Published in final edited form as:

Mach Vis Appl. 2013 October 1; 24(7): . doi:10.1007/s00138-012-0456-y.

Classification of small lesions in dynamic breast MRI: Eliminating the need for precise lesion segmentation through spatio-temporal analysis of contrast enhancement over time

Mahesh B. Nagarajan,

Department of Imaging Sciences and Biomedical Engineering, University of Rochester, 207 Robert B. Goergen Hall, Rochester NY 14627, USA

Markus B. Huber,

Department of Imaging Sciences and Biomedical Engineering, University of Rochester, 207 Robert B. Goergen Hall, Rochester NY 14627, USA

Thomas Schlossbauer,

Department of Radiology, Ludwig Maximilians Universität, Klinikum Innenstadt, Ziemssenstr. 1, 80336 München, Germany

Gerda Leinsinger,

Department of Radiology, Ludwig Maximilians Universität, Klinikum Innenstadt, Ziemssenstr. 1, 80336 München, Germany

Andrzej Krol, and

Department of Radiology, SUNY Upstate Medical University, 750 E. Adams Street, Syracuse NY 13210, USA

Axel Wismüller

Department of Imaging Sciences and Biomedical Engineering, University of Rochester, 207 Robert B. Goergen Hall, Rochester NY 14627, USA. Department of Radiology, Ludwig Maximilians Universität, Klinikum Innenstadt, Ziemssenstr. 1, 80336 München, Germany

Mahesh B. Nagarajan: mahesh.nagarajan@rochester.edu

Abstract

Characterizing the dignity of breast lesions as benign or malignant is specifically difficult for *small* lesions; they don't exhibit typical characteristics of malignancy and are harder to segment since margins are harder to visualize. Previous attempts at using dynamic or morphologic criteria to classify small lesions (mean lesion diameter of about 1 cm) have not yielded satisfactory results. The goal of this work was to improve the classification performance in such small diagnostically challenging lesions while concurrently eliminating the need for precise lesion segmentation. To this end, we introduce a method for topological characterization of lesion enhancement patterns over time. Three Minkowski Functionals were extracted from all five post-contrast images of sixty annotated lesions on dynamic breast MRI exams. For each Minkowski Functional, topological features extracted from each post-contrast image of the lesions were combined into a high-dimensional texture feature vector. These feature vectors were classified in a machine learning task with support vector regression. For comparison, conventional Haralick texture features derived from gray-level co-occurrence matrices (GLCM) were also used. A new method for extracting thresholded GLCM features was also introduced and investigated here. The best classification performance was observed with Minkowski Functionals *area* and *perimeter*, thresholded GLCM features f8 and f9, and conventional GLCM features f4 and f6. However, both Minkowski Functionals and thresholded GLCM achieved such results *without lesion segmentation* while the performance of GLCM features significantly deteriorated when lesions were not segmented ($p < 0.05$). This suggests that such advanced spatio-temporal characterization can

improve the classification performance achieved in such small lesions, while simultaneously eliminating the need for precise segmentation.

Keywords

Minkowski Functionals; topological texture features; mutual information; support vector regression; dynamic breast MRI

1 Introduction

Breast cancer is among the leading causes of mortality for women in North America [1]. In this regard, dynamic contrast-enhanced MRI (DCE-MRI) is emerging as a promising diagnostic modality in the detection and evaluation of suspicious mammographic lesions. However, DCE-MRI exams typically require acquisitions in both spatial and temporal domains making subjective evaluation of clinical findings a challenging task for the radiologist. As a result, breast cancer diagnosis on DCE-MRI has been the subject of research in the area of computer aided diagnosis (CADx) [2–11].

Previous work has showcased the ability of CADx applications to achieve a high diagnostic sensitivity (up to 97%) and reasonable specificity (76.5%) in the task of classifying suspicious lesions on DCE-MRI through the use of dynamic criteria, morphologic criteria, or combinations of both [12]. However, not many studies have focused on evaluating the value of DCE-MRI in *small* lesions which may not exhibit typical characteristics of benign and malignant tumors [13]. Accurate diagnosis of such small lesions is clinically important for improving disease management in patients, where evaluating the dignity of breast lesions as being benign or malignant is specifically challenging.

In this regard, Leinsinger et al. 2006 reported a diagnostic accuracy of 75% in breast cancer diagnosis through cluster analysis of signal intensity (SI) time curves [14]. More recently, Schlossbauer et al. 2008, attempting to classify a dataset of small lesions (mean size 1.1 cm), reported AUC values of 0.76, 0.61 and 0.72 when using dynamic criteria, morphologic criteria and combinations of both, respectively [13]. The primary goal of this work is to improve the classification performance of such small, diagnostically challenging lesions on breast DCE-MRI. Improved classification performance can contribute to reducing (1) the likelihood of performing “false positive” biopsies of benign lesions thereby eliminating the surgical risks associated with the biopsy and (2) missed breast cancers developing from misdiagnosed malignant lesions, while also enabling earlier diagnosis of suspicious lesions.

In this work, we focus on the use of texture features to capture properties of the lesion enhancement pattern. Previous studies have investigated the use of standard second-order statistical features derived from gray-level co-occurrence matrices (GLCM) in capturing properties of lesion enhancement [8–10]. Other studies have focused on characterizing the lesion enhancement pattern through dynamic evaluation of spatial variation [15–17]. More recently, combining such GLCM-derived texture features from all post-contrast images of the lesion was also proposed to improve classification performance [18,19]. Most of these studies have noted that texture features capturing aspects of heterogeneity or homogeneity of the enhancement pattern contribute to the best classification performance. This is not surprising given that heterogeneity has been previously identified as a key criterion indicative of lesion malignancy [20,21].

