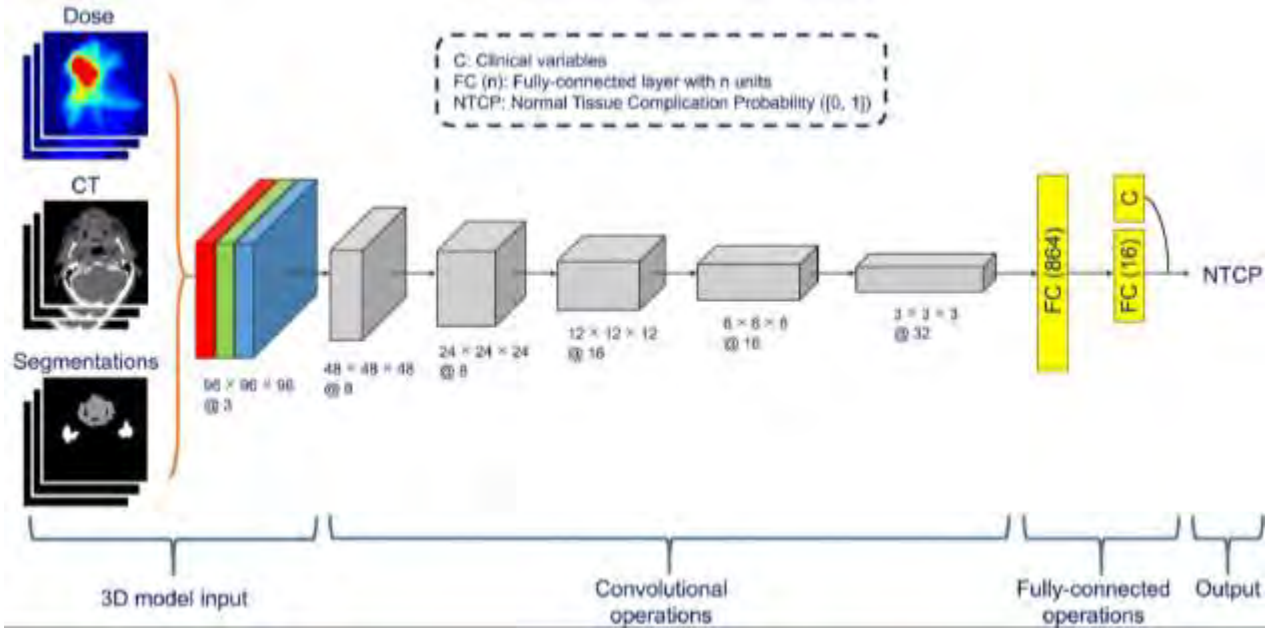


FULLER LAB

Advances in AI-Based Prediction Models: The Head and Neck Cancer Use-Case

DOI: 10.6084/m9.figshare.26817322

THE UNIVERSITY OF TEXAS
MD Anderson Cancer Center
Making Cancer History®

cdfuller@mdanderson.org

Funding Acknowledgment / Disclosures

- NIH NIDCR Small Research Grants for Oral Health Data Analysis and Statistical Methodology Development (R03DE033550)
- NIH-NCI Postdoctoral Training Program (T32CA261856)
- NIH- NIDCR Prospective Observational or Biomarker Validation Study Cooperative Agreement (U01DE032168)
- NIH-NCI Joint NSF/NIH Smart Connected Health Program Award (R01CA257814)
- NIH- NCI/BD2K Early-Stage Technologies in Biomedical Computing, Informatics, and Big Data Science Award (R01CA214825)
- NIH- NCI Joint NSF/NIH Quantitative Approaches to Biomedical Big Data (R01CA225190)
- NIH-NCI Early Phase Clinical Trials in Imaging and Image-Guided Interventions (R01CA218148)
- NIH- NIDCR Academic-Industrial Partnerships to Translate and Validate in vivo Cancer Imaging Systems Award (R01DE028290)
- NIH- NIBIB Research Education Programs for Residents and Clinical Fellows Award (R25EB025787)
- NIH NCI Parent Research Project Grant (R01CA258827)
- NIH NCI Early Phase Clinical Trials in Imaging and Image-Guided Interventions Program (1R01CA218148)
- NIH-NCI Cancer Center Support Grant (CCSG) (P30CA016672)
- NIH NCI Small Business Innovation Research Grant Program sub-award (R43CA254559)
- NIH HuBMAP Integration, Visualization & Engagement (HIVE) Initiative (OT2OD026675) sub-award
- NIH NIDCR Exploratory/Developmental Research Grant Program (R21DE031082)
- Patient-Centered Outcomes Research Institute (PCS-1609-36195) sub-award from Princess Margaret Hospital
- National Science Foundation (NSF) Division of Civil, Mechanical, and Manufacturing Innovation (CMMI) Award (1933369)
- Elekta AB/MD Anderson MRI-LinAc Consortium Seed Grant*
- Elekta AB Travel support & Honoraria*
- Licensing from the University of Texas from Kallisio, Inc.
- Honoraria/in-kind registration reimbursement from professional societies: ASCO, AAPM, ESTRO, ASTRO, RANZCR
- Charles & Daneen Steifel Oropharynx Research Fund

Federal funder | Industry/For-Profit | Philanthropic

FULLERLAB

David Rosenthal,
MD
Professor/Section Chief

Bill Morrison,
MD
Professor

Adam Garden,
MD
Professor

Steven Frank,
MD
Professor

Brandon Gunn
MD
Professor

Dave Fuller,
MD, PhD
Professor

THE UNIVERSITY OF TEXAS
MD Anderson ~~Cancer~~ Center
Making Cancer History

Radiation Oncology Head and Neck Section

Jack Phan,
MD, PhD
Assoc. Professor

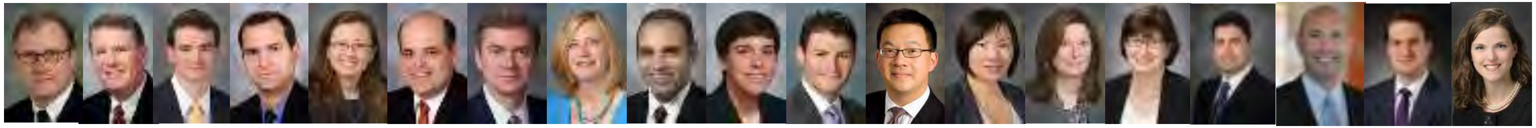
Mike Spiotto
MD, PhD
Assoc. Professor

Jay Reddy
MD
Asst. Professor

Amy Moreno
MD,
Asst. Professor

Anna Lee
MD, MPH
Asst. Professor

MDACC Head and Neck Team



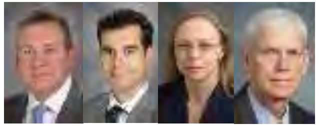
Head and Neck Surgery

Thoracic/Head and Neck Medical Oncology

Neuroradiology

Radiation Oncology/Medical Physics

Pathology



Oncologic Dentistry

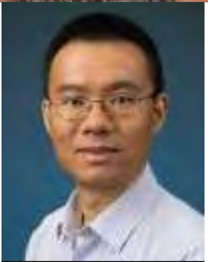
THE UNIVERSITY OF TEXAS
MD Anderson ~~Cancer~~ Center
Making Cancer History®

SMART-ACT: Spatial Methodologic Approaches for Risk Assessment and Therapeutic Adaptation in Cancer Treatment

NSF 1557679, R01CA258827, R01CA225190, R01CA214825

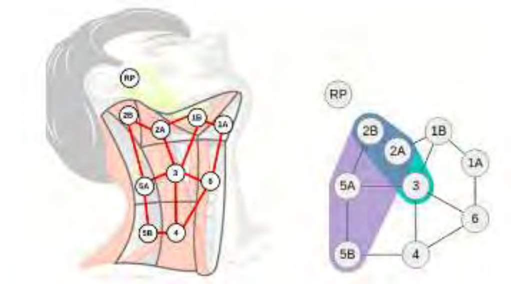
Liz Marai, PhD
Computer Science,
UIC

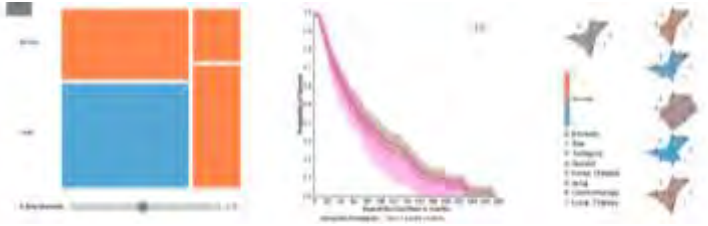
Guadalupe Canahuate, PhD
Computer Science,
UIowa



Xinhua Zhang, PhD
Computer Science,
UIC

Dave Fuller, MD, PhD
Radiation Oncology
MDACC





MD Anderson Multi-disciplinary Symptom Working Group

Stephen Lai
MD, PhD
Head
and Neck Surgery

Kate Hutcheson
PhD
Speech Pathology

Amy Moreno, MD
Radiation Oncology

Abdallah Mohamed
MD, MSc
Radiation Oncology

Jihong Wang
PhD
Radiation Oncology

Dave Fuller
MD, PhD
Radiation Oncology

R01DE025248; U01DE032168; R01CA218148; R03CA188162; R21CA226200; R01CA271223; R21DE031082; K01DE030524

John Christodouleas
MD
Elekta AB/UPenn

Dave Fuller
MD, PhD
MDACC

NIH Academic Industrial Partnership
(R01 DE028290-01)

Elekta

THE UNIVERSITY OF TEXAS
MD Anderson ~~Cancer~~ Center
Making Cancer History

PHILIPS

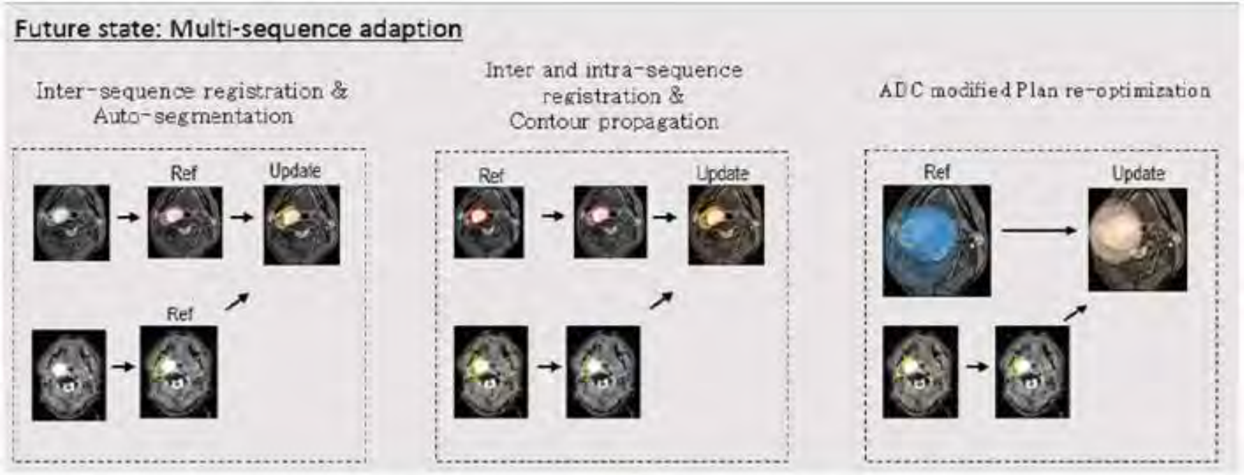


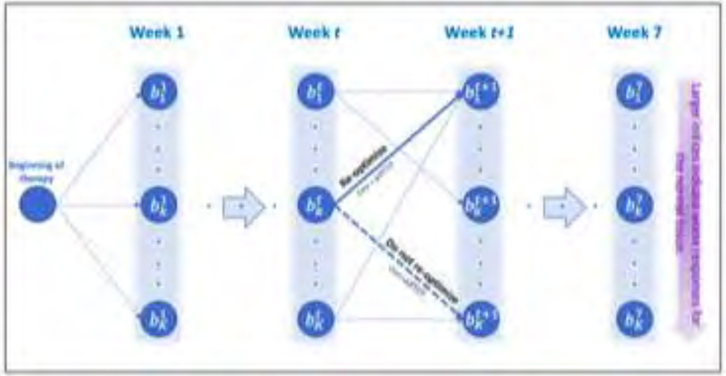
Figure 8: Diagram illustrating the conceptualized Elekta software development pipeline.

NSF-NIH Smart-Connected Health Program
Rice-MDACC Operations Research in Oncology

NSF 1933369, R01CA257814

Andrew Schaefer, PhD
Computational
and Applied Math
RiceU

Dave Fuller, MD, PhD
Radiation Oncology
MDACC



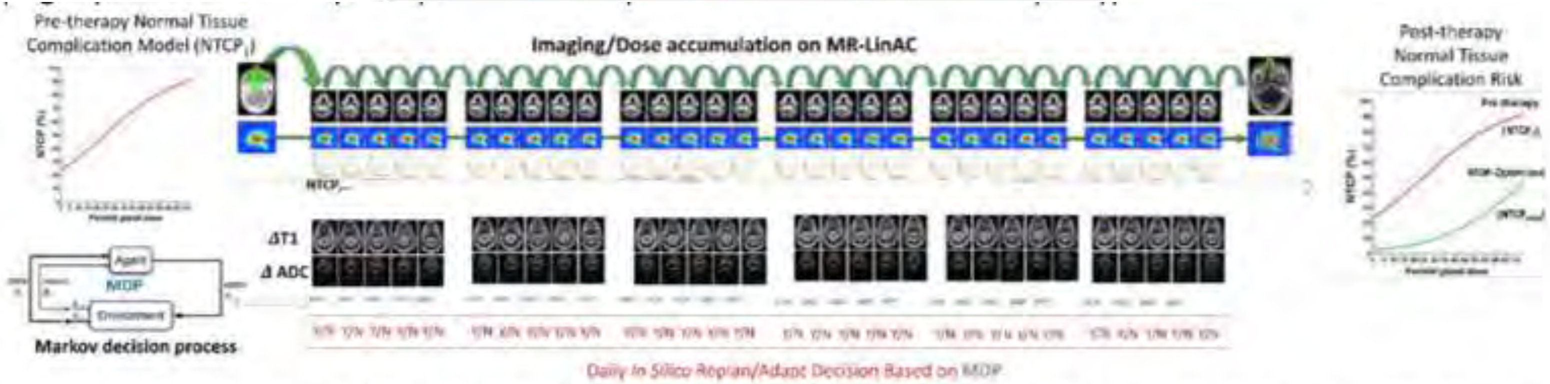


Figure 6. Graphical abstract of Aim 3, showing dose/NTCP and image implementation for adaptive radiotherapy decision support.

Image Guided Cancer Therapy Program

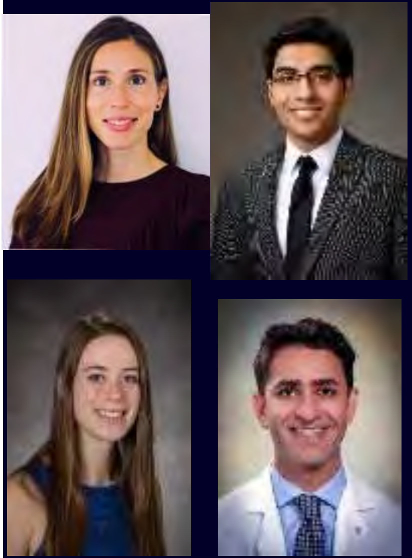
T32CA261856

Kristy Brock, PhD
Director,
IGCTR Program

Stephen Lai,
MD, PhD

Dave Fuller,
MD, PhD

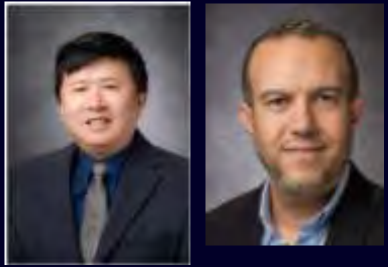
Physicians



Post-docs

FULLERLAB

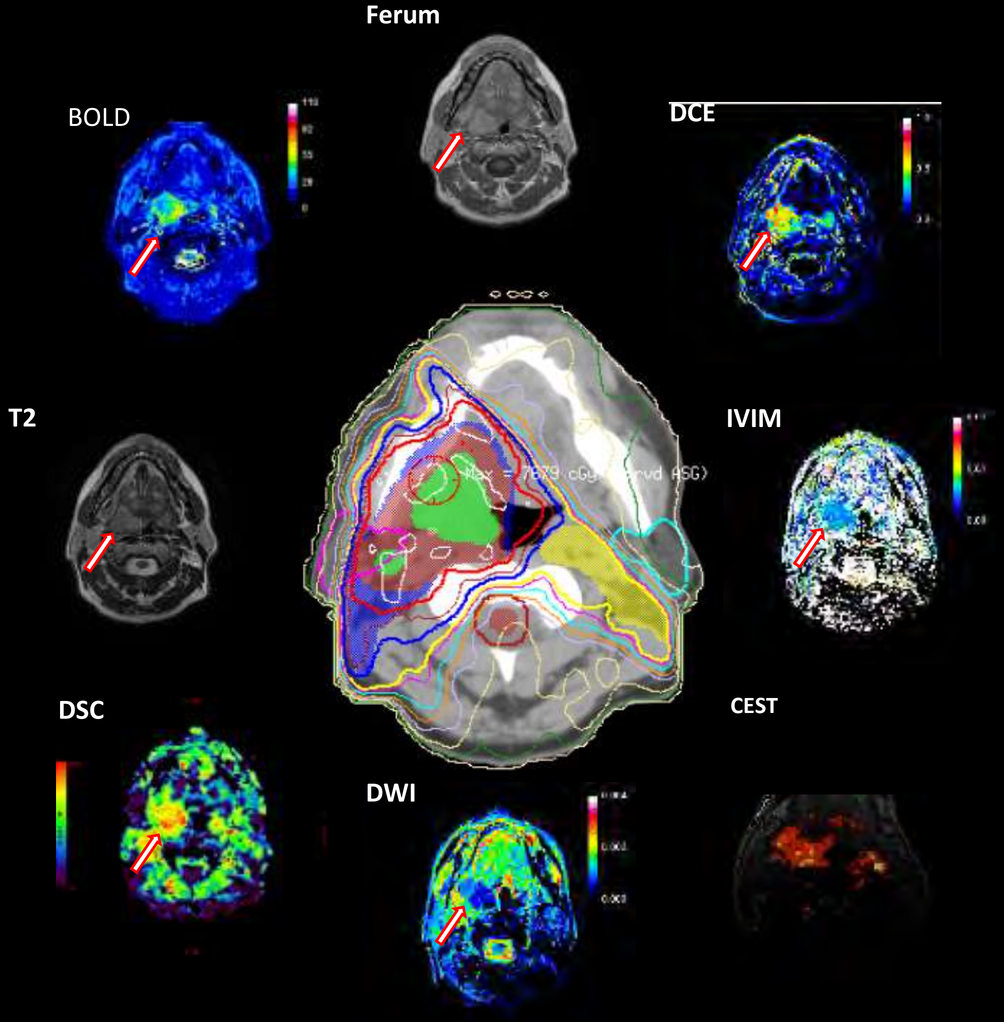
Grad students



Computational Scientists

Lab Manager

For radiation oncologists, *spatial* dose/response data is what separates us from other cancer paradigms



FULLERLAB

1954 PhD Berkeley (mathematics)

1960 -1967 UCLA (mathematics)

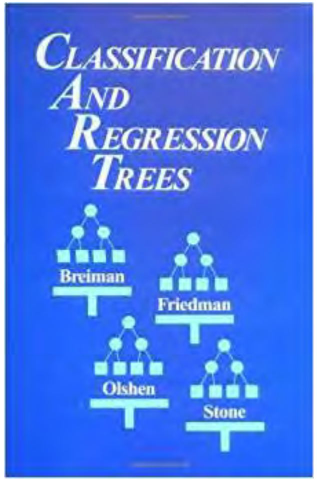
1969 -1982 Consultant

1982 - 1993 Berkeley (statistics)

1984 “Classification & Regression Trees”
(with Friedman, Olshen, Stone)

1996 “Bagging”

2001 “Random Forests”



RECURSIVE PARTITIONING ANALYSIS (RPA) OF PROGNOSTIC FACTORS IN THREE RADIATION THERAPY ONCOLOGY GROUP (RTOG) BRAIN METASTASES TRIALS

LAURIE GASPAR, M.D.,* CHARLES SCOTT, M.S.,† MARVIN ROTMAN, M.D.,‡ SUCHA ASBELL, M.D.,§ THEODORE PHILLIPS, M.D.,¶ TODD WASSERMAN, M.D.,# W. GILLIES MCKENNA, M.D., Ph.D.** AND ROGER BYHARDT, M.D.††

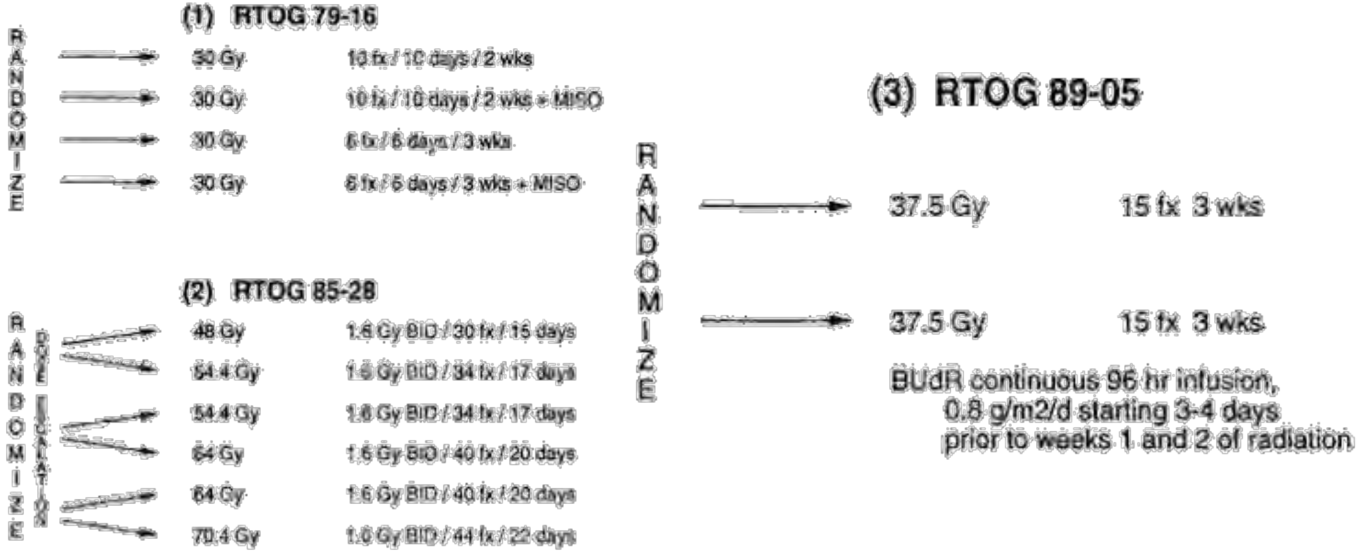


Fig. 1. Protocol schemas.

Int. J. Radiation Oncology Biol. Phys., Vol. 37, No. 4, pp. 745–751, 1997

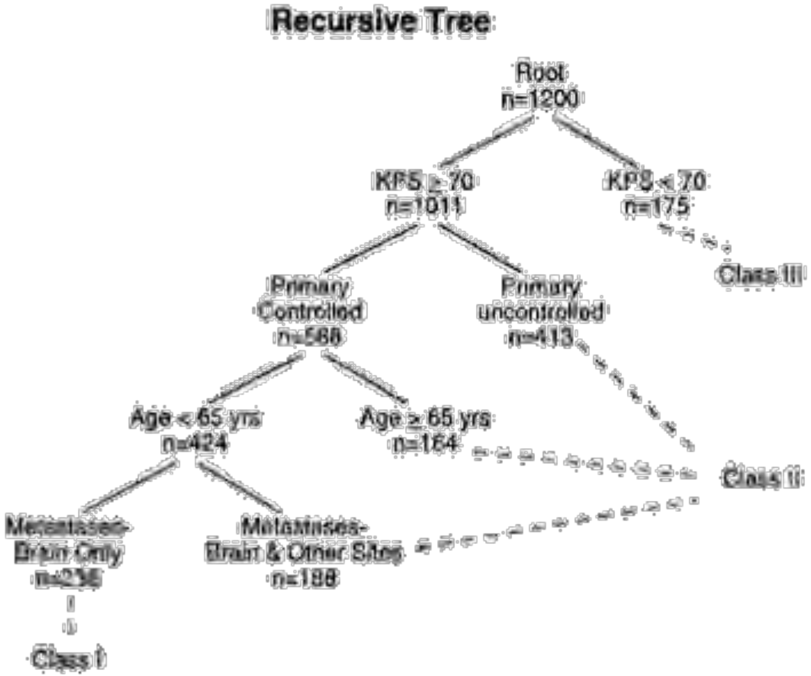


Fig. 2. Recursive tree.

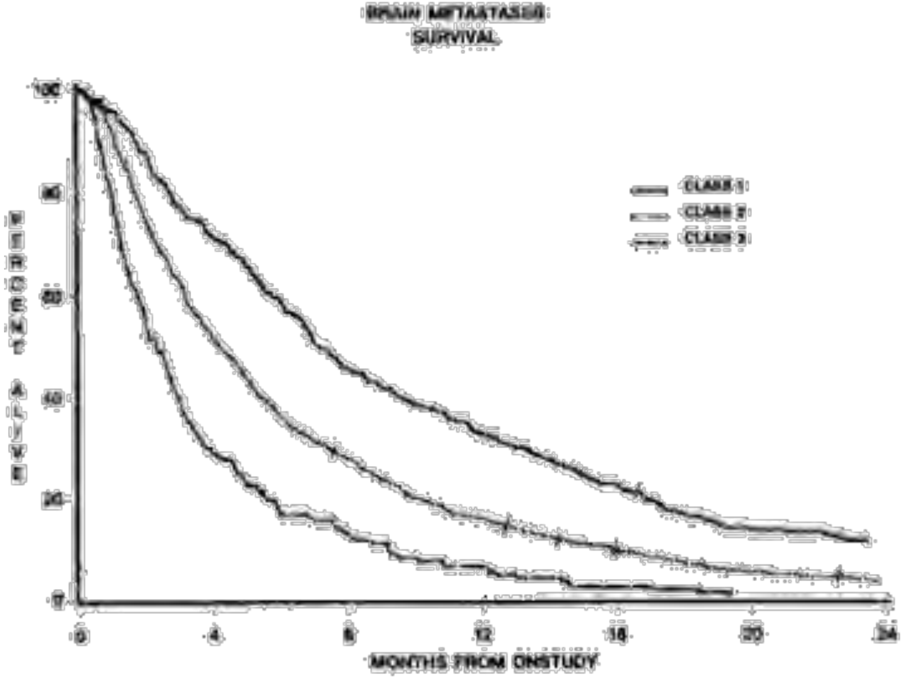


Fig. 3. Survival curves for Class I, II, III.

Int. J. Radiation Oncology Biol. Phys., Vol. 37, No. 4, pp. 745–751, 1997

Statistical Science
2001, Vol. 16, No. 3, 199–231

Statistical Modeling: The Two Cultures

Leo Breiman

Hypothesis Testers

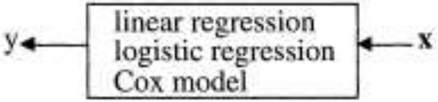
AI Modelers

The Data Modeling Culture

The analysis in this culture starts with assuming a stochastic data model for the inside of the black box. For example, a common data model is that data are generated by independent draws from

response variables = f(predictor variables, random noise, parameters)

The values of the parameters are estimated from the data and the model then used for information and/or prediction. Thus the black box is filled in like this:



Model validation. Yes–no using goodness-of-fit tests and residual examination.
Estimated culture population. 98% of all statisticians.

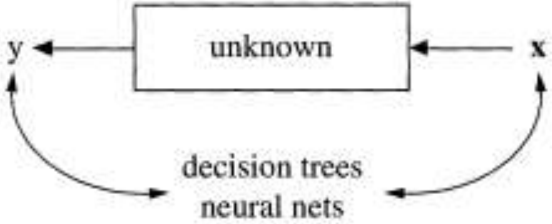
nature

oals in analy
able to predic
o future inpu
extract som
ssociating th
ables.

different appr

The Algorithmic Modeling Culture

The analysis in this culture considers the inside of the box complex and unknown. Their approach is to find a function $f(\mathbf{x})$—an algorithm that operates on $\mathbf{x}$ to predict the responses $\mathbf{y}$. Their black box looks like this:



Model validation. Measured by predictive accuracy.
Estimated culture population. 2% of statisticians, many in other fields.

Richard Bellman: "The curse of dimensionality"

Bellman, first editor of Mathematical Biosciences, was working in dynamic optimization

-Referred initially to issues that arise in higher-order analyses that are hard for humans to conceptualize as we move increase dimensions or add time-varying components

-Broadly, refers to typical increase in sparsity of data in high-dimensions and information reduction through dimensional summarization.

