IMAGING AI CRITICAL SUCCESS FACTORS

• Addressing the “Right” Problem
• Verifying Safety and Efficacy in a high-stakes environment
• Integrating into the Clinical Workflow
• Monitoring, Adapting, and Communicating Results
BACKGROUND
The Role of the ACR

• Founded in 1924, the American College of Radiology has been at the forefront of radiology evolution

• More than 38,000 radiologists, radiation oncologists, nuclear medicine physicians and medical physicists.

• Core Purpose:

  To serve patients and society by empowering members to advance the practice, science and professions of radiological care.
AI and Next Generation Technology

• The ACR Data Science Institute established May 2017

• Core Purpose:

  ACR Data Science Institute (DSI) empowers the advancement, validation, and implementation of artificial intelligence in medical imaging and the radiological sciences for the benefit of our patients, society, and the profession
REGULATORY CONSIDERATIONS (FDA)

• Objectives
  • Protect the public health
  • Help speed safe and effective innovation

• Medical Device Classification
  • Based on Risk
  • Based on Intended Use (what does your label say)
  • Based on Indications for Use (under what conditions will the product be used)

<table>
<thead>
<tr>
<th>Class</th>
<th>LOW RISK</th>
<th>HIGH RISK</th>
</tr>
</thead>
<tbody>
<tr>
<td>Class I</td>
<td>General Controls</td>
<td>General Controls + Pre Market Approval</td>
</tr>
<tr>
<td>Class II</td>
<td>General Controls + Special Controls</td>
<td>General Controls + Special Controls</td>
</tr>
<tr>
<td>Class III</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Where Does AI fall?

• CADe - Detection
  
  Devices intended to identify, mark, highlight, or in any other manner direct attention to portions of an image, or aspects of radiology device data, that may reveal specific abnormalities during interpretation of patient radiology images or patient radiology device data by the clinician

• CADx – Diagnosis
  
  Devices go beyond CADe and include those that are intended to provide an assessment of disease or other conditions in terms of the likelihood of the presence or absence of disease, or are intended to specify disease type (i.e., specific diagnosis or differential diagnosis), severity, stage, or intervention recommended

• 9/17 – Ruling classified CADx with AI as Class II. Vendors with similar products can apply for 510k clearance and avoid Pre-Market Approval (PMA)
Opportunities to Accelerate the Process

• Software as a Medical Device (SaMD)
  • 21st Century Cures Act provides guidance of medical device software
  • FDA is developing guidance for implementation

• Medical Device Development Tools
  • Promotes innovation in medical device development and regulatory science to help bridge the gap between research of medical devices and the delivery of devices to patients.

• National Evaluation System For Health technology (NEST)
  • Intended to shorten the time to market for new technology health care products by developing a system for more robust post-market surveillance
Establishing NEST Will Enable The Pre-Post Market Shift

Graphic courtesy of Greg Pappas, Assistant Director FDA NEST
NEST program chooses ACR AI use case as demonstration project

February 02, 2018 | Nicholas Leider

The National Evaluation System for Health Technology (NEST) Coordinating Center announced Feb. 2 it had selected a use case from the American College of Radiology (ACR) Data Science Institute (DSI) as one of its first demonstration projects.

NEST chose “Lung-RADS Assist: Advanced Radiology Guidance, Reporting and Monitoring” to determine workflow via artificial intelligence (AI).
**LUNG CANCER**

Leading cause of cancer related deaths in men and women:
- 1.59 Million worldwide \(^{(2012)}\)
- 158,000 United States \(^{(2016)}\)
- 75% present symptomatically with incurable disease

**USPSTF RECOMMENDATION**

Annual screening for lung cancer with low-dose computed tomography (LDCT) in adults **aged 55 to 80 years** who have a **30 pack-year smoking history** and currently smoke or have quit within the past 15 years.

