

Determinación del grado de agresividad tumoral mediante técnicas de inteligencia artificial aplicadas a imágenes de resonancia magnética

Facundo Segura, Pablo Segura, Florencio Segura

Centro Privado Ortopedia y Traumatología Segura, 2.^a Cátedra de Ortopedia y Traumatología, Hospital Tránsito Cáceres de Allende, Universidad Nacional de Córdoba, Córdoba, Argentina

RESUMEN

Objetivo: Determinar el grado de agresividad tumoral mediante técnicas de inteligencia artificial utilizando imágenes de resonancia magnética de sarcomas con grado histológico comprobado. **Materiales y Métodos:** Dos cohortes retrospectivas independientes de pacientes con sarcomas de partes blandas. Para cada paciente de las dos cohortes se adquirieron tres tipos de secuencias de imágenes como indican los protocolos clínicos: potenciadas en T1, en T2 con supresión grasa (T2FS) y STIR. A fin de desarrollar el modelo de inteligencia artificial, se utilizaron 134 imágenes, tanto las de alto grado como las de bajo grado, en T1 y T2 tomando la imagen más representativa del tumor en cualquier corte. Esto se traduce en 36 millones de píxeles que serán analizados por el programa Landing AI. **Resultados:** La precisión promedio del modelo fue del 84,3% y la sensibilidad, del 73,3%, con un umbral de confianza de 0,66, lo que demuestra inicialmente que se obtuvo un modelo de buena calidad para predecir con imágenes de resonancia magnética el grado de agresividad de un sarcoma de partes blandas antes de la biopsia. **Conclusiones:** Se presenta un enfoque novedoso para abordar un tipo de enfermedad infrecuente usando técnicas de inteligencia artificial para determinar el grado tumoral en imágenes de resonancia magnética. Según los resultados de nuestro modelo, se lo puede considerar como una segunda opinión experta al realizar los estudios por imágenes antes de la biopsia.

Palabras clave: Sarcoma de partes blandas; grado histológico; inteligencia artificial; aprendizaje automático; visión artificial.

Nivel de Evidencia: III

Measuring Tumor Aggressiveness Through Artificial Intelligence

ABSTRACT

Objective: To determine the degree of tumor aggressiveness by means of artificial intelligence techniques using magnetic resonance images of sarcomas with proven histological grade. **Materials and Methods:** Two independent cohorts of patients with soft tissue sarcomas (STS) were retrospectively collected. For each patient in the two cohorts, three types of imaging sequences were acquired as indicated by the clinical protocols: T1-weighted (T1), fat-suppressed T2-weighted (FST2) and STIR. For the development of the artificial intelligence model, 134 images were used, both high-grade and low-grade T1 and T2 images, taking the most representative image of the tumor at any slice. This translated into more than 36 million pixels that were analyzed by the Landing AI program. **Results:** To determine the degree of tumor aggressiveness by means of artificial intelligence techniques using magnetic resonance The model's average accuracy was 84.3%, and its sensitivity was 73.3%, with a confidence threshold of 0.66, indicating that a good quality model was generated for predicting the grade of aggressiveness of an STS prior to biopsy using MRI scans. **Conclusions:** A novel approach is presented to address a rare pathology using artificial intelligence techniques to determine the tumor grade based on nuclear magnetic resonance images. Based on the results of our model, it can be considered as a second expert opinion when performing imaging studies prior to biopsy.

Keywords: Soft tissue sarcoma; artificial intelligence; histological grade; deep learning; automated machine learning; machine vision.

Level of Evidence: III

Recibido el 14-3-2023. Aceptado luego de la evaluación el 1-9-2023 • Dr. FACUNDO SEGURA • facusegura@gmail.com  <https://orcid.org/0009-0000-7101-9145>

Cómo citar este artículo: Segura F, Segura P, Segura F. Determinación del grado de agresividad tumoral mediante técnicas de inteligencia artificial aplicadas a imágenes de resonancia magnética. *Rev Asoc Argent Ortop Traumatol* 2023;88(6):653-661. <https://doi.org/10.15417/issn.1852-7434.2023.88.6.1738>

INTRODUCCIÓN

Hace 60 años, John McCarthy definió originalmente a la inteligencia artificial (IA) como “la ciencia y la ingeniería para fabricar máquinas inteligentes”, y predijo que estas máquinas, algún día, serían capaces de hacer hazañas que anteriormente se pensaba que eran de exclusivo dominio de la inteligencia humana, como resolución avanzada de problemas.¹ Jerrold S. Maxmen, profesor de psiquiatría en la Universidad de Columbia, predijo que la IA traería consigo la “era posmédica” para el siglo XXI, y describió el cambio como “posible, inevitable y deseable”.²

La visión artificial (*computer vision*) es una función de la IA mediante la cual se pueden aplicar técnicas de aprendizaje automático y redes neuronales complejas para permitir que las computadoras capturen, analicen e interpreten información de imágenes clínicas, utilizando algoritmos que se entrenan para hacer clasificaciones y predicciones basadas en patrones que nos brindan los datos.³ En el campo de la Ortopedia y Traumatología, el uso de estas técnicas se está expandiendo rápidamente. La visión artificial ha progresado, de manera notable, en el análisis de imágenes médicas en una variedad de modalidades que se usan comúnmente en ortopedia, incluidas radiografías, tomografía computarizada y resonancia magnética (RM). Para ello es de gran ayuda el aprendizaje automático automatizado (AutoML). En este caso, nos referimos al proceso de automatizar una acción aplicando el aprendizaje automático común para resolver problemas reales.⁴ El alto grado de automatización en el AutoML permite que aquellos que no son expertos en IA puedan hacer uso de modelos de aprendizaje automático sin que sea necesario contar con una avanzada experiencia en dicho campo.⁵