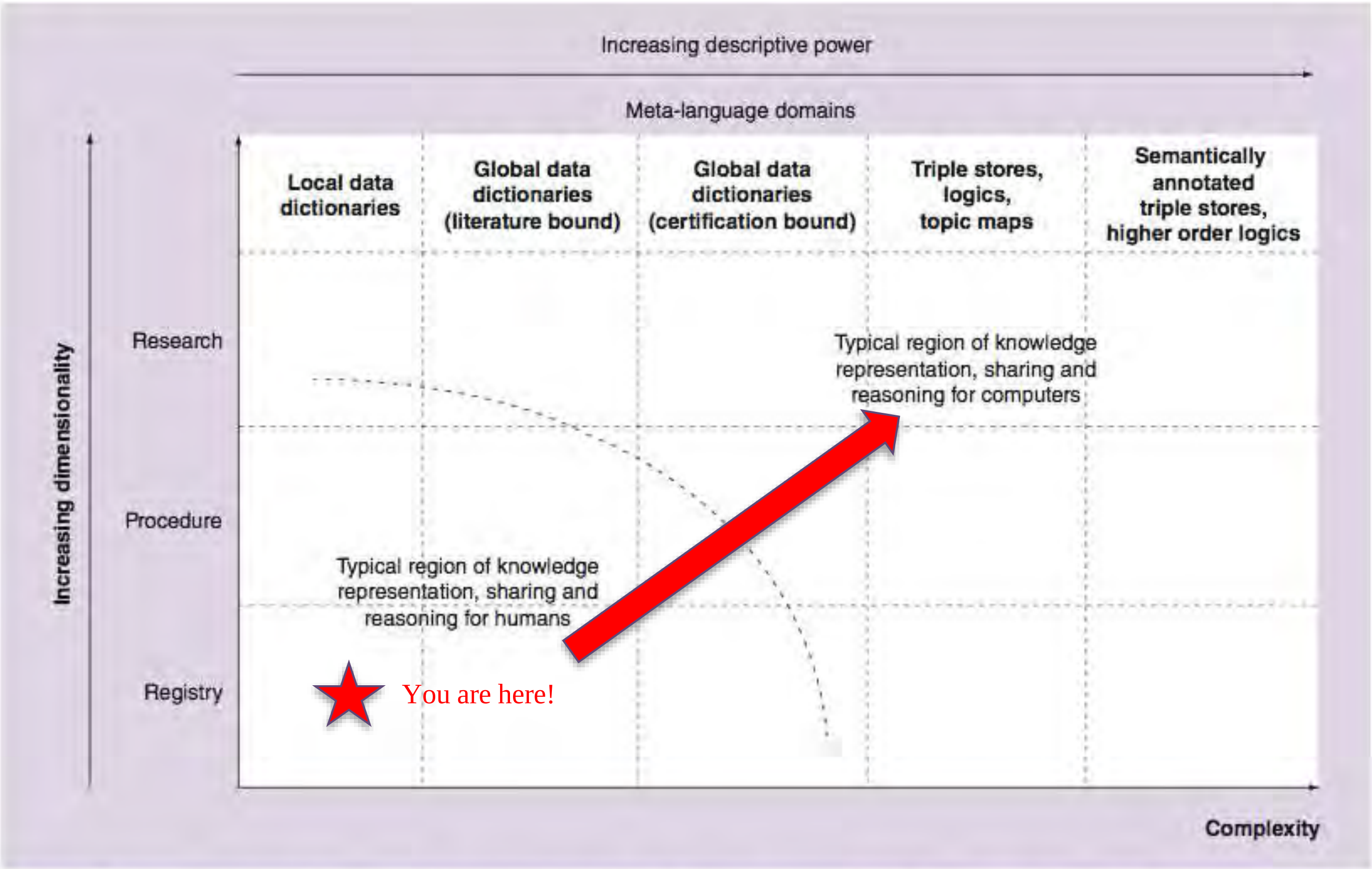
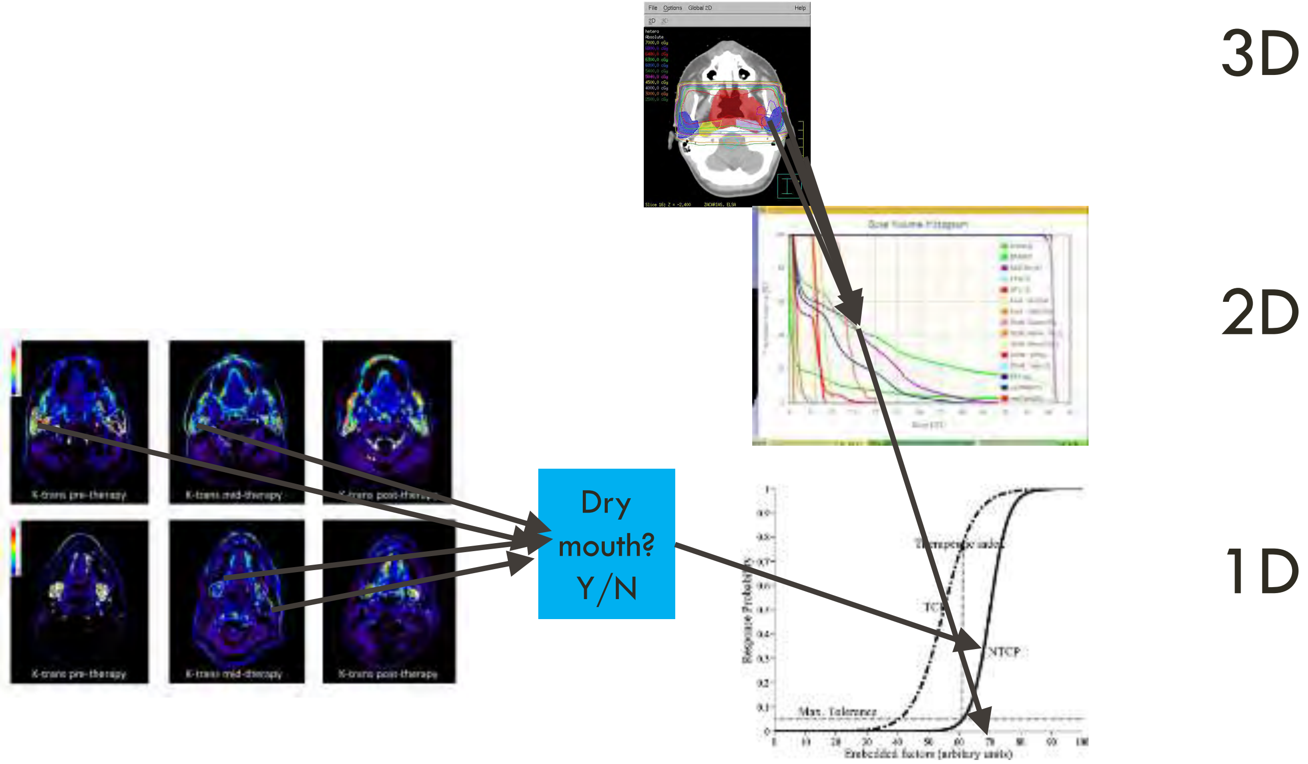


Figure 1. Possible evolution in knowledge representation, seen from the perspective of computer science, under a qualitative point of view.

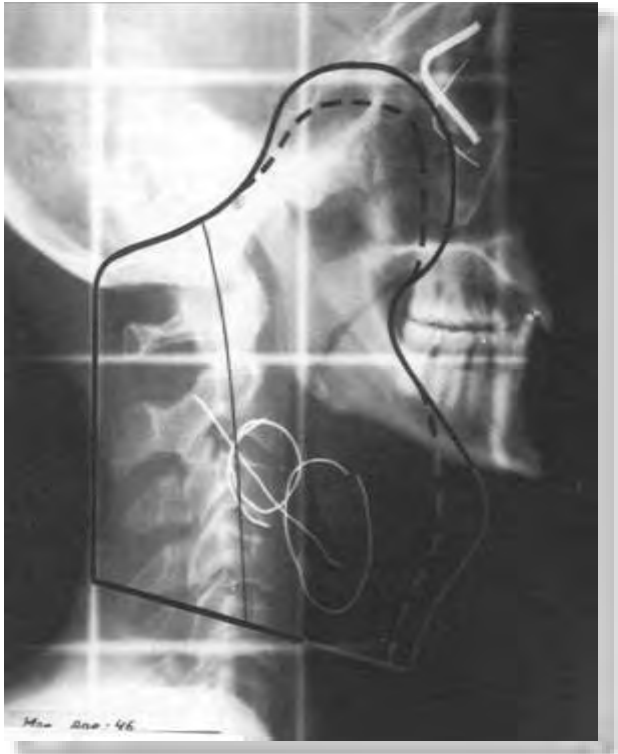
at MD Anderson

Example:
Information loss through summarization by dimensionality reduction

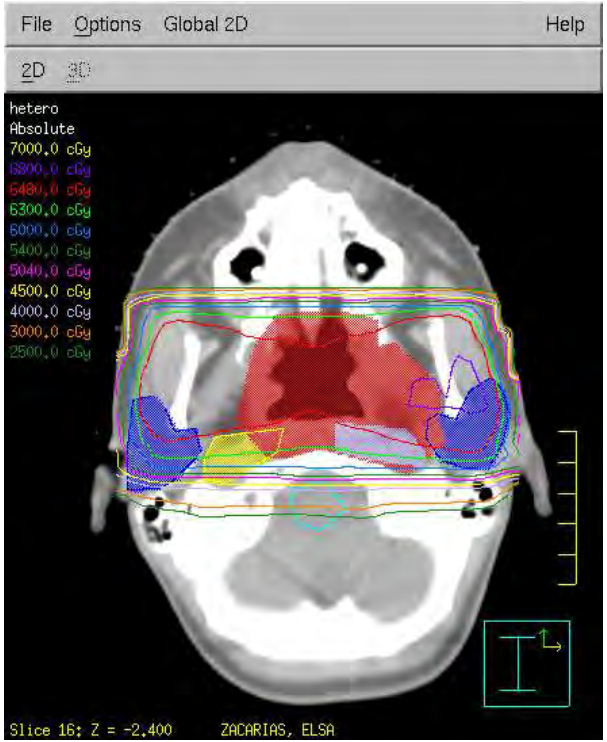


What NTCP models were built for…

1990



2000



PAROTID GLAND FUNCTION AFTER RADIOTHERAPY: THE COMBINED MICHIGAN AND UTRECHT EXPERIENCE

Tim Dijkema, M.D.,* Cornelis P. J. Raaijmakers, Ph.D.,* Randall K. Ten Haken, Ph.D.,† Judith M. Roesink, M.D., Ph.D.,* Pètra M. Braam, M.D., Ph.D.,* Anette C. Houweling, M.Sc.,* Marinus A. Moerland, Ph.D.,* Avraham Eisbruch, M.D.,† and Chris H. J. Terhaard, M.D. Ph.D.*

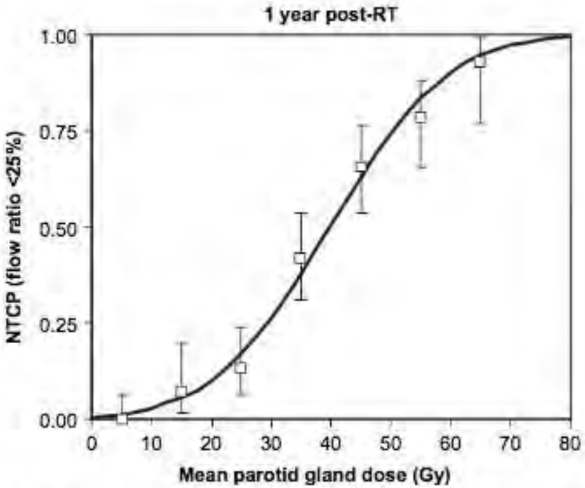


Fig. 3. Combined Michigan and Utrecht normal tissue complication probability (NTCP) curve as a function of the mean parotid gland dose. Clinical NTCP values (using mean dose bins of 10 Gy) are shown, including 95% confidence intervals. RT = radiotherapy.

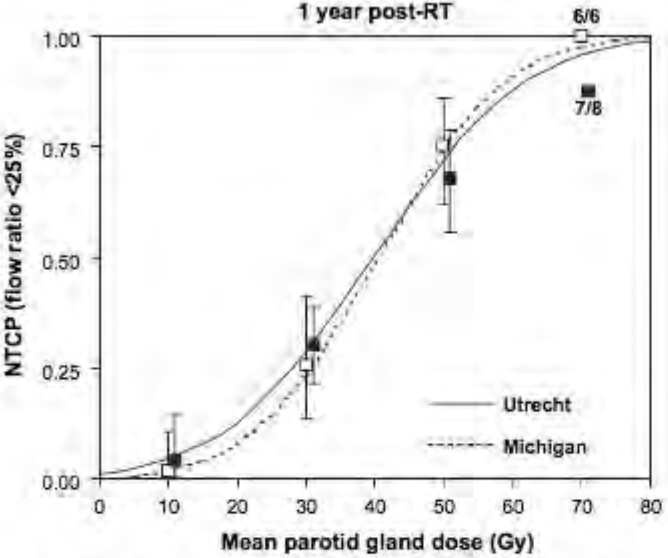


Fig. 2. Normal tissue complication probability (NTCP) curves as a function of the mean parotid gland dose for Michigan (dashed line) and Utrecht (solid line). Clinical NTCP values (using mean dose bins of 20 Gy) are shown for Michigan (open squares) and Utrecht (black squares), including 95% confidence intervals. RT = radiotherapy.

doi:10.1016/j.ijrobp.2009.07.1708

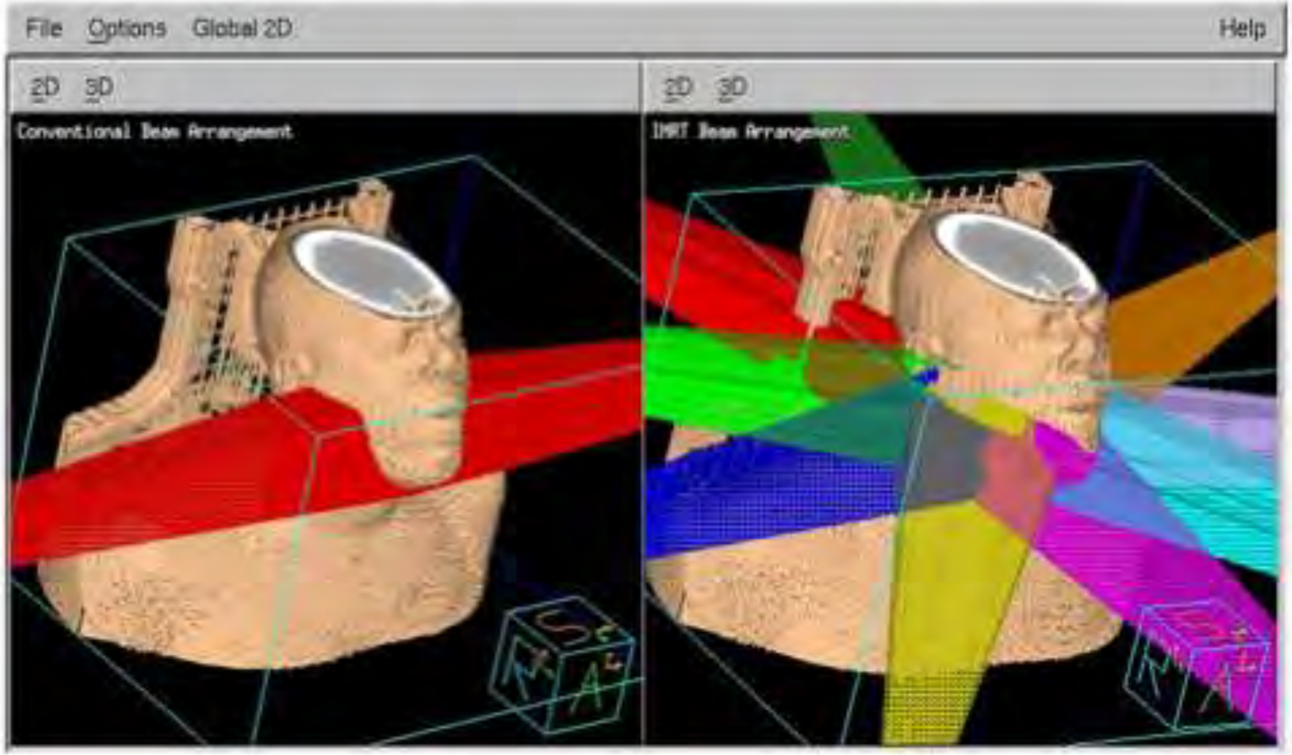
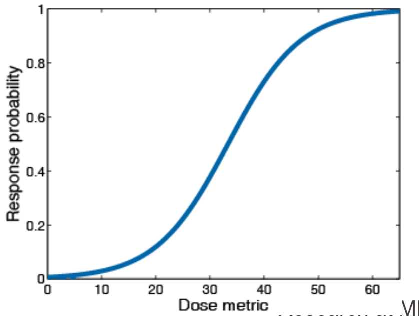


Fig. 1. Comparison of nontarget beam paths in intensity-modulated radiotherapy (top) vs. conventional three-dimensional technique (bottom).

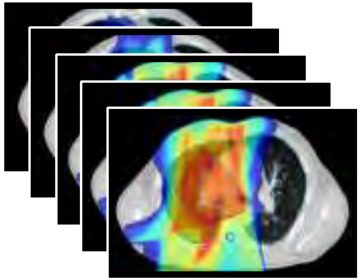
Standard phenomenological modelling methodology

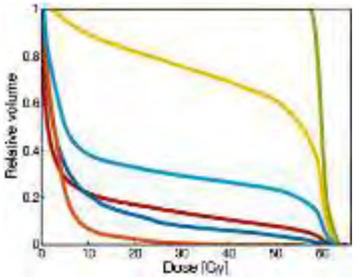
"Dose"

Generalised linear modelling



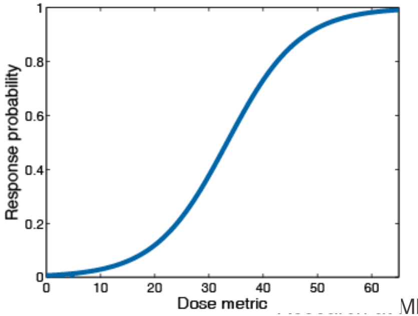
Standard phenomenological modelling methodology



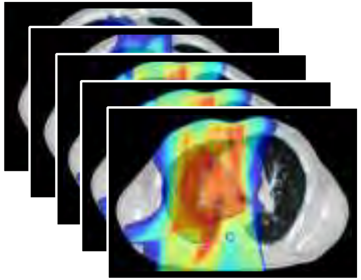


”Dose”

Generalised linear modelling



Standard phenomenological modelling methodology

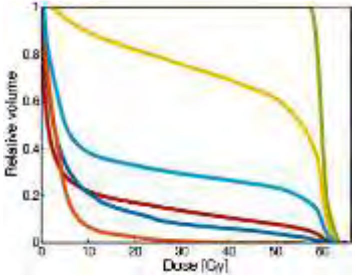


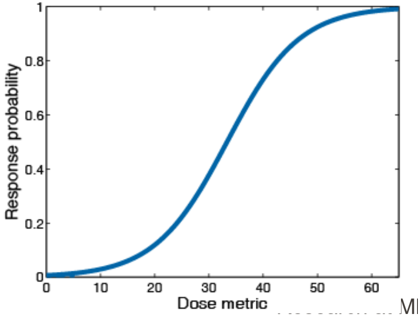
EUD

V_x

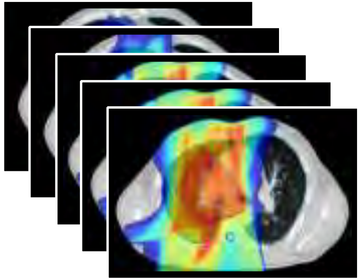
D_{mean}

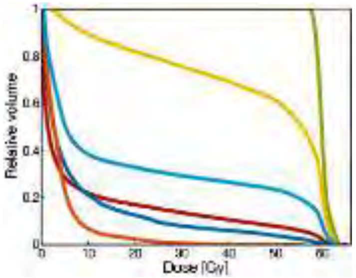
Generalised linear modelling





Standard phenomenological modelling methodology





EUD

V_x

D_{mean}

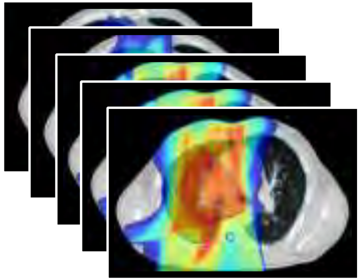
Reduce DVH to one (or a limited number of) dose metrics

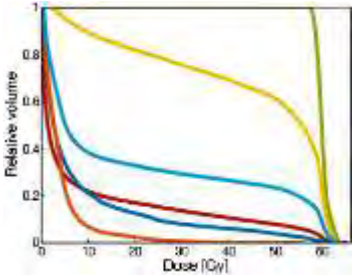
$$EUD = \left(\sum_k d_k^a \frac{v_k}{V_{tot}} \right)^{1/a}$$

$$V_x = \sum_k E(d_k) v_k \qquad E(d_k) = \begin{cases} 0 \;\; for\, d_k < x\, Gy \\ 1 \;\; for\, d_k \geq x\, Gy \end{cases}$$

Find the dose representation that best correlates with toxicity

Potential problems with the standard dose-reduction approach





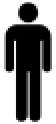
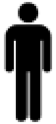
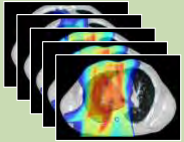
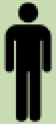
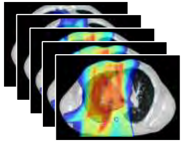
Reduce dose distribution to DVH

- Removes all spatial information
- Assumes equal sensitivity/response of all parts of OAR

Alternatives:

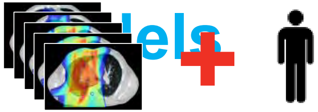
- Divide into anatomical substructures
- Dose surface histograms
- Consider (and/or explicitly model) local response on voxel-to-voxel basis

Adding spatial information to (N)TCP models – general strategies

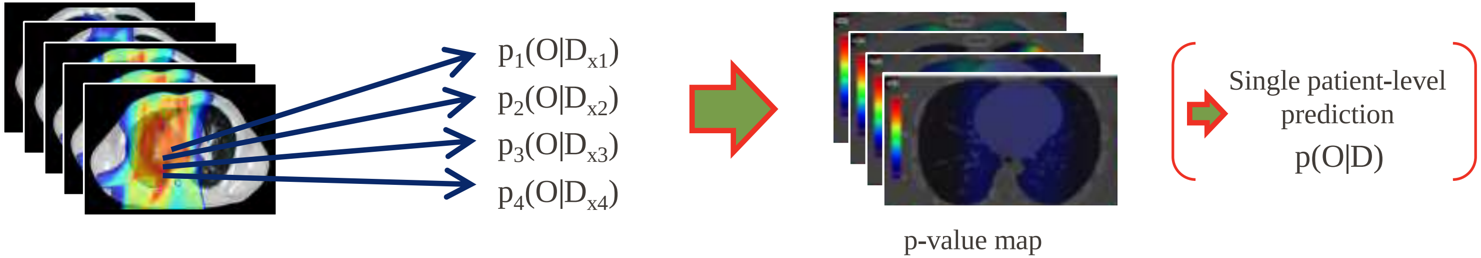
	Dose variable(s)		Response measure	
Traditional NTCP modelling		One per patient		One per patient
Voxel-based analysis (VBA), convolutional neural networks (CNN)		Many per patient (2D or 3D data)		One per patient
Image-based response models		Many per patient (2D or 3D data)		Many per patient (2D or 3D data)

Palma et al. Cancers 2021;13(14):3553. Palma et al. Phys Med 2020;69:192-204. Appelt et al. Clin Oncol 2022;34(2):e87-e96

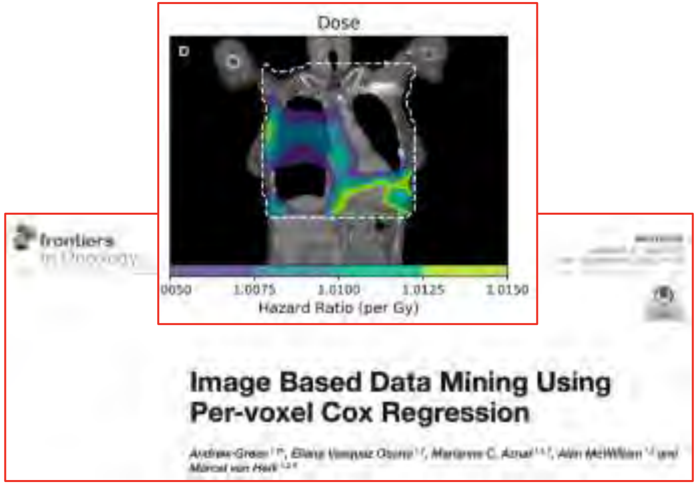
Adding spatial information to (N)TCP models



VBA



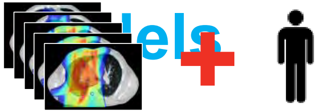
New anatomical insights from voxel-based analysis of dose?

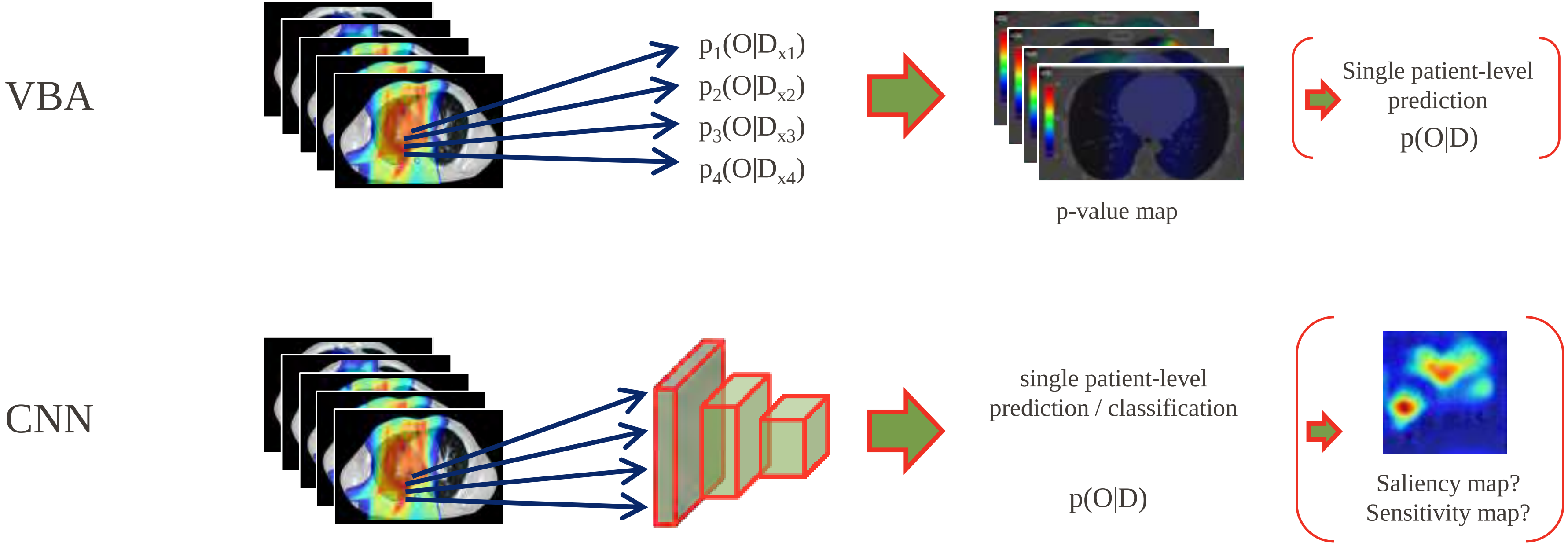


Generally for VBA based studies:

- How dependent are the results by structures in the dose data (e.g. dose gradients and robustness of planned relative to delivered dose)?
- Issues with statistical analysis in some parts of the published literature
  - Shortall et al. Flogging a Dead Salmon? IJROBP 2021

Adding spatial information to (N)TCP





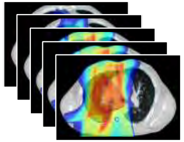
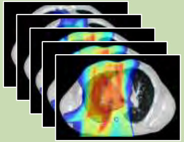
Palma et al. Cancers 2021;13(14):3553. Palma et al. Phys Med 2020;69:192-204. Appelt et al. Clin Oncol 2022;34(2):e87-e96

Improved toxicity prediction with voxel-based analysis?

Patient number	Cancer site	Ref	Improvement over GLM	External validation
42	Cervical	Zhen 2017	✚	✖
125	Liver	Ibragimov 2018	✚	✖
784	Head and neck	Men 2019	✚	✖
120	Liver	Ibragimov 2019	✚	✖
122	Liver	Ibragimov 2020	-	✖
160	Oropharyngeal	Welch 2020	✖	✖
70	NSCLC	Liang 2019	✚	✖
66	Oropharyngeal	Wang 2020	-	✖
52	Post-prostatectomy	Yang 2021	-	✖
217	Thoracic	Liang 2021	✚	✖

Appelt et al. Deep Learning for Radiotherapy Outcome Prediction Using Dose Data - A Review. Clin Oncol 2022

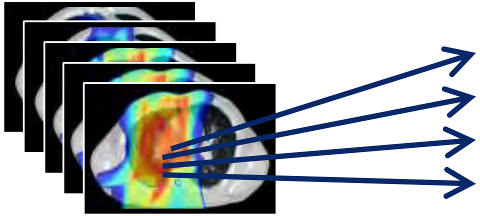
Adding spatial information to (N)TCP models – general strategies

	Dose variable(s)		Response measure	
Traditional NTCP modelling		One per patient		One per patient
Voxel-based analysis (VBA), convolutional neural networks (CNN)		Many per patient (2D or 3D data)		One per patient
Image-based response models		Many per patient (2D or 3D data)		Many per patient (2D or 3D data)

Palma et al. Cancers 2021;13(14):3553. Palma et al. Phys Med 2020;69:192-204. Appelt et al. Clin Oncol 2022;34(2):e87-e96

Adding spatial information to (N)TCP models

Image-based response models



$$\begin{pmatrix} p(R_{x1}|D_{x1}) \\ p(R_{x2}|D_{x2}) \\ p(R_{x3}|D_{x3}) \\ p(R_{x4}|D_{x4}) \end{pmatrix}$$

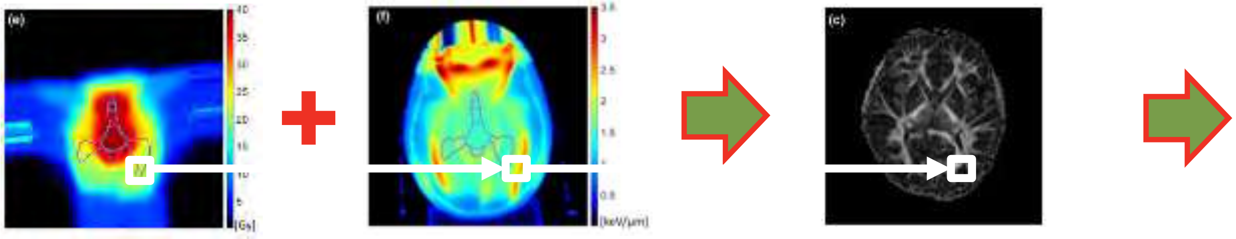
Multilevel mixed effect model

single model linking dose & local response

p(O|D)

Better or novel biological insights from voxel-based analysis of dose?

