School on Medical Physics for Radiation Therapy: Dosimetry, Treatment Planning and Delivery for Advanced Applications

11 - 22 September 2023 An ICTP Meeting Trieste, Italy

Further information: http://indico.ictp.it/event/10205/ smr3871@ictp.it





Introduction to Artificial Intelligence

Michele Avanzo Centro di Riferimento Oncologico IRCCS Aviano (PN), Italy, mavanzo@cro.it

Historical background

• Alan Turing (London, 1912 – Manchester, 1954)





Historical background



Tape is infinite in both directions. Only finitely many nonblank cells at any time.

Turing machine

Vol. LIX. No. 236.]

[October, 1950

MIND

A QUARTERLY REVIEW

OF

PSYCHOLOGY AND PHILOSOPHY

I.—COMPUTING MACHINERY AND INTELLIGENCE

By A. M. TURING

1. The Imitation Game.

I PROPOSE to consider the question, 'Can machines think?' This should begin with definitions of the meaning of the terms 'machine' and 'think'. The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous. If the meaning of the words 'machine' and 'think' are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, 'Can machines think?' is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

The new form of the problem can be described in terms of a game which we call the 'imitation game'. It is played with three people, a man (A), a woman (B), and an interrogator (C) who may be of either sex. The interrogator stays in a room apart from the other two. The object of the game for the interrogator is to determine which of the other two is the man and which is the woman. He knows them by labels X and Y, and at the end of the game he says either 'X is A and Y is B' or 'X is B and Y is A'. The interrogator is allowed to put questions to A and B thus:

C: Will X please tell me the length of his or her hair ? Now suppose X is actually A, then A must answer. It is A's 28 433

Historical background

 L'Al was born at the conference at Dartmouth College (Hanover, NH, USA) in 1956 where the the term "artificial intelligence" was coined



Artificial Intelligence

• Artificial Intelligence (AI) is a group of technologies aiming at giving to machines abilities requiring human intelligence



Machine Learning

Machine learning (ML) is the discipline which aims at giving machines the ability to learn to perform tasks, without being explicitly programmed to conduct these tasks (Arthur Samuel 1959)



Supervised Machine learning

• The ML algorithm is trained in a known dataset in order to yield a model which can make predictions on an unknown dataset



Classification

• Dataset of iris flowers (by Roland Fisher, 1936)

Iris setosa



Iris versicolor



Iris virginica



V Sepalo

Decision Trees





https://victorzhou.com/blog/gini-impurity/

Metrics for performance



- A highly sensitive test is preferable for screening, as it gives less false negatives
- A highly specific test gives less *false positives*, so it is useful to confirm a pathology

Curva Receiver Operating Characteristic (ROC)

- Consider a classifiers that, instead of a binary output, provides the probability of being positive
- ROC curve shows true positives vs false positives for different thresholds of the continuous index





Statistics vs machine learning



Machine learning algorithms <u>analyze a dataset and then extract correlations</u>, a reversal of traditional data analysis, where the hypothesis is chosen first and then the data queried to test the hypothesis.

• For the same reason, ML <u>needs large datasets</u> to learn, rather than just a statistically relevant sample.

Neural networks



FIGURE 1

Mccullough e Pitts (1943) describe the brain using an abstract model of neural network

Harold Hebb (1949): "Cells that fire together wire together"

Neural networks

• Perceptron by Frank Rosenblatt (1959)



Cathegories of machine learning



Physica Medica Volume 83, March 2021, Pages 221-241

Al applied to imaging

• Lodwick, 1963





Bayesian Bone Tumor Diagnosis

Michael L. Richardson, M.D. University of Washington Department of Radiology

rick in Radiol Clin N Am 1965:3:487-497 and in Radiology 1963;80:273-275.

sental in nature, and was designed to be used as a teaching aid for the education of medical students and radiology residents. It should NOT be used for diagnosis or treatment planning in actual patients.

Please input the following information about your bone tumor

What is the patient's age?

What is the maximum diameter of the tumor in cm?

