Introducción al desarrollo de aplicaciones de IA

Ana Jiménez Pastor

Head Strategic Projects and Frontiers in AI

Quibim SL

INTELIGENCIA ARTIFICIAL

sefm Sociedad Española de Física Médica

Índice

- Creación de la base de datos
- Anotación
- Estrategias de entrenamiento:
  - Modelos basados en características
  - Modelos basados en imagen
- Estrategias de validación
- Despliegue de modelos de IA

Índice

- **Creación de la base de datos**
- Anotación
- Estrategias de entrenamiento:
  - Modelos basados en características
  - Modelos basados en imagen
- Estrategias de validación
- Despliegue de modelos de IA

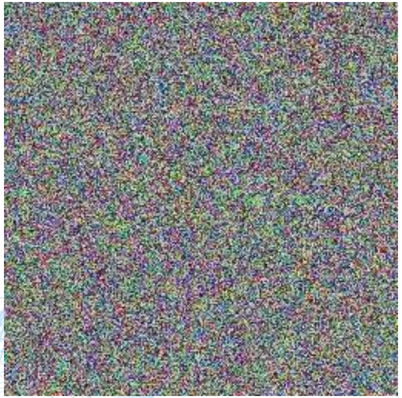
Creación de la base de datos

Imagen original

Tumor

+

Ruido



=

Imagen con ruido

Sano

Los modelos de IA son muy sensibles a pequeñas variaciones de la imagen de entrada

Creación de la base de datos

Requerimientos para el desarrollo de un modelo de IA robusto:

Creación de la base de datos

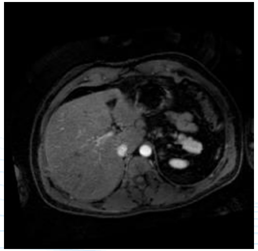
Requerimientos para el desarrollo de un modelo de IA robusto:

1. **Base de datos heterogénea:**
   - Mujeres y hombres
   - Diferentes edades
   - Diferentes nacionalidades
   - Multi-céntrico
   - Diferentes escáneres
   - Diferentes protocolos de adquisición
   - …

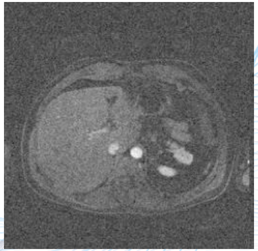
Creación de la base de datos

Requerimientos para el desarrollo de un modelo de IA robusto:

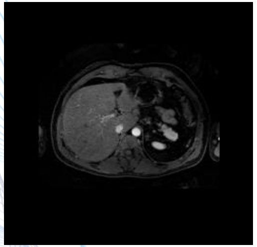
1. Base de datos heterogénea
2. **Técnicas de aumento de datos:**



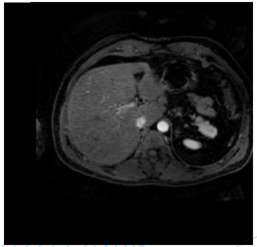
Rotaciones



Ruido



Zoom



Translaciones

…

Creación de la base de datos

Requerimientos para el desarrollo de un modelo de IA robusto:

1. Base de datos heterogénea
2. Técnicas de aumento de datos
3. **Validación externa**

Design Characteristics of Studies Reporting the Performance of Artificial Intelligence Algorithms for Diagnostic Analysis of Medical Images: Results from Recently Published Papers

Dong Wook Kim, MD[1]*, Hye Young Jang, MD[2]*, Kyung Won Kim, MD, PhD[2], Youngbin Shin, MS[2], Seong Ho Park, MD, PhD[2]

[1]Department of Radiology, Taean-gun Health Center and County Hospital, Taean-gun, Korea; [2]Department of Radiology and Research Institute of Radiology, University of Ulsan College of Medicine, Asan Medical Center, Seoul, Korea

Results: Of 516 eligible published studies, only 6% (31 studies) performed external validation. None of the 31 studies adopted all three design features: diagnostic cohort design, the inclusion of multiple institutions, and prospective data collection for external validation. No significant difference was found between medical and non-medical journals.

Creación de la base de datos

Requerimientos para el desarrollo de un modelo de IA robusto:

1. Base de datos heterogénea
2. Técnicas de aumento de datos
3. **Validación externa**

Checklist for Artificial Intelligence in Medical Imaging (CLAIM): A Guide for Authors and Reviewers

John Mongan, Linda Moy, Charles E. Kahn, Jr

Author Affiliations

Published Online: Mar 25 2020 | https://doi.org/10.1148/ryai.2020200029

Checklist for Artificial Intelligence in Medical Imaging (CLAIM)

Section/Topic	No.	Item
TITLE or ABSTRACT		
	1	Identification as a study of AI methodology, specifying the category of technology used (eg, deep learning)
ABSTRACT		
	2	Structured summary of study design, methods, results, and conclusions
INTRODUCTION		
	3	Scientific and clinical background, including the intended use and clinical role of the AI approach
	4	Study objectives and hypotheses
METHODS		
Study Design	5	Prospective or retrospective study
	6	Study goal, such as model creation, exploratory study, feasibility study, noninferiority trial
Data	7	Data sources
	8	Eligibility criteria: how, where, and when potentially eligible participants or studies were identified (eg, symptoms, results from previous tests, inclusion in registry, patient-care setting, location, dates)
	9	Data preprocessing steps
	10	Selection of data subsets, if applicable
	11	Definitions of data elements, with references to common data elements
	12	De-identification methods
	13	How missing data were handled
Ground Truth	14	Definition of ground truth reference standard, in sufficient detail to allow replication
	15	Rationale for choosing the reference standard (if alternatives exist)
	16	Source of ground truth annotations; qualifications and preparation of annotators
	17	Annotation tools
	18	Measurement of inter- and intrarater variability; methods to mitigate variability and/or resolve discrepancies
Data Partitions	19	Intended sample size and how it was determined
	20	How data were assigned to partitions; specify proportions
	21	Level at which partitions are disjoint (eg, image, study, patient, institution)
Model	22	Detailed description of model, including inputs, outputs, all intermediate layers and connections
	23	Software libraries, frameworks, and packages
	24	Initialization of model parameters (eg, randomization, transfer learning)
Training	25	Details of training approach, including data augmentation, hyperparameters, number of models trained
	26	Method of selecting the final model
	27	Ensembling techniques, if applicable
Evaluation	28	Metrics of model performance
	29	Statistical measures of significance and uncertainty (eg, confidence intervals)
	30	Robustness or sensitivity analysis
	31	Methods for explainability or interpretability (eg, saliency maps) and how they were validated
	32	Validation or testing on external data
RESULTS		
Data	33	Flow of participants or cases, using a diagram to indicate inclusion and exclusion
	34	Demographic and clinical characteristics of cases in each partition
Model performance	35	Performance metrics for optimal model(s) on all data partitions
	36	Estimates of diagnostic accuracy and their precision (such as 95% confidence intervals)
	37	Failure analysis of incorrectly classified cases
DISCUSSION		
	38	Study limitations, including potential bias, statistical uncertainty, and generalizability
	39	Implications for practice, including the intended use and/or clinical role
OTHER INFORMATION		
	40	Registration number and name of registry
	41	Where the full study protocol can be accessed
	42	Sources of funding and other support; role of funders

Creación de la base de datos

Requerimientos para el desarrollo de un modelo de IA robusto:

1. Base de datos heterogénea
2. Técnicas de aumento de datos
3. **Validación externa**

Ground Truth	14	Definition of ground truth reference standard, in sufficient detail to allow replication
	15	Rationale for choosing the reference standard (if alternatives exist)
	16	Source of ground truth annotations; qualifications and preparation of annotators
	17	Annotation tools
	18	Measurement of inter- and intrarater variability; methods to mitigate variability and/or resolve discrepancies
Data Partitions	19	Intended sample size and how it was determined
	20	How data were assigned to partitions; specify proportions
	21	Level at which partitions are disjoint (eg, image, study, patient, institution)
Model	22	Detailed description of model, including inputs, outputs, all intermediate layers and connections
	23	Software libraries, frameworks, and packages
	24	Initialization of model parameters (eg, randomization, transfer learning)
Training	25	Details of training approach, including data augmentation, hyperparameters, number of models trained
	26	Method of selecting the final model
	27	Ensembling techniques, if applicable
Evaluation	28	Metrics of model performance
	29	Statistical measures of significance and uncertainty (eg, confidence intervals)
	30	Robustness or sensitivity analysis
	31	Methods for explainability or interpretability (eg, saliency maps) and how they were validated
	32	Validation or testing on external data

