

Article

Not peer-reviewed version

Developing a Radiomics and Machine Learning Model Applied to Magnetic Resonance Images for Differential Diagnosis of Myometrial Tumors

[Valentina Chiappa](#) , [Giulia Gremmo](#) , [Matteo Interlenghi](#) , [Christian Salvatore](#) , [Giorgio Bogani](#) ,
[Simona Palladino](#) , [Umberto Leone Roberti Maggiore](#) , Giuseppina Calareso , [Biagio Paolini](#) , [Lucia Zanchi](#) ^{*} ,
Giulia Brugiavini , [Francesco Raspagliesi](#) , [Isabella Castiglioni](#)

Posted Date: 11 March 2026

doi: 10.20944/preprints202603.0895.v1

Keywords: radiomics; machine learning; leiomyoma; sarcoma; magnetic resonance imaging; diagnosis; differential



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Developing a Radiomics and Machine Learning Model Applied to Magnetic Resonance Images for Differential Diagnosis of Myometrial Tumors

Valentina Chiappa ¹, Giulia Gremmo ², Matteo Interlenghi ³, Cristian Salvatore ³, Giorgio Bogani ¹, Simona Palladino ¹, Umberto Leone Roberti Maggiore ¹, Giuseppina Calareso ⁴, Biagio Paolini ⁵, Lucia Zanchi ^{1,*}, Giulia Brugiavini ⁶, Francesco Raspagliesi ¹ and Isabella Castiglioni ⁷

- ¹ Gynecologic Oncology Unit, Fondazione IRCCS Istituto Nazionale Dei Tumori, Milan, Italy
- ² Department of Obstetrics and Gynecology, 'Filippo Del Ponte' Hospital, University of Insubria, Varese, Italy and Ospedale Michele e Pietro Ferrero, Division of Gynecologic Oncology, Verduno, Italy
- ³ DeepTrace Technologies S.R.L., Via Conservatorio 17, 20122 Milano, Italy
- ⁴ Radiology, Fondazione IRCCS Istituto Nazionale Dei Tumori, Milan, Italy
- ⁵ Pathology, Fondazione IRCCS Istituto Nazionale Dei Tumori, Milan, Italy
- ⁶ Department of Medical Sciences, Obstetrics and Gynecology Unit, Sant'Anna University Hospital, University of Ferrara, Via Aldo Moro, 8, 44124, Ferrara, Italy
- ⁷ Department of Physics "G. Occhialini", Università degli Studi di Milano-Bicocca, Milan, Italy
- * Correspondence: lucia.zanchi@istitutotumori.mi.it; Tel.: +39 02 2390 2507

Abstract

Background. The preoperative differential diagnosis of myometrial lesions remains a significant challenge when using conventional imaging techniques, such as ultrasound (US) and magnetic resonance imaging (MRI). Radiomics and machine learning, which leverage quantitative features beyond human visual perception, are increasingly recognized as promising tools for improving differential diagnosis in gynecology. **Methods.** This retrospective study included patients who underwent surgery for uterine masses and had preoperative MR. A machine learning model was developed to analyze radiomic features extracted from T2-weighted and diffusion-weighted MR images. **Results.** 44 subjects were included: 19 (43.2%) classified as "sarcoma" and 25 (56.8%) as "fibroid" based on histology after surgery. This dataset was used for training and cross-validation of different models. Three models, comprising ensembles of machine learning classifiers (random forests, support vector machines, and k-nearest neighbors), were developed for binary classification using histological diagnosis as reference standard. The best-performing model achieved the following results: AUC 90%, accuracy 82%, sensitivity 95%, specificity 72%, PPV 72%, and NPV 95%. **Conclusions.** Our model demonstrated high sensitivity and moderate accuracy, suggesting its potential as a valuable tool for assisting clinicians in the preliminary assessment of myometrial lesions and guiding decision-making toward conservative management in cases of non-suspicious masses.

Keywords: radiomics; machine learning; leiomyoma; sarcoma; magnetic resonance imaging; diagnosis; differential

1. Introduction

Uterine fibroids and sarcomas are distinct pathological entities with significant clinical implications [1]. Fibroids, also known as leiomyomas, are benign lesions originating from the smooth muscle cells of the uterine wall and represent the most common uterine tumors. Although benign,

fibroids can cause symptoms such as pelvic pain, abnormal uterine bleeding, and reproductive dysfunction, significantly affecting the quality of life of women. On the other hand, uterine sarcomas account for 3% to 5% of malignant uterine tumors originating from mesenchymal tissue and are associated with a poor prognosis [2–4]. Transvaginal ultrasound (US) is the first-line imaging technique for evaluating myometrial lesions. However, despite many efforts to identify pathognomonic ultrasound features of uterine sarcomas, it remains challenging to reliably distinguish benign from malignant lesions prior to surgery [5–7]. For this reason, magnetic resonance imaging (MR) plays a pivotal role in the evaluation of uterine masses due to its superior soft tissue contrast and multiplanar imaging capabilities, which are essential for preoperative assessment. MR facilitates detailed evaluation of tumor morphology, vascularity, and tissue characteristics, crucial for differentiating between fibroids and sarcomas. On MR, sarcomas may present as large, heterogeneous enhancing masses with central T2 hyperintensity in cases of necrosis, or as homogeneous low-signal masses, resembling leiomyomas [8–12]. Due to these varied features, distinguishing between these entities based solely on conventional MR characteristics can be challenging, as both fibroids and sarcomas may exhibit overlapping findings [13]. In this context, the differential diagnosis between fibroids and sarcomas remains a complex issue. Despite the rarity of sarcomas, even the remote possibility of a malignant myometrial lesion has a significant impact on therapeutic decisions for women. Specifically, the choice to pursue conservative or minimally invasive surgery (including morcellation) for large myometrial lesions in women of reproductive age could be made with greater confidence if a robust and objective diagnostic algorithm were available. In recent years, there has been increasing interest in the use of radiomics [14]—a quantitative imaging analysis technique—to improve the diagnostic accuracy of MR in differentiating uterine fibroids from sarcomas [14,15]. Radiomics involves the extraction of numerous quantitative features from medical images, which can be integrated into machine learning algorithms to assist clinicians in their decision-making processes. The rationale for utilizing radiomics in the context of uterine masses lies in its ability to identify subtle imaging biomarkers that are not easily detectable through visual inspection. By analyzing a broad range of radiomic features, including texture, shape, and intensity, radiomics has the potential to uncover hidden information within MR image data, thereby improving the differential diagnosis between benign and malignant lesions.