Dose LET Imaging changes



Model of RBE dependence on dose and LET

A systematic review of clinical studies on proton Relative Biological Effectiveness (RBE)

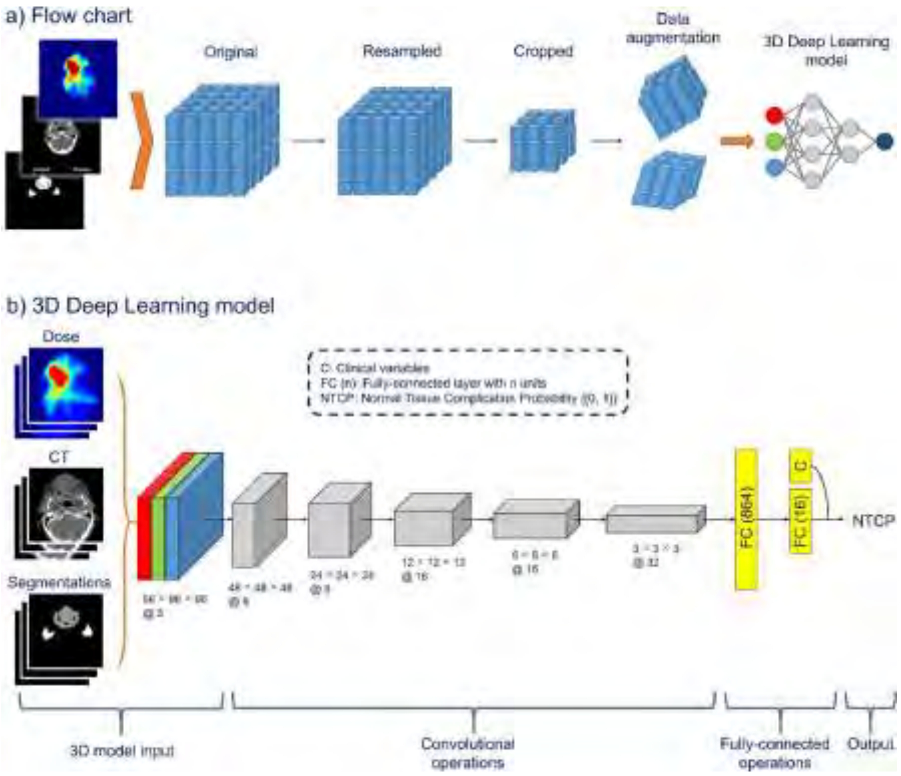
- 13 studies using voxel-wise analyses of patient effects versus dose and LET
  - **3/13: No effect of LET on RBE**
  - **6/16: Maybe effect of LET on RBE**
  - **4/13: Effect of LET on RBE**
- Significant methodological modelling issues
  - E.g. no consideration of nested / multi-level data

Underwood et al. A systematic review of clinical studies on variable proton Relative Biological Effectiveness (RBE). Radiother Oncol. 2022

3D deep learning Normal Tissue Complication Probability model to predic late xerostomia in head and neck cancer patients

Hung Chu MSc [1], Suzanne P.M. de Vette MSc [1], Hendrike Neh MSc [1], Nanna M. Sijtsema PhD [1], Roel J.H.M. Steenbakkers MD, PhD [1], Amy Moreno MD [2], Johannes A. Langendijk MD, PhD [1], Peter M.A. van Ooijen PhD [1], Clifton D. Fuller MD, PhD [2], Lisanne V. van Dijk PhD [1]

https://doi.org/10.1016/j.ijrobp.2024.07.2334



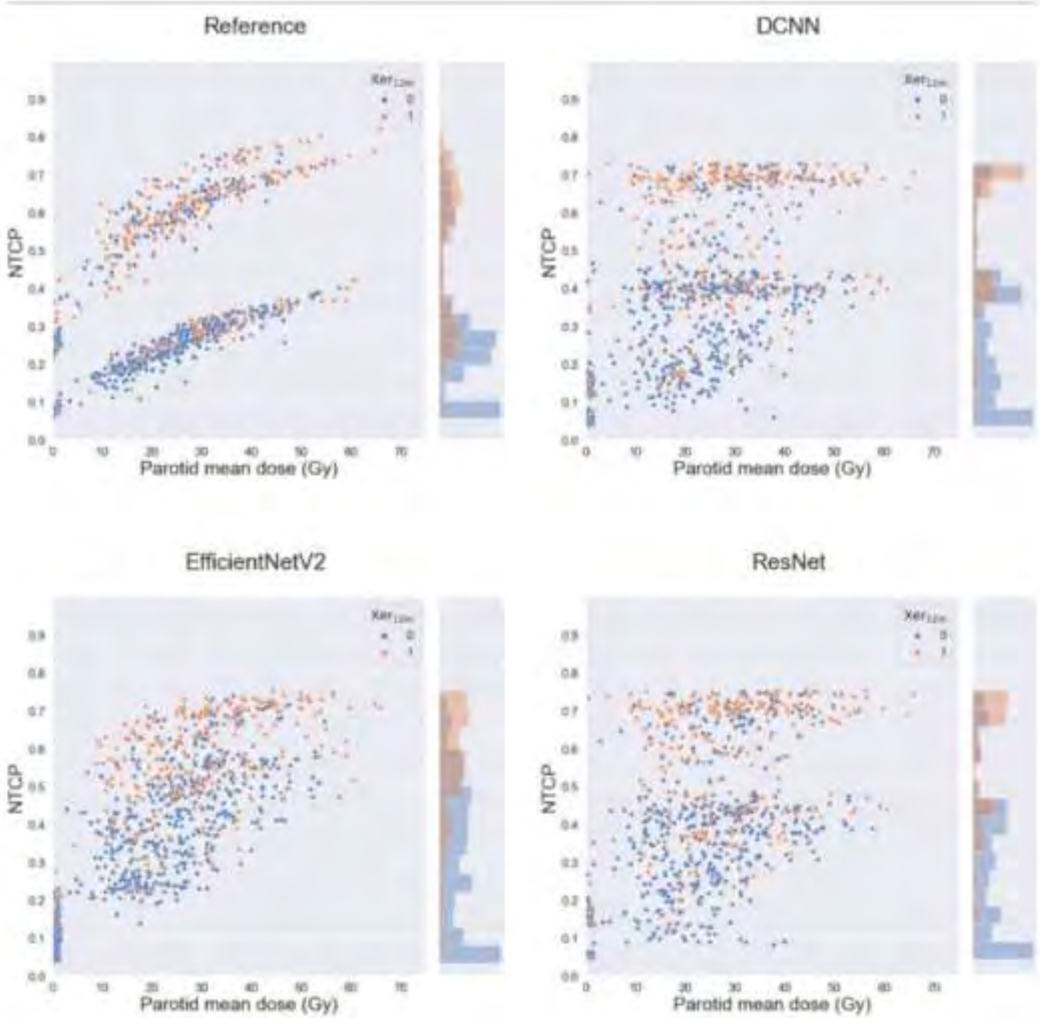
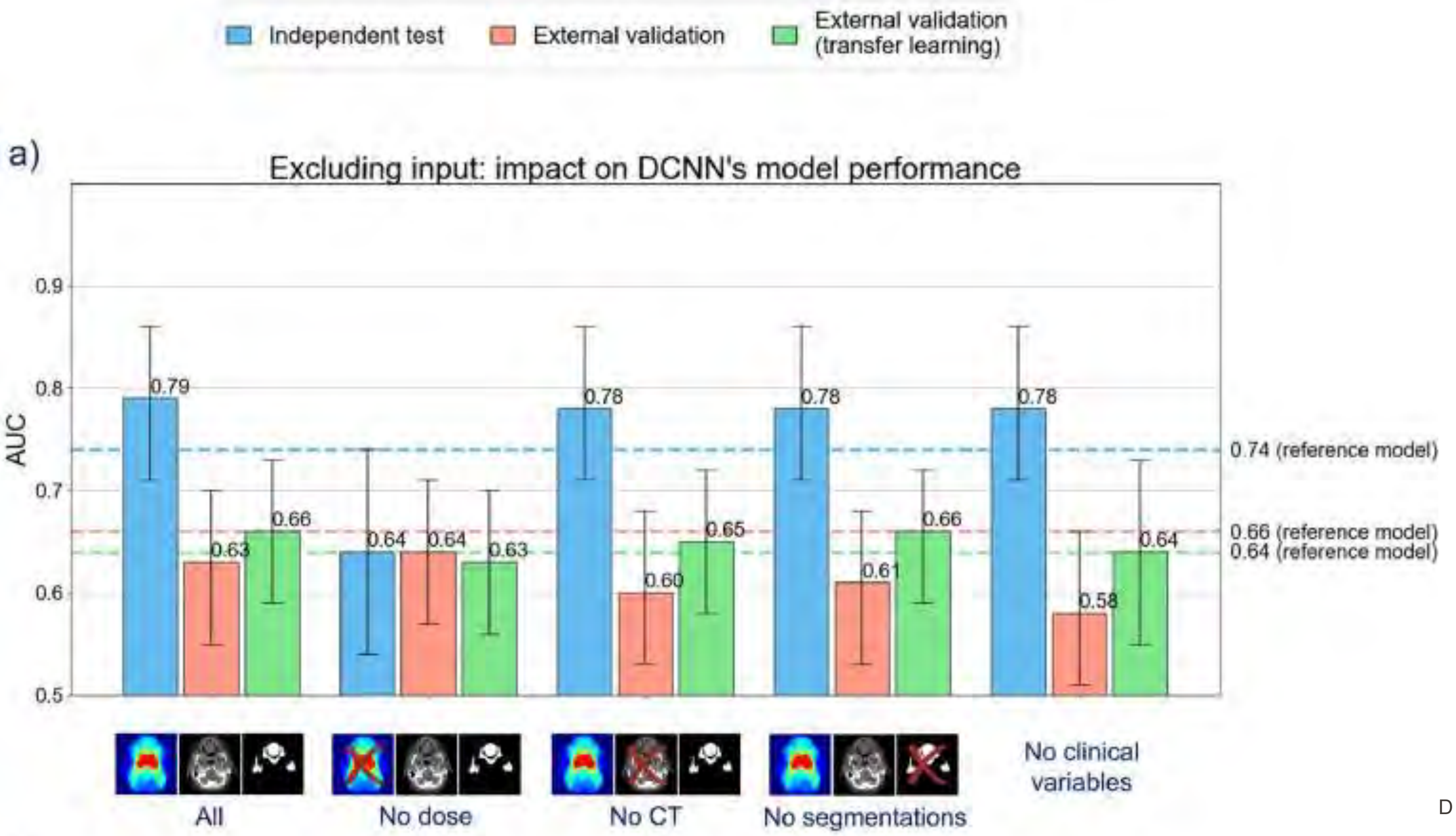


Figure 2. These scatterplots display the relationship between parotid mean dose (in Gy) and NTCP (Normal Tissue Complication Probability) value for all models. Patients who experienced moderate-to-severe xerostomia 12 months post-radiotherapy are represented by orange, while the remaining patients are represented by blue. The accompanying histogram illustrates the distribution of the NTCP values.

	Training	Independent test	External validation
Total	759	138	311



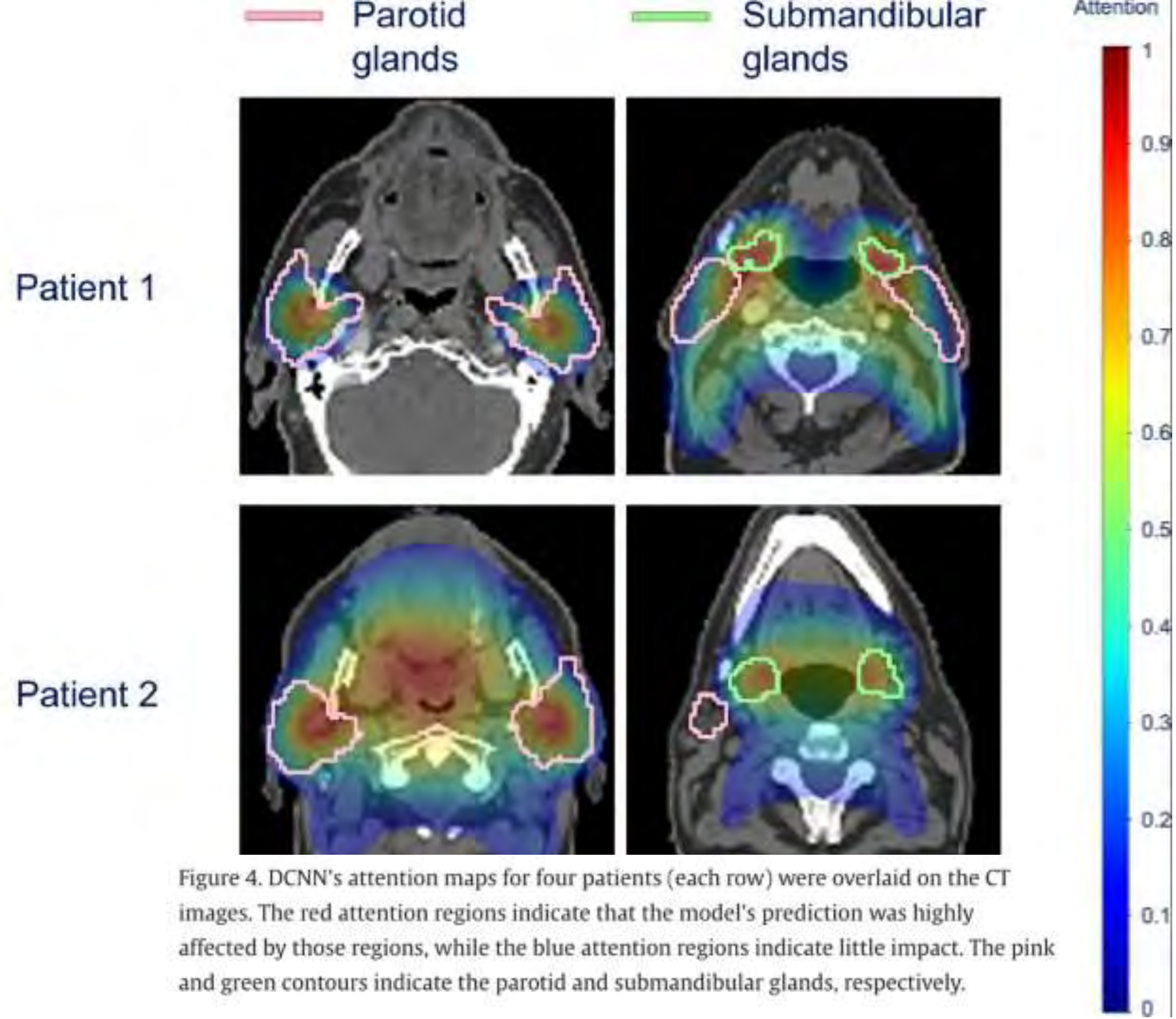


Figure 4. DCNN's attention maps for four patients (each row) were overlaid on the CT images. The red attention regions indicate that the model's prediction was highly affected by those regions, while the blue attention regions indicate little impact. The pink and green contours indicate the parotid and submandibular glands, respectively.

Original article

Beyond mean pharyngeal constrictor dose for beam path toxicity in non-target swallowing muscles: Dose–volume correlates of chronic radiation-associated dysphagia (RAD) after oropharyngeal intensity modulated radiotherapy ☆

MD Anderson Head and Neck Cancer Symptom Working Group (Contributing authors Timothy Dale, Katherine Hutcheson, Abdallah S.R. Mohamed, Jan S. Lewin, G. Brandon Gunn, Arvind U.K. Rao, Jayashree Kalpathy-Cramer, Steven J. Frank, Adam S. Garden, Jay A. Messer, Benjamin Warren, Stephen Y. Lai, Beth M. Beadle, William H. Morrison, Jack Phan, Heath Skinner, Neil Gross, Renata Ferrarotto, Randal S. Weber, David I. Rosenthal, Clifton D. Fuller*)

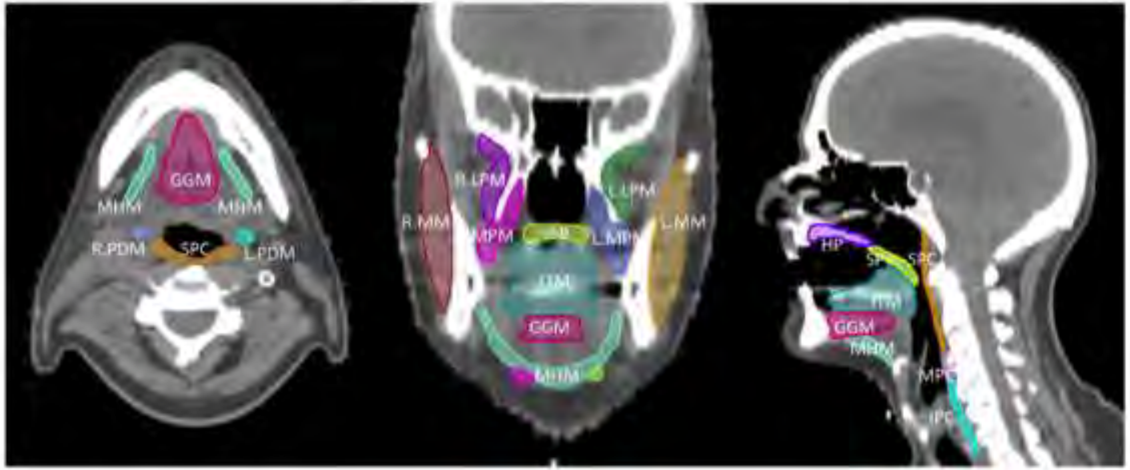
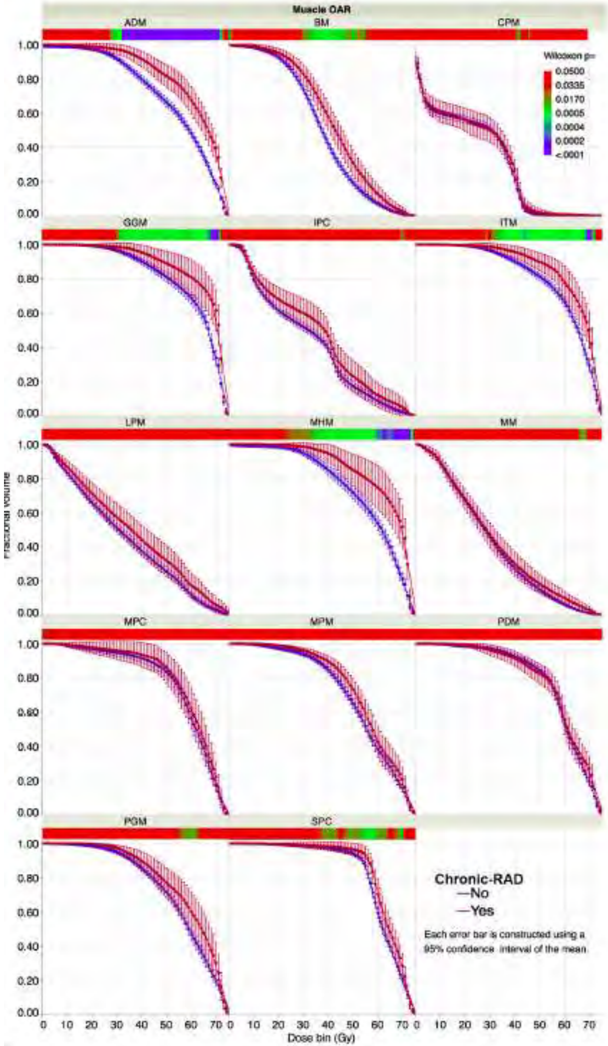
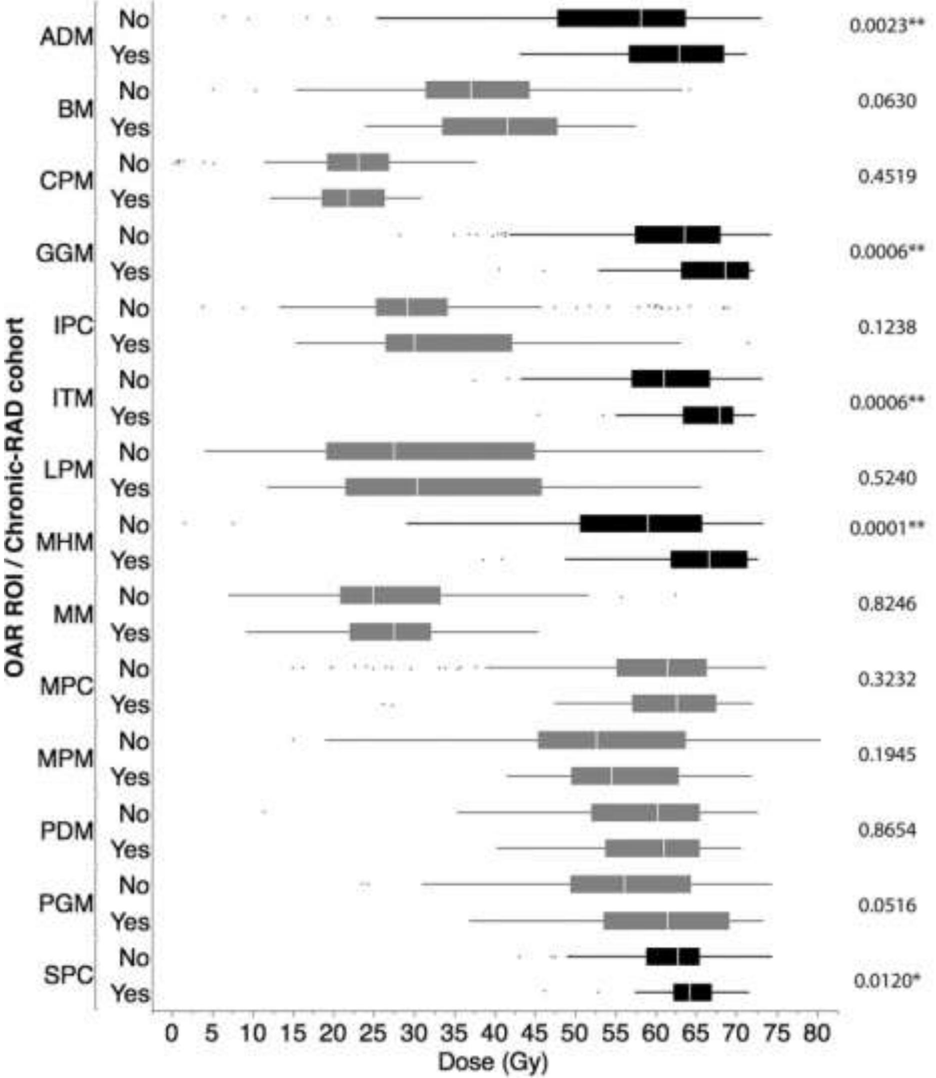
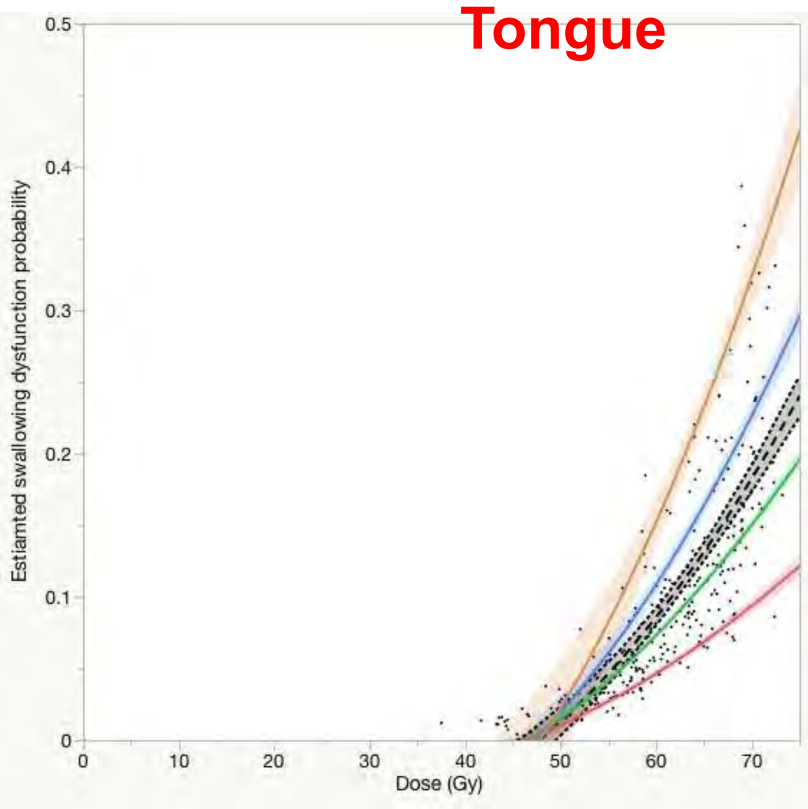


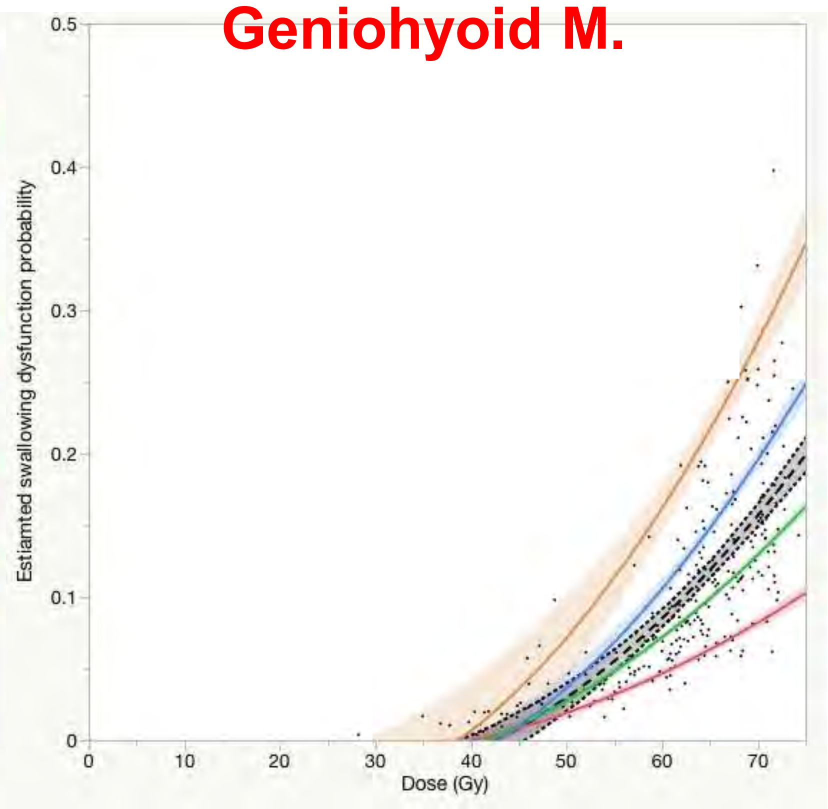
Fig. 1. Exemplar swallow-related ROI. Axial, coronal, and sagittal images of the contoured segments. *Abbreviations:* GGM – genioglossus muscle; HP – hard palate; IPC – inferior pharyngeal constrictor; ITM – intrinsic tongue muscles; LPM – lateral pterygoid muscle; MHM – mylo/geniohyoid complex; MM – masseter muscle; MPM – medial pterygoid muscle; PDM – posterior digastric muscle; SP – soft palate; SPC – superior pharyngeal constrictor; R.-right, L.-left.


Muscle OAR
ADM
BM
CPM
Wilcoxon p=
0.0500
0.0335
0.0170
0.0005
0.0004
0.0002
<.0001
GGM
IPC
ITM
LPM
MHM
MM
MPC
MPM
PDM
PGM
SPC
Fractional volume
Dose bin (Gy)
Chronic-RAD
No
Yes
Each error bar is constructed using a 95% confidence interval of the mean.

