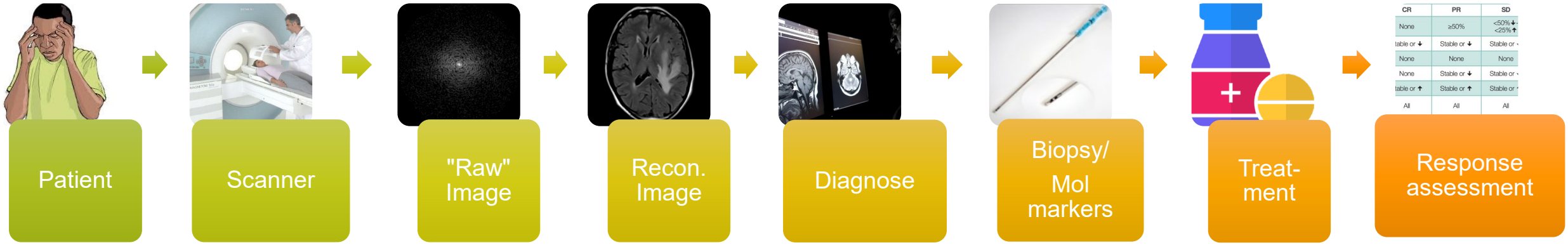
University of Colorado **Anschutz Medical Campus**

AI and Quantitative Imaging

Jayashree Kalpathy-Cramer, PhD
Chief, Division of Artificial Medical Intelligence
University of Colorado

AI/ML is being used widely throughout the entire patient journey



AI can be applied throughout the workflow

Deep Learning in Radiological Imaging

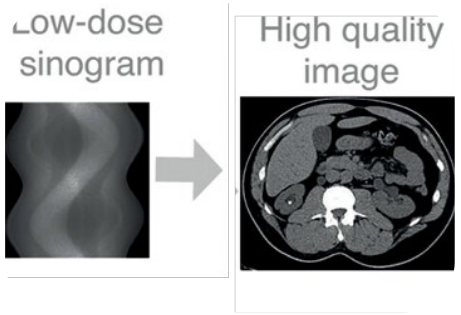
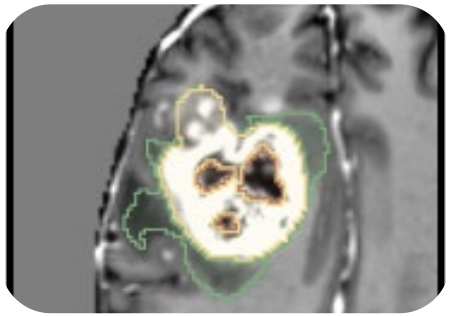
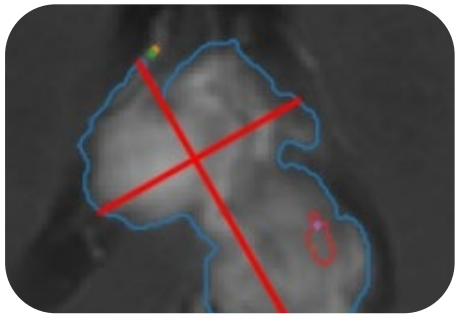


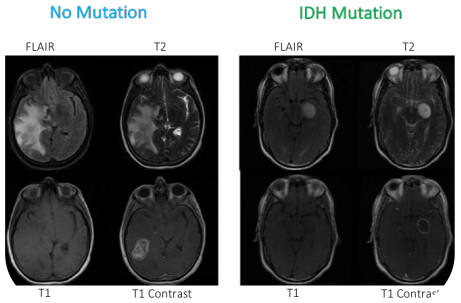
Image reconstruction



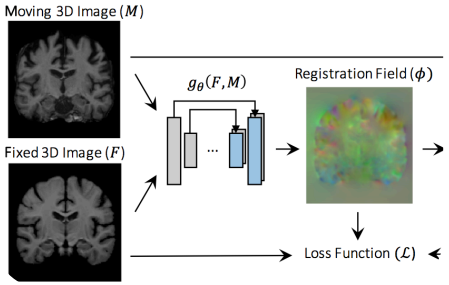
Segmentation



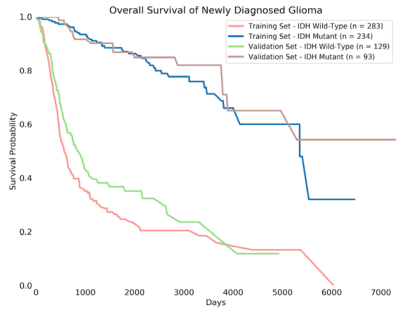
Response Assessment



Radiogenomics



Registration

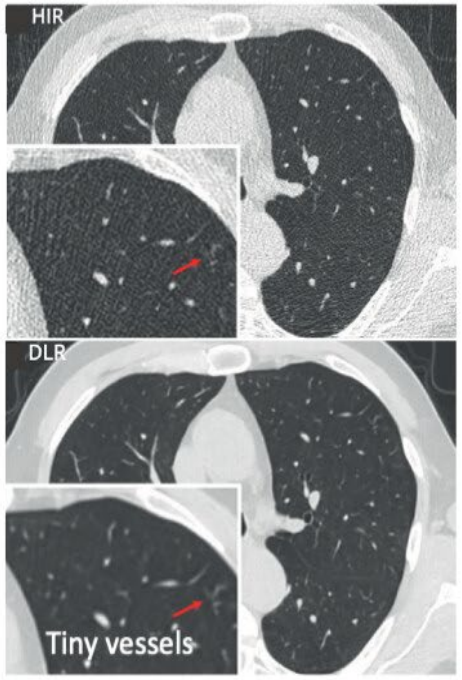


Survival Prediction

University of Colorado
Anschutz Medical Campus

AI is being used extensively in image reconstruction

Deep Learning Image Reconstruction for CT: Technical Principles and Clinical Prospects

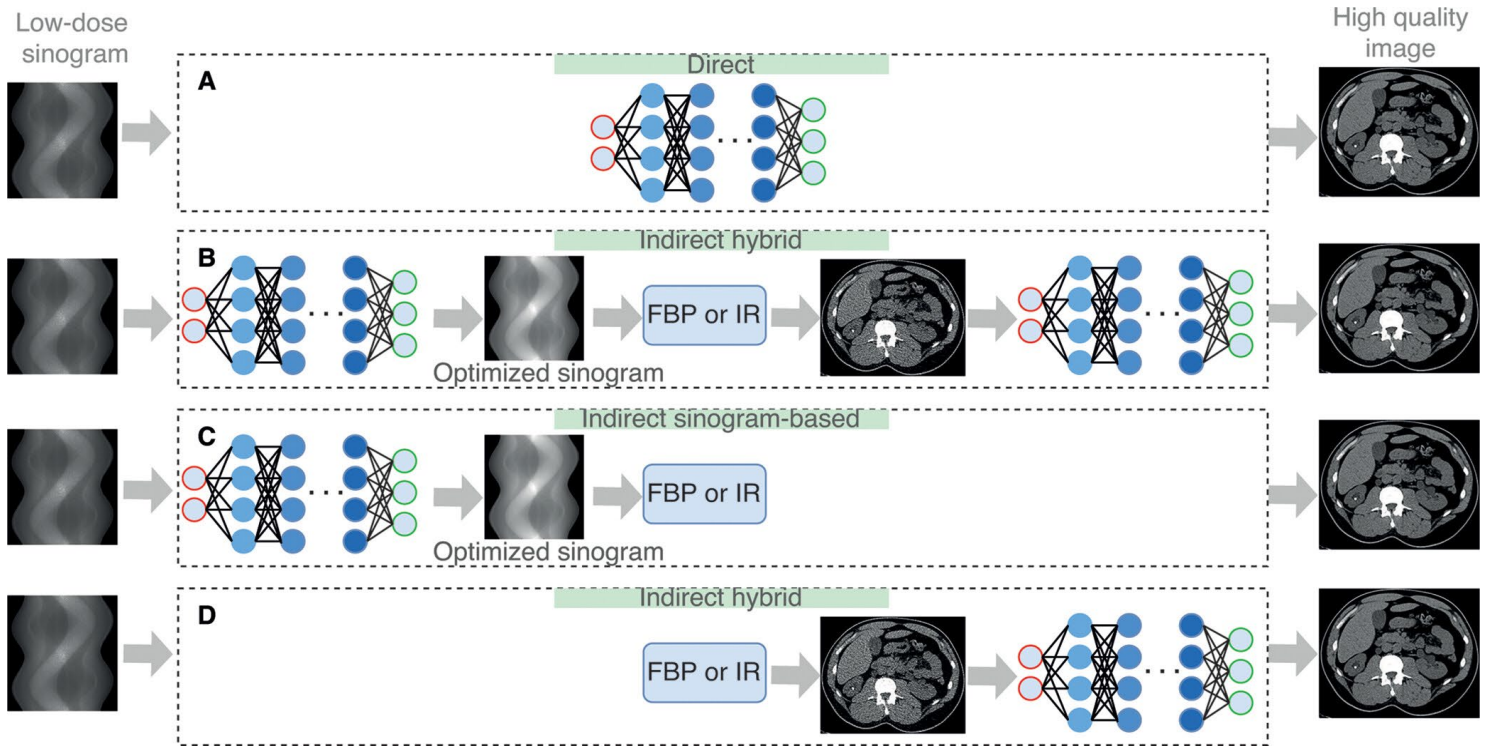


- Deep learning reconstruction (DLR) algorithms can be applied in the raw data domain, image domain, or both.
- Compared with filtered back projection and hybrid iterative reconstruction (HIR), DLR provides improved image quality.
- DLR allows for radiation dose reductions of 30%–71% compared with HIR with maintained image quality due to noise reduction.
- Deep learning–based metal artifact reduction may remove metal artifacts more accurately than current state-of-the-art methods.

Koetzier LR and Mastrodicasa D et al. Published Online: January 31, 2023
https://doi.org/10.1148/radiol.221257

Radiology

AI is being used extensively in image reconstruction



Example products

- TrueFidelity
- AiCE
- Precise Image
- PixelShine
- ClariCT.AI
- Air Recon DL
- Subte

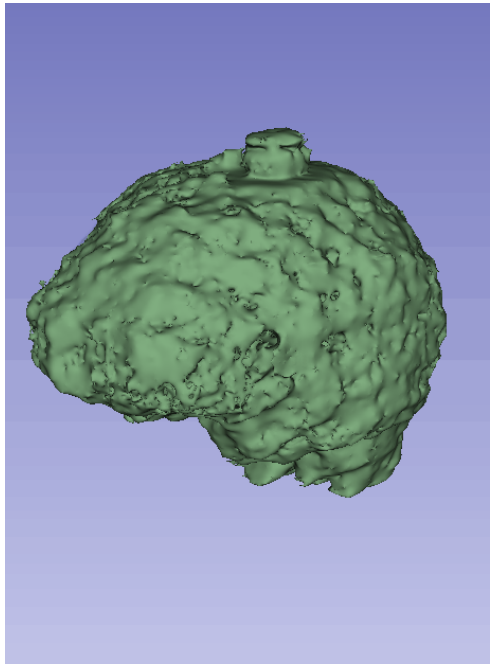
Koetzier LR. Published Online: January 31, 2023
https://doi.org/10.1148/radiol.221257

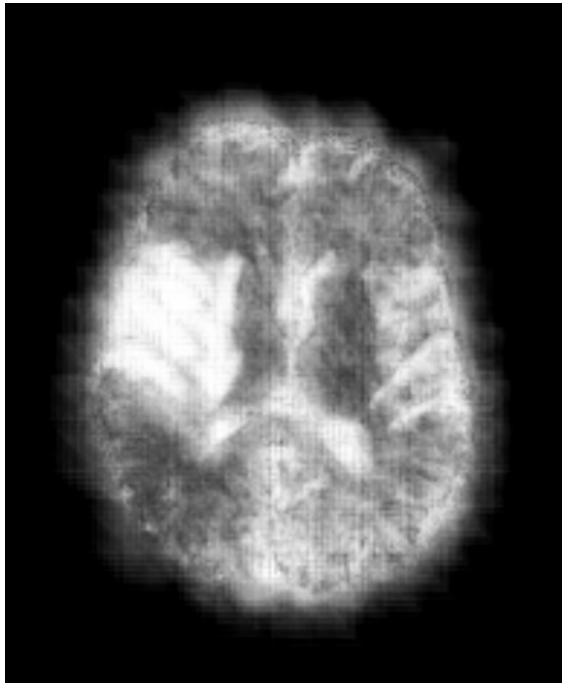
Radiology

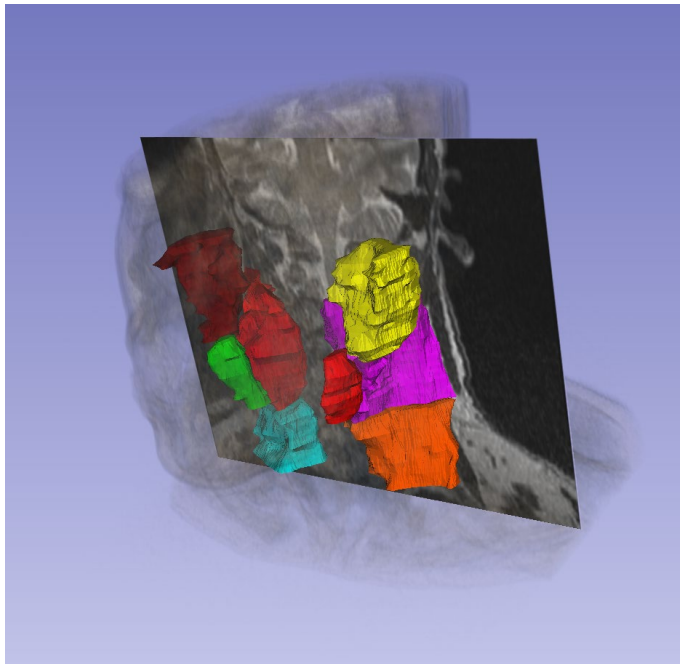
Segmentation (delineation of object boundary) is often used in oncology and radiation oncology

Quantifying tumor burden at a single time point and longitudinally

Contouring of tumors and organs at risk is key in radiation therapy planning







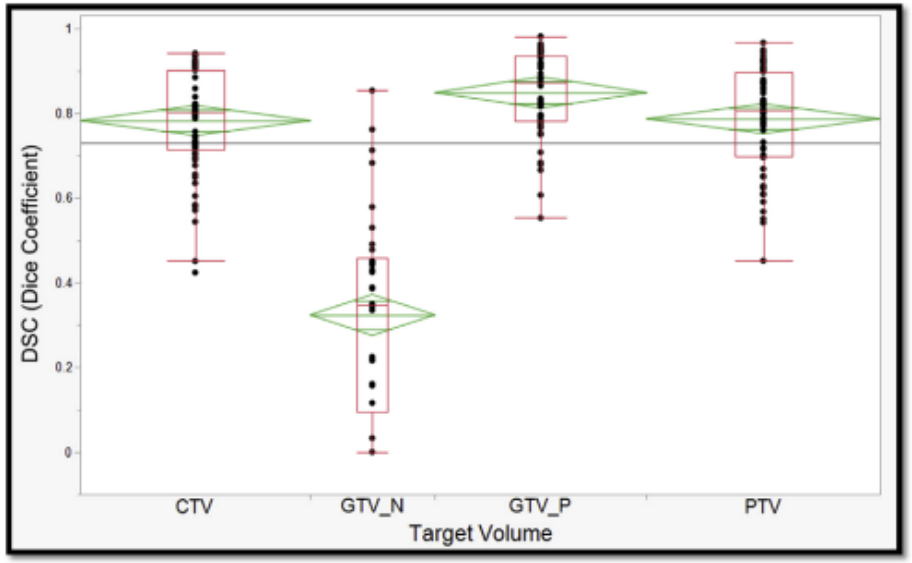
University of Colorado
Anschutz Medical Campus

Radiation Oncology Research: Contouring Variability

Original Research Article

An *in-silico* quality assurance study of contouring target volumes in thoracic tumors within a cooperative group setting ☆

Hesham Elhalawani [a,*], Baher Elgohari [a], Timothy A. Lin [a,b], Abdallah S.R. Mohamed [a,c], Thomas J. Fitzgerald [d], Fran Laurie [d], Kenneth Ulin [d], Jayashree Kalpathy-Cramer [e], Thomas Guerrero [f], Emma B. Holliday [a], Gregory Russo [g,1], Abhilasha Patel [h], William Jones [h], Gary V. Walker [a,i], Musaddiq Awan [j,2], Mehee Choi [k,3], Roi Dagan [l], Omar Mahmoud [m,4], Anna Shapiro [n], Feng-Ming (Spring) Kong [o], Daniel Gomez [a], Jing Zeng [p], Roy Decker [q], Femke O.B. Spoelstra [r], Laurie E. Gaspar [s], Lisa A. Kachnic [t], Charles R. Thomas Jr. [u,*], Paul Okunieff [v], Clifton D. Fuller [a,*]



University of Colorado
Anschutz Medical Campus

Original Research Article

A prospective in silico analysis of interdisciplinary and interobserver spatial variability in post-operative target delineation of high-risk oral cavity cancers: Does physician specialty matter?

