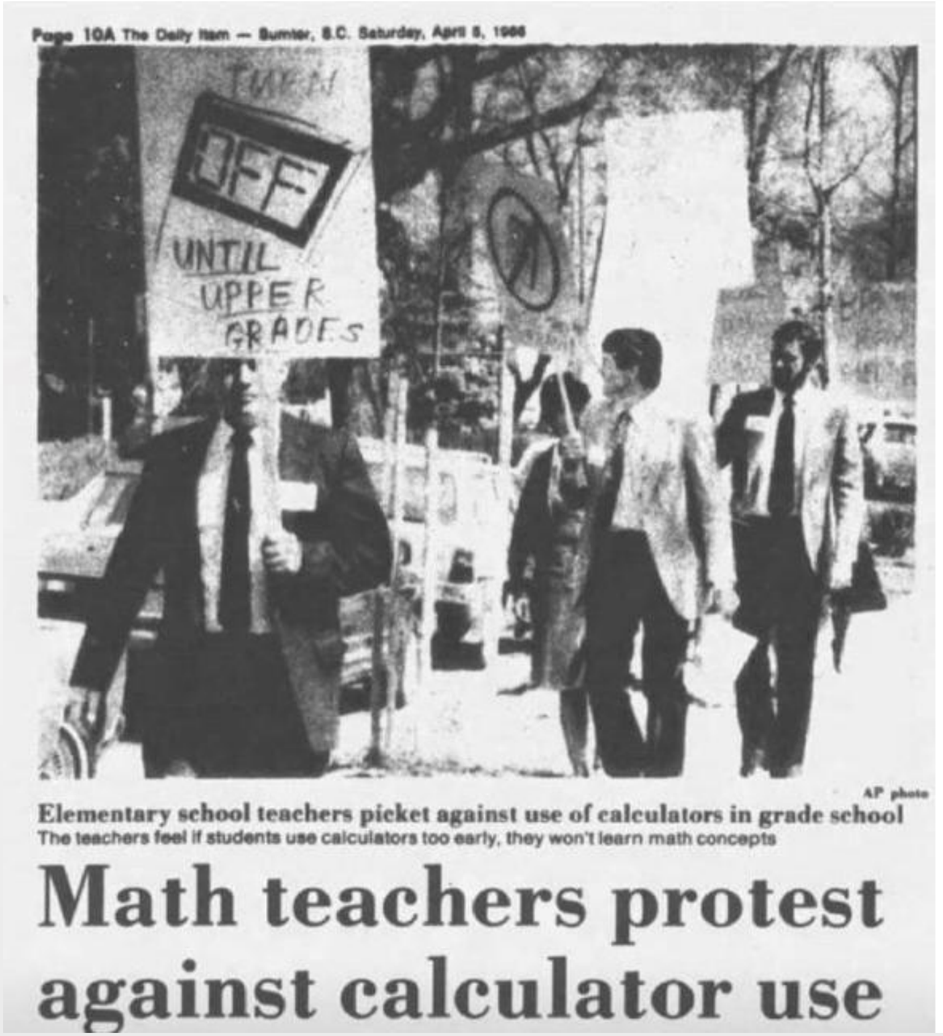
JORNADAS SARH
Inteligencia Artificial Aplicada a la Física Médica
Antequera, 14 de noviembre 2025

Integración de la inteligencia artificial en Radioterapia: avances tecnológicos y retos clínicos

Eva Ambroa Rey

Especialista en Radiofísica Hospitalaria
Hospital del Mar (Barcelona)
evamaria.ambroa.rey@hmar.cat

¿LA IA REEMPLAZARÁ NUESTRO TRABAJO?



Page 10A The Daily Item — Sumter, S.C. Saturday, April 9, 1966

AP photo

Elementary school teachers picket against use of calculators in grade school

The teachers feel if students use calculators too early, they won't learn math concepts

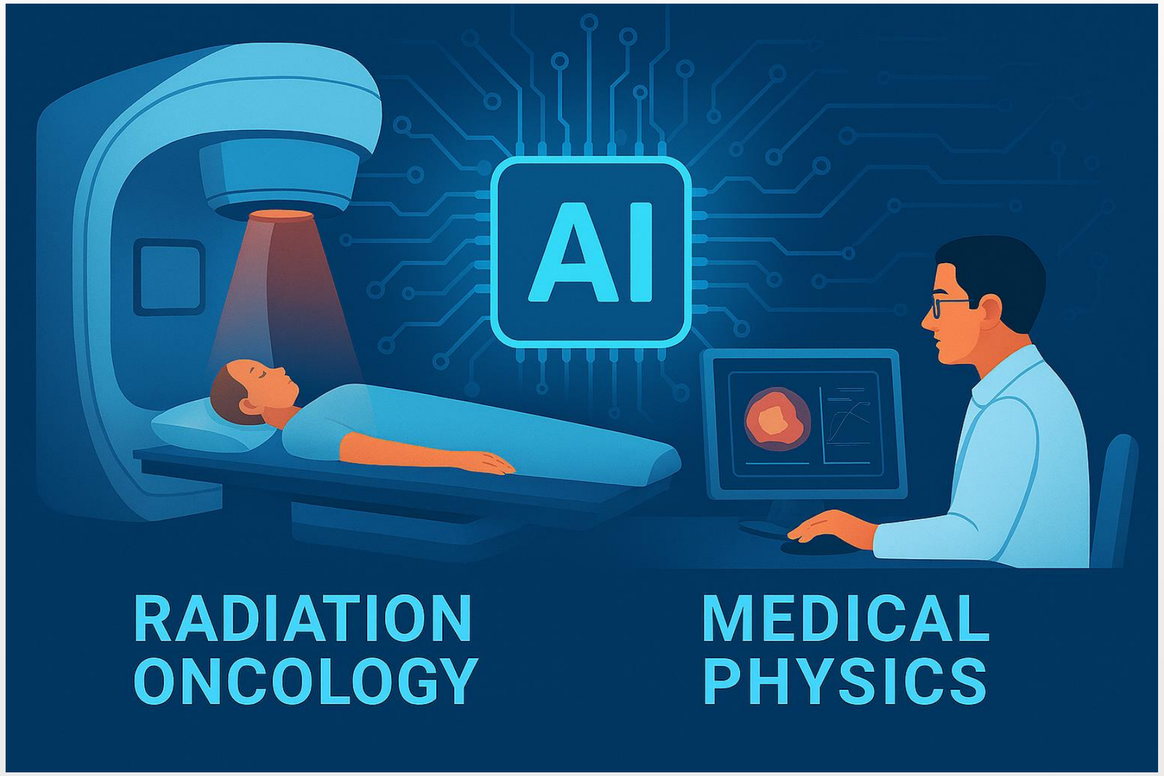
Math teachers protest against calculator use

1966

https://miroslao7.medium.com/chatgpt-wont-take-your-job-away-a96d9a7b6a8

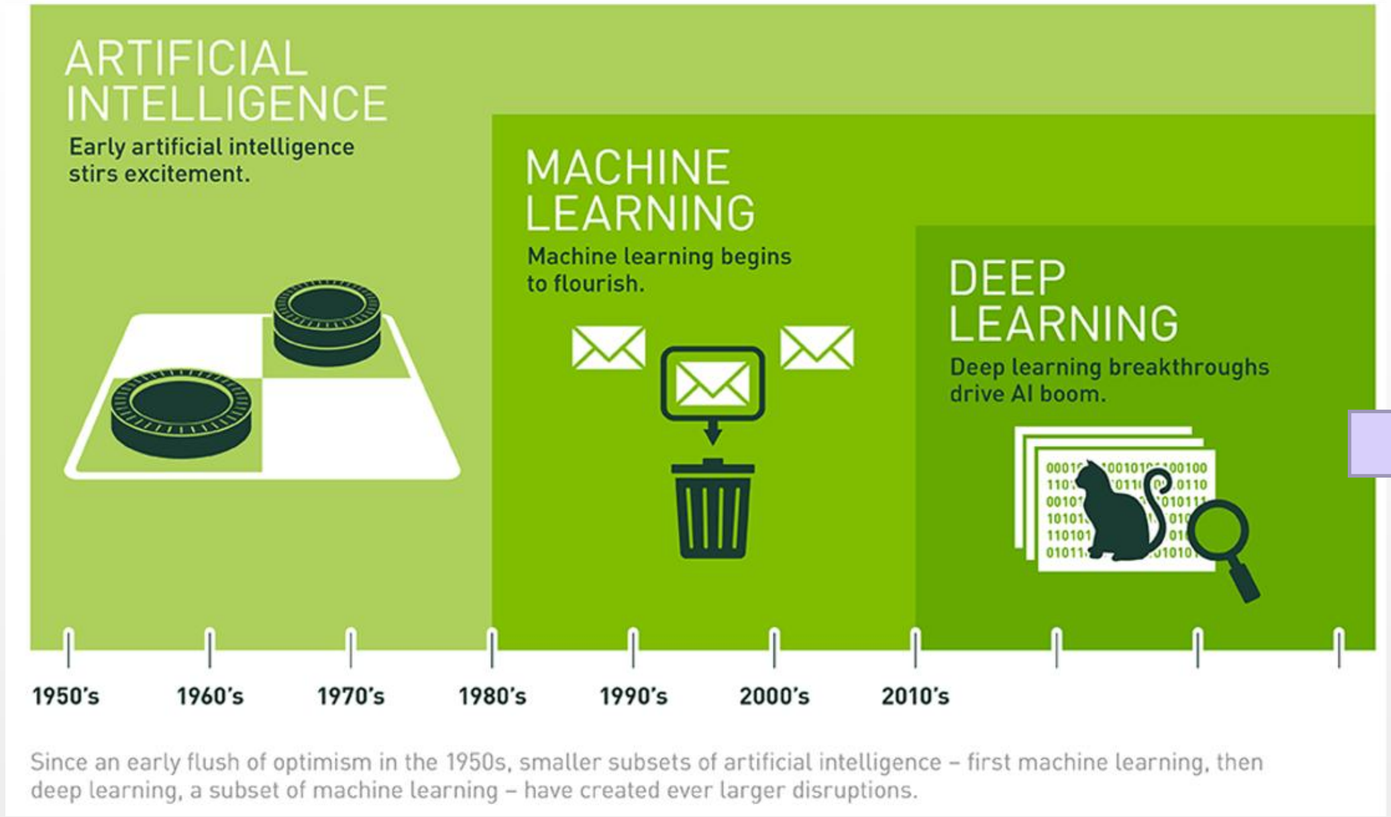
FUTURO DE LA IA EN RT Y FÍSICA MÉDICA

- Prompt Co-pilot: puedes crearme una imagen de cómo influirá la IA en el ámbito de la oncología radioterápica y la física médica?



- Prompt Co-pilot: puedes crearme una imagen de cúal será la promesa de la IA en sanidad?

Evolución histórica



Fuente: Michael Copeland What's The Difference Between AI, Machine Learning And Deep Learning, NVIDIA Blogs

GENERATIVE AI

- Diffusion models
- Large Language Models (LLM)

Midjourney

ChatGPT

DATOS Y ESTRUCTURA

- 97% de los datos clínicos son ‘dark data’ perdidos
- Cuando se estructuran los datos y se comparten:

Benedict S, et al. *Introduction to Big Data in Radiation Oncology*. (2016)

Big data initiatives

OncoSpace – AI-based predictive planning

Momentum – MRgRT

UniCancer – CANTO-RT

AAPM Data Science Committee – Big Data Subcommittee

TCIA - RTStruct data

MROCQ – registry for benchmarking and practice improvement

NSIR-RT – nation-wide incident reporting

E^2-RADIatE – EORTC/ESTRO pan-European platform for prospective registries

Grand Challenge datasets

IAEA DIRAC – comprehensive database on radiotherapy resources

ONCOSPACE

MOMENTUM

unicancer

AAPM

EORTC ESTRO ...the E^2-RADIatE platform.

MRQC MICHIGAN RADIATION ONCOLOGY QUALITY CONSORTIUM

TCGA

THE CANCER IMAGING ARCHIVE

APLICACIONES IA EN RT

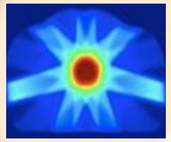
01. DECISIÓN TTO

Herramientas que combinan datos clínicos con genómicos e imagen para dar soporte a la decisión.

02. IMAGEN

Reconstrucción, mejora de imagen, generación de imágenes sintéticas, auto-contorneo…

03. PLANIFICACIÓN DE TTO



Predicción dosis, DVH, técnica, …

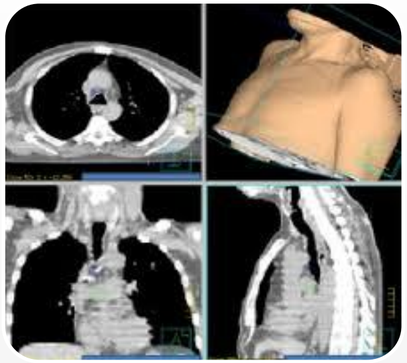
04. ADMINISTRACIÓN TTO

Posicionamiento, monitorización durante el tto. predecir cambios anatómicos, adaptativa sobre CBCT, …

05. CONTROL DE CALIDAD

QA paciente, errores MLC, output, simetría, …

IMAGEN



Generación/Reconstrucción CT/RM

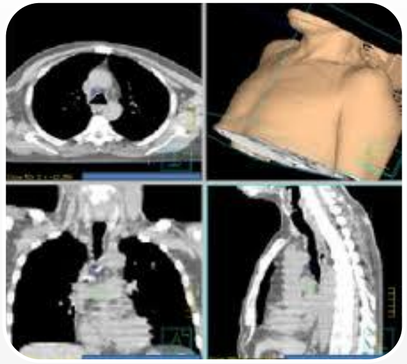
- ↑ calidad imagen
- ↓ ruido
- ↓ artefactos
- Supervisados
- U-Net, GAN

Table 1. Available commercial solution for simulation CT generation in October 2023.