Creación de la base de datos

Requerimientos para el desarrollo de un modelo de IA robusto:

1. Base de datos heterogénea
2. Técnicas de aumento de datos
3. **Validación externa**

Ground Truth	14	Definition of ground truth reference standard, in sufficient detail to allow replication
	15	Rationale for choosing the reference standard (if alternatives exist)
	16	Source of ground truth annotations; qualifications and preparation of annotators
	17	Annotation tools
	18	Measurement of inter- and intrarater variability; methods to mitigate variability and/or resolve discrepancies
Data Partitions	19	Intended sample size and how it was determined
	20	How data were assigned to partitions; specify proportions
	21	Level at which partitions are disjoint (eg, image, study, patient, institution)
Model	22	Detailed description of model, including inputs, outputs, all intermediate layers and connections
	23	Software libraries, frameworks, and packages
	24	Initialization of model parameters (eg, randomization, transfer learning)
Training	25	Details of training approach, including data augmentation, hyperparameters, number of models trained
	26	Method of selecting the final model
	27	Ensembling techniques, if applicable
Evaluation	28	Metrics of model performance
	29	Statistical measures of significance and uncertainty (eg, confidence intervals)
	30	Robustness or sensitivity analysis
	31	Methods for explainability or interpretability (eg, saliency maps) and how they were validated
	32	Validation or testing on external data

Creación de la base de datos

Requerimientos para el desarrollo de un modelo de IA robusto:

1. Base de datos heterogénea
2. Técnicas de aumento de datos
3. **Validación externa**

Ground Truth	14	Definition of ground truth reference standard, in sufficient detail to allow replication
	15	Rationale for choosing the reference standard (if alternatives exist)
	16	Source of ground truth annotations; qualifications and preparation of annotators
	17	Annotation tools
	18	Measurement of inter- and intrarater variability; methods to mitigate variability and/or resolve discrepancies
Data Partitions	19	Intended sample size and how it was determined
	20	How data were assigned to partitions; specify proportions
	21	Level at which partitions are disjoint (eg, image, study, patient, institution)
Model	22	Detailed description of model, including inputs, outputs, all intermediate layers and connections
	23	Software libraries, frameworks, and packages
	24	Initialization of model parameters (eg, randomization, transfer learning)
Training	25	Details of training approach, including data augmentation, hyperparameters, number of models trained
	26	Method of selecting the final model
	27	Ensembling techniques, if applicable
Evaluation	28	Metrics of model performance
	29	Statistical measures of significance and uncertainty (eg, confidence intervals)
	30	Robustness or sensitivity analysis
	31	Methods for explainability or interpretability (eg, saliency maps) and how they were validated
	32	Validation or testing on external data

Creación de la base de datos

Requerimientos para el desarrollo de un modelo de IA robusto:

1. Base de datos heterogénea
2. Técnicas de aumento de datos
3. **Validación externa**

Ground Truth	14	Definition of ground truth reference standard, in sufficient detail to allow replication
	15	Rationale for choosing the reference standard (if alternatives exist)
	16	Source of ground truth annotations; qualifications and preparation of annotators
	17	Annotation tools
	18	Measurement of inter- and intrarater variability; methods to mitigate variability and/or resolve discrepancies
Data Partitions	19	Intended sample size and how it was determined
	20	How data were assigned to partitions; specify proportions
	21	Level at which partitions are disjoint (eg, image, study, patient, institution)
Model	22	Detailed description of model, including inputs, outputs, all intermediate layers and connections
	23	Software libraries, frameworks, and packages
	24	Initialization of model parameters (eg, randomization, transfer learning)
Training	25	Details of training approach, including data augmentation, hyperparameters, number of models trained
	26	Method of selecting the final model
	27	Ensembling techniques, if applicable
Evaluation	28	Metrics of model performance
	29	Statistical measures of significance and uncertainty (eg, confidence intervals)
	30	Robustness or sensitivity analysis
	31	Methods for explainability or interpretability (eg, saliency maps) and how they were validated
	32	Validation or testing on external data

Creación de la base de datos

Requerimientos para el desarrollo de un modelo de IA robusto:

1. Base de datos heterogénea
2. Técnicas de aumento de datos
3. **Validación externa**

Ground Truth	14	Definition of ground truth reference standard, in sufficient detail to allow replication
	15	Rationale for choosing the reference standard (if alternatives exist)
	16	Source of ground truth annotations; qualifications and preparation of annotators
	17	Annotation tools
	18	Measurement of inter- and intrarater variability; methods to mitigate variability and/or resolve discrepancies
Data Partitions	19	Intended sample size and how it was determined
	20	How data were assigned to partitions; specify proportions
	21	Level at which partitions are disjoint (eg, image, study, patient, institution)
Model	22	Detailed description of model, including inputs, outputs, all intermediate layers and connections
	23	Software libraries, frameworks, and packages
	24	Initialization of model parameters (eg, randomization, transfer learning)
Training	25	Details of training approach, including data augmentation, hyperparameters, number of models trained
	26	Method of selecting the final model
	27	Ensembling techniques, if applicable
Evaluation	28	Metrics of model performance
	29	Statistical measures of significance and uncertainty (eg, confidence intervals)
	30	Robustness or sensitivity analysis
	31	Methods for explainability or interpretability (eg, saliency maps) and how they were validated
	32	Validation or testing on external data

Creación de la base de datos

Requerimientos para el desarrollo de un modelo de IA robusto:

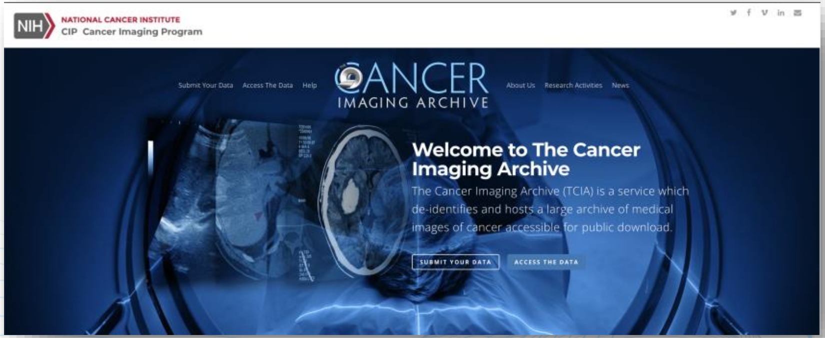
1. Base de datos heterogénea
2. Técnicas de aumento de datos
3. **Validación externa**

Certificados para poder vender como producto sanitario

Creación de la base de datos

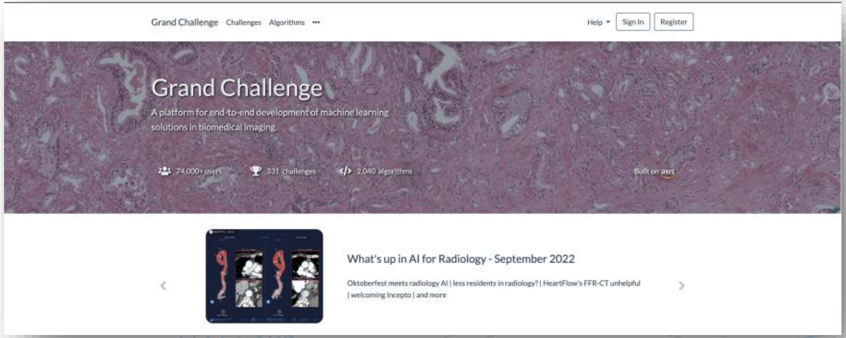
Bases de datos públicas

The Cancer Imaging Archive (TCIA)



