# AAPM Symposium --


1. Overview of MIDRC and AI of COVID-19
   - Maryellen Giger (University of Chicago)
2. Image Quality and Harmonization of MIDRC Imaging Data
   - Paul Kinahan (University of Washington)
3. Performance Metrics for Evaluation of COVID-19 AI
   - Michael McNitt-Gray (UCLA)
4. Role of Sequestered Datasets in MIDRC
   - Kyle Myers (FDA)
5. Grand Challenges with MIDRC
   - Samuel Armato, III (University of Chicago)

RSNA Radiological Society of North America

ACR American College of Radiology

AAPM American Association of Physicists in Medicine

MIDRC Medical Imaging and Data Resource Center

# Overview of MIDRC -- and AI of COVID-19

**Maryellen Giger, Ph.D.**

**University of Chicago**

NIH National Institute of Biomedical Imaging and Bioengineering

The University of Chicago

# Medical Imaging and Data Resource Center
# THE ORIGIN

**NIBIB** convenes medical imaging academics, professional organizations, industry and other Federal agencies with workshops on **AI in Medical Imaging** (August 2018) & **Acceleration of Clinical Applications of Machine Intelligence in MI** (November 2019), which identified critical gaps:

1) Absence of large & diverse medical image datasets
2) Need to integrate siloed databases & knowledgebases
3) Need to develop non-redundant efficient AI tools
4) Need to create an ecosystem of stakeholders to develop clinically validated AI applications that improve patient management and clinical outcomes

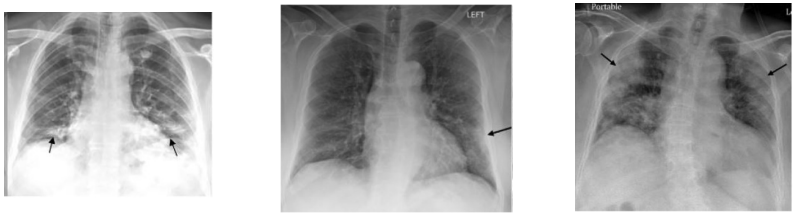
Parallel workshops by ARBIR/NIST, RSNA and AAPM in 2018-2019 confirm the developing consensus

Slide courtesy of Kris Kardarpa, NIBIB

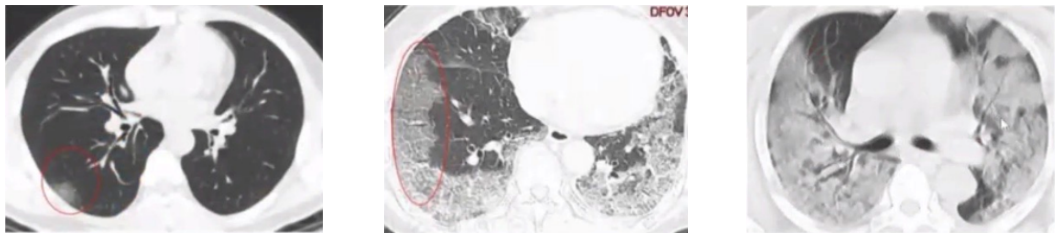
# Then there was the COVID-19 Pandemic

MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER

- COVID-19 impacts **lungs, as well as heart, vessels, & brain**.
- Need to **collect medical images** and **develop artificial intelligence** (AI) methods to aid in the analysis / interpretation of medical images



Various presentations of COVID-19 on chest radiographs from three different patients



Early Stage | Progressive Stage | Severe Stage

Various stages of COVID-19 shown on chest CTs from three different patients

# Rapid Response to COVID-19 Pandemic

MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER.
Established August 21, 2020

**University of Chicago** NIBIB Contract PI: Maryellen Giger

**American Association of Physicists in Medicine** (AAPM) PIs:

- Maryellen Giger (University of Chicago & AAPM Data Science Committee Chair)
- Paul Kinahan (University of Washington & AAPM Research Committee Chair)

**Radiological Society of North America** (RSNA) PIs:

- Curtis Langlotz (Stanford University & RSNA Board Liaison for IT & Annual Meeting)
- Adam Flanders (Thomas Jefferson University & Member RSNA CDE Committee)

**American College of Radiology** (ACR) PIs:

- Etta Pisano (ACR Chief Research Officer & Harvard University)
- Michael Tilkin (ACR Chief Information Officer)

**Gen3** PI: Robert Grossman

NIH National Institute of Biomedical Imaging and Bioengineering

THE UNIVERSITY OF CHICAGO

AAPM AMERICAN ASSOCIATION of PHYSICISTS IN MEDICINE

ACR AMERICAN COLLEGE OF RADIOLOGY

RSNA Radiological Society of North America

# Goals

- Progress from data to deployment & hypothesis to discovery
- Accelerate the creation and transfer of knowledge for clinical management of COVID-19

**A. Open Discovery Data Repository Commons**

- Creation, testing, quality assurance, diversity, and data connectivity
- Large scale
- Think “ImageNet” for medical imaging/radiology

