

Armonización de radiografías de tórax para su uso en Inteligencia Artificial

Sara Gómez Peña¹, Ángel Nava¹,
Nerea Encina^{1,2}, Sohana Hassanaly^{2,3},
Joaquin L. Herraiz^{1,2}, Juan Arrazola¹

¹Instituto de investigación Sanitaria del Hospital Clínico San Carlos, Madrid

² Grupo de Física Nuclear. Dpto. EMFTEL e IPARCOS. Univ. Complutense de Madrid, Madrid

³ Leonard de Vinci Graduate School of Engineering, Paris

Objetivos

1. Mostrar uno de los principales retos de la Inteligencia Artificial (IA) en radiología: necesitar imágenes de calidad y propiedades muy distintas para conseguir un número suficiente de casos para entrenarse, requiere hacer estas imágenes compatibles
2. Destacar la importancia de tener en cuenta el origen y características de los casos empleados en el entrenamiento de herramientas de IA, para evitar introducir posibles sesgos de manera implícita
3. Mostrar técnicas de armonización de las imágenes empleadas en IA para disminuir esos sesgos

Motivación

- Si una herramienta de IA se entrena con casos de origen heterogéneo y no todas las patologías consideradas se adquirieron con los mismos equipos o siguiendo el mismo protocolo, la herramienta de IA **puede aprender a distinguir el equipo o protocolo en vez de las patologías en sí mismas**
- Este problema se ha visto claramente en el caso del análisis con IA de radiografías de pacientes de COVID-19, donde muchos resultados publicados presentan importantes sesgos

nature
machine intelligence

ANALYSIS

<https://doi.org/10.1038/s42256-021-00307-0>

Check for updates

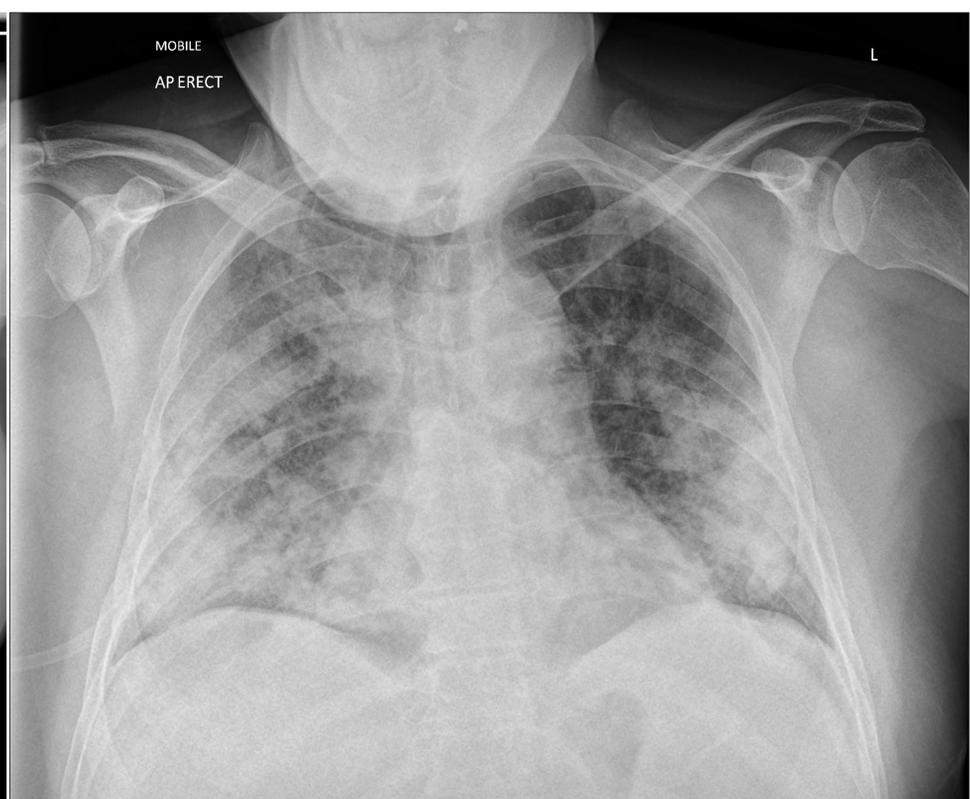
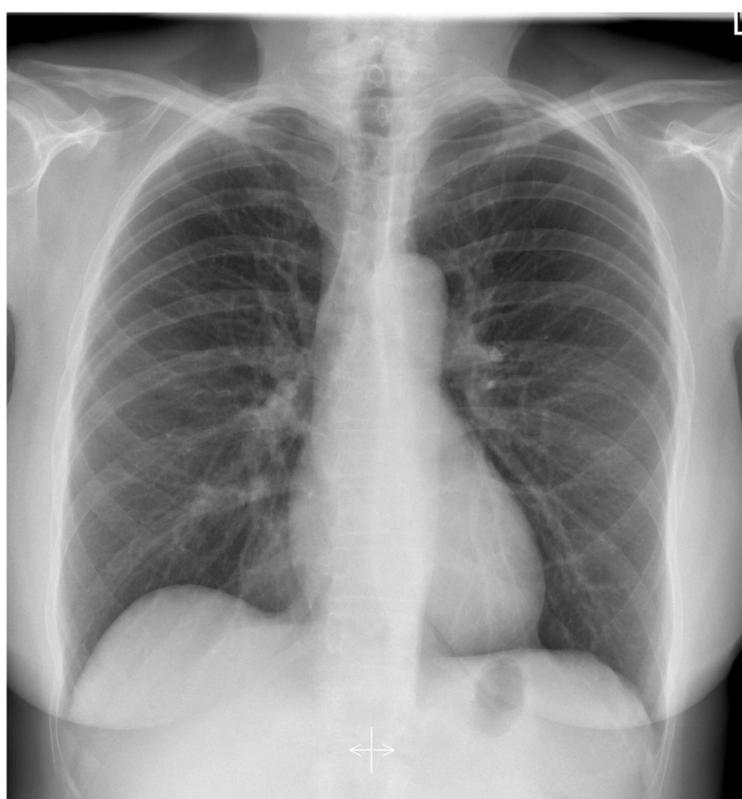
OPEN

Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans

Michael Roberts^{1,2}✉, Derek Driggs¹, Matthew Thorpe³, Julian Gilbey¹, Michael Yeung⁴, Stephan Ursprung^{4,5}, Angelica I. Aviles-Rivero¹, Christian Etmann¹, Cathal McCague^{4,5}, Lucian Beer⁴, Jonathan R. Weir-McCall^{4,6}, Zhongzhao Teng⁴, Effrossyni Gkrania-Klotsas⁷, AIX-COVNET*, James H. F. Rudd^{8,36}, Evis Sala^{4,5,36} and Carola-Bibiane Schönlieb^{1,36}

Motivación: Ejemplo

- Si se emplean radiografías de tórax de pacientes de COVID-19 adquiridas en el momento de su ingreso en el hospital:
 - *Radiografías con proyección posterior-anterior y en bipedestación: pacientes “más sanos”*
 - *Radiografías portátiles: pacientes “más enfermos” o con alta sospecha*
- La IA “*aprenderá*” a asociar una mayor gravedad a los pacientes que tengan una radiografía de equipo portátil, independientemente de lo que se detecte en esa radiografía



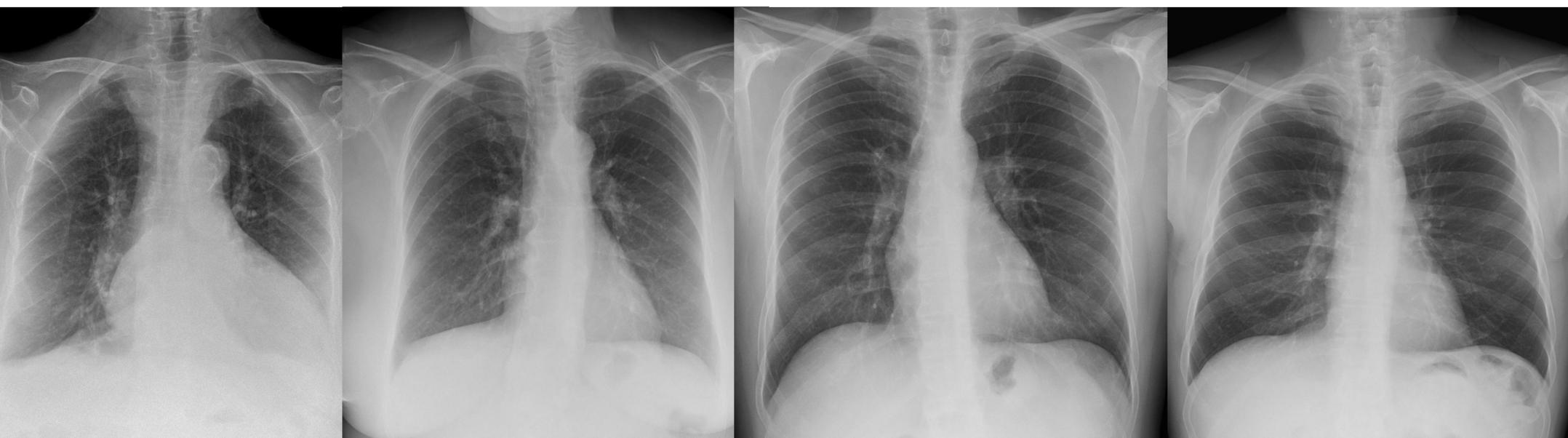
Introducción: La IA necesita miles de casos para aprender

- Existen grandes bases de datos públicas de radiografías **grandes bases de datos:**
 - CheXpert (Stanford AIMI Center)
 - Physionet (BIDMC, Boston, Massachusetts)
 - National Institutes of Health (NIH)
