AI in the Clinic:
Error Detection and Prevention

AAPM | COMP 2020

Radiation Therapy in the Era of Artificial Intelligence

Leigh Conroy, PhD

Radiation Oncology
UNIVERSITY OF TORONTO

UHN Princess Margaret Cancer Centre

Acknowledgements

Princess Margaret Cancer Centre

- Tom Purdie
- Chris McIntosh
- Andrea McNiven

University of Washington

- Eric Ford

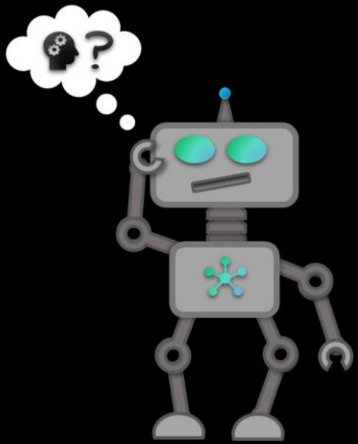
I have no conflict of interest to disclose

Outline

AI in the Clinic: Error Detection and Prevention

AI in the Clinic	Error Detection AI for QA	Error Prevention QA for AI
Accuracy Interpretability Integration	AI for patient-specific QA Treatment plan QA AI for quality improvement and incident learning	Validation: Technical & Clinical Risk-based analysis

AI in the Clinic



Decision Support System (DSS): guide judgements & actions

DSS for oncologists:

- Predict disease presence (diagnosis) or outcomes (prognosis)

DSS for physicists:

- Does this machine require maintenance?
- Is this treatment plan acceptable and deliverable?
- Are there gaps in my quality program?

AI *for* the Clinic ➡ AI *in* the Clinic

- Tension between Accuracy and Interpretability
- Machine Prediction meets Human Judgement
- Workflow Integration

Radiation Oncology
UNIVERSITY OF TORONTO
UHN Princess Margaret Cancer Centre

Error Detection

Traditional Rule-based

Compare with historical or reference values

Statistical outlier methods: mean, standard deviation

Statistical Process Control

Machine Learning-based

Linear regression

Classification model

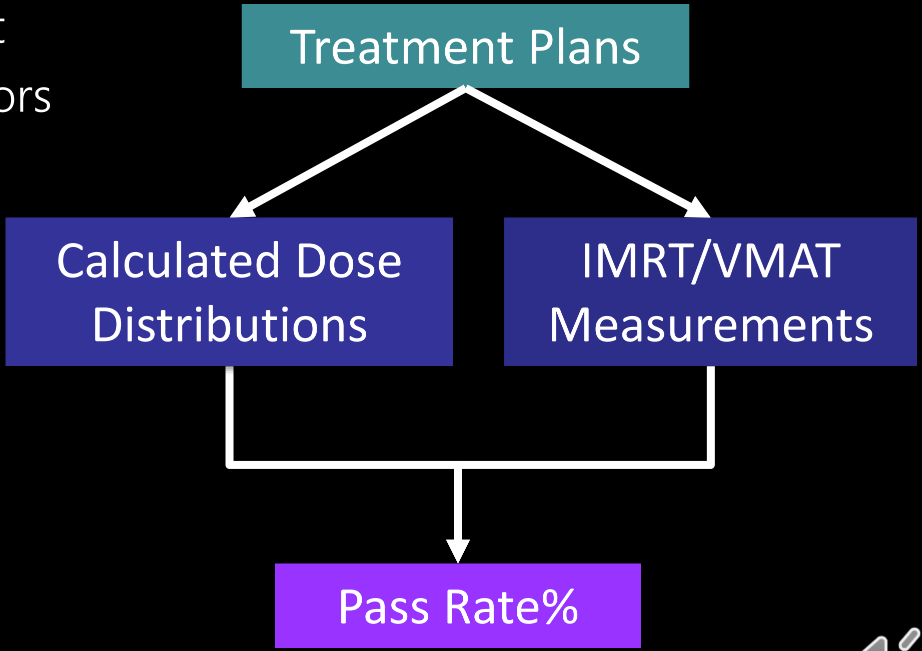
Random Forests, Isolation Forests

Convolutional Neural Networks

Error Detection in Patient Specific QA

Patient specific QA

- Time consuming, difficult to interpret
- Insensitive, often unable to catch errors
- Occurs late in the planning process



