Towards Trustworthy AI for Clinical Oncology

Issam El Naqa, PhD, DABR, FAAPM, FIEEE, FAIMBE

Chair of Machine Learning Department, Moffitt

Senior member of Radiation Oncology & GI Departments, Moffitt

Physics and Medical Engineering, University of South Florida (adjunct)

issam.elnaqa@moffitt.org

XAI^3 Workshop at ECAI 2023

October 1st, 2023, Krakow, Poland

Acknowledgements

REVIEW ARTICLE

AI and machine learning ethics, law, diversity, and global impact

Katherine Drabiak, Skylar Kyzer, Valerie Nemov and Issam El Naqa

Published Online: 23 May 2023 · https://doi.org/10.1259/bjr.20220934

International Journal of Radiation Oncology biology • physics
www.redjournal.org

23 March 2021

Radiomic and radiogenomic modeling for radiotherapy: strategies, pitfalls, and challenges

James T. T. Coates, Giacomo Pirovano, Issam El Naqa

Author Affiliations +

J. of Medical Imaging, 8(3), 031902 (2021). https://doi.org/10.1117/1.JMI.8.3.031902

Radiation Therapy Outcomes Models in the Era of Radiomics and Radiogenomics: Uncertainties and Validation

Issam El Naqa, PhD,* Gaurav Pandey, PhD,† Hugo Aerts, PhD,‡,§ Jen-Tzung Chien, PhD,||,¶ Christian Nicolaj Andreassen, MD, PhD,# Andrzej Niemierko, PhD,** and Randall K. Ten Haken, PhD*

Oncogene www.nature.com/onc

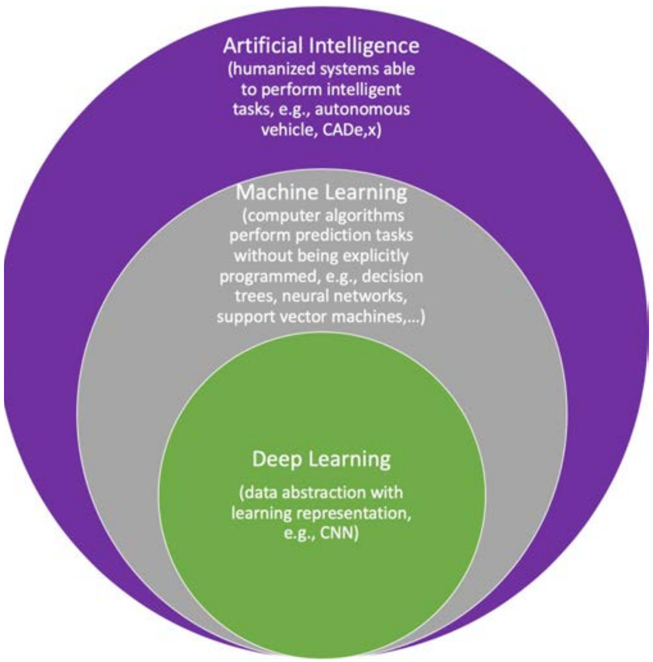
REVIEW ARTICLE

Check for updates

Translation of AI into oncology clinical practice

Issam El Naqa[1], Aleksandra Karolak[1], Yi Luo[1], Les Folio[2], Ahmad A. Tarhini[3], Dana Rollison[4] and Katia Parodi[5]

What is AI/ML/DL?



El Naqa, BJR 125th Annv., 2020

Artificial Intelligence

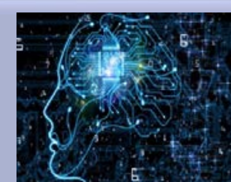
Originated in the 1950s

Build machines that think like humans

Machine Learning

Originated in the 1960s

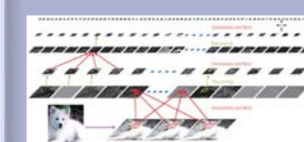
Computer algorithms that learn from data



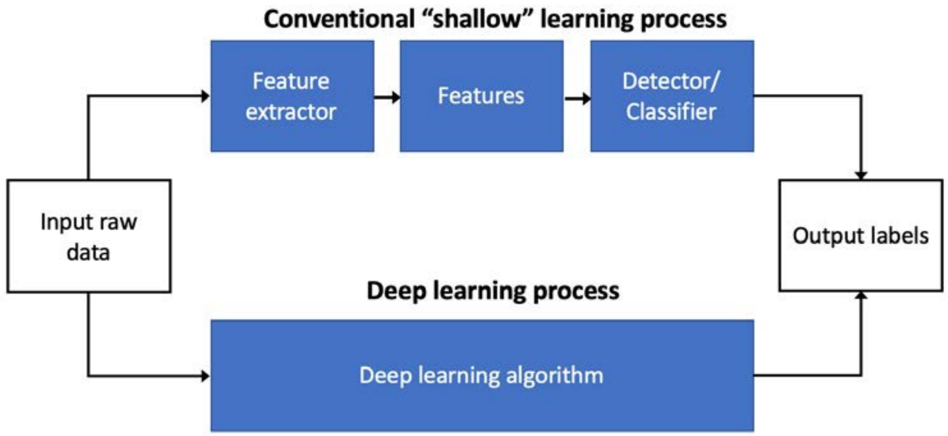
Deep Learning

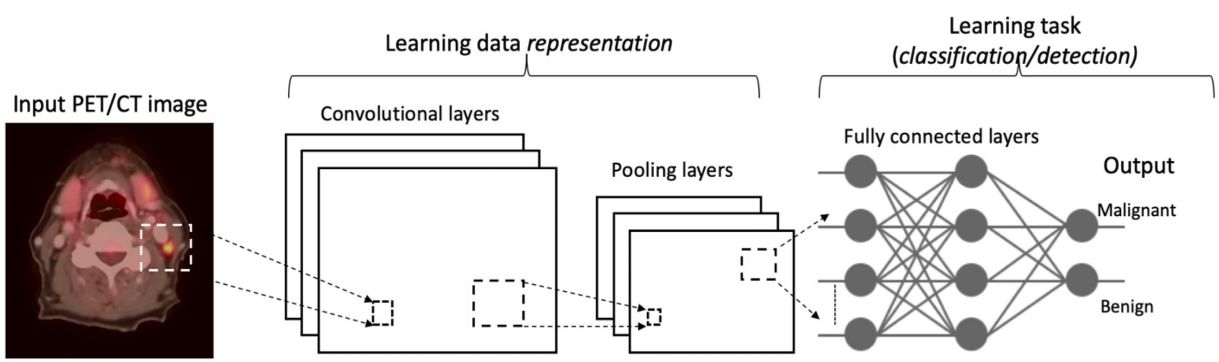
Originated in the 1970s

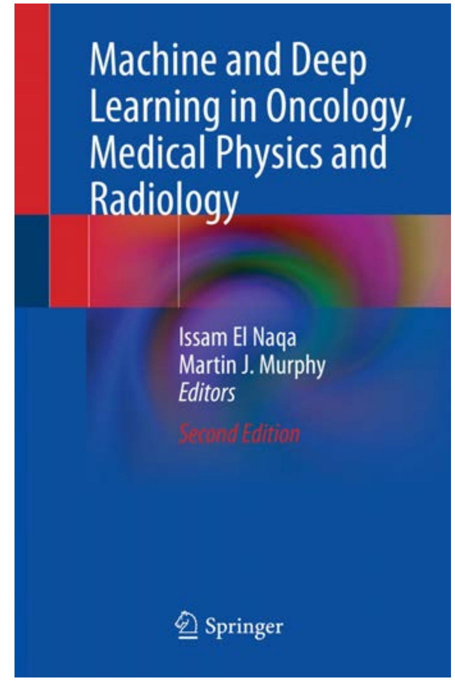
Based on neural networks that learn features



Deep vs conventional machine learning







Zaidi and El Naqa, Annu. Rev. Biomed. Eng., 2021

National and Global AI/ML interest

National AI Initiative Act of 2020 (NAIIA)

Became law on January 1, 2021

As part of the "*William M. (Mac) Thornberry National Defense Authorization Act for Fiscal Year 2021*", *H.R. 6395, Division E.*

DIVISION E—NATIONAL ARTIFICIAL INTELLIGENCE INITIATIVE ACT OF 2020

SEC. 5001. SHORT TITLE.

This division may be cited as the "National Artificial Intelligence Initiative Act of 2020".

Microsoft
Moffitt Cancer Center
NASA
National Center for Atmospheric Research
National Center for Supercomputing Applications at the University of Illinois at Urbana-Champaign
National Energy Technology Laboratory

EUROPEAN COMMISSION

Brussels, 21.4.2021

COM(2021) 206 final

2021/0106(COD)

Proposal for a

REGULATION OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

LAYING DOWN HARMONISED RULES ON ARTIFICIAL INTELLIGENCE (ARTIFICIAL INTELLIGENCE ACT) AND AMENDING CERTAIN UNION LEGISLATIVE ACTS

AI/ML-Enabled Devices By Primary Medical Specialty

- Radiology
- Cardiovascular
- Hematology
- Neurology
- Ophthalmic
- Clinical Chemistry
- General And Plastic Surgery
- Microbiology
- Gastroenterology-Urology
- Anesthesiology
- General Hospital
- Obstetrics And Gynecology
- Pathology
- Dental
- Orthopedic

Primary Medical Specialty	*Number of Devices in Primary Medical Specialty*	*Number of Devices with Oncology Applications*
Radiology	241	157
Cardiovascular	41	0
Hematology	13	10
Neurology	12	1
Ophthalmic	6	0
Clinical Chemistry	5	0
General And Plastic Surgery	5	3
Microbiology	5	0
Gastroenterology-Urology	4	3
Anesthesiology	4	0
General Hospital	3	0
Obstetrics And Gynecology	1	0
Pathology	1	1
Dental	1	0
Orthopedic	1	0
	Total: 343	Total: 175

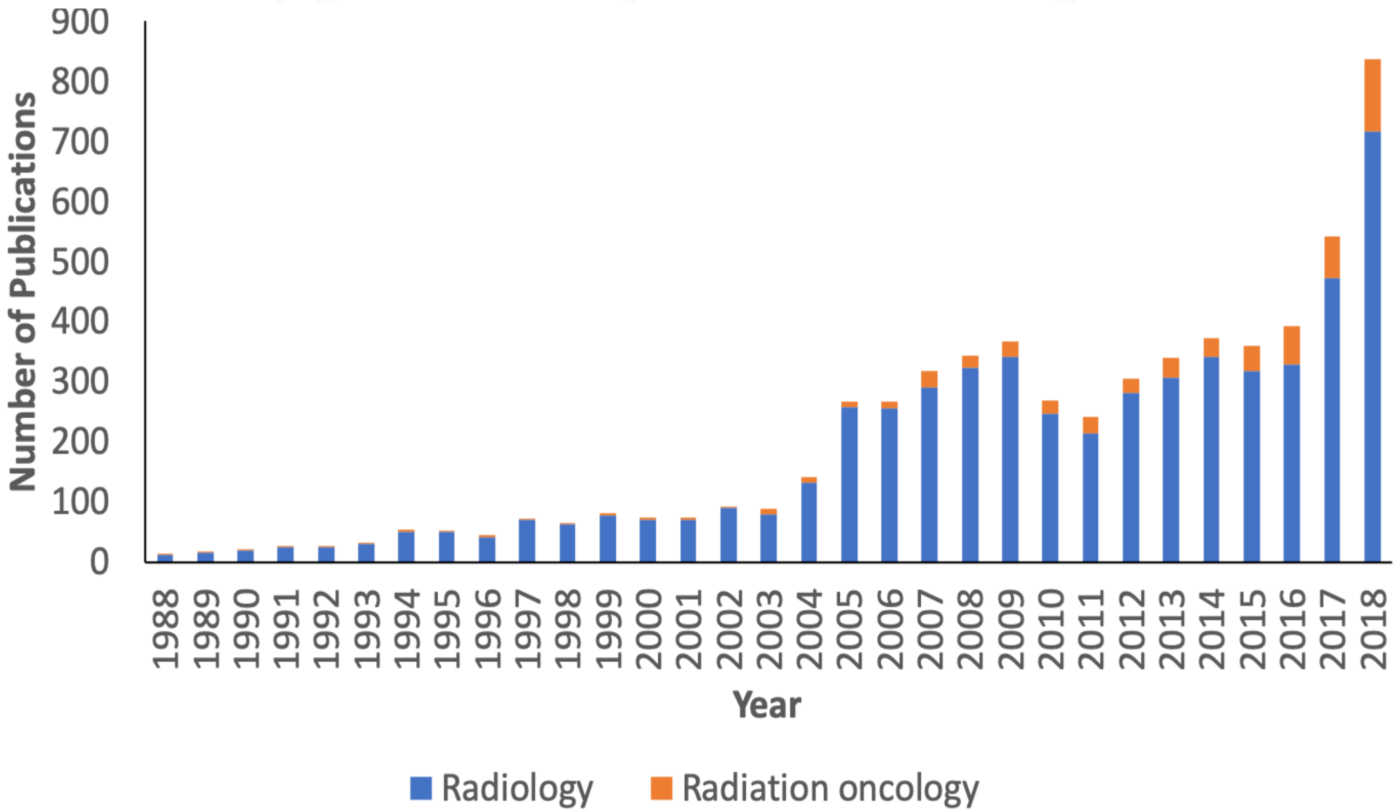
https://www.ai.gov/wp-content/uploads/2023/01/NAIRR-TF-Final-Report-2023.pdf

Drabiak K., BJR, 2023

BJR 125TH ANNIVERSARY SPECIAL FEATURE: REVIEW ARTICLE

Artificial Intelligence: reshaping the practice of radiological sciences in the 21st century

[1]ISSAM EL NAQA, PhD, [2]MASOOM A HAIDER, MD, [3]MARYELLEN L GIGER, PhD and [1]RANDALL K TEN HAKEN, PhD



Why AI/ML for Oncology?

The NEW ENGLAND JOURNAL of MEDICINE

REVIEW ARTICLE

FRONTIERS IN MEDICINE

Machine Learning in Medicine

Alvin Rajkomar, M.D., Jeffrey Dean, Ph.D., and Isaac Kohane, M.D., Ph.D.