- Empresas como Google, Microsoft... también están firmando acuerdos con hospitales para aumentar sus bases de datos



Introducción: Existe una gran variabilidad en la calidad de las radiografías de tórax

- Para conseguir un gran número de casos con los que entrenar una herramienta de IA, se necesita usar adquisiciones obtenidas en **múltiples centros, en distintos equipos de rayos X, con distintos protocolos...**
- **Las radiografías de tórax presentan mucha más variabilidad** (tamaño, contraste, resolución..) que otras técnicas como CT o MR



Introducción: Armonización de imágenes médicas

- Existe un esfuerzo creciente por **armonizar adquisiciones y bases de datos de imagen médica** para reducir la dependencia de los resultados con el protocolo de adquisición, tipo de equipo...
- Esto es **especialmente importante en IA y en estudios multicentro**
- Se busca evitar los sesgos que podría haber si determinada patología se ha estudiado más en un centro o con un equipo determinado

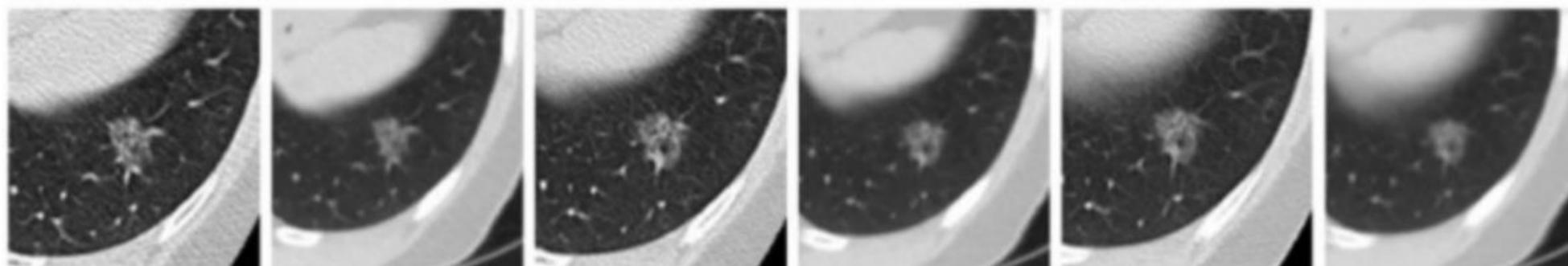


Figura 1. Tumor pulmonar identificado en 6 estudios de TC realizados en el mismo equipo pero con protocolos de adquisición diferentes: espesor de corte de 1,25 mm y reconstrucción con ventana pulmonar. Foto tomada de [3].