What ture of home is involved?	small or flat \bigcirc
what type of oone is involved:	long 💿
	epiphysis 🔾
Where is the factor factor is relation to the abovic?	physis 🔾
vinere is the festion focated in relation to the physis;	metaphysis 💿
	diaphysis 🔾
	osteoid 💿
What type of calcified matrix is present?	cartilaginous 🔾
	none 🔾

Calculate differential diagnosis

Radiomic variables or "features"

Histogram

 $kurtosis = \frac{\frac{1}{N} \sum_{i} \left(X(i) - \overline{X} \right)^{4}}{\left(\frac{1}{N} \sum_{i} \left(X(i) - \overline{X} \right)^{2} \right)^{2}}$

 $entropy = \sum_{i=1}^{n} \left(P(i) \log_2 P(i) \right)$

Shape

 $compactness 2 = 36\pi \frac{A^2}{V^3}$

Texture

cluster shade =
$$\sum_{i,j} (i + j - 2\mu)^3 * P(i, j)$$

coarseness = $\frac{1}{\varepsilon + \sum_i P(i)s(i)}$

Avanzo et al. Phys Med 38 (2017) 122-139







• Analizzare e capire le variabili selezionate e perchè



post-RT











EGFR	СТ	Volume	Radius Std	Shape SI6	Gabor_Energy-	Gabor_Energy-	Laws_Energy	Laws_Ene
status	acquisition				dir135-w3	dir45-w9	-10	rgy-13
EGFR positive	Baseline (Fig 1-a)	7766.5	1.522	0.145	5337.9	419770.4	475.2	1369.6
	Followup (1- b)	7195.8	1.657	0.151	4043.5	327365.1	512.0	1352.9
	Change	-570.6	0.135	0.006	-1294.4	-92405.3	36.8	-16.6
	Baseline (Fig 1-c)	3502.4	1.422	0.173	11601.7	419578.9	367.7	353.9
Wild type	Followup (1- d)	4522.8	1.251	0.165	10605.5	361191.5	326.3	349.3
	Change	1020.4	-0.171	-0.009	-996.2	-58387.4	-41.5	-4.5

Aerts et al. Sci Rep 6 (2016) 33860

Radiomic profile



Prognosis

ML in radiomics

- High number of often redundant variables:

29 shape + (174-shape=145) * (filters LoG, wavelet...) * (imaging modalities: PET, CT, MRI)

- Complex, non-linear dependencies on outcome



Histogram (1st Order)

Shape

Textural (2nd order)



II. Contouring

Patients

III. Pre- Processing, filtering

IV. Features



Characterization of lesion

- Molecular subtype of breast cancer
- stepwise feature selection and linear discriminant analysis



		-	-	
	-			

	ER Positive Case (a)	ER Negative Case (b)
Cancer Subtype	Luminal A	HER2-enriched
MRI CEIP Size (Effective Diameter) Range [7.8 54.0]	12.9 mm	23.8 mm
MRI CEIP Shape (Irregularity) Range [0.40 0.84]	0.452	0.602

https://www.nature.com/articles/npjbcancer201612



Classification task	Number of tumors	MRI phenotype	Mean value (s.d.) positive versus negative	P-value	Significance level ($\alpha^T = 0.05$)
ER+ versus ER-	91 (77 vs. 14)	Effective diameter	17.6 mm (5.6) vs. 24.8 mm (10.4)	0.001 ^a	0.0167
		Irregularity	0.61 (0.11) vs. 0.65 (0.11)	0.23	0.05
		Entropy	6.40 (0.11) vs. 6.45 (0.09)	0.08	0.025
PR+ versus PR-	91 (72 vs.19)	Effective diameter	18.0 mm (5.6) vs. 21.6 mm (10.5)	0.14	0.025
		Irregularity	0.61 (0.11) vs. 0.63 (0.10)	0.43	0.05
		Entropy	6.39 (0.11) vs. 6.45 (0.07)	0.03	0.0167
HER2+ versus HER2-	91 (19 vs. 72)	Effective diameter	18.4 mm (5.7) vs. 18.8 mm (7.3)	1.0	0.05
		Irregularity	0.59 (0.12) vs. 0.62 (0.11)	0.36	0.0167
		Entropy	6.41 (0.10) vs. 6.40 (0.11)	0.93	0.025
TN versus others	91 (11 vs. 80)	Effective diameter	17.8 mm (5.6) vs. 25.6 (11.5)	0.006 ^a	0.0167
		Irregularity	0.60 (0.11) vs. 0.68 (0.09)	0.03	0.025
		Entropy	6.40 (0.10) vs. 6.45 (0.10)	0.13	0.05

Abbreviations: ER, estrogen receptor; HER2, human epidermal growth factor receptor 2; MRI, magnetic resonance imaging; PR, progesterone receptor; TN, triple negative.

^aIndicates statistical significance was achieved after correction for multiple comparisons.