Courtesy of Andrea McNiven

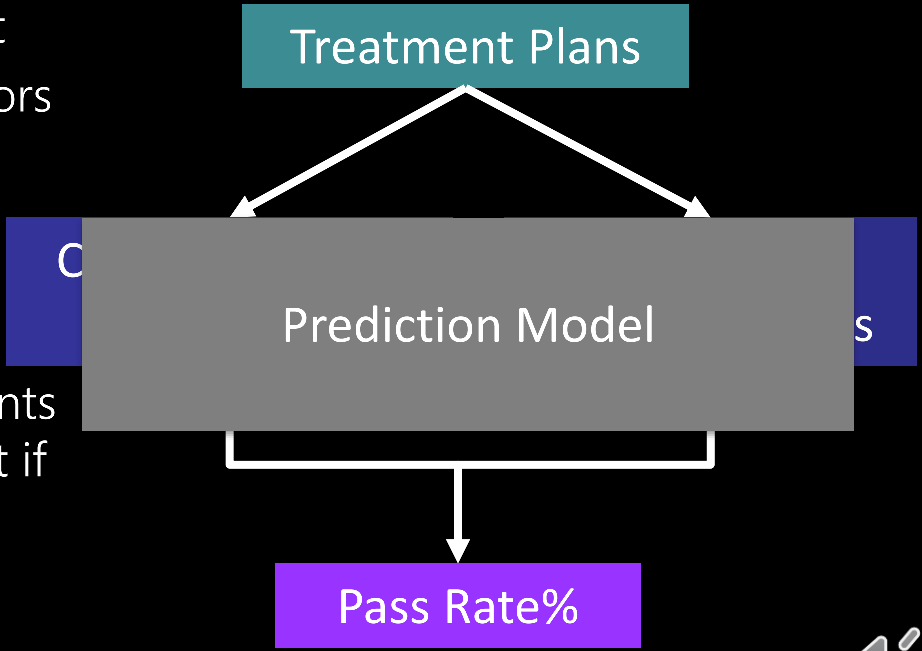
Error Detection in Patient Specific QA

Patient specific QA

- Time consuming, difficult to interpret
- Insensitive, often unable to catch errors
- Occurs late in the planning process

AI for patient specific QA

- Use past patient specific measurements to train and test the model to predict if a plan will pass QA



Courtesy of Andrea McNiven

Radiation Oncology UNIVERSITY OF TORONTO

UHN

AI for PSQA

Detectors

- Planar diode array [1,2,7], Film [5]
- 3D Diode Array [3,8]
- Portal Dosimetry or EPID [2,4]

Models

- Poisson regression [1,2,7]
- Random forest [7,8]
- Support vector classifier [3,6]
- CNN [5,6,8]

Predicted Values

- Gamma pass rate [1,2,5,7,8]
- Mean dose difference [3]
- Errors/Outliers [4,6]

Features

- Plan based & complexity [1,2,3,5,7,8]
- Machine QC results [3]
- Radiomics [4,6]

1. G. Valdes et al 2016 Med Phys 43(7): 4323 – 4334.
2. G. Valdes et al 2017 J Appl Clin Med Phys 18(5): 279-284.
3. D. Granville et al 2019 Phys Med Biol 64: 095017.
4. L. Wootton et al 2018 IJROBP 22(1): 219- 228.
5. S. Tomori et al 2018 Med Phys 45(9): 4055 – 4065.
6. M Nyflot et al 2019 Med Phys 46(2): 456 - 464
7. J Li et al 2019 IJROBP 105(4): 893 – 902.
8. T. Ono et al 2019 Med Phys 46(6): 3823 – 3832.

…

AI-facilitated interpretation

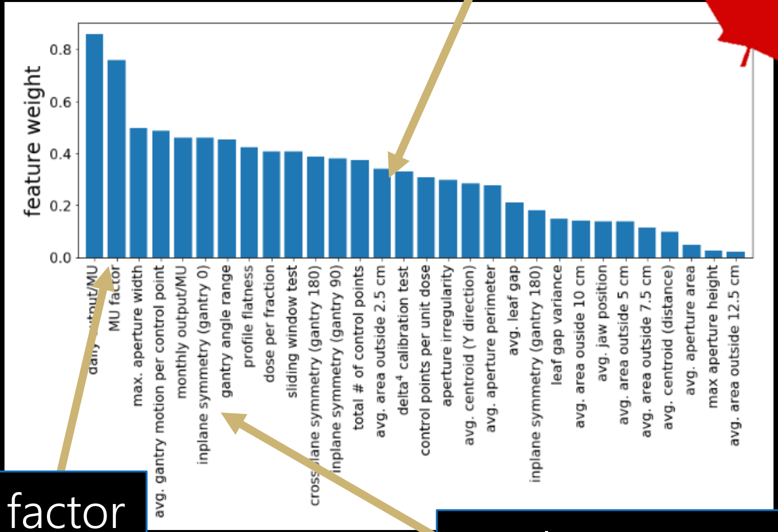
Unlike traditional measurements, predictions can be interrogated to determine most relevant features.

This inferred knowledge can be fed back into continuous quality improvement.

- Planning
- Machine QA
- Detector limitations

Detector Calibration

Top 30 most predictive features



MU factor

In-plane symmetry

D. Granville et al 2019 Phys Med Biol 64: 095017.

Radiation Oncology
UNIVERSITY OF TORONTO

AI *in a different clinic*

RADIATION ONCOLOGY PHYSICS

IMRT QA using machine learning: A multi-institutional validation

Gilmer Valdes[1,3,a] | Maria F. Chan[2,a] | Seng Boh Lim[2] | Ryan Scheuermann[3] | Joseph O. Deasy[2] | Timothy D. Solberg[1,3]

- "Virtual IMRT QA" (Valdes 2016)
- ~80-200 plans to train model
- Complimentary to measurement-based program

Model

Features

>90 plan complexity metrics (CIAO, modulation factor, irregularity factor...)

