



Disponible en línea en [www.sciencedirect.com](http://www.sciencedirect.com)

CienciaDirecta

página de inicio de la revista: [www.ejancer.com](http://www.ejancer.com)



## Investigación original

# Un modelo de inteligencia artificial predice la supervivencia de pacientes con tumores sólidos a partir de imágenes y datos clínicos



Kathryn Schutte<sup>a,\*</sup>, Fabien Brulport<sup>a</sup>, Sana Harguem-Zayani<sup>b</sup>, Jean-Baptiste Schiratti<sup>a</sup>, Riduane Ghermi<sup>a</sup>, Pablo Jehanno<sup>b</sup>, Alejandro Jaeger<sup>un,e</sup>, Talal Alamri<sup>b</sup>, Rafael Naccache<sup>c</sup>, Leila Haddag-Miliani<sup>b</sup>, teresa orsi<sup>b</sup>, Jean Philippe Lamarque<sup>c</sup>, Isaline Hoferer<sup>b, d</sup>, Littisha Lawrance<sup>b, d</sup>, Baya Benatsou<sup>b, d</sup>, Imad Bousaid<sup>c</sup>, mikael azoulay<sup>c</sup>, antoine verdon<sup>c</sup>, Francis Bidault<sup>b, d</sup>, Corinne Balleyguier<sup>b, d</sup>, Víctor Aubert<sup>a</sup>, Etienne Bendjebbar<sup>a</sup>, Carlos Maussion<sup>a</sup>, Nicolás Loiseau<sup>a</sup>, Benoit Schmauch<sup>a</sup>, Meriem Sefta<sup>a</sup>, Gilles Wainrib<sup>a</sup>, Tomas Clozel<sup>a</sup>, Samy Ammari<sup>b, d</sup>, Nathalie Lassau<sup>b, d</sup>

<sup>a</sup> Owkin Lab, Owkin, Inc., 10003, Nueva York, NY, EE. UU.

<sup>b</sup> Departamento de Imágenes, Gustave Roussy, Université Paris Saclay, 94805, Villejuif, Francia

<sup>c</sup> Dirección de Transformación Digital y Sistemas de Información, Gustave Roussy, Université Paris-Saclay, 94805, Villejuif, Francia

<sup>d</sup> Biomapas, UMR1281 INSERM, CEA, CNRS, Université Paris-Saclay, 94805, Villejuif, Francia

<sup>e</sup> Calypso Consulting, 75002, París, Francia

Recibido el 24 de mayo de 2022; recibido en forma revisada el 27 de junio de 2022; aceptado el 28 de junio de 2022

Disponible en línea el 16 de agosto de 2022

## PALABRAS CLAVE

Inteligencia artificial;  
Imágenes;  
biomarcador;  
Pronóstico;  
tratamiento  
antiangiogénico;  
Tumor sólido

**Resumen Antecedentes:** La necesidad de desarrollar nuevos biomarcadores está aumentando con la aparición de muchas terapias dirigidas. Los algoritmos de inteligencia artificial (IA) se han mostrado muy prometedores en el campo de las imágenes médicas para construir modelos predictivos. Desarrollamos un modelo de pronóstico para pacientes con tumores sólidos utilizando IA en datos multimodales.

**Pacientes y métodos:** nuestro estudio retrospectivo incluyó exámenes de pacientes con siete tipos diferentes de cáncer realizados entre 2003 y 2017 en 17 hospitales diferentes. Los radiólogos anotaron todas las metástasis en las imágenes iniciales de tomografía computarizada (TC) y ultrasonido (US).

Las características de las imágenes se extrajeron mediante modelos de IA y se usaron junto con los metadatos de los pacientes y los tratamientos. Se ajustó una regresión de Cox para predecir el pronóstico. Se evaluó el desempeño

\* Autor para correspondencia:  
Dirección de correo electrónico: [kathryn.schutte@owkin.com](mailto:kathryn.schutte@owkin.com) (K. Schutte).

<https://doi.org/10.1016/j.ejca.2022.06.055> 0959-8049/

© 2022 El(las) Autor(es). Publicado por Elsevier Ltd. Este es un artículo de acceso abierto bajo la licencia CC BY-NC (<http://creativecommons.org/licenses/by-nc/4.0/>).

en un conjunto de prueba omitido con 1000 bootstraps.

Resultados: El modelo se construyó en 436 pacientes y se probó en 196 pacientes (edad media 59, IQR: 51 e6, 411 hombres de 616 pacientes). En total, se anotaron 1147 imágenes de EE. UU. con la delimitación de las lesiones, y 632 TC de tórax, abdomen y pelvis (un total de 301 975 cortes) se anotaron completamente con un total de 9516 lesiones. El modelo desarrollado alcanza un índice de concordancia medio de 0,71 (0,67e0,76, IC 95%). Utilizando la mediana del riesgo predicho como valor umbral, el modelo puede aislar significativamente (valor de P de la prueba de rango logarítmico < 0,001) a los pacientes de alto riesgo de los pacientes de bajo riesgo (mediana de SG de 11 y 31 meses respectivamente) con un cociente de riesgos instantáneos de 3,5 (2,4 e5,2, IC 95%).

Conclusión: la IA pudo extraer características pronósticas de los datos de imágenes y, junto con los datos clínicos, permite una estratificación precisa de los pronósticos de los pacientes.

## 1. Introducción

Los antiangiogénicos se usan ampliamente para tratar tumores sólidos y mejorar el resultado de los pacientes al bloquear la angiogénesis [1,2]. Se ha aprobado la prescripción de tratamientos antiangiogénicos para muchos tumores sólidos metastásicos como el cáncer de riñón, colorrectal, pulmón, hígado, GIST y mama avanzado, entre otros. Se están realizando ensayos clínicos sobre muchos otros tipos de cáncer.