Sweet Ping Ng [a,*], Brandon A. Dyer [b], Jayashree Kalpathy-Cramer [c], Abdallah Sherif Radwan Mohamed [a], Musaddiq J. Awan [d], G. Brandon Gunn [a], Jack Phan [a], Mark Zafereo [e], J. Matthew Debnam [f], Carol M. Lewis [e], Rivka R. Colen [f], Michael E. Kupferman [e], Nandita Guha-Thakurta [f], Guadalupe Canahuate [g], G. Elisabeta Marai [h], David Vock [i], Bronwyn Hamilton [j], John Holland [k], Carlos E. Cardenas [l], Stephen Lai [e], David Rosenthal [a], Clifton David Fuller [a]

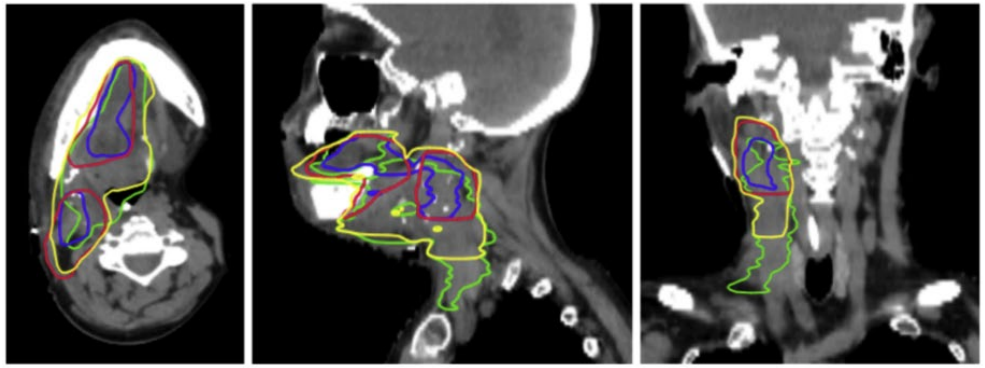
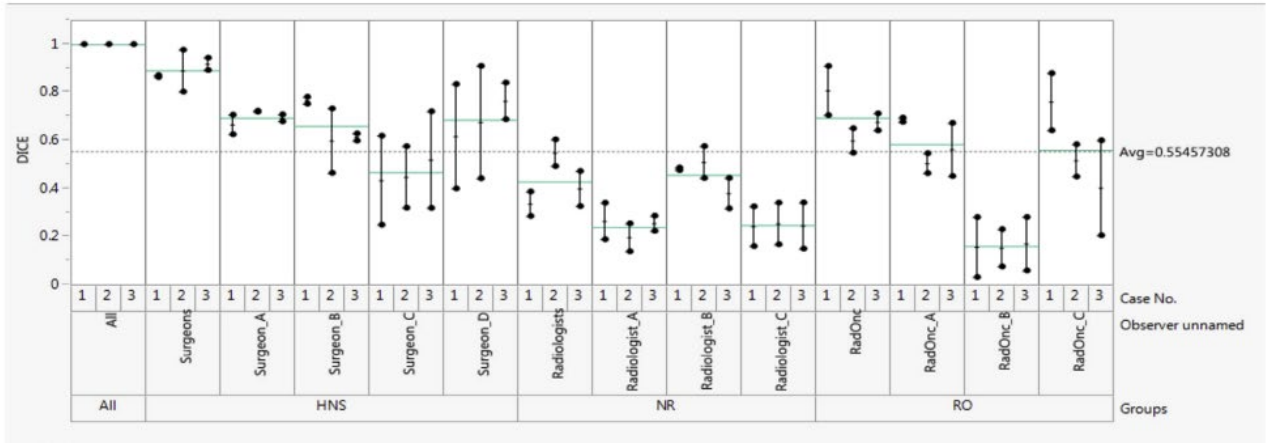


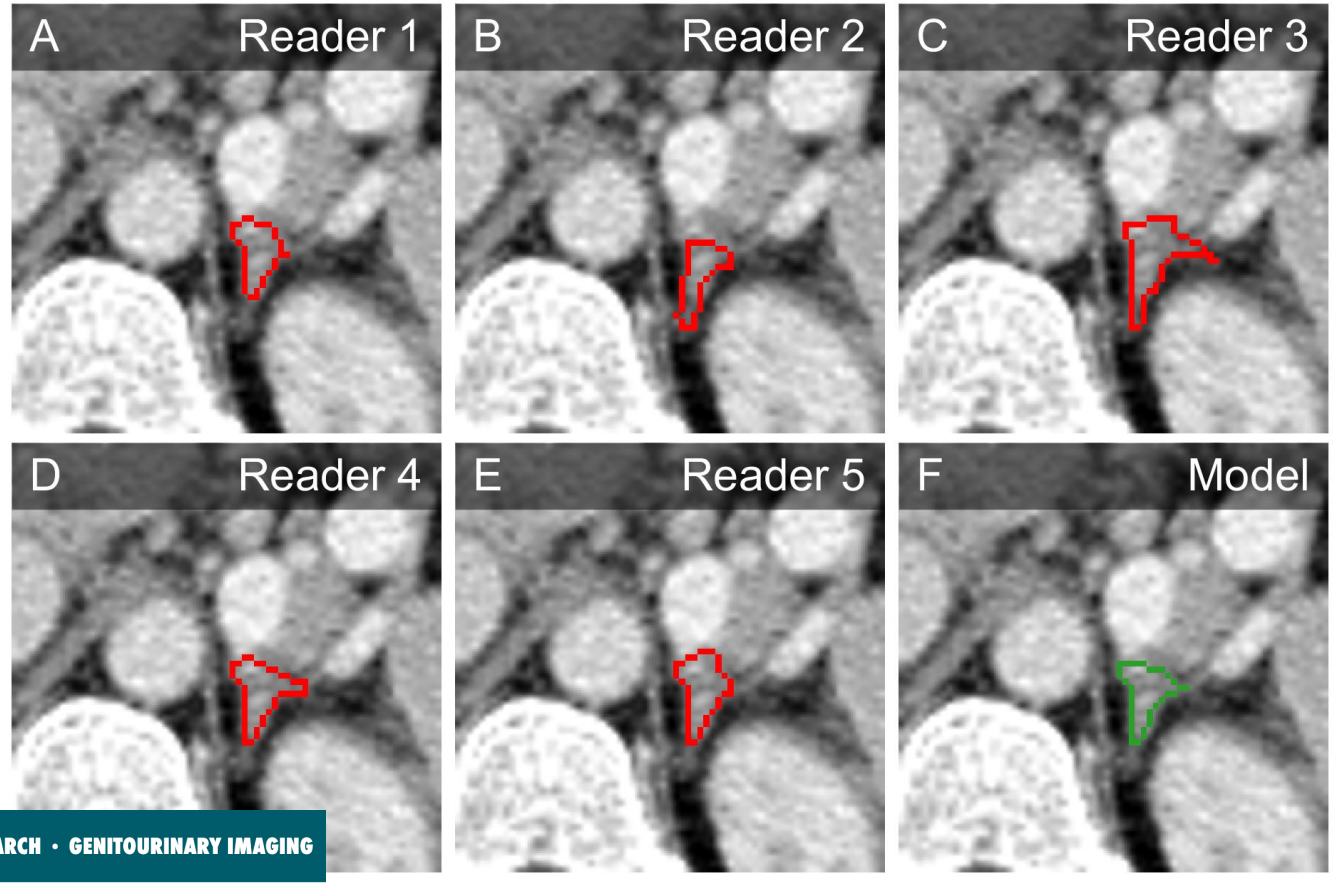
Fig. 4. An example of contours delineated by radiation oncologist (red), radiologist (blue), and surgeon (green), and STAPLE contour (yellow).



Adrenal gland segmentation

5 expert radiologists vs. automatic model

Inter-reader dice score coefficient (DSC) was *not* statistically significant from model-reader DSC (p = 0.35)



Radiology

ORIGINAL RESEARCH • GENITOURINARY IMAGING

Machine Learning for Adrenal Gland Segmentation and Classification of Normal and Adrenal Masses at CT

Cory Robinson-Weiss, MD • Jay Patel, BS* • Bernardo C. Bizzo, MD • Daniel I. Glazer, MD • Christopher P. Bridge, DPhil • Katherine P. Andriole, PhD • Borna Dabiri, MD, PhD • John K. Chin, MD •*

Automatic Segmentation

1412

Neuro-Oncology

21(11), 1412–1422, 2019 | doi:10.1093/neuonc/noz106 | Advance Access date 13 June 2019

Automatic assessment of glioma burden: a deep learning algorithm for fully automated volumetric and bidimensional measurement

Ken Chang,† Andrew L. Beers,† Harrison X. Bai,† James M. Brown, K. Ina Ly, Xuejun Li, Joeky T. Senders, Vasileios K. Kavouridis, Alessandro Boaro, Chang Su, Wenya Linda Bi, Otto Rapalino, Weihua Liao, Qin Shen, Hao Zhou, Bo Xiao, Yinyan Wang, Paul J. Zhang, Marco C. Pinho, Patrick Y. Wen, Tracy T. Batchelor, Jerrold L. Boxerman, Omar Arnaout, Bruce R. Rosen, Elizabeth R. Gerstner, Li Yang, Raymond Y. Huang, and Jayashree Kalpathy-Cramer

Neuroinformatics (2021) 19:127–140
https://doi.org/10.1007/s12021-020-09477-5

SOFTWARE ORIGINAL ARTICLE

Check for updates

DeepNeuro: an open-source deep learning toolbox for neuroimaging

Andrew Beers[1] · James Brown[1] · Ken Chang[1] · Katharina Hoebel[1] · Jay Patel[1] · K. Ina Ly[1,3] · Sara M. Tolaney[2] · Priscilla Brastianos[3] · Bruce Rosen[1] · Elizabeth R. Gerstner[1,3] · Jayashree Kalpathy-Cramer[1]

Published online: 23 June 2020

Anschutz Medical Campus

Neuro-Oncology

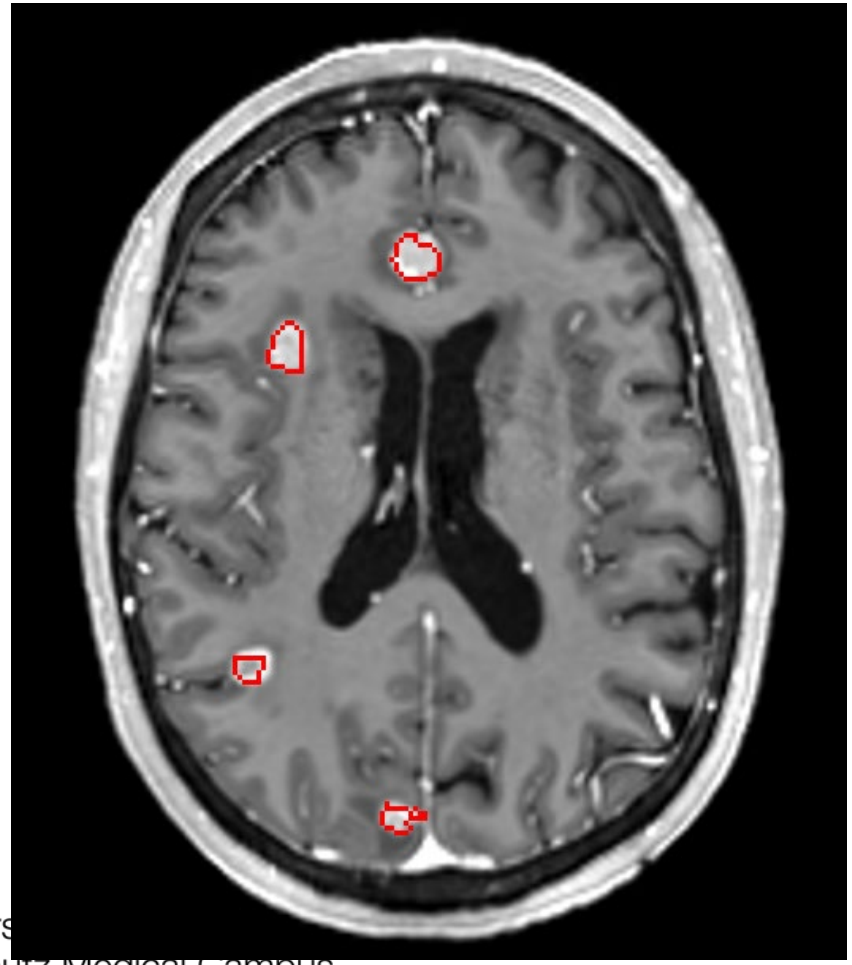
24(2), 289–299, 2022 | https://doi.org/10.1093/neuonc/noab151 | Advance Access date 26 June 2021

Deep learning-based automatic tumor burden assessment of pediatric high-grade gliomas, medulloblastomas, and other leptomeningeal seeding tumors

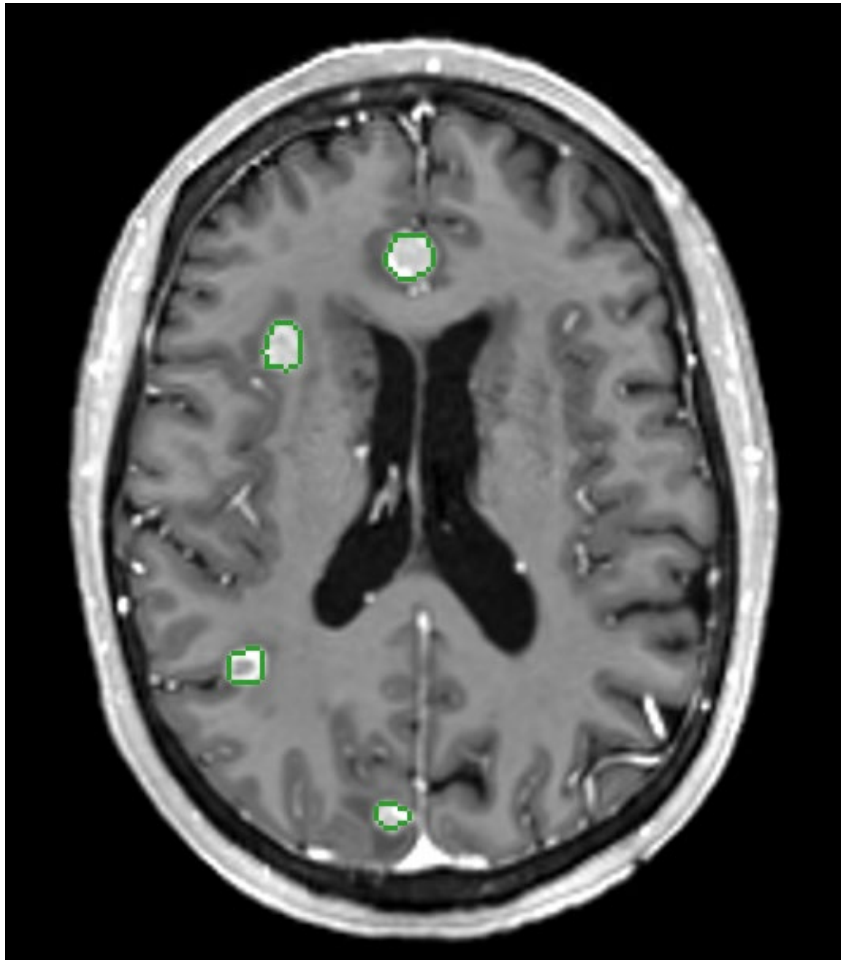
Jian Peng,† Daniel D. Kim,† Jay B. Patel, Xiaowei Zeng, Jiaer Huang, Ken Chang, Xinping Xun, Chen Zhang, John Sollee, Jing Wu, Deepa J. Dalal, Xue Feng, Hao Zhou, Chengzhang Zhu, Beiji Zou, Ke Jin, Patrick Y. Wen, Jerrold L. Boxerman, Katherine E. Warren, Tina Y. Poussaint, Lisa J. States, Jayashree Kalpathy-Cramer, Li Yang, Raymond Y. Huang, and Harrison X. Bai