Given these observations, we investigate the use of more novel and advanced integral geometry texture analysis techniques that could be better suited to capturing such aspects of lesion enhancement heterogeneity. Specifically, we investigate the use of Minkowski

Functionals, which can capture topological properties of the underlying gray-level pattern [22]. This approach has been previously investigated in pattern recognition problems in medical imaging, specifically related to fibrotic tissue classification on lung CT [23,24] and prediction of hip-fracture in post-menopausal women [25]. The primary motivation for investigating the use of such topological texture features in the problem of lesion character classification is illustrated in Figure 1. As seen here, when subject to several thresholds, malignant and benign lesions exhibit different changes to topology as a function of the gray-level threshold. Minkowski Functionals could be used to capture such differences and hence distinguish between benign and malignant lesions. While this figure shows larger lesions, we investigate the effectiveness of such techniques using a dataset of small lesions (mean lesion diameter 1.05 cm) as motivated by the clinical need described earlier.

Previous studies have also shown that pursuing standard GLCM approaches to texture analysis for purposes of lesion classification on breast MRI require *prior segmentation* of the lesion for satisfactory classification performance. However, the precise segmentation of small lesions, such as those used in this study (example shown in Figure 2), is a challenging task for the radiologist. Manual segmentation of such lesions is subject to significant inter-reader variability between different radiologists owing to difficulties in visualizing lesion margins, combined with the limited spatial resolution provided by MRI. Automated segmentation techniques could assist in this task, but these methods usually have free parameters that need to be optimized with respect to the dataset being analyzed. Moreover, verification of results achieved by automated segmentation is challenging in itself, given the difficulty in establishing a 'ground truth' for segmentation stemming from the absence of accepted reference standards for such clinical imaging data [26].

For these reasons, we aim to pursue a texture analytic strategy that characterizes the lesion enhancement pattern without segmenting the lesion *prior to analysis*. Specifically, every annotated ROI (on all post-contrast images) is subjected to several thresholds and each such threshold is used to create a binary image. Strictly speaking, this may be considered as a special case of segmentation. However, it is accomplished as part of the texture feature extraction process itself without requiring any additional human or algorithmic input. The goal is then to quantify the morphological changes observed in the ROI as a function of the gray-level threshold. Extraction of such spatio-temporal features is achieved in this study using two approaches - (1) topological texture features derived from Minkowski Functionals and (2) thresholded GLCM, i.e. GLCM features computed from such binary images as a function of the gray-level threshold. This thresholding suppresses the effects of healthy surrounding tissue, thereby removing the need for precise segmentation of the lesion. We investigate the classification performance of such spatio-temporal texture features when extracted from unsegmented lesions, and compare this to conventional GLCM features extracted from segmented lesions.

In this work, texture analysis using Minkowski Functionals and GLCM is performed on all five post-contrast images of a dynamic breast MRI exam and the extracted features are combined to form lesion characterizing high-dimensional feature vectors which are used for lesion classification. For the machine learning task, support vector regression (SVR) is used [27]; SVR extends the use of support vector machines to regression analysis and is used in this study as a function approximator that predicts the class label of texture feature vectors extracted from lesions of unknown character. While support vector machines can themselves be used in a classification task [28], they provide binary outputs for class predictions. SVR provides fuzzy class labels that can subsequently be used to generate ROC curves, which is the accepted metric of classification performance in previous studies. These processing steps were used to optimize the free parameters of texture features derived from Minkowski Functionals and GLCM, study the effects of segmentation on the classification performance

of such texture features, and compare their performance to that achieved by conventionally used GLCM-derived texture features, as discussed in the following sections.

2 Data

Sixty lesions were identified from a representative set of DCE-MRI exams from 54 female patients by two experienced radiologists who came to a consensus on evaluation of clinical findings. The mean patient age was 52 with a standard deviation of 12 and a range of 27 to 78. In all cases, histo-pathologically confirmed diagnosis from needle aspiration/excision biopsy was available prior to this study; 32 of the lesions were diagnosed as benign and the remaining 28 as malignant. Mean lesion diameter was 1.05 cm (standard deviation of 0.73 cm). An example of such a lesion is shown in Figure 2. The histological distribution of the 32 benign lesions is as follows - 3 fibroadenoma, 10 fibrocystic change, 5 fibrolipomatous change, 7 adenosis, 1 papilloma and 6 non-typical benign disease. The histological distribution of the 28 malignant lesions is as follows - 18 invasive ductal carcinoma, 5 invasive lobular carcinoma, 3 ductal carcinoma in-situ and 2 nontypical malignant disease.

Patients were scanned in the prone position using a 1.5T MR system (Magnetom Vision™, Siemens, Erlangen, Germany) with a dedicated surface coil to enable simultaneous imaging of both breasts. Images were acquired in the axial slice orientation using a T1-weighted 3D spoiled gradient echo sequence with the following imaging parameters; echo repetition time (TR) = 9.1 ms, echo time (TE) = 4.76 ms and flip angle (FA) = 25°. Acquisition of the pre-contrast series was followed by subsequent administration of 0.1 mmol/kg body weight of paramagnetic contrast agent (gadopentate dimeglumine, Magnevist™, Schering, Berlin, Germany). Five post-contrast series were then acquired, each with a measurement time of 83 seconds, and at intervals of 110 seconds. All MR datasets were acquired, based on routine clinical indication only, with informed consent from the patients. Purely retrospective use of strictly anonymized MR data was performed according to the IRB guidelines of Ludwig Maximilians University, Munich, Germany.

