

DATOS EN EL ENTORNO DE LA FÍSICA MÉDICA

Jaime Martí Asenjo
Hospital HM Sanchinarro

INTELIGENCIA ARTIFICIAL

¿Algoritmos?



STORY BY DATA

Warren McCulloch & Walter Pitts, wrote a paper on how neurons might work; they modeled a simple neural network with electrical circuits.

1943

1949

1950s

1956

1957

1958

Donald Hebb reinforced the concept of neurons in his book, *The Organization of Behavior*. It pointed out that neural pathways are strengthened each time they are used.

The Dartmouth Summer Research Project on Artificial Intelligence provided a boost to both artificial intelligence and neural networks.

Frank Rosenblatt began work on the Perceptron; the oldest neural network still in use today.

1982

1981

1969

1959

1982

John Hopfield presented a paper to the national Academy of Sciences. His approach to create useful devices; he was likeable, articulate, and charismatic.

Progress on neural network research halted due fear, unfulfilled claims, etc.

Marvin Minsky & Seymour Papert proved the Perceptron to be limited in their book, *Perceptrons*.

Bernard Widrow & Marcian Hoff of Stanford developed models they called ADALINE and MADALINE; the first neural network to be applied to a real world problem.

1982

1985

1997

1998

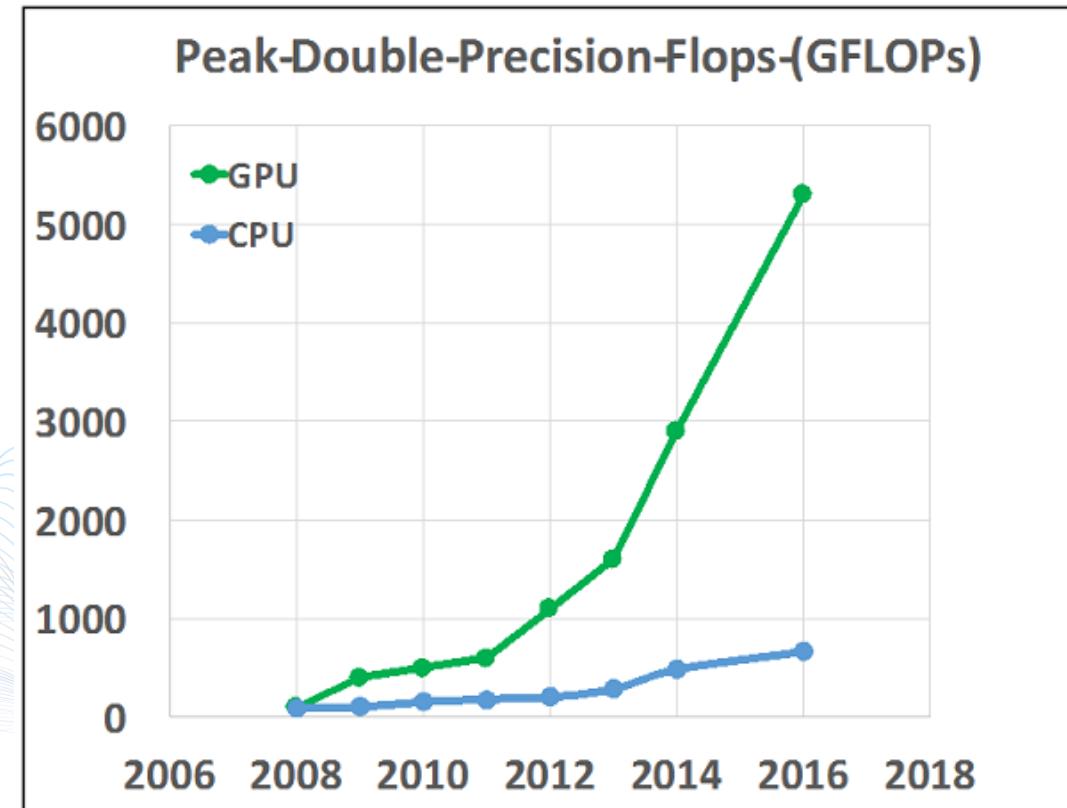
NOW

HISTORY OF NEURAL NETWORKS

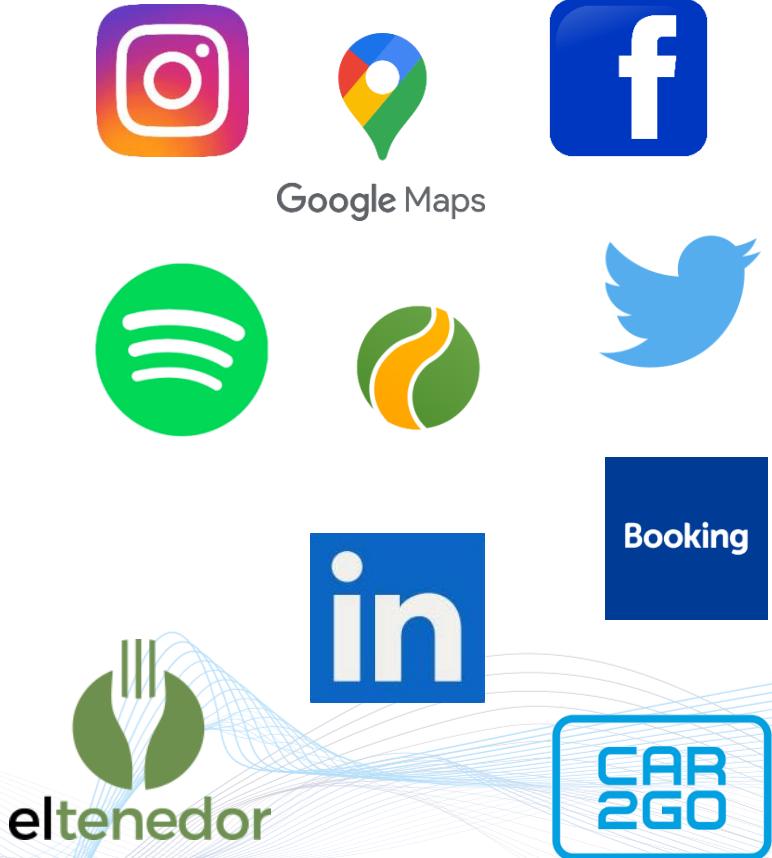
1943-2019

1943-2019

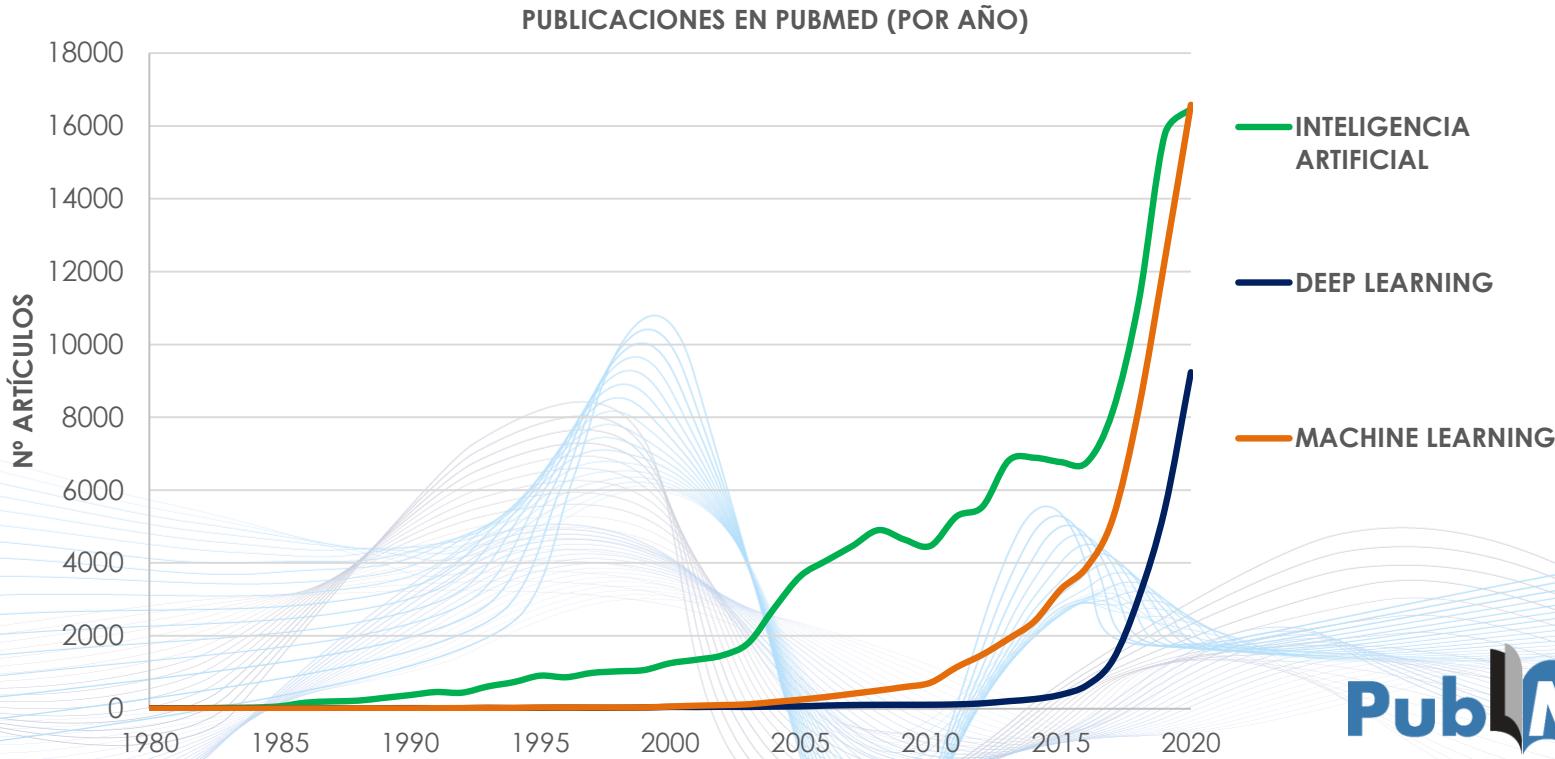
Computación



Datos



IA en investigación clínica



Datos en el entorno de la física médica

- **Sistemas de registro de dosis a paciente en radiodiagnóstico**
-
- **Control de calidad del equipamiento**
- **Verificación de tratamientos**
- **Registro de imágenes**
- **Segmentación de imágenes**
- **Dosimetrías**
- **Planes de tratamiento**
- **Sistemas de IGRT**
- **Sistemas de SGRT**
- **Sistemas de control respiratorio**
-



Tipos de datos

- **TEXTUAL**

Historias clínicas, informes radiológicos...

- **AUDIOVISUAL**

Imágenes diagnósticas, procedimientos...

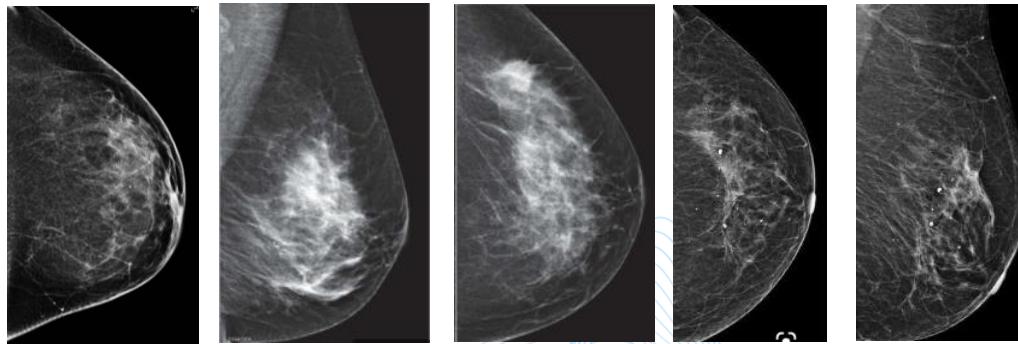
- **NUMÉRICO**

Sistemas que generan datos numéricos, bases de datos clínicas...

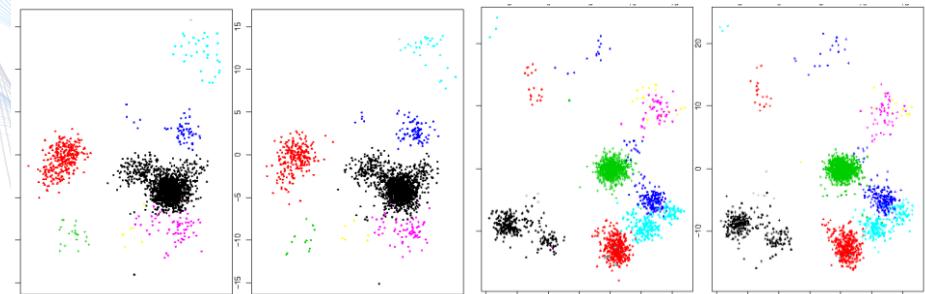
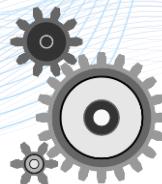


Tipos de datos

- No Anotados (Aprendizaje No Supervisado)

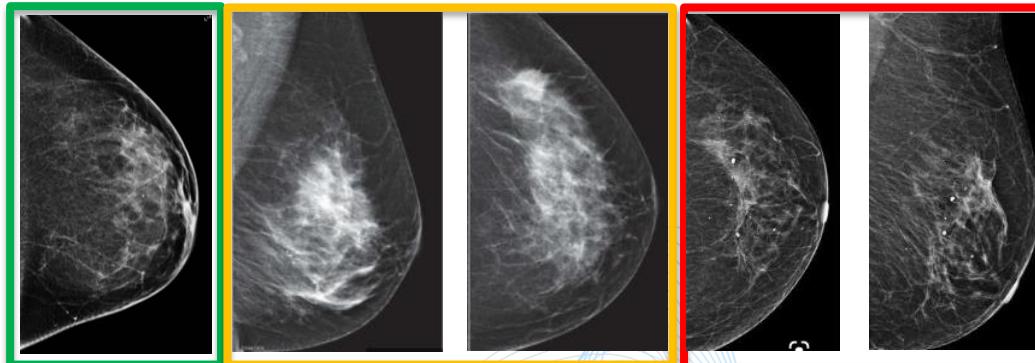


DATOS



Tipos de datos

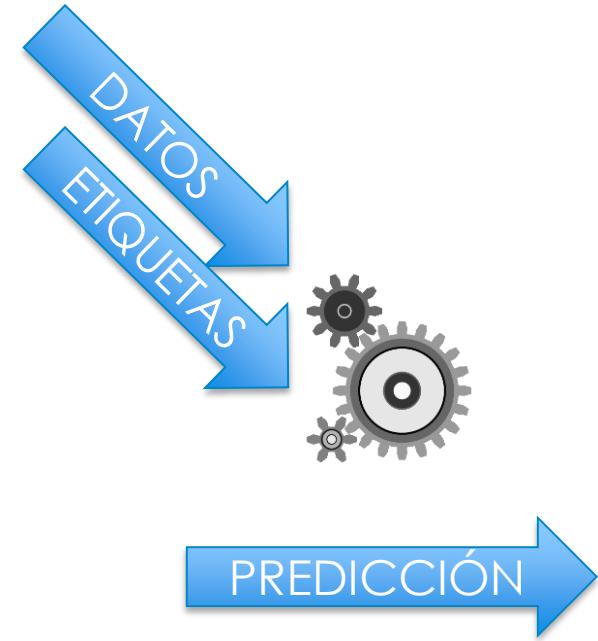
- Anotados (Aprendizaje Supervisado)



Normal

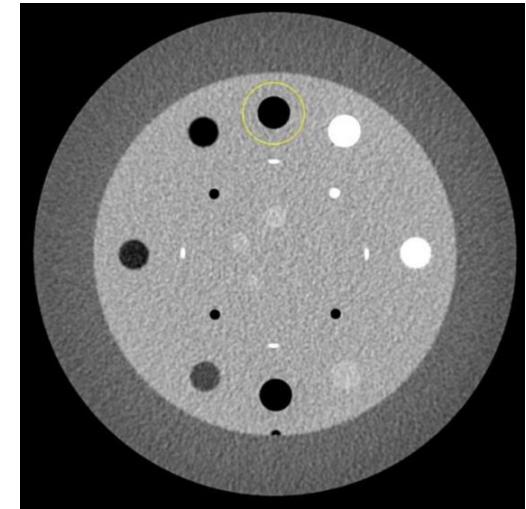
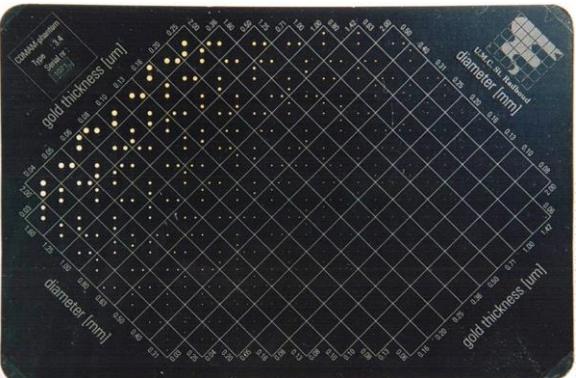
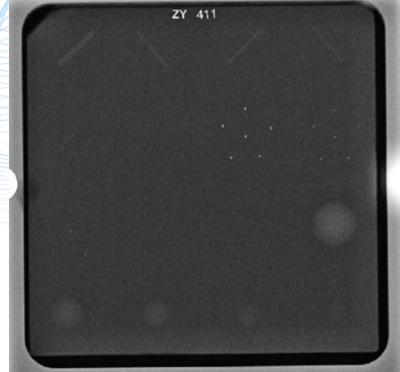
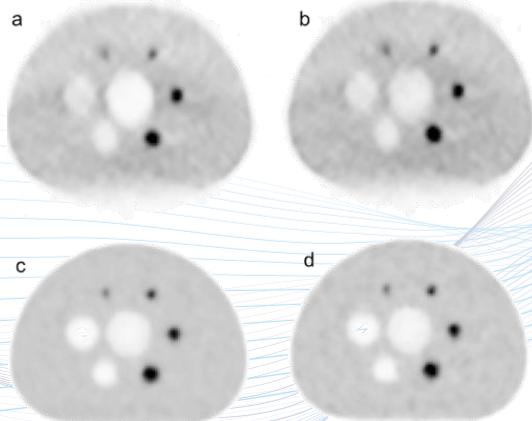
Anormal

Patológica



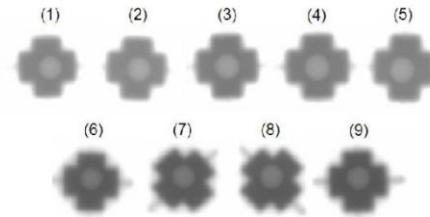
Control de calidad

- Radiodiagnóstico, Medicina Nuclear, Radioterapia
- Datos mixtos (numéricos, imágenes...)