Automatic segmentation and tracking of brain metastases

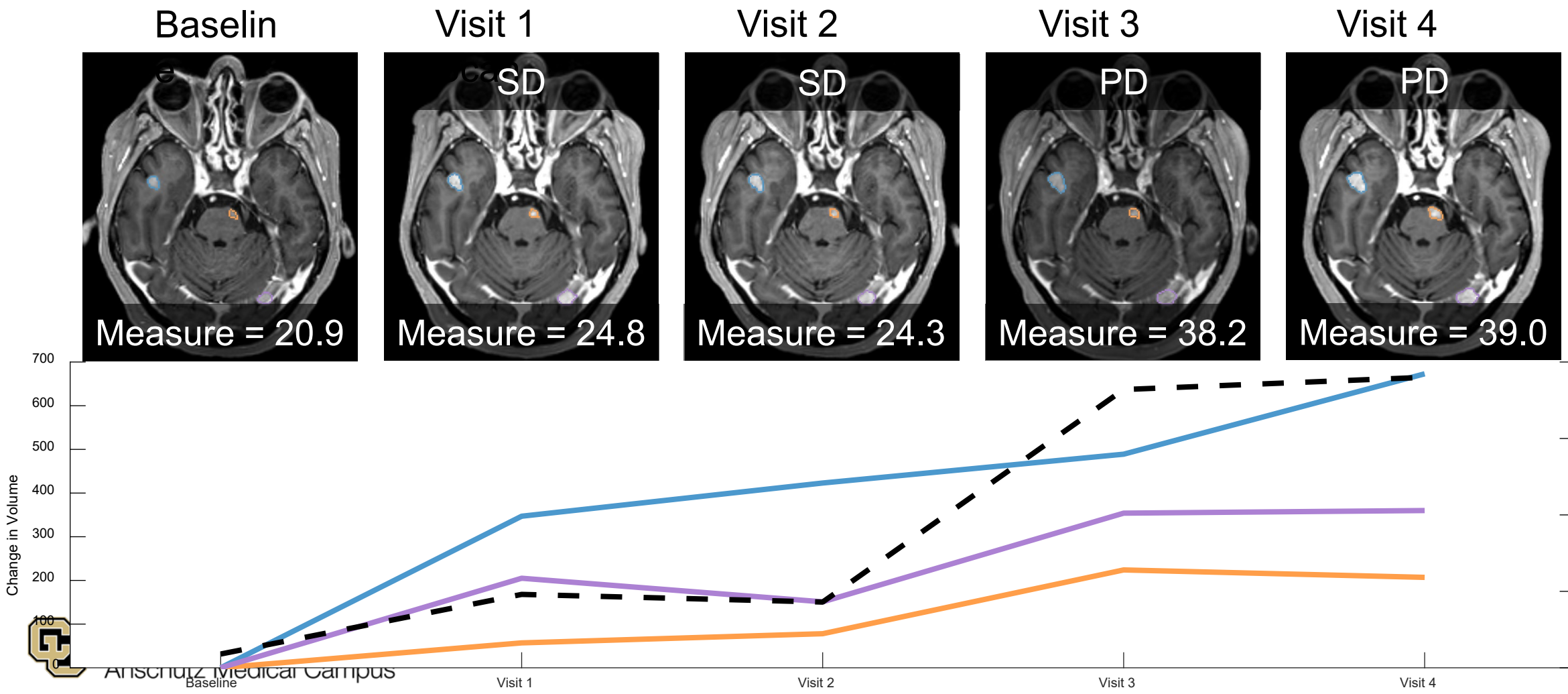
Manual



Automatic

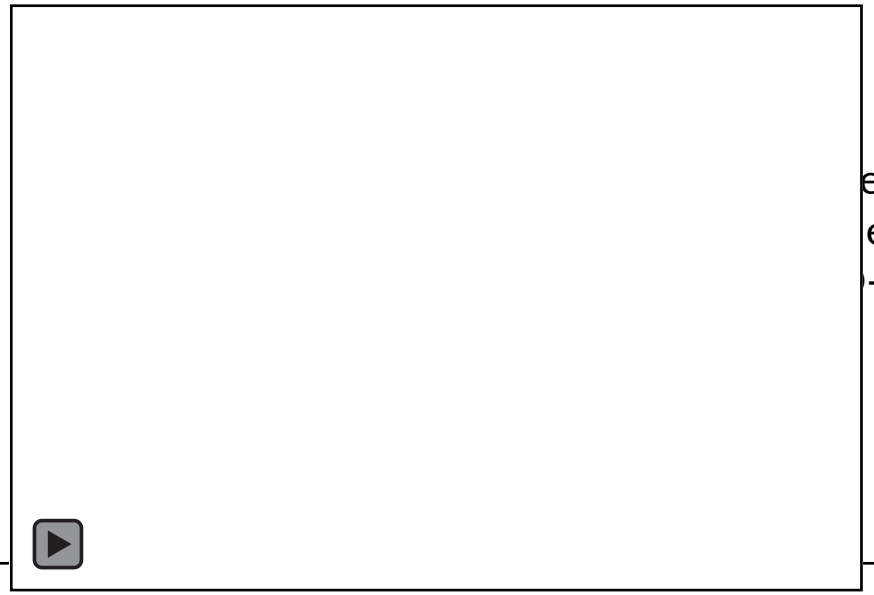


Response Assessment

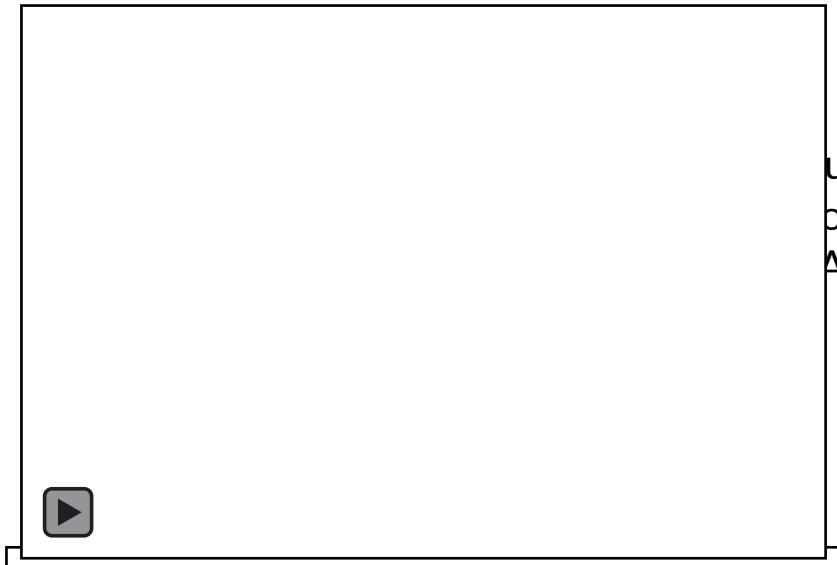


Anschutz Medical Campus

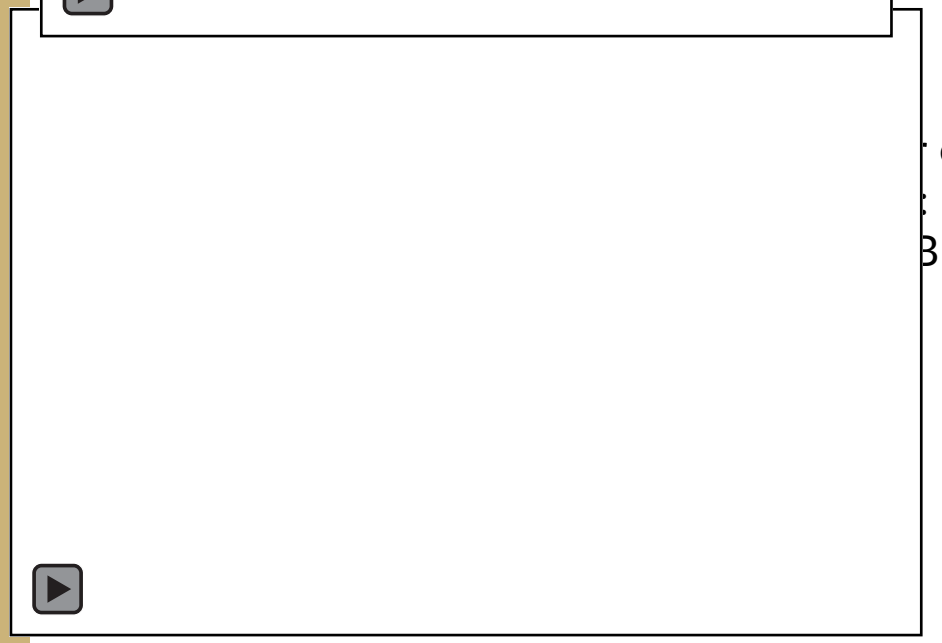
Volume vs AutoRANO-BM



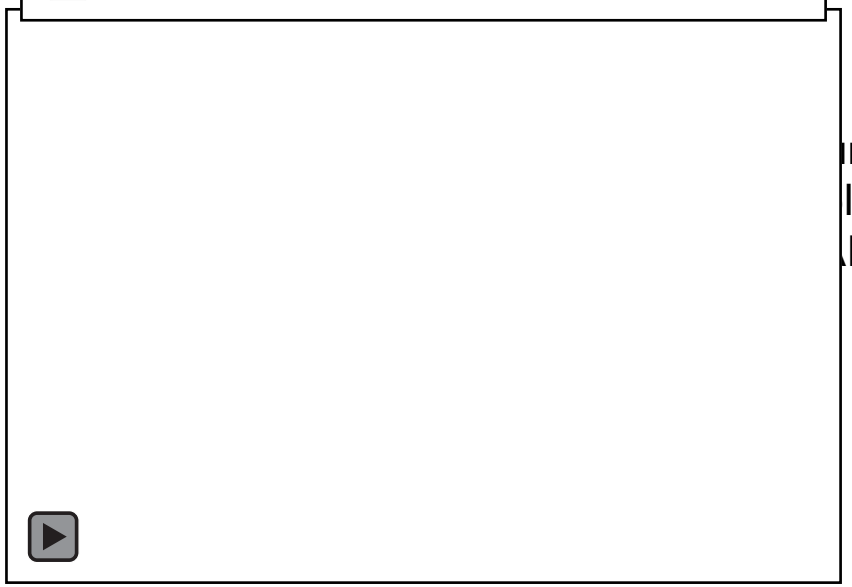
er of Mets: 1
e: 1730mm^3
-BM: 21.6mm



umber of Mets: 7
olume: 1720mm^3
ANO-BM: 39.4mm

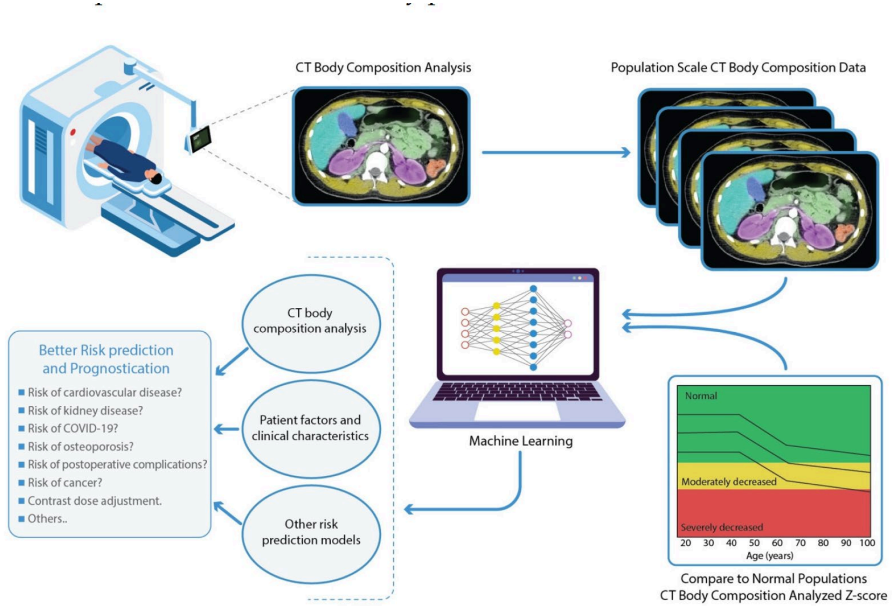


of Mets: 2
1750mm^3
3M: 16.8mm



mber of Mets: 15
lume: 1700mm^3
NO-BM: 0.0mm

Opportunistic Screening



Review

Role of Machine Learning-Based CT Body Composition in Risk Prediction and Prognostication: Current State and Future Directions

Tarig Elhakim [1,2,*], Kelly Trinh [3], Arian Mansur [4], Christopher Bridge [2,4] and Dania Daye [2,4,*]

nature medicine

Article

https://doi.org/10.1038/s41591-023-02232-8

Body composition and lung cancer-associated cachexia in TRACERx

ORIGINAL RESEARCH | Annals of Internal Medicine

Opportunistic Screening for Osteoporosis Using Abdominal Computed Tomography Scans Obtained for Other Indications

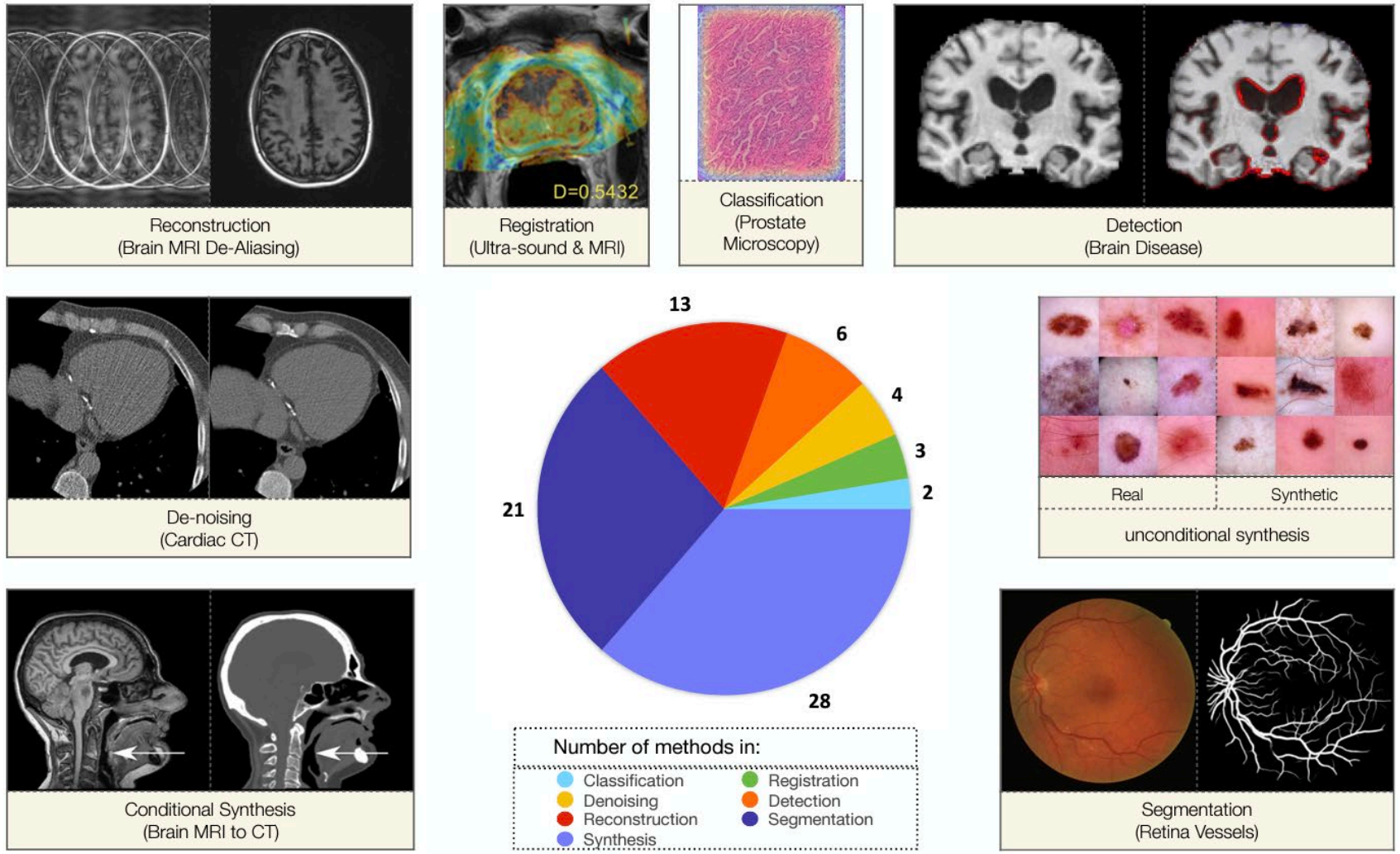
Perry J. Pickhardt, MD; B. Dustin Pooler, MD; Travis Lauder, BS; Alejandro Muñoz del Rio, PhD; Richard J. Bruce, MD; and Neil Binkley, MD

Opportunistic Screening: *Radiology* Scientific Expert Panel

Perry J. Pickhardt, MD • Ronald M. Summers, MD, PhD • John W. Garrett, PhD • Arun Krishnaraj, MD • Sheela Agarwal, MD • Keith J. Dreyer, DO, PhD • Gregory N. Nicola, MD