Varios estudios preliminares sobre cohortes específicas de cáncer [3,4] han demostrado que las mediciones de la perfusión tumoral con contraste dinámico (DCE US) de la perfusión tumoral [5] en las primeras etapas del tratamiento con antiangiogénicos se asocian con criterios de valoración estándar. En 2014, el estudio STIC (Soutien aux Techniques Innovantes Couteuses) [6] confirmó en una gran cohorte multicéntrica prospectiva que el análisis de DCE-US permite detectar la respuesta temprana a los antiangiogénicos para tumores sólidos. El criterio recientemente descubierto se agregó luego a las Directrices internacionales y Recomendaciones de buenas prácticas clínicas para la ecografía con contraste [7].

Sin embargo, a pesar de los esfuerzos adicionales para encontrar asociaciones pronósticas en etapas más tempranas [8], no se encontró una asociación significativa entre los valores de los criterios iniciales y los criterios de valoración estándar, como la supervivencia libre de progresión (PFS) y la supervivencia general (OS).

Encontrar un marcador pronóstico basal tendría un gran impacto en la planificación del tratamiento al identificar pacientes con alto y bajo riesgo de supervivencia bajo tratamiento antiangiogénico.

El uso de la inteligencia artificial (IA) en el campo de las imágenes médicas se ha mostrado recientemente muy prometedor [9,10]. Schmauch et al. [11] utilizaron algoritmos de aprendizaje profundo en imágenes de ultrasonido para detectar y caracterizar con precisión las lesiones hepáticas focales. De manera similar, P. Blanc-Durand et al. demostraron que los métodos de aprendizaje profundo pueden segmentar con precisión la masa corporal muscular a partir de imágenes de tomografía computarizada (TC) para cuantificar la sarcopenia [12]. Más recientemente, Lassau et al. [13] hizo uso de la capacidad de los algoritmos de IA para agregar información contenida en múltiples modalidades para predecir la gravedad de los pacientes con COVID-19.

Sin embargo, la calidad y la cantidad de conjuntos de datos disponibles en el campo de la medicina representan un obstáculo importante para el desarrollo de algoritmos robustos.

En el estudio multicéntrico retrospectivo actual, recopilamos y anotamos una gran cohorte de imágenes de TC y ultrasonido (US) y aprovechamos los algoritmos de aprendizaje profundo para desarrollar un modelo multimodal predictivo de respuesta a los antiangiogénicos para tumores sólidos.

## 2. Materiales y métodos

### 2.1. Recopilación de datos

Nuestro estudio retrospectivo utiliza una cohorte recolectada desde 2003 hasta 2017, que incluye pacientes con tumores sólidos tratados con antiangiogénicos durante diferentes ensayos clínicos de múltiples centros que tuvieron seguimiento por DCE-US [14].

Las imágenes de referencia se tomaron dentro de los 60 días anteriores al inicio del tratamiento y 7 días después del inicio del tratamiento. Se recopilaron metadatos de los pacientes y del tratamiento, así como datos de seguimiento, incluida la SLP y la SG.

Todos los pacientes procuraron su consentimiento escrito. El estudio fue aprobado por el comité de ética de nuestra institución y declarado a la Comisión Francesa Nationale Informatique et Liberté' (CNIL MR-004).

Todos los pacientes consistieron en una o dos imágenes de EE. UU. (ortogonales) centradas en una lesión diana y una tomografía computarizada de tórax, abdomen y pelvis. Radiólogos expertos dibujaron una delineación de la lesión visible en cada imagen de ultrasonido y anotaron todas las lesiones visibles en cada TC con un cuadro delimitador.

Tanto en las imágenes de TC como en las de EE. UU., las características de la IA se extrajeron de las regiones de las lesiones.

Finalmente, se extrajeron características adicionales relacionadas con la carga tumoral de las anotaciones de todas las lesiones en la TC: el número de lesiones y el volumen de la carga tumoral en el hígado, los pulmones y otros lugares se aproximaron mediante un elipsoide inscrito en los cuadros delimitadores.

Las características clínicas utilizadas en este estudio incluyeron metadatos de los pacientes (sexo, edad, peso), tipo de cáncer y tratamiento

metadatos (línea de tratamiento, molécula de tratamiento y recuento de moléculas).

Más detalles sobre la anotación y el preprocesamiento de datos se puede encontrar en el material complementario.

2.2. Modelado

Todos los modelos de pronóstico son modelos de regresión de Cox, implementados en el paquete de python Lifelines [15].

Las características de TC se agregaron a través de las lesiones utilizando la combinación máxima para obtener un solo conjunto de características (2048 características de IA). Las funciones de IA se normalizaron mediante escalado estándar y se utilizó el análisis de componentes principales (PCA) para obtener 10 funciones por modalidad. Las características de entrada del modelo multimodal consisten en la concatenación de solo cinco componentes principales por modalidad de imagen.

2.3. análisis estadístico

Un conjunto de prueba de 196 pacientes obtenidos de una división de prueba de tren estratificado en pacientes y tipos de cáncer se dejó fuera del estudio con fines de validación.