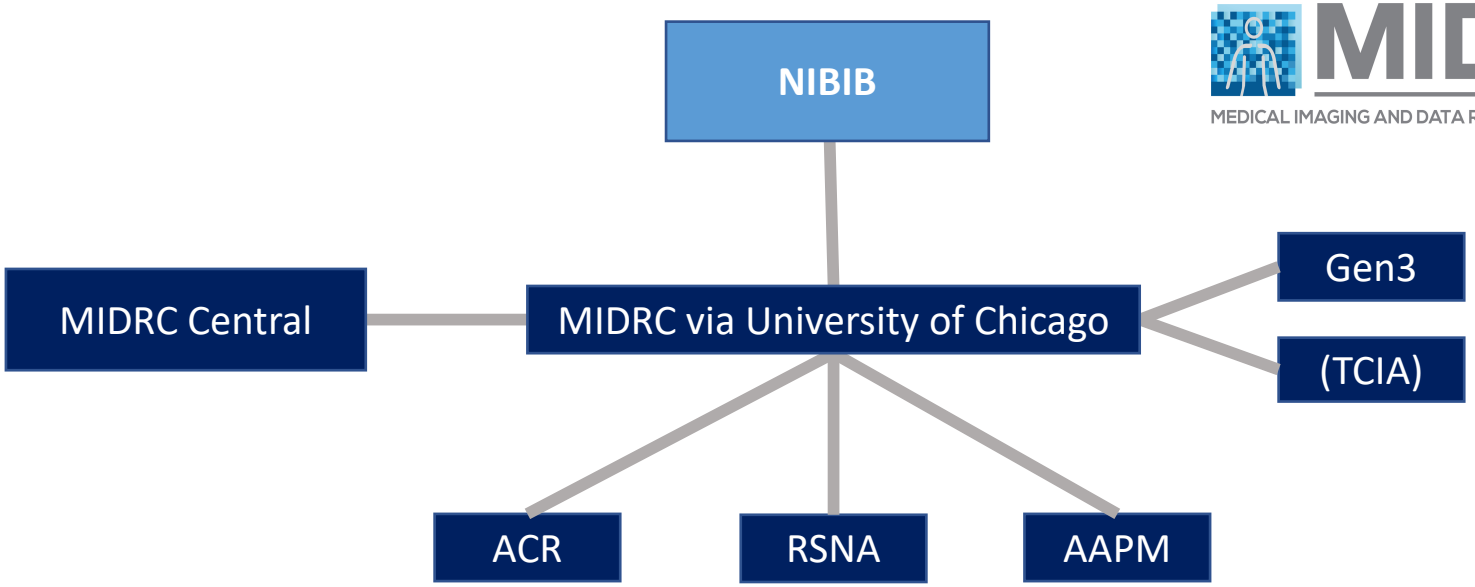
**B. Machine Intelligence Computational Capabilities**

- clinically relevant algorithms and software tools

MIDRC

MEDICAL IMAGING AND DATA RESOURCE CENTER.


- MIDRC -- radiologists & medical imaging scientists from across the nation
  - 23 institutions from academia, community practices, FDA
  - Expert collaboration with community engagement
- See website for listing of all investigators
  - **https://www.midrc.org**
- High-quality and diverse data commons enabling researchers to address topics no single archive could yield independently




MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER.

## Two Major Scientific Components

**Creation of Open Discovery Data Repository: 5 Technology Development Projects along with three data science subcommittees and advisory committees**

**Machine Intelligence Computational Capabilities: 12 Collaborative Research Projects along with multiple trans-MIDRC scientific workgroups**

# MIDRC: Technology Development Projects

The **MIDRC infrastructure and processes** will be created through five **Technology Development Projects**, which will be conducted collaboratively:

1. Creating an open discovery platform for COVID-19 imaging and associated data (**led by RSNA**).
2. Creating a real-world testing and implementation platform with direct real-time connections to health care delivery organizations (**led by ACR**).
3. Developing and implementing quality assurance and evaluation procedures for usage across the MIDRC (**led by AAPM**).
4. Enabling data intake, access and distribution via a world-facing data commons portal (**led by all three plus Gen3**).
5. Linking the MIDRC to other clinical and research data registries (**led by all three plus Gen3**).

**Three MIDRC Data Science Subcommittees**

- DSIT - Data Standards and Information Technology Subcommittee
  - led by RSNA
- DPP - Data Policy and Procedures Subcommittee
  - led by ACR
- DQH - Data Quality and Harmonization Subcommittee
  - led by AAPM

# MIDRC

MEDICAL IMAGING AND DATA RESOURCE CENTER.

## Two Data Intake Portals

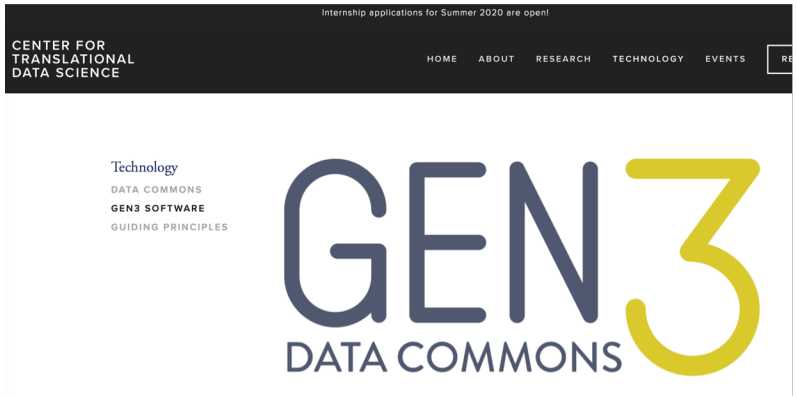
RSNA
COVID-19 Database
RICORD

ACR
COVID-19 Imaging Research REGISTRY™

## One Output User Portal

University of Chicago


- To be accessed by hundreds of researchers and developers of AI.
- For training and testing of AI to reduce bias and enhance diversity.
- To expedite translation of AI to clinical care.




https://gen3.org

- Gen3 Data Commons at the University of Chicago is a cloud-based software platform for managing, analyzing, harmonizing, and sharing large datasets.
- Gen3 is an open source platform for developing data commons.
- Data commons accelerate and democratize the process of scientific discovery, especially over large or complex datasets.

Gen3 hosts over a dozen Data Commons including, for example

NIH National Heart, Lung, and Blood Institute | BioData CATALYST

| | |
|---|---|
| 240,460 | Subjects |
| 992 | Attributes |
| 24,291,972 | Files |
| Total Size | 2.93 PB |

NIH NATIONAL CANCER INSTITUTE
Cancer Research Data Commons

| | |
|---|---|
| 83,709 | Subjects |
| 622 | Attributes |
| 5,108,909 | Files |
| Total Size | 3.53 PB |

MIDRC

MEDICAL IMAGING AND DATA RESOURCE CENTER.

