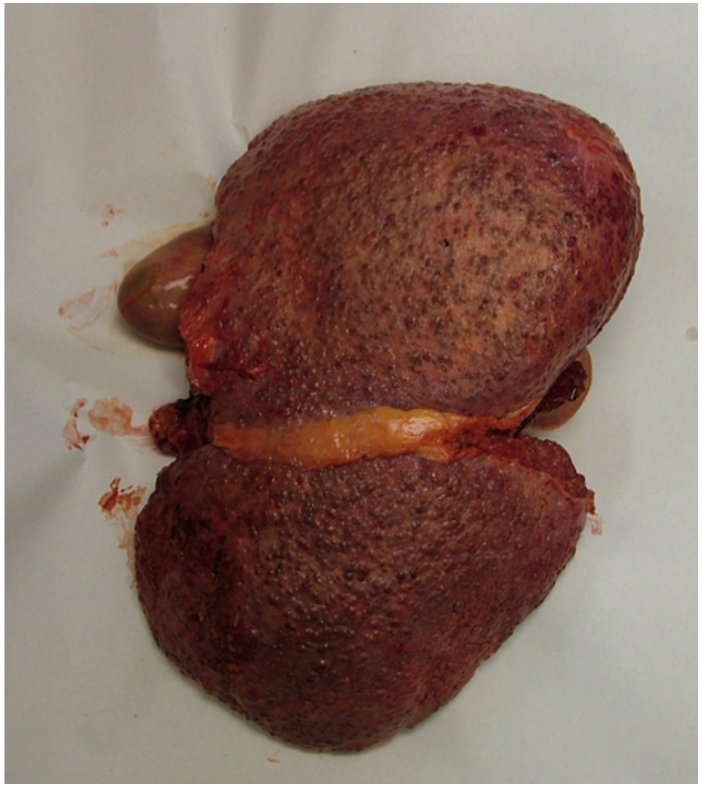


Figure 22: explanted liver

Validation: Example

Figure 23: "systematic" evaluation of liver

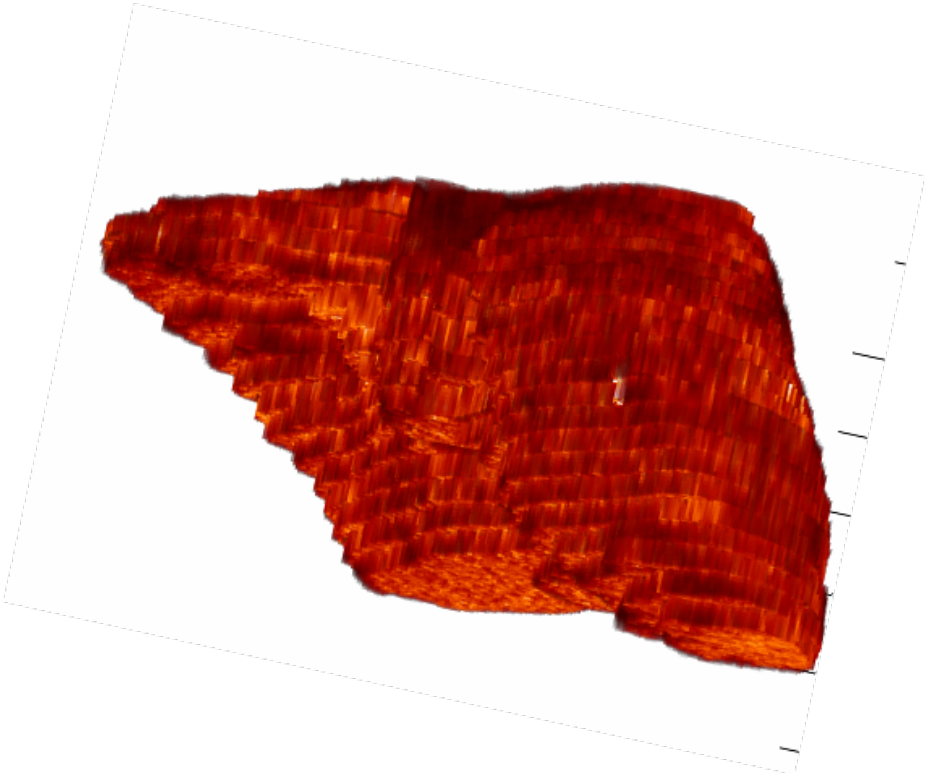


Figure 24: 3D model of liver

“Failure to integrate clinical information...”

- ML relevant summary of patient medical record

“The unreliability of site interpretations…”

References I

Bick, U., R. M. Trimboli, A. Athanasiou, C. Balleyguier, P. A. T. Baltzer, M. Bernathova, K. Borbély, et al. 2020. "Image-guided breast biopsy and localisation: recommendations for information to women and referring physicians by the European Society of Breast Imaging." *Insights Imaging* 11 (1): 12.

Dyro, J. F., and J. Dyro. 2004. *Clinical Engineering Handbook*. Academic Press Series in Biomedical Engineering. Elsevier Science. https://books.google.com.au/books?id=EIeQhdrW2VMC.

Elfgen, C., B. Papassotiropoulos, Z. Varga, L. Moskovszky, M. Nap, U. Güth, A. Baege, E. Amann, F. Chiesa, and C. Tausch. 2019. "Comparative analysis of confocal microscopy on fresh breast core needle biopsies and conventional histology." *Diagn Pathol* 14 (1): 58.

Jarrahi, Mohammad Hossein, Ali Memariani, and Shion Guha. 2023. "The Principles of Data-Centric AI." *Commun. ACM* 66 (8): 84–92. https://doi.org/10.1145/3571724.

References II

Kevles, B. 1997. *Naked to the Bone: Medical Imaging in the Twentieth Century*. Sloan Technology Series. Rutgers University Press. https://books.google.com.au/books?id=et2k_o-K-fQC.

Kjelle, E., E. R. Andersen, A. M. Krokeide, L. J. J. Soril, L. van Bodegom-Vos, F. M. Clement, and B. M. Hofmann. 2022. "Characterizing and quantifying low-value diagnostic imaging internationally: a scoping review." *BMC Med Imaging* 22 (1): 73.

Naidich, J. J., A. Boltyenkov, J. J. Wang, J. Chusid, D. Hughes, and P. C. Sanelli. 2020. "Coronavirus Disease 2019 (COVID-19) Pandemic Shifts Inpatient Imaging Utilization." *J Am Coll Radiol* 17 (10): 1289–98.

Smith-Bindman, R., M. L. Kwan, E. C. Marlow, M. K. Theis, W. Bolch, S. Y. Cheng, E. J. A. Bowles, et al. 2019. "Trends in Use of Medical Imaging in US Health Care Systems and in Ontario, Canada, 2000-2016." *JAMA* 322 (9): 843–56.

References III

Winder, M., A. J. Owczarek, J. Chudek, J. Pilch-Kowalczyk, and J. Baron. 2021. "Are We Overdoing It? Changes in Diagnostic Imaging Workload during the Years 2010-2020 including the Impact of the SARS-CoV-2 Pandemic." *Healthcare (Basel)* 9 (11).