The aim of our study is to develop a machine learning model based on radiomics applied to MR images for the differential diagnosis of myometrial lesions. This model is intended to provide clinicians with a decision-support tool, which will require external validation in a multicenter, prospective setting. Through a systematic analysis of radiomic features extracted from MR scans of patients with histologically confirmed fibroids and sarcomas, we aim to identify robust radiomic signatures that can aid in the non-invasive differential diagnosis. Such findings may significantly influence patient management, including treatment planning and prognostication, ultimately improving outcomes for women with uterine masses.

2. Materials and Methods

2.1. Dataset

This is a retrospective study based on the prospective evaluation of patients enrolled in research protocol INT 155/20, approved by the Institutional Review Board (IRB) of the Fondazione IRCCS Istituto Nazionale dei Tumori di Milano. We included consecutive patients with uterine masses who underwent surgery at the Gynecologic Oncology Unit of the Fondazione IRCCS Istituto Nazionale dei Tumori between January 1, 2021, and January 1, 2024. All patients provided informed consent and authorized the collection of data and imaging for research purposes. The inclusion criteria were: (i) diagnosis of a uterine mass, (ii) preoperative MR examination performed within 4 weeks before surgery, (iii) surgery performed, and (iv) definitive histological diagnosis of either leiomyoma (including variants) or sarcoma. The exclusion criteria were: (i) age <18 years, (ii) absence of evaluable MR images, (iii) poor-quality images, and (iv) withdrawal of consent.

Additionally, to minimize potential confounding factors, patients with smooth muscle tumors of uncertain malignant potential (STUMP) were excluded. All MR imaging and surgeries were performed at the Fondazione IRCCS Istituto Nazionale dei Tumori, and histological examinations were reviewed by a dedicated pathologist with over 15 years of experience (P.B.), as required by the study protocol. MR images were stored in DICOM format at the time of preoperative evaluation.

2.2. Machine Learning Model

The machine learning model training workflow (summarized in Figure 1), performed with the Trace4ResearchTM platform [16], is the following:

- 1) Segmentation: manual segmentation of the volume of interest (VOI) was performed by an expert radiologist (C.G.) using the "Segmentation" tool of Trace4ResearchTM [16] to extract from T2-weighted image (T2WI), store and display the segmented portion of images based on the VOI.
- 2) Preprocessing: segmented VOI preprocessing included resampling to isotropic voxel spacing (2 mm).
- 3) Features computation: Radiomics features from the following feature families were computed from the segmented VOI, following the IBSI guidelines (definition, computation, and nomenclature) [17]: Morphological features, Intensity-based Statistic features, Intensity Histogram features, and Texture features: Gray-Level Co-occurrence Matrix (GLCM), Gray-Level Run Length Matrix (GLRLM), Gray-Level Size Zone Matrix (GLSZM), Neighborhood Gray Tone Difference Matrix (NGTDM), and Neighboring Gray Level Dependence Matrix (NGLDM). Intensity discretization was applied to the VOI, using a fixed number of 64 bins, to compute Intensity Histogram features and Texture features. In addition to the IBSI-compliant Radiomics features, a set of 2048 features were computed using the convolutional layers of a pre-trained Convolutional Neural Network (ResNet50) (DeepLearning features). VOI were pre-processed with an intensity discretization using a fixed number of 256 bins and resampled to a dimension of 224x224x20 voxels.
- 4) Feature reduction: Radiomics methodology was applied to MR images selecting informative and not redundant features according to the Image Biomarker Standardization Initiative (IBSI) guidelines [17], to maximize useful information and minimize redundancy. Features characterized by a low variance (below 0.1) and features with low mutual-information with the class label (based on a mutual-information analysis, with mutual information < 0.5), were removed.

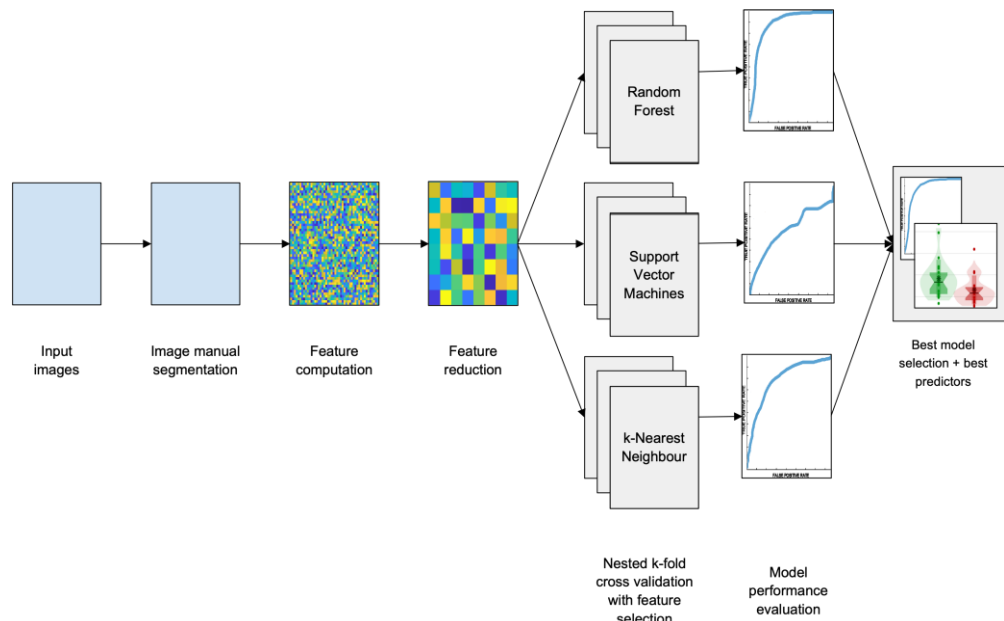


Figure 1. Machine learning model training workflow.

5) Machine learning models training: Three different models of machine-learning classifiers (Random Forest - RF, Support Vector Machines - SVM, and k nearest neighbor - kNN) were trained, validated, and internally tested, for the binary classification task of interest ("sarcoma" vs. "fibroid"), based on supervised learning, using histological diagnosis after surgery as reference standard. Each model was trained using a nested 10-fold cross validation method (i.e., using 8 folds for training, 1-fold for validation, and 1-fold for internal testing for each iteration). The first model consisted of 3 ensembles of 10x10 RF classifiers each, optimized with Gini Index; the second model consisted of 3 ensembles of 10x10 SVM classifiers; the third model consisted of 3 ensembles of 10x10 kNN classifiers. While the RF classifiers do not explicitly need further processing, since they are not affected by inter-correlation and perform autonomous feature selection, additional steps are required for SVM and kNN classifiers. For such models, to further select useful features and to eliminate the effects of inter-correlation, Principal Component Analysis (PCA) and Fisher Discriminant Ratio (FDR) ranking were performed in combination with a forward feature selection algorithm. For each iteration of the nested cross-validation, principal components were computed and ranked according to their FDR only on the training data; the number of components to be retained was then optimized by maximizing the Area Under the Receiver Operating Characteristic Curve (ROC-AUC) on the validation data.