Brand name (Company)	Imaging technique	Anatomical site	Certifications
MRCAT® (Philips Healthcare)	T1-weighted GRE + Dixon	Pelvis, brain, head & neck	CE, FDA March 2016
MRI Planner® (Spectronics Medical)	T2-weighted	Pelvis	FDA, CE June 2016
	T1-weighted GRE	Brain, head & neck	
SyngoVia.syntheticCT® (Siemens Healthcare)	Multiple MRI sequences	Brain	FDA, CE January 2018
	T1-weighted GRE + Dixon	Pelvis	
MR-Box® (Therapanacea)	T1-weighted GRE	Brain	CE, FDA 2021–2022
	T1-weighted GRE, T2-weighted	Pelvis	
Adapt-Box® (Therapanacea)	Cone beam CT	Pelvis	CE 2023, FDA pending

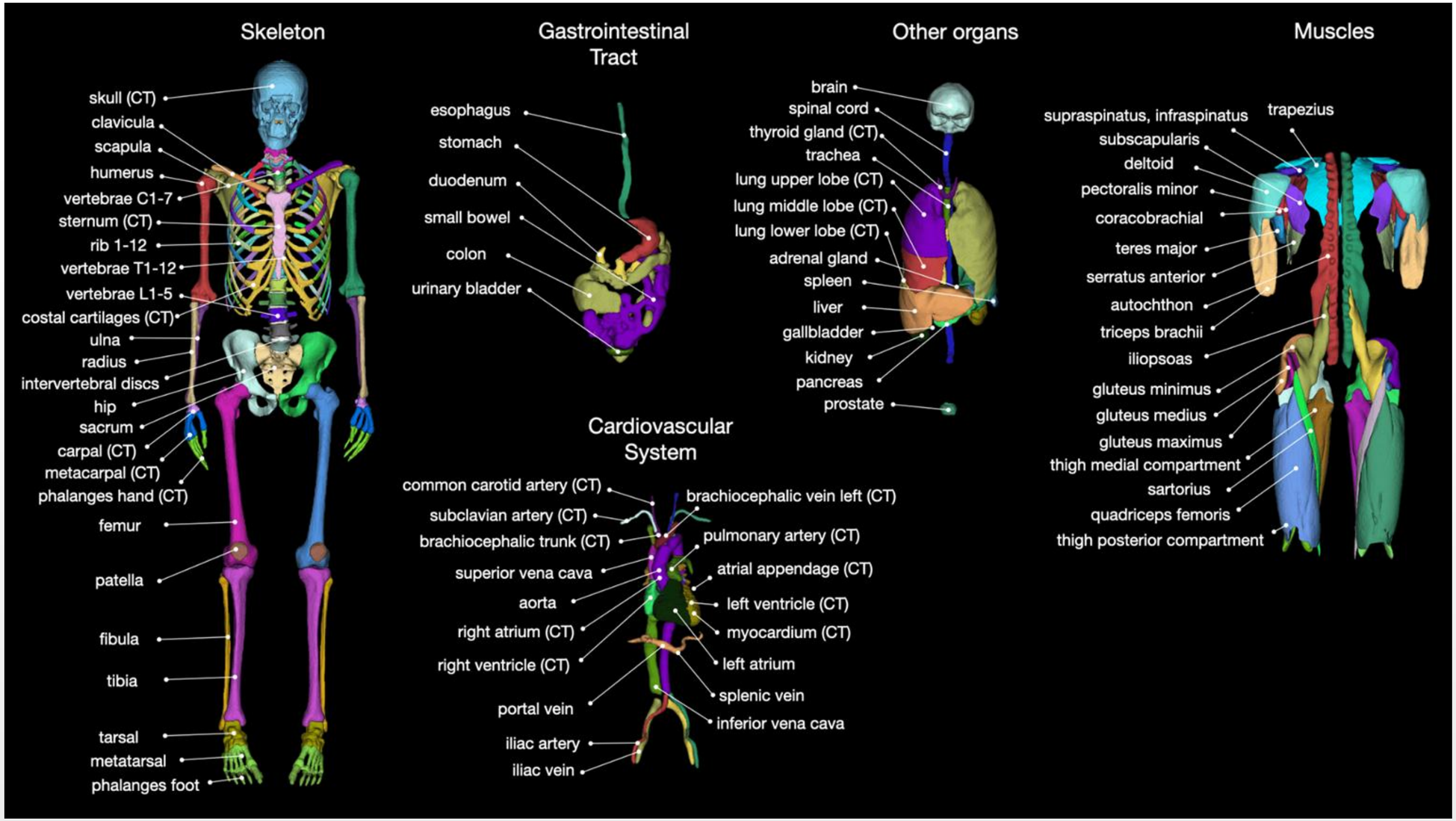
Artificial intelligence in radiotherapy: Current applications and future trends,Diagnostic and Interventional Imaging, Paul Giraud, Jean-Emmanuel Bibault,2024

IMAGEN



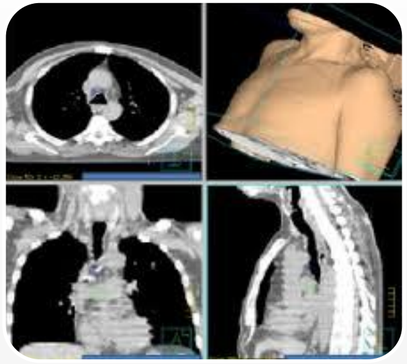
Reconstrucción CT/RM

- ↑ calidad imagen
- ↓ ruido
- ↓ artefactos
- Supervisados
- U-Net, GAN



https://github.com/wasserth/TotalSegmentator

IMAGEN



Generación/Reconstrucción CT/RM

- ↑ calidad imagen
- ↓ ruido
- ↓ artefactos
- Supervisados
- U-Net, GAN

Segmentación

- CT, PET, RM
- ↓ tiempo
- ↑ estandarización
- Supervisados
- U-Net, ResNet, DenseNet

NO SOLO OARS!

Radiotherapy and Oncology

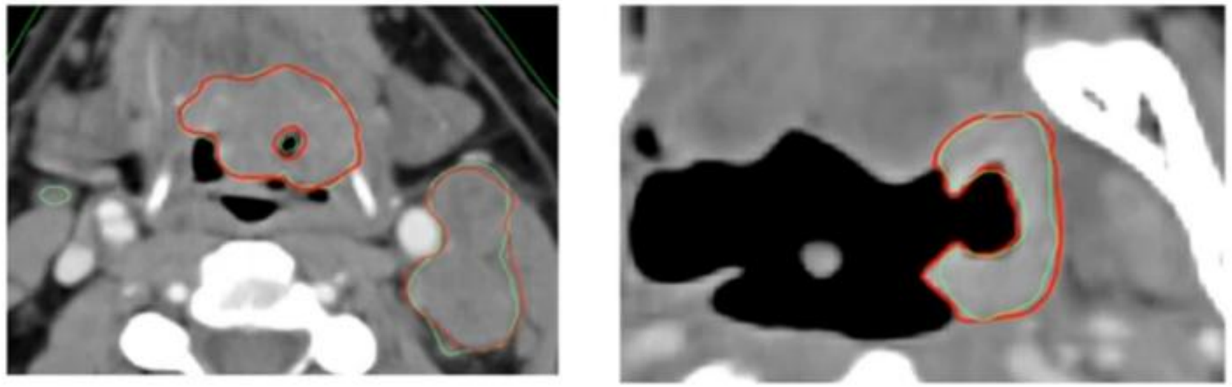
journal homepage: www.thegreenjournal.com

Original Article

Benefits of automated gross tumor volume segmentation in head and neck cancer using multi-modality information

Heleen Bollen [a,*,1], Siri Willems [b,1], Marilyn Wegge [a], Frederik Maes [b], Sandra Nuyts [a]

[a] KU Leuven, Dept. Oncology, Laboratory of Experimental Radiotherapy, & UZ Leuven, Radiation Oncology; and [b] KU Leuven, Dept. ESAT, Processing Speech and Images (PSI), & UZ Leuven, Medical Imaging Research Center, B-3000 Leuven, Belgium



(g)

(h)

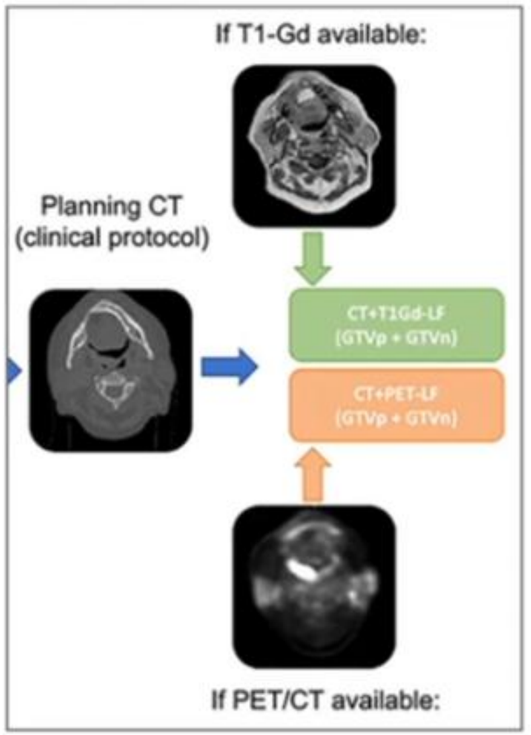


Fig. 2. Examples of automated GTV delineations by the LF CNN for CT + PET (a-d) and CT + MRI (e-h). Red. manual ground truth; green. automated delineation.

- 150 casos clínicos H&N
- CT, CT+PET o CT+MRI
- CNN con multimodalidad mejor que CNN solo con CT

The time necessary to manually delineate the whole GTV was on average 15,5 min per case, while correcting the automated delineations by applying only necessary corrections took 8 min on average per case, i.e. an increase in efficiency of 48%.

DL auto-segmentation (1 año uso clínico)

1. Evaluar el impacto del autocontorneo en el tiempo
2. Evaluar el impacto sobre el tiempo total para completar la planificación del tto (desde el CT hasta el inicio del TTO)

1. Se ahorró un 70% aprox
2. El tto no se inició antes a pesar de reducirse el tiempo de contorneo

Necesidad de revisar el workflow y adaptarlo a las nuevas necesidades

Radiotherapy and Oncology 209 (2025) 110977

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Real world AI-driven segmentation: Efficiency gains and workflow challenges in radiotherapy

Ciaran Malone [a,*], Jill Nicholson [a,b], Samantha Ryan [a], Pierre Thirion [a,c], Ruth Woods [a], Peter McBride [a], Orla McArdle [a,c], Frances Duane [a,b], Gerard G. Hanna [a,b], Brendan McClean [a,d], Sinead Brennan [a,b]

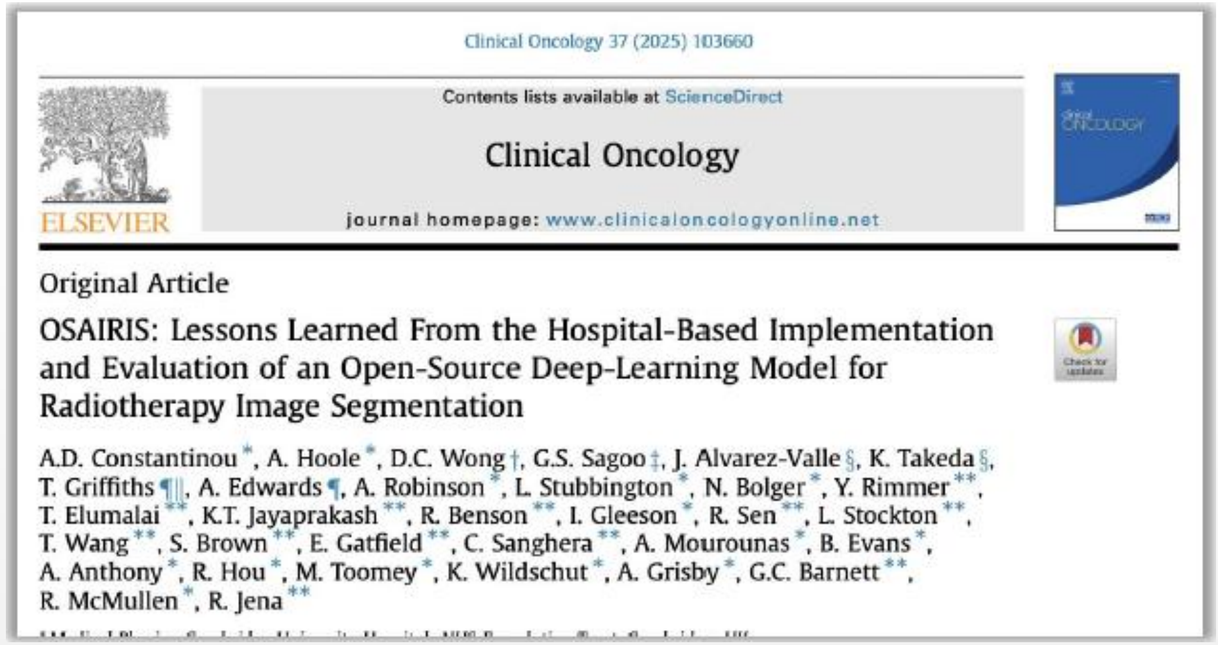
[a] St.Luke's Radiation Oncology Network, Dublin, Ireland
[b] Applied Radiation Therapy Trinity, Discipline of Radiation Therapy & Trinity St James's Cancer Institute, Trinity College Dublin, Dublin, Ireland
[c] School of Medicine, Discipline of Radiation Therapy & Trinity St James's Cancer Institute, Trinity College Dublin, Dublin, Ireland
[d] Centre for Physics in Health and Medicine, School of Physics, UCD, Dublin, Ireland
[e] Beaumont RSCI Cancer Centre, Royal College of Surgeons, Ireland