This framing emphasizes that machine learning is not just a new tool, such as a new drug or medical device. Rather, it is the fundamental technology required to meaningfully process data that exceed the capacity of the human brain to comprehend; increasingly, this overwhelming store of information pertains to both vast clinical databases and even the data generated regarding a single patient.[7]

Nearly 50 years ago, a Special Article in the *Journal* stated that computing would be "augmenting and, in some cases, largely replacing the intellectual functions of the physician."[8] Yet, in early 2019, surprisingly little in health care is driven by machine learning. Rather than report the myriad proof-of-concept models (of retrospective data) that have been tested, here we describe the core structural changes and paradigm shifts in the health care system that are necessary to enable the full promise of machine learning in medicine (see video).

Artificial intelligence in cancer research, diagnosis and therapy

Olivier Elemento, Christina Leslie, Johan Lundin & Georgia Tourassi

Nature Reviews Cancer 21, 747–752 (2021) | Cite this article

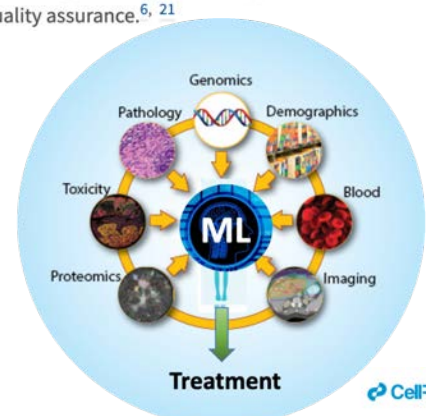
Artificial intelligence and machine learning techniques are breaking into biomedical research and health care, which importantly includes cancer research and oncology, where the potential applications are vast. These include detection and diagnosis of cancer, subtype classification, optimization of cancer treatment and identification of new therapeutic targets in drug discovery. While big data used to train machine learning models may already exist, leveraging this opportunity to realize the full promise of artificial intelligence in both the cancer research space and the clinical space will first require significant obstacles to be surmounted. In this Viewpoint article, we asked four experts for their opinions on how we can begin to implement artificial intelligence while ensuring standards are maintained so as transform cancer diagnosis and the prognosis and treatment of patients with cancer and to drive biological discovery.

The *Lancet* Commission on cancer and health systems: harnessing synergies to achieve solutions

Felicia Marie Knaul • Patricia J Garcia • Mary Gospodarowicz • Beverley M Essue • Naomi Lee • Richard Horton

Published: August 19, 2021 • DOI: https://doi.org/10.1016/S0140-6736(21)01895-X • Check for updates

The data science revolution makes it affordable to develop, digitalise, synthesise, analyse, store, and share vast quantities of information that anchor machine learning. Additionally, artificial intelligence could improve health-care quality and efficiency in all resource settings, alleviating workforce and equipment shortages, and facilitating clinical decision support tools and remote technic[al] and quality assurance.[6, 21]



Cell

Leading Edge

Commentary

Precision medicine in 2030—seven ways to transform healthcare

Joshua C. Denny[1,3,*] and Francis S. Collins[2]
[1]*All of Us* Research Program, National Institutes of Health, Bethesda, MD, USA
[2]National Institutes of Health, Bethesda, MD, USA
[3]Present address: Bldg. 1 Room 228, 1 Center Drive, Bethesda, MD 20814, USA
*Correspondence: joshua.denny@nih.gov
https://doi.org/10.1016/j.cell.2021.01.015

Precision medicine promises improved health by accounting for individual variability in genes, environment, and lifestyle. Precision medicine will continue to transform healthcare in the coming decade as it expands in key areas: huge cohorts, artificial intelligence (AI), routine clinical genomics, phenomics and environment, and returning value across diverse populations.

Progress in the Application of Machine Learning Algorithms to Cancer Research and Care

Neal J. Meropol, MD[1]; Janet Donegan, BSN, MA[1]; Alexander S. Rich, PhD[1]

» Author Affiliations | Article Information

JAMA Netw Open. 2021;4(7):e2116063. doi:10.1001/jamanetworkopen.2021.16063

The application of artificial intelligence in medical care has lagged behind its use in finance, advertising, and other consumer industries. This contrast is associated, in part, with the high stakes involved in developing tools that will ultimately affect patients. Given the expanding evidence gaps in oncology and the growing complexity of medical decisions, the imperative to apply available technologies has never been greater. In this context, careful consideration must be given to model development and scientific validation.[5,6] Large-scale appropriate training data and rigorous downstream validation, with transparency to permit reproducibility, may provide researchers the ability to use machine-based variables in appropriate clinical settings. In addition, explainability of model features may also be required if broad adoption by nontechnical clinical users is expected. The true promise of machine-based approaches is in enabling a learning health care system in which patient data are used for research and clinical applications and evolving care patterns and outcomes measurements are incorporated in a continuous feedback loop.[7] Success demands a broad recognition of the importance of high-quality data collection, data standards, and the benefits of data sharing for patients and public health.

BJR 125TH ANNIVERSARY SPECIAL FEATURE: REVIEW ARTICLE

Artificial Intelligence: reshaping the practice of radiological sciences in the 21st century

[1]ISSAM EL NAQA, PhD, [2]MASOOM A HAIDER, MD, [3]MARYELLEN L GIGER, PhD and [1]RANDALL K TEN HAKEN, PhD

Perspective | Published: 17 May 2018

OPINION

Artificial intelligence in radiology

Ahmed Hosny, Chintan Parmar, John Quackenbush, Lawrence H. Schwartz & Hugo J. W. L. Aerts

Nature Reviews Cancer 18, 500–510 (2018) | Cite this article

Non-invasive decision support for NSCLC treatment using PET/CT radiomics

Wei Mu, Lei Jiang, JianYuan Zhang, Yu Shi, Jhanelle E. Gray, Ilke Tunali, Chao Gao, Yingying Sun, Jie Tian, Xinming Zhao, Xilin Sun, Robert J. Gillies & Matthew B. Schabath

Nature Communications 11, Article number: 5228 (2020) | Cite this article

Personalized vaccines for cancer immunotherapy

UGUR SAHIN AND ÖZLEM TÜRECI

SCIENCE • 23 Mar 2018 • Vol 359, Issue 6382 • pp. 1355-1360 • DOI: 10.1126/science.aar7112

COVER STORY

GUEST EDITORIAL

Moffitt Cancer Center: Why we are building the first machine learning department in oncology

By Issam El Naqa and Dana Rollison

VISION

To transform personalized cancer care and accelerate scientific discovery in cancer research with machine/deep learning

MISSION

To design, develop, and translate state-of-the-art patient-centered machine and deep learning algorithms

VALUE

Patient-centered ML/DL for facilitating cancer care and research

VALUE

Unbiased, generalizable, and *interpretable* ML/DL from blended data

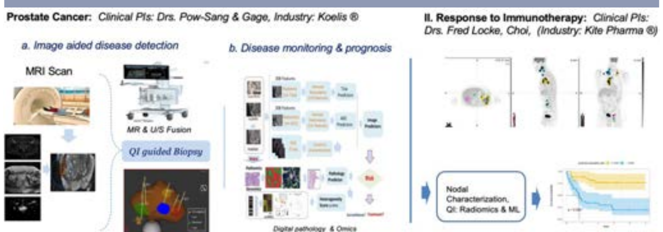
VALUE

Translate ML/DL findings into the clinic to improve cancer care and research

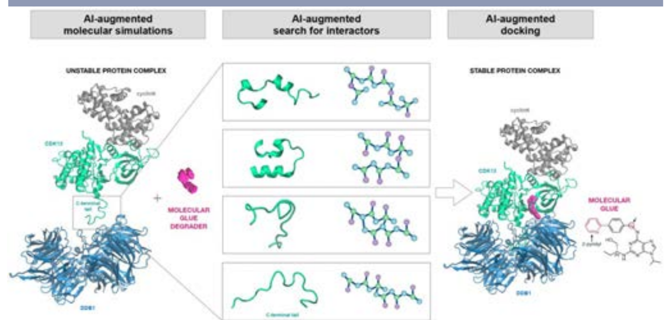
Moffitt.org/MachineLearning

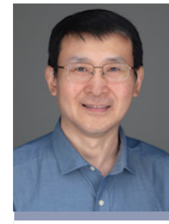
(@ml4onco)

Yoga Balagurunathan, PhD
Quantitative Imaging & AI Lab
Research Focus: Disease detection, monitoring & prognosis

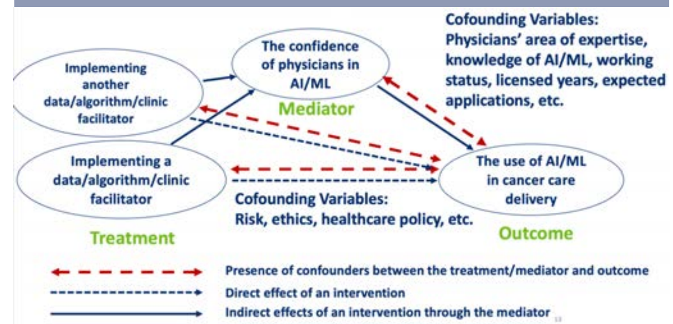


Aleks Karolak, PhD
Molecular AI Lab
Research Focus: Molecular interactions, drug discovery

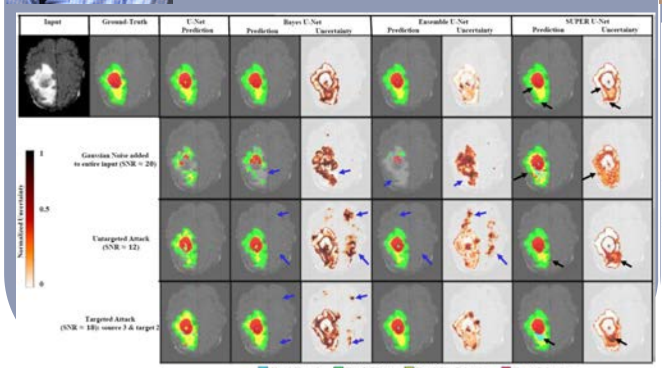




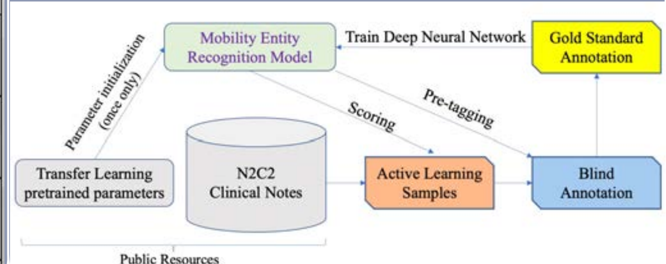
Yi Luo, PhD
Fair AI Lab
Research Focus: Patient outcomes, public health, social determinant



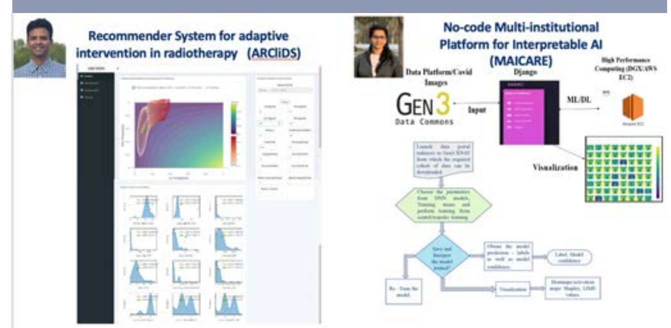
Ghulam Rasool, PhD
Robust Multimodal AI Lab
Research Focus: AI uncertainty, multimodal data modeling



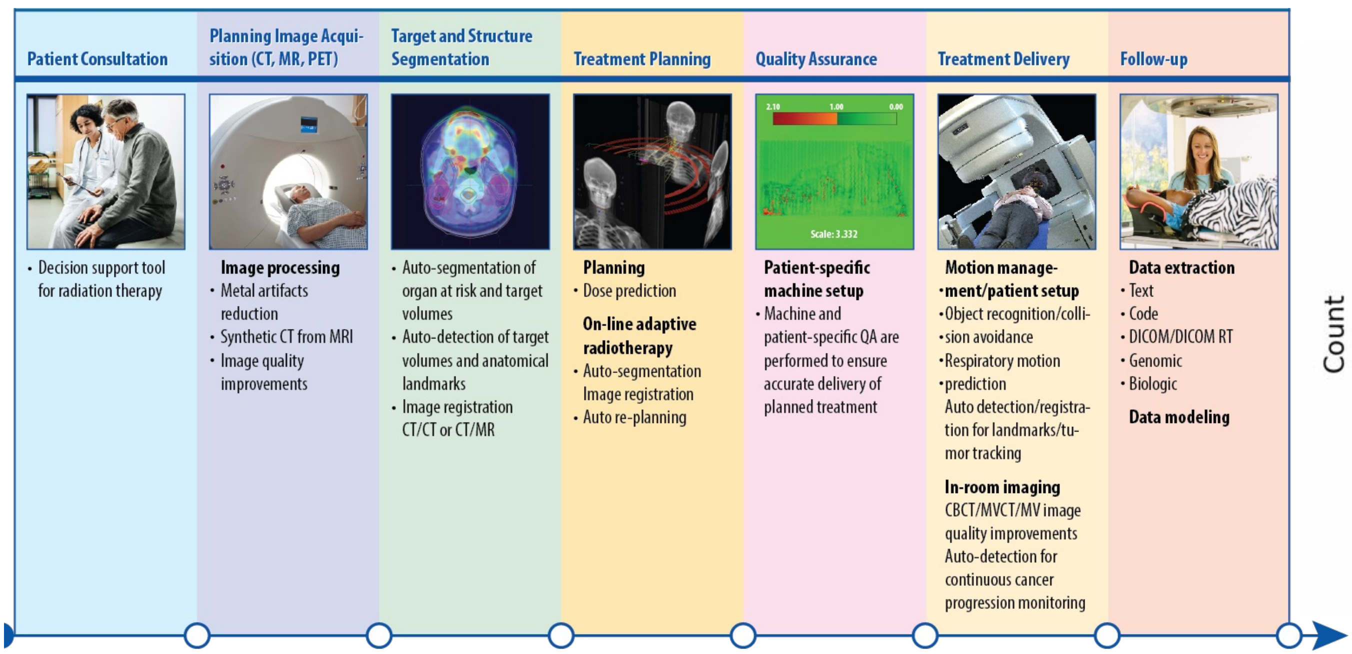
Thanh Thieu, PhD
Language And Intelligence Laboratory (LAILab)
Research Focus: NLP, language models, functional mobility

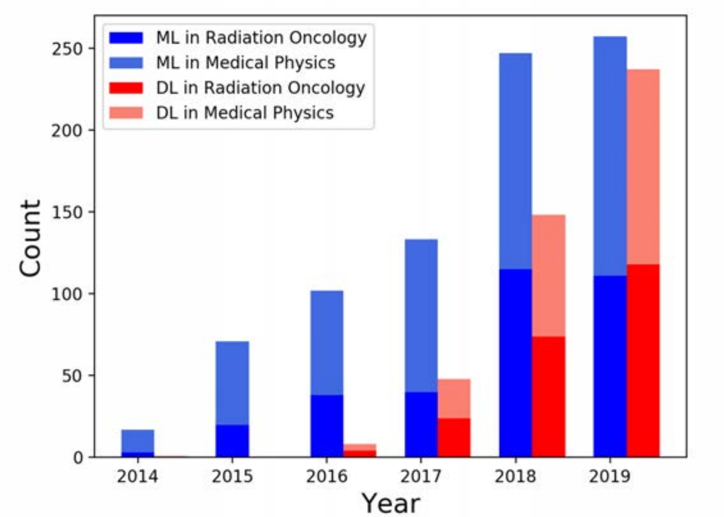


Issam El Naqa, PhD
Decision support & outcome modeling
Research Focus: image analytics, medical physics, human factors



Applications of ML/DL in Medical Physics and Radiation Oncology

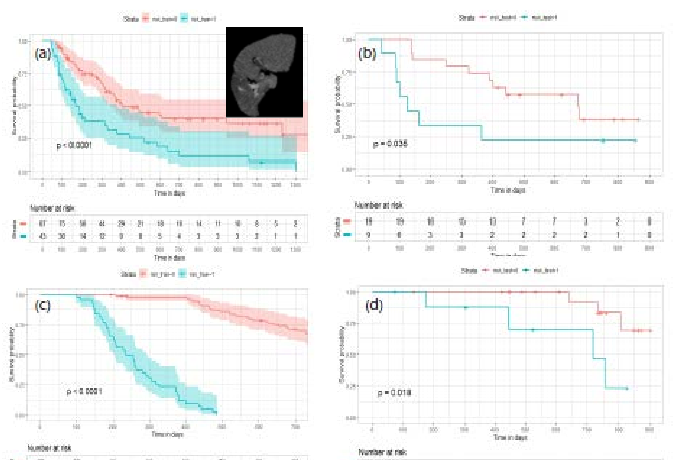




Cui, Med Phys, 2020

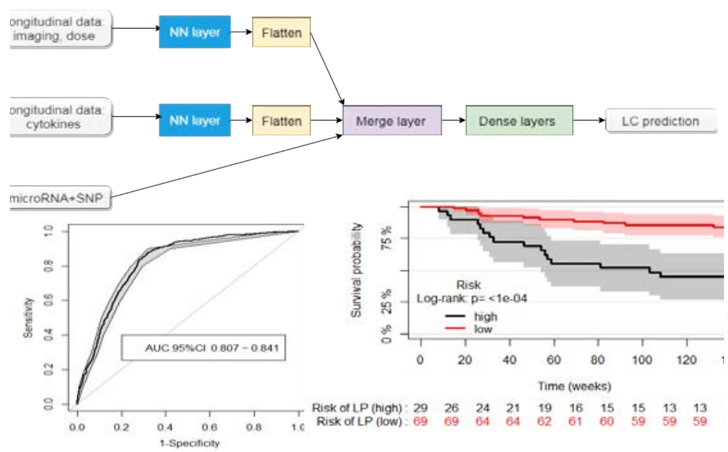
Sample Applications in Radiotherapy

Imaging biomarkers



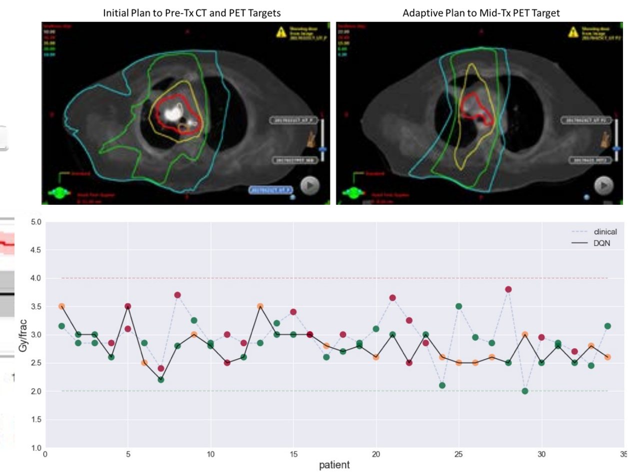
Wei, Physica Medica , 2021

Outcome modeling



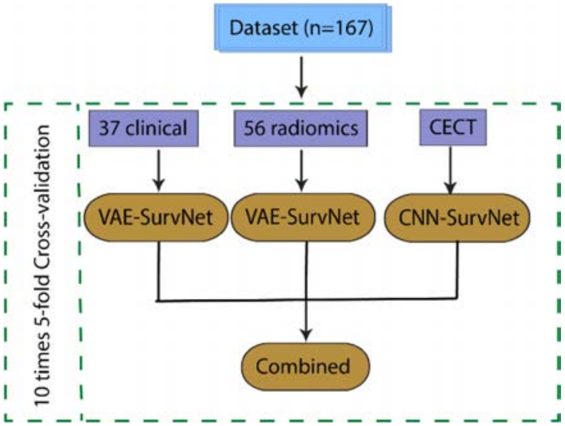
Cui, IEEE TRMPS, 2018
(Best of ASTRO)

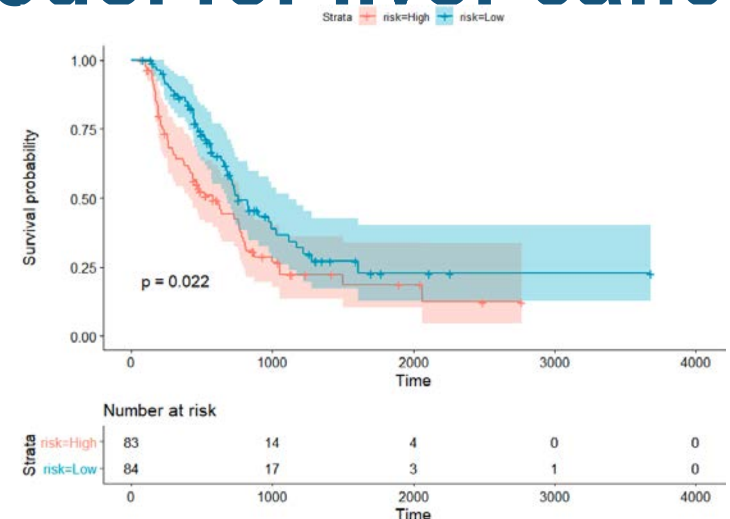
Adaptive RT

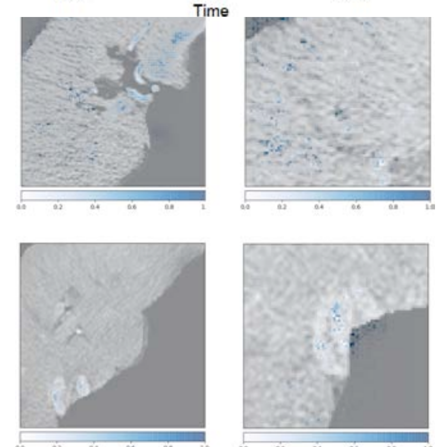


Tseng, Med Phys, 2017
(Best paper in Medical Physics)

Radiomics deep survival model for liver cancer

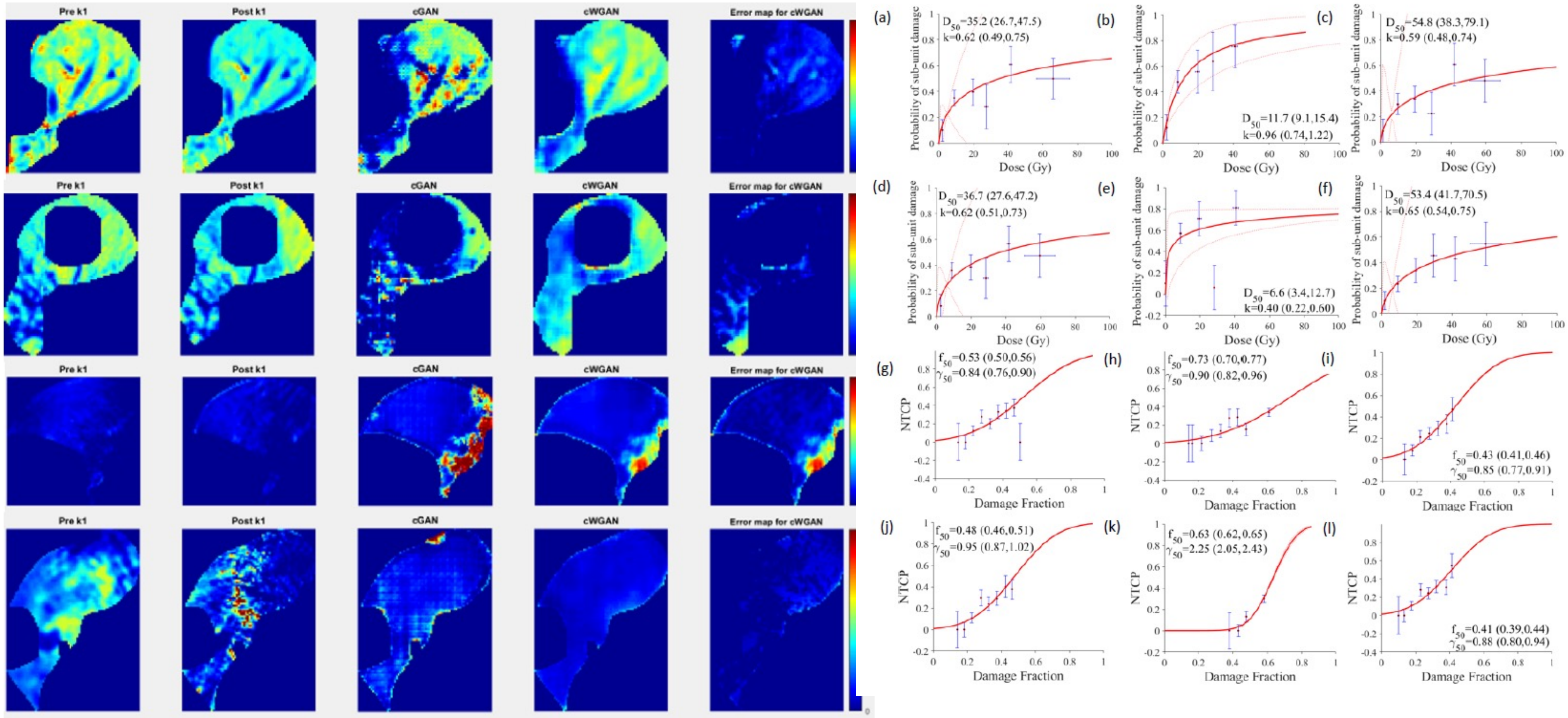






Wei et al, Physica Medica, 2021

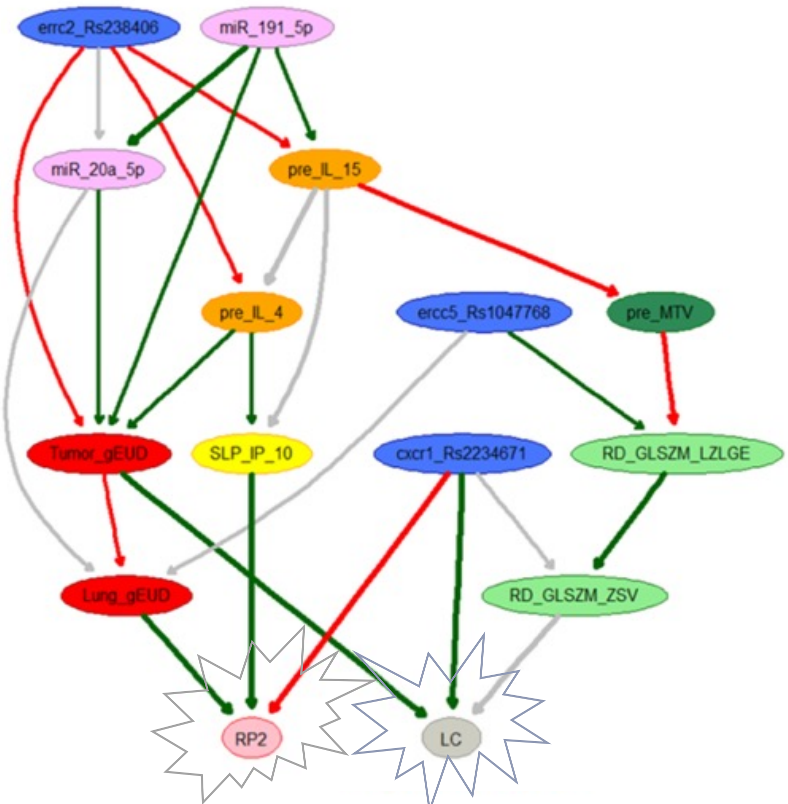
Deep Learning Prediction of post-SBRT Liver Function Changes and NTCP Modeling in HCC based on DGAE-MRI

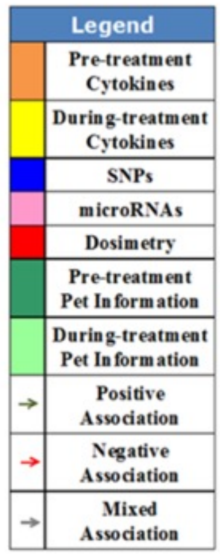


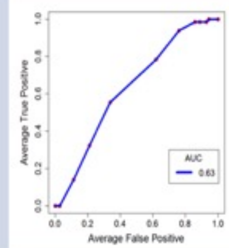
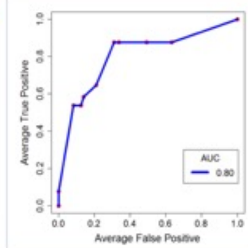
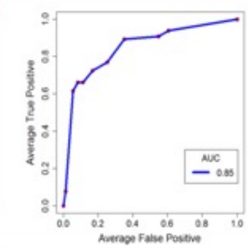
Wei et al, Med Phys, 2023

Multi-Objective radiogenomics model with generative ML

A multi-objective Bayesian networks can be used to predict multiple radiation outcomes simultaneously, which provides opportunities of finding appropriate treatment plans to solve the trade-off between competing risks.

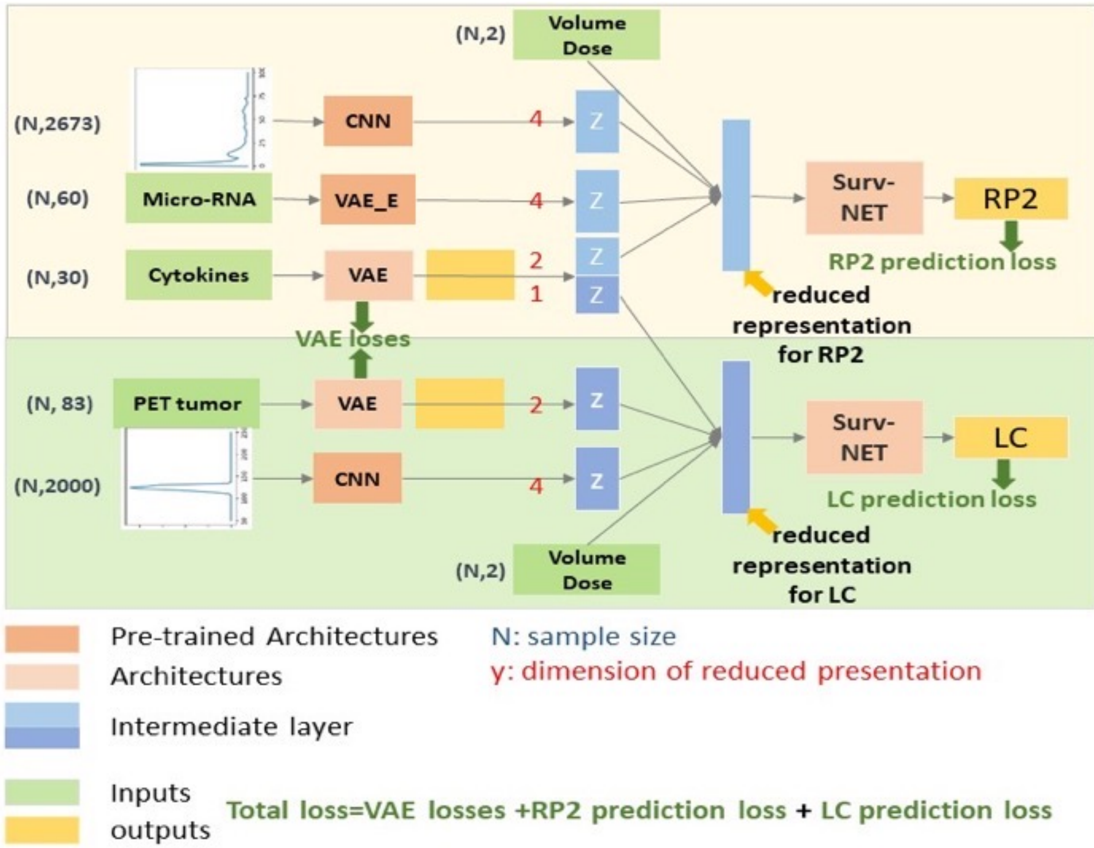




	Lung_gEUD and Tumor_gEUD	Pre-Treatment BN for joint prediction of LC and RP2	During-Treatment BN for joint prediction of LC and RP2
AU-FROC	0.63	0.8	0.85
95% CI	0.53-0.77	0.66-0.85	0.71-0.89
FROC			

Luo et al, Med Phys, 2018 (Editor's Choice)

Multi-objective response model with deep survival neural networks



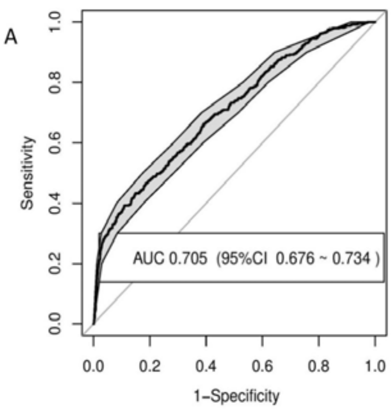
20 times of 5-fold cross validations

C-index (95%CI)	RP2	LC
NN-com	*0.705 (0.676~0.734)*	*0.740 (0.715 ~0.765)*
NN-DVH	0.660 (0.630~0.690)	0.727 (0.700~0.753)
Lyman/log-logistic	0.613 (0.583~0.643)	0.569 (0.545~0.594)

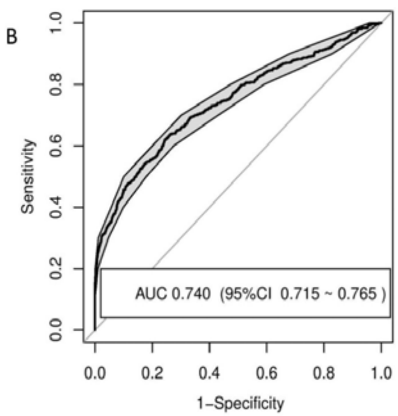
Independent test on 25 newly treated patients

C-index (95%CI)	RP2	LC
NN-composite	0.692	0.721
NN-DVH	0.684	0.706
Lyman/log-logistic	0.588	0.573

A



B

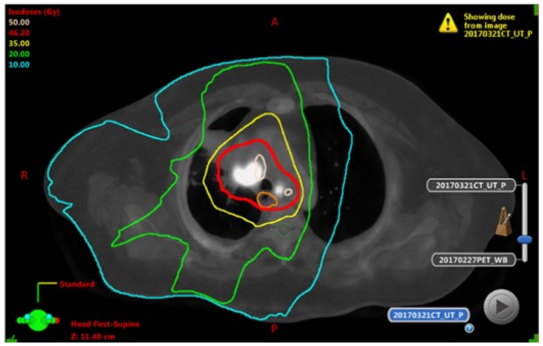


Cui et al, IJROBP, 2021

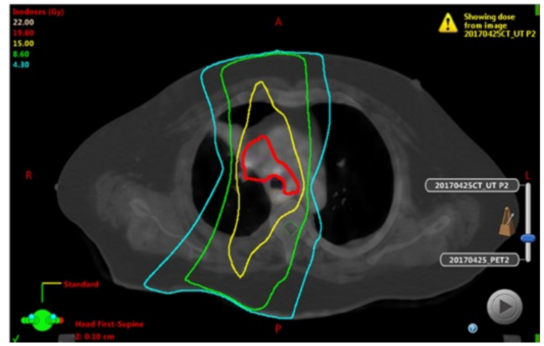
Adaptive Radiation Oncology Decision Making with Deep Learning

- Decision support tool for radiation therapy

Initial Plan to Pre-Tx CT and PET Targets

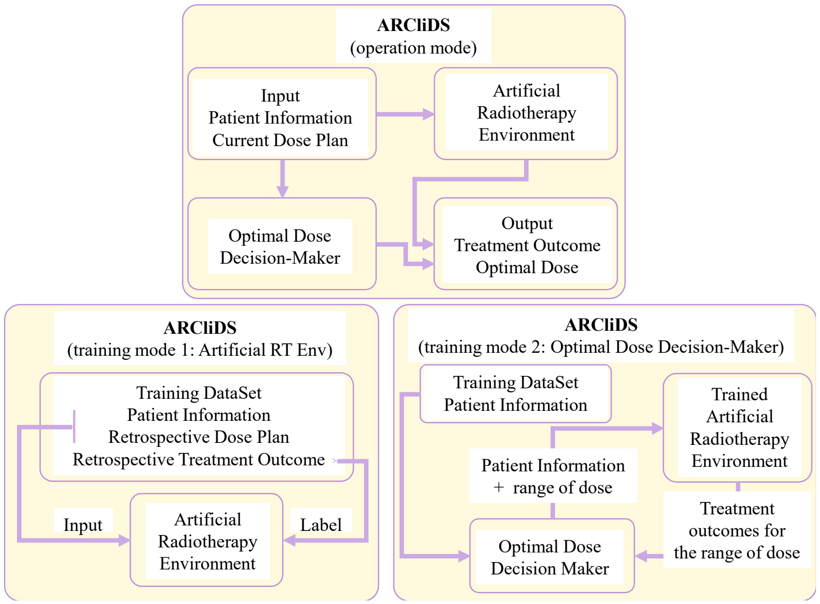


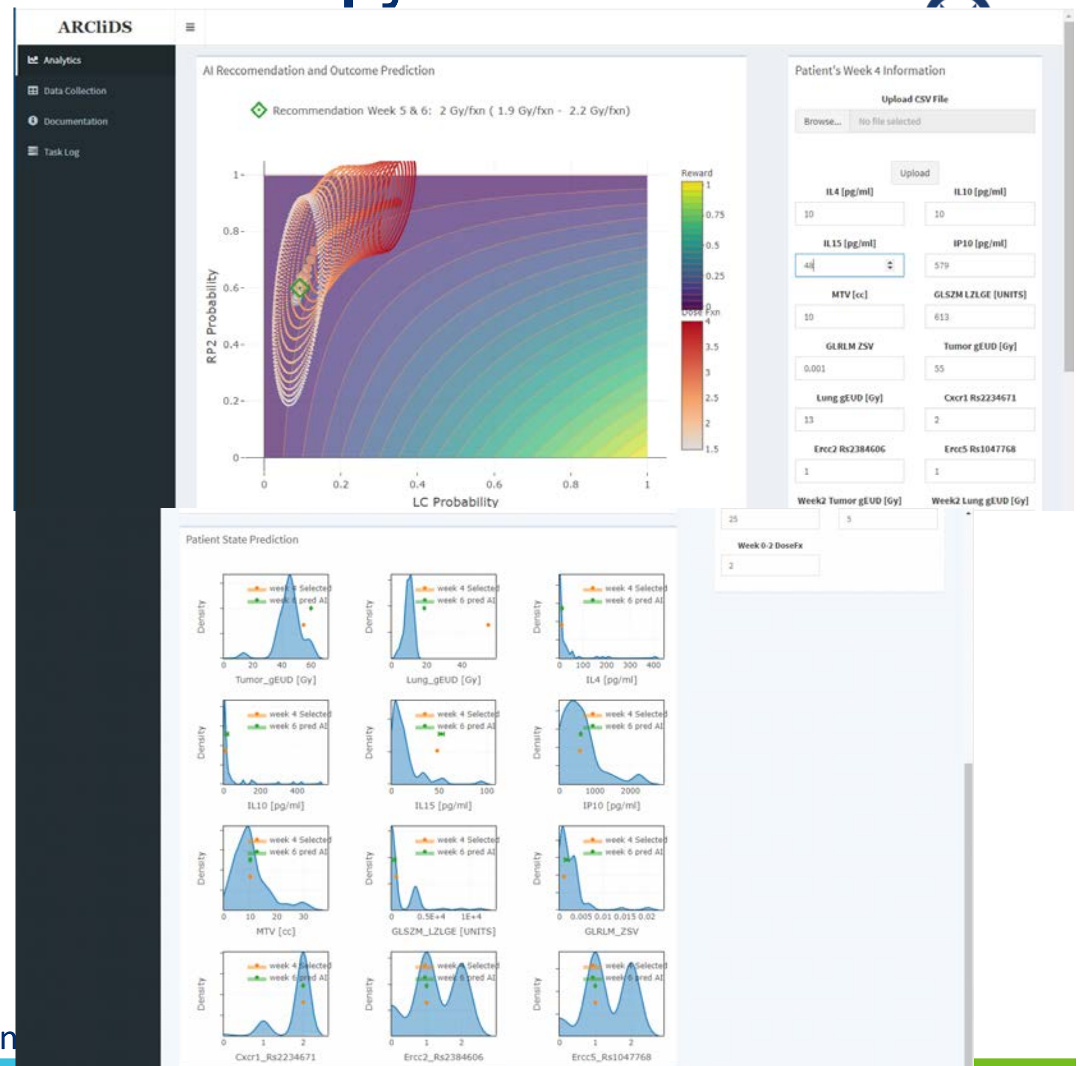
Adaptive Plan to Mid-Tx PET Target



Tseng, Medical Physics, 2017 (Farrington Daniels Award)

Software tools for Adaptive Radiotherapy Clinical Decision Support (ARClIDS)





- ➢ User Factors in AI implementation

Niraula, Nature Sci Rep, 2023; Sun CMPB, 2022; patent pendin

AI/ML is nothing but perfect!

- Google Flu Trends (GFT) (Ginsberg, 2009)
  - GFT called out sick 2013 due to overestimation!
- Predicting pneumonia risk (Caruana, 2015)
  - Patients with pneumonia and asthma to be at a lower risk of death from pneumonia than patients with pneumonia but without asthma!
- Skin cancer risk prediction (Esteva, 2017)
  - Presence of a ruler as a sign of high risk would skew prediction
- Lung disease prediction from xray (Rajpurkar, 2017)
  - Presence of tube can indicate high risk
- Covid-19 infection of AI (Deshpande, 2020; Roberts, 2021, El Naqa, 2021)
  - Unreliable AI models for Covid-19 prediction

⇒Data quality and context matters

COMPUTING

Racial Bias Found in a Major Health Care Risk Algorithm

Black patients lose out on critical care when systems equate health needs with costs

By Starre Vartan on October 24, 2019

Man is to Computer Programmer as Woman is to Homemaker? Debiasing Word Embeddings

Amazon scraps secret AI recruiting tool that showed bias against women

Study finds gender and skin-type bias in commercial artificial-intelligence systems

Examination of facial-analysis software shows error rate of 0.8 percent for light-skinned men, 34.7 percent for dark-skinned women.

External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients

Andrew Wong, MD[1]; Erkin Otles, MEng[2,3]; John P. Donnelly, PhD[4]; et al

EPIC's Sepsis Model Is Not Ready for Prime Time

Aaron J. Calderon, MD, FACP, SFHM, reviewing Wong A et al. JAMA Intern Med 2021 Aug

Despite its widespread use, the proprietary electronic health record system missed sepsis 67% of the time.

Issues in ML application in Oncology

Data modeling

- Availability and sharing
- Ethics and compliance

Algorithmic modeling

- Models' validation
- Models' interpretability

MEDICAL PHYSICS
The International Journal of Medical Physics Research and Practice

Special Issue Paper | Free Access

Machine learning and modeling: Data, validation, communication challenges

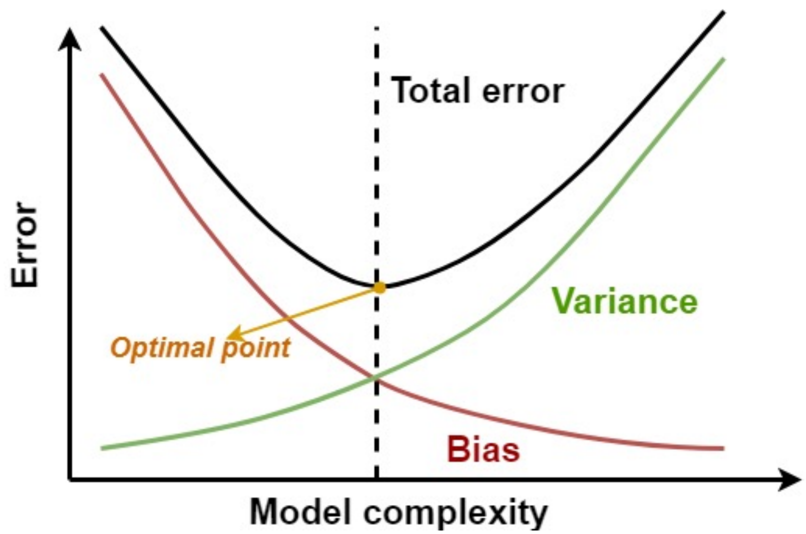
Issam El Naqa, Dan Ruan, Gilmer Valdes, Andre Dekker, Todd McNutt, Yaorong Ge, Q. Jackie Wu, Jung Hun Oh, Maria Thor, Wade Smith, Arvind Rao, Clifton Fuller, Ying Xiao, Frank Manion, Matthew Schipper, Charles Mayo, Jean M. Moran, Randall Ten Haken

First published: 24 August 2018 | https://doi.org/10.1002/mp.12811

What training data sample size is required?

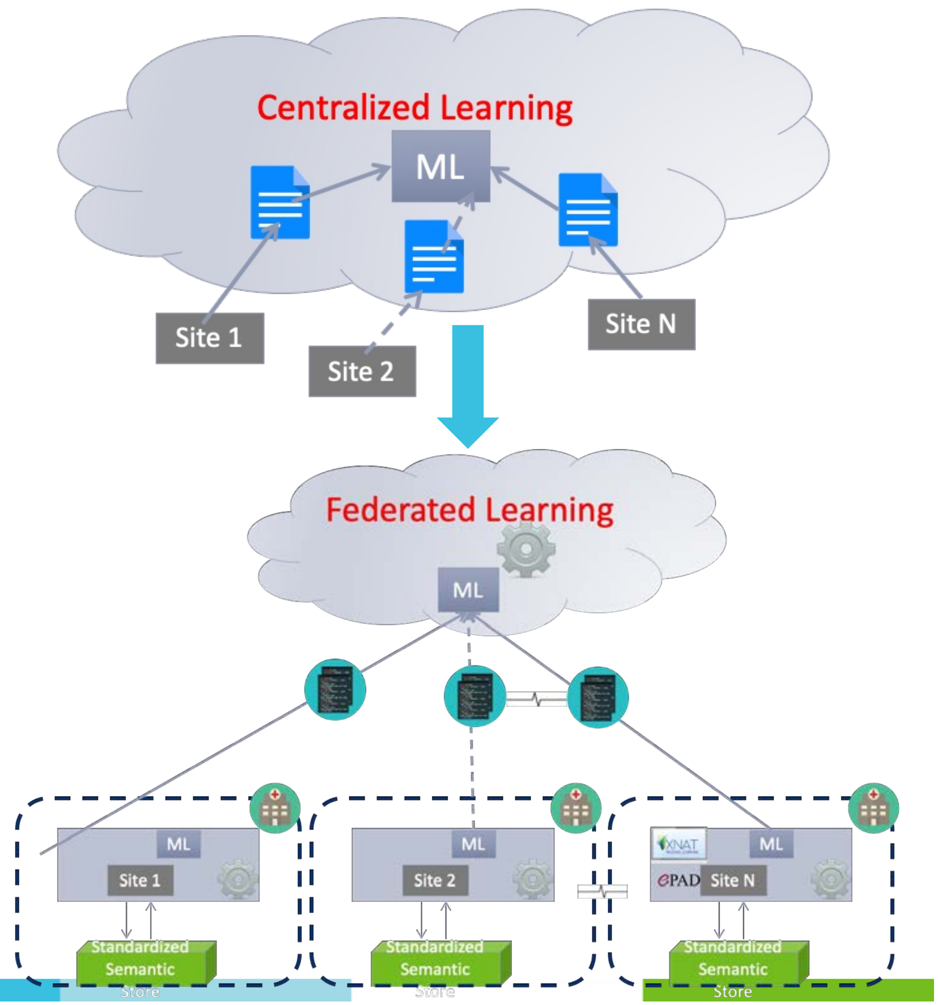
Introduction to Machine and Deep Learning for Medical Physicists

Sunan Cui,* Huan-Hsin Tseng,† Julia Pakela,‡ Randall K. Ten Haken,§ and Issam El Naqa¶
Department of Radiation Oncology, University of Michigan, Ann Arbor, MI 48103, USA



Cui, Medical Physics, 2020

Ethical Challenge of Data Access



Jochems, IJROBP, 2017

Data Democratization!

MIDRC

MEDICAL IMAGING AND DATA RESOURCE CENTER.

Lessons learned in transitioning to AI in the medical imaging of COVID-19

Issam M. El Naqa, Hui Li, Jordan D. Fuhrman, Qiyuan Hu, Naveena Gorre, Weijie Chen, Maryellen L. Giger

Author Affiliations +

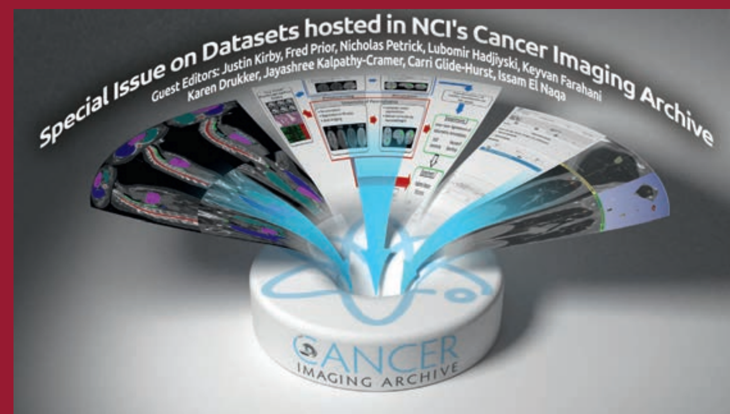
J. of Medical Imaging, 8(S1), 010902 (2021). https://doi.org/10.1117/1.JMI.8.S1.010902

December 2020 | Volume 47, Iss

Check for updates

MEDICAL PHYSICS

The International Journal of Medical Physics Research and Practice



Collage of illustrations from papers from the Special Issue on Datasets hosted in The Cancer Imaging Archive (TCIA).
Thanks to Jeff Tobler, University of Arkansas, for creating this collage.

Medical Physics is an official journal of the AAPM,
the International Organization for Medical Physics (IOMP), and
the Canadian Organization of Medical Physicists (COMP).

AMERICAN ASSOCIATION *of* PHYSICISTS IN MEDICINE

WILEY

Machine and Federated Learning Infrastructure (API)

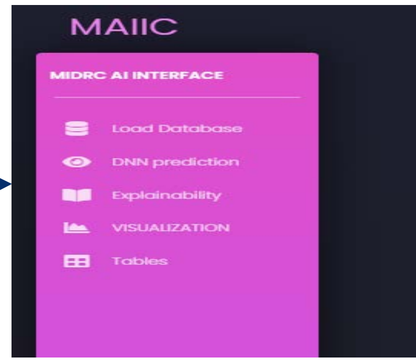
High Performance Computing (DGX/AWS EC2)

Django

Data Platform/Covid Images

GEN3 Data Commons

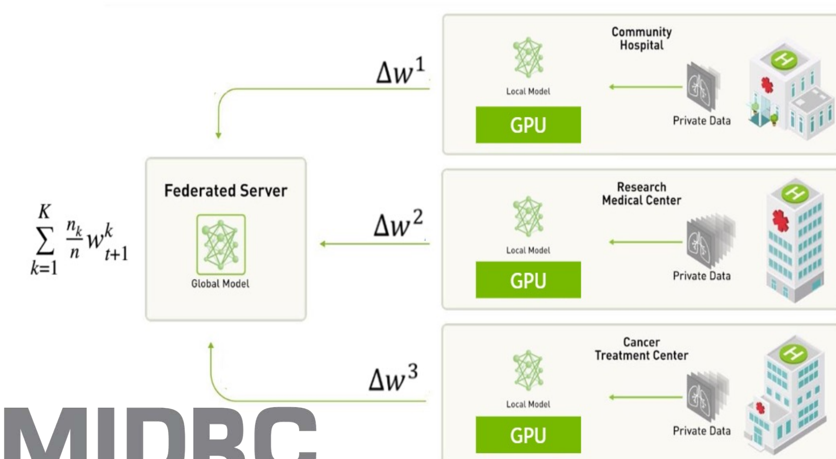
Input

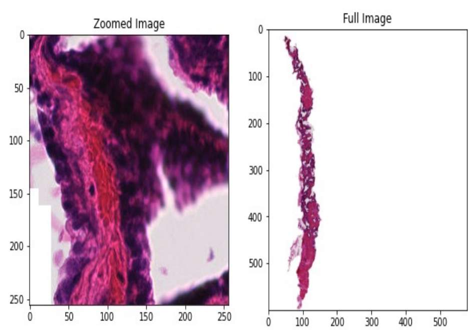


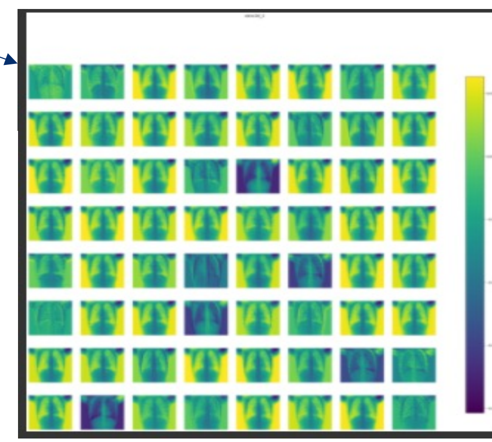
ML/DL

aws

Amazon EC2



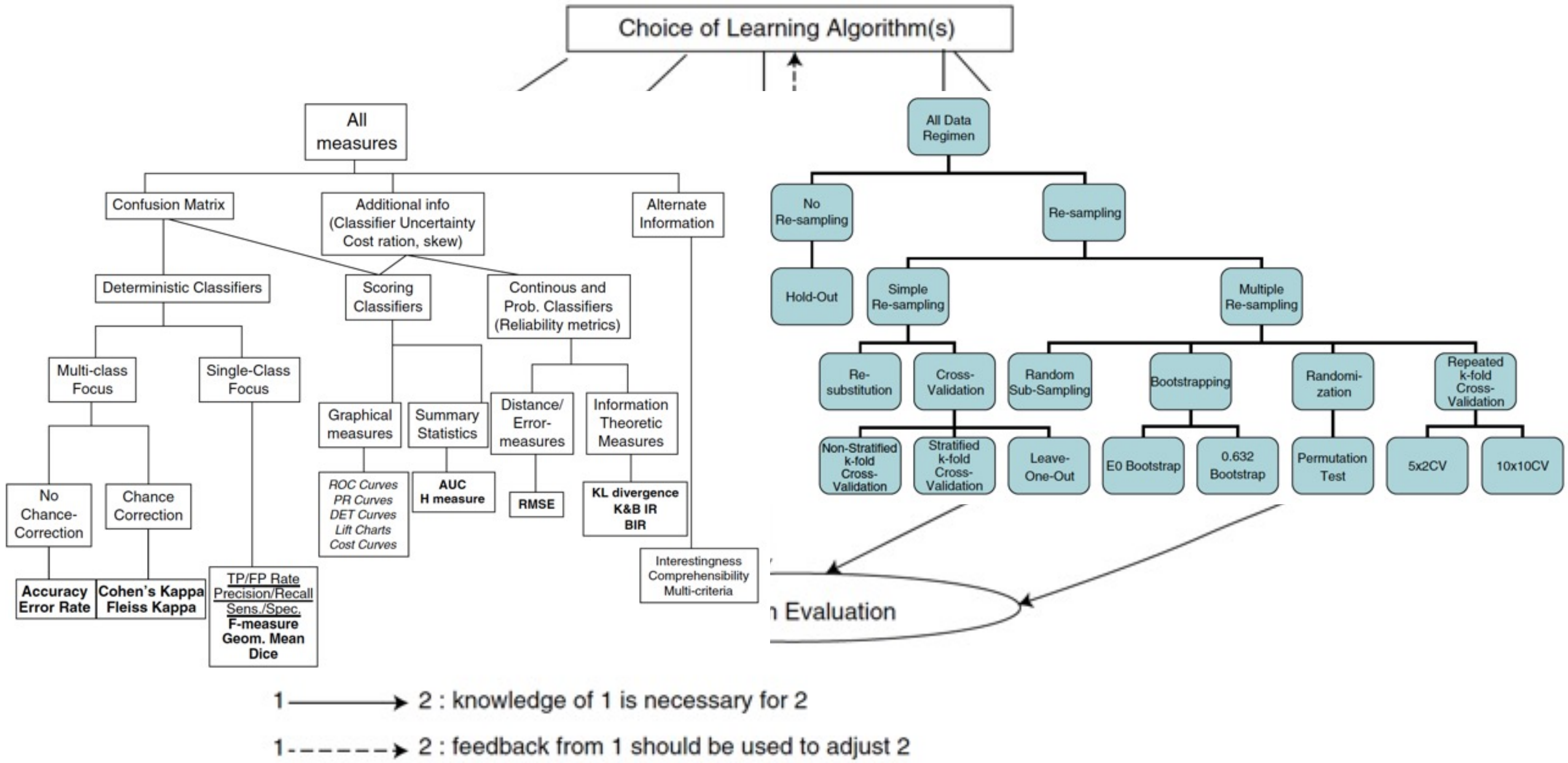




MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER.

Gorre et al, PMB, 2023

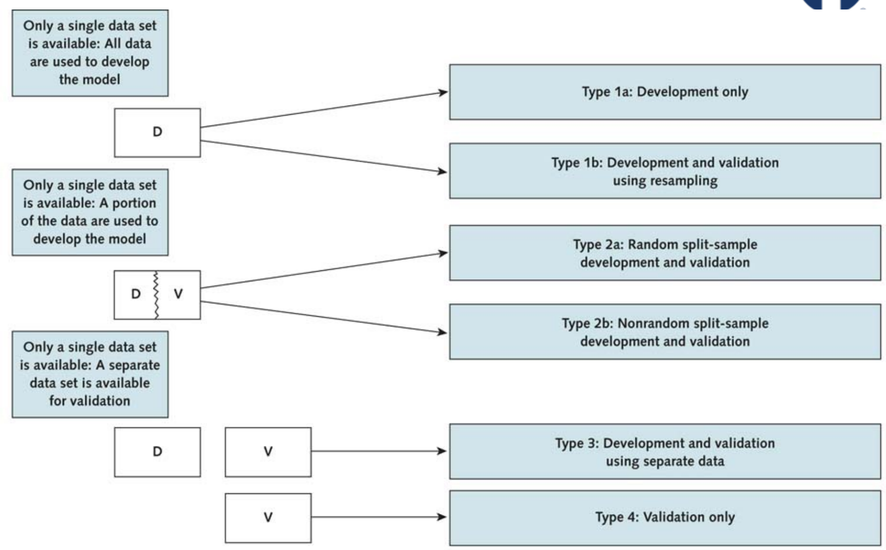
What evaluation plan for AI/ML?



Japkowicz and Shah, 2015

How to validate an ML/DL model?

Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD)



Analysis Type	Description
Type 1a	Development of a prediction model where predictive performance is then directly evaluated using exactly the same data (apparent performance).
Type 1b	Development of a prediction model using the entire data set, but then using resampling (e.g., bootstrapping or cross-validation) techniques to evaluate the performance and optimism of the developed model. Resampling techniques, generally referred to as "internal validation", are recommended as a prerequisite for prediction model development, particularly if data are limited (6, 14, 15).
Type 2a	The data are randomly split into 2 groups: one to develop the prediction model, and one to evaluate its predictive performance. This design is generally not recommended or better than type 1b, particularly in case of limited data, because it leads to lack of power during model development and validation (14, 15, 16).
Type 2b	The data are nonrandomly split (e.g., by location or time) into 2 groups: one to develop the prediction model and one to evaluate its predictive performance. Type 2b is a stronger design for evaluating model performance than type 2a, because allows for nonrandom variation between the 2 data sets (6, 13, 17).
Type 3	Development of a prediction model using 1 data set and an evaluation of its performance on separate data (e.g., from a different study).
Type 4	The evaluation of the predictive performance of an existing (published) prediction model on separate data (13).

Types 3 and 4 are commonly referred to as "external validation studies." Arguably type 2b is as well, although it may be considered an intermediary between internal and external validation.

Depending on the level of evidence

- Selection appropriate learning algorithms
- Validation and evaluation (TRIPOD criteria)
  - Internally (cross-validation schemes)
  - Externally (independent datasets)
- Provide interpretation of machine learning prediction

Radiology: Artificial Intelligence

lical Im

ers

Minimum information about clinical artificial intelligence modeling: the MI-CLAIM checklist

Beau Norgeot, Giorgio Quer, Brett K. Beaulieu-Jones, Ali Torkamani, Raquel Dias, Milena Gianfrancesco, Rima Arnaout, Isaac S. Kohane, Suchi Saria, Eric Topol, Ziad Obermeyer, Bin Yu & Atul J. Butte

Nature Medicine **26**, 1320–1324(2020) | Cite this article

Che
Mec
(CL

- Purp
  algo
- Data
  size,
  tunir
- ML n
  - O
    a
  - P
    (i
- Signi
  - Ir
  - C

Novelty	Please briefly (150 words or less) describe the novelty and/or significance of your study.: N/A If there is anything you wish to tell the editor that is not covered in this submission questionnaire, please enter it here: N/A
Artificial Intelligence and Machine Learning	Is this article on the topic of artificial intelligence or machine learning?: Yes The number of training, validation, and test sets are described in the Abstract. The number of input data and output results, along with the type of data (e.g. MRI images, CT images, etc.) are mentioned in the Abstract.: No The stage of development is described in the manuscript Introduction.: Yes The data, its source, and data composition are described in detail in the Materials section.: No The details of the machine learning algorithm, including pre-processing and training method, are provided in the Methods section. All major results are accompanied by appropriate tests of statistical significance.: No The innovation, significance, and/or contributions to the field of medical physics are discussed in the Discussion section.: Yes
Author ORCID Status	0 of 1 ORCIDs available.
NIH Funding	No funding has been received from NIH
CrossCheck Manuscript	Never Processed / Send File

Manuscript Items

Issam El Naqa[1]
John M. Boone[2]
Stanley H. Benedict[3]
Mitchell M. Goodsitt[4]
Heang-Ping Chan[4]
Karen Drukker[5]
Lubomir Hadjiiski[4]
Dan Ruan[6]
Berkman Sahiner[7]

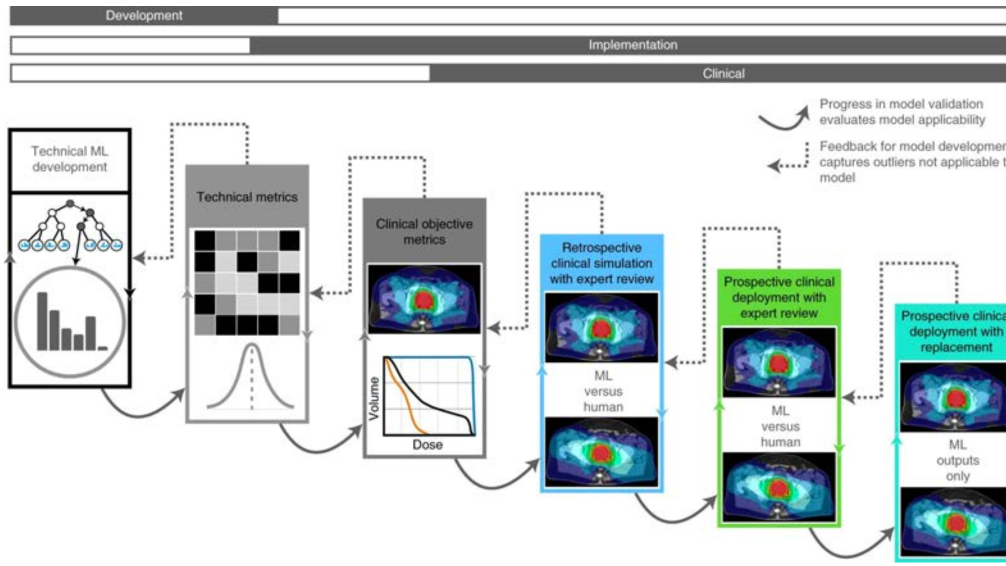
AI/ML in the real-world!

Letter | Published: 03 June 2021

Clinical integration of machine learning for curative-intent radiation treatment of patients with prostate cancer

Chris McIntosh, Leigh Conroy, Michael C. Tjong, Tim Craig, Andrew Bayley, Charles Catton, Mary Gospodarowicz, Joelle Helou, Naghmeh Isfahanian, Vickie Kong, Tony Lam, Srinivas Raman, Padraig Warde, Peter Chung, Alejandro Berlin & Thomas G. Purdie

Nature Medicine **27**, 999–1005 (2021) | Cite this article



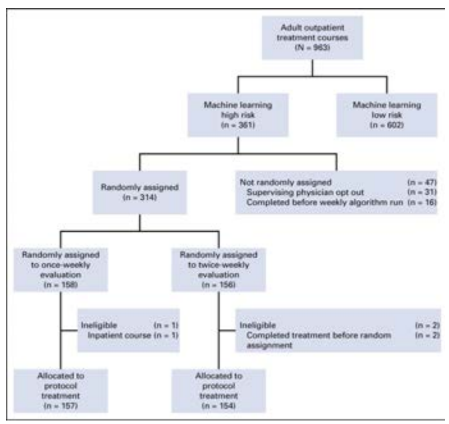
ORIGINAL REPORTS | Radiation Oncology

System for High-Intensity Evaluation During Radiation Therapy (SHIELD-RT): A Prospective Randomized Study of Machine Learning–Directed Clinical Evaluations During Radiation and Chemoradiation

Check for updates

Julian C. Hong, MD, MS[1,2,3]; Neville C. W. Eclov, PhD[3]; Nicole H. Dalal, MD[4]; Samantha M. Thomas, MS[5,6]; Sarah J. Stephens, MD[3]; Mary Malicki, MSN, ACNP[3]; Stacey Shields, ANP-BC[3]; Alyssa Cobb, RN, BSN[3]; Yvonne M. Mowery, MD, PhD[3,6]; Donna Niedzwiecki, PhD[5,6]; Jessica D. Tenenbaum, PhD[5]; and Manisha Palta, MD[3,6]

[1]Department of Radiation Oncology, University of California, San Francisco, San Francisco, CA
[2]Bakar Computational Health Sciences Institute, University of California, San Francisco, San Francisco, CA
[3]Department of Radiation Oncology, Duke University, Durham, NC
[4]Department of Medicine, University of California, San Francisco, San Francisco, CA
[5]Department of Biostatistics and Bioinformatics, Duke University, Durham, NC
[6]Duke Cancer Institute, Duke University, Durham, NC



News & Views | Published: 09 July 2021

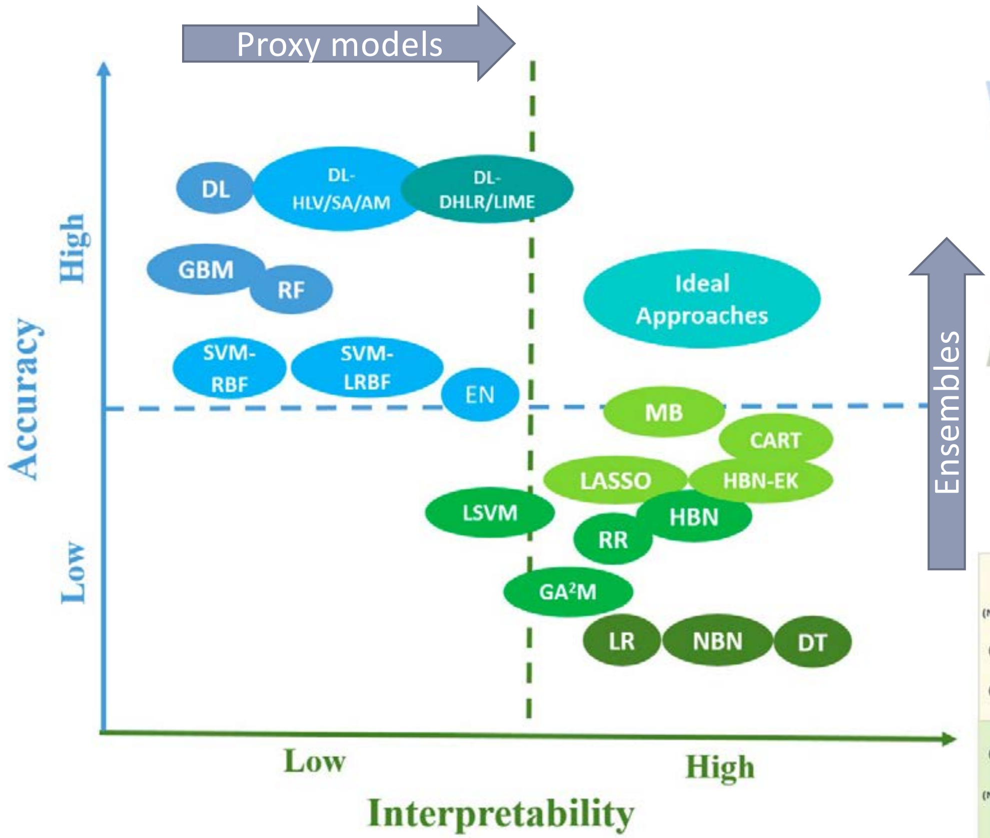
RADIOTHERAPY

Prospective clinical deployment of machine learning in radiation oncology

Issam El Naqa

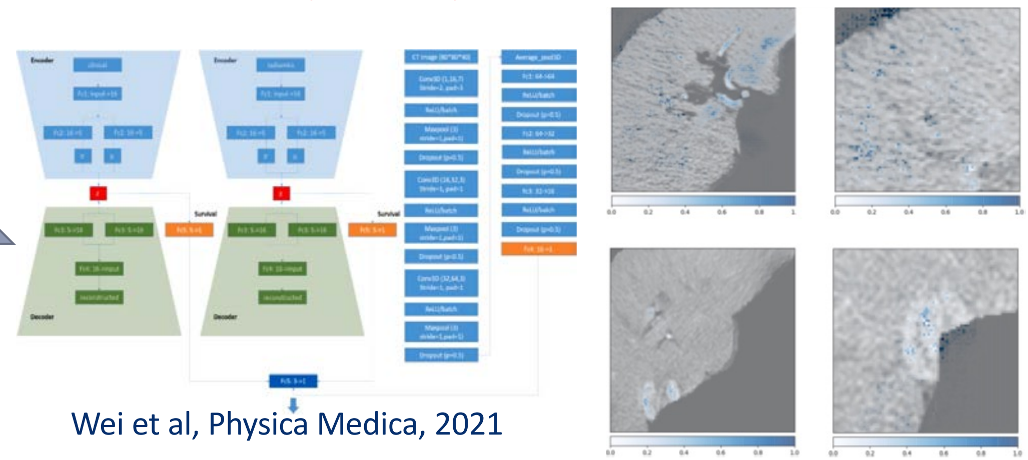
Nature Reviews Clinical Oncology (2021) | Cite this article

ML Accuracy versus interpretability



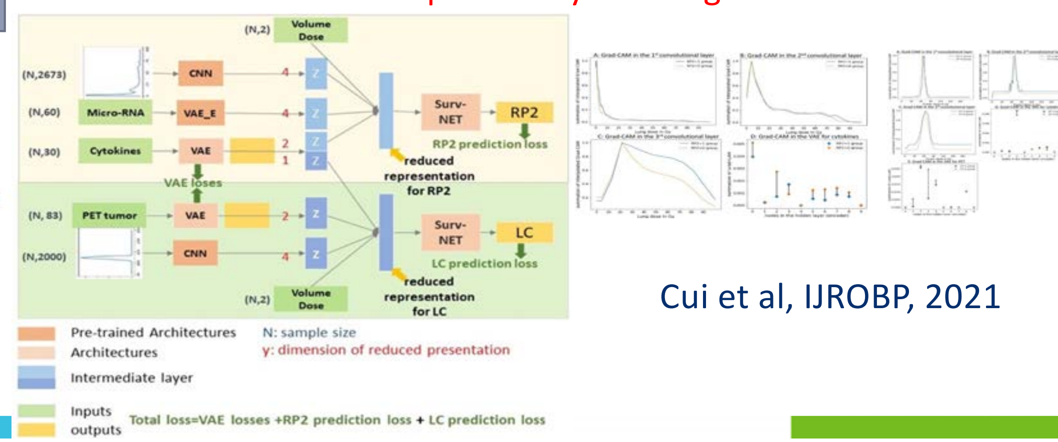
Luo, BJR-O, 2019

Radiomics Interpretability for Liver Cancer (Grad-CAM)



Wei et al, Physica Medica, 2021

Multi-omics interpretability for Lung Cancer



Cui et al, IJROBP, 2021

MEDICAL PHYSICS
The International Journal of Medical Physics Research and Practice

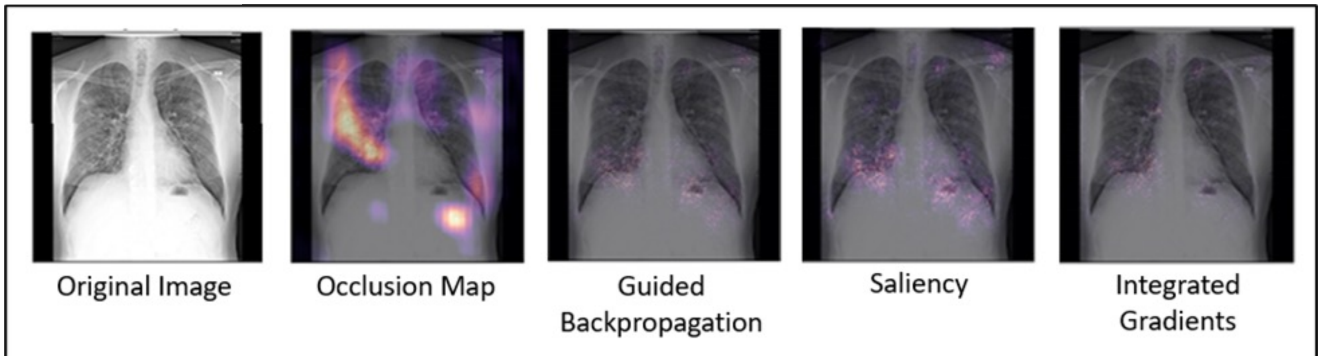
REVIEW ARTICLE | Free Access

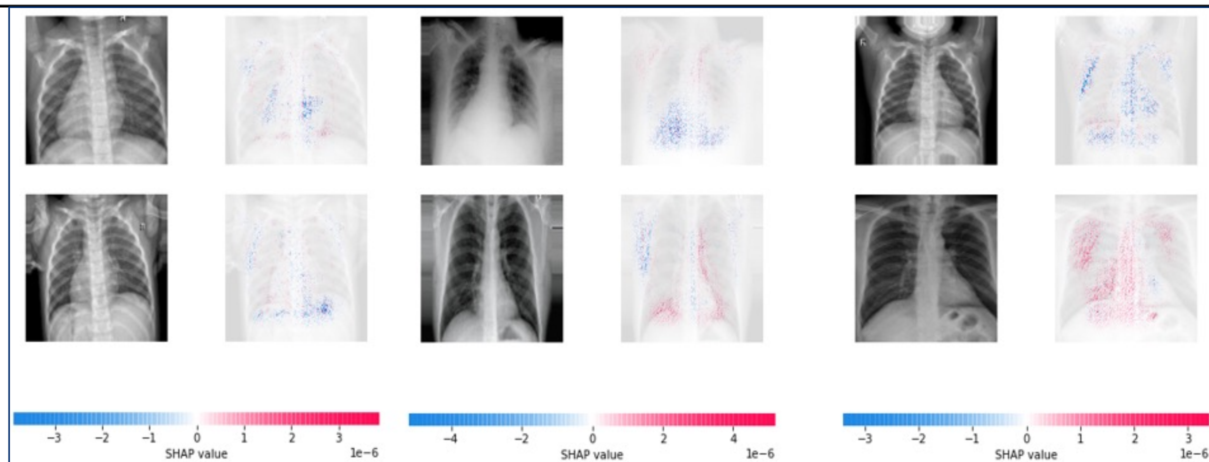
A review of explainable and interpretable AI with applications in COVID-19 imaging

Jordan D. Fuhrman, Naveena Gorre, Qiyuan Hu, Hui Li, Issam El Naqa, Maryellen L. Giger

First published: 18 November 2021 | https://doi.org/10.1002/mp.15359

Senior author: Maryellen L. Giger m-giger@uchicago.edu





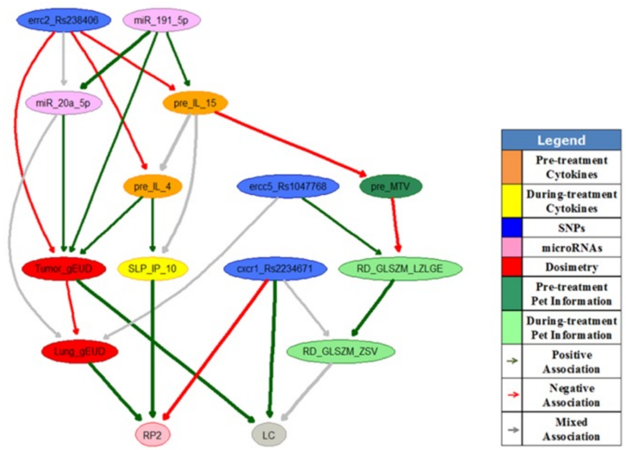
Intelligence augmentation (IA) instead of AI

Human in the loop!

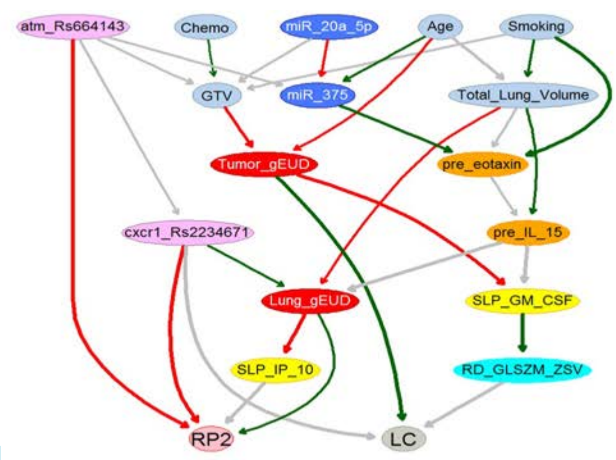
Figure 1. A "Fundamental Theorem" of informatics.

(C. Friedman)

Data-driven ML



Human + Data-driven ML

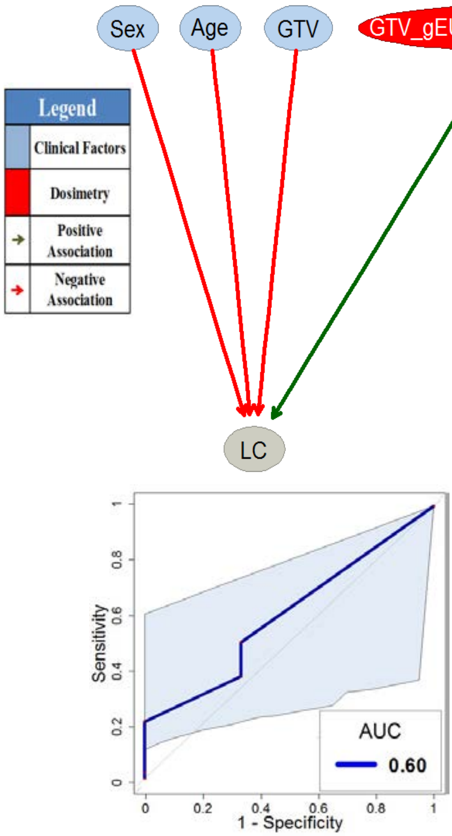


Tighter CIs but similar predictions!

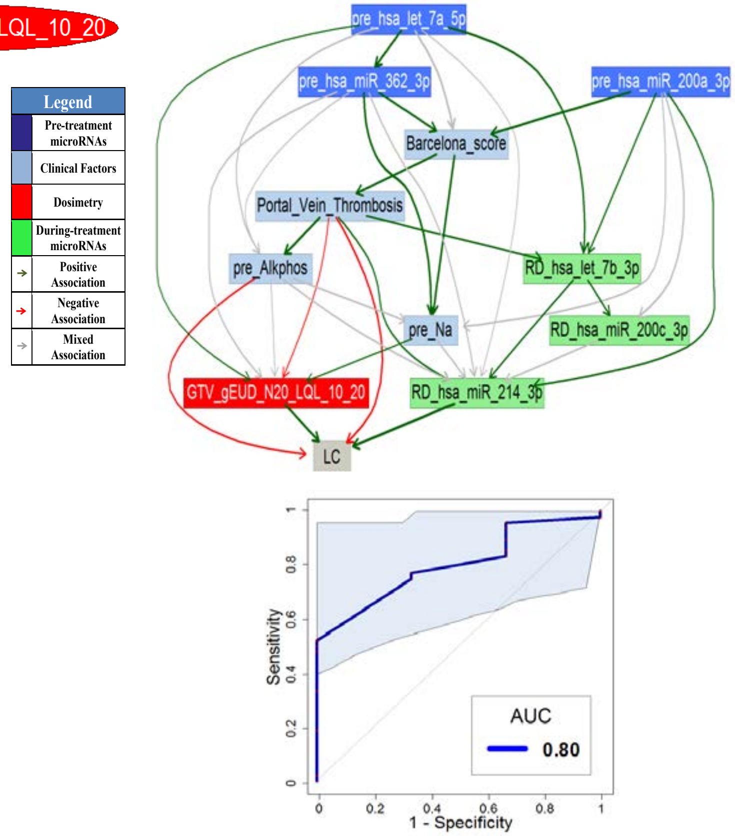
Luo, Physica Medica (Editor Choice), 2021

Human-in-the loop: Predicting Local Control in Liver Cancer

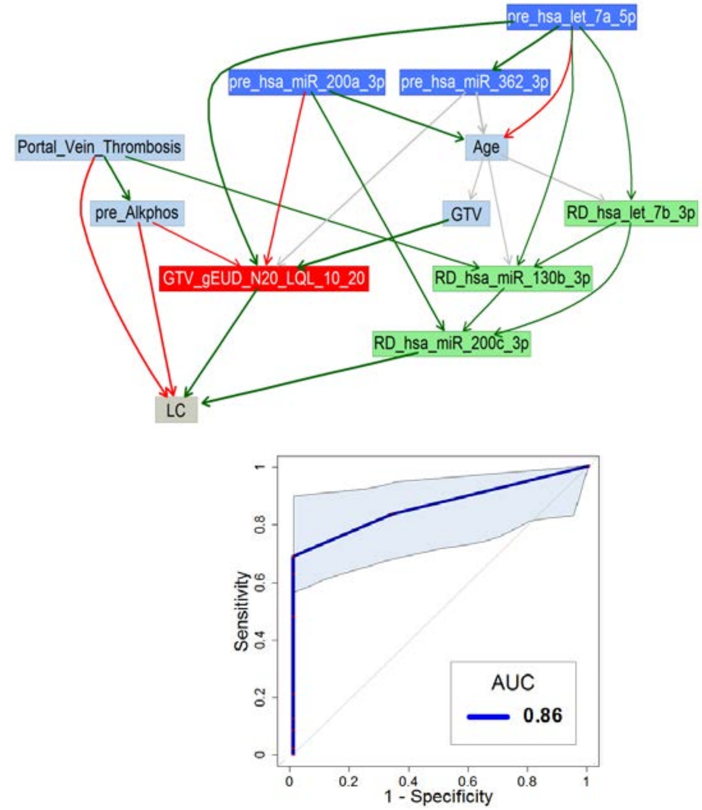
Human only



Machine only



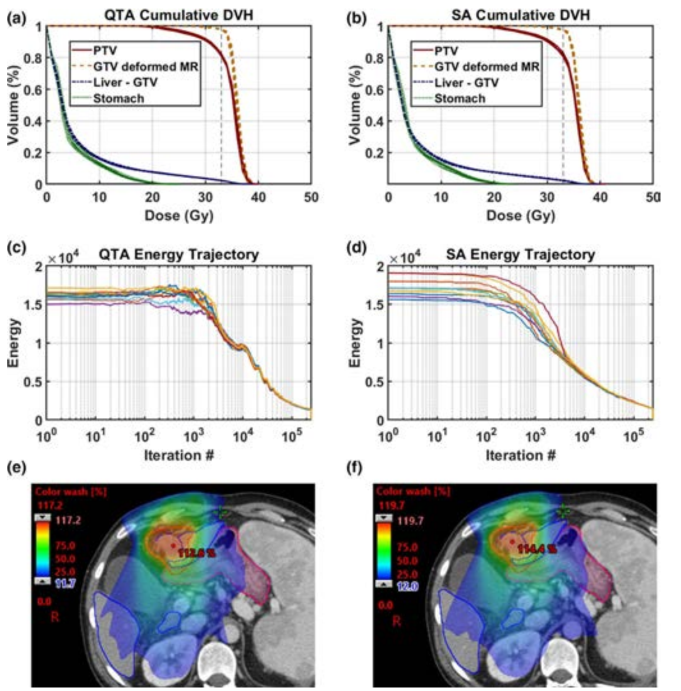
Human + Machine



Luo et al, Front Oncol, 2022

Can Quantum theory help develop more robust AI/ML algorithms?