Los modelos se compararon utilizando el índice de concordancia (índice c) en 1000 conjuntos de 196 pacientes extraídos del conjunto de prueba con reemplazo (método de arranque). Se considera que el rendimiento de un modelo es significativamente

superior a otro modelo si es superior en más del 95% de las 1000 comparaciones. Para evaluar el rendimiento de los modelos para estratificar a los pacientes, utilizamos el valor de riesgo medio predicho como umbral. Los cocientes de riesgos instantáneos (HR) se compararon utilizando el mismo método utilizado para las comparaciones del índice c. Finalmente, se realizó una prueba de rango logarítmico en el conjunto de prueba y se dibujaron las curvas de KaplanMeier.

Se realizó un análisis de subgrupos en los dos grupos de pacientes de alto y bajo riesgo identificados por el modelo.

Las variables continuas para cada subgrupo se compararon mediante una prueba T y las variables binarias se compararon mediante una prueba Z de proporciones. Los valores de p se corrigieron por el número de pruebas y se consideraron significativos si  $<0,05$ .

3. Resultados

3.1. Descripción de la cohorte

Este estudio incluyó un total de 632 pacientes de la cohorte inicial de 1034 pacientes. Se realizó una primera selección sobre el tipo de cáncer para incluir solo los tipos de cáncer suficientemente representados: carcinoma hepatocelular primario (CHC), carcinoma metastático de células renales (RCC), carcinoma colorrectal (CRC), tumor del estroma gastrointestinal (GIST), melanoma, cáncer de mama, cáncer o sarcoma, que resultó en la

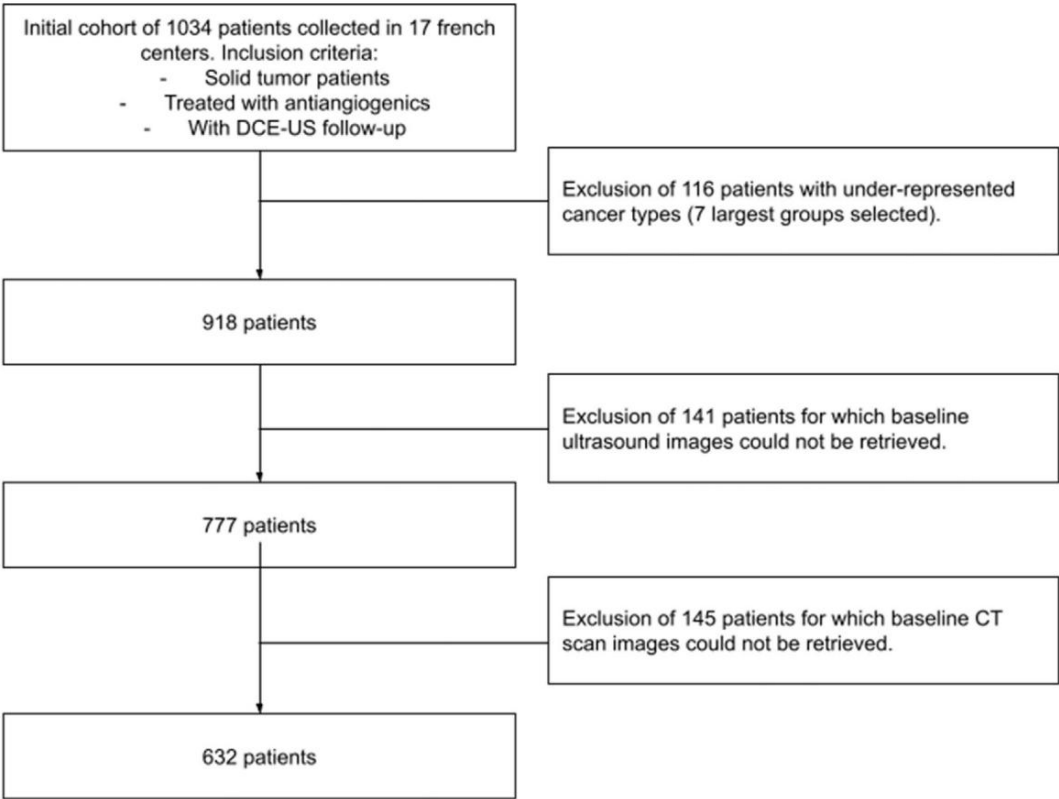


Figura 1. Diagrama de flujo de los criterios de inclusión y exclusión del estudio. Los siete tipos de cáncer más representados se seleccionaron de la cohorte original: carcinoma hepatocelular, carcinoma de células renales, carcinoma colorrectal, tumor del estroma gastrointestinal, melanoma, cáncer de mama y sarcoma. Se seleccionaron los pacientes para los que se tomaron imágenes de referencia dentro de los 60 días antes del inicio del tratamiento y 7 días después del inicio del tratamiento.

exclusión de 116 pacientes. Luego, excluimos a 286 pacientes para los que no se pudieron recuperar imágenes basales de US o CT. Los criterios de inclusión y exclusión se describen en el diagrama de flujo de la [figura 1](#).

La mediana de SLP de la cohorte es de 9 meses y la mediana de SG es de 14 meses. Durante el período observado, ocurrieron 480 (75,9% de los pacientes) eventos de muerte. Las características clínicas de los pacientes se describen en la [tabla 1](#): 411 hombres y una mediana de edad de 59 años (RIC: 51-67).

El proceso de anotación resultó en 632 pacientes completamente anotados; Se anotaron 1147 imágenes de EE. UU. con delineación de lesiones, y 632 TC (301 975 cortes) se anotaron completamente con hasta 150 lesiones por paciente para un total de 9516 lesiones anotadas.