Los sarcomas de partes blandas (SPB) representan una enfermedad rara y heterogénea que comprende el 1% de todos los tumores malignos en adultos. En 2019, se diagnosticó SPB a unas 13.500 personas en los Estados Unidos.⁶ Abarcan, al menos, 100 subtipos histológicos y moleculares diferentes, y cada subtipo muestra un comportamiento clínico variable.⁷ Los casos infrecuentes, junto con la heterogeneidad de la enfermedad y su tratamiento complejo resaltan la importancia del manejo interdisciplinario de estas lesiones que involucra a equipos especializados en diagnóstico por imágenes, anatomía patológica, radioterapia, oncólogos y cirujanos ortopedistas con experiencia en el manejo de SPB.

Ante un paciente con SPB, la toma de decisiones en cuanto al tratamiento óptimo del tumor depende, en gran medida, del tamaño y la ubicación del SPB, y del resultado del estudio de anatomía patológica que evalúa la agresividad del tumor.

En la actualidad, el sistema que utilizamos para clasificar el grado histológico del SPB es el propuesto por la *French Fédération Nationale des Centres de Lutte Contre le Cancer* (FNCLCC) de Francia. El sistema de la FNCLCC evalúa únicamente tres parámetros: la diferenciación tumoral, el índice mitótico y la cantidad de necrosis tumoral (Tabla 1).⁸

El grado del tumor constituye el factor más importante para adoptar la mejor decisión terapéutica. La biopsia de la lesión es fundamental para determinar el grado y establecer el diagnóstico de certeza. Mediante el análisis de imágenes de RM y utilizando las técnicas de IA basadas en el aprendizaje profundo, es posible plantear una forma alternativa de determinar el grado de agresividad de los SPB.

El objetivo de este estudio fue clasificar, de forma no invasiva, los SPB en: de bajo grado (G1) y alto grado (G2/G3) utilizando imágenes de RM de sarcomas con grado histológico comprobado mediante el AutoML.

MATERIALES Y MÉTODOS

Se formaron dos cohortes retrospectivas independientes de pacientes. Los criterios de inclusión fueron: SPB con grado histológico comprobado con información disponible de la clasificación de tumores de la FNCLCC. Los criterios de exclusión fueron: artefactos de RM, radioterapia previa, sarcoma óseo primario o sarcoma de Ewing.

La primera cohorte estaba formada por el conjunto de datos del *The Cancer Image Archive* (TCIA) que se encuentra disponible públicamente. Este conjunto de datos contiene imágenes de tomografías computarizada/por emisión de positrones con fluor-18-fluorodesoxiglucosa y RM (potenciada en T1, potenciada en T2 con supresión grasa [T2FS]) de 51 pacientes con sarcomas de tejido blando histológicamente comprobados en las extremidades. Todos se habían sometido a exploraciones por tomografía computarizada/por emisión de positrones con fluor-18-fluorodesoxiglucosa y RM antes del tratamiento, entre noviembre de 2004 y noviembre de 2011. Además, posee los datos demográficos de todos los pacientes analizados.

La segunda cohorte incluyó imágenes de la base de datos del Centro Privado de Ortopedia y Traumatología Segura. En este caso, se tomaron las imágenes de 22 pacientes con SPB con grado histológico comprobado, operados entre 2020 y 2022.

Tabla 1. Grado de malignidad de los sarcomas de partes blandas

Parámetro	Criterio		
	Diferenciación tumoral		
1 punto	El sarcoma se asemeja con proximidad al tejido mesenquimal adulto normal		
2 puntos	El tipo histológico del sarcoma está definido		
3 puntos	Sarcomas embrionarios o desdiferenciados, sarcoma sinovial y del tipo incierto		
	Índice mitótico		
1 punto	0-9 por 10 campos de gran aumento		
2 puntos	10-19 por 10 campos de gran aumento		
3 puntos	>20 por 10 campos de gran aumento		
	Necrosis tumoral		
1 punto	Sin necrosis		
2 puntos	≤50%		
3 puntos	≥50%		
Grado tumoral			
Grado 1: Bajo	Puntos totales	2, 3	
Grado 2: Intermedio	Puntos totales	4, 5	
Grado 3: Alto	Puntos totales	6, 7, 8	

Puntaje del 1 al 3 según el grado de diferenciación tumoral, el índice mitótico y el porcentaje de necrosis tumoral. Un valor mínimo total <3 indica bajo grado, entre 3 y 6, grado intermedio y >6, alto grado.

