

Table S1. Several properties of the acquisition protocols of the 116 T1-weighted MRI sequences of patients with lipoma or well-differentiated liposarcoma (WDLPS) that were used to build the radiomics model.

Property	N	%		
Magnetic field strength				
1T	10	9.6		
1.5T	98	84.5		
3T	8	6.9		
Manufacturer				
Siemens	45	38.8		
Philips	45	38.8		
GE	26	22.4		
Setting (Unit)	Mean	Std.	Min.	Max.
Slice Thickness (mm)	4.77	1.14	2.5	10.0
Repetition time (ms)	555	108	280	831
Echo time (ms)	13.2	4.3	5.7	37
Available MRI sequences	N	%		
T1	116	100		
T1-FS	55	47.4		
T1-GD	42	36.2		
T1-FS-GD	80	69.0		
T2	76	65.5		
T2-FS	92	79.3		

Std.: standard deviation, min.: minimum value, max.: maximum value, mm: millimeters, ms: milliseconds, FS: Fat Saturation, GD: gadolinium contrast.

Table S2. Performance of the radiomics models based on T1 imaging features only; patient features only; manually scored features only; the combination of T1 imaging and manually scored features; and of volume only on the full dataset. Performance for the radiomics models is reported for each experiment as mean [95% confidence interval] over the cross-validation iterations.

	Model 1 T1 imaging features	Model 2 Patient features	Model 3 Manually scored features	Model 4 T1 imaging + manually scored features	Model 5 Volume
AUC	0.83 [0.75, 0.90]	0.75 [0.64, 0.85]	0.72 [0.62, 0.81]	0.69 [0.58, 0.79]	0.83 [0.75, 0.91]
Accuracy	0.68 [0.67, 0.84]	0.68 [0.59, 0.76]	0.67 [0.57, 0.76]	0.61 [0.51, 0.70]	0.76 [0.67, 0.84]
Sensitivity	0.68 [0.53, 0.82]	0.77 [0.63, 0.90]	0.76 [0.58, 0.94]	0.53 [0.37, 0.68]	0.67 [0.52, 0.81]
Specificity	0.84 [0.72, 0.95]	0.59 [0.45, 0.72]	0.57 [0.43, 0.71]	0.69 [0.54, 0.84]	0.84 [0.71, 0.97]
NPV	0.73 [0.63, 0.82]	0.73 [0.61, 0.85]	0.73 [0.59, 0.86]	0.60 [0.50, 0.69]	0.75 [0.66, 0.83]
PPV	0.82 [0.70, 0.93]	0.66 [0.58, 0.73]	0.64 [0.54, 0.74]	0.64 [0.51, 0.76]	0.81 [0.69, 0.93]

AUC: area under the curve, NPV: negative predictive value, PPV: positive predictive value

Table S3. Performance of radiomics models trained on features extracted from various MRI sequences on the full dataset. Performance is reported as mean [95% confidence interval] over the cross-validation iterations.

	T1	T1 + T1-FS	T1 + T1-GD	T1 + T1-FS-GD	T1 + T2	T1 + T2-FS
AUC	0.83 [0.75, 0.90]	0.84 [0.75, 0.92]	0.81 [0.72, 0.90]	0.81 [0.73, 0.89]	0.89 [0.83, 0.95]	0.81 [0.73, 0.88]
Accuracy	0.68 [0.67, 0.84]	0.77 [0.69, 0.85]	0.76 [0.67, 0.84]	0.75 [0.66, 0.83]	0.81 [0.74, 0.87]	0.74 [0.66, 0.81]
Sensitivity	0.68 [0.53, 0.82]	0.69 [0.56, 0.82]	0.69 [0.56, 0.82]	0.66 [0.51, 0.81]	0.74 [0.61, 0.86]	0.66 [0.53, 0.79]
Specificity	0.84 [0.72, 0.95]	0.84 [0.73, 0.95]	0.77 [0.71, 0.83]	0.84 [0.72, 0.95]	0.88 [0.78, 0.98]	0.82 [0.70, 0.93]
NPV	0.73 [0.63, 0.82]	0.74 [0.65, 0.82]	0.73 [0.64, 0.82]	0.72 [0.63, 0.81]	0.78 [0.69, 0.86]	0.72 [0.63, 0.80]
PPV	0.82 [0.70, 0.93]	0.83 [0.72, 0.93]	0.80 [0.69, 0.91]	0.81 [0.69, 0.93]	0.88 [0.78, 0.97]	0.79 [0.68, 0.90]

AUC: area under the curve, NPV: negative predictive value, PPV: positive predictive value, FS: Fat Saturation, GD: gadolinium contrast

Table S4. Performance of the three radiologists in differentiating between well-differentiated liposarcomas and lipomas on both the full and volume-matched cohort, and in differentiating dedifferentiated liposarcoma (DDLPS) and non-DDLPS (well-differentiated liposarcoma [WDLPS]/lipomas).

	Full cohort			Volume-matched cohort			DDLPS vs. non-DDLPS		
	Rad. 1	Rad. 2	Rad.3	Rad. 1	Rad. 2	Rad. 3	Rad. 1	Rad. 2	Rad. 3
AUC	0.74	0.72	0.61	0.68	0.74	0.55	0.97	0.91	0.90
Accuracy	0.64	0.64	0.61	0.62	0.63	0.55	0.95	0.62	0.89
Sensitivity	0.74	0.91	0.64	0.65	0.88	0.60	0.95	0.95	0.91
Specificity	0.55	0.36	0.59	0.58	0.37	0.50	0.95	0.56	0.89
NPV	0.68	0.81	0.62	0.61	0.74	0.54	0.99	0.98	0.98
PPV	0.62	0.59	0.61	0.62	0.59	0.56	0.78	0.29	0.61

AUC: area under the curve, NPV: negative predictive value, PPV: positive predictive value, rad.: radiologist