Explainable Artificial Intelligence to Identify Dosimetric Predictors of Toxicity in Patients with LA-NSCLC: A Secondary Analysis of RTOG 0617

Jeff Ryckman, MD, MSMP
Assistant Professor
WVU Cancer Institute

Disclosures

- Editor, Rad Onc Review
- Content Developer, Rad Onc Calc
- Panel Member, CancerRetreatment.org
- Non-Voting Member, TG263U1
- Advisor, NewLeaf AI

INTERNATIONAL JOURNAL OF
RADIATION ONCOLOGY • BIOLOGY • PHYSICS
www.redjournal.org

PHYSICS CONTRIBUTION

Explainable Artificial Intelligence to Identify Dosimetric Predictors of Toxicity in Patients with Locally Advanced Non-Small Cell Lung Cancer: A Secondary Analysis of RTOG 0617

Colton Ladbury, MD,* Richard Li, MD,† Anseh Danesharasteh, MS,‡ Zeynep Ertem, PhD,‡ Andrew Tam, MD,* Jason Liu, MD,* Claire Hao,* Rose Li, MD, PhD,* Heather McGee, MD, PhD,* Sagus Sampath, MD,* Terence Williams, MD, PhD,* Scott Glaser, MD,* Mohammad Khasawneh, PhD,‖ Zhongxing Liao, MD,‖ Percy Lee, MD,‖ Jeff Ryckman, MD,¶ Parvez Shaikh, MD,# and Arya Amini, MD*

*Department of Radiation Oncology, City of Hope National Medical Center, Duarte, California; †Department of Radiation Oncology, Partners in Health Whittier Hospital, Whittier, California; ‡Department of Systems Science and Industrial Engineering, Binghamton University, Binghamton, New York; ‖Department of Radiation Oncology, MD Anderson Cancer Center, Houston, Texas; ¶Department of Radiation Oncology, City of Hope Orange County Lennar Foundation Cancer Center, Irvine, California; #Department of Radiation Oncology, West Virginia University Medicine Camden Clark Medical Center, Parkersburg, West Virginia; and ‖Department of Radiation Oncology, West Virginia University School of Medicine, Morgantown, West Virginia

Received Feb 28, 2023; Accepted for publication Jun 13, 2023

Outline

Understand was XAI represents

Learn how to interpret SHAP visualization

Understand trade-offs in modern radiation planning when pushing low dose volumes

An aside on Historic vs. Modern Planning

Background

“…The greatest cellucidal effect is obtained by single -dose fractionation; however, as a rule, the concomitant damage to normal tissues…is not well tolerated…and we are forced to fractionate.” *Marciel V. Time-dose fractionation relationships in radiation therapy. Natl Cancer Inst Monogr 1967; 24: 187-203*

"Because of the sparseness of long-term follow-up for SBRT, it should be recognized that the data in both Table III and the published reports represent, at best, a first approximation of normal tissue tolerance." *TG-101*

"I was recently informed that I won a contest that I did not actually enter." *Robert Timmerman*

➢ Historically, we were forced to fractionationate. Constraints continue to be a moving target, especially as technology and planning techniques continue to improve.

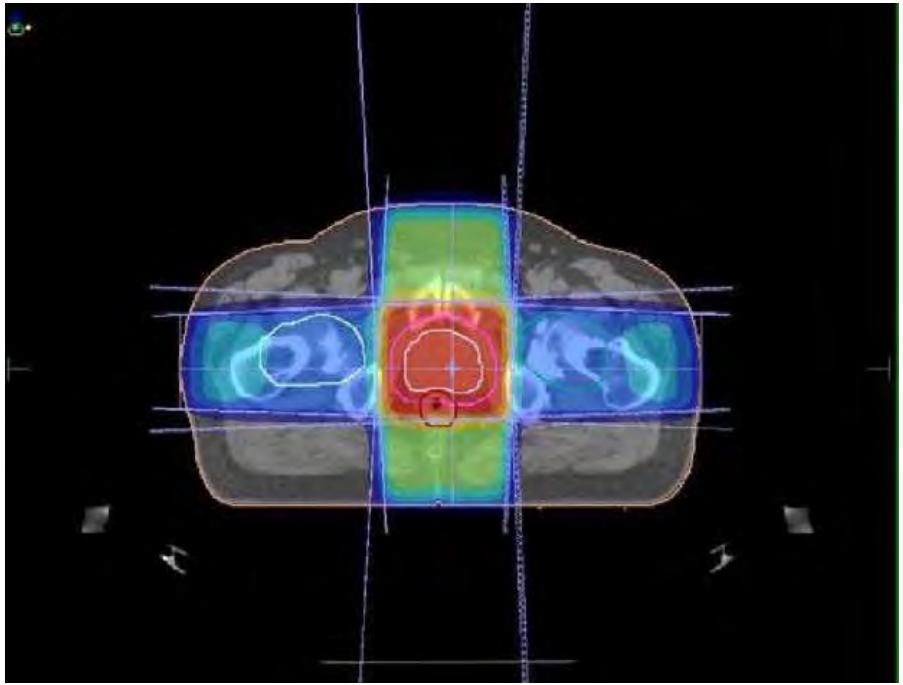
The Pace of Technology

"Radiotherapy is a treatment that has been around for one hundred years now, and certainly, if you think about the advances in your mobile phone even over the last decade, you can translate into medical technology for radiation. Our outcomes are getting exceptional now, and, I think one thing to impress to the community, is how advanced we have come in the delivery and how safe radiation treatment is. I think in the future we will be using it for more and more indications."

-Shankar Siva

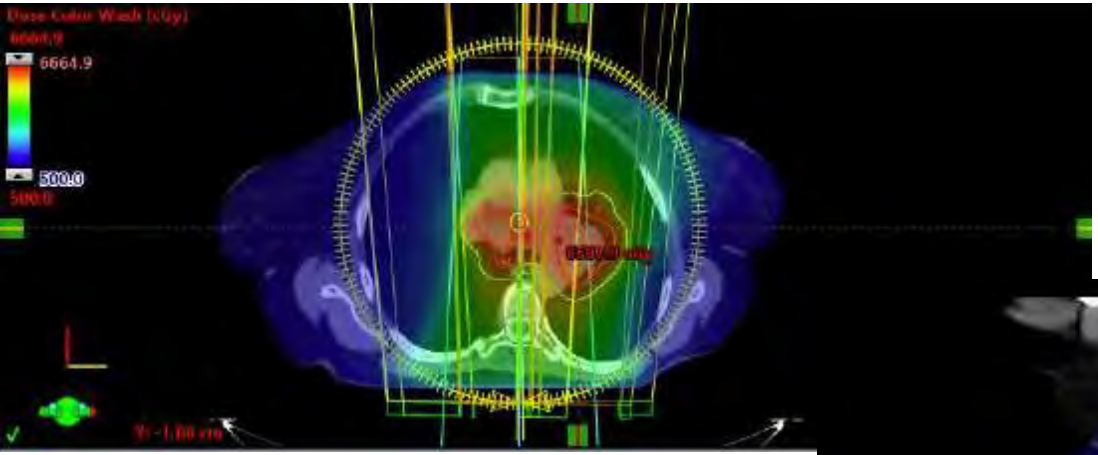
➢ Modern radiation planning varies greatly compared to historical radiation planning. How come we still apply conventional metrics to modern planning?

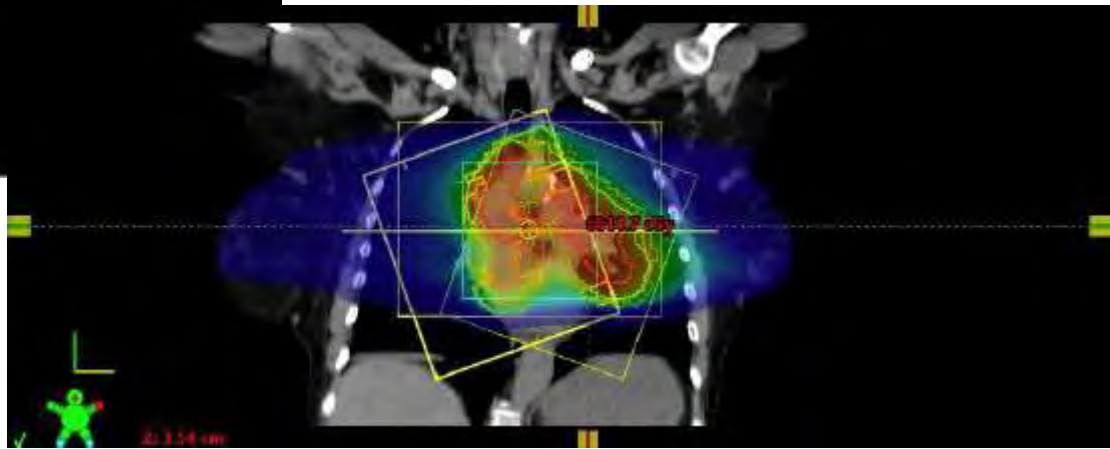
Historic RT



➢ Four field box represents historical radiotherapy planning. In this context, low dose wash may *directly* correlate with intermediate and high dose exposure.

