May 19, 2018

Defining Use Case Standards For Diagnostic Radiology Artificial Intelligence Algorithms

Bibb Allen, MD FACR
Chief Medical Officer
ACR Data Science Institute

Keith Dreyer, DO PhD FACR
Chief Science Officer
ACR Data Science Institute

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

Authors: Affiliations And Conflicts

Bibb Allen, MD FACR
Department of Radiology
Grandview Medical Center
Birmingham, Alabama
Chief Medical Officer
ACR Data Science Institute

Keith Dreyer, DO, PhD FACR FSIIM
Chief Data Science Officer, MGH, BWH, Partners Healthcare
Associate Professor of Radiology, Harvard Medical School
Boston, Massachusetts
Chief Science Officer
ACR Data Science Institute
ACR Board of Chancellors, Chair Informatics Commission

No Commercial Conflicts Of Interest

Neither the authors nor their immediate families have a financial relationship with a commercial organization that may have a direct or indirect interest in the content of this presentation

Standardized Use Cases For Development Of AI Algorithms

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

AI Use Cases

Ensure Value
Establish Standards
Promote Interoperability
Respect Population Diversity
Reduce Unintended Bias

- Many uses cases for artificial intelligence (AI) in diagnostic radiology have been broadly and inconsistently defined with variable definitions of how algorithms will be developed, validated, adopted and monitored in clinical practice[1].

- While a number of studies have demonstrated the diagnostic performance of AI algorithms is high, developers are frequently using single institution imaging datasets for training and testing the algorithm, and the algorithm output is specifically tailored to that site's perspective of the clinical workflow. It remains to be seen whether the effectiveness of these algorithms will be generalizable to widespread clinical practice and how they will be integrated into clinical workflows across a variety of practice settings.

- The American College of Radiology's Data Science Institute has developed a standardized process for AI use case development to help achieve the goal of widespread use of clinically relevant, safe and effective AI algorithms in routine radiological practice[2].

TOUCH-AI: Technology Oriented Use Cases in Healthcare

- **ACR DSI use case authoring system**
- **Open framework**
- **Integrates with CARDS for an end to end AI use case development process**
- **Establishes common data elements to help standardize AI authoring**

- TOUCH-AI (**Technology Oriented Use Cases in Healthcare**) is an open framework authoring system for defining clinical and operational AI use cases for the radiological sciences intersect high clinical value with problems solvable by AI.

- TOUCH-AI provides a framework that includes narrative descriptions and flowcharts that specify the goals the algorithm should meet, the required clinical inputs, how it should integrate into the clinical workflow, and how it should interface with both human end-users and an array of electronic resources, such as reporting software, PACS, and electronic health records.

- Combined with CARDS (**Computer Assisted Reporting Data Science**)[3], the ACR's existing open framework for authoring and implementing computer assisted reporting tools in clinical workflows, TOUCH-AI and CARDS provide an end to end AI use case authoring tool for the development of ACR DSI Use Cases for the AI developer community.