Métodos: Datos

- Se emplearon radiografías de tórax de pacientes con neumonía y “sanos” en formato DICOM de la base de datos anonimizada de **Physionet (MIMIC-CXR Database)** [4]:
 - ~377.000 radiografías obtenidas principalmente en el Beth Israel Deaconess Medical Center (BIDMC, Boston, Massachusetts)
 - Etiquetadas por radiólogos, de las cuales se dispone, así como del informe radiológico

PhysioNet

The Research Resource for Complex Physiologic Signals

[Data](#)[Software](#)[Challenges](#)[Tutorials](#)[Database](#) [Credentialed Access](#)

MIMIC-CXR Database

Alistair Johnson¹, Tom Pollard¹, Roger Mark¹, Seth Berkowitz¹, Steven Horng¹

Published: Sept. 19, 2019. Version: 2.0.0

MIMIC-CXR paper published! (Feb. 10, 2020, 4:06 p.m.)

SCIENTIFIC DATA

OPEN

DATA DESCRIPTOR

MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports

Alistair E. W. Johnson¹, Tom J. Pollard¹, Seth J. Berkowitz², Nathaniel R. Greenbaum³, Matthew P. Lungren⁴, Chih-ying Deng⁵, Roger G. Mark⁶ & Steven Horng⁷

Métodos: Anonimización

- Al ser radiografías anonimizadas **hemos perdido información útil presente en las etiquetas del DICOM original**
- Campos útiles para tener en cuenta a la hora de armonizar estudios:
 - ✓ Tipo de escáner
 - ✓ Campo de visión
 - ✓ Voltaje empleado
 - ✓ Protocolo de adquisición
- Sin esta información, puede parecer que todas las imágenes de una base de datos se han adquirido de forma similar, que son totalmente compatibles

