

# A Multichannel Markov Random Field Framework for Tumor Segmentation With an Application to Classification of Gene Expression-Based Breast Cancer Recurrence Risk

Ahmed B. Ashraf\*, *Member, IEEE*, Sara C. Gavenonis, Dania Daye, Carolyn Mies, Mark A. Rosen, and Despina Kontos, *Member, IEEE*

**Abstract**—We present a methodological framework for multichannel Markov random fields (MRFs). We show that conditional independence allows loopy belief propagation to solve a multichannel MRF as a single channel MRF. We use conditional mutual information to search for features that satisfy conditional independence assumptions. Using this framework we incorporate kinetic feature maps derived from breast dynamic contrast enhanced magnetic resonance imaging as observation channels in MRF for tumor segmentation. Our algorithm based on multichannel MRF achieves an receiver operating characteristic area under curve (AUC) of 0.97 for tumor segmentation when using a radiologist's manual delineation as ground truth. Single channel MRF based on the best feature chosen from the same pool of features as used by the multichannel MRF achieved a lower AUC of 0.89. We also present a comparison against the well established normalized cuts segmentation algorithm along with commonly used approaches for breast tumor segmentation including fuzzy C-means (FCM) and the more recent method of running FCM on enhancement variance features (FCM-VES). These previous methods give a lower AUC of 0.92, 0.88, and 0.60, respectively. Finally, we also investigate the role of superior segmentation in feature extraction and tumor characterization. Specifically, we examine the effect of improved segmentation on predicting the probability of breast cancer recurrence as determined by a validated tumor gene expression assay. We demonstrate that an support vector machine classifier trained on kinetic statistics extracted from tumors as segmented by our algorithm gives a significant improvement in distinguishing between women with high and low recurrence risk, giving an AUC of 0.88 as compared to 0.79, 0.76, 0.75, and 0.66 when using normalized cuts, single channel MRF, FCM, and FCM-VES, respectively, for segmentation.

**Index Terms**—Breast cancer recurrence prediction, breast dynamic contrast enhanced (DCE) magnetic resonance imaging (MRI), breast tumor segmentation, tumor characterization.

## I. INTRODUCTION

CRUCIAL to the performance of a feature extraction and image classification system is the availability of a reliable and accurate segmentation approach for the object of interest (e.g., tumor). In most medical imaging applications, the automation of this step is particularly important because of the large amount of images to be analyzed. This makes the manual segmentation approach tedious and prohibitively expensive. As a result, extensive research has been done in the medical imaging community for improving the quality of automated segmentation. Specifically, there is an abundance of methods geared towards segmenting well defined anatomical structures (e.g., parts of the brain). Notable among them are the variants of active contours [1] and active shape models [2]. There are two main drawbacks associated with this class of methods. First, they require manual initialization that should be very close to the actual structure to be segmented. Efforts for automating the initialization step have been restricted to structures for which precise *a priori* knowledge is available [3]. Second, they aim at segmenting known anatomical structures that have well defined control points in their shape, which is often not the case for arbitrarily shaped lesions. Fuzzy C means (FCM) clustering and its variants are also prevalent for the segmentation of well defined structures [4]–[6]. Markov random field (MRF) based approaches have also been used for segmentation. For instance, in [7], the authors model brain magnetic resonance imaging (MRI) images as a MRF and use prior information for segmenting anatomical structures. Another approach for segmentation includes spectral graph clustering methods [8]–[10]. Notable among them is the normalized cuts algorithm [8] which is a highly robust segmentation method and is widely used for segmentation applications [11]–[14]. Specifically, [11] uses normalized cuts for MRI image segmentation.

As compared to anatomically well defined structures, there is relatively less work in the literature on segmenting arbitrarily shaped structures (e.g., breast tumors). Among the work done on breast tumor segmentation the most popular approach is

Manuscript received June 15, 2012; revised September 09, 2012; accepted September 10, 2012. Date of publication September 19, 2012; date of current version March 29, 2013. This work was supported by the Institute of Translational Medicine and Therapeutics (ITMAT) Transdisciplinary Program in Translational Medicine and Therapeutics under Grant UL1RR024134 from the National Center for Research Resources. The content of this work is the sole responsibility of the authors and does not necessarily represent the official views of the National Center for Research Resources or the National Institute of Health. *Asterisk indicates corresponding author.*

\*A. B. Ashraf is with the Computational Breast Imaging Group (CBIG), Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: ahmed.ashraf@uphs.upenn.edu).

