

Temporal Analysis of Tumor Heterogeneity and Volume for Cervical Cancer Treatment Outcome Prediction: Preliminary Evaluation

Jeffrey W. Prescott,¹ Dongqing Zhang,² Jian Z. Wang,² Nina A. Mayr,² William T.C. Yuh,³ Joel Saltz,¹ and Metin Gurcan¹

In this paper, we present a method of quantifying the heterogeneity of cervical cancer tumors for use in radiation treatment outcome prediction. Features based on the distribution of masked wavelet decomposition coefficients in the tumor region of interest (ROI) of temporal dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) studies were used along with the imaged tumor volume to assess the response of the tumors to treatment. The wavelet decomposition combined with ROI masking was used to extract local intensity variations in the tumor. The developed method was tested on a data set consisting of 23 patients with advanced cervical cancer who underwent radiation therapy; 18 of these patients had local control of the tumor, and five had local recurrence. Each patient participated in two DCE-MRI studies: one prior to treatment and another early into treatment (2–4 weeks). An outcome of local control or local recurrence of the tumor was assigned to each patient based on a posttherapy follow-up at least 2 years after the end of treatment. Three different supervised classifiers were trained on combinational subsets of the full wavelet and volume feature set. The best-performing linear discriminant analysis (LDA) and support vector machine (SVM) classifiers each had mean prediction accuracies of 95.7%, with the LDA classifier being more sensitive (100% vs. 80%) and the SVM classifier being more specific (100% vs. 94.4%) in those cases. The K-nearest neighbor classifier performed the best out of all three classifiers, having multiple feature sets that were used to achieve 100% prediction accuracy. The use of distribution measures of the masked wavelet coefficients as features resulted in much better predictive performance than those of previous approaches based on tumor intensity values and their distributions or tumor volume alone.

KEY WORDS: Cervical cancer, treatment outcome prediction, dynamic contrast-enhanced MRI, wavelet

INTRODUCTION

Cervical cancer is the second most incident neoplastic disease worldwide¹, trailing only

breast cancer. Even with an aggressive treatment plan, the mortality rate is 34% in the USA and other developed countries and 50% in developing countries. Radiation therapy is the current modality recommended for treatment of advanced cervical cancer (stages IB₂–IVB of the International Federation of Gynecology and Obstetrics guidelines)².

The use of tumor-imaging studies, such as dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), can provide information about the tumor environment that is important for treatment planning and delivery³. It has been shown that the intensity responses of a DCE-MRI scan are related to the oxygen content, and therefore the vascularity, in different areas of the tumor⁴. The lower intensity regions on a DCE-MRI scan correspond to areas of decreased vascularity or even necrosis. The resultant reduction in perfusion hinders oxygen delivery to the tumor. This inhibits the production of oxygen-free

¹From the Department of Biomedical Informatics, The Ohio State University, 333 W. 10th Ave., Columbus, OH, 43210, USA.

²From the Department of Radiation Medicine, The Ohio State University Medical Center, 300 W. 10th Ave, Columbus, OH, 43210, USA.

³From the Department of Radiology, The Ohio State University Medical Center, 607 Means Hall, 1654 Upham Dr., Columbus, OH, 43210, USA.

Correspondence to: Metin Gurcan, Department of Biomedical Informatics, The Ohio State University, 333 W. 10th Ave., Columbus, OH, 43210, USA; tel: +1-614-2921084; fax: +1-614-6886600; e-mail: gurcan.1@osu.edu

Copyright © 2009 by Society for Imaging Informatics in Medicine

Online publication 27 January 2009

doi: 10.1007/s10278-009-9179-7

radicals which are the effectors of radiation therapy. Hence, the tumor may not respond to treatment. The early identification of poor treatment response allows for more aggressive treatment with the current treatment modality or the possible addition of adjuvant therapies.

The use of DCE-MRI for quantification of tumor vascularity, tumor staging, or treatment efficacy has been explored for cancers of the prostate^{5,6}, colon⁷, liver⁸, breast⁹, and cervix. For the case of cervical cancer, the change in volume over the first 2–4 weeks of therapy has been shown to be related to treatment outcome¹⁰. In addition, the kinetics of DCE-MRI response has also been employed for outcome prediction^{11,12}. These parameters are calculated either voxel by voxel or using an entire region of interest (ROI). It should be noted that, for our current data set, a thorough kinetic analysis was not feasible since the DCE-MRI studies for the participating patients consisted of only two time points—one just prior to contrast injection and another 75 s after injection.

Previous work on this data set focused on the calculation of intensity distribution values, such as mean, standard deviation, skewness, and kurtosis, with the addition of fractal dimension as a measure of texture^{13,14}. These features were then used, along with the temporal relationships between the multiple studies and series for each patient, to predict the outcome of radiation treatment. The use of intensity distribution measures allowed for the quantification of the overall response to tumor perfusion. However, much of the tumor micro-environment information is captured in DCE-MRI by small, local variations in image intensity. A drawback of basic intensity distribution calculations is that local variations may be inadequately quantified. The fractal dimension measure was included with the intensity distribution characteristics in order to offset this shortcoming. As a discriminating feature for treatment outcome prediction, it performed well. However, the overall prediction performance of the feature sets was marginal. It was concluded that a more thorough texture analysis may be better suited to capture the local variation rather than a simple intensity distribution analysis. In this paper, the wavelet decomposition is used to robustly extract information about the small, local variations or texture of the DCE-MRI images. In order to accurately analyze the variation only within the tumor, we

proposed the use of tumor masks to identify texture components only within the region of interest, while not contaminating the results from information coming from the surrounding tissue.

MATERIALS AND METHODS

Data

The data set consisted of 23 patients with cervical cancer who underwent radiation therapy. The data was collected with a protocol approved by the Ohio State Institutional Review Board. The data set included 18 patients with an outcome of local control of the tumor and five with local recurrence. Local control and local recurrence were defined as the absence or presence, respectively, of the cancer on biopsy during posttherapy follow-up. The minimum follow-up for patients in this study was 2 years.

The patients underwent multiple DCE-MRI studies (Table 1). Two T1-weighted DCE-MRI studies, consisting of sagittal scans of the abdominal/pelvic area (resolution, 256×256), were performed for each patient—one pretreatment and the other during early treatment (2–4 weeks). The volume of the gadolinium-chelated contrast agent used for the studies was 0.1 mmol/kg, injected at a rate of 5 mL/s. There were two subsets of patients with respect to the study setup. The first subset consisted of patients who had two series in each study—one before injection of contrast agent and another 75 s after injection. The second subset

Table 1. Summary of the Data Set Used in the Analysis

Patients	23 (18 local control, 5 local recurrence)
Patient age (years)	Minimum, 31; maximum, 88; mean, 51
Studies	Pretreatment Early treatment (2–4 weeks)
Series	Precontrast injection (0 s) Postcontrast injection (75 s)
Number of slices	Minimum, 12; maximum, 14
Size of slices (voxels × voxels)	256 × 256
Resolution (mm/voxel)	X, 1.5625; Y, 1.5625; Z, 8.0000
Contrast agent	Gadolinium chelate, 0.1 mmol/kg; rate 5 mL/s
Signal weighting	T1

consisted of patients who had between 13 and 14 series per study, with a pre-injection series followed by multiple post-injection series, each 25 s apart. All other study protocols were the same between the two patient subsets. In order to provide uniformity across the database, the data set that was used for the analysis consisted of the DCE-MRI series directly before injection of contrast agent and a second series 75 s after the injection from each patient subset (Fig. 1).

