



Predicting Soft Tissue Sarcoma Response to Neoadjuvant Chemotherapy Using an MRI-Based Delta-Radiomics Approach

Brandon K. K. Fields¹ · Natalie L. Demirjian² · Steven Y. Cen³ · Bino A. Varghese³ · Darryl H. Hwang³ · Xiaomeng Lei³ · Bhushan Desai³ · Vinay Duddalwar³ · George R. Matcuk Jr.⁴ 

Received: 27 October 2022 / Revised: 14 January 2023 / Accepted: 16 January 2023 / Published online: 25 January 2023
© The Author(s) 2023

Abstract

Objectives To evaluate the performance of machine learning–augmented MRI-based radiomics models for predicting response to neoadjuvant chemotherapy (NAC) in soft tissue sarcomas.

Methods Forty-four subjects were identified retrospectively from patients who received NAC at our institution for pathologically proven soft tissue sarcomas. Only subjects who had both a baseline MRI prior to initiating chemotherapy and a post-treatment scan at least 2 months after initiating chemotherapy and prior to surgical resection were included. 3D ROIs were used to delineate whole-tumor volumes on pre- and post-treatment scans, from which 1708 radiomics features were extracted. Delta-radiomics features were calculated by subtraction of baseline from post-treatment values and used to distinguish treatment response through univariate analyses as well as machine learning–augmented radiomics analyses.

Results Though only 4.74% of variables overall reached significance at $p \leq 0.05$ in univariate analyses, Laws Texture Energy (LTE)-derived metrics represented 46.04% of all such features reaching statistical significance. ROC analyses similarly failed to predict NAC response, with AUCs of 0.40 (95% CI 0.22–0.58) and 0.44 (95% CI 0.26–0.62) for RF and AdaBoost, respectively.

Conclusion Overall, while our result was not able to separate NAC responders from non-responders, our analyses did identify a subset of LTE-derived metrics that show promise for further investigations. Future studies will likely benefit from larger sample size constructions so as to avoid the need for data filtering and feature selection techniques, which have the potential to significantly bias the machine learning procedures.

Key words Sarcoma · Soft tissue · Magnetic resonance imaging · Radiomics · Machine learning · Neoadjuvant chemotherapy

Key Points

- Our decision classifiers constructed using machine learning–augmented MRI-based radiomics data were not able to separate neoadjuvant chemotherapy responders from non-responders in a cohort of soft tissue sarcomas, with AUCs of 0.40 (95% CI 0.22–0.58) and 0.44 (95% CI 0.26–0.62) for RF and AdaBoost, respectively.
- Our univariate analyses revealed that 46.04% of features reaching statistical significance at the $p \leq 0.05$ level were derived from Laws Texture Energy (LTE)-based computations, which is in line with existing literature suggesting a promising role for spatial filtering metrics in identifying features of tumor heterogeneity.
- Though frequently reported in the literature, we advocate against the routine use of feature reduction and data filtering methods in radiomics studies as these methods are highly prone to introducing bias when working with modern machine learning algorithms.

This manuscript is based on Scientific Exhibit No. E1500 presented at the 2021 Annual Meeting of the American Roentgen Ray Society.

Extended author information available on the last page of the article

Abbreviations

STS	Soft tissue sarcoma
RECIST	Response Evaluation Criteria In Solid Tumors
NAC	Neoadjuvant chemotherapy
RF	Random forest
AdaBoost	Real adaptive boosting
AUC	Area under the curve
ROC	Receiver operating characteristic
LTE	Laws Texture Energy

Introduction

Assessment of treatment response in soft tissue sarcomas (STS) by conventional radiologic imaging has long posed unique set of challenges for clinicians [1–4]. Owing to their highly variable internal compositions, tumors undergoing a

biologic response to chemotherapy may not actually diminish in size due to factors such as cystic degeneration, hyalinization, fibrosis, centralized necrosis, and intratumoral hemorrhage, all of which have the potential to affect estimations of whole-tumor volume [2, 3, 5–10]. Thus, appraisals of treatment response that depend on evaluations of tumor size—including the World Health Organization (WHO) response evaluation criteria and the oft-cited Response Evaluation Criteria In Solid Tumors (RECIST)—may fail to appreciate satisfactory biologic response to chemotherapy in tumors that do not demonstrate macroscopic shrinkage on radiologic imaging [2, 3, 7, 8, 11–16]. The Choi criteria and modified Choi criteria, which were later proposed in an effort to incorporate additional features such as changes in attenuation or signal intensity on CT or MRI, were shown to better correlate with pathologic response [7, 8, 14, 17–19]. However, the Choi criteria were notably not originally designed for STS and still rely heavily on size-based estimations, thus calling into question in their ability to accurately resolve complex architectural changes in STS, particularly in cases of synovial sarcoma [4, 7, 20]. In the age of targeted molecular therapies, there exists a growing need for modernized response criteria that more accurately reflect the scope of phenotypic heterogeneity [6, 13, 21, 22].

Radiomics is defined as the conversion of medical imaging into multi-dimensional mineable data for clinical decision support to bolster accurate diagnosis, prognostication, and prediction of treatment response [4, 23–28]. In comparison with standard biopsy techniques, radiomics analysis offers the advantage of being able to non-invasively quantify heterogeneity of entire tumor volumes at given time points of interest, which in theory should allow for better characterization of chemotherapeutic response than use of size-based criteria alone [6, 20–22, 26, 27, 29–31]. Radiomics has already been successfully applied to a variety of clinical applications related to STS, including stratification of benign from malignant soft tissue neoplasms, prediction of histologic grade, and assessment of metastatic risk [27, 31–34], though lack of standardized protocols has hindered widespread adoption of radiomics workflows in clinical practice [16, 24, 25, 35].

Standard-of-care typically encourages the use of anthracycline-based regimens as first-line chemotherapy in patients with newly diagnosed STS, which have demonstrated improved overall and metastasis-free survival in phase 3 clinical trials [2, 3, 29, 30]. Yet, ongoing research in sarcoma care remains limited in part due to the previously detailed shortcomings of traditional size-based response criteria, which calls into question their appropriateness for use as endpoints in clinical trials [2, 3, 20]. Thus, ongoing collaborations between leading agencies including the US Food and Drug Administration and the US National Cancer Institute have since called for the validation of quantitative imaging

techniques to serve as surrogate biomarkers, as these may in fact more accurately reflect early biological changes in tumor physiology [6, 13, 20, 22]. In a previous pilot study [20], we were able to demonstrate that quantitative-MRI (q-MRI) evaluation of enhancing tumor volume was able to accurately stratify responders from non-responders in a small cohort of patients with histopathologically diagnosed STS treated with standard-of-care neoadjuvant chemotherapy (NAC). Therefore, based on studies correlating intratumoral heterogeneity on radiologic imaging with higher histologic grade and poorer patient outcomes [5, 23, 24, 29, 32, 33, 36], we hypothesized that change from baseline of radiomics metrics taken pre- and post-NAC (i.e., delta-radiomics) might be able to better predict response to NAC in STS. While a small body of evidence does suggest a role for radiomics-based predictive modeling in stratifying response to neoadjuvant therapy [30, 34, 37, 38], these studies are not repeatable due to the application of data filtering and feature reduction techniques prior to model training and cross-validation, which has been shown to bias model performance [28, 39]. Thus, we further aimed to investigate whether these previous findings could be replicated under more rigorous test conditions.

Materials and Methods

This single-center retrospective study was approved by our university's institutional review board. The requirement for informed consent was waived due to the retrospective nature of our data collection.