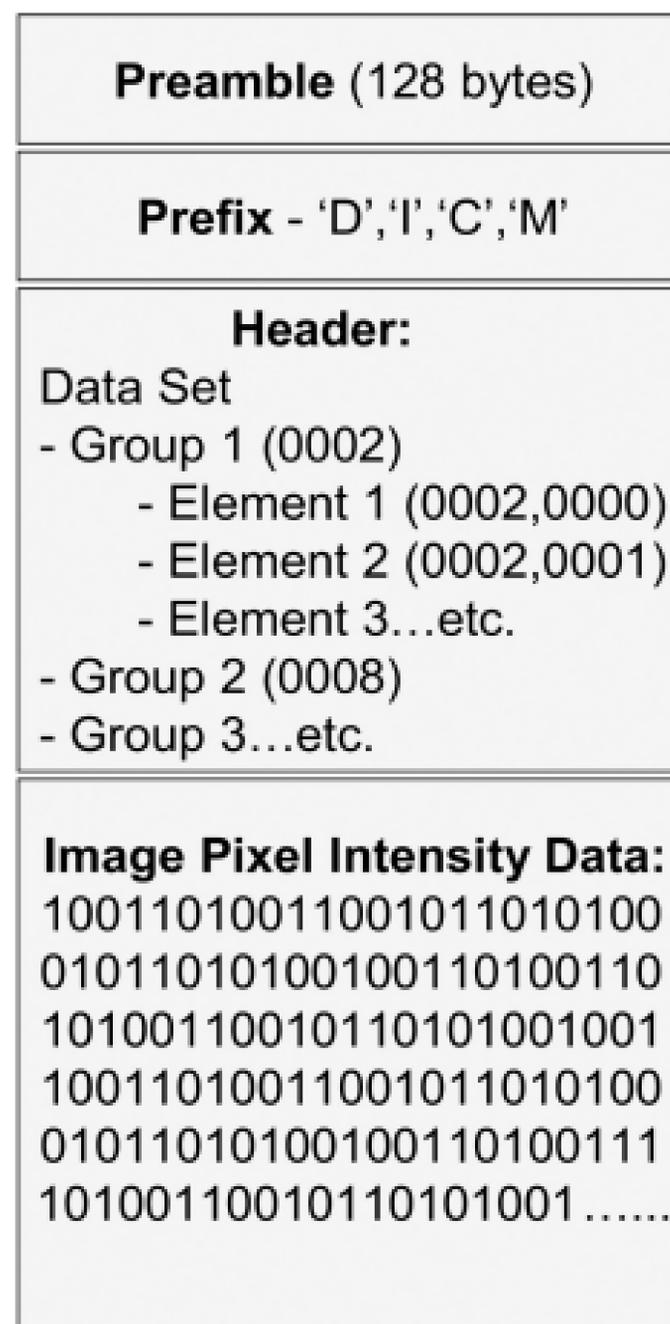


Figura 2. Estructura de un archivo de imagen DICOM, Imagen tomada de [6]

Métodos: Datos

- Se extrajeron **1.430 radiografías de tórax** de pacientes con neumonía y “*sanos*” en formato DICOM adquiridas con **3 equipos de rayos X fijos** en proyección posteroanterior
- Aunque la **información del tipo de equipo empleado no figuraba** en los metadatos de los ficheros, se pudo obtener analizando el número de píxeles de las radiografías en cada caso:

	Píxeles	N sanos	N neumonía
ESCÁNER A	3056 x 2544	27695	2028
ESCÁNER B	2021 x 2021	1055	415
ESCÁNER C	2022 x 1736	2588	445

Métodos: Armonización

1. Adaptación de las dimensiones y el campo de visión

Obteniendo un tamaño de pixel igual en todos los casos

2. Adaptación del rango dinámico de las radiografías para hacerlas similares entre sí

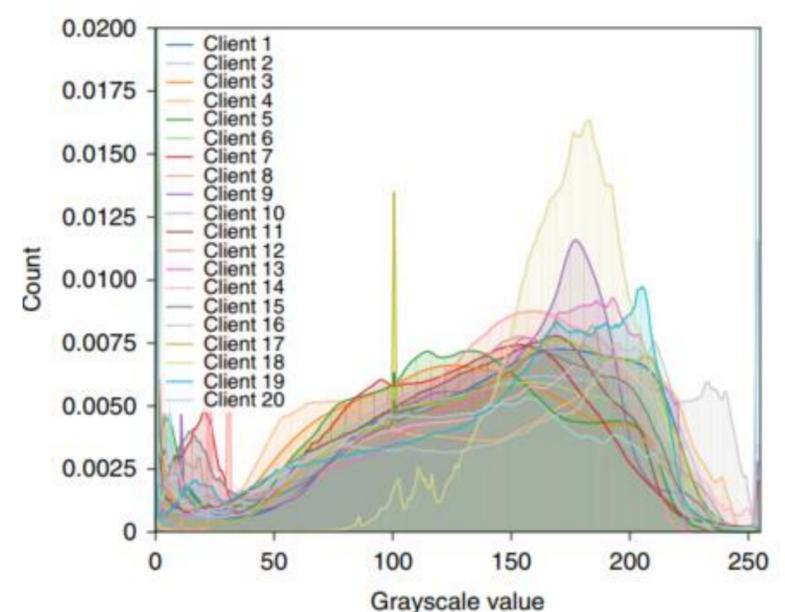
3. Métodos de ajuste de histograma

- CLAHE
- RAVEL
- CYCLE-GAN

4. Armonización de las características extraídas

- COMBAT

Figura 3. Distribución de la intensidad de la radiografía de tórax en cada equipo [7]