## Two Data Intake Portals

RSNA COVID-19 Databa

RICORD

ACR COVID-19 Imaging Research REGISTRY™

AAPM -- quality assurance and evaluation procedures

## One Output User Portal

University of Chicago


GEN3 DATA COMMONS

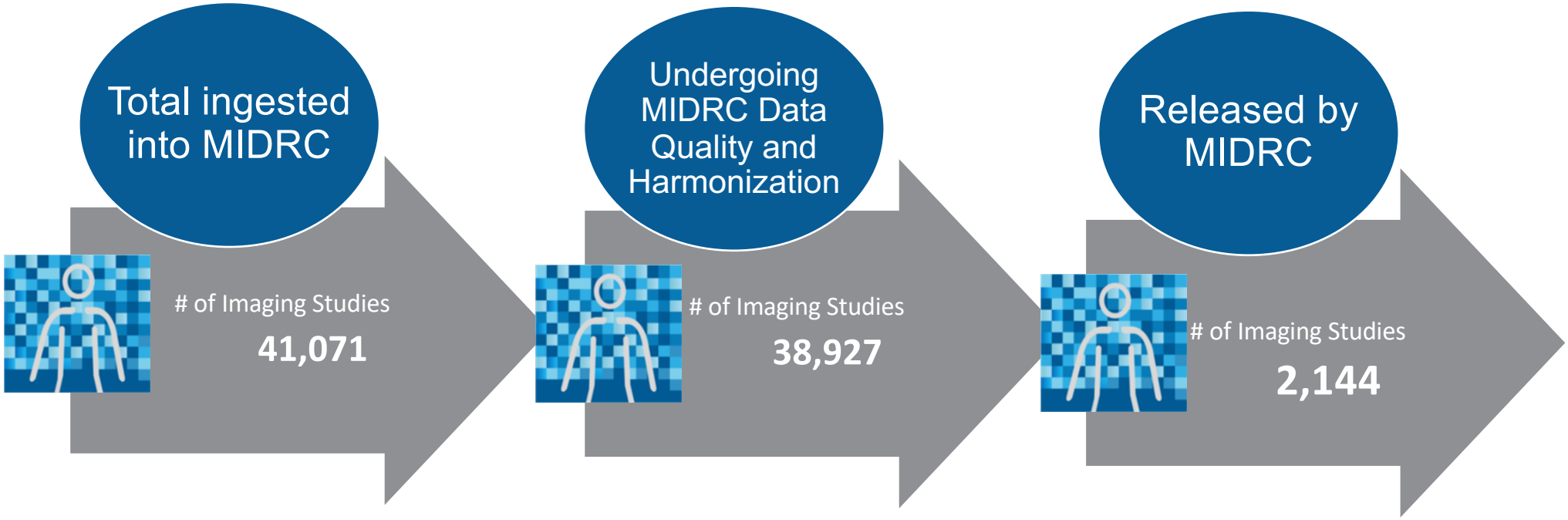
- To be accessed by hundreds of researchers and developers of AI.
- For training and testing of AI to reduce bias and enhance diversity.
- To expedite translation of AI to clinical care.

# TDP 3: Developing and implementing quality assurance and evaluation procedures for usage across the MIDRC (led by AAPM)

- TDP 3a: Development of digital and physical imaging phantoms for COVID data
  - John Boone, Paul Kinahan, Tony Siebert, Andrey Fedorov, Nicholas Bevins, Dan Sullivan
- TDP 3b: Assessment of image quality on ingestion into MIDRC
  - Paul Kinahan, John Boone, Tony Siebert, Andrey Fedorov, Nicholas Bevins, Dan Sullivan
- TDP 3c: Development of benchmarking methods for the various technology assessment and clinical tasks in COVID-19 research and translation
  - Michael McNitt-Gray, Berkman Sahiner, Karen Drukker, Maryellen Giger
- TDP 3d: Development of task-based distribution methods
  - Kyle Myers, Maryellen Giger, Heather Whitney, Natalie Baughan

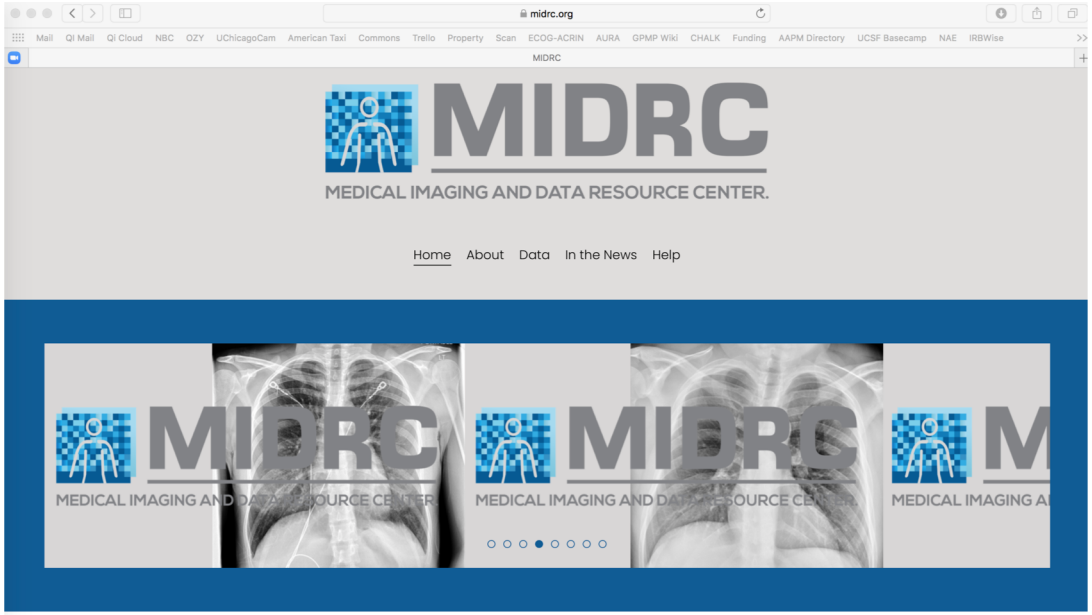
# MIDRC Data Dashboard

MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER.




**Quality checked**
**Diversity assessed**
**Clinical Task AI**

**Goal of 60,000 curated imaging studies to be released by MIDRC by Sept 2021**



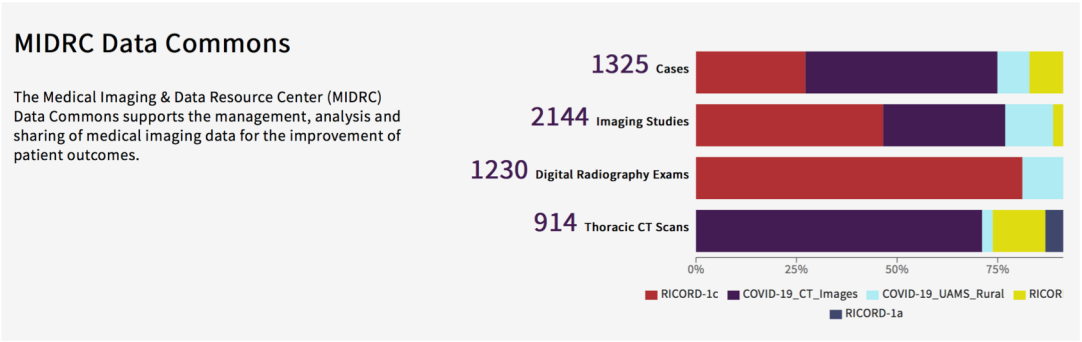

**midrc.org** is for:

- researchers
- data contributors
- the public

Subscribe to our newsletter.