Only one time point of the time-activity DCE-MRI curve was used for the data analysis. The imaging data was acquired using the DCE-MRI technique, and the plateau phase of the DCE curve

was based upon the single time point in the current study. This specific point was chosen to estimate the plateau signal intensity (SI) of the DCE-MRI curve after bolus injection of the contrast agent. The correlation of DCE-MRI and the vascular density and effectiveness of delivery of cytotoxic agents were mostly based upon the plateau SI of the DCE-MRI curve but not the rate of washout. The plateau SI usually becomes relatively steady between 50 and 100 s of the DCE curve¹¹. Therefore, we chose to analyze the DCE-MRI response at the 75-s time point.

The images in Figure 1 are taken from the central, largest slices of two patients' tumors in the

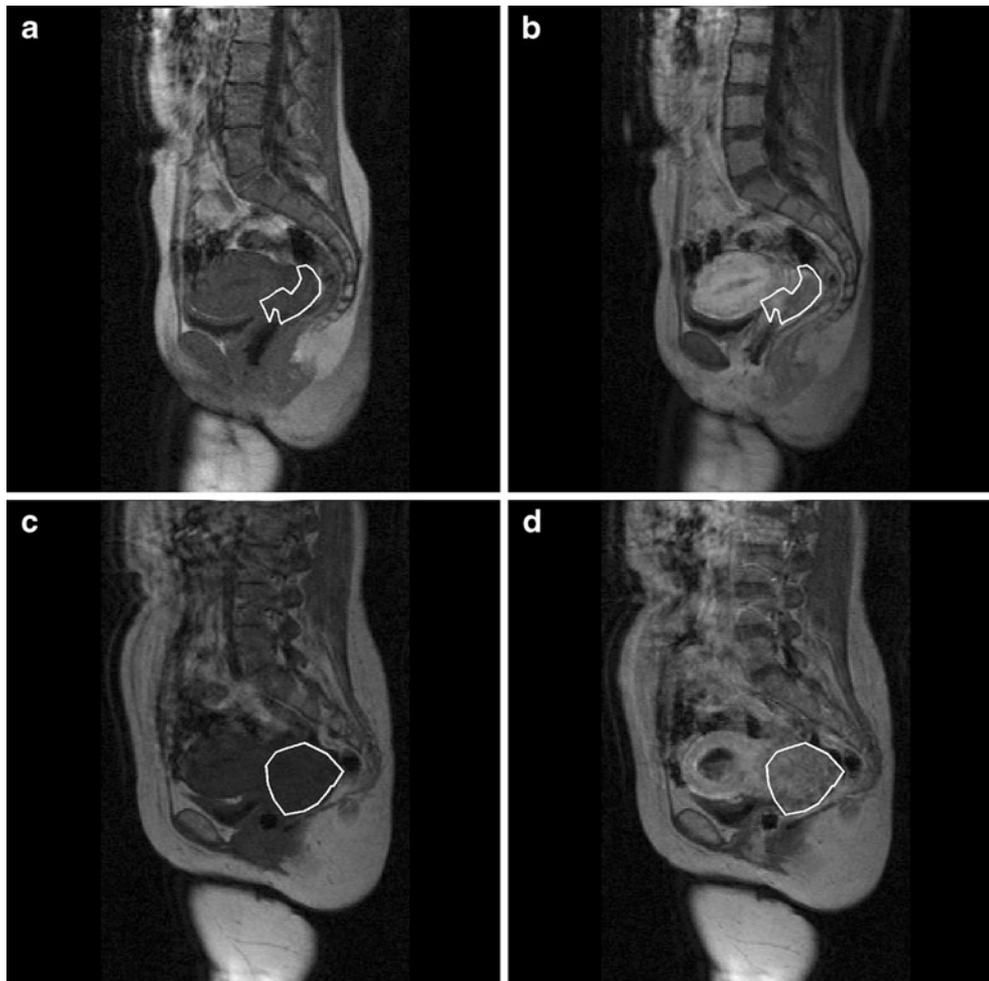


Fig. 1. Central sagittal slices of tumors from two different patients. a, b Pre- and postcontrast images, respectively, of patient with local recurrence of the tumor. c, d Pre- and postcontrast images, respectively, of patient with local control of the tumor. Notice that the tumor of the first patient, shown in a, b, is smaller than the tumor of the second patient, shown in c, d. This difference demonstrates that tumor volume information may not be sufficient for accurate treatment outcome prediction.

sagittal plane. Figure 1 demonstrates that the volume of a tumor is not sufficient for prediction of treatment outcome. In this figure, the patient with the smaller tumor volume (Fig. 1a, b) actually experienced local recurrence of the tumor, while the patient with the larger tumor (Fig. 1c, d) had local control. This difference highlights the importance of complementing volume information with spatial information in the prediction of tumor response to treatment.

A T2-weighted sagittal scan was also acquired for use in delineation of the tumor region of interest. The T2-weighted scan exhibited better contrast between the cervical tumor and the surrounding structures than the T1-weighted scan. This allowed for a more accurate manual segmentation of the tumor.

Preprocessing

For each study, the regions of interest were manually delineated for each anatomical scan image that included a portion of the tumor. The trained readers who marked the images were blinded to all patient information, including treatment plan and outcome. A single reader marked tumor ROIs for all 23 patients, while two others marked ten patients each in order to quantify inter- and intrareader reproducibility for tumor delineation. The mean interreader similarities for each combination of readers, measured using the Zijdenbos similarity index¹⁸, were 0.80, 0.72, and 0.76. The intrareader variability calculated for a single reader who created two separate markings was found to be 0.85. An index value greater than 0.7 corresponds to an “excellent” agreement. The remaining parts of the analysis were then carried out using the markings from the first reader, who marked all 23 cases.

The ROIs drawn on each anatomical scan slice were then matched with the corresponding slice in each of the DCE-MRI series. In addition, 12 studies required registration of the DCE-MR images to the anatomical images due to movement of the patient between the different scans. These Euclidean registrations were performed manually by identifying six salient points between the vertebral columns of the images to establish the correspondence, using Matlab’s (Mathworks, Natick, MA, USA) registration tool (Fig. 2). The amount of rotation and translation were then

estimated using each set of two two-point correspondences. The means of these rotations and translations were used as the final transformation for registration.

The Wavelet Transform

The discrete wavelet transform (DWT) provides a method of describing a signal in the set of square-integrable signals, L_2 , by use of a finite-length signal known as a *wavelet*. A wavelet can be scaled and shifted to produce an orthonormal basis for vector spaces V_k , which are subspaces (approximations) of L_2 , such that $\lim_{k \rightarrow -\infty} V_k = L_2$. In this way, the DWT is a *multiresolution* method of signal representation.

Given an L_2 signal, $x(t)$, and a wavelet basis $\{\psi_{k,n}(t) : k, n \in Z\}$, the DWT transform pair is

$$x(t) = \sum_{k=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} d_{k,n} \psi_{k,n}(t)$$

$$d_{k,n} = \langle \psi_{k,n}(t), x(t) \rangle = \int \psi_{k,n}^*(t) x(t) dt$$

where the $\{d_{k,n}\}$ are the wavelet coefficients. Therefore, the wavelet coefficients representing the original signal are found by projecting the original signal onto the wavelet basis. For the case of two-dimensional images (i.e., signals in R^2), local variations of particular orientations can be extracted by using the filterbank approach of separable multiresolution approximations¹⁵.

The single-level two-dimensional wavelet decomposition was calculated for each DCE-MR image (Fig. 3). Each decomposition resulted in four sets of coefficients: approximation (A), horizontal detail (H), vertical detail (V), and diagonal detail (D). The nomenclature of the coefficient sets (A, H, V, D) used in this paper is for descriptive purposes. The naming relates to the standard nomenclature as such: A = LL, H = LH, V = HL, D = HH, where LL represents the output from the lowpass–lowpass filter chain, LH from lowpass–highpass, HL from highpass–lowpass, and HH from highpass–highpass.