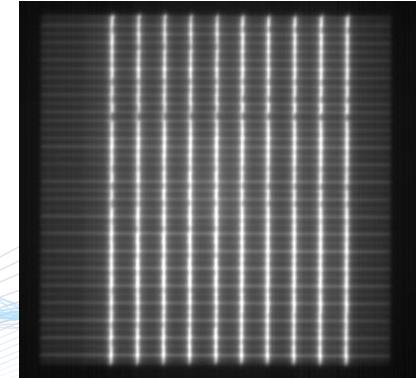
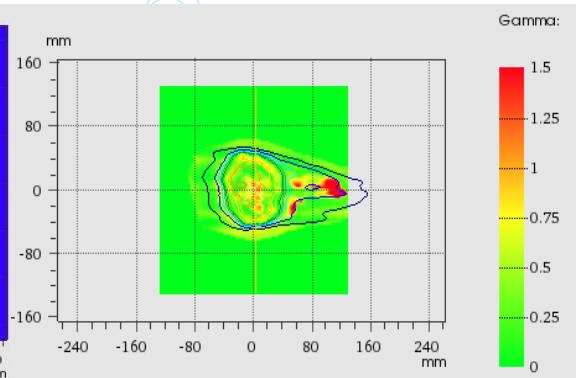
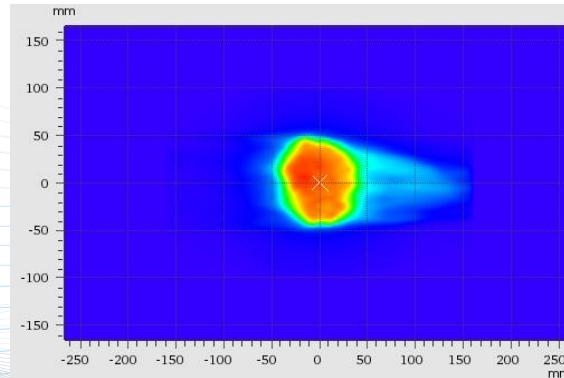


Control de calidad

- Radiodiagnóstico, Medicina Nuclear, Radioterapia
- Datos mixtos (numéricos, imágenes...)



Angle (deg)	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Gantry	0	50	130	230	310	0	0	0	0
Couch	0	0	0	0	0	270	315	45	90



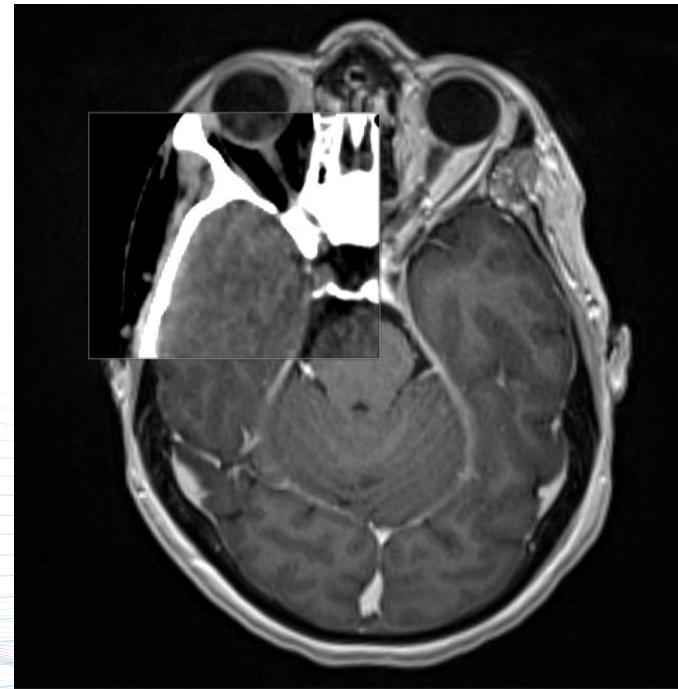
Sistemas de registro de dosis a paciente

- Gran cantidad de datos exportables y estructurados
- Etiquetados en cuanto a procedimiento, protocolo, equipo, operador...
- Variables categóricas y continuas

Descripción de la serie	Tipo de serie	Protocolo	kV	Corriente tubo (mA)	Corriente tubo máx. (mA)	Exposición (mAs)
ABDOMEN con CTE	Spiral	ABDOMEN con CTE	120	498	500	3710.10

Tiempo de exposición (ms)	mAs efectivos	Tiempo de exposición / Rotación (s)	Factor de paso	Anchura colim. único (mm)	Anchura total colim. (mm)	Índice de ruido
7450.00	306.27	0.50	0.81	5.00	40.00	10

Registro de imágenes



- Registro de imágenes para la definición de volúmenes
- Registro de imágenes para el posicionamiento en radioterapia

Transformation matrix:

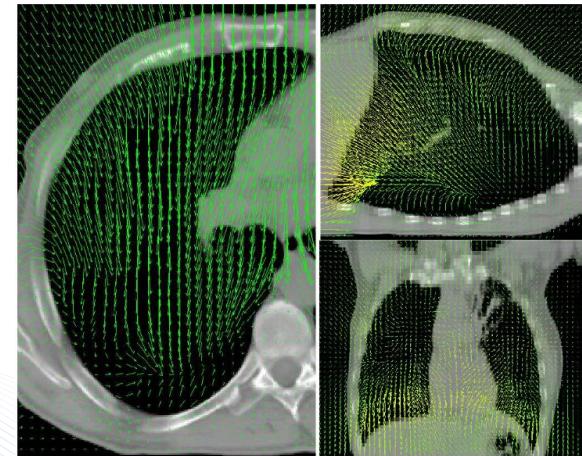
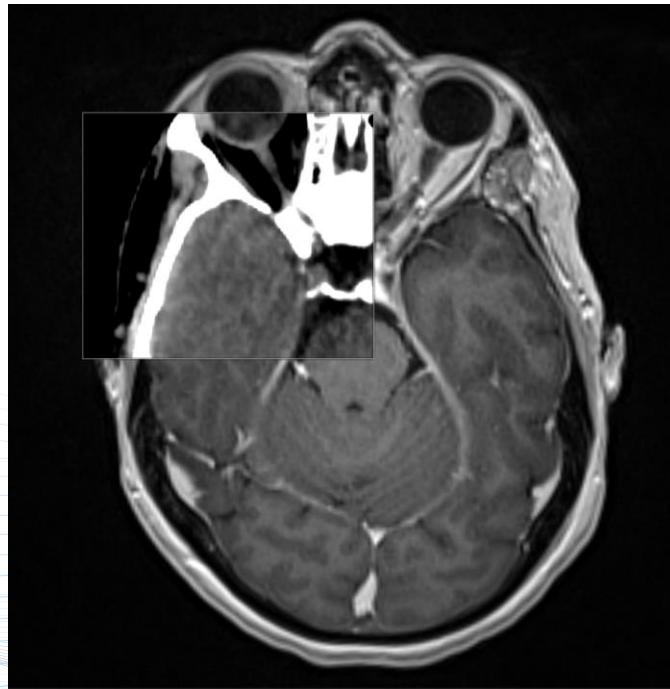
0.97	0.01	0.23	-0.33
-0.01	1.00	0.01	-3.15
-0.23	-0.02	0.97	-3.68
0.00	0.00	0.00	1.00

Rigid transform:

	Translation [cm]	Rotation [deg]
Right-Left:	-0.33	Pitch: -0.7
Inf-Sup:	-3.15	Roll: 13.3
Post-Ant:	-3.68	Yaw: -0.7

Registro de imágenes

- Registro de imágenes para la definición de volúmenes
- Registro de imágenes para el posicionamiento en radioterapia

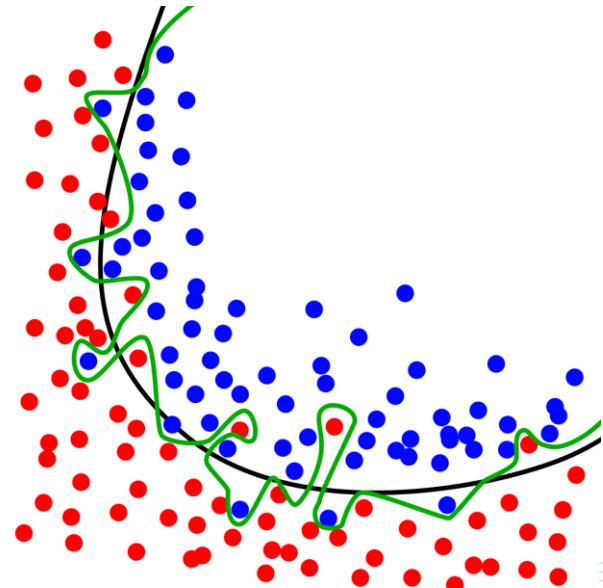


Datos en el entorno de la física médica

- **Accesibles:** Exportables, exportables en volumen, tipo de archivo (dicom, csv, propio del fabricante...?)
- **Estructurados:** Información accesible de forma iterativa
- **Anotados:** En el contexto del proyecto que queremos realizar
- **Volumen:** ¿Suficiente para el tipo de proyecto que queremos realizar?

Volumen de datos

- **Regla del 10 (ML):** Al menos 10 datos por cada característica que evaluemos
- **Clasificación de imágenes (DL):** Al menos 1k imágenes por clase
- **Segmentación imágenes (DL):** Al menos 5k imágenes por clase*

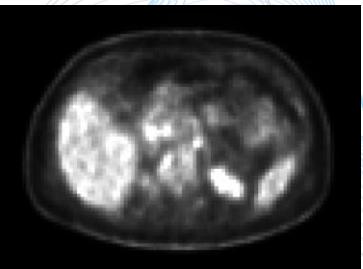
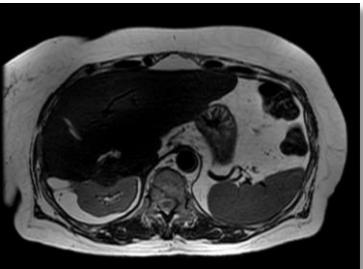
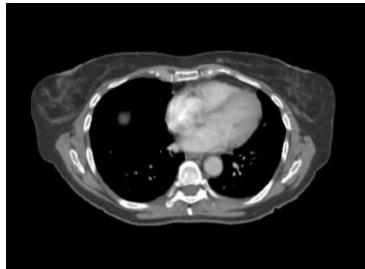


* Deep Learning (Ian J. Goodfellow, Yoshua Bengio and Aaron Courville), MIT Press, 2016.

Imagen diagnóstica

Formato estándar DICOM

IMAGEN + METADATOS (CABECERA)



0002,0002 Media Storage SOP Class UID: 1.2.840.10008.5.1.4.1.1.2
0002,0003 Media Storage SOP Inst UID: 1.2.392.200036.9116.2.6.1.44063.1797790409.1593385531.15309
0002,0010 Transfer Syntax UID: 1.2.840.10008.1
0002,0012 Implementation Class UID: 1.2.826.0.1.3680043.1.2.100.6.40.0.76
0002,0013 Implementation Version Name: DicomObjects.NET
0008,0005 Specific Character Set: ISO_IR_100
0008,0008 Image Type: ORIGINAL\PRIMARYAXIAL
0008,0016 SOP Class UID: 1.2.840.10008.5.1.4.1.1.2
0008,0018 SOP Instance UID: 1.2.392.200036.9116.2.6.1.44063.1797790409.1593385531.15309
0008,0020 Study Date: 20200629
0008,0021 Series Date: 20200629
0008,0022 Acquisition Date: 20200629
0008,0023 Image Date: 20200629
0008,0030 Study Time: 075807.000
0008,0031 Series Time: 080228.991
0008,0032 Acquisition Time: 080414.250
0008,0033 Image Time: 080415.487
0008,0050 Accession Number: 2388
0008,0060 Modality: CT
0008,0070 Manufacturer: TOSHIBA
0008,0080 Institution Name: HMS
0008,0081 Institution Address: TOSHIBA_MEC
0008,0090 Referring Physician's Name:
0008,1010 Station Name: HMS_TAC_TOS_3
0008,103E Series Description: PANCREAS ART RTP 3.0 CE
0008,1040 Institutional Department Name: RT
0008,1090 Manufacturer's Model Name: Aquilion/LB
0010,0010 Patient's Name: XXXXXXXXXX
0010,0020 Patient ID: XXXXXXXXXX
0010,0030 Patient's Birth Date: 19570922
0010,0040 Patient's Sex: F
0010,1010 Patient's Age: 062Y
0018,0010 Contrast/Bolus Agent: CE
0018,0015 Body Part Examined: PANCREAS
0018,0022 Scan Options: HELICAL_CT
0018,0050 Slice Thickness: 3.0
0018,0060 KVP: 120
0018,0090 Data Collection Diameter: 550.00

Cabecera DICOM

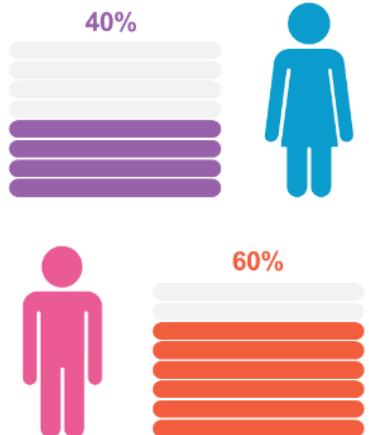
- **Información demográfica**

Edad

Sexo

Altura

Peso



- **Información adquisición**

Modalidad (TC, RM, PET-TC, MAM, GC...)

Descripción del estudio

Protocolo empleado

Técnica (kV, mA, tiempos de adquisición, pitch, secuencia, isótopo...)