Sign up with your email address to receive news and updates. Please note that subscription is a two-step process: a confirmation email will be sent to the email address provided for you to confirm subscription.

https://www.midrc.org/register-to-receive-newsletter




**data.midrc.org**
is for searching and downloading data

# MIDRC: Collaborative Research Projects (CRPs)

- Each of the three lead organizations, **RSNA, ACR, and AAPM, have major data science committees** that are already actively pursuing aspects of big data science in medical imaging and the role of machine intelligence.
  - RSNA has its Radiology Informatics Committee (RIC)
  - ACR has its Data Science Institute (DSI)
  - AAPM has its Data Science Committee (DSC)
- Thus, the MIDRC PIs, serving as the initial Executive Advisory Committee, developed the **twelve research projects** to be immediately funded through the MIDRC within the initial contract in order **to expedite the AI research** leading to **rapid translation** to the public and clinical impact.
- The investigators were selected from current members of the RSNA-RIC, ACR-DSI, or AAPM-DSC, thus, effectively **spanning the nation and spanning the medical imaging community**.

## Collaborative Research Projects – Investigators through the various Data Science Committees at ACR, RSNA, & AAPM

| Project | Title |
|---|---|
| 1 | Natural Language Processing of Radiology Reports for COVID-19 |
| 2 | Machine Intelligence Algorithms from Multi-Modal, Multi-institutional COVID-19 Data |
| 3 | Image Labeling and Annotation by a Crowd of Experts for COVID-19 |
| 4 | Efficient Training and Explainability of Machine Learning Methods from Multi-Institutional Data |
| 5 | COVID Pneumonia Machine Learning Algorithm Validation and Visualization |
| 6 | Safe Public Training Dataset for COVID-19 Machine Learning Algorithms |
| 7 | Leveraging Registry Data to Conduct Virtual Clinical Trials |
| 8 | Prediction of COVID Pneumonia Outcome using Radiomic Feature Analysis |
| 9 | Radiomics & Machine Intelligence of COVID-19 for detection and diagnosis on chest radiographs and thoracic CTs |
| 10 | Visualization & Explainability of Machine Intelligence of COVID-19 for prognosis and monitoring therapy |
| 11 | Investigation of image-based biomarkers for radiogenomics of COVID-19 |
| 12 | Determining COVID-19 image data quality, provenance, and harmonization |

**Trans-MIDRC scientific workgroups**

- Grand Challenges Work Group
  - Created to coordinate effort on all aspects of challenges
  - Potential to merge top performing algorithms to benefit the common good
- Bias and Diversity Work Group
  - Goal of assessing and mitigating bias in data and ML
  - Diversity in MIDRC investigators and users

| Project | MIDRC Collaborate Research Projects |
|---|---|
| 1 | Natural Language Processing of Radiology Reports for COVID-19 |
| 2 | Machine Intelligence Algorithms from Multi-Modal, Multi-institutional COVID-19 Data |
| 3 | Image Labeling and Annotation by a Crowd of Experts for COVID-19 |
| 4 | Efficient Training and Explainability of Machine Learning Methods from Multi-Institutional Data |
| 5 | COVID Pneumonia Machine Learning Algorithm Validation and Visualization |
| 6 | Safe Public Training Dataset for COVID-19 Machine Learning Algorithms |
| 7 | Leveraging Registry Data to Conduct Virtual Clinical Trials |
| 8 | Prediction of COVID Pneumonia Outcome using Radiomic Feature Analysis |
| 9 | Radiomics & Machine Intelligence of COVID-19 for detection and diagnosis on chest radiographs and thoracic CTs |
| 10 | Visualization & Explainability of Machine Intelligence of COVID-19 for prognosis and monitoring therapy |
| 11 | Investigation of image-based biomarkers for radiogenomics of COVID-19 |
| 12 | Determining COVID-19 image data quality, provenance, and harmonization |