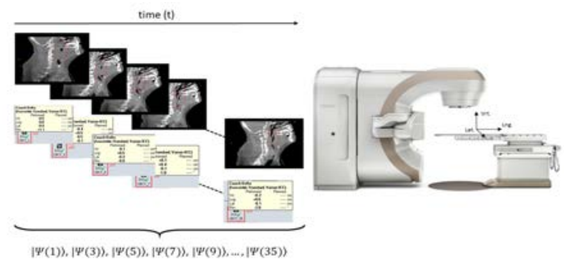
Treatment Planning

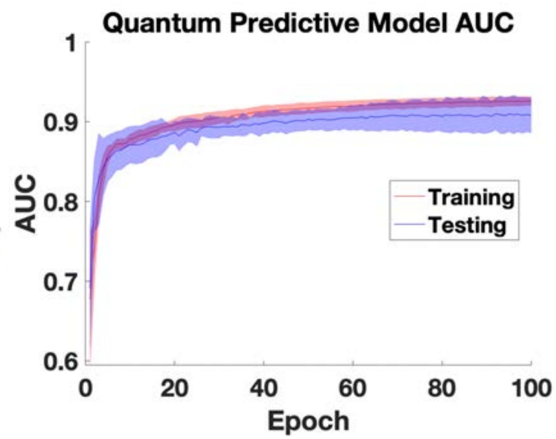


Algorithm	Width Function	Mean Convergence Rate (s)
SA	N/A	1157 ± 154.5
QTA	Hybrid	757.8 ± 162.3
QTA	MOCVD	622.1 ± 103.2
QTA	Sinusoid	526.2 ± 126.1

Pakela, Med Phys, 2020, (Editor's Choice)

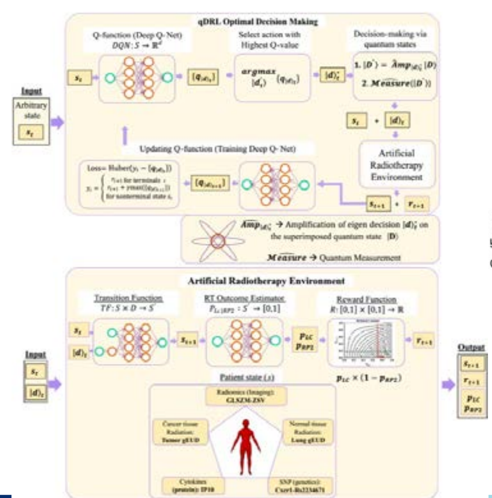
Image-guided radiotherapy

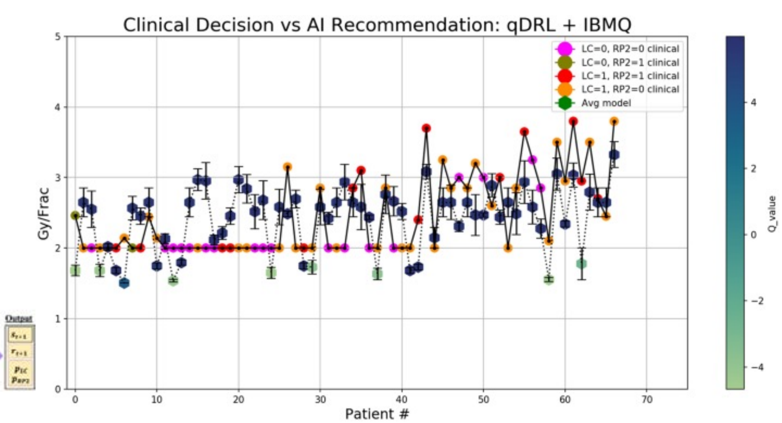




Pakela et al, PMB, 2021

Clinical Decision support





Niraula et al, Nature Sci Rep, 2021

REVIEW ARTICLE

AI and machine learning ethics, law, diversity, and global impact

[1]KATHERINE DRABIAK, JD, [1]SKYLAR KYZER, [1]VALERIE NEMOV, BS and [2]ISSAM EL NAQA, PhD

[1]Colleges of Public Health and Medicine, University of South Florida, Tampa, FL, USA
[2]Department of Machine Learning, Moffitt Cancer Center, Tampa, FL, USA

Table 3. Recommendations for trustworthy and ethical AI/ML.

Recommendation	Sources
Ethical requirements (IRB/HIPAA) are monitored in data aggregation and annotation	UK Data Protection Act 2018[43,73] EU General Data Protection Act[43,73] HIPAA[78] Mittelstadt 2021[25]
Transparency of training data characteristics, augmentation methods and ensuring proper inclusion of underrepresented groups (across age gender and race)	CLAIM, Consort-AI, CLAMP[50]
Transparency of training data model developments (architecture, loss function, optimization parameters)	CLAIM, Consort-AI, CLAMP[50]
Multilevel evaluation process (internal and external)	TRIPOD/Equator network[47]
Mitigation of explicit and implicit data leakage between training and testing	El Naqa et al, 2021[50]
Evaluation of human factors in evaluating real-world implementation and conduct prospective clinical trials if necessary	Luo et al. 2019[22] Mahadevaiah et al. 2020[37] Char et al. 2020[41] UK Department of Health and Social Care[43]
Continuous quality assurance and monitoring of deployed AI/ML models and live data incorporation	US FDA guidance[15] UK MHRA guidance[66] IMDRF[69]

AI, artificial intelligence; ML, machine learning.

Quality assurance for AI/ML application in the clinic

Acceptance Testing

- To ensure that the ML tool meets all applicable safety and performance standards (prediction) and that it meets contractual specifications
- Manufacturer includes an acceptance test procedure with the ML tool
  - Selection of evaluation endpoint and definition of performance criteria (e.g., AUC);
  - Selection of a benchmark data

Commissioning

- The process whereby the needed tool-specific data/parameters are acquired and operational procedures are defined
- May include:
  - Training data collection
  - Developing procedures
  - User training before first use

Quality Assurance (QA)

- Effort to ensure treatments are given accurately, safely and efficiently according to established tests and evaluations

Continuing Quality Improvement (CQI)

- Effort that seeks to make treatments and operations better by recognizing current weaknesses in the program, anticipating problems before they happen, streamlining tasks and responding to changes in practice

Table 10.1
Contemporary QA considerations for the current state of machine learning applications

TYPE OF MACHINE LEARNING APPLICATION	QA CONSIDERATIONS FOR THE CURRENT STATE			
	PERFORMED BY REVIEWED BY	COMMISSIONING	ROUTINE QA	RISK BEING MITIGATED
ML replaces human tasks: linear accelerator QA	Confirm functionality with sample QA data (Ritter et al. 2018)	• Evaluate ML against current clinic standards (Klein et al. 2009) • Test limits of analytics such as by inserting errors into delivery tests or datasets for analysis, e.g., intentional leaf offset present in the measurement result but missing in the delivery file • Document situations where the software passes and fails • Document situations where results differ by >5%	• Frequency: monthly • Monitor software settings for analysis • Repeat analysis of a subset of the commissioning dataset (e.g., dynamic leaf gap) including one at the limit • Expect identical results unless the software has changed. • If software has changed, determine if a new baseline is needed • Evaluate against a subset of the manual analysis for software update • Review trends	• Confirm that the analysis is performed correctly to avoid the hazards of expectation bias
ML supplemental to human tasks: treatment planning	• Confirm functionality with vendor-supplied treatment plans • Define scope of ML for planning	• Evaluate behavior against appropriate portions of original TPS commissioning results (if available) (Fraass et al. 1998) • Are clinical goals met? Is the agreement within ±5% for key metrics, such as mean dose for targets and max dose to a volume (e.g., 1 cc)? • Evaluate ML tools for a range of body sites and have site-specific rollout of techniques for at least a limited number of body sites • Evaluate permissions of different user types for applying ML techniques (e.g., physicist vs. dosimetrist) • Have different users perform the same test case—results within 5%? • Establish procedures for quality control steps post-application of ML, e.g., MD and physicist review of final dose distribution	• Repeat analysis of a subset of the commissioning dataset (e.g., dynamic leaf gap) including one at the limit • Monitor key dosimetric results from ML techniques using Big Data Analytical tools where available by body site: e.g. target coverage and maximum dose to a volume (e.g., 1cc) for OARs (Mayo et al., 2017) • Add extra scrutiny on key metrics for the first 5 patients per body site	• Monitor for any unintentional shift in clinical practice due to settings in the ML algorithm • Maintain evaluation of plan against MD-provided goals (planning objectives) (Evans et al. 2016; Marks et al. 2013)

(continued next page)

Table 10.1 *(continued)*
Contemporary QA considerations for the current state of machine learning applications

TYPE OF MACHINE LEARNING APPLICATION	QA CONSIDERATIONS FOR THE CURRENT STATE			
	PERFORMED BY REVIEWED BY	COMMISSIONING	ROUTINE QA	RISK BEING MITIGATED
ML/AI enhances human tasks: patient workflow, such as preparation for optimization	Confirm functionality and understand the scope of what is automated	• Define if ML tools will be applied and implemented for all patients or by body site • Create a commissioning dataset which includes manual preparation of the plan for optimization and automated preparation • Confirm reasonably concordant results between human and automated creation • Inspect the overlay of human vs. automated volumes to confirm expansions are correct • Verify volumes for optimization are within 5% or 2 cc (for optic and other small structures)	• Repeat a subset of the commissioning dataset • Confirm derivative structures such as optimization structures are consistent with those by humans (monthly) • Confirm that quality control steps post-application remain in place, such as review of the final dose distribution by MD and physicist	• Risk being mitigated is an incorrect expansion from target or OAR volumes to create optimization structures for dose coverage or sparing, respectively • Maintain evaluation of plan against MD provided goals (planning directive) (Evans et al., 2016; Marks et al. 2013)
ML additive: decision-making (El Naqa et al. 2018a)	• Evaluate with vendor-supplied dataset • Define size of training and testing dataset	• Partner with physicians to determine which disease types and staging are appropriate for the algorithm • Assess baseline variation in clinical practice among physicians within a practice, within a registry, or via publications before implementation • Assess sensitivity of the output of algorithms with training sets across the spectrum of limited variability to significant variability • Is the algorithm supporting implementation of a national practice standard? • Is the algorithm being used to apply new science in a clinical trial?	• Confirm that the input and expected output are consistent with the intent of the practice • Assess the frequency of patient type to determine how often the training dataset should be updated • Monitor the relationship between decisions with prior practice using Big Data Analytical tools where available by body site	

El Naqa, Moran, Ten Haken, The Modern Technology of Radiation Oncology, V4, Van Dyke

Oncogene www.nature.com/onc

REVIEW ARTICLE

Translation of AI into oncology clinical practice

Issam El Naqa[1]✉, Aleksandra Karolak[1], Yi Luo[1], Les Folio[2], Ahmad A. Tarhini[3], Dana Rollison[4] and Katia Parodi[5]

Table 1. Key requirements and their brief description for AI clinical implementation.

Requirement	Description
Technical	Hardware and software platforms to train and deploy AI/ML algorithms
Data modeling	Input training data is key influencer of AI performance. Further annotated data is needed for validation and testing retrospectively and possibly prospectively.
Regulations	Use in cancer care requires regulatory approvals (e.g., 510(k) by the food and drug administration (FDA) in the USA).
Ethics	AI is prone to bias and its implementation should be checked against societal ethical standards
Governance	A legal framework needs to be developed to monitor and ensure continued safe AI implementation.

Take home Messages

- Artificial intelligence/machine learning offers new opportunities to develop better understanding of oncology and therapeutic response
- ML/DL algorithms vary in accuracy and interpretability levels and choice of proper algorithm(s) is an application and data dependent
- Proper development and deployment of AI/ML involves following guidelines (CLAMP) with possible prospective validation while adhering to ethical AI standards to achieve trustworthiness
- Explainable AI (xAI) is key for trustworthiness & clinical translation
- To overcome current barriers in AI/ML for healthcare emerging methods include visualization for interpretability (Grad-CAM), behavioral science (human-in-the loop), physics-based (quantum computing) techniques
- Collaboration between stakeholders (data scientists, biologists, physicists, economists, clinical practitioners, regulators & vendors) will allow for safe and beneficial application of AI in biomedicine, radiology and oncology


COVER STORY
GUEST EDITORIAL
Moffitt Cancer Center: Why we are building the first machine learning department in oncology
By Issam El Naqa and Dana Rollison
Machine and Deep Learning in Oncology, Medical Physics and Radiology
A GUIDE TO OUTCOME MODELING IN RADIOTHERAPY AND ONCOLOGY
LISTENING TO THE DATA
Outcome Modeling
Utilization
Validation
Explanation
Reconstruction
Exploration
CRC Press
Emerging Developments
May 2020 | Volume 47, Issue 5
MEDICAL PHYSICS
The International Journal of Medical Physics Research and Practice
Special Issue on
The Role of Machine Learning in Modern Medical Physics
Guest Editors: Issam El Naqa and Shiva K. Das
WILEY
Special Issue on Datasets hosted in NCI's Cancer Imaging Archive
AMERICAN ASSOCIATION of PHYSICISTS IN MEDICINE
WILEY

THANK YOU!