Fecha del estudio

Centro del estudio

Cabecera DICOM

- **Información descriptiva imagen**

Profundidad valor de pixel

Tamaño de pixel/vóxel

Tamaño de la imagen

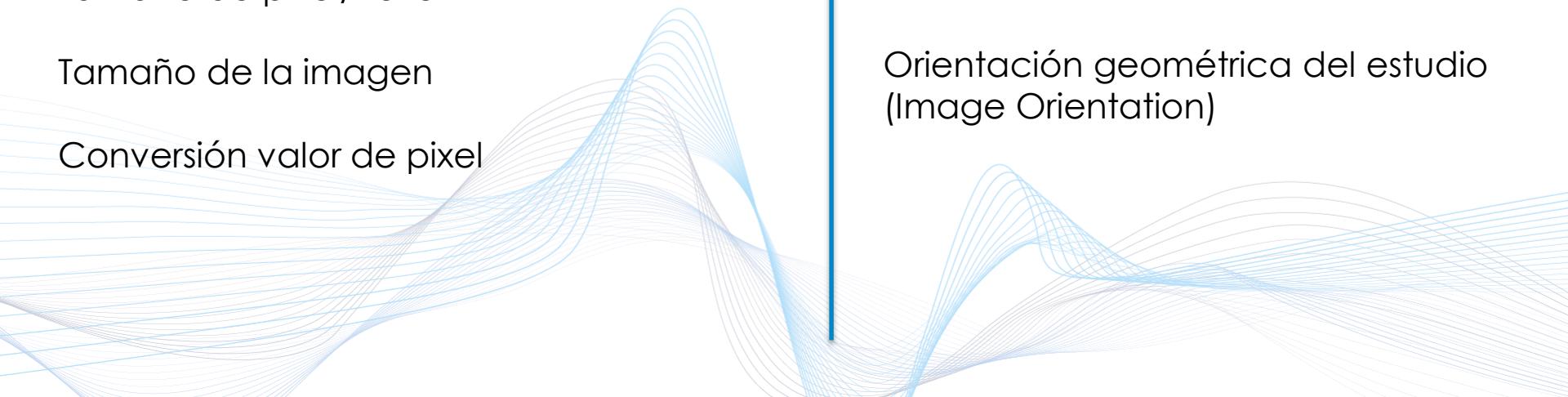
Conversión valor de pixel

- **Información geométrica**

Tipo de proyección

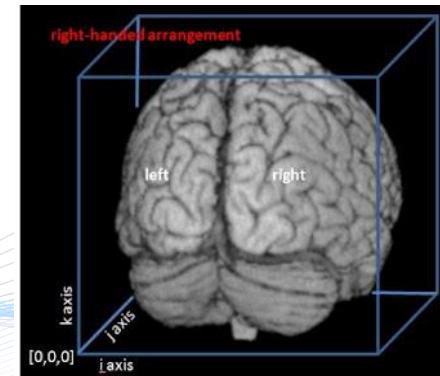
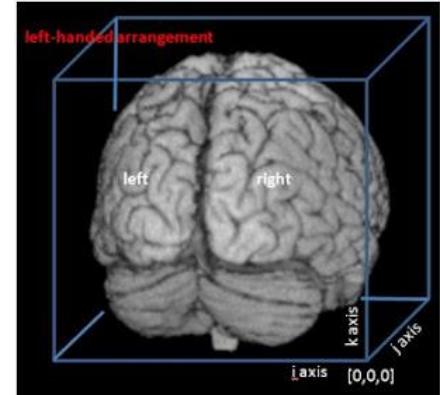
Origen del estudio (Image Position)

Orientación geométrica del estudio
(Image Orientation)



Otro tipo de archivos: NIFTI

- Neurolmaging Informatics Technology Initiative
- Formato de código abierto para estudios 3D
- Extensión (.nii / .nii.gz)
- Originalmente RM cerebral, también otros formatos
- Cabecera con metadatos



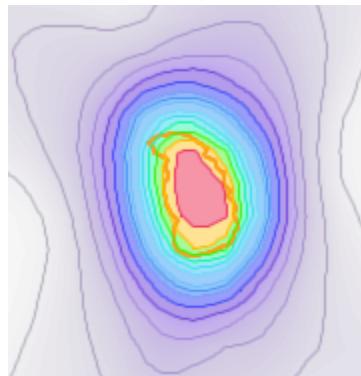
Otro tipo de archivos DICOM: RTDose

- Tratamiento similar a un archivo de imagen (Profundidad, escalado, unidades, tamaño y resolución de la malla de cálculo)
- Orientación y posición respecto al sistema de coordenadas de la sala (ImageOrientation y ImagePosition)
- Escalado: Conversión del pixel a dosis, unidades de dosis
- Asociado a un RTPlan y a un estudio imagen

Bitdepth (2^N)
Rows
Columns
PixelSpacing
SliceThickness

ImageOrientation
ImagePosition

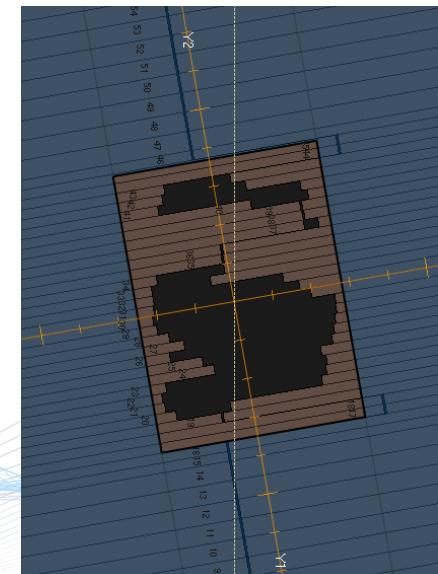
DoseGridScaling
DoseUnits



Otro tipo de archivos DICOM: RTPlan

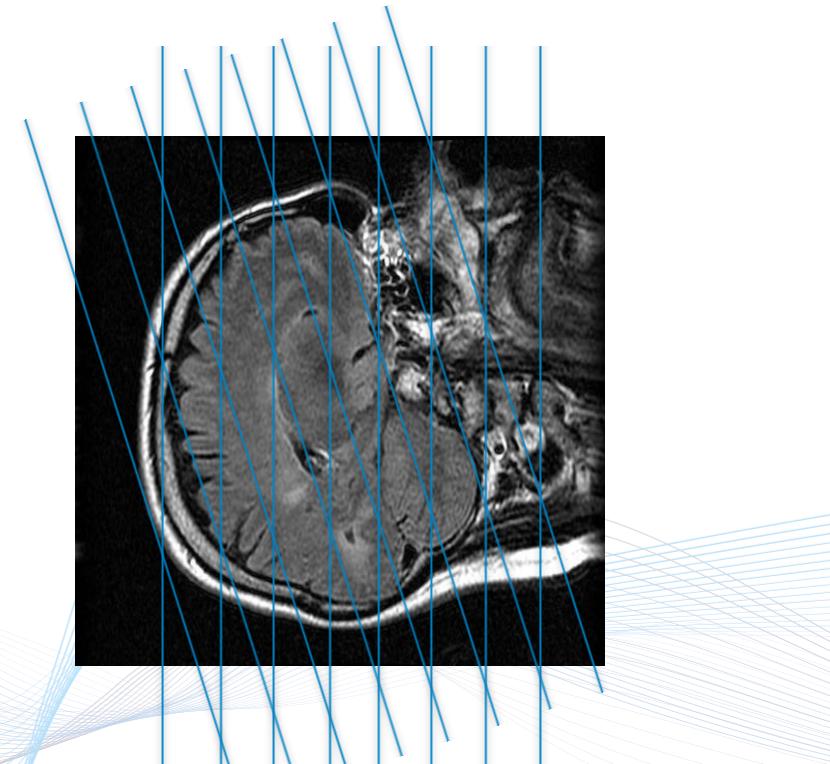
- Nombre del plan
- Tipo de plan (VMAT, IMRT, 3D...)
- Fraccionamiento
- Posición del Isocentro (sistema de coordenadas de la sala)
- Número de haces
- Parámetros del haz (UM, Gantry, colimador...)
- Posicionamiento del ML por puntos de control
- Asociado a un estudio de imagen

BeamSequence
FractionGroupSequence



Otro tipo de archivos DICOM: RTStruct

- Relación de estructuras definidas sobre el estudio de imagen
- Orientación y posición respecto al sistema de coordenadas de la sala (ImageOrientation y ImagePosition)
- Definidas a partir de vértices en el espacio de coordenadas de la sala.
- Asociado a un estudio de imagen



Otro tipo de archivos DICOM: RTStruct

- Relación de estructuras definidas sobre el estudio de imagen
- Orientación y posición respecto al sistema de coordenadas de la sala (ImageOrientation y ImagePosition)
- Definidas a partir de vértices en el espacio de coordenadas de la sala.
- Asociado a un estudio de imagen

StructureSetROISequence
ROIContourSequence
StructurSetROISequence

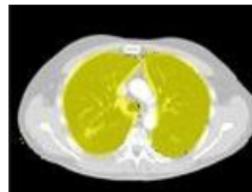


	1
1	13.3590
2	-134.4240
3	-541.5000
4	17.8983
5	-134.0235
6	-541.5000
7	19.4114
8	-133.6244
9	-541.5000
10	20.9245
11	-132.8253
12	-541.5000
13	22.4376
14	-132.1805
15	-541.5000
16	23.9507
17	-131.2568
18	-541.5000

Herramientas

File Exchange

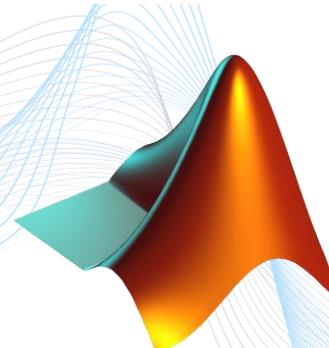
MATLAB Central ▾ | Files | Authors | My File Exchange ▾ | Publish | About



Use DICOM RT for 3D Semantic Segmentation of Medical images

version 1.0.1 (2.28 MB) by [Takuji Fukumoto](#) STAFF

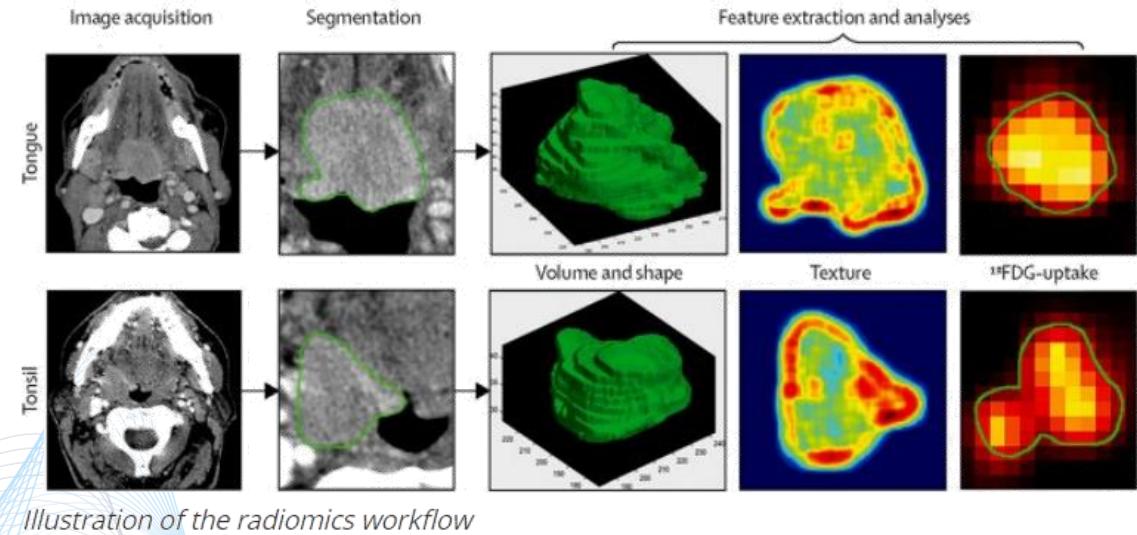
Apply 3D UNet (Semantic Segmentation) into medical CT image without wasting time for labeling.



A minimal Python library for RT Struct manipulation

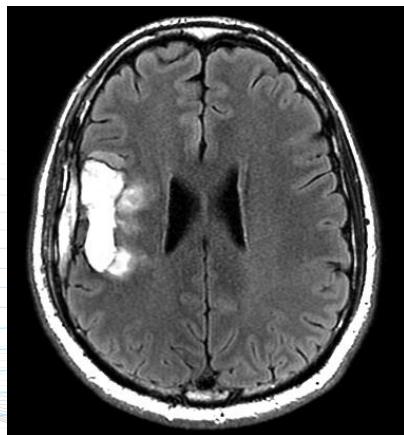


RADIÓMICA

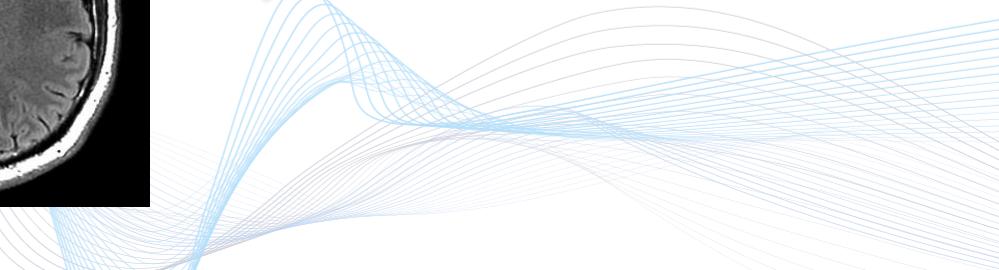


RADIÓMICA

Método para la extracción de un gran número de características a partir de una imagen médica. Estas características potencialmente pueden ocultar patrones ocultos al análisis visual y relevantes para el diagnóstico o la respuesta terapéutica a un determinado tratamiento.



$$(x_1, x_2, x_3, \dots, x_i)$$

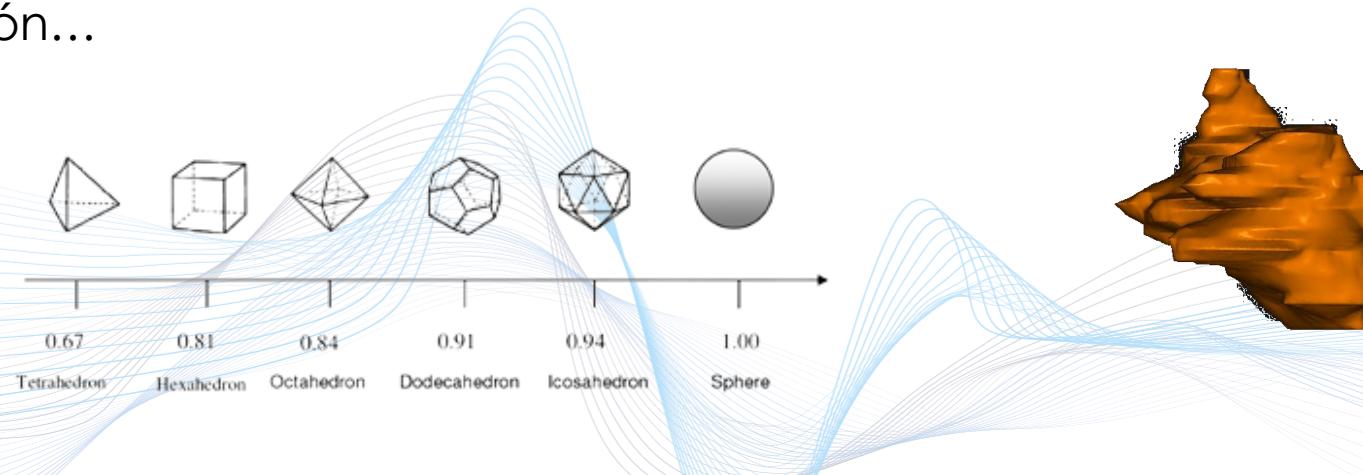


ESTADÍSTICOS:

Media, máximo, mínimo, Percentiles (10, 75, 90...), Mediana, RMS, SD, Kurtosis, Varianza...

GEOMÉTRICOS:

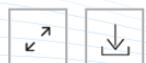
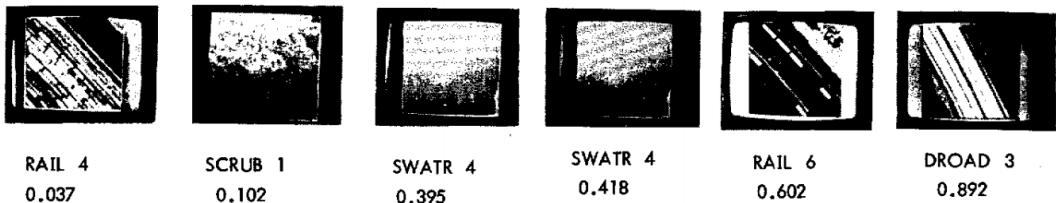
Superficie, volúmen, relación A/V, esfericidad, ejes principales, elongación...



TEXTURAS

Robert Haralick et al. 1973

- Identificación de objetos y regiones de interés (microfotografía, imágenes aéreas, satélite...)



