Measuring Tumor Aggressiveness Through Artificial Intelligence

Facundo Segura, Pablo Segura, Florencio Segura
Private Orthopedics and Traumatology Center. 2° Chair in Orthopedics and Traumatology, Hospital Tránsito Cáceres de Allende, Universidad Nacional de Córdoba, Córdoba, Argentina

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

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

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

Received on March 14th, 2023. Accepted after evaluation on September 1st, 2023 • Dr. FACUNDO SEGURA • facusegura@gmail.com https://orcid.org/0009-0000-7101-9145
How to cite this article: Segura F, Segura P, Segura F. Measuring Tumor Aggressiveness Through Artificial Intelligence. *Rev Asoc Argent Ortop Traumatol* 2023;88(6):653-661. https://doi.org/10.15417/issn.1852-7434.2023.88.6.1738

INTRODUCTION

60 years ago, John McCarthy originally defined artificial intelligence (AI) as "the science and engineering to make intelligent machines", and he predicted that these machines would one day be capable of performing feats that were previously thought to be the exclusive domain of human intelligence, such as advanced problem solving.[1] Jerrold S. Maxmen, professor of psychiatry at Columbia University, predicted that AI would bring with it the "post-medical era" for the 21st century, describing change as "possible, inevitable and desirable."[2]

Computer vision is a function of AI through which machine learning techniques and complex neural networks can be applied to allow computers to capture, analyze and interpret information from clinical images, using algorithms that are trained to make classifications and predictions based on patterns provided by the data.[3] In the field of Orthopedics and Traumatology, the use of these techniques is expanding rapidly. Computer vision has made significant progress in the analysis of medical images in a variety of modalities commonly used in orthopedics, including radiography, computed tomography and magnetic resonance imaging (MRI). Automated machine learning (AutoML) is of great help to do this. In this case, we refer to the process of automating an action by applying common machine learning to solve real problems.[4] The high degree of automation in AutoML allows those who are not experts in AI to make use of machine learning models without the need for advanced experience in that field.[5]

Soft tissue sarcomas (STS) represent a rare and heterogeneous disease that comprises 1% of all malignant tumors in adults. In 2019, about 13,500 people were diagnosed with STS in the United States.[6] They encompass at least 100 different histological and molecular subtypes, and each subtype shows a variable clinical behavior.[7] Rare cases, together with the heterogeneity of the disease and its complex treatment, highlight the importance of the interdisciplinary management of these injuries, involving teams specialized in diagnostic imaging, pathological anatomy, radiation therapy, oncologists and orthopedic surgeons with experience in the management of STS.

When faced with a patient with STS, decision-making regarding the optimal treatment of the tumor depends, to a large extent, on the size and location of the STS, and on the outcome of the pathological anatomy study that evaluates the aggressiveness of the tumor.

Currently, the system we use to classify the histological grade of STS is the one proposed by the *French Fédération Nationale des Centres de Lutte Contre le Cancer (FNCLCC)* of France. The FNCLCC system evaluates only three parameters: tumor differentiation, mitotic index and amount of tumor necrosis (Table 1).[8]

The grade of the tumor is the most important factor in making the best therapeutic decision. The biopsy of the lesion is essential to determine the degree and establish the diagnosis with certainty. By analyzing MRI images and using AI techniques based on deep learning, it is possible to propose an alternative way of determining the degree of aggressiveness of STSs.

The objective of this study was to classify STSs, in a non-invasive way, into: low-grade (G1) and high-grade (G2/G3) using MRI images of sarcomas with histological grade proven by AutoML.

MATERIALS AND METHODS

Two independent retrospective cohorts of patients were formed. The inclusion criteria were: STS with proven histological grade with available information on the classification of tumors from the FNCLCC. The exclusion criteria were: MRI artifacts, previous radiation therapy, primary bone sarcoma or Ewing sarcoma.