https://www.cancerimagingarchive.net/

Grand Challenge

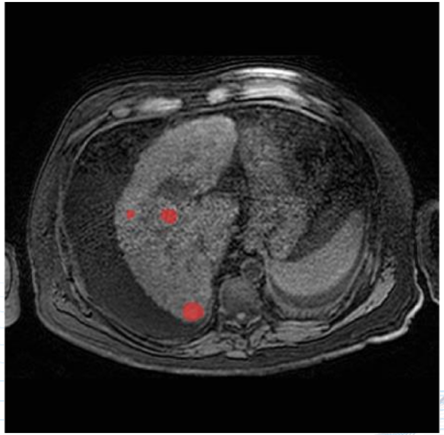


https://grand-challenge.org/

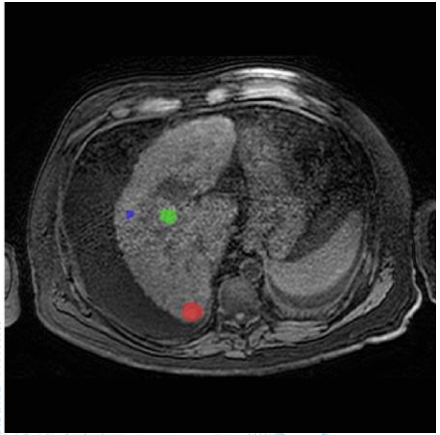
Índice

Anotación

Segmentación

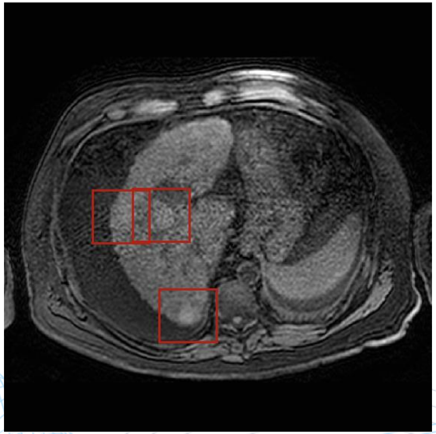


Segmentación semántica



Segmentación de instancias

Detección de objetos / Localización



Anotación

Herramientas

https://www.slicer.org/

http://www.itksnap.org/pmwiki/pmwiki.php

Anotación

Variabilidad intra- inter- observador → Guías de anotación y consenso

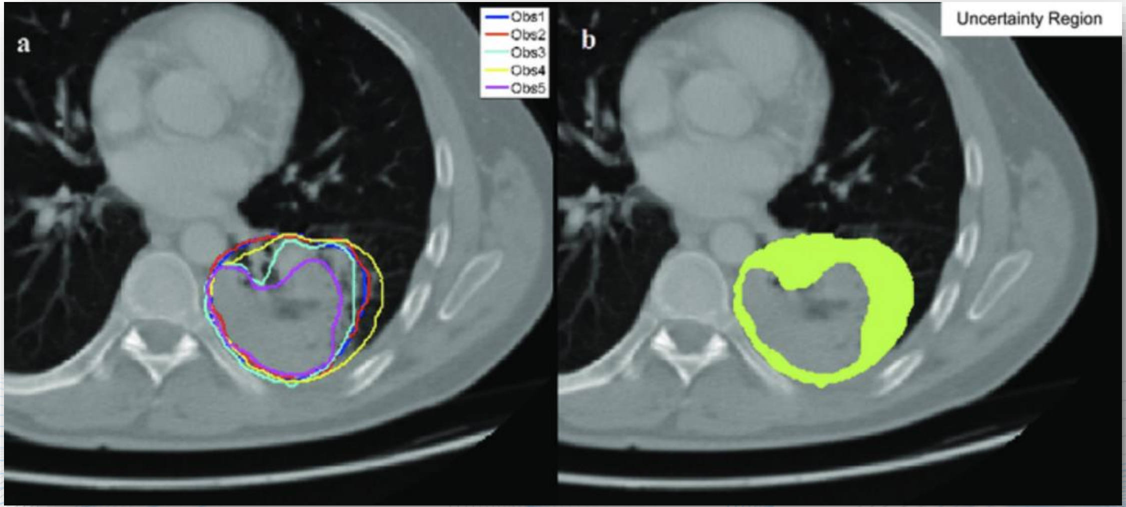
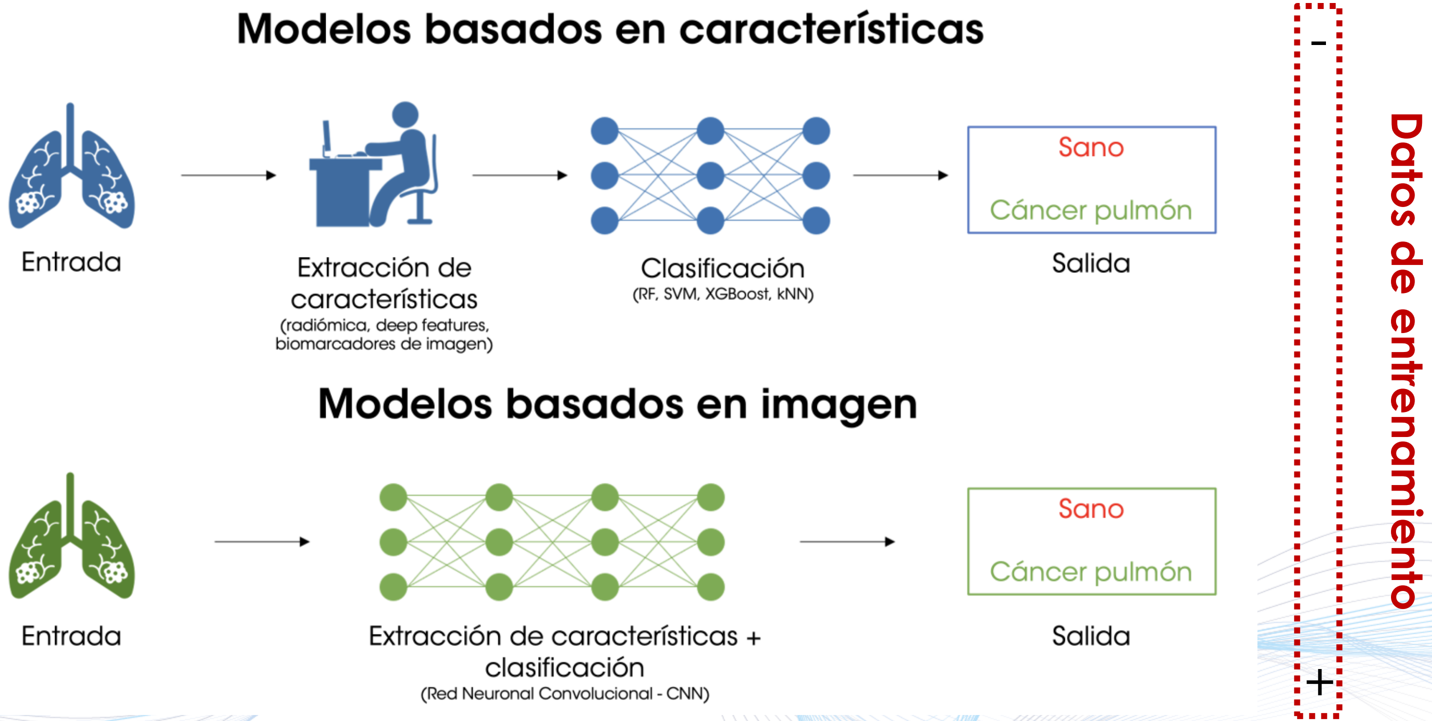


Image source: Rios Velazquez et al. Volumetric CT-based segmentation of NSCLC using 3D-slicer.

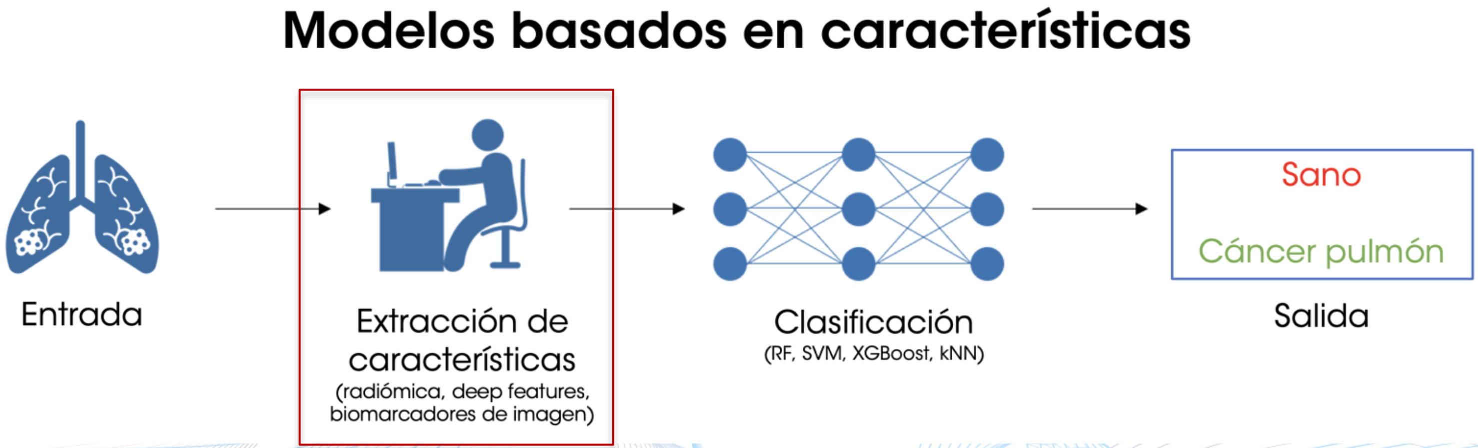
Índice

- Creación de la base de datos
- Anotación
- **Estrategias de entrenamiento:**
  - **Modelos basados en características**
  - **Modelos basados en imagen**
- Estrategias de validación
- Despliegue de modelos de IA