Site	N		Median (Hrs)			Significance
	Pre	*Post*	*Pre*	*Post*	*%*	p
All	3472	3820	3.3	1.6	-51.5%	<0.001
Thorax	594	489	4	1.3	-67.5%	<0.001
Pelvis	1021	1422	3.8	2.4	-36.8%	<0.001
H&N	821	912	3.4	1.1	-67.6%	<0.001
Breast	220	310	2.7	2.6	-3.7%	0.152
Abdomen	68	51	4	1.1	-72.5%	0.004
Rescan	369	292	1.8	1.4	-22.2%	0.022
Other	379	344	1.9	0.7	-63.2%	<0.001

A diferencia de otros trabajos, no se centra solo en métricas de precisión (como Dice), sino en **tiempo real de edición, utilidad práctica y seguridad clínica**.

Principales hallazgos

- **Ahorro de tiempo significativo:**
  - 36 % en próstata
  - 67 % en cabeza y cuello
- **Hausdorff distance** fue la métrica que **mejor correlacionó con el tiempo de edición** (ρ = 0.70), superando al Dice, que no refleja bien el esfuerzo clínico.
- Se detectó **“automation bias”**: los clínicos tienden a confiar en los contornos de la IA y a anclarse en ellos, aunque la mayoría de errores fueron corregidos (72 % en H&N, 81 % en próstata).
- En una evaluación ciega, los clínicos **prefirieron los contornos de la IA** frente a los de colegas humanos (“gold standard”).



Clinical Oncology 37 (2025) 103660

Contents lists available at ScienceDirect

Clinical Oncology

journal homepage: www.clinicaloncologyonline.net

Original Article

OSAIRIS: Lessons Learned From the Hospital-Based Implementation and Evaluation of an Open-Source Deep-Learning Model for Radiotherapy Image Segmentation

A.D. Constantinou *, A. Hoole *, D.C. Wong †, G.S. Sagoo ‡, J. Alvarez-Valle §, K. Takeda §, T. Griffiths ¶‖, A. Edwards ¶, A. Robinson *, L. Stubbington *, N. Bolger *, Y. Rimmer **, T. Elumalai **, K.T. Jayaprakash **, R. Benson **, I. Gleeson *, R. Sen **, L. Stockton **, T. Wang **, S. Brown **, E. Gatfield **, C. Sanghera **, A. Mourounas *, B. Evans *, A. Anthony *, R. Hou *, M. Toomey *, K. Wildschut *, A. Grisby *, G.C. Barnett **, R. McMullen *, R. Jena **

Article | Open access | Published: 24 October 2024

LLM-driven multimodal target volume contouring in radiation oncology

Yujin Oh, Sangjoon Park, Hwa Kyung Byun, Yeona Cho, Ik Jae Lee, Jin Sung Kim & Jong Chul Ye

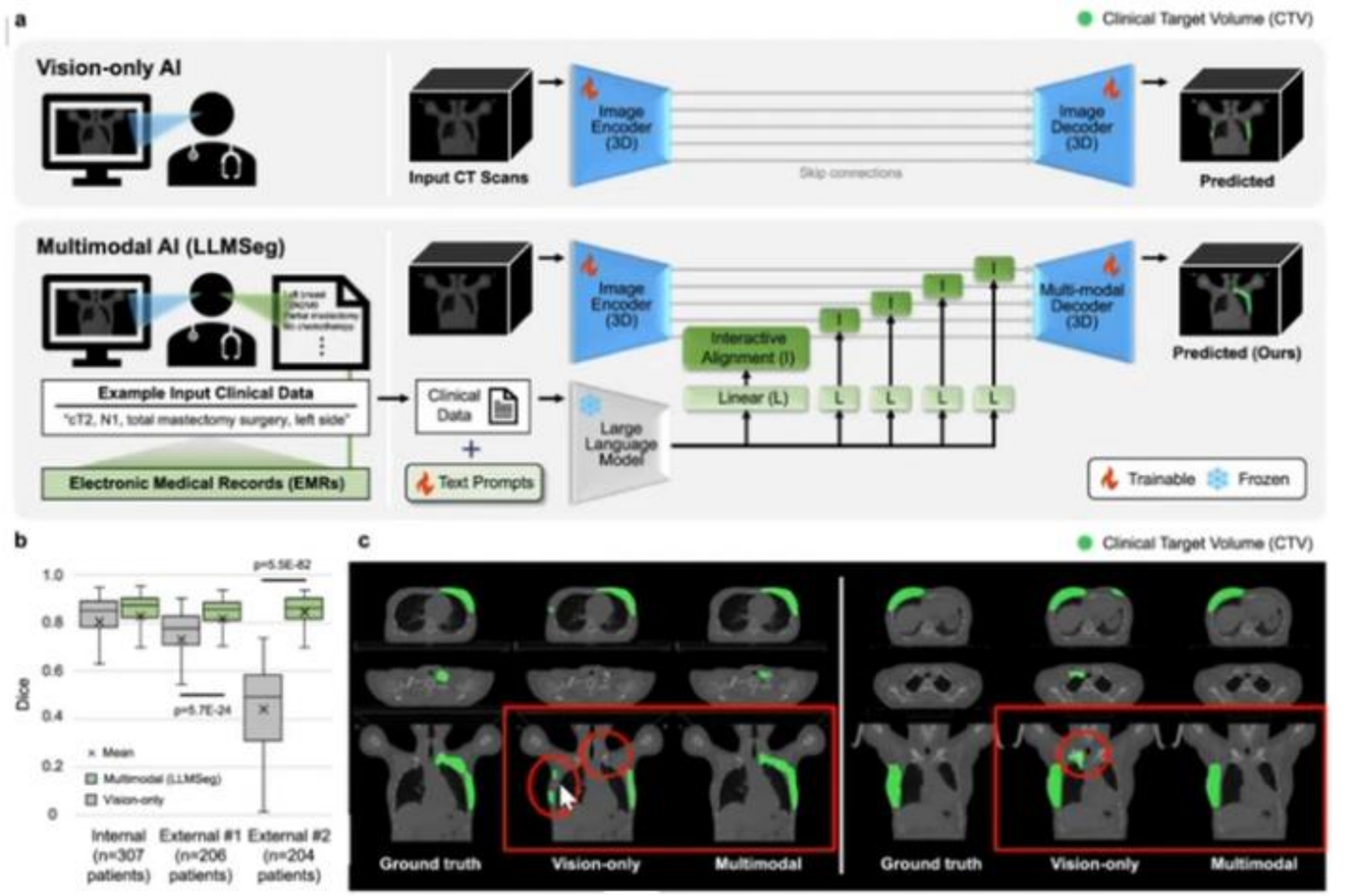
Nature Communications **15**, Article number: 9186 (2024) | Cite this article

Lo validan sobre un conjunto de datos de cáncer de mama (y también aplican a próstata) con imagen de simulación de radioterapia + información clínica.

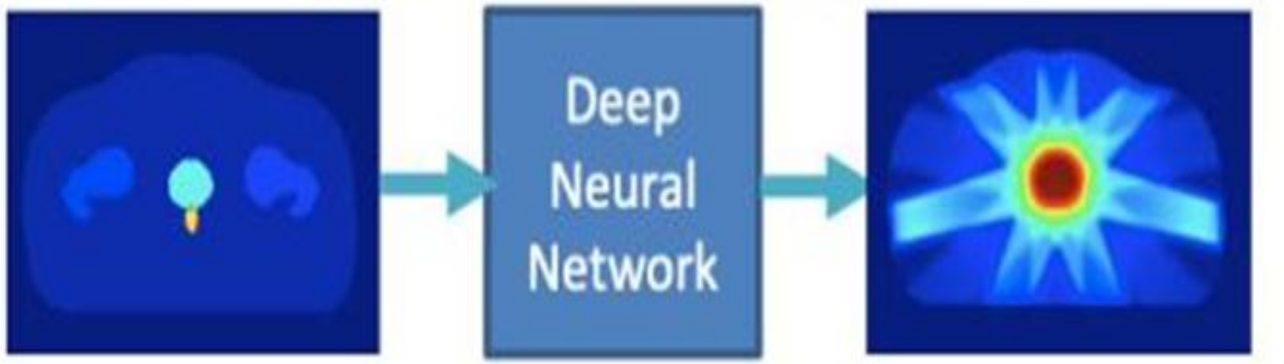
- Muestra que **la incorporación de información clínica textual** junto con imagen mejora la automatización de tareas de contorno,
- Refuerza la idea de que **menos datos pueden bastar** si se usan de forma multimodal — algo interesante si en tu entorno tienes limitaciones de datos de entrenamiento.

- Plantea un camino hacia modelos explicables que integran “razones clínicas” en la predicción, lo cual puede mejorar la confianza del clínico en la IA.

La validación se limita a ciertos tipos de cáncer (mama y próstata) con volúmenes de contorno relativamente estandarizados; no está validado aún para tumores más complejos o menos estructurados.



PREDICCIÓN DE DOSIS



- Guía para planificación de ttos
- Planificación automática
- Predicción de DVH, distribución dosis
- Comparación de modalidades (IMRT/VMAT/3D…)
- Cálculo de dosis
- Selección de ángulos
- Planificación interactiva con lenguaje natural

- Supervisados
- 3D
- 2D

> Med Phys. 2025 Mar;52(3):1798-1809. doi: 10.1002/mp.17527. Epub 2024 Dec 3.

Breast radiotherapy planning: A decision-making framework using deep learning

Pedro Gallego [1 2], Eva Ambroa [3], Jaime PérezAlija [1], Nuria Jornet [1], Cristina Anson [1]
Natalia Tejedor [1], H
Victor Riu [1], Javier F

Affiliations + expar

PMID: 39625151 PI

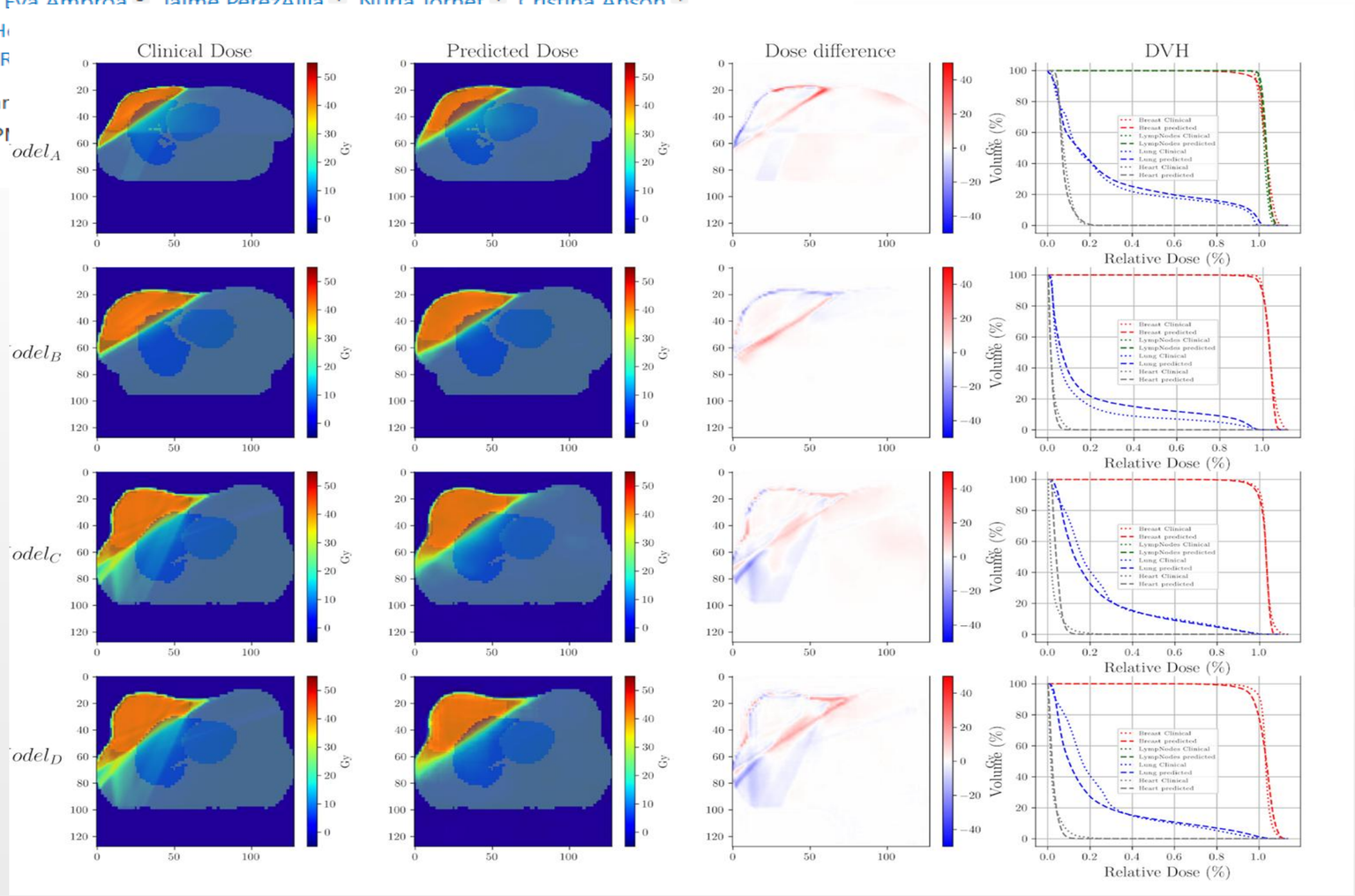


TABLE 3 Confusion matrix between the decisions made by the independent observer (ground truth) against the clinical decisions made historically.