PubMed

Textural Features for Image Classification

ROBERT M. HARALICK, K. SHANMUGAM, AND ITS'HAK DINSTEIN

Abstract—Texture is one of the important characteristics used in identifying objects and regions of interest in an image, whether the image be a photomicrograph, an aerial photograph, or a satellite image. This paper describes some easily computable texture measures based on gray-tone spatial dependencies, and illustrates their applications in category-identification tasks of three different kinds of images data: photomicrographs of five kinds of sandstones, 1:20 000 panoramic aerial photographs of eight land-use categories, and Earth Resources Technology Satellite (ERTS) multispectral imagery containing seven land-use categories. We use two kinds of decision rules: one for which the decision regions are convex polyhedra (a piecewise linear decision rule), and one for which the decision regions are rectangular parallelpipedoids (a min-max decision rule). In each experiment the data set was divided into two parts, a training set and a test set. Test set identification accuracy is 89 percent for the photomicrographs, 82 percent for the aerial photographic imagery, and 83 percent for the satellite imagery. These results indicate that the easily computable textural features probably have a general applicability for a wide variety of image-classification applications.

I. INTRODUCTION

WITH THE ADVENT of high-speed general-purpose digital computers it is becoming possible to perform mathematical or algorithmic processes on pictorial data from images of photographic quality. In most of these processes, the pictorial information is represented as a function of two variables (x, y). The image in its digital form is usually stored in the computer as a two-dimensional

Manuscript received April 25, 1973; revised June 12, 1973. This work was supported by the NASA Goddard Space Flight Center under Contract NAS-5-1882, and the Geographic Sciences Laboratory of the U.S. Army Engineer Topographic Laboratories under Contract DA-AG-02-70-C-0388.

R. M. Haralick and I. Dinstein are with the School of Electrical Engineering and the Center for Research, Inc., University of Kansas, Lawrence, Kansas, 66045.

K. Shanmugam was with the Center for Research, Inc., University of Kansas, Lawrence, Kansas. He is now with the Department of Electrical Engineering, Wichita State University, Wichita, Kansas, 67218.

In a search for meaningful features for describing pictorial information, it is only natural to look toward the types of features which human beings use in interpreting pictorial information. Spectral, textural, and contextual features are three fundamental pattern elements used in human interpretation of color photographs. Spectral features describe the average tonal variations in various bands of the visible and/or infrared portion of an electromagnetic spectrum, whereas textural features contain information

TEXTURAS

HARALICK *et al.*: TEXTURAL FEATURES FOR IMAGE CLASSIFICATION

Textural Features

1) Angular Second Moment:

$$f_1 = \sum_i \sum_j \{p(i,j)\}^2.$$

2) Contrast:

$$f_2 = \sum_{n=0}^{N_x-1} n^2 \left(\sum_{\substack{i=1 \\ |i-j|=n}}^{N_x} \sum_{j=1}^{N_y} p(i,j) \right).$$

3) Correlation:

$$f_3 = \frac{\sum_i \sum_j (ij)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

where μ_x , μ_y , σ_x , and σ_y are the means and standard deviations of p_x and p_y .

4) Sum of Squares: Variance

$$f_4 = \sum_i \sum_j (i - \mu)^2 p(i,j).$$

5) Inverse Difference Moment:

$$f_5 = \sum_i \sum_j \frac{1}{1 + (i - j)^2} p(i,j).$$

6) Sum Average:

$$f_6 = \sum_{i=2}^{2N_x} i p_{x+y}(i).$$

7) Sum Variance:

$$f_7 = \sum_{i=2}^{2N_x} (i - f_8)^2 p_{x+y}(i).$$

8) Sum Entropy:²

$$f_8 = - \sum_{i=2}^{2N_x} p_{x+y}(i) \log \{ p_{x+y}(i) \}.$$

9) Entropy:

$$f_9 = - \sum_i \sum_j p(i,j) \log \{ p(i,j) \}.$$

10) Difference Variance:

$$f_{10} = \text{variance of } p_{x-y}.$$

11) Difference Entropy:

$$f_{11} = - \sum_{i=0}^{N_x-1} p_{x-y}(i) \log \{ p_{x-y}(i) \}.$$

12), 13) Information Measures of Correlation:

$$f_{12} = \frac{HXY - HXY1}{\max \{ HX, HY \}}$$

$$f_{13} = (1 - \exp [-2.0(HXY2 - HXY)])^{1/2}$$

$$HXY = - \sum_i \sum_j p(i,j) \log \{ p(i,j) \}$$

where HX and HY are entropies of p_x and p_y , and

$$HXY1 = - \sum_i \sum_j p(i,j) \log \{ p_x(i)p_y(j) \}$$

$$HXY2 = - \sum_i \sum_j p_x(i)p_y(j) \log \{ p_x(i)p_y(j) \}.$$

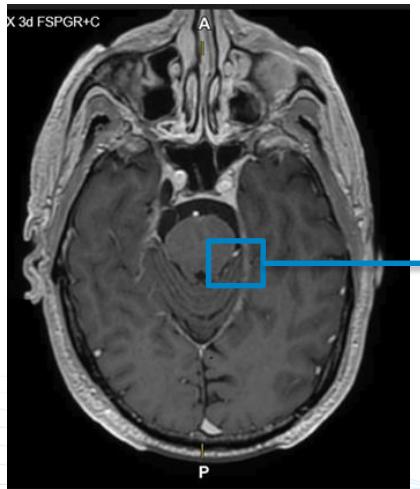
14) Maximal Correlation Coefficient:

$$f_{14} = (\text{Second largest eigenvalue of } Q)^{1/2}$$

where

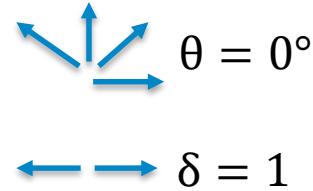
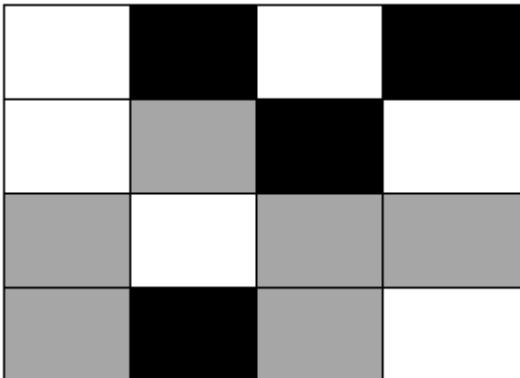
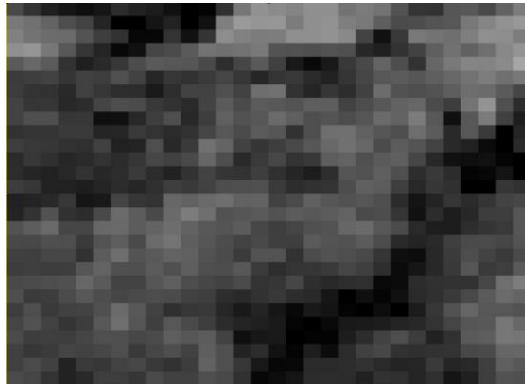
$$Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}.$$

Gray-Level Co-Occurrence Matrix (GLCM)



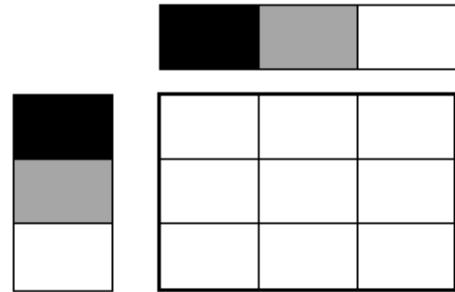
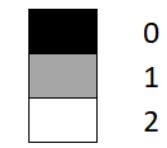
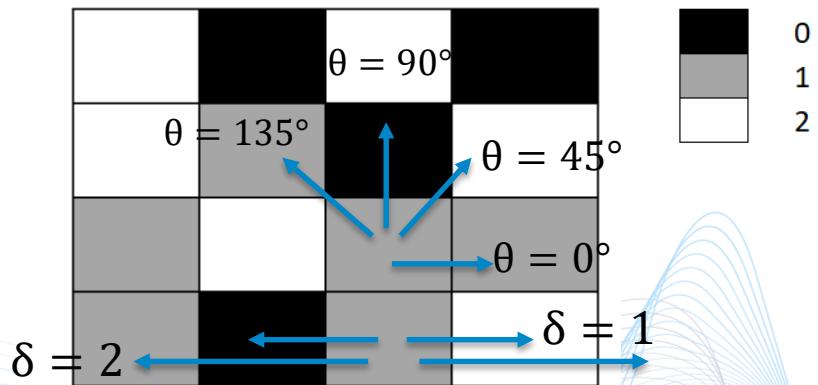
Matriz cuadrada y simétrica, dimensiones $N \times N$ donde N es el número de valores de pixel discretizados

Gray-Level Co-Occurrence Matrix (GLCM)

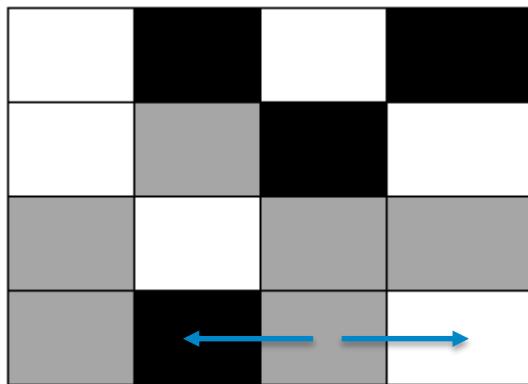


N: N° valores de pixel
δ: Distancia de comparación(pxl)
θ: Ángulo de comparación

Gray-Level Co-Ocurrence Matrix (GLCM)

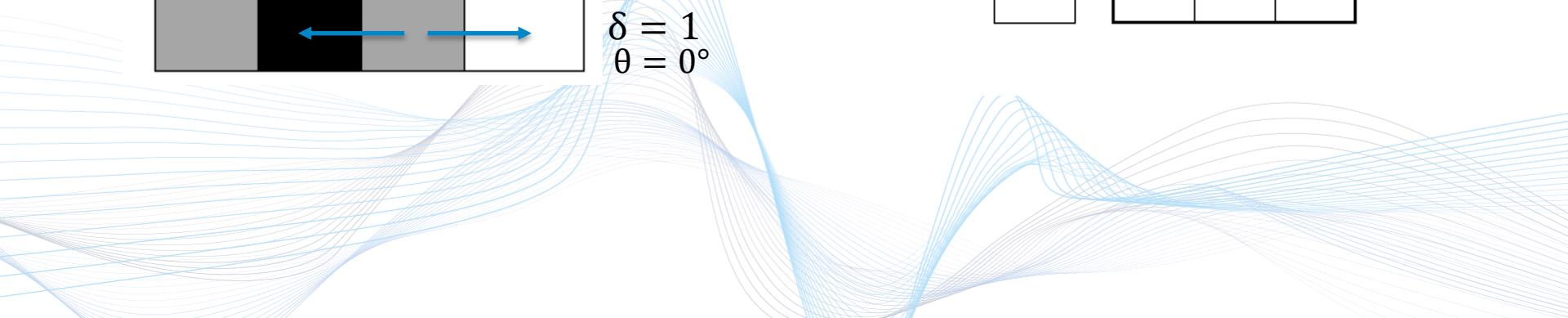
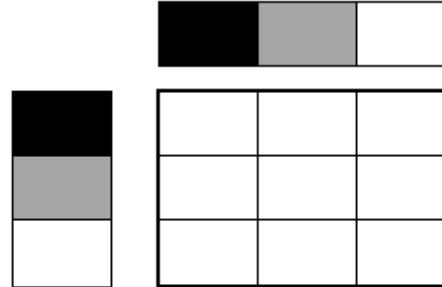


Gray-Level Co-Ocurrence Matrix (GLCM)



0
1
2

$$\delta = 1
θ = 0^\circ$$



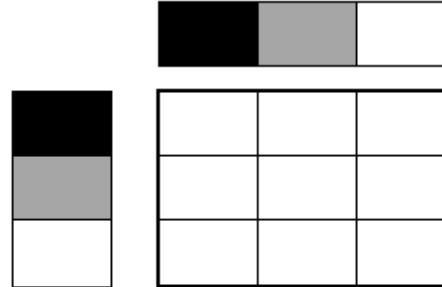
Gray-Level Co-Ocurrence Matrix (GLCM)

2	0	2	0
2	1	0	2
1	2	1	1
1	0	1	2



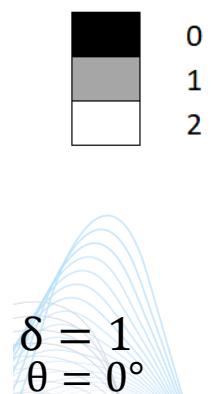
0
1
2

$\delta = 1$
 $\theta = 0^\circ$



Gray-Level Co-Ocurrence Matrix (GLCM)

2	0	2	0
2	1	0	2
1	2	1	1
1	0	1	2



0	3	4
3	2	4
4	4	0

TEXTURAS

HARALICK *et al.*: TEXTURAL FEATURES FOR IMAGE CLASSIFICATION

Textural Features

1) Angular Second Moment:

$$f_1 = \sum_i \sum_j \{p(i,j)\}^2.$$

2) Contrast:

$$f_2 = \sum_{n=0}^{N_x-1} n^2 \left(\sum_{\substack{i=1 \\ |i-j|=n}}^{N_x} \sum_{j=1}^{N_y} p(i,j) \right).$$

3) Correlation:

$$f_3 = \frac{\sum_i \sum_j (ij)p(i,j) - \mu_x \mu_y}{\sigma_x \sigma_y}$$

where μ_x , μ_y , σ_x , and σ_y are the means and standard deviations of p_x and p_y .

4) Sum of Squares: Variance

$$f_4 = \sum_i \sum_j (i - \mu)^2 p(i,j).$$

5) Inverse Difference Moment:

$$f_5 = \sum_i \sum_j \frac{1}{1 + (i - j)^2} p(i,j).$$

6) Sum Average:

$$f_6 = \sum_{i=2}^{2N_x} i p_{x+y}(i).$$

7) Sum Variance:

$$f_7 = \sum_{i=2}^{2N_x} (i - f_8)^2 p_{x+y}(i).$$

8) Sum Entropy:²

$$f_8 = - \sum_{i=2}^{2N_x} p_{x+y}(i) \log \{ p_{x+y}(i) \}.$$

9) Entropy:

$$f_9 = - \sum_i \sum_j p(i,j) \log \{ p(i,j) \}.$$

10) Difference Variance:

$$f_{10} = \text{variance of } p_{x-y}.$$

11) Difference Entropy:

$$f_{11} = - \sum_{i=0}^{N_x-1} p_{x-y}(i) \log \{ p_{x-y}(i) \}.$$

12), 13) Information Measures of Correlation:

$$f_{12} = \frac{HXY - HXY1}{\max \{ HX, HY \}}$$

$$f_{13} = (1 - \exp [-2.0(HXY2 - HXY)])^{1/2}$$

$$HXY = - \sum_i \sum_j p(i,j) \log \{ p(i,j) \}$$

where HX and HY are entropies of p_x and p_y , and

$$HXY1 = - \sum_i \sum_j p(i,j) \log \{ p_x(i)p_y(j) \}$$

$$HXY2 = - \sum_i \sum_j p_x(i)p_y(j) \log \{ p_x(i)p_y(j) \}.$$

14) Maximal Correlation Coefficient:

$$f_{14} = (\text{Second largest eigenvalue of } Q)^{1/2}$$

where

$$Q(i,j) = \sum_k \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}.$$

Otras matrices de texturas

Gray Level Size Zone Matrix (GLSZM)

$P(i,j)$ = Zonas de la imagen con el valor de gris i y el tamaño j

Gray Level Run Length Matrix (GLRLM)

$P(i,j)$ = Zonas de la imagen con el valor de gris i y la longitud j (en el ángulo indicado)

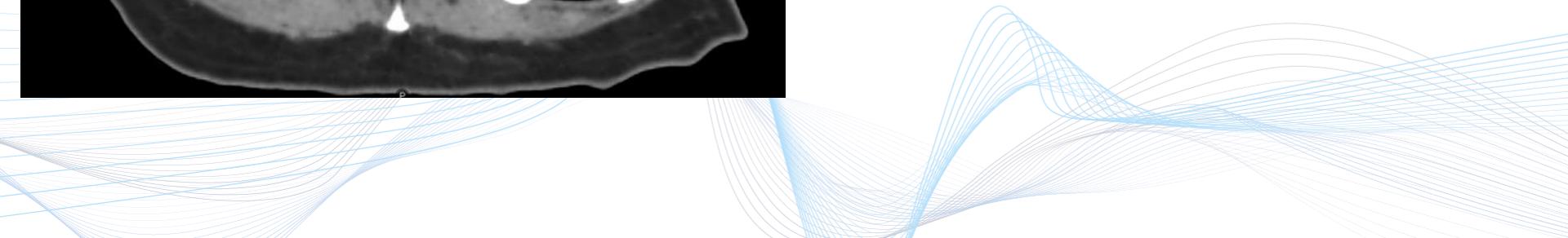
Neighbouring Gray Tone Difference Matrix (NGTDM)