The first cohort consisted of the publicly available *The Cancer Image Archive* (TCIA) dataset. This dataset contains computed/positron emission tomography images with fluorine-18-fluorodeoxyglucose and MRI (T1-weighted, T2-weighted with fat suppression [T2FS]) of 51 patients with histologically proven soft-tissue sarcomas in the extremities. All had undergone computed tomography/positron emission explorations with fluorine-18-fluorodeoxyglucose and MRI before treatment, between November 2004 and November 2011. In addition, the dataset contains the demographic data of all the patients analyzed.

The second cohort included images from the Private Center for Safe Orthopedics and Traumatology database. In this case, images were taken of 22 patients with STS with a proven histological grade, operated between 2020 and 2022.

Table 1. Degree of malignancy of soft tissue sarcomas

Parameter	Criterion	
	Tumor differentiation	
1 point	The sarcoma closely resembles normal adult mesenchymal tissue	
2 points	The histological type of sarcoma is defined	
3 points	Embryonic or dedifferentiated sarcomas, synovial sarcoma and sarcoma of the uncertain type	
	Mitotic Index	
1 point	0-9 by 10 high-power fields	
2 points	10-19 by 10 high-power fields	
3 points	>20 by 10 high-power fields	
	Tumor necrosis	
1 point	No necrosis	
2 points	≤50%	
3 points	≥50%	
Tumor grade		
Grade 1: Low	Total points	2, 3
Grade 2: Intermediate	Total points	4, 5
Grade 3: High	Total points	6, 7, 8

Score from 1 to 3 depending on the degree of tumor differentiation, the mitotic index and the percentage of tumor necrosis. A total minimum value <3 indicates low grade; between 3 and 6, intermediate grade; and >6, high grade.

In each patient of the two cohorts, three types of image sequences used as indicated by clinical protocols were acquired: T1-weighted, T2FS and STIR, which allows the signal of certain elements or tissues to be suppressed in a specific way (fat, water). The T1 sequences were acquired in the axial plane, while the T2FS and STIR sequences were taken in different orientations (axial, sagittal and coronal). In addition, the slice thickness was 5.5 mm for the T1-weighted sequence and 5 mm for the T2FS sequence. The in-plane resolution was 0.63 mm^2; 0.74 mm^2 and 0.86 mm^2 for the T2FS, T1 and STIR sequences, respectively.[9] From each complete MRI study of the patient, the most representative image was extracted in T1 and T2, this was the image where the largest diameter of the tumor was observed.

In total, 37 images of high-grade sarcomas were obtained in the T1 sequence and 37 images in the T2 sequence. For intermediate-grade sarcomas, 15 images were obtained in the T1 sequence and 15 images in the T2 sequence. The images of low-grade sarcomas were 20 for both the T1 sequence and the T2 sequence.

Grade 2 injuries have an uncertain prognosis and are therefore not useful for decision-making.[10] Taking into account that both high-grade and intermediate-grade sarcomas have similar pre-surgical and post-surgical development, prognosis, and treatment, for the purposes of this project, high-grade and intermediate-grade sarcomas were unified in the same dataset, leaving low-grade sarcomas separately.[11]

The analysis of the evaluated patients was performed using Google Colab using the Python language with the Numpy, Pandas, Matplotlib and Seaborn libraries to graph the results obtained. The implementation of the project's predictive model was carried out with the Landing Lens tool, an AI platform on the Landing.ai website. Before the data training stage, data augmentation techniques were used to increase the number of images to be analyzed. In this case, the image size was changed to a height of 1000 and a width of 1000 pixels. As an augmentation technique, a horizontal and vertical flip was performed with a 0.5 probability of being applied to each image.

The lesions in the MRI images were manually classified as high or low grade, with red indicating high-grade lesions and light blue indicating low-grade lesions.

Assessment

We discuss the most significant metrics to consider when evaluating the performance of a machine learning model in the context of a binary classification.[12]

Confusion matrix: it is used to have a more complete view when evaluating the performance of a model. It is a tool that allows the visualization of the performance of an algorithm, in this case, the analysis of millions of pixels that the model uses to learn.