En cada paciente de las dos cohortes, se adquirieron tres tipos de secuencias de imágenes utilizadas como indican los protocolos clínicos: potenciada en T1, potenciada en T2FS y STIR que permite suprimir la señal de determinados elementos o tejidos de forma específica (grasa, agua). Las secuencias T1 se adquirieron en el plano axial, mientras que las secuencias T2FS y STIR se tomaron en diferentes orientaciones (axial, sagital y coronal). Además, el grosor del corte fue de 5,5 mm para la secuencia T1 y de 5 mm para la T2FS. La resolución en el plano fue de 0,63 mm²; 0,74 mm² y 0,86 mm² para las secuencias T2FS, T1 y STIR, respectivamente.⁹ De cada estudio de RM completo del paciente, se extrajo la imagen más representativa en T1 y T2, esta fue la imagen donde se observó el mayor diámetro del tumor.

En total, se obtuvieron 37 imágenes de sarcomas de alto grado en la secuencia T1 y 37 imágenes en la secuencia T2. Para los sarcomas de grado intermedio se obtuvieron 15 imágenes en la secuencia T1 y 15 imágenes en la secuencia T2. Las imágenes de sarcomas de bajo grado fueron 20 tanto para la secuencia T1 como para la secuencia T2.

Las lesiones grado 2 son de pronóstico incierto y, por lo tanto, no son útiles para la toma de decisiones.¹⁰ Teniendo en cuenta que tanto los sarcomas de alto grado como los de grado intermedio tienen un desarrollo, un pronóstico y un tratamiento quirúrgico y posquirúrgico similares, a los fines de este proyecto, se unificaron los sarcomas de alto grado y grado intermedio en el mismo conjunto de datos (*dataset*), dejando por separado a los de bajo grado.¹¹

El análisis de los pacientes evaluados se realizó mediante Google Colab utilizando el lenguaje Python con las librerías de Numpy, Pandas, Matplotlib y Seaborn para graficar los resultados obtenidos. La implementación del modelo predictivo del proyecto se llevó a cabo con la herramienta Landing Lens, plataforma de IA del sitio web Landing.ai. Antes de la etapa de entrenamiento de los datos, se utilizaron técnicas de aumentación y transformación (*data augmentation*) para aumentar la cantidad de imágenes para analizar. En este caso, se modificó el tamaño de la imagen a una altura de 1000 y un ancho de 1000 píxeles. Como técnica de aumentación, se realizó un *flip* horizontal y vertical de 0,5 de probabilidad de aplicarlo a cada imagen.

Las lesiones en las imágenes de RM se clasificaron en alto y bajo grado manualmente, delimitando con rojo las lesiones etiquetadas como de alto grado y, en celeste, las de bajo grado.

Evaluación

En el contexto de una clasificación binaria, se describen cuáles son las principales métricas importantes por seguir para evaluar el rendimiento del modelo de aprendizaje automático.¹²

Matriz de confusión (*confussion matrix*): se utiliza para tener una visión más completa al evaluar el rendimiento de un modelo. Es una herramienta que permite la visualización del desempeño de un algoritmo, en este caso, el análisis de millones de píxeles que utiliza el modelo para aprender.

Precisión (*precision*): es la proporción de predicciones correctas con un conjunto de datos de prueba. Es la relación entre el número de predicciones correctas y el número total de ejemplos de entrada. En forma práctica, es el porcentaje de casos positivos detectados.

Sensibilidad (*recall*): es la proporción de casos positivos que fueron correctamente identificados por el algoritmo. Se representa como la fracción de verdaderos positivos. También se conoce como tasa de verdaderos positivos (*true positive rate*).

Precisión y sensibilidad por umbral de confianza: el modelo decidirá si la clasificación es de alto grado o bajo grado si el valor devuelto por el modelo supera un umbral de decisión. Si aumentamos este valor, estaremos aumentando la precisión e, inversamente, si lo disminuimos, aumentamos la sensibilidad.

RESULTADOS

El primer análisis se realizó sobre la distribución de la población y la enfermedad. El promedio de edad en ambos grupos era de 54.82 años (± 16.98) (rango 16-83). La distribución por sexo de la población era similar (sexo femenino 52,9%, sexo masculino 47,1% [estadística Z: 0,594, p 0,552]).

En la **Figura 1**, se detallan los casos analizados según el resultado anatomopatológico.