Métodos: Estudios IA

ESTUDIO 1. Tras los distintos pasos de armonización de las radiografías, se entrenó una herramienta de IA para determinar si diferenciaba el tipo de equipo en el que se había realizado cada radiografía

ESTUDIO 2. Se entrenó una herramienta de IA usando sólo casos de un equipo determinado para que clasificase las radiografías como pacientes “*sanos*” o con neumonía

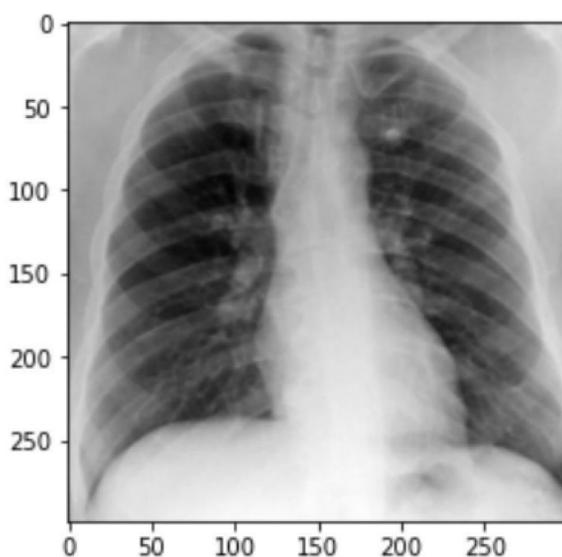
ESTUDIO 3. Con las imágenes ya armonizadas, se evaluó si una red entrenada para distinguir pacientes “*sanos*” frente a pacientes con neumonía, usando sólo casos de un equipo determinado, funcionaba bien cuando se aplicaba a radiografías realizadas en otro equipo

Resultados: Estudios 1

“Distinguir tipo de equipo”

- Aunque la información respecto al equipo fue eliminada de los ficheros DICOM al anonimizarlos, **se puede estimar a partir del número de píxeles** de las imágenes
- Aunque se adapten las dimensiones de las CXR de cada equipo para hacerlas compatibles, todavía es posible diferenciarlas con una red neuronal. En muchos casos es sencillo, porque sus histogramas de valores son muy distintos

Equipo A



Equipo C

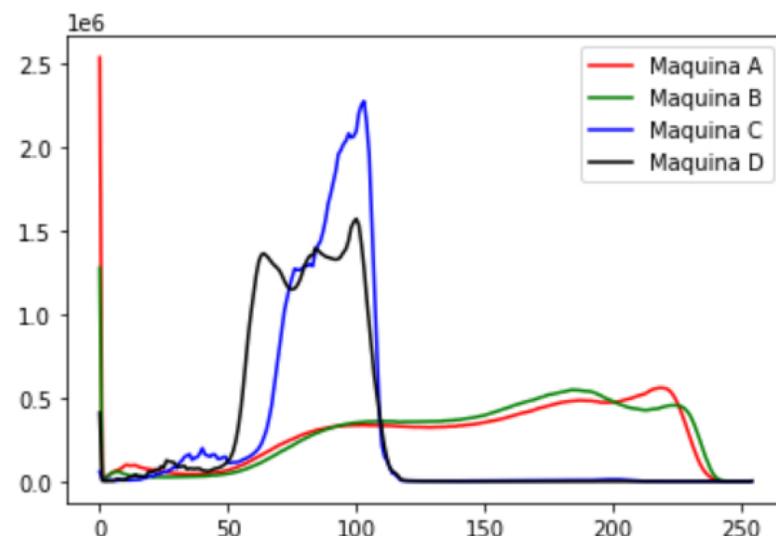
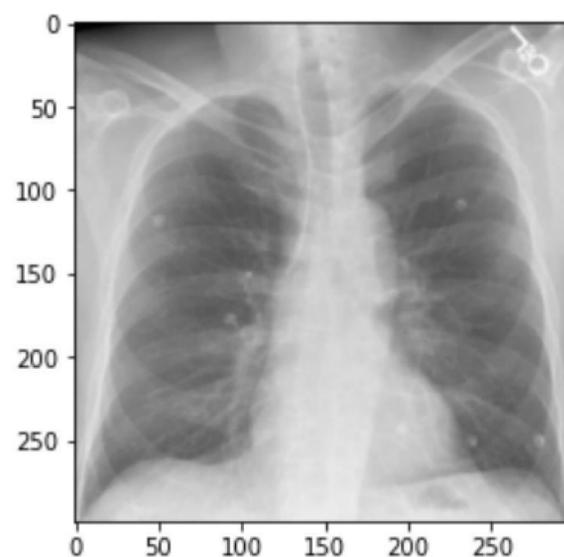