Precision: The ratio of correct predictions with a test dataset. It is the ratio between the number of correct predictions and the total number of input examples. Practically speaking, it is the percentage of positive cases detected.

Recall: It is the proportion of positive cases that were correctly identified by the algorithm. It is represented as the fraction of true positives. It is also known as the true positive rate.

Precision and recall by confidence threshold: the model will decide whether the classification is high-grade or low-grade if the value returned by the model exceeds a decision threshold. If we increase this value, we will be increasing precision and, conversely, if we decrease it, we increase recall.

RESULTS

The first analysis was carried out on the distribution of the population and the disease. The average age in both groups was 54.82 years (± 16.98) (range 16-83). The distribution of the population by sex was similar (female gender 52.9%, male gender 47.1% [Z-statistic: 0.594, p 0.552]).

Figure 1 details the cases analyzed according to the pathological analysis result.

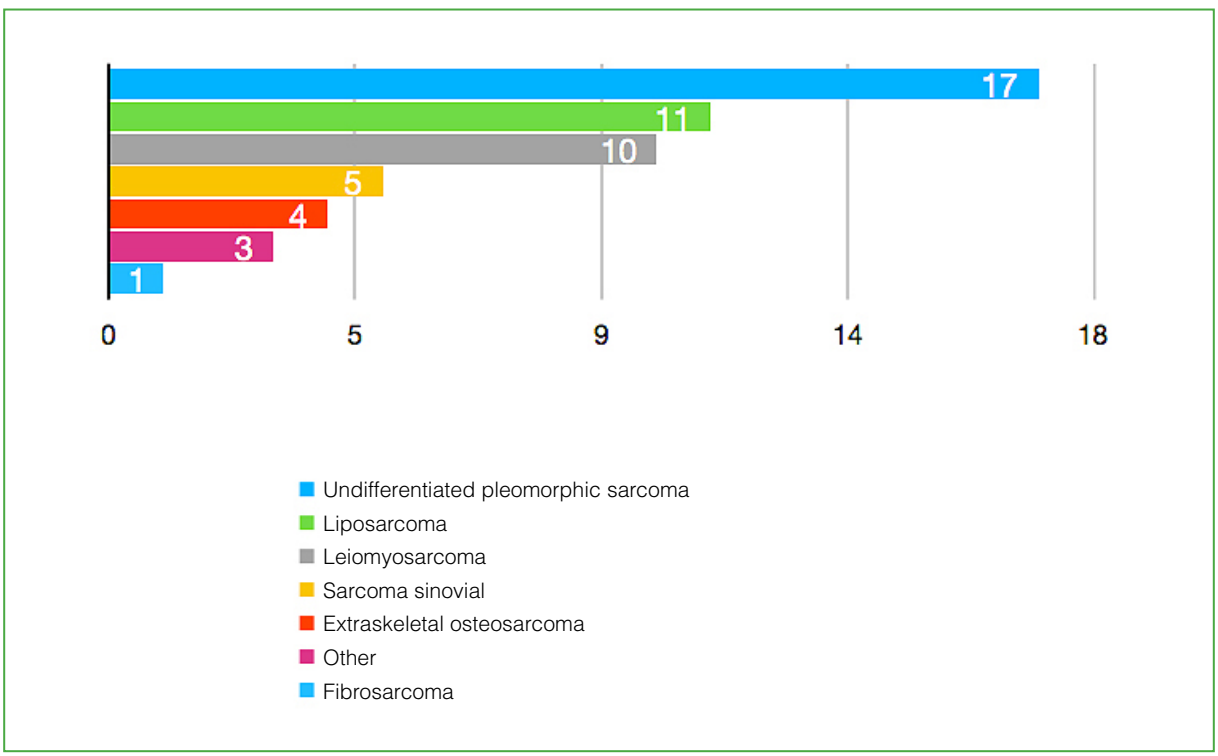

Figure 1. Distribution of cases according to the results of the pathology study.