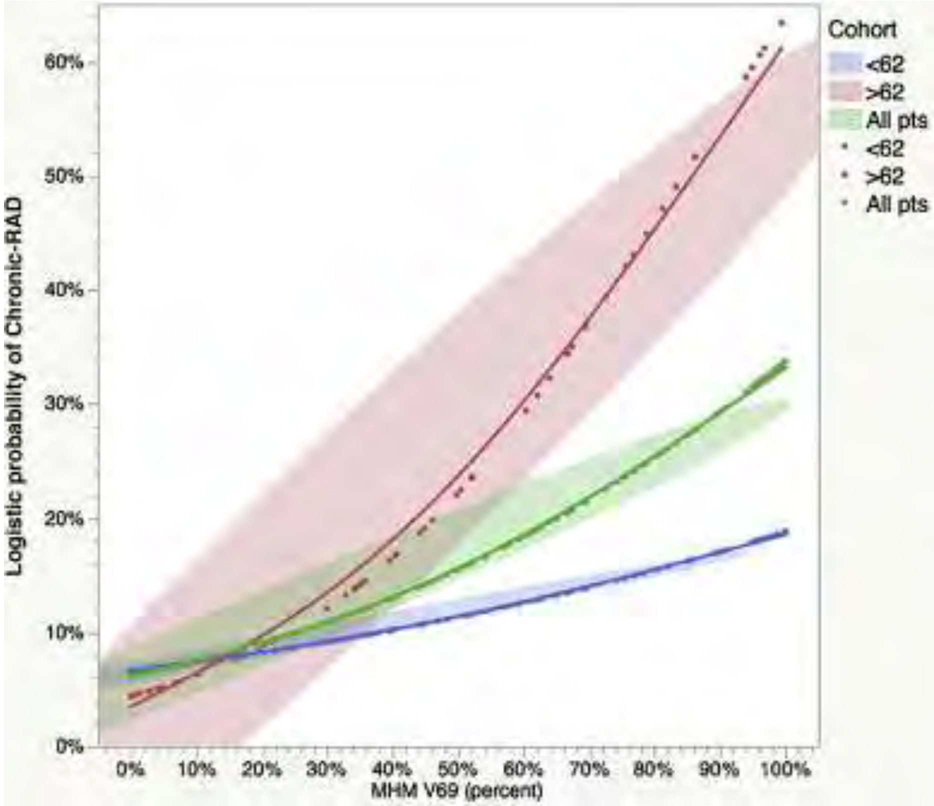

OAR ROI / Chronic-RAD cohort
ADM
BM
CPM
GGM
IPC
ITM
LPM
MHM
MM
MPC
MPM
PDM
PGM
SPC
No
Yes
0.0023**
0.0630
0.4519
0.0006**
0.1238
0.0006**
0.5240
0.0001**
0.8246
0.3232
0.1945
0.8654
0.0516
0.0120*
Dose (Gy)

Example: Age and dysphagia





Optimum OPC model includes mylohyoid/geniohyoid dose & age



Adding spatial data…

Magnetic resonance imaging of swallowing-related structures in nasopharyngeal carcinoma patients receiving IMRT: Longitudinal dose–response characterization of quantitative signal kinetics

Jay A. Messer, Abdallah S.R. Mohamed, Katherine A. Hutcheson, Yao Ding, Jan S. Lewin, Jihong Wang, Stephen Y. Lai, Steven J. Frank, Adam S. Garden, Vlad Sandulache, Hillary Eichelberger, Chloe C. French, Rivka R. Colen, Jack Phan, Jayashree Kalpathy-Cramer, John D. Hazle, David I. Rosenthal, G. Brandon Gunn, Clifton D. Fuller

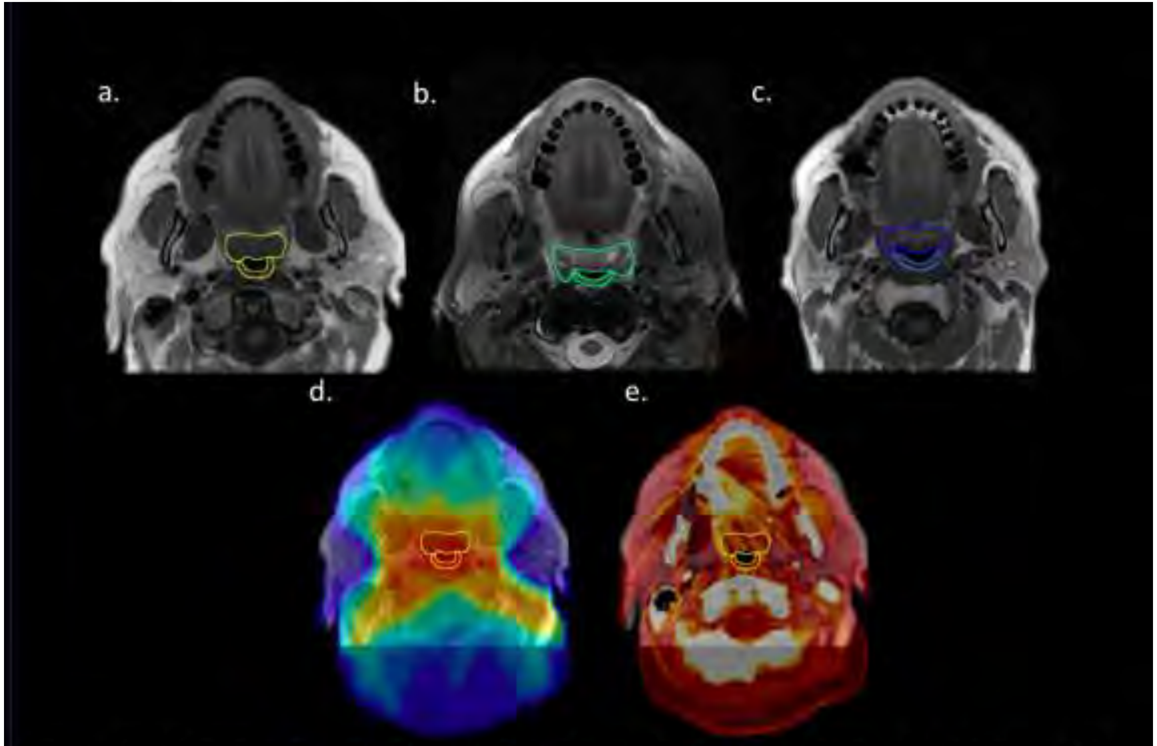


Figure 1. 1a) T1 Baseline. 1b) T2 Early Post-RT after 3 months 1c) T1 Late Post-RT after 29 months 1d) Radiation dose grid 1e) Co-registration of MRI and planning CT

T1W Muscle damage/dose biomarker

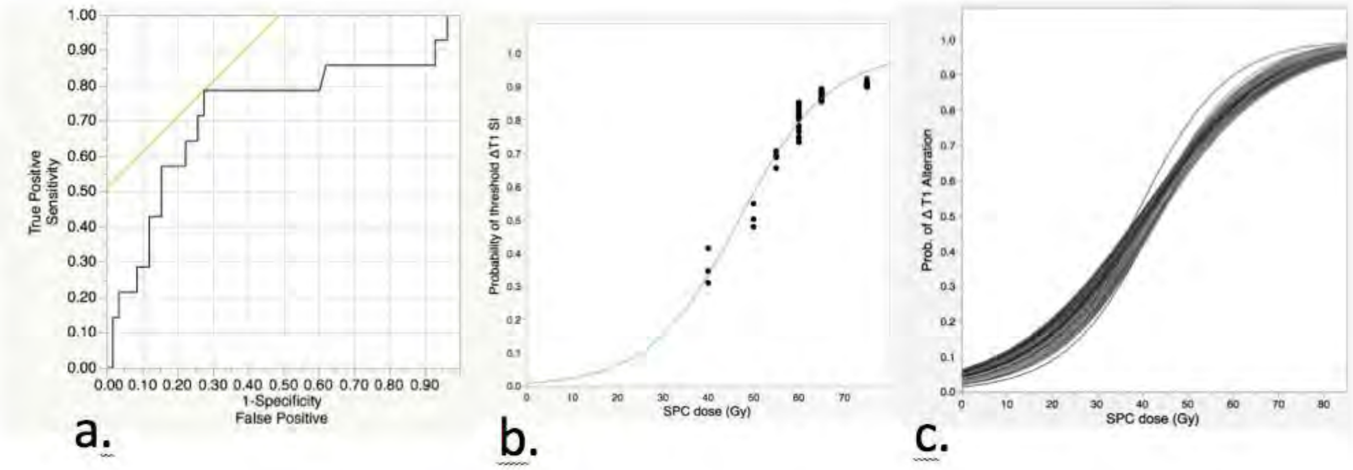
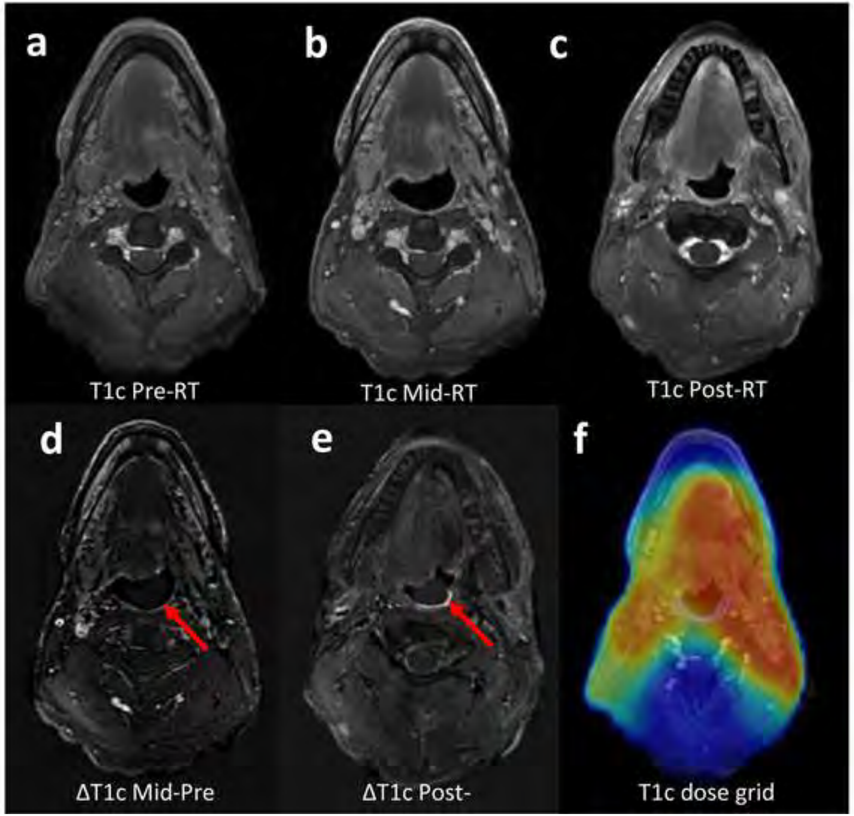
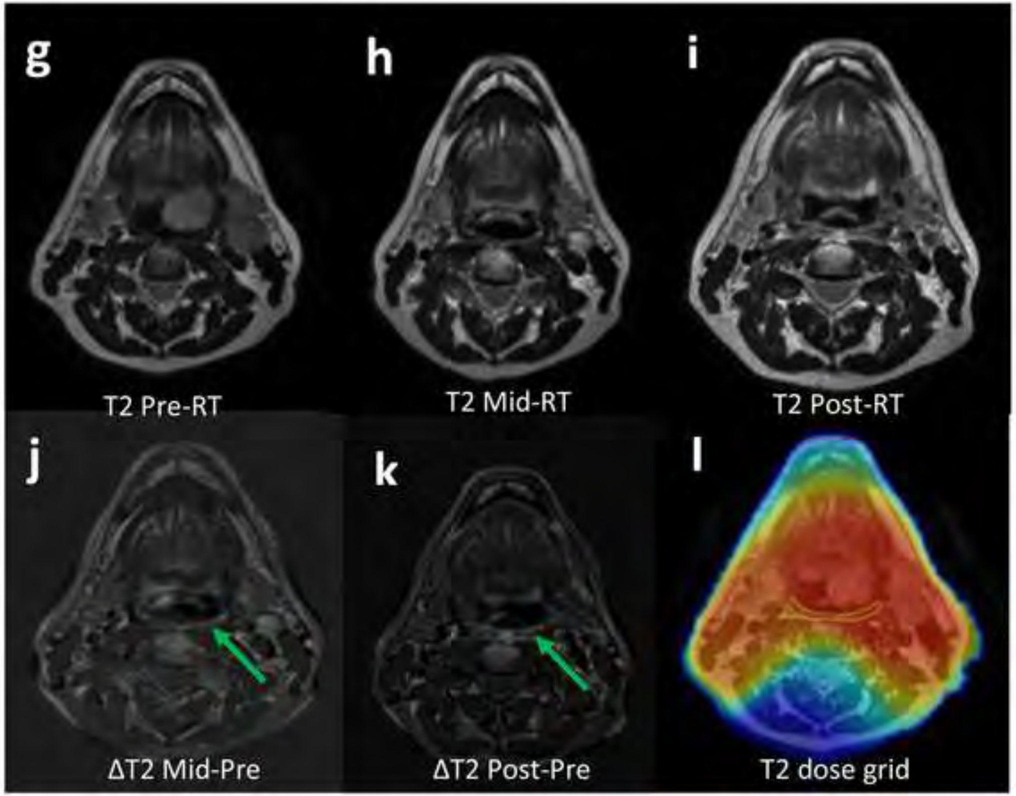


Figure 5: Continuous (non-linear) dose response characterization of late T1 superior pharyngeal constrictor signal alteration from baseline. 5a. Confirmatory analysis of RPA-derived dose-threshold; Receiver operator characteristic curve (ROC), showing split performance for T1 signal intensity changes of greater than or less than 0.57 in the superior pharyngeal constrictors, as a function of D_{mean}, with area-under the curve (AUC) of 0.72 (P=0.013). 5b. Sigmoidal fit of observed probability of threshold T1 signal alteration as function of D_{mean} to superior pharyngeal constrictor muscles (R2=0.93). 5c. Incidence–resampled bootstrap predicted probability of threshold T1 alteration as a function of dose; 10^4 - independently-resampled distributions were individually fit using a maximum likelihood 2P-sigmoidal function, representing the range of possible dose-response normal tissue complication probability curves in order to best approximate a "true population incidence."

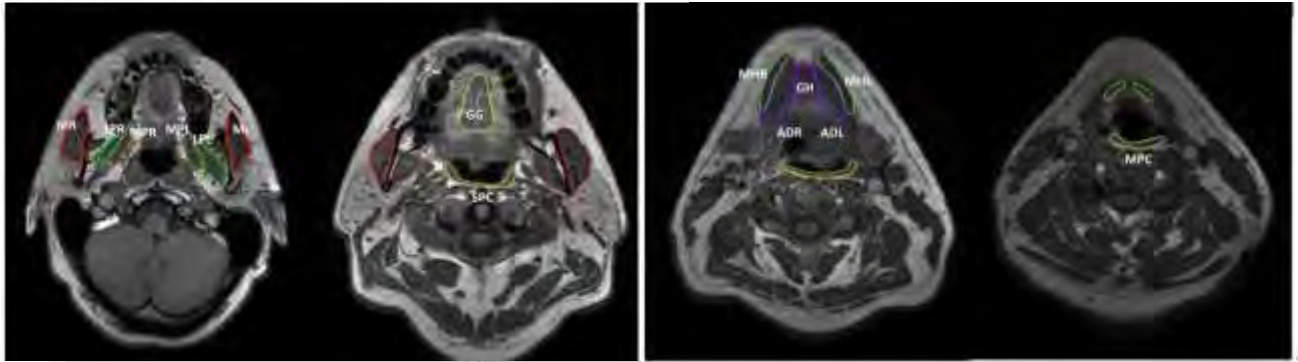
What if we just used standardized T1W/T2W MRI?

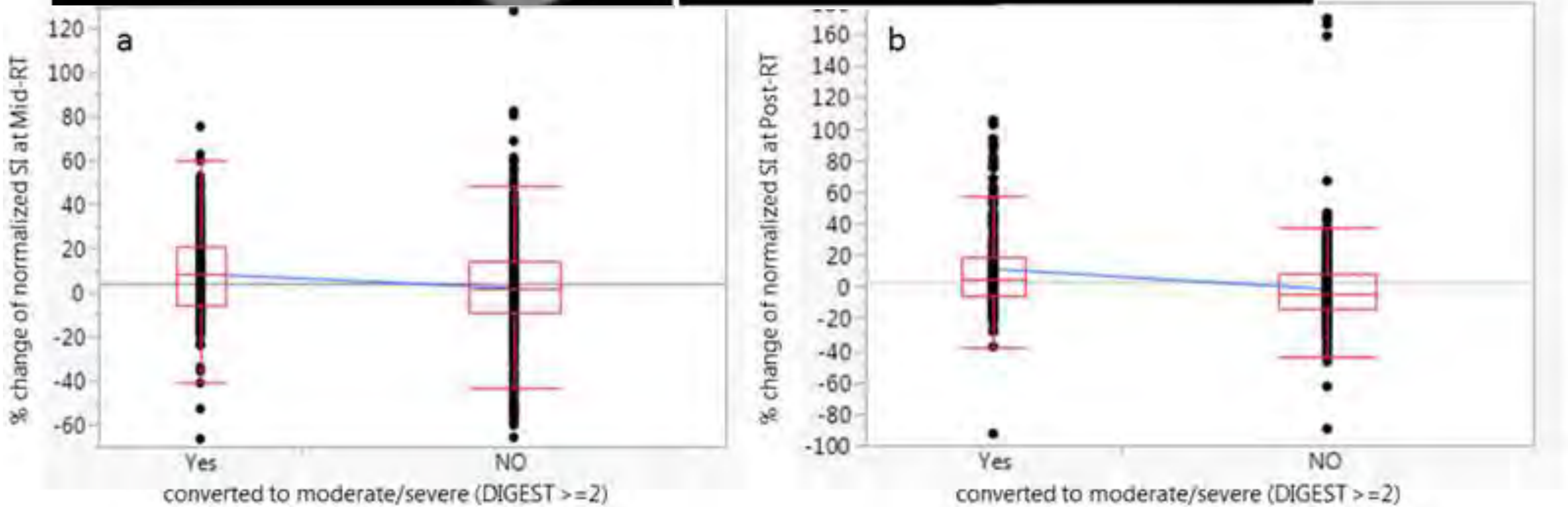


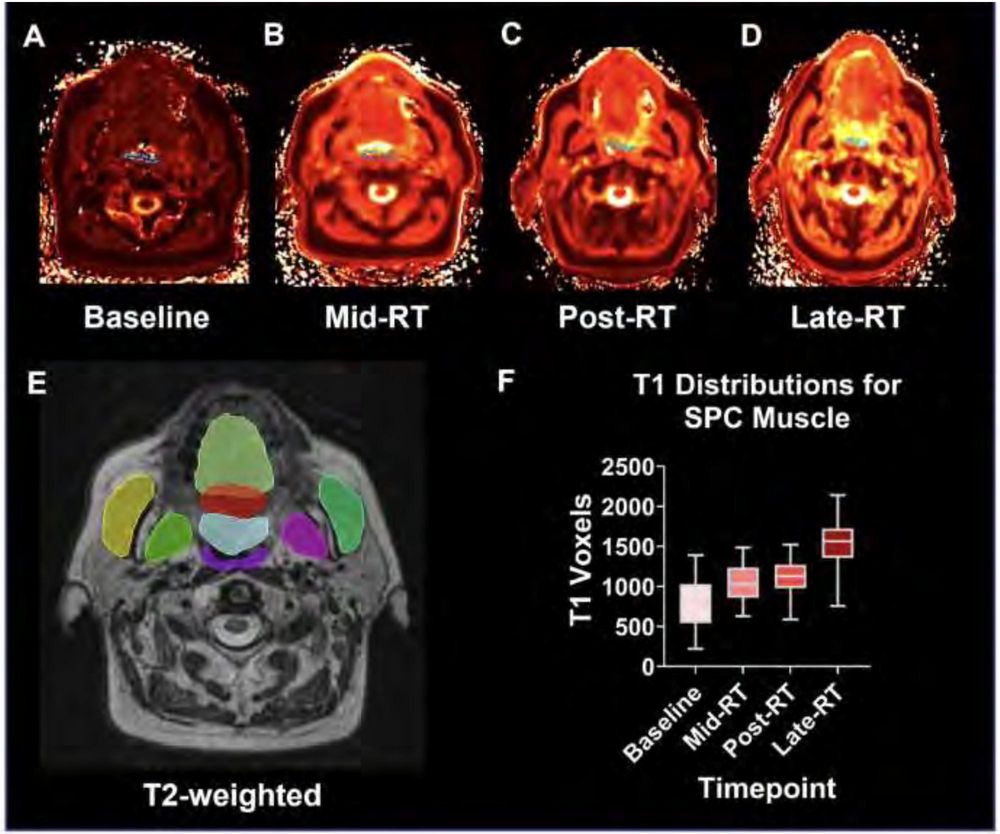


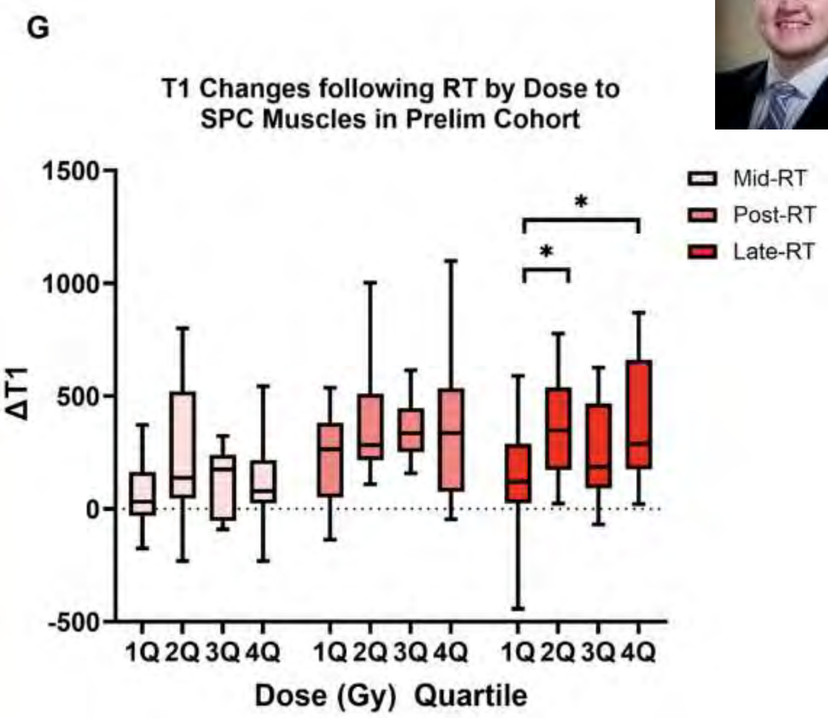
A prospective longitudinal assessment of MRI signal intensity kinetics of non-target muscles in patients with advanced stage oropharyngeal cancer in relationship to radiotherapy dose and post-treatment radiation-associated dysphagia: Preliminary findings from a randomized trial

Radiotherapy and Oncology 130 (2019) 46–55

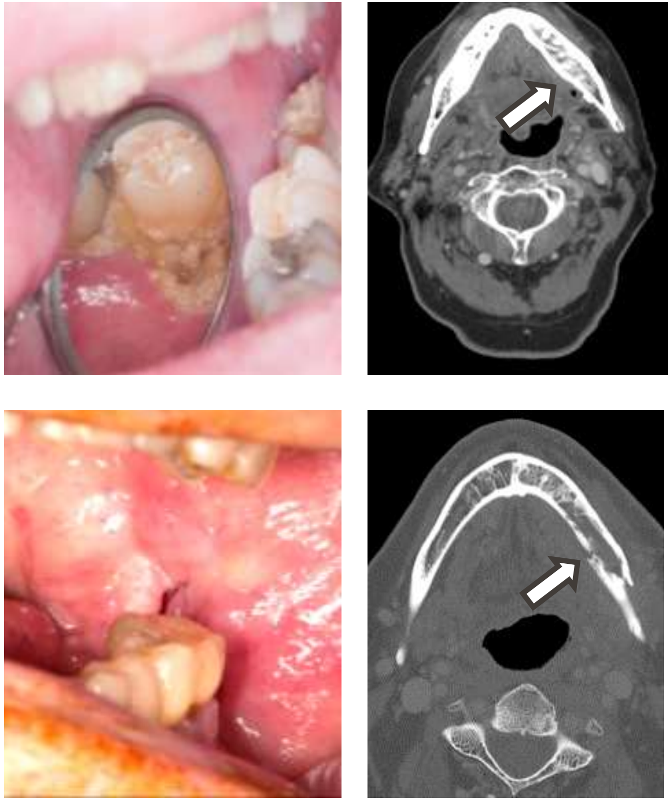





A
B
C
D
Baseline
Mid-RT
Post-RT
Late-RT
E
T2-weighted
F
T1 Distributions for SPC Muscle
T1 Voxels
2500
2000
1500
1000
500
0
Baseline
Mid-RT
Post-RT
Late-RT
Timepoint


G
T1 Changes following RT by Dose to SPC Muscles in Prelim Cohort
ΔT1
1500
1000
500
0
-500
Mid-RT
Post-RT
Late-RT
*
*
1Q 2Q 3Q 4Q
1Q 2Q 3Q 4Q
1Q 2Q 3Q 4Q
Dose (Gy) Quartile

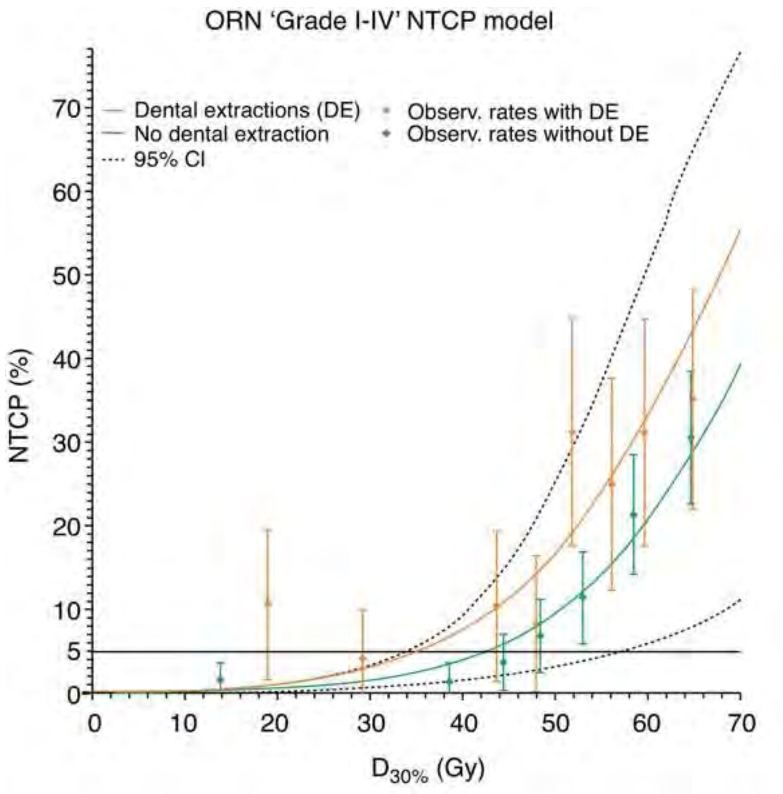
Osteoradionecrosis (ORN)



“Exposed bone in a field of irradiation.”

MDACC rate ~6-7%, which means about 65 cases/year

Normal Tissue Complication Probability (NTCP) For ORN



Research at MD Anderson

adapted from van Dijk et al IJ*ROBP* 2021

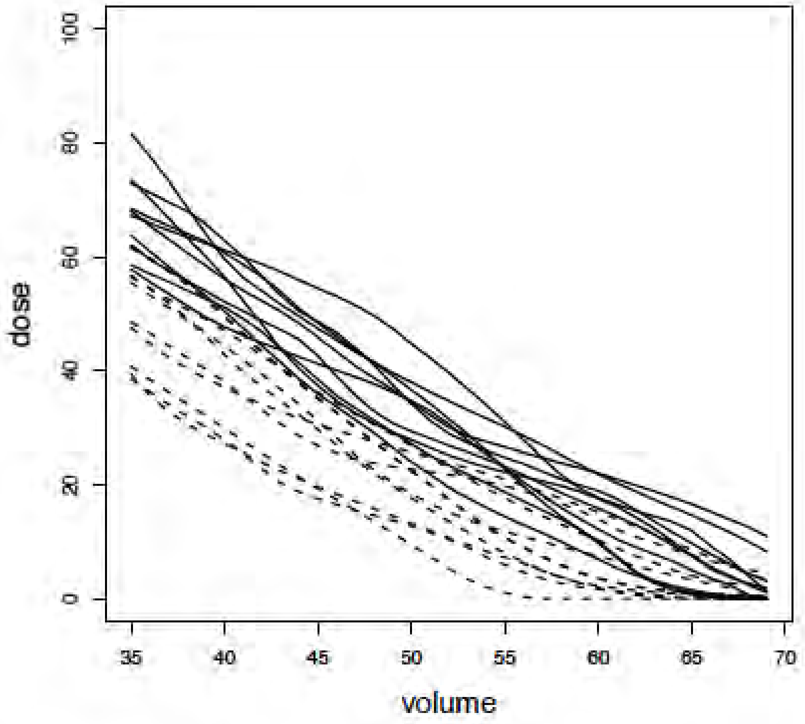
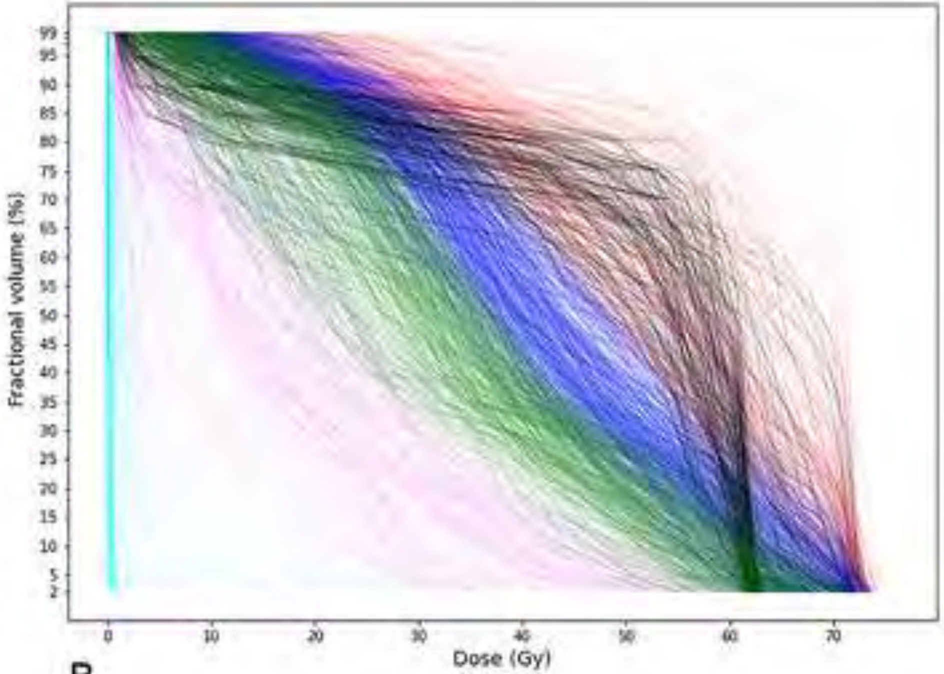


Figure 1: The dose-volumes curves for patients with ORN (solid) versus controls (dotted).



B

	PDE = 0		PDE = 1	
Cluster no.	ORN incidence	Risk index (95% CI)	ORN incidence	Risk index (95% CI)
1	0 out of 58	0.0%	0 out of 9	0.0%
2	2 out of 68	2.9% (0.0%, 6.9%)	0 out of 8	0.0%
3	8 out of 273	2.9% (0.9%, 4.9%)	10 out of 86	11.6% (4.8%, 18.4%)
4	39 out of 318	12.3% (8.7%, 15.9%)	34 out of 144	23.6% (16.7%, 30.5%)
5	31 out of 118	26.3% (18.3%, 34.3%)	14 out of 47	29.8% (16.7%, 42.9%)
6	21 out of 82	25.6% (16.1%, 35.1%)	14 out of 48	29.2% (16.3%, 42.1%)

Abbreviations: ORN = osteoradionecrosis; PDE = preradiation dental extraction (0 = no/edentulous, 1 = dental extractions).

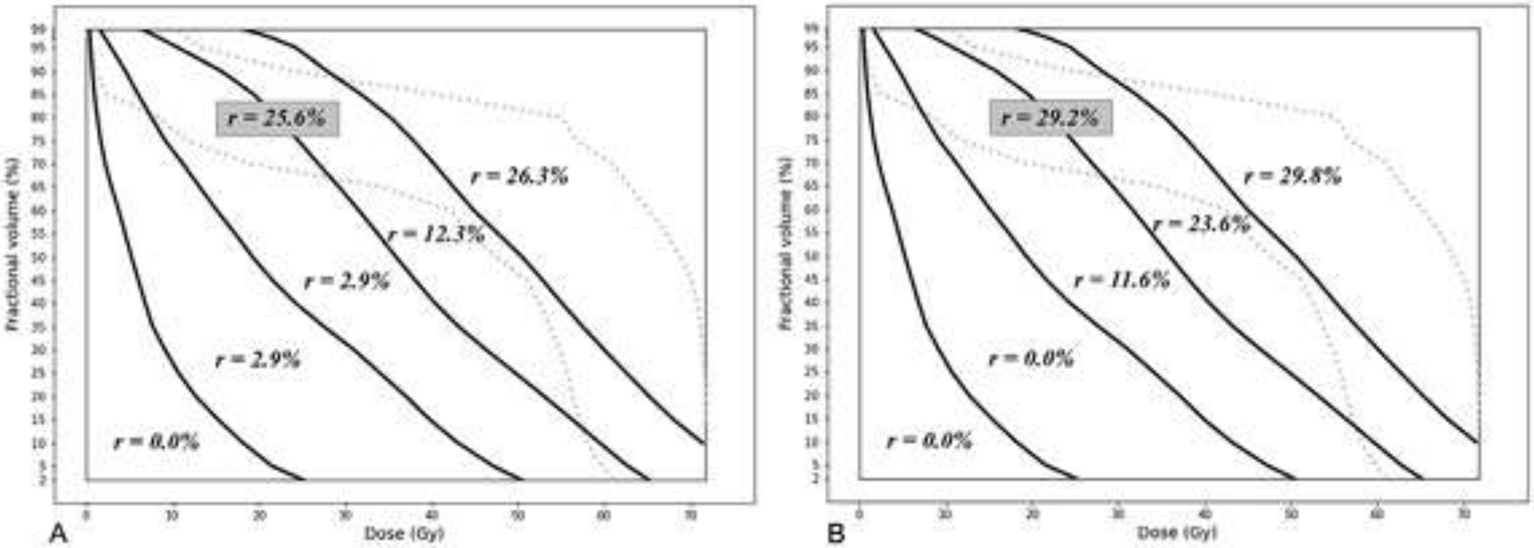


Fig. 4. Risk indices of dose-volume regions for K = 6. (A) No/edentulous dental extractions (PDE = 0). (B) With dental extractions (PDE = 1).