Study Participants

A total of 44 subjects (mean age 53.70 years; range 16–80 years) who received NAC at our institution for histologically diagnosed STS were included in this study. Enrollments were restricted to only subjects who had both a baseline MRI obtained prior to initiation of NAC and a post-treatment MRI obtained at least 2 months following NAC initiation and prior to surgical resection. Patients were identified by chart review of cases discussed at our institution's Orthopedic and Sarcoma Tumor Boards from January 2010 to January 2017. All 44 subjects were previously reported as part of a related study investigating the utility of radiomics analysis in stratifying benign and malignant soft tissue neoplasms [27]. Seven of these subjects were additionally included in previous pilot studies [1, 20]. In our cohort, the most common pathologic diagnoses were undifferentiated pleomorphic sarcoma ($n=17$), synovial sarcoma ($n=6$), myxoid liposarcoma ($n=4$), and leiomyosarcoma ($n=4$) (Table 1). Lesions were most often encountered in the thigh ($n=21$), followed by the arm ($n=4$), pelvis/buttock ($n=4$), and calf ($n=3$).

Table 1 Summary features of neoplasms analyzed for machine learning-augmented radiomics analysis

Characteristic	All tumors (<i>n</i> = 44)
Age (years)*	53.70 ± 16.05 (16–80)
Sex	
Males	19/44 (43.18%)
Females	25/44 (56.82%)
Anatomic location	
Shoulder	1/44 (2.27%)
Arm	4/44 (9.09%)
Forearm	1/44 (2.27%)
Wrist	1/44 (2.27%)
Groin	2/44 (4.55%)
Hip	1/44 (2.27%)
Pelvis/buttock	4/44 (9.09%)
Thigh	21/44 (47.23%)
Calf	3/44 (6.82%)
Knee	1/44 (2.27%)
Ankle	1/44 (2.27%)
Foot	2/44 (4.55%)
Retroperitoneum	2/44 (4.55%)
Histologic subtypes	
Undifferentiated pleomorphic sarcoma	17/44 (38.64%)
Dedifferentiated liposarcoma	2/44 (4.55%)
Myxoid liposarcoma	4/44 (9.09%)
Pleomorphic liposarcoma	1/44 (2.27%)
Leiomyosarcoma	4/44 (9.09%)
Myxofibrosarcoma	2/44 (4.55%)
Embryonal rhabdomyosarcoma	1/44 (2.27%)
Pleomorphic rhabdomyosarcoma	1/44 (2.27%)
Extraskeletal Ewing sarcoma	2/44 (4.55%)
Extraskeletal osteosarcoma	2/44 (4.55%)
Malignant peripheral nerve sheath tumor	2/44 (4.55%)
Synovial sarcoma	6/44 (13.64%)
Histologic grade	
High	41/44 (93.18%)
Low	3/44 (6.82%)
Surgical margins	
Positive	26/44 (59.09%)
Marginal	7/44 (15.91%)
Negative	6/44 (13.64%)
Not specified	5/44 (11.36%)
NAC response	
Responder	23/44 (52.27%)
Non-responder	21/44 (47.73%)

all data are presented as numerators and denominators with percentages in parentheses unless otherwise specified

*Data presented as mean age in years ± standard deviation with range reported in parentheses

NAC neoadjuvant chemotherapy

Sequence Acquisitions

Two MRI studies were analyzed per subject for a total of 88 scans, 37 of which were acquired at our institution and 51 of which were acquired at outside facilities. A total of 29 institutions contributed to the acquisition pool, which featured radiomics data extracted from 11 unique MRI sequences (Table 2). All studies were uploaded to and available for review through our institution's PACS at the time of subject enrollment.

Volumetric Segmentations and Radiomics Data Extraction

The workflow for image segmentation and radiomics data extraction has been previously described in detail [27]; briefly, images were loaded onto server-deployed Synapse 3D software (Fujifilm Medical Systems), after which tumor volumes were manually delineated on one sequence of interest. 3D regions of interest (ROIs) were then transferred onto additional sequences of interest from the same MRI study following sequence co-registration using statistical parametric mapping (SPM) software [40] (Fig. 1). Subsequently, 1708 radiomics features were extracted from the 3D-ROIs using MATLAB® (MathWorks) software running our comprehensive institutional radiomics pipeline, which has been rigorously benchmarked against an Image Biomarkers Standardization Initiative (IBSI) phantom and reference values [41] (Fig. 2). Delta values were then calculated from the extracted radiomics features. Generally speaking, delta-radiomics capture either the change or the percent change in radiomics features across different points in time [4, 30, 34, 37, 38]. The quantification of delta change over time in our study was thus calculated as post-NAC minus pre-NAC feature numeric values, from which subsequent analyses and machine learning algorithms were derived.

Statistical Analyses

Univariate comparisons were performed using independent *t*-test or Wilcoxon rank sum test depending on data normality. Descriptive analyses included mean, standard deviation, and interquartile ranges displayed as box plots. Benjamini–Hochberg Procedure was used to adjust for multiple comparisons errors [42]. The percentages of radiomics features with unadjusted and Benjamini–Hochberg adjusted $p \leq 0.05$ and $p \leq 0.01$ within each radiomics family were calculated as an assessment of the overall signal strength of a given family.

Two machine learning algorithms, namely random forest (RF) and real adaptive boosting (AdaBoost), were trained with the aim of distinguishing between responders and

Table 2 Overview of MRI scans obtained at our institution and at outside facilities broken down by plane of section

Sequence	Total (<i>n</i> = 88)	Acquired at our institution (<i>n</i> = 37)	Acquired at an outside facility (<i>n</i> = 51)
T1			
Axial	85.23%	97.30%	76.47%
Coronal	13.64%	2.70%	21.57%
Sagittal	–	–	–
T1 post-contrast			
Axial	5.68%	2.70%	7.84%
Coronal	2.27%	–	3.92%
Sagittal	–	–	–
T1 fat-saturated			
Axial	63.64%	97.30%	39.22%
Coronal	2.27%	–	3.92%
Sagittal	1.14%	–	1.96%
T1 fat-saturated post-contrast			
Axial	84.09%	100.00%	72.55%
Coronal	2.27%	–	3.92%
Sagittal	1.14%	–	1.96%
STIR			
Axial	56.82%	75.68%	43.14%
Coronal	18.18%	8.11%	25.49%
Sagittal	10.23%	10.81%	9.80%
T2			
Axial	13.64%	13.51%	13.73%
Coronal	3.41%	–	5.88%
Sagittal	6.82%	–	11.76%
T2 fat-saturated			
Axial	15.91%	2.70%	25.49%
Coronal	3.41%	–	5.88%
Sagittal	1.14%	2.70%	–
PD			
Axial	6.82%	2.70%	9.80%
Coronal	3.41%	–	5.88%
Sagittal	1.14%	–	1.96%
PD fat-saturated			
Axial	22.73%	8.11%	33.33%
Coronal	–	–	–
Sagittal	4.55%	8.11%	1.96%
HASTE			
Axial	–	–	–
Coronal	1.14%	–	1.96%
Sagittal	–	–	–
FIESTA			
Axial	1.14%	–	1.96%
Coronal	–	–	–
Sagittal	–	–	–

all data are presented as percentages of MRI scans for which the indicated sequence was available for analysis. Unless otherwise specified, all sequences were obtained prior to contrast administration. *STIR* short-tau inversion recovery, *PD* proton density, *HASTE* half-Fourier acquisition single-shot turbo spin-echo, *FIESTA* fast imaging employing steady-state acquisition

non-responders using delta-radiomics values [43]. Both RF and AdaBoost are decision tree-based methods that are robust to non-normal distributions, missing data, and outliers, though RF in particular has performed exceptionally well in related radiomics studies [44–46]. Model performance was evaluated using a tenfold cross-validation. *K*-fold cross-validation is a commonly employed validation technique in radiomics studies [4, 27, 30, 31, 34, 47, 48], as the systematic formulation of multiple training and testing cohorts renders many unanticipated confounders essentially inert. For the purposes of our study, the full dataset was first divided equally into 10 folds. Subsequently, the learning process was re-iterated 10 times, during which a given classifier was applied in turn to each of the testing samples. In this way, each study sample was allowed to serve as an independent test case.

