ADC-derived spatial features can accurately classify adnexal lesions

Article *in* Journal of Magnetic Resonance Imaging · September 2017

DOI: 10.1002/jmri.25854

CITATIONS
4

READS
124

8 authors, including:

Anahita Fathi Kazerooni
University of Pennsylvania
48 PUBLICATIONS **159** CITATIONS
SEE PROFILE

Masoumeh Gity
Tehran University of Medical Sciences, Iranian Society of Radiology,
91 PUBLICATIONS **408** CITATIONS
SEE PROFILE

Mahrooz Malek
Tehran University of Medical Sciences
24 PUBLICATIONS **124** CITATIONS
SEE PROFILE

Hamidreza Saligheh Rad
Tehran University of Medical Sciences
132 PUBLICATIONS **710** CITATIONS
SEE PROFILE

ADC-Derived Spatial Features Can Accurately Classify Adnexal Lesions

Anahita Fathi Kazerooni, PhD [1,2] Mahnaz Nabil, PhD,[3]
Hamidreza Haghighat Khah, MD,[4] Mohammadreza Alviri, MSc,[1]
Maryam Heidari-Sooreshjaani, MD,[5] Masoumeh Gity, MD,[6,7] Mahrooz Malek, MD,[6,7]
and Hamidreza Saligheh Rad, PhD [1,2]*

Background: The role of quantitative apparent diffusion coefficient (ADC) maps in differentiating adnexal masses is unresolved.
Purpose/Hypothesis: To propose an objective diagnostic method devised based on spatial features for predicting benignity/malignancy of adnexal masses in ADC maps.
Study Type: Prospective.
Population: In all, 70 women with sonographically indeterminate and histopathologically confirmed adnexal masses (38 benign, 3 borderline, and 29 malignant) were considered for this study.
Field Strength/Sequence: Conventional and diffusion-weighted magnetic resonance (MR) images (b-values = 50, 400, 1000 s/mm^2) were acquired on a 3T scanner.
Assessment: For each patient, two radiologists in consensus manually delineated lesion borders in whole ADC map volumes, which were consequently analyzed using spatial models (first-order histogram [FOH], gray-level co-occurrence matrix [GLCM], run-length matrix [RLM], and Gabor filters). Two independent radiologists were asked to identify the attributed (benign/malignant) classes of adnexal masses based on morphological features on conventional MRI.
Statistical Tests: Leave-one-out cross-validated feature selection followed by cross-validated classification were applied to the feature space to choose the spatial models that best discriminate benign from malignant adnexal lesions. Two schemes of feature selection/classification were evaluated: 1) including all benign and malignant masses, and 2) scheme 1 after excluding endometrioma, hemorrhagic cysts, and teratoma (14 benign, 29 malignant masses). The constructed feature subspaces for benign/malignant lesion differentiation were tested for classification of benign/borderline/malignant and also borderline/malignant adnexal lesions.
Results: The selected feature subspace consisting of RLM features differentiated benign from malignant adnexal masses with a classification accuracy of ~92%. The same model discriminated benign, borderline, and malignant lesions with 87% and borderline from malignant with 100% accuracy. Qualitative assessment of the radiologists based on conventional MRI features reached an accuracy of 80%.
Data Conclusion: The spatial quantification methodology proposed in this study, which works based on cellular distributions within ADC maps of adnexal masses, may provide a helpful computer-aided strategy for objective characterization of adnexal masses.
Level of Evidence: 1
Technical Efficacy: Stage 2

J. MAGN. RESON. IMAGING 2017;00:000–000.

Preoperative differential diagnosis of sonographically indeterminate adnexal masses, comprising 15–25% of adnexal masses overall, is crucial to avoid unnecessary surgeries in women with benign masses, to reduce the risk of missing malignant lesions, and to more reliably refer patients to the proper specialist.[1,2] Magnetic resonance imaging (MRI)

View this article online at wileyonlinelibrary.com. DOI: 10.1002/jmri.25854

Received Mar 21, 2017, Accepted for publication Aug 29, 2017.

*Address reprint requests to: H.S.R., Medical Physics and Biomedical Engineering, Quantitative MR Imaging and Spectroscopy Group, Research Center for Molecular and Cellular Imaging, Tehran University of Medical Sciences, Tehran, Iran, PO. 1419733141. E-mail h-salighehrad@tums.ac.ir

From the [1]Quantitative MR Imaging and Spectroscopy Group, Research Center for Molecular and Cellular Imaging, Tehran University of Medical Sciences, Iran; [2]Department of Medical Physics and Biomedical Engineering, School of Medicine, Tehran University of Medical Sciences, Iran; [3]Department of Mathematics, Islamic Azad University, Qazvin Branch, Qazvin, Iran; [4]Department of Diagnostic Imaging, Shohada-e-Tajrish Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran; [5]Department of Radiology, Modarres Hospital, Shahid Beheshti University of Medical Sciences, Tehran, Iran; [6]Advanced Diagnostic and Interventional Radiology Research Center (ADIR), Tehran University of Medical Sciences, Tehran, Iran; and [7]Department of Radiology, Medical Imaging Center, Tehran University of Medical Sciences, Tehran, Iran

is a key diagnostic tool for reliable characterization of sonographically indeterminate adnexal masses[2–4] and has been shown to reduce "overoperation" and "underoperation" in benign and malignant adnexal lesions, respectively.[4,5] Yet conventional MRI findings are insufficient in diagnosis of adnexal masses and complementary techniques must be incorporated with the examination to reach a more reliable diagnosis.[6,7]

Diffusion-weighted imaging (DWI) and its derived quantitative map, apparent diffusion coefficient (ADC), provide tractable biomarkers of tumor progression.[8,9] DWI is sensitive to the microscopic movement of water molecules, called Brownian motion, in the extracellular, intracellular, and intravascular spaces. Tissue regions with higher cellularity, like malignant tumors, provide smaller extracellular space for diffusion of water molecules and defective cell membranes permit motion of water molecules among extracellular and intracellular spaces. Therefore, as abnormal processes within the tissue are accompanied with changes in water mobility, DWI can show tissue abnormalities. ADC values reduce when restriction of water diffusion increases with higher cellular density.[8]

In the context of adnexal masses, it is now accepted that qualitative DWI assessment in combination with conventional MRI is useful for identifying benign masses, as low signal intensity within a solid component on T_2-weighted MRI with low signal intensity on high b-value DWI is highly suggestive of benignity.[7,10,11] But the accuracy of the qualitative assessment of DW and conventional MR images depends on the expertise of the readers.[12]

In this regard, quantitative approaches may aid the radiologists to achieve more accurate and objective diagnosis. Nevertheless, in adnexal lesions the contribution of quantitative ADC in discrimination of benign and malignant masses is still doubtful and several studies have reported large overlaps of ADC values among benign and malignant lesions, when typical mean ADC has been considered the discriminating feature.[10,11,13] These outcomes may have been achieved as a result of quantifying the lesion area with conventional features, like mean ADC, that average out the interrelationships of the pixels within the lesion area. A preliminary study by Kierans et al reported the diagnostic value of ADC entropy in comparison with ADC mean in differentiating benign from malignant adnexal masses, but it was suggested that the benefit of ADC entropy decreases when the reader has more experience.[12]

The purpose of the current study was to propose an objective decision-making approach for discrimination of benign from malignant adnexal masses, by implementing a computer-aided diagnosis method based on spatial analysis and automatic classification of ADC maps as quantitative imaging data.

Materials and Methods

Patients

This prospective study was compliant with the Health Insurance Portability and Accountability Act (HIPAA), and approved by the Institutional Review Board (IRB). The patients were included only if they provided informed consent. Between June 2012 and November 2015, patients who presented with indeterminate masses on first-line sonography examination based on "Risk of Malignancy Index" (RMI = ultrasound findings × CA125 × menopausal status)[14] were included. Ultrasound findings (U) were presence of multilocular cysts, solid components, bilateralism, metastases, and ascites. A U-score equal or above 1 (U = 0: no abnormality; U = 1: single abnormality; U = 2: more than one abnormality), in conjunction with the concentration of CA125 (directly entered in the formula) and patient's menopausal status (M = 1: premenopausal; M = 3: postmenopausal) were used for calculating RMI. The criteria for referring the patients for MRI was an RMI value passing a threshold of 200, as the discriminative point for segregating benign and malignant masses for RMI 1.[15,16]

MRI Acquisition Parameters

MR images were acquired on a 3T scanner (MAGNETOM Tim TRIO, Siemens, Erlangen, Germany), with patients placed in a surface phased-array coil in the supine position. A sagittal T_2-weighted fast spin-echo image from one femoral head to the other, an axial T_2-weighted fast spin-echo images from the renal hilum to below the symphysis pubis, an axial T_1-weighted gradient-echo with breathhold, and pre- and postcontrast axial fat-suppressed spoiled T_1-weighted gradient-echo before and after injection of 0.2 mmol/kg of contrast agent (Dotarem; Guerbet, Aulnay, France) were acquired.

DW-MRI was acquired using a fat-suppressed single-shot GE-EPI sequence at b-values of 50, 400, and 1000 s/mm^2. ADC maps were generated automatically on the system workstation from DW images using all three b-values based on monoexponential fitting. The MRI parameters are summarized in Table 1.

Histopathological Assessment

A total of 91 patients with the aforementioned indications for MRI were imaged. Inclusion criteria for this study consisted of: 1) patients who underwent surgery with documented histopathological results within 1 month after their MRI examination ($n = 88$), 2) availability of a whole set of MR images as well as artifact-free DWI ($n = 76$). Exclusion criteria were: 1) history of gynecological surgeries or infections prior to MRI examination ($n = 5$), 2) known primary cancer prior to the MRI ($n = 1$).

The final dataset included 70 patients (mean age, 37.7; age range, 14–67), consisting of 38 benign, 3 borderline, and 29 malignant cases, which were eligible to be considered in the analysis. Histopathological findings of the included patients are detailed in Table 2.

Qualitative Image Analysis Based on Conventional MRI

All images were transferred to an offline workstation and were reviewed independently by two radiologists (reader 1 [H.H.] with 16 years of experience in pelvic MRI and reader 2 [M.G.] with 10 years of experience in abdominal imaging) who were aware of the ages but blinded to clinical and histopathological information of the patients. The radiologists were provided with conventional MR images, were asked to assess morphological features of the lesions

TABLE 1. MRI Parameters

	Coronal and/or sagittal T2-w	Axial T2-w	Axial T1-w, and fat-saturated pre- and post- contrast T1-w	Axial DWI (b-values = 50, 400, 1000)
TR (msec)	4000	4470	840	4800
TE (msec)	80	84	14	97
Flip angle (°)	120	150	140	90
Field of view (mm^2)	200×250	336×340	336×340	300×300
Slice thickness (mm)	6	5	5	4
Gap (mm)	1	1	1	5
Matrix	280×280	384×380	384×380	128×128

based on signal intensity characteristics on T_1/T_2-weighted images and postcontrast T_1-weighted MRI, and to provide their diagnosis about the category of masses as benign or malignant based on their impression about the presence of multilocular cyst, thick septa, enhancing mural nodule, solid component, papillary projections, ascites, peritoneal seeding, and pelvic adenopathy.

Region of Interest (ROI) Delineation

Two radiologists in consensus (M.H.S. with 1 year of experience in body MRI and reader 1, H.H.) manually delineated volumetric polygonal whole-lesion borders of the adnexal masses throughout the whole ADC slices. To avoid partial volume effects, the borders were selected immediately inside the outer margin of the lesion.

TABLE 2. Histopathological Characteristics of the Recruited Patients With Adnexal Masses

	Histologic subtype	Number of patients	Prevalence[a] (%)
Benign		38	54.2
	Broad ligament fibroma	4	5.7
	Chronic tuboovarian abcess	1	1.4
	Fibrotechoma	1	1.4
	Serous cystadenoma	4	5.7
	Mucinous cystadenoma	2	2.9
	Mature cystic teratoma	4	5.7
	Endometrioid cyst	17	24.3
	Hemorrhagic cyst	3	4.3
	Hydrosalpynx	2	2.9
Borderline		3	4.3
	Mucinous tumor	2	2.9
	Serous tumor	1	1.4
Malignant		29	41.4
	Serous cystadenocarcinoma	14	20.0
	Mucinous cystadenocarcinoma	4	5.7
	Endometrioid cystadenocarcinoma	3	4.3
	Dysgerminoma	4	5.7
	Lymphoma	1	1.4
	Yolk-sac tumor	1	1.4
	Krukenburg tumor	2	2.9

Unless otherwise stated, data are numbers of patients.
[a]The percentages may not add up to 100% due to rounding.

TABLE 3. Summary of Quantitative Features Used in Our Work

Feature category	No. of features	Feature names
First-Order Histogram (FOH)	10	Mean, Standard Deviation (SD), Variance (Smoothness), Energy (Uniformity), Entropy (Irregularity), Skewness, Kurtosis, 25th, 75th, and 95th Percentiles
Gray-Level Co-occurrence Matrix (GLCM)	23	Autocorrelation, Contrast, Correlation ($n = 2$), Cluster Prominence, Cluster Shade, Dissimilarity, Energy, Entropy, Homogeneity ($n = 2$),[a] Maximum Probability, Variance, Sum Average, Sum Variance, Sum Entropy, Difference Variance, Difference Entropy, Information measure of correlation ($n = 2$), Homomorphic Inverse Difference, Inverse Difference Normalized, Inverse Difference Moment Normalized
Run-Length Matrix (RLM)	44	11 Features in 4 Directions (0°, 45°, 90°, 135°): Short-Run Emphasis (SRE), Long-Run Emphasis (LRE), Gray-level Nonuniformity (GLN), Run-Length Nonuniformity (RLN), Run Percentage (RP), Low Gray-level Run Emphasis (LGRE), High Gray-level Run Emphasis (HGRE), Short-Run Low Gray-level Emphasis (SRLGE), Short-Run High Gray-level Emphasis (SRHGE), Long-Run Low Gray-level Emphasis (LRLGE), Long-Run High Gray-level Emphasis (LRHGE)
Gabor Texture	25	Rotation-invariant Gabor filters calculated by taking magnitudes of Fourier coefficients in 8 orientations and 5 radial frequencies. Due to Fourier transformation, only the first 5 Fourier coefficients are unique, resulting in 25 texture features (5 coefficients for 5 radial frequencies).

[a]The values in the parentheses indicate the number of features with the same concept but calculated with different formulations.

Spatial Quantification

To characterize cellular distributions within adnexal lesions, we exploited four types of regional quantification models to fit the ADC maps. A number of textural features were derived from each model (overall $n = 102$ features): 1) first-order histogram (FOH) model ($n = 10$); 2) gray-level co-occurrence matrix (GLCM) ($n = 23$); 3) run-length matrix (RLM) ($n = 44$); and 4) Gabor texture model ($n = 25$). The aforementioned texture models are described as follows (a summary of the derived quantitative features is provided in Table 3).

1) FOH: Image histogram is the simplest statistical function that globally measures the number of pixels with the same ADC value within the whole ROI. The characteristics of the tumor can be assessed through several descriptive features of FOH.[17] In a study by Kierans et al, it was shown that ADC entropy as a histogram-derived parameter can outperform ADC mean in discriminating benign from malignant adnexal lesions.[12] Figure 1 illustrates histograms of benign and malignant populations included in our study. The histograms are generated from ADC values within volumes of interest (VOIs) of all patients within each of benign and malignant groups. From the figure, it is apparent that a classification model solely constructed based on descriptors of histogram shape cannot separate the two classes of adnexal lesions due to large overlap of histogram appearances. Thus, more complicated classifiers based on textural features are required to make a distinction among these groups.

2) GLCM: Global FOH statistics do not account for the interrelationships of the local neighboring pixels and ignore the probability of co-occurrence of pixel values. GLCM is a statistical texture extraction method that allows for measuring the distribution of pixel pairs, separated by a specific distance and direction. The number of pixel pairs with a specified distribution is calculated to construct the co-occurrence matrix. Here the ADC values within ROIs were quantized to 256 gray-levels with distances of 1 pixel apart averaged over the four directions to extract Haralick GLCM features.[18]

3) RLM: Run-length features represent the coarseness of texture in a given direction. A run comprises of consecutive pixels of the same gray-level in a predetermined direction. For example, larger values in shorter runs indicate fine textures, for which the changes of the distribution with distance rapidly occur.[19]

4) Gabor textural features: Gabor filters, as a transform-based texture analysis model, optimally localize image properties in spatial and frequency domains. Gabor filters imitate the human visual cortex system like that in the retina, and decompose a scene into several filtered images with signal-intensity variations occurring in limited ranges of frequencies and orientations.[20]

The mentioned features from FOH, GLCM, RLM, and Gabor categories were calculated using an in-house software developed in MatLab 2016a (MathWorks, Natick, MA).

Statistical Analysis: Feature Selection and Automatic Classification

The features extracted from the categories mentioned in the previous section construct a feature space with a large dimensionality, which should be reduced through a feature selection step. The best

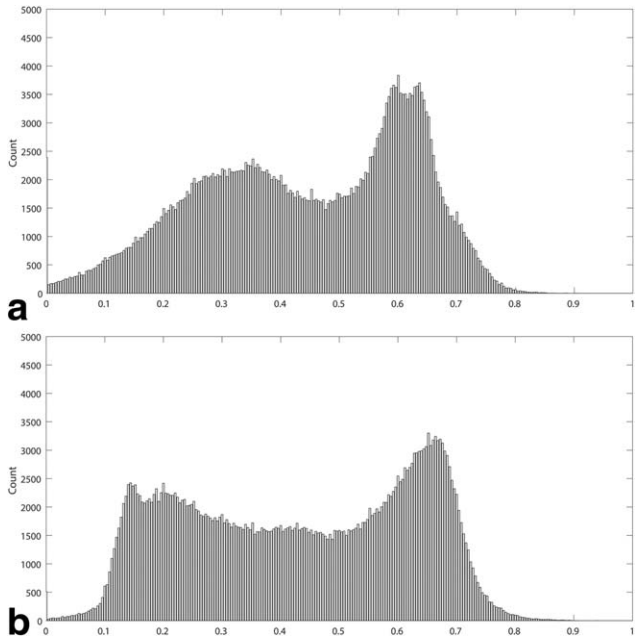

FIGURE 1: The population histograms for (a) 38 benign and (b) 29 malignant patients. ADC values within whole volume of interest for all patients in each group are accumulated to visualize the degree of intra- and interpatient heterogeneity and large overlap of ADC values among benign and malignant adnexal masses. The gray levels are normalized between 0 and 1 for better visualization. As can be observed, both histograms are bimodal and the distributions are widely spread in both groups. Differentiation of benign and malignant lesions purely based on histogram shape characteristics does not seem to be feasible. FOH features are not selected in the final classification model, as the heterogeneity of adnexal lesions cannot be sufficiently represented by the global spatial features extracted from the FOH. Therefore, more complicated spatial features, which account for interrelationships of the neighboring pixels, are necessary to be able to accurately model textural distributions of adnexal lesions and discriminate benign from malignant cases.

feature subset selected is the one that best approximates cellular distribution of adnexal lesions and minimizes the loss of information while controlling overfitting of the features to the data. As information theory is founded on the basis of quantifying the expected value of information, here we employed and compared two statistical information-based feature selection criteria: Akaike Information Criterion (AIC)[21] and Schwarz Bayesian Information Criterion (BIC).[22] Both criteria were applied in forward, backward, and stepwise selection strategies; therefore, six feature selection methods were implemented to find the best subset of features.

We implemented an algorithm by which all possible feature subsets constructed within each individual textural category (FOH, GLCM, RLM, and Gabor) using feature selection methods were assessed for their classification performance with the aim to find the best representative textural category for classifying adnexal masses. Furthermore, the feature space comprising all features of four categories was explored by the feature selection methods to find the best combination subset.

Classification based on the selected texture features was performed using Fischer's linear discriminant analysis (LDA) method. We selected this nonparametric classifier as it seeks a line that optimally discriminates the lesion groups and, therefore, can reveal the underlying pattern of data based on inherent discriminative potential of features/feature subsets.[23]

To avoid overfitting and decreasing the possible bias effects of outliers on diagnostic performance of the model selection criterion and classifier, feature selection was applied in loops of leave-one-out crossvalidation (keeping the number of iterations less than the number of data points). The most frequently selected feature subset with highest accuracy was considered as the final feature subset. Furthermore, a leave-one-out crossvalidation using the selected feature subset was performed over 50 iterations to validate the performance of classification in terms of sensitivity, specificity, accuracy, and area under the receiver operating (ROC) curve (AUC) measures. The reported results were calculated by taking averages of these measures over 50 iterations.

All statistical procedures, including feature selection and classification, were carried out in R Statistical Software (R v. 3.0.2, Vienna, Austria).

In summary, in an attempt to identify the spatial heterogeneity model of adnexal lesions, AIC and BIC feature selection methods were implemented in forward selection, backward elimination, and stepwise selection strategies (overall, six model selection methods) and applied on a space consisting of the features extracted from four categories: FOH, GLCM, RLM, and Gabor. The feature selection step was followed by LDA classification to categorize the patients into benign and malignant groups.

Due to the small sample size of borderline tumors ($n = 3$), we only considered benign and malignant groups (ie, borderline tumors excluded) for feature selection and classification. Two classification schemes were explored: 1) including all benign and malignant masses, and 2) scheme 1 excluding endometrioma, hemorrhagic cysts, and teratoma masses. Scheme 1 was implemented to build a generic spatial quantification method that can objectively classify diverse masses in their attributing classes. Therefore, all adnexal masses were included to find the best spatial textural subsets that can reliably discriminate sonographically indeterminate masses into benign and malignant categories. As endometrioma, hemorrhagic cysts, and teratoma show highly specific features on conventional MR images,[12] to ensure the absence of any biases in feature selection, in Scheme 2, the analysis was repeated for the situation when these masses were excluded from the Scheme 1.

In each scheme, the feature subsets were generated for classification of benign from malignant adnexal lesions, and were then tested for their capability in discriminating benign, borderline, and malignant masses. Regarding the importance of differentiating borderline from malignant lesions, the previously selected feature subsets were again examined considering two classes of borderline and malignant tumors.

Results

Qualitative Analysis

The diagnostic accuracy of qualitative analysis, performed by two radiologists based on conventional MR images, was assessed. When including all masses, reader 1 correctly diagnosed 29 patients (of 38) as benign (TN) and 25 patients

TABLE 4. Classification Performance of the Selected Feature Subset for Discrimination of Benign From Malignant Adnexal Masses Including Endometrioma, Hemorrhagic Cysts, and Teratoma (38 Benign, 29 Malignant)

Quantitative analysis—evaluation in terms of classification performance (crossvalidated)[a]

	Selected features	Sens.	Spec.	Acc.	AUC (95% CI[b])
Subset (1)	$N = 43$, from RLM Category	90.4	92.6	91.7	97.5 (94.6–100)
Subset (2)	$N = 67$, from RLM and Gabor Categories	100	100	100	100 (100–100)
Subset (3)	$N = 9$: FOH-Skewness, FOH-Smoothness, FOH-95th Percentile, GLCM-Contrast, RLM-SRLGE in 90°, RLM-LGRE in 90°, 3 from Gabor	83.8	82.9	84.1	90.8 (83.6–98.1)

Qualitative Analysis by Radiologists

Readers	Sens.	Spec.	Acc.	AUC (95% CI[b])
Reader #1	86.2	76.3	80.1	80.1 (69.2–96.1)
Reader #2	65.6	68.4	67.1	65.2 (52.3–79.2)

[a]The assessment results are all in percentages (%). Sens. = Sensitivity; Spec. = Specificity; Acc. = Accuracy; AUC = Area Under the ROC Curve.
[b]95% Confidence Interval (CI).

(of 29) as malignant (TP), while nine benign patients were incorrectly classified as malignant (FP), and four malignant patients were mistakenly diagnosed as benign (FN), resulting in overall sensitivity of 86.2%, specificity of 76.3%, and accuracy of 80.5% (Table 4). False positives included endometrioma ($n = 3$), teratoma ($n = 2$), hydrosalpynx ($n = 1$), fibroma ($n = 1$), mucinous cystadenoma ($n = 1$), and fibrotechoma ($n = 1$). False negatives comprised of serous cystadenocarcinoma ($n = 2$), dysgerminoma ($n = 1$), and endometrioid cystadenocarcinoma ($n = 1$). When excluding endometrioma, hemorrhagic cysts, and teratoma, the qualitative analysis by reader 1 resulted in TP = 25 (of 29), TN = 8 (of 14), FP = 6, FN = 4 (sensitivity, 86.2%, specificity, 57.1%, accuracy, 76.7%) (Table 5).

Evaluating the interpretation of reader 2 when including all masses indicated the following results: TP = 19 (of

TABLE 5. Classification Performance of the Selected Feature Subset for Discrimination of Benign From Malignant Adnexal Masses Excluding Endometrioma, Hemorrhagic Cysts, and Teratoma (14 Benign, 29 Malignant)

Quantitative assessment of classification performance (crossvalidated)[a]

Feature selection method	Selected features	Sens.	Spec.	Acc.	AUC (95% CI[b])
Subset (4)	$N = 43$, from RLM Features	100	100	100	100 (100–100)
Subset (5)	$N = 15$, from Gabor	90.2	98.5	92.6	96.9 (92.1–100)
Subset (6)	$N = 12$; FOH-Standard Deviation, GLCM-Max Probability, GLCM-Cluster Shade, GLCM-Difference Entropy, RLM-LRE in 0°, RLM-LGRE in 45°, RLM-HGLRE in 90°, 5 from Gabor	99.9	82.9	94.9	98.0 (94.2–100)

Qualitative analysis by radiologists

Readers	Sens.	Spec.	Acc.	AUC (95% CI[b])
Reader #1	86.2	57.1	76.7	67.6 (50.5–85.6)
Reader #2	65.6	42.9	58.1	49.3 (31.2–68.4)

[a]The assessment results are all in percentages (%). Sens. = Sensitivity; Spec. = Specificity; Acc. = Accuracy; AUC = Area Under the ROC Curve.
[b]95% Confidence Interval (CI).

TABLE 6. Classification Performance of the Selected Feature Subset for Discrimination of Benign, Borderline, and Malignant Adnexal Masses Including Endometrioma, Hemorrhagic Cysts, and Teratoma (38 Benign, 3 Borderline, 29 Malignant)

Benign/borderline/malignant				
Feature selection method			**Acc.**	**AUC (95% CI[a])**
Subset (1)			87.3	88.5 (78.0–99.0)
Subset (2)			99.8	99.7 (99.3–100)
Subset (3)			80.6	75.1 (41.8–99.8)
Borderline/malignant				
Feature selection method	**Sens.**	**Spec.**	**Acc.**	**AUC (95% CI[b])**
Subset (1)	100	100	100	100 (100–100)
Subset (2)	100	100	100	100 (100–100)
Subset (3)	100	33.3	93.7	96.9 (90.0–100)

[a]The assessment results are all in percentages (%). Sens. = Sensitivity; Spec. = Specificity; Acc. = Accuracy; AUC = Area Under the ROC Curve.
[b]95% Confidence Interval (CI).

29), TN = 26 (of 38), FP = 12, and FN = 10, yielding an overall sensitivity of 65.5%, specificity of 68.4%, and accuracy of 67.1% (Table 4). False positives included endometrioma ($n = 3$), teratoma ($n = 2$), fibroma ($n = 3$), mucinous cystadenoma ($n = 2$), serous cystadenofibroma ($n = 1$), hydrosalpynx ($n = 1$), fibrotechoma ($n = 1$). False negatives consisted of serous cystadenocarcinoma ($n = 4$), mucinous cystadenocarcinoma ($n = 2$), endometrioid cystadenocarcinoma ($n = 2$), dysgerminoma ($n = 1$). For this reader, excluding endometrioma, hemorrhagic cysts, and teratoma, the following diagnostic performance resulted: TP = 19 (of 29), TN = 6 (of 14), FP = 8, FN = 10 (sensitivity, 65.5%, specificity, 42.9%, accuracy, 58.1%) (Table 5).

Both radiologists diagnosed two of the borderline tumors (one serous and one mucinous) as malignant and one mucinous tumor as benign.

Spatial Quantification and Automatic Classification

As the number of generated subsets was large, in what follows we just report those resulting in classification performance of over 80%.

SCHEME 1: INCLUDING ALL BENIGN AND MALIGNANT MASSES. In this scheme, feature selection was designed for discrimination of benign ($n = 38$) from malignant ($n = 29$) adnexal lesions, and then tested for classification of benign ($n = 38$), borderline ($n = 3$), and malignant ($n = 29$), and finally, for differentiation of borderline from malignant adnexal tumors. The results are presented in Tables 4 and 6.

Table 4 shows the three best feature subsets constructed for classification of benign and malignant adnexal masses. As can be observed, the subset consisting of RLM textural features (subset 1) resulted in sensitivity, 90.4%, specificity, 92.6%, accuracy, 91.7%, and AUC, 97.5%. The FPs for this classification model consisted of three patients with serous cystadenoma ($n = 1$), hydrosalpinx ($n = 1$), and fibroma ($n = 1$). The two latter were also misdiagnosed by readers 1 and 2. The FNs included Krukenburg tumor ($n = 1$), mucinous cystadenocarcinoma ($n = 1$), and endometrioid cystadenocarcinoma ($n = 1$). The two radiologists misdiagnosed the endometrioid cystadenocarcinoma; reader 2 misclassified the mucinous cystadenocarcinoma. Another subset built based on Gabor and RLM features (subset 2) resulted in 100% classification accuracy. Figure 2 illustrates an example of a benign lesion correctly classified in the benign group, by subsets 1 and 2.

Finally, a subset including nine features (subset 3) derived from different textural categories returned sensitivity, 83.8%, specificity, 82.9%, accuracy, 84.1%, and AUC, 90.8%.

The same subsets were tested on three classes of benign ($n = 38$), borderline ($n = 3$), and malignant lesions ($n = 29$) (Table 6), indicating accuracy and AUC of 87.3% and 88.5% for subset 1, 100% and 100% for subset 2, and 80.6% and 75.1% for subset 3. The misclassified borderline tumor in subset 1 was the serous tumor, which was mistakenly categorized as malignant. In subset 3, two borderline tumors, including the serous and one of the mucinous tumors, were misclassified in the malignant class. When applying the derived feature subsets only on borderline and

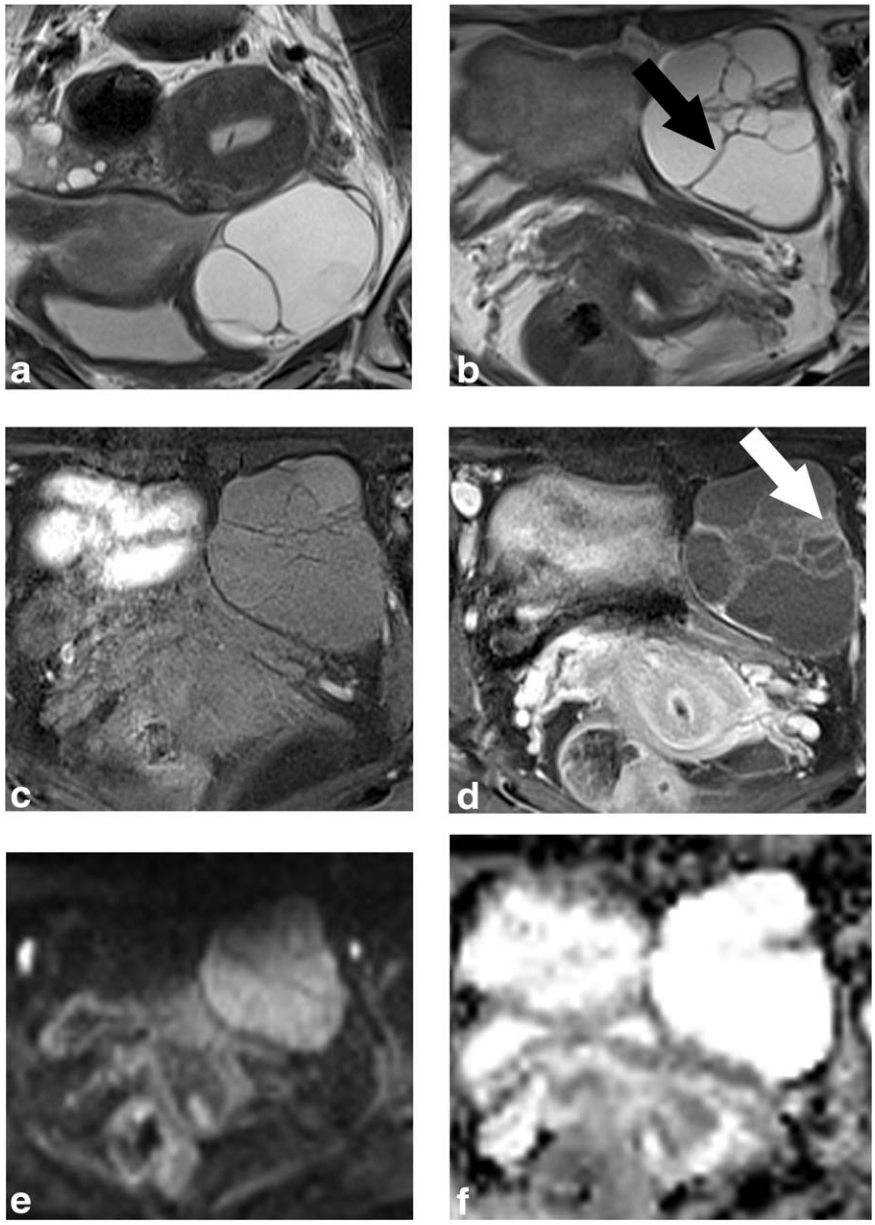

FIGURE 2: A 34-year-old female histopathologically confirmed with benign mucinous cystadenoma. (a) Coronal T_2w image. (b) Axial FS T_2w. (c) Axial precontrast FS T_1w. (d) Axial postcontrast FS T_1w. (e) Axial DWI (b = 1000). (f) ADC map. The images show left ovarian cystic lesion with thin walls and multiple thin internal septa (black arrow). The signal intensity of the fluid within cyst is slightly high on T_1w (pre- and postcontrast), and high on T_2w. A solid enhancing mural nodule with 7.9 mm diameter can be observed (white arrow). The radiologists suggested the possibility of a malignancy for this patient. Our proposed method correctly classified it in the benign group.

malignant tumor classes, subsets 1 and 2 showed sensitivity, specificity, accuracy, and AUC of 100% and subset 3 again misclassified the same two borderline tumors and demonstrated sensitivity, 100%, specificity, 33.3%, accuracy, 93.7%, and AUC, 96.9%.

SCHEME 2: EXCLUDING ENDOMETRIOMA, HEMORRHAGIC CYSTS, AND TERATOMA. We investigated the classification scheme when endometrioma, hemorrhagic cysts, and teratoma ($n = 24$) were excluded from the benign group in the scheme 1, to test the capability of spatial quantification of ADC maps for diagnosing the benign lesions without clear morphological appearances.

As the results in Table 5 indicate, a model constructed based on RLM features could discriminate benign ($n = 14$) from malignant ($n = 29$) lesions with 100% accuracy. Based on the Gabor model, classification resulted in sensitivity, 90.2%, specificity, 98.5%, accuracy, 92.6%, and AUC, 96.9%. This model did not show false negatives in most of the iterations; in a few iterations, one of the fibroma cases (misdiagnosed by readers 1 and 2) was mistakenly classified as malignant. False positives for this Gabor model were

TABLE 7. Classification Performance of the Selected Feature Subset for Discrimination of Benign, Borderline, and Malignant Adnexal Masses Excluding Endometrioma, Hemorrhagic Cysts, and Teratoma (14 Benign, 3 Borderline, 29 Malignant)

Benign/borderline/malignant				
Feature selection method			**Acc.**	**AUC (95% CI[b])**
Subset (4)			100	100 (100–100)
Subset (5)			90.1	95.4 (86.6–100)
Subset (6)			92.7	85.4 (65.6–100)
Borderline/malignant				
Feature selection method	**Sens.**	**Spec.**	**Acc.**	**AUC (95% CI[b])**
Subset (4)	100	100	100	100 (100–100)
Subset (5)	96.6	67.7	94.0	98.0 (92.5–100)
Subset (6)	100	67.7	96.9	96.0 (87.0–100)

[a]The assessment results are all in percentages (%). Sens. = Sensitivity; Spec. = Specificity; Acc. = Accuracy; AUC = Area Under the ROC Curve.
[b]95% Confidence Interval (CI).

dysgerminoma ($n = 1$), endometrioid adenocarcinoma ($n = 1$), and lymphoma ($n = 1$). The dysgerminoma tumor was misdiagnosed by our radiologists, while this case of endometriod adenocarcinoma and the lymphoma were correctly diagnosed by readers 1 and 2.

Furthermore, a combination textural model was formed using 12 features resulted in sensitivity, 99.9%, specificity, 82.9%, accuracy 94.9%, and AUC, 98%. False positives were one fibroma and one serous cystadenoma; readers 1 and 2 misdiagnosed the fibroma.

The same feature subsets were employed for discrimination of three groups of adnexal lesions, ie, benign, borderline ($n = 3$), and malignant, resulting in accuracy and AUC of 100% for RLM model, 90.1% and 95.4% for Gabor model, and 92.7% and 85.4% for the combination model (Table 7). Borderline tumors could be differentiated from malignant tumors with sensitivity, specificity, accuracy, and AUC of 100% for the RLM model. The Gabor model resulted in sensitivity, 96.6%, specificity, 67.7%, accuracy, 94.0%, and AUC, 98.0%, while the combination model returned sensitivity, 100%, specificity, 67.7%, accuracy, 96.6%, and AUC, 96.0%. The misclassified borderline tumor in Gabor and combination model (false positive) was the serous tumor.

Discussion

In this work, by mathematical spatial analysis and classification of adnexal lesions, we showed that through an objective and automated diagnosis approach, the information provided by ADC maps about cellular distribution of adnexal lesions may be predictive of malignancy or benignity.

While several studies have reported on the insignificance of mean ADC for differentiating benign from malignant adnexal masses,[24,25] it has been shown that qualitative assessment of high b-value DW images in conjunction with conventional MRI may reach 90% accuracy.[12,26] However, qualitative assessment is dependent on the experience of radiologists for accurate interpretation of the images to avoid over- or undertreatments of the patients.[2] Furthermore, employing an algorithmic qualitative approach depends on the presence of a supervising radiologist to tailor the MRI examination and to advise problem-solving MRI sequences.[7] Here, we showed that qualitative assessment of the two radiologists involved in this work, based on conventional MRI features, could at most reach 80% accuracy, while our proposed spatial classification method could provide over 90% accuracy and diagnostic performance (AUC). Therefore, objective and quantitative approaches may produce higher diagnostic accuracy when the acquisition of problem-solving MR images is not plausible or radiologists are not expert in reading and interpreting gynecological MR images.

In a study performed by Kierans et al on 39 (30 benign and 9 malignant) adnexal lesions, it was shown that ADC entropy, as a first-order histogram measure, is capable of differentiating benign from malignant adnexal masses, while mean ADC was not statistically significant among benign and malignant lesions.[12] This suggests that mean ADC, as a typical ADC-derived measure, is incapable of resolving benign from malignant adnexal masses, since it averages out the heterogeneity within the mass; whereas spatial quantification methods, like texture analysis, may

account for spatial differences existing among lesions. Accordingly, we indicated that differentiation of benign and malignant adnexal lesions based on spatial quantification and automatic classification achieves high accuracy (>90%).

In several studies, it is presumed that the existing overlap among benign and malignant adnexal lesions is partly due to abnormal DWI signal intensity within some benign lesions, particularly endometrioma and teratoma.[25] Endometriomas with high blood and hemosiderin concentrations[27] and teratomas with keratinoid components[28] show significantly reduced mean ADC values close to those of malignant lesions. As endometriomas and teratomas indicate highly specific morphological features on conventional MRI, some studies have excluded these lesions from their analysis.[10,29] However, isolating these lesions could have biased the derived conclusions.

To test capability of spatial quantification techniques in recognizing a different environment of benign masses as a whole group from malignant masses, in the current work, two schemes were assessed for classification of benign and malignant adnexal lesions, one including and one excluding endometrioma, hemorrhagic cysts, and teratoma masses. We showed that in both classification schemes the RLM textural model could accurately separate benign and malignant groups (accuracy ~92% including all masses and 100% excluding highly specific benign masses), and benign, borderline, and malignant groups (accuracy ~88%). When considering all of the adnexal masses, addition of Gabor textural features to the RLM model increased the diagnostic accuracy to 100% (for classification of benign/malignant, benign/borderline/malignant, and borderline/malignant). Therefore, selection of the right analytical model is essential for identifying intra- and interclass differences and proposing a generative diagnostic approach, which could most probably diagnose any cases accurately. Discrimination of borderline from malignant tumors is important for preoperative decision-making about the treatment strategy undertaken for the patients with borderline tumors, as they can be scheduled for conservative surgery.[30] Nonetheless, this result may change when tested on a larger cohort.

Accordingly, this work was designed to find the model that best represents the spatial heterogeneity of adnexal masses. The models with higher accuracies were primarily constructed based on RLM and Gabor textural features, and the subset comprised purely of histogram features, such as ADC entropy,[12] could not perform accurately. The reason for the suboptimal performance of histogram in contrast to RLM and Gabor is that the resulting descriptive features can only globally assess the region under study and the local interrelationships of adjacent voxels are ignored.[31] RLM features in this application characterize the coarseness of cellular distributions in the specified directions and seem to realize the complexity of the underlying microstructure. Transformational Gabor features are derived by decomposing the ADC values into a range of constitutive frequencies, and therefore have the potential to detect differences in cellular distribution between masses. Gabor features could also provide reliable diagnosis of adnexal masses when endometriomas, hemorrhagic cysts, and teratomas were excluded, but only in combination with RLM features can an accurate model be formed when considering all adnexal masses.

Limitations of mean ADC for characterization of adnexal masses has encouraged many researchers to interrogate the role of other advanced MRI techniques, such as qualitative[7,26,29] and quantitative[24,32,33] dynamic contrast-enhanced (DCE) MRI. While these additional techniques have shown potential, employing a well-designed DWI acquisition integrated with an accurate postprocessing spatial quantification technique for helping the radiologists with their diagnosis may obviate the requirement for administration of contrast agents and could reduce the scan time.

Our study had a number of limitations to be addressed. This study was carried out on a relatively small patient population and validation was not performed on an independent dataset. Thus, to find a generic classification model, this method should be applied on a large population including a variety of tumor subtypes being imaged by different scanners. Furthermore, due to small sample size of the borderline patients ($n = 3$) included in our study, the performance of this method for differentiating borderline tumors should be confirmed in future works.

It should also be pointed out that DWI and ADC images are usually susceptible to lower quality than conventional MRI because of susceptibility artifact and the resulting low signal-to-noise ratio acquisitions. This challenge increases when the imaging site is outside the brain, where more tissue inhomogeneity as well as natural physiologic motion generate signal dephasing and image distortion. The authors checked the basic considerations to maintain the minimum necessary image quality for acceptable DW images by accurate application of shimming to increase field homogeneity and choosing the right EPI parameters for DWI acquisitions to avoid such susceptibility and motion artifacts.

In conclusion, spatial quantification of cellular heterogeneity showed the potential to appraise the attributing class of adnexal masses. This may allow for devising objective and accurate pattern recognition computer-aided diagnosis systems for differentiation of adnexal lesions.

Acknowledgments

Contract grant sponsor: Tehran University of Medical Sciences & Health Services; contract grant numbers: 21499, 32917

References

1. Spencer JA, Forstner R, Cunha TM, Kinkel K, Sub-Committee EFI. ESUR guidelines for MR imaging of the sonographically indeterminate adnexal mass: an algorithmic approach. Eur Radiol 2010;20:25–35.
2. Spencer JA, Ghattamaneni S. MR imaging of the sonographically indeterminate adnexal mass. Radiology 2010;256:677–694.
3. Adusumilli S, Hussain HK, Caoili EM, et al. MRI of sonographically indeterminate adnexal masses. AJR Am J Roentgenol 2006;187:732–740.
4. Sohaib S, Mills T, Sahdev A, et al. The role of magnetic resonance imaging and ultrasound in patients with adnexal masses. Clin Radiol 2005;60:340–348.
5. Spencer JA, Weston MJ, Wilkinson N. Integration of imaging and pathology in the multidisciplinary process. In: Wilkinson N, ed. Pathology of the ovary, fallopian tube and peritoneum. London: Springer; 2014. p 109–131.
6. Thomassin-Naggara I, Fedida B, Kermarrec E. Adnexal masses: Characterization of benign adnexal masses. In: Kauczor HU, Hricak H, Essig M, Brady LW, Lu JJ, eds. Medical radiology (continuation of: Handbuch der medizinischen Radiologie / Encyclopedia of Medical Radiology). Heidelberg, Berlin: Springer; 2017. p 1–13.
7. Forstner R, Thomassin-Naggara I, Cunha TM, et al. ESUR recommendations for MR imaging of the sonographically indeterminate adnexal mass: An update. Eur Radiol 2016:1–10.
8. Koh D-M, Collins DJ. Diffusion-weighted MRI in the body: Applications and challenges in oncology. AJR Am J Roentgenol 2007;188: 1622–1635.
9. Padhani AR, Liu G, Mu-Koh D, et al. Diffusion-weighted magnetic resonance imaging as a cancer biomarker: consensus and recommendations. Neoplasia 2009;11:102–125.
10. Thomassin-Naggara I, Daraï E, Cuenod CA, et al. Contribution of diffusion-weighted MR imaging for predicting benignity of complex adnexal masses. Eur Radiol 2009;19:1544–1552.
11. Bakir B, Bakan S, Tunaci M, et al. Diffusion-weighted imaging of solid or predominantly solid gynaecological adnexial masses: Is it useful in the differential diagnosis? Br J Radiol 2011;84:600–611.
12. Kierans AS, Bennett GL, Mussi TC, et al. Characterization of malignancy of adnexal lesions using ADC entropy: Comparison with mean ADC and qualitative DWI assessment. J Magn Reson Imaging 2013;37:164–171.
13. Rockall AG. Diffusion weighted MRI in ovarian cancer. Curr Opin Oncol 2014;26:529–535.
14. Yamamoto Y, Yamada R, Oguri H, Maeda N, Fukaya T. Comparison of four malignancy risk indices in the preoperative evaluation of patients with pelvic masses. Eur J Obstet Gynecol 2009;144:163–167.
15. Tingulstad S, Hagen B, Skjeldestad FE, et al. Evaluation of a risk of malignancy index based on serum CA125, ultrasound findings and menopausal status in the pre-operative diagnosis of pelvic masses. BJOG 1996;103:826–831.
16. Håkansson F, Høgdall EV, Nedergaard L, et al. Risk of malignancy index used as a diagnostic tool in a tertiary centre for patients with a pelvic mass. Acta Obstet Gynecol Scand 2012;91:496–502.
17. Just N. Improving tumour heterogeneity MRI assessment with histograms. Br J Cancer 2014;111:2205–2213.
18. Haralick RM. Statistical and structural approaches to texture. Proc IEEE 1979;67:786–804.
19. Chu A, Sehgal CM, Greenleaf JF. Use of gray value distribution of run lengths for texture analysis. Pattern Recognit Lett 1990;11:415–419.
20. Zacharaki EI, Wang S, Chawla S, et al. Classification of brain tumor type and grade using MRI texture and shape in a machine learning scheme. Magn Reson Med 2009;62:1609–1618.
21. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Selected papers of Hirotugu Akaike. Berlin: Springer; 1998. p 199–213.
22. Schwarz G. Estimating the dimension of a model. Ann Stat 1978;6: 461–464.
23. Kazerooni AF, Malek M, Haghighatkhah H, et al. Semiquantitative dynamic contrast-enhanced MRI for accurate classification of complex adnexal masses. J Magn Reson Imaging 2017;45:418–427.
24. Carter JS, Koopmeiners JS, Kuehn-Hajder JE, et al. Quantitative multiparametric MRI of ovarian cancer. J Magn Reson Imaging 2013;38: 1501–1509.
25. Kim H-J, Lee S-Y, Shin YR, Park CS, Kim K. The value of diffusion-weighted imaging in the differential diagnosis of ovarian lesions: A meta-analysis. PLoS One 2016;11:e0149465.
26. Thomassin-Naggara I, Aubert E, Rockall A, et al. Adnexal masses: Development and preliminary validation of an MR imaging scoring system. Radiology 2013;267:432–443.
27. Katayama M, Masui T, Kobayashi S, et al. Diffusion-weighted echo planar imaging of ovarian tumors: Is it useful to measure apparent diffusion coefficients? J Comput Assist Tomogr 2002;26:250–256.
28. Nakayama T, Yoshimitsu K, Irie H, et al. Diffusion-weighted echoplanar MR imaging and ADC mapping in the differential diagnosis of ovarian cystic masses: Usefulness of detecting keratinoid substances in mature cystic teratomas. J Magn Reson Imaging 2005;22:271–278.
29. Thomassin-Naggara I, Toussaint I, Perrot N, et al. Characterization of complex adnexal masses: Value of adding perfusion-and diffusion-weighted MR imaging to conventional MR imaging. Radiology 2011; 258:793–803.
30. Bentivegna E, Gouy S, Maulard A, et al. Fertility-sparing surgery in epithelial ovarian cancer: a systematic review of oncological issues. Ann Oncol 2016;27:1994–2004.
31. Kassner A, Thornhill R. Texture analysis: A review of neurologic MR imaging applications. AJNR Am J Neuroradiol 2010;31:809–816.
32. Fathi Kazerooni A, Nabil M, Haghighat Khah H, Parviz S, Gity M, Saligheh Rad H. A one-step biomarker quantification methodology for DCE-MRI of adnexal masses: Capturing kinetic pattern from early to late enhancement. Magn Reson Med 2017; in press. doi: 10.1002/mrm.26743.
33. Thomassin-Naggara I, Balvay D, Aubert E, et al. Quantitative dynamic contrast-enhanced MR imaging analysis of complex adnexal masses: a preliminary study. Eur Radiol 2012;22:738–745.