Research Clinician Perspective

Ronald M. Summers, M.D., Ph.D., FSAR, FAIMBE, FSPIE

Senior Investigator

Imaging Biomarkers and CAD Laboratory
Radiology and Imaging Sciences
NIH Clinical Center, Bethesda, MD
rms@nih.gov
github.com/rsummers11
www.cc.nih.gov/drd/summers.html

Disclosures and Disclaimer

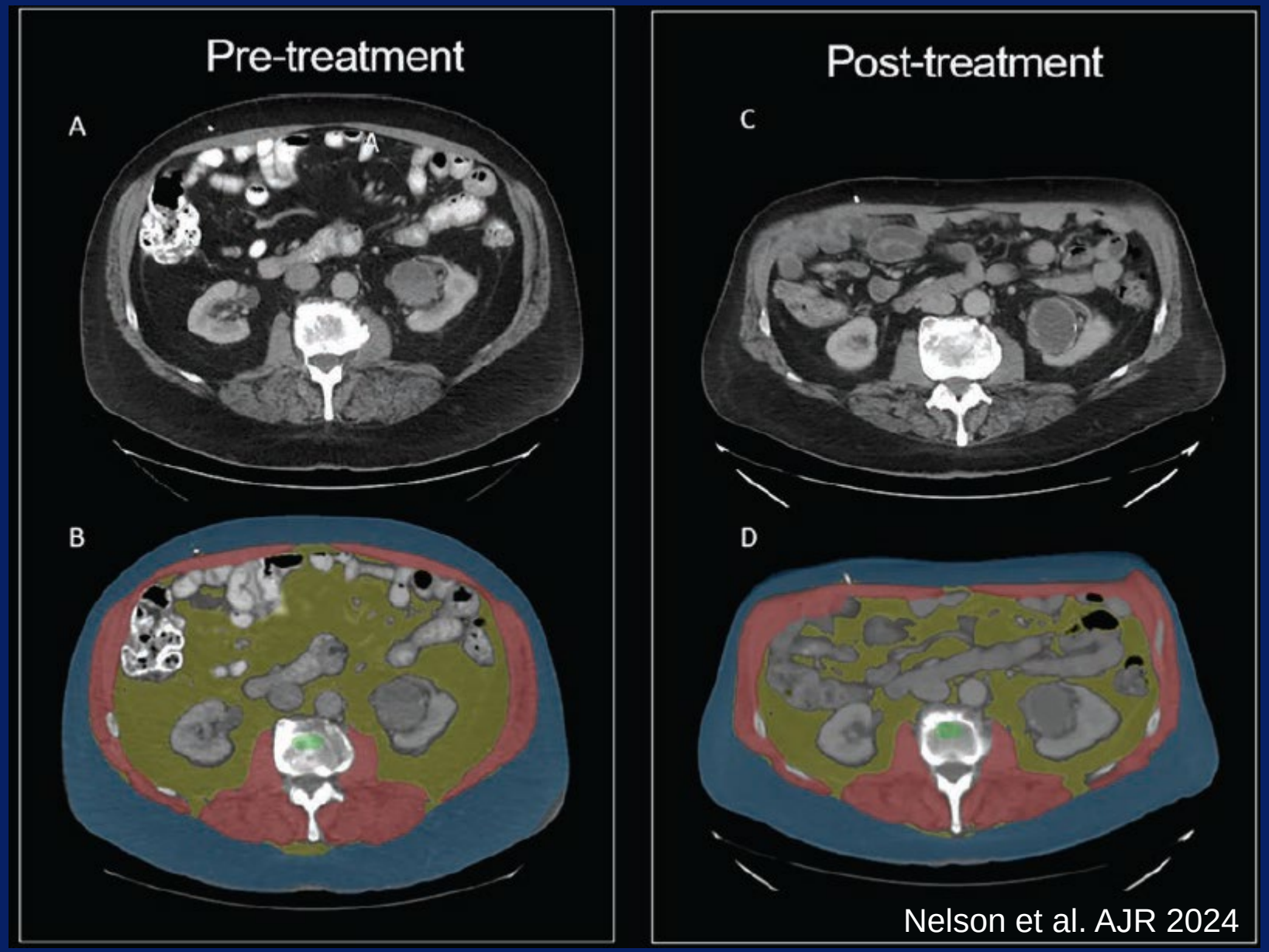
- Patent royalties from iCAD, ScanMed, PingAn, Philips, Translation Holdings, MGB
- Past research support from Ping An

