MRI Radiomics Analysis in the Diagnostic Differentiation of Malignant Soft Tissue Myxoid Sarcomas From Benign Soft Tissue Musculoskeletal Myxomas

Hadas Benhabib, MD, MSc, FRCPC,[1,2] Daniel Brandenberger, MD,[1,2,3] Katherine Lajkosz, MSc,[4] Elizabeth G. Demicco, MD, PhD,[5,6] Kim M. Tsoi, MD, PhD, FRCSC,[5,6,7] Jay S. Wunder, MD, FRCSC,[7,8] Peter C. Ferguson, MD, MSc, FRCSC,[7,8] Anthony M. Griffin, MSc,[8] Ali Naraghi, MBBS, MRCP, FRCR,[1,2] Masoom A. Haider, MD, FRCPC,[1,2] and Lawrence M. White, MD, FRCPC[1,2,7]*

Background: Differentiation of benign myxomas and malignant myxoid sarcomas can be difficult with an overlapping spectrum of morphologic MR findings.
Purpose: To assess the diagnostic utility of MRI radiomics in the differentiation of musculoskeletal myxomas and myxoid sarcomas.
Study Type: Retrospective.
Population: A total of 523 patients were included; histologically proven myxomas ($N = 201$) and myxoid sarcomas ($N = 322$), randomly divided (70:30) into training:test subsets.
Sequence/Field Strength: T1-weighted (T1W), T2-weighted fat-suppressed (fluid-sensitive), and T1-weighted post-contrast (T1W + C) sequences at 1.0 T, 1.5 T, or 3.0 T.
Assessment: Seven semantic (qualitative) tumor features were assessed in each case. Manual 3D tumor segmentations performed with radiomics features extracted from T1W, fluid-sensitive, and T1W + C acquisitions. Models were constructed based on radiomic features from individual sequences and from their combination, both with and without the addition of qualitative tumor features.
Statistical Tests: Intraclass correlation evaluated in 60 cases segmented by three readers. Features with intraclass correlation <0.7 excluded from further analysis. Boruta feature selection and Random Forest modeling performed using the training-dataset, with resultant models used to assess class discrimination (myxoma vs. myxoid sarcoma) in the test dataset. Radiomics score defined as probability class = myxoma. Logistic regression modeling employed to estimate performance of the radiomics score. Area under the receiver operating characteristic curve (AUC) was used to assess diagnostic performance, and DeLong's test to assess performance between constructed models. A *P*-value <0.05 was considered significant.
Results: Four qualitative semantic features showed significant predictive power in class discrimination. Radiomic models demonstrated excellent differentiation of myxomas from myxoid sarcomas: AUC of 0.9271 (T1W), 0.9049 (fluid-sensitive), and 0.9179 (T1W + C). Incorporation of multiparametric data or semantic features did not significantly improve model

View this article online at wileyonlinelibrary.com. DOI: 10.1002/jmri.29691

Received Oct 1, 2024, Accepted for publication Dec 12, 2024.

*Address reprint requests to: L.M.W., JDMI—Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, Canada M5G 1X5. E-mail: lawrence.white@uhn.ca
Hadas Benhabib and Daniel Brandenberger contributed equally as cofirst authors.
Lawrence M. White and Masoom A. Haider contributed equally as co-senior authors.

From the [1]Department of Medical Imaging, University of Toronto, Toronto, Ontario, Canada; [2]Joint Department of Medical Imaging, University Health Network, Sinai Health System, Women's College Hospital, Mount Sinai Hospital, Toronto, Ontario, Canada; [3]Institut für Radiologie und Nuklearmedizin, Kantonsspital Baselland, Liestal, Switzerland; [4]Department of Biostatistics, University Health Network, Toronto, Ontario, Canada; [5]Department of Pathology and Laboratory Medicine, Mount Sinai Hospital, Toronto, Ontario, Canada; [6]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada; [7]Department of Surgery, University of Toronto, Toronto, Ontario, Canada; and [8]University Musculoskeletal Oncology Unit, Division of Orthopedic Surgery, Mount Sinai Hospital, Toronto, Ontario, Canada

Additional supporting information may be found in the online version of this article

performance ($P \geq 0.08$) compared to radiomic models derived from any individual MRI sequence alone.
Data Conclusion: MRI radiomics appears to be accurate in the differentiation of myxomas from myxoid sarcomas. Classification performance did not improve when incorporating qualitative features or multiparametric imaging data.
Plain Language Summary: Accurately distinguishing between benign soft tissue myxomas and malignant myxoid sarcomas is essential for guiding appropriate management but remains challenging with conventional MRI interpretation. This study utilized radiomics, a method that extracts quantitative mathematically derived features from images, to develop predictive models based on routine MRI examination. Analyzing over 500 cases, MRI radiomics demonstrated excellent diagnostic accuracy in differentiating between benign myxomas and malignant myxoid sarcomas, highlighting the potential of the technique, as a powerful non-invasive tool that could complement current diagnostic approaches, and enhance clinical decision-making in patients with soft tissue myxoid tumors of the musculoskeletal system.
Level of Evidence: 3
Technical Efficacy: Stage 2

J. MAGN. RESON. IMAGING 2025;61:2630–2641.

Myxoid soft tissue tumors comprise a heterogenous group of mesenchymal neoplasms, including benign, locally aggressive, and malignant entities, which are characterized by the presence of substantial extracellular myxoid matrix.[1–3] Benign intramuscular myxomas and malignant myxoid soft tissue sarcomas share many overlapping clinical, histopathological, and imaging characteristics, complicating diagnosis. The prominent mucoid stromal component of these tumors results in a notably elevated water content, leading to their hypoechoic presentation on ultrasound, low attenuation on CT, and high signal intensity on T2-weighted MR imaging.[4,5] Signal heterogeneity on MRI, secondary to intrinsic cellular elements, hemorrhage, or necrosis, tends to be associated with malignant sarcomatous lesions.[6,7]

While MR imaging plays an important role in the initial evaluation of a suspected soft tissue tumor, it often lacks the specificity required to differentiate benign myxomas from malignant myxoid soft tissue sarcomas.[8,9] As a result, patients presenting with a myxoid neoplasm at tertiary sarcoma referral centers regularly undergo biopsy to achieve tissue diagnosis.[10] Myxomas are treated conservatively, either with clinical surveillance or marginal excision.[11] In contrast, myxoid sarcomas require aggressive treatment with (neo)adjuvant radiotherapy and negative-margin surgical resection.[12] The distinction between these entities is important for guiding clinicians in tailoring the most effective treatment. More accurate MRI analysis has potential to circumvent the requirement for biopsy and to improve patient care and healthcare resource utilization.[11,13]

Radiomics provides a robust and high-throughput method for investigating a multitude of image-related features, with results that can be applied to create predictive models that have clinically relevant outcomes.[14] Recent studies have highlighted the potential utility of radiomic models as predictive biomarkers of tumor histology and grade, therapeutic response of tumors to neoadjuvant therapies, as well as clinical outcomes including systemic spread of disease, local recurrence, and patient survival.[15,16]

Thus the purpose of this study was to retrospectively investigate the diagnostic accuracy of radiomics-based models derived from pre-treatment MRI examinations, in the differentiation of soft tissue myxomas vs. myxoid sarcomas of the musculoskeletal system.