Estrategias de entrenamiento



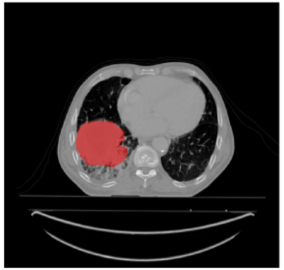
Estrategias de entrenamiento



Estrategias de entrenamiento

Modelos basados en características

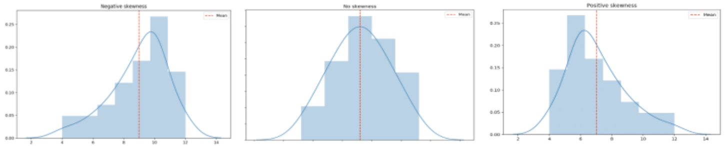
Características de radiómica



Características para el estudio de la heterogeneidad del tumor.

PyRadiomics*

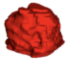
1. First order: Distribution of individual voxel values



2. Second order: Statistical inter-relationships between neighboring voxels

GLCM GLRLM GLSZM GLDM NGTDM

3. Shape: Geometric properties of the delineated ROI



4. Higher order features

Square, exponential, logarithm, wavelet, LoG

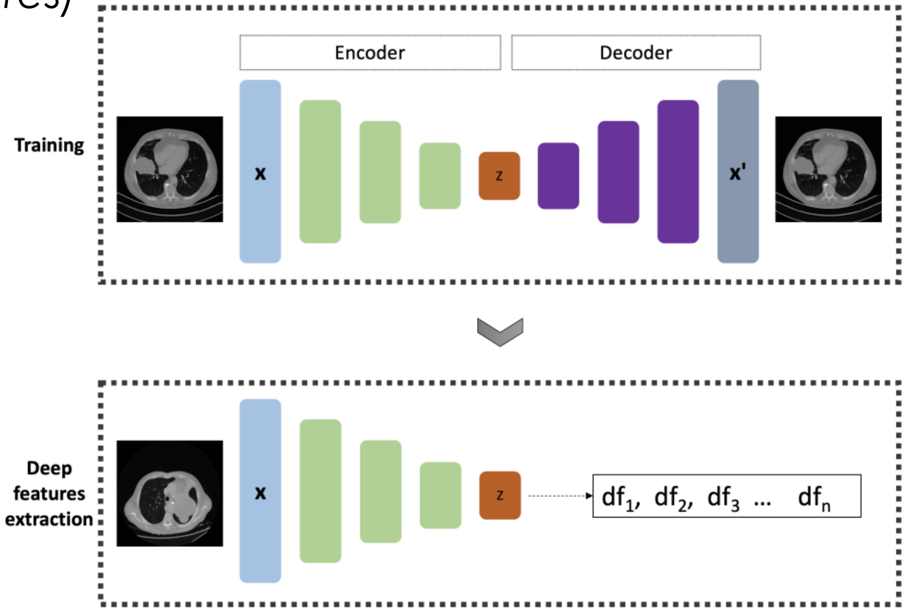
**van Griethuysen, J. J. M. et al. (2017). Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Research*

Estrategias de entrenamiento

Modelos basados en características

Características profundas (Deep features)

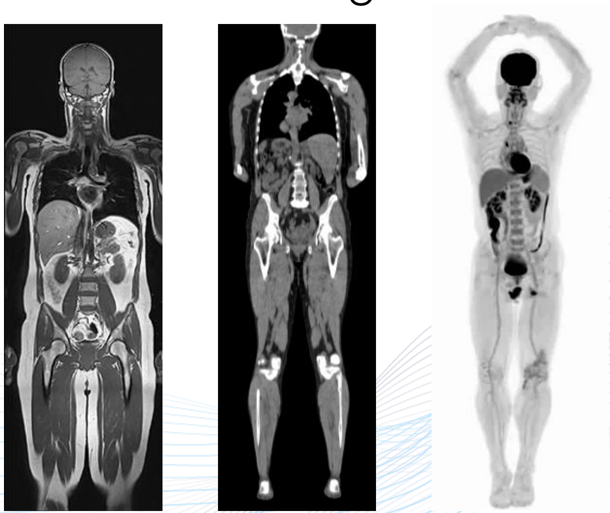
Se emplea la capacidad de aprendizaje de las **CNN**. Así, se entrena una **autocodificador** capaz de extraer la **información más relevante** de la imagen de entrada (características).
Una vez entrenado, se emplea el **codificador** para extraer dichas **características** de imágenes nuevas.



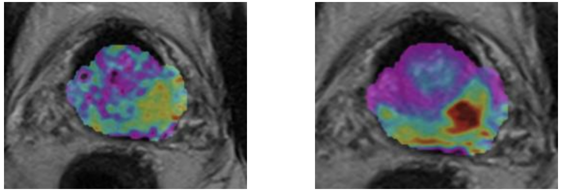
Estrategias de entrenamiento

Modelos basados en características

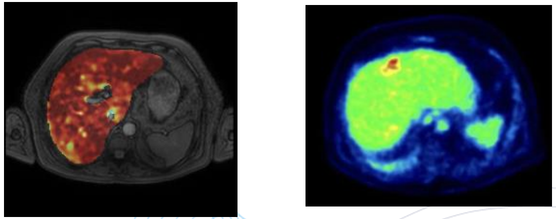
Biomarcadores de imagen



RM TC PET



ADC K^{trans}



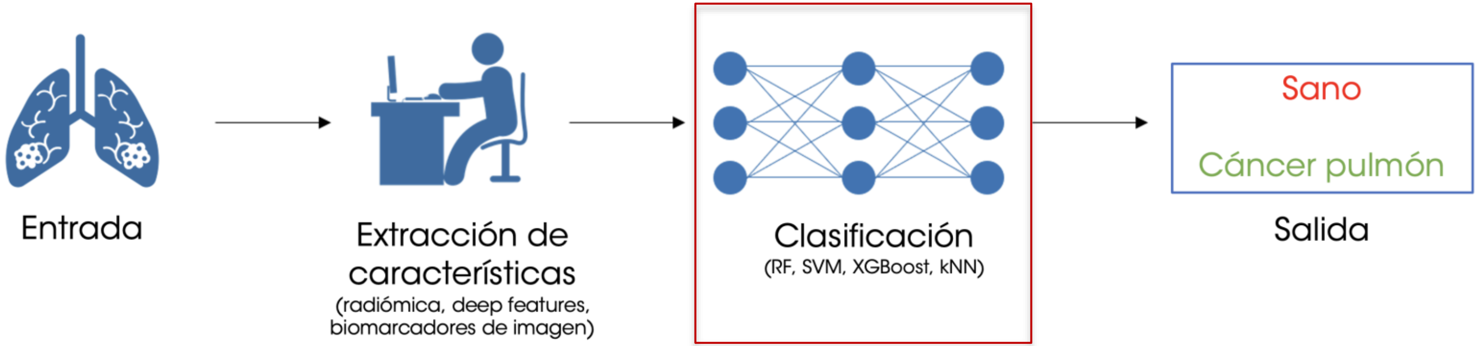
PDFF SUV

...