The machine learning models were constructed as previously described [27]. In the case of RF, the model was constructed using 800 trees with a leaf size of 16. Maximal depth was set at 50. The square root of the variable number was taken as the variable to try. Given that AdaBoost is comparatively more efficient, only 25 trees with a maximal depth of 3 were used during model construction [49]. For both models, Gini impurity index served as the loss function. Prior correction as described by King et al. was used to adjust for imbalanced outcomes [50]. The accuracies of the predictive models were quantitatively assessed by taking the areas under the curve (AUCs) of the receiver operating characteristic (ROC) curves generated from the predicted probabilities of the 10 testing datasets combined.

Variables of importance were selected and ranked using out-of-bag Gini index. The cut-off for top performing variables of importance was taken to be the “cliff” of the out-of-bag Gini ranking, i.e., a sudden large change from previous ranking position. The variables of importance selection procedure were repeated 10 times, with the final ranking based on the sum of out-of-bag Gini rankings over the tenfold cross-validation.

SAS Enterprise Miner 15.1: High-Performance Procedures were used for machine learning. SAS v9.4 statistical computing software was used for all other statistical analysis.

Results

Univariate analyses revealed that only 4.74% (*n* = 265) of variables showed significant differences in delta-radiomics metrics at the $p \leq 0.05$ level between NAC responders vs. non-responders. Though only a small percentage of metrics overall showed statistical significance, an increased representation of Laws Texture Energy (LTE)-derived features was notably observed as compared to other texture families,

accounting for 46.04% ($n=122$) of all features reaching statistical significance at the $p \leq 0.05$ level (Fig. 3). Likewise, only 1.34% ($n=75$) of variables showed statistically significant differences at the $p \leq 0.01$ level. Concordantly, both machine learning methods failed to predict NAC response by ROC analysis, with AUCs of 0.40 (95% CI 0.22–0.58) and 0.44 (95% CI 0.26–0.62) for RF and AdaBoost, respectively (Fig. 4).

As demonstration of proof of concept, we repeated our machine learning procedure on 2 restricted datasets of pre-selected radiomics features filtered by univariate p -values. When run on a restricted dataset of only features found to be significant at the $p \leq 0.05$ level, RF and AdaBoost yielded AUCs of 0.74 (95% CI 0.59–0.89) and 0.75 (95% CI 0.60–0.89), respectively. Similarly, when run on a restricted dataset of only features found to be significant at the $p \leq 0.01$ level, RF and AdaBoost yielded AUCs of 0.78 (95% CI 0.64–0.92) and 0.82 (95% CI 0.70–0.95), respectively. This exercise was conducted solely for the purposes of comparison and discussion and is not routinely recommended by the authors as a method of bolstering machine learning results.

Discussion

MRI remains the preeminent method for serial evaluation of STS progression and treatment response [4, 9, 19, 37, 38]. Yet, it has been clear for some time that historical definitions of treatment response—which have tended to rely on size-based criteria—are severely lacking, as they often fail to account for non-dimensional changes in tumors exhibiting a biologic response to NAC [2–4, 7, 8, 20, 30]. In this study, we investigated the utility of a delta-radiomics approach to predict chemotherapeutic response in STS by assessing for temporal changes in the radiomics feature parameters of MRI scans taken pre- and post-NAC. In contrast to previously published findings [30], we do not find an ability for decision classifiers constructed using an MRI-based delta-radiomics approach to accurately predict treatment response in STS treated with NAC.

Despite a concerted push for the validation of novel therapeutic biomarkers in oncologic care [6, 21, 22], few studies have thus far investigated the utility of MRI-based radiomics features to serve as surrogate predictors of neoadjuvant response in STS [4] (Table 3). To the best of our knowledge, only one previously published study by Crombé et al. similarly utilized an MRI-based delta-radiomics approach for predicting treatment response specifically to NAC. In their procedure, the authors calculated the absolute change in 33 radiomics features in 65 patients with STS following anthracycline-based NAC, from which only a subset of pre-selected delta features was used to train 4 decision classifiers [30]. Likewise, though Peeken et al., Gao et al., and

Miao et al. all suggest an ability for delta-radiomics-based decision classifiers to predict STS response to radiotherapy [34, 37, 38], these studies also employed feature reduction or recalling techniques prior to model training. While data filtering has become an unfortunately common practice to address high dimensionality in radiomics datasets, these approaches have the potential to induce information leakage. Information leakage further leads to disruption of test data independency, thereby resulting in problems of overfitting [28, 39]. We demonstrate these phenomena explicitly through the results of our filtered analyses, whereby restricting our machine learning inputs to only variables which were significant at the $p \leq 0.05$ and $p \leq 0.01$ levels in our univariate analyses yielded comparable AUCs to those reported by Crombé et al., Peeken et al., Gao et al., and Miao et al. [30, 34, 37, 38].

Publication bias has emerged as a growing area of concern among radiomics studies. As recently as 2018, Buvat et al. reported that a mere 6% of all PET radiomics studies in the published literature explicitly reported negative results [51]. Moreover, in a systematic review of 52 sarcoma-specific radiomics studies, Crombé et al. found that no studies specifically described negative findings [36], further highlighting the need for more balanced publication practices within the field. As discussed above, our result was not able to reproduce separation of neoadjuvant responders from non-responders using machine learning augmented MRI-based radiomics analyses [30, 34, 37, 38]. We believe this is in large part due to our more rigorous approach to our machine learning methodologies without reliance on data filtering and feature selection techniques featured in related works [27, 28, 39]. In particular, Crombé et al. even further report that they constructed their models by first selecting one feature per category and then increasing the number of included features in a “forward stepwise fashion” as determined by univariate p -values [30]. Such steps are not only unnecessary but actually bias and invalidate the results of modern machine learning approaches such as RF—which was notably their top performing classifier—as these algorithms are designed to work with high dimensionality datasets without pre-selection of so-called candidate features [27, 47, 52].