The decomposition was calculated separately for five different mother wavelets: Daubechies (db) wavelets: 1, 2, 4, 8, and 16. Each patient had two series, precontrast (PrC) and postcontrast (PoC), in each of two studies, pretreatment (PT) and early

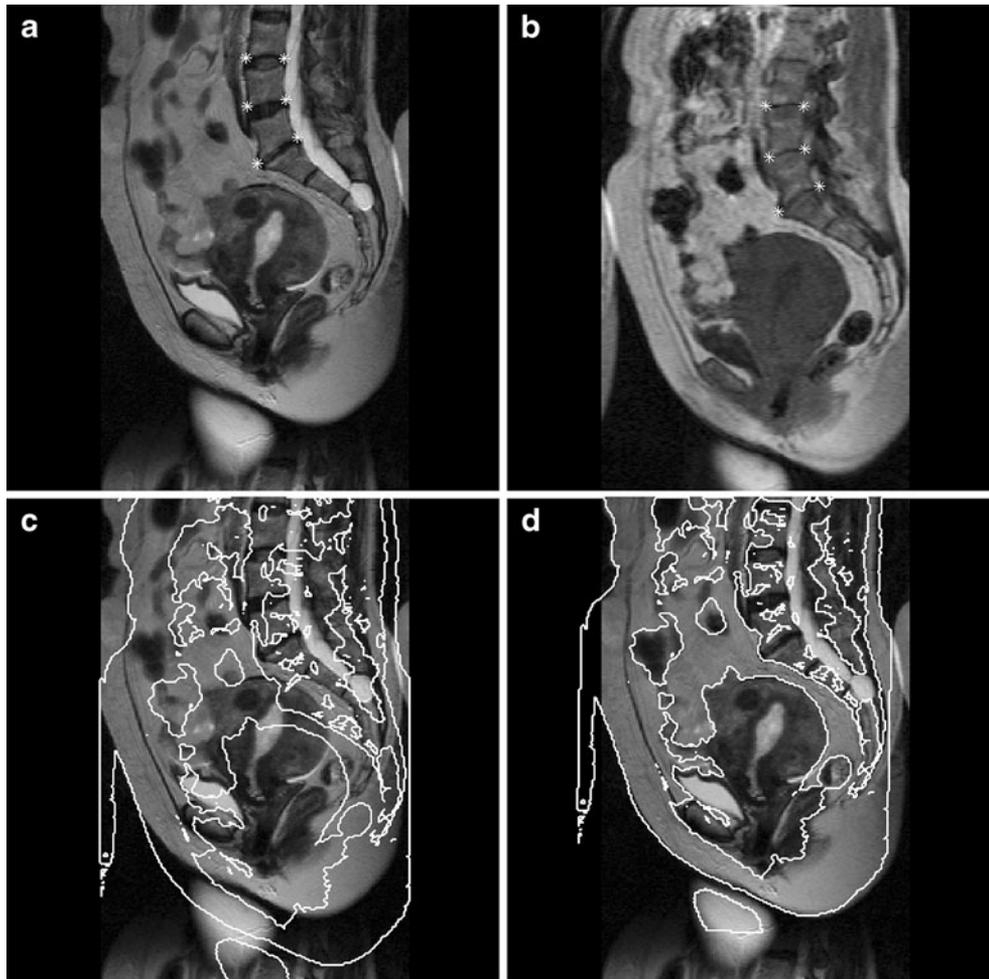


Fig. 2. a, b Salient point correspondences (marked by asterisks) between anatomical and DCE scans, respectively; c initial alignment; d registered alignment using transformation calculated from a, b point correspondences.

treatment (ET). Therefore, each patient had four series of images which underwent wavelet decomposition for each mother wavelet. The wavelet analysis was performed with Matlab's wavelet toolbox `dwt2` function, using symmetric padding.

Feature Extraction

A diagram of the feature extraction process is shown in Figure 4. The wavelet transform is performed on each image in a series in which a portion of the tumor appears. The result is four sets of coefficients—A, H, V, and D—for each image. These coefficient sets retain the same two-dimensional structure as the original images. However, each of the sets are one fourth of the size of the

original image as a result of the downsampling of the original image by 2 in both the vertical and horizontal dimensions. The ROIs that were defined for each image were also downsampled by 2 in each dimension and used to mask the wavelet transform coefficients. The result was a collection of masked wavelet coefficient sets which corresponded to the tumor ROIs. The coefficients for all images in a series were then pooled into A, H, V, and D coefficient sets for the entire tumor volume. Statistics were calculated for these pooled distributions. The results were the wavelet features used for classifier training.

It should be emphasized that the wavelet decomposition was performed before masking of the tumor ROI. It was found that ROI masking

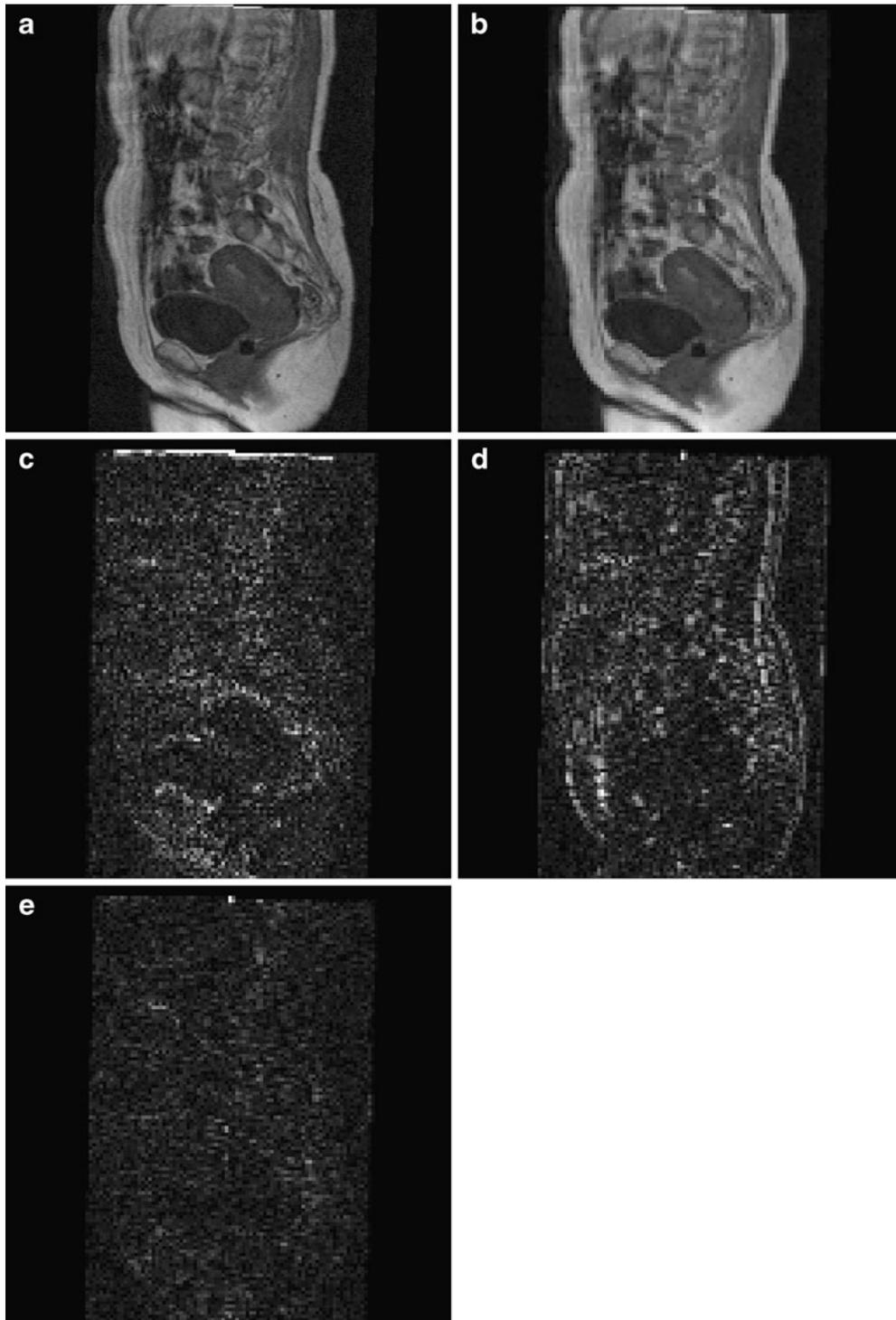


Fig. 3. Example of the wavelet decomposition performed on a DCE-MRI scan a using the db2 mother wavelet. b Absolute values of the approximation coefficients, A, c horizontal detail coefficients, H, d vertical detail coefficients, V, and e diagonal detail coefficients, D.