To reduce the imbalance between the two classes, an oversampling technique for the minority class ("sarcoma") was applied, on training data, using an adaptive synthetic sampling method (ADASYN).

The following performances of the 3 models were measured on the internal testing data, across all iterations, both in terms of majority-vote and mean: ROC-AUC, Accuracy, Sensitivity, Specificity, Positive predictive value (PPV), Negative predictive value (NPV), and corresponding 95% confidence interval (CI).

The metrics are defined as follows:

$$\begin{aligned} \text{Accuracy (\%)} &= 100 \times \frac{\text{True Positive} + \text{True Negative}}{\text{Positive} + \text{Negative}} \\ \text{Sensitivity (\%)} &= 100 \times \frac{\text{True Positive}}{\text{Positive}} \\ \text{Specificity (\%)} &= 100 \times \frac{\text{True Negative}}{\text{Negative}} \\ \text{PPV (\%)} &= 100 \times \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \\ \text{NPV (\%)} &= 100 \times \frac{\text{True Negative}}{\text{True Negative} + \text{False Negative}} \end{aligned}$$

Positive and Negative are the number of subjects of the positive class ("sarcoma") and the negative class ("fibroid"), respectively. True Positive and True Negative are the number of subjects correctly classified by the model as belonging to the positive class and the negative class, respectively. False Positive and False Negative are the number of subjects wrongly classified by the model as belonging to the positive class and the negative class, respectively. The ROC curve plots true positive rate (i.e. Sensitivity) against false positive rate (i.e. 1-Specificity) at each threshold setting.

Finally, the model with the best performance, according to ROC-AUC, was chosen as the best classification model for the binary task of interest ("sarcoma" vs. "fibroid"). Information extracted was used to define radiomic biomarker profiles to discriminate malignant masses from benign masses. Histological subtypes were reported according to the 2020 World Health Organization (WHO) classification [18]. Statistical analysis was conducted with embedded tools of the Trace4ResearchTM platform [16]. To describe the distribution of each of the most relevant features in the "sarcoma" and "fibroid" classes, we calculated their medians with 95% CI and presented graphically violin and box plots for intuitive visualization and interpretation. A non-parametric univariate Wilcoxon rank-sum test (Mann-Whitney U test) was performed for each of the relevant radiomic predictors to verify its significance in discriminating "sarcoma" and "fibroid" classes. To account for multiple comparisons, the *p*-values were adjusted using the Bonferroni-Holm method and the significance levels were set at 0.05 (*) and 0.005 (**).

3. Results

After developing, testing and validating several models internally, we selected the one with the best performance.

3.1. Dataset

A total of 49 patients were identified for the analysis. Two patients were excluded due to withdrawal of consent, and three were excluded because of poor image quality (Figure 2). Overall, 44 patients were included in the final study population, and texture analysis was performed on stored MR images. Among these subjects, 19 women (43.2%) belonged to the class "sarcoma" (2 adenosarcoma, 14 leiomyosarcoma, 3 undifferentiated sarcoma) and 25 patients (56.8%) belonged to the class "fibroid" (intravascular leiomyomatosis, apoplectic leiomyomatosis, angioleiomyoma, leiomyoma, fibrosclerotic leiomyoma), according to histological diagnosis after surgery. Median age of "sarcoma" group patients was 58 years old (range 40 – 72), while median age of "fibroid" group patients was 46 years old (range 30 - 67). This image set was used for the training, cross-validation and internal testing of 3 machine-learning models.

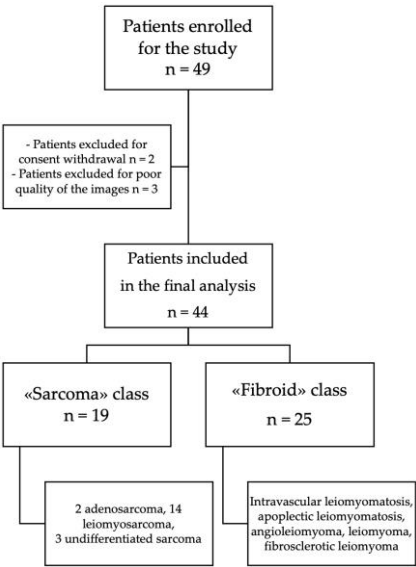


Figure 2. Study design.

3.2. Machine Learning Model

From each segmented VOI of each image considered in this study, IBSI-compliant radiomic features and deep learning-based features were computed, for a total of 3738 features. Among those, a total of 11 features were obtained after feature reduction, and used as predictors, according to the feature selection of each training iteration.

For the classification task of interest (19 images from class "sarcoma" vs. 25 images from class "fibroid"), these features were used for training, cross-validation and internal testing (nested 10-fold cross validation) of 3 different models of machine-learning classifiers considered in this work. Tables 1A-B-C show ROC-AUC, Accuracy, Sensitivity, Specificity, PPV and NPV as obtained from the training, cross-validation and internal testing of the 3 models consisting of 3 ensembles of machine-learning classifiers. Internal testing performances are reported as mean values across all iterations, and as majority vote with two different classification thresholds (balanced - 50%, and to maximize Sensitivity).

Table 1A. Model of 3 ensembles of random forest classifiers.

	Training	Validation	Internal testing (mean)	Internal testing ¹	Internal testing ²
ROC-AUC (%) [95% CI]	100* [99-100]	89** [85-92]	90** [87-92]	90	90
Accuracy (%) [95% CI]	100* [99-100]	82** [81-83]	83** [80-87]	82	82
Sensitivity (%) [95% CI]	100* [99-100]	76** [73-80]	79** [79-79]	74	95
Specificity (%) [95% CI]	100* [99-100]	86** [86-86]	87** [81-92]	88	72
PPV (%) [95% CI]	100* [99-100]	84** [81-87]	82** [76-88]	82	72
NPV (%) [95% CI]	100* [99-100]	87** [85-89]	84** [84-85]	81	95

¹ majority vote - 50% threshold; ² majority vote - 36% threshold that maximize sensitivity. Classification performance in terms of AUC, Accuracy, Sensitivity, Specificity, PPV, NPV, corresponding 95% confidence interval, and statistical significance with respect to chance/random classification (p-value). Performance is reported for training, validation and internal testing (mean, majority vote at 50% and majority vote to maximize Sensitivity). Statistical significance is set with * p-value <0.05 / ** p-value<0.005. ROC-AUC: Area Under the Receiver Operating Characteristic Curve, PPV: positive predictive value, NPV: negative predictive value.