S. C. Gavenonis, D. Daye, M. A. Rosen, and D. Kontos are with the Computational Breast Imaging Group (CBIG), Department of Radiology, University of Pennsylvania, Philadelphia, PA 19104 USA.

C. Mies is with the Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104 USA.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TMI.2012.2219589

FCM clustering employed by many researchers due to its simplicity (e.g., [15], [16]). More recently Lee *et al.* [17] presented a method that runs FCM on the variance map of enhancement kinetics (FCM-VES). Although useful for segmentation, these methods do not take into account the overlap of the feature values of the tumor and nontumor pixels. As a result they have to settle with a manually set threshold on FCM membership probabilities, leading to poor generalization. In order to address these issues we propose to incorporate a component of learning for breast tumor segmentation. We present a framework for multichannel extension of MRFs to make maximal usage of multiple feature streams derived from the imaging data, here specifically related to kinetic analysis of dynamic contrast enhanced (DCE) breast MR images.

We have previously shown that a multichannel extension of MRFs is feasible [18]. In this paper, we present a complete methodological framework, selection of conditionally independent feature channels for MRF, extended comparison with other segmentation algorithms, and assessment of the potential impact of improved segmentation on classification and characterization of breast cancer tumors. We explore how inference methods like loopy belief propagation [19] may be extended for a multichannel MRF. There have been considerable work in the literature that uses MRF for segmentation. In a representative work [7], authors address the problem of brain image segmentation using MRFs and present an expectation maximization framework to solve the MRF. However, the approach in [7] is limited to a single channel MRF. Often, it is desirable to integrate information from different channels, if available, to achieve better segmentation. For example, in our specific application, multiple channels would allow to take maximum advantage of DCE MRI kinetic features that represent the properties of contrast agent uptake by the tissue, which are expected to be different for tumor and background tissues [15], [20]. In this work we exploit conditional independence for solving an MRF via loopy belief propagation [19] that reduces a multichannel MRF to a single channel MRF for inference queries, allowing to capitalize on information coming in from different channels. We also present a principled method to choose features satisfying conditional independence to be included as multiple channels in the MRF observation model.

Key contributions of this paper are as follows.

- We show that conditional independence allows loopy belief propagation to solve a multichannel MRF as a single channel MRF thereby avoiding the modeling of joint distributions (Section II).
- Using conditional mutual information we present a principled method for choosing conditionally independent features as different channels in the MRF observation model (Section III).
- To elaborate on this premise, we introduce multiple feature channels derived from the kinetic analysis of DCE magnetic resonance (MR) images in the observation model of MRF (Section III).
- We show that our segmentation algorithm yields an area under curve (AUC) of 0.97 under the ROC curve for breast tumor segmentation compared to 0.89 for single channel MRF using the best feature from the same pool of fea-

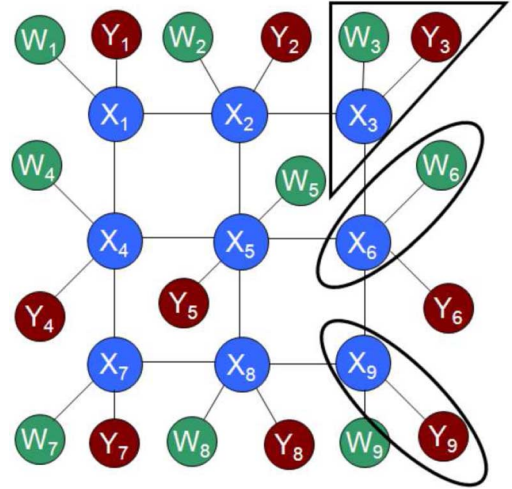


Fig. 1. Multichannel MRF with two observations for each node. The variables inscribed in the triangle represent the joint distribution of the MRF model. If the observation variables ( $Y, W$ ) are conditionally independent given the hidden variable ( $X$ ), the joint distribution can be factorized into two parts, represented by ellipses, allowing loopy belief propagation to solve a multichannel MRF as a single channel MRF (see Section II).

tures as used by our multichannel MRF. Also, the normalized cuts segmentation algorithm and commonly used approaches for breast tumor segmentation, [15], [17] gave lower AUCs (0.92, 0.88, and 0.60) (Section IV).