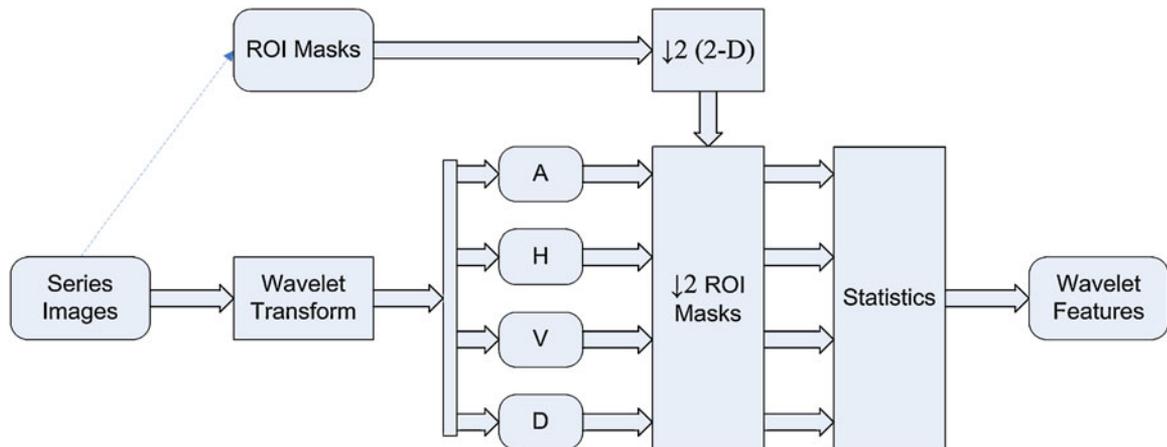


Fig. 4. Diagram of methodology of wavelet feature extraction from a DCE-MRI series.

followed by the wavelet decomposition produced heightened wavelet coefficient values around the edge of the delineated tumor (Fig. 5). This was due to the artificial edge around the tumor created by

the masking. However, masking the ROI after decomposition could lead to inclusion of pixels outside of the ROI in the calculation of decomposition coefficients—an effect that would be in-

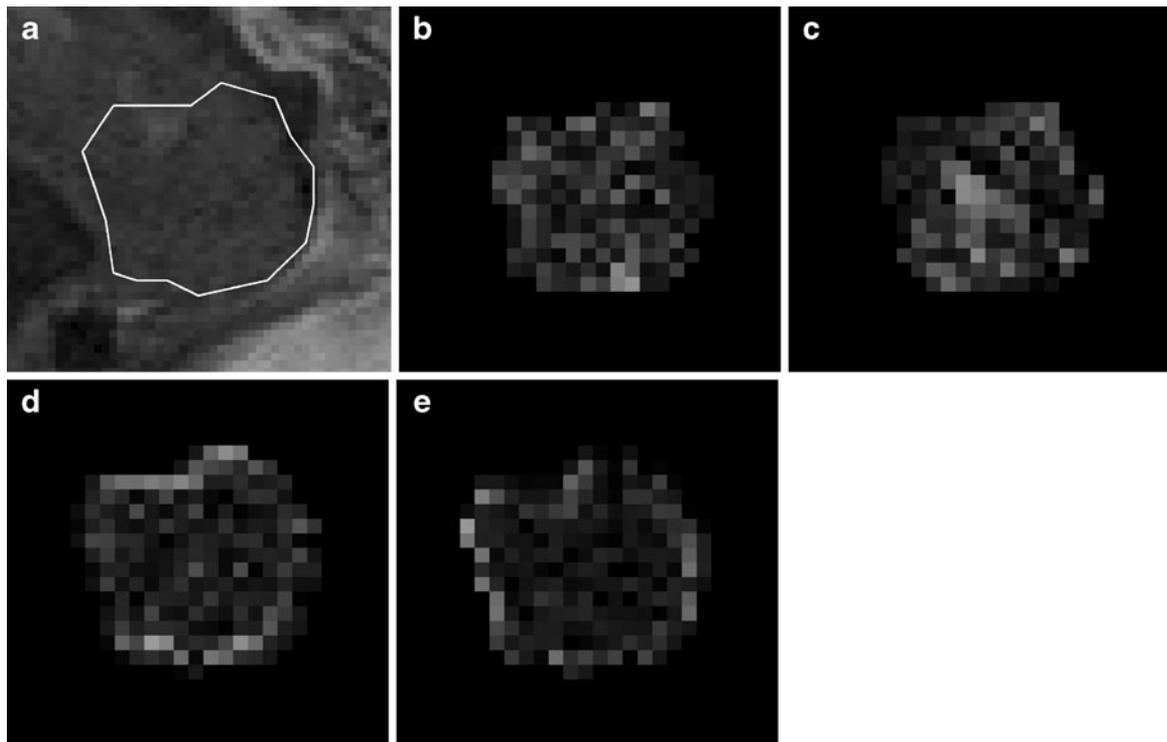


Fig. 5. a Tumor ROI shown on a T1-weighted DCE scan, 75 s after contrast injection; b, c absolute values of the horizontal and vertical detail coefficients, respectively, when wavelet decomposition is performed *before* masking of the tumor ROI; d, e absolute values of the horizontal and vertical detail coefficients, respectively, when wavelet decomposition is performed *after* masking of the tumor ROI. Notice that the outer pixels of d, e have greater intensity than the pixels in the rest of the tumor. This is due to the “false edge” created by the masking of the tumor before calculating the wavelet decomposition. This is in contrast to the coefficient maps in b, c, which show no heightened intensity edges.

creased with increasing length of the filter used to represent the mother wavelet (e.g., db2 filter length is 4, while db1 filter length is 2). This effect was partially resolved by removing the outer edge of pixels in the delineated tumor ROI region. However, the inclusion of intensity responses of structures outside the ROI is not altogether unwarranted, due to the intimate relationship between a tumor and its surrounding environment. Therefore, masking after decomposition focused mainly on the tumor ROI without introducing “artificial” responses, while at the same time including a small amount of information about the surrounding environment. Performing the masking after the wavelet decomposition and subsequently removing the outer edge of pixels in the ROI is a novel method of focusing on the interior of the tumor for the feature extraction while including some information about the immediate surrounding environment.

The different decomposition coefficients (A, H, V, D) from each image in a given series were pooled with the decomposition coefficients from other images in that series to produce four sets of coefficient distributions representing the entire tumor volume. In addition, the detail coefficients

(H, V, D) were pooled to create a fifth coefficient distribution. Therefore, each series had a set of five different coefficient distributions for a given mother wavelet.

Since each patient had four series, and each series underwent wavelet decomposition with five different mother wavelets, each producing five coefficient distributions, the total number of wavelet distributions produced for a given patient was 100. Characteristics of these distributions were then calculated, to be used as features in the prediction of treatment outcome. The distribution characteristics calculated were mean, median, mode, standard deviation, skewness, and kurtosis (Fig. 6). The higher-order statistics (skewness and kurtosis) were included as previous studies have shown the efficacy of these distribution measures in characterization of pathological images¹⁶. The calculation of six features from each of 100 wavelet coefficient distributions resulted in 600 total wavelet features. With the addition of the two volume measurements (one PT and one ET), each patient was associated with 602 *total features*.

