ICADS

Integrated Program in Cancer and Data Science

FAIR Artificial Intelligence/Machine Learning (AI/ML) Course

Part of the T32 Integrated Cancer Data Science Program

Spring 2022

Co-directors: Issam El Naqa and AC Tan
Co-PIs: Cress Fridley, Flores

MOFFITT CANCER CENTER

(http://lab.moffitt.org/elnaqa)

ICADS

Integrated Program in Cancer and Data Science

Clinical and Research Data Made AI FAIR

- Findable
- Accessible
- Interoperable
- Reusable

- Coursework via USF
- Workshops via RET
- Public Videos via YouTube

WORKFORCE

Members Impacted

- Ph.D Students
- Postdoctoral Fellows
- Moffitt Faculty and Staff
- Potential Clinical Trainees
- Researchers world-wide
- Clinicians world-wide

Machine and Deep Learning Algorithm

Clinical Decision Making

Improvement in Health and Patient Outcome

Course Schedule

5/02 **Introduction to AI/ML/DL algorithms** (Dr. El Naqa)

5/09 **Training requirements for ML** (Dr. El Naqa)

5/16 **Methods Assessment, Uncertainty and bias estimation** (Dr. Tan)

5/23 **FAIR Principles** (Dr. Teer and Dr. Luo)

6/06 **Best practices** (Dr. Teer and Dr. Luo)

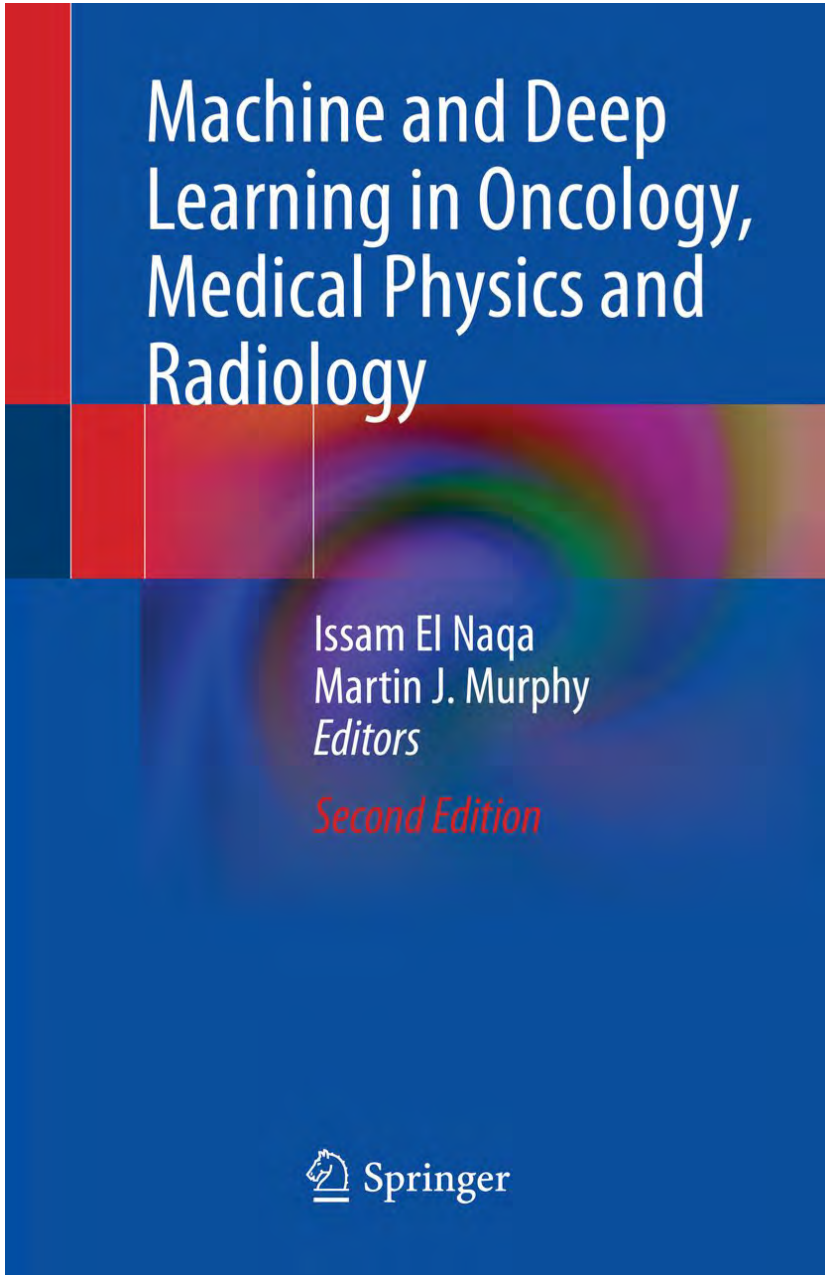
6/13 **Data resources** (Dr. Tan)

6/20 **Hands-on Workshop 1 – Radiomics** (Ms. Gorre and Mr. Carranza)

6/27 **Hands-on Workshop 2 – Cancer genomics** (Dr. Tan)

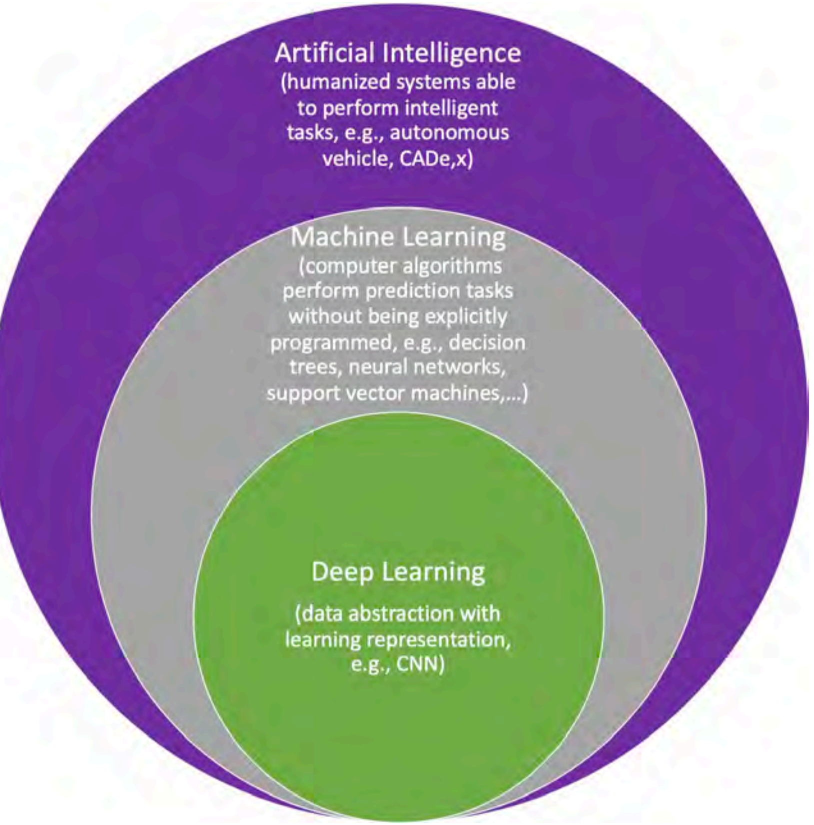
Lecture 1: Introduction to AI/ML/DL algorithms

Issam El Naqa, PhD



https://link.springer.com/book/10.1007/978-3-030-83047-2

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

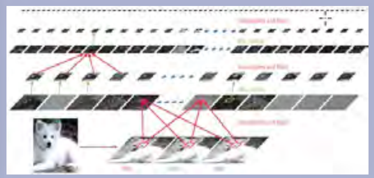
Originated in the 1960s

Computer algorithms that learn from data

Deep Learning

Originated in the 1970s

Based on neural networks that learn features



Formal definition of Machine Learning (ML)

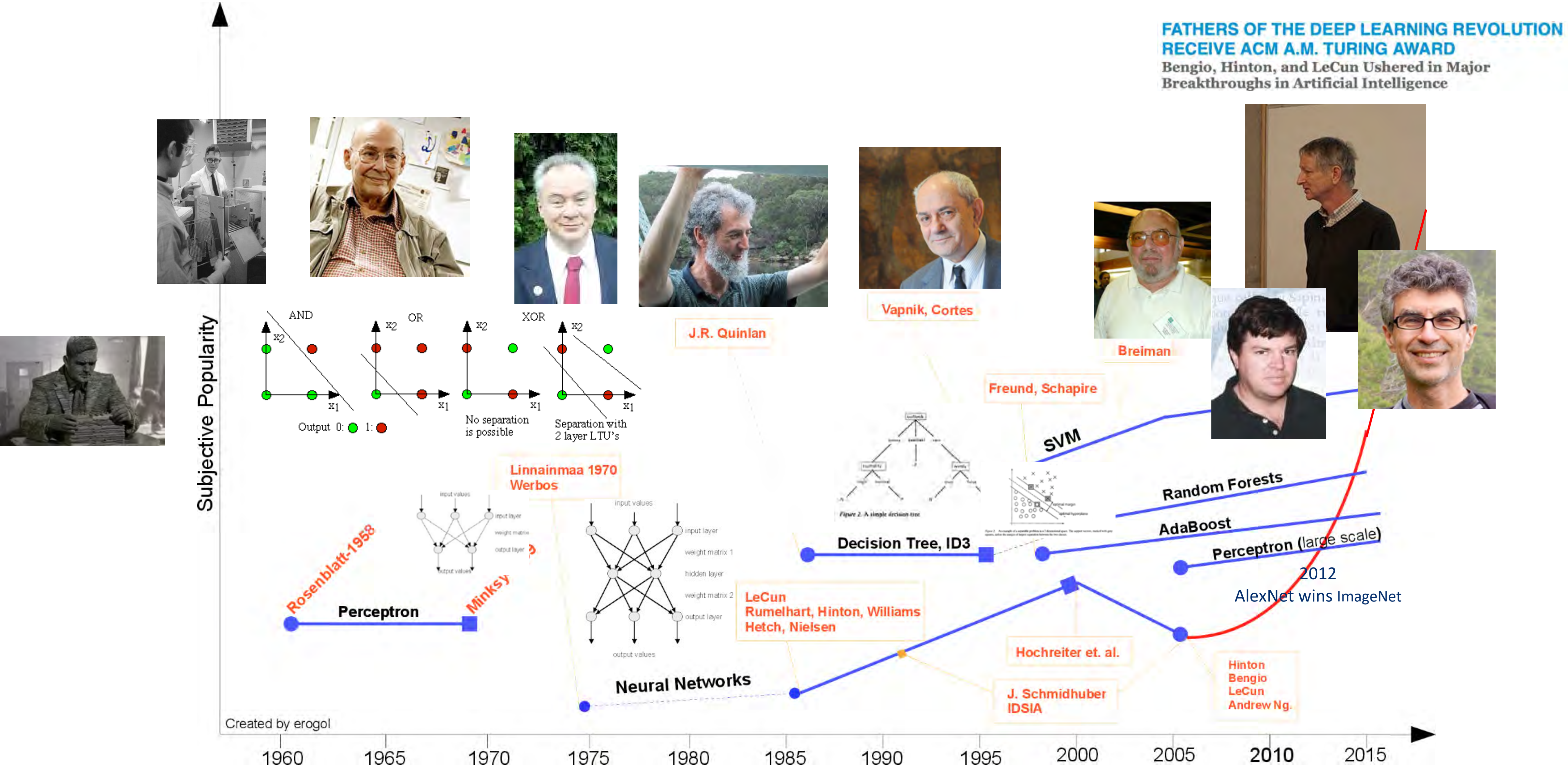
ML is:

- Field of study that gives computers the ability to learn without being explicitly programmed (Samuel, 1959)
- Computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E (Mitchell, 1998)
- Programming computers to optimize a performance criterion using example data or past experience (Alpaydin, 2009)

Related fields: ML builds on expertise from

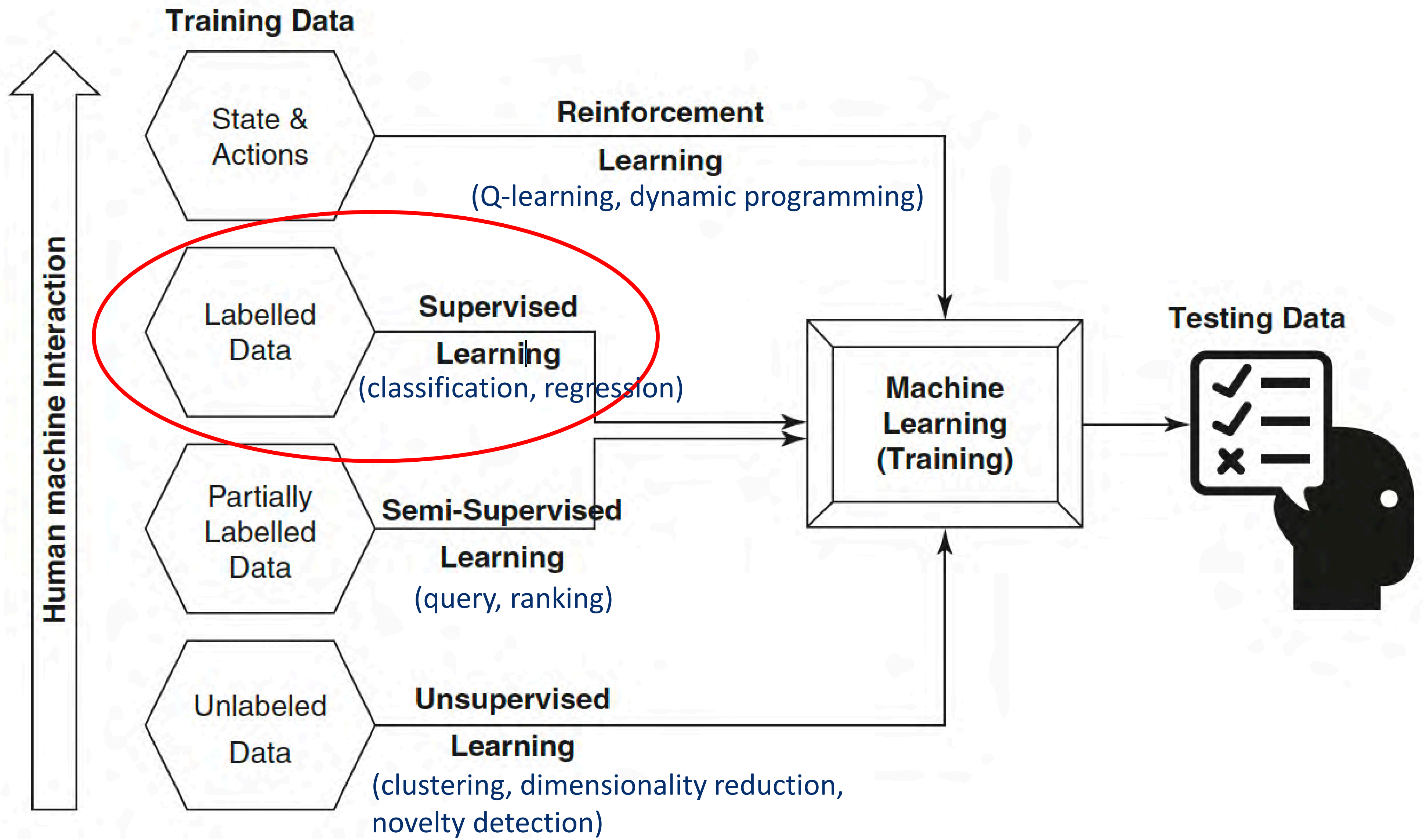
- artificial intelligence, probability and statistics, computer science, information theory, neuroscience, psychology, control theory, and philosophy

Machine Learning Evolution Timeline



(Adapted from E Golge)

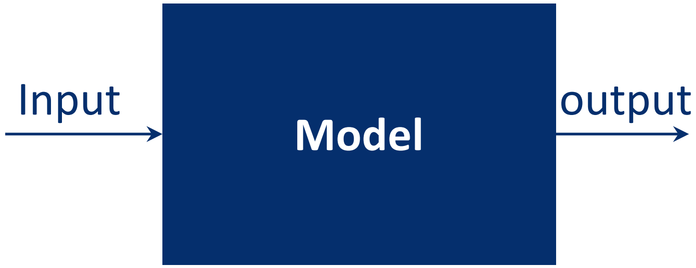
Machine learning by *tasks*



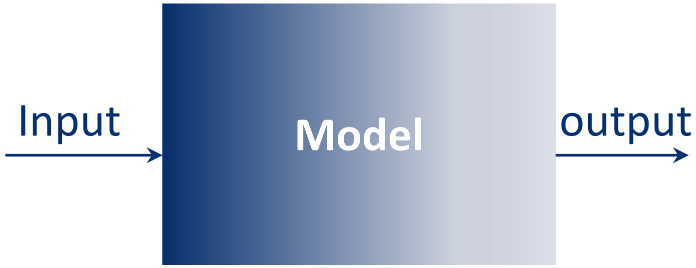
Machine learning by *models*

Probabilistic models could be divided:

- *Discriminant* models
  - Directly estimate posterior probabilities (logistic regression, neural networks, convolutional neural networks, random forests, support vector machines)
    - *Predict without knowing the system*



- *Generative* models
  - Model class-conditional probability and priors s (Bayesian networks, Markov models)
    - *To predict you need to know the system*