- Finally we demonstrate that superior segmentation leads to improved feature extraction and tumor characterization. In the specific application presented here, we show that our segmentation method leads to improvement in image-based prediction of breast cancer prognosis by distinguishing between patients with high and low recurrence risks as determined by a validated gene expression assay (Section V).

## II. MULTICHANNEL MRF

To elaborate on the MRF concept, Fig. 1 shows how an image can be modeled as a MRF. Each node ( $X$ ) represents the class of a pixel, and neighboring pixels are connected via edges. In the context of segmentation the goal is to infer the class label for each pixel (e.g., foreground versus background). For a detailed review on single channel MRFs see [21]. Here, we present a multichannel extension of MRFs and elaborate how to incorporate multiple observations in the MRF model. In Fig. 1 each node emits two observations,  $Y$  and  $W$  (similar notion could be extended to more than two observations). The joint probability of the pixel class and the two observations over the entire image is given below

$$P(\mathcal{X}, \mathcal{Y}, \mathcal{W}) \propto \prod_i \phi_{i(\mathcal{MC})}(x_i, y_i, w_i) \prod_{i,j \in E} \psi_{ij}(x_i, x_j) \quad (1)$$

where

$$\phi_{i(\mathcal{MC})}(x_i = c, y_i, w_i) = P(y_i, w_i | \mathcal{G}, \mathcal{M}, \mathcal{M}_c) \quad (2)$$

$$\psi_{ij}(x_i, x_j) = e^{-\beta \mathcal{I}(x_i \neq x_j)}. \quad (3)$$

In (1),  $\phi_{i(\mathcal{MC})}$  is the multichannel node potential for the two observations, and  $\psi$  represents the edge potential, and  $E$  models the adjacency of nodes including only those nodes that have edges between them. The node potential captures the correlation between the features and the class label, indicating the likelihood of  $x_i$  coming from class  $c$  based on the feature values of  $y_i, w_i$  and the feature distribution of class  $c$ . In (3),  $\mathcal{I}(x_i \neq x_j)$  is an indicator function that equals 1 when  $x_i \neq x_j$  and 0 otherwise. As a result,  $\psi$  biases neighboring nodes to have the same class label via the parameter  $\beta$ . In (2),  $\mathcal{GMM}_c$  is a Gaussian mixture model learned for class  $c$  for features  $y, w$  which can be learned from labeled training data. It is clear from (2) that with increasing number of observations the task of modeling the node potential as a joint distribution would become difficult. However, if the features in different channels are conditionally independent given the hidden variable ( $X$ ), we can factorize the right-hand side of (2) as follows:

$$P(y_i, w_i | \mathcal{GMM}_c) = P(y_i | \mathcal{GMM}_c) P(w_i | \mathcal{GMM}_c) \quad \text{if } Y \perp W | X. \quad (4)$$

Equivalently (4) can be written as

$$\begin{aligned} \phi_{i(\mathcal{MC})}(x_i = c, y_i, w_i) \\ = \phi_i^{(y)}(x_i = c, y_i) \cdot \phi_i^{(w)}(x_i = c, w_i) \end{aligned} \quad (5)$$

where  $\phi_i^{(y)}$  and  $\phi_i^{(w)}$  are the individual node potentials for the features  $y$  and  $w$  respectively, which can be learned from labeled training data (see Section III-D). The individual node potentials,  $\phi_i^{(y)}$  and  $\phi_i^{(w)}$ , can be multiplied together to give a single potential representing the multichannel potential. In Section III-C, we present a principled method to choose features satisfying conditional independence to invoke the factorization given in (4). Using (4) the inference machinery for solving a single channel MRF can be reused for multichannel MRFs as well, described below.