Due to this large number of features, the analysis was separated into smaller parts based on the mother wavelet and the coefficient set. For

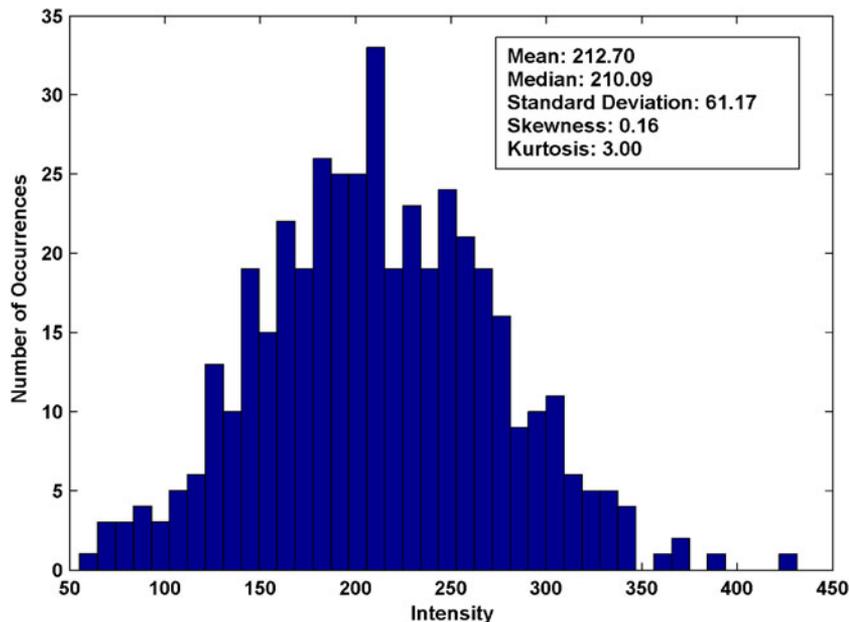


Fig. 6. Histogram and distribution statistics of approximation coefficients, A, from a wavelet decomposition using the db2 mother wavelet. The original images were from a precontrast series of an early treatment study.

each patient, each set of 24 series–wavelet–coefficient features (four series, one mother wavelet, one coefficient set, six features) was paired with the PT and ET volume values for use in classifier training, testing, and comparison. For example, the distribution parameters (mean, median, etc.) calculated from the approximation coefficients (A) of the wavelet analysis using the db1 mother wavelet on each of the four series (PT-PrC, PT-PoC, ET-PrC, ET-PoC) along with the PT and ET tumor volumes of each patient were used as one set of 26 features to train the classifiers.

Dimensionality Reduction and Feature Selection

The feature sets used for classifier training were pulled from the complete set of 24 wavelet and two volume features for each mother wavelet and coefficient set. Each possible combination of one, two, or three features was used in order to exhaustively search for the best-performing feature subset. The feature space was limited to three dimensions due to the small sample size and in order to focus on strong relationships between the input features and the output treatment outcome classification, thereby avoiding the “curse of dimensionality.”¹⁷ In the case of the three-feature sets, the dimensionality of the set was first reduced from three to two using principal components analysis (PCA) before running the exhaustive search for the best-performing set of features¹⁷.

Classifiers

The linear discriminant analysis (LDA), support vector machine (SVM), and K-nearest neighbor (KNN) classifiers were used for treatment outcome prediction using the calculated wavelet features¹⁷. The classifiers were trained using a leave-one-out methodology, in which a single sample was left out for testing, while the rest were used in classifier training. The training and testing followed a round-robin format in which each of the 23 patients was tested once on a classifier trained using the remaining patients.

The KNN classifier assigns a class to a sample based on a majority vote of the K closest samples (nearest neighbors), where “closeness” is defined by a distance metric. In the simplest case, the vote of a neighbor is simply its assigned class. In this

analysis, the Euclidean distance metric was used, and the number of neighbors was set to one. Therefore, the tested sample was assigned the class of its closest neighbor in Euclidean space.

The LDA classifier is based on a discriminant function, g , which has the following form

$$g(\mathbf{x}) = \mathbf{w}^t \mathbf{x} + w_0$$

Where \mathbf{x} is the *feature vector*, \mathbf{w} is the *weighting vector*, and w_0 is the *bias or threshold weight*. Given two classes, w_1 and w_2 , the weighting vector is the normal to the maximal class-separating hyperplane. It is calculated with the assumption that the underlying distributions of the classes are normal. For testing, a sample, \mathbf{x}_t , is assigned to class w_1 if $g(\mathbf{x}_t) > 0$ or class w_2 if $g(\mathbf{x}_t) < 0$. For a point lying on the hyperplane ($g(\mathbf{x}_t) = 0$), the sample can be assigned to either class.

The SVM classifier is based on a preprocessing of the data to a higher dimension, using a nonlinear mapping function. A discriminant function is then calculated by maximizing the *margin* or the distance between the discriminant function and the closest samples in either class. A set of support vectors are then chosen to represent the separating function. For our analysis, the SVM classifier was implemented using a linear kernel.

Performance Measures

The performance of each feature set used to train the classifiers was quantified by calculating the prediction accuracy, sensitivity, and specificity for the set of 23 leave-one-out tests. The prediction accuracy was calculated as the total number of correct classifications divided by the total number of tests. The sensitivity was calculated as the number of correctly classified local recurrence patients divided by the total number of actual local recurrence patients in the data set. The specificity was calculated as the number of correctly classified local control patients divided by the total number of actual local control patients in the data set.

RESULTS

In this study, it is important to select the best-performing set of features. The features change

Table 2. Conversion Table, for Use with Table 3, with Each Feature and Corresponding Feature Number

Study	Series	Feature	Index
PT	n/a	Volume	1
		PrC	Mean
	Median		3
	Mode		4
	Std		5
	Skewness		6
	Kurtosis		7
	PoC	Mean	8
		Median	9
		Mode	10
		Std	11
		Skewness	12
		Kurtosis	13
ET	n/a	Volume	14
		PrC	Mean
	Median		16
	Mode		17
	Std		18
	Skewness		19
	Kurtosis		20
	PoC	Mean	21
		Median	22
		Mode	23
		Std	24
		Skewness	25
		Kurtosis	26

All features, except for volume, are based on the distribution of the wavelet coefficients

PT pretreatment, ET early treatment, PrC precontrast, PoC postcontrast, Std standard deviation, n/a not applicable

Fig. 7. Histograms showing the number of occurrences of each feature, listed in Table 2, that appeared in the best-performing feature sets (of Tables 3, 4, and 5) for the a LDA, b SVM, and c KNN classifiers.

based on the set of coefficients and mother wavelets. Table 2 lists all combinations of features used and their corresponding indices. Table 3 shows the best-performing feature sets, given a particular set of wavelet coefficients.

Due to the size of the data set, we selected up to three features for classification. Experimentally, we observed that, regardless of the selected features, the classifiers trained and tested using three features performed better than the classifiers trained using sets of one or two features. The performance was measured in terms of mean prediction accuracy, sensitivity, and specificity. Therefore, the data presented in this section will focus on the case of three-feature sets. In the case of multiple feature sets having equivalent mean prediction accuracy, the feature set which had the highest sensitivity was chosen as the best set. If there were still equivalent sets, then the specificity was used as a tie-breaker. Finally, those that performed equally in areas of mean prediction accuracy, sensitivity, and specificity were listed in the tables as co-best performers (in curly brackets).