Prediction

Gamma Pass Rate

3%/3 mm Threshold: 10%

Institution 1
Diode array
498 plans

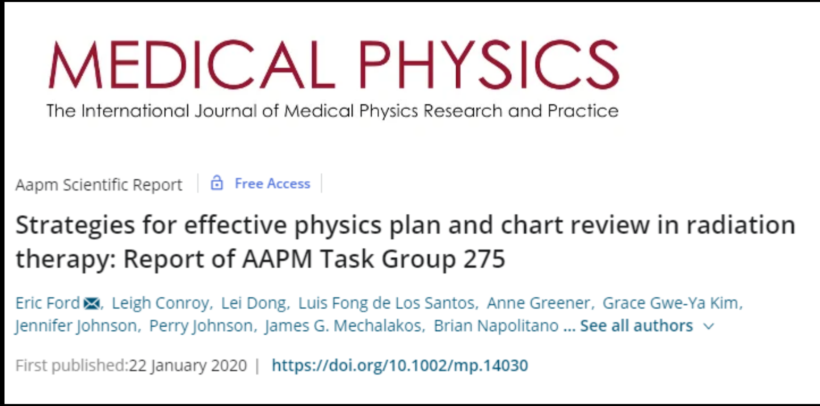
3% accuracy

Institution 2
Portal dosimetry
139 plans

3.5% accuracy

1. G. Valdes et al 2016 Med Phys 43(7): 4323 – 4334.
2. G. Valdes et al 2017 J Appl Clin Med Phys 18(5): 279-284

Error Detection in Treatment Planning

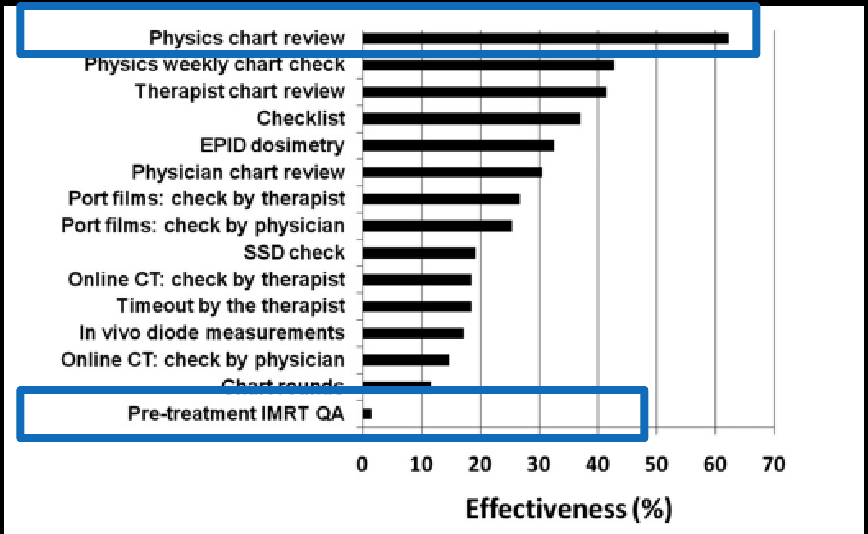

MEDICAL PHYSICS
The International Journal of Medical Physics Research and Practice

Aapm Scientific Report | Free Access

Strategies for effective physics plan and chart review in radiation therapy: Report of AAPM Task Group 275

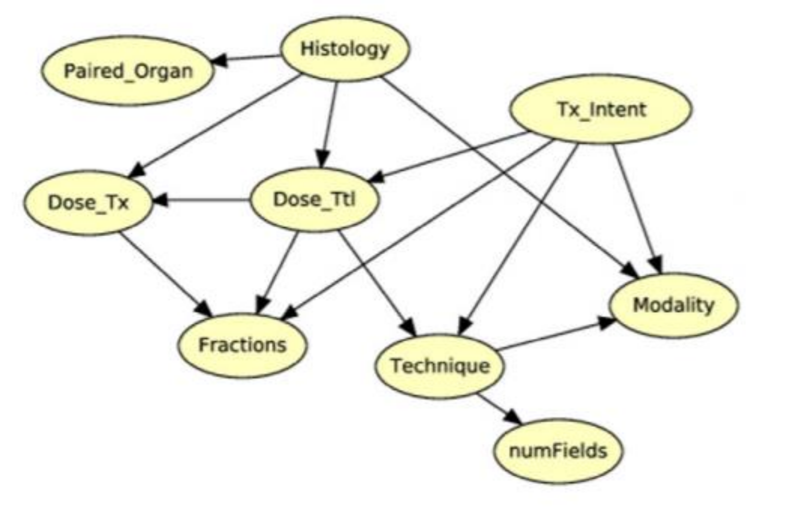
Eric Ford, Leigh Conroy, Lei Dong, Luis Fong de Los Santos, Anne Greener, Grace Gwe-Ya Kim, Jennifer Johnson, Perry Johnson, James G. Mechalakos, Brian Napolitano ... See all authors