Fuente imágenes: ADC, Ktrans and PDFF (Quibim SL, Valencia, Spain). SUV (Dimitrakopoulou-Strauss, A. et al. Kinetic modeling and parametric imaging with dynamic PET for oncological applications: general considerations, current clinical applications, and future perspectives. Eur J Nucl Med Mol Imaging (2021)

Estrategias de entrenamiento

Modelos basados en características



Estrategias de entrenamiento

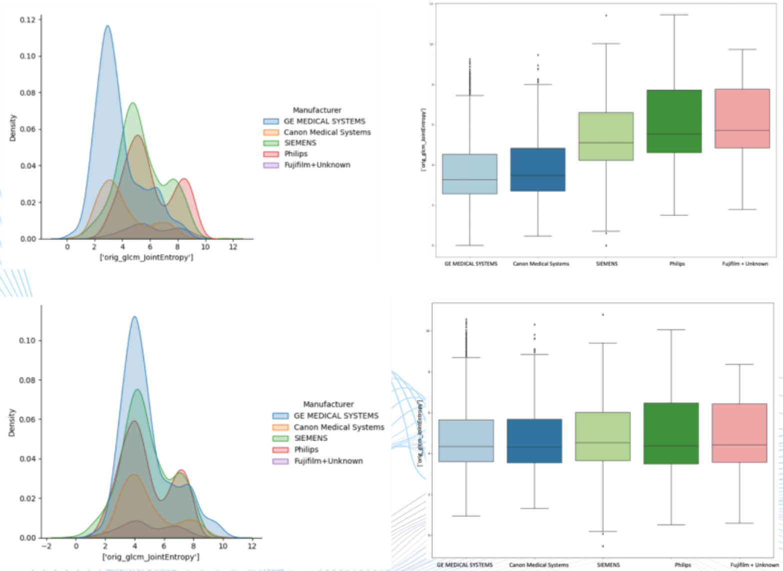
Modelos basados en características

Entrenamiento del modelo

1. Armonización de características

Las técnicas de armonización permiten reducir las diferencias entre fabricantes y protocolos de adquisición.

ComBat* es una técnica extendida que se empleó previamente en genómica.

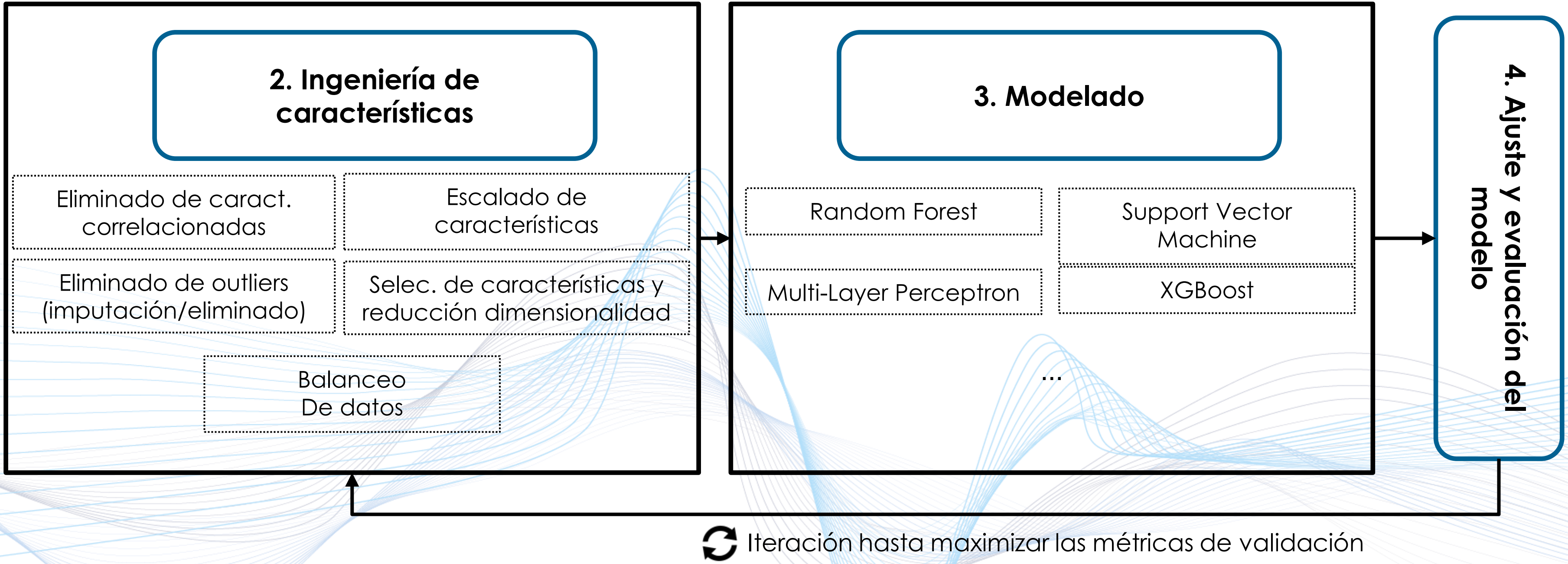


Johnson WE et. al. Adjusting batch effects in microarray expression data using empirical Bayes methods. Biostatistics. (2007)

Estrategias de entrenamiento

Modelos basados en características

Entrenamiento del modelo

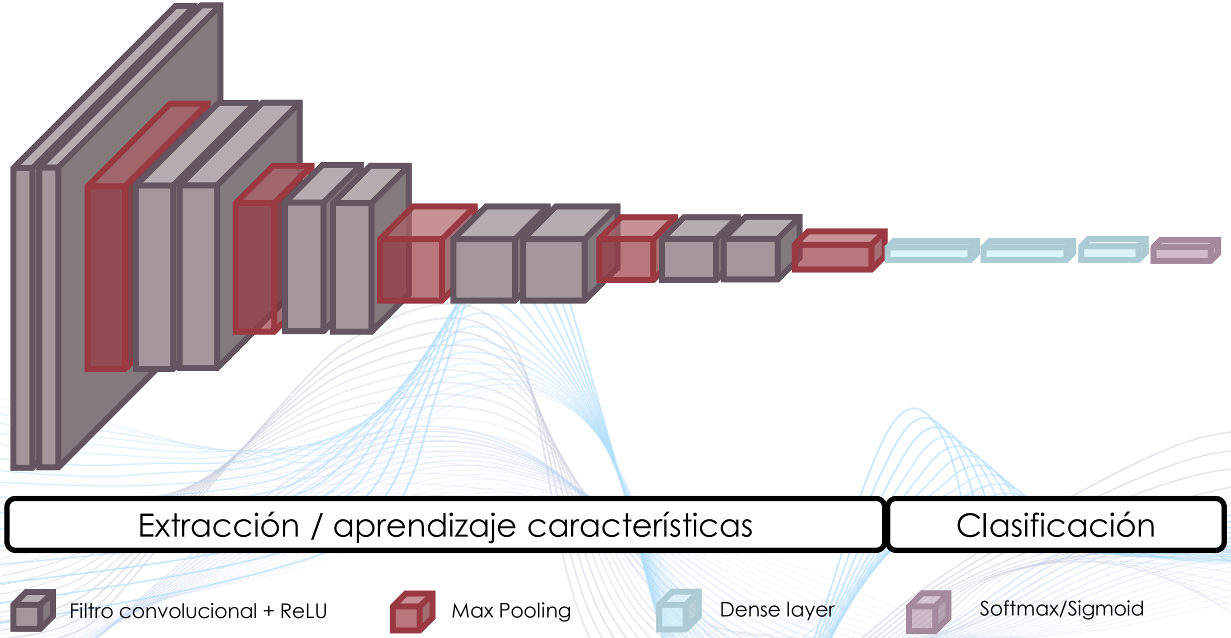


Estrategias de entrenamiento

Modelos basados en imagen

Estrategias de entrenamiento

Modelos basados en imagen



Índice

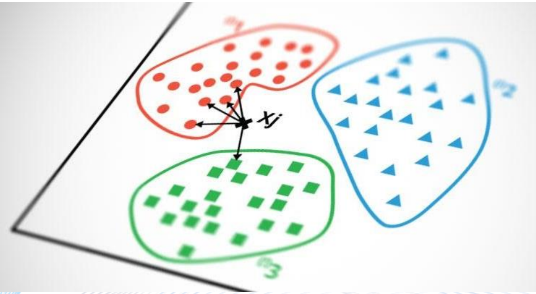
Estrategias de validación

Validación técnica

Comparación de los resultados contra una referencia (fantoma, DRO, dispositivo de referencia, anotaciones, etc.). Se emplea para la validación de los algoritmos de extracción de características.

Validación clínica

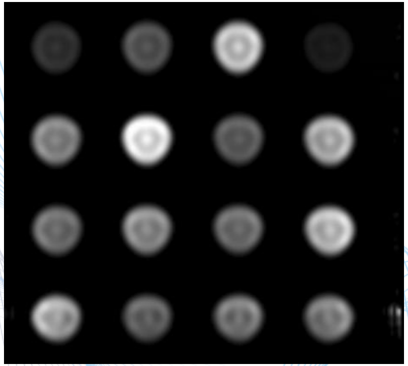
Correlación con un objetivo clínico (e.g., estadiaje, respuesta al tratamiento, tiempo de supervivencia, etc.).