TOUCH-AI: Common Data Elements (CDE) and RadElements

DATA SCIENCE INSTITUTE™
AMERICAN COLLEGE OF RADIOLOGY

AI Data Elements

RadElement.org

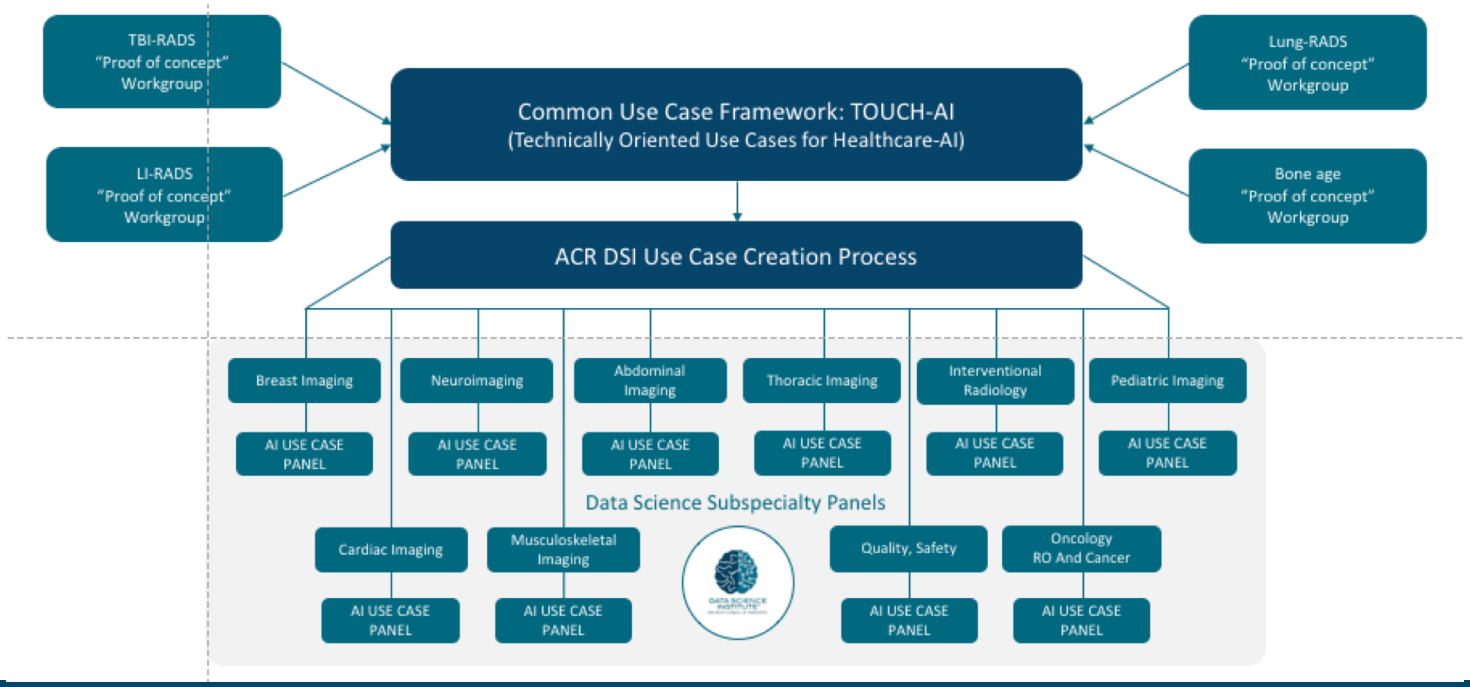
RSNA Informatics™

- Using TOUCH-AI and CARDS for AI use case development creates uniform data elements that allow standardization of:
  - Data elements for creation and annotation of datasets for algorithm testing and training
  - Data elements and statistical metrics for algorithm validation
  - Application programming interfaces (APIs) for algorithm deployment into clinical and departmental workflows
  - Data elements for monitoring the algorithm's performance in widespread clinical practice.
- Ensures patient safety with use cases that have data elements for algorithm validation pre-FDA approval and monitoring real world performance after deployment in routine clinical practice.
- Ensures use cases have data elements that will allow effective clinical integration using our workflow tools such as reporting software, the modalities, PACS and EHR.
- Creating CDEs for AI in the radiological sciences is part of a joint ACR RSNA informatics project – RadElements.[4]

ACR Data Science Institute Use Case Development: Data Science Subspecialty Panels

- ACR DSI has established **Data Science Subspecialty Panels** composed of clinical experts, many with data science backgrounds, to evaluate and choose the highest value use case proposals for development.

- The ACR DSI staff work with the panels to provide the proper data encoding to translate narrative descriptions and flowcharts of the use case into machine readable formats for AI developers.