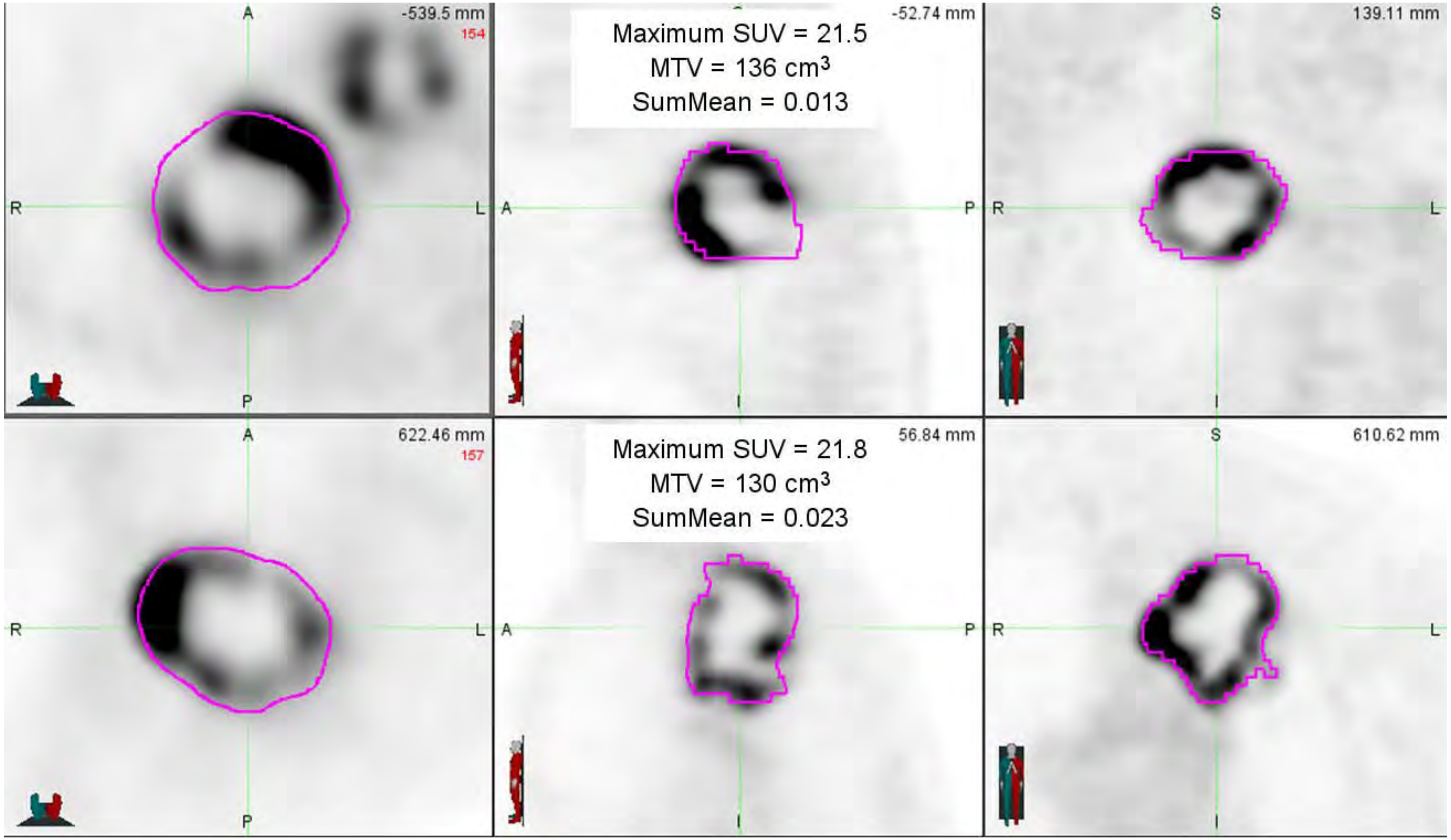


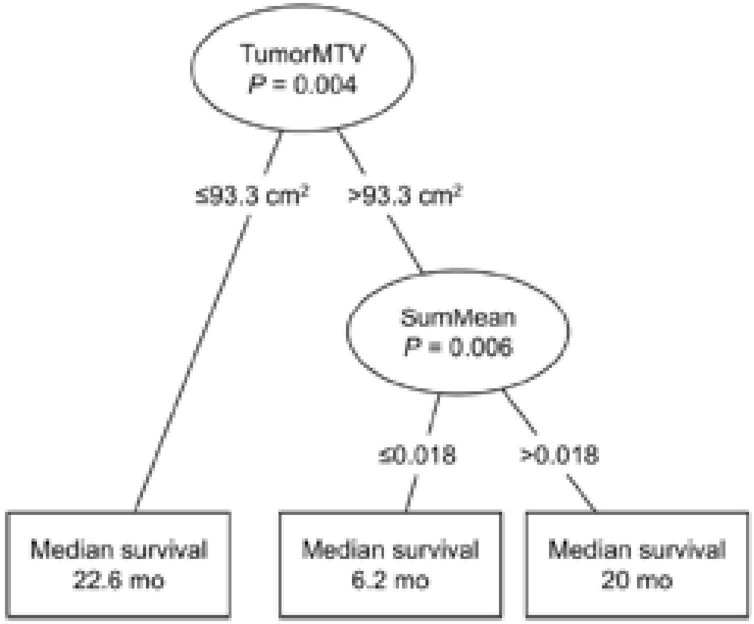
Tseng, Oncology, 2018

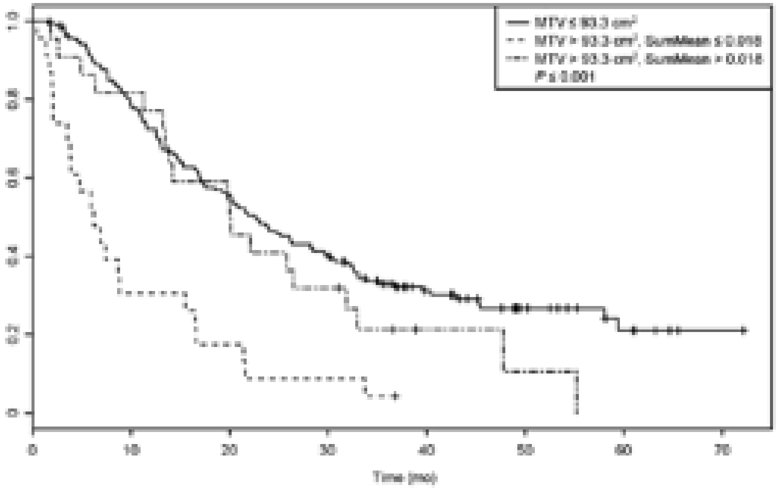
The Universe of Machine Learning (ML)

Adapted from: Brownlee, 2013

Radiomics for lung cancer outcome modeling with decision trees

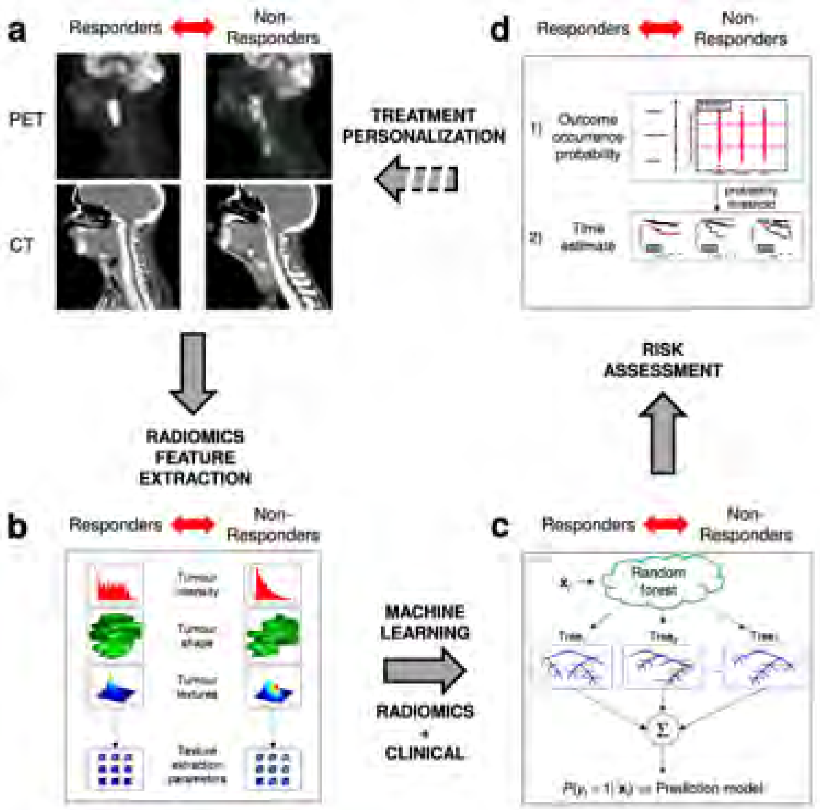


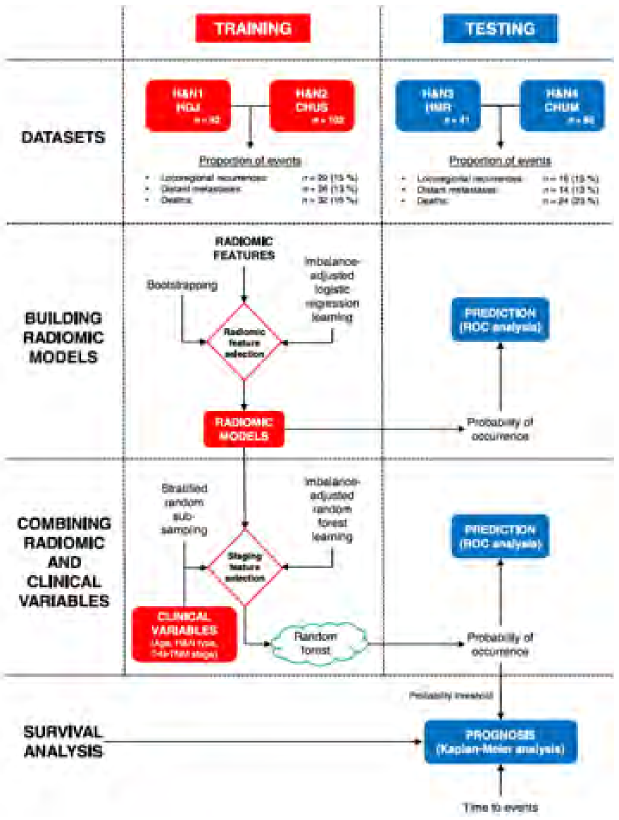


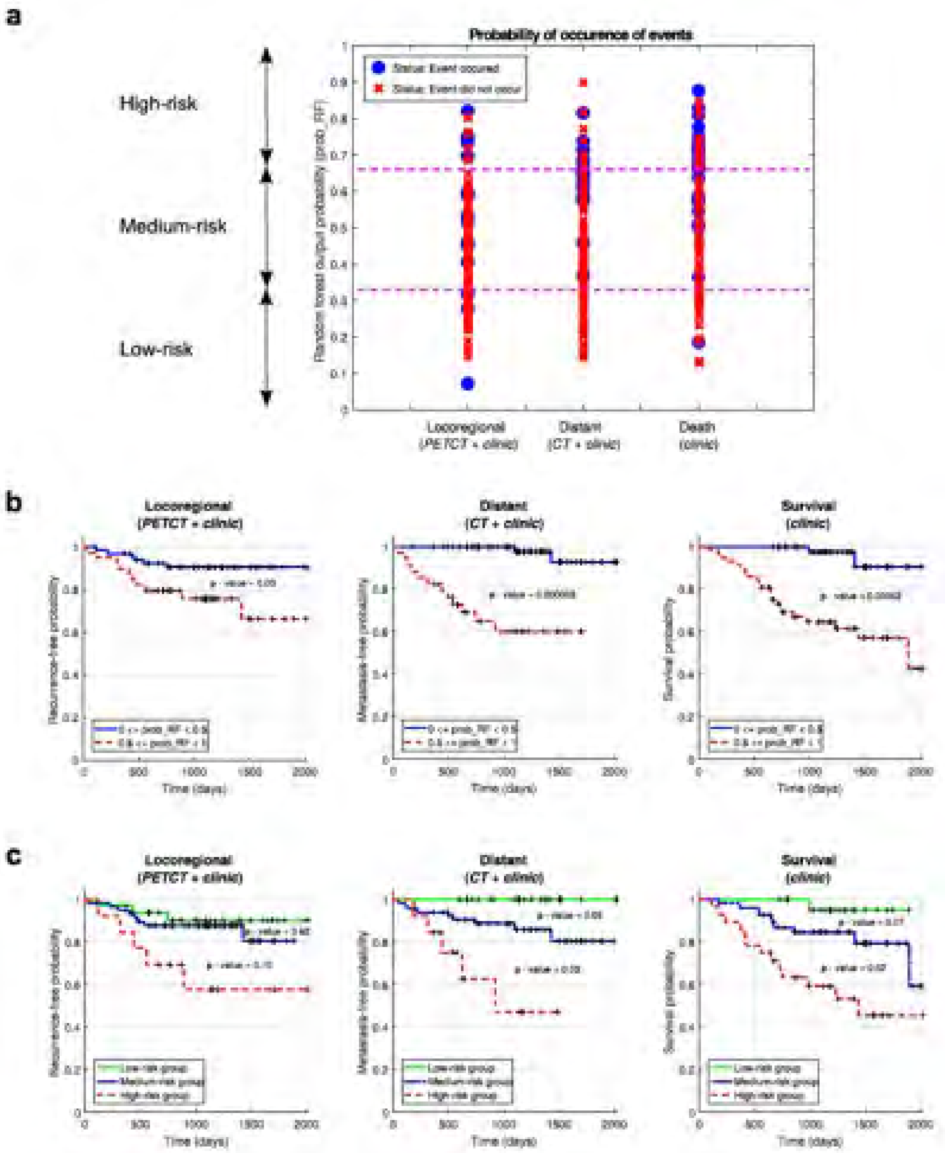


Ohri et al, JNM, 2016

Risk Assessment in Head & Neck

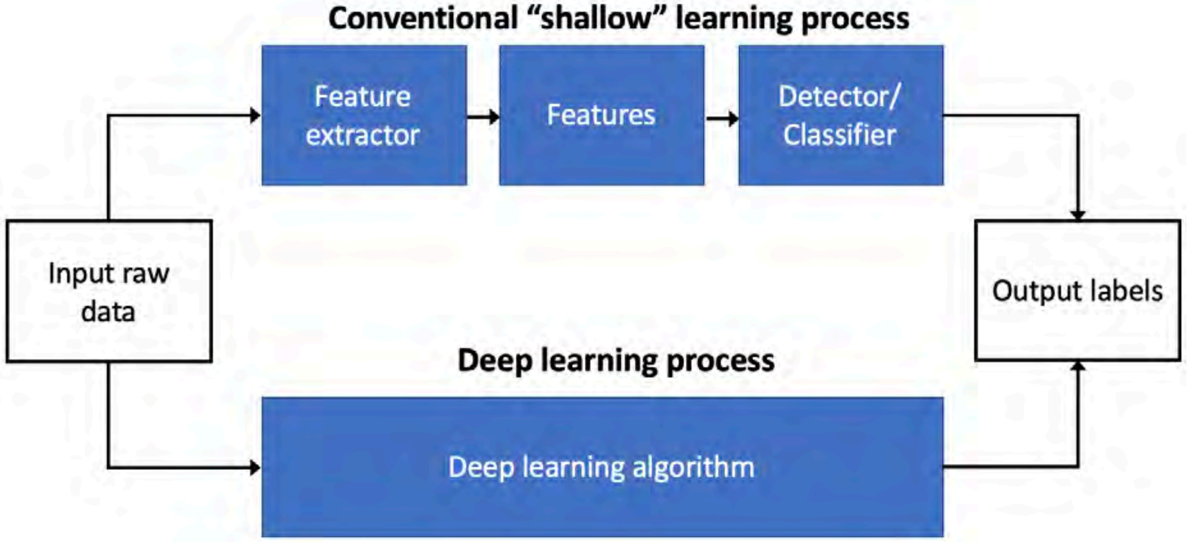


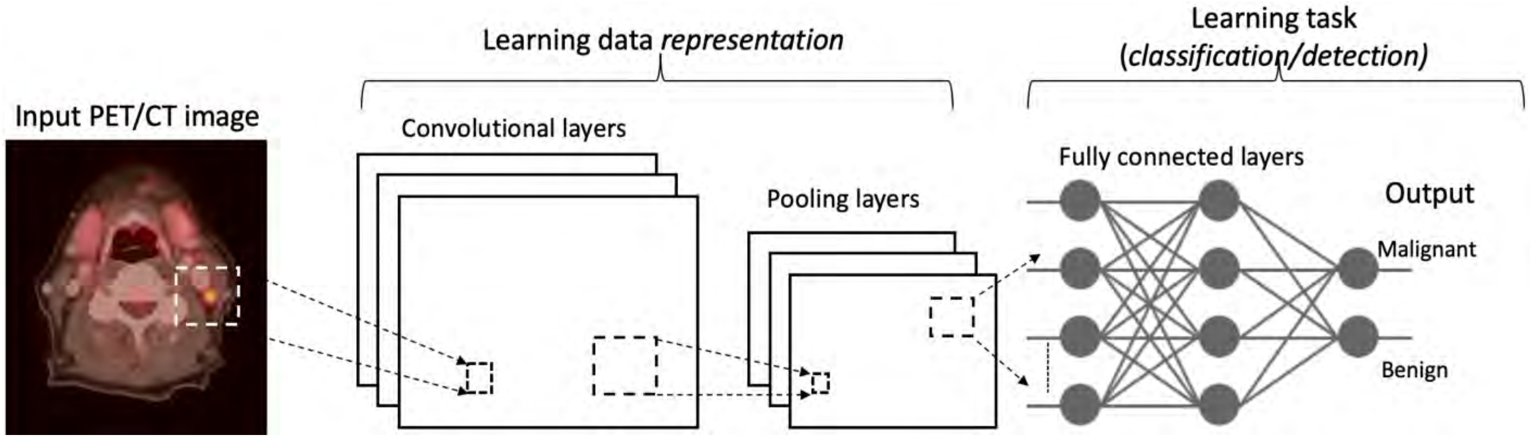




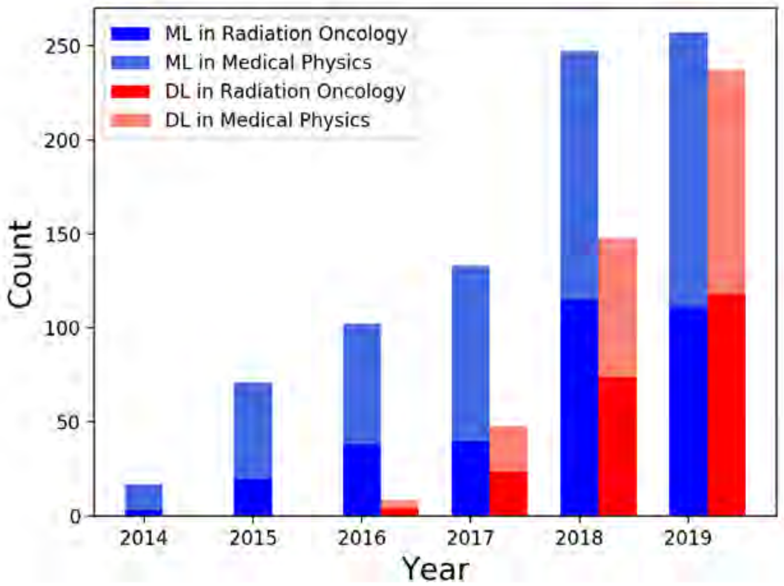
Carrier-Vallieres et al., Scientific report, 2017

Deep vs conventional machine learning





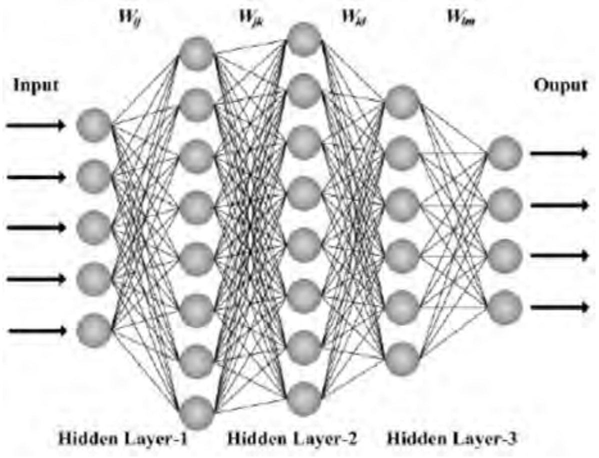
Zaidi and El Naqa, Annu. Rev. Biomed. Eng., 2021