Methods

Study Cohort

Institutional research ethics board approval was obtained, with waiver of informed consent for this retrospective study. The study cohort was derived from a database of consecutive cases, managed through the University of Toronto Musculoskeletal Oncology Unit (Mount Sinai Hospital, Sinai Health System, Toronto, Canada) between April 2002 and June 2021. Study design and inclusion criteria are detailed in Fig. 1, with the total study cohort comprised of 523 histologically proven cases in 523 unique patients; 201 myxomas (201/523, 38.4%) and 322 myxoid sarcomas (322/523, 61.6%). Demographic details of the study cohort, as well as final histology tissue diagnoses, are summarized in Table 1.

MRI examinations in the 523 patients included 516 T1-weighted (T1W) acquisitions, 518 T2-weighted fat suppressed or short tau inversion recovery (fluid-sensitive) acquisitions, and 391 post-contrast enhanced T1-weighted fat suppressed (T1W + C) acquisitions. A summary of MRI systems on which examinations were performed, pulse sequences from which radiomics features were extracted, and original spatial resolution (voxel volumes) of the acquisitions, based on DICOM header data, are provided in Tables S1 and S2 in the Supplemental Material.

Segmentation

Three-dimensional segmentations of tumors were performed using 3D slicer, version 5.0.3 (https://www.slicer.org/), an open-source image computing platform.[17] Tumors were

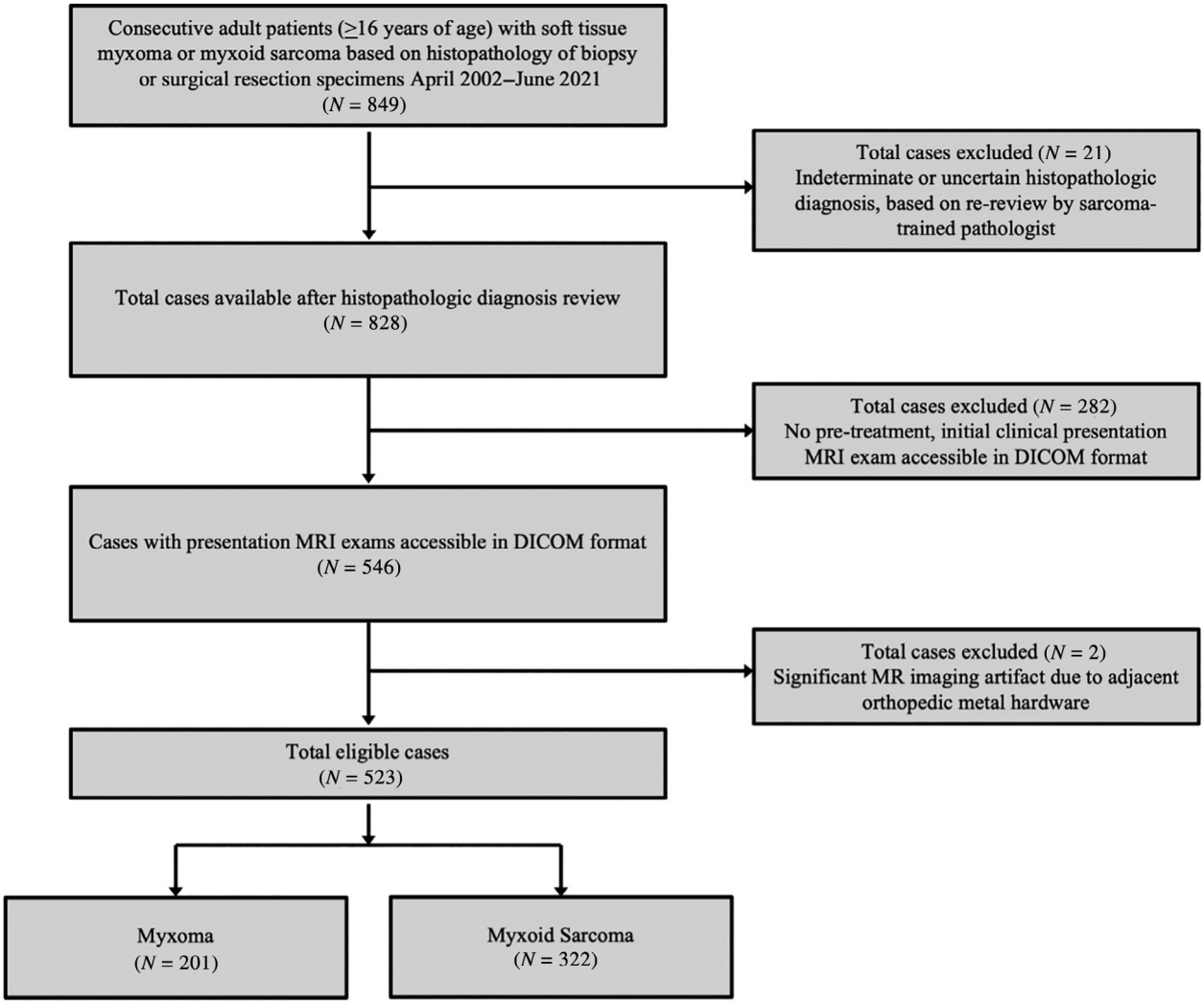

FIGURE 1: Flowchart illustrating the study design and the number of eligible cases. Participants were excluded if they did not have a definitive histopathologic diagnosis of myxoma or myxoid sarcoma, or a presentation MRI available of adequate quality.

manually segmented with lesion tissue delineated on every axial slice of all available T1W, fluid-sensitive, and T1W + C MRI acquisitions. If an axial acquisition was not available, a corresponding coronal or sagittal acquisition, if present, was used for lesion segmentation. Perilesional edema was excluded from segmentation volumes. All operations to generate and manipulate the segmentations were achieved leveraging built-in functions of 3D slicer (Fig. 2). Image interpolation was disabled during lesion segmentation to facilitate segmentation reproducibility and consistency across readers by allowing contouring on original unaltered slices only.

Segmentations of tumors were performed by one of three fellowship trained subspecialist musculoskeletal radiologists; with 5 (HB), 6 (BD), and 35 (L.M.W.) years of experience in musculoskeletal sarcoma imaging.

A randomly selected subgroup of 60 cases (30 myxomas and 30 myxoid sarcomas) were independently segmented by each of the three readers. Readers were blinded to patient information and segmentation results of other readers.