# CRPs led by AAPM members

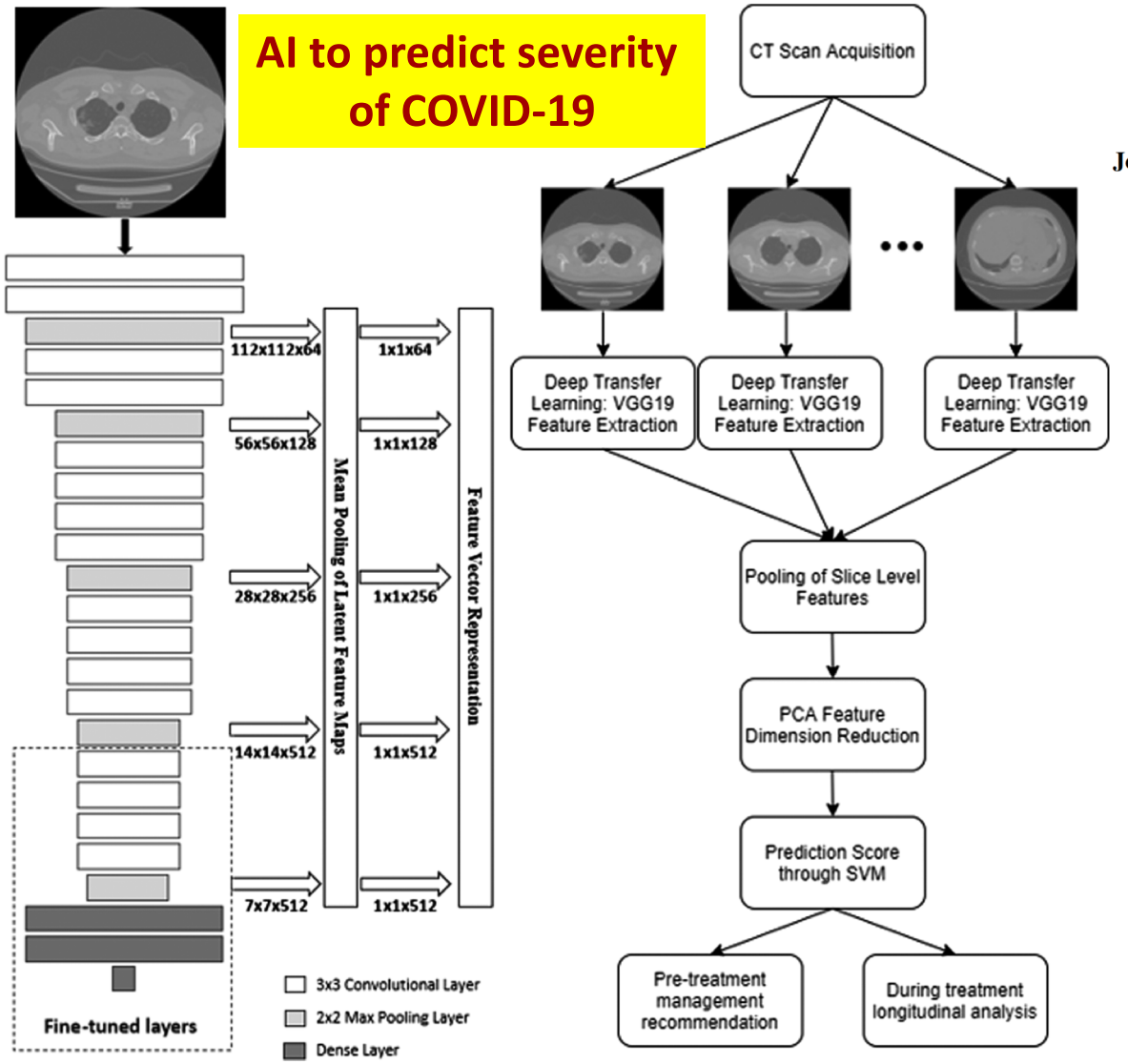
- CRP 9: Radiomics & Machine Intelligence of COVID-19 for detection and diagnosis on chest radiographs and thoracic CTs
  - Sam Armato, Lubomir Hadjiski, Karen Drukker
- CRP 10: Visualization & Explainability of Machine Intelligence of COVID-19 for prognosis and monitoring therapy
  - Maryellen Giger, Hui Li, Issam El-Naqa, Jonathan Fuhrman, Isabelle Hu, Naveena Gorre
- CRP 11: Investigation of image-based biomarkers for radiogenomics of COVID-19
  - Weijie Chen, Sandy Napel, Maryellen Giger, Diane Lauderdale
- CRP 12:Determining COVID-19 image data quality, provenance, and harmonization
  - Paul Kinahan, Andrey Fedorov, Dan Sullivan

Example of MIDRC CRP Activity:

# CRP10: Visualization & Explainability of Machine Intelligence of COVID-19 for prognosis and monitoring therapy

- Developing white papers to aid the AI investigators better understand AI for medical imaging
- Developing machine learning methods **to predict severity of COVID-19**
  - Accurate prognosis of COVID-19 is crucial as it enables implementation of appropriate treatment for individual patients and medical resource allocation optimization
  - **Use "surrogate markers" of severity in the training**
  - Fuhrman et al.
    - Assess severity (prognosis) of COVID-19 in CT scans.
    - Predict that a COVID-19 patient would be recommended for steroid treatment or not.
    - Temporal evaluation for monitoring treatment.
  - Hu et al.
    - Assess severity (prognosis) of COVID-19 in Chest Radiographs
    - Predict patients' needs for intensive care, defined as intubation (invasive mechanical ventilation) and/or intensive care unit (ICU) admission

# AI to predict severity of COVID-19

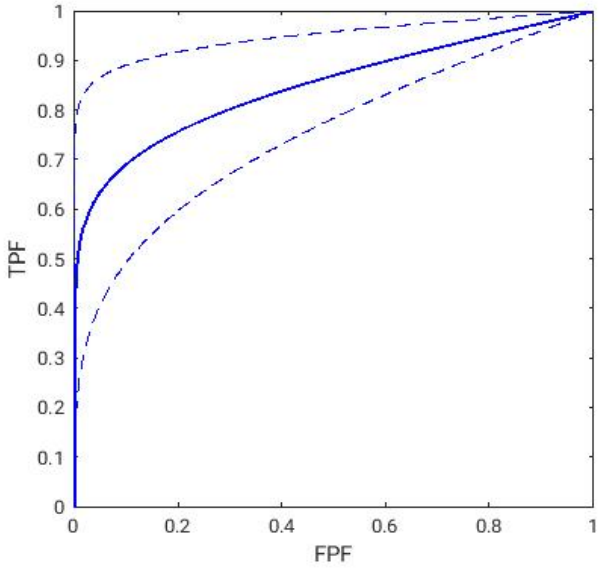



**Cascaded deep transfer learning on thoracic CT in COVID-19 patients treated with steroids**

Jordan D. Fuhrman,[a] Jun Chen,[b] Zegang Dong,[c] Fleming Y. M. Lure,[c] Zhe Luo,[d,e,*] and Maryellen L. Giger[a,*]

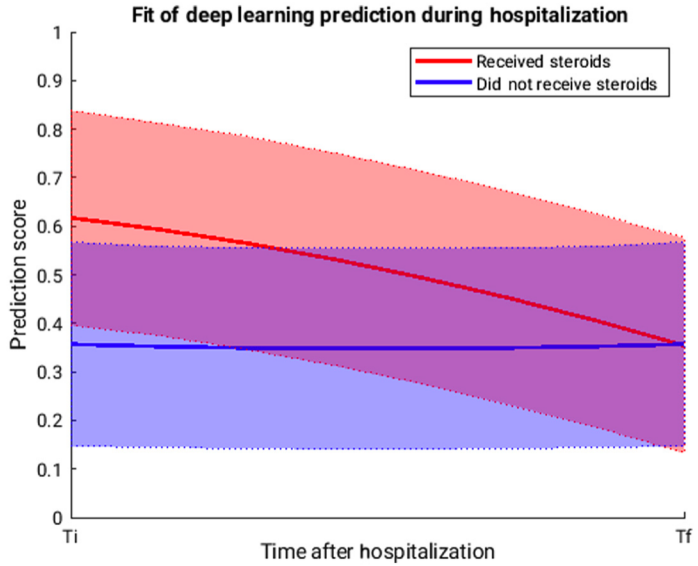
- Pretrained VGG19 network feature extraction approach operating on a two-dimensional (2D) CT section.
  - Max pooling layer features with the given dimensions were averaged and concatenated to produce a representative feature vector for each slice.
- Full cascaded transfer learning workflow for pretreatment assessment and during-treatment monitoring analysis.
- The feature extraction scheme displayed on the left is utilized at the "Deep Transfer Learning: VGG19 Feature Extraction" stage of right.