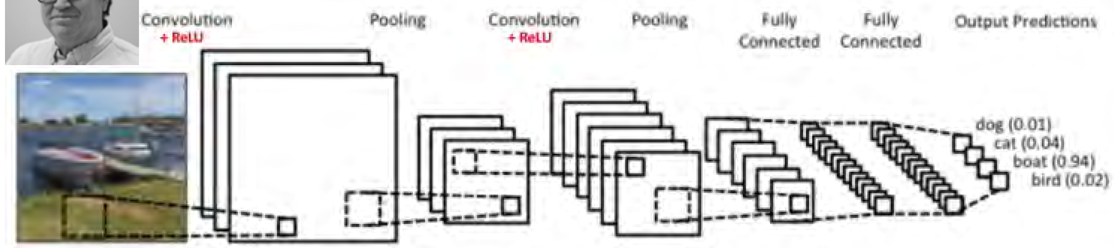
Cui, Med Phys, 2020

Deep Learning (NN) Architectures

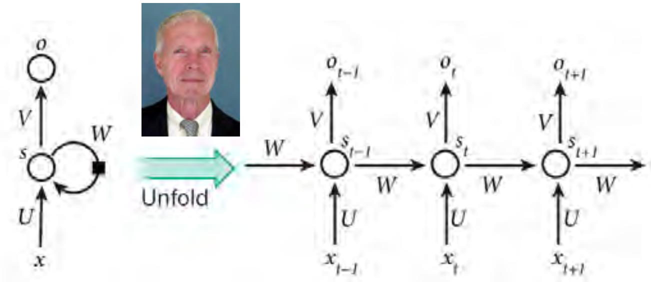
Multi-layer neural perceptron (MLP)



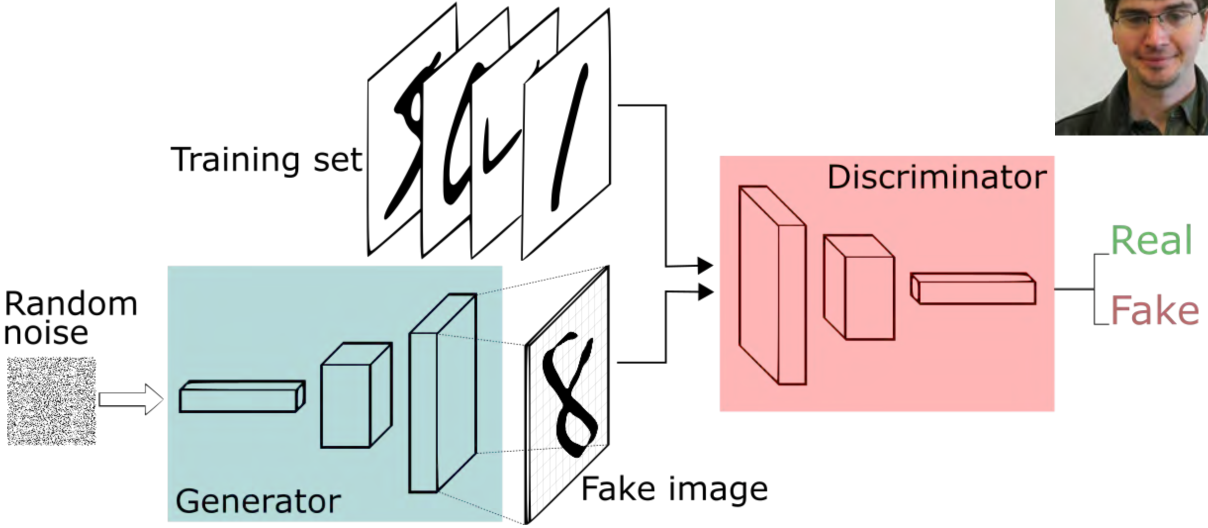
Convolutional Neural Network (CNN)



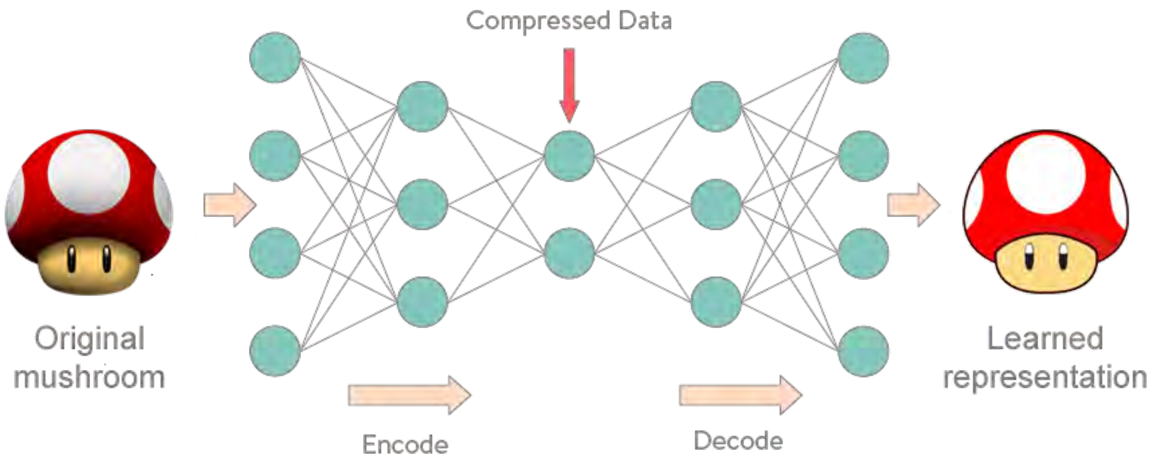
Recurrent Neural Network (RNN)



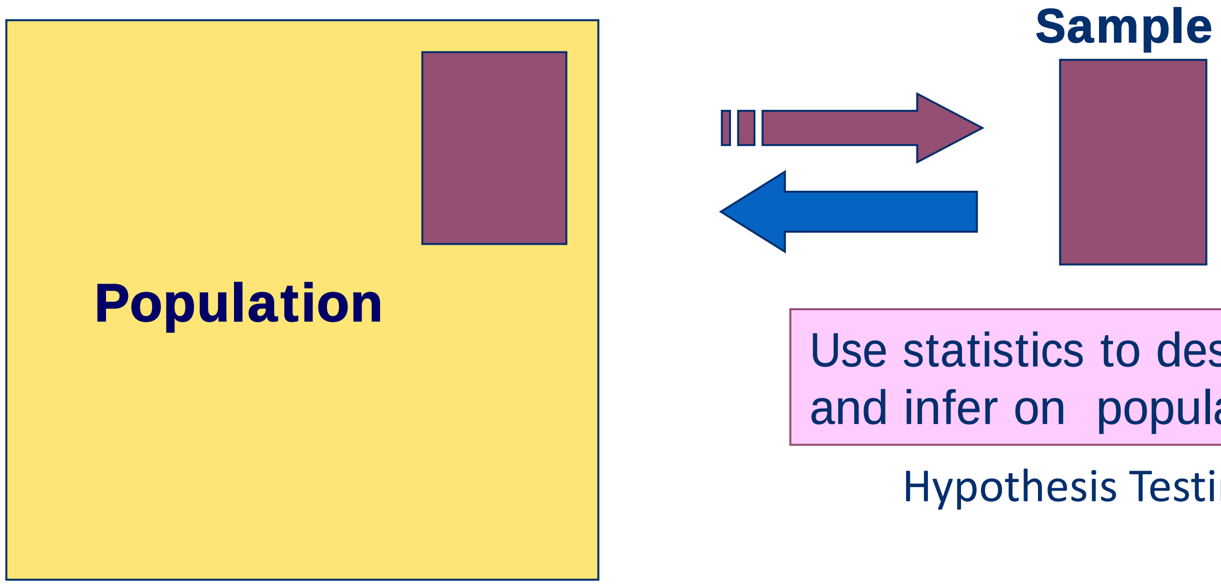
Generative adversarial networks (GAN)

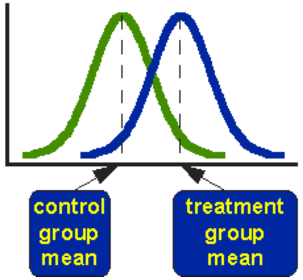


Autoencoders (AE)



Machine Learning vs. Epidemiology/Statistics





Use statistics to describe data and infer on population

Hypothesis Testing (inference)

Use computational modeling to describe data and summarize features

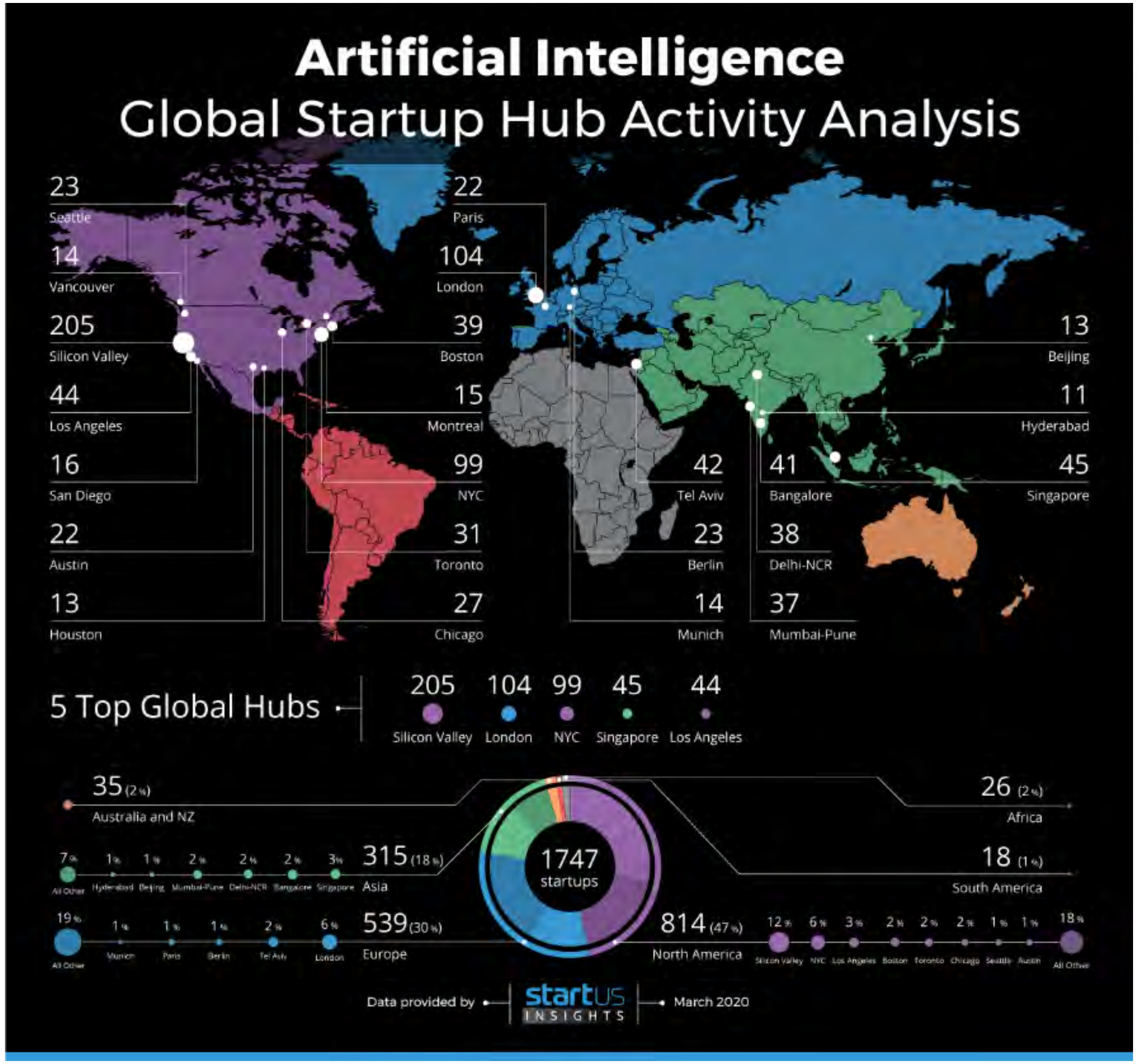
Hypothesis Generation (Prediction)

REVIEW ARTICLES

Prospects and Challenges for Clinical Decision Support in the Era of Big Data

Issam El Naqa, Michael R. Kosorok, Judy Jin, Michelle Mierzwa, and Randall K. Ten Haken

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".

Prioritize AI R&D
Grow and sustain U.S. research leadership and capacity

Leverage AI for Government and National Security
Apply AI to improve provision of government services and national security

Strengthen AI Research Infrastructure
Enhance access to high quality data, models, and computing resources

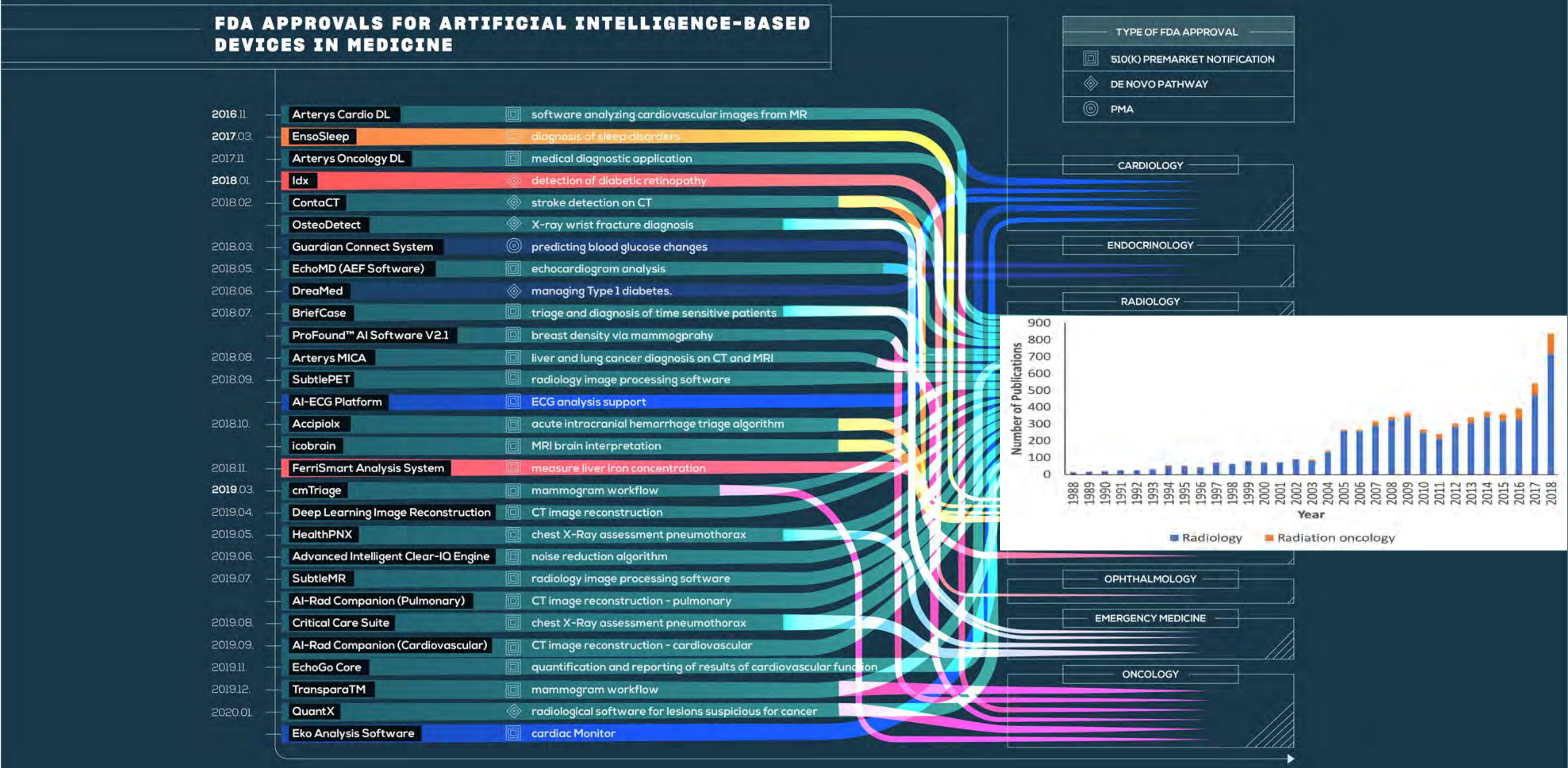
Promote International AI Engagement
Engage with like-minded allies to promote a global AI environment supportive of democratic values

Advance Trustworthy AI
Modernize governance and technical standards for AI-powered technologies, protecting privacy, civil rights, civil liberties, and other democratic values

Train AI-Ready Workforce
Provide AI-ready education at all levels: K-12, college, re-training, re-skilling, R&D workforce

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

https://www.ai.gov/nairrtf/86-fr-39081-responses/



Benjamens, S., Digital medicine, 2020

Why AI/ML in 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

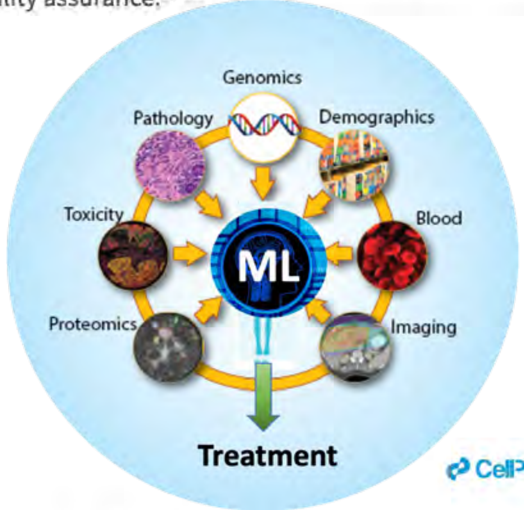
Artificial intelligence and machine learning techniques are breaking into biomedical research and health care, which importantly includes cancer research and oncology, where the potential applications are vast. These include detection and diagnosis of cancer, subtype classification, optimization of cancer 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 technica 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.