**UNITED STATES ELIGIBLE POPULATION**

20 Million individuals require annual screening
Lung Nodule Detection Algorithms

1. Data Management
2. Machine Learning
3. Clinical Validation
4. AI Model

Data Acquisition | Ground Truth | Algorithm Training | Clinical Validation

Nodule – Description

Inferencing
Lung Nodule Detection Algorithms

MODEL 1

Data Acquisition | Ground Truth | Training | Validation

MODEL 2

Data Acquisition | Ground Truth | Training | Validation

MODEL 3

Data Acquisition | Ground Truth | Training | Validation

Inferencing
VALIDATION, CERTIFICATION & COMPLIANCE
Challenges in the AI Life Cycle

- How generalizable is the inference model?
- Is there hidden sample bias?
- What is the appropriate threshold for clinical use?
- How do we ensure ongoing performance?
- How robust is the model to changes in the environment?
Challenges in the AI Life Cycle

- Do models solving the same problem yield consistent, comparable outputs?
- Does the customer understand potential differences in the implicit use cases?
- How do we establish standard, consistent performance metrics?
Establish Standards & Certification Criteria

- Establish common expectations for addressing specific clinical scenarios (e.g. BI-RADS)
- Create well-qualified data sets that address explicit concerns about bias
- Define standard performance metrics that establish a quality threshold
- Validate models that address a specific clinical condition against these standards
Monitoring and Communication

- Monitor Ongoing Performance to Ensure Ongoing Quality and Safety
- Provide Feedback Loop to Providers, Regulators, Vendors, Content Creators
- Match continuous learning with continuous assessment, monitoring, and feedback
TOUCH-AI
Detecting Lisfranc Joint Injury

Lisfranc joint injury is common and easily missed. AI that segments and detects abnormality would prove valuable and help reduce false negative rate, patient risk, and medical-legal risk for the radiologists.
DSI Use Cases Clinical Guidance for Developers

Example: Lisfranc Joint Injury

Expected Clinical Inputs/Outputs

Conditions for launch

Data Considerations for Training/Testing
Use Case Development Status

• All ACR DSI Subspecialty Data Panels underway
  - 19 Use Cases in drafting stage
  - 9 Use Cases in the review stage

• Examples of use cases under development
  - Pediatric Bone Age classification
  - Lisfranc fracture detection and classification
  - Colon polyp detection
  - TBI-RADS

• Industry collaborations
LungRads Assist - Demonstration Project
<table>
<thead>
<tr>
<th>Category Descriptor</th>
<th>Category</th>
<th>Findings</th>
<th>Management</th>
<th>Probability of Malignancy</th>
<th>Population Prevalence</th>
</tr>
</thead>
</table>
| Probably benign finding(s) - short term follow up suggested; includes nodules with a low likelihood of becoming a clinically active cancer | 3 | solid nodule(s):  
   - ≥ 6 to < 8 mm at baseline OR  
   - new 4 mm to < 6 mm  
   - part solid nodule(s):  
   - ≥ 6 mm total diameter with solid component < 6 mm OR  
   - new < 6 mm total diameter  
   - non solid nodule(s) (GGN) ≥ 20 mm on baseline CT or new | 6 month LDCT | 1-2% | 5% |
| Findings for which additional diagnostic testing and/or tissue sampling is recommended | 4A | solid nodule(s):  
   - ≥ 8 to < 15 mm at baseline OR  
   - growing < 8 mm OR  
   - new 6 to < 8 mm  
   - part solid nodule(s):  
   - ≥ 6 mm with solid component ≥ 6 mm to < 8 mm OR  
   - with a new or growing < 4 mm solid component  
   - endobronchial nodule  
   - solid nodule(s):  
   - ≥ 15 mm OR | 3 month LDCT; PET/CT may be used when there is a ≥ 8 mm solid component | 5-15% | 2% |

Prior Lung Cancer  
Consider the presence with a prior diagnosis of lung cancer who return to screening  
C  
modifier - may add on to category 0-4 coding  
-  
-  
-
LungRads Assist - Demonstration Project
Certification Data Sets (e.g. LDCT for Lung Screening)