Modern RT





➢ VMAT is one representation of modern radiation planning. In the context of full arcs, low dose wash may *inversely* correlate with intermediate and high dose wash.

Low dose vs. High dose tradeoff

Partial arcs:

- Best for when targets are not centrally located (e.g., N1/hilar lymph node)
- Low dose wash may *directly* correlate with intermediate and high dose wash.

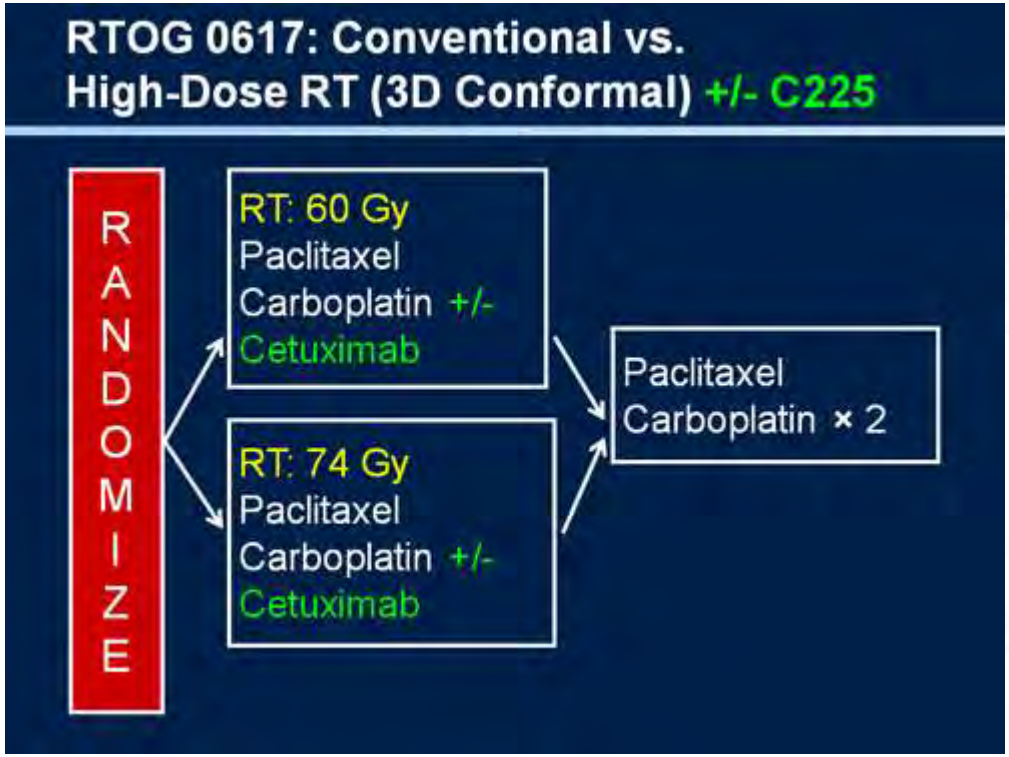
Full arcs:

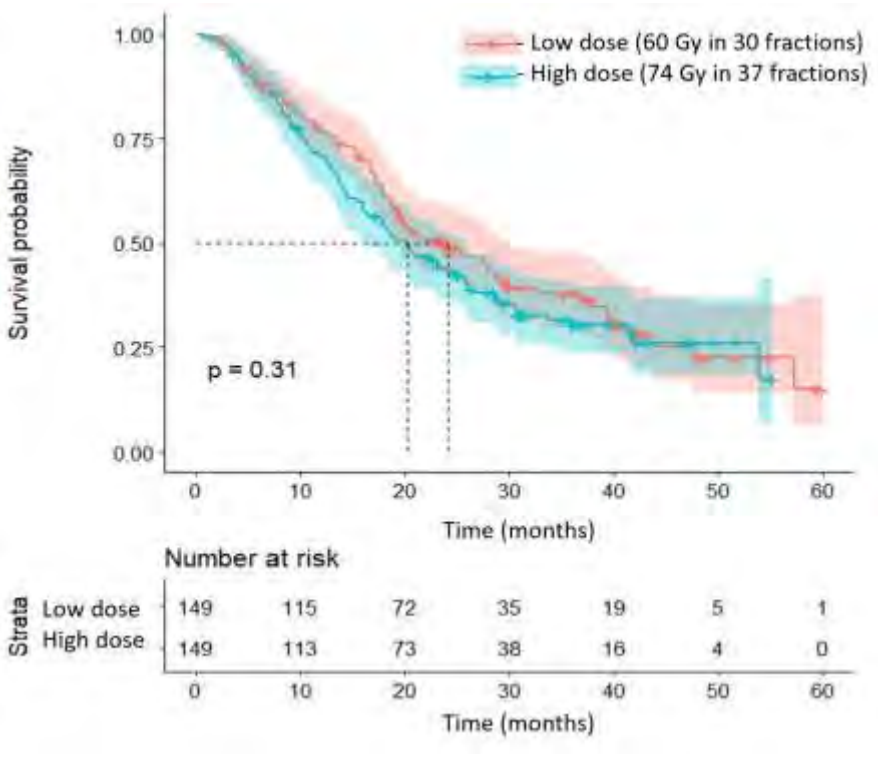
- Best for when targets are located behind the heart (e.g., esophageal primaries or NSCLC where components of targets may extend centrally, e.g. N2 disease)
- Low dose wash may *indirectly* correlate with intermediate and high dose wash

➢ Physics 101: Allowing for more liberal low dose wash allows for intermediate and high dose ALARA in the context of full arcs. Intermediate and high dose drive toxicity.

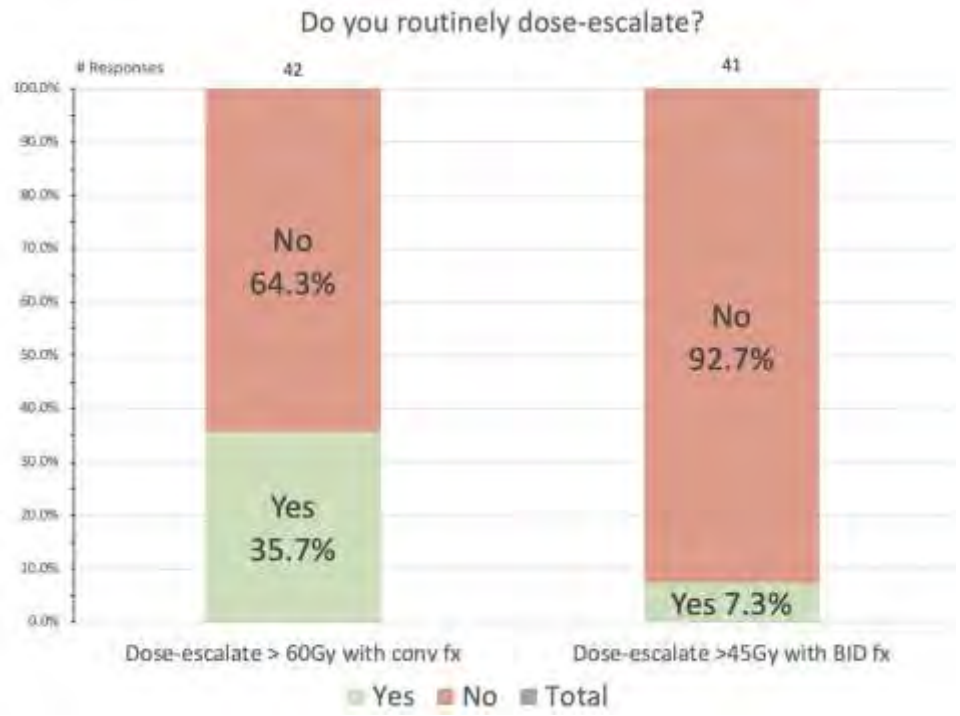
Review of RTOG 0617 and Treatment of Node Positive LA-NSCLC

RTOG 0617: Brief Overview





Qd CCRT: 36% of thoracic experts dose escalate



Dose Planning and Radiation Optimization for Thoracic Conventional, Twice Daily, and Stereotactic Radiation Therapy: A Delphi Consensus from a National Survey of Practitioners

Julius Weng[1], Jeff Ryckman[2], Matthew S. Katz[3], Hina Saeed[4], Christopher Estes[5], Issam El Naqa[6], Amy Moreno[1], Sue S. Yom[7], for the Dose Planning and Radiation Optimization (Dose-PRO) Consensus Group. Submitted for review to IJROBP in 2024.