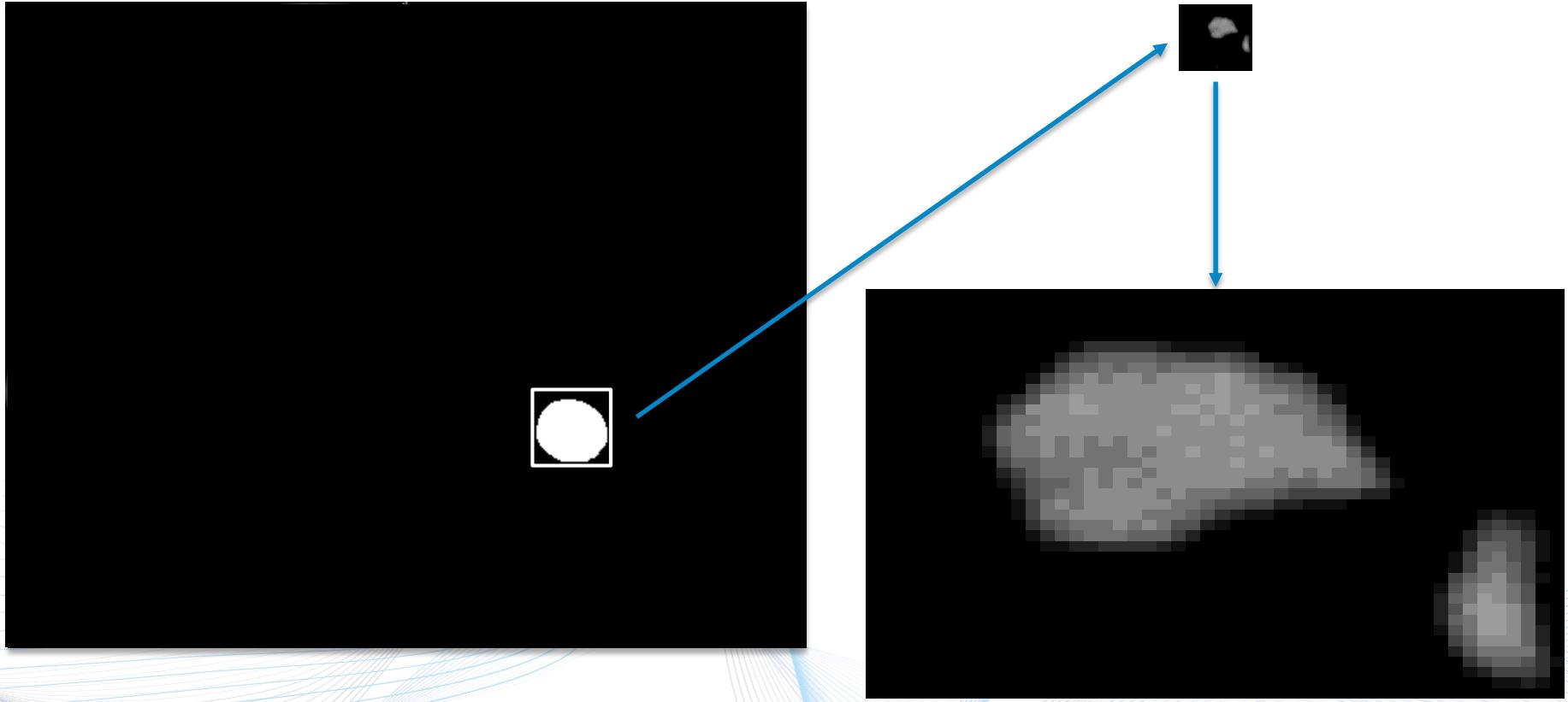
Cuantifica la diferencia de valor de gris entre un valor concreto y el promedio de su entorno

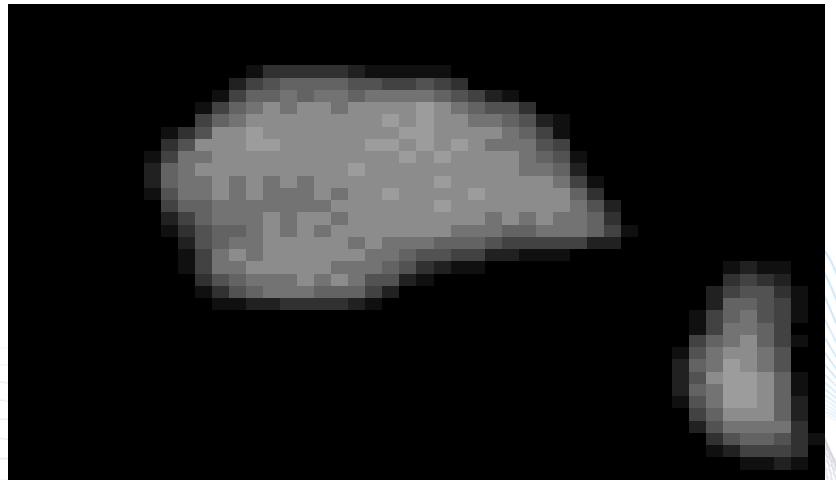
Gray Level Dependence Matrix (GLDM)

Número de voxels con un valor de gris similar ($abs(i-j) < a$) al voxel central

<https://pyradiomics.readthedocs.io/en/latest/features.html>





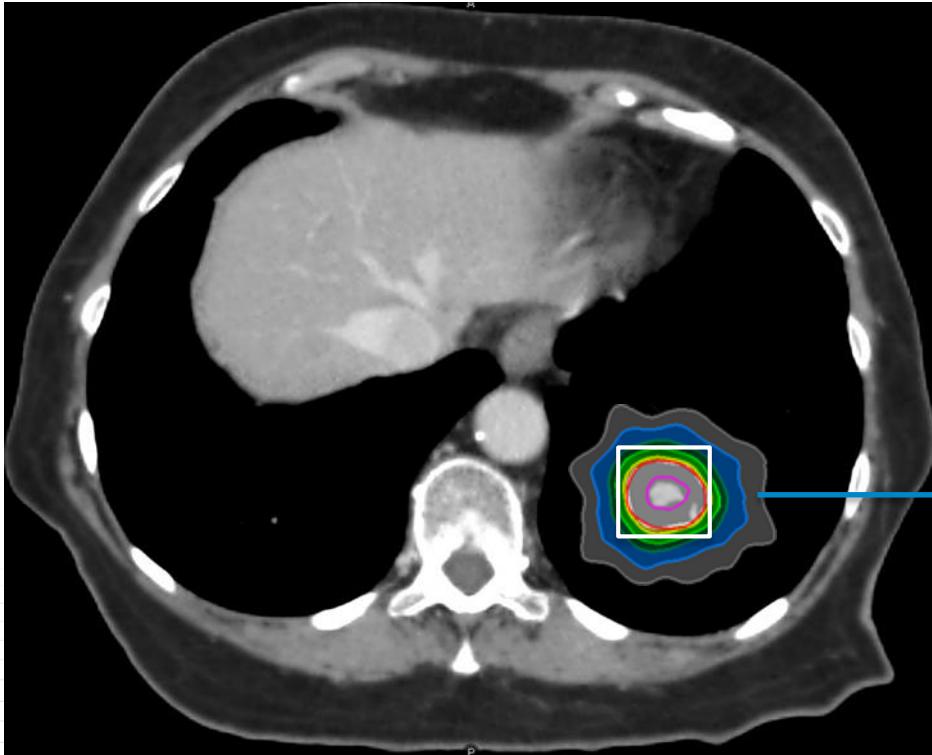


Matriz original discretizada según límites

GLCM

736	5	3	4	5	1	8	4	5	2	6	3
5	2	1	1	0	1	2	1	0	2	2	1
3	1	0	1	2	0	1	2	0	0	1	1
4	1	1	2	1	2	2	0	2	2	1	0
5	0	2	1	0	3	1	3	3	2	2	0
1	1	0	2	3	0	1	1	2	1	2	0
8	2	1	2	1	1	4	2	2	3	2	2
4	1	2	0	3	1	2	2	0	3	2	0
5	0	0	2	3	2	2	0	2	2	1	1
2	2	0	2	2	1	3	3	2	0	2	1
6	2	1	1	2	2	2	2	1	2	0	1
3	1	1	0	0	0	2	0	1	1	1	0

```
[G, I]=graycomatrix(A, 'offset', [0 1], 'Symmetric', true, 'NumLevels', 12, 'GrayLimits', [-6, 5]);
```



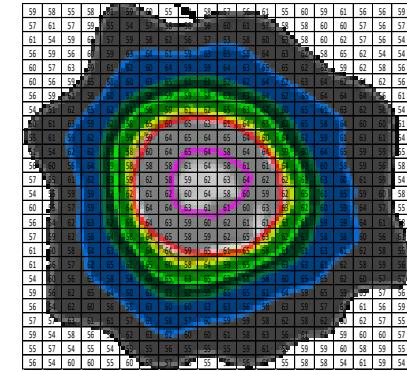
Dosiómica

Tratamiento de la distribución de dosis como si se tratara de una imagen radiológica

59	58	55	58	61	59	60	55	54	58	57	56	61	55	60	59	61	56	56	59
57	61	57	59	55	54	57	58	58	56	60	61	57	58	58	60	60	57	56	57
61	54	59	63	57	59	58	62	56	57	63	58	60	63	58	60	62	57	56	54
56	59	56	60	59	63	64	61	59	63	65	58	64	63	62	58	65	62	54	54
60	57	63	59	61	62	60	64	59	59	64	63	65	65	62	62	59	62	58	56
60	56	59	65	58	60	63	65	61	61	62	62	64	62	63	64	62	61	62	56
56	59	61	58	60	62	58	59	59	58	63	63	64	62	65	64	64	64	56	61
54	61	62	65	62	61	58	59	61	62	65	63	62	60	65	61	63	62	58	54
60	61	63	59	62	58	65	60	65	63	61	64	63	62	65	58	64	60	60	61
55	61	59	62	60	59	59	64	65	64	65	64	60	64	62	59	61	63	61	54
61	54	62	62	59	58	60	64	65	64	58	64	64	59	60	64	65	59	59	55
56	60	56	64	59	58	58	61	64	58	61	61	64	63	60	58	59	56	58	58
57	55	61	62	63	59	62	58	59	62	63	64	62	62	63	63	60	61	59	54
54	55	59	59	62	62	61	62	60	64	58	60	59	62	65	61	65	59	60	58
60	55	57	59	60	64	64	64	63	61	61	60	63	63	62	60	59	64	57	55
56	54	62	63	62	65	64	63	59	60	62	61	61	63	61	59	59	64	61	58
57	59	62	58	64	58	64	65	58	59	62	65	63	62	63	58	60	56	61	58
61	60	58	64	63	63	64	64	59	65	61	65	58	61	61	63	61	62	58	55
61	56	57	62	65	62	65	60	58	64	59	65	63	63	63	61	62	58	59	56
54	60	56	63	58	58	64	65	63	65	63	58	58	60	65	65	62	60	57	57
59	56	62	65	64	60	61	63	62	59	65	65	61	64	59	65	59	65	57	56
56	54	62	60	56	58	63	60	60	63	63	56	58	63	59	57	61	61	56	59
57	57	63	61	61	57	61	58	57	60	59	59	58	62	58	62	60	62	57	55
59	54	58	56	58	62	63	62	60	60	61	58	63	56	61	61	59	60	60	57
55	57	54	55	55	54	59	55	56	55	55	55	59	61	55	59	59	60	58	59
56	54	60	60	55	60	58	57	58	55	55	56	61	55	58	58	58	54	61	59

Dosiómica

- Toxicidad (Neumonitis, hipotiroidismo)
- Recurrencia locorregional
- Control local
- Linfopenia



Dosiomics: 28 entradas



Desafíos de la radiómica

- **Falta homogeneidad**

- Nomenclatura de las distintas variables
- Definición de las distintas variables

- **Falta de reproducibilidad**

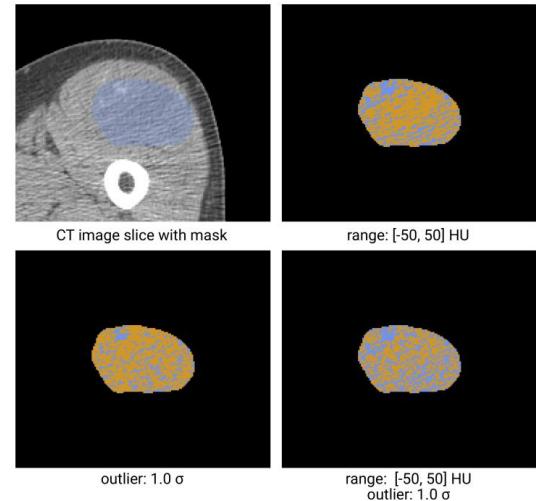
- Diferencias debidas a distintos fabricantes
- Diferencias debidas distinta resolución espacial, anisotropía
- Distintos tamaños imagen
- Distintos planos

Desafíos de la radiómica



Image biomarker standardisation initiative
Reference manual

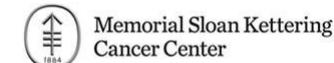
- IBSI 1 (finalizado Marzo 2020)
 - *Valores de referencia para 169 variables radiómicas*
 - *Esquema estándar para el pre-procesado de imágenes*
 - *Guía para el reporte y desarrollo de estudios de radiómica*



Desafíos de la radiómica



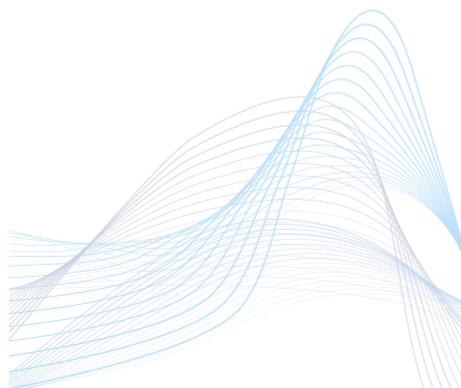
- [pyRadiomics](#)
- [CERR](#)
- [LIFEEx](#)
- [MITK](#)
- [QIFE](#)
- [radiomics-develop](#)
- [CaPTk](#)
- [RaCat](#)
- [SERA](#)
- [QuantImage](#)
- [MIRP](#)



Desafíos de la radiómica



- [pyRadiomics](#)
- [CERR](#)
- [LIFE](#)
- [MITK](#)
- [QIFE](#)
- [radiomics-develop](#)
- [CaPTk](#)
- [RaCat](#)
- [SERA](#)
- [QuantImage](#)
- [MIRP](#)



 MathWorks® [Productos](#) [Soluciones](#) [Educación](#) [Soporte](#) [Comunidad](#) [Eventos](#)

File Exchange

[MATLAB Central](#) | [Archivos](#) | [Autores](#) | [Mi File Exchange](#) | [Publicar](#) | [Acerca de](#)



RADIOMICS
versión 1.2.0.0 (1.38 MB) por [Martin Vallières](#)
MATLAB programming tools for radiomics analysis
<https://github.com/mvallieres/radiomics>

[Visión general](#) [Funciones](#) [Reseñas \(7\)](#) [Conversaciones \(17\)](#)

|<<https://github.com/mvallieres/radiomics>>|

-> A package providing MATLAB programming tools for radiomics analysis.