First published:22 January 2020 | https://doi.org/10.1002/mp.14030



E. Ford et al 2012 IJROBP 84(3): e263 - 269

Error Detection in Treatment Planning



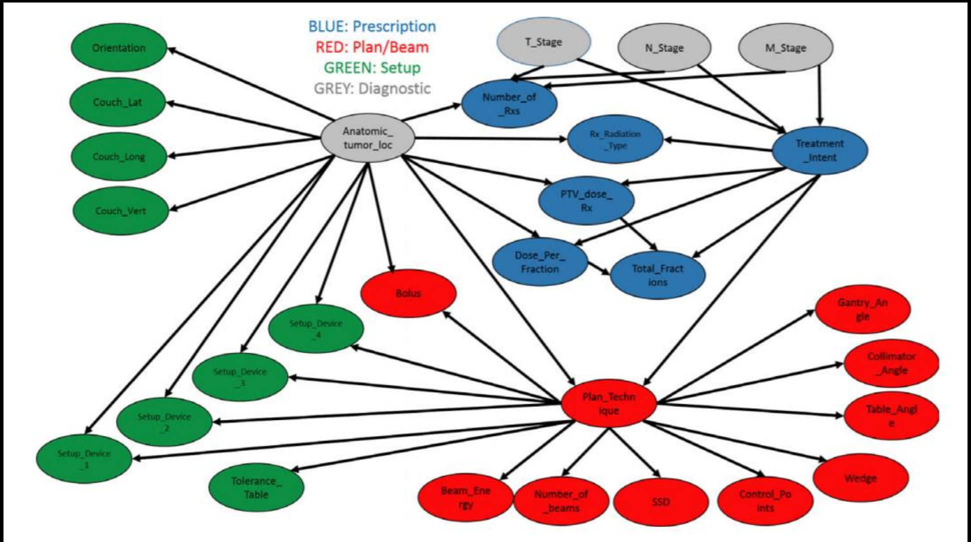
- Joint probability distributions: what is the probability of certain RT parameters, given set of clinical information
- Flag low probability events
- Mimics how humans check plans

Technical & Clinical Expert Validation

- Network AUCs = 0.88 – 0.98
- Human Expert AUCs = 0.90 +/- 0.01

Kalet et al 2015 Phys Med Biol 60: 2735

Error Detection in Treatment Planning



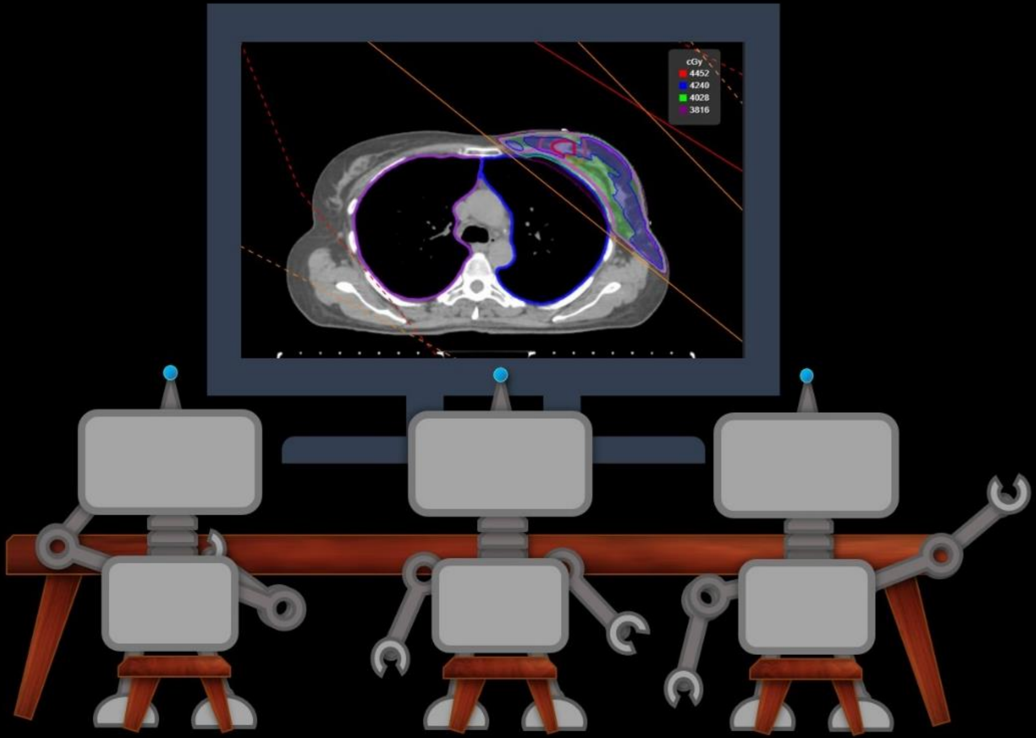
AUC by years of historical data trained on:

- 2 years: 0.82
- 3 years: 0.85
- 4 years: 0.89
- 5 years: 0.88

Recommendation
Train on 4 years of data, update model yearly

Luk et al 2019 Med Phys 46(5): 2006 – 2014.

AI for Peer Review Rounds



Radiation Oncology
UNIVERSITY OF TORONTO

AI for Peer Review Rounds

Plans from
Peer Review Rounds

Complexity Score [0,1]

↓

Testing Set
N = 202
(15 consecutive weeks)