- Views expressed are mine alone and not necessarily those of NIH or DHHS

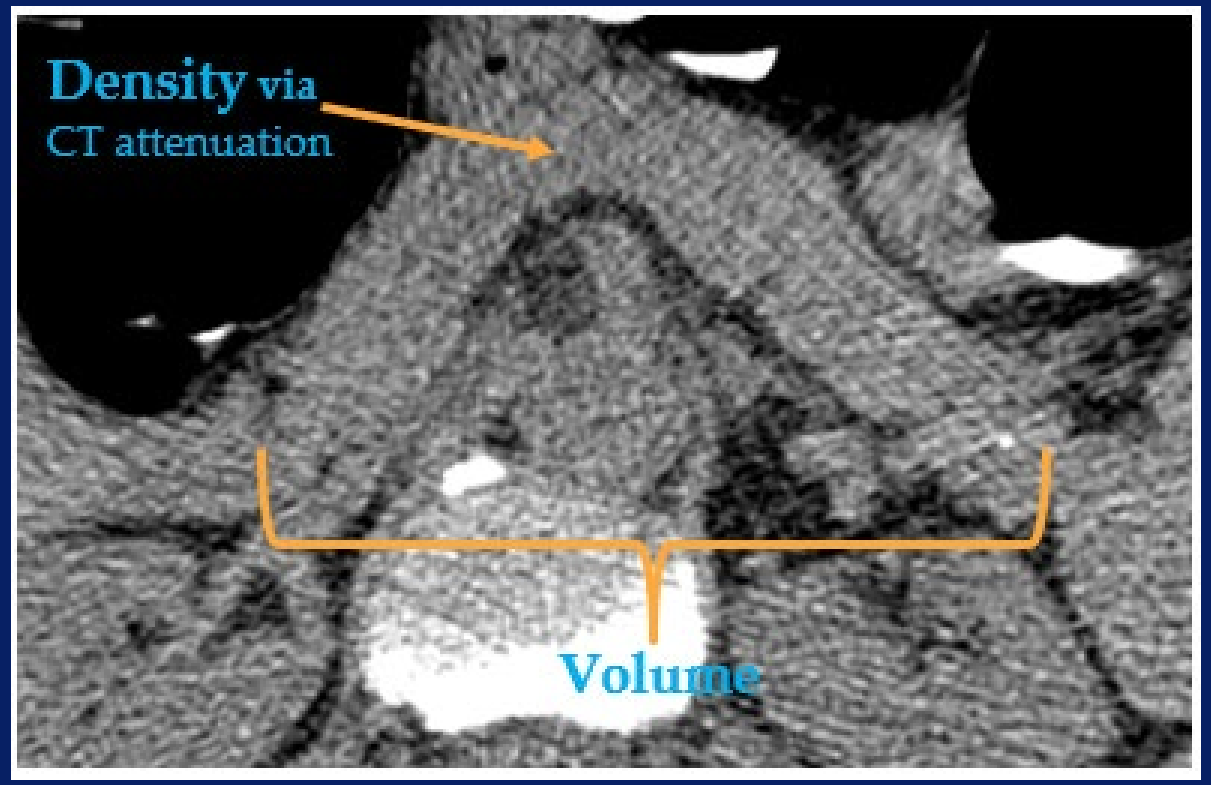
Imaging Phenotyping

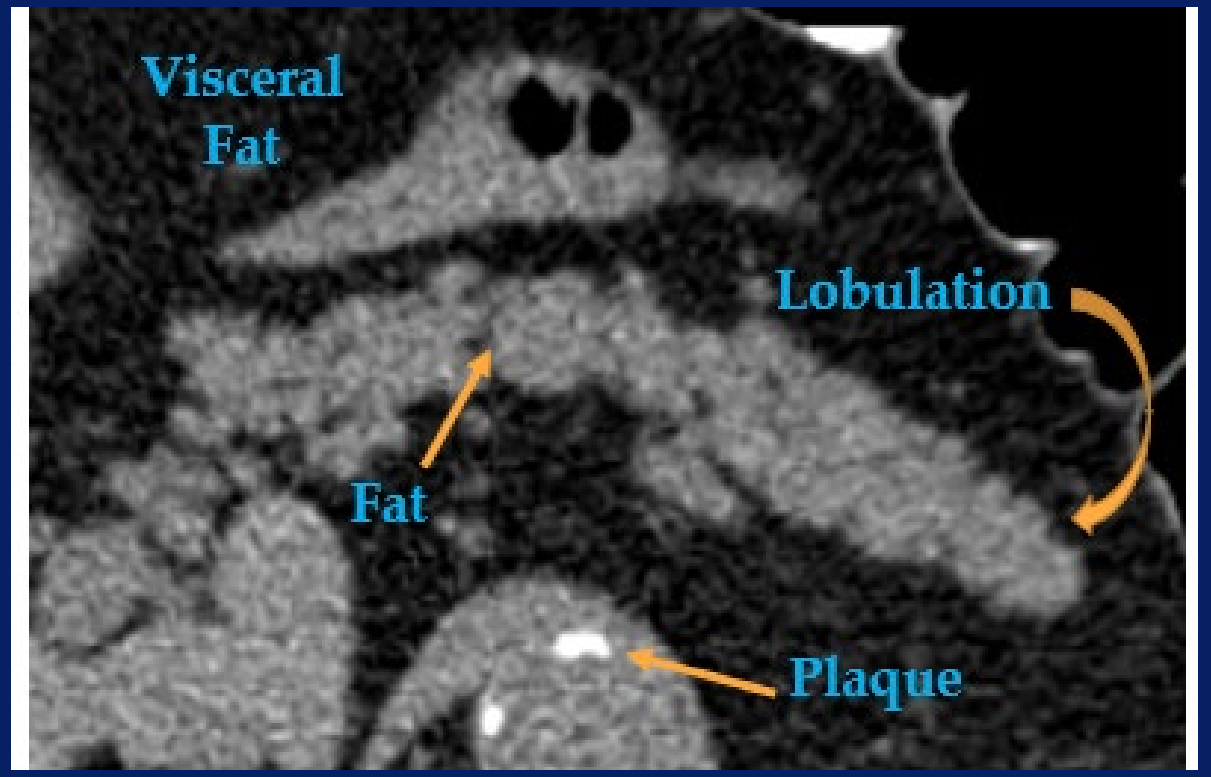
- Assessment of expressed traits as influenced by genetics and the environment
- Missing ingredient due to image complexity
- One of my lab’s major contributions to medicine
- Enables:
  - Assessment of treatment response
  - Data mining of unstructured imaging data
  - Integration with non-imaging data