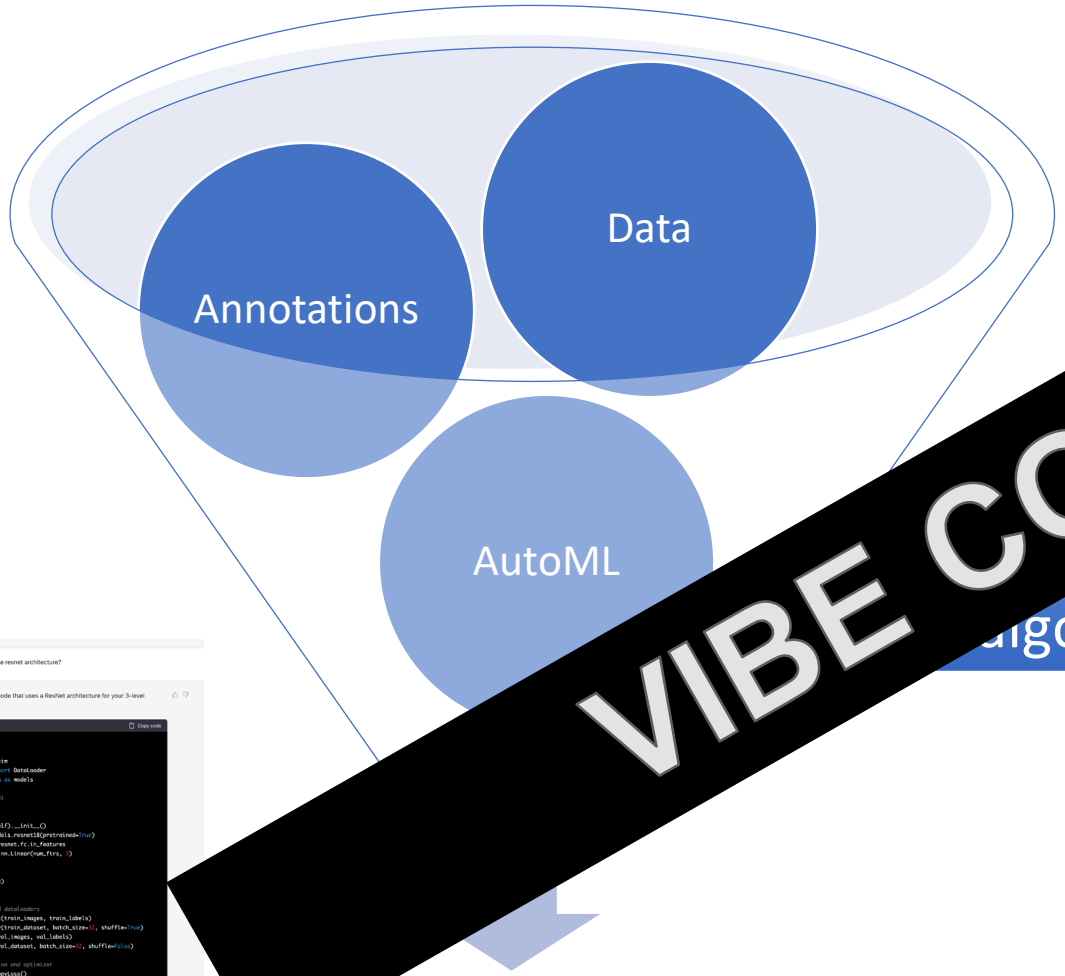
University of Colorado
Anschutz Medical Campus

Applications of Generative AI in Medical Imaging



Kazeminia et al, Artificial Intelligence In Medicine 2020

It is becoming *really easy* to create an AI algorithm today!



VIBE CODING?

AI Algorithm

Title

Performance of an AI al… …thy of prematurity

Abstr…

…otentially blinding eye disorder that affects …is and treatment of ROP is critical for preserving vision … In recent years, artificial intelligence (AI) algorithms have shown …omated diagnosis of ROP. In this study, we evaluated the performance …orithm for the diagnosis of ROP using a dataset of fundus images from …ature infants.

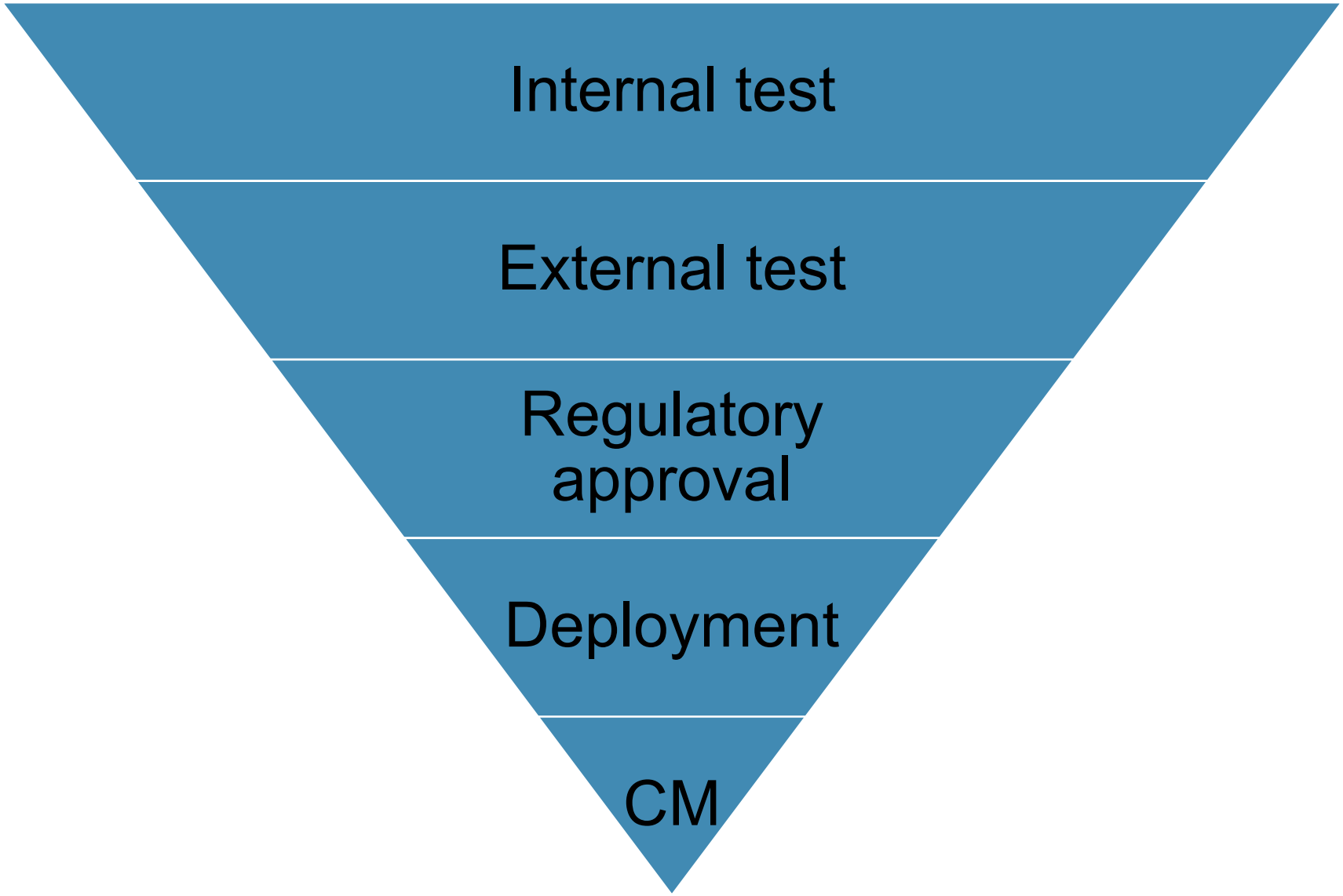
We trained a convolutional neural network (CNN) on a dataset of 5,000 fundus images from premature infants with and without ROP. We evaluated the performance of the CNN on a separate dataset of 1,000 fundus images from premature infants, including 500 images with ROP and 500 images without ROP. We measured the sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), and accuracy of the CNN for the diagnosis of ROP.

Our results showed that the CNN achieved a sensitivity of 95.2%, a specificity of 93.8%, a PPV of 92.1%, an NPV of 96.2%, and an accuracy of 94.5% for the diagnosis of ROP. The area under the receiver operating characteristic curve (AUC-ROC) was 0.96, indicating excellent diagnostic accuracy.

Our study demonstrates that an AI algorithm based on a CNN can achieve high diagnostic accuracy for the diagnosis of ROP. The use of AI algorithms for the automated diagnosis of ROP has the potential to improve the efficiency and accuracy of ROP screening programs, particularly in resource-limited settings where access to ophthalmologists and specialized equipment may be limited.

University of Colorado Anschutz Medical Campus

AI Algorithm Development Funnel



Challenges in AI model development/deployment

Generalizability– models are brittle and do not generalize across scanners, populations, disease presentation

Shortcut learning

Model predictions may not be repeatable!

Calibration- commonly used approaches for binary models can lead to poorly calibrated models

Silent failures – models may fail without indication (”confidently wrong”)

Overfitting – reported model performance can be over-optimistic

Explainability/interpretability

Models can be biased (in hard to detect ways)

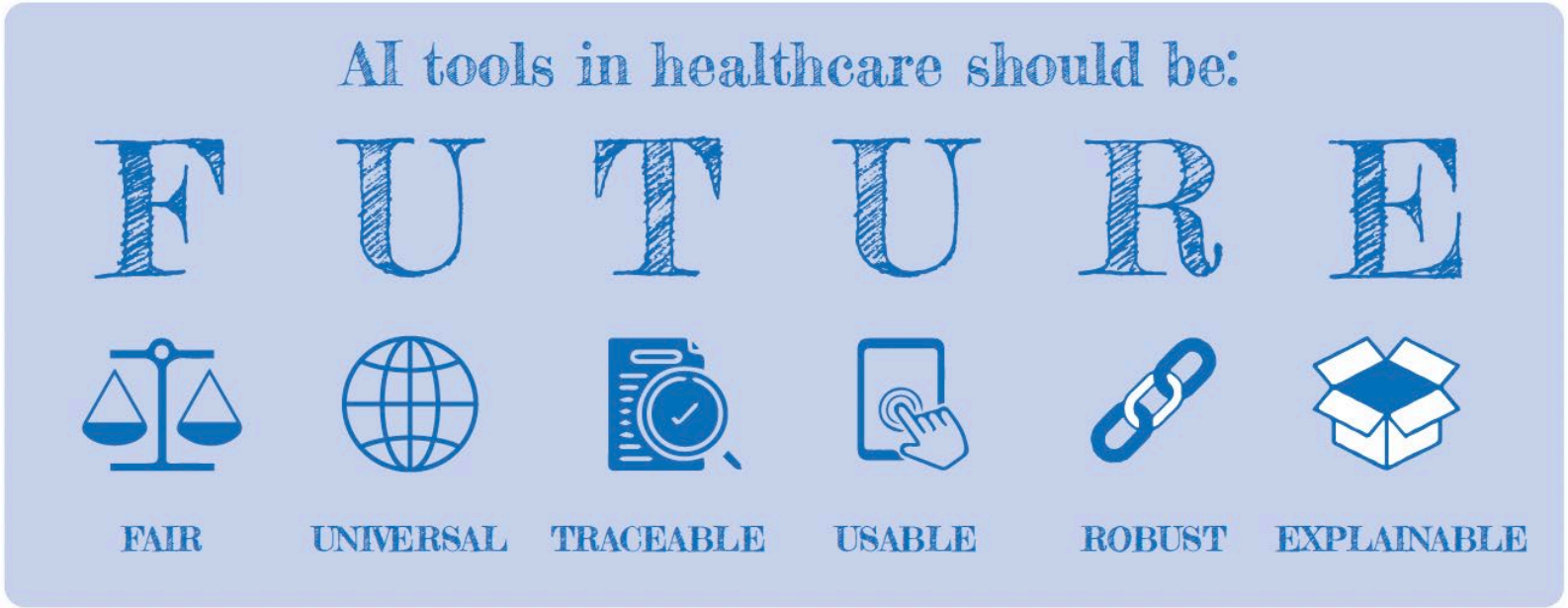
Incorrect metrics

Incorrect ground truth

Inadequate Testing and Validation in the field

FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare

Karim Lekadir,[1,2] Alejandro F Frangi,[3,4] Antonio R Porras,[5] Ben Glocker,[6] Celia Cintas,[7] Curtis P Langlotz,[8] Eva Weicken,[9] Folkert W Asselbergs,[10,11] Fred Prior,[12] Gary S Collins,[13] Georgios Kaissis,[14] Gianna Tsakou,[15] Irène Buvat,[16] Jayashree Kalpathy-Cramer,[17] John Mongan,[18] Julia A Schnabel,[19] Kaisar Kushibar,[1] Katrine Riklund,[20] Kostas Marias,[21] Lameck M Amugongo,[22] Lauren A Fromont,[23] Lena Maier-Hein,[24] Leonor Cerdá-Alberich,[25] Luis Martí-Bonmatí,[26] M Jorge Cardoso,[27] Maciej Bobowicz,[28] Mahsa Shabani,[29] Manolis Tsiknakis,[21] Maria A Zuluaga,[30] Marie-Christine Fritzsche,[31] Marina Camacho,[1] Marius George Linguraru,[32] Markus Wenzel,[9] Marleen De Bruijne,[33] Martin G Tolsgaard,[34] Melanie Goisauf,[35] Mónica Cano Abadía,[35] Nikolaos Papanikolaou,[36] Noussair Lazrak,[1] Oriol Pujol,[1] Richard Osuala,[1] Sandy Napel,[37] Sara Colantonio,[38] Smriti Joshi,[1] Stefan Klein,[33] Susanna Aussó,[39] Wendy
on behalf of the FUTURE-A



University of C
Anschutz Med

Fig 2 | Organisation of the FUTURE-AI framework for trustworthy artificial intelligence (AI) according to six guiding principles—fairness, universality, traceability, usability, robustness, and explainability

FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare

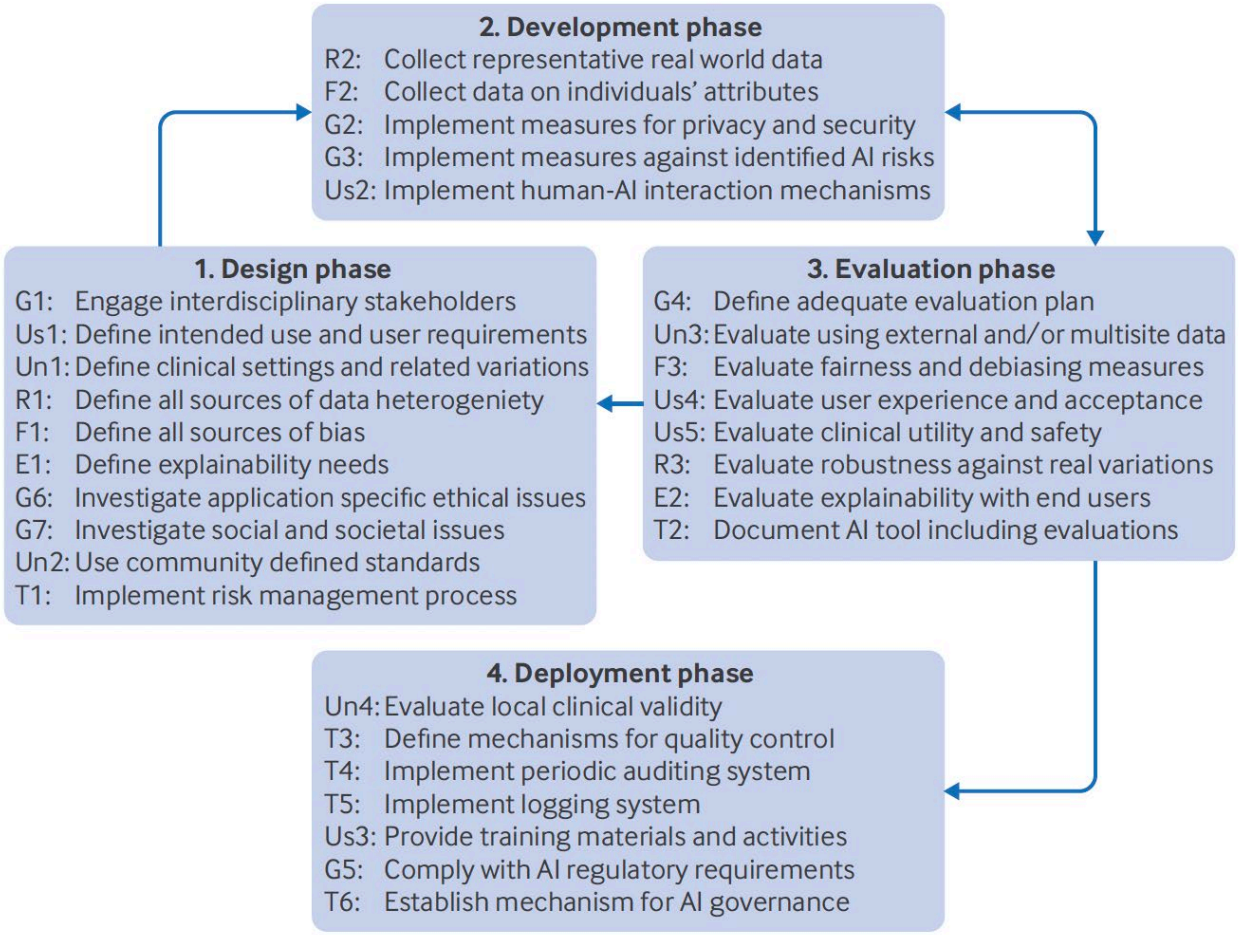


Fig 3 | Embedding the FUTURE-AI best practices into an agile process throughout the artificial intelligence (AI) lifecycle. E=explainability; F=fairness; G=general; R=robustness; T=traceability; Un=universality; Us=usability

University of Colorado
Anschutz Medical Campus

https://future-ai.eu/

Checklist before model deployment

- ✓ What is repeatability (test-retest performance) of the model?
- ✓ What is the reproducibility/ portability performance?
- ✓ Does the system have an “out of distribution” detector?
- ✓ How well is the model calibrated?
- ✓ How often does the model make grave errors? Is the model confidently wrong?
- ✓ Is image quality assessed?
- ✓ Does the image contain enough information to make a prediction?
- ✓ Can the model be adapted locally?
- ✓ What is the continuous monitoring plan?