## **Pretreatment** Cascaded Transfer Learning of CT for COVID-19 Management Recommendation




- Classification ability of the cascade transfer learning method for estimating the likelihood that a COVID-19 patient would be **recommended for steroid treatment or not.** AUC = 0.85

## **During-Treatment** Transfer Learning for COVID-19 Longitudinal Analysis of CT Scans




- The SVM-output prediction score assessed **temporally** across the duration of hospitalization
- The shaded regions denote one standard deviation above and below the fit line.

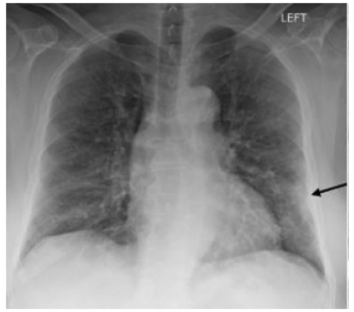
Fuhrman et al. 2021

# AI to predict severity of COVID-19

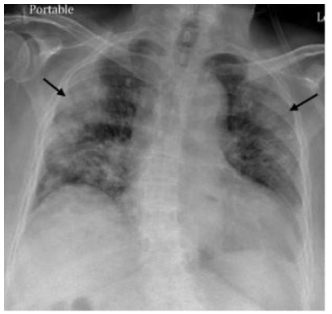
**Predicting the Need for Intensive Care for COVID-19 Patients using Deep Learning on Chest Radiography**

**Qiyuan Hu, Karen Drukker, Maryellen L. Giger**
Committee on Medical Physics, Department of Radiology, University of Chicago
{qhu,kdrukker,m-giger}@uchicago.edu

- Chest radiography (CXR) is recommended for triaging at patient presentation and disease monitoring due to its fast speed, relatively low cost, wide availability, and portability
- Develop an AI method to perform prognosis for COVID-19 patients using CXR, predicting patients' needs for intensive care, which we defined as intubation (invasive mechanical ventilation) and/or intensive care unit (ICU) admission
- CXR exams from adult patients who underwent COVID-19 RT-PCR tests
  - 1670 CXR from 1178 COVID-19+ patients
- Input: Standard or portable CXR of COVID-19+ patients
- Curriculum learning approach employed to train the model on a sequence of gradually more specific and complex tasks, mimicking the human learning process



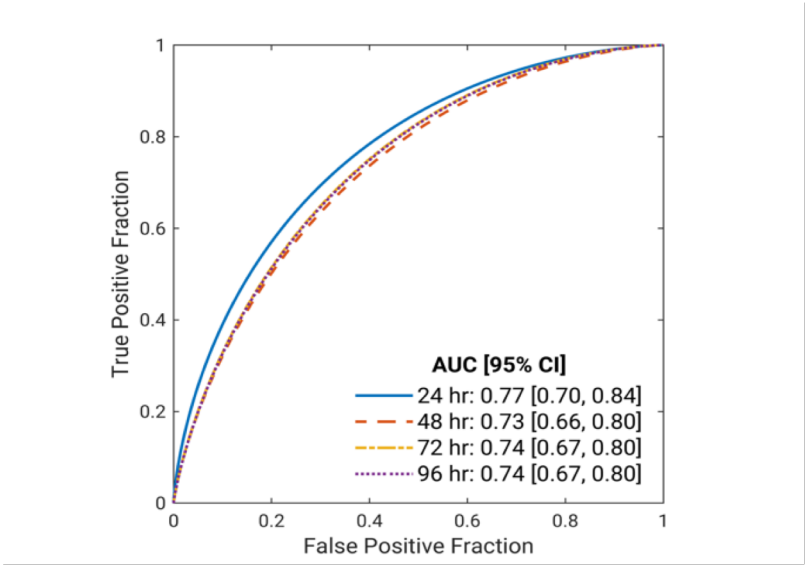

Patchy peripheral ground glass opacities




Peripheral air space opacities

# COVID-19 Prognosis Results

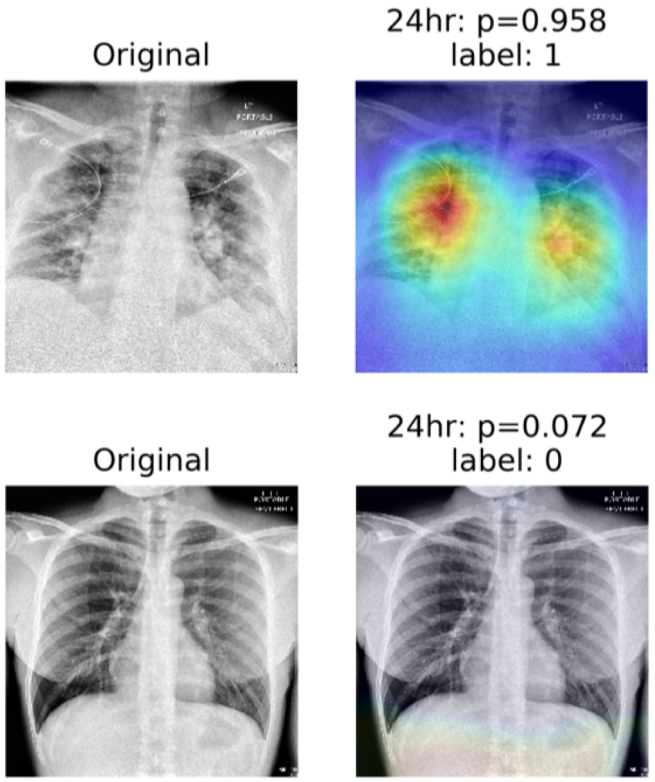
Phase 3: Fine-tuned DenseNet121 on COVID-19 dataset predict