↓

Isolation Forest Model
Trained on historical dataset
N= 3460 plans

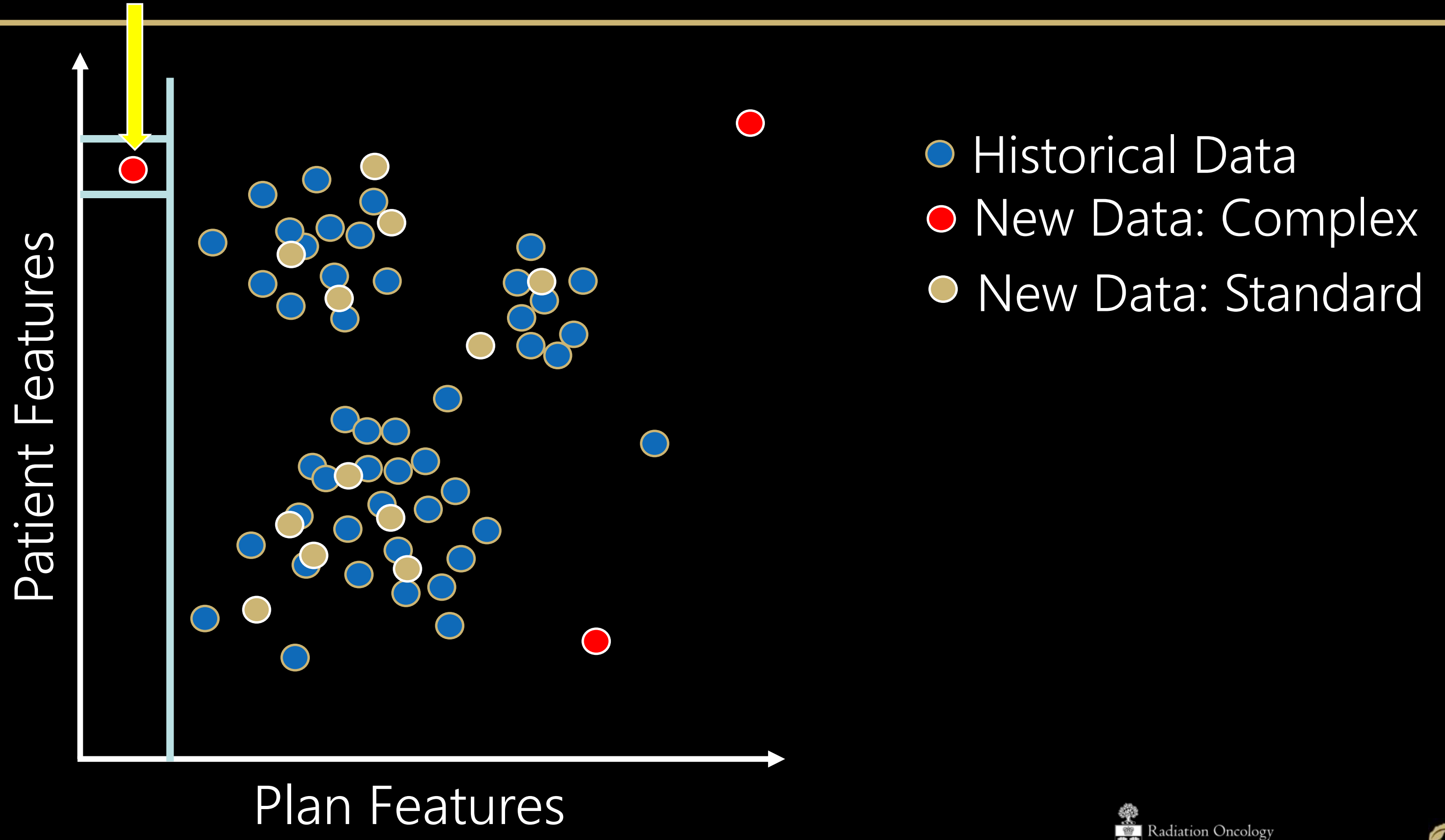
↓

Complexity Prioritization Score
[0 – 1]

Aim: Develop an automated framework for complex case prioritization in peer review rounds

Prospectively assigned binary complexity scores
Complex plan (discussion) = 1 [n=38]
Standard plan (no discussion) =0 [n=164]

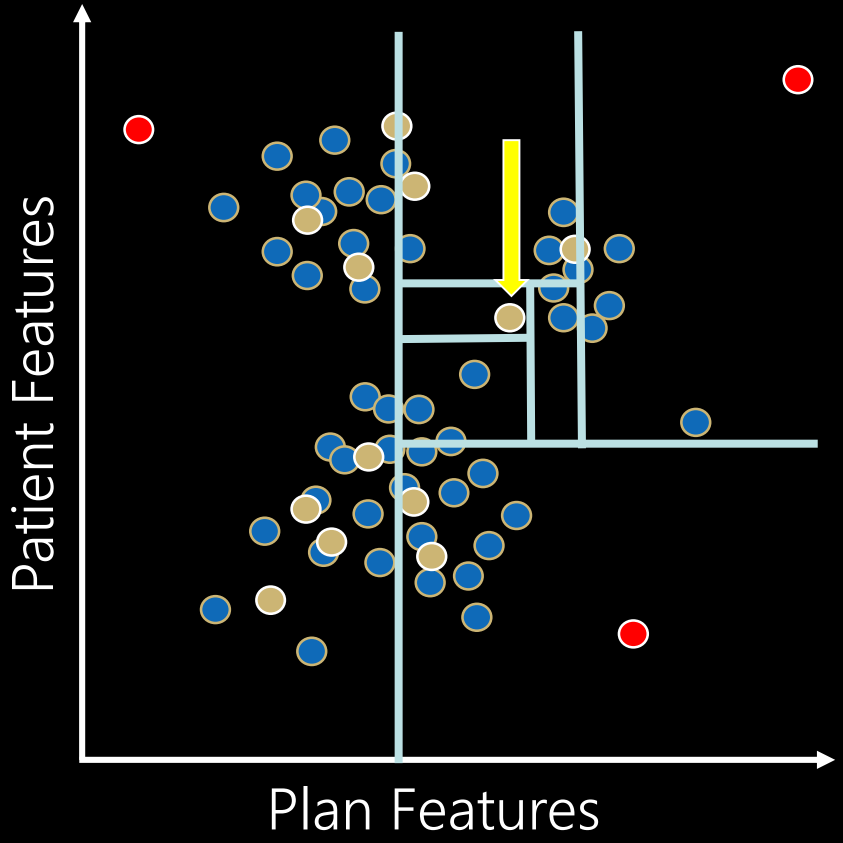
Isolation Forest for Outlier Detection



Radiation Oncology
UNIVERSITY OF TORONTO

UHN

Isolation Forest for Outlier Detection



- Historical Data
- New Data: Complex
- New Data: Standard

Predicting Complexity

Extracted Features Included:

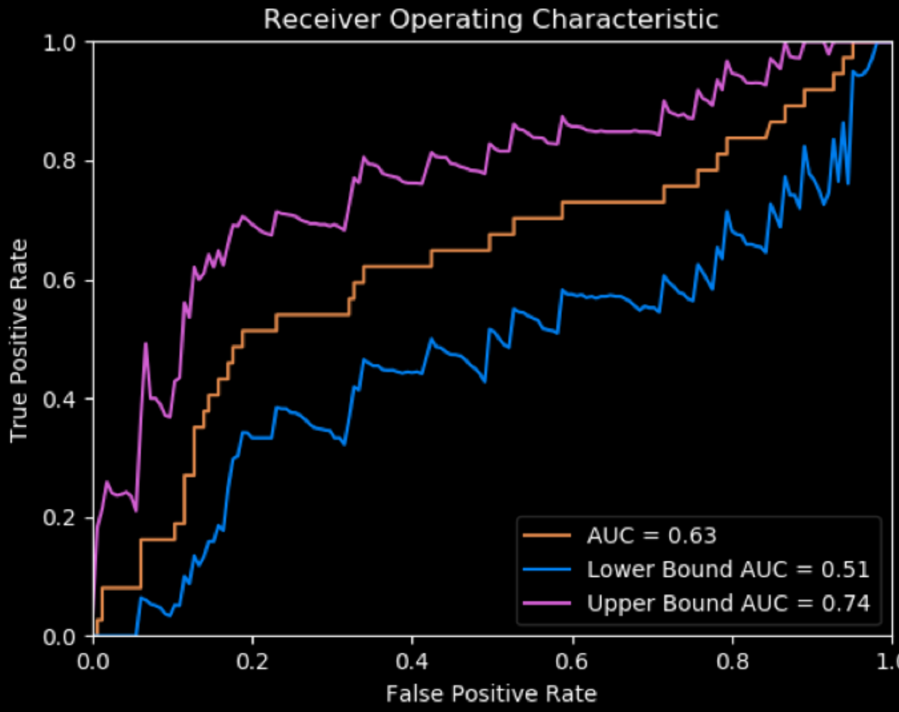
- Image and radiation dose features (filters/deep learning)
- Contour features
- Radiation beam features (angles, MUs)

Courtesy of Chris McIntosh

Predicting Complexity

Extracted Features Included:

- Image and radiation dose features (filters/deep learning)
- Contour features
- Radiation beam features (angles, MUs)

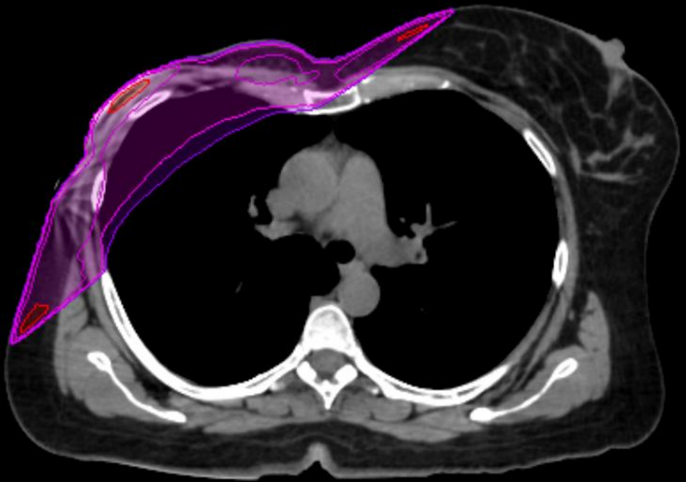


Courtesy of Chris McIntosh

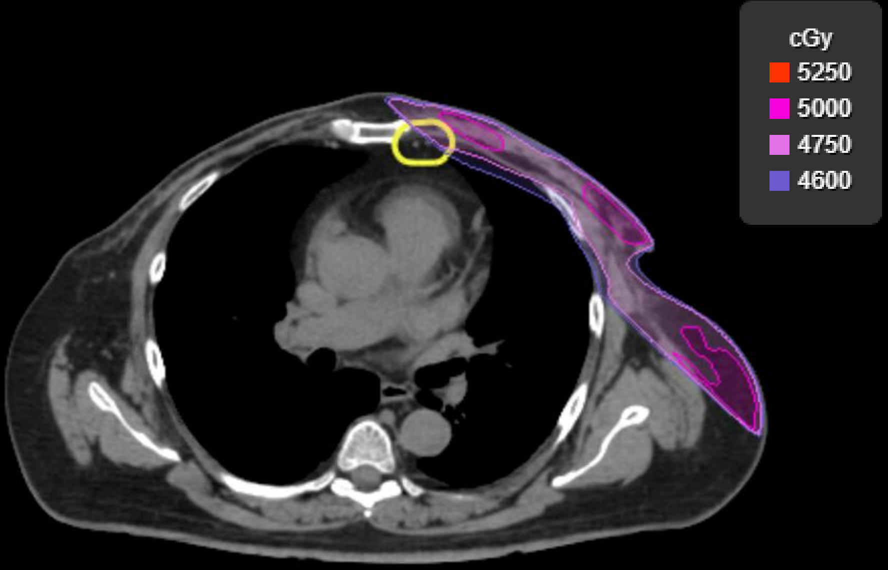
Radiation Oncology UNIVERSITY OF TORONTO