Patients on RTOG 0617

439 patients included in this analysis

- Grade 3+ esophageal toxicity: 12.1% (n=53).
- Grade 3+ cardiac toxicity: 9.3% (n=41).
- Grade 3+ pulmonary toxicity: 5.9% (n=23).

➢ All toxicities had class imbalances, where cases without toxicity far outnumbered cases with toxicity, which may prevent ML from adequately learning minority class.

Outline

Understand was XAI represents

Learn how to interpret SHAP visualization

Understand trade-offs in modern radiation planning when pushing low dose volumes

What is Explainable Artificial Intelligence (XAI)?

One common limitation of machine learning algorithms the inherent complexity and interpretability due to a "black box" approach

XAI, popularized by Shapley Additive Values (SHAP), provides visualizations of the inner workings of the "black box"

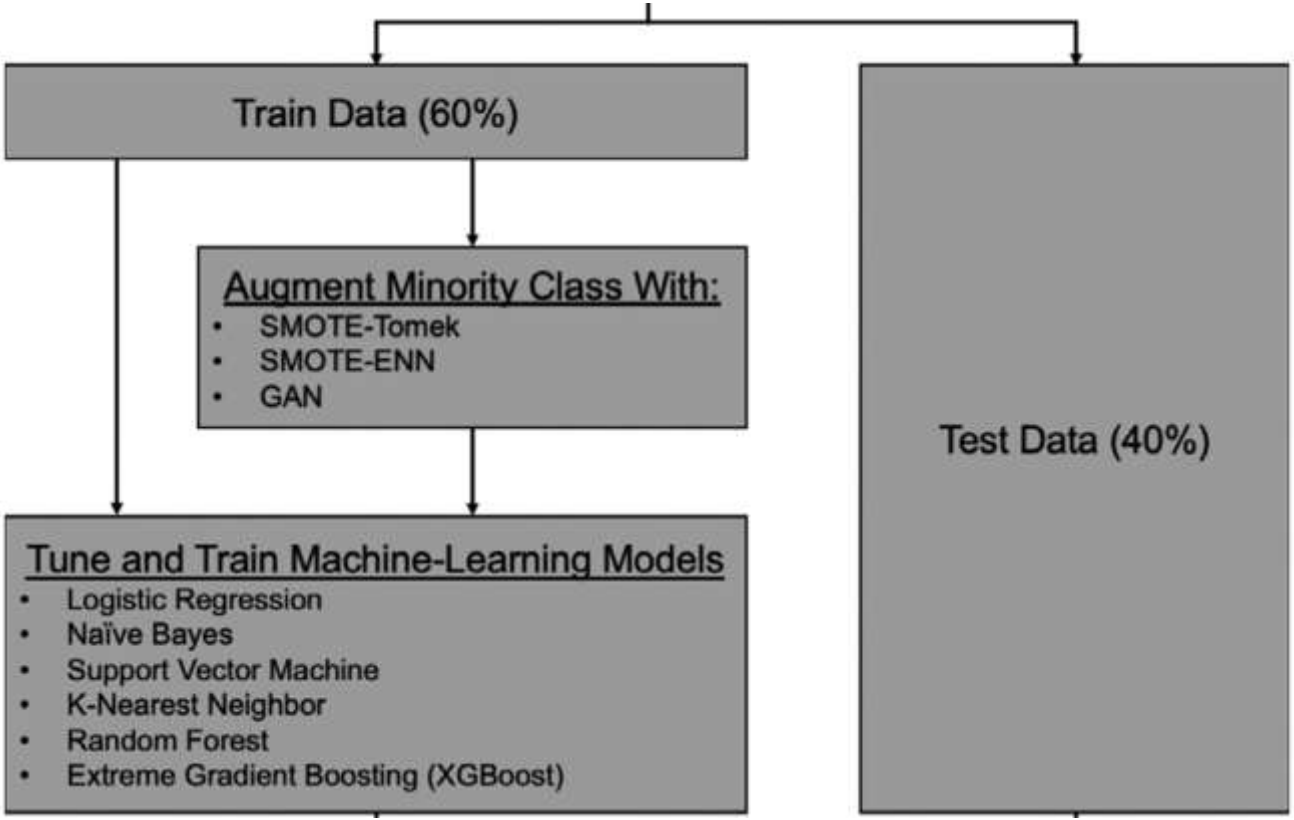
This report is the first report exploring dose-toxicity relationship to identify clinically useful dose constraints.

➢ XAI: Allows a "glass box" approach to see what is going on inside the black box.

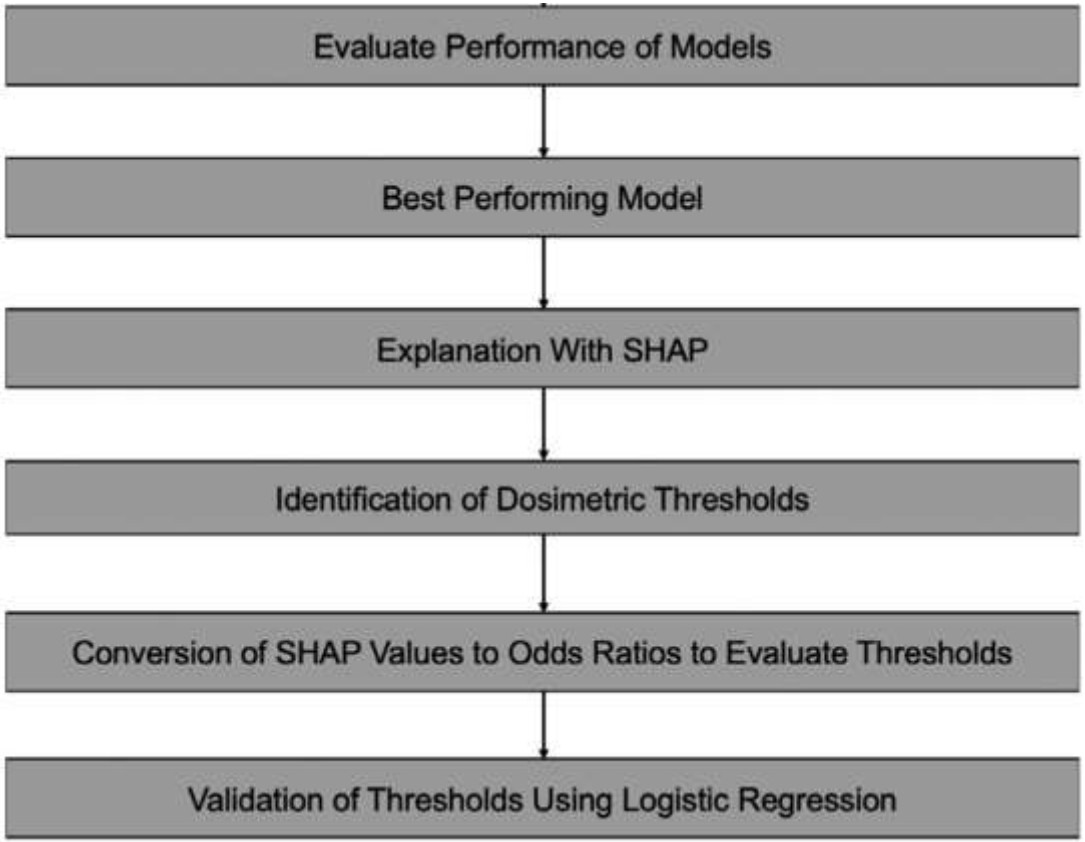
Design

DATASET		
X VARIABLES		Y VARIABLES
CATEGORICAL	CONTINUOUS	
• Zubrod Performance Score • T Stage • N Stage • Group Stage • Receipt of Cetuximab • Receipt of Consolidation Chemo • Radiation Treatment Arm • Radiation Technique	• Actual RT Dose Received • Lung D_{mean} • Lung V5 • Lung V20 • Esophagus • Esophagus D_{mean} • Esophagus V20 • Heart D_{mean} • Heart V40	Grade 3+ Pulmonary Toxicity Grade 3+ Esophageal Toxicity Grade 3+ Cardiac Toxicity

Design


Train Data (60%)
Augment Minority Class With:
• SMOTE-Tomek
• SMOTE-ENN
• GAN
Test Data (40%)
Tune and Train Machine-Learning Models
• Logistic Regression
• Naïve Bayes
• Support Vector Machine
• K-Nearest Neighbor
• Random Forest
• Extreme Gradient Boosting (XGBoost)

Design


Evaluate Performance of Models
Best Performing Model
Explanation With SHAP
Identification of Dosimetric Thresholds
Conversion of SHAP Values to Odds Ratios to Evaluate Thresholds
Validation of Thresholds Using Logistic Regression

SHAP Visualizations

Values > 0: Positive effect (event is more likely)

Values < 0: Negative effect (event is less likely).