Semantic Feature Assessment

At the time of tumor segmentations, readers assessed a series of semantic features for each tumor on MRI, including tumor location, perilesional fat and edema, osseous or neurovascular involvement, and transcompartmental extension. Each of the three readers independently assessed semantic features in the subgroup of 60 cases (30 myxomas and 30 myxoid sarcomas) randomly selected for segmentation by each reader, with inter-reader correlation assessed using Fleiss kappa analysis. Semantic feature assessments of the most experienced reader (L.M.W.) were used in further statistical analyses. For tumor location, lesions were defined as being situated "deep" or "superficial" to deep fascia, with lesion location defined as deep if any portion of the tumor extended deep to fascia. Tumors were categorized as being either intramuscular or extramuscular in location. The presence/absence of fat abutting any aspect of the tumor margin was categorized on T1W acquisitions and the presence/absence of any perilesional juxta-capsular edema was assessed on fluid-sensitive sequences. The presence/absence

TABLE 1. Demographic and Histopathologic Information for the Study Cohort

	Myxoma ($N = 201$)	Myxoid Sarcoma ($N = 322$)
Gender, *N* (%)		
Female	145 (72.1%)	131 (40.7%)
Male	56 (27.9%)	191 (59.3%)
Age, years		
Mean (SD)	59 (12.57)	57 (18.42)
Range	16–93	16–96
Anatomic location, *N* (%)		
Upper extremity	59 (29.4%)	56 (17.4%)
Shoulder	7 (3.5%)	13 (4.0%)
Arm	39 (19.4%)	23 (7.1%)
Forearm	12 (6.0%)	20 (6.2%)
Wrist/Hand	1 (0.5%)	0
Trunk	44 (21.9%)	48 (14.9%)
Chest wall	5 (2.5%)	22 (6.8%)
Paraspinal	3 (1.5%)	7 (2.2%)
Pelvis/Gluteal	36 (17.9%)	19 (5.9%)
Lower extremity	98 (48.8%)	218 (67.7%)
Thigh	85 (42.3%)	173 (53.7%)
Calf	10 (5.0%)	41 (12.7%)
Ankle/Foot	3 (1.5%)	4 (1.2%)
Histopathologic diagnosis, *N* (%)		
Myxoma	201 (100%)	-
Myxofibrosarcoma	-	201 (62.4%)
Myxoid liposarcoma	-	104 (32.3%)
Extraskeletal myxoid chondrosarcoma	-	12 (3.7%)
Ossifying fibromyxoid tumor	-	4 (1.2%)
Malignant myxoid mesenchymoma	-	1 (0.3%)

SD = standard deviation.

of osseous abutment, defined as direct contiguity of tumor with adjacent osseous structures without intervening normal tissue planes, was assessed.[18] Similarly, the presence/absence of tumor abutment of major neurovascular structures was evaluated, with abutment characterized by loss of normal perineural or perivascular fat planes at locations contiguous with tumor.[19] Lastly, the anatomic location of each tumor was catalogued, as was the presence/absence of transcompartmental tumor extension.

Preprocesssing and Radiomics Feature Extraction

Image preprocessing and radiomics feature extraction were performed using pyRadiomics version 3.0.1 (https://github.com/AIM-Harvard/pyradiomics), compliant with IBSI.[20] In all cases segmented image voxel signal intensities were normalized (scale set to 500) and resampled to an isotropic voxel size of $(1 \times 1 \times 1)$ mm^3 using a b-spline interpolator, for texture feature sets to be rotationally invariant. Gray-scale

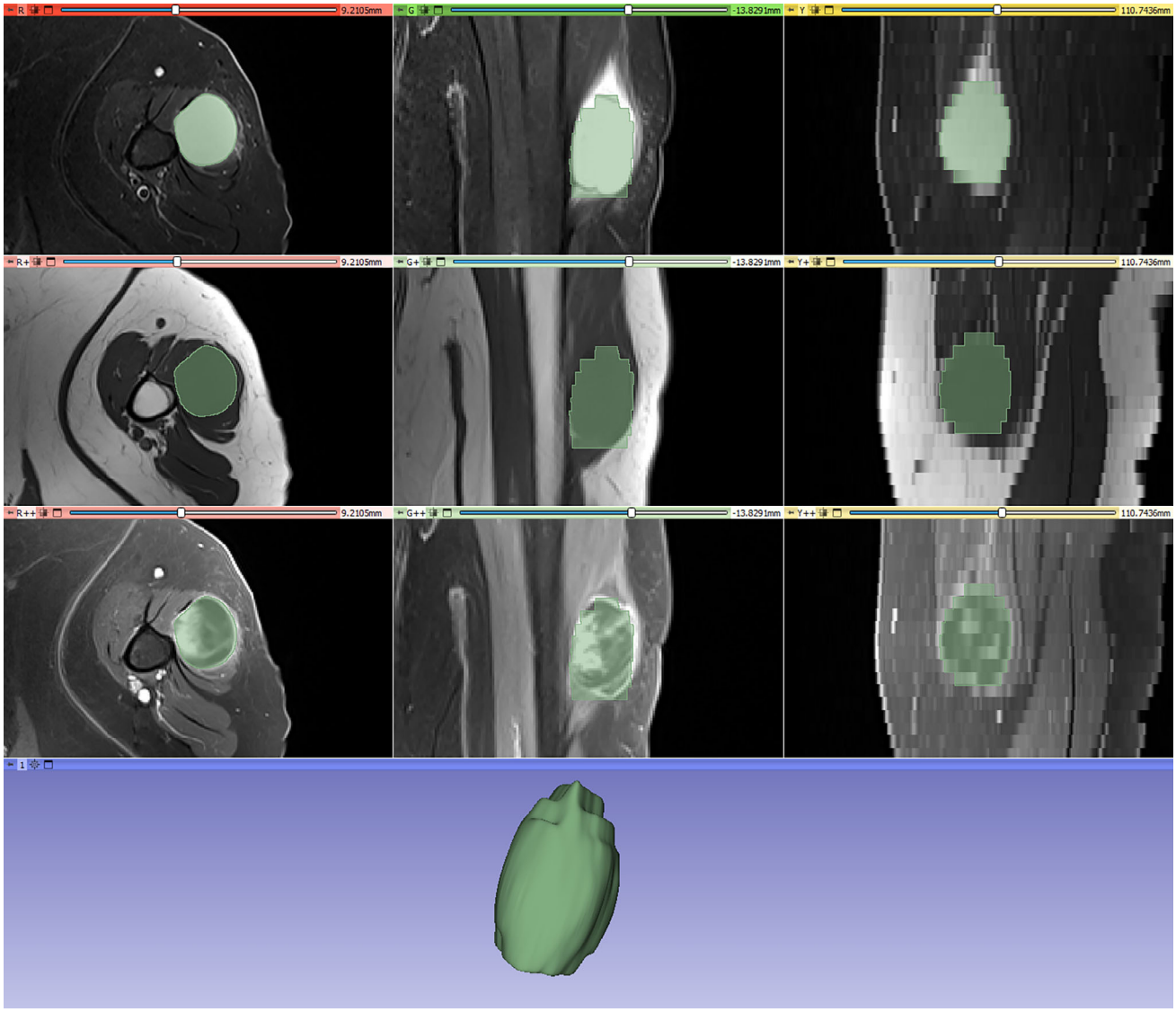

FIGURE 2: Manual segmentation process of a myxoma in 3D slicer using a custom view displaying T2-weighted fat-suppressed (top row), T1-weighted (middle row), and T1-weighted fat-suppressed contrast-enhanced (bottom row), in three planes. Segmentation (marked light green) was performed in the axial plane (left column). Corresponding coronal planes are shown in the middle column. Reformats of the axial slices are displayed (in this example in the sagittal plane, right column), when no original images in the corresponding orthogonal plane were available.

value discretization was set to a fixed bin width of 25. No image harmonization technique was applied.