DSI ACR Use Case Development: Nearly Endless Opportunities

- Image interpretation
  - Quantification of findings
  - Quantified comparison between multiple studies
  - Multiparametric analysis across multiple modalities
  - Volumetric analysis
  - Textural analysis
  - Automation of Region Of Interest targeting and measuring
- Patient care and safety
  - Detection and prioritization of potentially critical results
  - Radiation dose optimization
  - Pre-test probability assessment of patient risk of positive findings and contrast reactions
  - Cancer and mammography screening
  - Automatic protocoling of studies from EMR data

- Radiologist and practice optimization for productivity and quality
  - Automated transcription of audio narration
  - Automated population of structured reports
  - Optimization for case assignment across teams
  - Smarter PACS hanging protocols and synchronization protocols
  - Communication and tracking of primary and incidental findings
  - Decreased patient waiting times
  - Quality improvement in scanning
  - Prediction and prevention of missed patient appointments

Example Use Categories

- **Prioritize work lists**
- **Improving quality of reconstructed images**
- **Pre-analysis of cases to mitigate observer fatigue**
- **Extracting information from images that is not visually apparent**

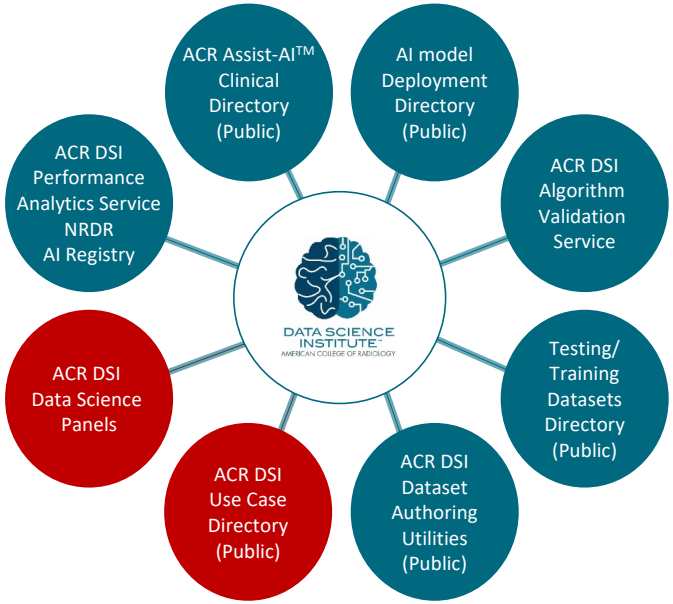
ACR DSI Use Case Development Status

- All ACR DSI Subspecialty Data Panels have started work
  - 25 use cases near completion
  - 7 use cases in review
  - 3 use case in draft stage
- Examples of use cases under development
  - Pediatric bone age classification
  - Lisfranc fracture detection and classification
  - Colon polyp detection
  - Lung-RADS classification
- Industry, institutional and FDA collaborations have been established to develop and demonstrate proof of concept prototypes.

Moving Clinically Effective AI Use Cases To Clinical Practice: Radiology AI Ecosystem

The Radiology AI Ecosystem

Ideas To Clinical Practice



Radiology's Value Proposition

- Trusted partnerships with industry and regulators
- Ensure patient safety
- Increase radiology professionals' value in healthcare

Healthcare AI Challenges

Although there has been rapid development of AI algorithms in healthcare, clinical deployment has been slow to occur

	Possible Reasons	Current Impact
1	Clinically effective uses for AI have been poorly defined	
2	No standards for clinical integration / care management	
3	Large, annotated training sets are difficult to create	
4	Currently no successful economic/business models	
5	Limitations in current AI/human UX/UI	
6	Inconsistent results and explicability between models	
7	Healthcare regulatory hurdles are challenging	
8	Resulting inference models are too brittle in practice	
9	Data science algorithms are limited for healthcare use	
10	Poor acceptance of technology in healthcare	

- ACR DSI Use Cases are designed to help mitigate these challenges so that AI use cases that benefit patients the most are readily available to AI developers.
- We envision a **Radiology AI Ecosystem** where radiology professionals collaborate with industry developers and government agencies to expeditiously bring safe, effective and clinically relevant algorithms to widespread radiology practice[5].

DSI Use Case Development: Narrative Descriptions To Machine Readable Language

DATA SCIENCE INSTITUTE™ AMERICAN COLLEGE OF RADIOLOGY

Detection of Polyps Larger than 6mm

Overview

Synonym	6mm Polyp Detection
Purpose	To detect polyps > 6 mm in size at CT colonography
Panel	TOUCH-AI ABDO
TOUCH-AI ID	<not yet defined>
Version	Draft

Clinical implementation

Value Proposition

Colorectal polyps are precursors to cancer, which can be detected with imaging and then removed at optical colonoscopy to prevent cancer development. CT colonography (CTC) provides a minimally invasive structural exam of the colon and rectum to detect clinically significant polyps, however less experienced radiologists can miss polyps and take excessive time to complete the exam. AI could help improve accuracy and efficiency of polyp detection at CTC, reduce false positives and reduce medical legal risk for radiologists.