One other notable aspect of our study’s methodology was our inclusion of scans from multiple image acquisition centers. Issues with reproducibility in radiomics studies has garnered progressively more attention in recent years, as it has become increasingly clear that radiomics-based machine learning procedures based on single-center, single-vendor datasets generalize poorly to multicentric data pools [36, 48, 53, 54]. Moreover, as we have discussed in our prior work [27], databases derived from single-center cohorts are poorly reflective of modern clinical practice models [29, 48]. Thus, our study is in line with literature supporting the use

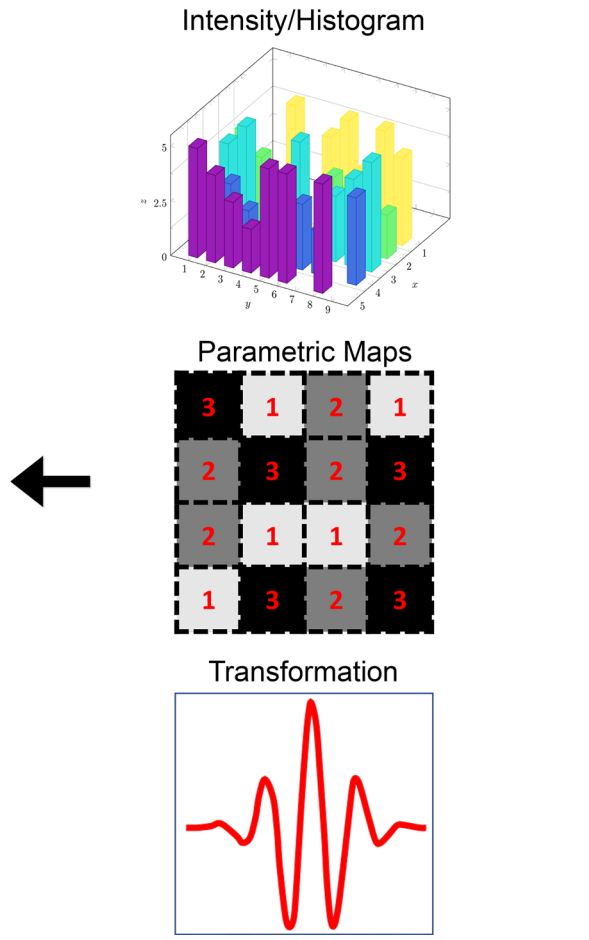
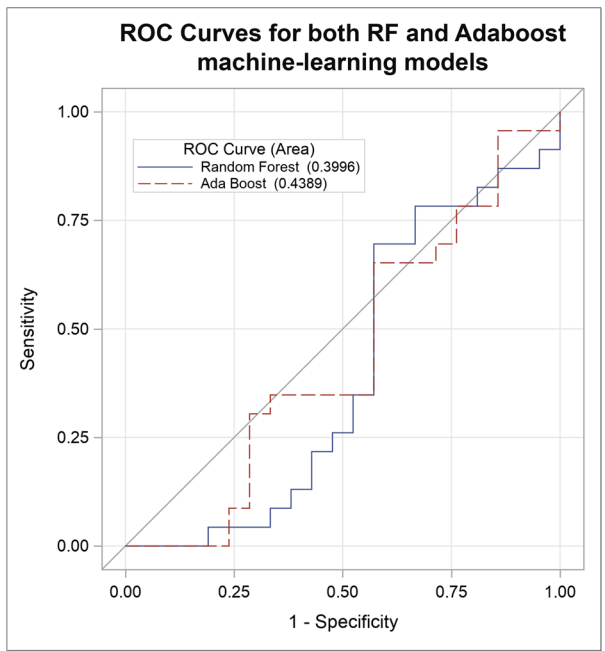
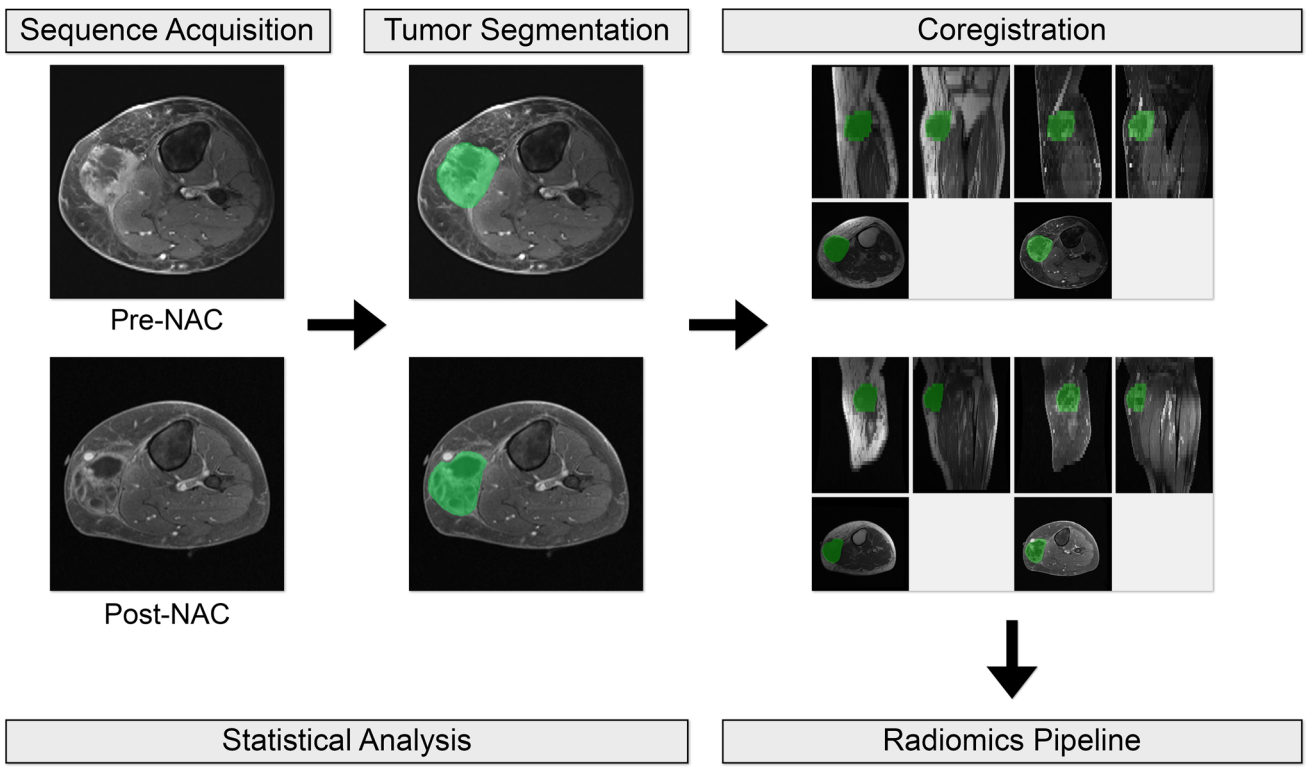


Fig. 1 Flowchart illustrating the workflow for radiomics feature extraction and machine learning–augmented analysis. Example depicts a 50-year-old female with undifferentiated pleomorphic sarcoma of the left calf pre- and post-NAC. Following manual segmentation of 3D-ROIs, data were then extracted from co-registered sequences of interest via our institutional radiomics pipeline. Machine learning models using random forest (RF) and real adaptive boosting (AdaBoost) methods were constructed using a tenfold cross-validation procedure. NAC, neoadjuvant chemotherapy; ROI, region of interest; RF, random forest; AdaBoost, real adaptive boosting.

of multicentric datasets in radiomics studies [16, 26, 27, 36, 53, 55], which theoretically would help mitigate confounding effects of signal noise introduced as a result of heterogeneity in acquisition parameters.

Finally, though the results of our machine learning process failed to reach overall statistical significance, we do note an increased representation of LTE-derived metrics in the univariate analyses, with 46.04% of all metrics reaching statistical significance at the $p \leq 0.05$ level deriving from LTE-based computations. LTE-based measures belong to a group of spatial filtering techniques that reflect the properties of $n \times n$ -sized “convolution kernels” [56–58]. Using this method, spatial domain filters are generated from the vector products of one-dimensional convolution masks, each representing a different texture feature [58]. In the case of our institutional

radiomics pipeline, LTE-based metrics accounted for 1472 individual radiomics features out of a total of 5585 features extracted from 9 separate texture families during the course of this study. This subset of our findings do support previously published data suggesting that spatial filtering techniques are well-suited to detect features indicative of tumor heterogeneity [26, 27], possibly as a consequence of more completely capturing voxel-to-voxel variation through the creation of neighborhood-based matrices [56, 58].