#### A. Inference in Multichannel MRFs

Inference in the above mentioned MRF involves the maximum a posteriori (MAP) estimate of the state of each node in the MRF. For tree structured distributions, the MAP estimate for the random fields can be computed efficiently by dynamic programming [22]. If the graphical model is a tree it can also be computed in polynomial time using graph cuts [23]. For more general settings, the belief propagation (BP) algorithm is an efficient way for solving inference problems in graphical models [24]. Specifically, in [24], the authors show that the fixed points arrived at by the BP algorithm correspond to the Bethe approximation of the free energy of a factor graph. Moreover, minimizing the free energy in turn corresponds to minimizing the Kullback–Leibler divergence between a “trial” distribution and the distribution that needs to be recovered [24], [25]. As such belief propagation offers an efficient solution toward estimating inference probabilities in graphical models. It should be noted that the BP algorithm solves inference problems exactly when the factor graph is a tree, and only approximately when the graph has cycles [24]. There are BP algorithms that give good approximate results even for graphical models with cycles [24]. If the

graph has cycles, the corresponding BP algorithm is often referred to as “loopy” belief propagation [19]. Freeman *et al.* [24] present a seminal description as to why BP algorithm works well which is primarily due to its equivalence with minimizing the free energy of the factor graph, thereby arriving at better approximation of the distribution that needs to be recovered. As such, the BP algorithm has significant advantages over other inference methods, [19], [24], [25].

To capitalize on the advantages of the BP algorithm we investigate the loopy belief propagation algorithm to solve our multichannel MRF. Loopy BP is a dynamic message passing algorithm used for doing inference in MRF.  $\delta_{i \rightarrow j}(\mathcal{X}_j)$  is defined to be an incoming message into node  $j$  from its neighboring node  $i$ . Intuitively,  $\delta_{i \rightarrow j}$  captures the degree of belief that node  $i$  has about node  $j$ . To start the process of message passing through the nodes, the node messages are initialized (typically to unity) and then the messages can be updated in the next iteration as follows:

$$\begin{aligned} \delta_{i \rightarrow j}(\mathcal{X}_j) = \frac{1}{Z_{i \rightarrow j}} \sum_{x_i} \phi_{i(\mathcal{MC})}(x_i) \cdot \psi_{ij}(x_i, x_j) \\ \times \prod_{k \in \mathcal{N}(i) - j} \delta_{k \rightarrow i}(x_i) \end{aligned} \quad (6)$$

where  $Z_{i \rightarrow j}$  is a normalization constant, and  $\mathcal{N}(i)$  is a set containing the neighbors of node  $i$ . Equation (6) is repeatedly invoked till the messages converge (the update in each message is less than  $\epsilon$  e.g.,  $10^{-6}$ ). Once the messages have converged the final inference is done by using

$$\hat{\mathcal{P}}(\mathcal{X}_i) = \frac{1}{Z_i} \phi_{i(\mathcal{MC})}(x_i) \prod_{k \in \mathcal{N}(i)} \delta_{k \rightarrow i}(\mathcal{X}_i). \quad (7)$$

The inference engine will output for each node a vector of size  $C \times 1$  ( $C = 2$  for two classes), representing the belief of this node coming from each class. Optimal class label is simply the class with the highest belief.

In the next section, we elaborate on the extraction of pixel-wise feature maps to build a kinetic observation model for our multichannel MRF.

### III. MULTICHANNEL MRF BASED ON DCE-MRI KINETICS

#### A. Candidate Feature Channels

Typically, for DCE MR images we have a precontrast image (captured prior to the injection of a contrast agent) and a number of postcontrast images, captured at different time points after the injection of the contrast agent [20]. The uptake of the contrast agent by different tissues manifests itself in the form of contrast enhancement in postcontrast MR images, and the enhancement patterns, in general, are different for tumor and background tissues [26], [15]. As such DCE MRI derived features hold promise for differentiating between tumor and background tissues and we propose to investigate them as candidate features for our multichannel MRF observation model.

A common way to quantify the enhancement pattern is to compute relative enhancements as compared to the precontrast image [17], [20]. By computing the relative enhancement on a