FDA Perspective is worth considering

Preparing for the Unknowns of Large Language Models and Generative AI

Applications of generative AI, such as large language models (LLMs), **present a unique challenge** because of the potential for unforeseen, emergent consequences; the FDA is yet to authorize an LLM.

Even “AI scribes” meant to summarize medical notes **can hallucinate or include diagnoses not discussed in the visit.**

Because we cannot unduly burden individual clinicians with such oversight, there is a need for specialized **tools that enable better assessment of LLMs** in the contexts and settings in which they will be used.

JAMA | Special Communication | AI IN MEDICINE

FDA Perspective on the Regulation of Artificial Intelligence in Health Care and Biomedicine

Haider J. Warraich, MD; Troy Tazbaz, BS; Robert M. Califf, MD

FDA Perspective is worth considering

The Central Importance of AI Life Cycle Management

"Given the capacity for "unlocked" models to evolve and AI's sensitivity to contextual changes, it is becoming increasingly evident that AI performance **should be monitored in the environment in which it is being used.**"

"health systems will need to provide an **information ecosystem** much like that monitoring a patient in the intensive care unit.

Finding the Balance Between Big Tech, Start-Ups, and Academia

The Tension Between Using AI to Optimize Financial Returns vs Improving Health Outcomes

"the relationship between optimizing finances and improving health outcomes for patients and communities is complex and at times at odds"

The Future of AI: Unlocking Possibilities


University of Colorado Anschutz Medical Campus
THANK YOU

“Brittleness” of machine learning models

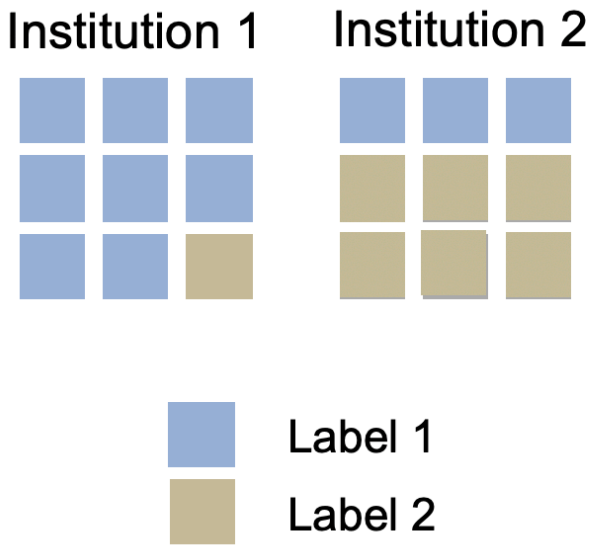
Deep learning models do not generalize well

Only 6% of published AI studies have external validation (Kim et al., KJR, 2019)

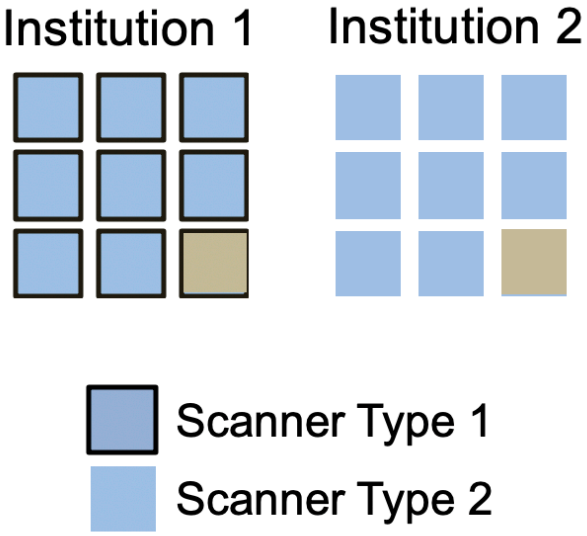
Data heterogeneity can lead to poor model performance on external datasets.

Distribution differences

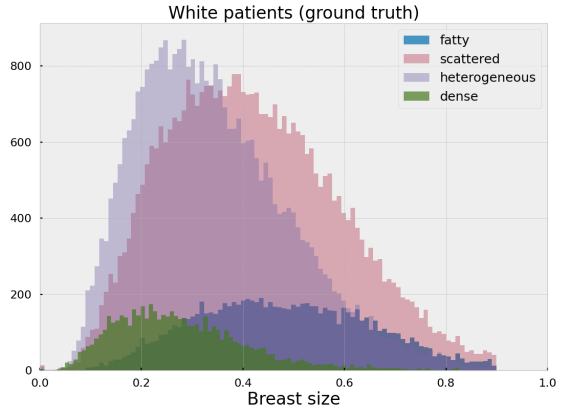
A

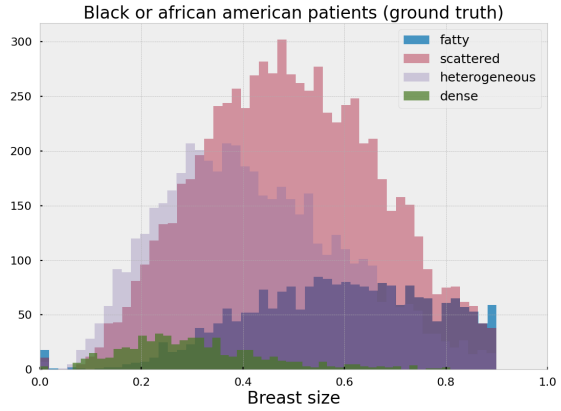


B



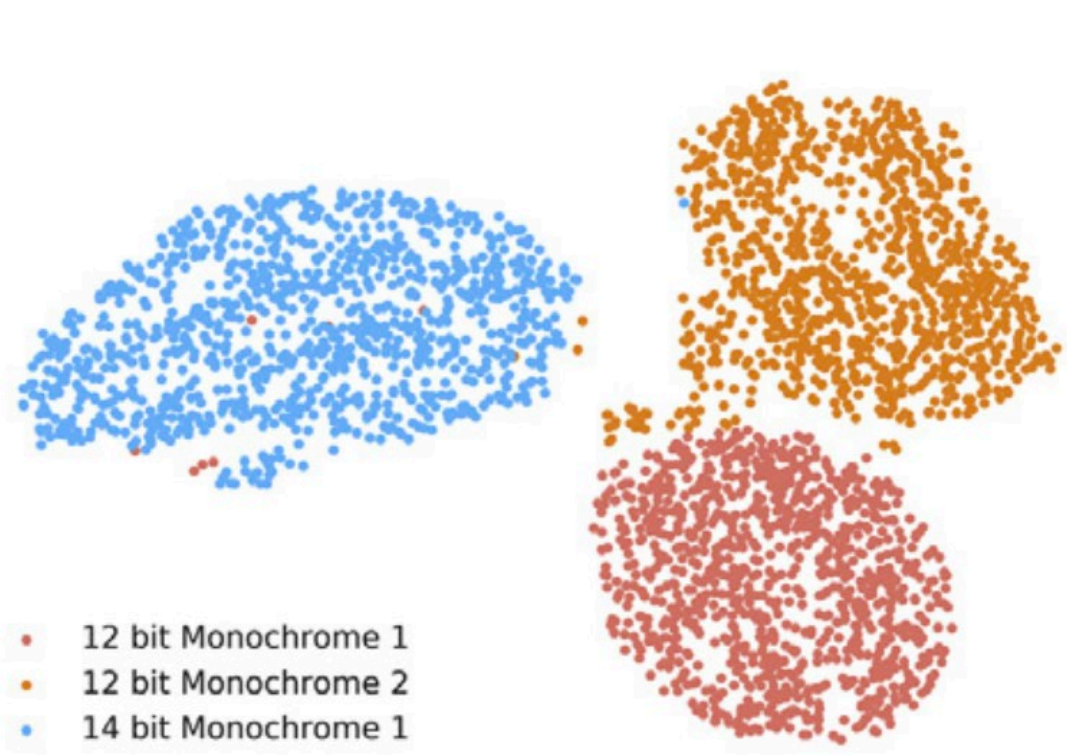
C

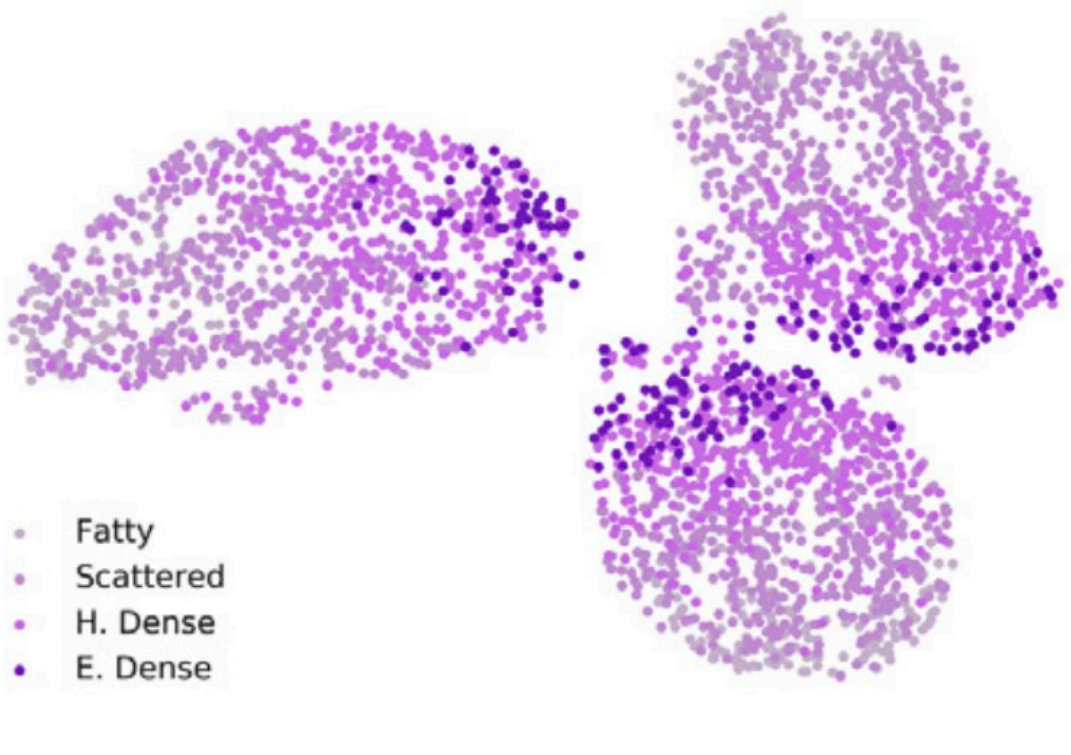




University of Colorado
Anschutz Medical Campus

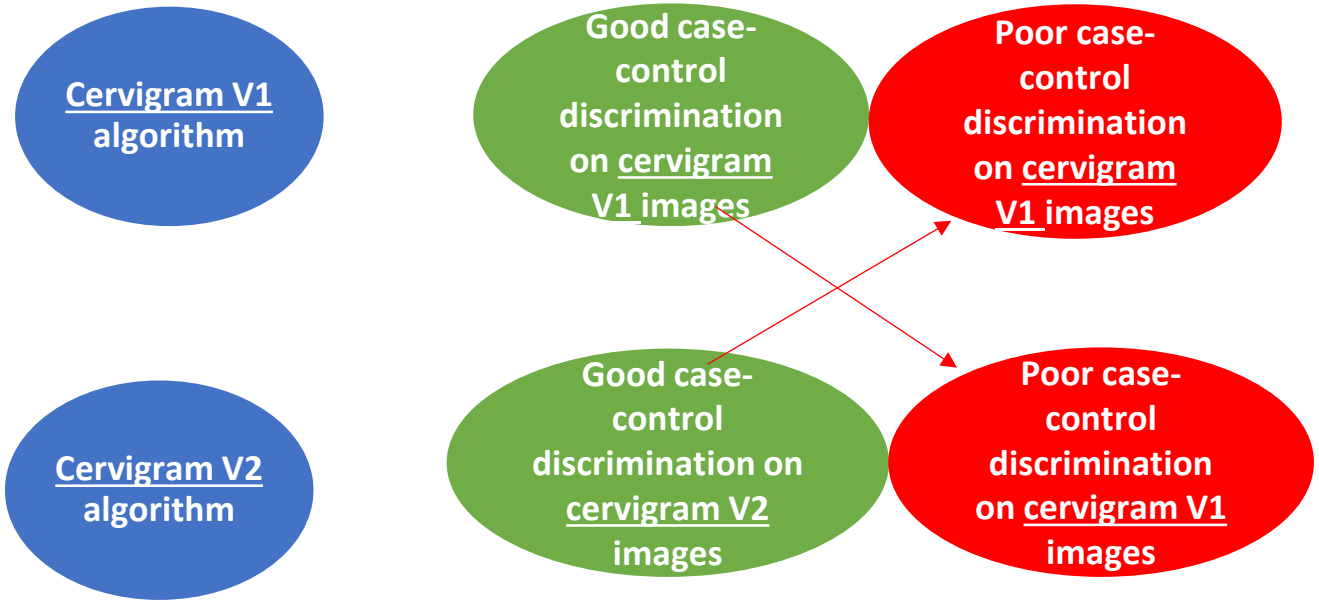
Impact of acquisition heterogeneity persists through the network

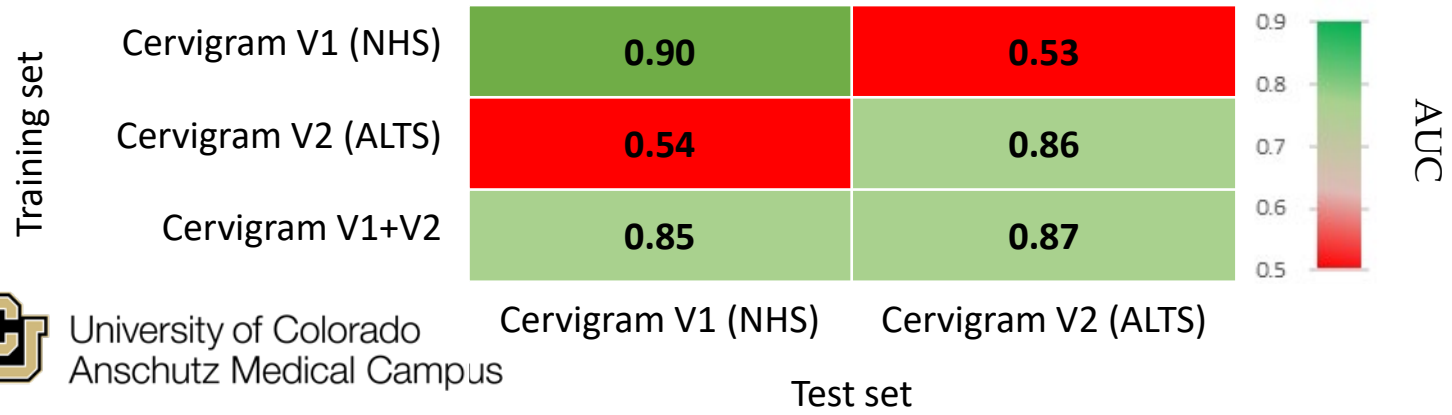




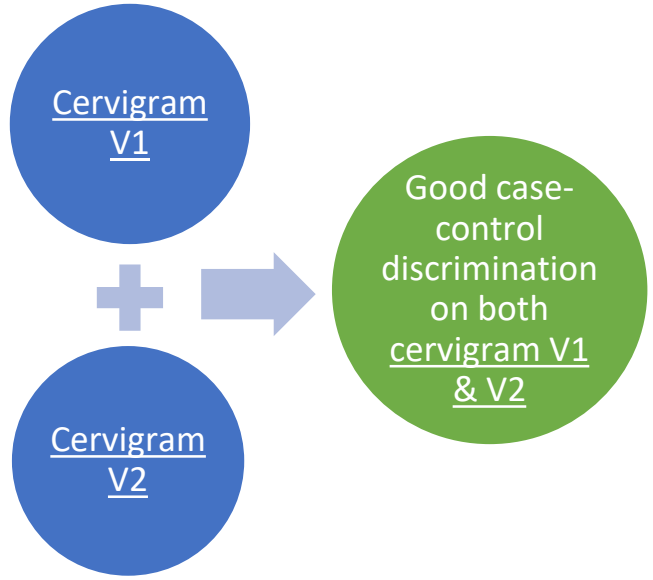
UMap of penultimate layer features shows distinct clusters by scanner type