Figures 2 and 3 show the distribution of the degree of aggressiveness of each type of tumor, as well as the distribution of the grade according to its location.

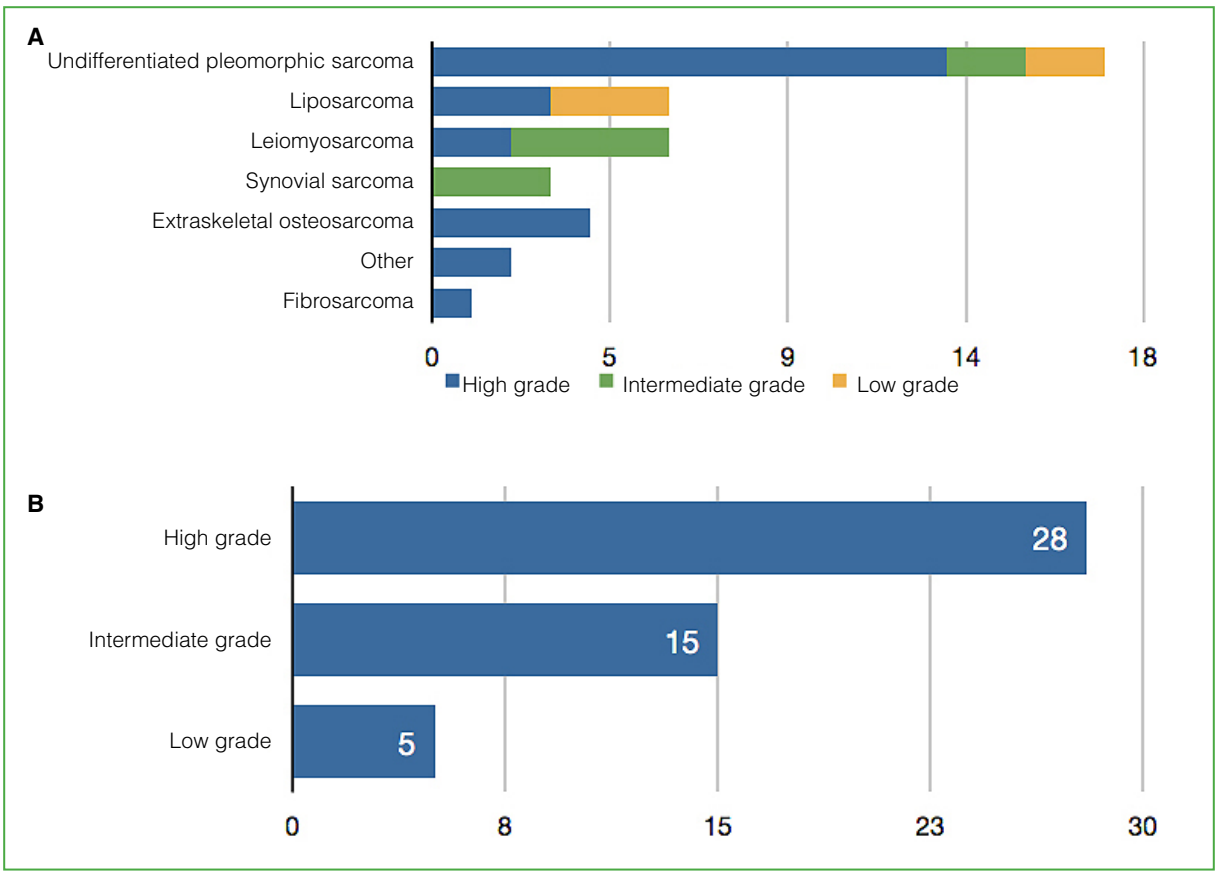

Figure 2. Distribution of soft-tissue sarcomas according to the degree of aggressiveness and their location.

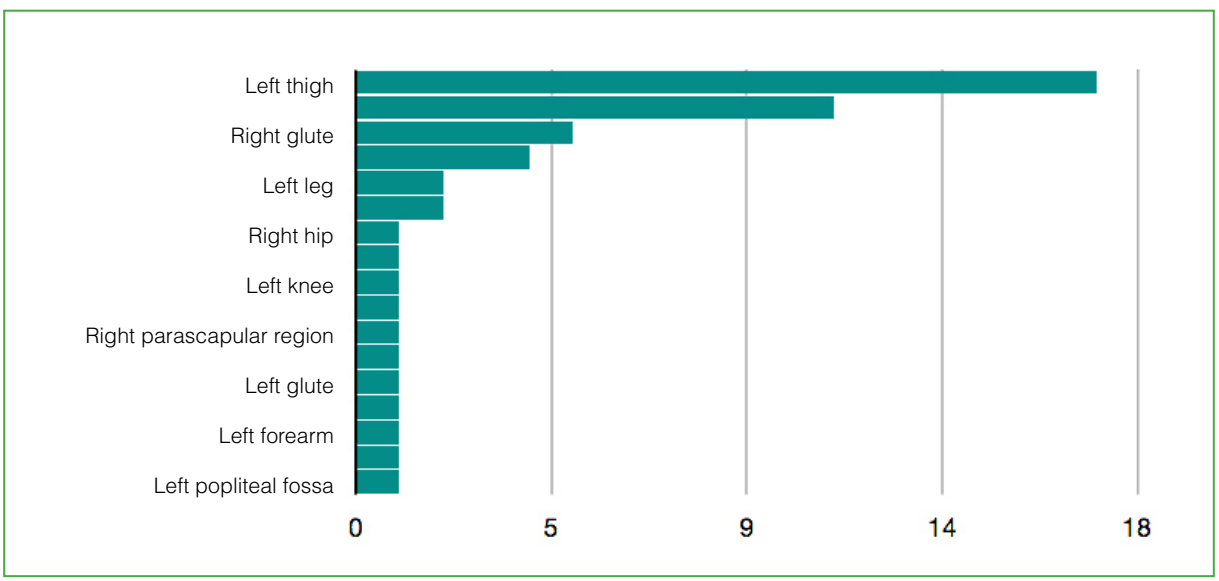

Figure 3. Distribution of lesions according to their location.

In a second analysis, to develop the AI model, 134 images available in the dataset were used, both high-grade and low-grade, and in T1 and T2 sequences according to the image study performed, taking the most representative image of the tumor in any section (Figure 4).

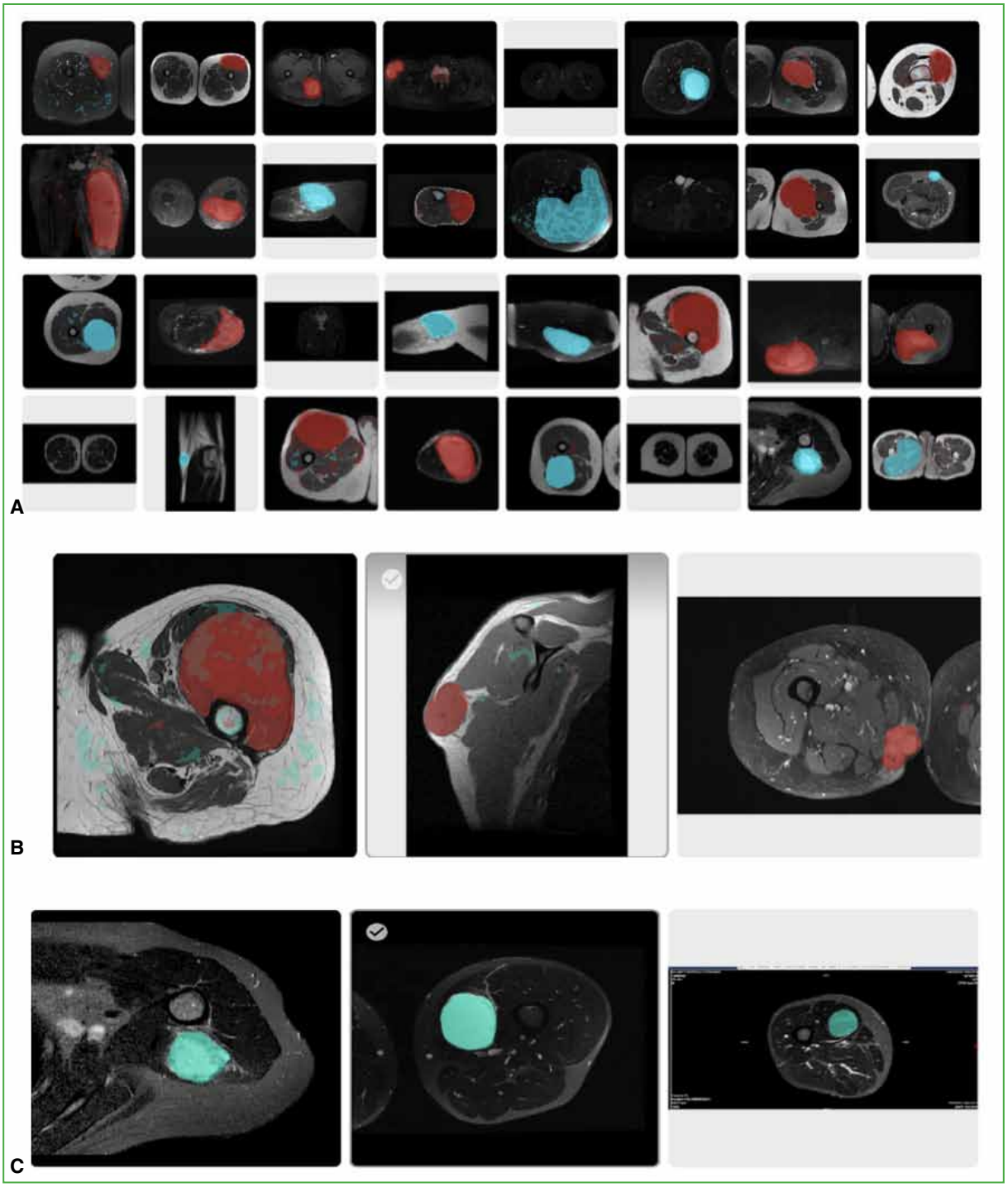

Figure 4. A. Labeling of images of high-grade (red) and low-grade (light blue) soft-tissue sarcomas. **B.** Labeling of high-grade injuries (red). This process must be done meticulously without marking the image outside of the tumor, as it can alter the final results. **C.** The same procedure for low-grade injuries (light blue).

The lesions were properly and meticulously labeled by 'coloring' them within the edges, because neural networks assess images based on the pixels assigned. Therefore, an incorrect marking of injuries can alter the final results.

Of the total, 95 were used as training samples (70.9%); 26 (19.4%), as validation images and 13, as test images (9.7%) for both groups of patients (Figure 5).