Body Composition and Semaglutide Therapy



Pancreatic CT and Diabetes





H Tallam et al. RSNA 2021 & Radiology 2022

Example Scenario

- NIDDK multi-institutional trial of acute and chronic pancreatitis
- CT scans vary from one institution to another
- Substantial effort involved in harmonizing the imaging data

Data Scarcity

- Data are the new oil
- AI model development requires large datasets from different populations
- Well-labeled data are helpful but expensive

Radiology

Editorials

Michael W. Vannier
Ronald M. Summers

Sharing Images[1]

Radiology 2003

Imaging Data Archives

- Different data formats for different archives
- Multiplicity of archives
- Wide variety of interfaces
- User agreements vary from permissive to very restrictive
- Metadata vary widely
- Costs vary widely

Image Formats

- Clinical images are voluminous in individual size and numbers
- DICOM is frequently advocated but is storage inefficient
- NIFTI is storage efficient, but metadata can be lost

Clinical Text and Metadata Formats

- Text anonymization is challenging
- Manual review often required
- CSV format for sharing wide variety of data types

What is Needed

- Data harmonization for biomedical images & metadata
- Consolidation of image archives
- Standardized infrastructure
- More publicly available clinical data, particularly those data paid for by government

AI Model Sharing

- Minority of AI publications share the AI model
- Code sharing is most common but of limited utility

Bigolin Lanfredi et al. MedIA 2025

Clinical Text

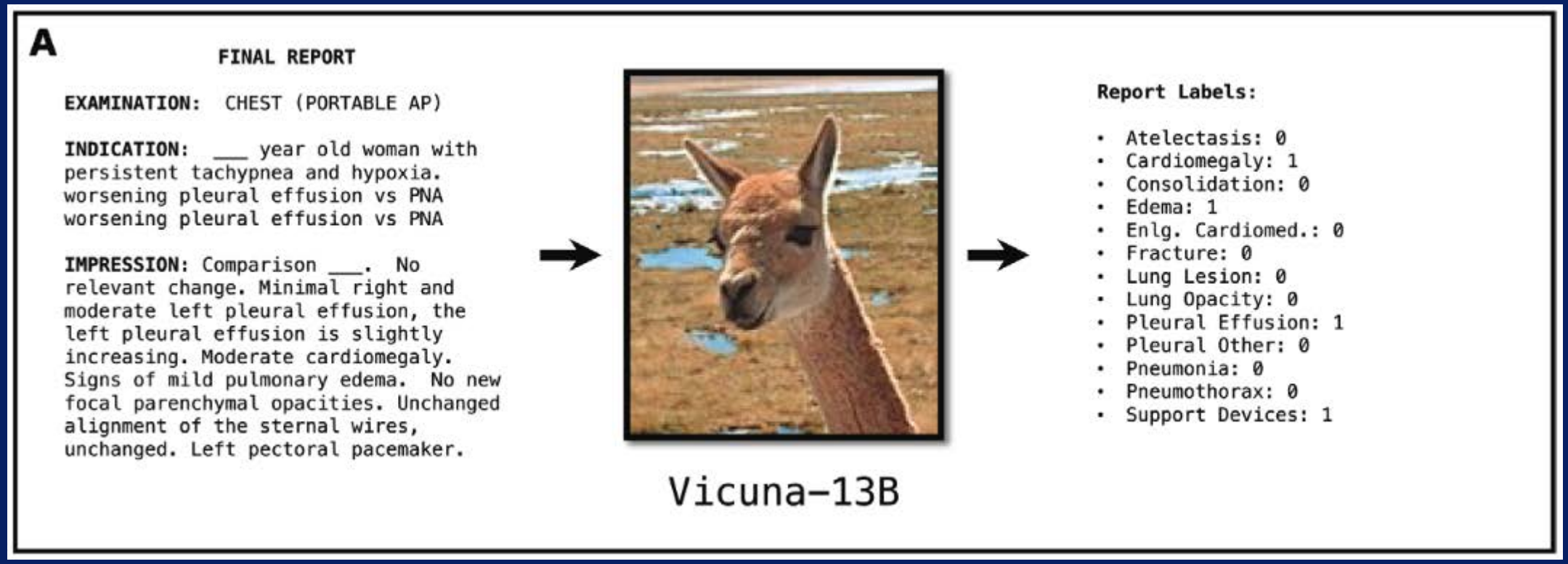
- Corpuses are limited
- Most research uses the same clinical text corpuses
- Given LLMs and Gen AI, more clinical text needed
- Standards for clinical text anonymization desirable