Our study was limited by several factors. First, while our study population was similar in size and composition to the cohort reported on by Crombé et al. [30], it is possible that our study was underpowered to detect a significant result, whereby 100 subjects is often regarded as the threshold sample size for radiomics studies [23]. Although feature selection can theoretically lower the cohort threshold size, we feel that routine use of these procedures should generally be avoided in radiomics studies for reasons as discussed thoroughly above. Thus, given the relative rarity of STS in the general population, multi-institutional collaborations may be necessary in future studies to accrue adequate sample sizes [4, 5, 10, 20, 27, 32, 37, 48]. Second, the retrospective nature of our data collection poses a risk for selection bias given that our subjects were screened for enrollment eligibility from a

Radiomics panel							
Transformation Filters (2D)	Histogram analysis (2D)	DCT (4 bands) (2D)	GLCM (2D)	GLDM (2D)	Fractal analysis (2D)	Grey-level run length matrix (2D)	Grey-level size zone matrix (2D)
<ul style="list-style-type: none"> Gabor Law Wavelets FFT Entropy of FFT Magnitude Entropy of FFT Phase Complexity Index	<ul style="list-style-type: none"> Mean Median Skewness Kurtosis Minimum Maximum Quartile Range Standard deviation 	Haralick Metrics <ul style="list-style-type: none"> Angular Second Moment Contrast Correlation Dissimilarity Entropy Homogeneity Inverse Difference Moment Mean IMC 1 and 2 Mean Measure of Correlation Coefficient Mean Square root of Variance Standard Deviation Uniformity Variance Difference Average and Entropy Sum of Average, Entropy, Variance Mean 			Fractal dimension Lacunarity Succolarity	<ul style="list-style-type: none"> Short-run emphasis (RE) Long-RE Low-grey RE High-grey RE Short run low-grey RE Short run high-grey RE Long run low-grey RE Long run high-grey RE Gray-level non-uniformity Run-length non-uniformity Run Percent 	<ul style="list-style-type: none"> Small Area Emphasis Large Area Emphasis Intensity Variability Size Zone Variance Zone Percentage Low Intensity Emphasis High Intensity Small Area Emphasis High Intensity Small Area Emphasis Low Intensity Large Area Emphasis High Intensity Large Area Emphasis

Fig. 2 Diagrammatic representation of features extracted by our institutional radiomics pipeline. In total, our radiomics pipeline was engineered to extract 1708 radiomics features from 9 distinct texture families. Reprinted by permission from Springer Nature Customer Service

Centre GmbH: Springer Nature European Radiology, [27] (Whole-tumor 3D volumetric MRI-based radiomics approach for distinguishing between benign and malignant soft tissue tumors, Fields et al., European Society of Radiology (2021).

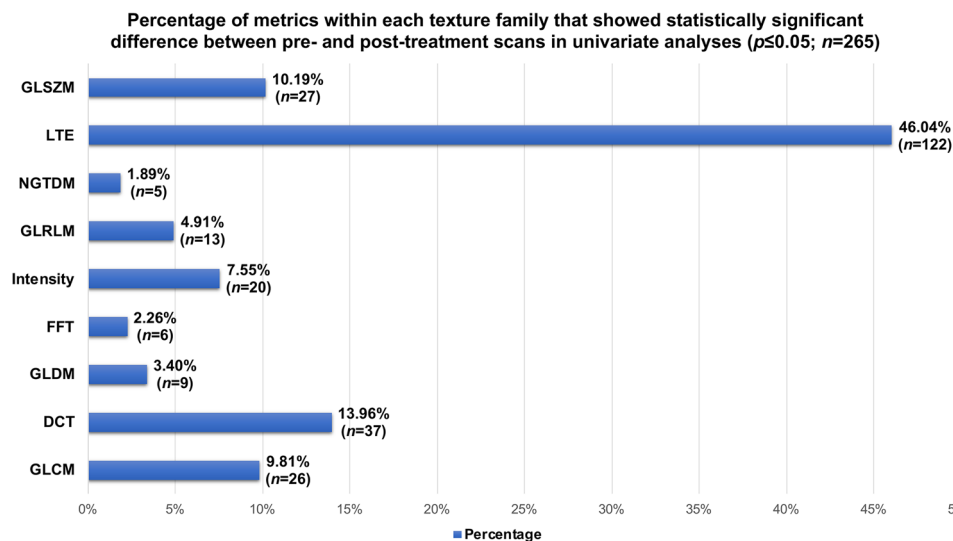


Fig. 3 Bar plots illustrating the percentage of radiomics features that were statistically significant at the $p \leq 0.05$ level broken down by texture family. Overall, only 4.74% of all variables reached statistical significance in univariate analyses; however, analyses did reveal increased representation of LTE-derived features. GLSZM, gray-level

size zone matrix; LTE, Laws Texture Energy; NGTDM, neighboring gray tone difference matrix; GLRLM, gray-level run length matrix; FFT, fast Fourier transform; GLDM, gray-level dependence matrix; DCT, discrete cosine transform; GLCM, gray-level co-occurrence matrix.

larger pool of cases discussed at our institution's Orthopedic and Sarcoma Tumor Boards [59]. Third, though efforts are currently being made to standardize post-acquisition harmonization techniques [10, 24, 31, 60, 61], such applications lack general consensus regarding proper implementation and execution [26, 27]. Furthermore, while post-processing data harmonization techniques such as ComBat have shown some ability to ameliorate scanner and protocol variabilities in multicentric studies, such batch adjustment methods have limitations when used in small sample sizes and rely on stringent data distribution assumptions. [62]. Thus, these methods were of limited applicability to our dataset given concerns for adverse effects due to outliers as well as missing and skewed data distributions. Future efforts to validate post-processing methods aimed at mitigating signal instability across heterogeneous acquisition parameters will undoubtedly aid in the construction of large, multicentric datasets for future research. Additional future directions may also include focused studies correlating delta-radiomics changes with histologic subtype and histopathologic findings of percent necrosis, as well as those specifically focused on stratifying post-treatment changes related to specific chemotherapeutic regimens.

In conclusion, though our machine learning analyses did not show statistically significant separation of NAC responders from non-responders, we were able to identify increased representation of LTE-derived metrics in univariate analyses. These and other spatial filtering metrics may pose a promising area for future radiomics research due to

ROC Curves for both RF and Adaboost machine-learning models

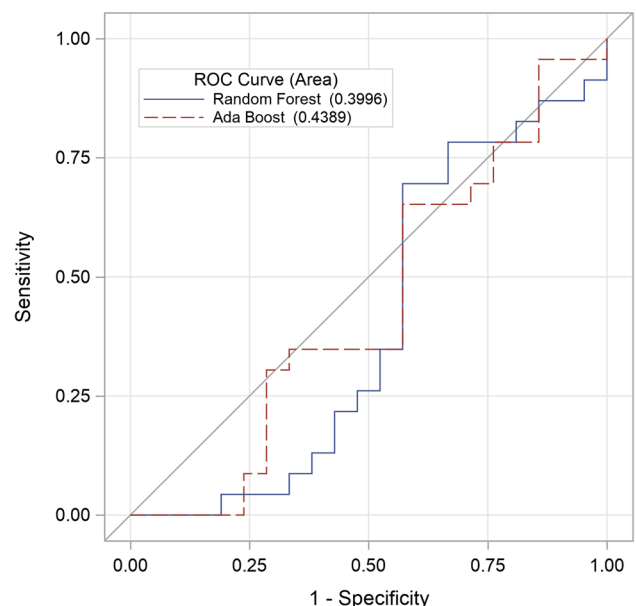


Fig. 4 Evaluation of prediction accuracies of the machine learning-augmented decision classifiers using ROC curve analysis. Prediction accuracy was evaluated by AUC. ROC, receiver operating characteristic; AUC, area under the curve; RF, random forest; AdaBoost, real adaptive boosting.

