

PAPER 51

Predicting Diagnosis in Myxoid Soft Tissue Tumors: Performance of Radiomics vs. Radiologists

Joshua M. Lawrenz, MD¹; Can Cui, MS²; Samuel R. Johnson, BS¹; Katherine S. Hajdu, BS¹; Joanna L. Shechtel, MD³; Nicholson S. Chadwick, MD³; Hakmook Kang, PhD⁴; David S. Smith, PhD³,
Herbert S. Schwartz, MD¹ & Ginger E. Holt, MD¹

Investigation performed at Vanderbilt University Medical Center, Nashville, Tennessee, USA

From the Department of Orthopaedic Surgery¹
 Vanderbilt Institute of Surgery and Engineering²
 Department of Radiology³
 Department of Biostatistics⁴
 Vanderbilt University Medical Center
 Nashville, Tennessee, USA

Correspondence:

Joshua M. Lawrenz, MD
Assistant Professor of Orthopaedic Surgery
Division of Musculoskeletal Oncology
Vanderbilt University Medical Center
MCE South Tower, Suite 4200
Nashville, TN 37232-8774, USA
Phone: 630-291-4417
Fax: 615-875-1079
Email: josh.lawrenz@gmail.com

Title. Predicting Diagnosis in Myxoid Soft Tissue Tumors: Performance of Radiomics vs. Radiologists

Background. Benign and malignant myxoid soft tissue tumors in the musculoskeletal system have overlapping clinical, imaging, and histologic features that can make establishing a diagnosis challenging. Our institutional data has suggested a 30-40% indeterminate preoperative diagnosis rate even after needle biopsy. As treatment strategies differ, accurate preoperative diagnosis is preferred. We have previously constructed and trained a radiomics based machine learning (ML) model to classify benign and malignant myxoid tumors. The purpose of this study is validation of this model in a separate cohort and comparison to musculoskeletal radiologist review of images.

Questions/Purposes.

1. How does a radiomics based ML model used to distinguish benign and malignant myxoid tumors perform in a separate validation cohort?
2. How does a radiomics based ML model compare to musculoskeletal radiologists in distinguishing benign and malignant myxoid tumors?

Patients and Methods. A retrospective review of 90 patients with a pre-treatment MRI and a histologically confirmed myxoid soft tissue tumor (45 myxomas and 45 myxofibrosarcomas) on final resection pathology was performed. Baseline clinical features collected were patient age, sex, tumor size, tumor depth, tumor location, pain, and tumor as an incidental finding. Manual image segmentation of tumors was performed by attending clinicians on the entire axial T1, and entire T2FS or STIR sequences. Eighty-seven radiomics features (shape, intensity, and gray level matrix features) were extracted using PyRadiomics, and a LASSO model was used for feature reduction. With the selected radiomics and clinical features, five ML models (random forest, logistic regression, neural network, XgBoost, and SVM) were trained to classify tumors as benign or malignant in an initial 40 patients, and then tested with a separate validation cohort of 50 patients using 8-fold cross validation. The best ML classifier model using the validation cohort based on area under the receiver operating characteristic curve (AUC) was compared against the radiologists. For radiologist review of images, the same baseline clinical features were provided to the radiologists for each tumor. Two attending musculoskeletal radiologists independently classified the 50 tumors in the validation cohort as benign or malignant; for discordant cases, a third senior musculoskeletal radiologist made the consensus diagnosis. Radiologists classified tumors as benign or malignant with three iterations of information: 1. T1 + T2/STIR images; 2. T1 + T2/STIR + post-contrast images; and 3. T1 + T2/STIR + post-contrast images + the probability-based ML output score generated from the ML model. Radiologists 1 and 2 rated their confidence in the diagnosis of each tumor on a three-level scale: equivocal, probably, or consistent with. Diagnostic performance was compared between the radiomics based ML model and radiologist interpretations of images using the McNemar test.

Results. The best ML classifier model using the validation cohort was a logistic regression model with an AUC of 0.792. The ML model classified 78% (39/50) of tumors correctly compared to 74% (37/50) by radiology consensus using T1 + T2/STIR images only ($P=0.789$). Radiologists 1 and 2 cumulatively performed marginally worse when given post-contrast images (73%, 61/84), and marginally better when given the ML output score (75%, 75/100). When radiologists 1 and 2 were discordant (14/50), the senior radiologist classified 50% of tumors (7/14) correctly compared to 86% (12/14) by the ML model. Radiologist 1 and 2 rated their confidence as 'consistent with' in 33% (33/100) of tumors when T1, T2/STIR and post-contrast images were available to them, compared to 39% (39/100) when also given the ML output score. When radiologists 1 and 2 rated their confidence in the diagnosis as 'consistent with', they had cumulative 95% accuracy (37/39) compared to 62% accuracy (38/61) when they rated their confidence as 'equivocal/probably' ($P<0.001$). When radiologists 1 and 2 rated their confidence as 'equivocal/probably', they had 62% accuracy (38/61) compared to 74% accuracy (45/61) by the ML model ($P=0.244$).