Narrative

A 50 year old patient presents for first time colorectal screening at CT colonography. Algorithm evaluates the colon for detection of polyps larger than 6 mm in maximal diameter. The radiologist is informed of this categorization at the time of the interpretation.

Algorithm Workflow

Images obtained at CT and sent to 3D workstation/PACS and AI engine. Images are analyzed by engine. The size, number, and location of polyps larger than 6 mm are identified. A message is sent to PACS from the engine with the classification information.

Execution criteria

Age	>50 years
Procedures	CT Colonography, pelvis/abdomen, low dose screening, without contrast
Indication	Colorectal Screening
Views	Supine OR Prone OR Right lateral decubitus OR Left lateral decubitus

Dataset considerations

Age	Age ≥ 50 years AND Age ≤ 75 years
Colon Anatomy	Nonredundant AND Redundant AND Very Redundant
Colon surgery	Colon segment(s) excised ≥ 1
Inflammatory Bowel Disease:	True AND False
Crohn's	
Ulcerative Colitis	
Stool tagging	True AND False
Views	Supine AND Prone AND Right lateral decubitus AND Left lateral decubitus

Technical specifications

Inputs

Study

Definition	The image data for the study
Format	DICOM
Value Set	Binary Dataset
Units	N/A
Max Length	N/A
Is Required	Required

Outputs

Detection of polyps >6 mm

Common Data Element ID	<not yet defined>
Definition	Detection of polyps ≥6 mm
Format	Categorical
Value Set	0 – Zero polyps ≥ 6 mm 1 – One polyp ≥ 6 mm 2 – Two or more polyps ≥ 6 mm 3 – Indeterminate
Units	N/A
Max Length	1
Finding Coordinates	Point coordinates of each lesion determined
Is Required	Required

Narrative Use Case

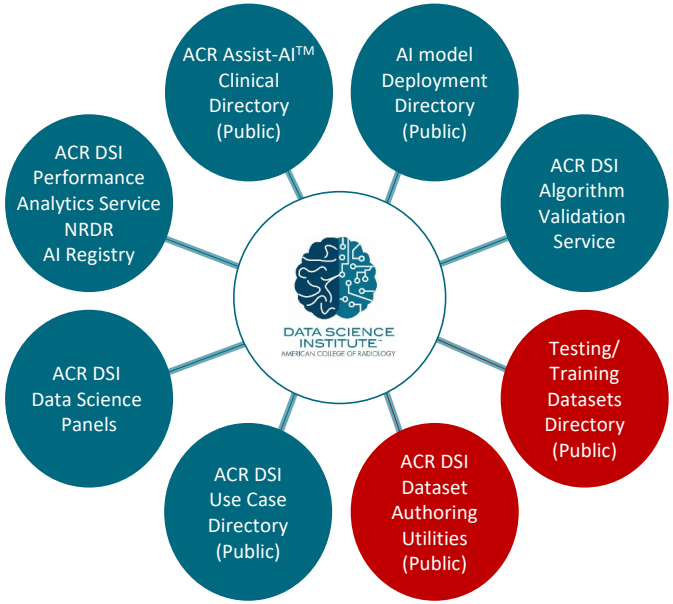
Example:
ACR DSI Use Case
Colon Polyp Detection

Machine Readable Programming Language For AI Developers

- Data elements for creation of datasets for training and testing AI algorithms.
- Data elements and statistical specifications for creation of datasets for algorithm validation prior to FDA review.
- Data elements and specifications for application programming interfaces (API) for deployment in clinical and operational workflows.
- Data elements for populating registries using real word evidence to monitor algorithm performance in clinical practice and provide feedback to developers and regulatory agencies.