### 3.2. Un modelo de pronóstico multimodal

Nuestro modelo de pronóstico multimodal (PULS-AI) se ajustó a la concatenación de cuatro conjuntos de variables: variables clínicas, características de IA extraídas de la lesión objetivo de EE. UU., características de IA extraídas de todas las lesiones en imágenes de TC y características de carga tumoral extraídas de anotaciones de TC ([Figura 2](#)). Este modelo alcanza un índice c promedio de 0.71 (0.67e0.76, 95% IC). El modelo desarrollado tiene un HR de 3,5 (2,4e5,2, IC del 95 %) y el análisis de KaplanMeier, como se muestra en la [Fig. 2](#), demuestra la capacidad de este modelo para estratificar significativamente (valor de P de prueba de rangos logarítmicos < 0,001) pacientes en un grupo de alto riesgo (mediana de SG de 11 meses) y un grupo de bajo riesgo (mediana de SG de 31 meses).

### 3.3. Análisis de subgrupos de alto y bajo riesgo

Se realizó un análisis de subgrupos en los pacientes de alto riesgo y los pacientes de bajo riesgo según lo predicho por el modelo PULS AI. Las variables que difieren significativamente entre los dos subgrupos se describen en la [Tabla 2](#). Los pacientes en el grupo de alto riesgo tenían un peso significativamente menor (valor de P de la prueba T <0,001) (mediana de 70 kg, IQR: 60-72) que los pacientes en el grupo de bajo riesgo. grupo de riesgo (mediana 70 kg, IQR: 70e82). El número de lesiones también difirió significativamente, con una mediana de 13 lesiones (IQR: 7e25) en el grupo de alto riesgo y siete lesiones (IQR: 3e13) en el grupo de bajo riesgo (T test P valor <0,001). Asimismo, el volumen tumoral en hígado y pulmón fue mayor en el grupo de alto riesgo que en el de bajo riesgo: una mediana de 124 cm<sup>3</sup> (RIC: 4e529) frente a 15 cm<sup>3</sup> (RIC: 0e109) de tumor hepático (T -test P valor Z 0,01) y una mediana de 1 cm<sup>3</sup> (IQR: 0e33) frente a 0 cm<sup>3</sup> (IQR: 0e5) de tumor de pulmón (T-test P valor Z 0,02).

Con respecto a las características de las imágenes de IA, el primer y el tercer componente principal de las características de la TC y el primer componente principal de las características de los EE. UU. difirieron significativamente entre los dos grupos (valores de P de la prueba T Z 0,01). Como era de esperar, la distribución de los tipos de cáncer y los tratamientos diferían significativamente entre los dos subgrupos, como se describe en la [Tabla 2](#).

Tabla 1

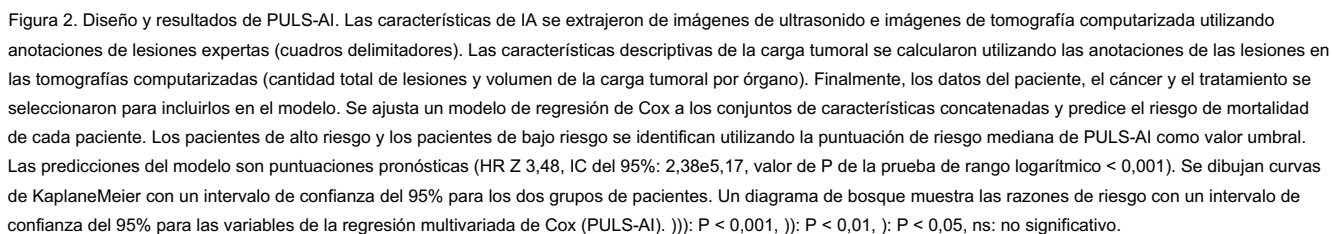
Descripción de la cohorte y asociación entre variables y supervivencia global.

Variable	N (%) Total Z 632	HORA	Asociación del valor de p con la supervivencia
Mujeres 221 (34,9)		1.1 (0.9e1.3) y 1.0	
Edad, mediana (RIC), años 59 (51e67)		(0.9e1.1) y 0.8	
Peso, mediana 71 (60e82)		(0.8e0.9) <0.001	
(IQR), kg mv: 183 Línea de			
tratamiento 1: 305 (48,3) 2: 167 (26,3) 3: 72		1.3 (1.2e1.4) <0.001	
(11,5) 4e10: 87			
(13,7) mv: 1			
1: 507			
(80,0) 2: 98			
Moléculas de tratamiento	(15,8) 3 : 24	1.0 (1.0e1.1) y	
contar	(3,8) 4: 3 (0,5)		
grupo de tratamiento			
TKI VEGFR	354 (56,2)	1,5 (1,2e1,7) <0,001	0,8
Mab anti-VEGFR	151 (23,8)	(0,7e1,0) y 0,3	
c-KIT	59 (9,3)	(0,2e0,5) <0,001	2,0 (1,5e2,7)
Otros	68 (10,7)	<0,001	
Cáncer			
RCC	253 (39,9)	0.9 (0.7e1.1) y 1.2	
CDN	97 (15,3)	(0.9e1.5) y 1.9	
CHC	93 (14,7)	(1.5e2.4) <0.001	0.4
ESENCIA	70 (11,0)	(0.3e0.5) <0.001	2.3 (1.7e3.2)
Melanoma	52 (8,5)	<0.001	0.8 (0.5e1.2) mi 1,6
Mama	37 (5,8)	(1,0e2,4) 0,04	
Sarcoma	30 (4,7)		

Las variables se describen como mediana (RIC) o como N (%). Las asociaciones con OS se evalúan con una regresión de Cox y se notifican con cocientes de riesgos instantáneos (HR) y valores de P cuando son significativos (<0,05). Para las variables continuas, los HR se calculan para un aumento de una desviación estándar de la variable continua. mv: valores faltantes, y: años, kg: kilogramos.