SHAP sorts from top to bottom based on mean absolute SHAP values.

SHAP values were used to create dependence plots to visualize the risk of an outcome as a function of a single independent feature (here, G3+ toxicity)

These plots can be used to identify pertinent dosimetric parameters by plotting dosimetric parameter and its corresponding SHAP values, focusing on where SHAP values cross 0.

➢ SHAP plots provide a general overview of which features most heavily influence ML model output.

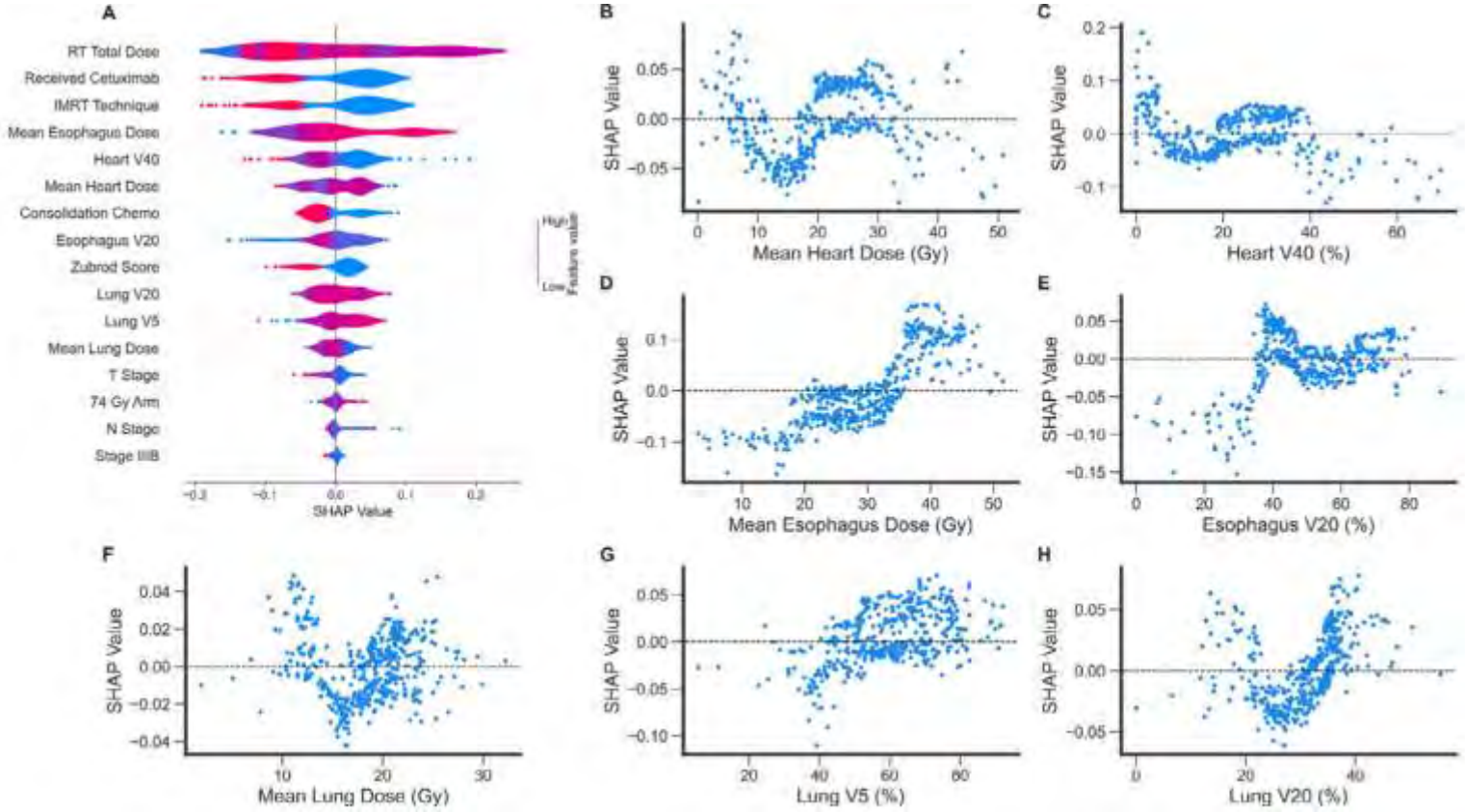
Evaluation of ML model performance

Supplementary Table 1: Performance of Machine Learning Models Measured Using Area Under the Curve With 95% Confidence Intervals

Model	Logistic Regression	Naïve Bayes	Support Vector Machine	K-Nearest Neighbors	Random Forest	Extreme Gradient Boosting
			Grade 3+ Pulmonary Toxicity			
No Imbalance Correction	0.539 (0.315-0.759)	0.557 (0.373-0.736)	0.604 (0.4-0.806)	0.54 (0.329-0.735)	0.633 (0.442-0.807)	0.586 (0.366-0.78)
SMOTE-Tomek	0.608 (0.413-0.796)	0.617 (0.456-0.784)	0.602 (0.388-0.794)	0.547 (0.378-0.727)	0.59 (0.389-0.761)	0.603 (0.41-0.786)
SMOTE-ENN	0.578 (0.375-0.775)	0.593 (0.371-0.798)	0.598 (0.393-0.781)	0.528 (0.35-0.704)	0.543 (0.344-0.736)	0.557 (0.376-0.743)
GAN	0.687 (0.47-0.872)	0.6 (0.403-0.787)	0.689 (0.482-0.882)	0.642 (0.44-0.833)	0.736 (0.574-0.879)	0.739 (0.579-0.878)*
			Grade 3+ Esophageal Toxicity			
No Imbalance Correction	0.672 (0.545-0.792)	0.647 (0.52-0.762)	0.598 (0.469-0.717)	0.663 (0.526-0.791)	0.668 (0.535-0.788)	0.652 (0.531-0.767)
SMOTE-Tomek	0.635 (0.514-0.75)	0.596 (0.474-0.719)	0.59 (0.464-0.71)	0.642 (0.516-0.765)	0.706 (0.593-0.806)*	0.684 (0.565-0.791)
SMOTE-ENN	0.682 (0.557-0.794)	0.624 (0.494-0.748)	0.689 (0.573-0.797)	0.572 (0.462-0.684)	0.703 (0.597-0.798)	0.691 (0.576-0.789)
GAN	0.682 (0.556-0.8)	0.645 (0.501-0.774)	0.68 (0.558-0.792)	0.668 (0.521-0.804)	0.668 (0.545-0.784)	0.691 (0.569-0.806)
			Grade 3+ Cardiac Toxicity			
No Imbalance Correction	0.568 (0.417-0.722)	0.71 (0.574-0.83)	0.476 (0.344-0.611)	0.466 (0.343-0.609)	0.527 (0.405-0.659)	0.52 (0.378-0.652)
SMOTE-Tomek	0.616 (0.486-0.74)	0.717 (0.564-0.853)	0.631 (0.504-0.76)	0.558 (0.399-0.705)	0.661 (0.498-0.82)	0.689 (0.541-0.821)
SMOTE-ENN	0.639 (0.51-0.765)	0.721 (0.57-0.854)*	0.662 (0.553-0.777)	0.556 (0.421-0.687)	0.712 (0.57-0.838)	0.675 (0.528-0.809)
GAN	0.543 (0.403-0.702)	0.634 (0.51-0.764)	0.512 (0.353-0.645)	0.573 (0.462-0.661)	0.621 (0.506-0.728)	0.502 (0.358-0.648)