REFERENCE:

[1] Vallières, M. et al. (2015). A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Physics in Medicine and Biology*, 60(14), 5471-5496. doi:10.1088/0031-9155/60/14/5471

AUTHOR: Martin Vallières <mart.vallieres@gmail.com>

HISTORY:

- Version 1.0: May 2015

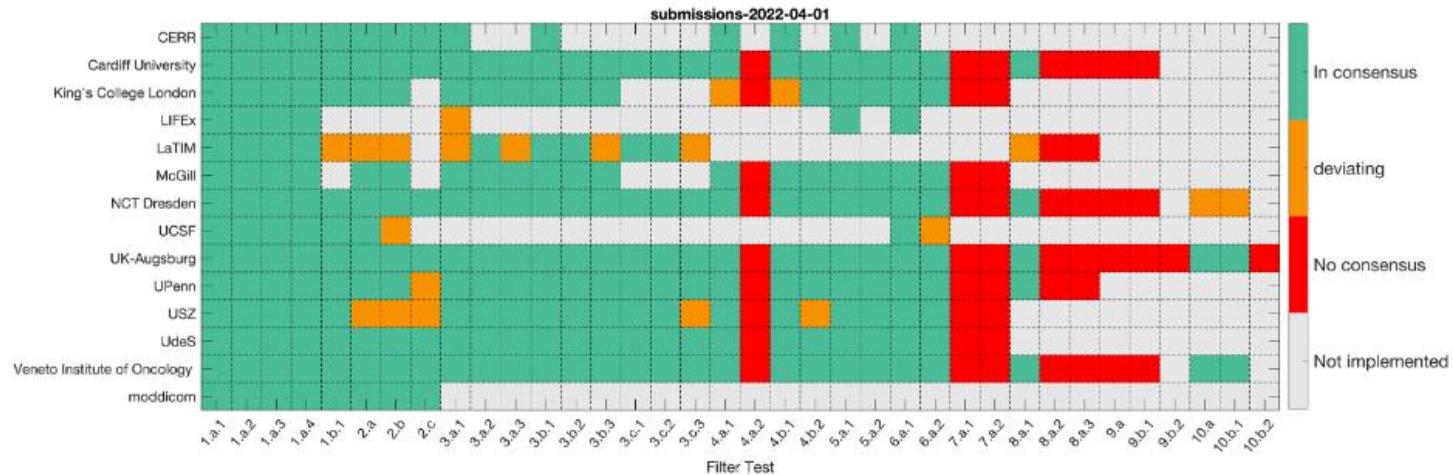
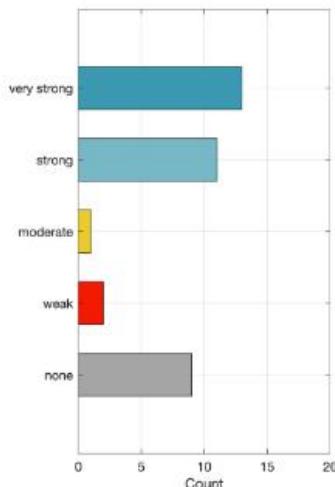
DISCLAIMER:

"I'm not a programmer, I'm just a scientist doing stuff!"

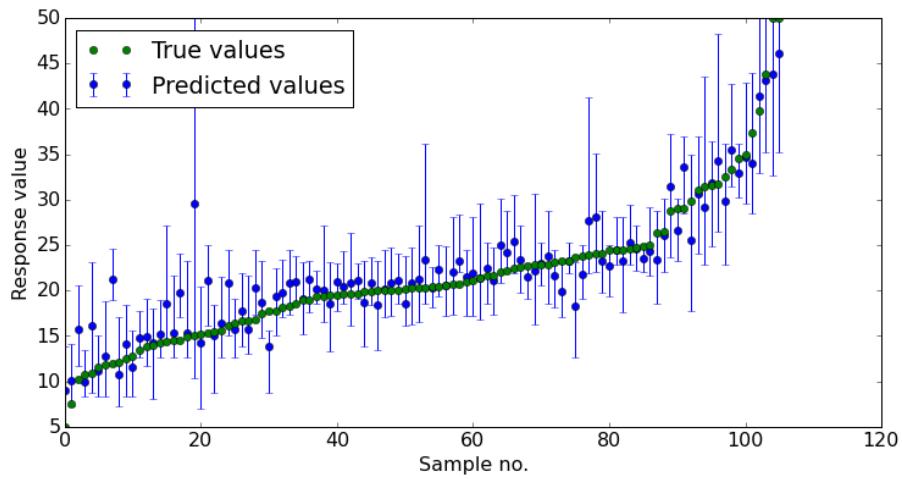
Desafíos de la radiómica



- IBSI 2 (en proceso)
 - *Estandarización de distintos filtros convolucionales para imágenes (wavelet, Laplacianos, Gaussianos...)*



MODELOS PREDICTIVOS

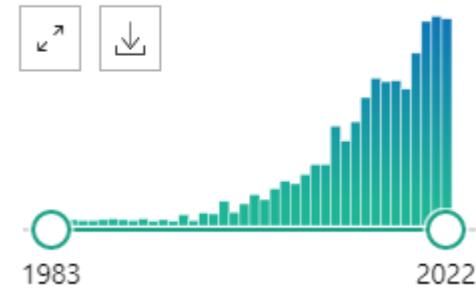


MODELOS CLÍNICOS PREDICTIVOS

CDSS: *Clinical Decision Support Systems*

- Árboles de decisiones
- Diagnóstico basado en experiencia previa
- Guías para el correcto manejo del paciente

RESULTS BY YEAR



MODELOS CLÍNICOS PREDICTIVOS

INFORMACIÓN DEMOGRÁFICA

INFORMACIÓN CLÍNICA

Imagen diagnóstica
Historial familiar
Análisis, biopsias..

Tratamiento

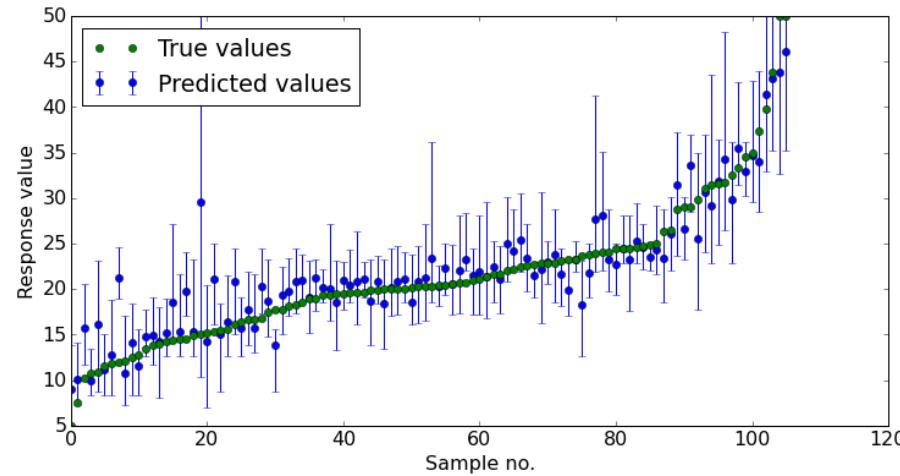
RESULTADO

....



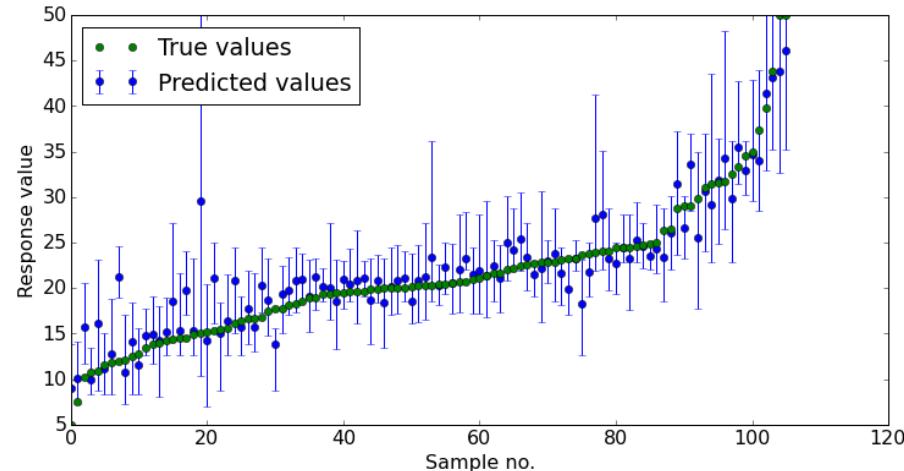
MODELOS CLÍNICOS PREDICTIVOS BASADOS EN IA

- Recolección de datos de forma retrospectiva
- Modelo predictivo de un determinado resultado (supervivencia, recidiva local..)
- En base a variables conocidas y variables nuevas extraídas de estudios



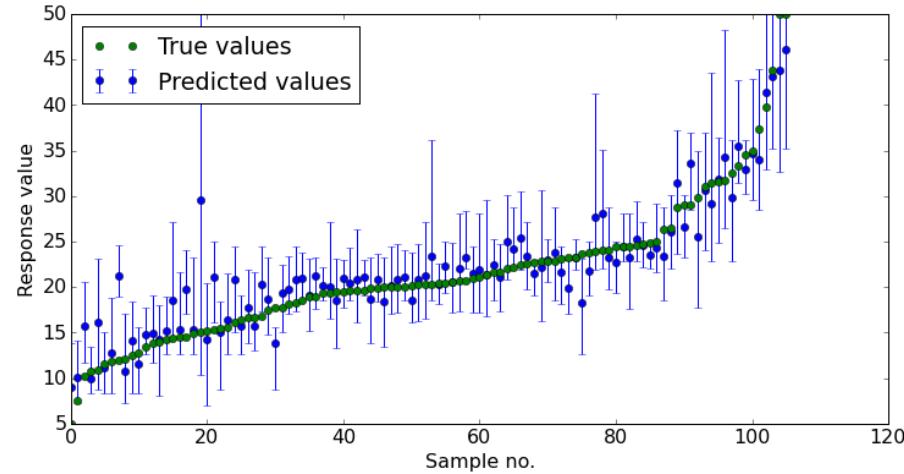
MODELOS CLÍNICOS PREDICTIVOS BASADOS EN IA

- Resultado más probable en base a la experiencia
- Detección de patrones subyacentes: Generación de nuevo conocimiento
- Aumento de la información disponible para la toma de decisiones clínicas
- Personalización del tratamiento



MODELOS CLÍNICOS PREDICTIVOS BASADOS EN IA

- ¿Es este el mejor tratamiento?
- ¿Existen parámetros dentro de este tratamiento ajustables para optimizarlo?
- ¿Qué variables parecen tener más peso en el resultado?



MODELOS CLÍNICOS PREDICTIVOS

INFORMACIÓN DEMOGRÁFICA

INFORMACIÓN CLÍNICA

RADIÓMICA

DOSIÓMICA

GENÓMICA

.....

Tratamiento

RESULTADO CLÍNICO

Respuesta al tratamiento

Recidiva local

Supervivencia

Toxicidad

...

MODELOS CLÍNICOS PREDICTIVOS

INFORMACIÓN DEMOGRÁFICA

INFORMACIÓN CLÍNICA

RADIÓMICA

DOSIÓMICA

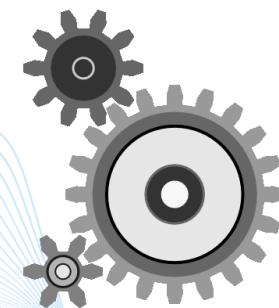
GENÓMICA

.....

RESULTADO CLÍNICO

Respuesta al tratamiento
Recidiva local
Supervivencia
Toxicidad

...



MODELOS PREDICTIVOS

Recidiva local en pacientes con diagnóstico de CA Orofaringe

VARIABLES DEMOGRÁFICAS: Edad, Sexo, fumador, bebedor, sobrepeso...

VARIABLES CLÍNICAS: Cirugía, resección total o parcial, quimioterapia, radioterapia, HPV positivo...

VARIABLES RADIÓMICAS: Calculadas sobre TC, sobre RM, sobre PET...

VARIABLES DOSIMÉTRICAS: Fraccionamiento, DBE, $V_{95\%}$, D_{MAX} , D_{min} ...

VARIABLES DOSIÓMICAS: GLCM, GLRLM...

Paciente 1

69 | 17 | 38 | 15 | 32 | 46 | 9 | 60 | 66 | 21 | 7 | 58 | 25 | 97 | 69 | 17 | 38 | 15 | 32 | 46 |

97 | 69 | 17 | 38 | 15 | 32 | 46 | 9 | 60 | 66 | 21 | 7 | 58 | 25 | 97 |

Paciente 2

69 | 17 | 38 | 15 | 32 | 46 | 9 | 60 | 66 | 21 | 7 | 58 | 25 | 97 | 69 | 17 | 38 | 15 | 32 | 46 |

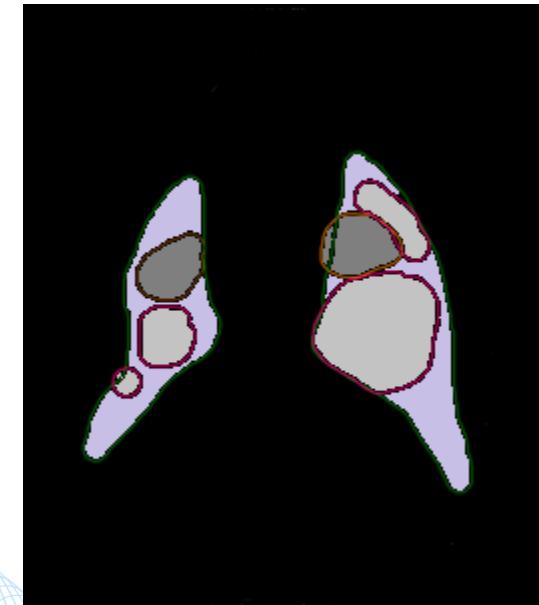
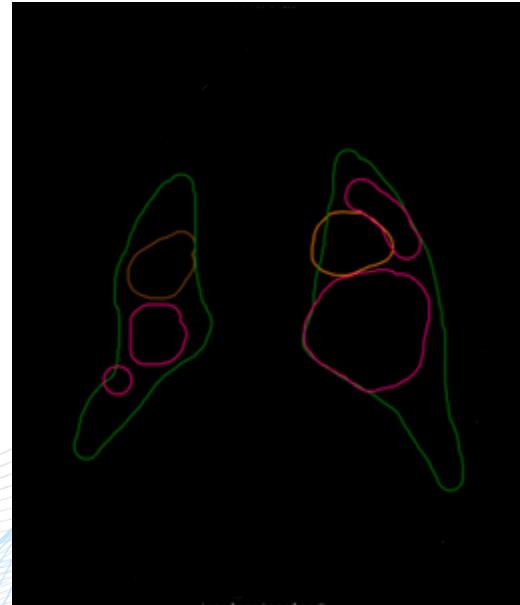
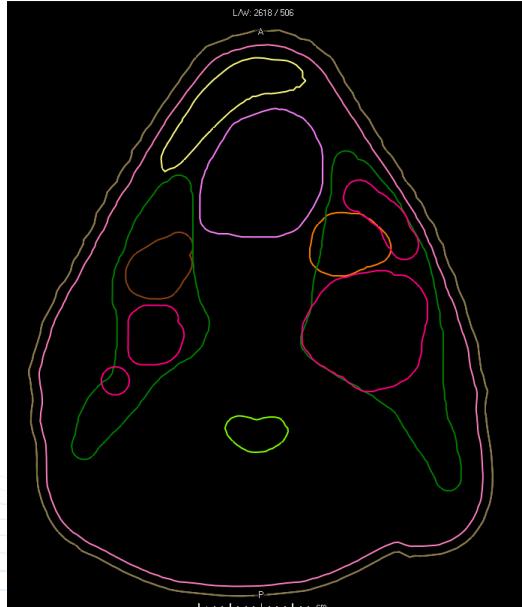
97 | 69 | 17 | 38 | 15 | 32 | 46 | 9 | 60 | 66 | 21 | 7 | 58 | 25 | 97 |

Paciente 3

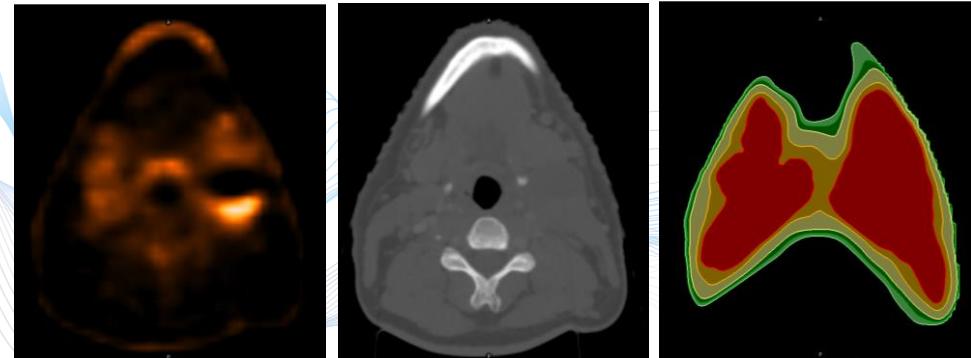
69 | 17 | 38 | 15 | 32 | 46 | 9 | 60 | 66 | 21 | 7 | 58 | 25 | 97 | 69 | 17 | 38 | 15 | 32 | 46 |