	Actually 3D-CRT	Actually IMRT
Historical 3D-CRT	11	3
Historical IMRT	12	4

Abbreviations: 3DCRT, three-dimensional conformal radiation therapy; IMRT, intensity-modulated radiation therapy.

To test the decision-making framework, $Model_B$ and $Model_C$ were used, with the latter chosen due to its better performance compared to $Model_D$ (Table 2, Figure 6).In comparing the historical decision and the decision-making framework, as shown in Table 3, the results exhibit a mix of true positives, true negatives, false positives, and false negatives. The recall rate for this comparison was 47.8%, precision was 78.6%, accuracy was 50%, and the F1 score was 59.5%.

TABLE 4 Confusion matrix between the decisions made by the independent observer (ground truth) against the decision making framework.

	Actually 3D-CRT	Actually IMRT
Predicted 3D-CRT	22	2
Predicted IMRT	1	5

Abbreviations: 3DCRT, three-dimensional conformal radiation therapy; IMRT, intensity-modulated radiation therapy.

In the second comparison between the independent observer's decision (ground truth) and the decision-making framework, represented in Table 4, the outcomes were more favorable. The recall rate in this comparison was 95.7%, precision was 91.7%, accuracy was 90.0%, and the F1 score was 93.6%.

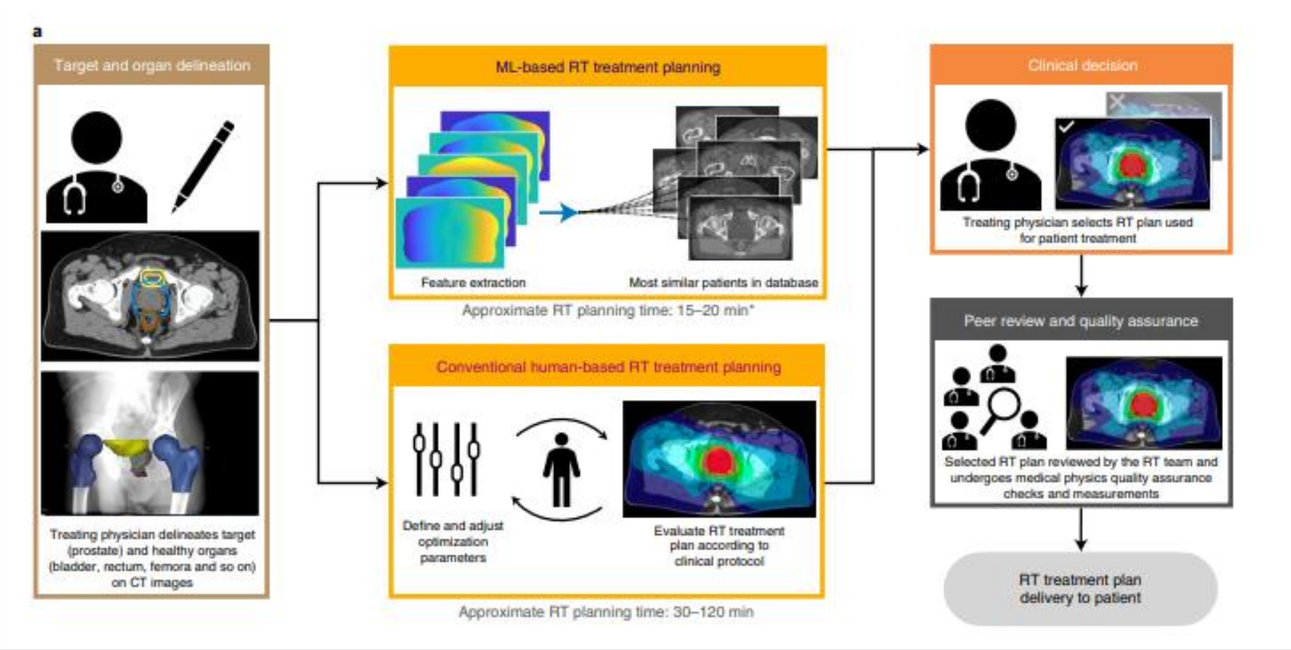
nature medicine
LETTERS
https://doi.org/10.1038/s41591-021-01359-w
Check for updates

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

Chris McIntosh[1,2,3,4,5,6,8], Leigh Conroy[1,2,7,8], Michael C. Tjong[1,7], Tim Craig[1,2,7], Andrew Bayley[1,7], Charles Catton[1,7], Mary Gospodarowicz[1,7], Joelle Helou[1,7], Naghmeh Isfahanian[1,7], Vickie Kong[1,7], Tony Lam[1,7], Srinivas Raman[1,7], Padraig Warde[1,7], Peter Chung[1,7], Alejandro Berlin[1,2,7,8] ✉ and Thomas G. Purdie[1,2,6,7,8] ✉

- **Random Forest** con 99 planes clínicamente aprobados.
- Se realizaron tres fases:
  1. **Factibilidad técnica** (n=17)
  2. **Simulación retrospectiva** (n=50)
  3. **Despliegue clínico prospectivo** (n=50), con selección ciega entre plan ML y plan humano; el plan elegido se utilizó realmente en el tratamiento.

- **Aceptabilidad clínica:** 89% de los planes ML fueron considerados aptos para tratar pacientes.
- **Selección plan tto:** el 72% ML vs humanos (83% en simulación, 61% en despliegue real).
- **Eficiencia:** el tiempo total de planificación se redujo en un **60%** (de 118 h a 47 h).
- **Calidad dosimétrica:** los planes ML mantuvieron calidad constante entre fases y se ajustaron a las guías clínicas.
- La menor selección de planes ML en la fase clínica no se debió a una pérdida de calidad, sino a **factores humanos** (percepción y confianza del médico tratante).



Radiotherapy and Oncology 200 (2024) 110522

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Original Article

Clinical implementation of deep learning robust IMPT planning in oropharyngeal cancer patients: A blinded clinical study

Ilse G. van Bruggen [a,*], Marije van Dijk [a], Minke J. Brinkman-Akker [a], Fredrik Löfman [b], Johannes A. Langendijk [a], Stefan Both [a], E.W. Korevaar [a]

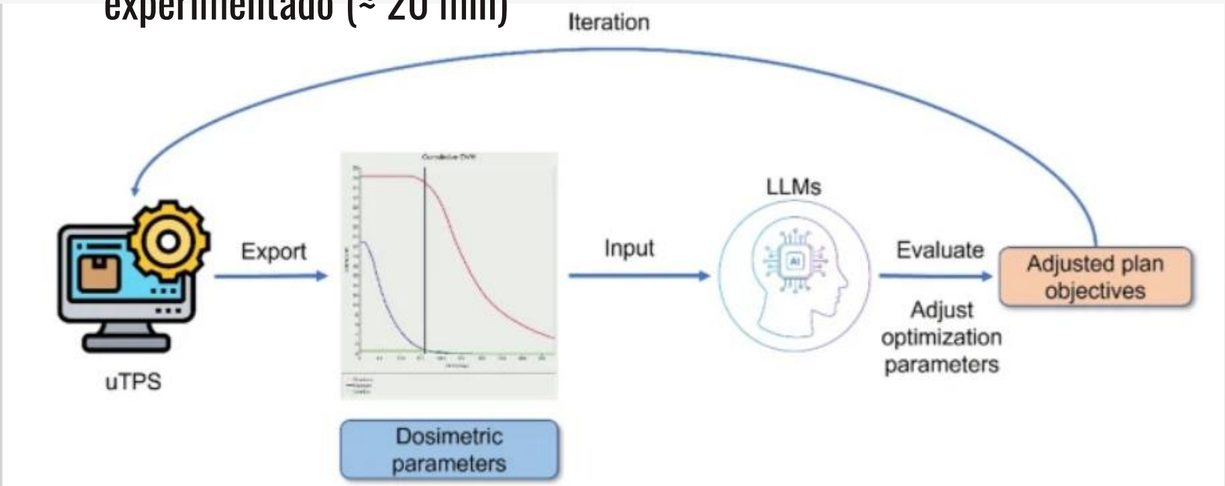
[a] Department of Radiation Oncology, University Medical Center Groningen, Groningen, the Netherlands
[b] RaySearch Laboratories, Stockholm, Sweden

Fig. 1. Overview of the DLO implementation workflow.

- Deep learning planning automates robust proton therapy planning for oropharyngeal cancer.
- Plan quality was evaluated in blinded retrospective and prospective studies.
- Deep learning plans were preferred over or comparable to manual plans in 92% of the patients.

- Aunque los resultados son prometedores, todavía se requiere supervisión humana para revisar y aprobar los planes generados automáticamente.
- La generalización a otros sitios anatómicos, otros centros con distintos sistemas de protonterapia o diferentes protocolos aún necesita validación adicional.
- Como siempre en IA aplicada al entorno clínico, la gestión del cambio, la aceptación del usuario y la integración en el flujo de trabajo son factores clave que deben considerarse más allá de la técnica.

- 35 pacientes cérvix
- Conformidad PTV y OARs
- Varios modelos LLM (2.5-max, LLama-3.2, Gemini 1.5 flash
- Un modelo generó halucinaciones
- Dos de los LLMs (Qwen-2.5-max y Llama-3.2) generaron planes aceptables en tiempos de ≈ 16.3 ± 5.0 min y ≈ 9.8 ± 2.1 min, respectivamente, mejorando el tiempo estimado de un físico experimentado (≈ 20 min)



The workflow of using LLM for radiotherapy treatment planning

Parameters fed into the LLMs comprised current doses, dose goals, constraint target values, ideal ranges of constraint target values for OAR optimization objectives, and PTV D95 values.