their ability to more accurately reflect subtle variations in the imaging grayscale [26, 27, 56, 58]. Larger sample sizes in future cohorts are warranted so as to obviate the need

Table 3 Summary of characteristics of comparator studies

First author [reference no.]	Year	Intervention	Sample size	Predictive models investigated	Validation procedure	Highest AUC reported
Crombé [30]	2019	NAC	<i>n</i> = 65	RF, support vector machine, k-nearest neighbors, logistic regression	10-fold stratified cross-validation	0.86
Peeken [34]	2021	Neoadjuvant RT with or without NAC	<i>n</i> = 156	RF, LogitBoost, elastic net regression	3-fold nested cross-validation	0.75
Gao [37]	2020	Neoadjuvant RT	<i>n</i> = 30	Support vector machine, logistic regression	5-fold cross-validation	0.91
Miao [38]	2022	Neoadjuvant RT with or without multi-targeted TKI	<i>n</i> = 30	Logistic regression	None	0.92

NAC neoadjuvant chemotherapy, RT radiotherapy, AUC area under the curve, RF random forest, TKI tyrosine kinase inhibitor

for data reductive techniques, which carry with them an inherent risk of introducing information leakage and thus biasing the decision classifiers [28, 39].

Acknowledgements The authors would like to thank Robert Fields CPA, MBA, for his assistance with restructuring the data output for interpretation and reporting. We thank the Radiological Society of North America's Research & Education Foundation for their support and funding of our work.

Funding Open access funding provided by SCELC, Statewide California Electronic Library Consortium This study was funded by the Radiological Society of North America Research Medical Student Grant RMS#1909.

Declarations

Ethics Approval This study was approved by the University of Southern California Institutional Review Board.

Conflict of Interest GRM is a consultant for Canon Medical Systems, USA. VD is a consultant for Radmetrix and Westat and serves on the advisory board for DeepTek. The authors declare that they have no other disclosures.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References


- Gulati M, Hu JS, Desai B, Hwang DH, Grant EG, Duddalwar VA (2015) Contrast-enhanced sonography for monitoring neoadjuvant chemotherapy in soft tissue sarcomas. *J Ultrasound Med* 34(8):1489–1499. <https://doi.org/10.7863/ultra.34.8.1489>
- Schuetze SM (2005) Imaging and response in soft tissue sarcomas. *Hematol Oncol Clin North Am* 19(3):471–487,vi. <https://doi.org/10.1016/j.hoc.2005.03.001>
- Schuetze SM, Baker LH, Benjamin RS, Canetta R (2008) Selection of response criteria for clinical trials of sarcoma treatment. *Oncologist* 13(Suppl 2):32–40. <https://doi.org/10.1634/theoncologist.13-S2-32>
- Spinnato P, Kind M, Le Loarer F, Bianchi G, Colangeli M, Sambri A, Ponti F, van Langevelde K, Crombe A (2021) Soft tissue sarcomas: the role of quantitative MRI in treatment response evaluation. *Acad Radiol*. <https://doi.org/10.1016/j.acra.2021.08.007>
- Baheti AD, O'Malley RB, Kim S, Keraliya AR, Tirumani SH, Ramaiya NH, Wang CL (2016) Soft-tissue sarcomas: an update for radiologists based on the revised 2013 World Health Organization Classification. *AJR Am J Roentgenol* 206(5):924–932. <https://doi.org/10.2214/AJR.15.15498>
- Kurland BF, Gerstner ER, Mountz JM, Schwartz LH, Ryan CW, Graham MM, Buatti JM, Fennessy FM, Eikman EA, Kumar V, Forster KM, Wahl RL, Lieberman FS (2012) Promise and pitfalls of quantitative imaging in oncology clinical trials. *Magn Reson Imaging* 30(9):1301–1312. <https://doi.org/10.1016/j.mri.2012.06.009>
- Stacchiotti S, Collini P, Messina A, Morosi C, Barisella M, Bertulli R, Piovesan C, Dileo P, Torri V, Gronchi A, Casali PG (2009) High-grade soft-tissue sarcomas: tumor response assessment—pilot study to assess the correlation between radiologic and pathologic response by using RECIST and Choi criteria. *Radiology* 251(2):447–456. <https://doi.org/10.1148/radiol.2512081403>
- Stacchiotti S, Verderio P, Messina A, Morosi C, Collini P, Lombart-Bosch A, Martin J, Comandone A, Cruz J, Ferraro A, Grignani G, Pizzamiglio S, Quagliuolo V, Picci P, Frustaci S, Dei Tos AP, Casali PG, Gronchi A (2012) Tumor response assessment by modified Choi criteria in localized high-risk soft tissue sarcoma treated with chemotherapy. *Cancer* 118(23):5857–5866. <https://doi.org/10.1002/cncr.27624>
- Patel DB, Matcuk GR Jr (2018) Imaging of soft tissue sarcomas. *Chin Clin Oncol* 7(4):35
- Kalisvaart GM, Bloem JL, Bovee J, van de Sande MAJ, Gelderblom H, van der Hage JA, Hartgrink HH, Krol ADG, de Geus-Oei LF, Grootjans W (2021) Personalising sarcoma care using quantitative multimodality imaging for response assessment. *Clin Radiol* 76(4):313.E311–313.E313. <https://doi.org/10.1016/j.crad.2020.12.009>
- Eisenhauer EA, Therasse P, Bogaerts J, Schwartz LH, Sargent D, Ford R, Dancy J, Arbuck S, Gwyther S, Mooney M, Rubinstein