“Portability challenges” in cervical cancer





Combined model

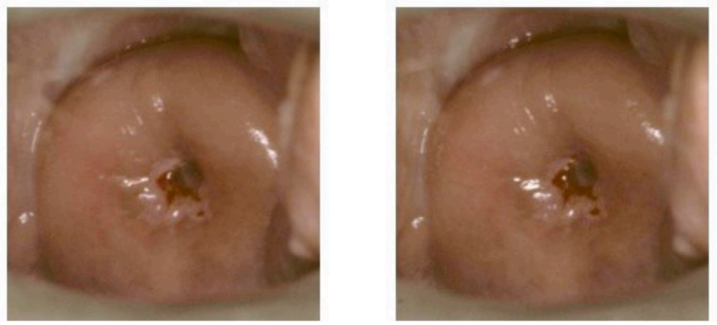


University of Colorado Anschutz Medical Campus

Slide: KananDesai/Mark Schiffman

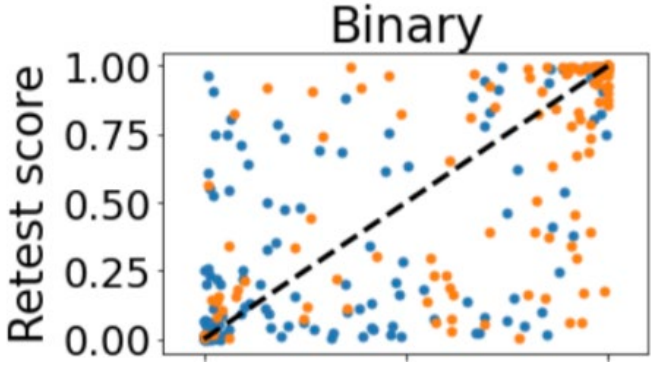
DL Model predictions are not repeatable!

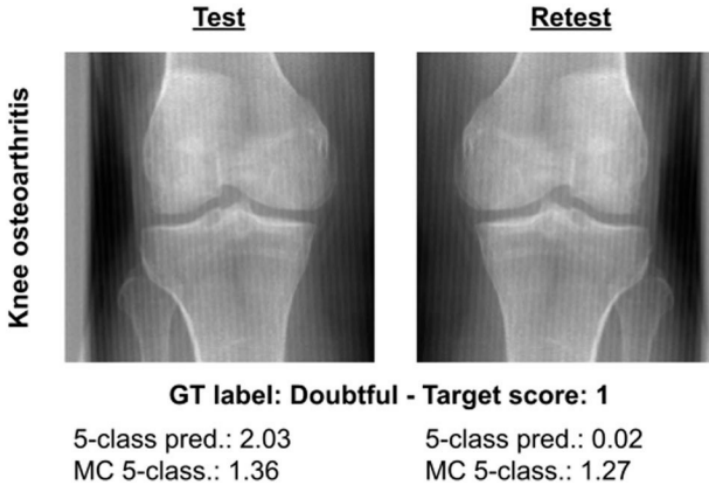
A replicate set of images yield different results (lack of repeatability)



(a) Model prediction: 0.01 (Normal) (b) Model prediction: 0.98 (Pre-cancer)

Fig. 1: Illustration of repeatability issues from deep learning models on different images of a cervix with precancerous lesions from the same patient taken the same day. A binary model without dropout layers generated the following outputs. (a) the binary model predicts a normal cervix (severity score: 0.01). (b) the binary model predicts pre-cancer (severity score: 0.98).





Little published literature on model repeatability/reproducibility

Many models are not repeatable when tested!.

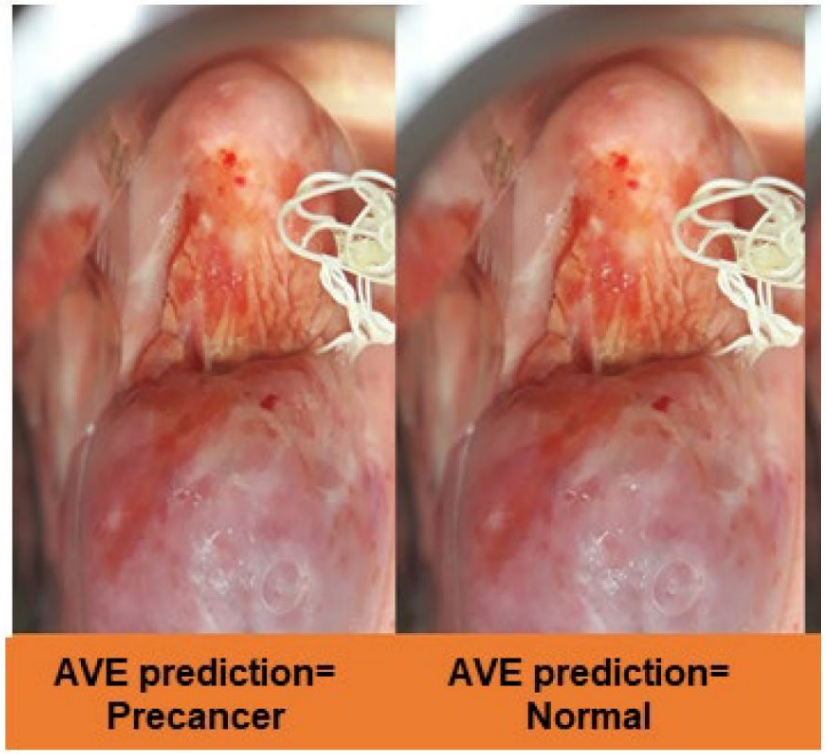
University of Colorado
Anschutz Medical Campus

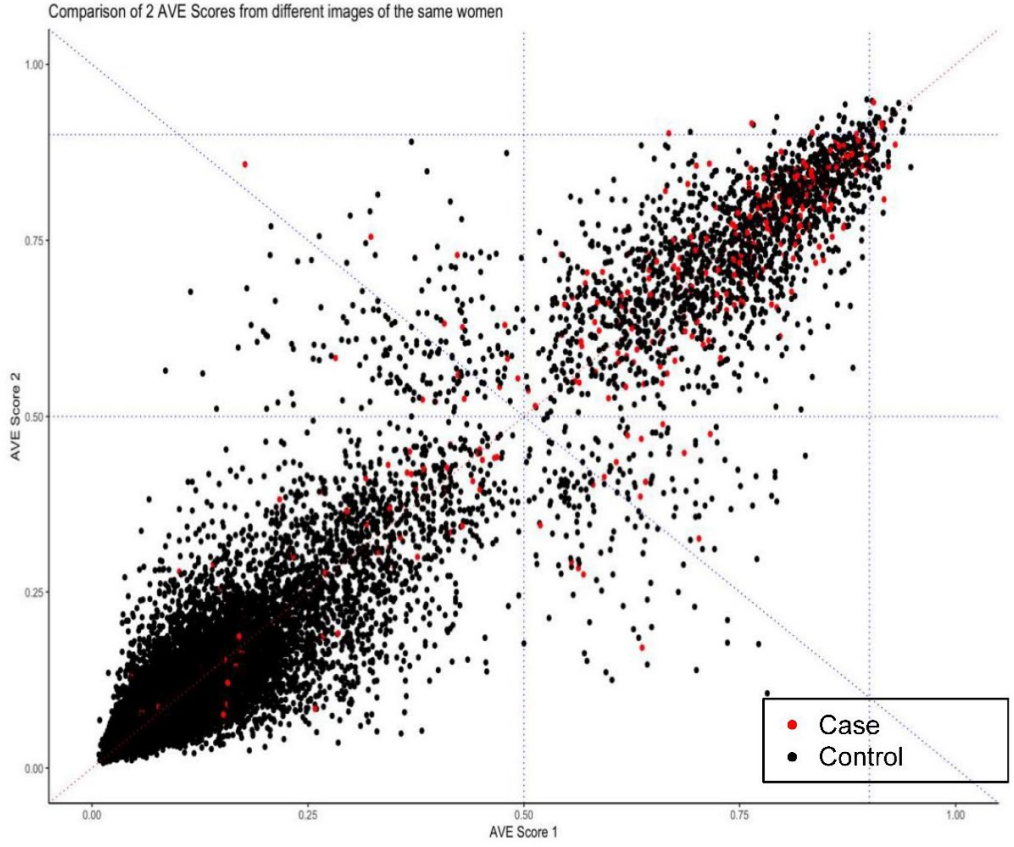
Desai K et al. IJC 2021;

Lemay et al, NPJ Digital Med, 2022

Problem 2: Test-retest repeatability can be an issue

Challenge: *A replicate set of images from a woman during same examination with same device, yielded different results* ***(lack of repeatability)***





This issue was seen soon after the initial publication

Desai K et al. IJC 2021;

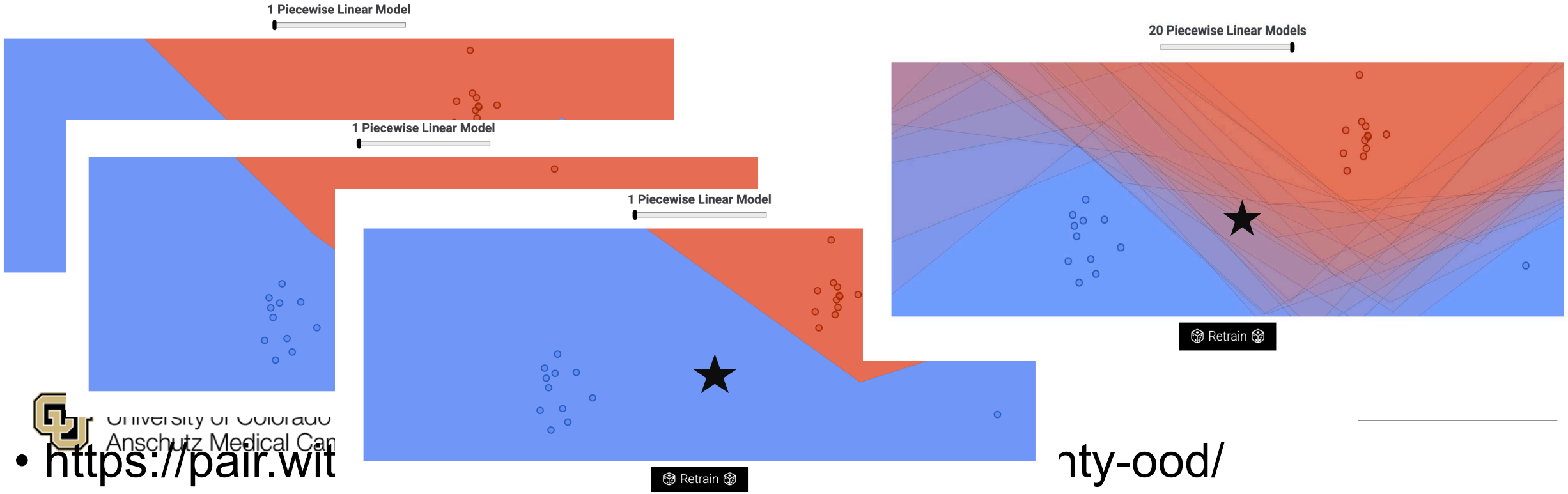
University of Colorado
Anschutz Medical Campus

PAIR EXPLORABLES

From Confidently Incorrect Models to Humble Ensembles

Combining Models Reduces Overconfidence

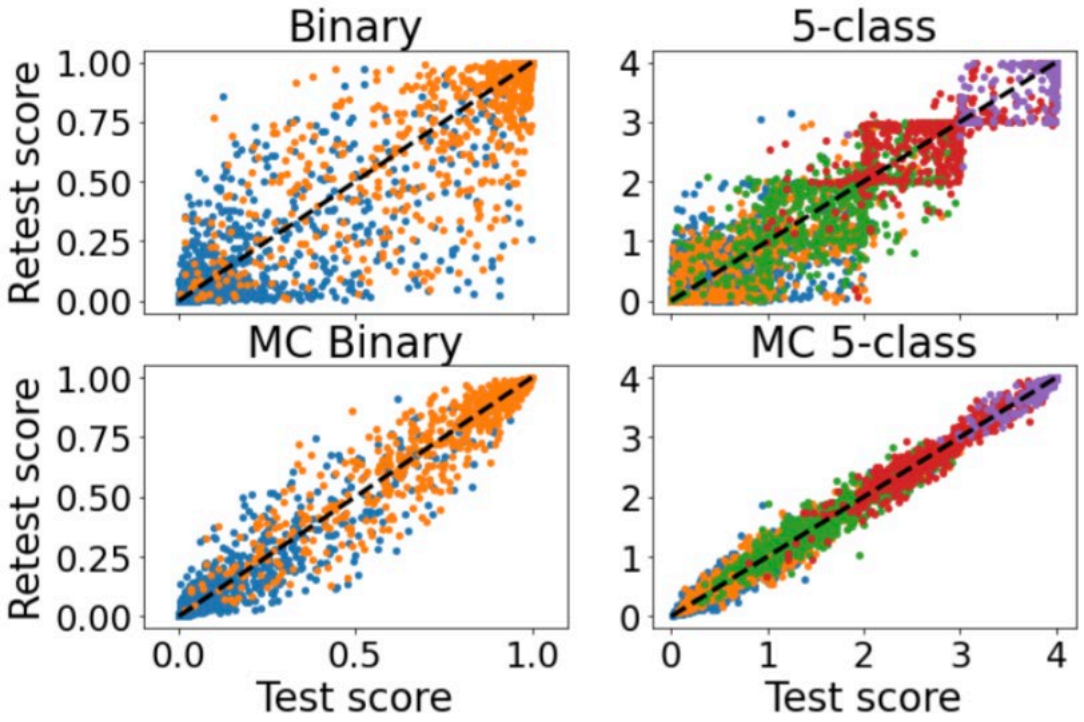
By averaging the output of multiple models, a technique known as **ensembling**, we can create a model



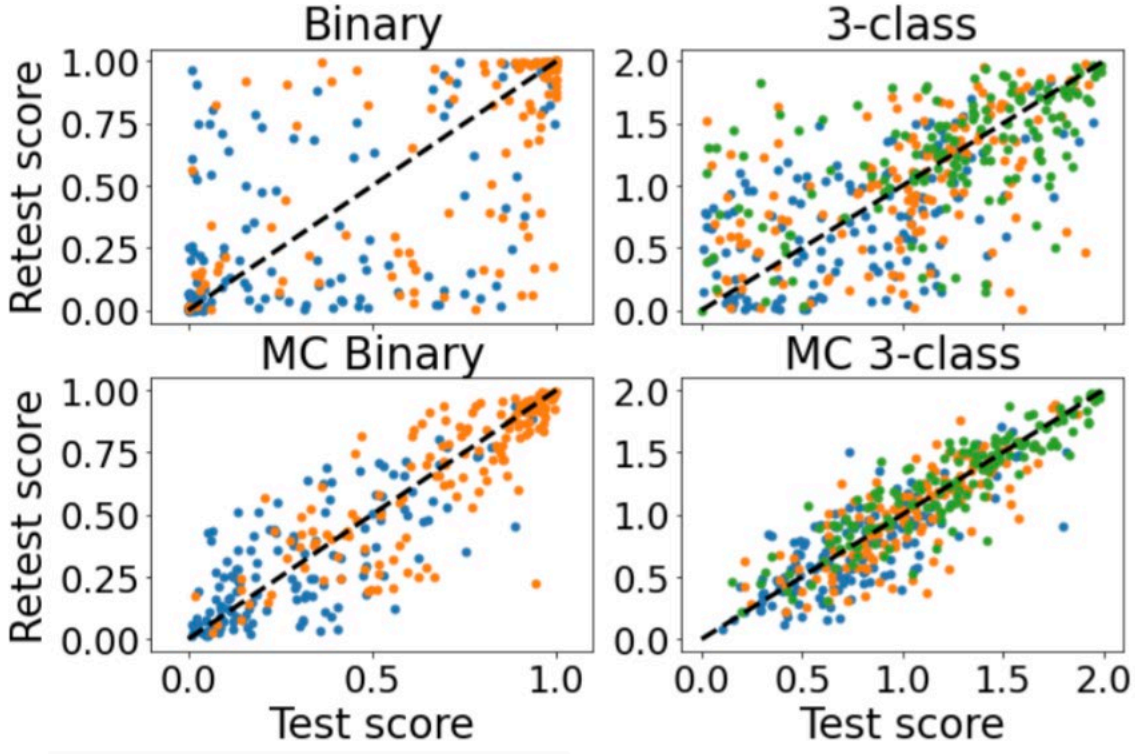
University of Colorado Anschutz Medical Can