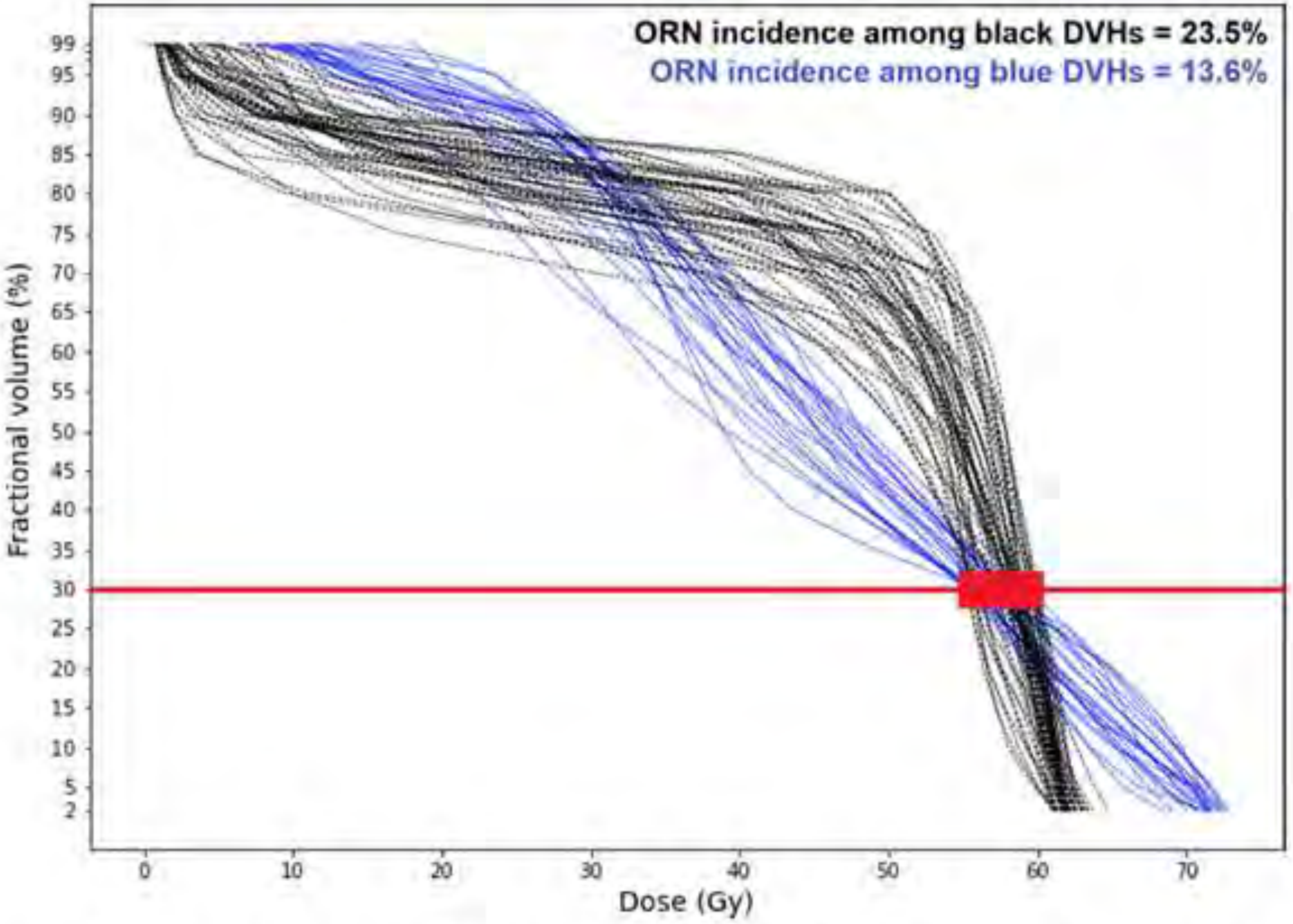
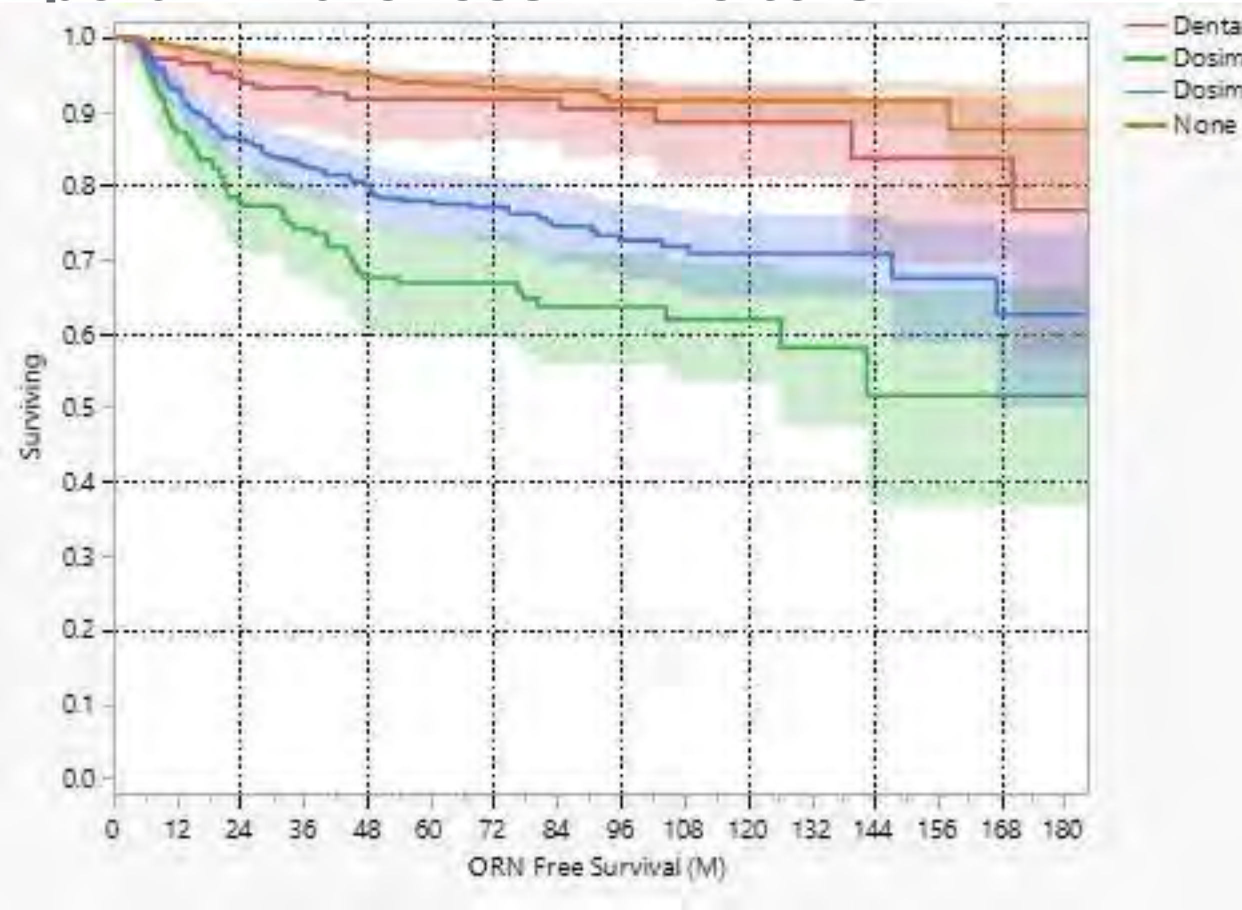
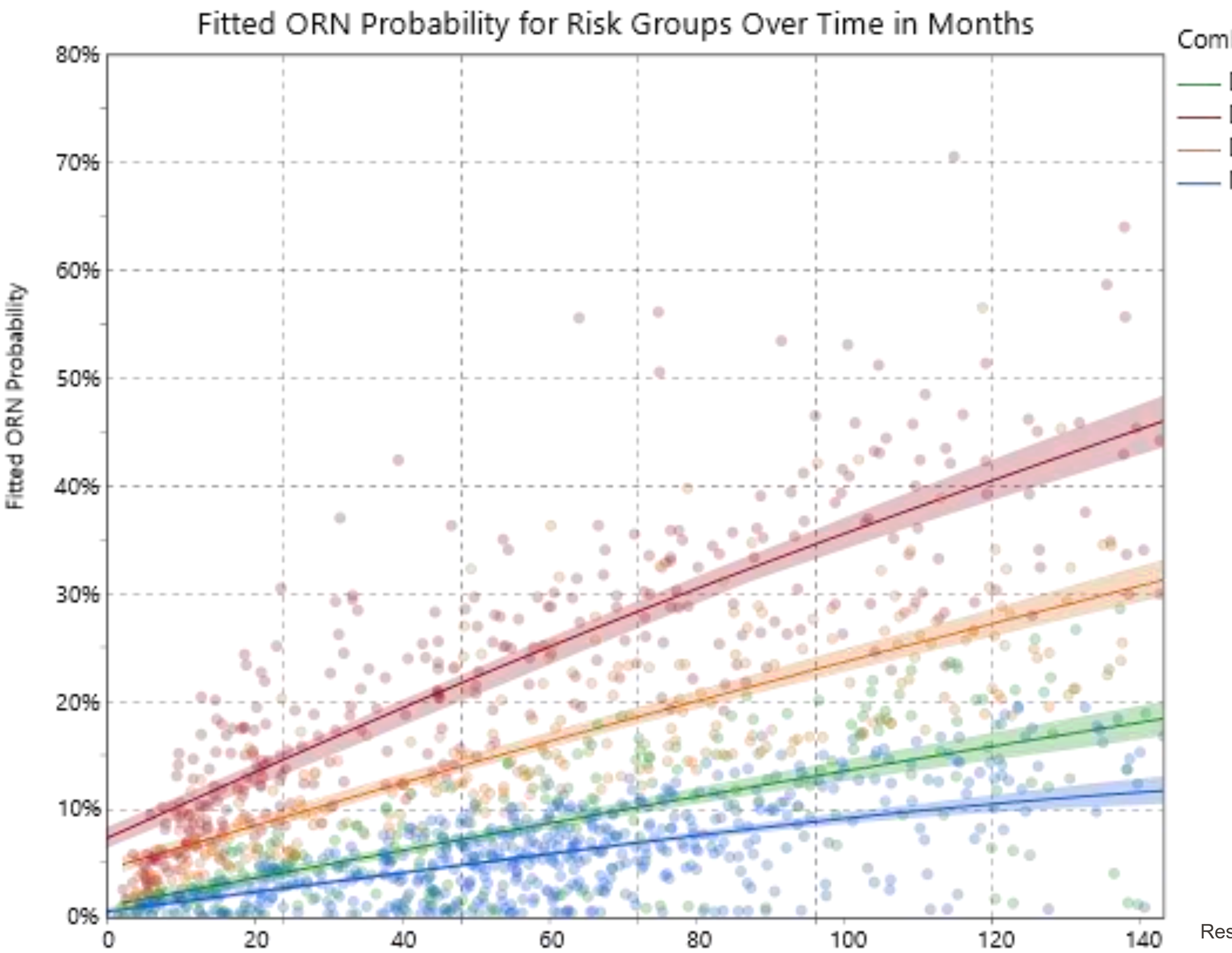


Fig. 5. Different osteoradionecrosis incidences among dose-volume histograms with the same D30% value.

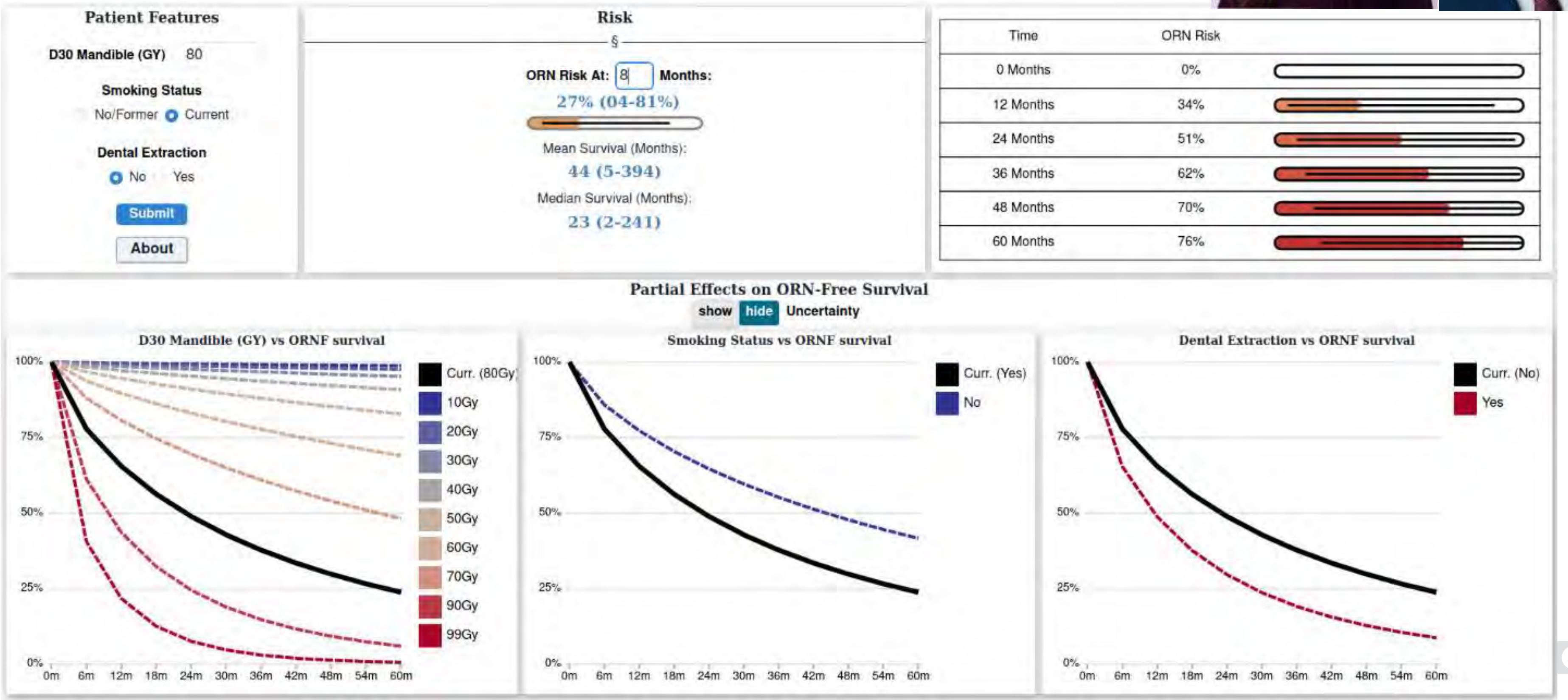
Temporal Awareness: Time to ORN





Research at MD Anderson

ORN Risk GUI


Patient Features
D30 Mandible (GY) 80
Smoking Status
No/Former Current
Dental Extraction
No Yes
Submit
About
Risk
§
ORN Risk At: 8 Months:
27% (04-81%)
Mean Survival (Months):
44 (5-394)
Median Survival (Months):
23 (2-241)
Time ORN Risk
0 Months 0%
12 Months 34%
24 Months 51%
36 Months 62%
48 Months 70%
60 Months 76%
Partial Effects on ORN-Free Survival
show hide Uncertainty
D30 Mandible (GY) vs ORNF survival
Curr. (80Gy)
10Gy
20Gy
30Gy
40Gy
50Gy
60Gy
70Gy
90Gy
99Gy
Smoking Status vs ORNF survival
Curr. (Yes)
No
Dental Extraction vs ORNF survival
Curr. (No)
Yes
100%
75%
50%
25%
0%
0m 6m 12m 18m 24m 30m 36m 42m 48m 54m 60m

Comparison of Machine-Learning and Deep-Learning Methods for the Prediction of Osteoradionecrosis Resulting From Head and Neck Cancer Radiation Therapy

Brandon Reber, BS,[a,*] Lisanne Van Dijk, PhD,[a,b] Brian Anderson, PhD,[a,c] Abdallah Sherif Radwan Mohamed, MD, PhD,[a] Clifton Fuller, MD, PhD,[a] Stephen Lai, MD, PhD,[a] and Kristy Brock, PhD[a]

https://doi.org/10.1016/j.adro.2022.101163

Table 1 Summary of subject demographics*

	ORN−	ORN+
Number of subjects	1086	173
Age, y, median	61	60
Sex, male, n (%)	894 (82%)	150 (87%)
Smoking, current, n (%)	153 (14%)	27 (16%)
Smoking, pack-years, median	7	8
Postoperative RT	172 (16%)	44 (25%)
Dental extraction pre-RT	270 (25%)	72 (42%)
Tumor site		
Oral cavity	146 (13%)	44 (25%)
Oropharynx	703 (65%)	123 (71%)
Hypopharynx/larynx/nasopharynx/unknown-primary	237 (22%)	6 (3%)

Abbreviations: ORN = osteoradionecrosis; RT = radiation therapy.

* Percent signs within cells indicate the percent of the subject cohort for the ORN− and ORN + cases separately that have each row attribute.

Table 2 **Mean (±SD) metric values for the cross-validation withheld folds for the ML models***

Model	Accuracy	Balanced accuracy	Recall	Precision	F1 score	AUROC	AUPRC
Logistic regression	0.69 ± 0.05	0.70 ± 0.07	0.72 ± 0.14	0.27 ± 0.05	0.39 ± 0.07	0.74 ± 0.07	0.28 ± 0.08
Random forest	0.65 ± 0.05	0.69 ± 0.07	0.74 ± 0.14	0.25 ± 0.04	0.37 ± 0.06	0.69 ± 0.07	0.23 ± 0.04
Support vector machine	0.69 ± 0.04	0.70 ± 0.07	0.71 ± 0.13	0.27 ± 0.04	0.39 ± 0.06	0.70 ± 0.07	0.24 ± 0.04
Random classifier	0.52 ± 0.04	0.49 ± 0.08	0.45 ± 0.14	0.14 ± 0.04	0.21 ± 0.07	0.50 ± 0.00	0.14 ± 0.01

Abbreviations: AUPRC = area under the precision recall curve; AUROC = area under the receiver operating characteristic curve; ML = machine learning.

* Each cell shows the mean (±SD) of the metrics from the withheld folds of the stratified 10-fold cross-validation with 10 repeats.

Table 3 **Performance of the best DL models for each architecture type***

Architecture	Accuracy	Balanced accuracy	Recall	Precision	F1 score	AUROC	AUPRC
ResNet	0.87	0.69	0.04	0.50	0.07	0.57	0.23
DenseNet	0.83	0.54	0.10	0.21	0.14	0.58	0.17
Autoencoder	0.71	0.53	0.33	0.18	0.23	0.59	0.15
Random	0.49	0.46	0.46	0.11	0.17	0.49	0.13

Abbreviations: AUPRC = area under the precision recall curve; AUROC = area under the receiver operating characteristic curve; DL = deep learning.

* The reported metrics are from the withheld test set not used during model training or selection. Metrics sensitive to data imbalance, such balanced accuracy, F1 score, and AUPRC, were lower than those for the logistic regression model using the test set.

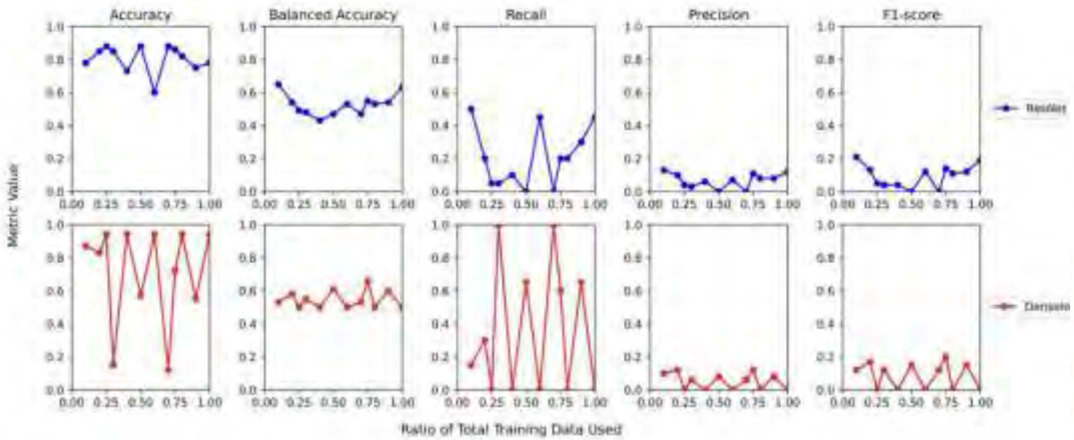
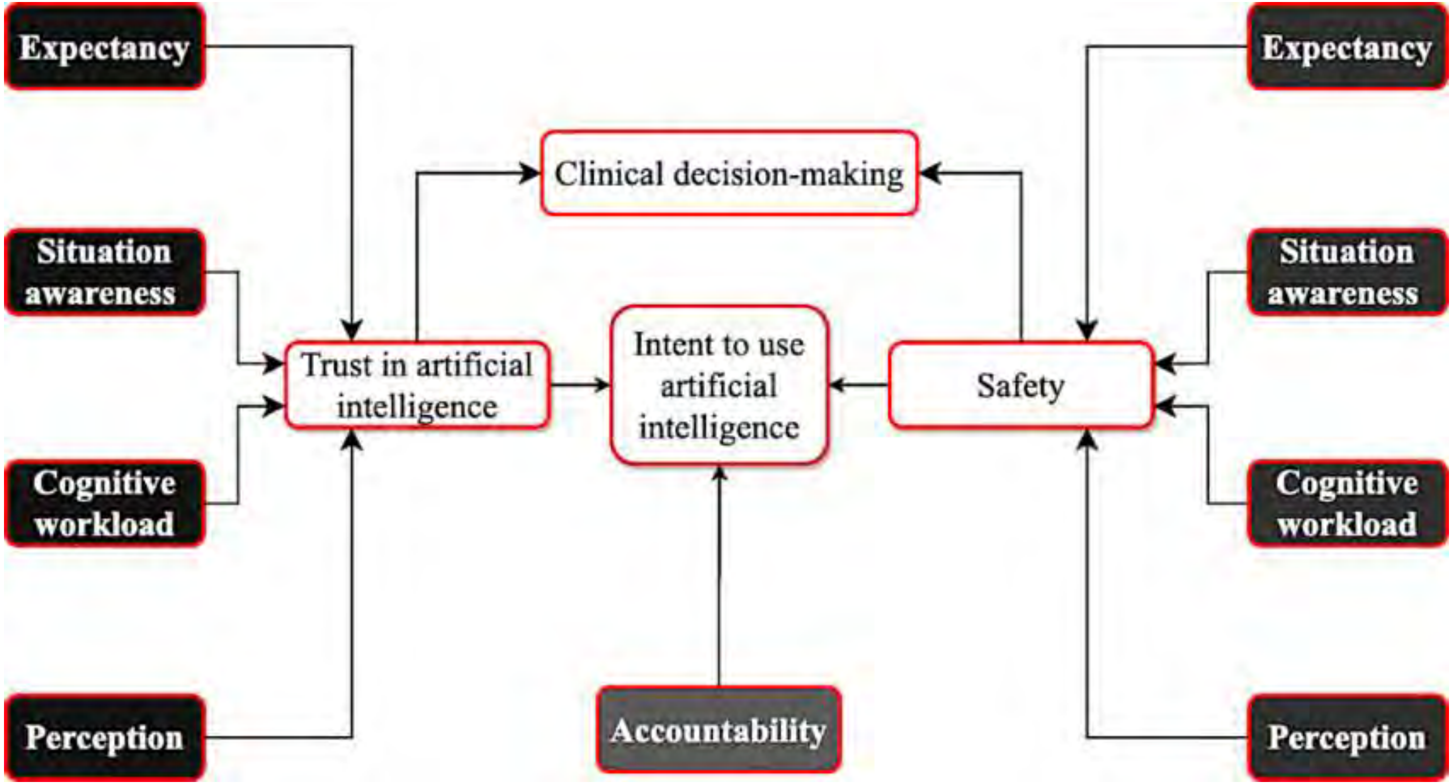


Figure 1 Deep-learning model performance with increasing amounts of training data.

Conclusion

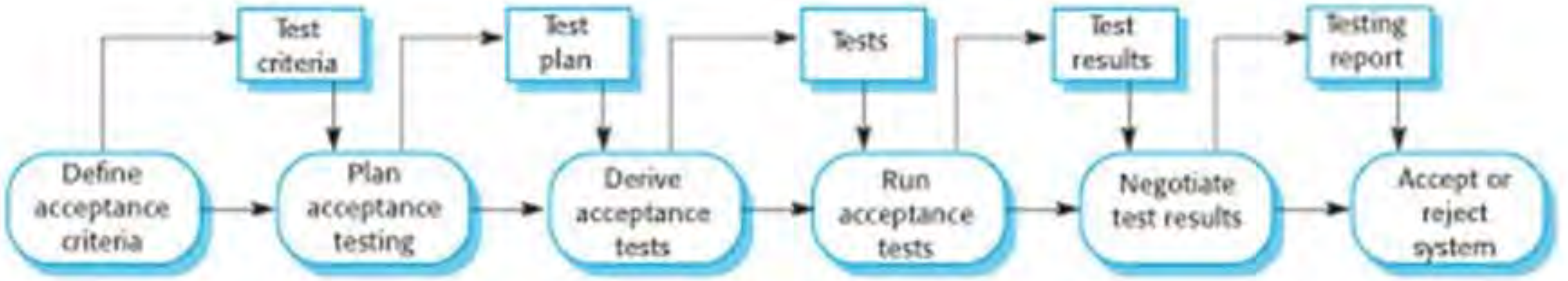
In this work, we compared traditional ML algorithms to DL algorithms for the prediction of mandible ORN resulting from HNC RT. The traditional ML algorithms performed similarly to each other when using cross-validation and were successful at predicting ORN. The performance of the ML models shows promise in clinical integration for future studies. Despite our use of different architectures and model ensembles, the DL models continued to underperform compared to the best-performing ML algorithm identified by cross-validation, logistic regression, when evaluated on the test set. When we used additional training data, no performance improvement trends were evident, suggesting that more data are needed despite the relatively large HNC patient cohort. In further work, researchers could use more

So how does AI model adoption practically occur?