Research | Open access | Published: 15 May 2025

Feasibility study of automatic radiotherapy treatment planning for cervical cancer using a large language model

Shuoyang Wei, Ankang Hu, Yongguang Liang, Jingru Yang, Lang Yu, Wenbo Li, Bo Yang & Jie Qiu

Radiation Oncology **20**, Article number: 77 (2025) | Cite this article

A feasibility study of automating radiotherapy planning with large language model agents

Qingxin Wang, Zhongqiu Wang, Minghua Li, Xinye Ni, Rong Tan, Wenwen Zhang, Maitudi Wubulaishan, Wei Wang, Zhiyong Yuan, Zhen Zhang Show full author list

Published 21 March 2025

Physics in Medicine & Biology, Volume 70, Number 7

3764 · Volume 123, Issue 1, Supplement , E799, September 01, 2025

Development and Performance Evaluation of an LLM-Based Automated Recommendation System for Radiotherapy Treatment Planning

Y.X. Yang [1] · L. Jia [2] · H. Li [2] · ... · G.Y. Wang [1] · G. Zhou [1] · Y. Sun [5] ... Show more

Affiliations & Notes Article Info

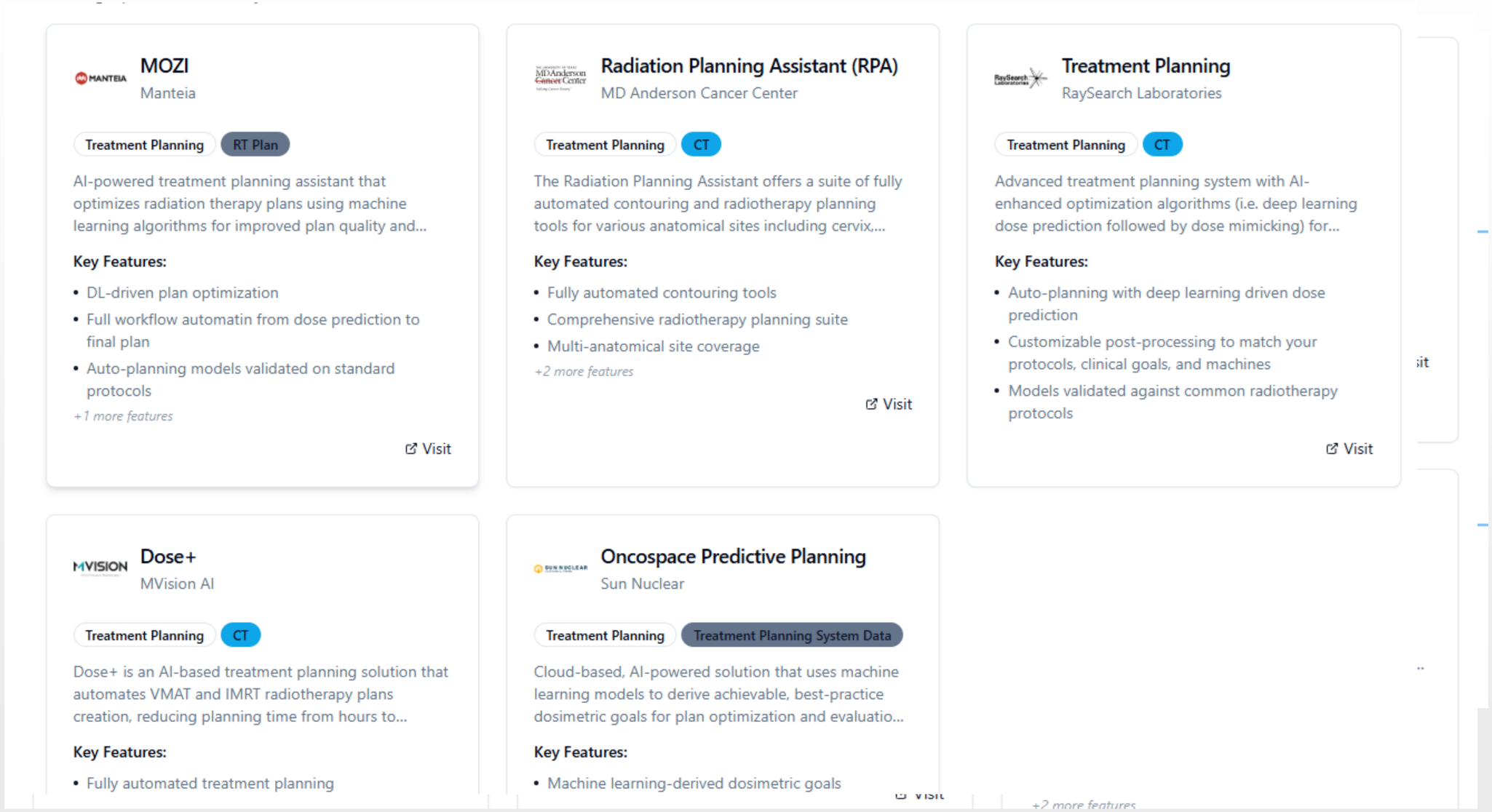
169 · Volume 123, Issue 1, Supplement , S68, September 01, 2025

Automating Head-and-Neck Cancer Intensity Modulated Radiation Therapy Treatment Planning with a ReAct Large Language Model Agent

D. Yang [1] · X. Wu [1] · Y. Xie [1] · ... · Q. Wu [1] · Q.J.J. Wu [2] · Y. Sheng [1] ... Show more

Aplicaciones clínicas desarrolladas

http://www.dlinrt.eu



Physics and Imaging in Radiation Oncology 16 (2020) 144–148

Contents lists available at ScienceDirect

Physics and Imaging in Radiation Oncology

journal homepage: www.sciencedirect.com/journal/physics-and-imaging-in-radiation-oncology

Machine learning applications in radiation oncology: Current use and needs to support clinical implementation

Charlotte L. Brouwer [a,*], Anna M. Dinkla [b], Liesbeth Vandewinckele [c,d], Wouter Crijns [c,d], Michaël Claessens [e,f], Dirk Verellen [e,f], Wouter van Elmpt [g]

- Survey
  - 213 responders in 202 radiotherapy clinics
  - 37% of responders implemented at least 1 ML application clinically
- Main reasons introducing AI:
  - Time saving
  - Quality improvement (also consistency)
- Main reasons **not** introducing AI:
  - Lack of knowledge how to implement
  - No software available & lack of time
  - Resistance or scepticism against AI

Inconvenientes principales de la adopción de la IA en la clínica

- Ética
- Transparencia de los modelos de IA (models cards)
- De-skilling (pérdida de conocimiento)
- Opiniones de los pacientes ante el uso de IA
- Educación y formación del personal

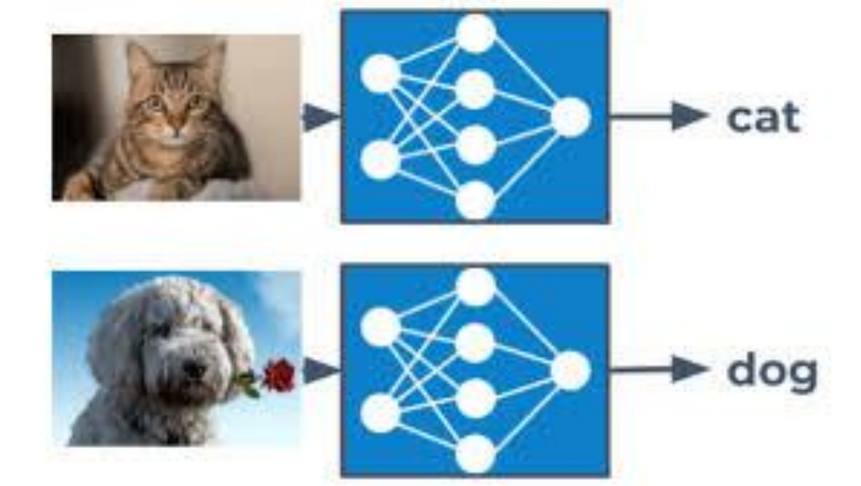
Intelligent Machines

The Dark Secret at the Heart of AI

No one really knows how the most advanced algorithms do what they do. That could be a problem.

by Will Knight April 11, 2017

Ética



- toma de decisiones sobre pacientes
- necesidad de saber las razones y generar confianza sobre predicciones
- herramientas para evaluar las predicciones

Official Journal
of the European Union

EN
L series

2024/1689

12.7.2024

REGULATION (EU) 2024/1689 OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL

of 13 June 2024

laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)

(Text with EEA relevance)

THE EUROPEAN PARLIAMENT AND THE COUNCIL OF THE EUROPEAN UNION,

Transparencia: MODEL CARDS

- Son como el prospecto de un medicamento, pero para un modelo de IA:
- qué hace,
- con qué datos se entrenó,
- para qué sirve
- para qué no.

Articles

Standardization of Artificial Intelligence Development in Radiotherapy

Alessia de Biase † [1], Nikos Sourlos ‡ [1], Peter M.A. van Ooijen § ‖

Model Card	Model Fact Labels	Fact Sheets
Model Details	**Summary/ Mechanism**	**Statement of purpose**
Person or organization developing model: It was developed in 2021 by a group of PhD students in Deep Learning in Medical Imaging from the University Medical Hospital of Groningen. With the implemented models the group participated in the HECKTOR 2021 challenge. **Model type:** Co-learning method and a 3D Skip Spatial and Channel Squeeze and Excitation Multi-Scale Attention method (Skip-scSE-M), both based on Convolutional Neural Networks **Paper or other resource for more information:** De Biase, A., et al.: Skip-SCSE multi-scale attention and co-learning method for oropharyngeal tumor segmentation on multi-modal PET-CT images. In: Andrearczyk, V., Oreiller, V., Hatt, M., Depeursinge, A. (eds.) HECKTOR 2021. LNCS, vol. 13209, pp. 109–120. Springer, Cham (2022)		
Intended Use/ Factors/Ethical considerations	**Uses and directions**	**Statement of purpose**
Intended Use: intended to be used **by radiation oncologists before radiotherapy treatment planning** for the automatic delineation of the gross tumor volume (GTVt) **Target Population:** oropharyngeal cancer patients **Benefits: manual delineation of contours is time consuming and prone to errors** **Use case: after PET and CT images are collected for a new patient, the primary tumor contour is immediately available and the treatment planning can start** **Appropriate decision support: the model identifies a primary tumor for the patient X and the oncologist discusses the possible treatment options** **Before using this model: test the model on internal data to establish the model validity on a different dataset** **Safety and efficacy evaluation: results are promising for this task when comparing with oncologists' manual contours**		
Metrics/ Quantitative Analysis	**Validation and Performance**	**Basic Performance**
The metrics used are the Dice Similarity Coefficient (DSC) and the Hausdorf distance at 95% (HD95) between the ground-truth annotations and the predictions of the implemented algorithms. All metrics are reported at a threshold of 0.5 The best results we obtained with the proposed methods on the test set were a mean DSC of 0.762 and median HD of 3.143		
Evaluation Data	**< no separate section >**	**< no separate section >**
3D PET/CT scans of 101 from 2 centers (CHUP - which is also present in the training set - and CHUV) Data was provided by the organizers of the challenge in August 2021 as 3D bounding boxes of size 144x144x144 containing the tumor region. As a pre-processing technique we used a z score normalization.		
Training Data	**Mechanism**	**Lineage**
3D PET/CT scans and GTVt manual contours of 224 from 5 centers (CHGJ, CHMR, CHUM, CHUS and CHUP) Data was provided by the organizers of the challenge in June 2021 as 3D bounding boxes of size 144x144x144 containing the tumor region. As a pre-processing technique we used a z score normalization.		
Caveats and Recommendations	**Warnings**	**Safety & Security**
Risks: The segmentation model can output false positive or false negative. A missed primary tumor could lead to mortality. A patient without tumor receiving radiation therapy leads to undesired side-effects and toxicities. **Inappropriate Settings: this model was trained on imaging data used for radiotherapy planning. Do not use on other image modalities or images collected for other purposes.** **Clinical Rationale: The model can not be interpret and does not provide rationale for good or bad segmentation results. The model outputs need to be approved by a doctor before they can be used.** **Inappropriate decision support: This model was trained and tested on patients from 5 different cohorts. It may not be accurate outside the target population.**		

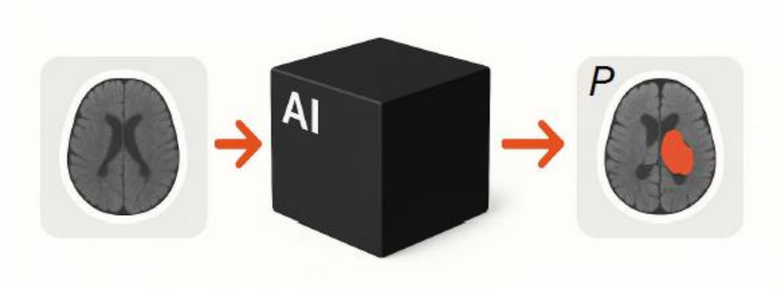
De-skilling

- *La IA: aprende rápido, comete errores y necesita supervisión (constante).*

- podría ser necesaria la **intervención humana**, pero esto se convierte en un problema si las personas ya no son competentes en las tareas que deben realizar.
- es fundamental **mantener la capacidad de supervisar las actividades de la IA** e intervenir cuando sea necesario, lo que exige una comprensión profunda de sus propias áreas de competencia.

