THE AI LIFECYCLE
EVALUATING AND MONITORING AI ALGORITHMS

Evolving Role of AI in Radiological Imaging
FDA Public Workshop
Washington, DC
February 25, 2020

Keith J. Dreyer DO, PhD, FACR
Laura P. Coombs, PhD

MGH & BWH Center for Clinical Data Science
Massachusetts General Hospital
Brigham and Women's Hospital
Harvard Medical School

Data Science Institute
American College of Radiology
**PRESENTER AFFILIATIONS**

**KEITH J. DREYER, DO, PhD, FACR**
- **Massachusetts General Brigham**
  - Chief Data Science Officer
  - Chief Imaging Information Officer
  - Vice Chairman of Radiology
  - Founding Director, CCDS
- **Harvard Medical School**
  - Associate Professor of Radiology
- **American College of Radiology**
  - Board of Chancellors
  - DSI - Chief Science Officer
- **No Commercial Conflicts**

**LAURA P. COOMBS, PhD**
- **American College of Radiology**
  - VP of Data Science and Informatics
  - Data Science Institute
  - Data Scientist
- **No Commercial Conflicts**
THE AI LIFECYCLE

DATA
- Use Case Definition
- Cohort Creation
- Data Preparation

MODELING
- Model Architecture
- Parameter Configuration
- Train, Test, Validate Cycle

REGULATORY
- Technical Testing
- Clinical Testing
- FDA Clearance
THE AI LIFECYCLE

DATA
- USE CASE DEFINITION
- COHORT CREATION
- DATA PREPARATION

MODELING
- MODEL ARCHITECTURE
- PARAMETER CONFIGURATION
- TRAIN, TEST, VALIDATE CYCLE

REGULATORY
- TECHNICAL TESTING
- CLINICAL TESTING
- FDA CLEARANCE

MARKET
- MULTISITE (DISTRIBUTED) VALIDATION
- SITE SHADOW VALIDATION
- SITE CLINICAL DEPLOYMENT

MAINTENANCE
- CONTINUOUS (REAL WORLD) MONITORING
- ACTIVE LEARNING
- UPGRADE & REPLACEMENT
The AI Lifecycle

Data
- Use Case Definition
- Cohort Creation
- Data Preparation

Modeling
- Model Architecture
- Parameter Configuration
- Train, Test, Validate Cycle

Regulatory
- Technical Testing
- Clinical Testing
- FDA Clearance

Market
- Multisite (Distributed) Validation
- Site Shadow Validation
- Site Clinical Deployment

Maintenance
- Continuous (Real World) Monitoring
- Active Learning
- Upgrade & Replacement
Define Use Case

Clinical Standard Use Case:

Incidental Pulmonary Nodules on Chest Radiograph

Purpose
Detect and characterize of incidental pulmonary nodules on chest radiographs (CR). These are nodules that are detected on CRs performed for reasons other than lung cancer screening.

Tag(s)
Thesauric

Panel
Diagnostic Imaging

Clinical vs. Standard Use Case:

Custom created:
P(Lung Cancer) = 0.45

Clinical Implementation

Lung cancer, the leading cause of cancer-related death in both women and men, frequently presents as a pulmonary nodule on chest radiographs (CR) or CT scans. While CT is utilized for lung cancer screening, chest radiography, among the most highly utilized diagnostic imaging procedures worldwide, is the most common thoracic imaging study in which incidental lung cancer is discovered. Nonetheless, interpretation of chest radiography is challenging and prone to many reading errors. Thus, nodules are frequently misread on CRs, with studies showing approximately 20% of nodules seen only in retrospect. The causes for these frequent errors are multifaceted, including overlapping anatomic structures such as ribs, clavicles, thoracic spine, pulmonary vessels, heart, mediastinum, and diaphragm; errors in visual search; lesion recognition or decision-making; and subtitles image quality. Small, indistinct nodules with low attenuation and conspicuity are particularly susceptible to being overlooked. At early stage of lung cancer, tumor morphology, initiation or delayed diagnosis due to these limitations may negatively impact patient survival.

