

Overcoming Barriers to Data Sharing

October 11, 2022

Curtis P. Langlotz, MD, PhD

Professor of Radiology and Biomedical Informatics, Stanford University

Director, Center for Artificial Intelligence in Medicine & Imaging

Associate Chair, Information Systems, Department of Radiology

Medical Informatics Director, Stanford Health Care

@curtlanglotz



Stanford | MEDICINE



- Anesthesia
- Bioengineering
- Biomedical Data Science
- Cardiothoracic Surgery
- Computer Science
- Dermatology
- Emergency Medicine
- Genetics
- Medicine
- Neurology & Neurological Sciences
- Neurosurgery
- Ophthalmology
- Pathology
- Pediatrics
- Psychiatry & Behavioral Sciences
- Psychology
- Radiation Oncology
- Radiology
- Surgery
- Urology



Stanford AIMI

Center for Artificial Intelligence
in Medicine and Imaging



Key Computer Science Faculty Collaborators



Fei Fei Li, PhD
Creator of ImageNet



Chris Manning, PhD
Stanford AI Lab Director



Andrew Ng, PhD
Deep Learning Pioneer



Chris Re', PhD
Founder, Snorkel AI

A Roadmap for Foundational Research on Artificial Intelligence in Medical Imaging: From the 2018 NIH/RSNA/ACR/The Academy Workshop

Curtis P. Langlotz, MD, PhD • Bibb Allen, MD • Bradley J. Erickson, MD, PhD • Jayashree Kalpathy-Cramer, PhD • Keith Bigelow, BA • Tessa S. Cook, MD, PhD • Adam E. Flanders, MD • Matthew P. Lungren, MD, MPH • David S. Mendelson, MD • Jeffrey D. Rudie, MD, PhD • Ge Wang, PhD • Krishna Kandarpa, MD, PhD

From the Department of Radiology, Stanford University, Stanford, CA 94305 (C.P.L., M.P.L.); Department (B.A.); Department of Radiology, Mayo Clinic, Rochester, Minn (B.J.E.); Department of Radiology, Massachusetts (J.K.C.); GE Healthcare, Chicago, Ill (K.B.); Department of Radiology, Hospital of the University of Radiology, Thomas Jefferson University Hospital, Philadelphia, Pa (A.E.F.); Department of Radiology, Icahn Biomedical Imaging Center, Rensselaer Polytechnic Institute, Troy, NY (G.W.); and National Institute of Health, Washington, DC (K.K.). Received March 17, 2019; revision requested March 19; revision received C.P.L. (e-mail: langlotz@stanford.edu).

Conflicts of interest are listed at the end of this article.

Radiology 2019; 00:1–11 • <https://doi.org/10.1148/radiol.2019190613> • Content code: **IN**

Area	Current State of the Art
Data needs for machine learning research	Few public image data sets are available, mostly small in size and lacking real-world variation.

Area	Current State
Software use cases for AI	AI algorithms are being created based on use cases developed at single institutions working with single developers, limiting diversity and generalizability to widespread clinical practice.

ORIGINAL ARTICLE

<https://doi.org/10.1016/j.jacr.2019.04.014>

A Road Map for Translational Research on Artificial Intelligence in Medical Imaging: From the 2018 National Institutes of Health/RSNA/ACR/The Academy Workshop