In the collection of patient data used in this study, images in the dynamic series were acquired with two different settings of spatial parameters; for 19 patients, the images were acquired as 32 slices per series with a 512×512 matrix, 0.684×0.684 mm² in-plane resolution and 4 mm slice thickness, while in other cases, the same images were acquired as 64 slices per series with a 256×256 matrix, 1.37×1.37 mm² and 2 mm slice thickness. To maintain uniform image data for texture analysis, the images acquired with a 512×512 matrix were reduced to a 256×256 matrix through bilinear interpolation.

3 Methods

3.1 Lesion Annotation and Segmentation

With the exception of two patients, where two separate lesions were chosen for analysis, one primary lesion was selected from all patients for analysis. In four cases, these lesions were captured with two non-overlapping ROIs; a single encapsulating ROI was used to capture the lesion in the rest. Each identified lesion was extracted as a 2-D square region of interest (ROI) of dimensions 11×11 pixels on the central slice of the lesion from the pre-contrast and all post-contrast images of the T1 dynamic series. This choice of ROI size was utilized by the radiologists to reduce the amount of healthy tissue surrounding the lesion captured in the ROI. The ROI annotations were made on difference images created by subtracting the fourth post-contrast image with the pre-contrast image (acquired as part of the imaging protocol); the ROI was subsequently translated to all five post-contrast images.

The annotated lesions were then segmented to ensure that the surrounding healthy tissue did not adversely affect subsequent texture analysis, specifically that involving GLCM. A fuzzy C-Means (FCM) approach previously proposed for lesion segmentation in DCE-MRI was used to accomplish this task [29]. The FCM algorithm is an unsupervised learning technique that creates fuzzy clustering assignments to separate an input set of data points into a specified number of clusters. In the lesion segmentation problem, the FCM algorithm was used to evaluate fuzzy cluster assignments for each pixel time series (from five post-contrast images of the lesion) in the 2D square ROI to partition them into two clusters i.e. lesion and healthy tissue. A detailed description of the FCM application to lesion segmentation can be found in [29]. In this study, the FCM approach failed to segment two lesions; these were manually segmented by consensus of the two expert radiologists.

3.2 Pre-processing Steps

Lesions were enhanced on each post-contrast ROI (S_i) by subtracting and dividing the i^{th} post-contrast ROI S_i , $i = \{1, 2, 3, 4, 5\}$, with the corresponding ROI annotated on the pre-contrast lesion (S_0), $S_i = (S_i - S_0)/S_0$. This step effectively suppresses the healthy tissue that surrounded the lesion in the ROI but can be problematic if patient motion during the acquisition results in improper registration between the various post-contrast and pre-contrast images. Datasets used in this study had only negligible motion artifacts over time and thus, compensatory image registration steps were not required.

3.3 Texture Analysis

Minkowski Functionals were computed by first binarizing each ROI through the application of several thresholds between its minimum and maximum intensity limits [22]. The number of thresholds applied was a free parameter; a suitable choice for this parameter was investigated here. Once the binary images were obtained, three Minkowski Functionals, i.e. area, perimeter and Euler characteristic, were computed from each ROI as follows -

$$MF_{area} = n_s, \quad (1)$$

$$MF_{perimeter} = -4n_s + 2n_e, \quad (2)$$

$$MF_{Euler} = n_s - n_e + n_v, \quad (3)$$

where n_s was the total number of white pixels, n_e was the total number of edges and n_v was the number of vertices. The area feature recorded the area of the white pixel regions in the binary image, the perimeter measure calculated the length of the boundary of white pixel areas and the Euler characteristic was a measure of connectivity between the white pixel regions. Once computed for every binary image derived from a specific ROI, these values were stored in three high-dimensional vectors corresponding to each Minkowski Functional. Such computations were performed for the same ROI on all five post-contrast images of the lesion and then combined. The dimensionality of such texture feature vectors was given by $N \times D$, where N was the number of thresholds used to binarize each ROI and D was the number of post-contrast images (5 in this study).

Gray-level co-occurrence matrices (GLCM) were extracted from the lesion ROIs as described in [30,31]. An inter-pixel distance of $d = 1$ was used in generating the GLCMs. On each ROI, GLCMs were generated in the four principal directions and then summed up element-wise resulting in one non-directional GLCM from which 11 Haralick features were computed. While 14 such texture features were described by Haralick et al. 1973, features

f3, f12 and f14 were undefined owing to the small lesion size in the dataset and were omitted from the analysis. Each texture feature was computed on every post-contrast image and then combined into a texture feature vector; thirteen such texture feature vectors with a dimensionality of 5 were computed in this manner for every individual lesion ROI.

Prior to GLCM texture analysis, the ROIs were re-binned to 32 gray-level histogram bins between the minimum and maximum intensity limits following lesion enhancement. The minimum and maximum intensity limits were defined 'globally', i.e. from all lesion ROIs, within each post-contrast set. The choice of 32 bins between 'globally' chosen intensity limits was recommended by previous work [9].

Thresholded GLCM features were computed in this study using thresholded binarization as an initial processing step, similar to the computation of Minkowski Functionals. Each ROI was first binarized through the application of several thresholds between its minimum and maximum intensity limits. Conventional GLCM features were then extracted from each of the binary images created. Such computations were performed for the same ROI on all five post-contrast images of the lesion and then combined. The dimension of such texture feature vectors was given by $N \times D$, where N was the number of thresholds used to binarize each ROI and D was the number of post-contrast images (5 in this study). The number of thresholds applied was a free parameter; a suitable choice for this parameter was guided by experiments performed with features derived from Minkowski Functionals. Re-binning, normally a pre-processing step for the extraction of conventional GLCM features described above, was not required here since GLCM features were extracted from binary images.