CellPress

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

Some Deep/Machine Learning medical applications

Applications of machine learning in drug discovery and development

Jessica Vamathevan, Dominic Clark, Paul Czodrowski, Ian Dunham, Edgardo Ferran, George Lee, Bin Li, Anant Madabhushi, Parantu Shah, Michaela Spitzer & Shanrong Zhao

Nature Reviews Drug Discovery **18**, 463–477(2019) | Cite this article

BRIEF COMMUNICATION
https://doi.org/10.1038/s41587-019-0224-x
nature biotechnology

Deep learning enables rapid identification of potent DDR1 kinase inhibitors

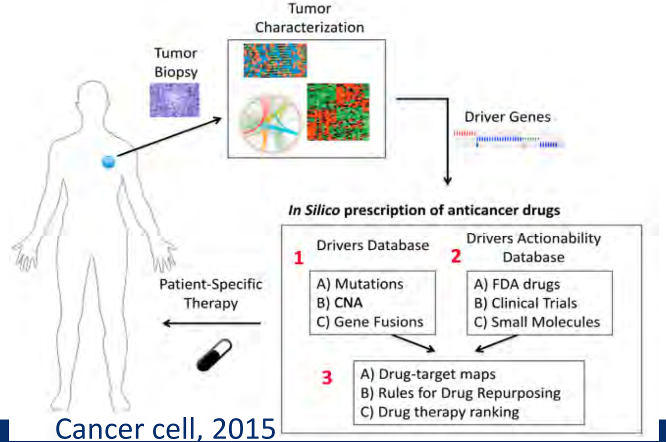
Highly accurate protein structure prediction with AlphaFold

John Jumper, Richard Evans, ... Demis Hassabis + Show authors

Nature **596**, 583–589 (2021) | Cite this article

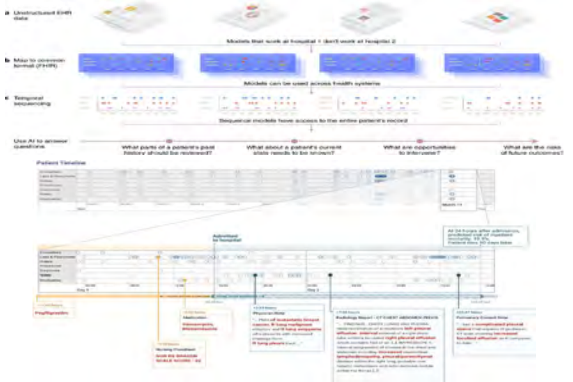
643k Accesses | **1590** Citations | **3071** Altmetric | Metrics

In silico trials



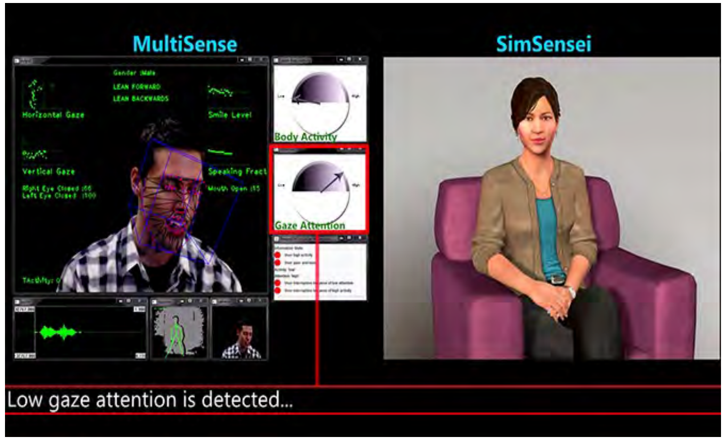
Cancer cell, 2015

Unlocking the *blackhole* of Electronic health records

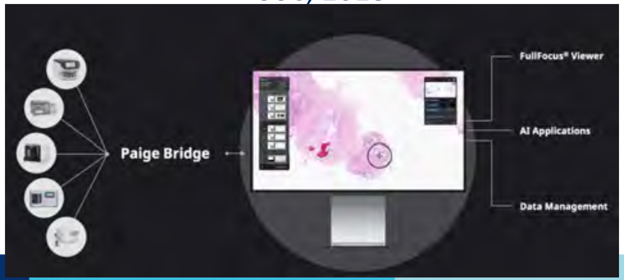


Digital medicine, 2018

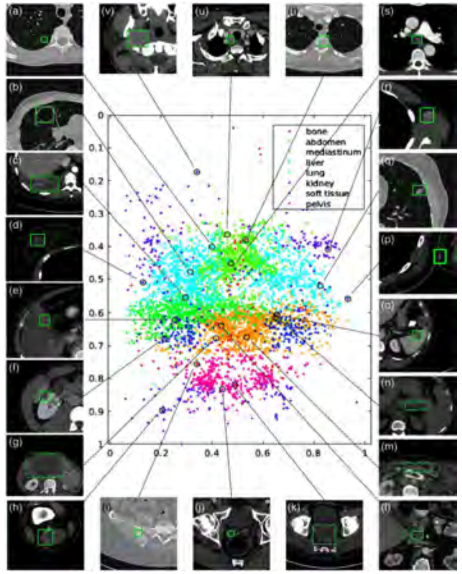
Virtual Counseling



USC, 2018

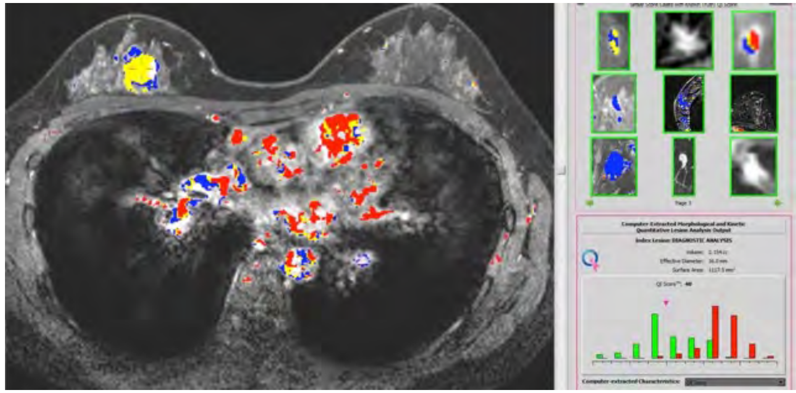


Annotation of radiological images



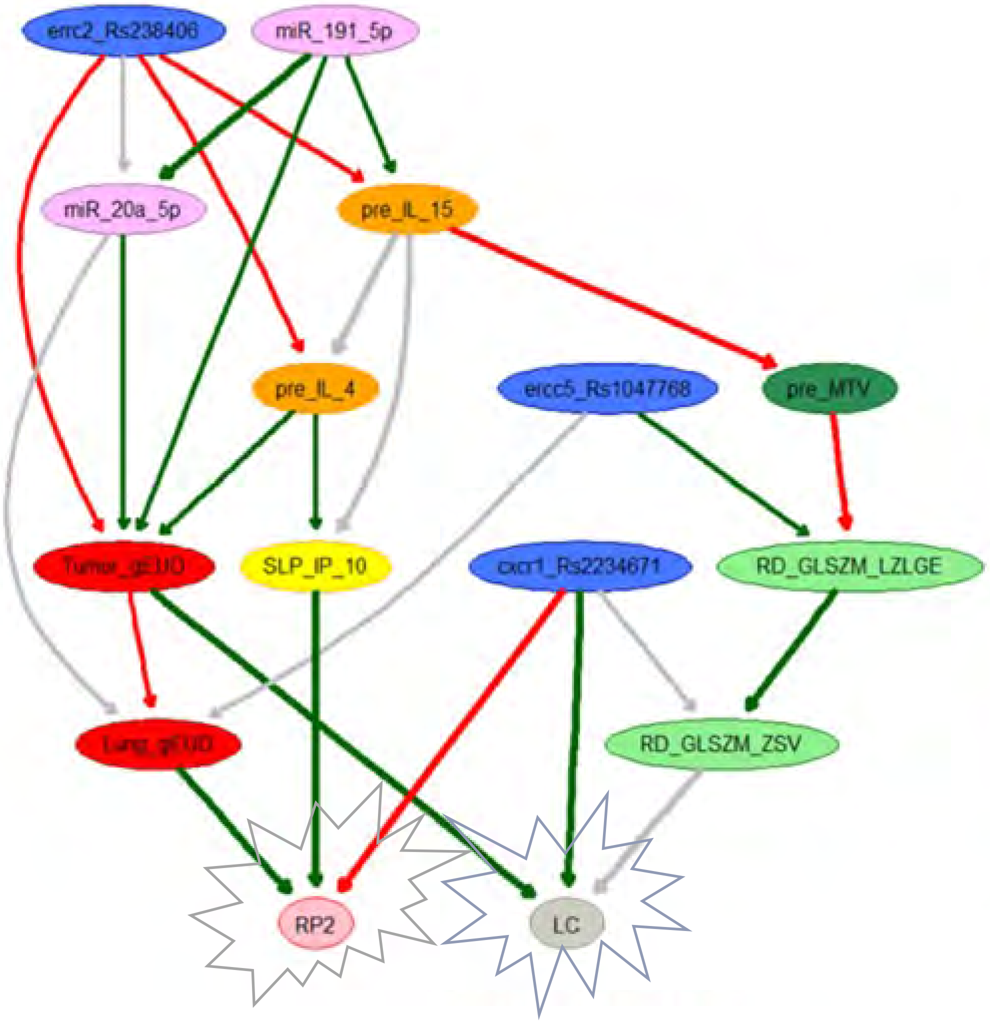
Yan, JMI, 2018

QuantX

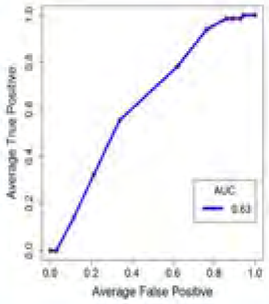
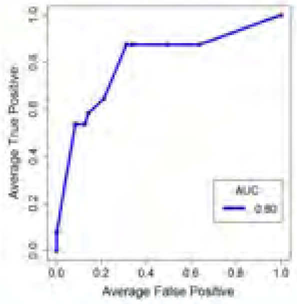
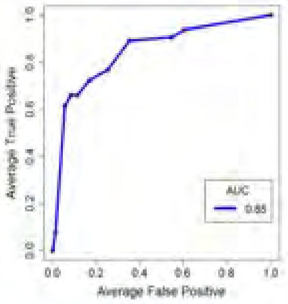


Multi-Objective Generative Models

A MO-BN 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.

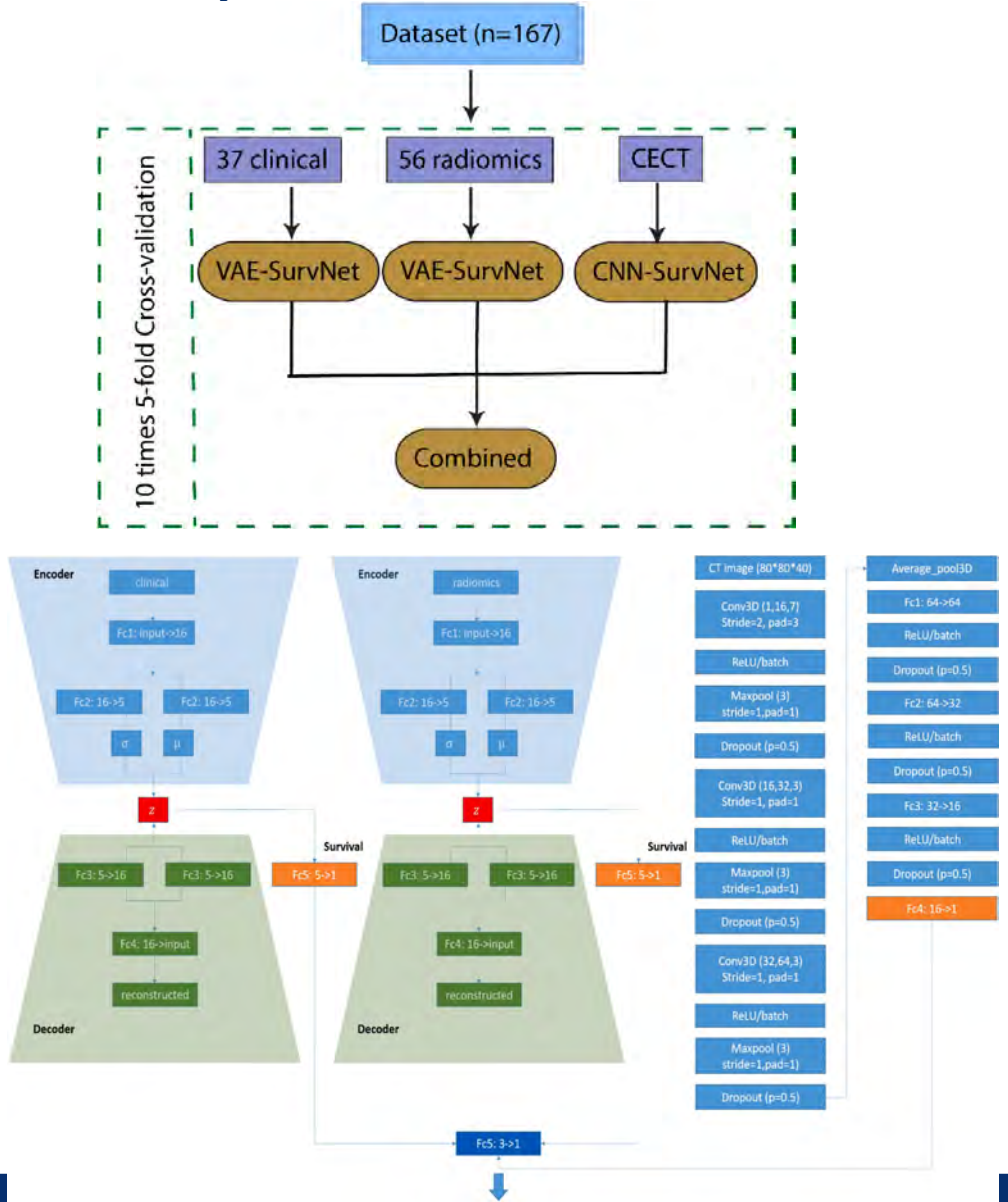


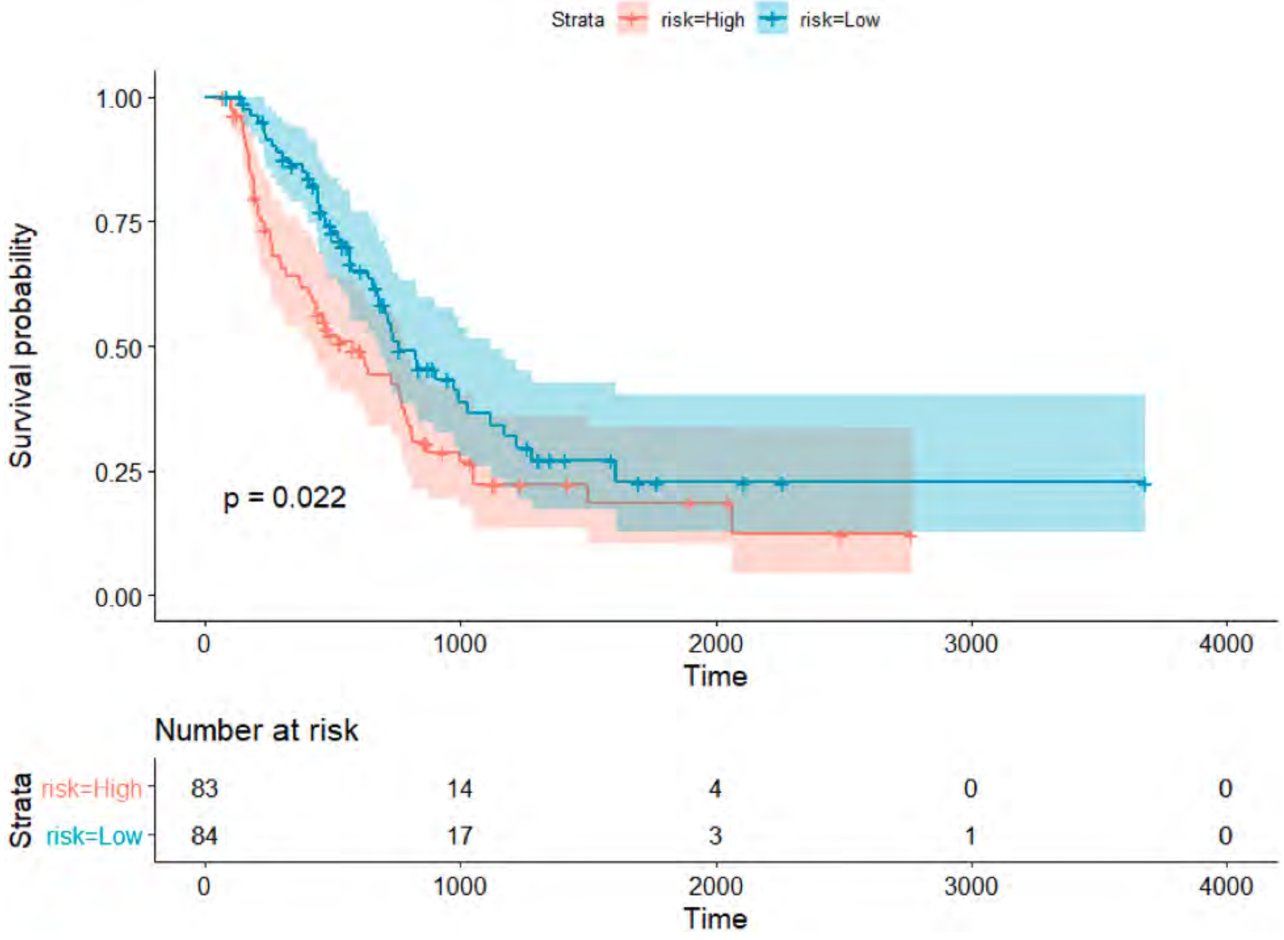
Legend	
	Pre-treatment Cytokines
	During-treatment Cytokines
	SNPs
	microRNAs
	Dosimetry
	Pre-treatment Pet Information
	During-treatment Pet Information
→	Positive Association
→	Negative Association
→	Mixed Association

	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, Med Phys, 2018 (Editor's Choice)

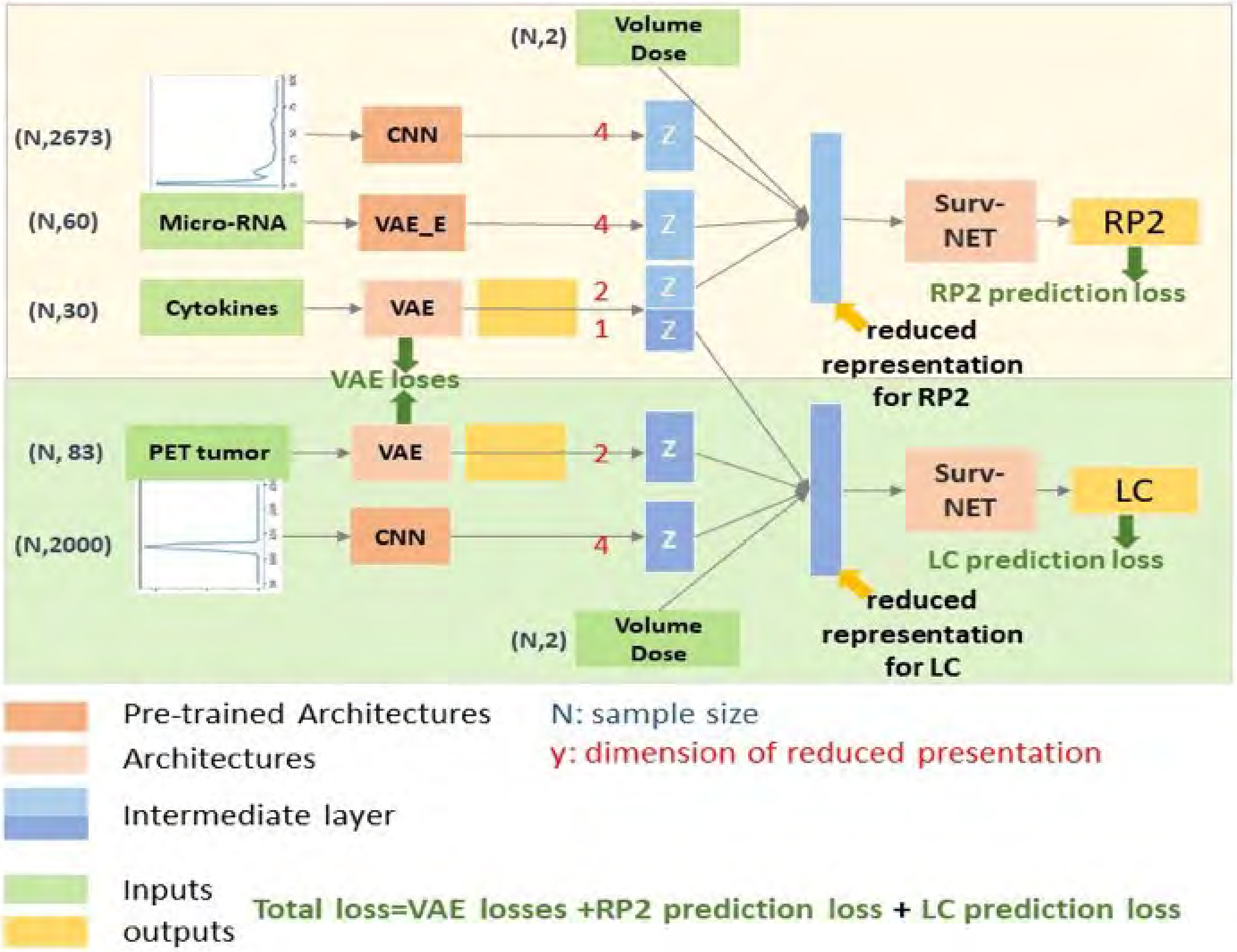
Radiomics model for Liver Cancer by Deep Survival