Table 1B. Model of 3 ensembles of support vector machine classifiers.

	Training	Validation	Internal testing (mean)	Internal testing ¹	Internal testing ²
ROC-AUC (%) [95% CI]	89** [87-91]	83** [79-87]	73** [59-87]	73	73
Accuracy (%) [95% CI]	80** [77-82]	70** [64-75]	61 [32-91]	66	59
Sensitivity (%) [95% CI]	75** [71-78]	64** [57-71]	56 [18-94]	53	95
Specificity (%) [95% CI]	83** [81-85]	73** [66-80]	65* [42-88]	76	32

PPV (%) [95% CI]	79** [77-82]	69** [61-78]	55 [22-87]	63	51
NPV (%) [95% CI]	80** [78-82]	77** [70-83]	66 [39-94]	68	89

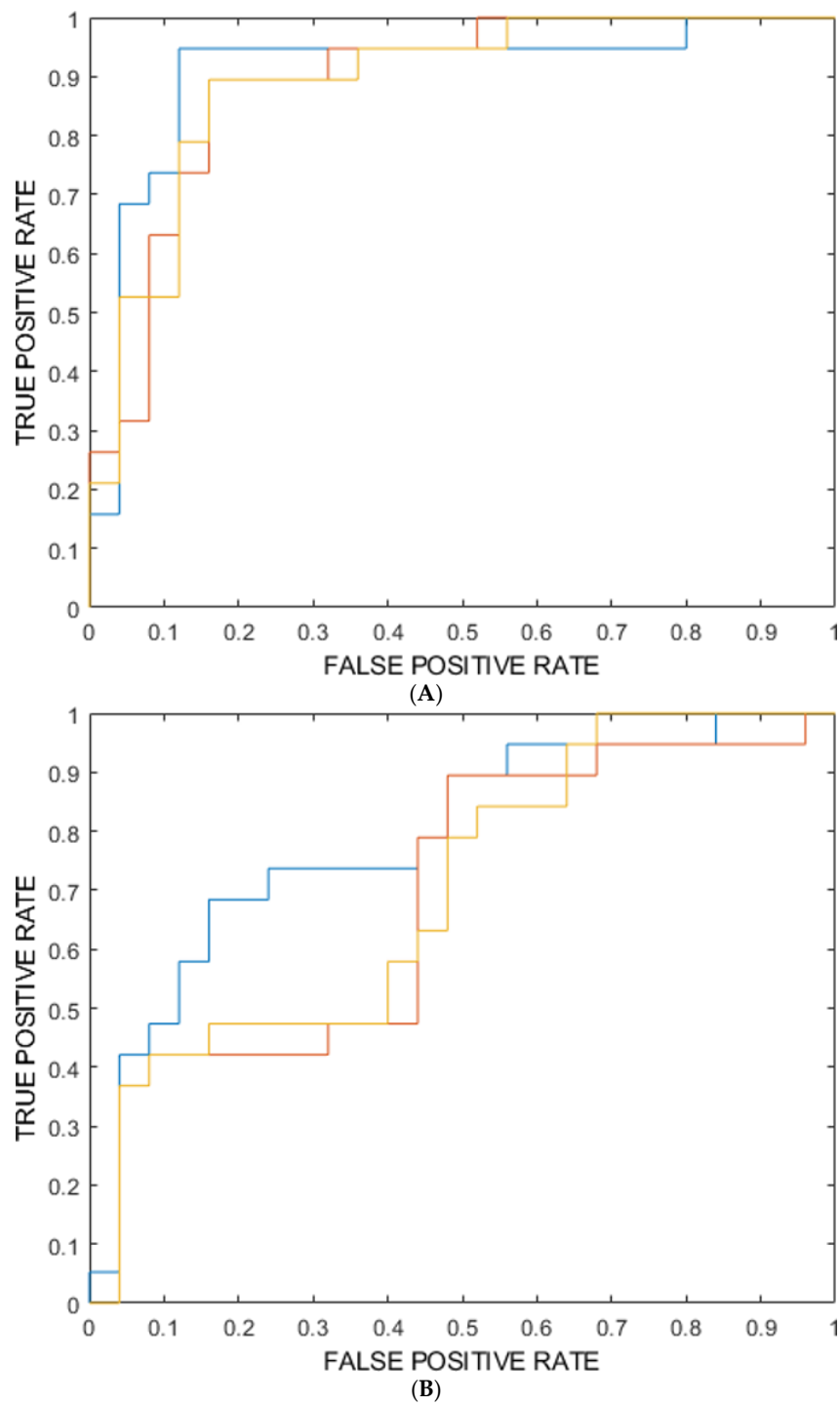
¹ majority vote - 50% threshold; ² majority vote - 13% threshold that maximize sensitivity. Classification performance in terms of AUC, Accuracy, Sensitivity, Specificity, PPV, NPV, corresponding 95% confidence interval, and statistical significance with respect to chance/random classification (p-value). Performance is reported for training, validation and internal testing (mean, majority vote at 50% and majority vote to maximize Sensitivity). Statistical significance is set with * p-value <0.05 / ** p-value<0.005. ROC-AUC: Area Under the Receiver Operating Characteristic Curve, PPV: positive predictive value, NPV: negative predictive value.

Table 1C. Model of 3 ensembles of support vector machine classifiers.