Conclusions. A radiomics based ML model predicted benign or malignant diagnosis in musculoskeletal myxoid soft tissue tumors similarly to consensus imaging interpretation by three musculoskeletal radiologists. Radiologist confidence in the diagnosis strongly correlated with their diagnostic accuracy. These data suggest that though radiomics and radiologists perform similarly overall, radiomics may provide novel diagnostic utility when radiologist confidence is low, or when radiologists disagree. The clinical significance and specific application of these findings needs further validation in a larger external cohort as we seek to create an easy-to-use tool that will assist clinicians in real-time with advanced image interpretation when the diagnosis of soft tissue tumors is in question.

Figure 1. Study Design.

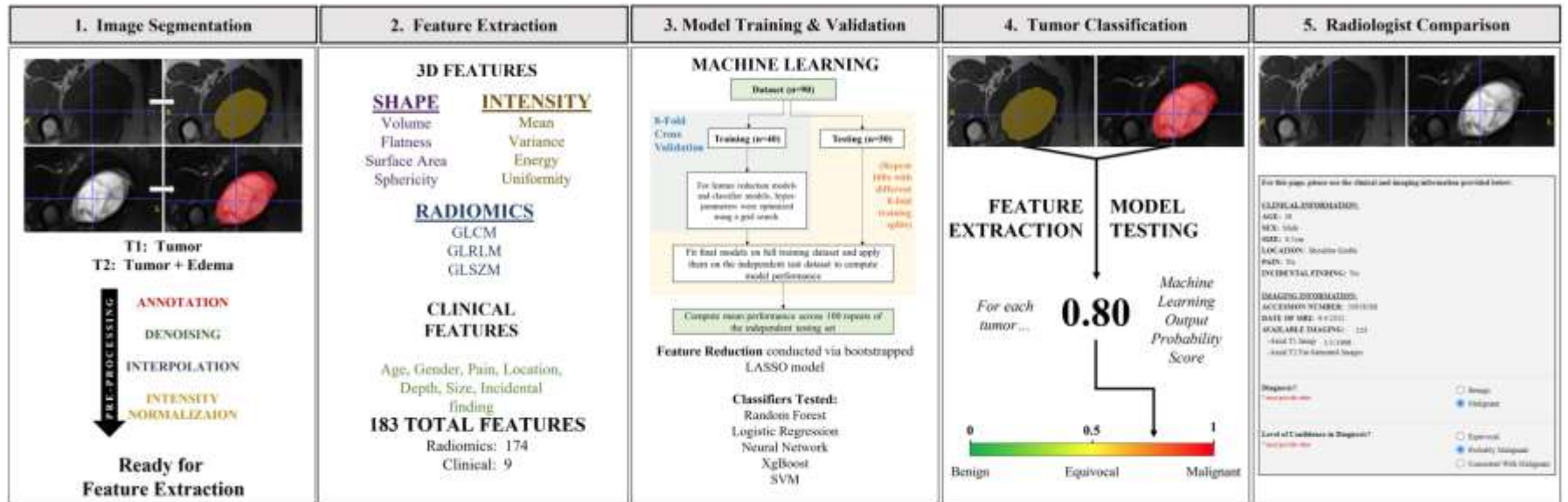


Table 1. Performance of radiomics vs. radiologists in distinguishing benign and malignant myxoid tumors.

Performance				
	Radiomics (%)	Radiologist 1 (%)	Radiologist 2 (%)	Consensus (%)
Accuracy	78.0	72.0	76.0	74.0
Sensitivity	72.0	80.0	64.0	72.0
Specificity	84.0	64.0	88.0	76.0
PPV	81.8	69.0	84.2	75.0
NPV	75.0	76.2	71.0	73.1
Radiomics vs. Radiologists				
	Comparison^a		P-Value	
	Radiomics vs. Radiologist 1		0.146	
	Radiomics vs. Radiologist 2		0.606	
	Radiomics vs. Consensus		0.789	

PPV = positive predictive value; NPV = negative predictive value; ^a = using McNemar test.