Wei, Physica Medica, 2021

Multi-objective multi-omics model with deep survival neural networks

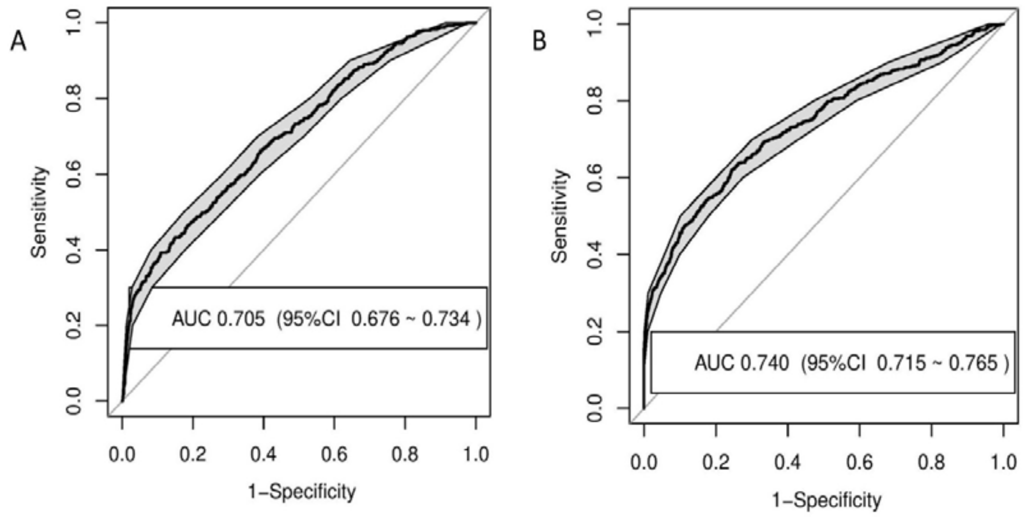


20 times of 5-fold cross validations

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)

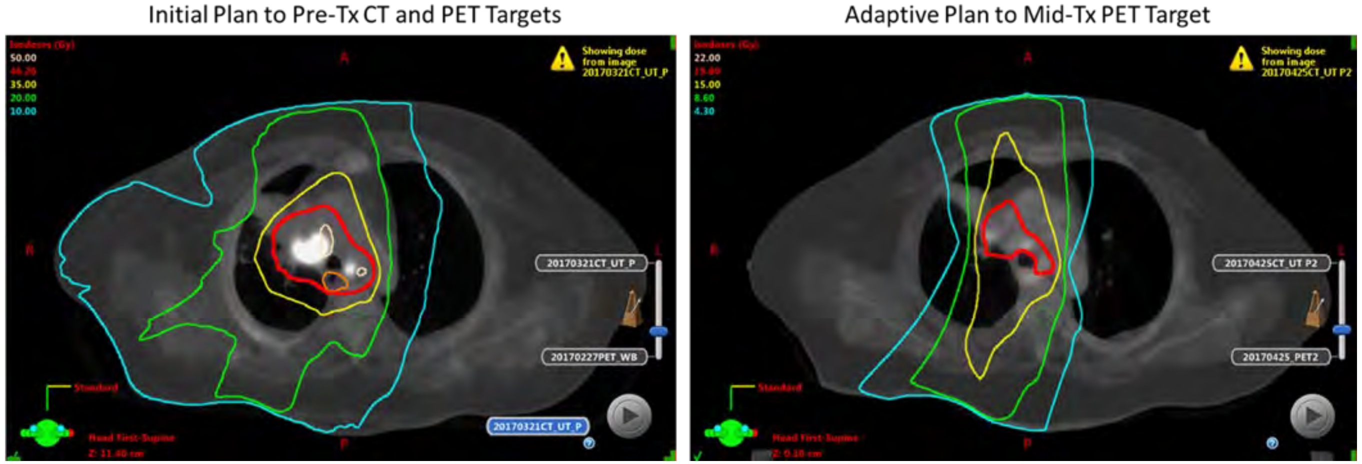
Independent test on 25 newly treated patients

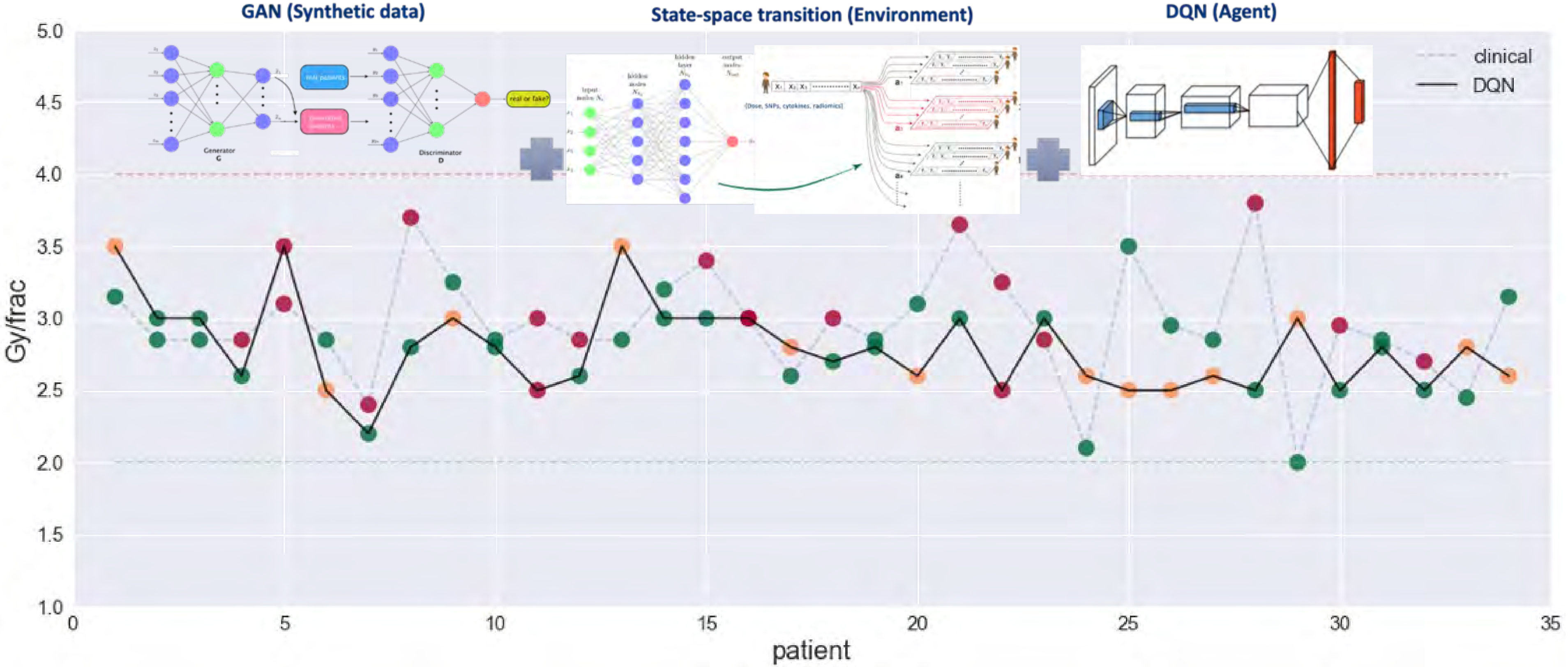
C-index (95%CI)	RP2	LC
NN-composite	0.692	0.721
NN-DVH	0.684	0.706
Lyman/log-logistic	0.588	0.573



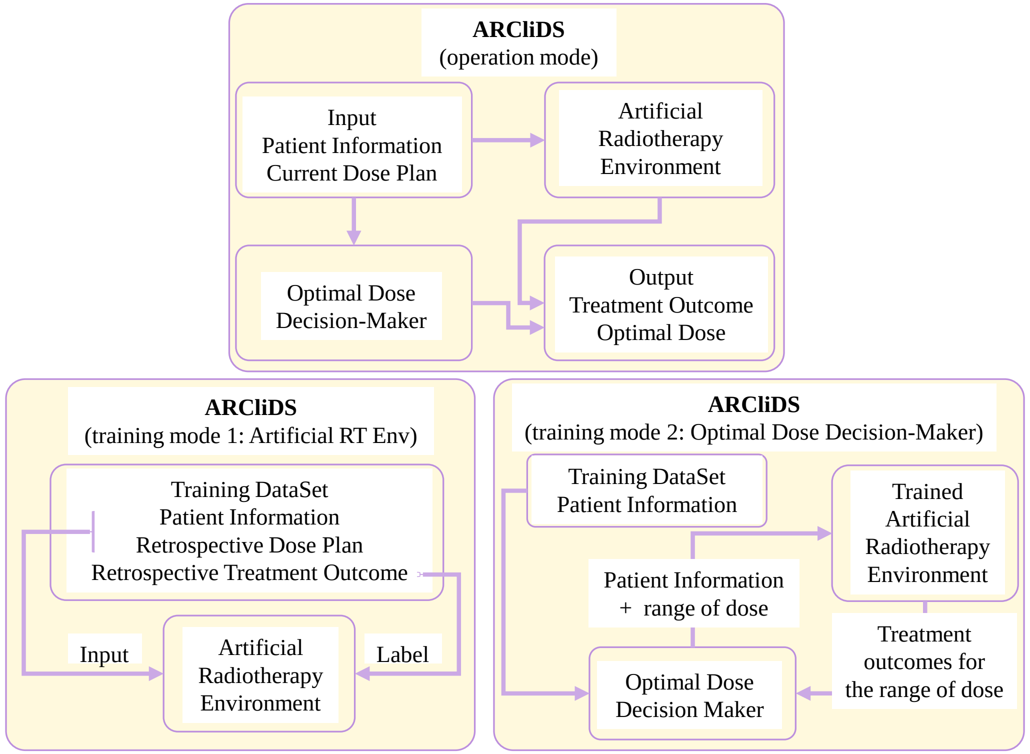
Cui et al, IJROBP, 2021

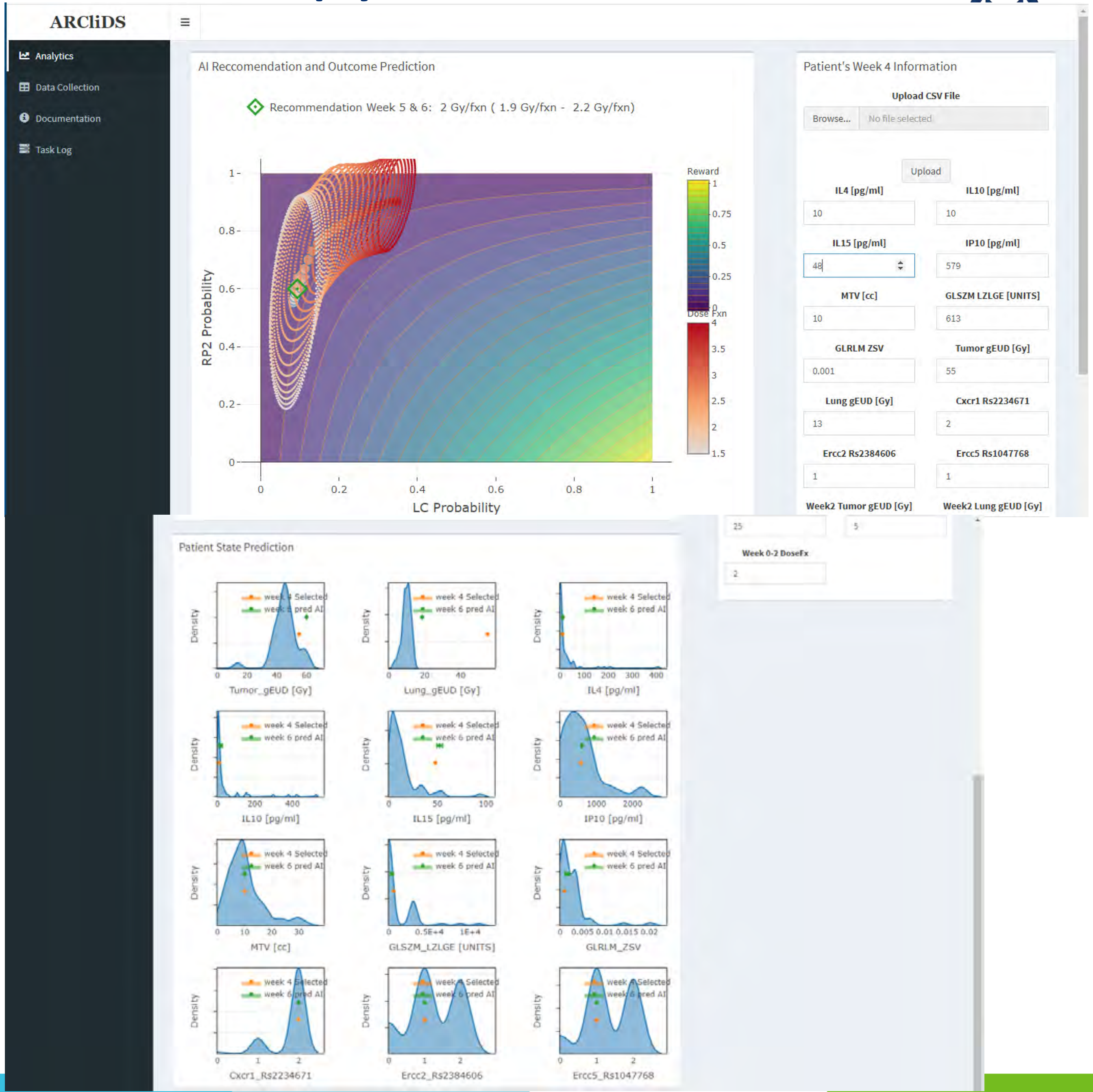
Adaptive Radiation Oncology Decision Making with Deep Learning





Software tools for Adaptive Radiotherapy Clinical Decision Support (ARClIDS)

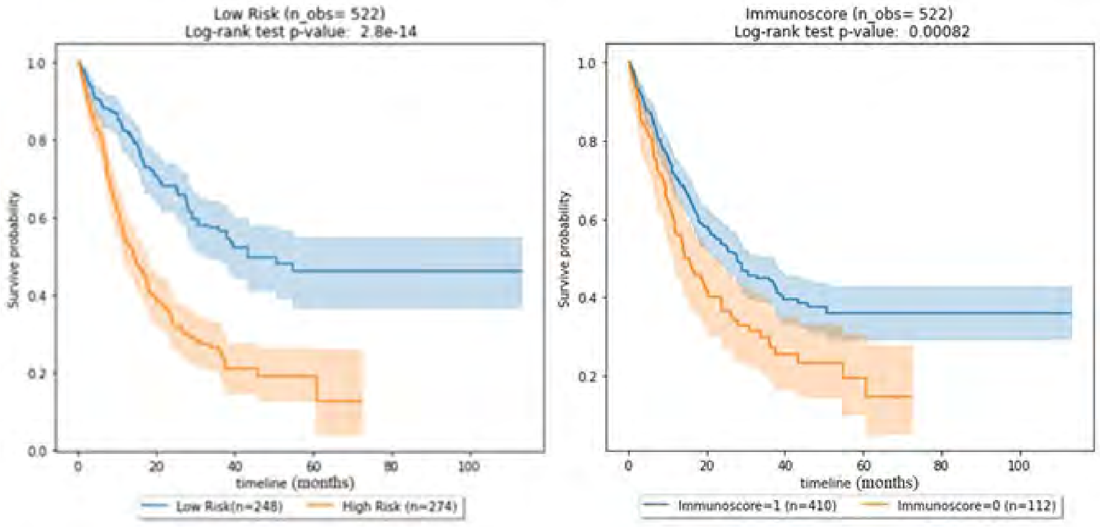




Niraula et al, patent pending

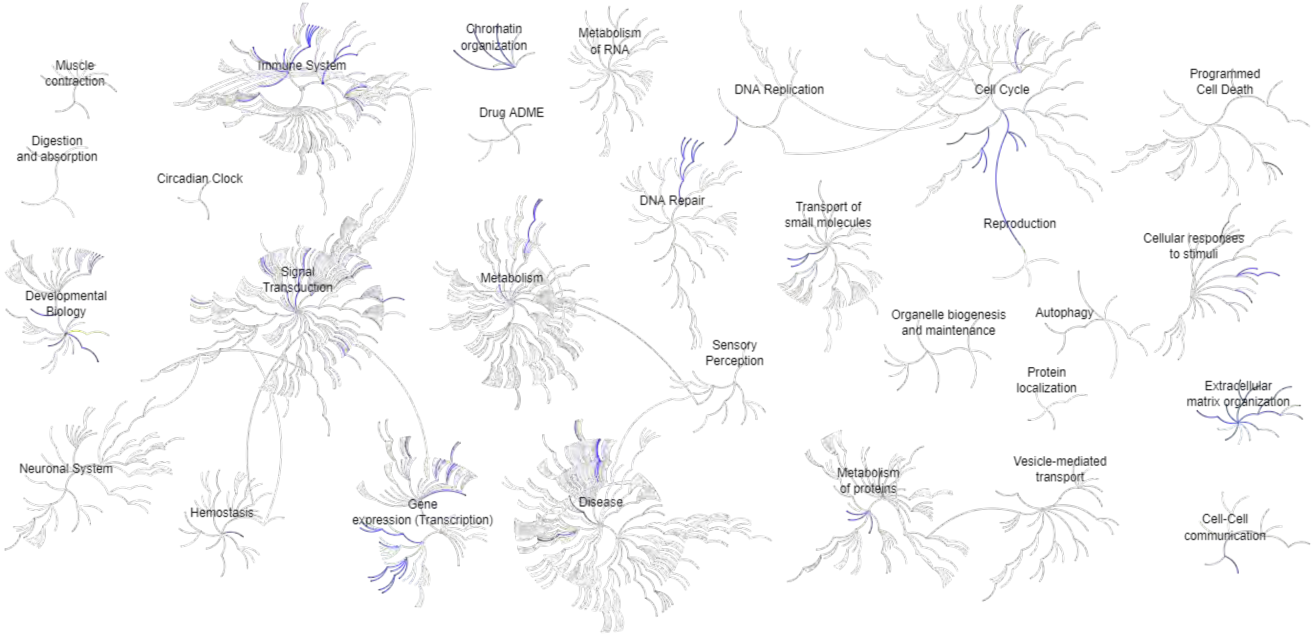
Deep learning for developing pan-cancer prognostic biomarkers for immunotherapy from RNAseq

- Collaboration with the Oncology Research Information Exchange Network® (ORIEN)– 18 centers
- Patients (n=522) with 4 primary cancer types
  - melanoma (n=125), renal cell carcinoma (n=149), non-small cell lung cancer (n=128) and head and neck cancer (n=120) treated with 6 immune checkpoint inhibitors
  - Deep learning: Auto-Encoder Survival Deep Network (AE-SDN) architecture



AE Deep

Immuno-score



Ghasemi, Tarhini, et al, ASCO

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.