Advantages of radiomics

- Standardization
- Interpretability
- Less computational resources needed





IBSI digital phantom

Shape -

Histogram (1st Order)

Textural (2nd order)



Neural networks and deep learning

• Deep learning was introduced to model outputs with more complex, non-linear dependencies from input



https://www.pnas.org/content/116/4/1074

Convolutional networks

• Layers to extract image features through convolution with learable filters ("kernels") Vertical





Sobel filter

+1

0

Horizontal Sobel kernel



https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5479722/

Other layers

- Pooling layers reduce size of feature map e.g. by calculating maximum or average in a group of voxels
- Dropout layers zero random inputs to prevent overfitting
- Softmax layers calculate probability of classes







Dropout



Convolutional neural networks (CNN)



Classification of objects



I. Image acquisition

Chenyang Shen et al 2020 Phys. Med. Biol. 65 05TR01



Segmentation by CNN

3x256x256

image





Rete neurale "U-net"



Contouring



Example: tumor detection and segmentation

- Patient data from public databases, IVO and prostate X
- Comparison against PIRADS 4+



Figure 6: ROC curve of the model for significance criterion $GGG \ge 2$, evaluated at the lesion level (left) and the patient level (right). For comparison, triangular marks represent the radiologist-assigned pre-biopsy PI-RADS.

Figure 8: Output of the model evaluated on three ProstateX test patients. First image from the left shows the GT on the T2; the rest show the output predictions of the model on different sequences (from left to right: T2, b800, ADC, K^{trans} .

Pellicer-Valero et al, arXiv:2103.12650

CNN "Variational autoencoders"

- **Encoder**: neural network producing a representation of data
- **Decoder**: neural network trying to reconstruct the original image from the representation of data



Cuiet al.: ML/DL for medical physicists , Medical Physics, 2020

Artifact correction



Superresolution



Reconstruction



Generative Adversarial Network







Signal analysis



Figure 6.1: A plot of the 30 minute predictions of LSTM-2 (in red) versus the target values (in blue) for one of SP92's test days. The grey area around the prediction line shows the standard deviation of the output distribution.

http://publications.lib.chalmers.se/records/fulltext/251317/251317.pdf

Wang et al. DOI: 10.1109/ACCESS.2018.2869780

Availability of Open Source ML/DL tools



Machine Learning

NomeOpen
Source?InterfaceVPython, C++VPython, C++K KerasYYPython, RVPython (Keras), C/C++,
Java, JavaScript, R,IJul
ia, Swift

Deep Learning

Distributions of python:



Web interface:

Data repositories:

Overfitting

• ML algorithms are highly opportunistic: they learn what boosts their performance during training, without worrying for generalizability



Covid diagnosis challenge

- A popular open-source dataset, COVIDx, was used to develop many deep learning tools to diagnose COVID from chest radiographies
- It was found that these models performed poorly on other datasets



Prediction probability: 0.938

Figure 5: Grad-CAM saliency map generated by DarkCovidNet COVIDx test data predictions. This is a correct prediction of COVID-19, with a prediction probability of 0.938 *Table 1– DarkCovidNet test data performance metrics.*

		1	5	
Test set	Precision	Recall	F1 score	Accuracy
COVIDx	0.87 ± 0.00	0.80 ± 0.00	0.82±0.00	0.88±0.00
External	0.44 ± 0.00	0.43±0.00	0.41 ± 0.00	0.43±0.00
LTHT	0.47±0.01	0.46 ± 0.00	0.44±0.01	0.45 ± 0.00

Table 2 - CoroNet test data performance metrics.

Test set	Precision	Recall	F1 score	Accuracy
COVIDx	0.81±0.05	0.90±0.01	0.84±0.05	0.88±0.03
External	0.18±0.07	0.34±0.02	0.19±0.03	0.35±0.01
LTHT	0.24±0.01	0.30±0.00	0.15±0.01	0.22 ± 0.00

Table 3 – COVIDNet test data performance metrics.