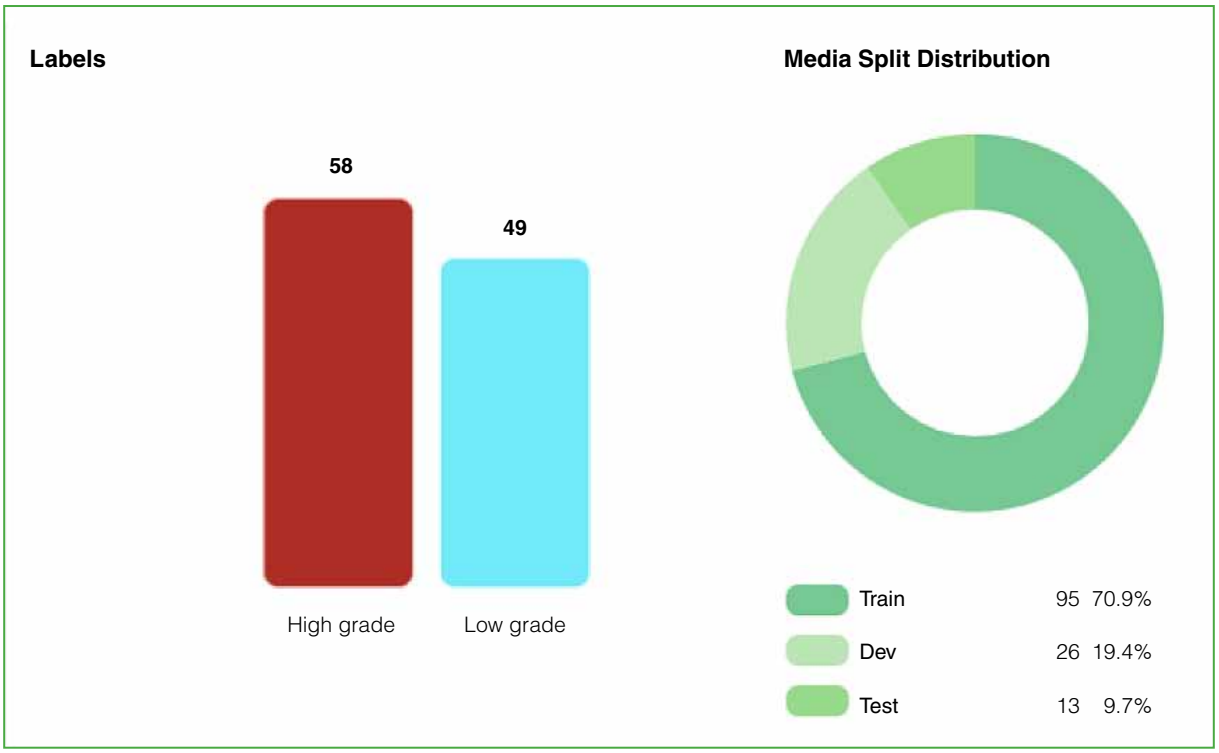

Figure 5. Labeling of the images and division into training group, validation group, and testing group. It is important to note that the precision and recall values are performed on the testing group.

The total number of pixels that will be used by the model for the 134 images is 36 million. The confusion matrix allows us to initially evaluate the performance of our model. The results are detailed in Table 2. In segmentation projects, the objective is to minimize the number of false positives and false negatives, and to maximize the number of true positives and true negatives.

The average precision of the model was 84.3% and the recall was 73.3%, with a confidence threshold of 0.66, which initially shows that a good-quality model was obtained with MRI images (Figure 6).

Table 2. Confusion matrix with the analysis of the 36 million pixels that the neural network analyzed

True positives	False positives
1,096,183	175,493
False negatives	**True negatives**
315,394	34,830,034

The confusion matrix allows us to initially evaluate the performance of our model. In image segmentation projects, the objective is to minimize the number of false positives and false negatives, and to maximize the number of true positives and true negatives.