Figura 4. Histograma de valores promedio de cada máquina

Resultados: Estudios 1

“Distinguir tipo de equipo”

		EQUIPO ESTIMADO			
		A	B	C	D
EQUIPO REAL	A	212	40	1	0
	B	31	245	0	0
	C	2	0	265	0
	D	0	0	7	253

Figura 5. Matriz de confusión de la estimación obtenida por una red neuronal en un conjunto de casos seleccionados de cada equipo

Resultados: Estudios 2

“Clasificar las radiografías de un mismo equipo”

- Se entrenó una red neuronal con el 80% de los casos y se usó el otro 20% (812 casos) para evaluar el modelo

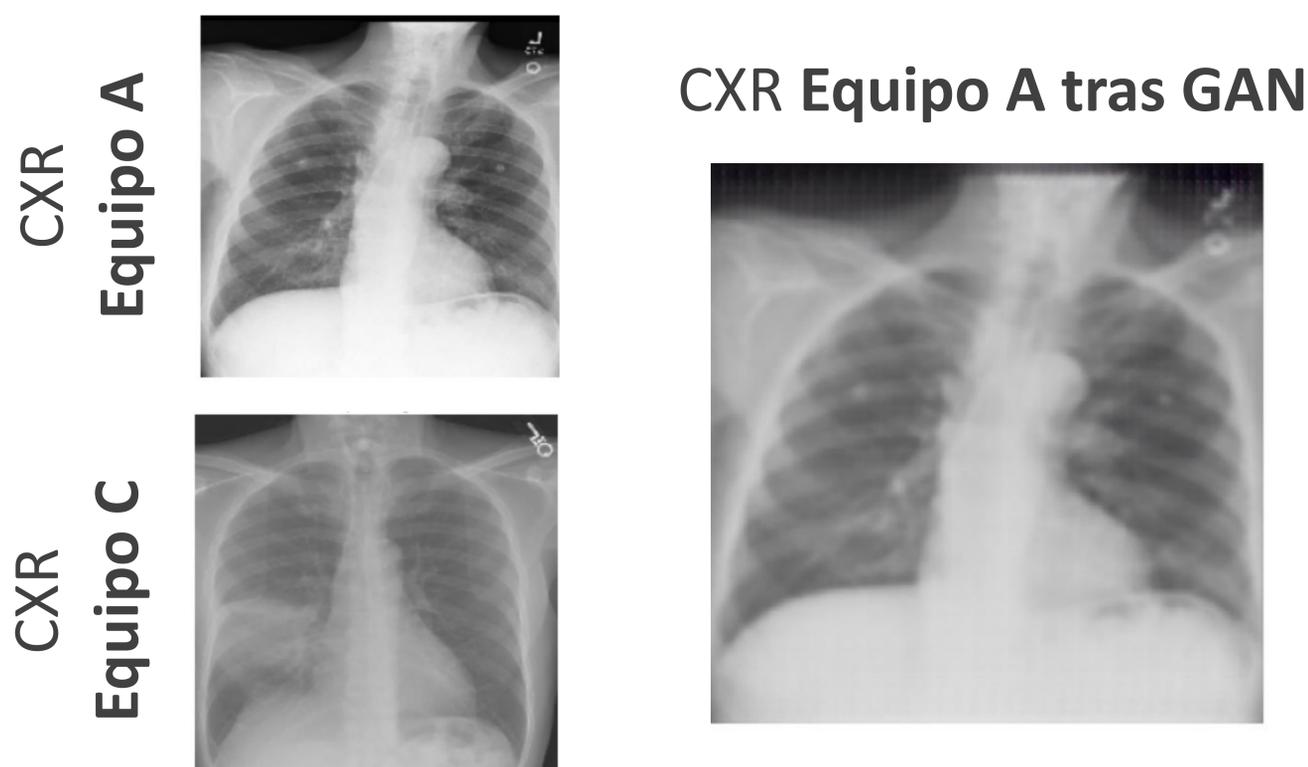
Accuracy = 80%		ESTIMACIÓN	
		SANO	NEUMONÍA
ESTADO DEL PACIENTE	SANO	360	76
	NEUMONÍA	89	287