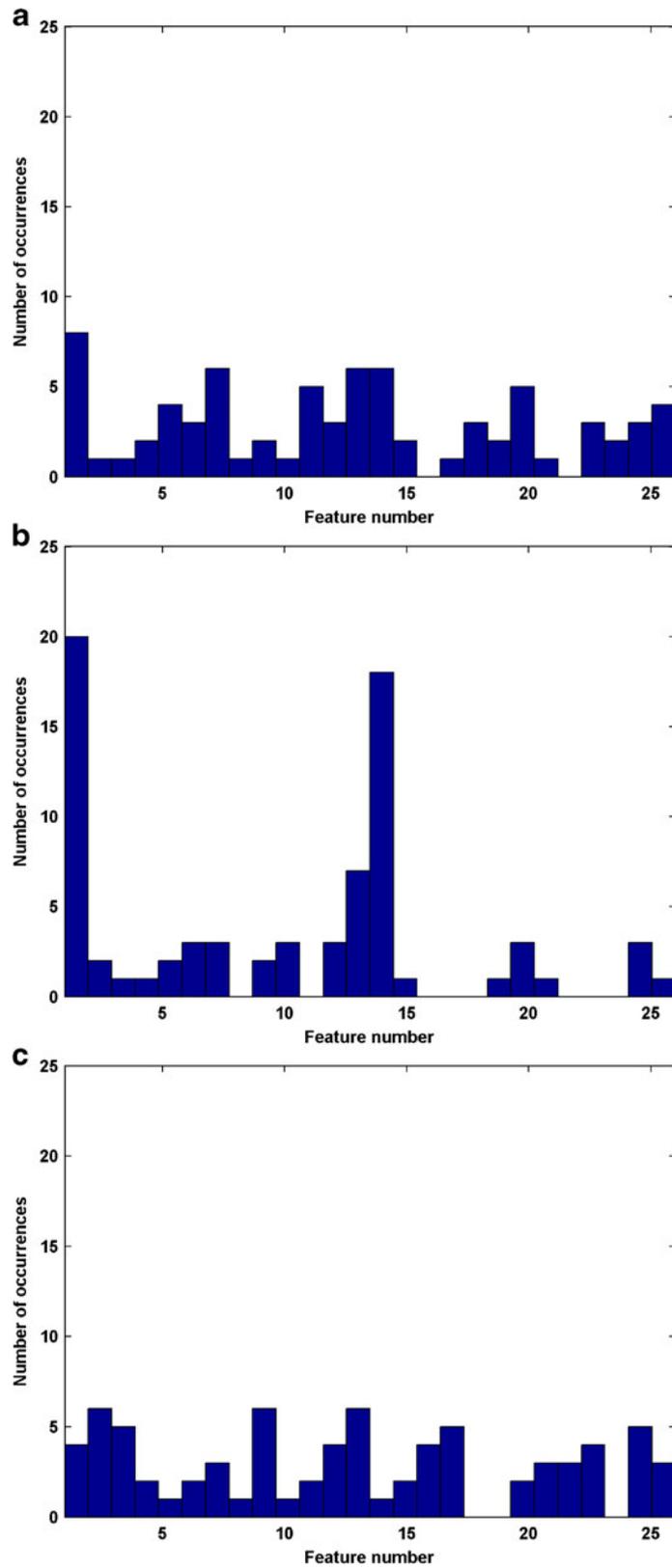
The histograms in Figure 7 show the distribution of features that participated in the best-performing

Table 3. LDA, SVM, and KNN Classifiers, Best-Performing Mother Wavelets, and Feature Sets for a Given Coefficient Set, Based on Best Mean Prediction Accuracy

Classifier	Coefficient set	Wavelet(s)	Prediction accuracy			Feature set(s)
			Mean (%)	Sensitivity (%)	Specificity (%)	
LDA	A	db2	91.3	100.0	88.9	6,12,20
	H	db2	91.3	100.0	88.9	13,18,24
	V	db2	95.7	100.0	94.4	11,20,23
	D	db1	91.3	80.0	94.4	1,6,14
	H, V, D	db8	95.7	100.0	94.4	1,9,14
SVM	A	db2	95.7	80.0	100.0	12,20,26
	H	db{1;4}	91.3	60.0	100.0	1,13,{14;7}
	V	db{2,8;16}	87.0	60.0	94.4	{9,10,20; 4,7,12; 1,6,14}
	D	db8	95.7	80.0	100.0	2,10,19
	H, V, D	db8	91.3	80.0	94.4	1,9,14
KNN	A	db{2;4}	95.7	100.0	94.4	12,13,{9; 10}
	H	db{8;16}	95.7	100.0	94.4	{7,9,16; 3,4,25}
	V	db8	100.0	100.0	100.0	6,11,12
	D	db{1;4}	100.0	100.0	100.0	{1,4,23; 3,9,16}
	H, V, D	db{1;4;8;16}	95.7	80.0	100.0	{2,15,17; 8,17,20; 9,22,25; 17,22,25}

See Table 2 for feature hash table

A approximation, H horizontal, V vertical, D diagonal, H, V, D horizontal-vertical-diagonal



classifiers in Table 3. For the LDA and SVM classifiers, the pretreatment and early treatment tumor volumes are well performing discriminatory features for local control or local recurrence. The pretreatment kurtoses, both precontrast and post-contrast, are also shown to be important features in the training of LDA classifiers. The distribution of discriminatory features for the KNN classifier is more evenly spread than for the LDA and SVM classifiers, suggesting that most of the feature data may have a greater propensity to clustering than to linear separation.

The LDA classifier had two best-performing feature sets. The first set was calculated using the vertical detail coefficients (V) of the wavelet decomposition with the db2 mother wavelet, with measures of PT-PoC standard deviation, ET-PrC kurtosis, and ET-PoC mode. The second set was calculated from the db8 wavelet's horizontal-vertical-diagonal (H, V, D) combined detail coefficient distribution and consisted of the PT volume, ET volume, and ET-PoC median of the coefficient distribution. Both feature sets were used to train LDA classifiers with a mean prediction accuracy of 95.7%, a sensitivity of 100% (5/5 positives), and a specificity of 94.4% (17/18 negatives). There were also two best-performing SVM feature sets. The first was created by the db2 wavelet's approximation (A) coefficient distribution and included the PT-PoC skewness, ET-PrC kurtosis, and ET-PoC kurtosis. The second feature set was calculated from the db8 wavelet's diagonal detail coefficients (D) and included the PT-PrC mean, PT-PoC mode, and ET-PrC skewness measures. The mean prediction accuracy was 95.7% with 80% (4/5) sensitivity and 100% (18/18) specificity using either feature set. Finally, the KNN classifier had several best-performing sets, which are more easily referenced in Table 3 than reproduced here.

The tests were also rerun, leaving the volume feature out, in order to analyze the texture features as discriminating features in their own right. It was found that the KNN classifier performed just as well as with volume measurements, which was expected. The histogram in Figure 7c shows that volume was a relatively less important feature for this classifier. The LDA and SVM classifiers fared worse, with both classifiers' average performance decreasing due to approximately one to two more misclassifications. Finally, simulations were run

that included the mean intensity of the entire tumor volume along with the wavelet and volume features. It was found that the addition of the mean intensity did not improve the performance of the classifiers, i.e., there was no change in mean prediction performance, sensitivity, and specificity.

DISCUSSION

The LDA and SVM classifiers exhibit similar performances, in some cases using the same feature sets. However, it is interesting to note that, while both classifiers in general exhibit nearly equivalent mean prediction accuracies, the LDA classifier has better sensitivity, and the SVM has better specificity. The KNN classifier produces better results than both the LDA and SVM classifiers when judged by the combination of mean prediction accuracy, sensitivity, and specificity.