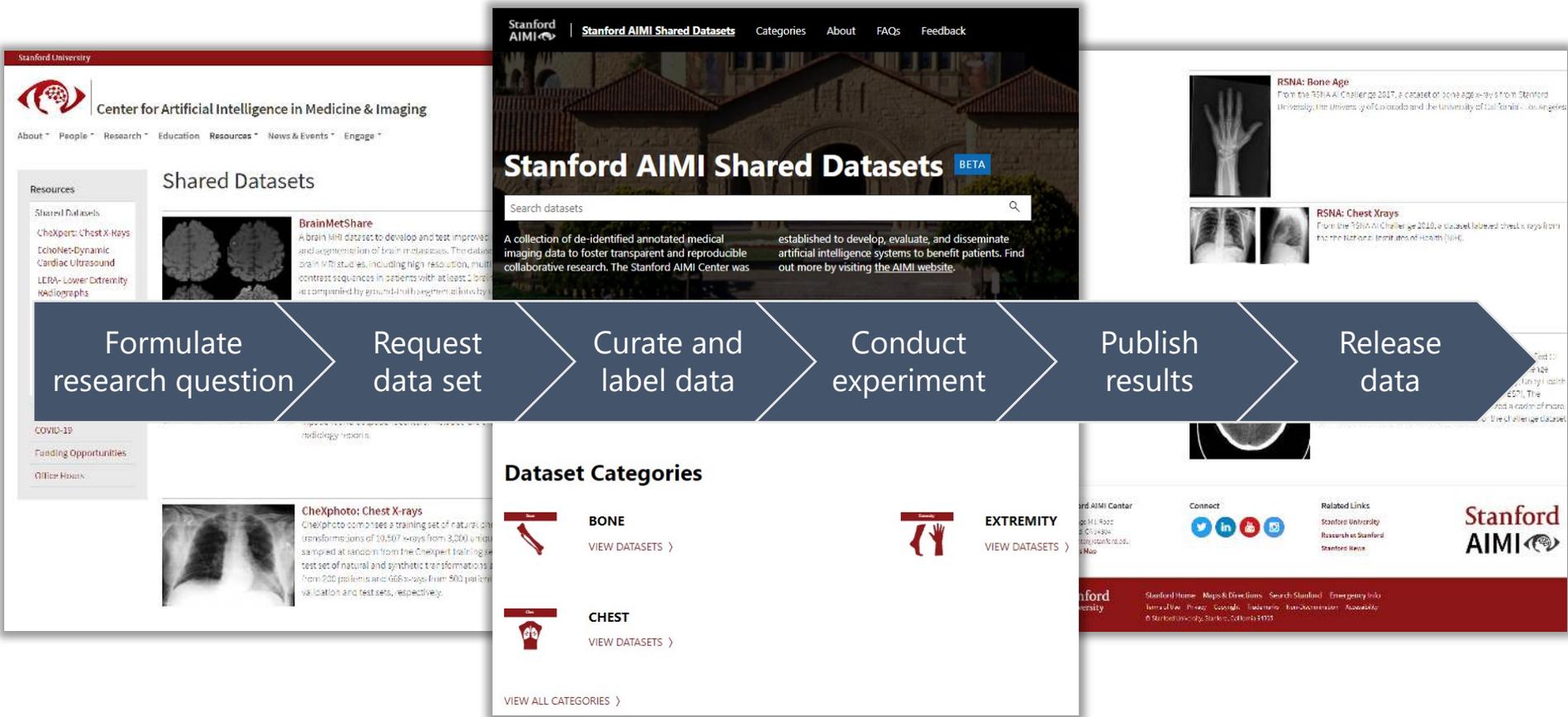
Bibb Allen Jr, MD^a, Steven E. Seltzer, MD^{b,c}, Curtis P. Langlotz, MD, PhD^d, Keith P. Dreyer, DO, PhD^e, Ronald M. Summers, MD, PhD^f, Nicholas Petrick, PhD^g, Danica Marinac-Dabic, MD, PhD, MMS^h, Marisa Cruz, MDⁱ, Tarik K. Alkasab, MD, PhD^e, Robert J. Hanisch, PhD^j, Wendy J. Nilsen, PhD^k, Judy Burleson, BSW, MHSA^l, Kevin Lyman, BS^m, Krishna Kandarpa, MD, PhDⁿ

Ethics of Using and Sharing Clinical Imaging Data for Artificial Intelligence: A Proposed Framework

*David B. Larson, MD, MBA • David C. Magnus, PhD • Matthew P. Lungren, MD, MPH •
Nigam H. Shah, MBBS, PhD • Curtis P. Langlotz, MD, PhD*

“After clinical data are used to provide care, the primary purpose for acquiring the data is fulfilled. At that point, clinical data should be treated as a form of public good, to be used for the benefit of future patients.”