	Training	Validation	Internal testing (mean)	Internal testing ¹	Internal testing ²
ROC-AUC (%) [95% CI]	88** [86-89]	62** [57-68]	69* [50-88]	73	73
Accuracy (%) [95% CI]	80** [78-82]	59** [52-65]	62** [50-74]	68	66
Sensitivity (%) [95% CI]	75** [73-77]	51** [47-55]	42** [19-65]	47	95
Specificity (%) [95% CI]	84** [81-88]	64** [53-74]	77** [72-83]	84	44
PPV (%) [95% CI]	82** [78-86]	53** [44-62]	58** [40-76]	69	56
NPV (%) [95% CI]	81** [79-83]	66** [58-74]	64** [54-74]	68	92

¹ majority vote - 50% threshold; ² majority vote - 32% threshold that maximize sensitivity. Classification performance in terms of AUC, Accuracy, Sensitivity, Specificity, PPV, NPV, corresponding 95% confidence interval, and statistical significance with respect to chance/random classification (p-value). Performance is reported for training, validation and internal testing (mean, majority vote at 50% and majority vote to maximize Sensitivity). Statistical significance is set with * p-value <0.05 / ** p-value<0.005. ROC-AUC: Area Under the Receiver Operating Characteristic Curve, PPV: positive predictive value, NPV: negative predictive value.

Furthermore, for each model, ROC curves for the 3 ensembles are plotted in Figures 3A-B-C. Based on ROC-AUC, the model of random forest classifiers resulted to be the best model for the task of interest (19 images from class "sarcoma" vs. 25 images from class "fibroid").



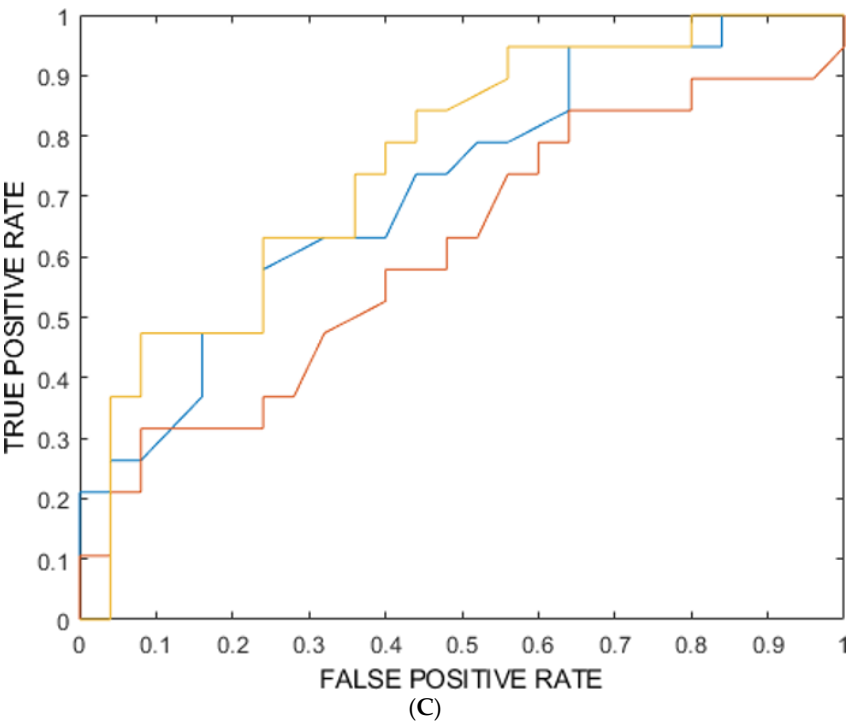


Figure 3. A. ROC Curve for the model consisting of 3 ensembles of random forest classifiers (from Internal Testing). B. ROC Curve for the model consisting of 3 ensembles of support vector machine classifiers (from Internal Testing). C. ROC Curve for the model consisting of 3 ensembles of k nearest neighbors classifiers (from Internal Testing).

The 11 radiomic predictors are shown in Table 2, together with their IBSI feature family and feature nomenclature. Predictors are ranked according to their statistical significance and to their frequencies among the most relevant ones in the ensemble of random forest classifiers. Median values of each feature, 95% CIs and results from univariate statistical sum rank tests are also reported with adjusted p-values (without and with Bonferroni correction).

Table 2. Ensemble of random forest classifiers. The 11 predictors are sorted in descending order according to their statistical significance and relevance.

#	Feature family	Feature nomenclature	Median in the positive class (95% CI)	Median in the negative class (95% CI)	Uncorrected p-value	Corrected p-value
1	Texture - Neighbouring Grey Level Dependence Matrix	High Dependence High Grey Level Emphasis - Exponential filter	761.86 [316.76 - 1206.96]	562.32 [400.75 - 723.89]	< 0.05	0.12

2	Texture - Grey-Level Co-Occurrence Matrix	Second Measure Of Information Correlation - Wavelet LLH filter	0.62 [0.55 - 0.68]	0.55 [0.5 - 0.6]	< 0.05	0.17
3	Deep Learning-Based	DeepFeature 770	0.33 [0.3 - 0.35]	0.29 [0.26 - 0.31]	< 0.05	0.22
4	Intensity Histogram	Median - Wavelet LHL filter	33 [31.02 - 34.98]	37 [35.43 - 38.57]	< 0.05	0.29
5	Intensity-Based Statistics	Coefficient Of Variation - Square filter	0.58 [0.46 - 0.69]	0.83 [0.67 - 1]	< 0.05	0.3
6	Deep Learning-Based	DeepFeature 1523	0.18 [8.06e-02 - 0.28]	0.29 [0.17 - 0.4]	0.11	1
7	Texture - Grey-Level Size Zone Matrix	Small Zone High Grey Level Emphasis - Exponential filter	166.66 [62.09 - 271.23]	102.89 [74.87 - 130.9]	0.19	1
8	Intensity-Based Statistics	Mean Absolut Deviation - Wavelet LLL filter	333.36 [289.84 - 376.87]	168.41 [36.94 - 299.88]	0.25	1
9	Intensity-Based Statistics	Median Absolute Deviation - Wavelet LLL filter	321.55 [276.28 - 366.82]	167.4 [47.2 - 287.59]	0.27	1
10	Intensity-Based Statistics	Variance - Wavelet LLL filter	1.55e+05 [1.17e+05 - 1.93e+05]	5.00e+04 [-9.63e+04 - 1.96e+05]	0.29	1

11	Intensity-Based Statistics	Variance - Logarithm filter	3159.3 [1362.72 - 4955.88]	3843.17 [1091.45 - 6594.89]	0.46	1
----	----------------------------	-----------------------------	-------------------------------	--------------------------------	------	---

Radiomics predictor distributions are also reported in violin plots and boxplots (Figure 4).