Test set	Precision	Recall	F1 score	Accuracy
COVIDx	0.86±0.03	0.69 ± 0.05	0.72±0.05	0.86±0.02
External	0.34±0.05	0.36±0.01	0.29±0.02	0.38±0.01
LTHT	0.43±0.01	0.39±0.00	0.37±0.01	0.44±0.03

Harkness et al, https://arxiv.org/ftp/arxiv/papers/2109/2109.08020.pdf

Covid diagnosis challenge

- the pneumonia class within the COVIDx contains other pathologies, including, pleural effusion, infiltration, consolidation, emphysema and masses.
- The non-COVID data included paediatric chest Xray images which were the only significant source of paediatric images within COVIDx
- the images had different resolutions: 1024x1024 and 299x299. Models resize images (smaller images up-sampled and larger images must be down-sampled). This risks generation of artefacts.
- A dataset had presence of disk-shaped markers in COVID-19 chest Xrays

Black box

- Al methods are perceived as a "black box" which limits their use
- The clinician does not use a decision system when it's not clear how it reaches its decision



Machine learning: Look at the features selected and used by ML



Activation maps for deep learning



medicine chest, 0.13478 vacuum, 0.10249 water jug, 0.069547



Bias in dataset

- Bias in a dataset may include: confounding factors (e.g. comorbidities) but also dataset imbalance in factors such as gender, ethnic, social, environmental, or economic factors.
- A biased training dataset produces biased model ("garbage in, garbage out")



Data curation

- The MP can ensure that the images are acquired according to the protocol required for correct AI use free from relevant imaging artifacts, and correctly preprocessed and harmonized to reduce variability.
- MPs can significantly aid in the management of aggregate data from multiple modalities (PET, CT, radiography, MRI, ultrasound, daily CBCT, hybrid imaging, such as PET/CT and PET/MRI, 3D/4D and image time series, dose from RT etc.)
- MP should be involved in the development of metrics to assess the quality and completeness of data, methods to curate data, and QA programs of data archives

Physica Medica 83 (2021) 25-37

Data preparation for artificial intelligence in medical imaging: A comprehensive guide to open-access platforms and tools

Oliver Diaz^{a,*}, Kaisar Kushibar^a, Richard Osuala^a, Akis Linardos^a, Lidia Garrucho^a, Laura Igual^a, Petia Radeva^a, Fred Prior^b, Polyxeni Gkontra^a, Karim Lekadir^a



Data preparation pipeline prior developing and/or evaluation of AI solution.

Why physicists in AI?

• MP knows the basic physical mechanisms at the root of signal change and image contrast

Ehsan Samei, Thomas M. Grist, Physica Medica 64(1):319-322

• MPs are also trained in mathematics and can understand the principles of ML and DL better than other healthcare professionals

M. Avanzo

 "a change of focus from equipment to operation; from quality to consistency of quality, from testing performance to estimating outcome – and doing this with objective, standardisable and quantitative methods"

Editorial

The European Federation of Organisations for Medical Physics (EFOMP) White Paper: Big data and deep learning in medical imaging and in relation to medical physics profession

Mika Kortesniemi , Virginia Tsapaki, Annalisa Trianni, Paolo Russo, Ad Maas, , Hans-Erik Källman , Marco Brambilla, John Damilakis

Regulations

- The MP involved in research will be required to apply the statements and recommendations released by governmental agencies regarding privacy, security, secure access to health information, deidentification of sensitive data, and obtaining informed consent
- General data protection regulation (GDPR) in Europe, HIPAA in the US
- Understand medical devices regulations regarding AI (CE marking, FDA etc.)

Quality assurance of Al

International Journal of Medical Informatics 102 (2017) 71-79

Decaying relevance of clinical data towards future decisions in data-driven inpatient clinical order sets

Jonathan H. Chen^{a,*}, Muthuraman Alagappan^d, Mary K. Goldstein^{b,c}, Steven M. Asch^{a,f}, Russ B. Altman^{a,e}

^a Department of Medicine, Stanford University, Stanford, CA, USA

^b Geriatrics Research Education and Clinical Center, Veteran Affairs Palo Alto Health Care System, Palo Alto, CA, USA

^c Primary Care and Outcomes Research (PCOR), Stanford University, Stanford, CA, USA

^d Internal Medicine Residency Program, Beth Israel Deaconess Medical Center, Boston, MA USA

^e Departments of Bioengineering and Genetics, Stanford University, Stanford, CA, USA

^f Center for Innovation to Implementation (Ci2i), Veteran Affairs Palo Alto Health Care System, Palo Alto, CA, USA

Training with more longitudinal data (2009–2012) was no better than using only the most recent (2012)data, unless applying a decaying weighting scheme with a "half-life" of data relevance about 4 months.

Table 1. Summary of stages in the adoption of a AI-based Clinical Decision Support System.