3.4 Feature Selection

Feature selection involves identifying a subset of features from the input feature space that makes the most relevant contribution to separating the two different classes of data points in the machine learning step. This study used mutual information analysis, which measures the information content of each feature with regards to the decision task to be performed [32]. In this study, mutual information was used to identify a subset of features for each of the high-dimensional texture feature vectors (extracted from all five post-contrast images) that best contributed to lesion character classification.

Mutual information (MI) is a measure of general independence between random variables [33]. For two random variables X and Y , the MI is defined as -

$$I(X, Y) = H(X) + H(Y) - H(X, Y), \quad (4)$$

where entropy $H(\cdot)$ measures the uncertainty associated with a random variable. The MI $I(X, Y)$ estimates how the uncertainty of X is reduced when Y has been observed. If X and Y are independent, their MI is zero.

For the N texture feature vectors of dimension D , the MI between the each dimension of the texture feature vectors f_s , $s = \{1, 2, \dots, D\}$, and the corresponding class labels y_i , $i = \{1, 2, \dots, N\}$, was calculated by approximating the probability density function of each variable using histograms $P(\cdot)$ -

$$I(f_s, y_i) = \sum_{n_c} \sum_{n_f} P(f_s, y_i) \log_2 \frac{P(f_s, y_i)}{P(f_s)P(y_i)}, \quad (5)$$

Here, the number of classes $n_c = 2$ was used; the number of histogram bins for the texture features n_f was determined adaptively according to

$$n_f = \log_2 N + 1 + \log_2(1 + \kappa \sqrt{N/6}), \quad (6)$$

where κ is the estimated kurtosis and N the number of ROIs in the data set [32].

3.5 Classification

The extraction of texture features and subsequent feature selection was followed by a supervised learning step where the lesion patterns were classified as benign or malignant. In this work, support vector regression with a radial basis function kernel was used for the machine learning task [27]. Here, SVR treats the texture features as independent variables and their labels as the dependent variable and acts as a function approximator; this function is then used in conjunction with the texture features of the test data points to predict their labels. The SVR implementation was taken from the libSVM library [34].

In this study, 70% of the data was used for the training phase while the remaining 30% served as an independent test set. The training data was sub-sampled from the complete dataset in such a manner that at least 40% of each class (benign and malignant) were represented. Special care was taken to ensure that lesion ROIs extracted from the same patient was used either as training or test data to prevent any potential for biased training. To ensure the integrity of the independent test set, global intensity limits for pre-processing (for GLCM texture features) were determined using lesion ROIs from the training data alone. The best features of the texture feature vectors were selected by evaluating the mutual information criteria of the training data alone; this ensured that label information for the test data was not used prior to the classification task.

In the training phase, models are created from labeled data by employing a random sub-sampling cross-validation strategy where the training set is further split into 70% training samples and 30% validation samples - the purpose of the training was to determine the optimal classifier parameters, i.e. those that best capture the boundaries between the two classes of lesion patterns. The free parameters for the classifier used in this study were the cost parameter for SVR and the shape parameter of the radial basis function kernel. A wide range of values were investigated for these free parameters; for the cost parameter, the values were in the range of 0.001 – 20 while for the shape parameter, the values were in the range of 0.001 – 1. The specific value chosen for each of these parameters was determined by optimization through random sub-sampling cross-validation. During each run of the cross-validation process, the optimum value for the free parameter was chosen from the specified range to yield the best separation between benign and malignant classes of the training data alone. Then, during the testing phase, the optimized classifier predicted the label (benign or malignant) of lesion ROIs in the independent test dataset; an ROC curve was generated and used to compute the area under the ROC curve (AUC) which served as a measure of classifier performance. This process was repeated 100 times resulting in an AUC distribution for each feature set.

3.6 Statistical Analysis

A Wilcoxon signed-rank test was used to compare two AUC distributions corresponding to different texture features. Significance thresholds were adjusted for multiple comparisons using the Holm-Bonferroni correction to achieve an overall type I error rate (significance level) less than α (where $\alpha = 0.05$) [35,36].

Texture, classifier and statistical analysis were implemented using Matlab 2008b (The MathWorks, Natick, MA).

4 Results

Figure 3 shows the classification performance obtained for the three Minkowski Functionals when extracted with different number of binarizing thresholds. The best classification performance was observed when 20 such thresholds were used to binarize the ROI of the lesion (on each post-contrast image). The classification performance did not improve when more thresholds were used; all other results reported in this study used 20 thresholds when computing topological texture features derived from Minkowski Functionals from each post-contrast image. The overall dimension of each Minkowski Functional feature vector was 100.

Figure 4 shows a comparison in classifier performance when Minkowski Functional feature vectors were extracted from segmented and unsegmented lesions. The best classification performance was observed with area (0.83 ± 0.09) and perimeter (0.82 ± 0.09) when the feature vectors were extracted from unsegmented lesions. Such features significantly outperformed their counterparts extracted from segmented lesions ($p < 0.001$). The Minkowski Functional Euler Characteristic, however, improved in classification performance when extracted from segmented lesions but failed to achieve comparable performance in distinguishing between benign and malignant lesions.

Figure 5 shows a comparison classifier performance when thresholded GLCM feature vectors were extracted from segmented and unsegmented lesions. Twenty thresholds were used to compute thresholded GLCM features from each post-contrast image and the overall dimension of each thresholded GLCM feature vector was 100. The best classification performance was observed with f8 (0.82 ± 0.09) and f9 (0.83 ± 0.09) when the feature vectors were extracted from unsegmented lesions. Such features significantly outperformed their counterparts extracted from segmented lesions ($p < 0.001$). In all, 6 of 11 features showed significantly improved performance when extracted from unsegmented lesions ($p < 0.05$). Only two features, f5 and f6, significantly improved in performance when extracted from segmented lesions but they still failed to achieve comparable performance in distinguishing between benign and malignant lesions.

