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
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.
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.

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.
• 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), 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.
• 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.
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.

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
The Radiology AI Ecosystem

Ideas To Clinical Practice

- ACR DSI Data Science Panels
- ACR DSI Use Case Directory (Public)
- ACR DSI Dataset Authoring Utilities (Public)
- ACR DSI Algorithm Validation Service
- Testing/Training Datasets Directory (Public)
- ACR DSI Performance Analytics Service NRDR AI Registry
- ACR DSI Use Case Directory (Public)

Radiology’s Value Proposition

- Trusted partnerships with industry and regulators
- Ensure patient safety
- Increase radiology professionals’ value 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.

Healthcare AI Challenges

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

Although there has been rapid development of AI algorithms in healthcare, clinical deployment has been slow to occur.
**Detection of Polyps Larger than 6mm**

**Overview**
- **Synonym**: 6mm Polyp Detection
- **Purpose**: To detect polyps > 6 mm in size at CT colonography
- **Panel**: TOUCHE
- **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. The algorithm evaluates the colon for 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**: Crohn’s AND Ulcerative Colitis True AND False
- **Stool tagging**: True AND False
- **Views**: Supine AND Prone AND Right lateral decubitus AND Left lateral decubitus

**Technical specifications**

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Outputs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study</td>
<td>Detection of polyps &gt;6mm</td>
</tr>
<tr>
<td>Number</td>
<td>Disease Status, Polyp Count, Polyp Size, Polyp Location</td>
</tr>
<tr>
<td>Number of Polyps</td>
<td>Number of Polyps</td>
</tr>
<tr>
<td>Size</td>
<td>Size, Number, Location of Polyps</td>
</tr>
</tbody>
</table>

**Finding coordinates**
- Point coordinates of each polyp are determined

**Is Required**
- Required

**Example:**
**ACR DSI Use Case**

**Colon Polyp Detection**

**DSI USE CASE DEVELOPMENT: NARRATIVE DESCRIPTIONS TO MACHINE READABLE LANGUAGE**

**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 world 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

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.
Specifications For Algorithm Validation

- Centralized assessment of algorithm performance will be performed according 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.

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

**AI Monitoring Program**
- Patient safety and FDA surveillance
- Algorithm transparency and radiologist acceptance
- Developer improvements
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.
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

- Leverage the value of radiology professionals as AI evolves through the development of appropriate use cases and workflow integration
- Establish industry relationships by providing credible use cases, help with FDA and other government agencies, and pathways for clinical integration
- Protect patients through leadership roles in the regulatory process with government agencies and verification of algorithms
- Educate radiology professionals, other physicians and all stakeholders about AI and the ACR's role in data science for the good of our patients

ACR Data Science Institute Tools To Support The Radiology AI Ecosystem

References: