European Journal of Radiology 110 (2019) 203–211

Contents lists available at ScienceDirect

European Journal of Radiology

journal homepage: www.elsevier.com/locate/ejrad

Research article

A machine learning approach for distinguishing uterine sarcoma from leiomyomas based on perfusion weighted MRI parameters

Mahrooz Malek[a], Masoumeh Gity[a], Azadeh Alidoosti[a,*], Zeinab Oghabian[b], Pariya Rahimifar[a], Seyede Mahdieh Seyed Ebrahimi[a], Elnaz Tabibian[a], Mohammad Ali Oghabian[b]

[a] Department of Radiology, Medical Imaging Center, Imam Khomeini Hospital Complex (IKHC), Tehran University of Medical Sciences (TUMS), Tehran, Iran
[b] Neuroimaging and Analysis Group, Research Center for Molecular and Cellular Imaging, Department of Medical Physics and Biomedical Engineering, Tehran University of Medical Sciences (TUMS), Tehran, Iran

ARTICLE INFO

Keywords:
Perfusion weighted imaging
Computer-aided diagnosis
Leiomyomas
Magnetic resonance imaging
Uterine sarcoma

ABSTRACT

Purpose: To propose a computer-assisted method for distinguishing uterine sarcoma from leiomyomas based on perfusion weighted magnetic resonance imaging (PWI).

Materials and methods: Forty-two women confirmed to have a total of 60 masses (10 uterine sarcomas and 50 benign leiomyomas) were included. The reference diagnosis was based on postoperative histopathological examination. All women underwent the standard MRI protocol with 3-Tesla MR imager (Magnetom Trio, Siemens, Erlangen, Germany) for assessment of myometrial masses, followed by PWI. For each mass, two regions of interest (ROI) were outlined manually by an experienced radiologist; one (ROI_L) represented the entire tumor while the other (ROIs) was placed on the area of the lesion with the most marked contrast enhancement. Two additional ROI_s with diameters similar to ROI_s (3.0 to 3.1 mm) were placed on psoas muscle (ROI_P) and myometrium (ROI_M) in order to provide baselines for comparisons. The obtained ROIs of PWI images were then analyzed using the DCE Tool plug-in (version 2.0SP1) within ClearCanvas (Toronto, Ontario, Canada) framework. The DCE Tool provides seven parameters (K^{trans}, k_{ep}, V_b, IAUC, initial slope, peak, the mean squared error) for modelling contrast uptake within an ROI using the modified Tofts model. Parameters extracted from the ROIs were fed into a decision tree ensemble, which classified the corresponding lesions either as malignant or benign. The leave-one-out cross validation (LOOCV) was utilized to evaluate the performance of the classifier.

Results: None of the parameters extracted from ROI_L or ROI_s differed significantly between uterine sarcoma and benign leiomyomas (all $p > 0.05$). The overall accuracy of 66.7% was obtained by feeding seven parameters extracted from ROI_L to the classifier. When 21 features extracted from ROI_L, ROI_M, and ROI_P were fed into the classifier an accuracy of 91.7%, sensitivity of 100%, and specificity of 90% were achieved in the optimal operating point of classifier.

Conclusion: Although none of the PWI parameters differed significantly between benign and malignant lesions, when the information provided by the extracted features was aggregated using a machine learning method, a promising discriminative power was obtained. This suggests that the proposed model for combining the PWI parameters is potentially useful for differentiating uterine sarcoma from leiomyomas.

1. Introduction

Uterine sarcoma is a rare and heterogeneous disease accounting for 3%–9% of uterine corpus malignancies [1–8]. Unlike adenocarcinoma, which is the most common type of uterine cancer, the prognosis from the uterine sarcoma is poor [5] and about 26% of all deaths caused by uterine malignancies were caused by uterine sarcoma [5]. Currently there is no reliable diagnostic criteria for differentiating uterine sarcomas from other uterine masses and usually the ultimate diagnosis can be made only after the surgery based on postoperative histopathological assessment [4].

Depending on the tissue of origin and cancerous cell types, uterine sarcoma can exhibit different clinical features, but usually they manifest as rapidly growing masses with pelvic or abdominal pain [2]. Leiomyosarcoma is the most frequent subtype of uterine sarcoma and shares many common clinical grounds with leiomyoma, which is in turn

* Corresponding author at: Department of Radiology, Medical Imaging Center, Imam Khomeini Hospital Complex (IKHC), End of Keshavarz Blvd, Tehran, 1419733141, Iran.
E-mail address: azadeh.alidousti@gmail.com (A. Alidoosti).

https://doi.org/10.1016/j.ejrad.2018.11.009
Received 9 March 2018; Received in revised form 9 November 2018; Accepted 11 November 2018

the most common solid benign uterine neoplasm [1–3]. With the recent advents in minimally invasive techniques for managing leiomyomas, accurate preoperative diagnosis of uterine masses has become more important to select the safest therapeutic option [4]. Therefore, different imaging modalities, particularly magnetic resonance imaging (MRI), plays a key role in management of uterine masses [1].

The representation of uterine sarcoma and leiomyomas on T1- and T2-weighted MRI has been widely investigated [2]. Usually on T2-weighted sequences, leiomyomas appear as well-circumscribed hypointense masses with homogeneous density [1–3] On the other hand, typical sarcomas demonstrate as large ill-defined masses, with heterogeneously increased signal intensity on T2-weighted images and central hyperintensity indicative of central necrosis [1–3]. However, some of the degenerated and cellular leiomyomas might mimic sarcomas and exhibit regions of high signal intensity on T2-weighted images [9,10]. Therefore, it is not possible to make the ultimate differential diagnosis between sarcomas and their benign mimickers only based on conventional MRI sequences [1–3].

To improve the diagnostic accuracy, the efficacy of combining other imaging techniques such as diffusion-weighted imaging (DWI), magnetic resonance spectroscopy (MRS, and dynamic contrast-enhanced magnetic resonance imaging (PWI) with conventional MRI sequences has been studied. Tamai et al. [10] indicated that DWI, which provides quantitative measurements of apparent diffusion coefficient (ADC), may have a potential ability to distinguish degenerated leiomyomas from uterine sarcomas. However, the ADC of sarcoma overlapped with those of ordinary leiomyomas and cellular leiomyomas [10]. The MRS, particularly the evaluation of the choline concentration, can also provide valuable information in distinguishing benign and malignant uterine masses [11].

The efficacy of PWI for assessment of uterine masses was also previously investigated. PWI is a functional imaging technique involves the acquisition of serial MRI images before, during, and after the administration of an MR contrast agent. Zheng et al. [12] compared DCW-MRI with T2 images for the preoperative staging of early endometrial carcinoma. Visual assessment was done by two experts for evaluating both PWI and T2 images. In another study, Thomassin-Naggara et al [13] categorized the enhancement of the tumor solid tissue using visual classification of the time intensity curve obtained from PWI [14]. The study indicated that PWI was not useful for differentiating uncertain or malignant myometrial tumors from benign ones. In both studies [12,13], the visual assessment of an expert was used for the classification. However, quantitative contrast agent kinetic parameters can provide more reproducible measures for assessing the PWI. Previously it was shown that the quantitative PWI parameters differed between symptomatic uterine fibroids and normal uterus [15]. These parameters were also shown to be useful in predicting the immediate therapeutic response to MR-guided high-intensity focused ultrasound therapy among patients with symptomatic leiomyomas [16].

In this study, we investigated whether the quantitative parameters extracted from lesions using perfusion weighted magnetic resonance imaging (PWI) differed significantly between uterine sarcomas and Leiomyoma. Besides analysis of each feature on its own, we utilized a machine learning classifier to combine multiple PWI parameters and categorize lesions as benign or malignant. Previously, machine learning has been successfully utilized to combine four parameters extracted from PWI with T2 signal intensity and ADC to distinguish malignant from normal prostate tissue [17]. Our study only focused on seven parameters from PWI. We hypothesized that, by integrating multiple parameters from PWI, we could achieve higher accuracy in differentiating uterine sarcoma from benign uterine masses. We also explored whether the performance of the classifier could be improved by combining the PWI features extracted from the tumor area with the corresponding features calculated for the healthy tissue of myometrium and psoas muscle.

Table 1
The characteristics of the dataset.

	Number of patients (%)	Number of lesions (%)	Number of patients with multiple tumors	Age(mean ± STD, range)
Benign	33 (78.57%)	50 (83.33%)	10 (30.30%)	45.60 ± 11.19 (22-69)
Malignant	9 (21.43%)	10 (16.67 %)	1 (11.11%)	40.10 ± 15.60 (18-69)
Total	42	60	11 (26.19%)	44.68 ± 12.06 (18-69)

STD stands for standard deviation.

2. Material and methods

2.1. Study design and population

The institutional review board approved the study. Informed patient consent forms were obtained from all women recruited to participate in the study. We included all patients who were referred to our department for a pre-operative MRI between July 2015 and February 2017 and underwent hysterectomy or uterine tumor resection surgery. All women for whom MRI was contraindicated (e.g. those with an electronic implant or an acute respiratory disease) were excluded. The reference diagnosis for each tumor was made based on a pathologic specimen, obtained after tumor resection surgery or hysterectomy.

Forty-two women confirmed to have a total of 60 masses were included in the study. Overall, 11 women (26.2% of patients) had more than two tumors. Nine women (21.43% of women) had a diagnosis of malignancy while rest of them had a benign mass. None of the patients had both benign and malignant masses. Table 1 summarizes the characteristics of our dataset.

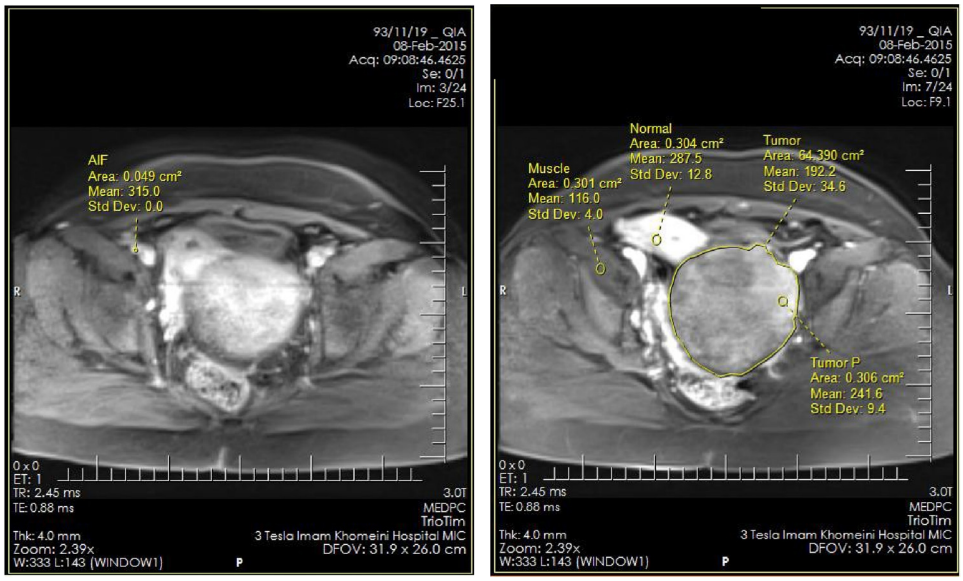
2.2. Imaging protocol

The MRI examinations were performed using a 3-Tesla MR imager (Magnetom Trio, Siemens, Erlangen, Germany). All women underwent the standard MRI protocol for the assessment of uterine masses followed by perfusion weighted imaging. Before undergoing the MRI scan, patients were advised not to eat or drink for at least three hours. Prior to the commencement of the scan, 20 mg of intramuscular hyoscine butylbromide was administered as an antispasmodic agent. All subjects were positioned supine on the MR scanner table with the phased-array coil positioned over the pelvis. The scan range in the craniocaudal direction was from the umbilicus level to the pubic symphysis. PWI was done in both groups whereby dynamic sequences were acquired in axial planes using volumetric interpolated breath-hold examination (VIBE) sequence with high temporal resolution.

For all patients, a single dose (0.2 cc/kg) of gadolinium contrast agent (DOTAREM; Guerbet, Aulnay, France) was administered at an injection rate of 3 cc/minute, followed by a saline chase of 20 cc. The PWI were acquired using volumetric interpolated breath-hold examination (VIBE) at a five-second temporal resolution for 350 s, beginning instantly after the injection.

2.3. Image analysis

Two Regions of Interest (ROI) were outlined manually by an experienced radiologist (M.K). The first ROI (ROI_L) encompassed the entire tumor while avoiding myometrium and surrounding healthy tissues. On the other hand, the second one (ROI_s) placed on the solid part of the lesion where the most marked contrast enhancement was observed. The diameter of ROI_s ranged from 3.0 to 3.1 mm. The magnitude of enhancement varies across the tumor and the quantitative



(a)

Tumor

Solution Parameter	Value	Units
K trans	0.531	mL/g/min
K ep	0.531	1/min
Vb	0.715	mL/g
IAUC	10.583	mmol/L*Sec
Initial Slope	5.742e-3	mmol/L/Sec
Peak	0.554	mmol/L
Mean Square Error	0.015	(mmol/L)^2

Tumor P

Solution Parameter	Value	Units
K trans	0.914	mL/g/min
K ep	0.914	1/min
Vb	1	mL/g
IAUC	20.129	mmol/L*Sec
Initial Slope	0.017	mmol/L/Sec
Peak	0.784	mmol/L
Mean Square Error	0.047	(mmol/L)^2

Normal

Solution Parameter	Value	Units
K trans	0.746	mL/g/min
K ep	0.746	1/min
Vb	0.976	mL/g
IAUC	21.263	mmol/L*Sec
Initial Slope	0.028	mmol/L/Sec
Peak	0.599	mmol/L
Mean Square Error	6.092e-3	(mmol/L)^2

Muscle

Solution Parameter	Value	Units
K trans	0.221	mL/g/min
K ep	0.284	1/min
Vb	0.104	mL/g
IAUC	4.62	mmol/L*Sec
Initial Slope	6.373e-3	mmol/L/Sec
Peak	0.217	mmol/L
Mean Square Error	2.273e-4	(mmol/L)^2

(b)

Fig. 1. (a) A sample screenshot of the software environment and (b) its output.

Table 2
The considered feature sets; each set included all seven features extracted from a certain group of ROIs. ROI_L and ROI_s represent the entire tumor area and the solid part of tumor respectively. ROI_P and ROI_M indicate areas within psoas muscle and myometrium where the most marked contrast enhancement was observed.

No.	ROIs for feature extraction	Number of features
1	ROI_L	7
2	ROI_L + ROI_M	14
3	ROI_L + ROI_P	14
4	ROI_L + ROI_M + ROI_P	21
5	ROI_s	7
6	ROI_s + ROI_M	14
7	ROI_s + ROI_P	14
8	ROI_s + ROI_M + ROI_P	21

features could be extracted either from the entire tumor area or only from the tumor area where the most marked contrast enhancement was observed. We extracted features from both ROI_L and ROI_s to compare these two strategies for extracting feature from the lesions. Two additional ROIs with similar diameters were placed on psoas muscle (ROI_P) and myometrium (ROI_M) in order to provide baselines for comparisons.

The obtained ROIs of PWI images were then analyzed using the DCE Tool plug-in (version 2.0SP1) within ClearCanvas (Toronto, Ontario, Canada) framework. The DCE Tool provides options to model contrast uptake within an ROI using the modified Tofts model [18]. Briefly, the model considers the blood plasma and the extracellular extravascular space as two compartments and assumes that the contrast agent uptake is governed by three tracer kinetic parameters, namely forward volume transfer constant or wash-in constant (K^{trans}), reverse reflux rate constant or wash-out constant (k_{ep}), and blood volume (V_b). The DCE Tool computed these three metrics for each ROIs. The most important physiologic parameter of the tissue is K^{trans}, which demonstrates a combination of both tissue perfusion and permeability. To assure that the injection rates and patients' cardiovascular states do not affect the kinetic parameters, the arterial input function (AIF) was incorporated to the two-compartment model. Here, the AIF corrected the kinetic analysis by measuring the signal from an ROI in the internal iliac artery on T1-weighted images and adjusting the results accordingly.

In addition to the above-mentioned parameters, the DCE Tool provides four semi-quantitative parameters for characterizing the analyzed ROIs. The first measure was the initial area under curve (IAUC), which is calculated by integrating the area under the contrast-enhancement time curve over the first 60 s after the contrast uptake. The second measure, called initial slope, represents the slope of the contrast enhancement-time curve in the first 10 s. The peak was the third metric, which simply indicates the absolute maximum contrast enhancement. Finally the last measure was the mean squared error. It shows the goodness of fit of the model fittings. For all four ROIs, these metrics were also recorded. Hence, in total seven features (three parameters based on the modified Tofts model and four semi-quantitative parameters) were from each ROI using the DCE tool. A sample screenshot of the software environment and its output is demonstrated in Fig. 1.

2.4. Ensemble of decision trees for malignant/benign classification

An ensemble of decision trees, a popular machine learning algorithm, was utilized to create a model which classify lesions as benign or malignant based on the features extracted in the previous section. The input of the model was the features of a lesion and its output is a label assigned by the model to that lesion. We utilized decision tree ensemble as it automatically estimates the feature importance while training and no further feature selection step was required [19]. We used MATLAB 2017a (Mathwork, Natick, MA, USA) software for implementation of the decision tree ensemble. The adaptive logistic regression [20] was used for aggregating the outputs of 200 decision tree. The minimum observations per leaf was set to 5.

For each lesion, four sets of features from four ROIs (ROI_L, ROI_s, ROI_M, and ROI_P) were extracted. We investigated the performance of the classifier for eight different combinations of the features. As shown in Table 2, the first and fifth feature sets comprised of features extracted from the entire tumor area (ROI_L) and those extracted from the solid part of the tumor (ROI_s), respectively. The second, third, and forth sets included the combination of the features extracted from ROI_L and baseline ROIs (ROI_P and ROI_M). On the other hand, three last feature sets contained the features extracted from ROI_s and two baseline ROIs.

2.5. Statistical analyses and leave-one-out cross validation

Patients' age in the malignant group was compared with that of the benign group using two independent samples *t*-test to ensure that the two groups were matched. The Mann–Whitney U test was utilized to compare Tofts model parameters between the benign and malignant groups. The parameters extracted from both ROI_s and ROI_L were evaluated. We also computed the ratio of Tofts model parameters for ROI_s and ROI_L to the corresponding parameters extracted from ROI_M and ROI_P. Therefore, four ratios were calculated for each parameter. The Mann–Whitney U test was also performed to explore if these ratio differed significantly between benign and malignant lesions. A p-value less than 0.05 was considered significant. All statistical analyses were performed using SPSS, version 22.0 (IBM Corp., Armonk, NY, USA).

As explained in section 2.4, the extracted features were fed into a machine learning classifier that categorized the lesions as benign or malignant. The ensemble of decision trees (the classifier) used the learned rule set for associating the features to the proper label. As these rules were learned from the data, one might argue that the classifier has been over-fitted to our dataset. To avoid this problem, leave-one-out cross validation (LOOCV) was utilized. In LOOCV, a single tumor was used as the test data (validation set) while the rest of the lesions were utilized as the training set to construct the decision tree ensemble. This was repeated such that each lesion served once as the test set. The confidence of the decision tree ensemble's classification of each lesion into the malignant group was recorded in each repetition. The LOOCV was performed using MATLAB 2017a (Mathwork, Natick, MA, USA).

To evaluate the performance of classifiers based on different feature sets, we generated a receiver operating characteristic (ROC) curve for each metric. The area under ROC curve (AUC) was then calculated in each case. The AUC values ranged from 1 (perfect classifier) to 0.5 (chance-level), indicating how well a classifier could differentiate benign lesions from malignant ones. In addition, the optimal operational point, which had the least distance to the upper left corner (true positive rate of 1 and false positive rate of 0), was also found for each ROC curve. By applying the cut-off threshold corresponding to the optimal operational point, the lesions were classified as benign or malignant. Seven accuracy measures including sensitivity, specificity, Negative Predictive Value (NPV), Positive Predictive Value (PPV), Positive Likelihood Ratio (LR+), Negative Likelihood Ratio (LR–), and overall accuracy were extracted for each feature set.

3. Results

As explained in section 2.3, the kinetic parameters of lesions based on both ROIL and ROIS were generated using the modified Tofts model. Fig. 2 (a) and 3 (b) show the ROIs overlaid on the original images acquired from two patients. The pixel-by-pixel pharmacokinetic parameter maps are indicated in Figs. 2 (b) and 3 (b). Table 3 shows the mean and standard deviation of each parameter for benign and malignant lesions. It also indicates the magnitude of difference in each parameter between the two groups. As shown in the table, none of these differences reached statistical significance ($p < 0.05$).

Table 4 indicates the ratio of kinetic parameters extracted from the

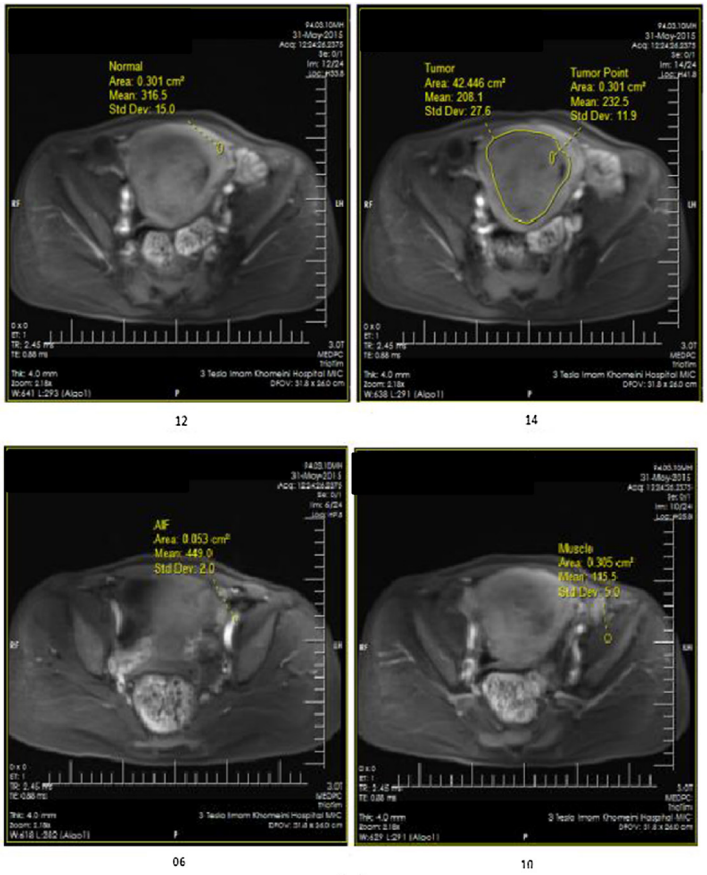
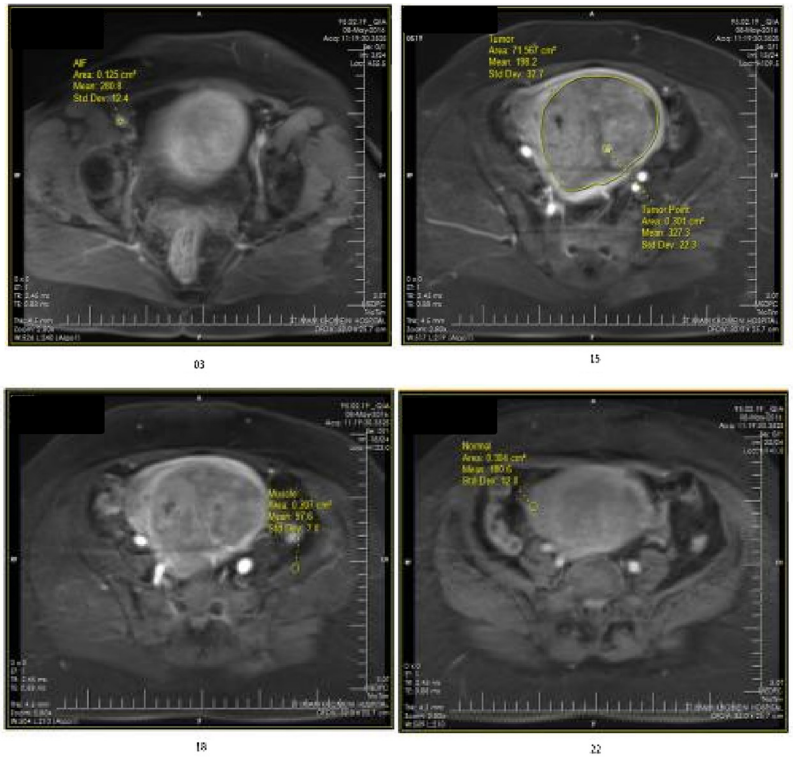


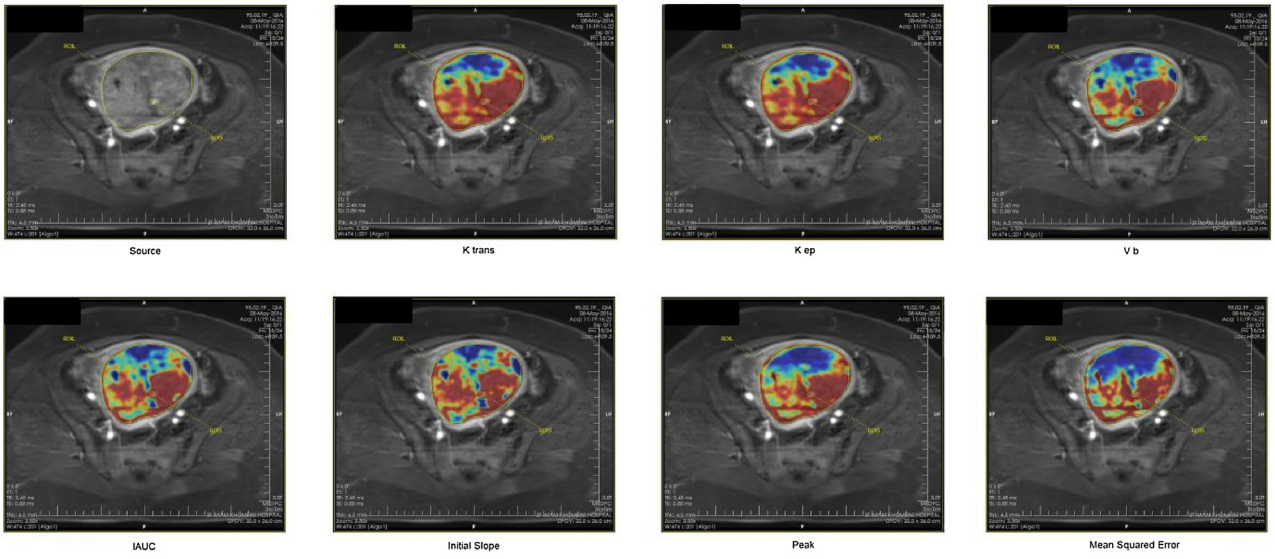
Fig. 2. (a) Segmented ROIs superimposed on the original images acquired from a 39 year-old patient with history of abnormal uterine bleeding and endometrial stromal sarcoma (b) pixel-by pixel pharmacokinetic parameter color maps.

tumor area to the corresponding parameters extracted from the two ROIs that served as the baseline for comparison, i.e., areas with the most marked enhancement in myometrium and psoas muscle. The table also shows similar ratios that were calculated using the parameters of the solid part of the tumor. As shown, although the ratios varied between the benign and malignant lesions, none of the differences were significant ($p < 0.05$).

Finally, we explored whether benign and malignant lesions could be differentiated by combining the extracted features from PWI. Using LOOCV, the performance of the decision tree ensemble for eight



(a)



(b)

Fig. 3. (a) Segmented ROIs superimposed on the original images acquired from a 51 year-old patient with history of abnormal uterine bleeding and leiomyomas (b) pixel-by-pixel pharmacokinetic parameter color maps.

considered feature sets (Table 2) in distinguishing malignant from benign lesions was evaluated. The ROC curve was generated using the confidence score of the classifier. The ROC curves are shown in Fig. 4 (a) and (b). The ROC curve corresponding to a random classifier (an AUC of 0.5) is also shown by the green line. The AUC values and their 95% confidence interval are listed in Table 5. As shown, the overall performance based on the AUC value was higher when the PWI parameters extracted from the entire tumor area (ROI_L) was considered compared to the condition where the parameters were extracted from the area with the most enhancement (ROI_S). The similar trends were observed for combination of ROI_L with ROI_M and combination of all three ROIs (ROI_L + ROI_M + ROI_P). However, when PWI parameters

Table 3
The comparison of the tracer kinetic parameters generated using the modified Tofts model between benign and malignant lesions. ROI column indicate the considered ROI for extracting Tofts model parameters. ROIL and ROIS represent the entire tumor area and the solid part of the tumor. The p-value is calculated using the Mann-Whitney U test.

ROI	Parameter	Status	Value (STD)	Difference	p-value
ROI_L	K^{trans}	Benign	0.577 (0.27)	−0.003	0.969
		Malignant	0.574 (0.296)		
	K_{ep}	Benign	0.646 (0.257)	0.011	0.903
		Malignant	0.657 (0.242)		
	V_b	Benign	0.455 (0.33)	0.158	0.323
		Malignant	0.613 (0.464)		
ROI_S	K^{trans}	Benign	0.690 (0.289)	0.052	0.579
		Malignant	0.742 (0.256)		
	K_{ep}	Benign	0.773 (0.27)	−0.027	0.797
		Malignant	0.746 (0.248)		
	V_b	Benign	0.570 (0.387)	0.056	0.700
		Malignant	0.626 (0.414)		
		Malignant	0.988 (0.038)		

STD stands for standard deviation.

Table 4
The comparison of the ratio of tracer kinetic parameters between tumor area and areas served as the baseline for comparisons (myometrium and psoas muscle). ROI_L and ROI_S correspond to the entire tumor area and its solid part while ROI_L and ROI_S represent the areas with the most marked enhancement in myometrium and psoas muscle. Generated using the modified Tofts model between benign and malignant lesions. P-values were calculated for the Mann-Whitney test statistic.

Metric	Comparison	Status	Value (STD)	Difference	p-value
K^{trans} Ratio	ROI_L to ROI_M	Benign	1.284 (0.856)	0.29	0.613
		Malignant	1.574 (2.005)		
	ROI_L to ROI_P	Benign	3.23 (2.34)	−0.337	0.394
		Malignant	2.893 (2.579)		
	ROI_S to ROI_M	Benign	1.519 (0.939)	0.249	0.662
		Malignant	1.768 (2.008)		
	ROI_S to ROI_P	Benign	3.924 (2.803)	−0.028	0.797
		Malignant	3.896 (2.936)		
K_{ep} Ratio	ROI_L to ROI_M	Benign	1.11 (0.599)	−0.083	0.713
		Malignant	1.027 (0.565)		
	ROI_L to ROI_P	Benign	1.462 (1.438)	−0.265	0.843
		Malignant	1.197 (0.489)		
	ROI_S to ROI_M	Benign	1.343 (0.697)	−0.204	0.513
		Malignant	1.139 (0.552)		
	ROI_S to ROI_P	Benign	1.77 (1.732)	−0.404	0.905
		Malignant	1.366 (0.518)		
V_b Ratio	ROI_L to ROI_M	Benign	2.061 (1.987)	3.212	0.990
		Malignant	5.273 (10.564)		
	ROI_L to ROI_P	Benign	9.381 (17.128)	0.154	0.797
		Malignant	9.535 (14.5)		
	ROI_S to ROI_M	Benign	2.435 (2.778)	2.809	0.863
		Malignant	5.244 (10.579)		
	ROI_S to ROI_P	Benign	11.808 (21.749)	−1.182	0.853
		Malignant	10.626 (14.949)		

STD stands for standard deviation.

were extracted from psoas muscle, the AUC values were almost identical. Among all eight feature sets, the one including PWI parameters from the tumor area, psoas muscle, and myometrium outperformed all other models in all eight performance metrics. It achieved a sensitivity of 100.0%, a specificity of 90.0%, and an accuracy of 91.7%. The AUC value was 0.972 (0.891-0.998), which was significantly than the AUC value obtained from the models relied only on the tumor parameters (ROI_L or ROI_S). The accuracy measures for the optimal operating point are also indicated in the table. As shown, the best results were obtained when features from the entire tumor area, myometrium, and psoas muscle classifier were combined.

We also computed estimates of feature importance for the ensemble of decision trees. To do so we used MATLAB built-in function (i.e. *predictorImportance*) [22]. Briefly, it generates the estimates by summing these importance estimates over all trees in the ensemble. To compute estimates of predictor importance for each tree, the summation of changes in the mean squared error due to splits on every feature was found and divided by the number of branch nodes. As we used LOOCV, we had 60 trained models. We took the average of importance estimates for all 60 trained models. The analysis showed that five most important features were IAUC of tumor, initial slope of myometrium, K_{ep} of myometrium, K_{trans} of tumor and psoas muscle. The five least important features were V_b of myometrium, initial slope of psoas muscle, V_b of psoas muscle, the mean squared error of myometrium and psoas muscle.

4. Discussion

In this study, we proposed a semi-automatic computer-assisted analysis framework for differentiating malignant uterine sarcoma from benign leiomyomas based on PWI. In the proposed framework the radiologist were asked to draw ROIs over tumor area, myometrium, and psoas muscle, and seven parameters were extracted from each ROI to characterize the contrast agent kinetics. The parameters were then fed into a decision tree ensemble classifier, which categorized lesions either as malignant uterine sarcoma or benign leiomyomas. The classifier achieved a sensitivity of 100% and specificity of 90% in its optimal operating point.

When the information provided by the extracted features was aggregated using a machine learning method, a promising discriminative power was reached. However, none of the features from the tumor area or the solid part of the tumor (i.e. the area with the most marked enhancement) differed significantly between benign and malignant lesions. Our results were in agreement with some previous studies, which reported that parameters based on DCE-MR perfusion imaging on their own demonstrated only a slight discriminative power in malignant tumor detection. For example, Riches et al. [21] investigated the efficacy of IAUC, kep and Ktrans in distinguishing malignant from normal prostate tissue. Similar to our results, the parameters for normal and malignant tissue did not differ; AUC values of 0.54, 0.55, and 0.57, respectively, were obtained. However, when IAUC was combined with ADC parameter, an AUC of 0.94 was obtained. Therefore, in agreement with our result, the combination of functional parameters improved malignant tissue detection. In another study, the results suggested that Vb, kep and Ktrans did not show a significant correlation with Gleason score of prostate cancer [23]. Thomassin-Naggara categorized the enhancement of the solid tissue using visual classification of the time intensity curve [14] and found that it was not useful for differentiating uncertain or malignant myometrial tumors from benign ones [13].

We also investigated the performance of the classifiers based on eight different sets of features. Based on the AUC and specificity measures, it was shown that added benefit of features from psoas muscle to the features extracted from tumor area are more than that of similar features extracted from myometrium. However, the model based on tumor and myometrium parameters had a higher sensitivity compared to the model based on tumor and psoas parameters. Our results showed that the classifier that utilized the most comprehensive set of features and combined features from tumor, myometrium, and psoas muscle outperformed the others. This was expected as we hypothesized that the model was able to predict the label of lesions more accurately by considering information about the magnitude of contrast enhancement on both lesion and comparison areas (myometrium, and psoas muscle). The results also indicated that considering the area of most marked enhancement is not sufficient to differentiate malignant from benign lesions, and the combination of features extracted from the entire tumor area resulted in higher accuracy compared to features extracted from a small area of the tumor with the highest enhancement level.

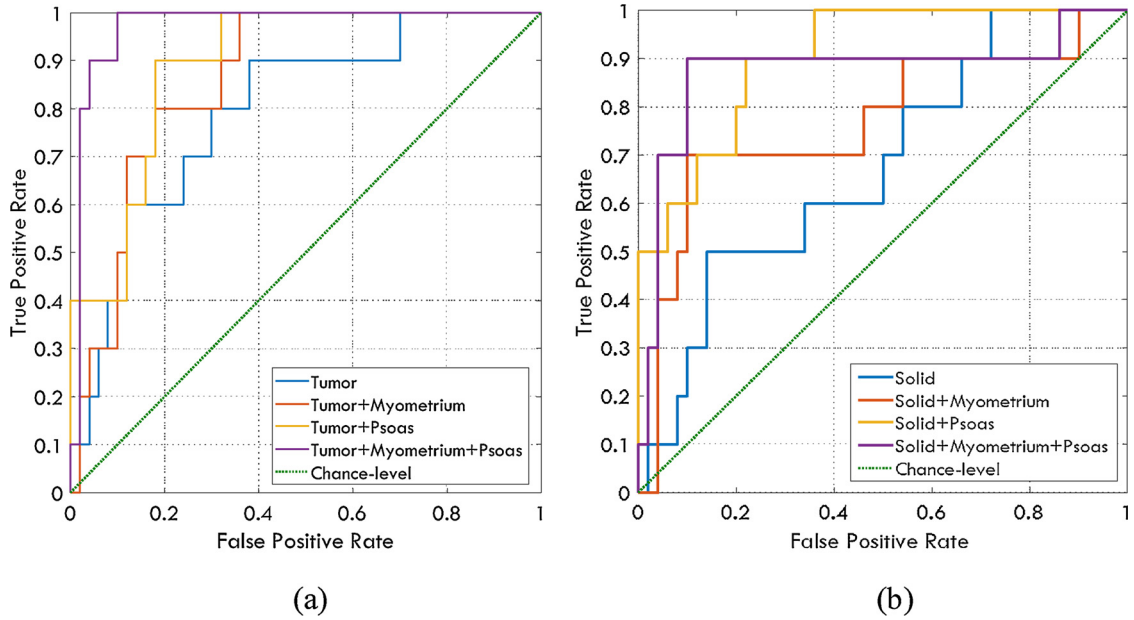


Fig. 4. The ROC curves obtained from the classifiers whose inputs included (a) the features extracted from the entire tumor area (b) the features extracted from the solid part of lesions where the most marked contrast enhancement was observed.

Table 5
Diagnostic performance of eight feature sets. Leave-one-out cross validation was used for all sets. The accuracy measures were reported for the optimal operating point of the classifiers. The best feature set based on each performance measure is shown in bold.

Metric	ROI_L	ROI_L + ROI_M	ROI_L + ROI_P	ROI_L + ROI_M + ROI_P	ROI_S	ROI_S + ROI_M	ROI_S + ROI_P	ROI_S + ROI_M + ROI_P
AUC (95% CI)[a]	0.766 (0.634-0.855)	0.862 (0.773-0.949)	0.892 (0.748-0.946)	**0.972 (0.891-0.998)**	0.676 (0.471-0.864)	0.766 (0.582-0.933)	0.904 (0.756-0.975)	0.874 (0.632-0.97)
Accuracy[b]	66.7%	70.0%	83.3%	**91.7%**	80.00%	86.67%	70.00%	90.00%
Sensitivity	90.0%	**100.0%**	90.0%	**100.0%**	50.00%	70.00%	**100.00%**	90.00%
Specificity	62.0%	64.0%	82.0%	**90.0%**	86.00%	**90.00%**	64.00%	**90.00%**
PPV	32.1%	35.7%	50.0%	**66.7%**	41.67%	58.33%	35.71%	64.29%
NPV	96.9%	**100.0%**	97.6%	**100.0%**	89.58%	93.75%	**100.00%**	97.83%
LR +	2.37	2.78	5.00	**10.00**	3.57	7.00	2.78	9.00
LR–	0.16	**0.00**	0.12	**0.00**	0.58	0.33	**0.00**	0.11

Confidence Interval (CI); Positive Predictive Value (PPV); Negative Predictive Value (NPV); Positive Likelihood Ratio (LR +); Negative Likelihood Ratio (LR–).
[a] Confidence Intervals were computed using bootstrap; the number of bootstrap replicas was set to 100.
[b] Overall accuracy.

The study has a number of limitations. First, our sample size was limited and included only ten malignant lesions. Although uterine sarcoma is a rare type of malignancy and our dataset was larger than those of many previous studies, further investigation of the proposed method using a larger dataset is required. Particularly a larger number of malignant lesions should be included to further validate the method. Also, the small sample size might limit the statistical power of our study. As shown in Table 3 and 4, none of the Tofts model parameters differed significantly between benign and malignant lesions, which might result from a lack of statistical power. Including more malignant cases might result in a statistically significant difference in some of the metrics. Further evaluation of the proposed method using a larger database is a potential avenue for future work. Second, our dataset might not represent all types of benign leiomyomas as only patients who underwent mass resection or hysterectomy were included in the study. Third, all ROIs were placed manually and therefore could be subject to inter- and intra-radiologist variability. In the future, the extent of these variabilities and the feasibility of automating ROI selection should be examined. Moreover, in this study, we only focused on the parameters extracted from PWI. A previous study showed that by combining features from T2-weighted MRI with DWI, a promising accuracy for differentiating benign from malignant myometrial tumors could be achieved [14]. Our study showed effectiveness of PWI parameters for this task. Investigating the added benefit of combining the PWI parameters extracted with other quantitative parameters obtained from T2-weighted sequences, DWI or MRS could be a potential future work for this study. Finally, the method proposed in this study required the computation of additional parameters and running an additional computer software to make a diagnosis. This could restrict the applicability of the proposed method in the clinical practice. Automating the ROI selections could facilitate the adoption of the proposed method as a decision support tool.

In summary, this study presented a model based on machine learning for classifying myometrial masses as benign or malignant. An ensemble of decision trees for combining quantitative parameters extracted from tumors and healthy tissue (myometrium and psoas muscle) on PWI was used. By feeding the PWI parameters from tumor, myometrium and psoas muscle to model, a sensitivity of 100% and a specificity of 90% was achieved. These preliminary results suggested that the proposed method could be potentially utilized along with conventional MRI sequences to differentiate between sarcomas and leiomyomas. As using this model will require computation of the PWI parameters, further studies on a larger sample size should be conducted to investigate cost-effectiveness of the method proposed in this study.

References

[1] P. Santos, T.M. Cunha, Uterine sarcomas: clinical presentation and MRI features, Diagn. Interv. Radiol. 21 (1) (2015) 4–9.
[2] T.I. Wu, T.C. Yen, C.H. Lai, Clinical presentation and diagnosis of uterine sarcoma, including imaging, Best Pract. Res. Clin. Obstet. Gynaecol. 25 (6) (2011) 681–689.
[3] E. D'Angelo, J. Prat, Uterine sarcomas: a review, Gynecol. Oncol. 116 (1) (2010) 131–139.
[4] Stewart, E.A., Differentiating uterine leiomyomas (fibroids) from uterine sarcomas.
[5] S.E. Brooks, et al., Surveillance, epidemiology, and end results analysis of 2677 cases of uterine sarcoma 1989-1999, Gynecol. Oncol. 93 (1) (2004) 204–208.
[6] D.D. Baird, et al., High cumulative incidence of uterine leiomyoma in black and white women: ultrasound evidence, Am. J. Obstet. Gynecol. 188 (1) (2003) 100–107.
[7] V.C. Buttram Jr, R.C. Reiter, Uterine leiomyomata: etiology, symptomatology, and management, Fertil. Steril. 36 (4) (1981) 433–445.
[8] S.P. Serden, P.G. Brooks, Treatment of abnormal uterine bleeding with the gynecologic resectoscope, J. Reprod. Med. 36 (10) (1991) 697–699.
[9] T. Namimoto, et al., Combined use of T2-weighted and diffusion-weighted 3-T MR imaging for differentiating uterine sarcomas from benign leiomyomas, Eur. Radiol. 19 (11) (2009) 2756–2764.
[10] K. Tamai, et al., The utility of diffusion-weighted MR imaging for differentiating uterine sarcomas from benign leiomyomas, Eur. Radiol. 18 (4) (2008) 723–730.
[11] M. Takeuchi, K. Matsuzaki, M. Harada, Differentiation of benign and malignant uterine corpus tumors by using proton MR spectroscopy at 3T: preliminary study, Eur. Radiol. 21 (4) (2011) 850–856.
[12] L. Zheng, et al., Comparison of dynamic contrast-enhanced magnetic resonance imaging with T2-weighted imaging for preoperative staging of early endometrial carcinoma, Onco. Ther. 8 (2015) 1743–1751.
[13] I. Thomassin-Naggara, et al., How to differentiate benign from malignant myometrial tumours using MR imaging, Eur. Radiol. 23 (8) (2013) 2306–2314.
[14] I. Thomassin-Naggara, et al., Dynamic contrast-enhanced MR imaging of ovarian neoplasms: current status and future perspectives, Magn. Reson. Imaging Clin. N. Am. 16 (4) (2008) 661–672.
[15] S. Majd, et al., Quantitative dynamic contrast-enhanced magnetic resonance imaging in symptomatic uterine fibroids and normal uterus: a feasibility study, Iran. J. Radiol. 14 (2) (2017).
[16] Y.-s. Kim, et al., Dynamic contrast-enhanced magnetic resonance imaging predicts immediate therapeutic response of magnetic resonance-guided high-intensity focused ultrasound ablation of symptomatic uterine fibroids, Invest. Radiol. 46 (10) (2011) 639–647.
[17] Y. Sun, et al., Predicting prostate tumour location from multiparametric MRI using Gaussian kernel support vector machines: a preliminary study, Australas. Phys. Eng. Sci. Med. 40 (1) (2017) 39–49.
[18] P.S. Tofts, et al., Estimating kinetic parameters from dynamic contrast-enhanced T 1-weighted MRI of a diffusable tracer: standardized quantities and symbols, J. Magn. Reson. Imaging 10 (3) (1999) 223–232.
[19] T.G. Dietterich, Ensemble learning, The Handbook of Brain Theory and Neural Networks Vol. 2 (2002), pp. 110–125.
[20] J. Friedman, T. Hastie, R. Tibshirani, Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors), Ann. Stat. 28 (2) (2000) 337–407.
[21] S.F. Riches, et al., MRI in the detection of prostate cancer: combined apparent diffusion coefficient, metabolite ratio, and vascular parameters, Am. J. Roentgenol. 193 (6) (2009) 1583–1591.
[22] W.Y. Loh, Regression trees with unbiased variable selection and interaction detection, Stat. Sin. 12 (2002) 361–386.
[23] A. Oto, et al., Diffusion-weighted and dynamic contrast-enhanced MRI of prostate cancer: correlation of quantitative MR parameters with Gleason score and tumor angiogenesis, Am. J. Roentgenol. 197 (6) (2011) 1382–1390.