Radiology

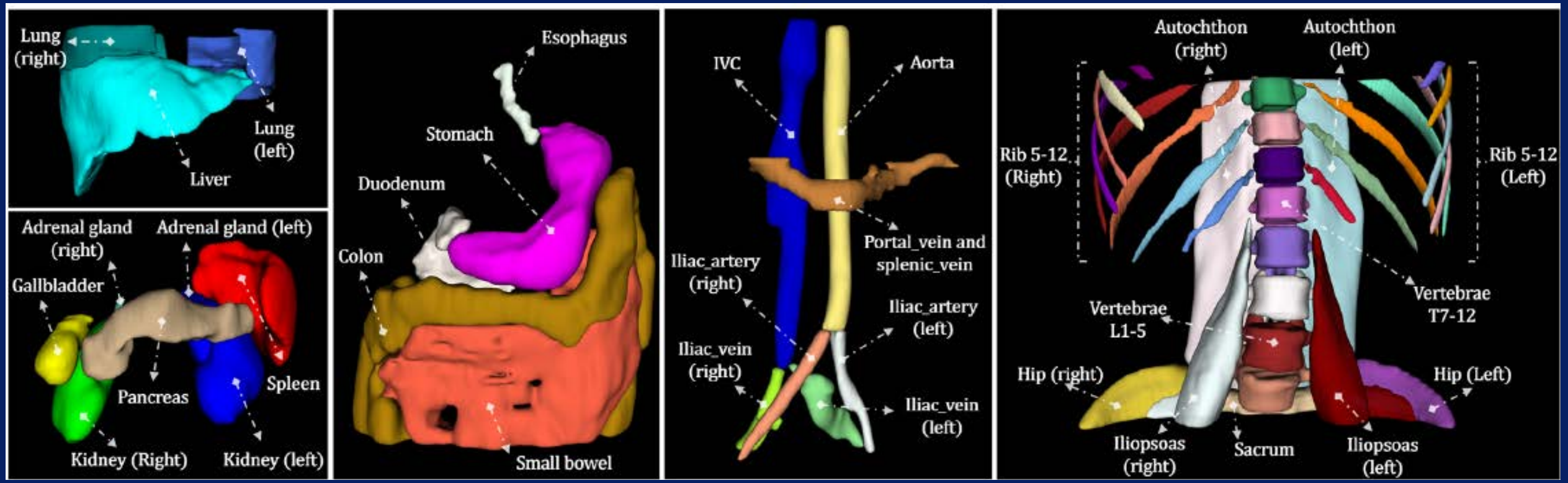
ORIGINAL RESEARCH • COMPUTER APPLICATIONS

Feasibility of Using the Privacy-preserving Large Language Model Vicuna for Labeling Radiology Reports

Pritam Mukherjee, PhD • Benjamin Hou, PhD • Ricardo B. Lanfredi, PhD • Ronald M. Summers, MD, PhD



Mukherjee et al. Radiology 2023



Yan Zhuang et al. Radiology (2025) & arXiv 2405.05944 (2024)

Summary

- AI has transformed biomedical image analysis and informatics
- Clinical images and text remain scarce
- Data archives must balance complexity with generalizability and cost
- Simplicity and standardization are desirable
- Making clinical data publicly available and AI ready will enable improved patient health