Proposed COVID-19 prognostic method achieved an AUC [95% CI] of 0.77 [0.70, 0.84] when predicting the need for intensive care 24 hours in advance, and at least 0.73 [0.66, 0.80] for earlier predictions (48 – 96 hours in advance) based on single CXR exams

**Predicting the Need for Intensive Care for COVID-19 Patients using Deep Learning on Chest Radiography**


**Qiyuan Hu, Karen Drukker, Maryellen L. Giger**
Committee on Medical Physics, Department of Radiology, University of Chicago
{qhu,kdrukker,m-giger}@uchicago.edu

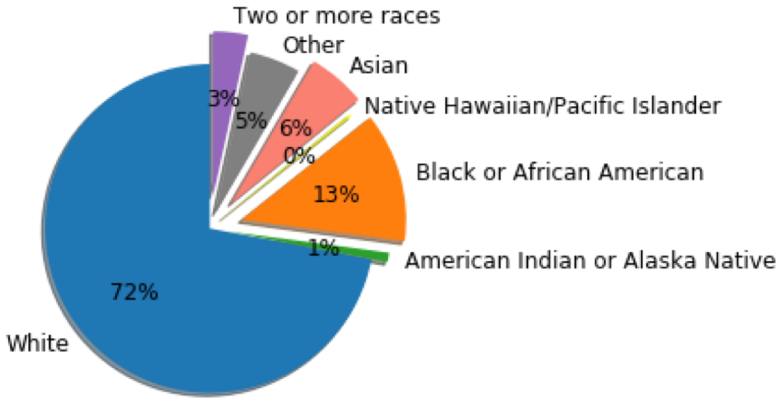
# Diversity of the data is an essential component in the developing and testing of unbiased data

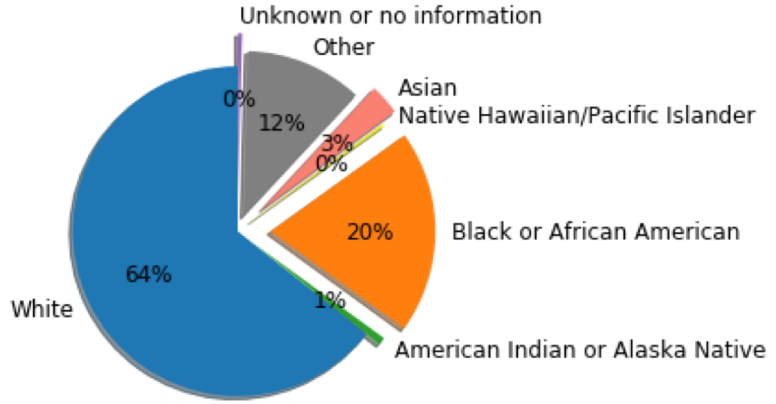
MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER.