Standard Specifications For Data Access

Making Datasets For AI Training Available To Developers



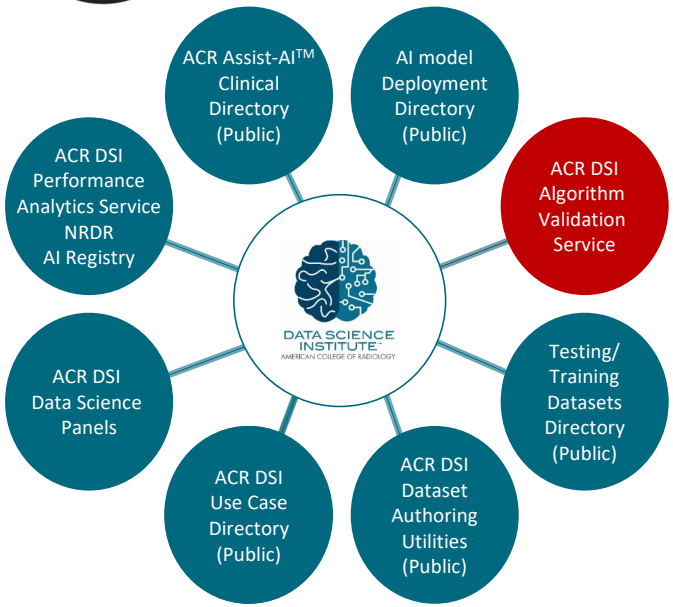
ACR DSI Data Access Directory

Specifications For Data Access

- Standardized definitions and data elements allow multiple institutions to use these standards to create datasets that developers can use for algorithm training and testing.
- Specifications include standardized tools and methods for image annotation.
- Using multiple sites as data sources for these datasets provides technical, geographic and patient diversity to prevent unintended bias in algorithm development.
- Allows more individuals and institutions to participate in AI development.
- The ACR DSI will house a freely available public directory of institutions that have created these datasets around ACR DSI Use Cases to inform the developer community.

Validating AI Algorithms

ACR Data Science Institute Certified Algorithms

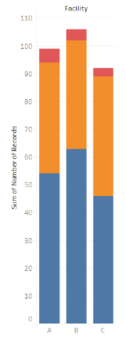


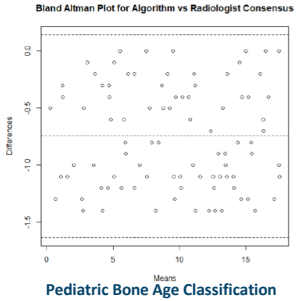
Specialty society certification of AI algorithms provides an "honest broker" partnership with radiology, developers and government regulators.

Specifications For Algorithm Validation

- Centralized assessment of algorithm performance will be performedaccording to the statistical metrics metrics specified in the use case using embargoed datasets.
- These validation datasets are created at multiple institutions to ensure geographic, technical and patient diversity within the validation dataset.
- Multiple readers and guidelines for data quality to ensure "ground truth" consistency between sites, consistent metrics for measuring performance across sites and standards to protect developers' intellectual property, ensure patient privacy and diminish bias.
- Reports are generated for developers to use in the FDA pre-market process.

Validation Dataset:	Facility A	Facility B	Facility C	Overall
Sample Size	99	107	91	297
% Exact Match	54%	59%	50%	55%
% within 6 months	40%	37%	47%	41%
% within 1 year	5%	4%	3%	4%
75th percentile	6 months	6 months	6 months	6 months
95th percentile	12 months	12 months	18 months	12 months
Kappa	.53	.58	.49	.54