Publicly-Released AI-Ready Radiology Datasets



Stanford University
Center for Artificial Intelligence in Medicine & Imaging

About People Research Education Resources News & Events Engage

Resources

Shared Datasets

BrainMetShare

CheXpert: Chest X-Rays
CohoNet-Dynamic
Cardiac Ultrasound
LEDA- Lower Extremity Radiographs

Shared Datasets

BrainMetShare

A brain MRI dataset to develop and test improved segmentation of brain metastases. The dataset brain MRI studies, including high resolution, multi-contrast sequences in patients with at least 2 brain metastases compared by ground truth segmentation by

Stanford AIMI Shared Datasets

Categories About FAQs Feedback

Stanford AIMI Shared Datasets BETA

Search datasets

A collection of de-identified annotated medical imaging data to foster transparent and reproducible collaborative research. The Stanford AIMI Center was established to develop, evaluate, and disseminate artificial intelligence systems to benefit patients. Find out more by visiting the AIMI website.

RSNA: Bone Age

From the RSNA AI Challenge 2017, a dataset of bone age x-rays from Stanford University, the University of Colorado and the University of California - Los Angeles

RSNA: Chest X-rays

From the RSNA AI Challenge 2010, a dataset labeled chest x-rays from the National Institutes of Health (NIH)

Formulate research question

Request data set

Curate and label data

Conduct experiment

Publish results

Release data

COVID-19

Funding Opportunities

Office Hours

CheXphoto: Chest X-Rays

CheXphoto comprises a training set of naturalistic transformations of 10,507 x-rays from 3,000 entities sampled at random from the CheXpert training set. A test set of natural and synthetic transformations is derived from 200 patients and 668 images from 500 patients for validation and test sets, respectively.

radiology experts.

Dataset Categories

BONE

VIEW DATASETS >

EXTREMITY

VIEW DATASETS >

CHEST

VIEW DATASETS >

VIEW ALL CATEGORIES >

Stanford AIMI Center

Connect

Related Links

Stanford University
Research at Stanford
Stanford Health

Stanford AIMI

Stanford Home | Maps & Directions | Search Stanford | Emergency Info
Terms of Use | Privacy | Contact | Accessibility | Non-Discrimination | News & Info
© Stanford University, Stanford, California 94305

RSNA® Data Science Challenges



Pediatric Bone Age Challenge

Organized by RSNA.org/organizing.committee - Current server

▶ Current

Test

Oct. 7, 2017, midnight UTC

Next

Leaderboard

Sept. 1, 2017, midnight UTC

2017

2020

RSNA Intracranial Hemorrhage Detection
Identify acute intracranial hemorrhage and its subtypes

\$25,000 Prize Money

RSNA Radiological Society of North America · 191 teams · 2 months to go (a month to go until merger deadline)

Join Competition

2018

RSNA Pneumonia Detection Challenge

Can you build an algorithm that automatically detects potential pneumonia cases?

\$30,000 Prize Money

RSNA Radiological Society of North America · 1,499 teams · 2 years ago

Join Competition

2021

RSNA-MICCAI Brain Tumor Radiogenomic Classification

Predict the status of a genetic biomarker important for brain cancer treatment

\$30,000 Prize Money

RSNA Radiological Society of North America · 124 teams · 3 months to go (3 months to go until merger deadline)

Join Competition

2019

RSNA STR Pulmonary Embolism Detection

Classify Pulmonary Embolism cases in chest CT scans

\$30,000 Prize Money

RSNA Radiological Society of North America · 488 teams · 17 days to go (10 days to go until merger deadline)

Join Competition

2021

RSNA-ASNR-MICCAI BraTS Challenge 2021

Click Here to Register

There are 476 registered participants. Join them now!