- L, Shankar L, Dodd L, Kaplan R, Lacombe D, Verweij J (2009) New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer* 45(2):228–247. <https://doi.org/10.1016/j.ejca.2008.10.026>
12. Therasse P, Arbutk SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, Verweij J, Van Glabbeke M, van Oosterom AT, Christian MC, Gwyther SG (2000) New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J Natl Cancer Inst* 92(3):205–216. <https://doi.org/10.1093/jnci/92.3.205>
 13. Jaffe CC (2008) Response assessment in clinical trials: implications for sarcoma clinical trial design. *Oncologist* 13(Suppl 2):14–18. <https://doi.org/10.1634/theoncologist.13-S2-14>
 14. Tirkes T, Hollar MA, Tann M, Kohli MD, Akisik F, Sandrasegaran K (2013) Response criteria in oncologic imaging: review of traditional and new criteria. *Radiographics* 33(5):1323–1341. <https://doi.org/10.1148/rg.335125214>
 15. World Health Organization (1979) WHO handbook for reporting results of cancer treatment. World Health Organization, Geneva, SZ
 16. Nardone V, Boldrini L, Grassi R, Franceschini D, Morelli I, Becherini C, Loi M, Greto D, Desideri I (2021) Radiomics in the setting of neoadjuvant radiotherapy: a new approach for tailored treatment. *Cancers (Basel)* 13(14). <https://doi.org/10.3390/cancers13143590>
 17. Choi H, Charnsangavej C, Faria SC, Macapinlac HA, Burgess MA, Patel SR, Chen LL, Podoloff DA, Benjamin RS (2007) Correlation of computed tomography and positron emission tomography in patients with metastatic gastrointestinal stromal tumor treated at a single institution with imatinib mesylate: proposal of new computed tomography response criteria. *J Clin Oncol* 25(13):1753–1759. <https://doi.org/10.1200/JCO.2006.07.3049>
 18. Crombe A, Le Loarer F, Cornelis F, Stoeckle E, Buy X, Cousin S, Italiano A, Kind M (2019) High-grade soft-tissue sarcoma: optimizing injection improves MRI evaluation of tumor response. *Eur Radiol* 29(2):545–555. <https://doi.org/10.1007/s00330-018-5635-4>
 19. Gennaro N, Reijers S, Bruining A, Messiou C, Haas R, Colombo P, Bodalal Z, Beets-Tan R, van Houdt W, van der Graaf WTA (2021) Imaging response evaluation after neoadjuvant treatment in soft tissue sarcomas: where do we stand? *Crit Rev Oncol Hematol* 160:103309. <https://doi.org/10.1016/j.critrevonc.2021.103309>
 20. Fields BKK, Hwang D, Cen S, Desai B, Gulati M, Hu J, Duddalwar V, Varghese B, Matcuk GR Jr (2020) Quantitative magnetic resonance imaging (q-MRI) for the assessment of soft-tissue sarcoma treatment response: a narrative case review of technique development. *Clin Imaging* 63:83–93. <https://doi.org/10.1016/j.clinimag.2020.02.016>
 21. Aerts HJ (2016) The potential of radiomic-based phenotyping in precision medicine: a review. *JAMA Oncol* 2(12):1636–1642. <https://doi.org/10.1001/jamaoncol.2016.2631>
 22. Buckler AJ, Bresolin L, Dunnick NR, Sullivan DC (2011) For the Group. A collaborative enterprise for multi-stakeholder participation in the advancement of quantitative imaging. *Radiology* 258(3):906–914. <https://doi.org/10.1148/radiol.10100799>
 23. Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. *Radiology* 278(2):563–577. <https://doi.org/10.1148/radiol.2015151169>
 24. Varghese BA, Cen SY, Hwang DH, Duddalwar VA (2019) Texture analysis of imaging: what radiologists need to know. *AJR Am J Roentgenol* 212(3):520–528. <https://doi.org/10.2214/AJR.18.20624>
 25. Hwang DH, Varghese BA, Chang M, Deng C, Ugweze C, Cen SY (2017) Duddalwar VA. Radiomics-based quantitative biomarker discovery: development of a robust image processing infrastructure. *Proc SPIE* 10160, 12th International Symposium on Medical Information Processing and Analysis, 1016017, January 26, 2017. <https://doi.org/10.1117/12.2256829>
 26. Demirjian NL, Varghese BA, Cen SY, Hwang DH, Aron M, Siddiqui I, Fields BKK, Lei X, Yap FY, Rivas M, Reddy SS, Zahoor H, Liu DH, Desai M, Rhie SK, Gill IS, Duddalwar V (2022) CT-based radiomics stratification of tumor grade and TNM stage of clear cell renal cell carcinoma. *Eur Radiol* 32(4):2552–2563. <https://doi.org/10.1007/s00330-021-08344-4>
 27. Fields BKK, Demirjian NL, Hwang DH, Varghese BA, Cen SY, Lei X, Desai B, Duddalwar V, Matcuk GR Jr (2021) Whole-tumor 3D volumetric MRI-based radiomics approach for distinguishing between benign and malignant soft tissue tumors. *Eur Radiol* 31(11):8522–8535. <https://doi.org/10.1007/s00330-021-07914-w>
 28. Demircioglu A (2021) Measuring the bias of incorrect application of feature selection when using cross-validation in radiomics. *Insights Imaging* 12(1):172. <https://doi.org/10.1186/s13244-021-01115-1>
 29. Crombe A, Marcellin PJ, Buy X, Stoeckle E, Brouste V, Italiano A, Le Loarer F, Kind M (2019) Soft-tissue sarcomas: assessment of MRI features correlating with histologic grade and patient outcome. *Radiology* 291(3):710–721. <https://doi.org/10.1148/radiol.2019181659>
 30. Crombe A, Perier C, Kind M, De Senneville BD, Le Loarer F, Italiano A, Buy X, Saut O (2019) T2-based MRI delta-radiomics improve response prediction in soft-tissue sarcomas treated by neoadjuvant chemotherapy. *J Magn Reson Imaging* 50(2):497–510. <https://doi.org/10.1002/jmri.26589>
 31. Peeken JC, Neumann J, Asadpour R, Leonhardt Y, Moreira JR, Hippe DS, Klymenko O, Foreman SC, von Schacky CE, Spraker MB, Schaub SK, Dapper H, Knebel C, Mayr NA, Woodruff HC, Lambin P, Nyflot MJ, Gersing AS, Combs SE (2021) Prognostic assessment in high-grade soft-tissue sarcoma patients: a comparison of semantic image analysis and radiomics. *Cancers (Basel)* 13(8). <https://doi.org/10.3390/cancers13081929>
 32. Corino VDA, Montin E, Messina A, Casali PG, Gronchi A, Marchiano A, Mainardi LT (2018) Radiomic analysis of soft tissues sarcomas can distinguish intermediate from high-grade lesions. *J Magn Reson Imaging* 47(3):829–840. <https://doi.org/10.1002/jmri.25791>
 33. Vallieres M, Freeman CR, Skamene SR, El Naqa I (2015) 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* 60(14):5471–5496. <https://doi.org/10.1088/0031-9155/60/14/5471>
 34. Peeken JC, Asadpour R, Specht K, Chen EY, Klymenko O, Akinkuoroye V, Hippe DS, Spraker MB, Schaub SK, Dapper H, Knebel C, Mayr NA, Gersing AS, Woodruff HC, Lambin P, Nyflot MJ, Combs SE (2021) MRI-based delta-radiomics predicts pathologic complete response in high-grade soft-tissue sarcoma patients treated with neoadjuvant therapy. *Radiother Oncol* 164:73–82. <https://doi.org/10.1016/j.radonc.2021.08.023>
 35. Baessler B, Weiss K, Pinto Dos Santos D (2019) Robustness and reproducibility of radiomics in magnetic resonance imaging: a phantom study. *Invest Radiol* 54(4):221–228. <https://doi.org/10.1097/RLI.0000000000000530>
 36. Crombe A, Fadli D, Italiano A, Saut O, Buy X, Kind M (2020) Systematic review of sarcomas radiomics studies: Bridging the gap between concepts and clinical applications? *Eur J Radiol* 132:109283. <https://doi.org/10.1016/j.ejrad.2020.109283>
 37. Gao Y, Kalbasi A, Hsu W, Ruan D, Fu J, Shao J, Cao M, Wang C, Eilber FC, Bernthal N, Bukata S, Dry SM, Nelson SD, Kamrava M, Lewis J, Low DA, Steinberg M, Hu P, Yang Y (2020) Treatment effect prediction for sarcoma patients treated with preoperative radiotherapy using radiomics features from longitudinal diffusion-weighted MRIs. *Phys Med Biol* 65(17):175006. <https://doi.org/10.1088/1361-6560/ab9e58>