- https://pair.wit nty-ood/

Solution 2: Monte Carlo approaches may improve repeatability



Knee osteoarthritis (xray)



Cervical cancer (photos)

Lemay et al, NPJ Digital Med, 2022

Metrics, ground truth

Accuracy is often used in ML publications but not a useful metrics especially in low prevalence settings

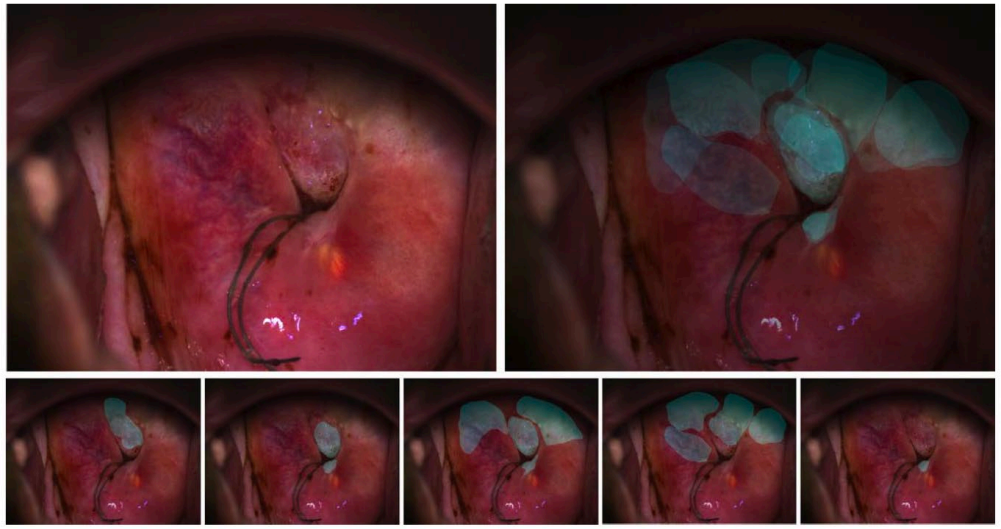
AUROC is often the metric of choice but does not capture the distribution of scores well

Technical metrics don’t often translate to clinical utility

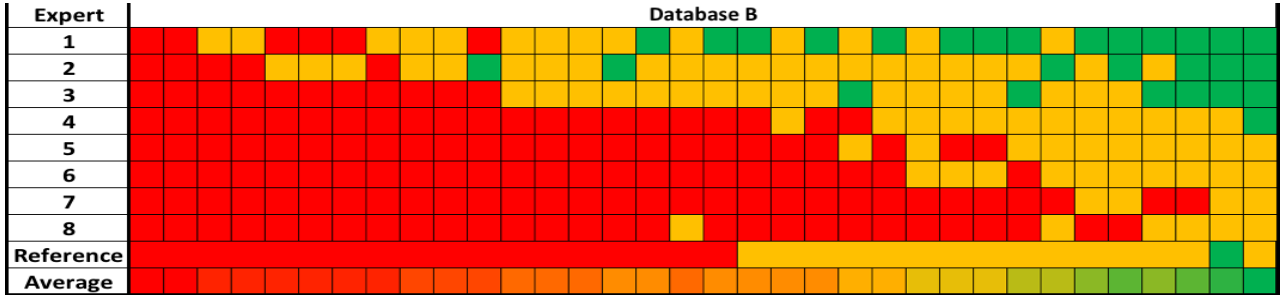
Ground truth

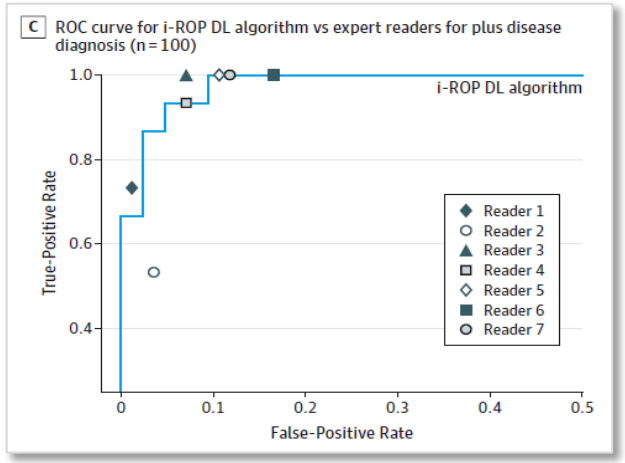
Ground truth can be murky

Human derived ground truth case be highly variable (and wrong)



Lycke et al, Journal of Lower Genital Tract Disease 2024



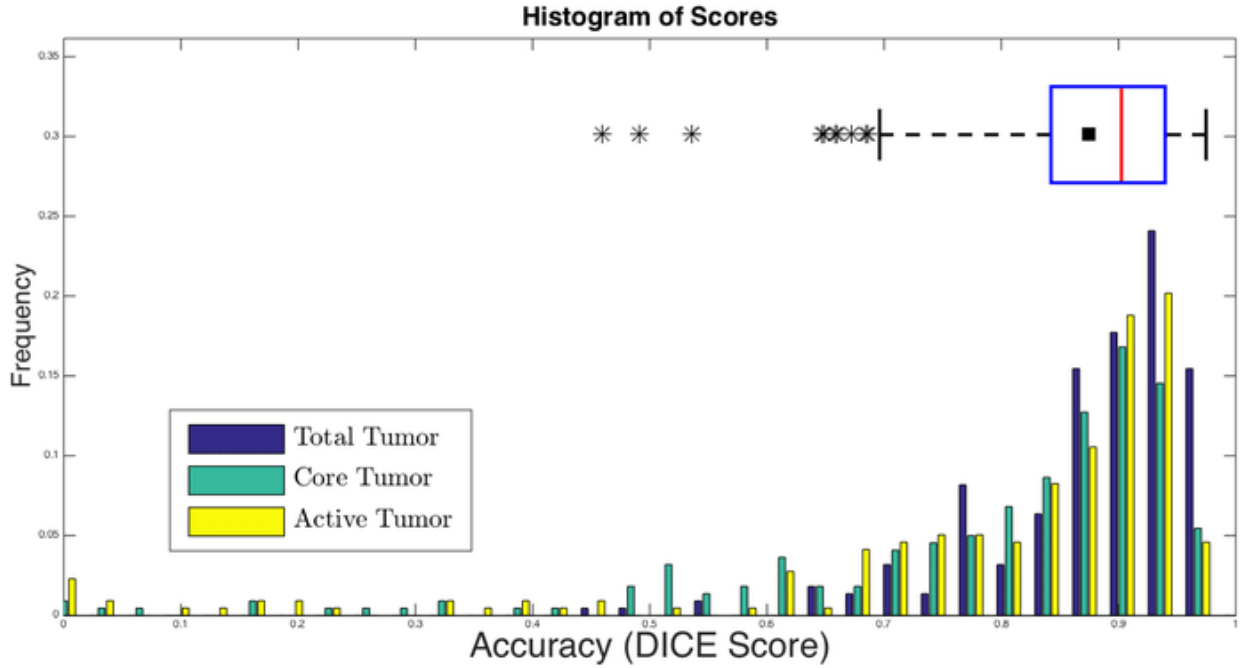


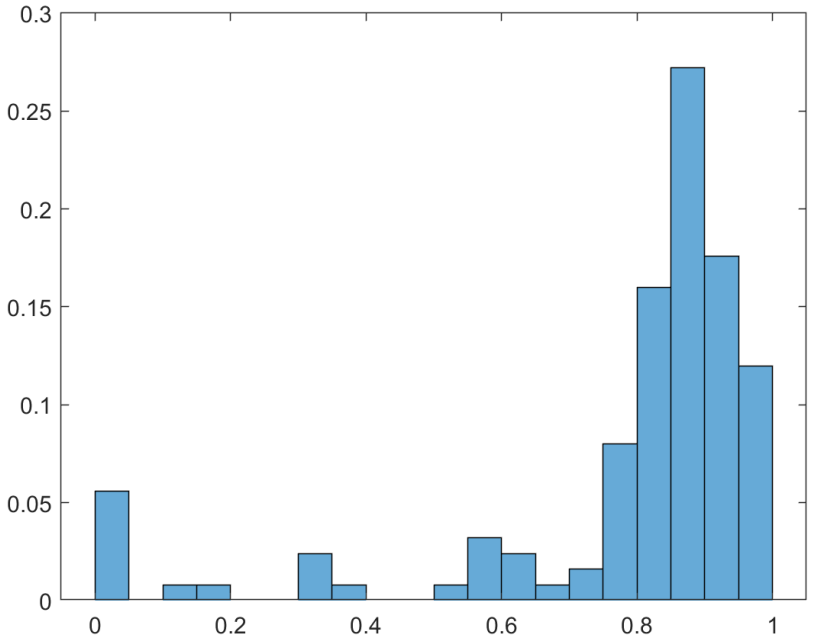
Campbell et al, Ophthalmology 2016;123:2338-44.

Models don’t inherently say “I don’t know” and may fail silently

Deep learning approaches (typically) do not provide measures of [segmentation] uncertainty

Example histogram of dice scores for segmentation shows long tail of low-quality segmentations





Generate continuous output variables instead of binary values, incorporate uncertainty, calibrate models

npj | Digital Medicine

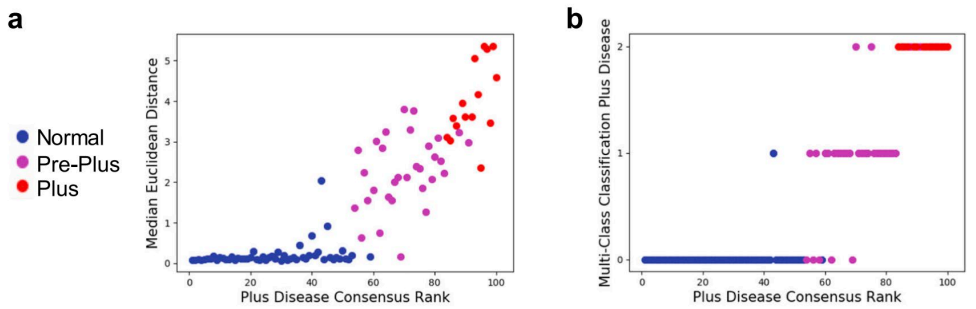
www.nature.com/npjdigitalmed

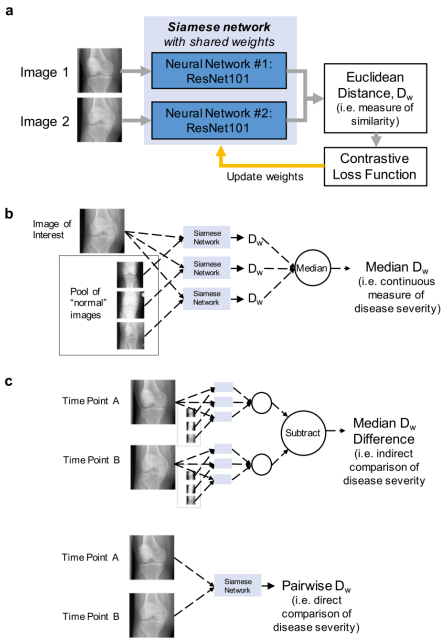
ARTICLE OPEN

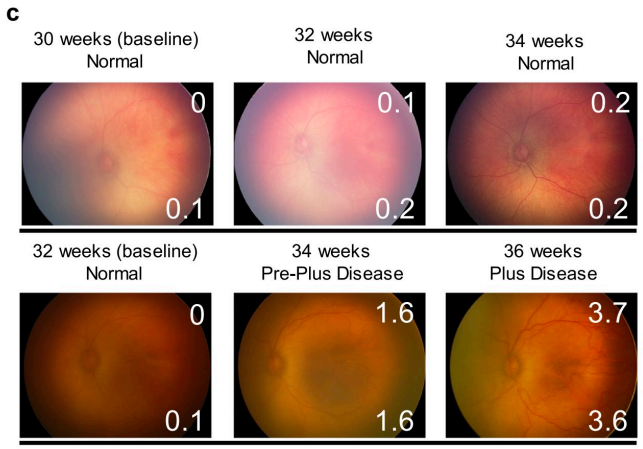
Check for updates

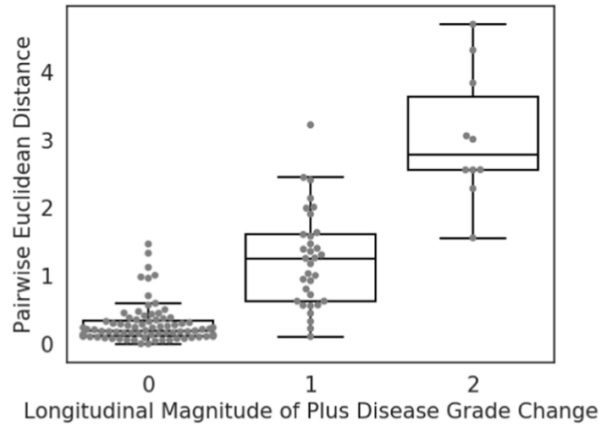
Siamese neural networks for continuous disease severity evaluation and change detection in medical imaging

Matthew D. Li[1], Ken Chang[1], Ben Bearce[1], Connie Y. Chang[2], Ambrose J. Huang[2], J. Peter Campbell[3], James M. Brown[4], Praveer Singh[1], Katharina V. Hoebel[1], Deniz Erdoğmuş[5], Stratis Ioannidis[5], William E. Palmer[2], Michael F. Chiang[3,6] and Jayashree Kalpathy-Cramer[1,7]✉