Overcoming Barriers to Data Sharing



Stanford AIMI

Center for Artificial Intelligence
in Medicine and Imaging

- Barriers to “AI-readiness” of data
 - Volume, labels, diversity
- Barriers to organizational readiness for data release
 - Organizational capacity
 - Many or no privacy offices, lack of an IRB
 - Risk aversion
 - HIPAA breach, public announcements
 - Legal deliberations
 - Patient consent
 - Data use agreements
 - Commercial use
 - Technical factors
 - De-identification
 - Format harmonization
 - Data hosting, user tracking
 - Cost

Persistence,
enlightened IRB

Mission-related
sharing, mandates

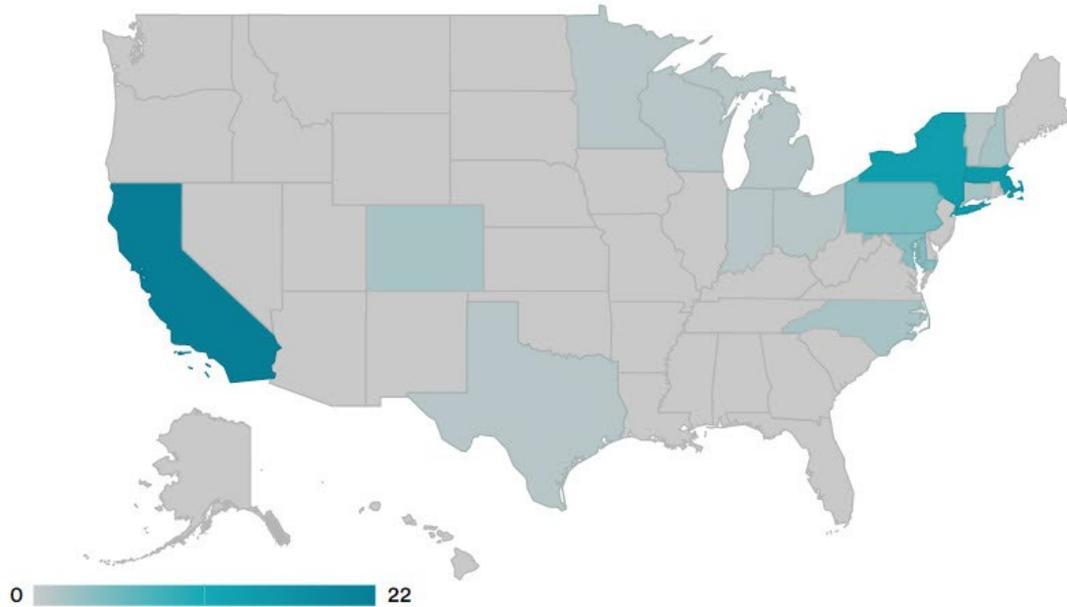
Current language
being expanded

Non-commercial
only

DICOM Anonymizer,
OMOP

Box→Microsoft,
RSNA, MIDRC

Geographic Distribution of Cohorts to Train Deep Learning Algorithms



REBECCA ROBBINS/STAT

SOURCE: "GEOGRAPHIC DISTRIBUTION OF US COHORTS USED TO TRAIN DEEP LEARNING ALGORITHMS," JAMA 2020.

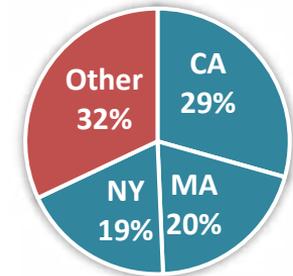
STAT

Table. US Patient Cohorts Used for Training Clinical Machine Learning Algorithms, by State^a

States	No. of studies
California	22
Massachusetts	15
New York	14
Pennsylvania	5
Maryland	4
Colorado	2
Connecticut	2
New Hampshire	2
North Carolina	2
Indiana	1
Michigan	1
Minnesota	1
Ohio	1
Texas	1
Vermont	1
Wisconsin	1

^a Fifty-six studies used 1 or more geographically identifiable US patient cohort in the training of their clinical machine learning algorithm. Thirty-four states were not represented in geographically identifiable cohorts: Alabama, Alaska, Arizona, Arkansas, Delaware, Florida, Georgia, Hawaii, Idaho, Illinois, Iowa, Kansas, Kentucky, Louisiana, Maine, Mississippi, Missouri, Montana, Nebraska, Nevada, New Jersey, New Mexico, North Dakota, Oklahoma, Oregon, Rhode Island, South Carolina, South Dakota, Tennessee, Utah, Virginia, Washington, West Virginia, and Wyoming.