UHN

Interpretation



True Positive



False Negative

Princess Margaret: Then and Now

	Historical Training Set (2016)	Testing Set (2018)
TPS	Pinnacle	RayStation
Nomenclature	Institutional	AAPM TG-263
Clinical Practice	Few IMNs treated	Many IMNs treated

- Changes in practice impact the applicability of trained models
- Clinical data has a half life
- Data curation and feature selection require domain expertise and many iterations

AI for near miss identification

Basic Original Report

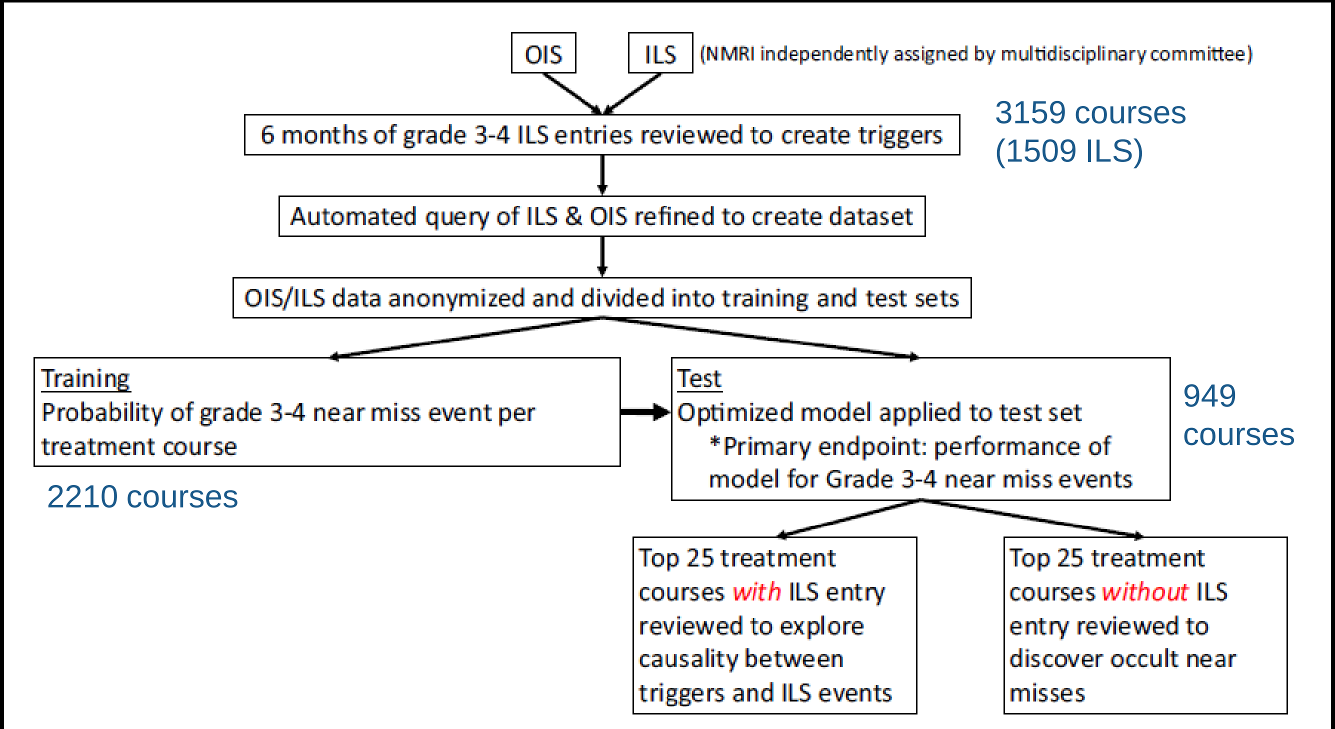
A Radiation Oncology—Specific Automated Trigger Indicator Tool for High-Risk, Near-Miss Safety Events

Pehr E. Hartvigson, MD,[a,b,*] Michael F. Gensheimer, MD,[c]
Phil K. Spady, BS,[a] Kimberly T. Evans, BA,[a] and Eric C. Ford, PhD[a]

[a]Department of Radiation Oncology, University of Washington School of Medicine, Seattle, Washington; [b]Department of Radiation Medicine, Oregon Health and Science University, Portland, Oregon; and [c]Department of Radiation Oncology, Stanford University, Stanford, California

- Limitations of Incident Learning Systems:
  - Voluntary
  - Require strong safety culture for high volume reporting
  - There will always be unreported events
- *Objective:* Develop a radiation oncology-specific trigger tool to estimate the probability of a grade 3-4 near-miss event for each treatment course using only the data available in the OIS.