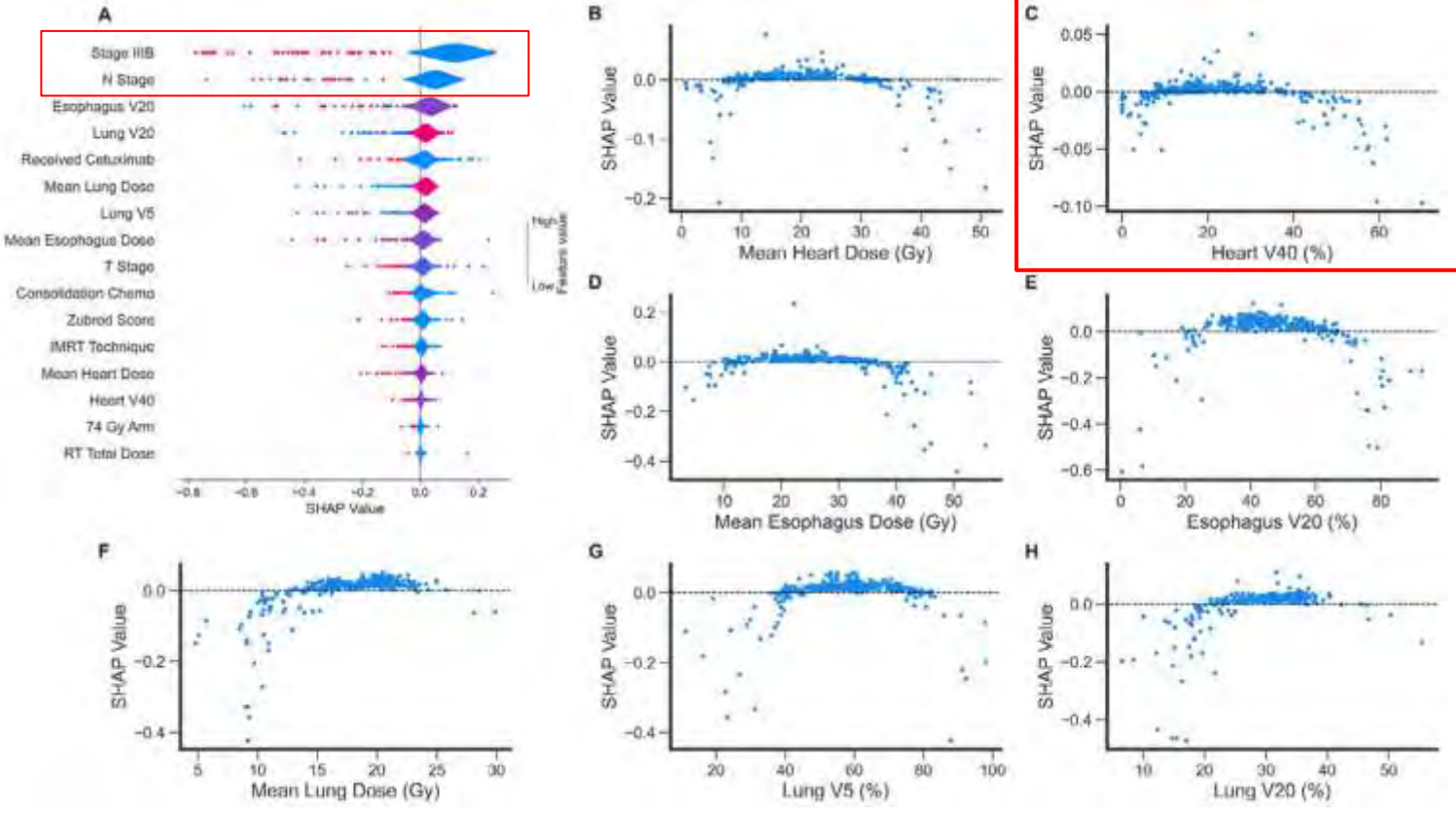
*Best performing model

Grade 3+ esophageal toxicity



Grade 3+ cardiac toxicity

Compare to PMID 38935373, which recommends heart V40 < 20%.



JAMA Oncology | Brief Report

Long-Term Prospective Outcomes of Intensity Modulated Radiotherapy for Locally Advanced Lung Cancer

A Secondary Analysis of a Randomized Clinical Trial

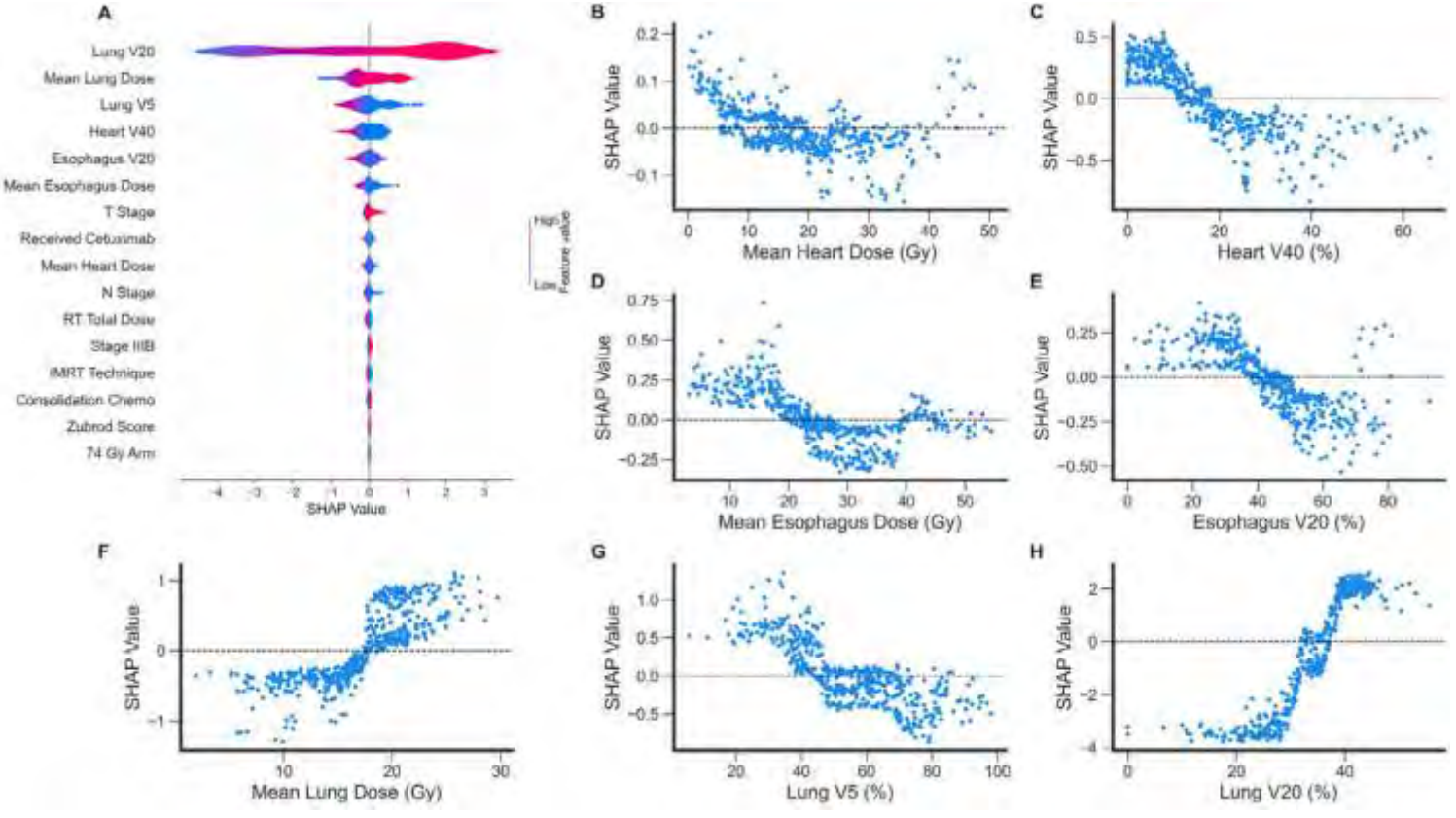
Stephen G. Chun, MD; Chen Hu, PhD; Ritsuko U. Komaki, MD; Robert D. Timmerman, MD; Steven E. Schild, MD; Jeffrey A. Bogart, MD; Michael C. Dobelbower, MD; Walter Bosch, DSc; Vivek S. Kavadi, MD; Samir Narayan, MD; Puneeth Iyengar, MD, PhD; Clifford Robinson, MD; Jan Rothman, MD; Adam Raben, MD; Mark E. Augspurger, MD; Robert M. MacRae, MD; Rebecca Paulus, BS; Jeffrey D. Bradley, MD

Heart V40 (< 20%) had better OS than V40 (20%) (median [IQR], 2.5 [2.1-3.1] years vs 1.7 [1.5-2.0] years; P < .001).

On multivariable analysis, heart V40 (20%), was associated with worse OS(hazard ratio, 1.34 [95% CI, 1.06-1.70]; P = .01), whereas lung V5 and age had no association with OS.

➢ Different models will give different dosimetric cutoffs. In this paper, stage (IIIB) and N-stage were not included as variables, which may also correlate with survival.