Considerations for Dataset Development

- Procedures: CR, DR, dual-energy, and bone suppression CRs
- Views: PA, lateral, AG, axial, oblique, oblique
- Age
  - 1-20 years old
- Sex at birth: Male, Female
- Nodule Validation
  - CR within 1 month of CR. Corresponding nodule location on CR confirmed by chest radiographers.
- Nodule attenuation based on CT confirmation
  - 10 HU
- Size (in mm)
  - ≥ 5 mm
- Shape: round, oval, triangular, lobular, irregular
- Margin: smooth, irregular, spiculated
- Location: broad sampling of lung regions, apex to base, central to peripheral
- Comorbidity
  - Smokers, non-smokers, COPD, travel/local history, other primary malignancy or history of primary malignancy, bronchitis, bronchiolitis, pneumonitis, tuberculosis, fungal and other pulmonary infections, focal inflammatory lesions, unifocal interstitial pneumonia and other diffuse lung diseases, pleural effusions
- Other Considerations
  - Review of CR technologist (CR, DR, dual-energy, bone)
The AI Lifecycle

Data
- Use Case Definition
- Cohort Creation
- Data Preparation

Modeling
- Model Architecture
- Parameter Configuration
- Train, Test, Validate Cycle

Regulatory
- Technical Testing
- Clinical Testing
- FDA Clearance

Market
- Multisite (Distributed) Validation
- Site Shadow Validation
- Site Clinical Deployment

Maintenance
- Continuous (Real World) Monitoring
- Active Learning
- Upgrade & Replacement

The AI Lifecycle Diagram
Create Cohort

• Representative of the target population
  • Disease Prevalence
  • Selection Bias or Spectrum Effect
Create Cohort

- Representative of the target population
- **Disease Prevalence**
- Selection Bias or Spectrum Effect

Disease prevalence = 50%
Create Cohort

• Representative of the target population
  • Disease Prevalence
  • Selection Bias or Spectrum Effect
THE AI LIFECYCLE

DATA
- USE CASE DEFINITION
- COHORT CREATION
- DATA PREPARATION

MODELING
- MODEL ARCHITECTURE
- PARAMETER CONFIGURATION
- TRAIN, TEST, VALIDATE CYCLE

REGULATORY
- TECHNICAL TESTING
- CLINICAL TESTING
- FDA CLEARANCE

MARKET
- MULTISITE (DISTRIBUTED) VALIDATION
- SITE SHADOW VALIDATION
- SITE CLINICAL DEPLOYMENT

MAINTENANCE
- CONTINUOUS (REAL WORLD) MONITORING
- ACTIVE LEARNING
- UPGRADE & REPLACEMENT
THE AI LIFECYCLE

Data
- Use Case Definition
- Cohort Creation
- Data Preparation

Modeling
- Model Architecture
- Parameter Configuration
- Train, Test, Validate Cycle

Regulatory
- Technical Testing
- Clinical Testing
- FDA Clearance

Market
- Multisite (Distributed) Validation
- Site Shadow Validation
- Site Clinical Deployment

Maintenance
- Continuous (Real World) Monitoring
- Active Learning
- Upgrade & Replacement

Clinic
- Deployment
- Continuous Monitoring
- Active Learning
- Upgrade & Replacement
Centralized model building, federated model shows higher accuracy
THE AI LIFECYCLE

DATA
- USE CASE DEFINITION
- COHORT CREATION
- DATA PREPARATION

MODELING
- MODEL ARCHITECTURE
- PARAMETER CONFIGURATION
- TRAIN, TEST, VALIDATE CYCLE

REGULATORY
- TECHNICAL TESTING
- CLINICAL TESTING
- FDA CLEARANCE

MARKET
- MULTI-SITE (DISTRIBUTED) VALIDATION
- SITE SHADOW VALIDATION
- SITE CLINICAL DEPLOYMENT

MAINTENANCE
- CONTINUOUS (REAL WORLD) MONITORING
- ACTIVE LEARNING
- UPGRADE & REPLACEMENT

MAINTENANCE
Algorithms are brittle

- Demographic variables
- Comorbidities
- Technology
- Disease severity

Artificial intelligence could revolutionize medical care. But don’t trust it to read your x-ray just yet

By Jennifer Couzin-Frankel | Jun. 17, 2019, 12:45 PM
The AI Lifecycle

Data
- Use Case Definition
- Cohort Creation
- Data Preparation

Modeling
- Model Architecture
- Parameter Configuration
- Train, Test, Validate Cycle

Regulatory
- Technical Testing
- Clinical Testing
- FDA Clearance

Market
- Multi-site (Distributed) Validation
- Site Shadow Validation
- Site Clinical Deployment

Maintenance
- Continuous (Real World) Monitoring
- Active Learning
- Upgrade & Replacement

Continuous (Real World) Monitoring
The AI Lifecycle

Data
Use Case Definition
Cohort Creation
Data Preparation

Modeling
Model Architecture
Parameter Configuration
Train, Test, Validate Cycle

Regulatory
Technical Testing
Clinical Testing
FDA Clearance

Market
Multisite (Distributed) Validation
Site Shadow Validation
Site Clinical Deployment

Maintenance
Continuous (Real World) Monitoring
Active Learning
Upgrade & Replacement
THE AI LIFECYCLE
INCONSISTENCIES BETWEEN ACQUISITION DEVICES
The AI Lifecycle
Inconsistencies between AI algorithms
Welcome to ACR AI-LAB™

The ACR Data Science Institute has developed the ACR AI-LAB™, a data science toolkit designed to democratize AI by empowering radiologists to develop algorithms at their own institutions, using their own patient data, to meet their own clinical needs.