As can be seen in Table 3, the best-performing classifiers were in general trained on features with wavelet coefficients that were calculated using the db1, db2, db4, and db8 mother wavelets. The increase in numbers in the Daubechies wavelet family ($n=1, 2, 4, \text{etc.}$) signifies the increasing filter length for implementation of the wavelet (filter length= $2n$). The increasing filter length can be thought of as an increasing neighborhood surrounding a given pixel that affects the wavelet coefficients associated with that pixel. Thus, it appears that a small neighborhood would provide the best local information for a given tumor pixel. This may also be due to better modeling of the tumor intensity profiles by these mother wavelets.

The size of the feature set was also important with respect to the classifier performance. The use of three features invariably resulted in classifiers which outperformed those classifiers trained on one or two features. This may be due to a synergistic combination of pretreatment and early treatment studies along with precontrast and post-contrast series, as exemplified by the case of the best-performing KNN classifier which used the feature set of PT-PrC median, PT-PoC median, and ET-PrC median. These temporal and perfusion response relationships cannot be extracted using only one or two features.

As can be seen in Table 3, the most discriminatory features among all the classifiers were the tumor volume and the median and kurtosis of the

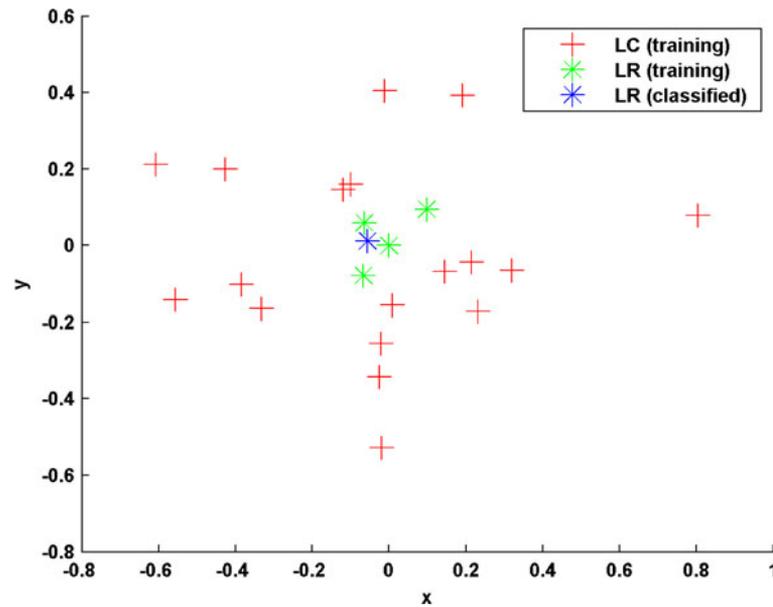


Fig. 8. Plot showing distribution of patient features of KNN classifier best-performing three-feature set of PT-PrC median, PT-PoC median, and ET-PrC median calculated from db4 wavelet decomposition diagonal coefficients. *LC* local control, *LR* local recurrence. The values on the *x*- and *y*-axes correspond to linear combinations of the original three-feature set after reduction to two dimensions by PCA.

wavelet coefficient distributions. The usefulness of tumor volume in prediction of radiation treatment outcome for cervical cancer has been shown previously¹⁰. The utility of the median and kurtosis of the wavelet coefficient distributions, however, is

a novel approach. The discriminatory power of both features suggests that the value of outliers is not as essential as the number of outliers (median, as opposed to mean) and the density about a central value (kurtosis).

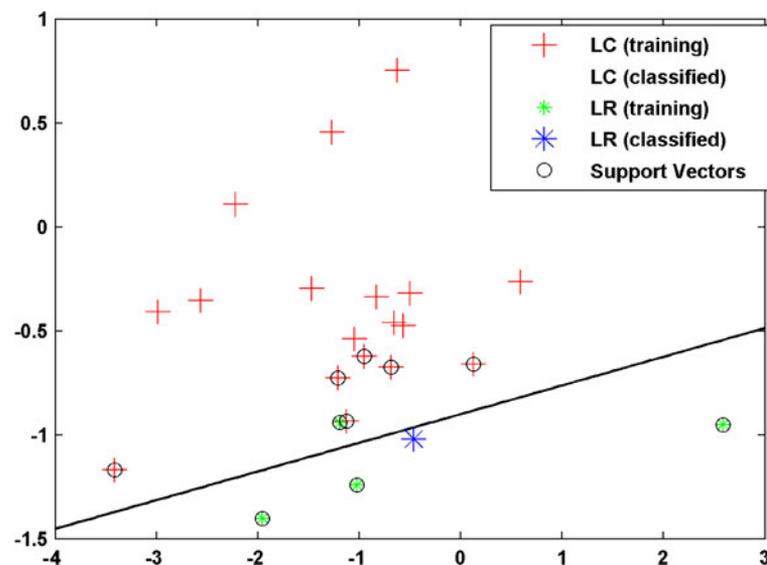


Fig. 9. Plot showing distribution of patient features of SVM classifier best-performing three-feature set of PT-PoC skewness, ET-PrC kurtosis, and ET-PoC kurtosis calculated from db2 wavelet decomposition approximation coefficients. *LC* local control, *LR* local recurrence. The values on the *x*- and *y*-axes correspond to linear combinations of the original three-feature set after reduction to two dimensions by PCA.

Table 4. Best-Performing Feature Sets for the LDA, SVM, and KNN Classifiers Using Measures of the Absolute Intensity Distribution

Classifier	Num features	Prediction accuracy			Feature set
		Mean (%)	Sensitivity (%)	Specificity (%)	
LDA	1	78.3	60.0	83.3	ET volume
	2	87.0	100.0	83.3	ET-PrC fractal dimension, ET-PoC kurtosis
	3	87.0	100.0	83.3	PT-PrC standard deviation, ET-PrC fractal dimension, ET-PoC kurtosis
SVM	1	87.0	40.0	100.0	ET volume
	2	87.0	40.0	100.0	PT volume, ET volume
	3	87.0	40.0	100.0	PT volume, ET volume, ET-PoC fractal dimension
KNN	1	82.6	40.0	94.4	ET volume
	2	87.0	40.0	100.0	PT-PoC kurtosis, ET-PoC fractal dimension
	3	95.7	80.0	100.0	ET-PoC mode, ET-PoC mean, ET-PoC median

The discriminatory power of the median of the wavelet coefficients is best demonstrated in the case of the KNN classifier. Using the db4 mother wavelet and the PT-PrC, PT-PoC, and ET-PrC medians of the diagonal detail coefficients, a prediction accuracy of 100% was achieved. An example of a KNN classifier trained using these features is shown in Figure 8. In this case, the three-dimensional feature set has been reduced to two dimensions by PCA. The concentration of the local recurrence feature set around the origin of the reduced feature space can easily be seen.

The use of the kurtosis measure is demonstrated by one of the best-performing SVM classifiers, which was trained on a feature set containing the PT-PoC skewness, ET-PrC kurtosis, and ET-PoC kurtosis calculated from the approximation coefficient distribution of the db2 wavelet decomposition. A plot of the classifier trained on the PCA-reduced feature set is presented in Figure 9. The two classes can be nearly cleanly separated due to the synergistic discriminatory effect of the three features (which, again, were reduced to two dimensions based on a linear combination calculated using PCA).

Previous work with this data set focused on distribution measures of the absolute intensities of the DCE-MR images^{13,14}. The methodology used to generate features, train classifiers, and predict the outcome of treatment was the same as that used in this paper, except that there were not as many features (only 26 total). The results from this work are shown in Table 4. Interestingly, the classifiers showed the same relative strengths in measures of prediction accuracy (mean, sensitivity, and specificity) as the classifiers in the wavelet analysis: the LDA classifier was the most sensitive, the SVM

classifier was the most specific, and the KNN classifier was the most accurate. However, none of these classifiers achieved absolute performance equivalent to the classifiers trained on feature sets containing wavelet statistics. The most discriminant features for these classifiers were measures of volume, kurtosis, and fractal dimension. The appearance of the fractal dimension as one of the most discriminant features suggested that quantification of texture could offer more powerful measures for treatment outcome prediction. In the previous analysis, the mean intensity of the tumor volume was not one of the most discriminatory features. For the current wavelet analysis, the mean intensity was added as a tested feature to see if it would complement the texture-based wavelet coefficients in predictive ability by adding a baseline reference value for the texture values. Interestingly, it was found that it did not improve the predictive ability of any of the classifiers when used as a feature.