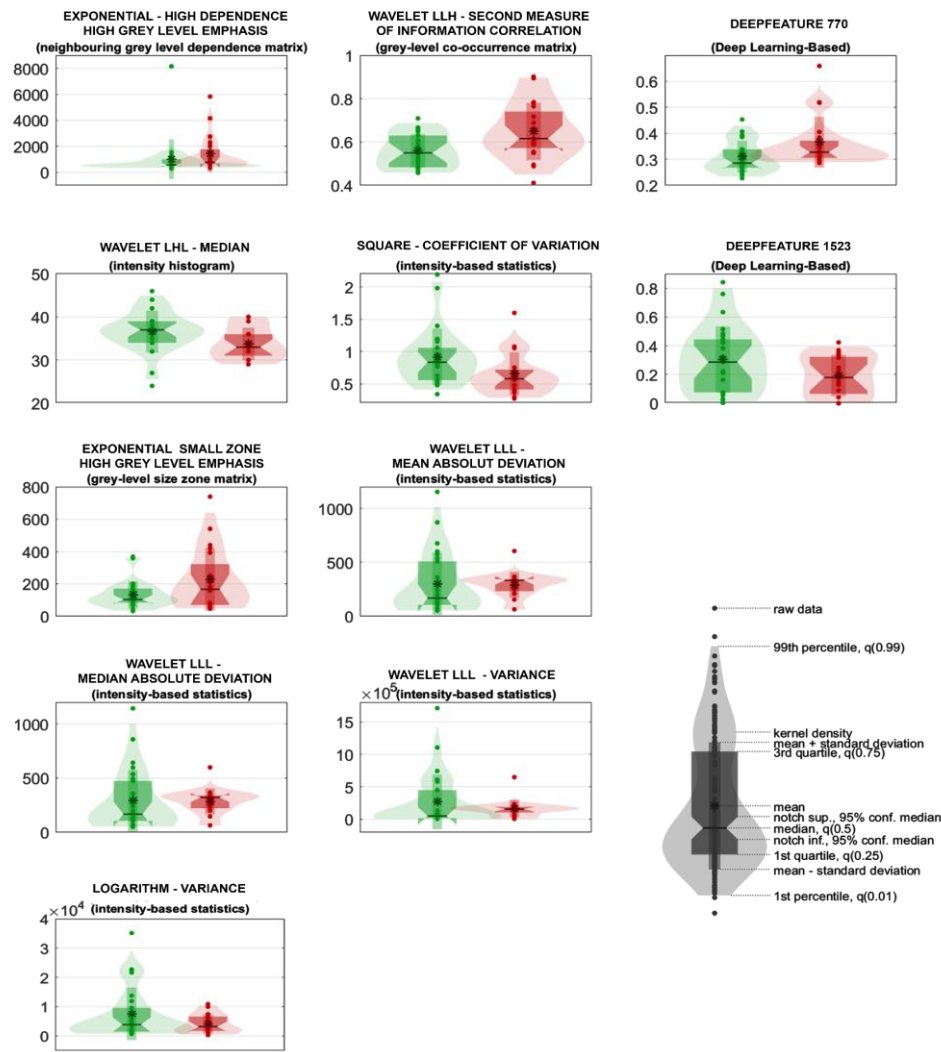


Figure 4. Ensemble of random forest. Violin and box plots of the radiomic predictors. Violin and box plots of "sarcoma" and "fibroid" classes are reported in red and green, respectively.

The best model for correctly classifying “sarcomas” vs “fibroids” showed: ROC-AUC (95% IC) of 90 (majority vote - 50% threshold), 90 (mean) [87-92], accuracy (%) of 82, 83**[80-87], sensitivity (%) of 74, 79** [79-79], specificity (%) of 88, 87** [81-92], PPV (%) of 82, 82** [76-88] and NPV (%) of 81, 84** [84-85] (*p<0,05, **p<0,005).

When a calibrated threshold was used to maximize sensitivity (i.e., 36% majority vote towards malignancy), the following internal testing performance were reached: ROC-AUC of 90, accuracy (%) of 82, sensitivity (%) of 95, specificity (%) of 72, PPV (%) of 72 and NPV (%) of 95.

4. Discussion

The differential diagnosis between uterine myomas and sarcomas is one of the most challenging in gynecological oncology. Uterine fibroids, on the one hand, are benign tumors commonly found in

premenopausal women. They are extremely prevalent, should not be over-treated, and can be managed conservatively or treated with conservative surgery. Uterine sarcomas, on the other hand, are rare malignant tumors with a very poor prognosis, particularly if not properly treated [19–21]. Conventional imaging methods for preoperative assessment, traditionally based on transvaginal US and MR, have several limitations. Over the years, various attempts have been made to identify pathognomonic imaging features of sarcomas and incorporate these into scoring systems that include demographic, clinical, and biochemical data (such as lactic dehydrogenase levels and isoforms). However, these efforts have not led to a significant improvement in clinical practice [22,23]. Implementing radiomics in a diagnostic algorithm may reduce operator-dependent errors and enhance the diagnostic accuracy of imaging in differentiating between benign and malignant myometrial lesions. In 2021, our group developed a machine learning model based on radiomics applied to transvaginal US images to discriminate between fibroids and sarcomas with promising results [24]. A recent systematic review by Raffone et al. [10] included 2495 MR images (2253 of uterine leiomyomas and 179 of uterine sarcomas) with a sensitivity of 0.90 and an area under the curve of 0.97. It confirms MR to be an adequate imaging modality to evaluate sarcomas, or rather, because of the heterogeneity of such lesions, as stated by Suzuki et al [25]. In their review, it can reliably extract cases for which the “possibility of sarcoma cannot be excluded.” By extracting and analyzing a multitude of quantitative imaging features, radiomics enables a more comprehensive evaluation of imaging data beyond what is perceptible to the naked eye. In the context of uterine sarcomas and fibroids, radiomics can potentially identify subtle imaging biomarkers indicative of malignancy, such as irregular margins, heterogeneous texture, and enhanced vascularity. When these features are quantified and analyzed systematically, they may contribute to a more nuanced understanding of the underlying tissue characteristics, aiding in the differentiation between benign and malignant lesions. Furthermore, the application of radiomics can address some of the limitations associated with conventional MR images interpretation, such as subjectivity and variability in qualitative assessments. By providing objective and quantitative metrics, radiomics may enhance inter-observer agreement and reproducibility in the diagnosis of uterine masses. This study achieved the best AUC of 0.90, comparable to that reported by Nakagawa et al. [26], who combined MR radiomics features with clinical parameters to build a machine learning model (AUC 0.93). Xie et al. also developed an artificial intelligence model for the differential diagnosis of fibroids and sarcomas using MR DWI, with promising results [27]; our study used only T2WI, which has more generalized clinical applicability, therefore making it a more widely available model. Our results are similar to those obtained by Roller et al. [28]: in their study, the Authors developed a model that integrates traditional MR features, clinical data, and radiomics, which performed better (AUC 0.989) than the model based on conventional MRI alone (AUC 0.929). This suggests that radiomics could serve as a valuable tool to complement the image assessment performed by an experienced radiologist.

The main advantages of our study were the high quality of data used to develop and train the model, both in terms of image quality and image processing by an experienced radiologist specializing in gynecology at an oncological referral center. Additionally, the quality of diagnosis, which is often challenging, was ensured by an experienced dedicated pathologist. A key advantage was the use of a platform for data analysis and model creation with a simple interface that can be fully customized by the clinician independently after brief training [16]. This allows the clinician to autonomously create diagnostic models using machine learning or deep learning on any topic, with various imaging methods.

The main limitations of the study are the small number of patients and its retrospective, single-center design (lack of external validation). To develop robust data within a model capable of significantly impacting clinical practice in this field, a much larger number of cases with available images would be required. Additionally, manual segmentation is very time-consuming for clinicians and we didn't assess inter-reader variability in segmentation among different radiologists with different levels of expertise. While this study is a retrospective analysis and has demonstrated good diagnostic performance, prospective multicentric studies are necessary to validate the current results.

Furthermore, while radiomics holds promise as a non-invasive tool for improving diagnostic accuracy, it should be viewed as a complementary rather than a replacement approach to conventional diagnostic methods. Clinical decision-making should still be based on a comprehensive evaluation that incorporates clinical history, physical examination, imaging findings, and, when necessary, histopathological analysis.

5. Conclusions

The radiomics analysis of MR imaging data in patients with myometrial lesions yields valuable insights. By leveraging advanced imaging techniques, we can discern subtle radiomic features that may distinguish between fibroids and sarcomas. This study underscores the potential of radiomics as a non-invasive tool for aiding in the differential diagnosis of uterine masses. However, further validation through larger prospective studies is warranted to establish the robustness and reliability of these findings in clinical practice. Integrating radiomics into routine MR evaluations could enhance diagnostic accuracy and facilitate more personalized treatment strategies for patients with uterine masses. This might be a precious aid before proposing conservative treatment to patients interested in fertility-sparing options.

Author Contributions: Conceptualization, V.C. and I.C.; data curation, S.P., G.G., G.C., B.P., L.Z., G.B. (Giulia Brugiavini), U.L.R.M., M.I., C.S.; formal analysis, M.I.; C.S.; investigation, S.P., G.G., G.C., B.P., L.Z., G.B. (Giulia Brugiavini), U.L.R.M., M.I., C.S.; methodology, I.C.; M.I., C.S.; project administration, I.C., V.C.; resources, F.R., I.C., G.C.; software, I.C.; M.I., C.S.; supervision, F.R., I.C.; validation, M.I., I.C.; visualization, G.G., G.B. (Giorgio Bogani), V.C., I.C., M.I.; writing—original draft, G.G.; G.B. (Giorgio Bogani), V.C., S.P.; writing – review and editing, V.C., M.I., C.S., G.B. (Giorgio Bogani), S.P., U.L.R.M., G.C., B.P., L.Z., I.C. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board of Fondazione IRCCS Istituto Nazionale dei Tumori, Milan (protocol code 155/20, 23/10/2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. Written informed consent has been obtained from the patient(s) to publish this paper.

Data Availability Statement: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Conflicts of Interest: C.S. and M.I. declare that they are employees of DeepTrace Technologies S.R.L., a spin-off of Scuola Universitaria Superiore IUSS, Pavia, Italy; I.C. and M.I. declare ownership of shares in DeepTrace Technologies S.R.L.; V.C. declares serving as a medical advisor for DeepTrace Technologies S.R.L. The other authors of this manuscript declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CI	Confidence interval
FDR	Fisher Discriminant Ratio
GLCM	Gray-Level Co-occurrence Matrix
GLRLM	Gray-Level Run Length Matrix
GLSZM	Gray-Level Size Zone Matrix
IBSI	Image Biomarker Standardization Initiative
IRB	Institutional Review Board
kNN	K-nearest neighbor
MR	Magnetic resonance
NGLDM	Neighboring Gray Level Dependence Matrix

NGTDM	Neighborhood Gray Tone Difference Matrix
NPV	Negative predictive value
PCA	Principal components analysis
PPV	Positive predictive value
RF	Random forest
ROC-AUC	Area Under the Receiver Operating Characteristic Curve
STUMP	Smooth muscle tumors of uncertain malignant potential
SVM	Support vector machines
T2WI	T2 weighted image
US	Ultrasound
VOI	Volume of interest
WHO	2020 World Health Organization