ELSEVIER

Physica Medica

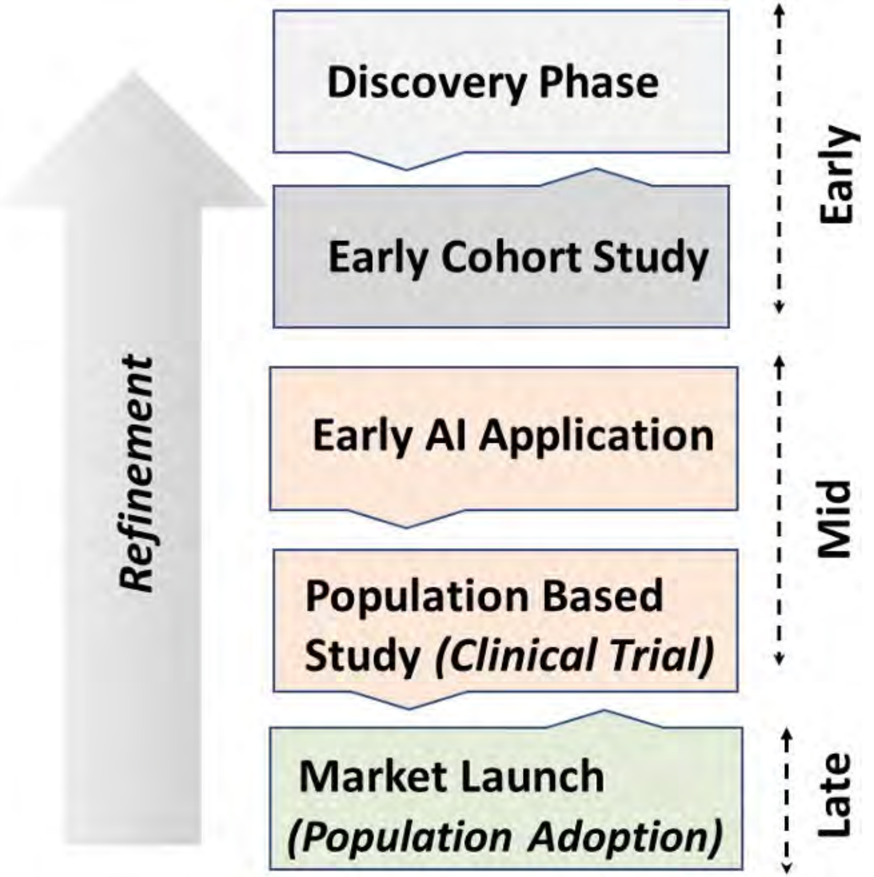
journal homepage: www.elsevier.com/locate/ejmp

Requirements and reliability of AI in the medical context

Yoganand Balagurunathan [a], Ross Mitchell [a,b], Issam El Naqa [a,*]

[a] Department of Machine Learning, H. Lee. Moffitt Cancer Center, Tampa, FL, USA
[b] Health Data Services, H. Lee. Moffitt Cancer Center, Tampa, FL, USA

- Diverse cohort of patient records for model training, achieved either through centralized or using federated/distributed learning models that uses silos of different data sources.
- Use of independent data cohort for testing, preferably in a distributed setting with diverse patient types.
- Transparency of deep network model architecture with confidence levels in its decisions.
- Ethically appropriate use of AI methods with some level of oversight.
- Assessment of reproducibility of AI models with test–retest type studies.
- Model transparent that discloses the architecture, data sets and trained weights for the network.
- Quality assurance program for implementation and continuous performance monitoring.



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

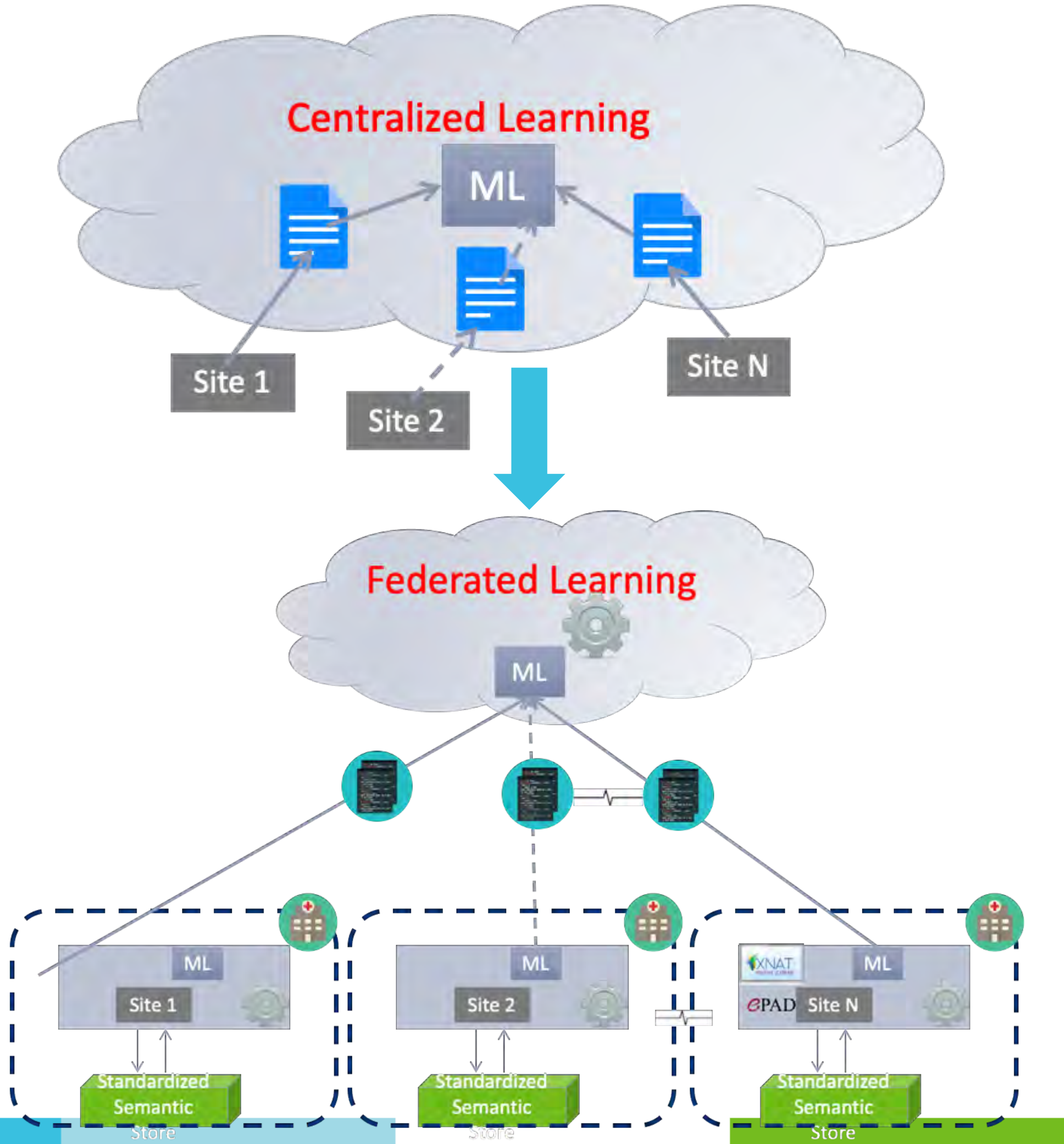
Machine learning and modeling: Data, validation, communication challenges

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

Ethical Challenge of Data Access

The company, Paige.AI, is one in a burgeoning field of start-ups that are applying artificial intelligence to health care, yet it has an advantage over many competitors: The company has an exclusive deal to use the cancer center's vast archive of 25 million patient tissue slides, along with decades of work by its world-renowned pathologists.



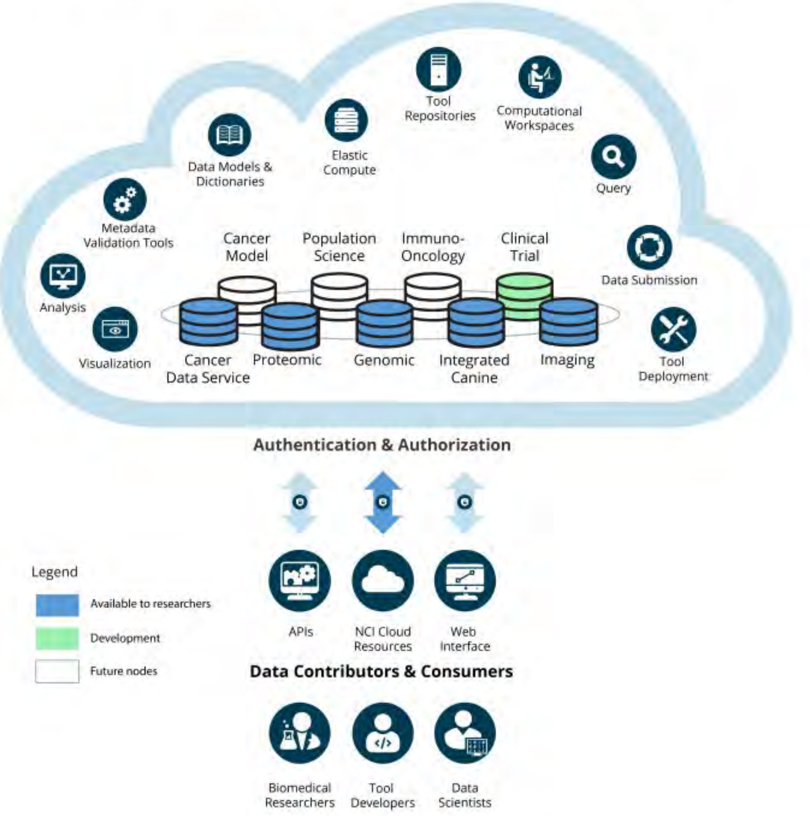
Jochems, IJROBP, 2017

Data Democratization!

MIDRC

MEDICAL IMAGING AND DATA RESOURCE CENTER.

NCI Cancer Research Data Commons (CRDC)

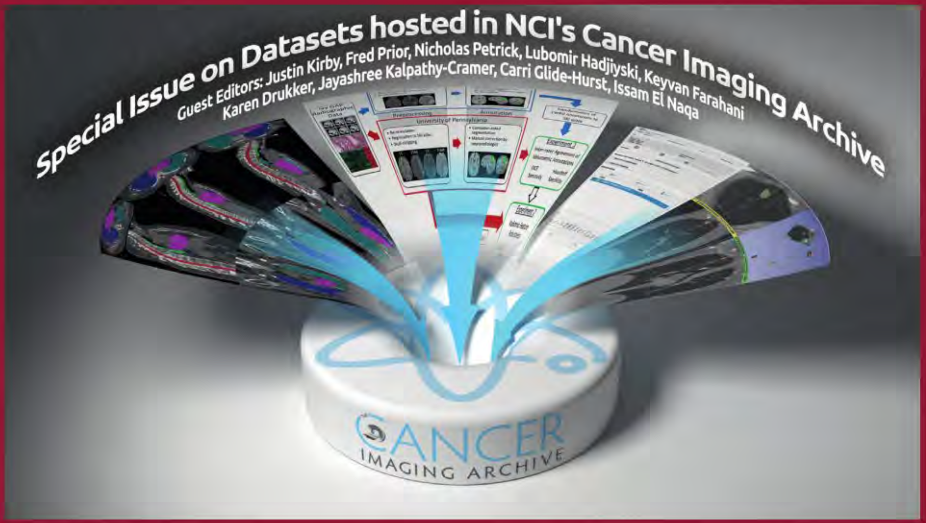


December 2020 | Volume 47, Issu

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.

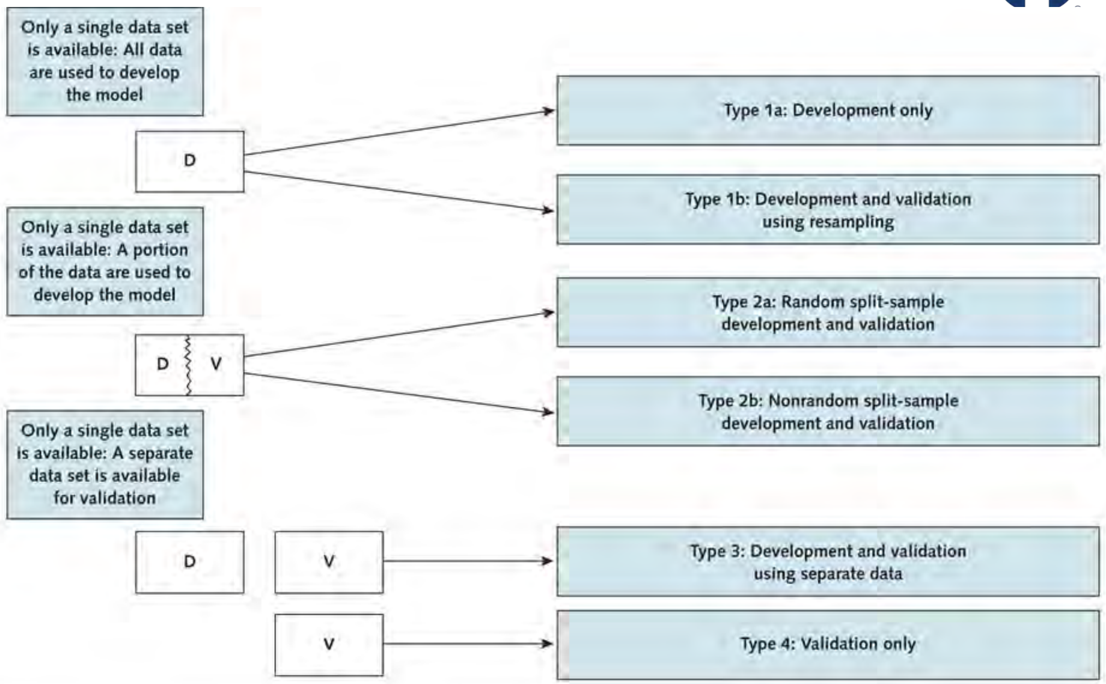
Medical Physics is an official journal of the AAPM, the International Organization for Medical Physics (IOMP), and the Canadian Organization of Medical Physicists (COMP).

AMERICAN ASSOCIATION *of* PHYSICISTS IN MEDICINE

WILEY

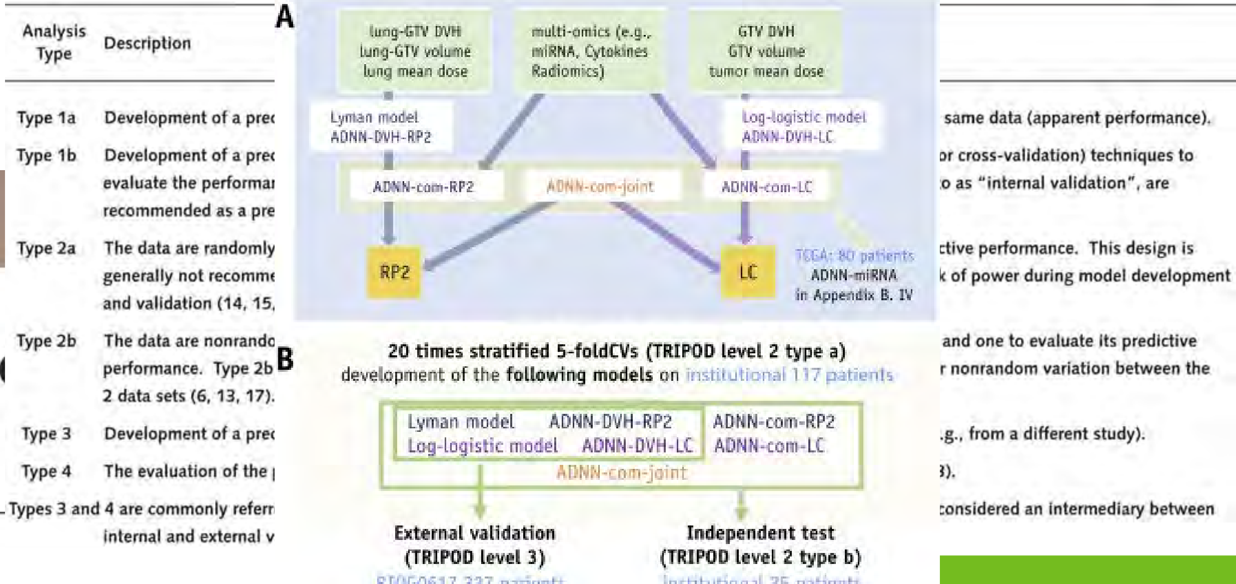
AI/ML Validation

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



Depending on the level of evidence

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



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

Check List for AI/ML in Medical Physics (CLAMP)

- Purpose and justification of AI/ML algorithm selection
- Dataset characteristics (acquisition, size, partitioning [3Ts: training, tuning, testing])
- ML methods
  - Optimization, loss function, augmentation, regularization
  - Performance metrics and evaluations (internal, external)
- Significance of results
  - Interpretation of ML performance
  - Clinical translation and actionability

TABLE 1 Checklist for AI in *Medical Physics* (CLAMP)

Indicate whether each section clearly summarizes or describes:	Checkboxes		
1. Abstract	**Yes**	**No**	**N/A**
a. Purpose, rationale, novelty or significance			
b. AI/ML methods and data type, dataset partitioning into training, validation (tuning), and test sets (include numbers used in training, validation, and test sets)			
c. Main results, including statistical analyses			
2. Introduction			
a. Purpose and justification of using AI/ML algorithm approach			
b. Contribution(s) of AI/ML to medical physics application			
c. Stage of development (e.g., pilot study, mature study)			
3. Materials			
a. Dataset characteristics including sample size and clinical acquisition sites			
b. Device(s) used for data acquisition (e.g., scanner makes), start-end dates of acquisition (or equivalent means with biotechnology generated data), and any data harmonization, augmentation, and enrichment strategies, or pre-processing are clearly described			
c. For imaging data: image or data acquisition modality, acquisition protocol, or parameter ranges are detailed			
d. For patient data: method to obtain the sample, representativeness of the population for the purpose of the study, IRB approval (or equivalent), and relevant patient demographics plus clinical variables such as prevalence(s) of disease(s) or lesion characteristics			
e. For phantom data: Type of phantom and method for generating phantom data			
f. Data composition appropriateness for AI/ML application			
g. Description of the "ground truth," that is, the reference standard, including the annotation process, level of subjectivity, and uncertainty			
h. Data partitioning into training, validation (tuning), and test sets including any criteria to mitigate bias and justification of sample sizes			
i. Final validation using public dataset or study dataset to be shared/made publicly available (desirable but not required).			
4.1 Methods: Machine learning algorithm			
a. Methodology in sufficient detail to allow replication, including model architecture, hyperparameters, inputs, dimensionality of the input (e.g., 2D or 3D images), pre-processing, output type and definition, and discretization/binning, if any.			
b. Training/optimization method including loss function, regularization approach, data imbalance mitigation process (if needed), measures to minimize overfitting and bias, and ablation studies, if any.			
c. AI/ML software code to be shared/made publicly available (desirable but not required).			
4.2 Methods: Performance and statistics			
a. Performance metric(s) including any postprocessing (such as scoring criteria, decision threshold, binning) of the AI/ML output.			
b. Method(s) to estimate the uncertainty (such as 95% confidence intervals) of the performance metric(s).			
c. Significance of the obtained results compared to the null hypothesis (if applicable) or compared to a suitable benchmark metric.			
d. Subgroup analyses for important subgroups (e.g., by age, lesion size).			
e. Demonstrative results for the training, validation (tuning), and test sets.			
5. Discussion			
a. Conclusions as supported by the results.			
b. Limitations of the study.			
c. Discussion/summary of innovation (algorithm or application), significance (clinical or scientific), and/or contributions to the field of medical physics.			