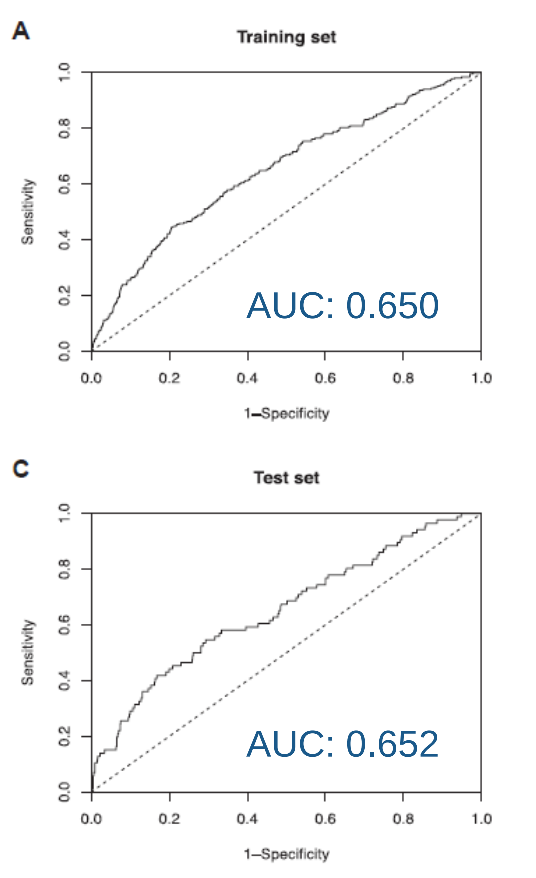
AI for near miss identification



Hartvigson, P., et al., A Radiation Oncology-Specific Automated Trigger Indicator Tool for High-Risk, Near-Miss Safety Events PRO (2019)

Radiation Oncology
UNIVERSITY OF TORONTO

AI for near miss identification



9/20 triggers significant:

- Hidden/new/deleted fields
- New prescription
- Documentation

- With ILS entry: causal effect in 50%

- Without ILS entry:
  - 5/25 flagged for additional review
  - 2/5 were unreported near miss

Radiation Oncology
UNIVERSITY OF TORONTO

UHN

Error Prevention: QA the AI!

Assess risk through systematic risk analysis (TG-100)
Mitigate risk with specialized AI-QA program design

A risk assessment of automated treatment planning and recommendations for clinical deployment

Kelly Kisling, and Jennifer L. Johnson
Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

Three key aspects of safe deployment:
- User training on potential failure modes
- Comprehensive manual review
- Automated QA (flag for human review)

Error Prevention: QA the AI!

Assess risk through systematic risk analysis (TG-100)
Mitigate risk with specialized AI-QA program design

A risk assessment of automated treatment planning and recommendations for clinical deployment

Kelly Kisling, and Jennifer L. Johnson
Department of Radiation Physics, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

Three key aspects of safe deployment:

- User training on potential failure modes
- Comprehensive manual review
- Automated QA (flag for human review)

AAPM Machine Learning Subcommittee (MLSC) Task group Proposal:

Quality Assurance for Machine Learning-based Clinical Technologies

Co-chairs: Dr. Habib Zaidi & Dr. Gilmer Valdes

Radiation Oncology
UNIVERSITY OF TORONTO

Where do Medical Physicists fit in?

Physicists are technical experts with clinical domain expertise

We are uniquely positioned to shape the future of AI in medicine

- Data: governance, collection, curation
- Problem definition
- Model development, testing, and tuning
- Workflow design, validation, and implementation
- Supervision, maintenance
- Development and execution of QA for AI

AI for Error Detection @ AAPM | COMP 2020

Thursday, 7/16/2020, 2:00 PM - 3:00 PM *[Eastern Time (GMT-4)]* Room: Track 5
A Generalizable Contour Validation Method Using Deep Learning-Based Image Classification
Y Zhang, F Ceballos, Y Liang, L Buchanan, X Li

E Posters:

Reducing IMRT QA Workload by 95% and Keeping the Same Level of Quality Control T Nano, M Descovich, E Hirata, Y Interian, G Valdes

Development and Validation of a Machine Learning Predictive Model of IMRT Patient-Specific Quality Assurance Approval Using Gamma-Radiomics C Yaly, J Lizar, P Santos, A Colello Bruno, G Viani, J Pavoni

Towards a Treatment Planning Optimization Framework Utilizing Predicted Quality Assurance Outcomes From a Machine Learning Model to Maximize Plan Quality and Deliverability P Wall, J Fontenot

Error Detection and Classification in Patient Specific IMRT QA with Dual Neural Networks N Potter, K Mund, J Andreozzi, J Li, C Liu, G Yan

Dose Prediction for Patient-Specific QA Using a Convolutional Neural Network K Mund, G Yan

Does Radiomics Have the Potential to Assess KV-CBCT Image Performance Acquired From Phantom Data Used for Daily QA?
M Shenouda, N Baughan, J Cruz Bastida, E Pearson, H Al-Hallaq

Out of Sample Performance of a Deep Learning Based Registration Quality Assurance Method X Zhou, S Galib, H Lee, G Hugo

Towards Quality Assurance for First AI-Driven Online Adaptive Radiotherapy Based On Failure Mode and Effect Analysis
J Booth, P Sibolt, E Laugeman, B Cai, D Sjostrom, S Mutic, M Perez

Radiation Oncology UNIVERSITY OF TORONTO

UHN Cancer Centre