Figure 6 shows a comparison in classification performance between the best Minkowski Functional and thresholded GLCM feature vectors extracted from *unsegmented*, and conventional GLCM-derived features extracted from *segmented* lesions. As seen here, comparable classification performance is achieved by Minkowski Functionals - area and perimeter, thresholded GLCM feature vectors - f8 (sum entropy) and f9 (entropy), and conventional GLCM feature vectors - f4 (sum of squares: variance) and f6 (sum average).

Figure 7 shows a comparison in classification performance between the best Minkowski Functional feature vectors, the best thresholded GLCM feature vectors, and conventional GLCM-derived feature vectors when all are extracted from *unsegmented* lesions. As seen here, the performance of conventional GLCM-derived texture features deteriorates when the lesions are not segmented. In fact, when unsegmented lesions are used for the analysis, all conventional GLCM-derived texture features are outperformed by Minkowski Functionals area and perimeter, and thresholded GLCM features f8 and f9 ($p < 0.05$). A summary of the best results achieved with features from all three approaches pursued in this study is shown in Table 1.

Figure 8 shows a comparison in classification performance achieved with different Minkowski Functionals, when different sizes of feature subsets as identified by mutual

information analysis are used in the machine learning task. As seen here, performance deteriorates when too few features are selected. While the best performance is observed when all features are used, comparable performance is achieved with Minkowski Functional area with a subset of 50 features (out of the 100 extracted). Similarly, subsets of 20 through 50 features provide comparable performance to using all features for Minkowski Functional perimeter.

Figure 9 shows a comparison in classification performance achieved with thresholded GLCM features f8 and f9, when different sizes of feature subsets as identified by mutual information analysis are used in the machine learning task. As seen here, performance deteriorates when too few features are selected. While the best performance is observed when all features are used, comparable performance is achieved with subsets of 30 features (out of the 100 extracted) for f8 and 50 features for f9.

5 Discussion

Given that previous studies have identified heterogeneity as an important criterion in distinguishing between benign and malignant lesions through texture analysis [10,17], the primary goal of this study was to evaluate the classification performance achieved by novel and advanced texture features that could be better suited to capturing aspects of lesion enhancement heterogeneity. Specifically, we investigate the use of Minkowski Functionals and thresholded GLCM to capture features of the underlying gray-level topology as a function of threshold. Such spatio-temporal texture features are used to classify a dataset of diagnostically challenging lesions, specifically those considered small (mean lesion diameter of 1.05 cm), on DCE-MRI. Our approaches show that these spatio-temporal approaches, while achieving comparable performance to two conventionally used GLCM features, eliminate the need for precise segmentation of such small lesions.

Previous work has shown that conventionally used GLCM-derived statistical texture features improve in classification performance when lesions are segmented prior to texture analysis [9]. However, precise segmentation of small lesions can be a challenging task for the radiologist. Borders of these lesions can be ill-defined and harder to visualize leading to imperfect manual segmentation and variability across different experts employed for this task. While automated and semiautomated segmentation techniques have been previously proposed for lesions [7,29], such methods often have free parameters that must be optimized with respect to the dataset. Verification of the segmentation results achieved by such methods is also problematic given the difficulty in establishing the ground truth for segmentation of such lesions. Given such shortcomings, we propose switching to a spatio-temporal characterization of the lesion enhancement pattern that *does not require prior lesion segmentation*. Here, we use texture features derived from Minkowski Functionals and thresholded GLCM, and extracted from lesions annotated by a square ROI alone. Our results in Figures 6 and 7 show that such spatio-temporal texture features, when applied to unsegmented lesions, achieve comparable performance to GLCM texture features extracted from segmented lesions.

Of the topological texture features derived from Minkowski Functionals, the best classifier performance was observed with area and perimeter, as shown in Figure 4. These features essentially track changes in area and perimeter of lesion pixel regions as a function of the gray-level threshold. The high classification performance observed with these features can be attributed to the heterogeneous/homogeneous nature of the lesion enhancement pattern. As previously indicated in Figure 1, benign and malignant lesions would respond differently as the gray-level threshold is varied and this is best captured by area and perimeter. Figure 4 also shows that the third Minkowski Functional, i.e. Euler Characteristic, performs much

poorly suggesting that connectivity is not a major distinguishing factor when such small lesions are analyzed. We hypothesize that Euler characteristic would exhibit better classification performance when used with larger lesions (such as the malignant lesion in Figure 1) where thresholding would significantly alter the connectivity of the analyzed region owing to the more pronounced appearance of heterogeneity. This will be investigated in further detail in future studies.

Of the texture features derived from thresholded GLCM, the best classifier performance was observed with f8 (sum entropy) and f9 (entropy), as shown in Figure 5. Thus, characterizing randomness in the topology through entropy measures as a function of the applied gray-level threshold helped achieve the best separation between benign and malignant lesions. This is yet another method of capturing the different behaviour of benign and malignant lesion ROIs exhibited when the gray-level threshold is varied, as shown in Figure 1. The performance achieved with such thresholded GLCM features was comparable to that achieved with Minkowski Functionals area and perimeter. It is noted that certain features, such as f6 (sum average), exhibit high performance when extracted conventionally rather than through the thresholded approach described in this paper, as seen in Table 1. Such features have been previously identified as being associated with the homogeneity of the gray-level intensity values. However, the thresholded approach described in this paper involves extracting such features from binary images. Thus the local variance of gray-level intensity values in the ROI is drastically reduced, which adversely affects the discriminative ability of such features.