- **Inclusion/Exclusion Criteria**
- **Sample Size** (number of cases, % positive)
- **Data Dictionary**
- **Dataset Stratifications**
- **Annotation**

### Data Collection

#### Inclusion criteria:
- Performed for lung cancer screening
- Low dose CT scanning technique
- Full inspiration study
- Patient weight < 90 kg (to avoid excess noise or artifacts)
- 1-1.25 mm in section thickness
- Any CT vendor or equipment
- Pathology proof of diagnosis (cancer type) for the Lung RADS 3 and/or 4.
- Follow up LDCT for non-biopsied nodules with stability of nodules or resolution of nodules when m is not met

#### Exclusion criteria:
- Motion artifacts
- Metal hardware
- Confounding findings: LDCT must not have diffuse lung disease or other abnormalities apart from nodules, or smoking related features (emphysema, bronchial wall thickening)

### Data Dictionary

**a. Per patient:**
- Patient weight
- Smoking history
- Presence of nodule (y/n)
- Number of nodules (integer)

**b. Per nodule:**
- Image number for each nodule
- Location (side and lobe; parenchymal, fissural, and endobronchial)
- Attenuation (solid, subsolid, part-solid, calcified (pattern), cavitary, cystic)
- Margins
- Size (maximum, minimum and average size in mm)
- Lung RADS category
- Pathology proof of diagnosis (cancer type) for the Lung RADS 3 and/or 4.

### Image Mark-up

- **a. Location**
- **b. Margins**
- **c. Size (maximum, minimum)**

### Criteria for establishing ground truth

**a. Detection** - Controlled reader study

**b. Size** - controlled reader study

**c. Lung RADS category**
- 50% Lung RADS 1
- 30% Lung RADS 2
- 30% Lung RADS 3
- 20% Lung RADS 4A
- 20% Lung RADS 4B
LungRads Assist - Demonstration Project

- LungRADS Use Case(s)
  - Vendor 1
  - Vendor 2
  - Vendor 3
- Certification Data Set
- Certify Model
- Assess Performance
## Threshold Considerations for Certification

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Examples</th>
<th>Eval Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification</td>
<td>*RADS, Nodule Type</td>
<td>AUC, logloss, MeanFScore</td>
</tr>
<tr>
<td>Segmentation</td>
<td>Nodule or organ seg.</td>
<td>DICE Coefficient</td>
</tr>
<tr>
<td>Estimation</td>
<td>Nodule Size, #, midline Shift</td>
<td>RMSE, RMSLE, NWRMSLE</td>
</tr>
<tr>
<td>Location</td>
<td>Nodule Detection</td>
<td>Dice Coefficient</td>
</tr>
</tbody>
</table>

### Use Case Evaluation Results

<table>
<thead>
<tr>
<th>Use Case</th>
<th>Evaluation Method</th>
<th>Possible Evaluation Outcome</th>
<th>Certified Use (FDA)</th>
<th>Possible Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location of nodule</td>
<td>Dice Coefficient</td>
<td>.90</td>
<td>Detection</td>
<td>Pass</td>
</tr>
<tr>
<td>Size of nodule</td>
<td>RMSE</td>
<td>5.6%</td>
<td>Detection</td>
<td>Pass</td>
</tr>
<tr>
<td>Attenuation of nodule</td>
<td>ROC AUC</td>
<td>.85</td>
<td>Detection</td>
<td>Pass</td>
</tr>
<tr>
<td>Lung-RADS category</td>
<td>ROC AUC</td>
<td>.80</td>
<td>Detection</td>
<td>Pass</td>
</tr>
</tbody>
</table>

### Clinical Use Risk Assessment

<table>
<thead>
<tr>
<th>Clinical Use</th>
<th>Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prioritization in Work list</td>
<td>Low</td>
</tr>
<tr>
<td>Detection and Classification</td>
<td>Med</td>
</tr>
<tr>
<td>Diagnosis</td>
<td>High</td>
</tr>
</tbody>
</table>
LungRads Assist - Demonstration Project
Monitoring and Feedback