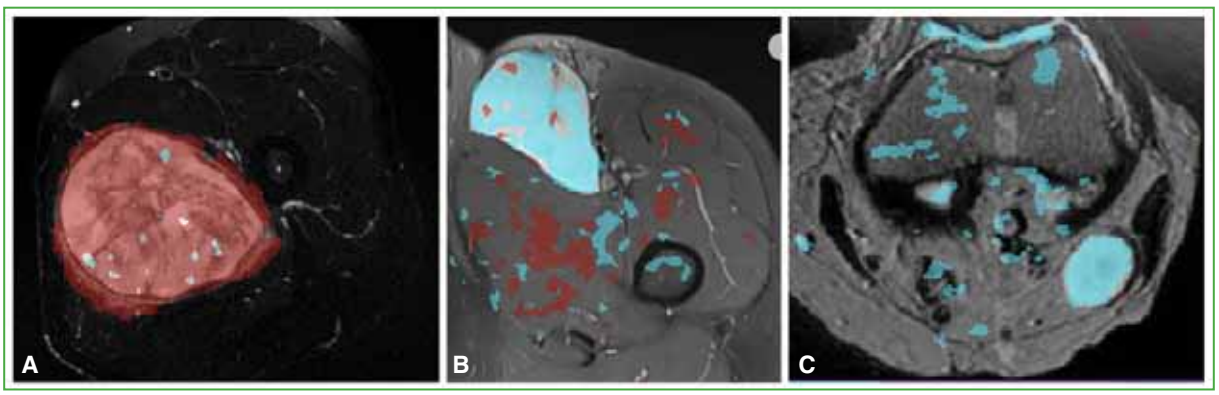

Figure 6. With a precision of 84.3% and a recall of 73.3% on images never seen by the program previously, the model can correctly classify: **A.** High-grade sarcoma (red). **B and C.** Low-grade sarcomas (light blue).

DISCUSSION

Due to the heterogeneity of this disease, the initial degree of aggressiveness of the biopsy may be underestimating the final degree of aggressiveness identified in the pathology analysis. This may lead to the need to search for accurate preoperative diagnostic tools to classify STS.[13]

STS is a rare type of cancer. For this reason, its treatment requires specialized interdisciplinary centers with professionals who are experts in the field, since a wrong diagnosis or a delay can have catastrophic consequences for the patient. In this case, the help of AutoML can be essential when imaging studies are being performed, as they allow the treating team to have a second qualified opinion on the degree of tumor malignancy in a short time, before the biopsy and the pathology analysis result.

The best AutoML model using Landing AI Landing Lens achieved a precision of 84.3% and a recall of 73.3%, based on MRI images in axial slices over the larger diameter of the lesion in T1 and T2 sequences.

Image analysis programs (e.g., Landing Lens from Landing AI; Vertex AI from Google Cloud Platform) allow professionals who are just starting out in data science and who still have flaws in implementing programming code, greater ease of use when implementing machine learning models, in addition to saving time on basic tasks. In our specialty, this is fundamental.

The study has its limitations. This is a retrospective study by a single center, which could have led to selection bias. However, the epidemiological characteristics of the series are comparable to those of other published articles.[14] In addition, as this is a rare disease, the number of images used by the model is still low. Although augmentation techniques can help to have a more robust data set, the contribution and collaboration of centers specialized in the treatment of sarcomas would be essential in order to increase the number of images for the study, trying to improve the results of the models (data sharing).

CONCLUSIONS

This study presents a novel approach to address a rare type of disease using automated deep learning techniques for the purpose of determining tumor grade based on MRI images. Although the true impact of AI in Orthopedics and Traumatology has yet to be demonstrated, there is much evidence to support the use of these technologies to generate value in healthcare. To the extent that these tools are easier to use, there will be more and more machine learning models built and implemented, allowing for new research and new results, with the aim of improving the early diagnosis of patients with this disease and others. These technologies should be seen as helping professionals, becoming a tool that can increase the capabilities of a doctor instead of replacing their responsibilities. In this case, with the results obtained, we could consider our AI model as a second expert opinion when performing imaging studies on patients with STS before the biopsy.

Conflict of interest: The authors declare no conflicts of interest.

P. Segura ORCID ID: https://orcid.org/0000-0002-2376-4834

F. Segura ORCID ID: https://orcid.org/0009-0004-0424-8334

REFERENCES

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. Scarlat A. *A machine learning primer for clinicians–part 1*. HIStalk. 2018. Available at: htpps://histalk2.com/2018/10/17/a-machine-learningprimer-for-clinicianspart-1/ [Consulted: August 5, 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. Available at: htpps://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-cncr1214>3.0.co;2-3