97 | 69 | 17 | 38 | 15 | 32 | 46 | 9 | 60 | 66 | 21 | 7 | 58 | 25 | 97 |

Recidiva SI
Recidiva NO
Recidiva NO



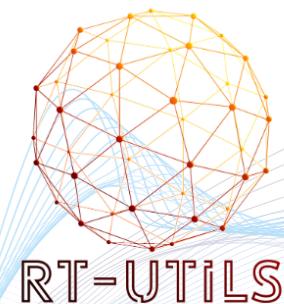
- Diferente tamaños de pixel / vóxel y vóxel no cúbico
- Diferentes orientaciones
- Diferentes tamaños de imagen
- Diferentes fabricantes / equipos



```

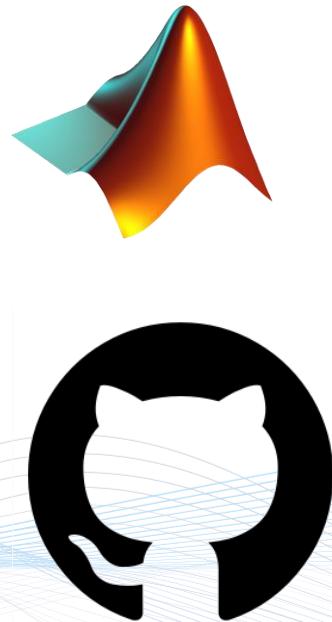
FileMetaInformationGroupLength          200
FileMetaInformationVersion            [0;1]
MediaStorageSOPClassUID              '1.2.840.10008.5.1.4.1.1.481.3'
MediaStorageSOPInstanceUID           '1.2.752.243.1.1.20190521175918738.1300.37082'
TransferSyntaxUID                   '1.2.840.10008.1.2'
ImplementationClassUID              '1.2.826.0.1.3680043.1.2.100.6.40.0.76'
ImplementationVersionName           'DicomObjects.NET'
SpecificCharacterSet                'ISO_IR 100'
InstanceCreationDate                '20190521'
InstanceCreationTime                '175918'
SOPClassUID                         '1.2.840.10008.5.1.4.1.1.481.3'
SOPInstanceUID                      '1.2.752.243.1.1.20190521175918738.1300.37082'
StudyDate                           '20190503'
SeriesDate                          '20190521'
StudyTime                           '111717'
SeriesTime                          '175918'
AccessionNumber                     '1251'
Modality                            'RTSTRUCT'
Manufacturer                         'RaySearch Laboratories'
ReferringPhysicianName              '1x1 struct'
SeriesDescription                   'RS: Approved Structure Set'
OperatorsName                       '1x1 struct'
ManufacturerModelName              'RayStation'
PatientName                         '1x1 struct'
PatientID                           '1'
PatientBirthDate                   '1.2.392.200036.9116.2.6.1.44063.1797790409.1556849836.84...'
PatientSex                           '1'
SoftwareVersions                    '8.0.1.10 (Dicom Export)'
StudyInstanceUID                    '1.2.392.200036.9116.2.6.1.44063.1797790409.1556849836.84...'
SeriesInstanceUID                   '1.2.752.243.1.1.20190521175918738.1300.37082.1'
StudyID                            '1251'
SeriesNumber                        '1'
InstanceNumber                      '1'
FrameOfReferenceUID                '1.2.392.200036.9116.2.6.1.44063.1797790409.1556849838.75...'
PositionReferenceIndicator         ''
StructureSetLabel                  'RS: Approved'
StructureSetDate                   '20190521'
StructureSetTime                   '175918'
ReferencedFrameOfReferenceSequence '1x1 struct'
StructureSetROISequence            '1x1 struct'
ROIContourSequence                 '1x1 struct'
RTROIObservationsSequence         '1x1 struct'
ApprovalStatus                     'APPROVED'
ReviewDate                          '20190521'
ReviewTime                          '183928'
ReviewerName                        '1x1 struct'

```



RT-UTILS

A minimal Python library for RT Struct manipulation



MODELOS PREDICTIVOS

Recidiva local en pacientes con diagnóstico de CA Orofaringe

VARIABLES DEMOGRÁFICAS: Edad, Sexo, fumador, bebedor, sobrepeso...

VARIABLES CLÍNICAS: Cirugía, resección total o parcial, quimioterapia, radioterapia, HPV positivo...

VARIABLES RADIÓMICAS: Calculadas sobre TC, sobre RM, sobre PET...

VARIABLES DOSIMÉTRICAS: Fraccionamiento, DBE, $V_{95\%}$, D_{MAX} , D_{min} ...

VARIABLES DOSIÓMICAS: GLCM, GLRLM...

Paciente 1

69 | 17 | 38 | 15 | 32 | 46 | 9 | 60 | 66 | 21 | 7 | 58 | 25 | 97 | 69 | 17 | 38 | 15 | 32 | 46 |

97 | 69 | 17 | 38 | 15 | 32 | 46 | 9 | 60 | 66 | 21 | 7 | 58 | 25 | 97 |

Paciente 2

69 | 17 | 38 | 15 | 32 | 46 | 9 | 60 | 66 | 21 | 7 | 58 | 25 | 97 | 69 | 17 | 38 | 15 | 32 | 46 |

97 | 69 | 17 | 38 | 15 | 32 | 46 | 9 | 60 | 66 | 21 | 7 | 58 | 25 | 97 |

Paciente 3

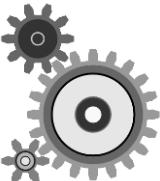
69 | 17 | 38 | 15 | 32 | 46 | 9 | 60 | 66 | 21 | 7 | 58 | 25 | 97 | 69 | 17 | 38 | 15 | 32 | 46 |

97 | 69 | 17 | 38 | 15 | 32 | 46 | 9 | 60 | 66 | 21 | 7 | 58 | 25 | 97 |

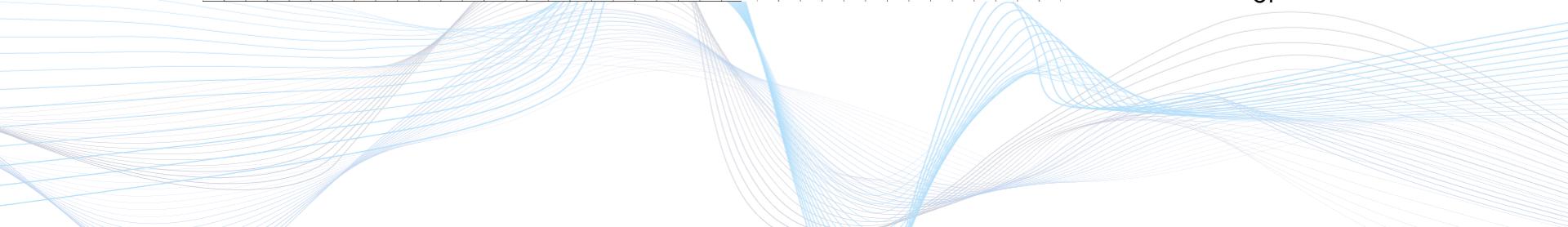
Recidiva SI
Recidiva NO
Recidiva NO

MODELOS PREDICTIVOS

1	54	42	48	36	7	78	42	21	44	35	15	28	7	80	54	42	48	36	7	78	42	21	44	35	15	28	7	80	
2	97	33	60	38	96	15	2	90	13	7	93	45	87	85	97	33	60	38	96	15	2	90	13	7	93	45	87	85	
3	81	48	67	66	88	22	79	99	87	83	73	40	66	96	81	48	67	66	88	22	79	99	87	83	73	40	66	96	
4	31	49	58	85	80	31	51	99	36	5	57	81	57	75	31	49	58	85	80	31	51	99	36	5	57	81	57	75	
5	21	55	65	17	59	15	20	19	88	74	0	27	26	35	21	55	65	17	59	15	20	19	88	74	0	27	26	35	
6	55	75	37	13	46	70	42	35	13	98	35	78	92	27	55	75	37	13	46	70	42	35	13	98	35	78	92	27	
7	52	60	81	38	56	56	79	89	6	43	71	67	24	66	52	60	81	38	56	56	79	89	6	43	71	67	24	66	
8	33	22	71	12	56	15	0	79	46	17	87	17	15	88	33	22	71	12	56	15	0	79	46	17	87	17	15	88	33
9	11	31	33	78	54	78	70	43	55	24	84	49	89	76	11	31	33	78	54	78	70	43	55	24	84	49	89	76	
10	52	66	93	53	9	33	23	51	23	90	27	98	74	82	52	66	93	53	9	33	23	51	23	90	27	98	74	82	
11	17	7	24	25	96	31	3	67	78	61	96	86	99	12	17	7	24	25	96	31	3	67	78	61	96	86	99	12	
12	86	55	81	70	7	61	48	39	13	64	38	37	40	93	86	55	81	70	7	61	48	39	13	64	38	37	40	93	
13	84	24	70	29	21	34	41	82	9	43	77	74	58	91	84	24	70	29	21	34	41	82	9	43	77	74	58	91	
14	69	17	38	15	32	46	9	60	66	21	7	58	25	97	69	17	38	15	32	46	9	60	66	21	7	58	25	97	
15	54	42	48	36	7	78	42	21	44	35	15	28	7	80	54	42	48	36	7	78	42	21	44	35	15	28	7	80	
16	97	33	60	38	96	15	2	90	13	7	93	45	87	85	97	33	60	38	96	15	2	90	13	7	93	45	87	85	
17	81	48	67	66	88	22	79	99	87	83	73	40	66	96	81	48	67	66	88	22	79	99	87	83	73	40	66	96	
18	31	49	58	85	80	31	51	99	36	5	57	81	57	75	31	49	58	85	80	31	51	99	36	5	57	81	57	75	
19	21	55	65	17	59	15	20	19	88	74	0	27	26	35	21	55	65	17	59	15	20	19	88	74	0	27	26	35	
20	55	75	37	13	46	70	42	35	13	98	35	78	92	27	55	75	37	13	46	70	42	35	13	98	35	78	92	27	
21	52	60	81	38	56	56	79	89	6	43	71	67	24	66	52	60	81	38	56	56	79	89	6	43	71	67	24	66	
22	33	22	71	12	56	15	0	79	46	17	87	17	15	88	33	22	71	12	56	15	0	79	46	17	87	17	15	88	33
23	11	31	33	78	54	78	70	43	55	24	84	49	89	76	11	31	33	78	54	78	70	43	55	24	84	49	89	76	
24	52	66	93	53	9	33	23	51	23	90	27	98	74	82	52	66	93	53	9	33	23	51	23	90	27	98	74	82	
25	17	7	24	25	96	31	3	67	78	61	96	86	99	12	17	7	24	25	96	31	3	67	78	61	96	86	99	12	
26	86	55	81	70	7	61	48	39	13	64	38	37	40	93	86	55	81	70	7	61	48	39	13	64	38	37	40	93	
27	84	24	70	29	21	34	41	82	9	43	77	74	58	91	84	24	70	29	21	34	41	82	9	43	77	74	58	91	
28	69	17	38	15	32	46	9	60	66	21	7	58	25	97	69	17	38	15	32	46	9	60	66	21	7	58	25	97	



Si
No
No
Si
No
No
No
No
Si



PRE-Procesado de datos

1. Variables categóricas convertirlas en numéricas
2. Normalización de variables continuas
3. Estrategias en el caso de características incompletas
4. Balance de clases
5. Limitar el número de características (feature selection)

PRE-Procesado de datos

1. Variables categóricas convertirlas en numéricas (en binarias si es posible)

Fumador: (Si / No) = (1 / 0)

Quimioterapia: (No / Abraxane / Nalirinoxx / Folfirinox) = (0 / 1 / 2 / 3)



PRE-Procesado de datos

2. Normalización de variables continuas

Evitar características con rangos muy amplios

Evitar características con órdenes de magnitud muy distintos entre si

1. Escalamiento a un rango
2. Escalamiento logarítmico
3. Normalización Z-Score

PRE-Procesado de datos: Normalización de variables continuas

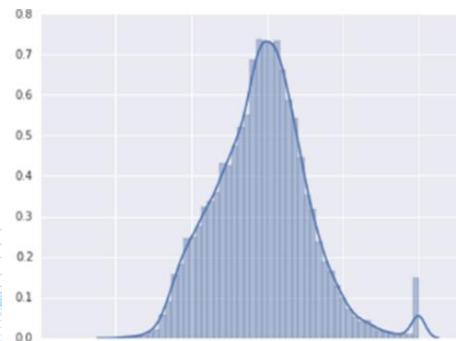
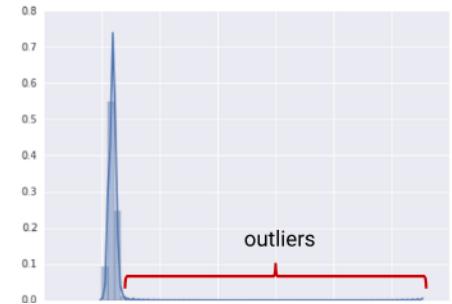
2.1 Escalamiento a un rango

Normalización a un rango $[0, 1]$ (También posible $[-1, 1]$)

Recomendado si los datos se mueven en un rango concreto conocido y de forma aproximadamente uniforme

Muy sensible a la existencia de outliers: Valorar un recorte de función

$$x' = \frac{(x - x_{min})}{(x_{max} - x_{min})}$$

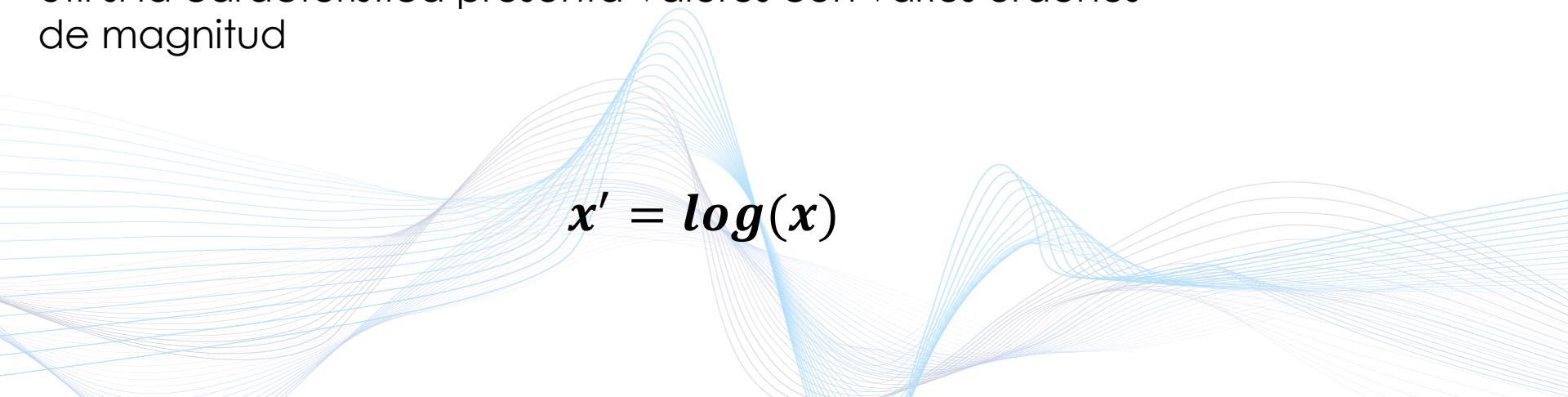


PRE-Procesado de datos: Normalización de variables continuas

2.2 Escalamiento logarítmico

Comprime una característica con un rango muy amplio en un rango más manejable.