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]

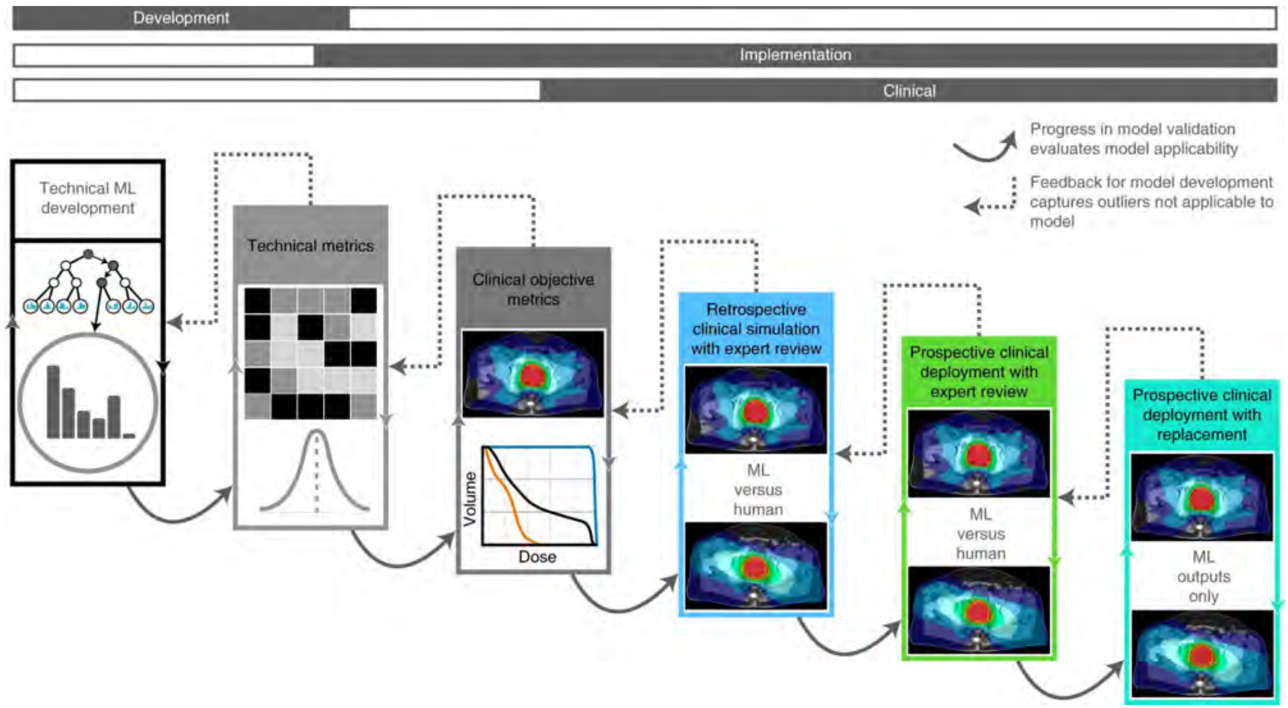
AI/ML in the real-world!

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

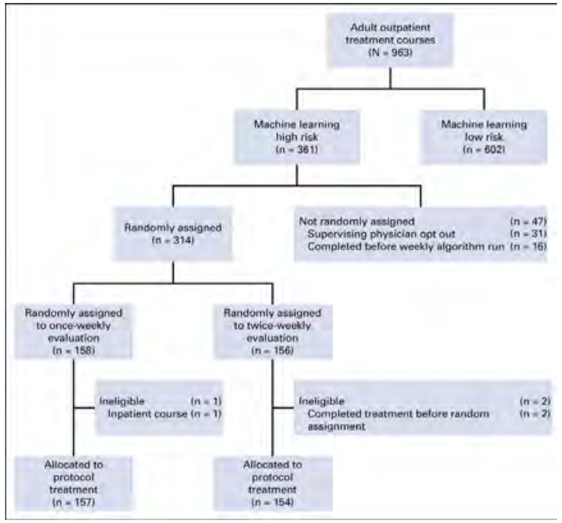


ORIGINAL REPORTS | Radiation Oncology

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

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

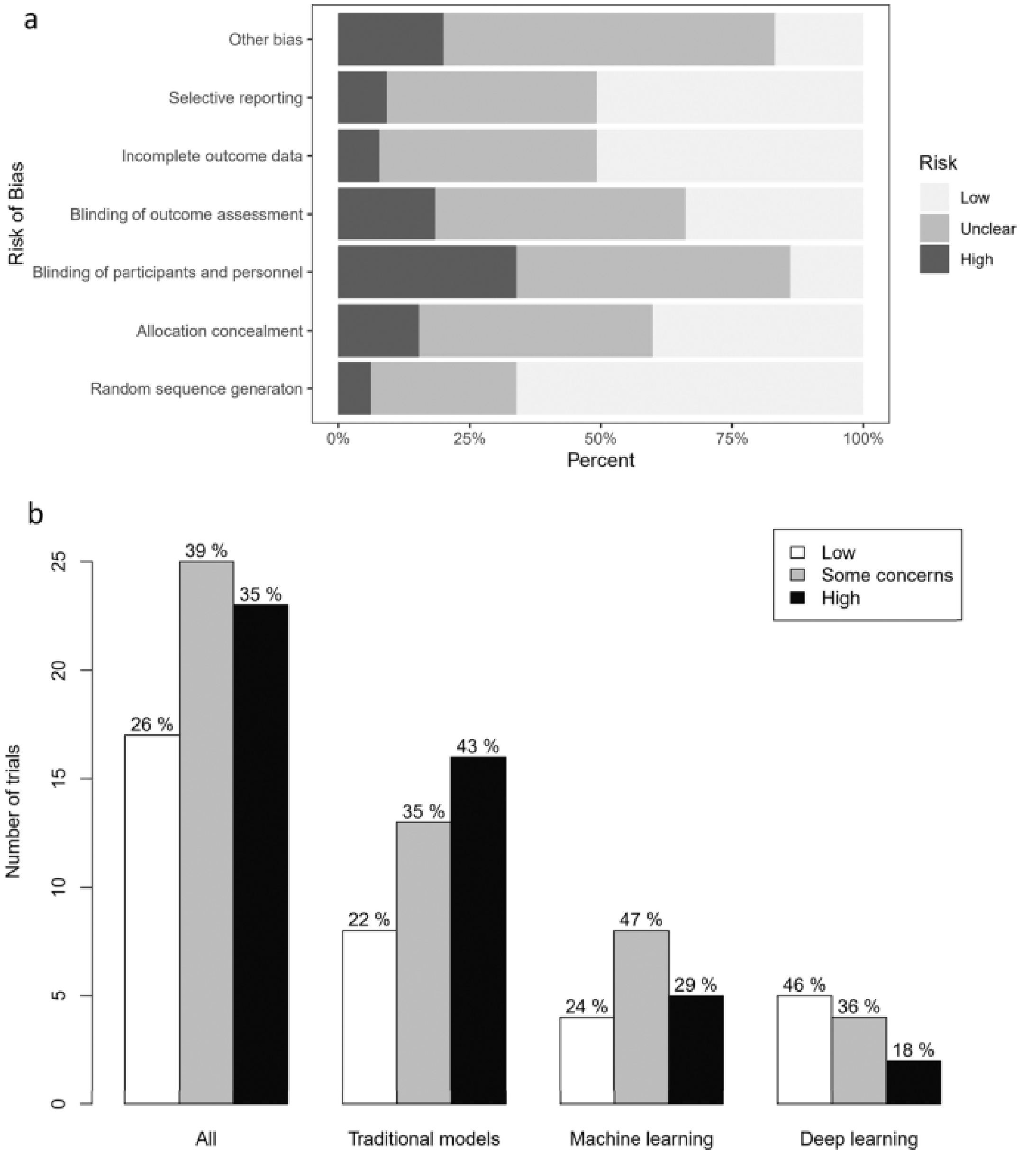
Randomized clinical trials with AI/ML/DL

Table 1 | Randomised trial registrations of deep learning algorithms

Trial registration	Title	Status	Record last updated	Country	Specialty	Planned sample size	Intervention	Control	Blinding	Primary outcome	Anticipated completion
ChiCTR-DDD-17012221	A colorectal polyps auto-detection system based on deep learning to increase polyp detection rate: a prospective clinical study	Completed, published	16 July 2018	China	Gastroenterology	1000	AI assisted colonoscopy	Standard colonoscopy	None	Polyp detection rate and adenoma detection rate	28 February 2018
NCT03240848	Comparison of artificial intelligent clinic and normal clinic	Completed, published	30 July 2018	China	Ophthalmology	350	AI assisted clinic	Normal clinic	Double (investigator and outcomes assessor)	Accuracy for congenital cataracts	25 May 2018
NCT03706534	Breast ultrasound image reviewed with assistance of deep learning algorithms	Recruiting	17 October 2018	US	Radiology	300	Computer aided detection system	Manual ultrasound imaging review	Double (participant and investigator)	Concordance rate	31 July 2019
NCT03840590	Adenoma detection rate using AI system in China	Not yet recruiting	15 February 2019	China	Gastroenterology	800	CSK AI system assisted colonoscopy	Standard colonoscopy	None	Adenoma detection rate	1 March 2020
NCT03842059	Computer-aided detection for colonoscopy	Not yet recruiting	15 February 2019	Taiwan	Gastroenterology	1000	Computer aided detection	Standard colonoscopy	Double (participant, care provider)	Adenoma detection rate	31 December 2021
ChiCTR1800017675	The impact of a computer aided diagnosis system based on deep learning on increasing polyp detection rate during colonoscopy, a prospective double blind study	Not yet recruiting	21 February 2019	China	Gastroenterology	1010	AI assisted colonoscopy	Standard colonoscopy	Double	Polyp detection rate and adenoma detection rate	31 January 2019
ChiCTR1900021984	A multicenter randomised controlled study for evaluating the effectiveness of artificial intelligence in improving colonoscopy quality	Recruiting	19 March 2019	China	Gastroenterology	1320	EndoAngel assisted colonoscopy	Colonoscopy	Double (participants and evaluators)	Polyp detection rate	31 December 2020
NCT03908645	Development and validation of a deep learning algorithm for bowel preparation quality scoring	Not yet recruiting	9 April 2019	China	Gastroenterology	100	AI assisted scoring group	Conventional human scoring group	Single (outcome assessor)	Adequate bowel preparation	15 April 2020
NCT03883035	Quality measurement of esophago-gastroduodenoscopy using deep learning models	Recruiting	17 April 2019	China	Gastroenterology	559	DCNN model assisted EGD	Conventional EGD	Double (participant, care provider)	Detection of upper gastrointestinal lesions	20 May 2020
ChiCTR1900023282	Prospective clinical study for artificial intelligence platform for lymph node pathology detection of gastric cancer	Not yet recruiting	20 May 2019	China	Gastroenterology	60	Pathological diagnosis of artificial intelligence	Traditional pathological diagnosis	Not stated	Clinical prognosis	31 August 2021

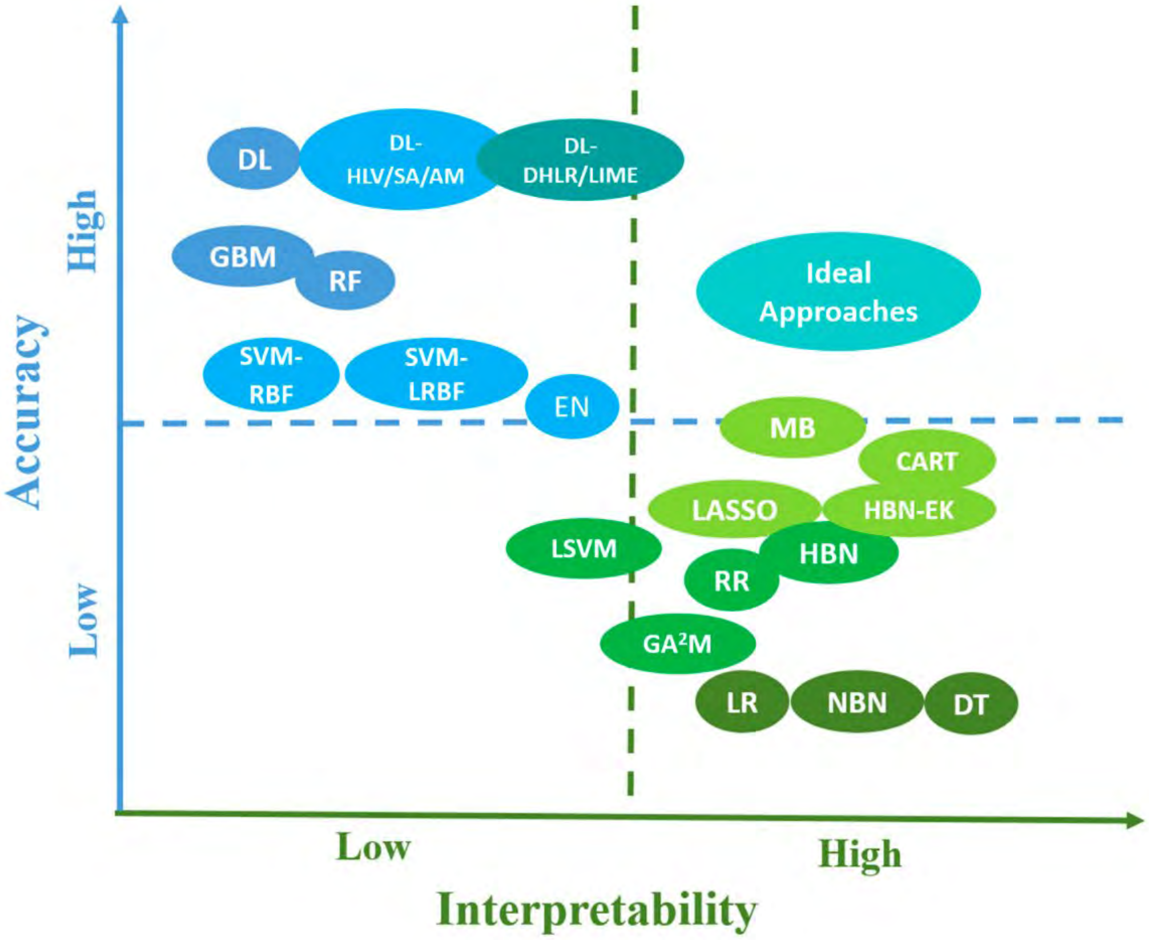
AI=artificial intelligence; CSK=commonsense knowledge; DCNN=deep convolutional neural network; EGD=esophagogastroduodenoscopy.

Nagendran, BMJ, 2020



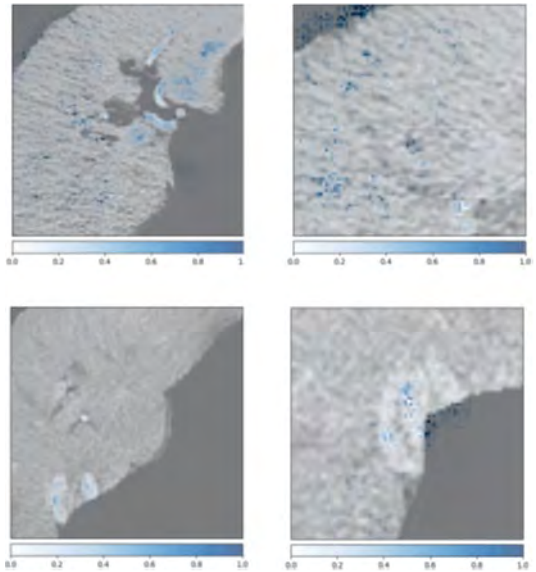
Zhou, Dig. Med., 2021

ML/DL Interpretability



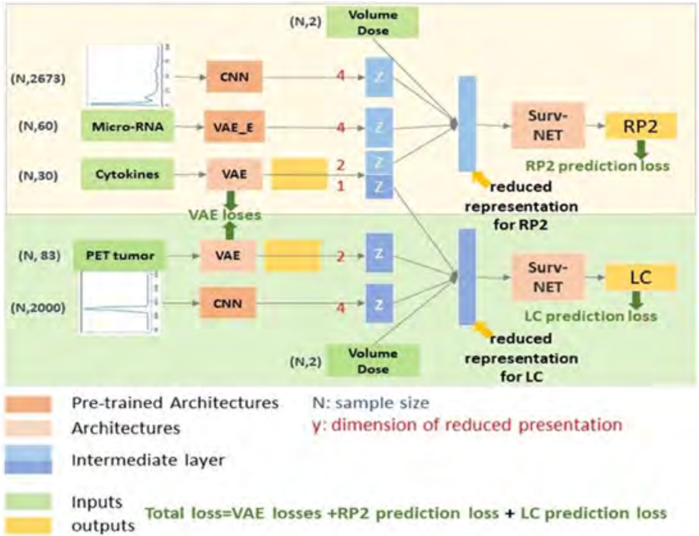
Luo, BJR-O, 2019

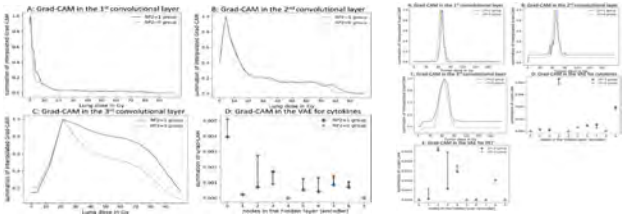
Radiomics Interpretability for Liver Cancer (Grad-CAM)



Wei et al, Physica Medica, 2021

Multi-omics interpretability for Lung Cancer





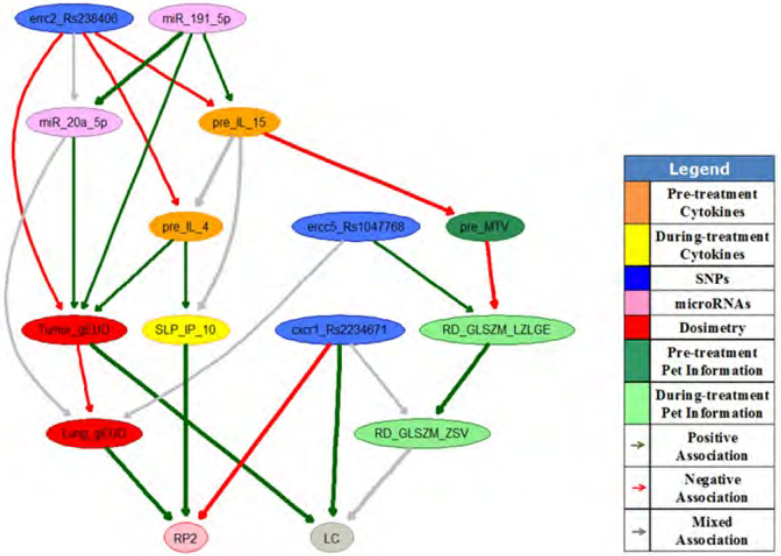
Cui et al, IJROBP, 2021

Intelligence augmentation (IA) instead of AI

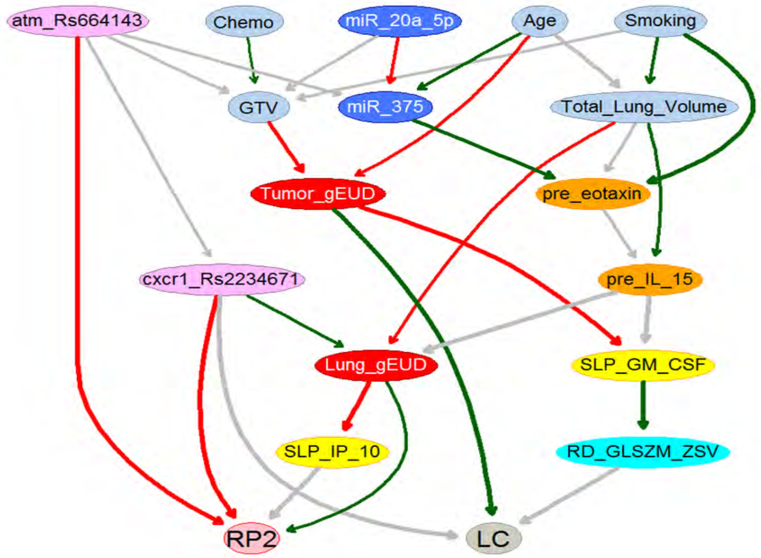
Human in the loop!