Hand-crafted radiomic features were extracted from original and filtered images according to pyRadiomics definitions (https://pyradiomics.readthedocs.io/en/latest/features.html). Feature classes included first-order statistics, shape-based (3D), gray-level co-occurrence matrix, gray-level size zone matrix, gray-level run length matrix, neighboring gray tone difference matrix, and gray-level dependence matrix (GLDM). A total of 1688 original and filtered features were extracted per segmentation, with original features used for model generation and analyses. All features were extracted in 3D space.

A list of software versions, preprocessing and extraction settings, and all extracted features is provided as supplemental data (Tables S3–S5 in the Supplemental Material).

Statistical Analysis

Analyses were performed using R v4.4.0 (R Foundation for Statistical Computing). The subgroup of 60 cases (myxoma $N = 30$ and myxoid sarcoma $N = 30$) independently segmented by each of the three readers was employed for assessment of intraclass correlation (ICC). Radiomics features were excluded from consideration for modeling if they had ICC <0.7 and variance <0.01, which included five features for T1W, one feature for fluid-sensitive, and no features for T1W + C acquisition-based models.

Due to incomplete data for certain imaging sequences (T1W, fluid-sensitive and T1W + C), six different models were developed. The first two models included all patients with complete T1W ($N = 516$) and fluid-sensitive ($N = 518$) data, and the remaining four focused on the

subset of patients ($N = 391$) with complete data for all three sequences ($N = 391$). In the subset, a separate model was developed for each image sequence, along with a multiparametric model that combined the three sequences.

To derive the models, the respective datasets were randomly split into training/test subsets using a 70/30 ratio. In the four subset models, the same patients were used in the respective training and test splits.

Within each training set, first a preliminary univariate analysis was conducted by summarizing the distribution of each feature stratified by class. Differences in distribution were assessed using Mann–Whitney U tests. P-values were adjusted for multiplicity using the Bonferroni correction.

Next, to select the radiomics features for inclusion in the final model, the pairwise Spearman correlation coefficient for all features were calculated in the training set. For pairs exhibiting correlation greater than |0.90|, the feature with the highest average absolute correlation with all other remaining features was removed, and the average absolute correlation was re-calculated with each feature removal. The process was repeated until all remaining features had pairwise correlations <|0.90|. The Boruta feature selection algorithm was then applied to the remaining features to identify the features with the highest variable importance for identifying histologic diagnosis (myxoma or myxoid sarcoma).

The selected features were incorporated into a random forest model using the training subset. To select the optimal hyperparameters for the final random forest model, a grid search was used to select the hyperparameter values that maximized the area under the curve (AUC). The hyperparameters included the total number of trees (ntree), ranging from 1000 to 2500 in increments of 50, and the number of features to be sampled as split criteria (mtry) ranging from 2 to 10 in increments of 1. Performance was assessed using 10-fold cross-validation. The final model utilized the hyperparameters that maximized the cross-validated AUC.

The fitted model was utilized to distinguish class in the test set. The radiomics score was defined as the probability of class = myxoma, as calculated by the model. Patients with a probability of ≥0.5 were classified as myxoma, and patients with a probability of <0.5 were classified as myxoid sarcomas. The radiomic score's AUC, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) were estimated by comparing predicted score/class to actual class.

Fleiss' kappa was used to evaluate interobserver agreement in sematic tumor feature assessment, in the subgroup of 60 cases independently assessed by each of the three readers. In the $N = 391$ subset the ability of semantic tumor features (location: deep/superficial, intramuscular/extramuscular; and presence/absence of perilesional fat and edema, osseous and neurovascular abutment, and transcompartmental extension) to differentiate myxoma vs. myxoid sarcoma was investigated with a univariable regression model, with odds ratio (OR) >1 denoting increased probability of myxoma. All semantic features with a P-value of <0.05 were retained for further analysis in two multivariable logistic regression models, the first incorporating the semantic tumor features only, and the second combining semantic tumor features with the radiomics score in a nomographic model which incorporated calculated radiomics scores and statistically significant semantic features. The fitted models were used to predict class in the test set, and each models' AUC was calculated.

In the four subset models, DeLong's test was used to assess differences in AUCs between 1) radiomics vs. radiomics + semantic features models, 2) semantic features vs. radiomics + semantic feature models, and 3) radiomics vs. semantic feature models, using T1W, fluid-sensitive, T1W + C, and multiparametric radiomic models.

Results

Cohort Demographics

Demographic details, source of tissue available for final histologic assessment and breakdown of final histopathologic diagnoses, are summarized in Table 1. Anatomic location of the lesions included the lower extremity in 49% of myxomas and 68% of myxoid sarcomas, the trunk in 22% of myxomas and 15% of myxoid sarcomas, and the upper extremity in 29% of myxomas and 17% of myxoid sarcomas.

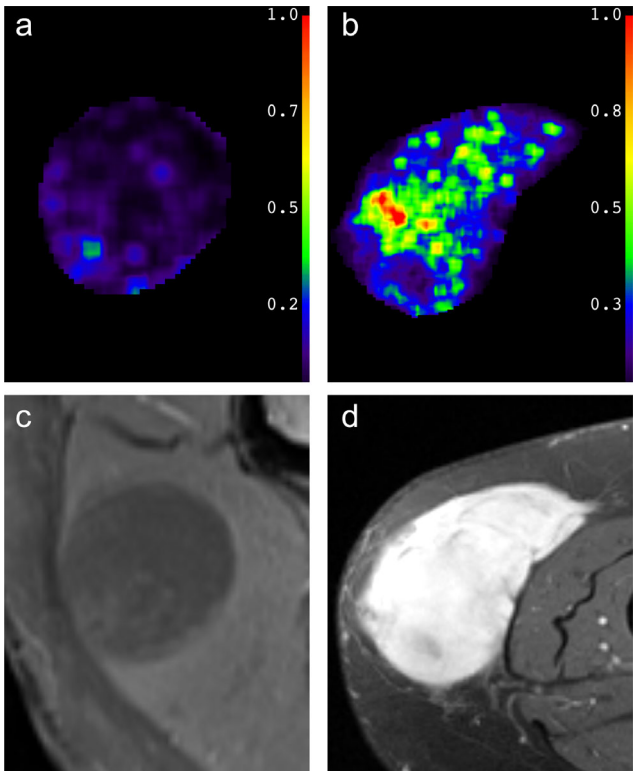

FIGURE 3: Color-scale images of voxel-wise maps of the extracted radiomics feature "original_gldm_large dependence high gray level emphasis" in an intramuscular myxoma (a) and myxofibrosarcoma (b). Corresponding original T1-weighted fat suppressed contrast enhanced images; myxoma (c) and myxofibrosarcoma (d).

Overview of Selected Radiomics Features

Radiomics feature summary statistics for each imaging sequence (T1W, fluid-sensitive, and T1W + C), stratified by diagnosis, are presented in Table S6 in the Supplemental Material. Radiomic features derived from each imaging sequence/model that optimized class prediction were identified (Figs. 3 and 4), with the relative predictive influence of features retained in models fit using the training sets shown in Fig. 4. The correlation between selected features are visualized in Fig. S1 in the Supplemental Material, and the selected hyperparameters for each model are displayed in Table S7 in the Supplemental Material.