Grade 3+ pulmonary toxicity



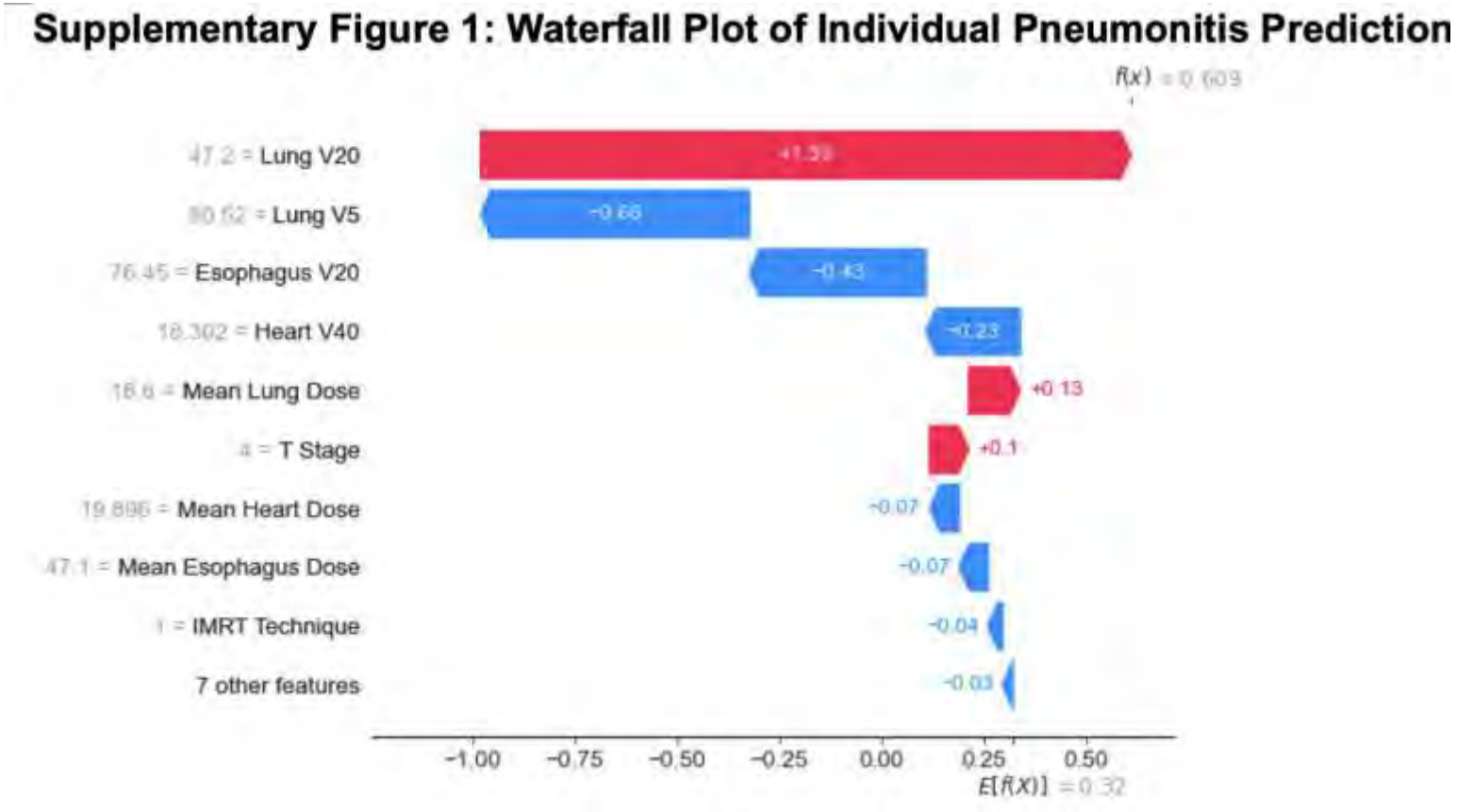
Dosimetric Predictors of Toxicity

Table 1 **Dosimetric predictors of toxicity from machine learning and logistic regression models**

	Machine learning		Univariable logistic regression		Multivariable logistic regression	
Category	**OR (95% CI)**	***P* value**	**OR (95% CI)**	***P* value**	**OR (95% CI)**	***P* value**
Grade ≥3 pulmonary toxicity						
Mean lung dose > 18 Gy	2.070 (1.508-2.967)	<.001	2.467 (1.049-5.800)	.038	1.907 (0.871-6.366)	.294
Lung V5 > 47%	0.551 (0.370-0.771)	.002	1.875 (0.632-5.566)	.258	1.059 (0.276-4.058)	.934
Lung V20 > 37%	22.268 (12.575-37.199)	<.001	2.722 (1.034-7.163)	.043	3.704 (1.111-12.345)	.033
Grade ≥3 esophageal toxicity						
Mean esophagus dose > 34 Gy	1.159 (1.092-1.249)	<.001	4.006 (2.183-7.354)	<.001	2.745 (1.321-5.706)	.007
Esophagus V20 > 37%	1.072 (1.022-1.132)	.008	3.725 (1.308-10.603)	.014	2.285 (0.721-7.244)	.16

Abbreviations: V5 = lung volume receiving ≥5 Gy; V20 = lung volume receiving ≥20 Gy.

Providing insight to high V20 OR with ML



Current Guidelines Suggest to Constrain V5

National Comprehensive Cancer Network®

NCCN Guidelines Version 7.2024
Non-Small Cell Lung Cancer

Table 5. Normal Tissue Dose-Volume Constraints for Conventionally Fractionated RT with Concurrent Chemotherapy[†,‡]

OAR	Constraints in 30–35 fractions
Spinal cord	Max ≤50 Gy
Lung	V20 ≤35%–40%;[§] MLD ≤20 Gy
Heart	V50 ≤25%; Mean ≤20 Gy
Esophagus	Mean ≤34 Gy; Max ≤105% of prescription dose; V60 ≤17%; contralateral sparing is desirable
Brachial plexus	Median dose ≤69 Gy

No lung V5!

NCCN Guidelines Version 4.2024
Esophageal and Esophagogastric Junction Cancers

Lungs[b]
- V_{40Gy} ≤10%
- V_{30Gy} ≤15%
- V_{20Gy} ≤20%
- V_{10Gy} ≤40%
- V_{05Gy} ≤50%
- Mean <20 Gy

Spinal Cord
- Max ≤45 Gy

Bowel
- Max dose <54 Gy
- V_{45Gy} <195 cc

Heart
- V_{30Gy} ≤30% (closer to 20% preferred)
- Mean <30 Gy (closer to 26 Gy preferred)

Left Kidney, Right Kidney (evaluate each one separately):
- V_{20Gy} ≤33%
- Mean <18 Gy

Liver
- V_{30Gy} ≤33%
- Mean <25 Gy

Stomach
- Mean <45 Gy
- Max dose <54 Gy

What does this mean for planning?

Liberalizing lung V5 with full arcs may be protective of radiation pneumonitis by enabling lung V20 and mean lung dose ALARA.

Machine learning with XAI is one way to enable us to understand dosimetric drivers of toxicity while also accounting for clinical factors.

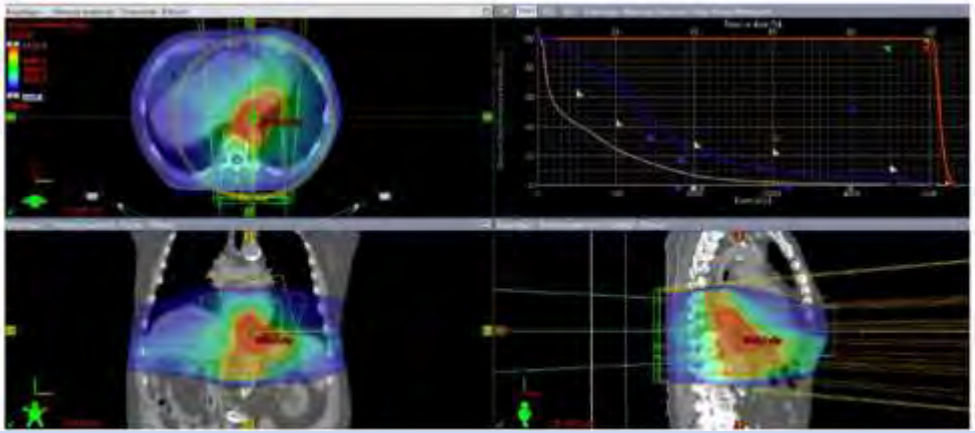
Many historic dose constraints may be upended by similar analyses in the future.

The Following Slides are from an Ongoing Study to be submitted in 2024. Please do not share publicly or on social media.