Figura 6. Matriz de confusión de la estimación obtenida por una red neurona en un conjunto de casos seleccionados de la máquina A

Resultados: Estudios 3

“Usar un modelo entrenando en otro equipo”

- La red entrenada en el estudio 2 (equipo A), aplicada a casos de otro equipo (equipo C) no funciona (50% de acierto)
- Con **herramientas de ecualización y segmentación** de la región pulmonar y cardiaca previa al entrenamiento se logra en ese caso un **acierto del 73%**



Las redes GAN permiten que las radiografías del equipo A sean más similares a las del equipo C y viceversa. Esto permite que modelos entrenados con un equipo funcionen mejor aplicados a casos del otro equipo

Resultados: Estudios 3

“Usar un modelo entrenando en otro equipo”

Accuracy = 73%		ESTIMACIÓN	
		SANO	NEUMONÍA
ESTADO DEL PACIENTE	SANO	338	98
	NEUMONÍA	123	253

Figura 7. Resultados de modelo entrenado con equipo A aplicado a equipo C

Conclusiones

1. Es muy importante **tener en cuenta el equipo de rayos X y el protocolo** con el que se adquieren las radiografías cuando se quieren emplear herramientas de IA
2. En los últimos años se están proponiendo muchas **técnicas de armonización** que pueden ayudar a realizar un preprocesado de las radiografías y **eliminar posibles sesgos**
3. Además de armonizar las imágenes, también se puede hacer la **armonización sobre las características (*features*)** extraídas de ellas
4. La armonización requiere tener acceso a un número suficiente de casos del nuevo equipo al que se quiere aplicar los modelos pre-entrenados, por lo que no siempre es factible

Bibliografía

1. Roberts M, Driggs D, Thorpe M, Gilbey J, Yeung M, Ursprung S et al. Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nature Machine Intelligence*. 2021;3(3):199-217
2. Nan Y, Ser J, Walsh S, Schönlieb C, Roberts M, Selby I et al. Data harmonisation for information fusion in digital healthcare: A state-of-the-art systematic review, meta-analysis and future research directions. *Information Fusion*. 2022;82:99-122
3. Zhao B, Tan Y, Tsai W, Qi J, Xie C, Lu L et al. Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Scientific Reports*. 2016;6(1)
4. Johnson, A., Lungren, M., Peng, Y., Lu, Z., Mark, R., Berkowitz, S., & Horng, S. (2019). MIMIC-CXR-JPG - chest radiographs with structured labels (version 2.0.0). *PhysioNet*.
5. Johnson A, Pollard T, Berkowitz S, Greenbaum N, Lungren M, Deng C et al. MIMIC-CXR, a de-identified publicly available database of chest radiographs with free-text reports. *Scientific Data*. 2019;6(1)
6. Varma D. Managing DICOM images: Tips and tricks for the radiologist. *Indian Journal of Radiology and Imaging*. 2012;22(01):4-13
7. Dayan, I., Roth, H.R., Zhong, A. et al. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat Med* 27, 1735–1743 (2021)
8. Tixier F, Jaouen V, Hognon C, Gallinato O, Colin T, Visvikis D. Evaluation of conventional and deep learning based image harmonization methods in radiomics studies. *Physics in Medicine & Biology*. 2021;66(24):245009