The effects of segmentation on the performance of spatio-temporal texture features were also analyzed. With regards to most texture features derived from Minkowski Functionals and thresholded GLCM, segmenting the lesions before feature extraction had a detrimental effect on classification performance, as shown in Figures 4 (Minkowski Functionals) and 5 (thresholded GLCM). This negative influence of segmentation on classification performance may be related to the thresholding process used in the extraction of texture features. Such binarization of the ROI at different thresholds, without any prior segmentation, allows the lesion margin to be analyzed at a range of *different* gray-level intensities. Such detailed morphological information about the lesion margin is lost when a *single* lesion boundary is annotated by a human expert or an automatic segmentation algorithm. However, we do note certain exceptions to this trend; specifically with Euler Characteristic in Minkowski Functionals and features f5–f7 in thresholded GLCM. This could be attributed to the fact that such features are more sensitive to noisy pixels that may be included in the unsegmented ROI. For instance, consider the thresholded images of the benign lesion in Figure 1; the bright spot on the top right of the ROI is some background artifact not related to the lesion. When computing the Minkowski Functionals, this artifact is considered as a connected component and subsequently affects the Euler Characteristic in a more dramatic fashion than area and perimeter, where it would simply account for a small increase in the number of pixels included (see equations 1–3).

Feature selection through mutual information analysis was used to evaluate the effect of reducing the feature set for each spatio-temporal texture feature vector on the overall classification performance achieved. As seen in Figure 8, comparable performance is achieved by Minkowski Functional area when up to 50 features are eliminated by mutual information analysis. Similar results are noted with perimeter as well. Figure 9 reveals similar observations about thresholded GLCM features f8 and f9. This suggests that while using 20 thresholds best captures the difference in behavior of such texture features extracted from benign and malignant lesions, not all of them equally contribute to the classification task. Figures 8 and 9 also show significant deterioration in classification performance when the feature subset is extremely small (i.e. when the best two, three or five

features were chosen from the extracted 100) suggesting that the difference between the two classes of lesions is not captured by single or few features but by observing trends over several features. Future studies will attempt to quantify these specific differences between the behavior of spatio-temporal texture feature vectors extracted from benign and malignant lesions as interpretable clinical findings.

One limitation of this study was that only those exams with negligible motion artifacts over time were included in texture analysis and classification tasks. Future studies with less stringent inclusion criteria for exams could incorporate sophisticated non-linear image registration methods as part of the pre-processing to compensate for motion artifacts over time [37]. Another limitation of this study regards the use of 2D lesion ROIs in texture analysis rather than 3D lesion volumes; volumetric analysis could not be performed with the image datasets used in this study owing to the anisotropy of the pixels involved. While previous research has shown that volumetric analysis of lesions improves classification performance [9], arguments have been made against acquiring breast images with isotropic voxels owing to the longer imaging time involved as well as the smaller coverage of the area being imaged [10]. This study also specifically focuses on evaluating the classification performance achieved with 2D Minkowski Functionals; future studies will investigate using 3D Minkowski Functionals which involve 3D counterparts of the those described here as well a fourth measure (mean breadth). However, even if the afore mentioned limitations are taken into account, this work demonstrates that introducing spatio-temporal feature descriptors, namely Minkowski Functionals and thresholded GLCM, has the potential to eliminate the need for breast lesion tissue segmentation prior to automated lesion classification.

6 Conclusion

This study evaluates the performance of spatio-temporal texture features derived from Minkowski Functionals and thresholded GLCM in the classification of diagnostically challenging lesions, specifically those considered *small* (mean lesion diameter of 1.05 cm). The results in this work suggest that such an approach provides comparable performance to conventionally used GLCM-derived statistical features while simultaneously eliminating the need for precise lesion segmentation prior to feature extraction. We hypothesize that such an approach would have clinical significance as segmentation is a challenging task for such small lesions. However, larger controlled trials need to be conducted in order to further validate the clinical applicability of our method.

Acknowledgments

This research was funded in part by the National Institute of Health (NIH) Award R01-DA-034977, the Clinical and Translational Science Award 5-28527 within the Upstate New York Translational Research Network (UNYTRN) of the Clinical and Translational Science Institute (CTSI), University of Rochester, and by the Center for Emerging and Innovative Sciences (CEIS), a NYSTAR-designated Center for Advanced Technology. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of Health.

The authors would also like to thank Benjamin Mintz for his assistance with developing the tool used for lesion annotation, and Prof. Dr. Maximilian Reiser, FACR, FRCR of the Department of Radiology, Ludwig Max-imilians University, for his continued support.