T1W and Fluid-Sensitive Complete Models

The complete T1W radiomic model for prediction of class across the test dataset, derived from T1W radiomics features demonstrated the highest test AUC of 0.9271 (sensitivity = 0.815, specificity = 0.903, PPV = 0.830, NPV = 0.894), and the complete fluid-sensitive model had a test AUC of 0.9049 (sensitivity = 0.833, specificity = 0.851, PPV = 0.763, NPV = 0.899). The T1W complete and subset models shared nine radiomics features, and the same seven features were selected for the fluid-sensitive complete and subset models (Fig. 4).

Subset Analysis

Seven qualitative semantic features were assessed on MRI, in each of the soft tissue myxomas and myxoid sarcomas cases (Table 2). Interobserver agreement in sematic tumor feature assessment, in the subgroup of 60 cases independently assessed by each of the three readers, was moderate to almost-perfect for all seven features (Table 3),[21] with semantic feature assessments of the most experienced reader (L.M.W.) used in further statistical analyses. Four out of the seven semantics features demonstrated significant predictive power

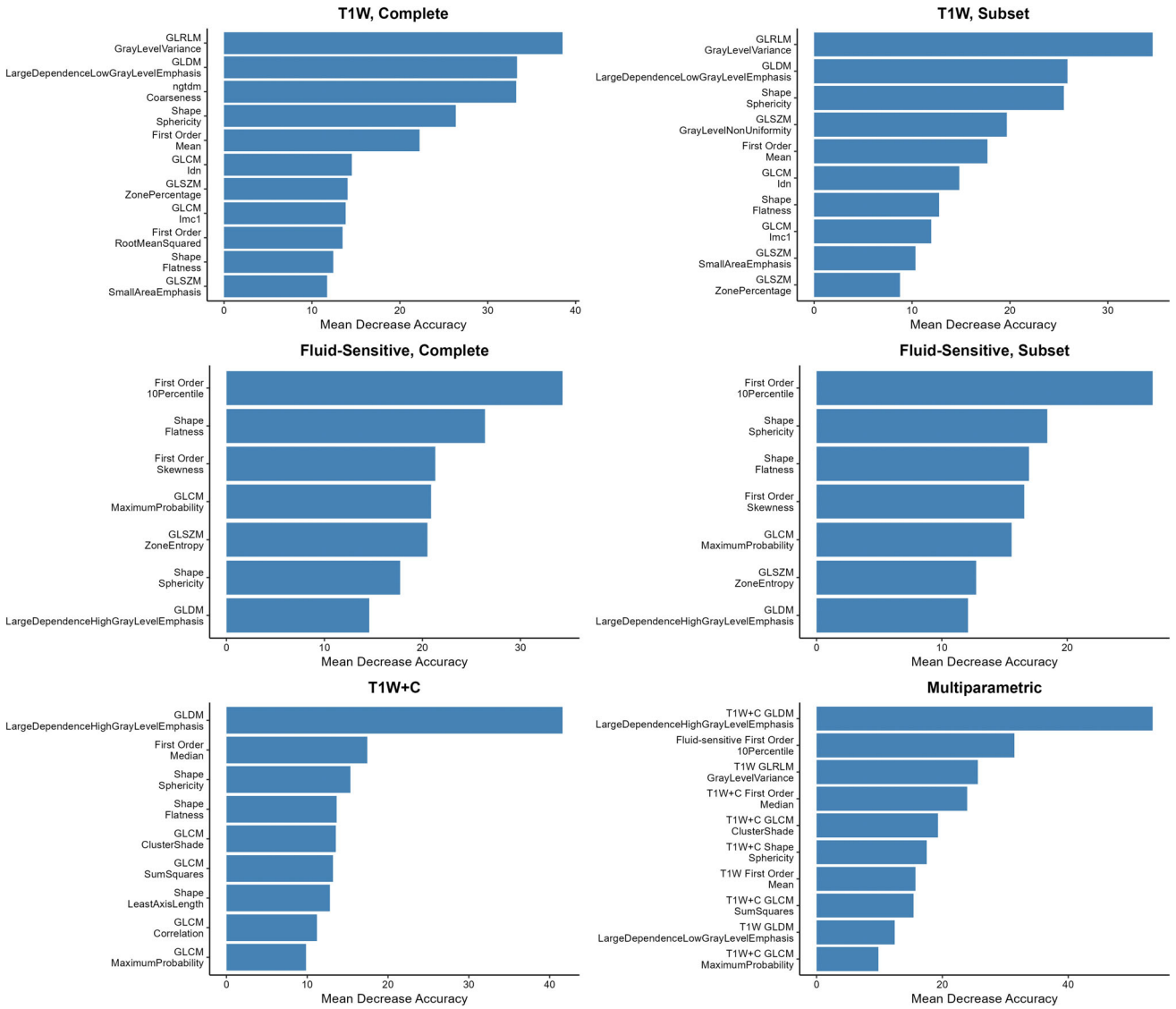

FIGURE 4: Radiomic features that optimized class prediction and were retained in models for each imaging sequence (T1W, fluid-sensitive, T1W + C) and their combination (multiparametric) with plots of their respective impact on model accuracy.

TABLE 2. Univariable Logistic Regression Model Results for the Semantic Tumor Features Restricted to the $N = 391$ Subset

	OR (95% CI)	*P*
Location relative to fascia		
Superficial	Reference	<0.001*
Deep	11.39 (4.46, 38.64)	
Extra-/Intramuscular		
Extramuscular	Reference	<0.001*
Intramuscular	9.06 (4.92, 17.73)	
Perilesional edema		
No	Reference	0.057
Yes	2.70 (1.04, 8.38)	
Perilesional fat		
No	Reference	0.38
Yes	0.55 (0.13, 2.12)	
Major NV bundle		
Abutment/Invasion	Reference	0.045*
Clear	2.07 (1.04, 4.36)	
Osseous extension		
Clear	Reference	0.33
Abutment/Invasion	1.31 (0.76, 2.23)	
Transcompartmental extension		
No	Reference	<0.001*
Yes	0.07 (0.01, 0.24)	

An odds ratio >1 denotes increased odds of myxoma.
*Features that were statistically significant ($P < 0.05$) in the prediction of tumor type.

by univariable analysis, in determination of tumor type (myxoma vs. myxoid sarcoma), including location deep to fascia, intramuscular, transcompartmental extension, and neurovascular abutment.

The subset T1W radiomic model had a test AUC of 0.9196 (sensitivity = 0.870, specificity = 0.889, PPV = 0.851, NPV = 0.903), the subset fluid-sensitive MRI radiomic model a test AUC of 0.9089 (sensitivity = 0.826, specificity = 0.794, PPV = 0.745, NPV = 0.862) and the T1W + C MRI radiomic model demonstrated a test AUC of 0.9179 (sensitivity = 0.804, specificity = 0.937, PPV = 0.902, NPV = 0.868). A multiparametric model based on radiomic features derived from all three sequences, did not exhibit significantly improved discrimination for tumor classification compared to any individual sequence on their own (T1W $P = 0.37$, fluid-sensitive $P = 0.21$, and T1W + C $P = 0.08$). All possible pairwise combinations were evaluated using DeLong's test, with no radiomic model (T1W, fluid-sensitive, T1W + C or multiparametric) demonstrating significant discriminatory superiority over another (Table 4). For each model, Figure S1 in the Supplemental Material displays correlation heat maps for all retained features.