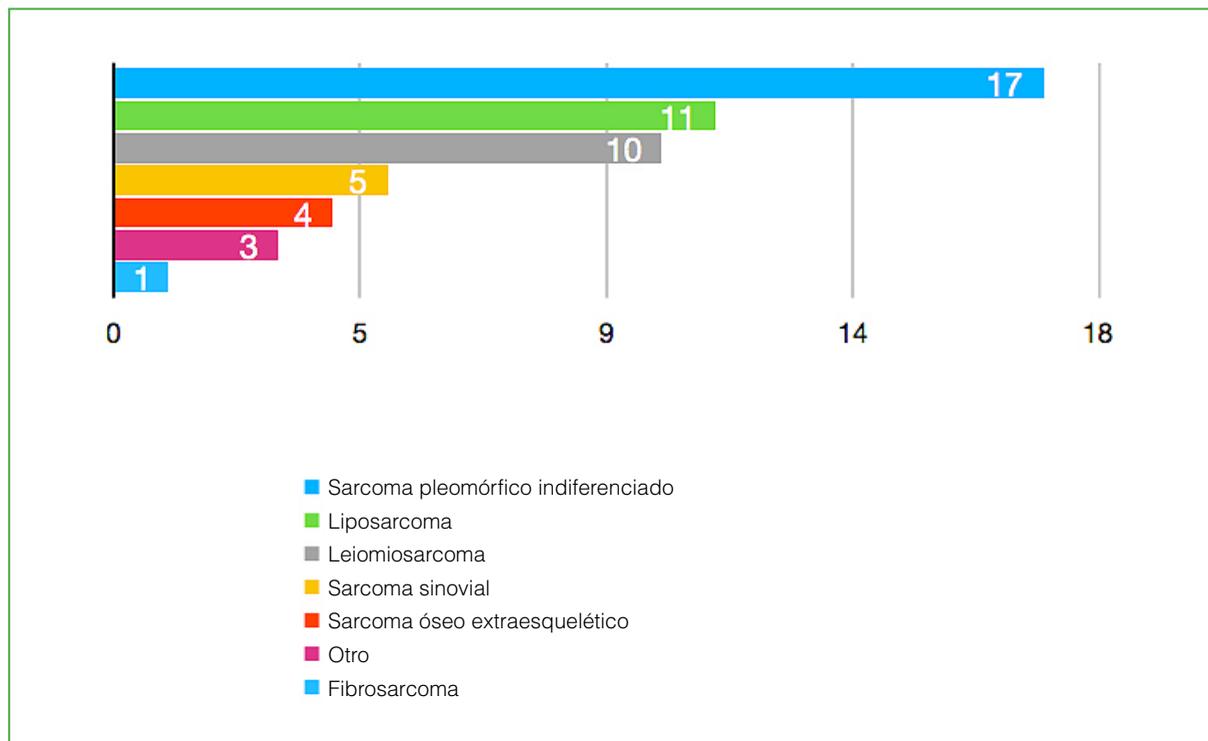


Figura 1. Distribución de los casos según el resultado del estudio anatomopatológico.

En las Figuras 2 y 3, se observan la distribución del grado de agresividad de cada tipo de tumor, y también la distribución del grado según su ubicación.

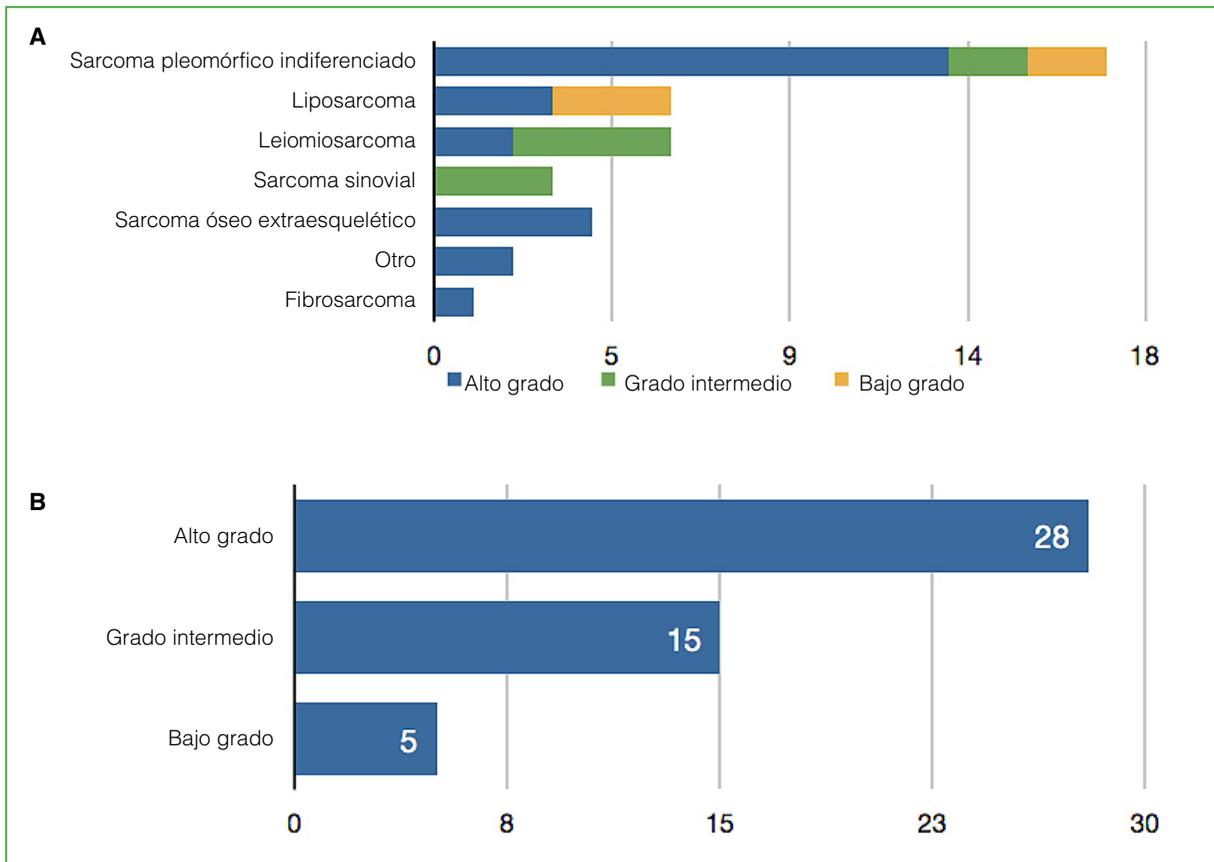


Figura 2. Distribución de los sarcomas de partes blandas según el grado de agresividad.

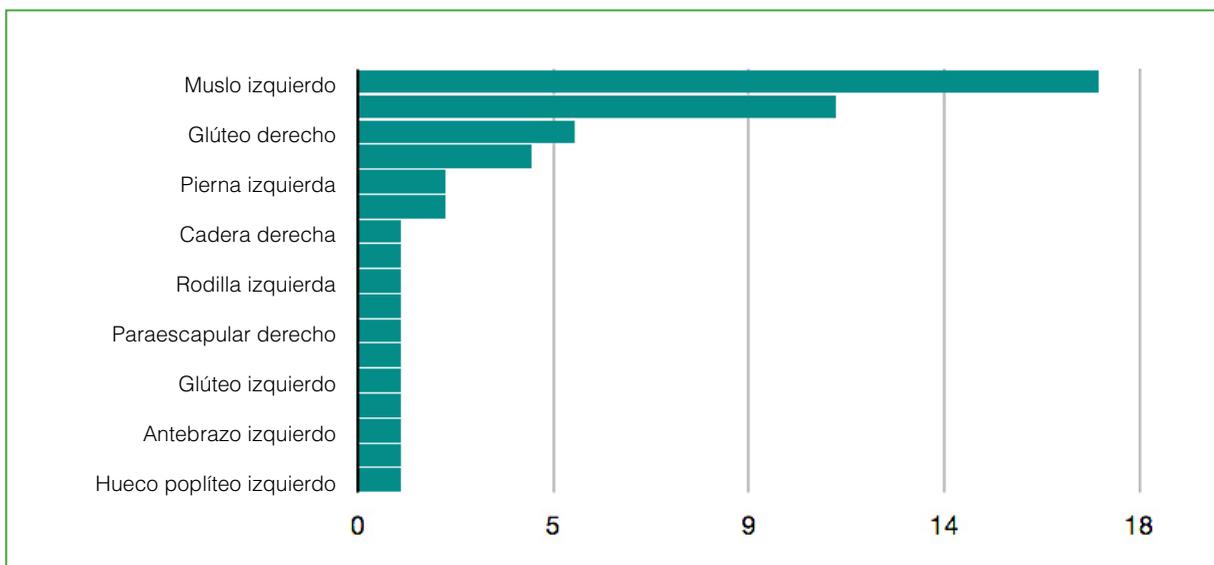


Figura 3. Distribución de las lesiones según su ubicación.

En un segundo análisis, para desarrollar el modelo de IA, se utilizaron 134 imágenes disponibles en el conjunto de datos, tanto las de alto grado como las de bajo grado, y en secuencias T1 y T2 de acuerdo con el estudio de imágenes realizado, tomando la imagen más representativa del tumor en cualquier corte (Figura 4).

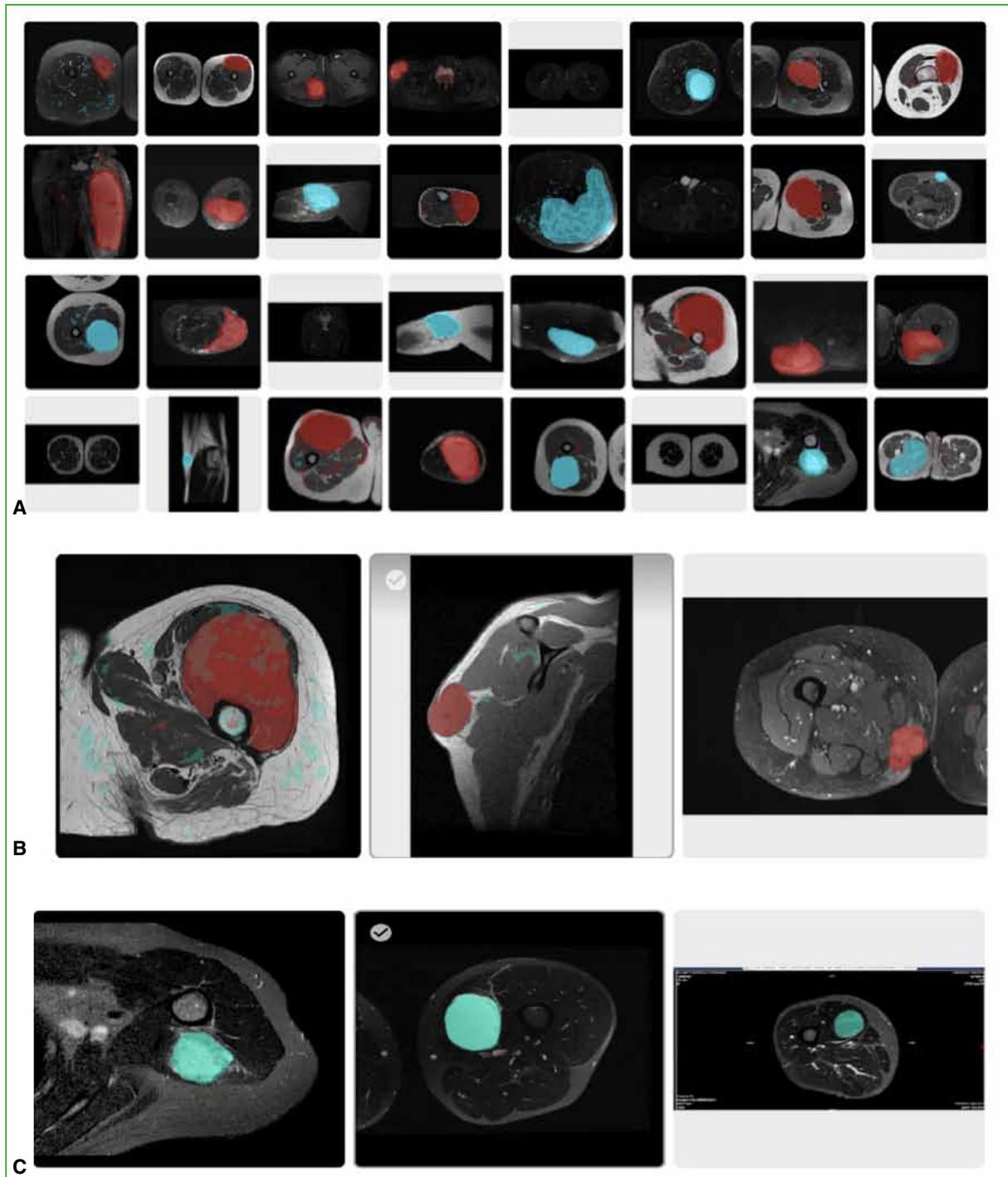


Figura 4. A. Etiquetado de las imágenes de los sarcomas de partes blandas de alto grado (rojo) y bajo grado (celeste).

B. Etiquetado de las lesiones de alto grado (rojo). Este proceso debe realizarse de manera meticulosa sin marcar la imagen por fuera del tumor, ya que puede alterar los resultados finales. **C.** El mismo procedimiento sobre las lesiones de bajo grado (celeste).

El adecuado y meticuloso etiquetado de las lesiones se realizó “pintando” la lesión por dentro de los bordes, debido a que las redes neuronales hacen el análisis de las imágenes de acuerdo con los píxeles marcados. Por lo tanto, una incorrecta marcación de las lesiones puede alterar los resultados finales.

Del total, 95 se utilizaron como muestras de entrenamiento (70,9%); 26 (19,4%), como imágenes de validación y 13, como imágenes de prueba (9,7%) para ambos grupos de pacientes (Figura 5).

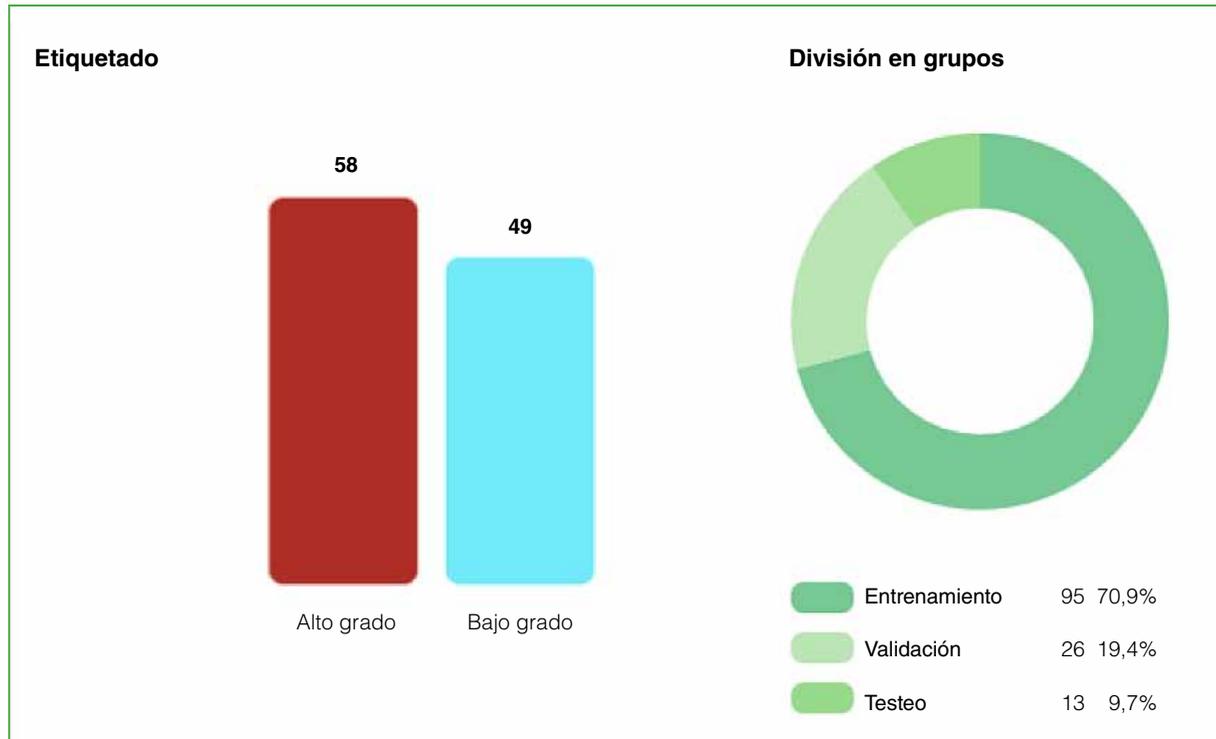


Figura 5. Etiquetado de las imágenes y división en grupo de entrenamiento, grupo de validación y grupo de testeo. Es importante destacar que los valores de precisión y sensibilidad se realizan sobre el grupo de testeo.