### 3.4. Estudio de ablación: evaluación de US y CT AI presenta valor predictivo

Como las lesiones diana se evaluaron en imágenes de ecografía y TC, queríamos evaluar la señal contenida en ambas imágenes por separado e investigar si ambas imágenes contienen información complementaria. Con ese fin, ajustamos un primer modelo (AI-US-Target) en las características de IA extraídas de las imágenes de EE. UU. de las lesiones objetivo, un segundo modelo (AI-CT-Target) en las características de IA extraídas de las imágenes de TC de las lesiones objetivo (excluyendo todas las demás lesiones) y finalmente un tercer modelo (AI-CTUS-Target) sobre la combinación de características de IA extraídas de ambos tipos de imágenes. Como se describe en la [Tabla 3](#), el modelo AI-CT-Target no funciona significativamente mejor que un predictor aleatorio, con un índice c de 0,53 (0,48e0,59, IC del 95 %). El análisis de KaplanMeier muestra que AI US-Target puede identificar significativamente a pacientes de alto riesgo y pacientes de bajo riesgo con un valor P de prueba de rango logarítmico de 0.01 y un HR de 1.7 (1.2e2.4). Los resultados dados muestran que las imágenes de US y CT no contienen información complementaria como el rendimiento obtenido al combinar



Para ello, ajustamos un modelo solo a las características de la carga tumoral (Cox-TB). El modelo Cox-TB alcanza un índice c de 0,63 (0,58e0,68, IC del 95%) y es capaz de estratificar significativamente a los pacientes en dos grupos de riesgo (valor de P de la prueba de rango logarítmico  $< 0,001$ ) con un HR de 2,32 (1,64e3 .36, IC del 95 %) como se muestra en la [Tabla 3](#). Curiosamente, el rendimiento de Cox-TB es similar al rendimiento de AI-CTp, lo que puede indicar que la información más valiosa derivada de las imágenes de TC se encuentra en la evaluación de

A continuación, queríamos medir el poder predictivo de las características de la carga tumoral extraídas de las imágenes.



Tabla 2  
Análisis de subgrupos de los pacientes de alto y bajo riesgo definidos por el modelo PULS-AI.

Característica	prueba t valor p	Grupo de alto riesgo mediana (RIC)	Mediana del grupo de bajo riesgo (RIC)
AI CT 0	<0,001	8,4 (12,3e32,2) 1,8	11,1 (30,0e10,7) 1,0
AI TC 2	<0,001	(6,9e3,3) 3,1	(4,2e6,7) 5,5
IA EE. UU. 0	0,01	(12,2e10,9) 1 13	(12,7e3,4) 70
Peso (kg)	<0,001		(70e82) 6
Cuenta de lesiones	<0,001	(7e25) 1	(3e13) 0
Volumen pulmonar (cm3)	0,02	(0e33) 124	(0e5) 15
Volumen hepático (cm3)	0,01	(4e529)	(0e109)
Característica	Proporción Z-test P valor	Alto riesgo (%)	Riesgo bajo (%)
Tratamiento TKI-VEGFR	0,01	63,3	48,7
Tratamiento C-KIT	<0,001	1,0	17,7
Tratamiento mab Anti-VEGFR	<0,001	15,8	32,0
CHC del cáncer	<0,001	25,3	4,1
Cáncer CCR	0,02	20,3	10,4
Cáncer GIST	<0,001	1,6	20,6
CCR de cáncer	<0,001	27,5	52,5
Sarcoma de cáncer	0,01	7,9	1,6

La tabla describe variables que son estadísticamente diferentes entre los dos grupos. Las variables continuas se probaron mediante una prueba T y están descritas por la mediana y el IQR. Las variables binarias se probaron mediante una prueba Z de proporción y se describen por el porcentaje correspondiente de pacientes dentro de cada grupo. Los valores de p se corrigieron para pruebas múltiples y se consideraron significativos si < 0,05.

3.6. Estudio de ablación: comparación de modelos monomodales frente a modelos multimodales

Finalmente, queríamos comparar el valor pronóstico de los datos de imagen y los datos clínicos y evaluar su complementariedad. Con ese fin, comparamos tres modelos: un modelo ajustado a las características de imágenes que incluyen características de IA extraídas de imágenes de EE. UU. de las lesiones objetivo, características de IA extraídas de imágenes de TC de todas las lesiones y características de carga tumoral artesanal (AI-USCT+) , un modelo ajustado a las variables clínicas solas (Clin-Cox) y el modelo de pronóstico multimodal (PULS-AI) presentado anteriormente. Los tres modelos mostrados en la Fig. 3 son pronósticos con un HR de 2.1 (1.5e3.1, 95% IC), 3.0 (2.1e4.4, 95% IC) y 3.5 (2.4e5.2, 95% IC) respectivamente. tivamente, y valores de P significativos de la prueba de rango logarítmico (<0,001). Sin embargo, PULS-AI supera significativamente (valores P < 0,001) AI-USCTp y Clin-Cox en términos de índice c, con 0,71 (0,67-0,76, IC del 95 %) puntos en comparación con 0,61

(0.56e0.66, 95% IC) y 0.68 (0.63e0.72, 95% IC) puntos respectivamente.

Las comparaciones estadísticas de los modelos se presentan en [Tabla complementaria 2](#).