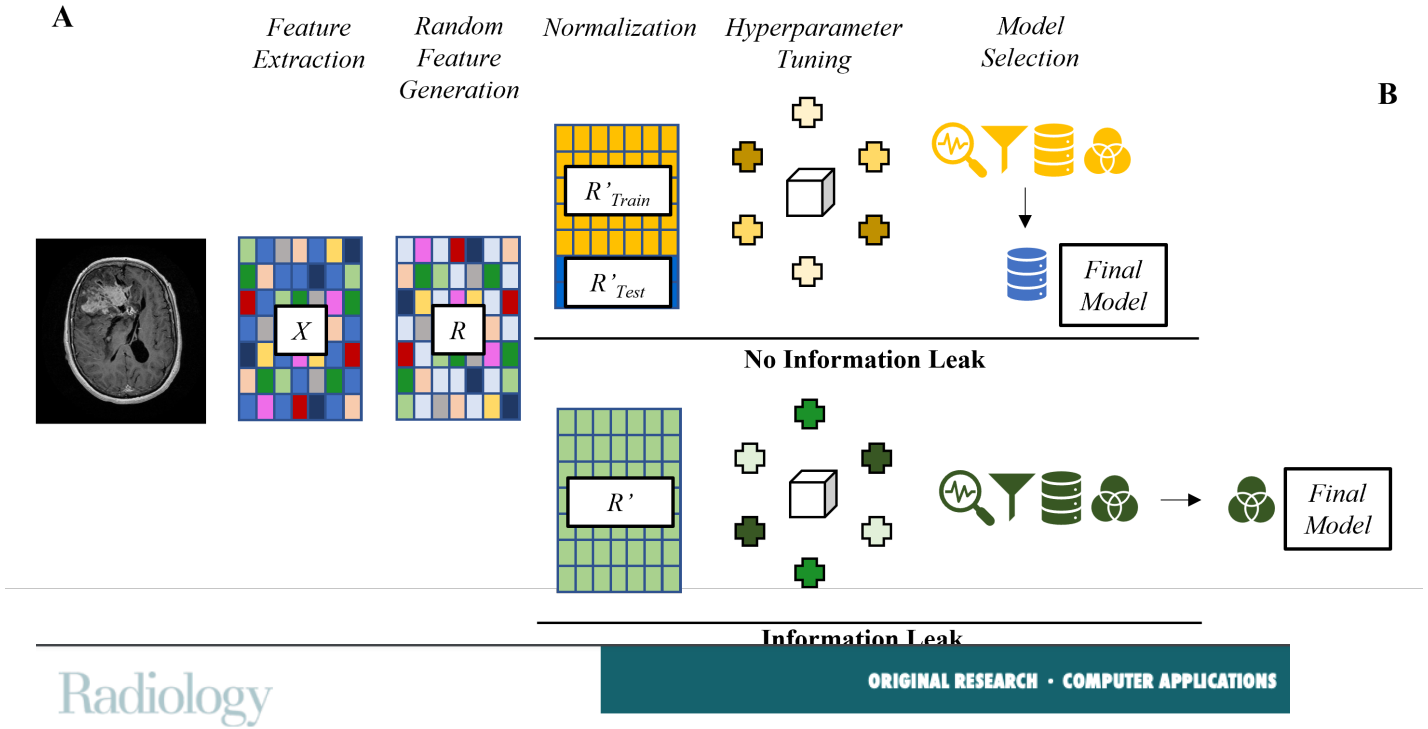


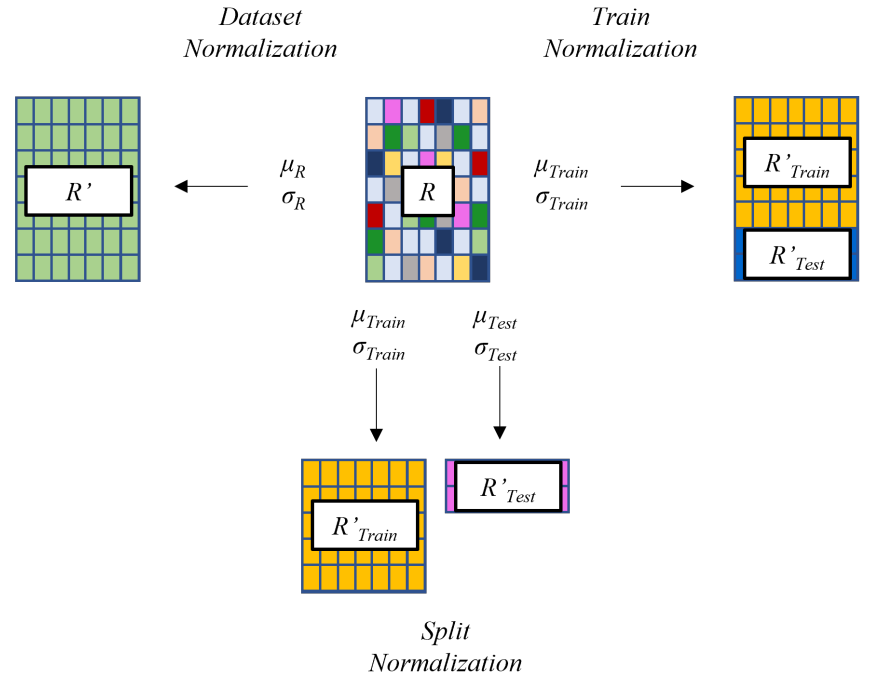




Overfitting is a common problem in the literature

The literature is rife with over-optimistic reported performance, primarily due to a lack of statistical rigor.





Radiology

ORIGINAL RESEARCH · COMPUTER APPLICATIONS

Inconsistent Partitioning and Unproductive Feature Associations Yield Idealized Radiomic Models

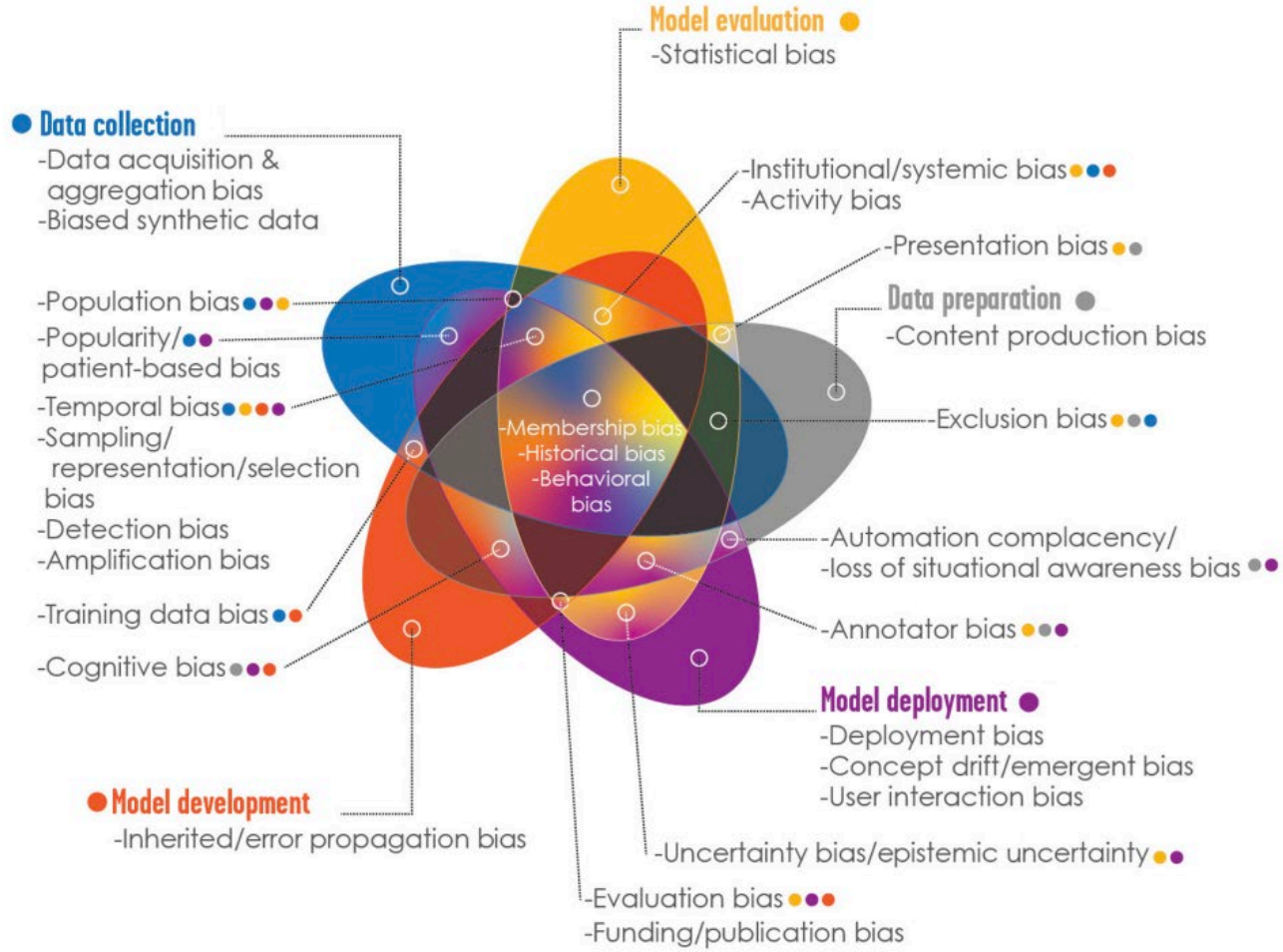
Mishka Gidwani, PhD • Ken Chang, MD, PhD • Jay Biren Patel, BS • Katharina Viktoria Hoebel, MD • Syed Rakin Ahmed, BA • Praveer Singh, PhD • Clifton David Fuller, MD, PhD • Jayashree Kalpathy-Cramer, PhD

Gidwani et al, Radiology, 2022

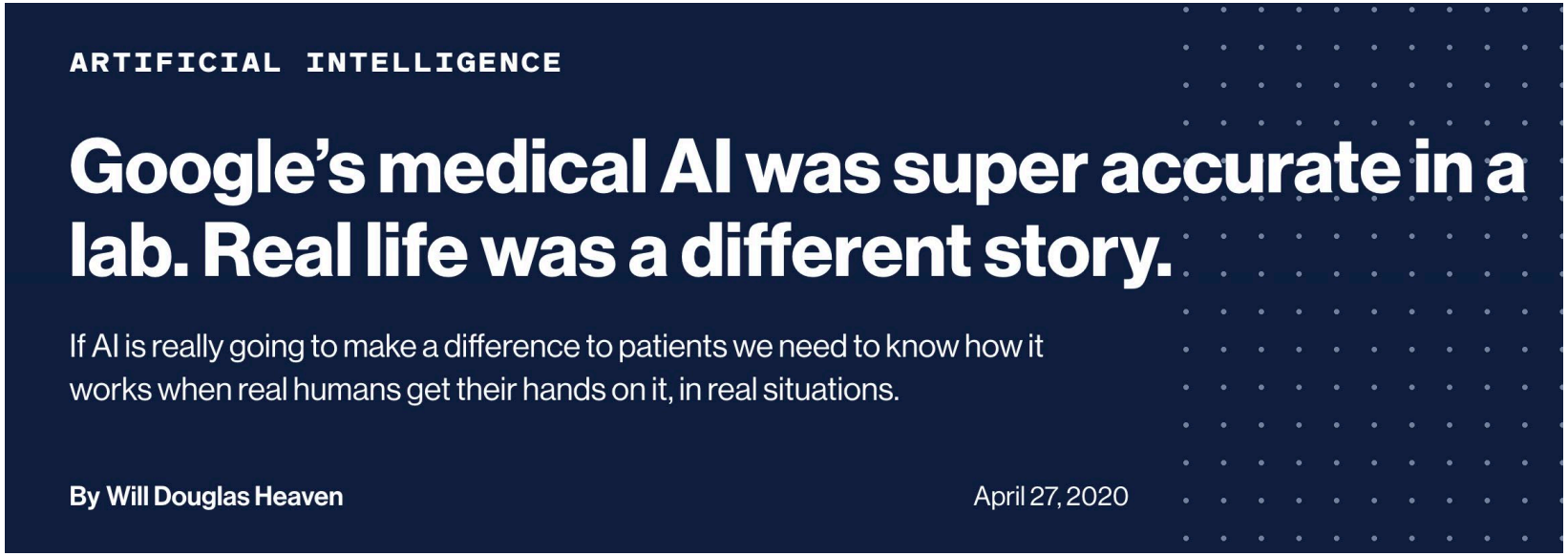
Bias and Fairness

Toward fairness in artificial intelligence for medical image analysis: identification and mitigation of potential biases in the roadmap from data collection to model deployment

Karen Drukker,[a,*] Weijie Chen,[b] Judy Gichoya,[c] Nicholas Gruszauskas,[a] Jayashree Kalpathy-Cramer,[d] Sanmi Koyejo,[e] Kyle Myers,[f] Rui C. Sá,[g,h] Berkman Sahiner,[b] Heather Whitney,[a] Zi Zhang,[i] and Maryellen Giger[a]



Challenges in real life deployment



- Low image quality in practice (AI was trained in high quality)
- Poor internet slowed workflow
- …

Checklist before model deployment

- ✓ What is repeatability (test-retest performance) of the model?
- ✓ What is the reproducibility/ portability performance?
- ✓ Does the system have an “out of distribution” detector?
- ✓ How well is the model calibrated?
- ✓ How often does the model make grave errors? Is the model confidently wrong?
- ✓ Is image quality assessed?
- ✓ Does the image contain enough information to make a prediction?
- ✓ Can the model be adapted locally?
- ✓ What is the continuous monitoring plan?

(Hard) Lessons Learned

- ✓ Continuous monitoring is imperative (how?)
- ✓ External validation (needed? Or hyperoptimze locally?)
- ✓ Continuous scores might be preferable to binary (or ordinal) where the disease lies on a severity spectrum
- ✓ Many commonly used explainability methods have issues – need rigorous evaluation
- ✓ Need to evaluate model repeatability

Conclusion

AI is here to stay (IMO) and impact all aspects of clinical care

There is tremendous potential, but we need to be vigilant before, during and after implementation

Implementing trustworthy AI is team science

Does AI have super-human capabilities?

Article | Published: 19 February 2018

Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning

Ryan Poplin, Avinash V. Varadarajan, Katy Blumer, Yun Liu, Michael V. McConnell, Greg S. Corrado, Lily Peng ✉ & Dale R. Webster

Nature Biomedical Engineering **2**, 158–164 (2018) | Cite this article

22k Accesses | **653** Citations | **2388** Altmetric | Metrics

Predicting sex from retinal fundus photographs using automated deep learning

Edward Korot[1], Nikolas Pontikos[1], Xiaoxuan Liu[1,2,3], Siegfried K. Wagner[1], Livia Faes[1,4], Josef Huemer[1,5], Konstantinos Balaskas[1], Alastair K. Denniston[1,2,3,6], Anthony Khawaja[1]✉ & Pearse A. Keane[1]✉

Predicting risk of breast cancer at one to five years from the mammogram.

Retinal microvasculature dysfunction is associated with Alzheimer's disease and mild cognitive impairment

Check for updates

Jacqueline Chua[1,2,3], Qinglan Hu[1,3], Mengyuan Ke[1,3], Bingyao Tan[1,3,4], Jimmy Hong[1], Xinwen Yao[1,3,4], Saima Hilal[5,6,7], Narayanaswamy Venketasubramanian[5,8], Gerhard Garhöfer[9], Carol Y. Cheung[10], Tien Yin Wong[1,2], Christopher Li-Hsian Chen[5] and Leopold Schmetterer[1,2,3,4,9,11,12*]

University of Colorado Anschutz Medical Campus

ORIGINAL REPORTS | Breast Cancer

Multi-Institutional Validation of a Mammography-Based Breast Cancer Risk Model

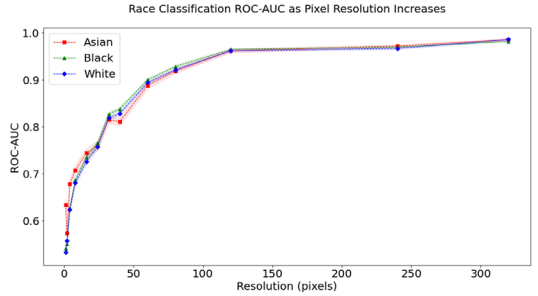
Check for updates

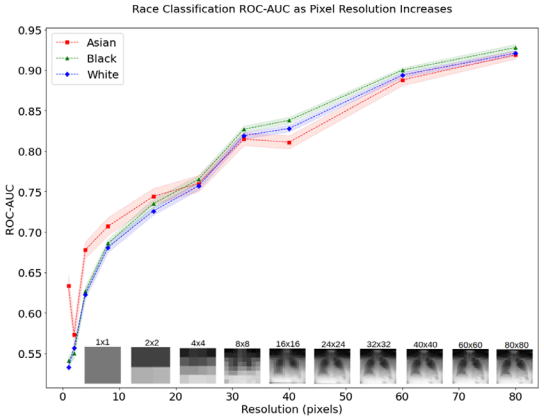
Adam Yala, MEng[1,2] ✉; Peter G. Mikhael, BS[1,2]; Fredrik Strand, MD, PhD[3,4]; Gigin Lin, MD, PhD[5]; Siddharth Satuluru, BS[6]; Thomas Kim, MS[7]; ...

Superhuman + risk of bias + not transparent –> need for continued vigilance?

AI recognition of patient race in medical imaging: a modelling study

Judy Wawira Gichoya, Imon Banerjee, Ananth Reddy Bhimireddy, John L Burns, Leo Anthony Celi, Li-Ching Chen, Ramon Correa, Natalie Dullerud, Marzyeh Ghassemi, Shih-Cheng Huang, Po-Chih Kuo, Matthew P Lungren, Lyle J Palmer, Brandon J Price, Saptarshi Purkayastha, Ayis T Pyrros, Lauren Oakden-Rayner, Chima Okechukwu, Laleh Seyyed-Kalantari, Hari Trivedi, Ryan Wang, Zachary Zaiman, Haoran Zhang





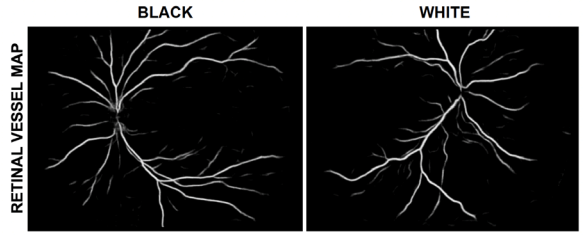
Anschutz Medical Campus

Not Color Blind: AI Predicts Racial Identity from Black and White Retinal Vessel Segmentations

Aaron S Coyner PhD[1,a], Praveer Singh PhD[2,3,a], James M Brown, PhD[4], Susan Ostmo MS[1], RV Paul Chan MD[5], Michael F Chiang MD, MA[6], Jayashree Kalpathy-Cramer PhD[2,3,b], J Peter Campbell MD, MPH[1,b]

Surprisingly:

Grayscale Retinal Vessel Maps Contain Information Associated with Self-Reported Race



	AUC-PR (image level)	AUC-ROC (image level)	AUC-PR (subject level)	AUC-ROC (subject level)
Grayscale RVM	0.938	0.959	0.995	0.995

Grayscale Retinal Vessel Maps Are Associated with Self-Reported Race

Implications for Artificial Intelligence Models

Aaron S. Coyner, PhD

OHSU