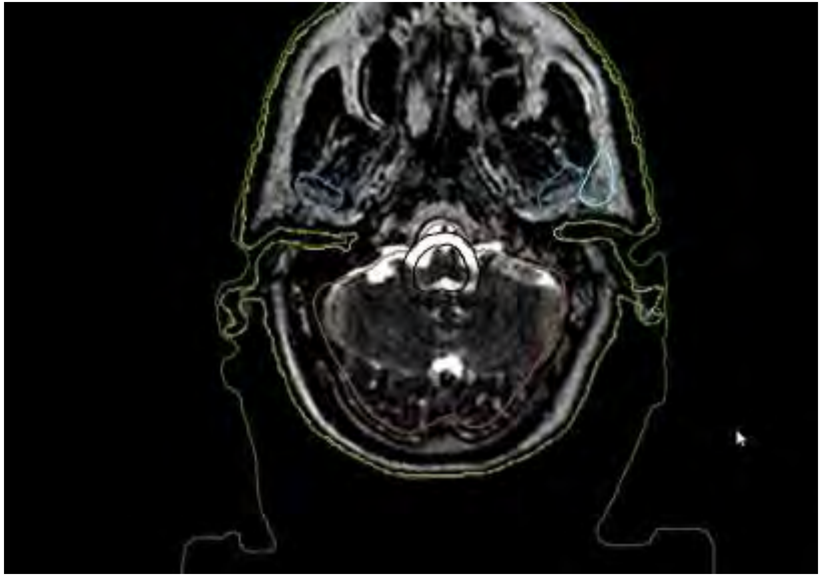
Choudhury A
Toward an Ecologically Valid Conceptual Framework for the Use of Artificial Intelligence in Clinical Settings: Need for Systems Thinking, Accountability, Decision-making, Trust, and Patient Safety Considerations in Safeguarding the Technology and Clinicians
JMIR Hum Factors 2022;9(2):e35421. doi: [10.2196/35421](10.2196/35421)

Real Life: Use-case specific acceptance testing



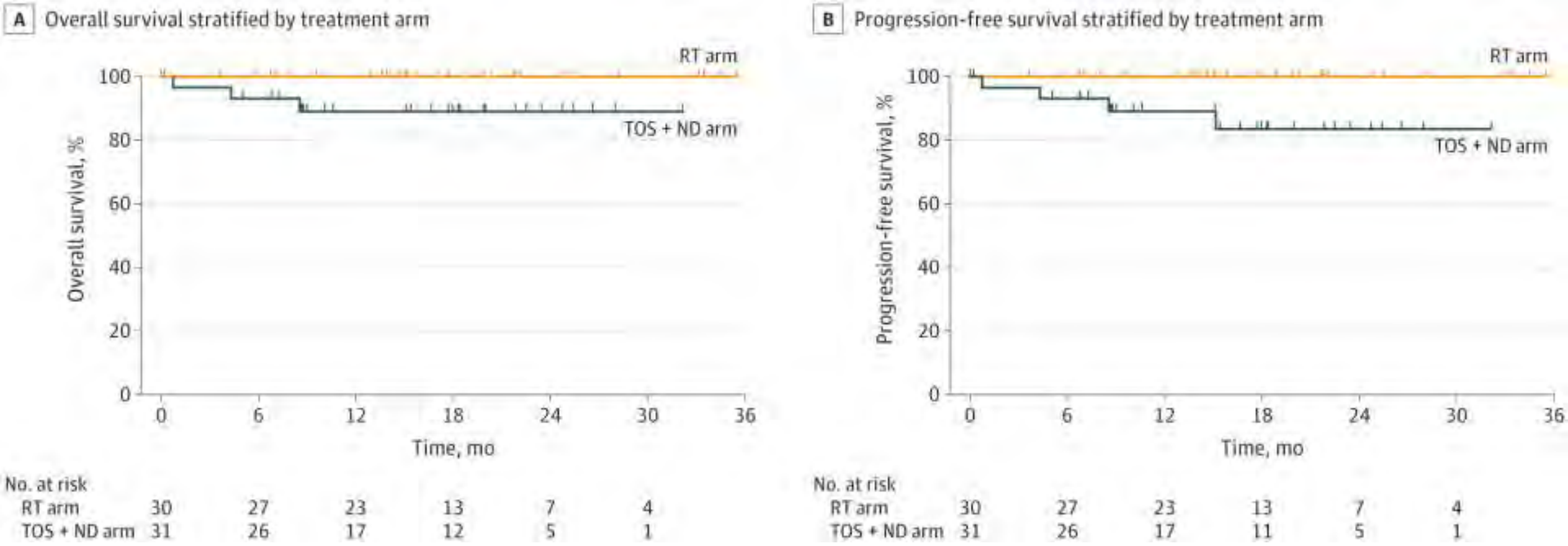
1. Define acceptance criteria
2. Plan acceptance testing
3. Derive acceptance tests
4. Run acceptance tests
5. Negotiate test results
6. Reject/accept system

Example: Decision Support Tools for Surgical vs. Non-surgical therapy selection



Example: Decision Support Tools

ORATOR2



doi:10.1001/jamaoncol.2022.0615

MDs/MDTs are bad at quantification of risk

If I do TORS, there is no PM or ECE >> Best outcome

- I have spared RT ☺

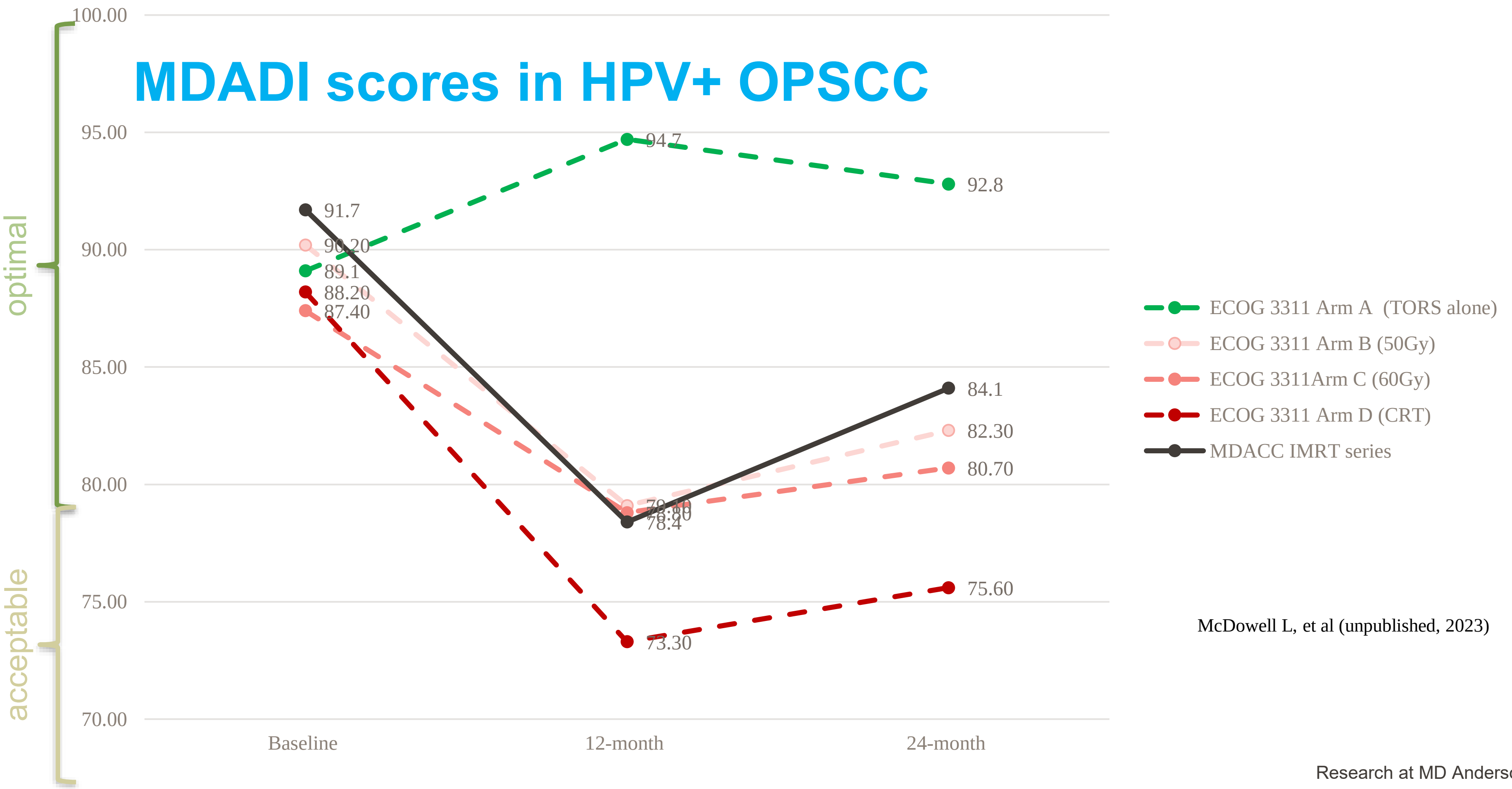
If I do TORS, and there is low volume ENE or close margin

- *Need adjuvant RT [bimodality]*
- MDADI is the same as RT alone,
- DIGEST is *worse* than RT alone 😐

If I do TORS, and there is PM or >2mm ENE

- *Need adjuvant chemoRT*
- MDADI/DIGEST is worse than chemo(RT) ☹

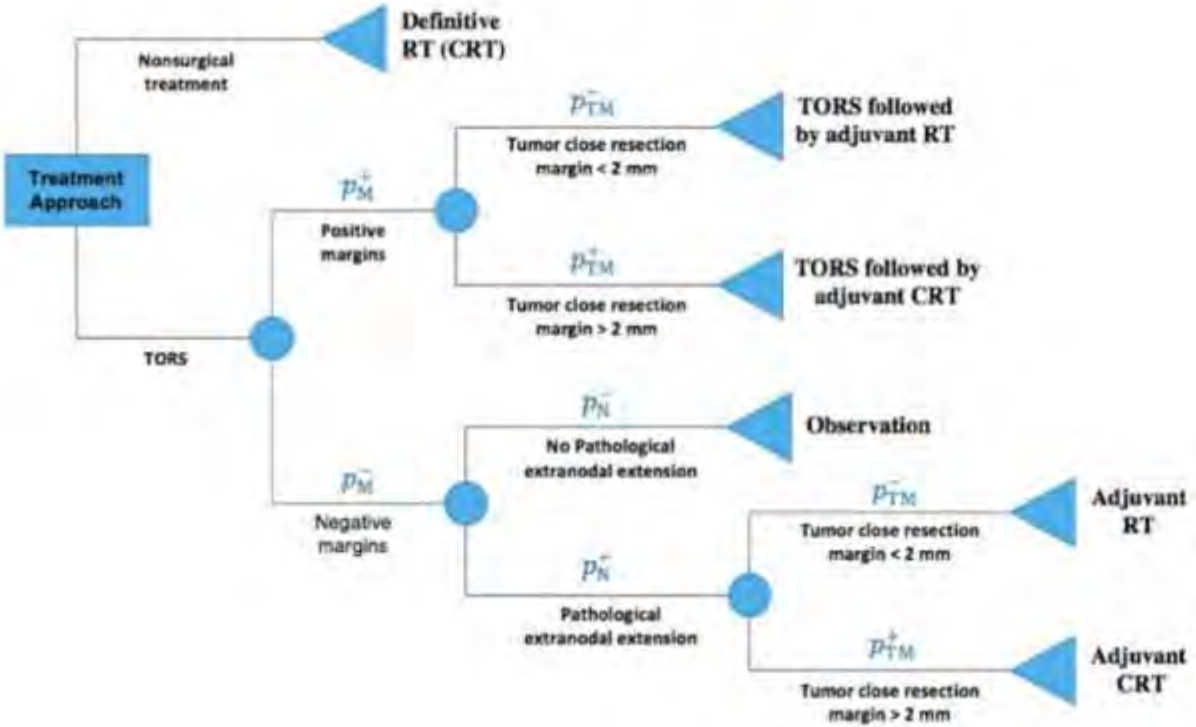
We are bad at quantification of risk



McDowell L, et al (unpublished, 2023)

DOI: 10.1002/cam4.5253

Optimized decision support for selection of transoral robotic surgery or (chemo)radiation therapy based on posttreatment swallowing toxicity



Research at MD Anderson

Optimized decision support for selection of transoral robotic surgery or (chemo)radiation therapy based on posttreatment swallowing toxicity

DOI: 10.1002/cam4.5253

TABLE 3 Range of likelihoods required for TORS and definitive therapies to become the optimal treatment under the second scenario

Instrument/measure	Confidence level of postoperative events for which TORS is optimal	Confidence level of postoperative events for which definitive RT is optimal
Scenario I		
MDADI		
Short term (3–6 months)	—	Any likelihood associated with ENE and/or PM
Long term (18–24 months)	When both ENE and PM have likelihood <70%	If either of ENE or PM has a likelihood >90%
MDASI		
Short term (3–6 months)	—	Any likelihood associated with ENE and/or PM
Long term (18–24 months)	Any likelihood associated with ENE and/or PM	—
DIGEST		
Short term (3–6 months)	When both ENE and PM have likelihood <40%	If either of ENE or PM has a likelihood >75%
Long term (18–24 months)	When both ENE and PM have likelihood <10%	If either of ENE or PM has a likelihood >25%

Instrument/measure	Confidence level of postoperative events for which TORS is optimal	Confidence level of postoperative events for which definitive CRT is optimal
Scenario II		
MDADI		
Short term (3–6 months)	Any likelihood associated with ENE and/or PM	—
Long term (18–24 months)	Any likelihood associated with ENE and/or PM	—
MDASI		
Short term (3–6 months)	Any likelihood associated with ENE and/or PM	—
Long term (18–24 months)	Any likelihood associated with ENE and/or PM	—
DIGEST		
Short term (3–6 months)	When both ENE and PM have likelihood <55%	If either of ENE or PM has a likelihood >80%
Long term (18–24 months)	When both ENE and PM have likelihood <20%	If either of ENE or PM has a likelihood >40%

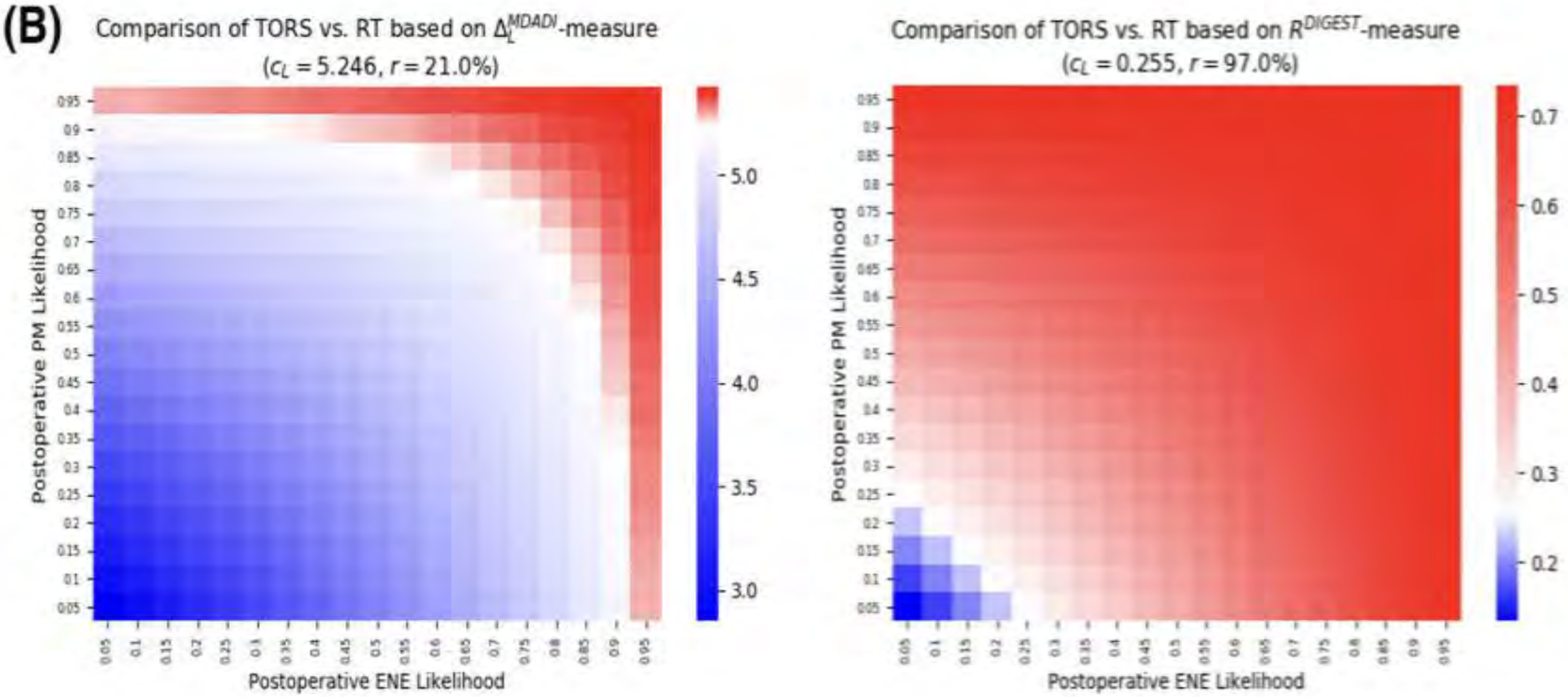
Abbreviations: ENE, postoperative extranodal extension; PM, postoperative positive margin.

Research at MD Anderson

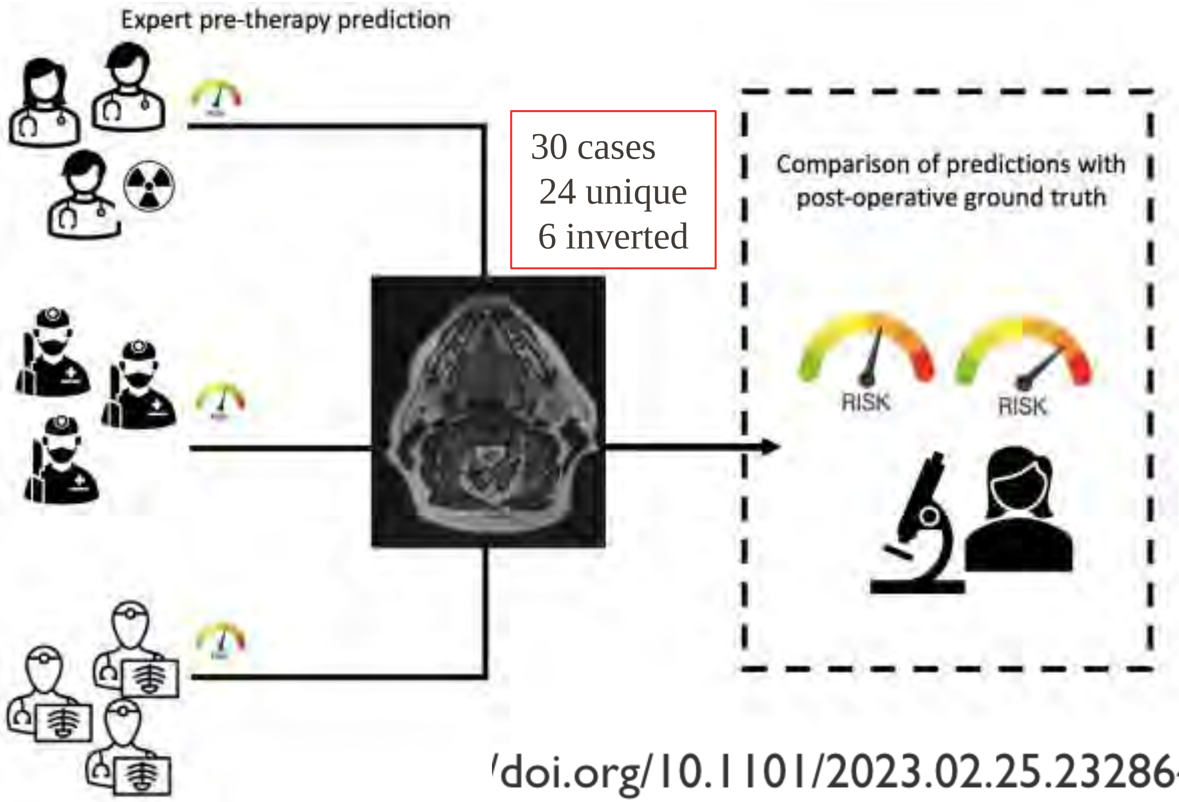
DOI: 10.1002/cam4.5253

Optimized decision support for selection of transoral robotic surgery or (chemo)radiation therapy based on posttreatment swallowing toxicity

Red==RT better Blue==TORS better



Multi-Specialty Expert Physician Identification of Extranodal Extension in Computed Tomography Scans of Oropharyngeal Cancer Patients: Prospective Blinded Human Inter-Observer Performance Evaluation

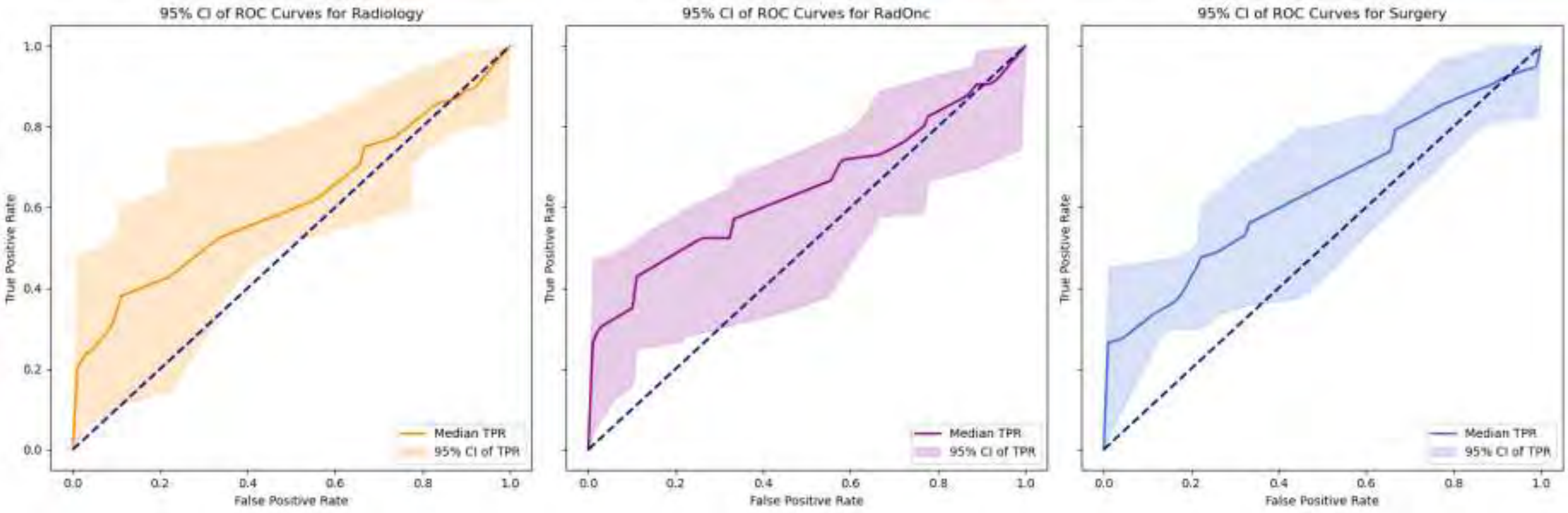


/doi.org/10.1101/2023.02.25.23286432

Expert head and neck physicians (n = 34):
- Radiation Oncologists (n = 11)
- Radiologists (n = 11)
- Surgeons (n = 11)

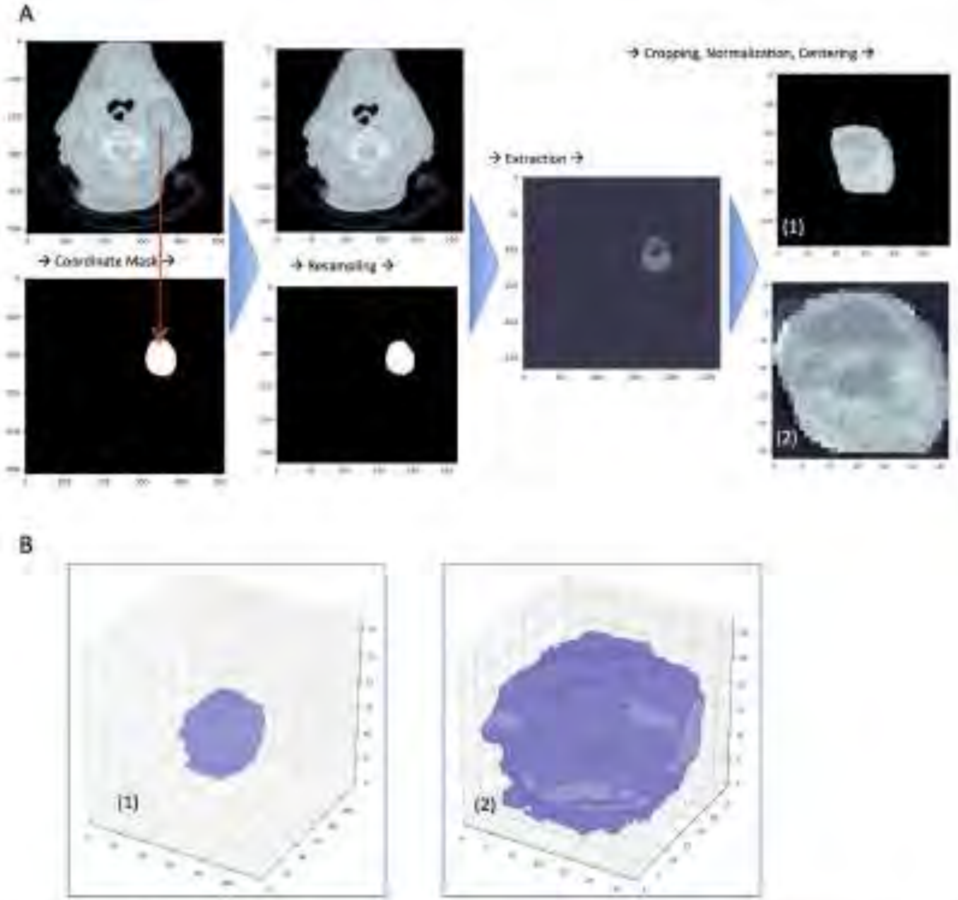
Radiographic criteria:
- Indistinct capsular contour
- Irregular lymph node margin
- Thick-walled enhancing nodal margin
- Perinodal fat stranding
- Perinodal fat plane or gross invasion
- Nodal necrosis
- Nodal matting

Problem:
Humans are crummy at pathologic ENE (pre)detection



Pretreatment Identification of Head and Neck Cancer Nodal Metastasis and Extranodal Extension Using Deep Learning Neural Networks

Benjamin H. Kann[1], Sanjay Aneja[1], Gokoulakrichenane V. Loganadane[1], Jacqueline R. Kelly[1], Stephen M. Smith[2], Roy H. Decker[1], James B. Yu[1], Henry S. Park[1], Wendell G. Yarbrough[3], Ajay Malhotra[4], Barbara A. Burtness[5] & Zain A. Husain[1]



SCIENTIFIC REPORTS | (2018) 8:14036 | DOI:10.1038/s41598-018-32441-y

Figure 1. (**A**,**B**) Lymph Node Region of Interest Preprocessing. (**A**) 2D representation of 3D lymph node segmentation preprocessing resulting in a dimension-preserving input (1) and a size-invariant, "zoomed-in" input (2). (**B**) Representation of actual 3D input arrays for dual-input deep learning neural network.

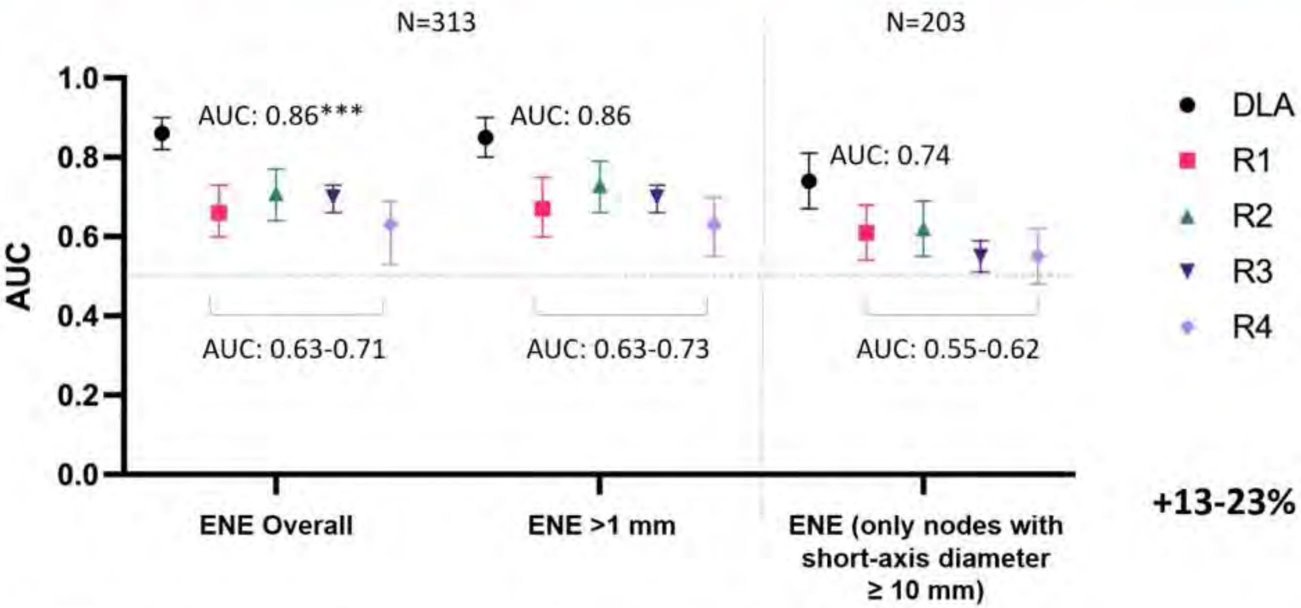
Pretreatment Identification of Head and Neck Cancer Nodal Metastasis and Extranodal Extension Using Deep Learning Neural Networks

SCIENTIFIC REPORTS | (2018) 8:14036 | DOI:10.1038/s41598-018-32441-y

Performance Metric	Extranodal Extension (ENE) ENE Test Set (n = 98*) DualNet DLNN	Random Forest	Benchmark Logistic	Nodal Metastasis (NM) Test Set (n = 131) DualNet DLNN	Random Forest	Benchmark Logistic
AUC	0.91	0.88	0.81	0.91	0.91	0.86
Accuracy	85.7%	82.6%	77.7%	85.5%	84.7%	76.1%
Sensitivity	0.88	0.79	0.72	0.84	0.75	0.79
Specificity	0.85	0.84	0.80	0.87	0.92	0.74
PPV	0.66	0.61	0.54	0.88	0.87	0.69
NPV	0.95	0.93	0.89	0.82	0.83	0.83
Youden Index	0.73	0.63	0.51	0.71	0.67	0.53

Table 3. Model Performance and Benchmark Comparisons on Independent Test Set By Lymph Node Feature. *Test set for ENE includes lymph nodes with region of interest diameters ≥ 1 cm. Abbreviations: AUC = area under the curve; PPV = positive predictive value; NPV = negative predictive value. Youden index = Sensitivity + Specificity − 1.

Deep learning outperforms radiologists for ENE prediction in E3311



+13-23% AUC for DLA

***All P-values < 0.001 for DLA vs R1-4 comparisons. Error bars represent 95% CIs

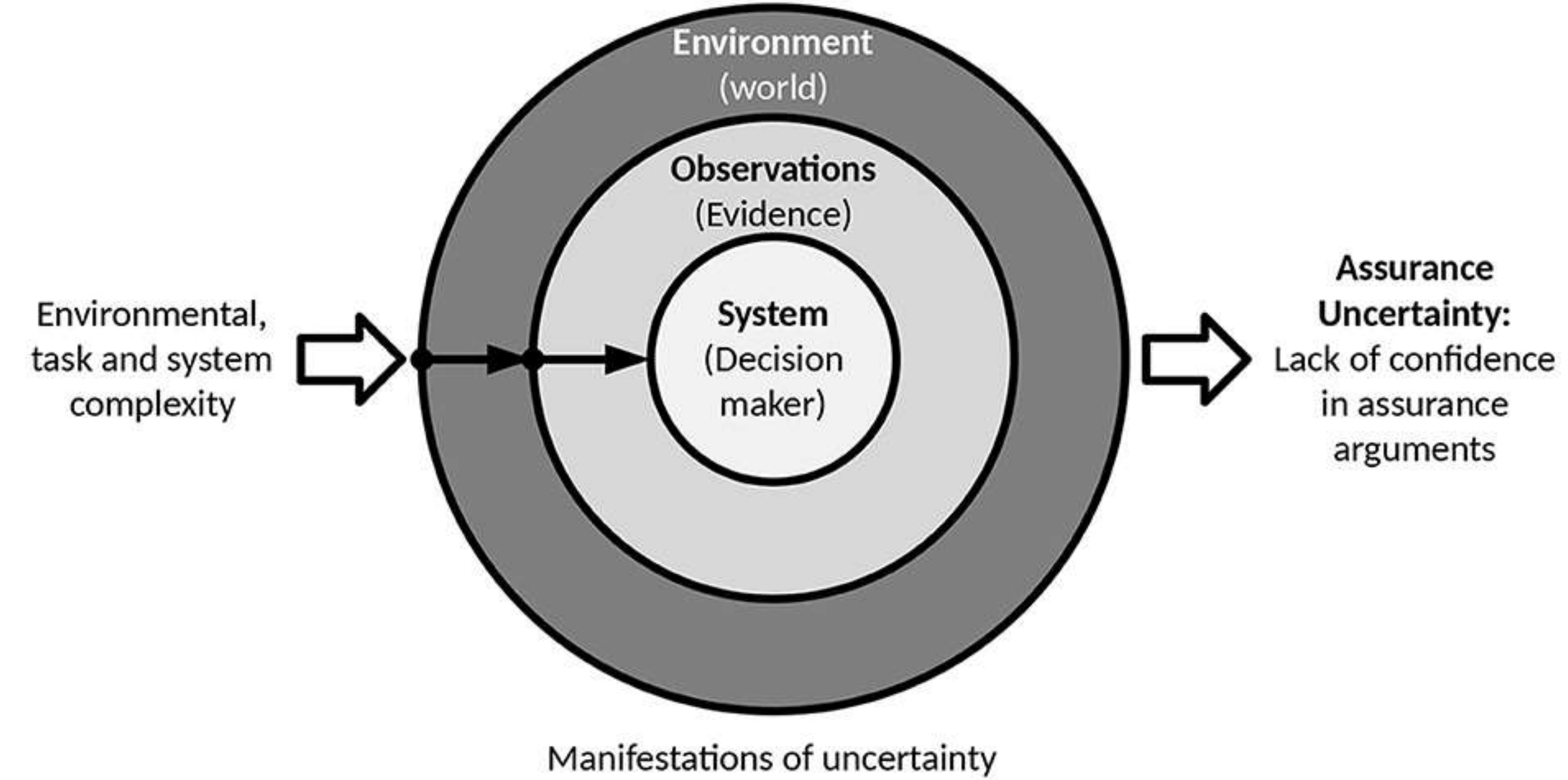
Manuscript under review

DLA outperformed expert radiologists and can have clinical utility in selection of patients appropriate for operative management and other de-escalation (or escalation) strategies for HPV + OPC

So, why aren't we using these tools?