Future Directions: Visualization of V5 impact

Esophagus Case #1

Distal/GEJ, 2 full arcs, 50/45 SIB, elective GH/celiac, lung V5 42%, lung V10 25%, lung V20 7%, heart Dmax 53Gy, heart mean 16Gy



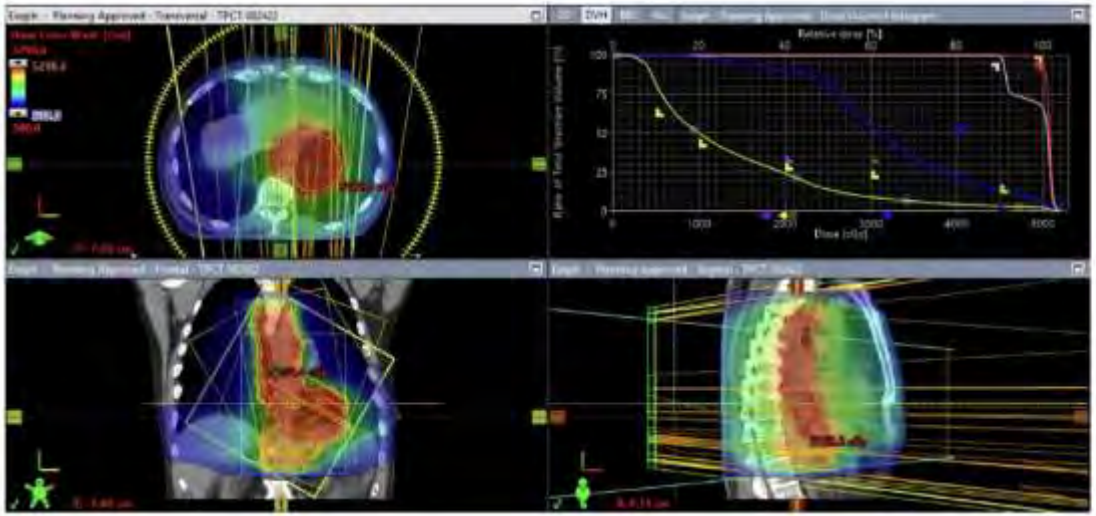
Plan			Esophagus
Total Dose			5000.0 cGy
Clinical Goal Summary			0 1 21
GTV_PET	P2	V 100.0 % ≥ 100.0 %	100.00 %
PTV50	P2	V 100.0 % ≥ 95.0 %	98.48 %
PTV_45_Eval	P2	V 4500 cGy ≥ 95.0 %	97.95 %
Bowel Bag	P3	Dmax ≤ 5200 cGy	3969.08 cGy
	P3	V 4500 cGy ≤ 135.0 cm³	0.00 cm³
	P3	Dmax ≤ 5500 cGy	3969.08 cGy
Heart	P3	Dmax ≤ 5200 cGy	5280.49 cGy
	P3	Dmean ≤ 3200 cGy	1607.52 cGy
	P3	V 4000 cGy ≤ 50.0 %	5.99 %
Kidney_L	P3	V 1800 cGy ≤ 15.0 %	3.54 %
	P3	V 1400 cGy ≤ 30.0 %	9.21 %
	P3	Dmean ≤ 1800 cGy	586.28 cGy
Liver	P3	V 3000 cGy ≤ 30.0 %	6.76 %
	P3	Dmean ≤ 2100 cGy	1740.75 cGy
Lungs	P3	V 3000 cGy ≤ 20.0 %	2.07 %
	P3	V 2000 cGy ≤ 25.0 %	6.65 %
	P3	V 1000 cGy ≤ 40.0 %	24.69 %
	P3	V 500 cGy ≤ 60.0 %	41.72 %
	P3	Dmean ≤ 2000 cGy	674.63 cGy
SpinalCanal	P1	Dmax ≤ 4500 cGy	2681.34 cGy
Stomach-PTV	P3	Dmax ≤ 5000 cGy	4733.25 cGy
	P3	Dmean ≤ 4000 cGy	3337.09 cGy

➢ Favor full arcs for central targets. Lung V5 may be easily met if GEJ lesion without much middle esophagus.

Future Directions: Visualization of V5 impact

Esophagus Case #2

Distal/GEJ, 4 full arcs, 50/45 SIB, elective GH/celiac, lung V5 81%, lung V20 24%, heart Dmax 52Gy, heart mean 32Gy



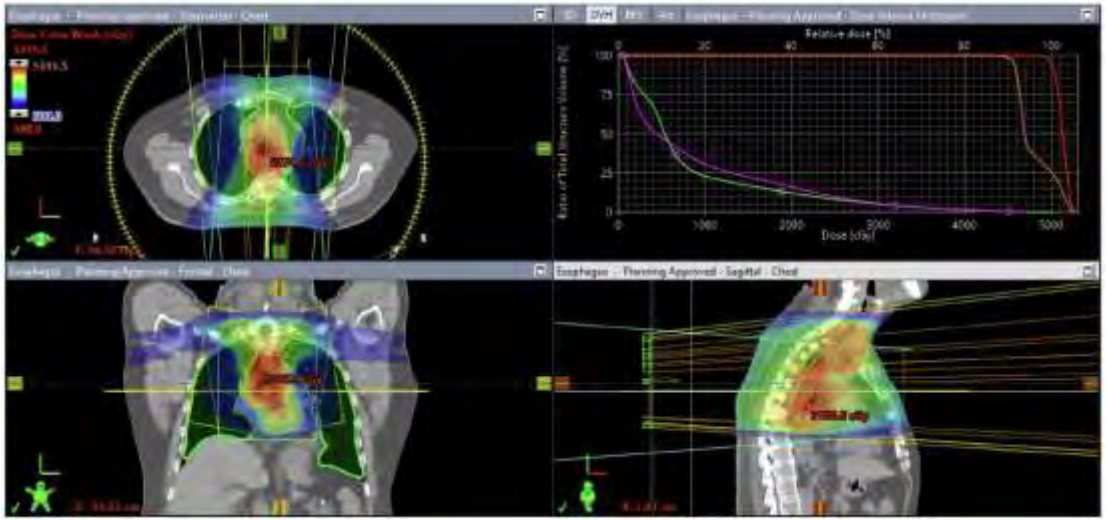
Plan			Esoph
Total Dose			5000.0 cGy
Clinical Goal Summary			1 3 15
PTV4500	P2	V 4500 cGy ≥ 95.0 %	89.58 %
PTV5000	P2	V 100.0 % ≥ 95.0 %	95.00 %
BowelBag	P3	Dmax ≤ 5200 cGy	5078.00 cGy
	P3	V 4500 cGy ≤ 135.0 cm³	13.52 cm³
Heart	P3	Dmax ≤ 5200 cGy	5225.71 cGy
	P3	Dmean ≤ 3200 cGy	3205.99 cGy
	P3	V 4000 cGy ≤ 50.0 %	21.36 %
ITVp	P2	V 100.0 % ≥ 100.0 %	100.00 %
		The volume that receives 100.0 % of the prescribed dose is larger than or equal to 100.0 % of the structure volume	
Kidneys_partial	P3	V 2000 cGy ≤ 30.0 %	12.17 %
Liver	P3	V 3000 cGy ≤ 30.0 %	3.34 %
	P3	Dmean ≤ 2100 cGy	1295.58 cGy
Lungs	P3	V 3000 cGy ≤ 20.0 %	8.95 %
	P3	V 2000 cGy ≤ 25.0 %	24.07 %
	P3	V 1000 cGy ≤ 40.0 %	49.98 %
	P3	V 500 cGy ≤ 60.0 %	81.37 %
	P3	Dmean ≤ 2000 cGy	1395.21 cGy
SpinalCanal	P1	Dmax ≤ 4500 cGy	4089.05 cGy

➢ Favor full arcs for central targets. Lung V5 < 50% is impossible to meet for middle esophagus or longer esophageal length fields.

Future Directions: Visualization of V5 impact

Esophagus Case #3

Upper SCC, 2 full arcs, 50/45 SIB, no positive nodes, elective SCV/mediastinal, 8.5cm (PTV50)/15.0cm (PTV45) CC length, lung V5 55%, lung V20 12%, mean lung 9Gy, heart Dmax 47Gy, heart mean 9Gy



➢ This case will be utilized to demonstrate how lateral avoidance sectors to meet lung V5 actually leads to increased cardiac exposure while inhibiting MLD/Lung V20 ALARA.

Future Directions: Visualization of V5 impact

Two lung cases are utilized.

- First case: N1 disease, partial arcs.
- Second case: N2 disease, including a bulky station 7 (behind heart)

➢ Be on the lookout for a PRO publication with visual representations of low dose trade-offs using sample esophageal and lung examples cases in 2024/2025!