ACR Assess Report for Vendor: AISolutions Version: 1.3

Agreement by Facility

<table>
<thead>
<tr>
<th>City</th>
<th>Facility</th>
<th>Agreement</th>
<th>Number of Records</th>
</tr>
</thead>
<tbody>
<tr>
<td>Atlanta</td>
<td>U</td>
<td>Exact Match</td>
<td>0</td>
</tr>
<tr>
<td>Birmingham</td>
<td>V</td>
<td>Exact Match</td>
<td>0</td>
</tr>
<tr>
<td>Boston</td>
<td>B</td>
<td>Exact Match</td>
<td>0</td>
</tr>
<tr>
<td>Chicago</td>
<td>N</td>
<td>Exact Match</td>
<td>0</td>
</tr>
<tr>
<td>Cleveland</td>
<td>I</td>
<td>Exact Match</td>
<td>0</td>
</tr>
<tr>
<td>Durham</td>
<td>F</td>
<td>Exact Match</td>
<td>0</td>
</tr>
<tr>
<td>Indianapolis</td>
<td>H</td>
<td>Exact Match</td>
<td>0</td>
</tr>
<tr>
<td>Iowa City</td>
<td>K</td>
<td>Exact Match</td>
<td>0</td>
</tr>
</tbody>
</table>

Agreement by Age Group

- Off by 2+
- Off by 1
- Exact Match

Category distance

n = 2405
Kappa = .74
WORKFLOW INTEGRATION
AI Opportunities Across the Imaging Life Cycle

- Imaging Order
- Scheduling
- Population Health
- Protocol
- Communication
- Image Acquisition
- Report Generation
- Assessment

Optimizing Patient Care

Business and Operations (e.g. worklist optimization)
**Features:** The elements of a described lesion will be used to determine the output of the algorithm. Includes synonyms of those features that might be used in reports.

**Decision Tree:** The logic which determines the output of the algorithm based on a lesion’s features.

**End Points:** Templates of the generated text to be inserted into the body, impression, and recommendations of reports.
**FINDINGS**

Pulmonary mass in the right upper lobe measuring 3 cm. A separate nodular mass is seen in the same lobe. No atelectasis/obstructive pneumonia is seen. There is no imaging evidence of bronchial involvement or invasion of local structures. No enlarged lymph nodes are seen.

**IMPRESSION**

Pulmonary mass in the right upper lobe measuring 3 cm is concerning for neoplasm. Histologic confirmation or short-term follow-up chest CT is recommended. Separate nodular mass in the same lobe, which could represent either metastasis or a metasynchronous primary. No bronchial involvement is seen. Mass is surrounded by lung or visceral pleura. No enlarged lymph nodes are seen, though this does not exclude nodal disease. No imaging evidence of intrathoracic metastasis is seen.
TOUCH-AI Use Case: Pediatric Bone Age

1. Define TOUCH-AI Use Case
2. Collect Training Data Set
3. Create Inference Model
4. Login to Workflow System, Select Imaging Study, Select Inference engine (Nuance)
5. Select Image and Submit to cloud service (Nuance, NVIDIA)
6. Open Reporting tool (Nuance PowerScribe)
7. Retrieve AI results and populate ACR Assist template
8. Review and Approve Report
9. Populate Review ACR Registry
ACR Annual Conference on Quality and Safety

Strategies for Continuously Improving Quality in Radiology

Oct. 13–14, 2017 | Westin Boston Waterfront Hotel

Stock Your Quality Toolbox

Earn CME, SAM, Category A and RLI Credits

Register today for best rates
Bring your team and SAVE!

acr.org/qualitymeeting

2018 Combined Quality and Safety And Artificial Intelligence Meeting

• Practicing Physicians
• Radiology Informaticians
• Developers
Thank You!