38. Miao L, Cao Y, Zuo L, Zhang H, Guo C, Yang Z, Shi Z, Jiang J, Wang S, Li Y, Wang Y, Xie L, Li M, Lu N (2022) Predicting pathological complete response of neoadjuvant radiotherapy and targeted therapy for soft tissue sarcoma by whole-tumor texture analysis of multisequence MRI imaging. *Eur Radiol*. <https://doi.org/10.1007/s00330-022-09362-6>
39. O'Neil C, Schutt R (2014) *Doing data science: straight talk from the frontline*, 1st edn. O'Reilly Media, Sebastopol, CA
40. Friston K, Ashburner J, Kiebel S, Nichols T, Penny W (2007) eds. *Statistical parametric mapping: the analysis of functional brain images*. 1st ed. London, UK: Academic Press, <https://doi.org/10.1016/b978-0-12-372560-8.X5000-1>
41. Lei M, Varghese B, Hwang D, Cen S, Lei X, Desai B, Azadikhah A, Oberai A, Duddalwar V (2021) Benchmarking Various radiomic toolkit features while applying the image biomarker standardization initiative toward clinical translation of radiomic analysis. *J Digit Imaging* 34(5):1156–1170. <https://doi.org/10.1007/s10278-021-00506-6>
42. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc: Ser B (Methodol)* 57(1):289–300. <https://doi.org/10.1111/j.2517-6161.1995.tb02031.x>
43. Friedman J, Hastie T, Tibshirani R (2000) Additive logistic regression: a statistical view of boosting (With discussion and a rejoinder by the authors). *Ann Stat* 28(2). <https://doi.org/10.1214/aos/1016218223>
44. Corso F, Tini G, Lo Presti G, Garau N, De Angelis SP, Bellerba F, Rinaldi L, Botta F, Rizzo S, Origgi D, Paganelli C, Cremonesi M, Rampinelli C, Bellomi M, Mazzarella L, Pelicci PG, Gandini S, Raimondi S (2021) The challenge of choosing the best classification method in radiomic analyses: recommendations and applications to lung cancer CT images. *Cancers (Basel)* 13(12). <https://doi.org/10.3390/cancers13123088>
45. Gu Q, Feng Z, Liang Q, Li M, Deng J, Ma M, Wang W, Liu J, Liu P, Rong P (2019) Machine learning-based radiomics strategy for prediction of cell proliferation in non-small cell lung cancer. *Eur J Radiol* 118:32–37. <https://doi.org/10.1016/j.ejrad.2019.06.025>
46. Parmar C, Grossmann P, Bussink J, Lambin P, Aerts H (2015) Machine learning methods for quantitative radiomic biomarkers. *Sci Rep* 5:13087. <https://doi.org/10.1038/srep13087>
47. Peeken JC, Bernhofer M, Wiestler B, Goldberg T, Cremers D, Rost B, Wilkens JJ, Combs SE, Nusslin F (2018) Radiomics in radiooncology - challenging the medical physicist. *Phys Med* 48:27–36. <https://doi.org/10.1016/j.ejmp.2018.03.012>
48. Gitto S, Cuocolo R, Albano D, Morelli F, Pescatori LC, Messina C, Imbriaco M, Sconfienza LM (2021) CT and MRI radiomics of bone and soft-tissue sarcomas: a systematic review of reproducibility and validation strategies. *Insights Imaging* 12(1):68. <https://doi.org/10.1186/s13244-021-01008-3>
49. Hastie T, Tibshirani R, Friedman J (2009) *The elements of statistical learning: data mining, inference, and prediction*, 2nd edn. Springer, New York, NY, p 363
50. King G, Zeng L (2017) Logistic regression in rare events data. *Polit Anal* 9(2):137–163. <https://doi.org/10.1093/oxfordjournals.pan.a004868>
51. Buvat I, Orlhac F (2019) The dark side of radiomics: on the paramount importance of publishing negative results. *J Nucl Med* 60(11):1543–1544. <https://doi.org/10.2967/jnumed.119.235325>
52. Couronne R, Probst P, Boulesteix AL (2018) Random forest versus logistic regression: a large-scale benchmark experiment. *BMC Bioinformatics* 19(1):270. <https://doi.org/10.1186/s12859-018-2264-5>
53. Bleker J, Yakar D, van Noort B, Rouw D, de Jong IJ, Dierckx R, Kwee TC, Huisman H (2021) Single-center versus multi-center biparametric MRI radiomics approach for clinically significant peripheral zone prostate cancer. *Insights Imaging* 12(1):150. <https://doi.org/10.1186/s13244-021-01099-y>
54. Castillo TJ, Starmans MPA, Arif M, Niessen WJ, Klein S, Bangma CH, Schoots IG, Veenland JF (2021) A multi-center, multi-vendor study to evaluate the generalizability of a radiomics model for classifying prostate cancer: high grade vs. low grade. *Diagnostics (Basel)* 11(2). <https://doi.org/10.3390/diagnostics11020369>
55. Wang H, Nie P, Wang Y, Xu W, Duan S, Chen H, Hao D, Liu J (2020) Radiomics nomogram for differentiating between benign and malignant soft-tissue masses of the extremities. *J Magn Reson Imaging* 51(1):155–163. <https://doi.org/10.1002/jmri.26818>
56. Chu Y, Li L, Goldhof DB, Qui Y, Clark RA (2003) Classification of masses on mammograms using support vector machine. *Proc SPIE 5032, Medical Imaging 2003: Image Processing*, May 15, 2003. <https://doi.org/10.1117/12.481142>
57. Laws KI (1980) Rapid texture identification. *Proc SPIE 0238, Image Processing for Missile Guidance*, December 23, 1980. <https://doi.org/10.1117/12.959169>
58. Parekh V, Jacobs MA (2016) Radiomics: a new application from established techniques. *Expert Rev Precis Med Drug Dev* 1(2):207–226. <https://doi.org/10.1080/23808993.2016.1164013>
59. Sica GT (2006) Bias in research studies. *Radiology* 238(3):780–789. <https://doi.org/10.1148/radiol.2383041109>
60. Varghese BA, Hwang D, Cen SY, Levy J, Liu D, Lau C, Rivas M, Desai B, Goodenough DJ, Duddalwar VA (2019) Reliability of CT-based texture features: phantom study. *J Appl Clin Med Phys* 20(8):155–163. <https://doi.org/10.1002/acm2.12666>
61. Whitney HM, Li H, Ji Y, Liu P, Giger ML (2020) Harmonization of radiomic features of breast lesions across international DCE-MRI datasets. *J Med Imaging (Bellingham)* 7(1):012707. <https://doi.org/10.1117/1.JMI.7.1.012707>
62. Cetin-Karayumak S, Stegmayer K, Walther S, Szeszko PR, Crow T, James A, Keshavan M, Kubicki M, Rathi Y (2020) Exploring the limits of ComBat method for multi-site diffusion MRI harmonization. *bioRxiv*. Preprint. <https://doi.org/10.1101/2020.11.20.390120>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Brandon K. K. Fields¹ · Natalie L. Demirjian² · Steven Y. Cen³ · Bino A. Varghese³ · Darryl H. Hwang³ · Xiaomeng Lei³ · Bhushan Desai³ · Vinay Duddalwar³ · George R. Matcuk Jr.⁴ 

✉ George R. Matcuk Jr.
George.Matcuk@cshs.org

¹ Department of Radiology & Biomedical Imaging, University of California, San Francisco, San Francisco, CA 94143, USA

² College of Medicine – Tucson, University of Arizona, Tucson, AZ 85724, USA

³ Department of Radiology, Keck School of Medicine of the University of Southern California, Los Angeles, CA 90033, USA

⁴ Department of Radiology, S Mark Taper Foundation Imaging Center, Cedars-Sinai Medical Center, 8700 Beverly Blvd., Ste M-335, CA 90048 Los Angeles, USA