## US Census

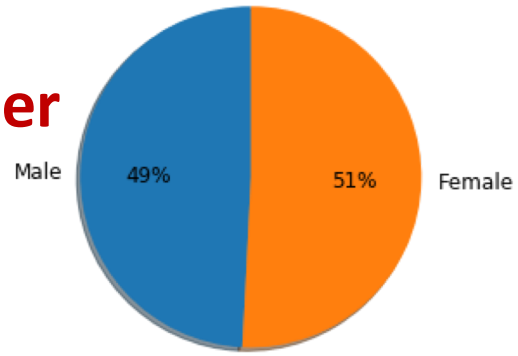
## MIDRC

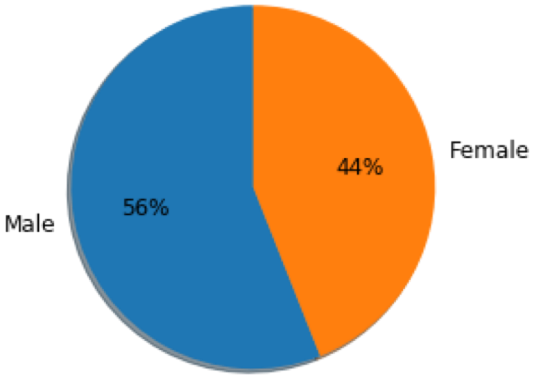
### Race




### Gender




As of incoming data
May 31, 2021

# Outreach & Collaborations: presentations, meetings, townhalls, …

MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER

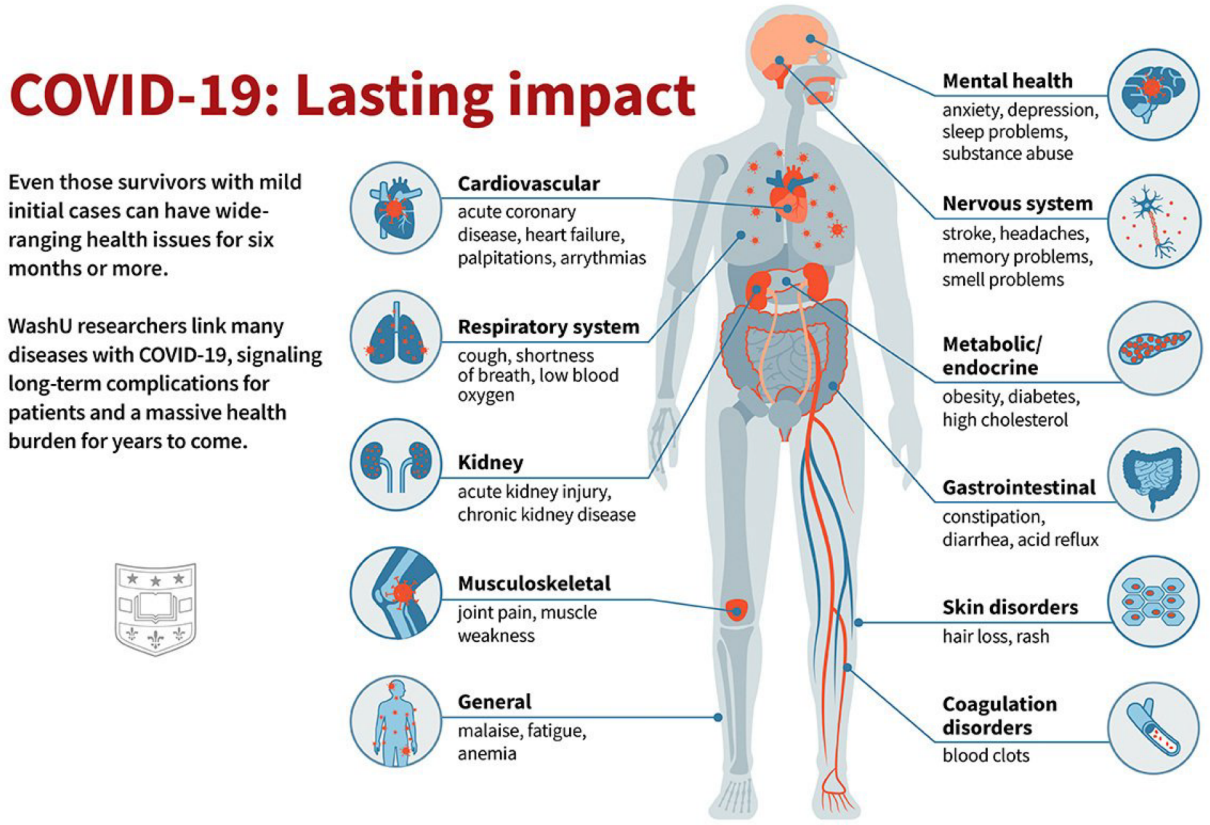
- Annual Meeting (Dec 2020)
- Data Contribution Townhall (March 2021) & **another Townhall planned (July 2021)**
- Congressional Briefing (Mar 2021) - with the Academy for Radiology & Biomedical Imaging Research
- NCATS N3C (multiple times)
- NHLBI BioData Catalyst (multiple times)
- ODSS (NIH Office of Data Science Strategy)
- SPIE (on supporting challenges)
- ARBIR (Academy for Radiology & Biomedical Imaging Research) (Dec 2020)
- AIUM (American Institute of Ultrasound in Medicine) (multiple times)
- ASNR (American Society of Neuroradiology) (multiple times)
- MDRIG (Medical Device Research Interest Group) (April 2021)
- MDIC (Medical Device Innovation Consortium) (April 2021)
- Various invited talks by MIDRC PIs

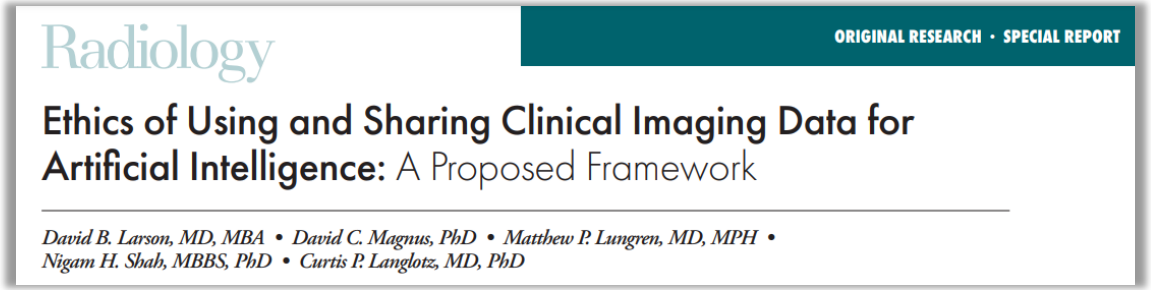
  Special symposium at AAPM in July 2021 and Special course at RSNA in November 2021

# What is next for MIDRC?

MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER.

- **Beyond chest radiographs and thoracic CTs for COVID-19**
  - Include images of the (**brain**)
  - Include images beyor images to **monitor p**
  - Collaborate with the across clinical data, ir
- **Beyond COVID-19**
  - MIDRC, with its deve tools, will be ready fc
- Thus, MIDRC will be a **n**
  - Currently funded for
  - Require additional fu






https://pubs.rsna.org/doi/pdf/10.1148/radiol.2020192536

MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER

"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. All who interact with or control the data have an obligation to ensure that the data are used for the benefit of future patients and of society."

Note, for a given medical image,

-- A patient has already benefited through medical care.

-- A hospital/medical center has already benefitted through reimbursement.

-- Now, the public can benefit with the MIDRC secondary usage of the images.

-- We all can help change the future of medical imaging and increase its impact on public health.

# Summary: MIDRC is more than just a data registry

MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER

- A **high-quality and diverse data commons** led by **major imaging societies** enabling researchers to address topics no single archive could yield independently.

Including:

- Proactive, encouraging system for **changing the culture of contributing data for the common good.**
- **User-friendly open output portal** for browsing/exploring the data, creating cohorts, and downloading data.
- **Research on AI development** including algorithms as well as techniques for digital reference phantoms, computer annotating & crowd sourcing for "truth", NLP of radiology reports, and virtual clinical trials.
- **Sequestered data commons** of diverse cases that will be used to independently evaluate AI algorithms for specific claims, clinical tasks, and populations (while retaining the integrity of the test set).
- **Public website** for **dissemination of information** on metrics of evaluation, task-based distribution methods, related software, and white papers such as ones on explainability of AI.
- Mechanism for hosting of **Grand Challenges.**
- **Welcoming scientific community** for collaboration with other societies, institutes, and organizations.

# Summary: MIDRC is more than just a data registry

- A **high-quality and diverse data commons** led by **major imaging societies** enabling researchers to address topics no single archive could yield independently.

Including:

- **User-friendly open output portal** for browsing/exploring the data, creating cohorts, and downloading data.
- **Research on AI development** including algorithms as well as techniques for digital reference phantoms, computer annotating & crowd sourcing for "truth", NLP of radiology reports, and virtual clinical trials.
- **Sequestered data commons** of diverse cases that will be used to independently eva
ret
- **Pul** sed
dist
explainability of AI.
- Mechanism for hosting of **Grand Challenges.**
- **Welcoming scientific community** for collaboration with other societies, institutes, and organizations.

Thank you, please donate imaging studies, and let's work together