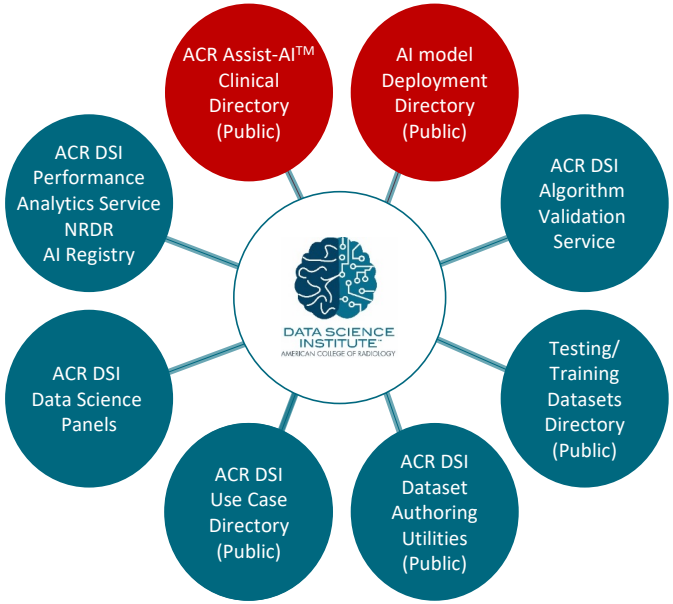


Pediatric Bone Age Classification ACR DSI Performance Report

Clinical Deployment

ACR Assist™

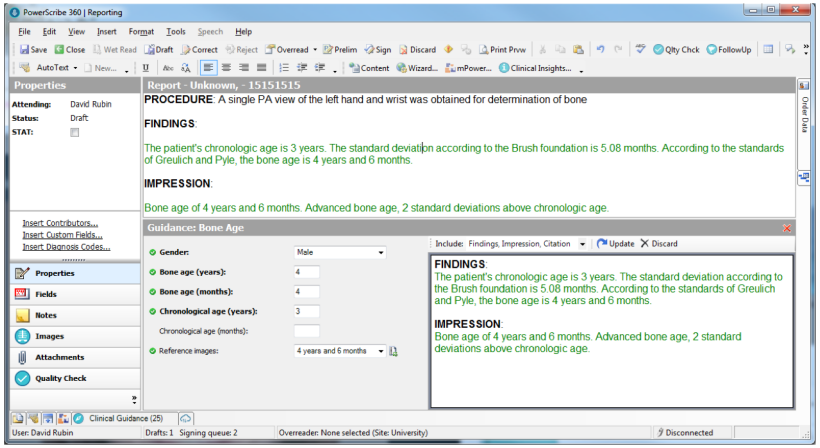
Modifications to the CARDS platform allows ACR Assist to capture output from AI algorithms and integrate it into clinical and departmental workflows.

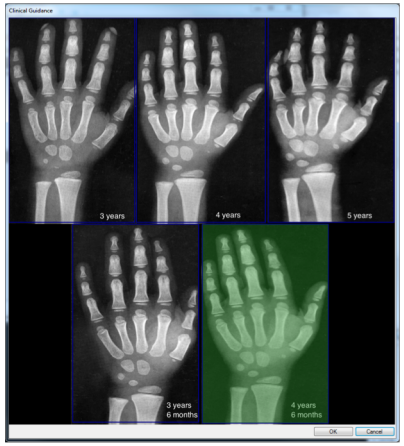


Vendor neutral output from AI algorithms incorporated into existing HIT tools including reporting software, the modalities and the EHR.

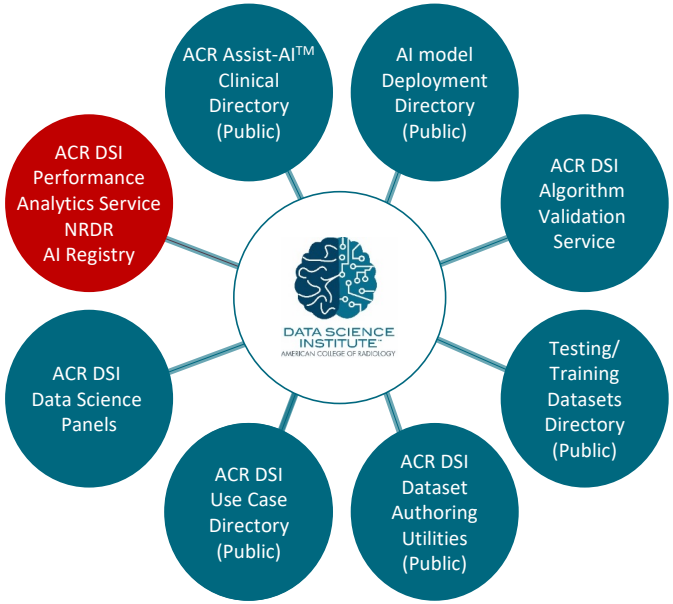
Specifications For Clinical Integration

- Data elements and specifications for application programming interfaces (API) allow algorithm output to be deployed in a vendor neutral environment for deployment in clinical and operational workflows.
- In this demonstration, the standardized output from a pediatric bone age algorithm is incorporated into reporting software, and saliency maps are shown ensuring algorithm transparency.