Útil si la característica presenta valores con varios órdenes de magnitud

$$x' = \log(x)$$


PRE-Procesado de datos: Normalización de variables continuas

2.3 Normalización Z-Score

La característica pasa a tener $\mu' = 0$ y desviación estándar $\sigma' = 1$

Poco sensible a los outliers

Útil si existen algunos outliers pero no tantos como para hacer recorte

$$x' = \frac{(x - \mu)}{\sigma}$$

PRE-Procesado de datos

3. Estrategias en el caso de observaciones incompletas:

1. Eliminar las observaciones incompletas
2. Establecer una categoría 'desconocida' (solo categóricas)
3. Establecer un valor según distribución de la variable
(media, mediana, moda...)
4. Intentar predecir los valores que faltan mediante otro modelo

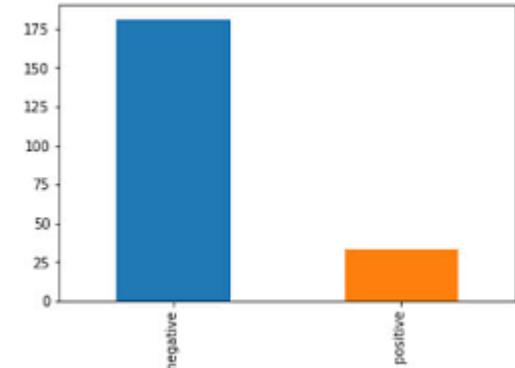
FUMADOR
Si
No
No
Si
No
No
?
?
No
Si
Si

PRE-Procesado de datos

4. Balance de clases

Existe una diferencia significativa entre las clases del modelo

La clase menos abundante puede no tener peso estadístico suficiente para que el modelo aprenda

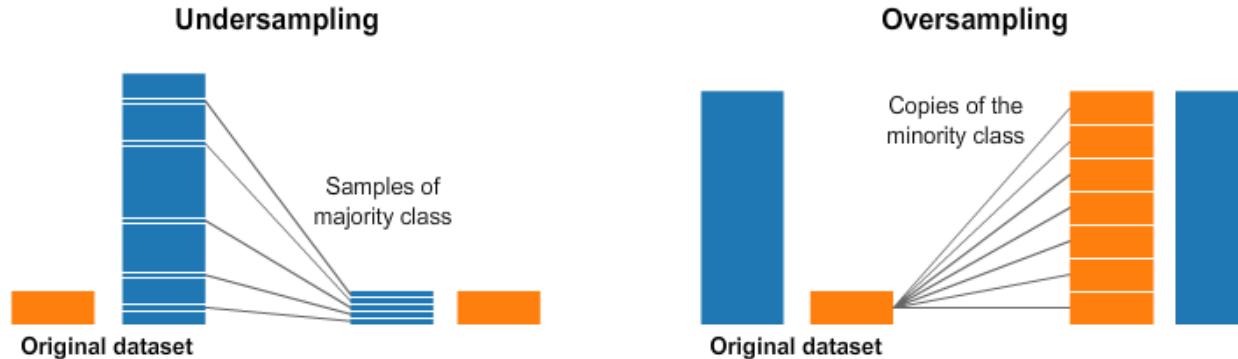


PRE-Procesado de datos

4. Balance de clases

Disminuir las observaciones de la clase más abundante (*undersampling*)

Aumentar las observaciones de la clase menos abundante (*oversampling*)

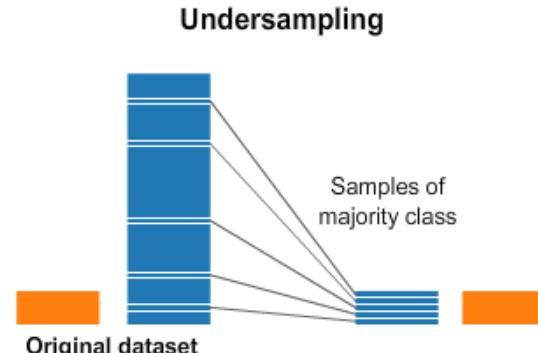
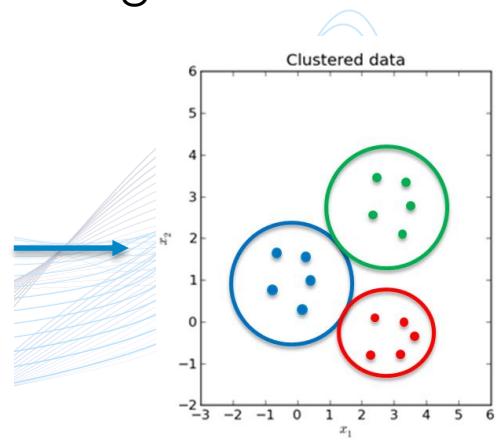
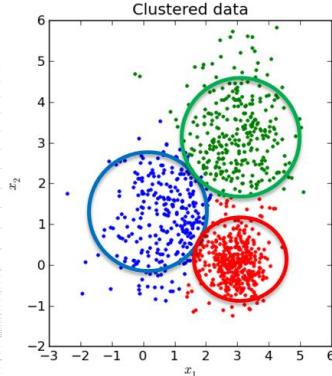


PRE-Procesado de datos

4. Balance de clases

Disminuir las observaciones de la clase más abundante (undersampling)

- Eliminar datos de forma aleatoria
- Técnicas de clustering



PRE-Procesado de datos

4. Balance de clases

Aumentar las observaciones de la clase menos abundante (oversampling)

- Data Warping: pequeñas variaciones sobre el dato original
- Datos sintéticos



PRE-Procesado de datos

5. Limitar el número de características (**feature selection**)

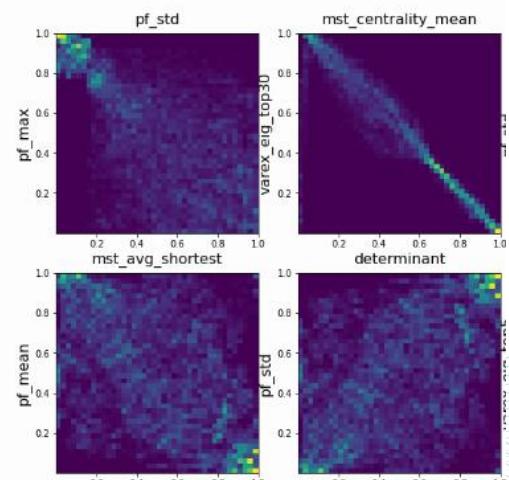
- Evitar un número muy elevado (*Curse of dimensionality*) (10:1)
- Eliminar aquellas con muy baja correlación con el GT
- Eliminar aquellas con varianza muy pequeña o nula
- Seleccionar entre aquellas con alta correlación entre si
- Otros métodos: *wrapping* y *embedding*

PRE-Procesado de datos

Limitar el número de características (feature selection)

Métodos de correlación

	Categórica	Continua
Categórica	Lambda, Cramer's V	Regresión logística, Point-Biserial (si dicotómica)
Continua	Regresión logística, Point-Biserial (si dicotómica)	Spearman, Kendall, Pearson



PRE-Procesado de datos

5. Limitar el número de características (feature selection)

54	42	48	36	7	78	42	21	44	35	15	28	7	80	54	42	48	36	7	78	42	21	44	35	15	28	7	80	
97	33	60	38	96	15	2	90	13	7	93	45	87	85	97	33	60	38	96	15	2	90	13	7	93	45	87	85	
81	48	67	66	88	22	79	99	87	83	73	40	66	96	81	48	67	66	88	22	79	99	87	83	73	40	66	96	
31	49	58	85	80	31	51	99	36	5	57	81	57	75	31	49	58	85	80	31	51	99	36	5	57	81	57	75	
21	55	65	17	59	15	20	19	88	74	0	27	35	21	55	65	17	59	15	20	19	88	74	0	27	26	35		
55	75	37	13	46	70	42	35	13	93	98	35	78	92	27	55	75	37	13	46	70	42	35	13	98	95	35	78	92
52	60	81	38	56	56	79	89	6	43	61	77	24	66	52	60	81	38	56	56	79	89	6	43	61	77	24	66	
33	22	71	12	56	15	0	79	46	17	87	17	15	88	33	22	71	12	56	15	0	79	46	17	87	17	15	88	
11	31	33	78	54	78	70	43	55	24	84	49	89	76	11	31	33	78	54	78	70	43	55	24	84	49	89	76	
52	66	93	53	9	33	23	51	23	90	27	98	74	82	52	66	93	53	9	33	23	51	23	90	27	98	74	82	
17	7	24	25	96	31	3	67	78	61	96	89	12	17	7	24	25	96	31	3	67	78	61	96	89	12	17	7	
86	55	81	70	7	61	48	39	13	64	36	37	40	93	86	55	81	70	7	61	48	39	13	64	36	37	40	93	
84	24	70	29	21	34	41	82	9	43	77	74	58	91	84	24	70	29	21	34	41	82	9	43	77	74	58	91	
69	17	38	15	32	46	9	60	66	21	7	58	25	97	69	17	38	15	32	46	9	60	66	21	7	58	25	97	
54	42	48	36	7	78	42	21	44	35	15	28	7	80	54	42	48	36	7	78	42	21	44	35	15	28	7	80	
97	33	60	38	96	15	2	90	13	7	93	45	87	85	97	33	60	38	96	15	2	90	13	7	93	45	87	85	
81	48	67	66	88	22	79	99	87	83	73	40	66	96	81	48	67	66	88	22	79	99	87	83	73	40	66	96	
31	49	58	85	80	31	51	99	36	5	57	81	57	75	31	49	58	85	80	31	51	99	36	5	57	81	57	75	
21	55	65	17	59	15	20	19	88	74	0	27	35	21	55	65	17	59	15	20	19	88	74	0	27	26	35		
55	75	37	13	46	70	42	35	13	93	98	35	78	92	27	55	75	37	13	46	70	42	35	13	98	95	35	78	92
52	60	81	38	56	56	79	89	6	43	61	77	24	66	52	60	81	38	56	56	79	89	6	43	61	77	24	66	
33	22	71	12	56	15	0	79	46	17	87	17	15	88	33	22	71	12	56	15	0	79	46	17	87	17	15	88	
11	31	33	78	54	78	70	43	55	24	84	49	89	76	11	31	33	78	54	78	70	43	55	24	84	49	89	76	
52	66	93	53	9	33	23	51	23	90	27	98	74	82	52	66	93	53	9	33	23	51	23	90	27	98	74	82	
17	7	24	25	96	31	3	67	78	61	96	89	12	17	7	24	25	96	31	3	67	78	61	96	89	12	17	7	
86	55	81	70	7	61	48	39	13	64	36	37	40	93	86	55	81	70	7	61	48	39	13	64	36	37	40	93	
84	24	70	29	21	34	41	82	9	43	77	74	58	91	84	24	70	29	21	34	41	82	9	43	77	74	58	91	
69	17	38	15	32	46	9	60	66	21	7	58	25	97	69	17	38	15	32	46	9	60	66	21	7	58	25	97	

Si
No
No
Si
No
No
.
.
.
Si

PRE-Procesado de datos

5. Limitar el número de características (*feature selection*)

1	54	42	48	36	7	78	42	21	44	35	15
2	97	33	60	38	96	15	2	90	13	7	93
3	81	48	67	66	88	22	79	99	87	83	73
4	31	49	58	85	80	31	51	99	36	5	57
5	21	55	65	17	59	15	20	19	88	74	0
6	55	75	37	13	46	70	42	35	13	98	35
7	52	60	81	38	56	56	79	89	6	43	71
8	33	22	71	12	56	15	0	79	46	17	87
9	11	31	33	78	54	78	70	43	55	24	84
10	52	66	93	53	9	33	23	51	23	90	27
11	17	7	24	25	96	31	3	67	78	61	96
12	86	55	81	70	7	61	48	39	13	64	38
13	84	24	70	29	21	34	41	82	9	43	77
14	69	17	38	15	32	46	9	60	66	21	7
15	54	42	48	36	7	78	42	21	44	35	15
16	97	33	60	38	96	15	2	90	13	7	93
17	81	48	67	66	88	22	79	99	87	83	73
18	31	49	58	85	80	31	51	99	36	5	57
19	21	55	65	17	59	15	20	19	88	74	0
20	55	75	37	13	46	70	42	35	13	98	35
21	52	60	81	38	56	56	79	89	6	43	71
22	33	22	71	12	56	15	0	79	46	17	87
23	11	31	33	78	54	78	70	43	55	24	84
24	52	66	93	53	9	33	23	51	23	90	27
25	17	7	24	25	96	31	3	67	78	61	96
26	86	55	81	70	7	61	48	39	13	64	38
27	84	24	70	29	21	34	41	82	9	43	77
28	69	17	38	15	32	46	9	60	66	21	7

Si
No
No
Si
No
No
•
•
•
N Si



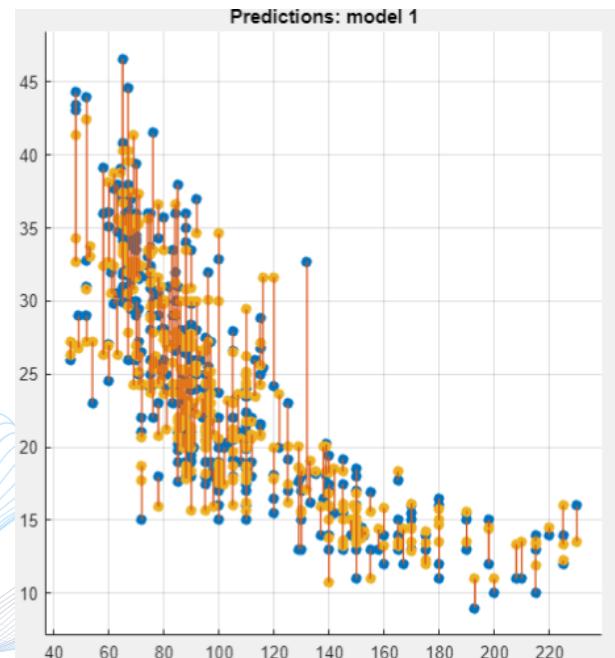
VALIDACIÓN DE UN MODELO PREDICTIVO

Modelos de regresión: Predicen un valor continuo, por ejemplo supervivencia, o tiempo libre de enfermedad.

Métricas basadas en el cálculo de errores:

$$RMSE = \sqrt{\frac{\sum (y - y_t)^2}{T}}$$

$$ECM = \frac{1}{T} \sum (y - y_t)^2$$



VALIDACIÓN DE UN MODELO PREDICTIVO

Modelos de clasificación: Predicen una clase, por ejemplo, larga, corta o media supervivencia.

Sensibilidad, especificidad, recall, precisión...

VALORES PREDICCIÓN	
VALORES REALES	
	Verdaderos positivos Falsos Positivos
Verdaderos Negativos	Falsos Negativos Verdaderos Negativos

sensitivity, recall, hit rate, or true positive rate (TPR)

$$\text{TPR} = \frac{\text{TP}}{\text{P}} = \frac{\text{TP}}{\text{TP} + \text{FN}} = 1 - \text{FNR}$$

specificity, selectivity or true negative rate (TNR)

$$\text{TNR} = \frac{\text{TN}}{\text{N}} = \frac{\text{TN}}{\text{TN} + \text{FP}} = 1 - \text{FPR}$$

precision or positive predictive value (PPV)

$$\text{PPV} = \frac{\text{TP}}{\text{TP} + \text{FP}} = 1 - \text{FDR}$$

negative predictive value (NPV)

$$\text{NPV} = \frac{\text{TN}}{\text{TN} + \text{FN}} = 1 - \text{FOR}$$

miss rate or false negative rate (FNR)

$$\text{FNR} = \frac{\text{FN}}{\text{P}} = \frac{\text{FN}}{\text{FN} + \text{TP}} = 1 - \text{TPR}$$

false out or false positive rate (FPR)

$$\text{FPR} = \frac{\text{FP}}{\text{N}} = \frac{\text{FP}}{\text{FP} + \text{TN}} = 1 - \text{TNR}$$

false discovery rate (FDR)

$$\text{FDR} = \frac{\text{FP}}{\text{FP} + \text{TP}} = 1 - \text{PPV}$$

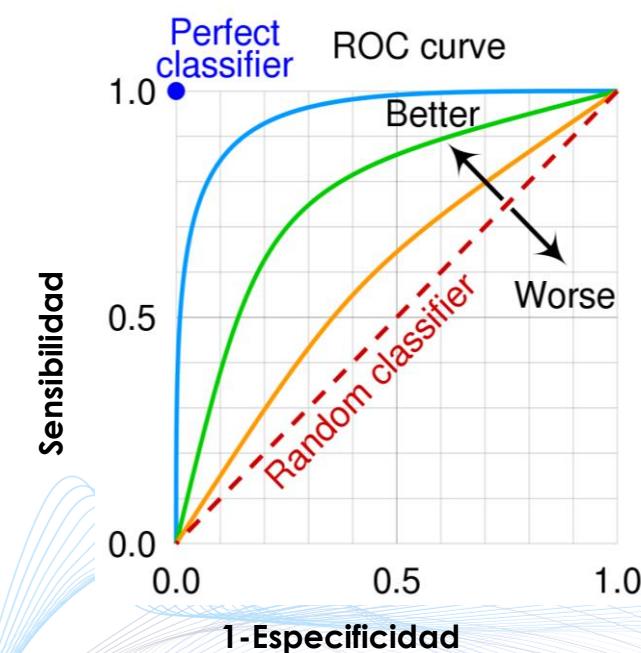
false omission rate (FOR)

$$\text{FOR} = \frac{\text{FN}}{\text{FN} + \text{TN}} = 1 - \text{NPV}$$

VALIDACIÓN DE UN MODELO PREDICTIVO

Modelos de clasificación: Predicen una clase, por ejemplo, larga, corta o media supervivencia.

Curva ROC y AUC



CONCLUSIONES

- **Los datos son la base de un Proyecto de IA**
 - Exactos
 - Coherentes
 - Consistentes
 - Proporcionados
 - Diversos
- **Existen muchas fuentes de datos en el entorno de la FM**
 - Accesibles
 - Estructurados
 - Etiquetados (o fácilmente etiquetables)

CONCLUSIONES

- La radiómica requiere un trabajo de procesado previo para asegurar la reproducibilidad de nuestro estudio
- Existen guías y código abierto (IBSI) para el correcto desarrollo de un proyecto de radiómica
- Los modelos predictivos están basados en la experiencia previa
- Utilizan características mixtas (demográficas, clínicas, radiómicas...)
- El pre-procesado del dataset es parte fundamental del proyecto



¡GRACIAS!