- *La clave para un futuro exitoso en la era de la IA reside en entender su naturaleza dual y en saber navegar con prudencia hacia una convivencia armoniosa entre la inteligencia humana y la artificial***.**

Retos clínicos

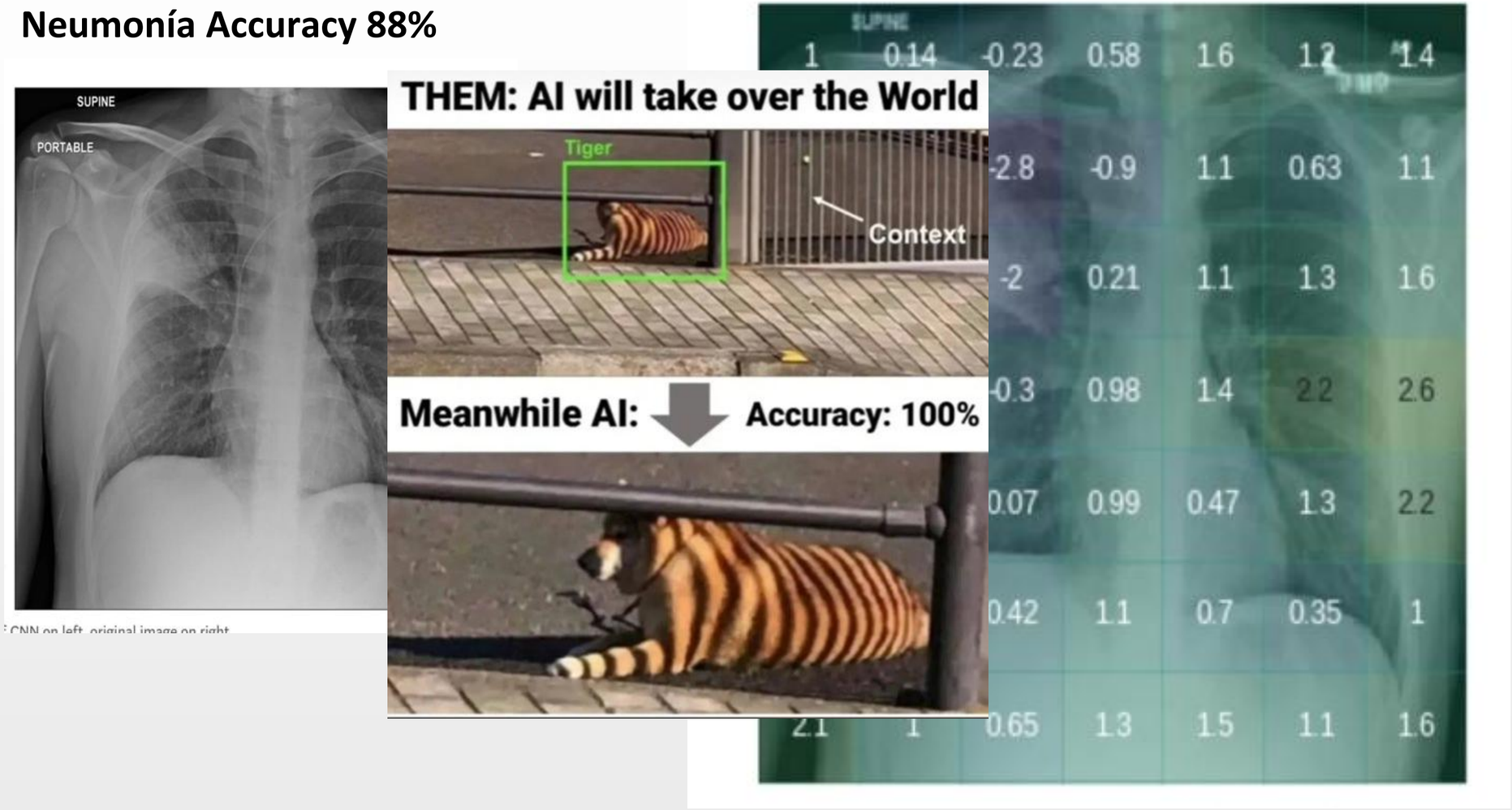


EU AI Act, Article 13:
"High-risk AI systems shall be designed and developed in such a way to ensure that their operation is sufficiently transparent to enable users to interpret the system's output and use it appropriately. An appropriate type and degree of transparency shall be ensured …"

- Interpretabilidad: el propio modelo (como funciona el modelo desde dentro)
- Explicabilidad: las razones de un predicción específica
- Incertidumbre: las limitaciones del modelo o de las predicciones

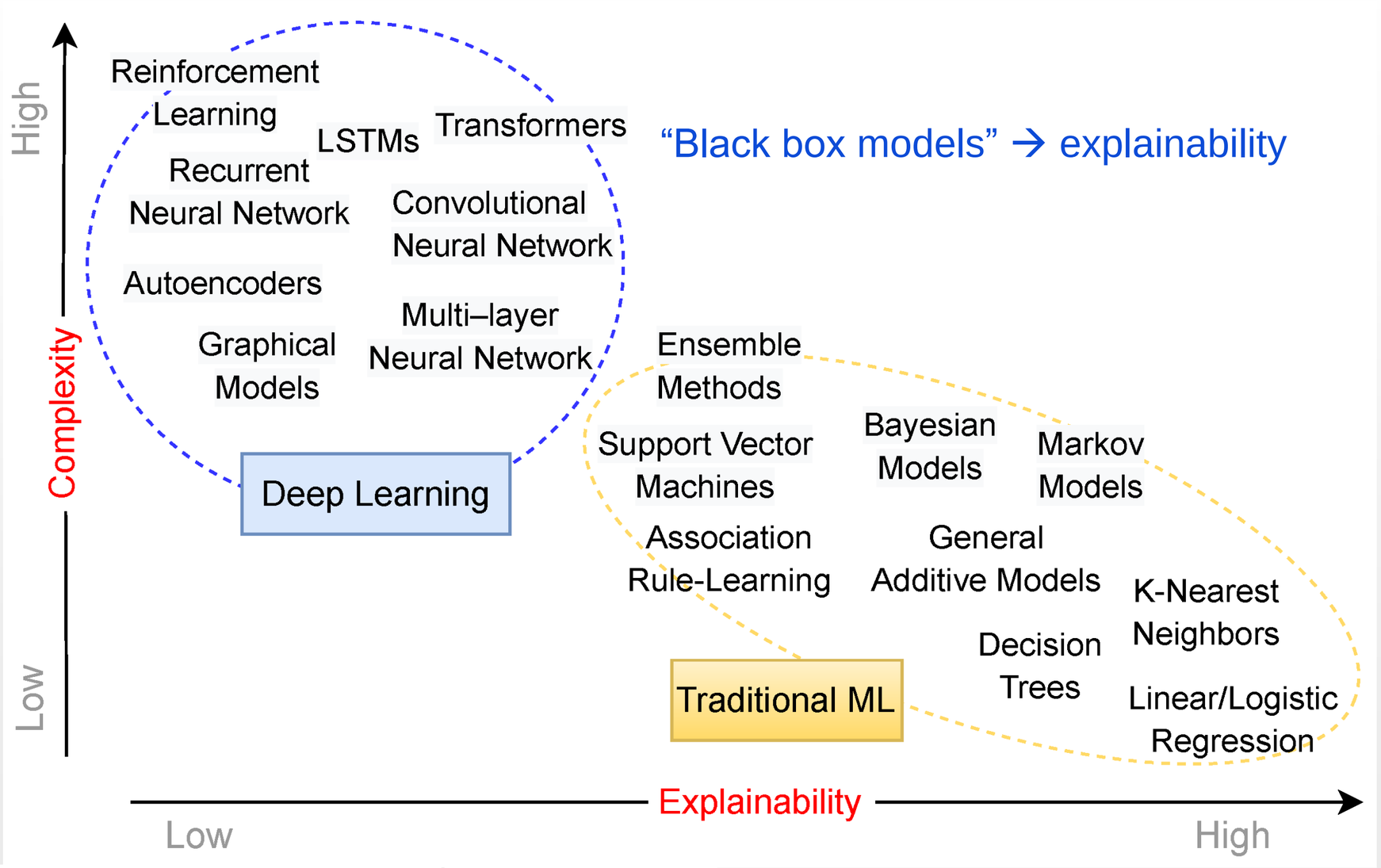
IA Robusta y accesible

Neumonía Accuracy 88%



CNN on left, original image on right

La interpretación depende del modelo

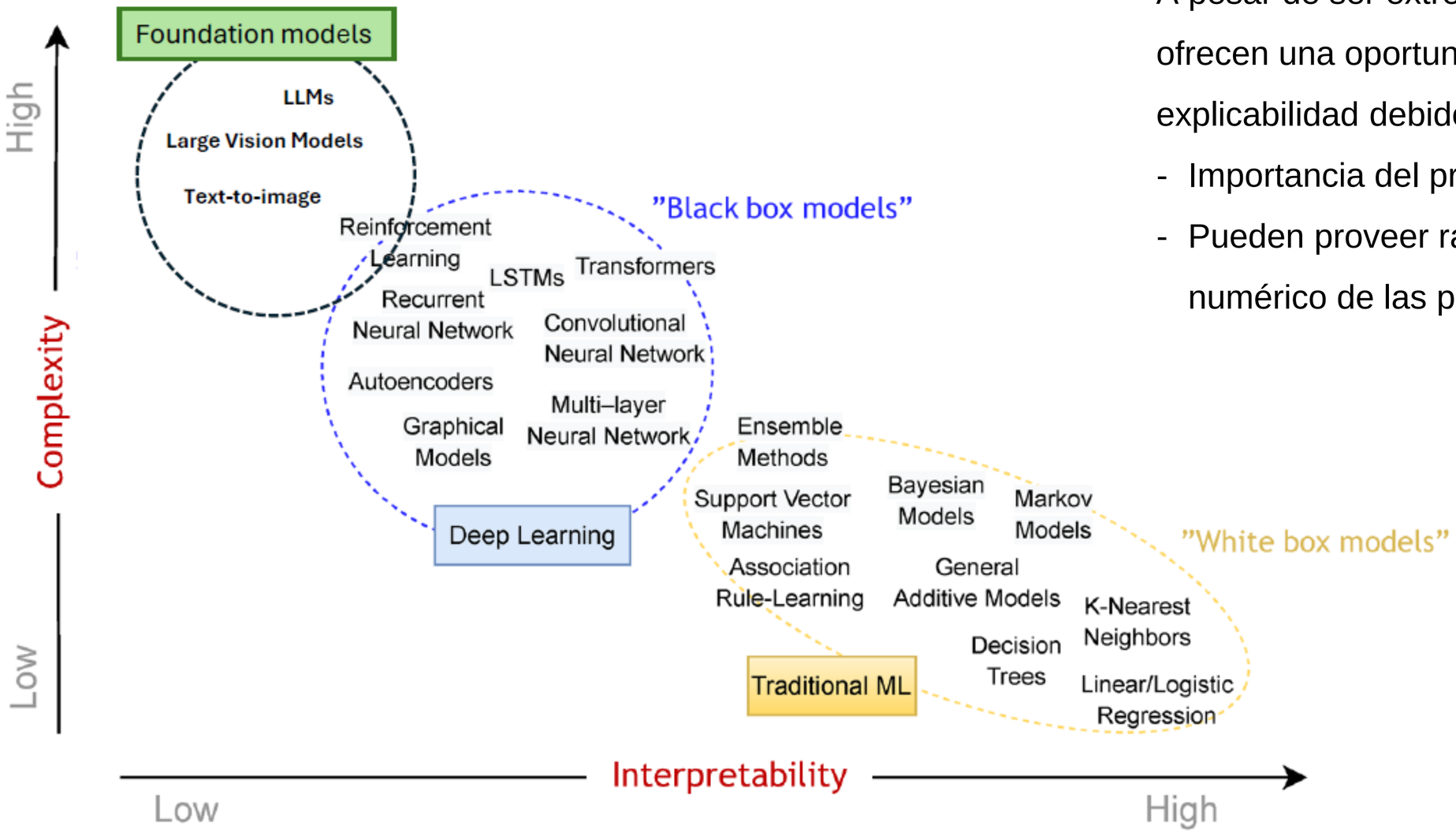


"Black box models" → explainability

"White box models"

Adapted from Clement et al, Mach. Learn. Knowl. Extr. 2023

La interpretación depende del modelo



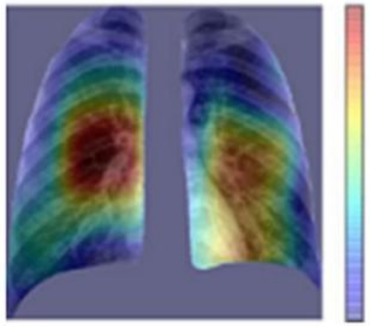
Foundation models

A pesar de ser extremadamente complejo, ofrecen una oportunidad para la explicabilidad debido a:

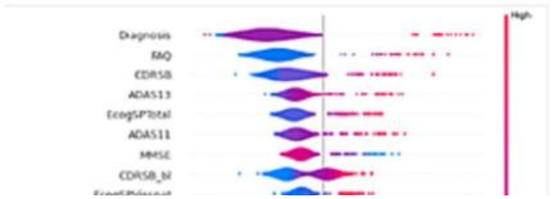
- Importancia del prompt usado
- Pueden proveer raciocinio textual/visual y numérico de las predicciones