El total de píxeles que serán utilizados por el modelo para las 134 imágenes es de 36 millones. La matriz de confusión permite evaluar inicialmente el rendimiento de nuestro modelo. Los resultados se detallan en la Tabla 2. En los proyectos de segmentación, el objetivo es minimizar el número de falsos positivos y falsos negativos, y maximizar el número de verdaderos positivos y verdaderos negativos.

La precisión promedio del modelo fue del 84,3% y la sensibilidad, del 73,3%, con un umbral de confianza de 0,66, lo que demuestra inicialmente que se obtuvo un modelo de buena calidad para la predicción del grado de agresividad de un SPB antes de la biopsia con imágenes de RM (Figura 6).

Tabla 2. Matriz de confusión con el análisis de los 36 millones de píxeles que la red neuronal analizó

Verdaderos positivos	Falsos positivos
1.096.183	175.493
Falsos negativos	Verdaderos negativos
315.394	34.830.034

La matriz de confusión permite evaluar inicialmente el rendimiento de nuestro modelo. En los proyectos de segmentación de imágenes, el objetivo es minimizar el número de falsos positivos y falsos negativos, y maximizar el número de verdaderos positivos y verdaderos negativos.

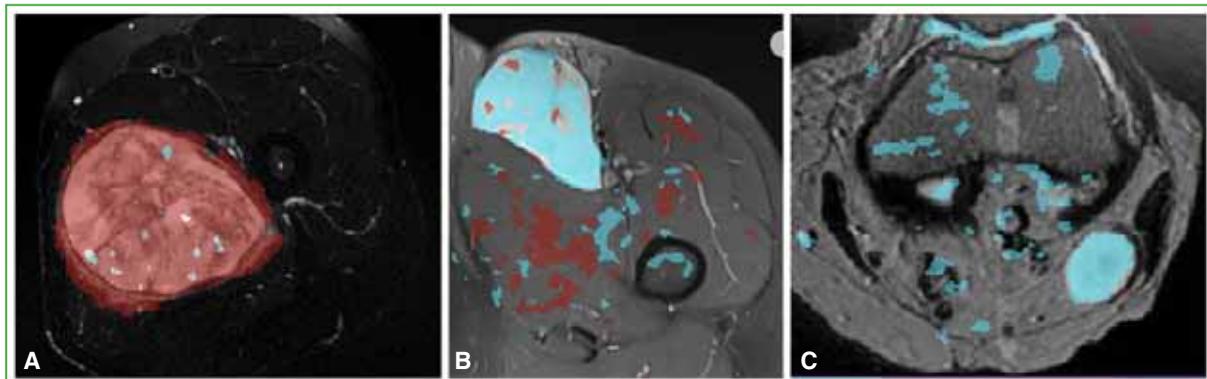


Figura 6. Despliegue del modelo, utilizando imágenes nunca visualizadas antes por el programa con una precisión del 84,3% y una sensibilidad del 73,3%. Clasifica correctamente: **A.** Sarcoma de alto grado (rojo). **B y C.** Sarcomas de bajo grado (celeste).

DISCUSIÓN

Debido a la heterogeneidad de esta enfermedad, el grado de agresividad inicial de la biopsia puede estar subestimando el grado de agresividad final identificado en el análisis del anatomopatólogo. Esto puede llevar a la necesidad de buscar herramientas de diagnóstico preoperatorio preciso para clasificar los SPB.¹³

Los SPB son un tipo infrecuente de cáncer. Por este motivo, su tratamiento requiere de centros interdisciplinarios especializados con profesionales expertos en el tema, ya que un diagnóstico equivocado o una demora pueden traer consecuencias catastróficas para el paciente. En este caso, la ayuda del AutoML puede resultar fundamental cuando se están realizando los estudios por imágenes, pues le permiten al equipo tratante tener una segunda opinión calificada del grado de malignidad tumoral en poco tiempo, antes de la biopsia y del resultado anatomopatológico.

El mejor modelo de AutoML utilizando Landing Lens de Landing AI logró una precisión del 84,3% y una sensibilidad del 73,3%, basándonos en imágenes de RM en cortes axiales sobre el diámetro mayor de la lesión en secuencias T1 y T2.

Los programas de análisis de imágenes (p. ej., Landing Lens de Landing AI; Vertex AI de Google Cloud Platform) permiten a los profesionales que nos estamos iniciando en la ciencia de datos y que todavía tenemos fallas en la implementación de código de programación, una mayor facilidad de uso en la implementación de modelos de aprendizaje automático, además de ahorrarnos tiempo en tareas básicas. En nuestra especialidad, esto resulta algo fundamental.