4. Discusión

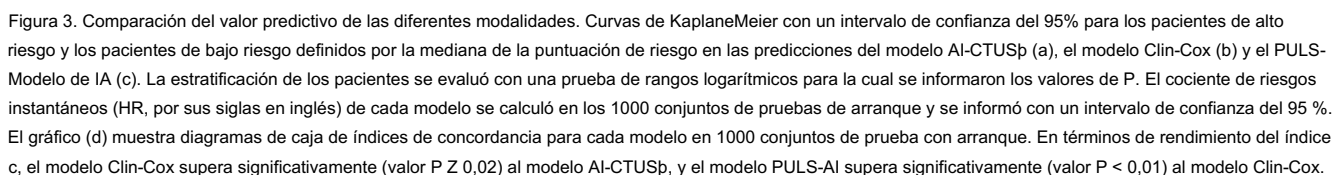
Las decisiones de tratamiento son pasos cruciales en el camino terapéutico de un paciente y tienen un gran impacto en el pronóstico del paciente. A medida que un paciente avanza en líneas sucesivas de tratamientos, las opciones y pautas terapéuticas son cada vez menos, lo que deja al personal médico con su propia experiencia para decidir sobre el tratamiento posterior. El desarrollo de marcadores pronósticos fiables y reproducibles es clave para mejorar los resultados de los pacientes.

A los efectos de este estudio, construimos una cohorte única de exámenes de radiología inicial anotados extensamente y manualmente de 632 pacientes recolectados de 17 centros franceses diferentes.

Tabla 3  
Estudio de ablación.

Modelo	Índice C (95 % IC) HR (95 % IC)		Prueba de rango logarítmico valor p	Funciones			Características de la carga tumoral en la TC
				Características clínicas	de EE. UU.	Características de IA CT	
Puls-AI	0,71 (0,67e0,76)	0,57	3.5 (2.4e5.2) 1.7	<0,001	X	X	X
Objetivo AI-EE. UU.	(0,52e0,62)	0,53	(1.2e2.4) 1.5	0,01	X		
Objetivo AI-CT	(0,48e0,59)	0,56	(1.1e2.0) 1.7	0,06			
Objetivo AI-US-CT	(0,51e0,62)	0,61	(1.2e2.4) 1.9	0,01		X	solo objetivo
AI-CTD	(0,56e0,66)	0,63	(1.3e2.7) 2.3	<0,001			X
TB-Cox	(0,58e0,68)	0,61	(1.6e3.4) 2.1	<0,001			X
AI-US-CTD	(0,56e0,67)	0,68	(1.5e3.1) 3.0	<0,001	X	X	X
Clin-Cox	(0.63e0.72)		(2.1e4.4)	<0,001	X		

El rendimiento de cada modelo se informa en términos de índice C y HR con un intervalo de confianza del 95 %. Además, se informan los valores de P de las pruebas de rango logarítmico. Para cada modelo, se especifican las características de entrada utilizadas.



Realizamos un análisis de subgrupos para identificar las variables que diferían significativamente entre los pacientes de alto riesgo y los pacientes de bajo riesgo. Los datos de imágenes revelan diferencias significativas en tres características extraídas por IA, así como en características de carga tumoral con un mayor número de metástasis y

Finalmente, hemos demostrado que un modelo en la combinación de datos de imágenes y datos clínicos produce resultados significativamente mejores que la construcción de modelos en cada modalidad por separado, demostrando la complementariedad

de las diferentes modalidades de datos. Este resultado respalda la necesidad de construir biomarcadores compuestos a partir de múltiples fuentes de datos.

Nuestro estudio tiene varias limitaciones. Primero, no pudimos constituir una cohorte de validación externa para validar nuestros resultados. Aunque el tamaño y el carácter multicéntrico de la cohorte deberían permitir resultados sólidos y mitigar el riesgo de descubrimiento falso que a menudo se discute con respecto a los biomarcadores de imágenes cuantitativas [16e18], el modelo desarrollado requiere una validación adicional en una cohorte externa antes de cualquier adopción clínica, de acuerdo con la hoja de ruta de biomarcadores de imágenes para la investigación del cáncer [19]. En segundo lugar, los datos clínicos recopilados no nos permitieron comparar el rendimiento de PULS-AI con las puntuaciones de pronóstico existentes, como el IMDC para CCR metastásico [20] o la puntuación de pronóstico de Glasgow para CCR [21] y CHC [22].

Finalmente, la aplicabilidad de PULS-AI es limitada hoy en día, ya que requiere la anotación de todas las lesiones visibles en las tomografías computarizadas de los pacientes, lo que requiere mucho tiempo y, por lo tanto, no se requiere en la rutina. Sin embargo, hemos demostrado el importante beneficio de evaluar la carga tumoral completa, y creemos que esta tarea se puede automatizar fácilmente y pronto se automatizará mediante un algoritmo de IA.

Otras investigaciones pueden centrarse en la interpretabilidad médica del modelo recientemente desarrollado. Recientemente se han desarrollado muchas técnicas para permitir una mejor comprensión de las predicciones basadas en el aprendizaje profundo. Una de estas técnicas, el método GradCam, permite visualizar áreas en una imagen dada que impactan en la predicción. Un enfoque más reciente utiliza modelos generativos [23] para mostrar cómo las características específicas de una imagen deben cambiar para producir diferentes predicciones [24].

Los resultados de este estudio de pronóstico muestran que los algoritmos de IA pueden extraer información relevante de las imágenes de radiología y agregar datos de múltiples modalidades para crear poderosas herramientas de pronóstico. Dichas herramientas podrían brindar asistencia a los médicos oncólogos en la toma de decisiones terapéuticas.

#### Soporte financiero

Este estudio fue apoyado por una subvención de la Región Ile De France.