Some personal thoughts

Lung V5 has never been investigated in prospective fashion.

Dosimetry strictly instructed to keep lung V5 out of the optimizer.

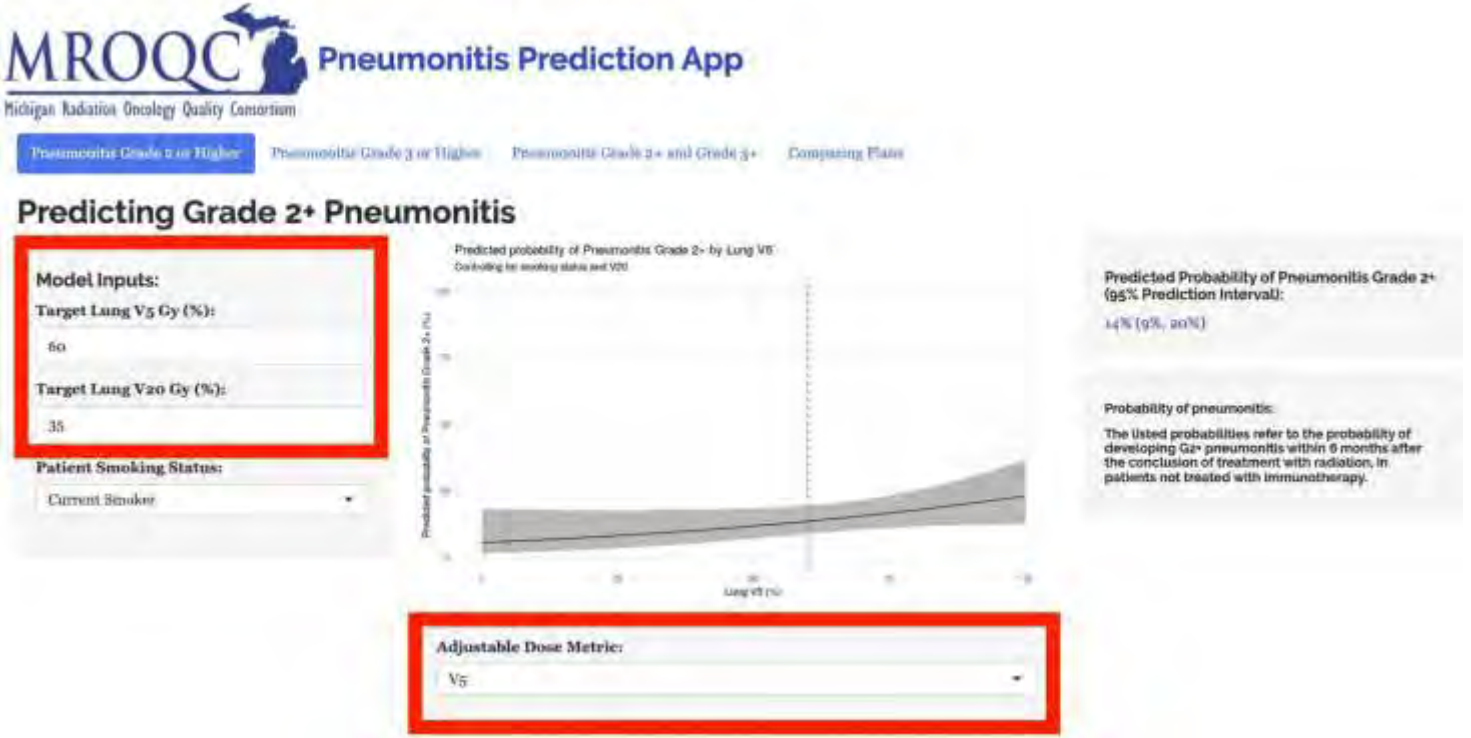
Lung scorecards: "Lung V5@" instead of "Lung V5 ≤ [Value]"

I do not flinch to accept lung V5 >80-90% if targets located behind the heart, especially in context of increasing craniocaudal length of target volumes.

➢ The art of what we do: Seeing what happens to low dose during planning when intermediate or high dose metrics are saturated.

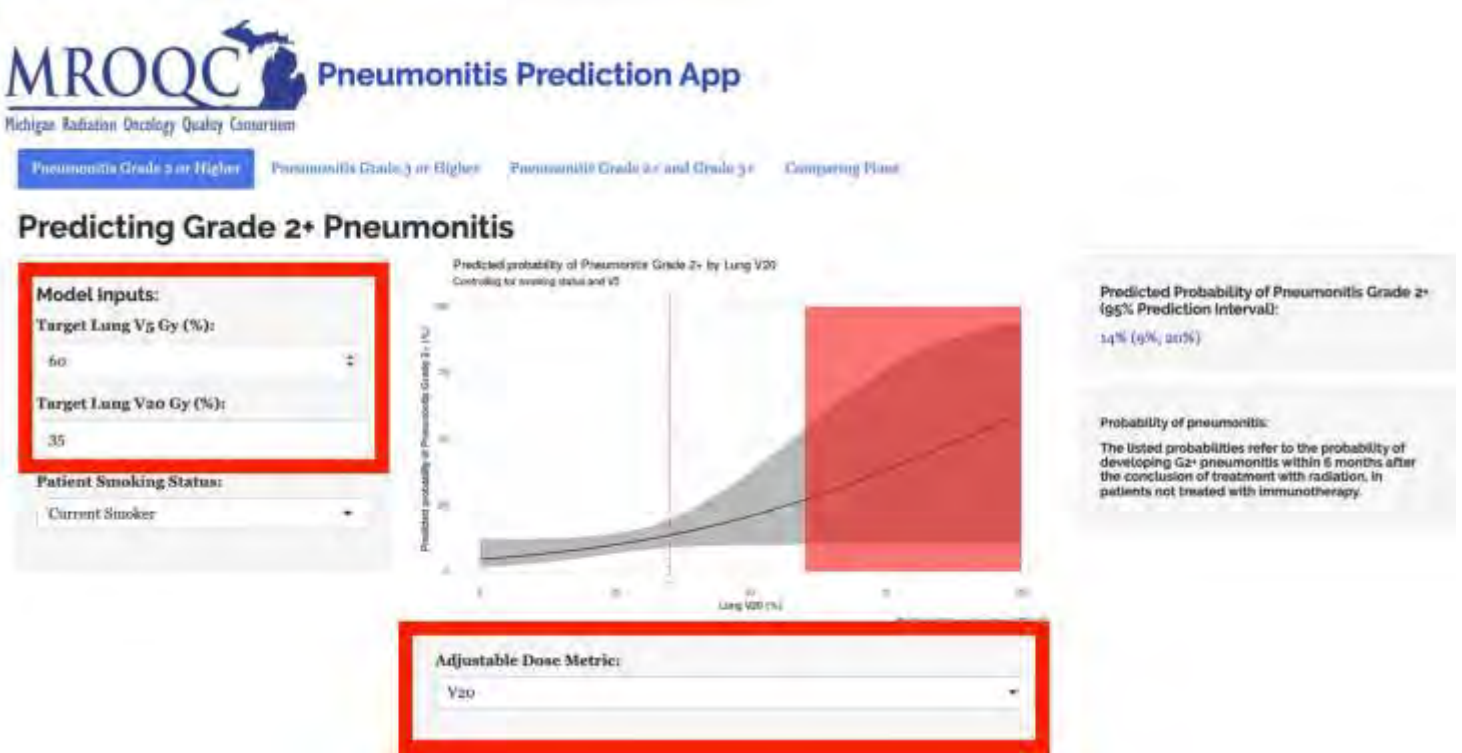
Real world data (MROQC) supports this approach

https://ppa.mroqc.org



Real world data (MROQC) supports this approach

https://ppa.mroqc.org



Summary

XAI aims to make AI decisions and processes transparent and understandable to humans.

SHAP provide a detailed explanation of how each feature impacts a model's output by illustrating which clinical and dosimetric factors contribute most to the predicted risk of toxicity.

When constraining low-dose exposure in context of modern IMRT, there may be an increased risk of toxicity due to higher integral intermediate and high dose.

Liberalizing low dose wash allows for intermediate and high dose ALARA (e.g., MLD, lung V20). Intermediate and high dose drive toxicity.

➢ For when targets are located behind heart and full arcs are utilized, low dose trade-offs may become important.

Questions?

Free Rad Onc Tools

Collection of nearly 3,000 constraint metrics

https://RadOncCalc.RadOncReview.org or Rad Onc Calc (Apple, Android)

Reirradiation EQD2 Calculator, including Michigan calculator

https://www.CancerRetreatment.org