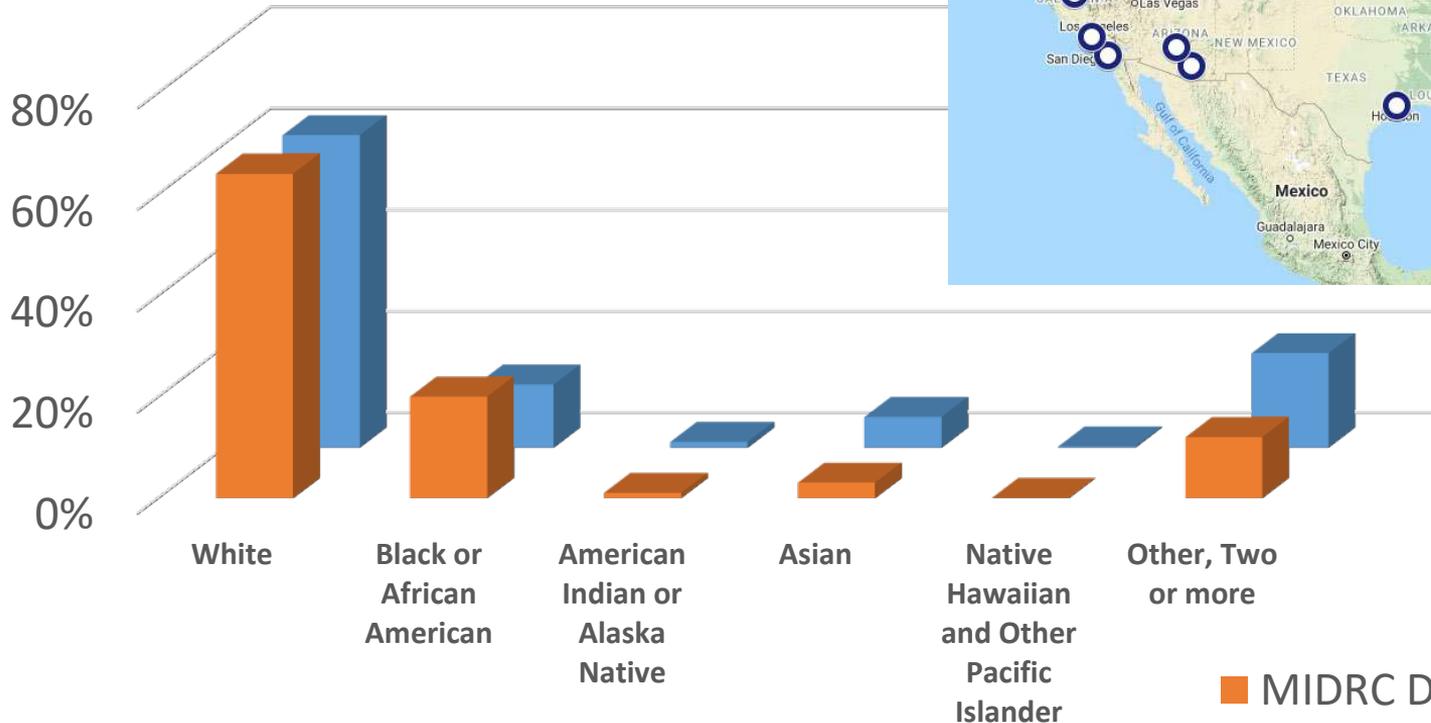
JAMA September 22/29, 2020 Volume 324, Number 12





MIDRC

MEDICAL IMAGING AND DATA RESOURCE CENTER.



Conclusions



- Lack of data is a key bottleneck.
- Wide public release of data can be accomplished.
- Data readiness and organizational readiness are distinct concepts.
- Recommendations:
 - Large organizations need guidance on best practices.
 - Inexperienced organizations need help from centers of excellence.
 - Sustainable models will require both central and distributed investment.



Stanford AIMI

Center for Artificial Intelligence
in Medicine and Imaging

Engage:

@curtlanglotz

@StanfordAIMI

LinkedIn: aimi-stanford

Join the mailing list: <https://aimi.stanford.edu/engage/subscribe>