Monitoring Algorithm Performance In Clinical Practice



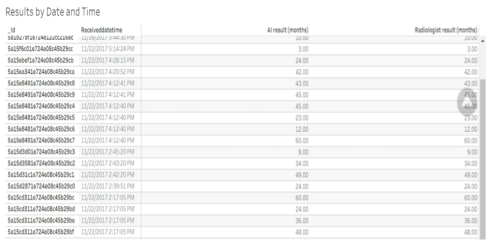
AI Monitoring Program

- Patient safety and FDA surveillance
- Algorithm transparency and radiologist acceptance
- Developer improvements

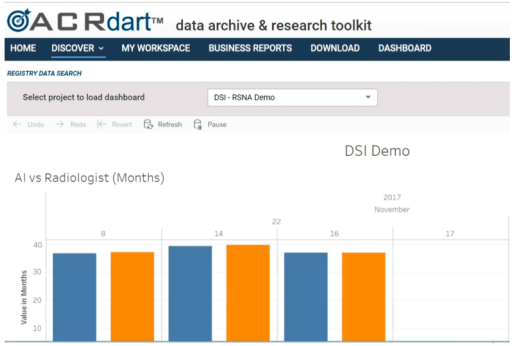
Specifications For Monitoring In Clinical Practice

- Data elements in each use case specify how the algorithm will be monitored in clinical practice.
- Radiologist input is gathered as the case is being reported, and if the radiologist does not incorporate the algorithm inferences into the report, this change is captured in the background by the reporting software. If the radiologists agrees changes the output of the agrees with algorithm, this is also noted and transmitted to the registry.
- Specified metadata about the exam such as equipment vendor, slice thickness and exposure exposure are also transmitted to the registry.
- Algorithm assessment reports include algorithm performance metrics and the exam parameters affecting the algorithms' performance.
- These reports are used by the developers to report to the FDA and for algorithm improvement.

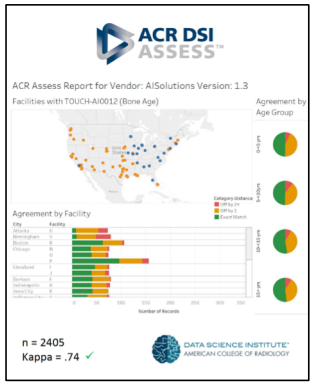
Monitoring Algorithm Performance In Clinical Practice