#### Contribuciones de autor

Conceptualización: Kathryn Schutte, Paul Jehanno, Samy Ammari, Nathalie Lassau, Victor Aubert, Etienne Bendjebbar, Charles Maussion, Meriem Sefta.

Metodología: Kathryn Schutte, Paul Jehanno, Samy Ammari, Nathalie Lassau, Charles Maussion, Fabien Brulport, Jean-Baptiste Schiratti, Ridouane Ghermi.

Software e investigaciones: Kathryn Schutte, Paul Jehanno, Fabien Brulport, Jean-Baptiste Schiratti, Ridouane Ghermi.

Análisis formal: Kathryn Schutte, Paul Jehanno, Fabien Brulport, Jean-Baptiste Schiratti, Ridouane Ghermi, Nicolas Loiseau.

Recursos: Sana Harguem-Zayani, Alexandre Jaeger, Talal Alamri, Raphaël Naccache, Leila Haddag-Miliani, Teresa Orsi, Jean-Philippe Lamarque, Isaline Hoferer, Littisha Lawrance, Baya Benatsou, Imad Bousaid, Mikael Azoulay, Antoine Verdon, François Bidault, Corinne Balleyguier, Gilles Wainrib, Thomas Clozel, Samy Ammari, Nathalie Lassau.

Curación de datos: Sana Harguem-Zayani, Alexandre Jaeger, Talal Alamri, Raphaël Naccache, Leila Haddag Miliani, Teresa Orsi, Jean-Philippe Lamarque, Isaline Hoferer, Littisha Lawrance, Baya Benatsou, Imad Bousaid, Mikael Azoulay, Antoine Verdon, François Bidault, Corinne Balleyguier, Gilles Wainrib, Thomas Clozel, Samy Ammari, Nathalie Lassau.

Redacción - borrador original: Kathryn Schutte, Fabien Brulport, Jean-Baptiste Schiratti, Ridouane Ghermi, Samy Ammari, Nathalie Lassau.

Redacción - revisión y edición: Kathryn Schutte, Fabien Brulport, Jean-Baptiste Schiratti, Ridouane Ghermi, Nicolas Loiseau, Samy Ammari, Nathalie Lassau, Charles Maussion, Benoit Schmauch.

Supervisión y administración del proyecto: Kathryn Schutte, Nathalie Lassau, Charles Maussion, Gilles Wainrib, Thomas Clozel.

#### Declaración de conflicto de intereses

Los autores declaran los siguientes intereses financieros/ relaciones personales, que pueden considerarse como posibles intereses en competencia: Los autores declaran los siguientes intereses en competencia: KS, FB, JBS, RG, PJ, AJ, VA, EB, CM, NL, BS, MS son empleados de Owkin. GW y TC son cofundadores de Owkin. NL informa que será ponente en Jazz Pharmaceuticals. NL informa de una subvención de Guerbet.

#### Apéndice A. Datos complementarios

Los datos complementarios a este artículo se pueden encontrar en línea en <https://doi.org/10.1016/j.ejca.2022.06.055>.

#### Referencias

- [1] Demetri GD, Reichardt P, Kang YK, et al. Eficacia y seguridad de regorafenib para tumores del estroma gastrointestinal avanzados después del fracaso de imatinib y sunitinib (GRID): un ensayo de fase 3 internacional, multicéntrico, aleatorizado, controlado con placebo. *Lancet Lond Engl* 2013;381(9863):295e302. [https://doi.org/10.1016/S0140-6736\(12\)61857-1](https://doi.org/10.1016/S0140-6736(12)61857-1).
- [2] Escudier B, Eisen T, Stadler WM, et al. Sorafenib en el carcinoma avanzado de células renales de células claras. *N Engl J Med* 2007;356(2):125e34. <https://doi.org/10.1056/NEJMoa060655>.
- [3] Lassau N, Koscielny S, Albiges L, et al. Carcinoma metastásico de células renales tratado con sunitinib: evaluación temprana de la respuesta al tratamiento mediante ultrasonografía dinámica con contraste. *clin*