The data set was large enough to show that the analysis of texture in DCE-MRI studies using

Table 5. Intraclass Mean and Standard Deviation for Patients with Local Recurrence (LR) or Local Control (LC)

Treatment Outcome	Class statistic	Feature set		
		PT-PoC skewness	ET-PrC kurtosis	ET-PoC kurtosis
LC	Mean	0.60	4.82	2.83
	Std	0.51	2.87	1.02
LR	Mean	0.06	3.49	3.09
	Std	0.53	1.72	1.61

The features are from the best-performing feature set used with the SVM classifier

wavelet decomposition allowed for better performance in prediction of treatment outcome than the comparable analyses using only measures of the absolute intensity of the tumor volume. Although we observed striking differences in outcome, our data set is too small to draw general conclusions on outcome correlations. A study with a larger data set could demonstrate the power of the wavelet analysis, especially in patients who had local recurrence.

The KNN classifier performs better than the linear SVM and LDA classifiers. This may simply be due to the fact that the features are in general not linearly separable and instead exhibit a more “cluster and outlier” distribution for the two classes (this behavior is seen in Fig. 8). However, the best-performing LDA and SVM classifiers only misclassify a single testing sample. An example of an SVM classifier trained using a best-performing feature set was presented in Figure 9. In that figure, there is a local recurrence patient whose features align more closely with the local control patients. Table 5 shows some intra-class statistics for the SVM classifier’s best-performing feature set. It can be seen that the main significant feature which discriminates between the two classes, judged by the separation of means and the corresponding class standard deviations, is the PT-PoC skewness. The principal component vectors used for the reduction of the feature space from three to two (as shown in Fig. 9), stated as (PT-PoC skewness coefficient, ET-PrC kurtosis coefficient, ET-PoC kurtosis coefficient), were $(-0.0971, 0.4486, -0.8884)$ and $(-0.9905, -0.1311, 0.0420)$. Therefore, the value along the x -axis in Figure 9 mainly incorporates the kurtosis measurements, while the value along the y -axis mainly incorporates the skewness. The single misclassification, then, is due to a small variation of that local recurrence patient’s skewness measure or combined kurtosis measure.

CONCLUSION

The use of the wavelet transform for analysis of local variations in DCE-MRI scans over time can complement the measure of changes in tumor volume for use in outcome prediction for radiation treatment of cervical cancer. The orthogonal separation of high-scale perfusion (the approxima-

tion coefficients) from the lower-scale variations in perfusion (the detail coefficients) offers a robust way of extracting local information for use in determining both intra- and intersubject differences in tumor microenvironment. This paper showed that the quantification of these tumor microenvironment differences using wavelet coefficient distribution statistical measures (especially the median and kurtosis) produces a good basis for prediction of radiation treatment outcome for cervical cancer. In addition, a novel method of calculating the wavelet coefficients before masking was shown to focus the analysis on the tumor ROI, without including surrounding tissue, and remove the partial volume effects. The analysis can still be extended, however, by investigating the spatial relationships between the wavelet parameters, without histogram binning and subsequent distribution calculations. In our future work, we will investigate methods for retaining these spatial relationships, along with developing fully automated tumor segmentation and registration routines.

REFERENCES

1. Franco EL, Schlecht NF, Saslow D: The epidemiology of cervical cancer. *Cancer J* 9(5):348–359, 2003
2. Khan FM Ed: Treatment planning in radiation oncology. 2nd ed. Philadelphia: Lippincott Williams & Wilkins, 2007
3. Choyke PL, Dwyer AJ, Knopp MV: Functional tumor imaging with dynamic contrast-enhanced magnetic resonance imaging. *J Magn Reson Imaging* 17(5):509–520, 2003
4. Hawighorst H, Knapstein PG, Knopp MV, Weikel W, Brix G, Zuna I, et al: Uterine cervical carcinoma: Comparison of standard and pharmacokinetic analysis of time-intensity curves for assessment of tumor angiogenesis and patient survival. *Cancer Res* 58(16):3598–3602, 1998, Aug
5. Noworolski SM, Henry RG, Vigneron DB, Kurhanewicz J: Dynamic contrast-enhanced MRI in normal and abnormal prostate tissues as defined by biopsy, MRI, and 3D MRSI. *Magn Reson Med* 53(2):249–255, 2005
6. Padhani AR, Gapinski CJ, Macvicar DA, Parker GJ, Suckling J, Revell PB, et al: Dynamic contrast enhanced MRI of prostate cancer: Correlation with morphology and tumour stage, histological grade and PSA. *Clin Radiol* 55(2):99–109, 2000
7. George ML, Dzik-Jurasz ASK, Padhani AR, Brown G, Tait DM, Eccles SA, et al: Non-invasive methods of assessing angiogenesis and their value in predicting response to treatment in colorectal cancer. *Br J Surg* 88(12):1628–1636, 2001
8. Wang B, Gao ZQ, Yan X: Correlative study of angiogenesis and dynamic contrast-enhanced magnetic resonance imaging features of hepatocellular carcinoma. *Acta Radiol* 46:353, 2005

9. Furman-Haran E, Schechtman E, Kelcz F, Kirshenbaum K, Degani H: Magnetic resonance imaging reveals functional diversity of the vasculature in benign and malignant breast lesions. *Cancer* 104(4):708–718, 2005
10. Mayr NA, Yuh WT, Magnotta VA, Ehrhardt JC, Wheeler JA, Sorosky JI, et al: Tumor perfusion studies using fast magnetic resonance imaging technique in advanced cervical cancer: A new noninvasive predictive assay. *Int J Radiat Oncol Biol Phys* 36(3):623–633, 1996, Oct 1
11. Mayr NA, Yuh WT, Arnholt JC, Ehrhardt JC, Sorosky JI, Magnotta VA, et al: Pixel analysis of MR perfusion imaging in predicting radiation therapy outcome in cervical cancer. *J Magn Reson Imaging* 12(6):1027–1033, 2000, Dec
12. Lancaster JA, Carrington BM, Sykes JR, Jones AP, Todd SM, Cooper R, et al: Prediction of radiotherapy outcome using dynamic contrast enhanced MRI of carcinoma of the cervix. *Int J Radiat Oncol Biol Phys* 54(3):759–767, 2002, Nov 1
13. Prescott JW, Zhang D, Wang JZ, Mayr NA, Yuh WTC, Saltz J, et al: In: Cancer treatment outcome prediction by assessing temporal change: Application to cervical cancer. *Proc. of SPIE Med Img, San Diego*, 2008
14. Prescott JW, Zhang D, Wang JZ, Mayr NA, Yuh WTC, Saltz J, et al: In: Outcome prediction for radiation treatment of cervical cancer by assessing tumor heterogeneity and temporal change. *OSUMC Research Day*, April 10, 2008
15. Mallat SG: A theory for multiresolution signal decomposition: The wavelet representation. *IEEE Trans Pattern Anal Mach Intell* 11(7):674–693, 1989
16. Gurcan MN: Detection of microcalcifications in mammograms using higher order statistics. *IEEE Signal Process Lett* 4(8):213–216, 1997
17. Duda RO, Hart PE, Stork DG: *Pattern classification*, Wiley: New York, 2001
18. Zijdenbos AP, Dawant BM, Margolin RA, Palmer AC: Morphometric analysis of white matter lesions in MR images: Method and validation. *IEEE Trans Med Imag* 13:716–724, 1994, Dec