References

1. American Cancer Society. Breast Cancer Facts & Figures 2005–2006. Atlanta: American Cancer Society, Inc; 2006.

2. Kuhl CK, Mielcareck P, Klaschik S, Leutner C, Wardelmann E, Gieseke J, Schild HH. Dynamic breast MR imaging: Are signal intensity time course data useful for differential diagnosis of enhancing lesions? *Radiology*. 1999; 211(1):101–110. [PubMed: 10189459]
3. Wedegärtner U, Bick U, Wörtler K, Rummeny E, Bongartz G. Differentiation between benign and malignant findings on MR-mammography: usefulness of morphological criteria. *European Radiology*. 2001; 11(9):1645–1650. [PubMed: 11511885]
4. Baum F, Fischer U, Vosschenrich R, Grabbe E. Classification of hypervascularized lesions in CE MR imaging of the breast. *European Radiology*. 2002; 12(5):1087–1092. [PubMed: 11976850]
5. Szabo B, Aspelin P, Wiberg M, Böne B. Dynamic MR imaging of the breast: Analysis of kinetic and morphologic diagnostic criteria. *Acta Radiologica*. 2003; 44(4):379–386. [PubMed: 12846687]
6. Chen W, Giger ML, Lan L, Bick U. Computerized interpretation of breast MRI: Investigation of enhancement-variance dynamics. *Medical Physics*. 2004; 31(5):1076–1082. [PubMed: 15191295]
7. Wismüller A, Meyer-Base A, Lange O, Schlossbauer T, Kallergi M, Reiser MF, Leinsinger G. Segmentation and classification of dynamic breast magnetic resonance image data. *Journal of Electronic Imaging*. 2006; 15(1)
8. Gibbs P, Turnbull LW. Textural analysis of contrast-enhanced MR images of the breast. *Magnetic Resonance in Medicine*. 2003; 50(1):92–98. [PubMed: 12815683]
9. Chen W, Giger ML, Li H, Bick U, Newstead GM. Volumetric texture analysis of breast lesions on contrast-enhanced magnetic resonance images. *Magnetic Resonance in Medicine*. 2007; 58(3):562–571. [PubMed: 17763361]
10. Nie K, Chen JH, Yu HJ, Chu Y, Nalcioğlu O, Su MY. Quantitative analysis of lesion morphology and texture features for diagnostic prediction in breast MRI. *Academic Radiology*. 2008; 15(12): 1513–1525. [PubMed: 19000868]
11. Meinel LA, Stolpen AH, Berbaum KS, Fajardo LL, Reinhardt JM. Breast MRI lesion classification: Improved performance of human readers with a backpropagation neural network computer-aided diagnosis (CAD) system. *Journal of Magnetic Resonance Imaging*. 2007; 25(1): 89–95. [PubMed: 17154399]
12. Fischer D, Wurdinger S, Boettcher J, Malich A, Kaiser W. Further signs in the evaluation of magnetic resonance mammography: a retrospective study. *Investigative Radiology*. 2005; 40(7): 430–435. [PubMed: 15973134]
13. Schlossbauer T, Leinsinger G, Wismüller A, Lange O, Scherr M, Meyer-Bae A, Reiser M. Classification of small contrast enhancing breast lesions in dynamic magnetic resonance imaging using a combination of morphological criteria and dynamic analysis based on unsupervised vector-quantization. *Investigative Radiology*. 2008; 43(1):56–64. [PubMed: 18097278]
14. Leinsinger G, Schlossbauer T, Scherr M, Lange O, Reiser M, Wismüller A. Cluster analysis of signal-intensity time course in dynamic breast MRI: does unsupervised vector quantization help to evaluate small mammographic lesions? *European Radiology*. 2006; 16(5):1138–1146. [PubMed: 16418862]
15. Gilhuijs KGA, Giger M, Bick U. Computerized analysis of breast lesions in three dimensions using dynamic magnetic-resonance imaging. *Medical Physics*. 1998; 25(9):1647–1654. [PubMed: 9775369]
16. Zheng Y, Englander S, Baloch S, Zacharaki EI, Fan Y, Schnall MD, Shen D. STEP: Spatiotemporal enhancement pattern for MR-based breast tumor diagnosis. *Medical Physics*. 2009; 36(7):3192–3204. [PubMed: 19673218]
17. Buelow T, Saalbach A, Bergtholdt M, Wiemker R, Buurman H, Meinel LA, Newstead G. Heterogeneity of kinetic curve parameters as indicator for the malignancy of breast lesions in DCE MRI. *Proceedings of SPIE Medical Imaging*. 2010; 7624
18. Agner S, Soman S, Libfield E, McDonald M, Thomas K, Englander S, Rosen M, Chin D, Noshier J, Madabhushi A. Textural kinetics: A novel dynamic contrast-enhanced (DCE)-MRI feature for breast lesion classification. *Journal of Digital Imaging*. 2011; 24(3):446–463. [PubMed: 20508965]
19. Nagarajan M, Huber M, Schlossbauer T, Leinsinger G, Krol A, Wismüller A. Classifying small lesions on breast mri through dynamic enhancement pattern characterization. *Lecture Notes in Computer Science*. 2011; 7009:352–359.

20. Ikeda DM, Hylton NM, Kinkel K, Hochman MG, Kuhl CK, Kaiser WA, Weinreb JC, Smazal SF, Degani H, Viehweg P, Barclay J, Schnall MD. Development, standardization, and testing of a lexicon for reporting contrast-enhanced breast magnetic resonance imaging studies. *Journal of Magnetic Resonance Imaging*. 2001; 13(6):889–895. [PubMed: 11382949]
21. D'Orsi, CJ.; Bassett, L.; Berg, W.; Feig, S.; Jackson, J.; Kopans, D. Breast imaging reporting and data system (BI-RADS) breast imaging atlas. American College of Radiology; Reston, VA: 2003.
22. Michielsen K, Raedt H. Integral-geometry morphological image analysis. *Physics Reports*. 2001; 347(6):461–538.
23. Boehm H, Fink C, Attenberger U, Becker C, Behr J, Reiser M. Automated classification of normal and pathologic pulmonary tissue by topological texture features extracted from multi-detector CT in 3D. *European Radiology*. 2008; 18(12):2745–2755. [PubMed: 18618121]
24. Huber M, Nagarajan M, Leinsinger G, Eibel R, Ray L, Wismüller A. Performance of topological texture features to classify fibrotic interstitial lung disease patterns. *Medical Physics*. 2011; 38(4): 2035–2044. [PubMed: 21626936]
25. Boehm H, Vogel T, Panteleon A, Burklein D, Bitterling H, Reiser M. Differentiation between post-menopausal women with and without hip fractures: enhanced evaluation of clinical DXA by topological analysis of the mineral distribution in the scan images. *Osteoporos International*. 2007; 18(6):779–787.
26. Warfield S, Zou K, Wells W. Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging*. 2004; 23(7):903–921. [PubMed: 15250643]
27. Drucker H, Burges C, Kaufman L, Smola A, Vapnik V. Support vector regression machines. *Advances in Neural Information Processing Systems*. 1996; 9:155–161.
28. Cortes C, Vapnik V. Support vector networks. *Machine Learning*. 1995; 20:273–297.
29. Chen WJ, Giger ML, Bick U. A fuzzy c-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images. *Academic Radiology*. 2006; 13(1):63–72. [PubMed: 16399033]
30. Haralick RM, Shanmuga K, Dinstein I. Textural features for image classification. *IEEE Transactions on Systems Man and Cybernetics Smc*. 1973; 3(6):610–621.
31. Haralick RM. Statistical and structural approaches to texture. *Proceedings of the IEEE*. 1979; 67(5):786–804.
32. Tourassi G, Frederick E, Markey M, Floyd C Jr. Application of the mutual information criterion for feature selection in computer-aided diagnosis. *Medical Physics*. 2001; 28(12):2394–2402. [PubMed: 11797941]
33. Duda, RO.; Hart, PE.; Stork, DG. *Pattern Classification*. Wiley-Interscience Publication; 2000.
34. Chang CC, Lin CJ. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*. 2011; 2:27:1–27:27. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.
35. Wright SP. Adjusted P-values for simultaneous inference. *Biometrics*. 1992; 48(4):1005–1013.
36. Holm S. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*. 1979; 6(2):65–70.
37. Rueckert D, Sonoda LI, Hayes C, Hill DL, Leach MO, Hawkes DJ. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*. 1999; 18(8):712–21. [PubMed: 10534053]