- Cancer Res Off J Am Assoc Cancer Res 2010;16(4):1216e25. <https://doi.org/10.1158/1078-0432.CCR-09-2175>.
- [4] Lassau N, Lamuraglia M, Vanel D, et al. Doppler US con software de perfusión e inyección de medio de contraste en la evaluación temprana de la perfusión aislada de extremidades de sarcomas de extremidades: estudio prospectivo de 49 casos. *Ann Oncol Off J Eur Soc Med Oncol* 2005; 16(7):1054e60. <https://doi.org/10.1093/annonc/mdi214>.
- [5] Dietrich CF, Averkiou MA, Correias JM, Lassau N, Leen E, Piscaglia F. Una introducción de EFSUMB al ultrasonido mejorado con contraste dinámico (DCE-US) para la cuantificación de la perfusión tumoral. *Ultraschall Med Stuttg Ger* 1980 2012;33(4):344e51. <https://doi.org/10.1055/s-0032-1313026>.
- [6] Lassau N, Bonastre J, Kind M, et al. Validación de la ecografía dinámica con contraste para predecir los resultados de la terapia antiangiogénica para tumores sólidos. *Invest Radiol* 2014;49(12): 794e800. <https://doi.org/10.1097/RLI.000000000000085>.
- [7] Sidhu PS, Cantisani V, Dietrich CF, et al. Las guías y recomendaciones de la EFSUMB para la práctica clínica del contraste ultrasonido mejorado (CEUS) en aplicaciones no hepáticas: actualización 2017 (versión larga). *Ultraschall Med Stuttg Alemania* 1980 2018; 39(2):e2e44. <https://doi.org/10.1055/a-0586-1107>.
- [8] Lassau N, Coiffier B, Kind M, et al. Selección de un biomarcador temprano para la normalización vascular mediante ultrasonografía dinámica mejorada con contraste para predecir los resultados de pacientes metastásicos tratados con bevacizumab. *Ann Oncol Off J Eur Soc Med Oncol* 2016; 27(10):1922e8. <https://doi.org/10.1093/annonc/mdw280>.
- [9] Oren O, Gersh BJ, Bhatt DL. Inteligencia artificial en imágenes médicas: cambio de datos patológicos radiográficos a puntos finales clínicamente significativos. *Lancet Digit Health* 2020;2(9): e486e8. [https://doi.org/10.1016/S2589-7500\(20\)30160-6](https://doi.org/10.1016/S2589-7500(20)30160-6).
- [10] Lee LIT, Kanthasamy S, Ayyalaraju RS, Ganatra R. El estado actual de la inteligencia artificial en imágenes médicas y medicina nuclear. *BJR|Abierto* 2019;1(1):20190037. <https://doi.org/10.1259/bjro.20190037>.
- [11] Schmauch B, Herent P, Jehanno P, et al. Diagnóstico de lesiones hepáticas focales por ecografía mediante deep learning. *Diagn Interv Imag* 2019;100(4):227e33. <https://doi.org/10.1016/j.diii.2019.02.009>.
- [12] Blanc-Durand P, Schiratti JB, Schutte K, et al. Segmentación de la musculatura abdominal y predicción de la superficie a partir de la TC mediante el aprendizaje profundo para la evaluación de la sarcopenia. *Diagnóstico Interv Imag* 2020; 101(12):789e94. <https://doi.org/10.1016/j.diii.2020.04.011>.
- [13] Lassau N, Ammari S, Chouzenoux E, et al. Integración del modelo de tomografía computarizada de aprendizaje profundo, variables biológicas y clínicas para predecir la gravedad de los pacientes con COVID-19. *Nat Commun* 2021;12(1):634. <https://doi.org/10.1038/s41467-020-20657-4>.
- [14] Lassau N, Chapotot L, Benatsou B, et al. Estandarización de la ecografía dinámica con contraste para la evaluación de terapias antiangiogénicas: el estudio multicéntrico francés Support for Innovative and Expensive Techniques. *Invest Radiol* 2012;47(12):711e6. <https://doi.org/10.1097/RLI.0b013e31826dc255>.
- [15] Davidson-Pilon C, Kalderstam J, Jacobson N, et al. CamDa vidsonPilon/lineas de vida: 0.26.0. <https://doi.org/10.5281/zenodo.2021>.
- [16] Fournier L, Costaridou L, Bidaut L, et al. Corrección para: incorporar la radiómica en los ensayos clínicos: consenso de expertos respaldado por la Sociedad Europea de Radiología sobre las consideraciones para los biomarcadores cuantitativos basados en datos en comparación con los basados en datos biológicos. *Eur Radiol* 2021;31(8):6408e9. <https://doi.org/10.1007/s00330-021-07721-3>.
- [17] Espinasse M, Pitre-Champagnat S, Charmettant B, et al. Desafíos del análisis de texturas por TC: influencia de los parámetros de adquisición y reconstrucción: una revisión exhaustiva. *Diagnóstico Basilea Suiza* 2020;10(5):E258. <https://doi.org/10.3390/diagnostics10050-258>.
- [18] Caramella C, Allorant A, Orlhac F, et al. ¿Podemos confiar en el cálculo de los índices de textura de las imágenes de TC? Un estudio fantasma. *Med Phys* 2018;45(4):1529e36. <https://doi.org/10.1002/mp.12809>.
- [19] O'Connor JPB, Aboagye EO, Adams JE, et al. Hoja de ruta de biomarcadores de imágenes para estudios de cáncer. *Nat Rev Clin Oncol* 2017; 14(3):169e86. <https://doi.org/10.1038/nrclinonc.2016.162>.
- [20] Factores pronósticos para la supervivencia general en pacientes con carcinoma metastásico de células renales tratados con agentes dirigidos contra el factor de crecimiento endotelial vascular: resultados de un estudio multicéntrico grande | Revista de Oncología Clínica nd <https://ascopubs.org/doi/10.1200/JCO.2008.21.4809> (consultado el 3 de noviembre de 2021).
- [21] Nozoe T, Matono R, Ijichi H, Ohga T, Ezaki T. La puntuación de pronóstico de Glasgow (GPS) puede ser un indicador útil para determinar el pronóstico de los pacientes con carcinoma colorrectal. *Int Surg* 2014;99(5): 512e7. <https://doi.org/10.9738/INTSURG-D-13-00118.1>.
- [22] Kinoshita A, Onoda H, Imai N, et al. La puntuación de pronóstico de Glasgow, una puntuación de pronóstico basada en la inflamación, predice la supervivencia en pacientes con carcinoma hepatocelular. *BMC Cáncer* 2013; 13(1):52. <https://doi.org/10.1186/1471-2407-13-52>.
- [23] Karras T, Laine S, Aila T. Una arquitectura generadora basada en estilos para redes antagonicas generativas. 2019. *ArXiv181204948 Cs Stat*.
- [24] Schutte K, Moindrot O, He'rent P, Schiratti JB, Je'gou S. Uso de StyleGAN para la interpretabilidad visual de modelos de aprendizaje profundo en imágenes médicas. 2021. *ArXiv210107563 CsEess*.