Métodos de explicabilidad

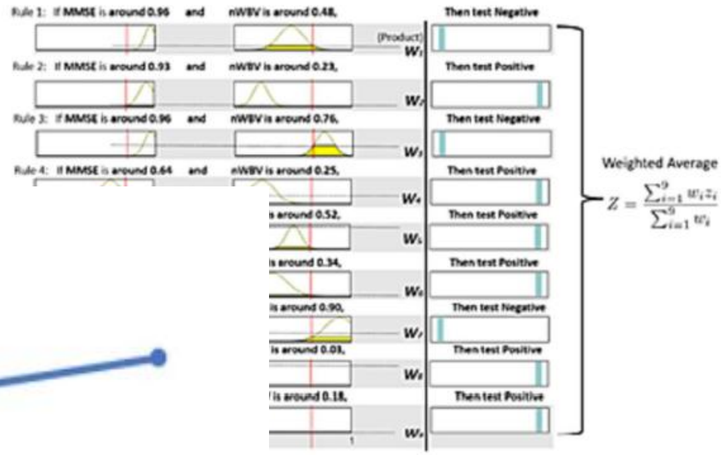
a. Visual

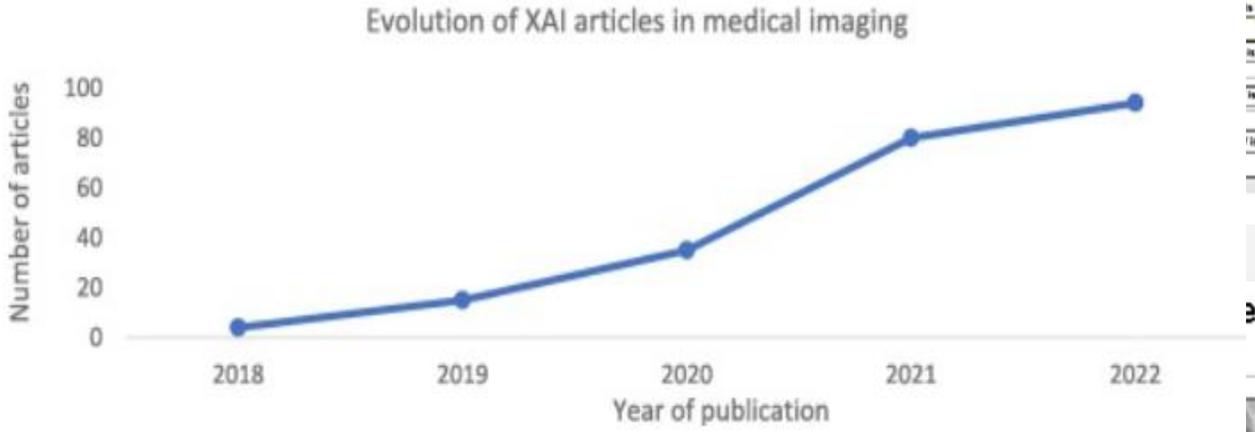


b. Numerical



c. Rule-based

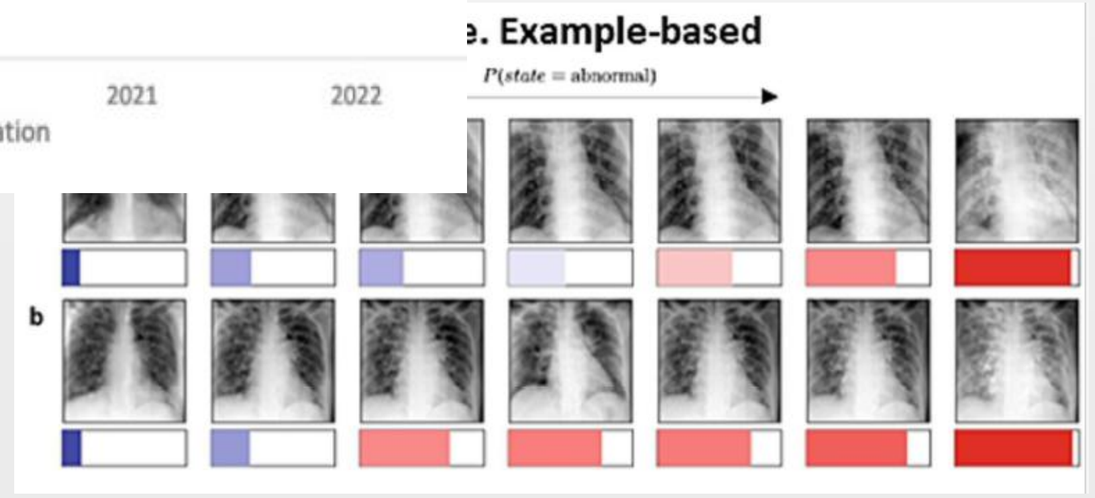




d.

Radiology output	Diagnosis: Alzheimer's (ALZ) Criteria in favour AD: 1. Hippocampal Volume is very Low. 2. Cluster Prominence is very Low in combination with Low Volume. Criteria in favour NC: 1. Entropy is High, Variance is Nominal and Contrast is not Very High The criteria supporting AD are in general stronger than those in favour for NC and therefore the predicted diagnosis is AD.

e. Example-based



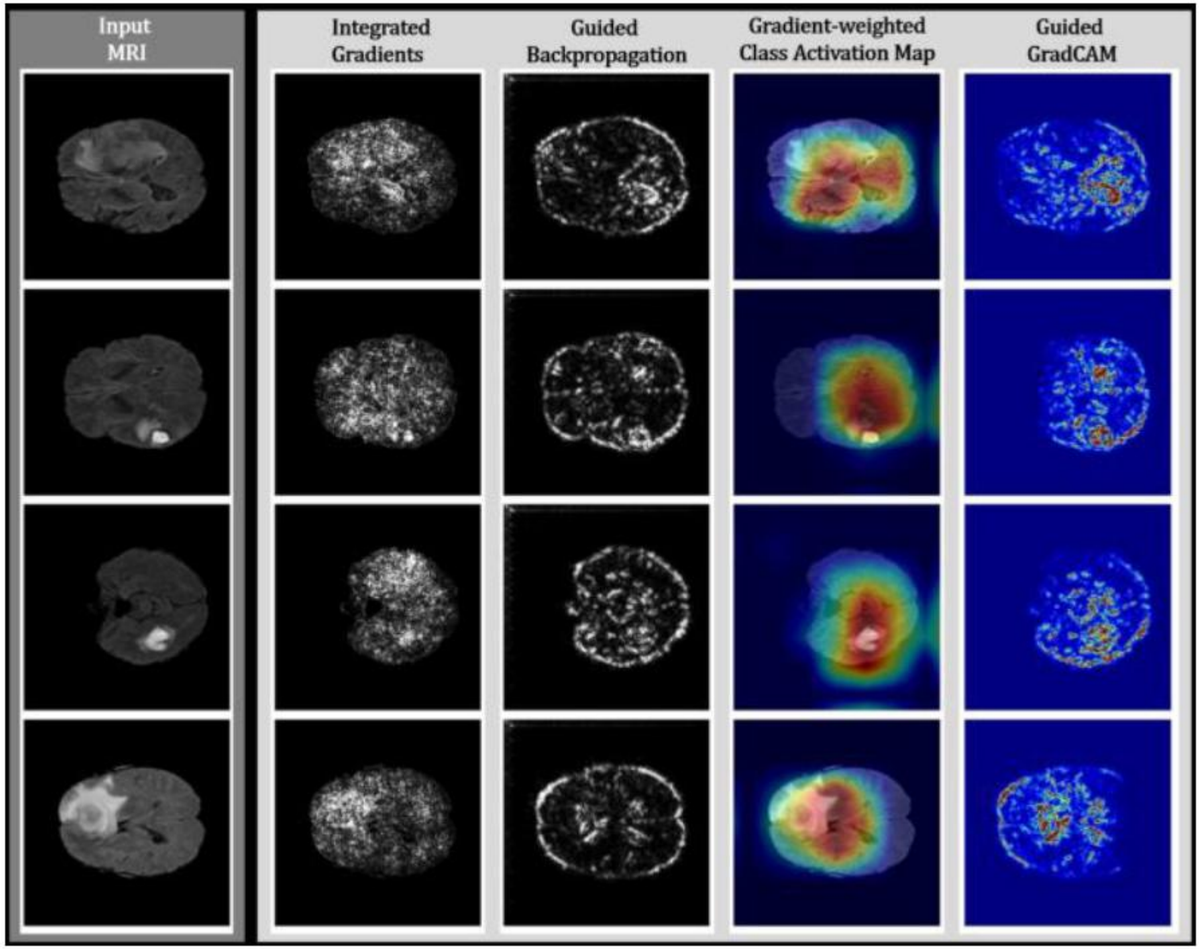
Champendal et al, EJR, 2023

Explicabilidad a través de la visualización

Mapas de atención (attention maps)

Señalan que voxeles o regiones de una imagen en la que se ha enfocado el modelo usando:

- Gradientes de imagen
- Pesos
- Bordes
- Texturas
- Patrones



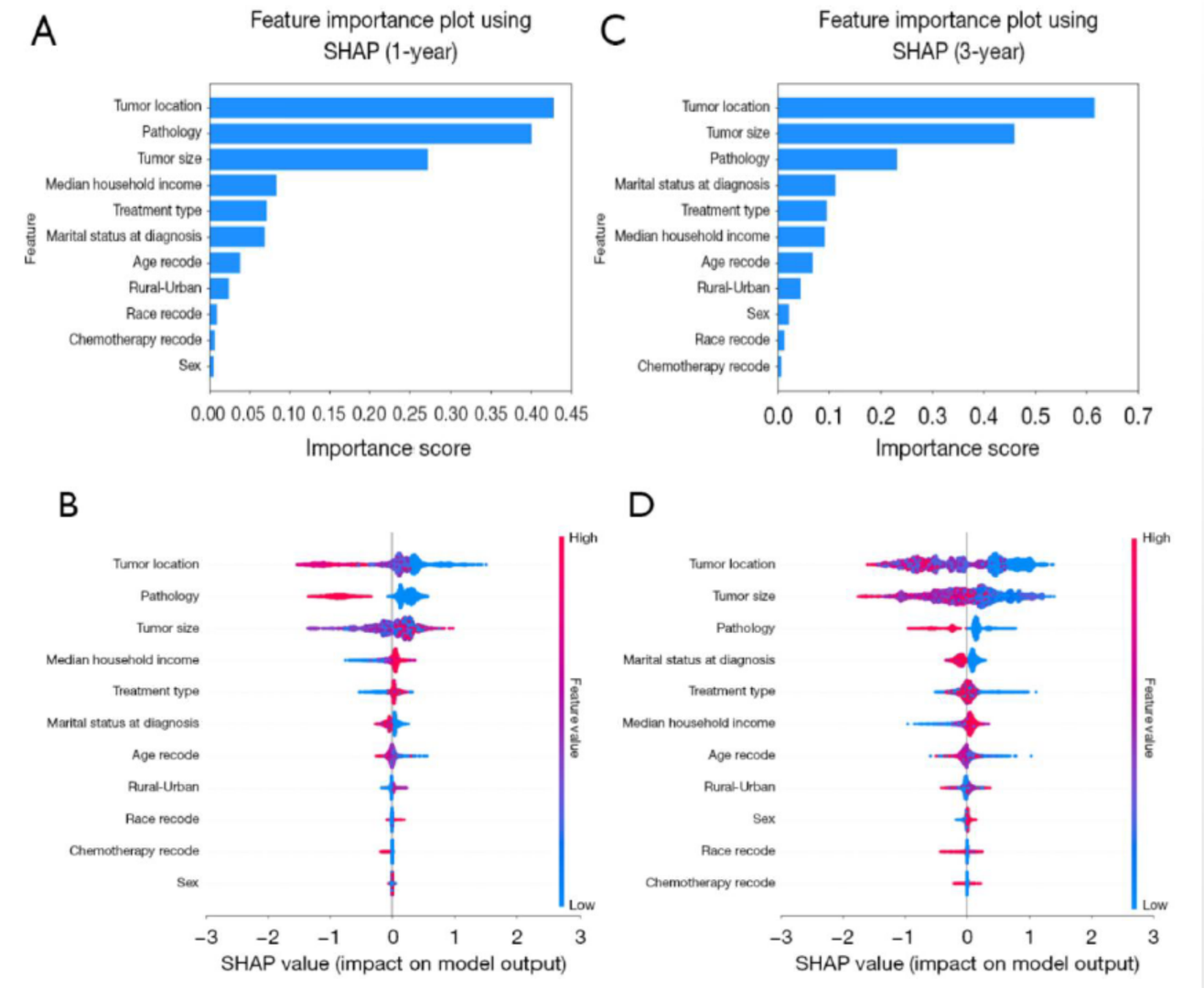
Erdur et al, Strahlenter Onk, 2024

Explicabilidad a través de rangos numéricos (feature weights)

SHAP (Shapley Additive exPlanations)

Señala la importancia y el impacto de las diferentes features:

- Datos clínicos
- Biomarcadores
- Características de pacientes
- Modalidades
- ...