Raw Data Captured In Reporting Software And Transmitted to an AI Registry



Raw Data Captured In Reporting Software And Transmitted to an AI registry

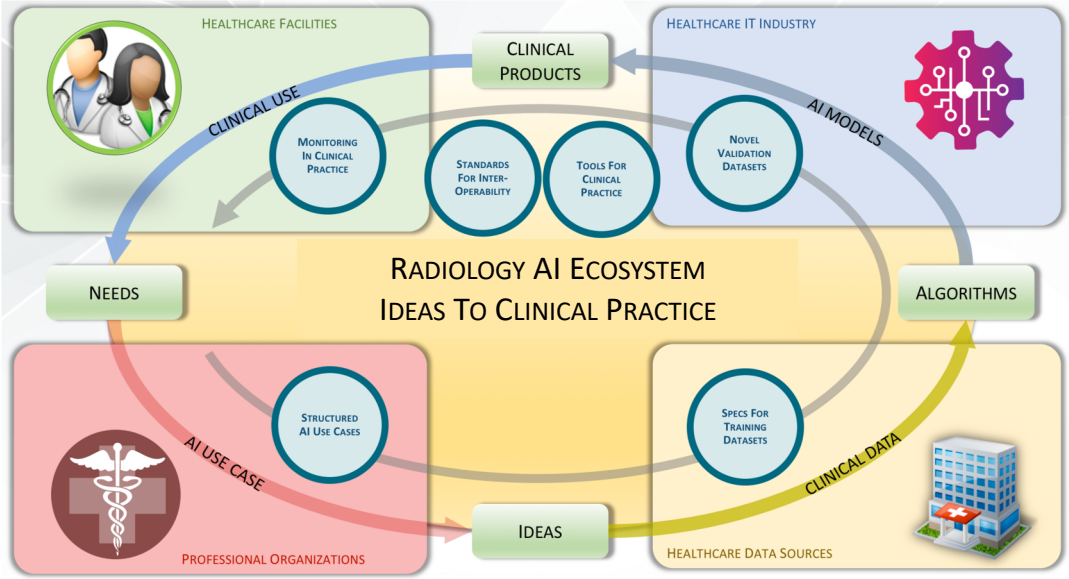


AI Algorithm Assessment Reports For Each Developer And Site

Working Example of Monitoring Algorithm Performance Using An AI Data Registry

- This example is from a pediatric bone age classification algorithm. The reporting software transmits information about the radiologist's agreement or disagreement with the algorithm along metadata about the examination to the AI data registry.
- The raw data are complied in the registry and reports are aggregate and developer specific generated for developers for use in FDA post-market surveillance reports and to improve algorithm performance.
- Site reports are provided to provide AI performance metrics to the clinical practices.

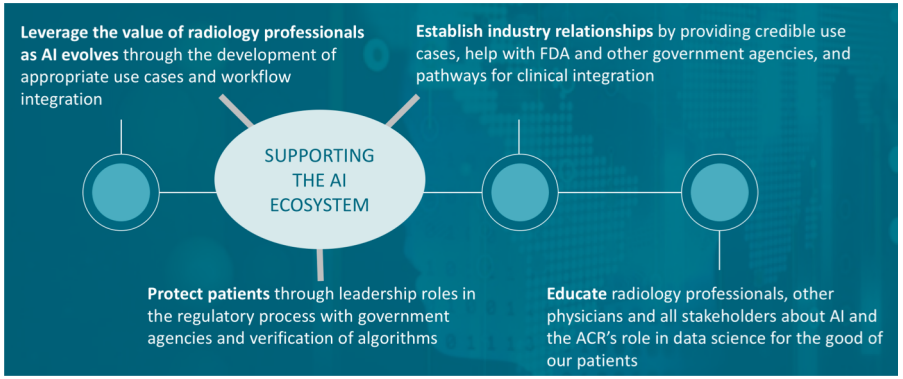
Conclusions



Development Of Robust AI Use Cases Can Mitigate The Challenges Of Bringing AI Algorithms To Routine Clinical Practice

- Professional organizations are best suited to bring ideas for AI algorithms and clinical needs of the healthcare community together.
- Standardized methods to annotate and aggregate data for AI model training and testing will improve model performance and mitigate bias.
- Standardized methods for AI model validation will enhance the regulatory process.
- Mechanisms to integrate and monitor AI models in clinical practice using real world experience will ensure patient safety, enhance algorithm transparency, inform the regulatory process and improve algorithm performance.

ACR Data Science Institute Missions



ACR Data Science Institute Tools To Support The Radiology AI Ecosystem

References:

1. Thrall JH, Li X, Li Q, Cruz C, Do S, Dreyer K, Brink J. Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. Journal of the American College of Radiology. 2018 Mar 1;15(3):504-8.
2. ACR Data Science Institute Use Case Development. http://www.acrdsi.org/use-case-development.html Accessed 5/12/2018
3. Alkasab TK, Bizzo BC, Berland LL, Nair S, Pandharipande PV, Harvey HB. Creation of an open framework for point-of-care computer-assisted reporting and decision support tools for radiologists. Journal of the American College of Radiology. 2017 Sep 1;14(9):1184-9.
4. RadElements. http://www.radelement.org Accessed 5/12/2018.
5. Allen B, Dreyer K. The Artificial Intelligence Ecosystem for the Radiological Sciences: Ideas to Clinical Practice. Journal of the American College of Radiology. 2018 In Press.