Nomographic models, derived from the addition of the four qualitative semantic features which demonstrated significant tumor-type predictive power, were investigated to determine whether there was a significant impact on class discrimination. AUCs from radiomic models alone, and nomographic models (incorporating radiomic and semantic features) were compared, with no significant improvement in AUC identified for any sequence (T1W $P = 0.37$, fluid-sensitive $P = 0.96$, T1W + C $P = 0.67$, and multiparametric $P = 0.88$) (Fig. 5). Furthermore, radiomic models significantly outperformed semantic features for T1W, T1W + C and multiparametric combined, with fluid-sensitive approaching statistical significance ($P = 0.054$) (Table 5).

Discussion

The current study focused on utilizing MRI radiomic models in the differentiation of myxoid sarcomas from myxomas. The results showed that radiomic models derived from T1W, fluid-sensitive (T2-FS or STIR), and T1W + C sequences all demonstrated excellent discrimination between benign myxomas and malignant myxoid sarcomas with AUCs greater than 0.9. T1-weighted imaging demonstrated the highest AUC, while also exhibiting high sensitivity and specificity and very high NPV. A multiparametric radiomics-based model, combining all three MRI sequences, also showed excellent differentiation between tumor types, but did not significantly outperform any of the individual sequence radiomic models.

Medical imaging, in particular MRI, plays an important role in the diagnostic assessment of soft tissue tumors including the identification of prognostic factors such as size, location, degree of tumor necrosis, and involvement of local-regional structures such as major nerves, vessels, and bone.[22,23] Classically, myxomas have been described as well-defined intramuscular lesions, with markedly high intralesional signal intensity and increased signal intensity of adjacent muscle, due to edema, on T2-weighted fat suppressed or short inversion time inversion-recovery sequences.[2] The slow-growing nature of these tumors has also been hypothesized to cause atrophy of surrounding muscle tissue with associated fatty infiltration, resulting in the prominent rind of fat often associated with myxomas.[4,24] However, distinguishing benign myxomas from malignant myxoid

TABLE 3. Fleiss Kappa Estimates and Concordance for Semantic Feature Assessments in the 60 Cases Evaluated Independently by Three Different Readers

Semantic Feature	Fleiss Kappa (95th CI)	Concordance
Superficial	0.76 (0.55–0.92)	54/60 (90%)
Extramuscular	0.90 (0.79–0.98)	56/60 (93.3%)
Perilesional edema	0.59 (0.02–1.0)	58/60 (96.7%)
Perilesional fat	−0.02 (−0.03–0)	57/60 (95%)
NV bundle abutment	0.84 (0.62–0.96)	56/60 (93.3%)
Osseous abutment	0.92 (0.78–1.0)	58/60 (96.7%)
Transcompartmental extension	1.0 (1–1)	60/60 (100%)

CI = confidence interval.

sarcomas can be challenging due to overlapping MR imaging features.[3,10]

Four out of seven semantic features evaluated in this study had significant predictive power by univariable regression analysis, in the differentiation between benign from malignant myxoid tumors. Specifically, myxomas were more likely to be intramuscular, and located deep to fascia. Myxoid-sarcomas, on the other hand, were prone to abut major neurovascular bundles and demonstrate transcompartmental extension. Interestingly, perilesional fat, which is classically used to characterize myxomas,[2] was a feature present in nearly all myxomas and myxoid-sarcomas. While qualitative semantic features can be promising predictors of diagnosis, models solely based on these features significantly underperformed compared to radiomics-based models in differentiating myxomas and myxoid sarcomas. Furthermore, models combining semantic features and radiomics features did not improve diagnostic performance in tumor discrimination. This finding is in concordance with a recent study investigating the association between qualitative semantic features and quantitative radiomics features in other malignancies, including non-small-cell lung cancer[25] and spinal metastases.[26]

Radiomics offers a novel data-driven method of mining a myriad of quantitative features from traditional tomographic images.[14,16] Radiomics-based models, either in isolation or in combination with patient data, can be applied to various musculoskeletal tumor outcome categories, such as response to treatment, patient survival, histologic grade of tumors, presence of molecular tumor markers, and systemic spread of disease.[27] Osteosarcomas represent one of the most extensively investigated group of tumors, with studies such as that of Wu et al establishing a prognostic non-invasive nomogram to predict survival at the time of diagnosis through the combination of both clinical risk factors and CT radiomics features,[28] and that of White et al employing radiomic modeling for prediction of histologic necrosis following neoadjuvant chemotherapy and patient survival outcomes based on initial presentation MRI.[29] The present study follows the precedent set by established literature that has focused on utilizing radiomics in the differentiation of soft tissue lesions, most commonly lipomatous tumors.[27]

In the current study, GLDM features, which are a subset of second order features reflective of image texture,[30] were determined to have the greatest influence in optimizing class prediction and were the most important category of pyradiomics features influencing model accuracy for T1W + C and multiparametric data, and ranked second for T1W models. While not the most important feature influencing the accuracy of the fluid-sensitive model, it was still significant, ranking seventh. GLDM, along with other texture features, are utilized to quantify and describe the heterogeneity and spatial pattern of intervoxel gray scale dependencies within an image, with changes in GLDM indicative of variations in tissue heterogeneity.[31] While there is considerable overlap in MR imaging features of myxoid tumors, studies have shown that higher-grade myxoid sarcomas more commonly exhibit heterogeneous signal compared to myxomas, which tend to be blander in cellularity.[1] This observed signal heterogeneity in malignant tumors may be due to the variable and often more complex components, such as intratumoral hemorrhage and necrosis.[1,10]

There are few studies investigating the discriminating MR imaging features of myxoid soft-tissue tumors, with many being either case reports or reviews. A prior study by Martin-Carreras et al performed radiomic analyses in a relatively small number of cases (myxoma $N = 29$ and myxofibrosarcoma $N = 27$) to distinguish between myxoma and myxofibrosarcoma.[32] The authors only performed radiomic analysis based on non-enhanced T1-weighted MRI sequences, and achieved an AUC of 0.885. However, a number of shortcomings of this study were noted, including a lack