References

1. Bell SW, Kempson RL, Hendrickson MR. Problematic uterine smooth muscle neoplasms. A clinicopathologic study of 213 cases. *Am J Surg Pathol.* 1994;18(6):535-58.
2. Taghzouti H, Saoud MK. 2022-RA-1158-ESGO Uterine sarcoma. *International Journal of Gynecological Cancer*2022.
3. Ferrandina G, Aristei C, Biondetti PR, Cananzi FCM, Casali P, Ciccarone F, et al. Italian consensus conference on management of uterine sarcomas on behalf of S.I.G.O. (Societa' italiana di Ginecologia E Ostetricia). *Eur J Cancer.* 2020;139:149-68. doi: 10.1016/j.ejca.2020.08.016.
4. D'Angelo E, Prat J. Uterine sarcomas: a review. *Gynecol Oncol.* 2010;116(1):131-9. doi: 10.1016/j.ygyno.2009.09.023.
5. Ludovisi M, Moro F, Pasciuto T, Di Noi S, Giunchi S, Savelli L, et al. Imaging in gynecological disease (15): clinical and ultrasound characteristics of uterine sarcoma. *Ultrasound Obstet Gynecol.* 2019;54(5):676-87. doi: 10.1002/uog.20270.
6. Exacoustos C, Romanini ME, Amadio A, Amoroso C, Szabolcs B, Zupi E, et al. Can gray-scale and color Doppler sonography differentiate between uterine leiomyosarcoma and leiomyoma? *J Clin Ultrasound.* 2007;35(8):449-57. doi: 10.1002/jcu.20386.
7. Aviram R, Ochshorn Y, Markovitch O, Fishman A, Cohen I, Altaras MM, et al. Uterine sarcomas versus leiomyomas: gray-scale and Doppler sonographic findings. *J Clin Ultrasound.* 2005;33(1):10-3. doi: 10.1002/jcu.20075.
8. Namimoto T, Yamashita Y, Awai K, Nakaura T, Yanaga Y, Hirai T, et al. Combined use of T2-weighted and diffusion-weighted 3-T MR imaging for differentiating uterine sarcomas from benign leiomyomas. *Eur Radiol.* 2009;19(11):2756-64. doi: 10.1007/s00330-009-1471-x.
9. Valletta R, Corato V, Lombardo F, Avesani G, Negri G, Steinkasserer M, et al. Leiomyoma or sarcoma? MRI performance in the differential diagnosis of sonographically suspicious uterine masses. *Eur J Radiol.* 2024;170:111217. doi: 10.1016/j.ejrad.2023.111217.
10. Raffone A, Raimondo D, Neola D, Travaglino A, Giorgi M, Lazzeri L, et al. Diagnostic accuracy of MRI in the differential diagnosis between uterine leiomyomas and sarcomas: A systematic review and meta-analysis. *Int J Gynaecol Obstet.* 2024;165(1):22-33. doi: 10.1002/ijgo.15136.
11. DeMulder D, Ascher SM. Uterine Leiomyosarcoma: Can MRI Differentiate Leiomyosarcoma From Benign Leiomyoma Before Treatment? *AJR Am J Roentgenol.* 2018;211(6):1405-15. doi: 10.2214/AJR.17.19234.
12. Tamai K, Koyama T, Saga T, Morisawa N, Fujimoto K, Mikami Y, et al. The utility of diffusion-weighted MR imaging for differentiating uterine sarcomas from benign leiomyomas. *Eur Radiol.* 2008;18(4):723-30. doi: 10.1007/s00330-007-0787-7.
13. Rosa F, Martinetti C, Magnaldi S, Rizzo S, Manganaro L, Migone S, et al. Uterine mesenchymal tumors: development and preliminary results of a magnetic resonance imaging (MRI) diagnostic algorithm. *Radiol Med.* 2023;128(7):853-68. doi: 10.1007/s11547-023-01654-1.
14. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology.* 2016;278(2):563-77. doi: 10.1148/radiol.2015151169.

15. Dai M, Liu Y, Hu Y, Li G, Zhang J, Xiao Z, et al. Combining multiparametric MRI features-based transfer learning and clinical parameters: application of machine learning for the differentiation of uterine sarcomas from atypical leiomyomas. *Eur Radiol.* 2022;32(11):7988-97. doi: 10.1007/s00330-022-08783-7.
16. Technologies D: https://www.deeptracetechnology.com/files/TechnicalSheet__TRACE4.pdf : 9-11. Accessed.
17. Zwanenburg A, Vallieres M, Abdalah MA, Aerts H, Andrearczyk V, Apte A, et al. The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping. *Radiology.* 2020;295(2):328-38. doi: 10.1148/radiol.2020191145.
18. Sbaraglia M, Bellan E, Dei Tos AP. The 2020 WHO Classification of Soft Tissue Tumours: news and perspectives. *Pathologica.* 2021;113(2):70-84. doi: 10.32074/1591-951X-213.
19. Raspagliesi F, Maltese G, Bogani G, Fuca G, Lepori S, De Iaco P, et al. Morcellation worsens survival outcomes in patients with undiagnosed uterine leiomyosarcomas: A retrospective MITO group study. *Gynecol Oncol.* 2017;144(1):90-5. doi: 10.1016/j.ygyno.2016.11.002.
20. Bogani G, Chiappa V, Ditto A, Martinelli F, Donfrancesco C, Indini A, et al. Morcellation of undiagnosed uterine sarcoma: A critical review. *Crit Rev Oncol Hematol.* 2016;98:302-8. doi: 10.1016/j.critrevonc.2015.11.015.
21. Bogani G, Cliby WA, Aletti GD. Impact of morcellation on survival outcomes of patients with unexpected uterine leiomyosarcoma: a systematic review and meta-analysis. *Gynecol Oncol.* 2015;137(1):167-72. doi: 10.1016/j.ygyno.2014.11.011.
22. Nagai T, Takai Y, Akahori T, Ishida H, Hanaoka T, Uotani T, et al. Novel uterine sarcoma preoperative diagnosis score predicts the need for surgery in patients presenting with a uterine mass. *Springerplus.* 2014;3:678. doi: 10.1186/2193-1801-3-678.
23. Nagai T, Takai Y, Akahori T, Ishida H, Hanaoka T, Uotani T, et al. Highly improved accuracy of the revised PREoperative sarcoma score (rPRESS) in the decision of performing surgery for patients presenting with a uterine mass. *Springerplus.* 2015;4:520. doi: 10.1186/s40064-015-1318-7.
24. Chiappa V, Interlenghi M, Salvatore C, Bertolina F, Bogani G, Ditto A, et al. Using rADiOMics and machine learning with ultrasonography for the differential diagnosis of myometrial tumors (the ADMIRAL pilot study). *Radiomics and differential diagnosis of myometrial tumors. Gynecol Oncol.* 2021;161(3):838-44. doi: 10.1016/j.ygyno.2021.04.004.
25. Suzuki A, Aoki M, Miyagawa C, Murakami K, Takaya H, Kotani Y, et al. Differential Diagnosis of Uterine Leiomyoma and Uterine Sarcoma using Magnetic Resonance Images: A Literature Review. *Healthcare (Basel).* 2019;7(4). doi: 10.3390/healthcare7040158.
26. Nakagawa M, Nakaura T, Namimoto T, Iyama Y, Kidoh M, Hirata K, et al. Machine Learning to Differentiate T2-Weighted Hyperintense Uterine Leiomyomas from Uterine Sarcomas by Utilizing Multiparametric Magnetic Resonance Quantitative Imaging Features. *Acad Radiol.* 2019;26(10):1390-9. doi: 10.1016/j.acra.2018.11.014.
27. Xie H, Hu J, Zhang X, Ma S, Liu Y, Wang X. Preliminary utilization of radiomics in differentiating uterine sarcoma from atypical leiomyoma: Comparison on diagnostic efficacy of MRI features and radiomic features. *Eur J Radiol.* 2019;115:39-45. doi: 10.1016/j.ejrad.2019.04.004.
28. Roller LA, Wan Q, Liu X, Qin L, Chapel D, Burk KS, et al. MRI, clinical, and radiomic models for differentiation of uterine leiomyosarcoma and leiomyoma. *Abdom Radiol (NY).* 2024;49(5):1522-33. doi: 10.1007/s00261-024-04198-8.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.