El estudio tiene sus limitaciones. Se trata de un trabajo retrospectivo de un solo centro, que podría haber llevado a un sesgo de selección. Sin embargo, las características epidemiológicas de la serie son comparables con las de otros artículos publicados.¹⁴ Además, al tratarse de una enfermedad infrecuente, el número de imágenes utilizadas por el modelo es todavía bajo. Si bien las técnicas de aumentación pueden ayudar a tener un conjunto de datos más robusto, resultaría fundamental el aporte y la colaboración de centros especializados en el tratamiento de sarcomas a fin de poder aumentar el número de imágenes para el estudio, intentando mejorar los resultados de los modelos (*data sharing*).

CONCLUSIONES

Este estudio presenta un enfoque novedoso para abordar un tipo de enfermedad infrecuente usando técnicas de aprendizaje profundo automatizado con el propósito de determinar el grado tumoral basándose en imágenes de RM. Aunque el verdadero impacto de la IA en la Ortopedia y Traumatología todavía está por demostrarse, existe mucha evidencia que avala el uso de estas tecnologías para generar valor en la atención médica. En la medida en que estas herramientas sean más fáciles de utilizar, habrá más y más modelos de aprendizaje automático construidos e implementados, lo que permitirá nuevas investigaciones y nuevos resultados, con el objetivo de mejorar el diagnóstico temprano de los pacientes con esta enfermedad y otras. Estas tecnologías deben verse como ayuda al profesional, transformándose en una herramienta que puede aumentar las capacidades de un médico en lugar de reemplazar sus responsabilidades. En este caso y con los resultados obtenidos, podríamos considerar a nuestro modelo de IA como una segunda opinión experta en el momento de realizar los estudios de imágenes en pacientes con SPB antes de la biopsia.

Conflicto de intereses: Los autores no declaran conflictos de intereses.

ORCID de P. Segura: <https://orcid.org/0000-0002-2376-4834>

ORCID de F. Segura: <https://orcid.org/0009-0004-0424-8334>

BIBLIOGRAFÍA

1. McCarthy J, Minsky ML, Rochester N, Shannon CE. A proposal for the Dartmouth Summer Research Project on Artificial Intelligence, August 31, 1955. *AI Magazine* 2006;27(4):12-4. <https://doi.org/10.1609/aimag.v27i4.1904>
2. Maxmen JS. *The post-physician era: Medicine in the 21st century*. Hoboken, New Jersey: Wiley; 1976, p. 2336-7.
3. Scarlet A. *A machine learning primer for clinicians—part 1*. HlStalk. 2018. Disponible en: <https://histalk2.com/2018/10/17/a-machine-learningprimer-for-clinicianspart-1/> [Consulta: 5 de agosto, 2019]
4. Chollet F. *Deep learning with Python*. Shelter Island, New York: Manning Publications; 2018.
5. Bini SA. Artificial intelligence, machine learning, deep learning, and cognitive computing: what do these terms mean and how will they impact health care? *J Arthroplasty* 2018;33(8):2358-61. <https://doi.org/10.1016/j.arth.2018.02.067>
6. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. *CA Cancer J Clin* 2019;69(1):7-34. <https://doi.org/10.3322/caac.21551>
7. Fletcher CDM, Bridge JA, Hogendoorn P, Mertens F. *WHO Classification of Tumors of Soft Tissue and Bone*. 4th ed., WHO Press; 2013, vol. 3.
8. Trojani M, Contesso G, Coindre JM, Rouesse J, Bui NB, de Mascarel A, et al. Soft tissue sarcomas in adults; study of pathological prognostic variables and definition of histopathological grading systems. *Int J Cancer* 1984;33(1):37-42. <https://doi.org/10.1002/ijc.2910330108>
9. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol* 2015;60(14):5471-96. <https://doi.org/10.1088/0031-9155/60/14/5471>
10. Refai F. The histopathological grading of soft tissue sarcomas: A review. *Saudi J Pathol Microbiol* 2019;4(8):581-5. <https://doi.org/10.21276/sjpm.2019.4.8.2>
11. Deyrup AT, Weiss SW. Grading of soft tissue sarcomas: the challenge of providing precise information in an imprecise world. *Histopathology* 2006;48(1):42-50. <https://doi.org/10.1111/j.1365-2559.2005.02288.x>
12. Amidi A, Amidi S. Hoja de referencia de consejos y trucos sobre Aprendizaje Automático. Métricas de clasificación. 2021. Disponible en: <https://stanford.edu/~shervine/l/es/teaching/cs-229/hoja-referencia-aprendizaje-automaEco-consejos-trucos>
13. Navarro F, Dapper H, Asadpour R, Knebel C, Spraker MB, Schwarze V, et al. Development and external validation of deep-learning-based tumor grading models in soft-tissue sarcoma patients using MR imaging. *Cancers (Basel)* 2021;13(12):2866. <https://doi.org/10.3390/cancers13122866>
14. Coindre JM, Terrier P, Guillou L, Le Doussal V, Collin F, Ranchère D, et al. Predictive value of grade for metastasis development in the main histologic types of adult soft tissue sarcomas: a study of 1240 patients from the French Federation of Cancer Centers Sarcoma Group. *Cancer* 2001;91(10):1914-26. [https://doi.org/10.1002/1097-0142\(20010515\)91:10<1914::aid-cnrc1214>3.0.co;2-3](https://doi.org/10.1002/1097-0142(20010515)91:10<1914::aid-cnrc1214>3.0.co;2-3)