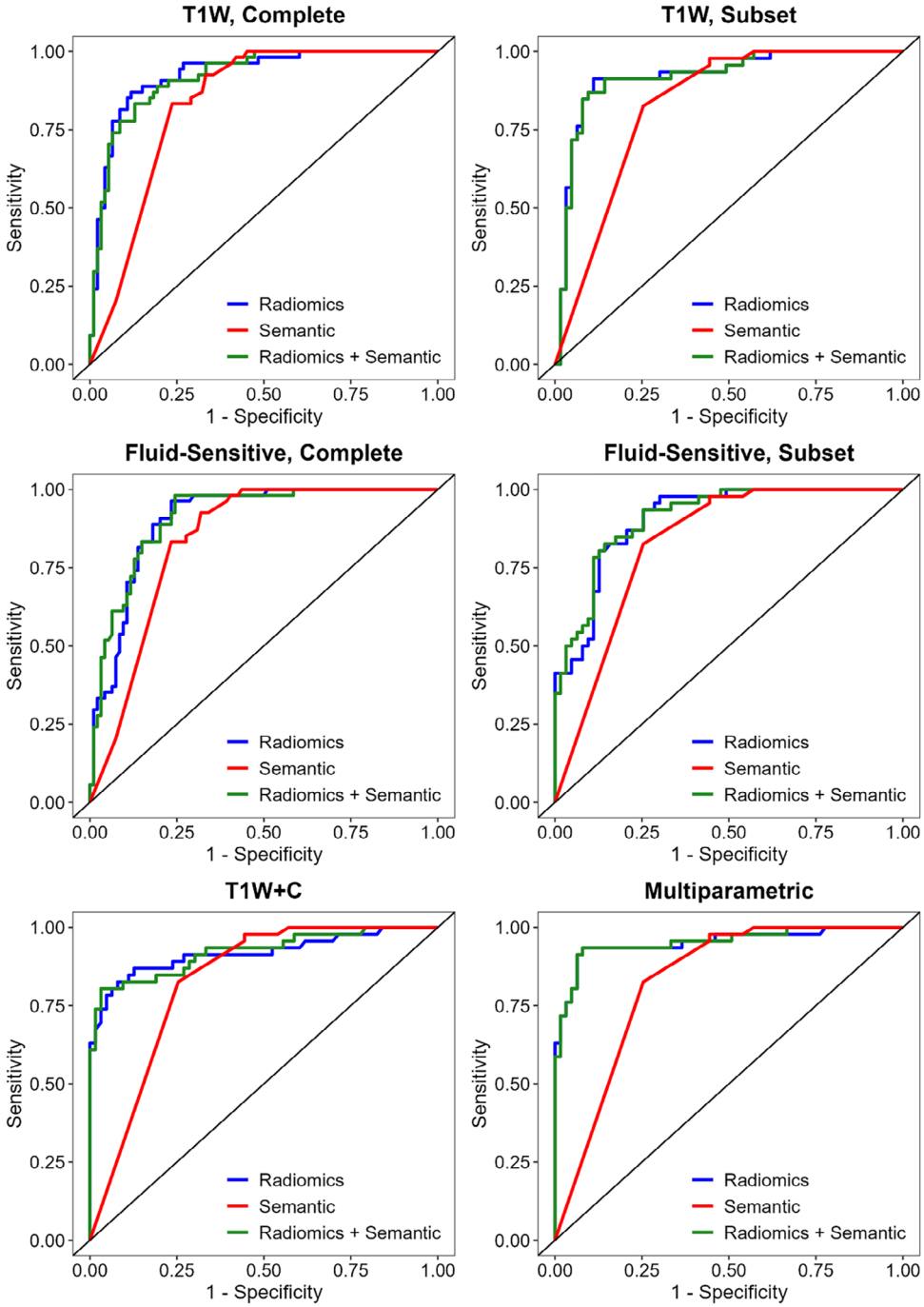

FIGURE 5: Receiver operating characteristic (ROC) curves of all models based on radiomic features, qualitative semantic features, and combinations of both radiomic and qualitative semantic features.

of training and validation test subsets, in addition to small study numbers that can severely limit artificial intelligence (AI)-based analyses due to model-overfitting.[14]

Integrating refined quantitative data-driven MRI radiomics analysis into the care pathway decision-making process of patients with musculoskeletal soft tissue tumors may be beneficial. The distinction between myxomas and myxoid sarcomas is clinically significant, as it directly influences patient management and choice of surgical approach and treatment strategies for patients. Accurate MRI radiomics based models, such as those provided in the current study, potentially provide a means of non-invasively differentiating the intrinsic histopathologic heterogeneity of myxoid lesions, often elusive to conventional imaging visual analysis and interpretation, and potentially avoiding the necessity of tissue sampling for diagnostic confirmation of benign disease.

Limitations

This study was performed at a tertiary referral center and the available pretreatment MRI examinations were heterogeneous

TABLE 4. DeLong Test Analysis of All Pairwise Combinations of MRI Sequence Radiomic Models

Pairwise Comparison	*P*
Multiparametric vs. T1W	0.37
Multiparametric vs. T1W + C	0.08
Multiparametric vs. fluid-sensitive	0.21
T1W vs. fluid-sensitive	0.79
T1W vs. T1W + C	0.97
T1W + C vs. fluid-sensitive	0.83

T1W = model based on radiomics features extracted from T1-weighted MRI acquisitions ($N = 391$ subset). Fluid-sensitive = model based on radiomics features extracted from T2-weighted fat suppressed, or short-tau inversion recovery, MRI acquisitions ($N = 391$ subset). T1W + C = model based on radiomics features extracted from T1-weighted fat-suppressed post gadolinium contrast enhancement, MRI acquisitions ($N = 391$ subset). Multiparametric = model based on radiomics features extracted from T1W, fluid-sensitive, and T1W + C MRI acquisitions ($N = 391$ subset).

TABLE 5. DeLong Test Analysis of Pairwise Combinations of MRI Sequence Radiomic Models, Nomographic Models, and Qualitative Semantic Features Models, All Within the *N* = 391 Subset

Pairwise Comparison	*P*
T1W vs. semantic features	0.01*
T1W vs. T1W + semantic features	0.37
Fluid-sensitive vs. semantic features	0.05
Fluid-sensitive vs. fluid-sensitive + semantic features	0.97
T1W + C vs. semantic features	0.04*
T1W + C vs. T1W + C + semantic features	0.67
Multiparametric vs. semantic features	≤0.001*
Multiparametric vs. multiparametric + semantic features	0.88

*Statistical significance of $P < 0.05$.

in nature. Many examinations were performed at outside imaging centers, with variable MRI system field strengths, vendors, models, operating software, pulse sequences, and sequence acquisition parameters. We did not evaluate or employ image harmonization techniques to address potential variables in MR imaging, including field strength, on extracted pyradiomic features and resultant radiomic models. While such heterogeneity could be seen as a constraint in the development of optimal radiomic modeling, heterogeneous MRI datasets help to address concerns and limitations of deep learning and artificial intelligence models developed using highly uniform homogeneous datasets obtained from a single center, and improve validity when applied to external datasets. This heterogeneity accurately reflects the variation seen in real-world practice where MR imaging of soft tissue tumors is often performed at different centers prior to referral to tertiary centers of expertise in musculoskeletal sarcoma care.[33] Inclusion of cases with MRI examinations performed at different sites also facilitated the analysis of a larger patient population, which is important for the accurate and reliable training and validation of radiomic and machine learning models.[34]

A drawback to studying a large population, however, is the substantial effort and time required to manually delineate and contour each tumor, which in the case of the current study required over 1600 individual manual 3D volumetric tumor segmentations. Auto-segmentation algorithms have rapidly evolved, offering high-throughput processing and enhancing the scalability of radiomic studies.[35] They provide several benefits such as improving consistency and reproducibility by minimizing interobserver variability, thereby ensuring reliable results across different radiologists and studies.[36,37] However, these algorithms may struggle with complex anatomic structures or subtle boundaries, impacting accuracy.[36] Further opportunities in follow-up to the current study include external validation of findings with large multi-institutional datasets and prospective clinical studies to further support translation to use in routine clinical practice.

Conclusion

This study demonstrated that MRI radiomics-based models based on routine clinical T1W, fluid-sensitive or T1W + C MRI acquisitions may accurately classify myxomas and myxoid sarcomas with excellent discrimination.