Figure 1. A "Fundamental Theorem" of informatics.
(C. Friedman)

Data-driven ML



Human + Data-driven ML

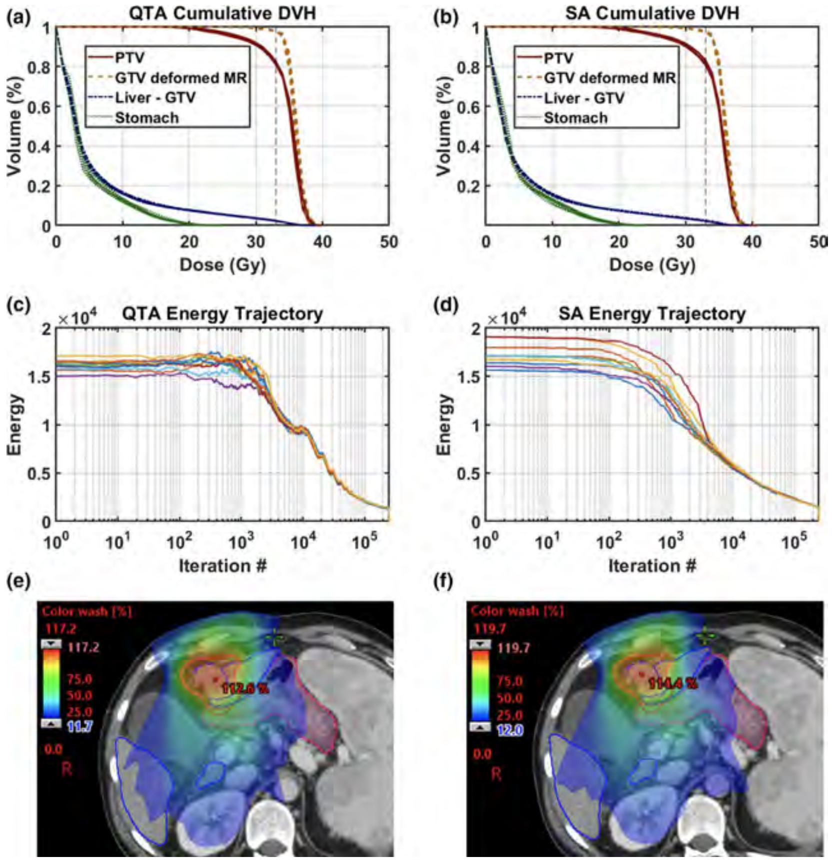


Tighter CIs but similar predictions!

Luo, Physica Medica (Editor Choice), 2021

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

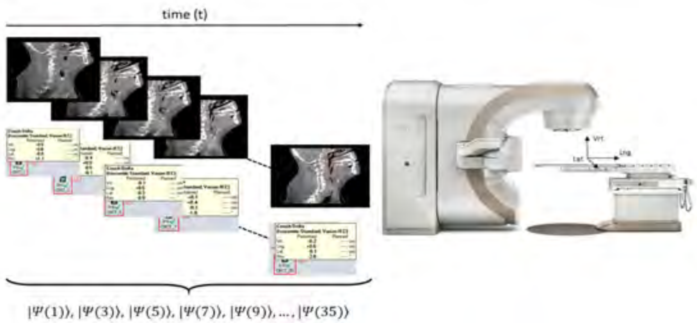
Treatment Planning

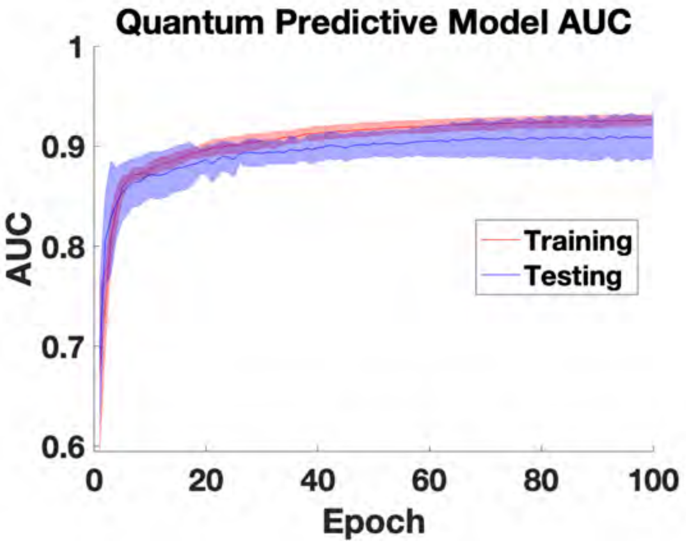


Algorithm	Width Function	Mean Convergence Rate (s)
SA	N/A	1157 ± 154.5
QTA	Hybrid	757.8 ± 162.3
QTA	MOCVD	622.1 ± 103.2
QTA	Sinusoid	526.2 ± 126.1

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

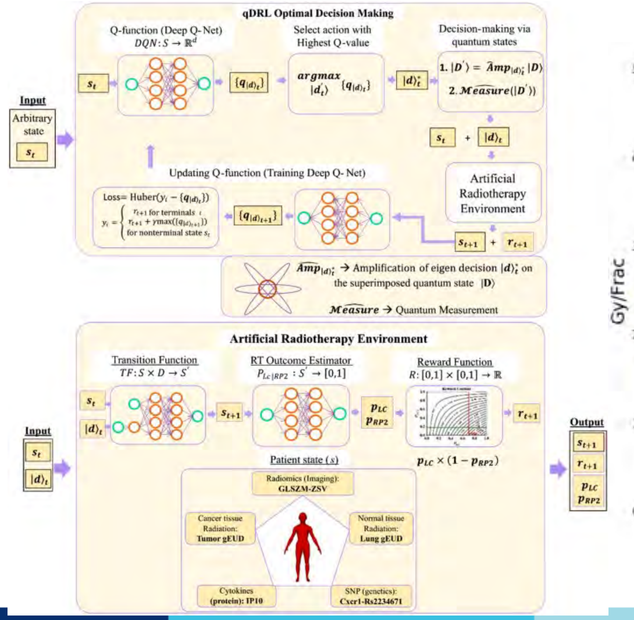
Image-guided radiotherapy

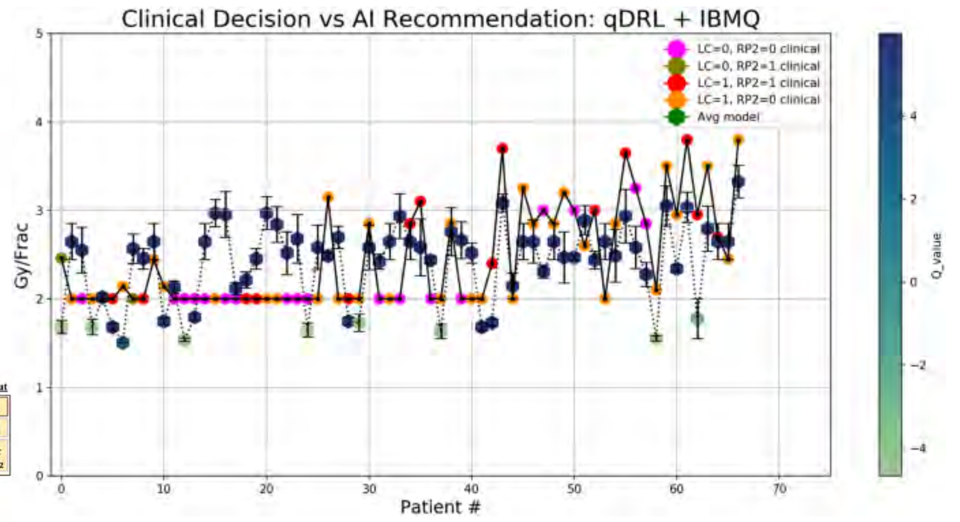




Pakela et al, PMB, 2021

Clinical Decision support





Niraula et al, Nature Sci Rep, 2021

Machine Learning at Moffitt

(@ml4onco)


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

ML Strategic Priorities @ Moffitt

Faculty

Strategic Priority	First Year				Second Year				Third Year				Fourth Year				Fifth Year			
	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4	Q1	Q2	Q3	Q4
1. Integration of ML into MCC research and clinical care																				
1.1 Develop a robust and secure ML infrastructure that also leverages existing MCC resources																				
1.2 Convert clinical care data into research data including linkage of unstructured data using NLP methods																				
1.3 Establish ML working group for R&D (Machine Learning League [MLL])																				
2. Establish translational ML research program in priority areas																				
2.1 Multimodality radiological and pathological imaging for diagnostic and outcomes																				
2.2 Information retrieval and annotation with natural language processing (NLP)																				
2.3 Outcome modeling and decision support by longitudinal integration of pan-omics data and using PROs for retrospective and prospective studies																				
2.4 Molecular and computational biology and in silico trial designs																				
3. Establish basic ML research programs in priority areas																				
3.1 Visual analytics, explainable and interpretable ML/AI																				
3.2 Automated ML architectures and evolutionary learning																				
3.3 Physics-based quantum ML, hybrid systems, and stochastic processes																				
4. Develop team science initiatives																				
4.1 Program project or center of excellence to address clinical ML role																				
4.2 Program project or biotechnology resource to address basic science ML role																				
5. Develop residency/training programs																				
5.1 PhD/Residency programs in ML for oncology																				

Staff (ML Engineers)

Moffitt.org/MachineLearning

Synopsis of Faculty Research Areas

Dr. Balagurunathan's research is focused on understanding the physiology of the tumor and its relationship to the underlying genome. His interests include data integration from various modalities (radiology, pathology, genome) to improve clinical decision support, His disease foci are prostate cancer, lung cancer and B-cell lymphomas.

Dr. Karolak's background includes applications and development of tools from the fields of computational and biophysical chemistry, structural biology, mathematical oncology, machine learning and information theory. Her interests focus on understanding cancer development, progression, and variability in the response to treatment using models that can be translated into the clinic.

Dr. Luo's research focuses on machine learning, systems informatics and their application to health outcomes, decision support, interpretable and credible models at both the individual and community levels for precision medicine, health equity and healthcare quality.

Dr. Thieu has been pursuing research in natural language processing and artificial intelligence with application in healthcare, education, and bioinformatics. His work involved standardization of mobility terminology from clinical notes, high throughput text mining, lexical complexity and language generation, and computer-assisted coding for medical billing.

Dr. Rasool focuses on developing robust and explainable machine learning models for various medical imaging and signal processing applications. He is also interested in exploring machine learning paradigms that can tackle datasets from multiple scales and learn to answer clinically relevant questions. Such models will be robust to day-to-day changes in the input data and must explain their decisions to users, thus building trustworthiness.

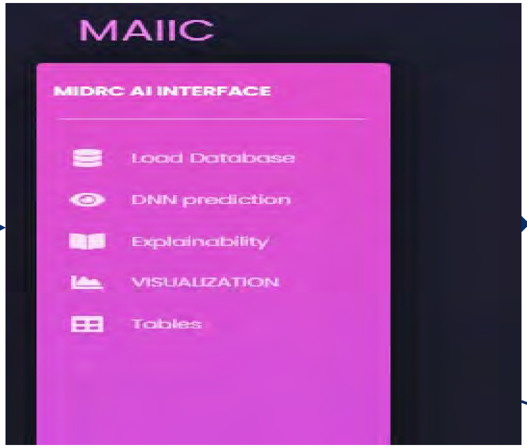
Machine and Federated Learning Infrastructure (API)

ta Platform/Covid Images

GEN 3
Data Commons

Input

Django

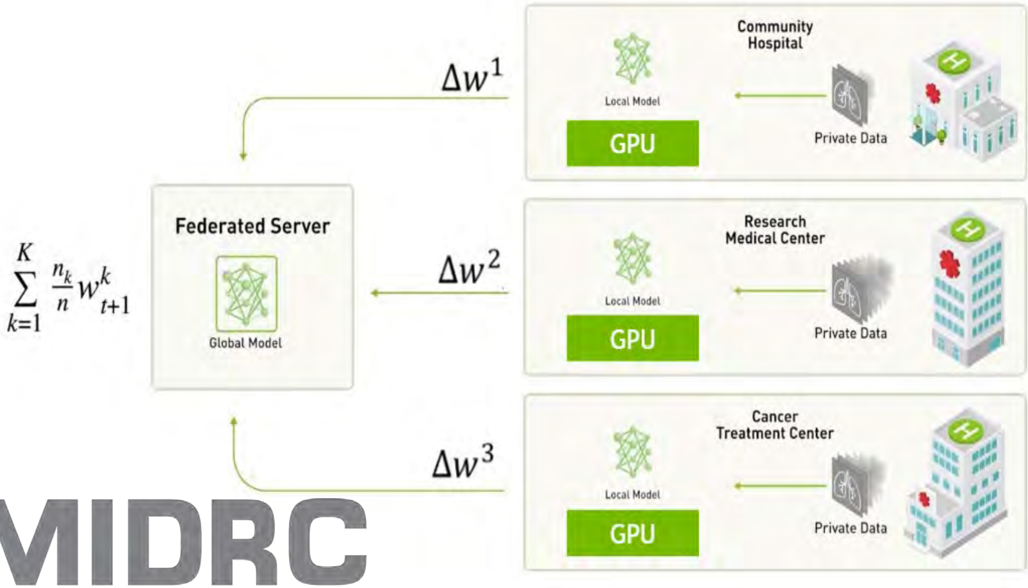


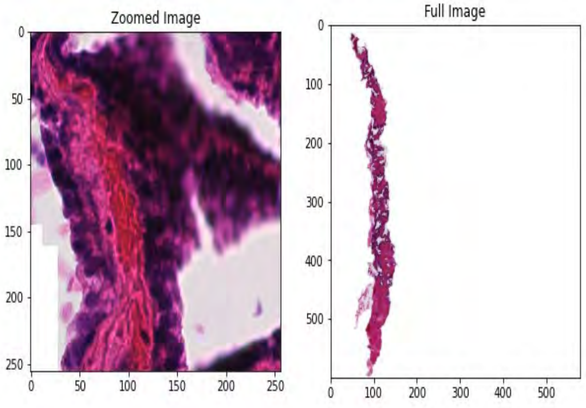
ML/DL

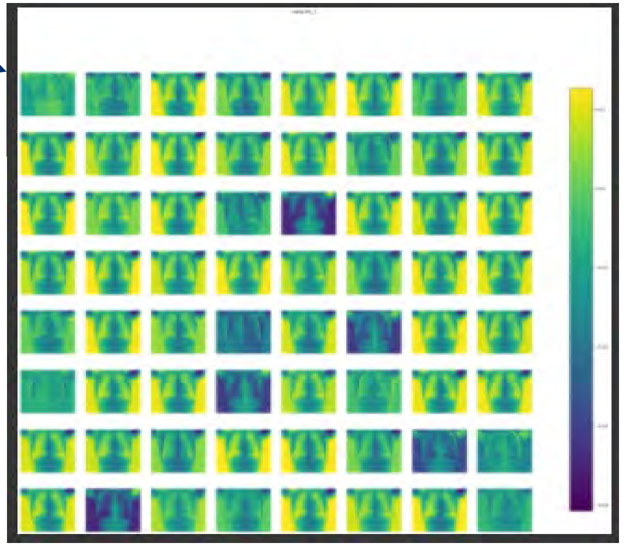
High Performance Computing (DGX/AWS EC2)

aws
Amazon EC2

Visualization







MIDRC
MEDICAL IMAGING AND DATA RESOURCE CENTER.

Community outreach: Machine Learning League

- Advance awareness and application of Machine Learning, Deep Learning, and Artificial Intelligence (AI) across the multiple disciplines in cancer research by hosting biweekly discussions on current machine learning research and tools

ihmc FLORIDA INSTITUTE FOR HUMAN & MACHINE COGNITION

CORE SCIENTIFIC

NVIDIA

capsico HEALTH

UNIVERSITY OF SOUTH FLORIDA

Take home Messages

- Artificial intelligence/machine learning offers new opportunities to develop better understanding of oncology processes and improve its workflow
- Current boom in AI/ML is driven by growth in patient-specific information (Big data) and advances in computer hardware/software resources
- AI/ML will touch every aspect of oncology from planning to decision making and we should be ready to embrace it
- To overcome current barriers in AI/ML for healthcare emerging methods include including visualization (Grad-CAM),behavioral science (human-in-the loop), and physics-based (quantum computing) techniques
- Collaboration between stakeholders (data scientists, biologists, clinical practitioners, & vendors) will allow for safe and beneficial application of AI in biomedicine and oncology