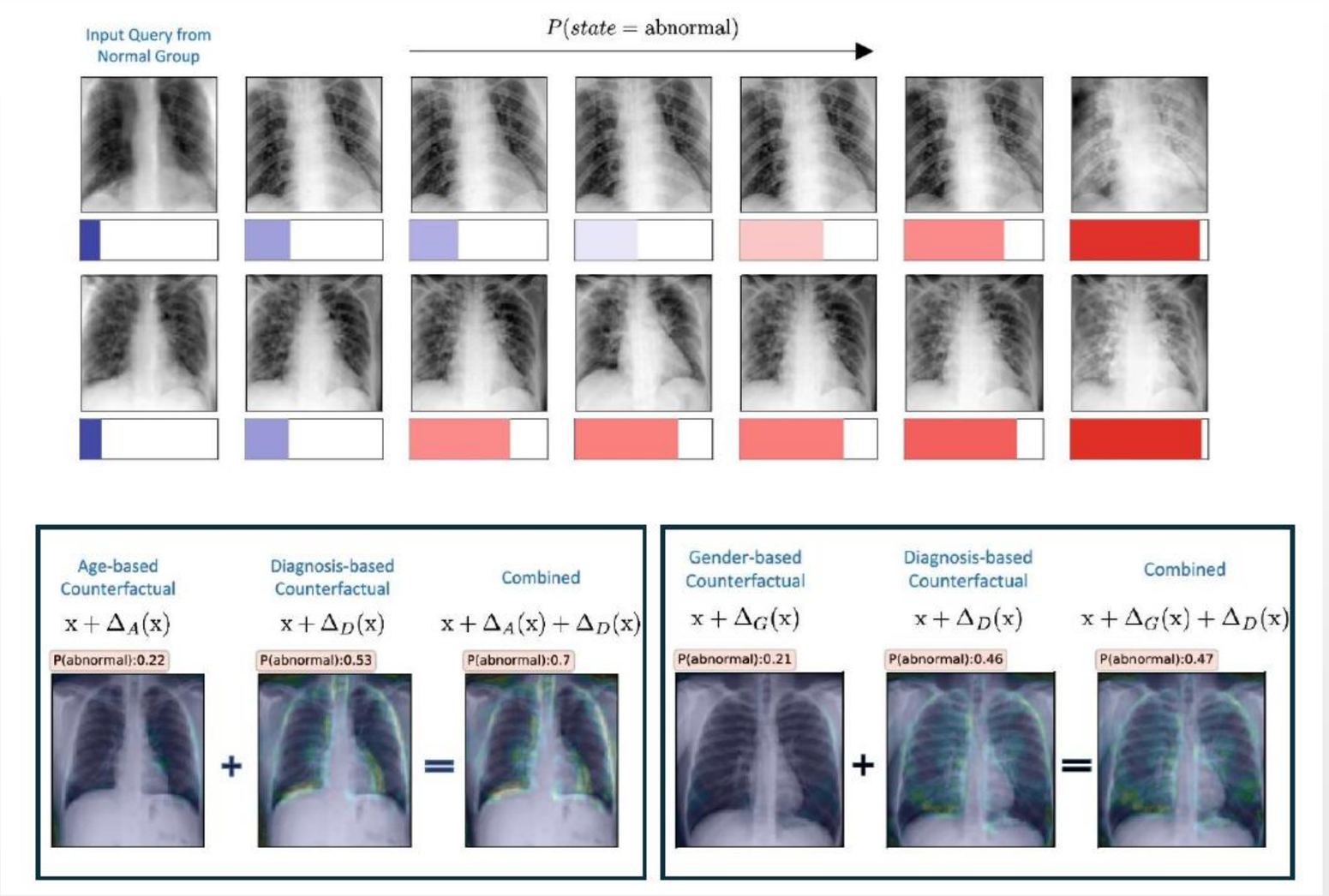
Ding et al, TLCR, 2025

Explicabilidad a través de la comparación

Genera imágenes sintéticas modificadas (“contrafactuales”) que muestran visualmente qué cambios harían que el modelo cambiara su predicción.

- Diagnóstico
- Género
- ...

Por ejemplo, si el modelo dice *“Neumonía”*, el contrafactual busca generar una imagen muy parecida, pero que el modelo clasifique como *“Normal”*.



Article | Open access | Published: 12 January 2022

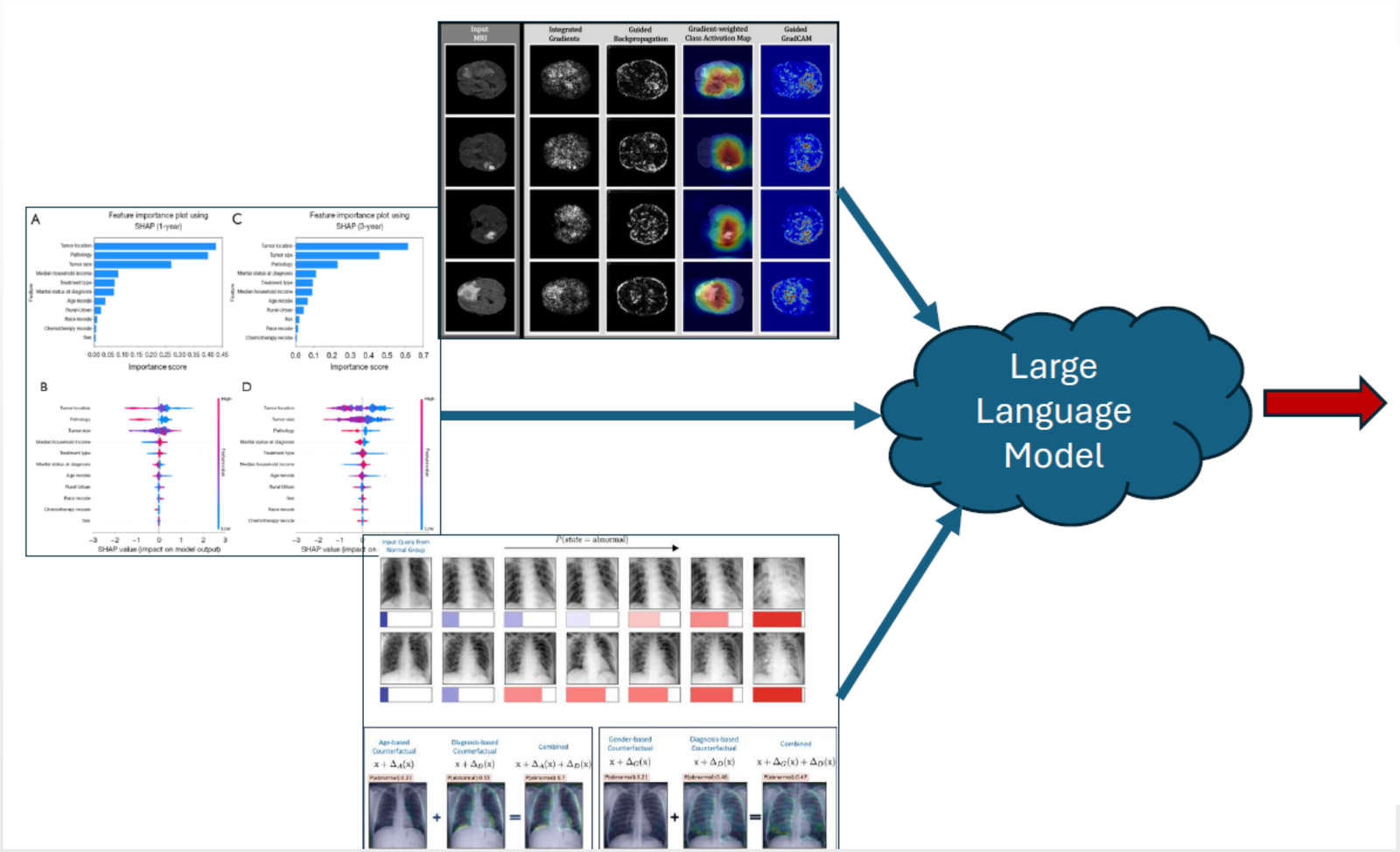
Training calibration-based counterfactual explainers for deep learning models in medical image analysis

Jayaraman J. Thiagarajan, Kowshik Thopalli, Deepta Rajan & Pavan Turaga

Explicabilidad a través de información textual

- Usa información textual para explicar las predicciones

¡Cuidado con los sesgos!



GTV primary tumor prediction:

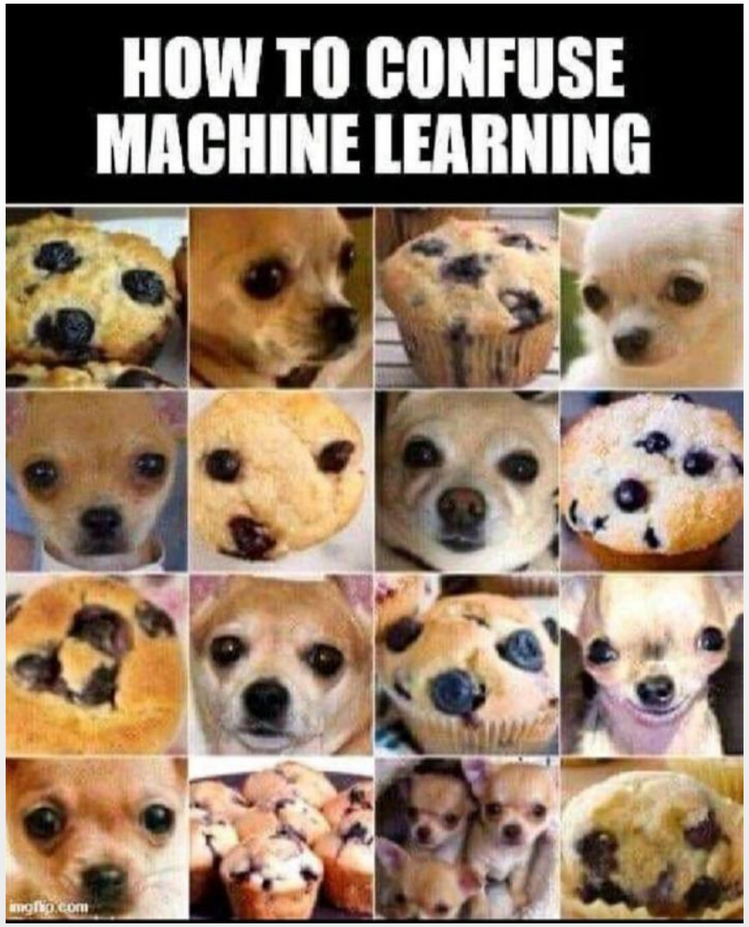
Primary drivers:
- Intense PET uptake within the right upper lobe
- Irregular pleural-based soft-tissue density on CT
- Vessel adjacency at medial border
- Spiculated edges

Exclusionary evidence:
- Preserved fissural planes
- Low-uptake contralateral lung regions
- The model reduced weight on necrotic core areas with homogeneous density

Caveats:
Attribution near shoulder artifacts and PET spillover were detected but contributed minimally. These require visual confirmation during contour QA.

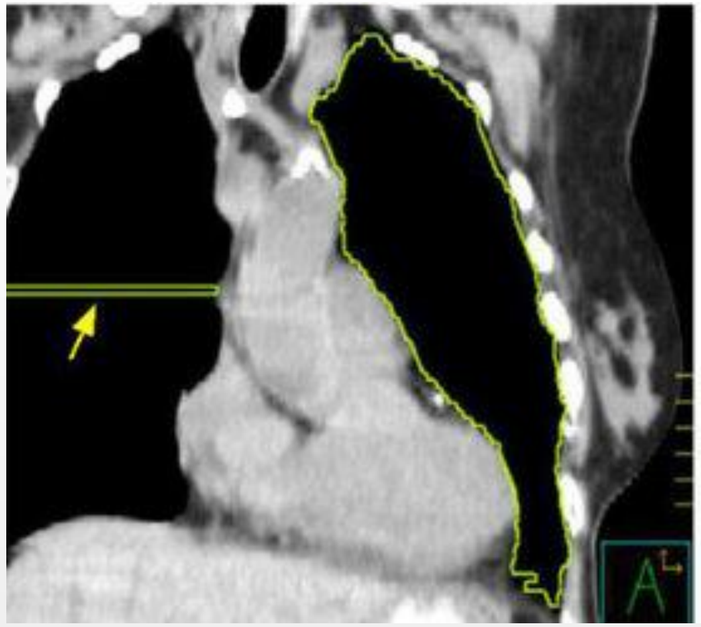
Behind the algorithm XAI. Korreman. ESTRO School

Interpretabilidad sí, pero...

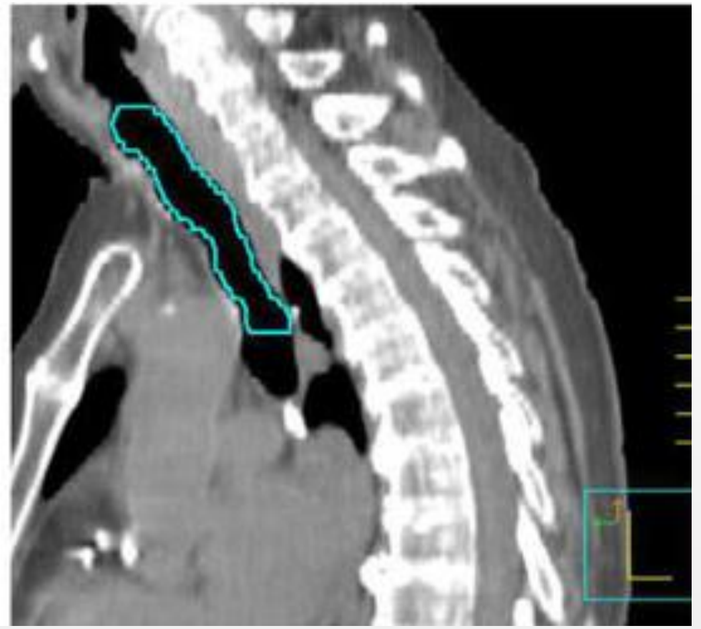


- Importancia de la calidad y cantidad de datos (positivos y negativos)
- Buena en reconocimiento de patrones, pero no entiende lo que está mirando: aprende por asociación de características.
- El contexto importa.

Errores IA vs humano



HUMANO



IA

El cerebro humano es extremadamente complejo

y tiene muy poca interpretabilidad!!

Muchas de las preocupaciones realmente son sobre las interacciones humanas con las IA, en vez de la IA por si sola...

Muchas gracias por vuestra atención