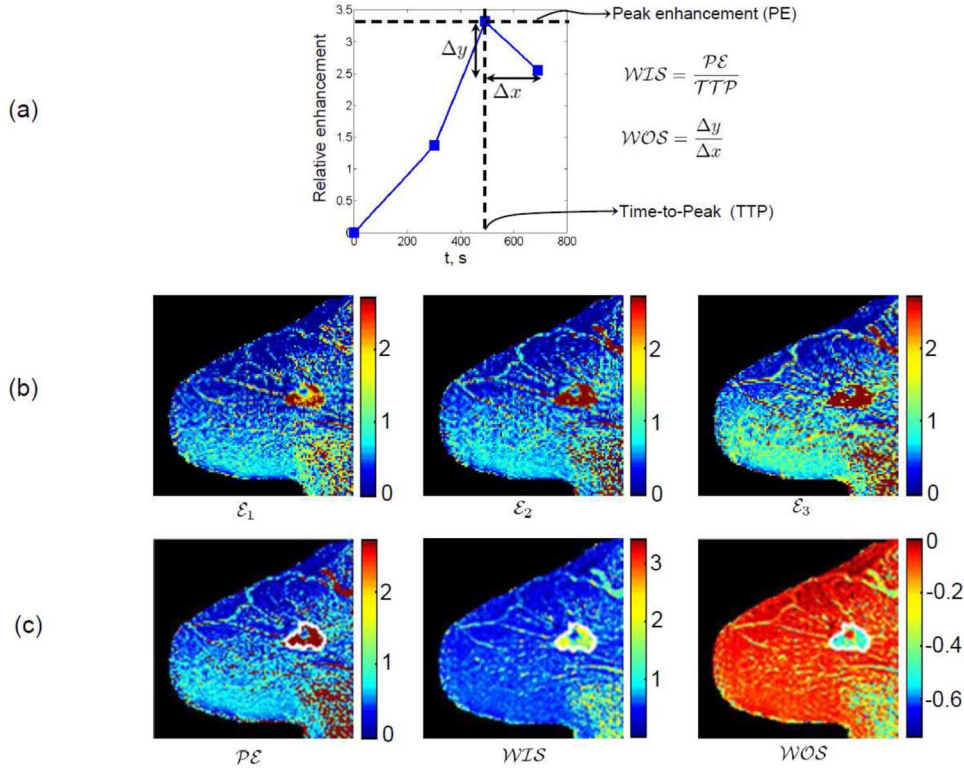


Fig. 2. Candidate features for different observation channels of our multichannel MRF. (a) Illustration of basic kinetic features for a single pixel. (b) Pixel wise maps of enhancement for three postcontrast time points. (c) Pixel wise maps of kinetic features: peak enhancement ( $\mathcal{PE}$ ), wash-in-slope ( $WIS$ ), wash-out-slope ( $WOS$ ).

pixel by pixel basis we can construct pixel wise maps of the relative contrast enhancement as follows:

$$\mathcal{E}(u, v, t) = \frac{I(u, v, t) - I(u, v, t_0)}{I(u, v, t_0)} \quad (8)$$

where  $I(u, v, t)$  represents the intensity of pixel  $(u, v)$  captured at time  $t$ , and  $t_0$  is the precontrast time instant. For a particular pixel, the relative enhancement plotted as a function of time is defined as the kinetic curve [20]. In the literature (e.g., [17], [20]) a number of basic features can be computed from this kinetic curve as illustrated in Fig. 2(a). Based on these features we can derive a rich kinetic feature set by computing the pixel-wise map for each feature as follows.

- Peak enhancement ( $\mathcal{PE}$ )

$$\mathcal{PE}(u, v) = \max_t \mathcal{E}(u, v, t) \quad (9)$$

which represents the peak relative enhancement for every pixel as computed over all postcontrast time points.

- Time to peak ( $TTP$ )

$$TTP(u, v) = \arg \max_t \mathcal{E}(u, v, t) \quad (10)$$

which represents the time at which peak enhancement is achieved.

- Wash-in slope ( $WIS$ )

$$WIS(u, v) = \begin{cases} \frac{\mathcal{PE}(u, v)}{TTP(u, v) - t_0}, & \text{if } TTP(u, v) \neq t_0 \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

which is a measure of the initial uptake rate of the contrast agent for every pixel.

- Wash-out slope ( $WOS$ )

$$WOS(u, v) = \begin{cases} \frac{\mathcal{E}(u, v, t_M) - \mathcal{PE}(u, v)}{t_M - TTP(u, v)}, & \text{if } t_M \neq TTP(u, v) \\ 0, & \text{otherwise} \end{cases} \quad (12)$$

where  $t_M$  is the last postcontrast time instant. The wash-out slope captures the drop in the uptake rate of the contrast agent after the peak enhancement is achieved.

Example pixel-wise maps of above features are depicted in Fig. 2(b). For the rest of the paper we use  $\mathcal{E}_i$  to represent the relative enhancement map for the  $i$ th postcontrast time point. We investigate the utility of these feature maps when included in our multichannel MRF observation model for the purpose of breast tumor segmentation.