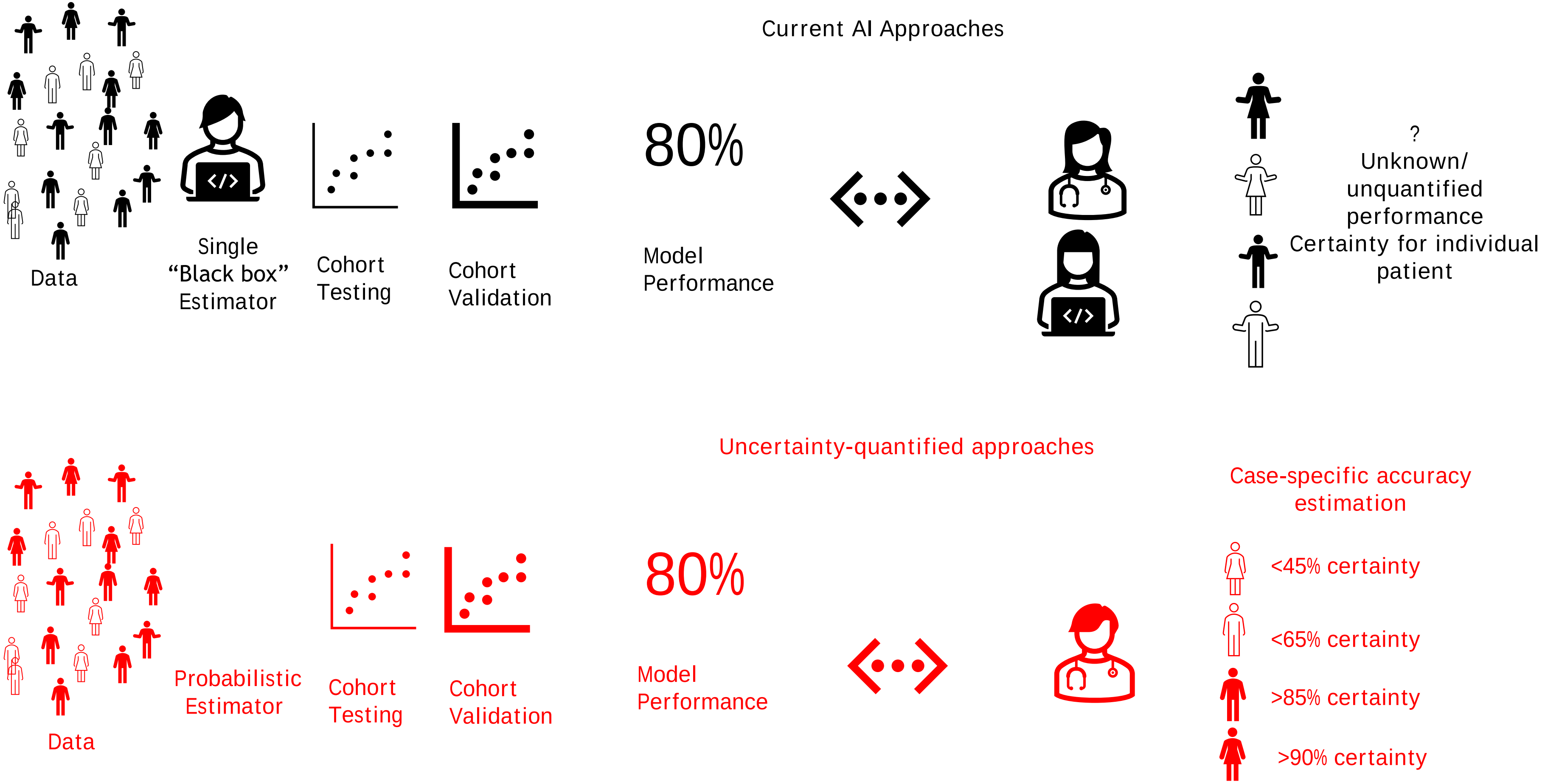
- **"I'm not sure about *this* case…"**
- **"What if it misses a node?"**
- **"I just don't trust it like I trust my colleagues…"**

The current clinical problem: Trustworthiness/Uncertainty Estimation



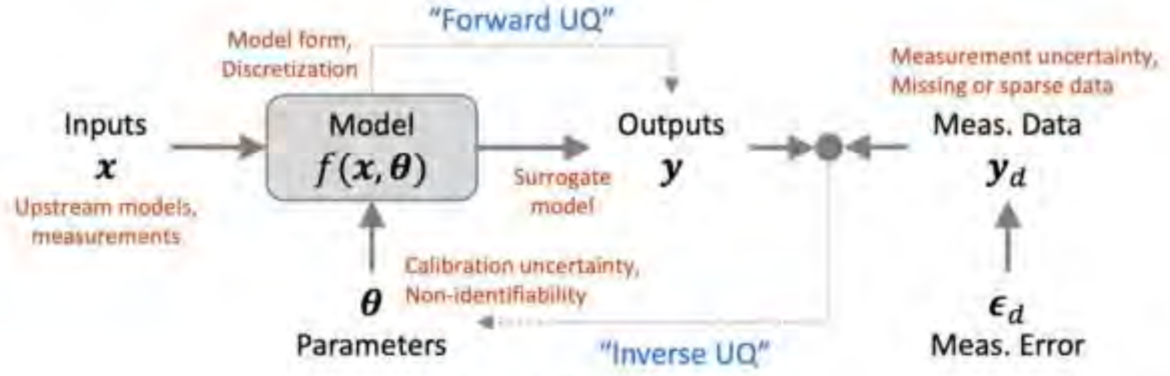
Burton & Herd, Addressing uncertainty in the safety assurance of machine-learning. Front. Comput. Sci., 06 April 2023
Sec. Software
Volume 5 - 2023 | https://doi.org/10.3389/fcomp.2023.1132580

The current clinical problem: Trustworthiness/Uncertainty Estimation



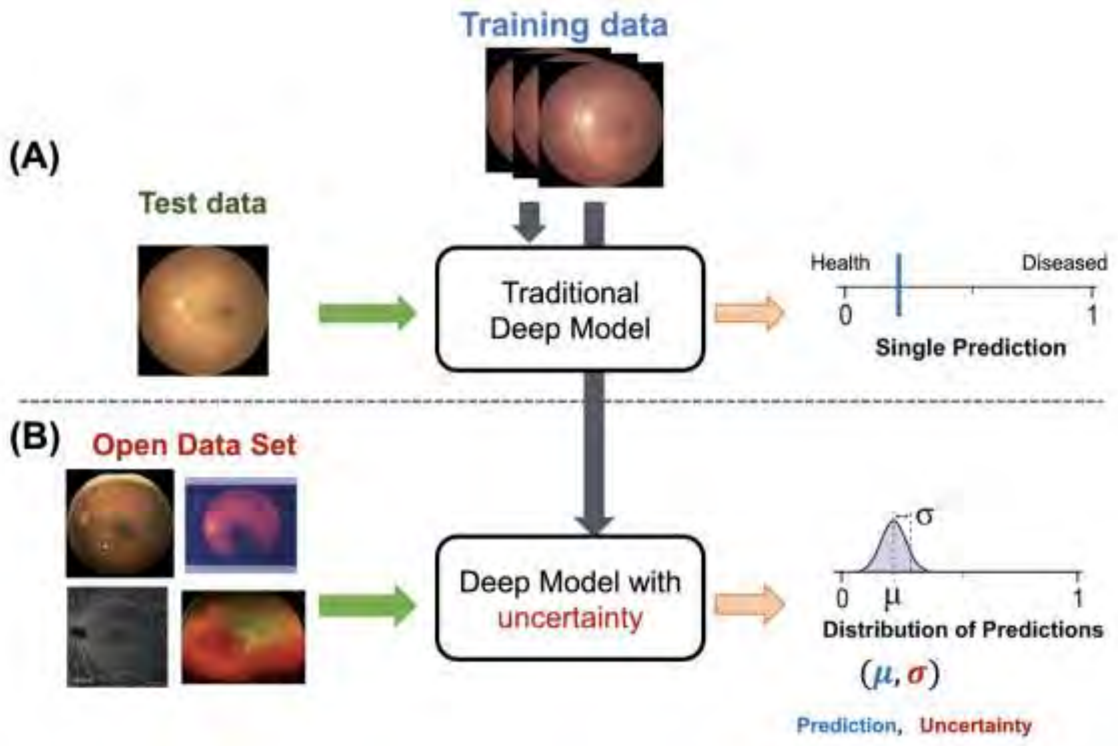
Statement: Without uncertainty quantification, we cannot move forward

UQ consists of activities such as model verification, sensitivity analysis, calibration, surrogate modeling, validation, and uncertainty propagation. *Forward UQ* quantifies uncertainty in the model output given uncertainties in the inputs, model parameters, and model errors. *Inverse UQ* is related to model calibration which updates model parameter uncertainty using measurements (which are also uncertain).



Sources of Uncertainty

The current clinical problem: Trustworthiness/Uncertainty Estimation



K. Zou et al. Meta-Radiology 1 (2023) 100003
https://doi.org/10.1016/j.metrad.2023.100003

The current clinical problem:
Tı

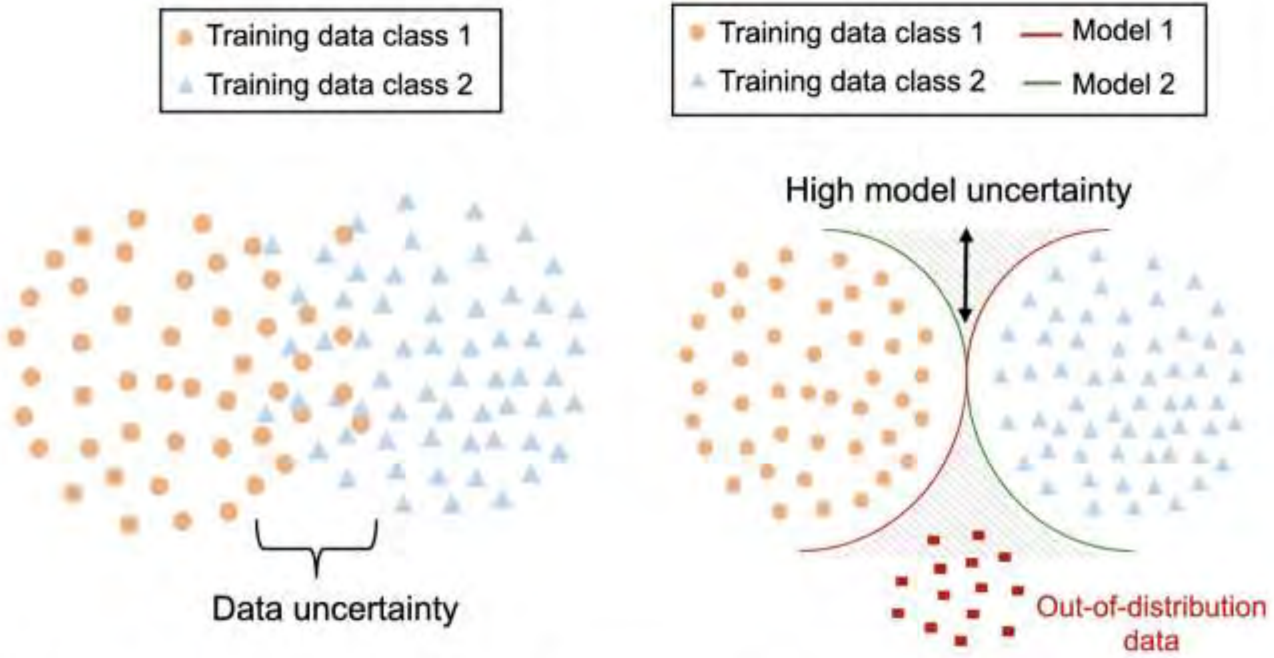


Fig. 2. Visualization of the aleatoric (data) and the epistemic (model) uncertainty for the classification model.

K. Zou et al. Meta-Radiology 1 (2023) 100003
https://doi.org/10.1016/j.metrad.2023.100003

The current clinical problem: Trustworthiness/Uncertainty Estimation

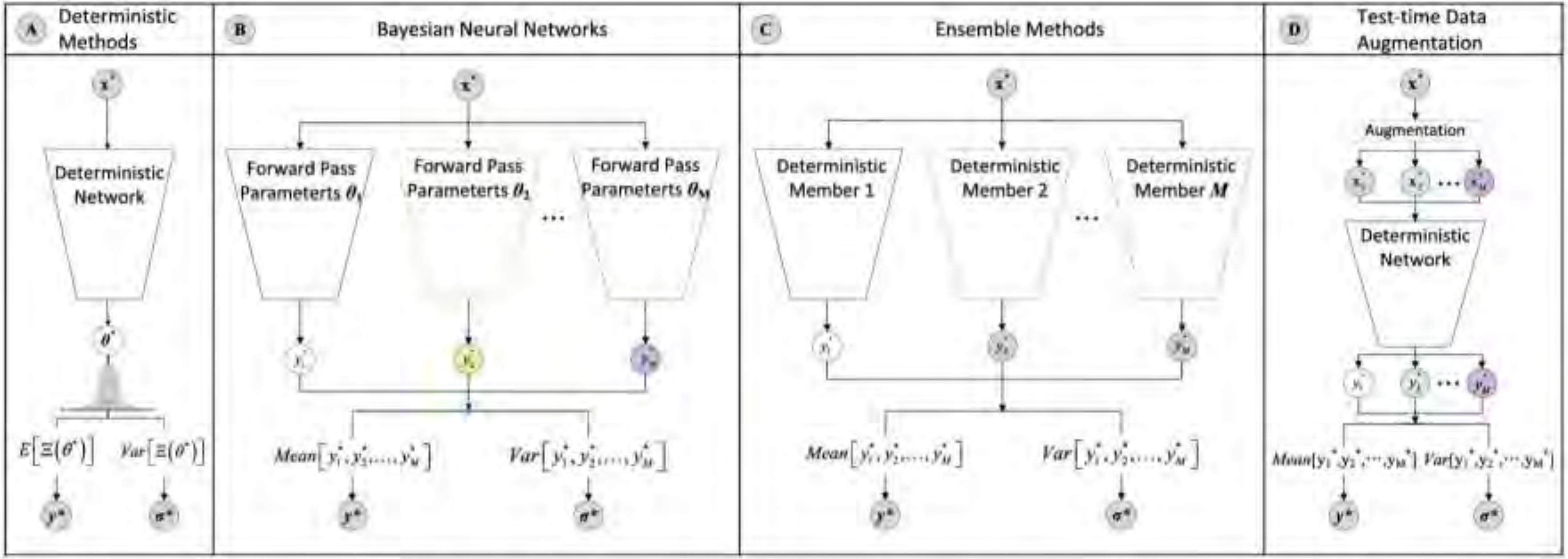
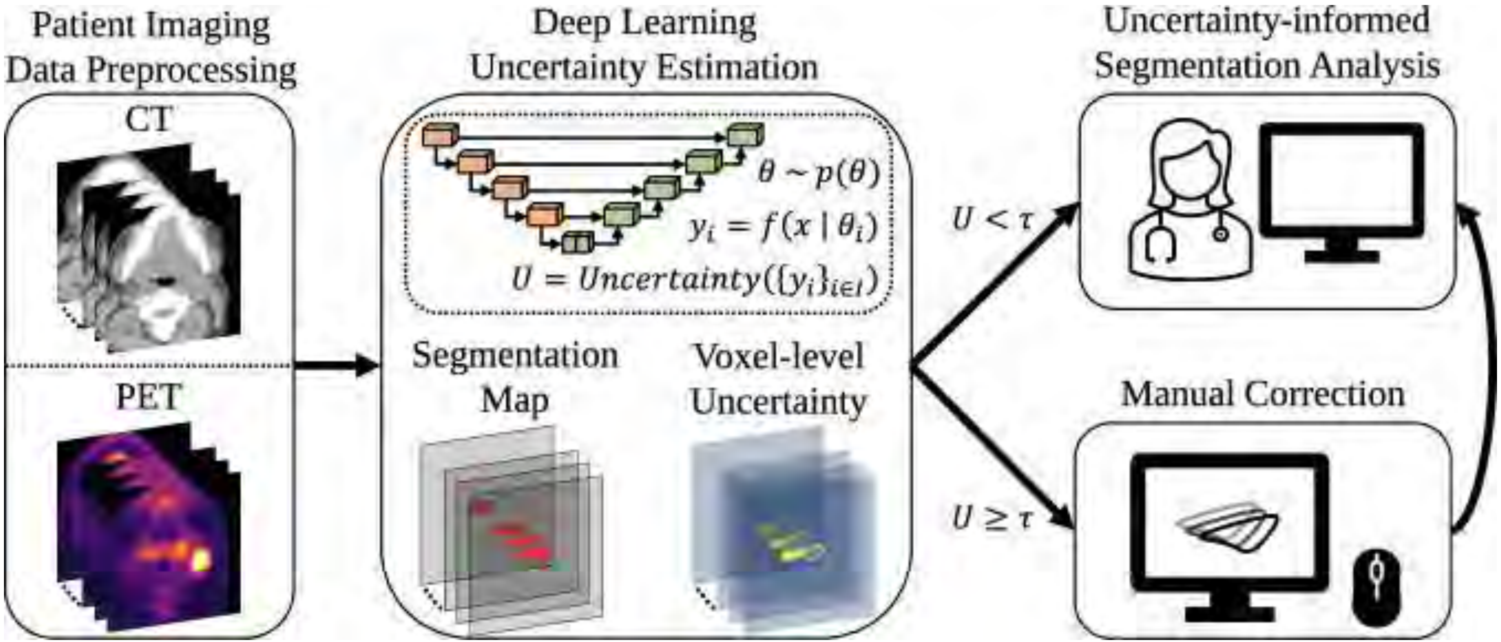


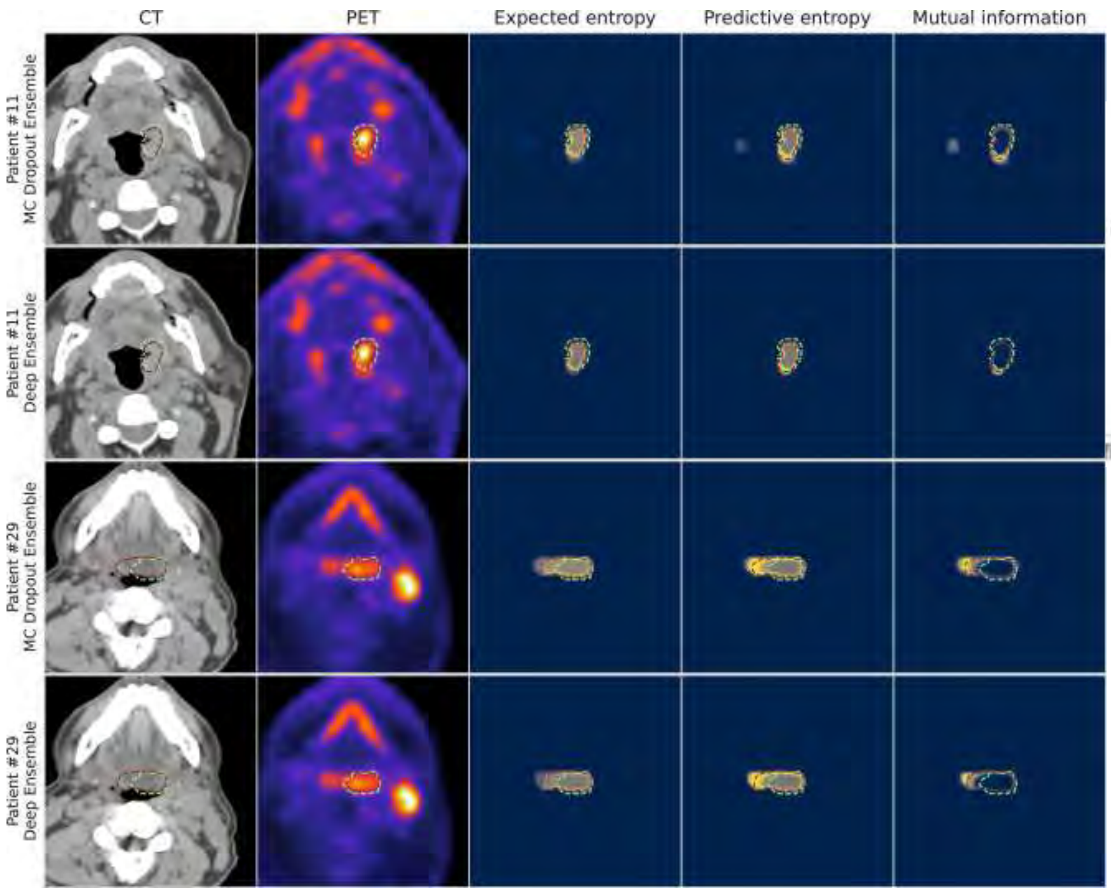
Fig. 3. The different methods of uncertainty estimation.

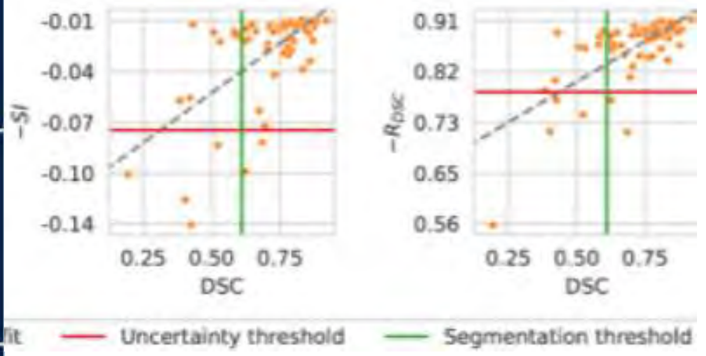
K. Zou et al. Meta-Radiology 1 (2023) 100003
https://doi.org/10.1016/j.metrad.2023.100003

Application of simultaneous uncertainty quantification and segmentation for oropharyngeal cancer use-case with Bayesian deep learning



> Commun Med (Lond). 2024 Jun 8;4(1):110. doi: 10.1038/s43856-024-00528-5.

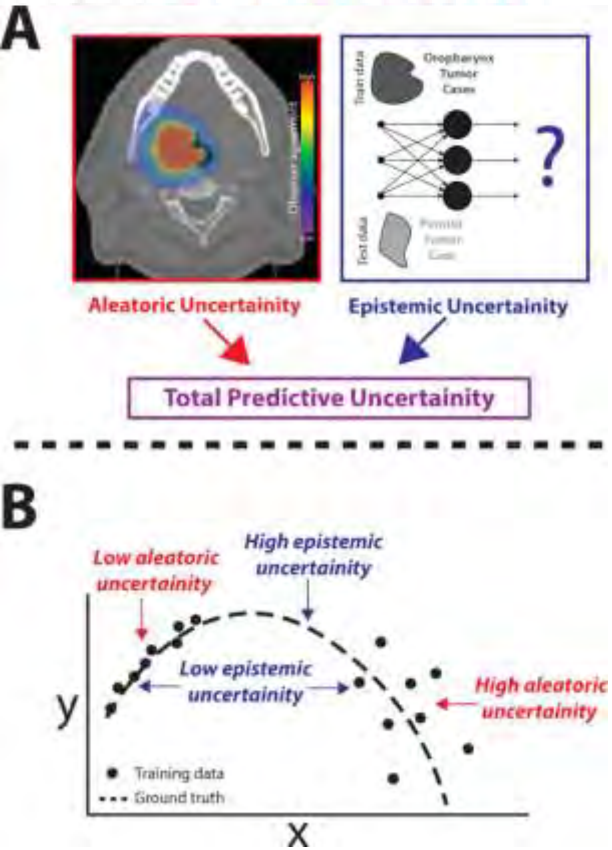


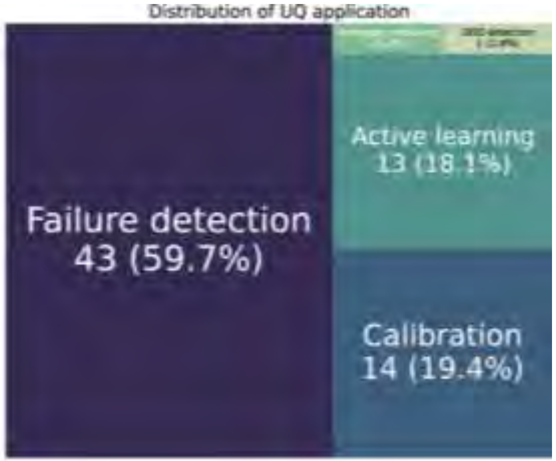


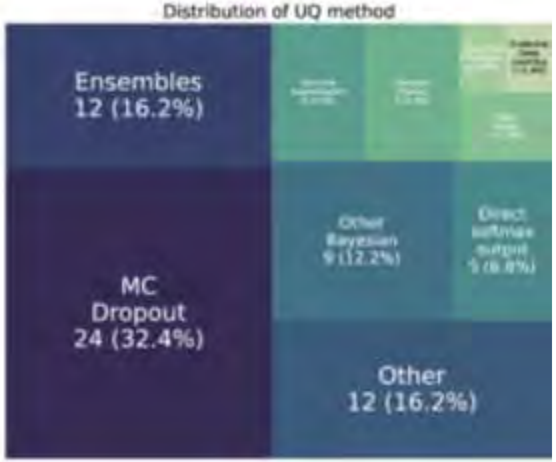
> Commun Med (Lond). 2024 Jun 8;4(1):110. doi: 10.1038/s43856-024-00528-5.

Artificial Intelligence Uncertainty Quantification in Radiotherapy Applications – A Scoping Review