Author Contributions

Hadas Benhabib and Daniel Brandenberger were involved in conceptualizing the research question; experimental design and methodology; data collection, analysis and interpretation; as well as manuscript draft preparation, revision, and approvals. Lawrence M. White and Masoom A. Haider were involved in conceptualizing the research question; experimental design and methodology; data collection, analysis and interpretation; manuscript preparation, revision, and approvals; and providing senior oversight and guidance throughout the project.

References

1. van Roggen G, Hogendoorn MF. Myxoid tumours of soft tissue. Histopathology 1999;35:291-312.
2. Petscavage-Thomas JM, Walker EA, Logie CI, Clarke LE, Duryea DM, Murphey MD. Soft-tissue myxomatous lesions: Review of salient imaging features with pathologic comparison. Radiographics 2014;34: 964-980.
3. Baheti AD, Tirumani SH, Rosenthal MH, et al. Myxoid soft-tissue neoplasms: Comprehensive update of the taxonomy and MRI features. Am J Roentgenol 2015;204:374-385.
4. Murphey MD, McRae GA, Fanburg-Smith JC, Temple HT, Levine AM, Aboulafia AJ. Imaging of soft-tissue myxoma with emphasis on CT and MR and comparison of radiologic and pathologic findings. Radiology 2002;225:215-224.
5. Luna A, Martinez S, Bossen E. Magnetic resonance imaging of intramuscular myxoma with histological comparison and a review of the literature. Skeletal Radiol 2005;34:19-28.
6. Ma LD, McCarthy EF, Bluemke DA, Frassica FJ. Differentiation of benign from malignant musculoskeletal lesions using MR imaging: Pitfalls in MR evaluation of lesions with a cystic appearance. Am J Roentgenol 1998;170:1251-1258.
7. Bermejo A, De Bustamante TD, Martinez A, Carrera R, Zabía E, Manjón P. MR imaging in the evaluation of cystic-appearing soft-tissue masses of the extremities. Radiographics 2013;33:833-855.
8. Chang H, Kang Y, Ahn JM, Lee E, Lee JW, Kang HS. Texture analysis of magnetic resonance image to differentiate benign from malignant myxoid soft tissue tumors: A retrospective comparative study. PLoS One 2022;17:e0267569.
9. Harish S, Lee JC, Ahmad M, Saifuddin A. Soft tissue masses with "cyst-like" appearance on MR imaging: Distinction of benign and malignant lesions. Eur Radiol 2006;16:2652-2660.
10. de Boer HC, Musson R. Imaging features of myxoid soft-tissue tumours. Clin Radiol 2023;78:635-643.
11. Reiter A, Trumm K, Ballhause TM, et al. Diagnostic and therapeutic pathways of intramuscular myxoma. Diagnostics 2022;12:1573.
12. Mutter RW, Singer S, Zhang Z, Brennan MF, Alektiar KM. The enigma of myxofibrosarcoma of the extremity. Cancer 2012;118:518-527.
13. Lunn BW, Littrell LA, Wenger DE, Broski SM. 18F-FDG PET/CT and MRI features of myxoid liposarcomas and intramuscular myxomas. Skeletal Radiol 2018;47:1641-1650.
14. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. Radiology 2016;278:563-577.
15. Crombé A, Spinnato P, Italiano A, et al. Radiomics and artificial intelligence for soft-tissue sarcomas: Current status and perspectives. Diagn Interv Imaging 2023;104:567-583.
16. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: The bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14:749-762.
17. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the quantitative imaging network. Magn Reson Imaging 2012;30:1323-1341.
18. Elias DA, White LM, Simpson DJ, et al. Osseous invasion by soft-tissue sarcoma: Assessment with MR imaging. Radiology 2003;229:145-152.
19. Panicek DM, Go SD, Healey JH, Leung DHY, Brennan MF, Lewis JJ. Soft-tissue sarcoma involving bone or neurovascular structures: MR imaging prognostic factors. Radiology 1997;205:871-875.
20. Zwanenburg A, Vallières M, Abdalah MA, et al. The image biomarker standardization initiative: Standardized quantitative radiomics for high-throughput image-based phenotyping. Radiology 2020;295:328-338.
21. Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics 1977;33:159-174.
22. Kransdorf MJ, Murphey MD. *Imaging of soft tissue tumors: Third edition*, Philadelphia: Lippincott Williams & Wilkins; 2014.
23. Balach T, Stacy GS, Haydon RC. The clinical evaluation of soft tissue tumors. Radiol Clin North Am 2011;49:1185-1196.
24. Bancroft LW, Kransdorf MJ, Menke DM, O'Connor MI, Foster WC. Intramuscular myxoma: Characteristic MR imaging features. Am J Roentgenol 2002;178:1255-1259.
25. Yip SSF, Liu Y, Parmar C, et al. Associations between radiologist-defined semantic and automatically computed radiomic features in non-small cell lung cancer. Sci Rep 2017;7:3519.
26. Liu K, Zhang Y, Wang Q, et al. Differentiation of predominantly osteolytic from osteoblastic spinal metastases based on standard magnetic resonance imaging sequences: A comparison of radiomics model versus semantic features logistic regression model findings. Quant Imaging Med Surg 2022;12:5004-5017.
27. Brandenberger D, White LM. Radiomics in musculoskeletal tumors. Semin Musculoskelet Radiol 2024;28:49-61.
28. Wu Y, Xu L, Yang P, et al. Survival prediction in high-grade osteosarcoma using radiomics of diagnostic computed tomography. EBioMedicine 2018;34:27-34.
29. White LM, Atinga A, Naraghi AM, et al. T2-weighted MRI radiomics in high-grade intramedullary osteosarcoma: Predictive accuracy in assessing histologic response to chemotherapy, overall survival, and disease-free survival. Skeletal Radiol 2023;52:553-564.
30. Mayerhoefer ME, Materka A, Langs G, et al. Introduction to radiomics. J Nucl Med 2020;61:488-495.
31. Varghese BA, Cen SY, Hwang DH, Duddalwar VA. Texture analysis of imaging: What radiologists need to know. Am J Roentgenol 2019;212: 520-528.
32. Martin-Carreras T, Li H, Cooper K, Fan Y, Sebro R. Radiomic features from MRI distinguish myxomas from myxofibrosarcomas. BMC Med Imaging 2019;19:67.
33. Rowbotham E, Bhuva S, Gupta H, Robinson P. Assessment of referrals into the soft tissue sarcoma service: Evaluation of imaging early in the pathway process. Sarcoma 2012;2012:781723.
34. Demircioğlu A. Are deep models in radiomics performing better than generic models? A systematic review. Eur Radiol Exp 2023;7:11.
35. Liu Z, Wang S, Dong D, et al. The applications of radiomics in precision diagnosis and treatment of oncology: Opportunities and challenges. Theranostics 2019;9:1303-1322.
36. Cardenas CE, Yang J, Anderson BM, Court LE, Brock KB. Advances in auto-segmentation. Semin Radiat Oncol 2019;29:185-197.
37. Heye T, Merkle EM, Reiner CS, et al. Reproducibility of dynamic contrast-enhanced MR imaging part II. Comparison of intra- and inter-observer variability with manual region of interest placement versus semiautomatic lesion segmentation and histogram analysis. Radiology 2013;266:812-821.