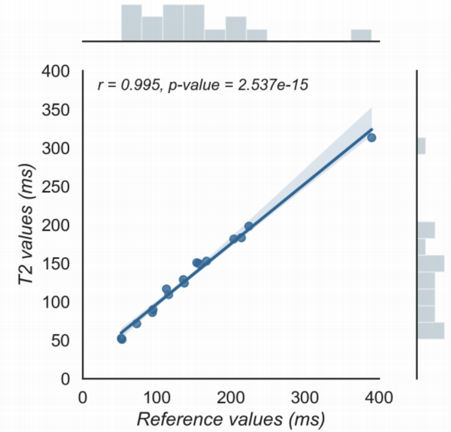


Estrategias de validación

Validación técnica. Fantoma

Objeto físico compuesto por tubos con valores conocidos (e.g., T1 y T2 en RM).



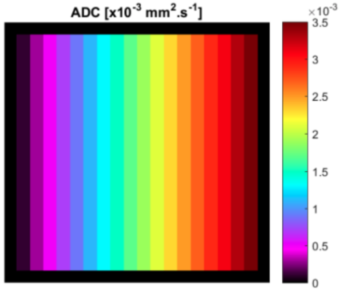


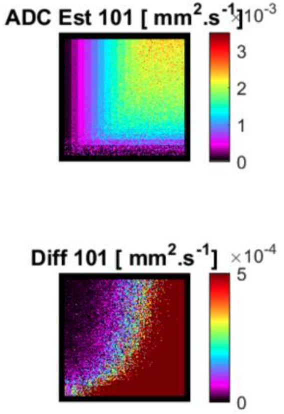
Validación de un algoritmo para la cuantificación del T2.

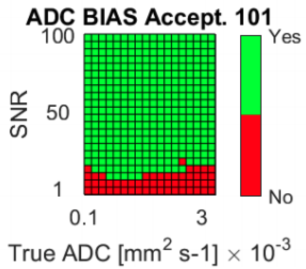
Estrategias de validación

Validación técnica. Objeto de referencia digital (DRO)

Imagen con los valores cuantitativos de referencia







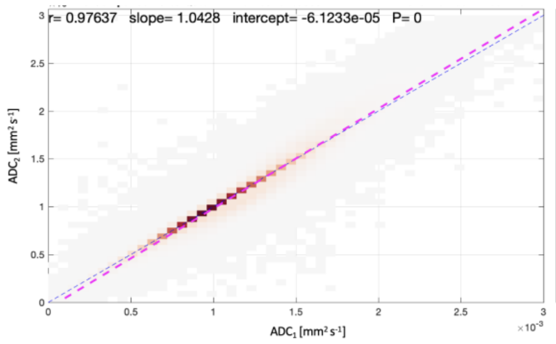
Perfil de QIBA : Imagen potenciada en diffusion (DWI). Validación de un algoritmo de cuantificación del coeficiente aparente de diffusion (ADC)

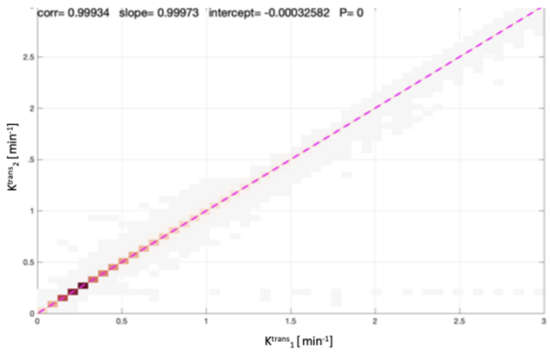
Quantitative Imaging Biomarkers Alliance RSNA

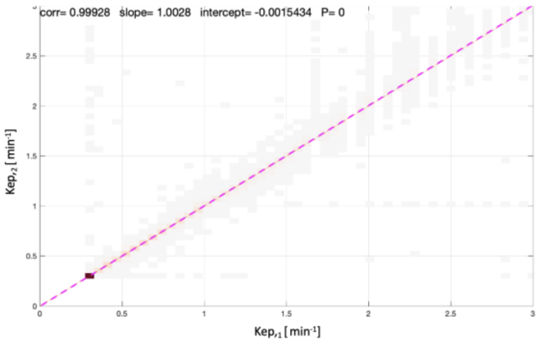
Estrategias de validación

Validación técnica. Dispositivo de referencia

Comparativa con un dispositivo médico previamente validado.





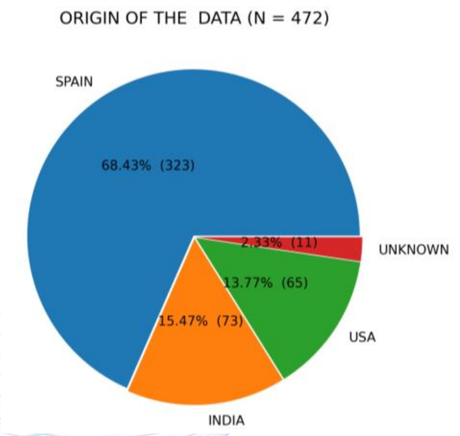


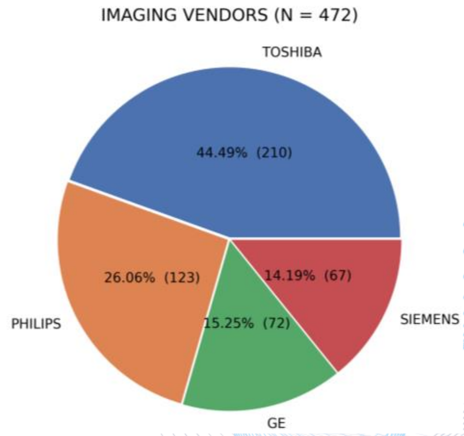
FDA 510k

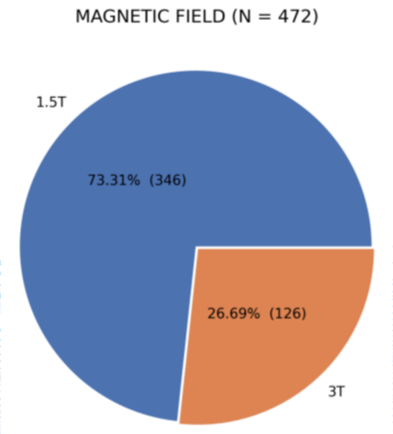
Estrategias de validación

Validación técnica. Anotaciones expertos

Comparativa con datos anotados manualmente y revisados por expertos.
Para garantizar la robustez del modelo hay que evaluarlo en diferentes escenarios (origen, fabricantes, campo magnético, etc.)

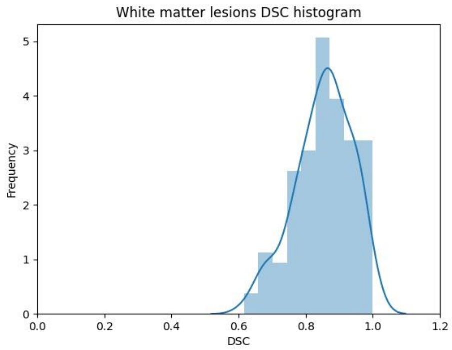


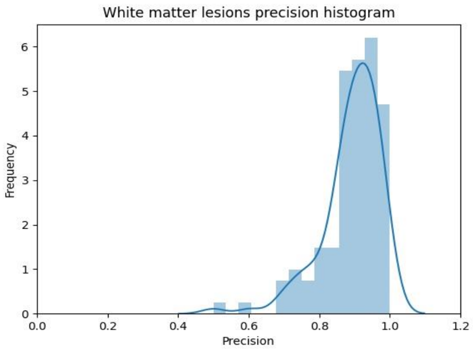




Estrategias de validación

Validación técnica. Anotaciones expertos. Segmentación





Métricas	Media	Std	Mediana
DSC	0.85 (0.84–0.87)	0.09 (0.06–0.10)	0.87 (0.85–0.88)
Precision	0.89 (0.88–0.91)	0.08 (0.07–0.10)	0.91 (89–0.92)
Recall	0.81 (0.79–0.83)	0.12 (0.10–0.14)	0.82 (0.80–0.84)

Validación de un algoritmo de segmentación automática de hiperintensidades de sustancia blanca.