Learn

Learn how AI applies to imaging through a series of detailed videos.

Define Use Cases

Explore existing use cases for AI in medical imaging, or propose your own idea for a use case.

For investigational use only. The performance characteristics of this product have not been established.
SUCCESSFUL AI ALGORITHMS REQUIRE MANY PHASES OF VALIDATION AND MONITORING.

FDA CLEARED AI ALGORITHMS FOR MEDICAL IMAGING ARE OFTEN VERY BRITTLE.

UNLIKE OTHER APPROVED DEVICES, AI ALGORITHMS’ INPUT DATACHANGES CONSTANTLY.

WITHOUT MONITORING, IT IS NOT KNOWN WHEN ALGORITHMS STOP PERFORMING TO SPEC.

METHODS FOR CLEARANCE AND LABELING ARE INCONSISTENT AND UNCLEAR TO THE PUBLIC.

AUTONOMOUS AI ALGORITHMS WILL HAVE FAR MORE SERIOUS CONSEQUENCES IF THEY FAIL.

42% not reported.
‘Rule-out’ Algorithm Considerations

Rule-outs require an extremely high negative predictive value

- Special Control: Society may require superhuman performance

Tests for safety and effectiveness should be different than other CAD

- Safe: Each false negative could represent a loss of life
- Effective: Positive predictive value as high as possible

Prevalence of radiologic findings for many diseases are often very low

- Special Control: Test with a large comprehensive population (distributed validation)

Should continuously monitor for false negatives in the released product

- Special Control: All cases should be continuously monitored for statistical changes
- Special Control: Humans should overread a fixed percentage of cases (maybe all initially?)

RadioLOGY exams can have more than one finding

- When AI is negative, would the recommendation be to not interpret for other findings?

Poor following of FDA labelling could lead to critical errors in diagnosis

### CT Pulmonary Angiogram

**Common Important Findings**

- Breast cancer
- Prostate cancer
- Colon and rectum cancer
- Cervical cancer
- Tracheal, bronchus, & lung cancer
- Uterine cancer
- Stomach cancer
- Bladder cancer
- Non-melanoma skin cancer
- Kidney cancer
- Thyroid cancer
- Brain & meninges malignant melanoma
- Esophageal cancer
- Lip & oral cavity cancer
- Overian cancer
- Larynx cancer
- Eosinophilic leukemia
- Liver cancer
- Renal cell cancer
- Non-Hodgkin lymphoma
- Pancreatic cancer
- Gallbladder & biliary tract cancer

**Share of population with cancer, World, 2017**

![Graph showing cancer rates worldwide](attachment:graph.png)

0.64%
The Road to Autonomous AI

Six Levels of Automation (SAE J3016)
U.S. National Highway Traffic Safety Administration
<table>
<thead>
<tr>
<th>CLINICAL FUNCTIONALITY</th>
<th>LEVEL 0</th>
<th>LEVEL 1</th>
<th>LEVEL 2</th>
<th>LEVEL 3</th>
<th>LEVEL 4</th>
<th>LEVEL 5</th>
</tr>
</thead>
<tbody>
<tr>
<td>DETERMINE IMAGE FITNESS</td>
<td>No Automation</td>
<td>Driver Assistance</td>
<td>Partial Automation</td>
<td>Conditional Automation</td>
<td>High Automation</td>
<td>Full Automation</td>
</tr>
<tr>
<td>IMPROVE IMAGE QUALITY (PostP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DETERMINE PRIORITY (CADt)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DETECT FINDING (CADe)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>QUANTIFY MEASUREMENTS (PostP)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SINGLE-SHOT DIAGNOSES (CADx)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
THE AI LIFECYCLE
EVALUATING AND MONITORING AI ALGORITHMS

Evolving Role of AI in Radiological Imaging
FDA Public Workshop
Washington, DC
February 25, 2020

Keith J. Dreyer DO, PhD, FACR
Laura P. Coombs, PhD