Fig. 1.

An example of binarization through application of gray-level thresholds as demonstrated with a benign and malignant lesion. Note the drastic changes in topology with the malignant lesion as the threshold changes as opposed to a fairly consistent topology exhibited by the benign lesion. Such differences can be captured by Minkowski Functionals.

Fig. 2. A small contrast-enhancing benign lesion (red circle) as observed on the fourth post-contrast image (left) of a dynamic breast MRI study. The annotated lesion is magnified and shown on the lower right corner of the image.

Fig. 3.

Comparison of classification performance achieved with Minkowski Functionals when different thresholds are used for computation of features. For each distribution, the central mark corresponds to the median and the edges are the 25th and 75th percentile. The best results are obtained with area and perimeter for 20 thresholds; no significant improvement in performance is observed when the thresholds are further increased.

Fig. 4.

Comparison of classification performance achieved with Minkowski Functionals when extracted from unsegmented and segmented lesion ROIs. For each distribution, the central mark corresponds to the median and the edges are the 25th and 75th percentile. The best results are obtained with area and perimeter when the texture features are extracted from unsegmented lesion ROIs.

Fig. 5. Comparison of classification performance achieved with thresholded GLCM when extracted from unsegmented and segmented lesion ROIs. For each distribution, the central mark corresponds to the median and the edges are the 25th and 75th percentile. The best results are obtained with features f8 (sum entropy) and f9 (entropy) when the texture features are extracted from unsegmented lesion ROIs.

Fig. 6.

Comparison of classification performance achieved with conventional GLCM-derived texture features (GLCM, blue) extracted from segmented lesions and the best spatio-temporal texture features, derived from Minkowski Functionals (MF, red) and thresholded GLCM (t-GLCM, red), extracted from unsegmented lesions. For each distribution, the central mark corresponds to the median and the edges are the 25th and 75th percentile. As seen here, the best Minkowski Functionals (area and perimeter) and best t-GLCM features (f8 and f9) exhibit comparable classifier performance to that achieved by conventional GLCM features f4 and f6.

Fig. 7.

Comparison of classification performance achieved with conventional GLCM-derived texture features (GLCM, blue) and the best spatio-temporal texture features, derived from Minkowski Functionals (MF, red) and thresholded GLCM (t-GLCM, red), when all features are extracted from unsegmented lesions. For each distribution, the central mark corresponds to the median and the edges are the 25th and 75th percentile. As seen here, the best Minkowski Functionals (area and perimeter) and best t-GLCM features (f8 and f9) outperform all conventional GLCM-derived features ($p < 0.05$).

Fig. 8.

Comparison of classification performance achieved with different sizes of feature subsets (out of 100 extracted features) selected through mutual information analysis for Minkowski Functionals when extracted from unsegmented lesions. For each distribution, the central mark corresponds to the median and the edges are the 25th and 75th percentile. As seen here, performance is comparable to using all features when a subset of 50 features is used for Minkowski Functional area, and when subsets of 20 through 50 features are used for perimeter.

Fig. 9.

Comparison of classification performance achieved with different sizes of feature subsets (out of 100 extracted features) selected through mutual information analysis for thresholded GLCM features f8 and f9 when extracted from unsegmented lesions. For each distribution, the central mark corresponds to the median and the edges are the 25th and 75th percentile. As seen here, performance is comparable to using all features when a subset of 30 features is used for f8, and 50 features for f9.

Table 1

Comparison of classification performance (mean AUC + standard deviation) between the best GLCM, thresholded GLCM (t-GLCM) and Minkowski Functionals (MF) when texture analysis uses features extracted from all five post-contrast images. Note that ROI state indicates whether the ROI was segmented prior to texture analysis.

Method	Feature	AUC	ROI state
GLCM	f4	0.83 ± 0.08	Segmented
GLCM	f6	0.82 ± 0.08	Segmented
t-GLCM	f8	0.74 ± 0.10	Unsegmented
t-GLCM	f9	0.70 ± 0.10	Unsegmented
MF	area	0.83 ± 0.09	Unsegmented
MF	perimeter	0.82 ± 0.10	Unsegmented