Estrategias de validación

Validación clínica

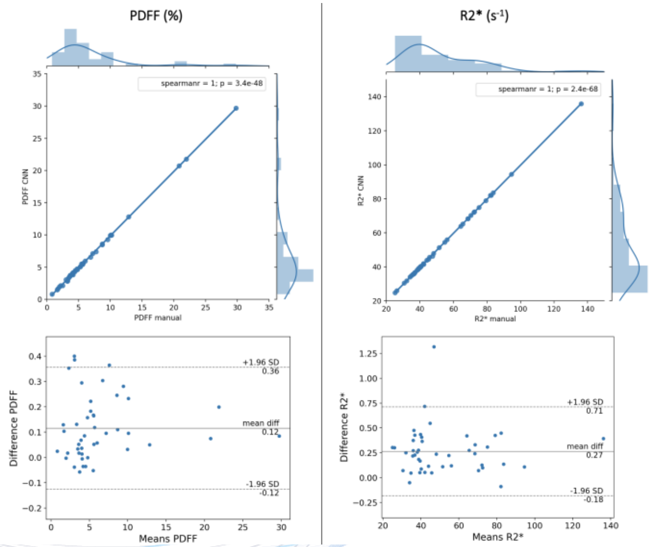
Comparación contra un objetivo clínico:

- **Modelos diagnósticos**: comparativa frente a estándar de referencia (e.g., biopsia). Estrategias de superioridad o no inferioridad.
- **Modelos predictivos**: análisis de supervivencia (e.g., supervivencia, tiempo libre de progresión), modelos de clasificación (e.g., response al tratamiento).

Estrategias de validación

Validación clínica

Valores cuantitativos: Segmentación CNN vs manual



Valores de PDFF vs. Grados esteatosis (biopsia)

Segmentation	Grades	Cut-off	AUC	95% CI	Sn	Sp
PDFF – Steatosis (%)						
ROI	S≥1	> 7.7	0.96	0.93-0.98	94	84
VOI		> 7.8	0.97	0.94-0.99	95	84
ROI	S≥2	> 10.2	0.95	0.92-0.98	92	90
VOI		> 10.8	0.96	0.93-0.99	88	91
ROI	S≥3	> 12.8	0.96	0.93-0.99	93	87
VOI		> 14.1	0.96	0.93-0.99	89	91

Validación de un algoritmo para la cuantificación de PDFF y R2 hepáticos*

Jimenez-Pastor, A. et al. Precise whole liver automatic segmentation and quantification of PDFF and R2* on MR images. Eur Radiol (2021).
Marti-Aguado D. et al. Automated Whole-Liver MRI Segmentation for Quantitative Assessment of Steatosis and Iron in Chronic Liver Disease: Correlations with Histologic Grades and Digital Image Analysis. Radiology (2022)

Estrategias de validación

Validación clínica

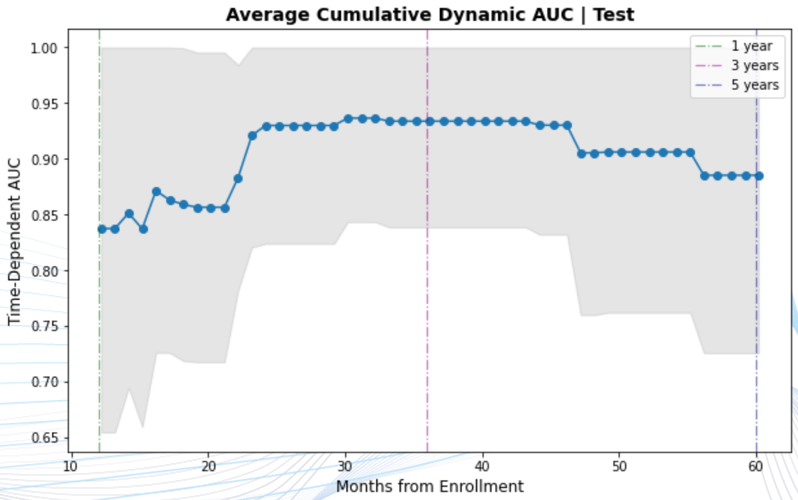
Validación de un algoritmo para la predicción del tiempo de supervivencia en pacientes con neuroblastoma

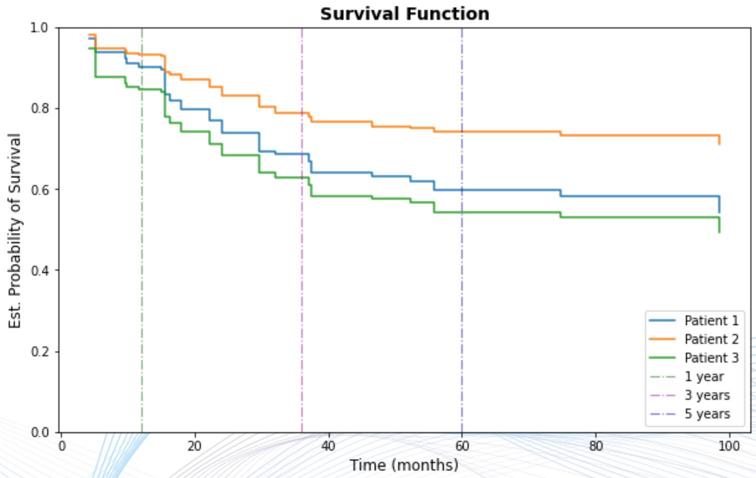
Model	Modifications	C. Index IPCW		C. Index		Mean AUC		Mean Brier Score	
		Train	Test	Train	Test	Train	Test	Train	Test
Cox	*Boruta*	0.745	0.709	0.785	0.728	0.829	0.875	0.108	0.117
ElasticNet Cox	*Boruta*	0.745	0.709	0.785	0.729	0.830	0.875	0.108	0.117
Random Survival Forest	***PCA***	**0.913**	**0.721**	**0.928**	**0.743**	**0.962**	**0.895**	**0.085**	**0.123**
Extra Survival Trees	*PCA*	0.869	0.718	0.893	0.747	0.928	0.827	0.093	0.127
Gradient Boosted Model	*PCA + Remove Outliers*	0.856	0.712	0.884	0.736	0.924	0.900	0.107	0.139

Estrategias de validación

Validación clínica

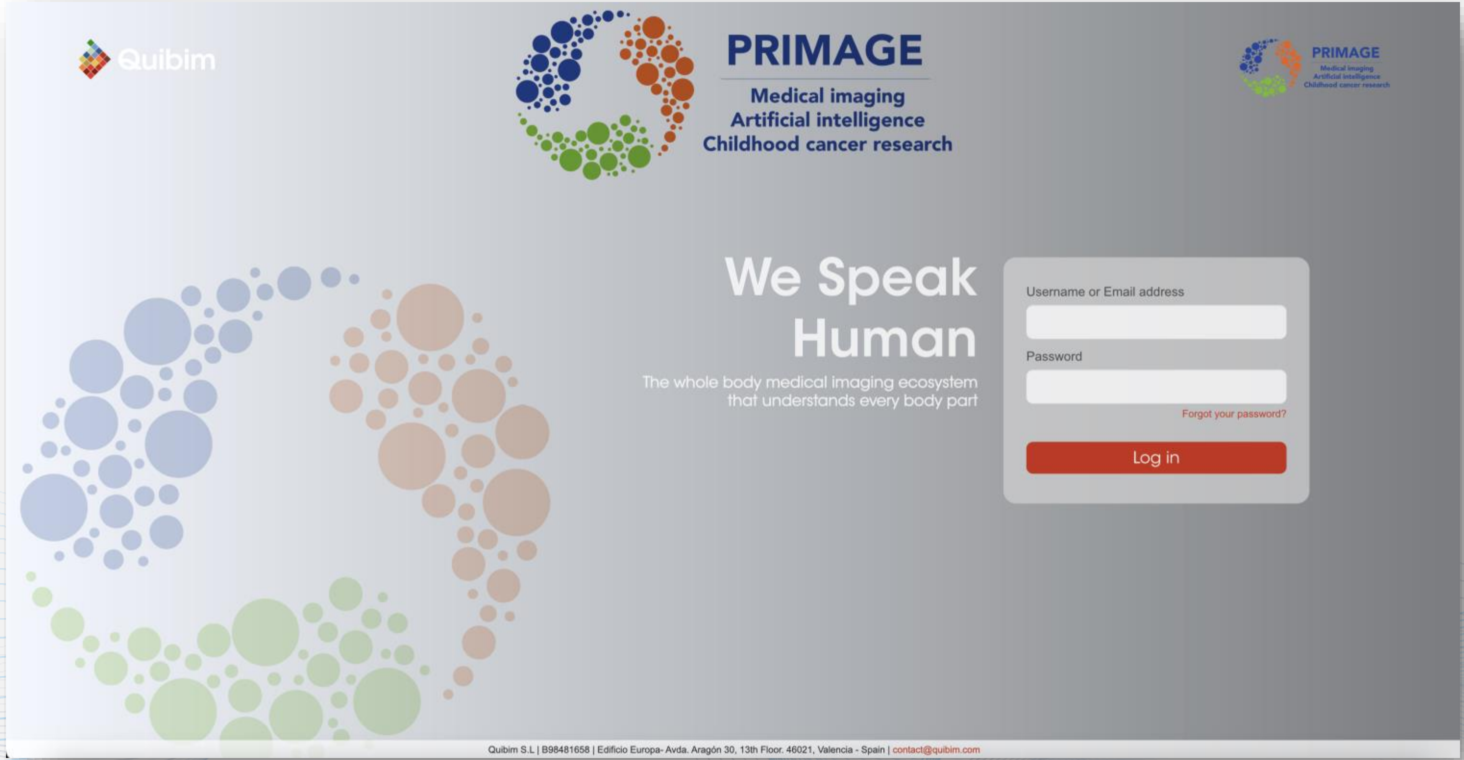
Validación de un algoritmo para la predicción del tiempo de supervivencia en pacientes con neuroblastoma