	Stages	Objective
	Selection	Pick most appropriate CDSS in terms of match with target use case and clinical workflow, five "rights," performance, and user acceptability
	Acceptance testing	Test that CDSS satisfies security, privacy, and safety requirements applicable to medical devices, covering typical error scenarios, exceptions, and unforeseen conditions
	Commissioning	Prepare the CDSS for optimized use in the clinic (including potential customization) and test its safety and performance within the local context
	Implementatio n	Roll out the CDSS and transition from the old workflow to the new after training the end users and managing their expectations
	Quality assurance	Ensure that the quality of the CDSS remains fit for purpose by monitoring internal and external updates as well as context drift

Artificial intelligence-based clinical decision support in modern medical physics: Selection, acceptance, commissioning, and quality assurance

Geetha Mahadevaiah and Prasad RV Philips Research India, Bangalore 560045, India

Inigo Bermejo^{a)}

Department of Radiation Oncology (MAASTRO), GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht 6229 ET, Netherlands,

David Jaffray

Princess Margaret Cancer Centre, University Health Network, Toronto, ON M5G 2M9, Canada

Andre Dekker and Leonard Wee

Department of Radiation Oncology (MAASTRO), GROW School for Oncology and Developmental Biology, Maastricht University Medical Centre, Maastricht 6229 ET, Netherlands

(Received 11 March 2019; revised 27 April 2019; accepted for publication 27 April 2019; published 15 May 2020)

Med. Phys. 47 (5), May 2020 0094-2405/2020/47(5)/e228/8

Quality assurance of AI

Phantoms





Fig. 1. (a) Digital ground truth phantom used as input to the MRI simulator.

Ground truth images

9	Zero noise added, R = 1			Noise = 1.0, R = 1		
Recon Type	Reconstructed Image	Error Map	SER (dB)	Reconstructed Image	Error Map	SER (dB)
iFFT		(La	20.40	×	(II)	20.27
CG		(Inc.)	20.40	×	(H)	20.23
тv	X		20.51	×	a la	20.43
WL	X	(II)	20.45	×		20.35

Physica Medica Volume 50, June 2018, Pages 26-36

https://www.nature.com/articles/s41598-018-31509-z

Quality assurance in Al

- "a <u>set of past cases</u>, which includes <u>difficult and rare cases</u> along with a <u>representative sample of</u> <u>the local case population</u> could be retrospectively tested if a database with past cases exists."
- At the same time, a "<u>repeated local validation</u>" cohort should be assembled from time to time or preferably continuously to critically reexamine the tests done during the commissioning stage.
- The repetition may help to ensure that the CDSS remains clinically valid"

Safety/Risk Management

- The responsibility to prevent errors due to use of AI falls to humans.
- Automation errors:
 - omission when humans do not notice the failure of an AI tool and
 - commission when an action is performed following wrong AI's decision when there is evidence
- One of the key activities of the MP is patient safety management that is the evaluation of medical devices and procedures to guarantee the safety of patients.
- MPs are trained to prevent and analyze accidents by using risk assessment
 - failure modes and effects analysis (FMEA),
 - incident reporting programs



EFOMP Policy Statement

EFOMP policy statement 16: The role and competences of medical physicists and medical physics experts under 2013/59/EURATOM

Carmel J. Caruana¹, Virginia Tsapaki, John Damilakis, Marco Brambilla, Guadalupe Martín Martín, Asen Dimov, Hilde Bosmans, Gillian Egan, Klaus Bacher, Brendan McClean

EFOMP, United Kingdom

Multidisciplinary team

- «A fundamental role of the medical physicist is team member working together with physicians, technologists, nurses, therapists, engineers, and even patients in the effective, efficient, and safe delivery of health care»
- "In a similar vein, another key role of the medical physicist is education and training, not only of junior level and medical physicists in training but also of other health care professionals including residents and fellows."

Artificial intelligence will soon change the landscape of medical physics research and practice

Lei Xing, Ph.D. Department of Radiation Oncology, Stanford University, Stanford, CA 94305, USA (Tel: 650-498-7896; E-mail: lei@stanford.edu)

Elizabeth A. Krupinski, Ph.D. Department of Radiology & Imaging Sciences, Emory University, Atlanta, GA 30322, USA (Tel: 404-712-3868; E-mail: ekrupin@emory.edu)

Jing Cai, Ph.D., Moderator

(Received 12 February 2018; revised 19 February 2018; accepted for publication 19 February 2018; published 13 March 2018)

[https://doi.org/10.1002/mp.12831]