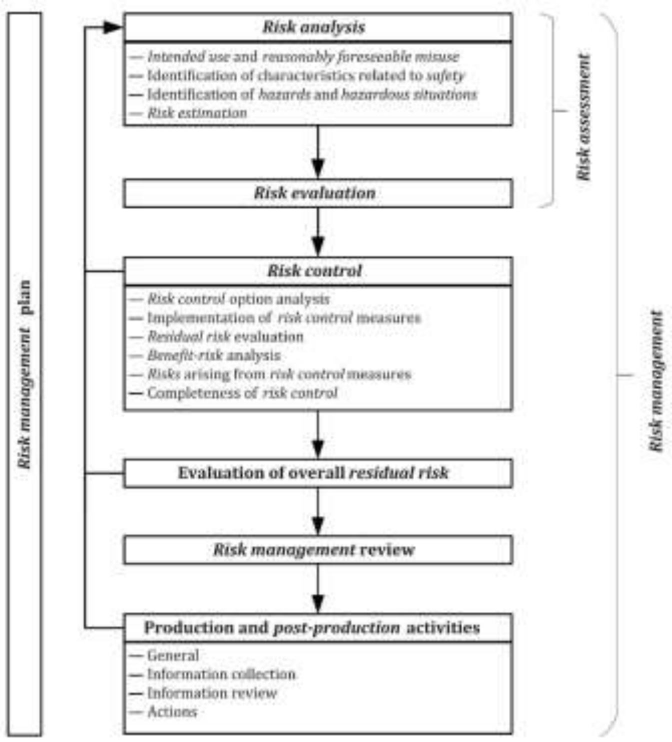
PMCID: PMC11118597

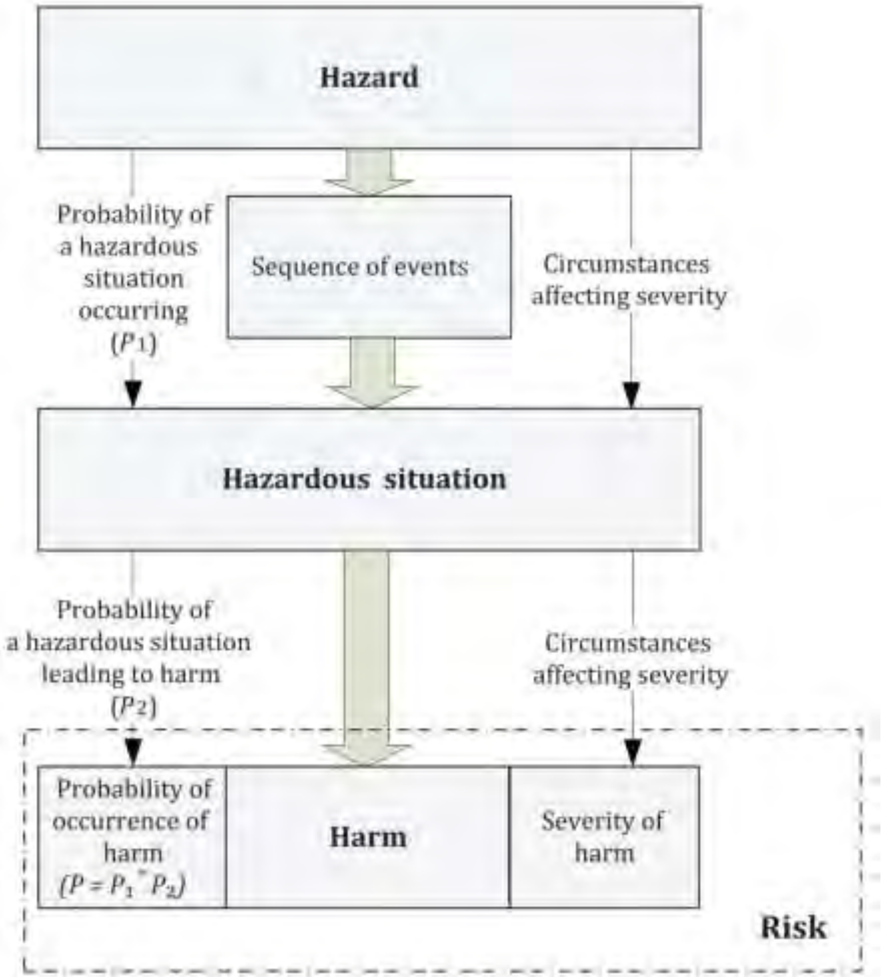






Uncertainty estimation allows **direct** safety assessment





Risk Estimation flow charts from ISO 14971:2019

Breiman's "Two Cultures" Revisited and Reconciled

Subhadeep Mukhopadhyay[1]
deep@unitedstatalgo.com

Kaijun Wang
kwang2@fredhutch.org

https://arxiv.org/abs/2005.13596

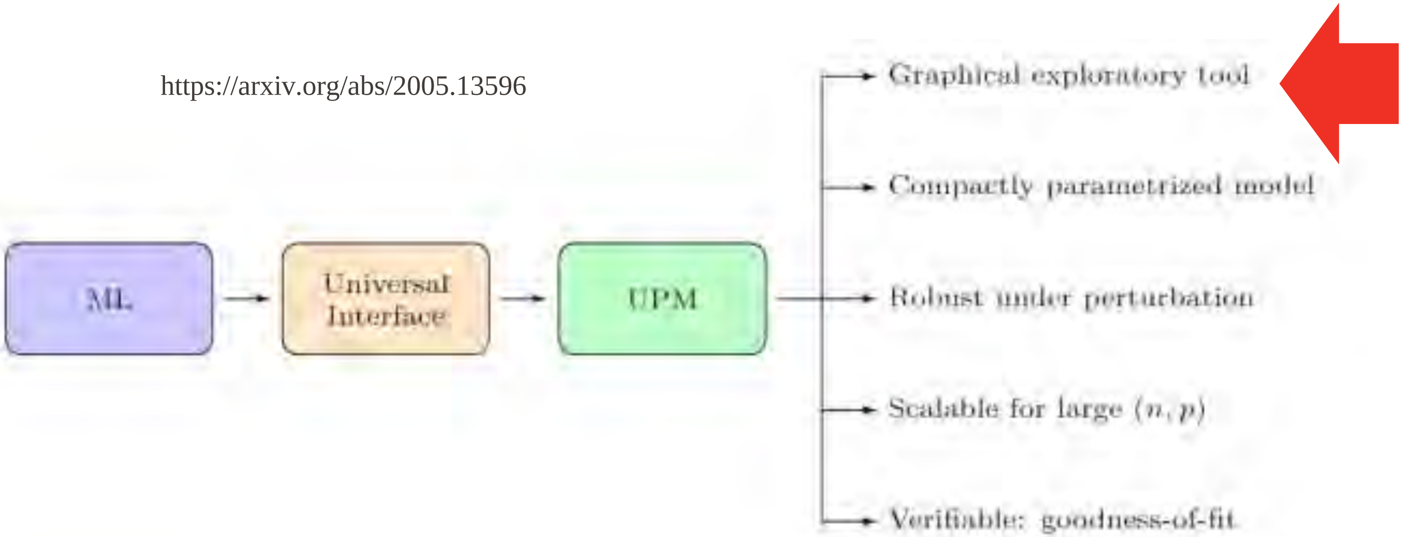
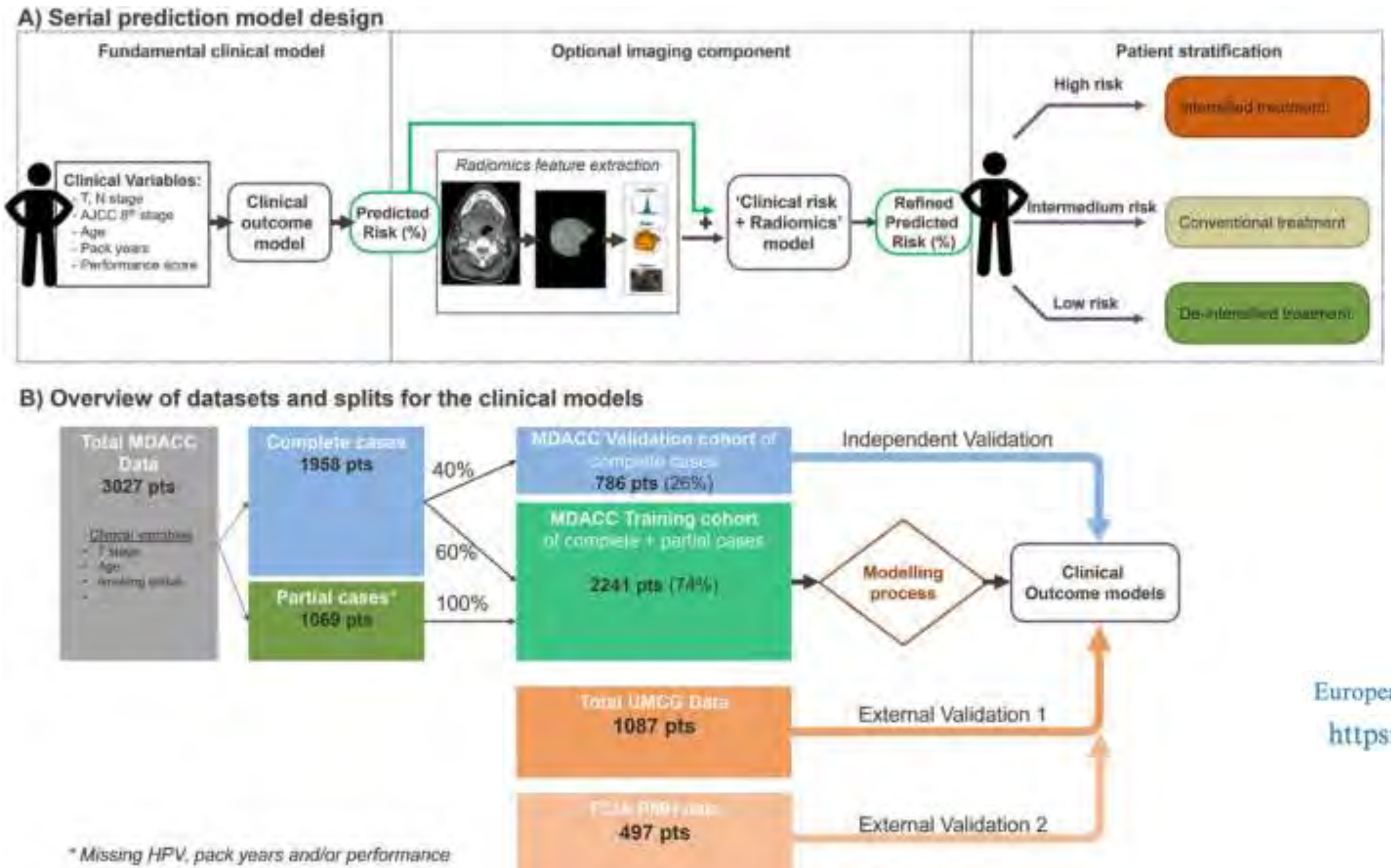


Figure 3: Integrated statistical learning framework at a glance; 'ML' stands for (an arbitrary) machine learning algorithm, and 'UPM' denotes uncertainty prediction machine.

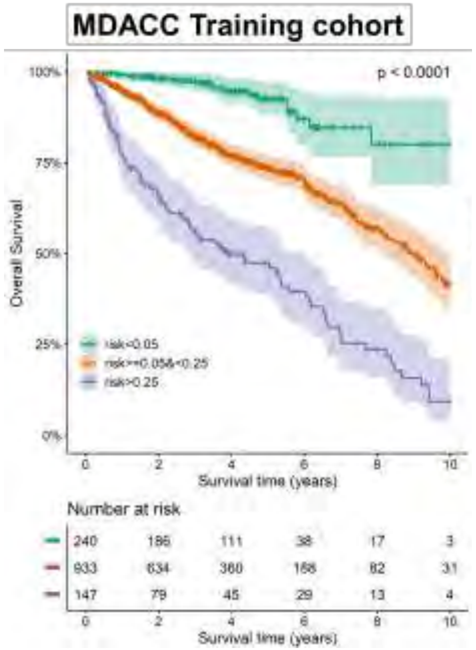
Oncologic prediction GUI



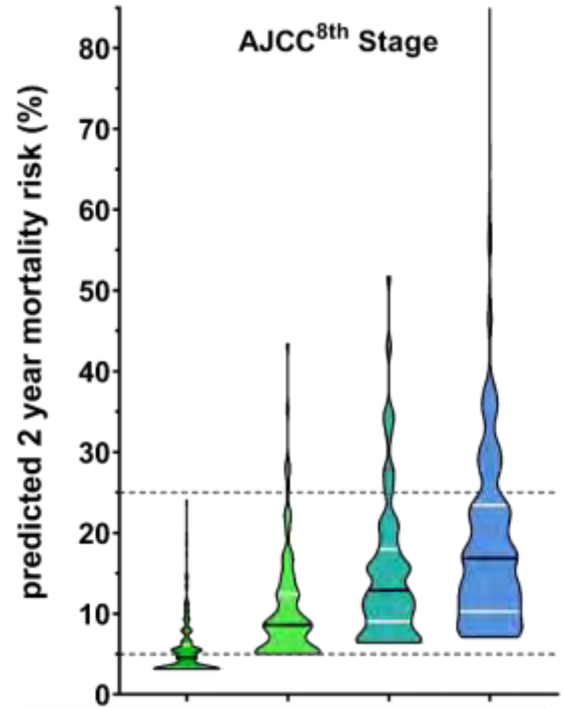
Sanne van Dijk, PhD
UMC Gronigen

European Journal of Cancer 178 (2023) 150–161
https://doi.org/10.1016/j.ejca.2022.10.011

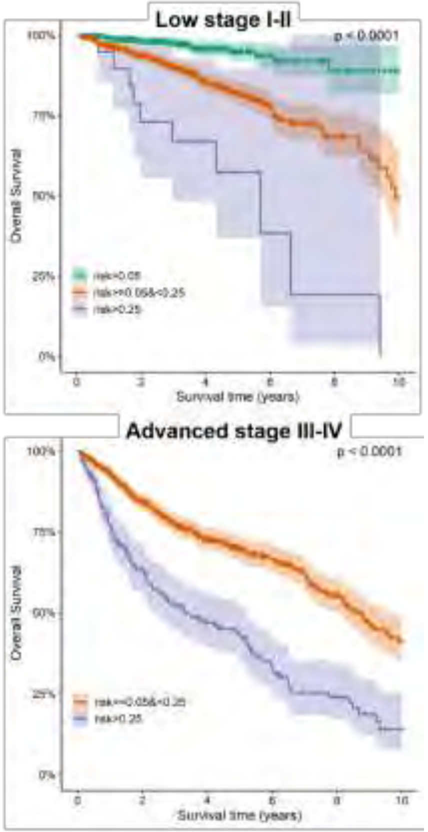
	MDACC Train
c-index	**0.71** [0.65-0.77]



https://doi.org/10.1016/j.ejca.2022.10.011



Stage:	I	II	III	IVa/b
2y-risk:	6%	8%	15%	18/35%



Sanne van Dijk, PhD
UMC Gronigen

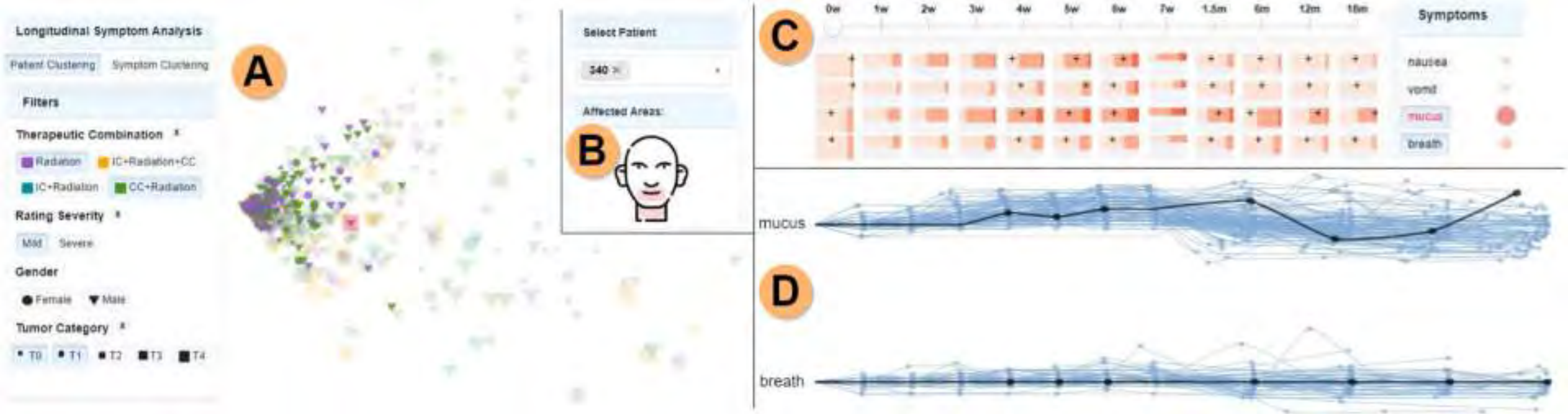
https://doi.org/10.1016/j.ejca.2022.10.011

Web-based individual OS risk prediction in new patients

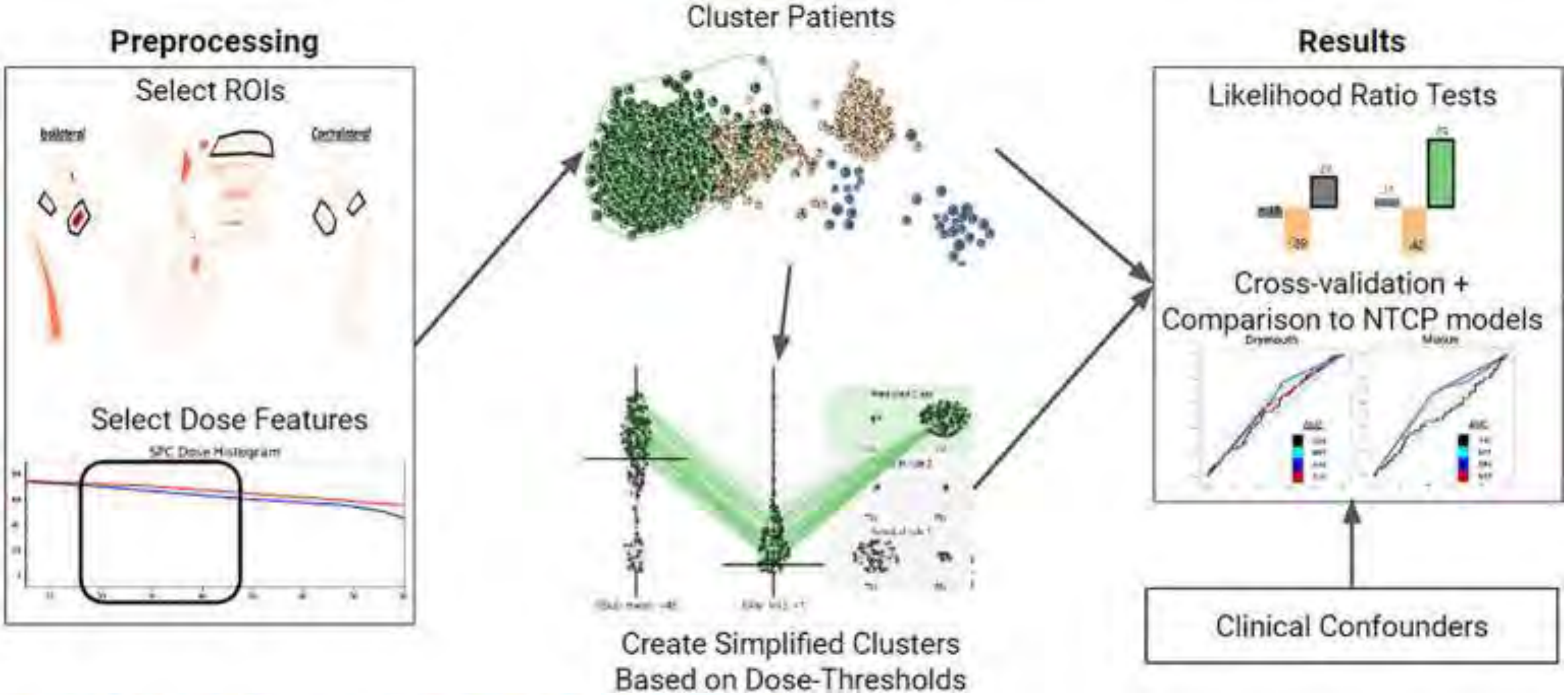
https://doi.org/10.1016/j.ejca.2022.10.011

THALIS: Human-Machine Analysis of Longitudinal Symptoms in Cancer Therapy

IEEE Trans Vis Comput Graph. 2022 Jan;28(1):151-161. doi: 10.1109/TVCG.2021.3114810.

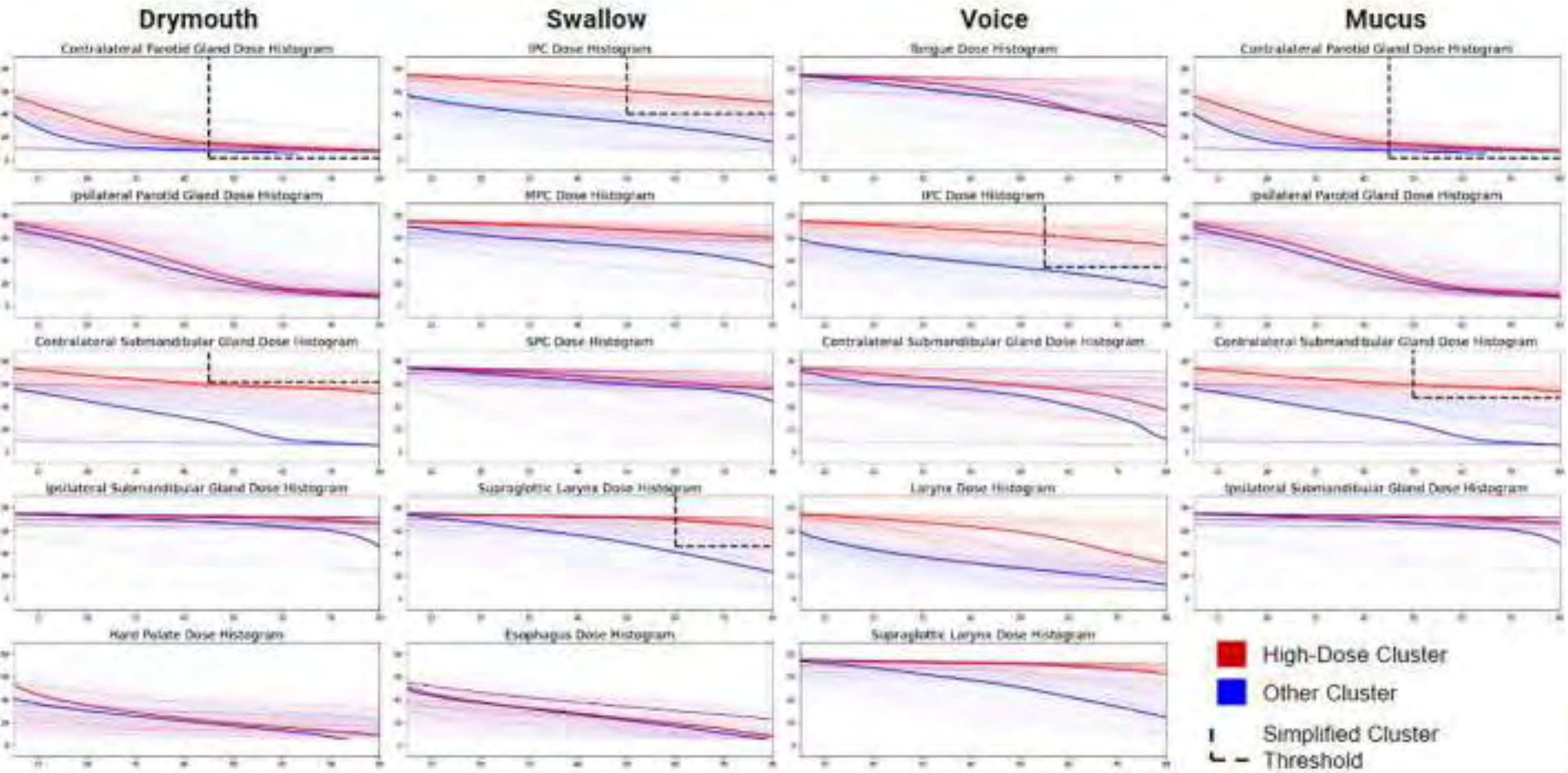


Multi-Organ Spatial Stratification of 3-D Dose Distributions Improves Risk Prediction of Long-Term Self-Reported Severe Symptoms in Oropharyngeal Cancer Patients Receiving Radiotherapy: Development of a Pre-Treatment Decision Support Tool.



DOI: 10.3389/fonc.2023.1210087

Predicting dynamic injury AND response kinetics

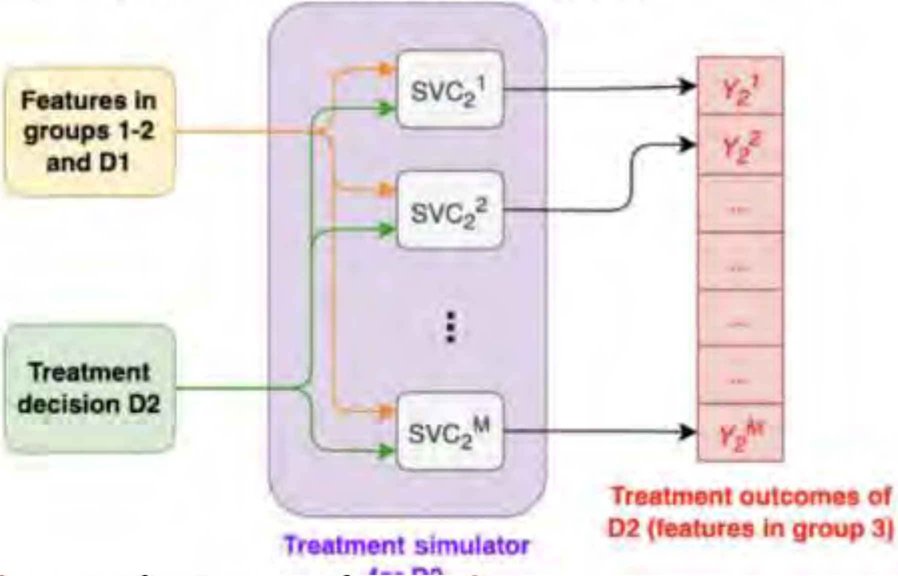


 doi: 10.2196/29455

Optimal Treatment Selection in Sequential Systemic and Locoregional Therapy of Oropharyngeal Squamous Carcinomas: Deep Q-Learning With a Patient-Physician Digital Twin Dyad

(J Med Internet Res 2022;24(4):e29455)

Figure 4. Illustration of the treatment simulator for D2. Those for D1 and D3 are similar, and their input features are from group 1 and groups 1-3, respectively. SVC: support vector classifier. D1: decision 1. D2: decision 2. D3: decision 3.



Research

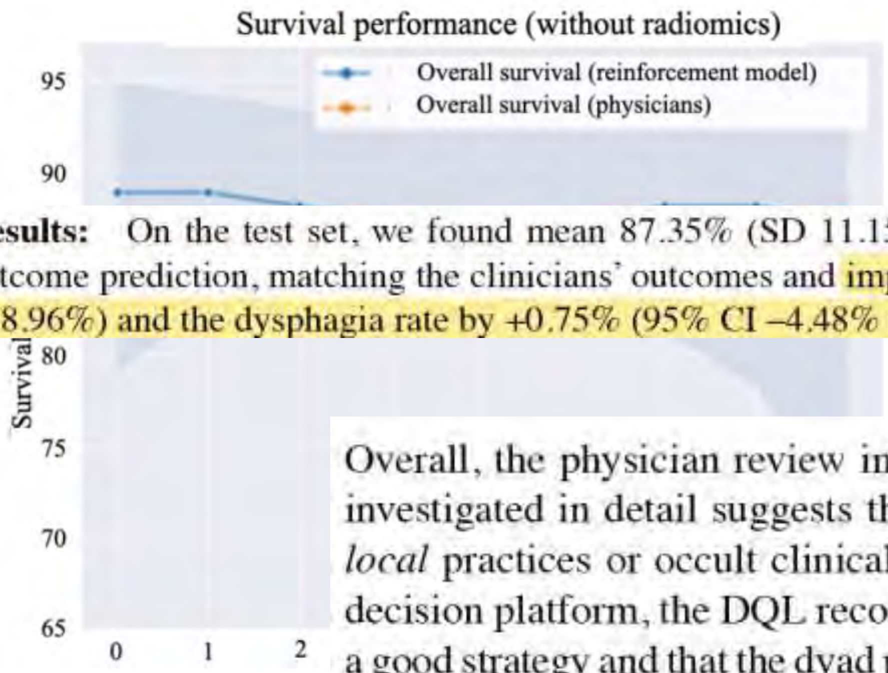
AI is good at survival prediction AND selecting therapy based on toxicity]

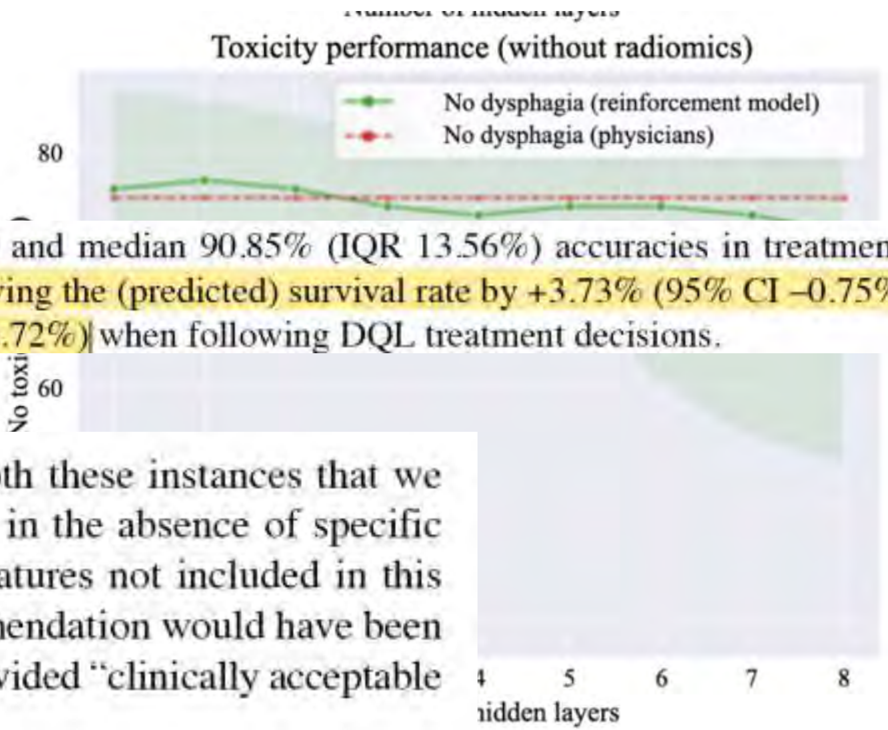
Optimal Treatment Selection in Sequential Systemic and Locoregional Therapy of Oropharyngeal Squamous Carcinomas: Deep Q-Learning With a Patient-Physician Digital Twin Dyad

Tardini et al

doi: 10.2196/29455

(J Med Internet Res 2022;24(4):e29455)





Results: On the test set, we found mean 87.35% (SD 11.15%) and median 90.85% (IQR 13.56%) accuracies in treatment outcome prediction, matching the clinicians' outcomes and improving the (predicted) survival rate by +3.73% (95% CI –0.75% to 8.96%) and the dysphagia rate by +0.75% (95% CI –4.48% to 6.72%) when following DQL treatment decisions.

Overall, the physician review in both these instances that we investigated in detail suggests that, in the absence of specific *local* practices or occult clinical features not included in this decision platform, the DQL recommendation would have been a good strategy and that the dyad provided "clinically acceptable recommendations."

DITTO: A Visual Digital-twin for Interventions and Temporal Treatment Outcomes in Head and Neck Cancer

Andrew Wentzel, Serageldin Attia, Xinhua Zhang, Guadalupe Canahuate, Clifton David Fuller, and G.Elisabeta Marai

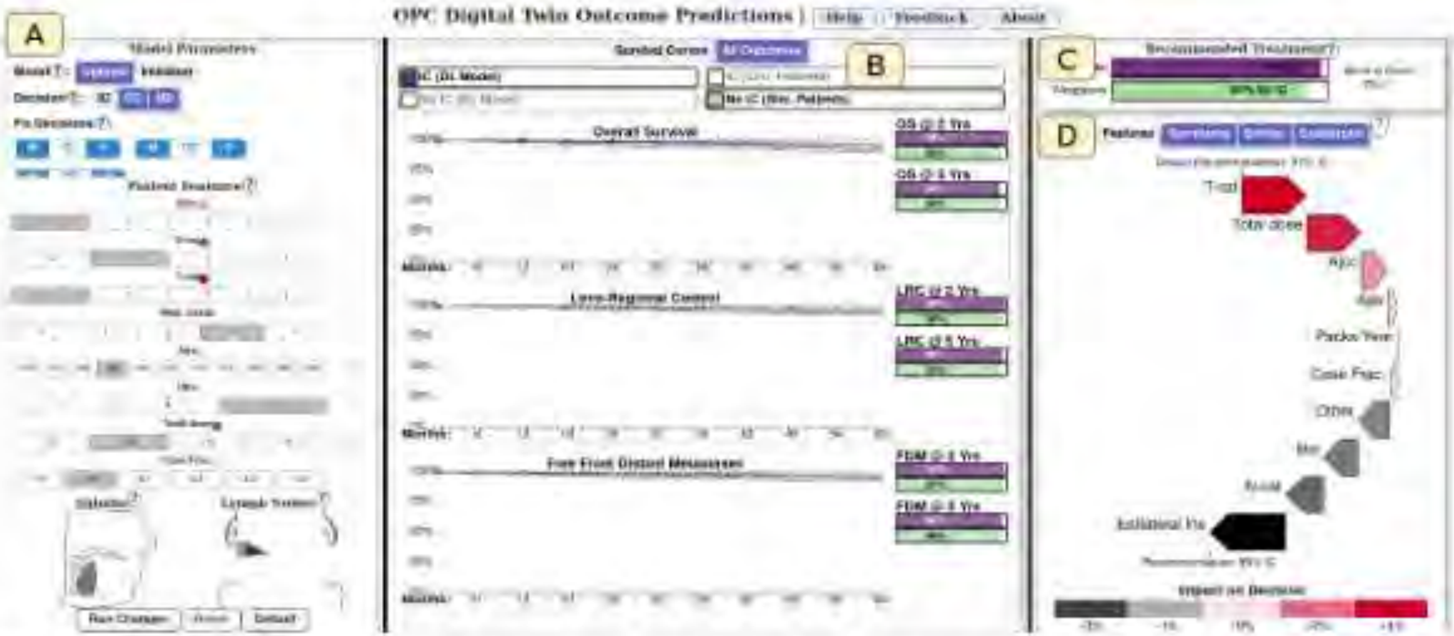


Fig. 1: Overview of DITTO. (A) Input panel to alter model parameters and input patient features. (B) Temporal outcome risk plots for the patient based on different models and treatment groups. (C) Treatment recommendation based on the twin model and similar patients. (D) Auxiliary data panel, currently showing a waterfall plot of how each feature cumulatively contributes to the model decision.

Username

andrewwentzel

Password

•••••••••••••••

Log in

But the view looks good for computational models in #RadOnc

Please email/visit soon!

cdfuller@mdanderson.org