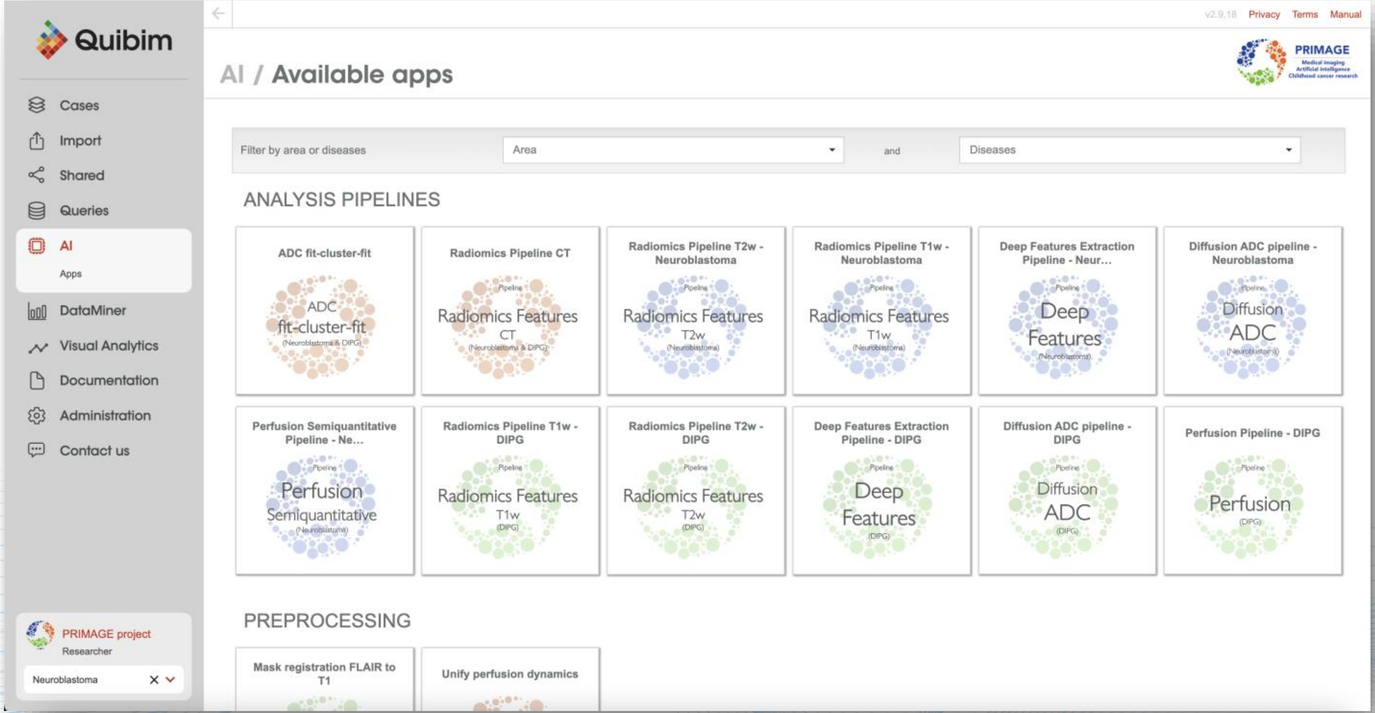
Índice

Despliegue de modelos de IA

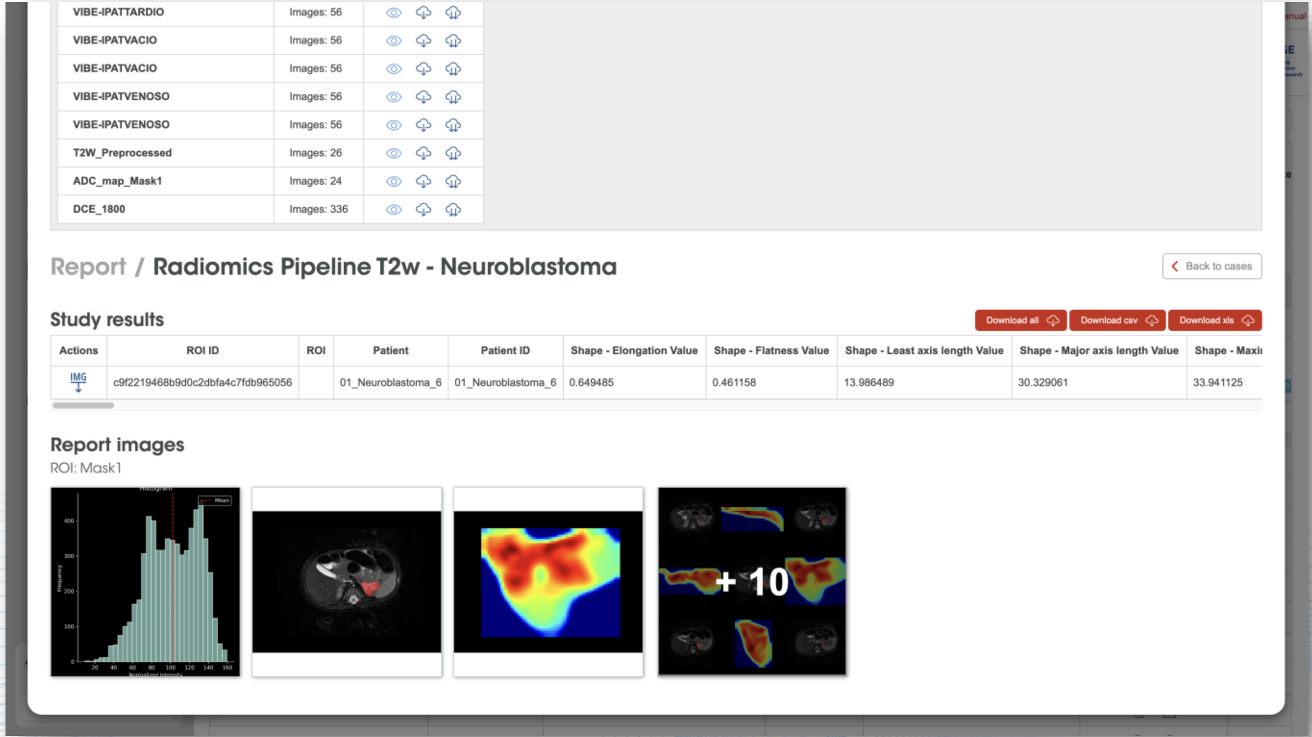


Sociedad Española de Física Médica

Despliegue de modelos de IA



Despliegue de modelos de IA



Conclusiones

- La colección y anotación de datos es un paso esencial para garantizar un buen rendimiento y generalización del modelo.

- Cuando se desarrolla un modelo de IA basado en imagen, podemos aproximarlo de dos formas: basado en características o basado en imagen.

- La validación externa del modelo es necesaria para garantizar la robustez y generalización del modelo.

- Hay diferentes maneras de validar un modelo que dependerá de su uso final.