## B. MRF Representation-Superpixels

Many computer vision algorithms use the pixel grid as an underlying representation. MRFs are also often defined on this regular grid. The pixel grid, however is not a natural representation of the imaged structure, but rather an artefact of the imaging process [27]. As a result there has been significant work in deriving efficient and perceptually meaningful entities from a low level grouping process [27]–[29]. A popular approach is to divide the image into segments of homogeneity called superpixels



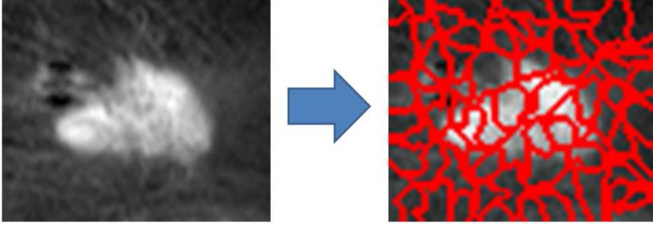


Fig. 3. Illustration of the oversegmentation step using superpixels. Here an image consisting of  $\sim 5000$  pixels has been divided into 100 superpixels.

[8]. Following the approach in [8], we first over-segmented the DCE MR images into superpixels. There are a number of advantages to defining the MRF over the superpixels rather than individual pixels. First, it reduces the complexity of images to a few hundred pixels. Second, they are representationally efficient in that they allow longer range interactions between superpixels, which in the case of a regular grid are restricted to adjacent pixels in the grid [28]. In order to define the MRF neighborhood for superpixels we scan the rows and columns of the superpixelized image and look for transitions. This enables to build the adjacency matrix for superpixels that captures the neighborhood. An example of the oversegmentation step using superpixels is shown in Fig. 3. For this step, we used the publicly available implementation<sup>1</sup> of superpixels based on [28], [29]. Throughout this work the input image was first segmented into 100 superpixels that were used as nodes in the MRF. If  $N$  is the number of MRF nodes, the per-iteration complexity of loopy belief propagation message update (6) ranges from  $O(N)$  to  $O(N^2)$ , depending on the connectedness of the nodes in the graph. As such defining the MRF over superpixels greatly reduces the complexity of the message update step.

### C. Feature Choice for Multichannel MRF

The multichannel MRF described in Section II is based on features that are conditionally independent of one another given the pixel class. It is well known that mutual information is a good indicator of statistical dependence (or lack thereof) between random variables [30]. Small values of mutual information between two random variables indicate little statistical dependence between the two. When certain variables are known, the conditional mutual information captures the influence of this knowledge on the statistical interaction between the variables in question. We thus use conditional mutual information as a criterion to assess the conditional independence of our features. Specifically, in the multichannel MRF of Fig. 1, we are interested in assessing the independence of  $Y$  and  $W$  given  $X$ . Their conditional mutual information is given by

$$MI(Y; W|X) = \sum_{x,y,w} p(x, y, w) \log \frac{p(y, w|x)}{p(y|x)p(w|x)}. \quad (13)$$

The variable  $X$  represents the class of the pixel and hence is discrete, whereas the features represented by  $Y$  and  $W$  are continuous. To compute the distributions required for evaluating the above mutual information, one option is to discretize the continuous variables and use histograms as approximation to the prob-

ability densities. However, the inefficiencies of histograms for modeling probability densities, especially the inherent discontinuities, are well recognized in the literature [31]–[33]. Therefore we employed kernel density estimation [32], [34] for estimating the densities required for computing the mutual information. For all computations, the Gaussian kernel was used. For bandwidth selection, a data driven approach for computing diagonal bandwidth matrix for multivariate Gaussian kernel was employed [34].

We computed  $I(Y; W|X)$  by sequentially setting  $(Y, W)$  to all possible pairs of the following six feature maps: postcontrast enhancements ( $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3$ ), peak enhancement ( $\mathcal{PE}$ ), wash-in-slope ( $\mathcal{WIS}$ ), and wash-out-slope ( $\mathcal{WOS}$ ). The features with the least conditional mutual information were selected as channels for the multichannel MRF (see Section IV-B).

### D. Training-Learning Node Potentials

The training step involves the computation of node potentials as given in (5). We model node potentials as a mixture of Gaussian distributions. A Gaussian mixture model ( $\mathcal{GMM}$ ) is a weighted sum of  $M$  component Gaussian distributions [35]. For example, we can express the node potential for feature  $y$  as follows:

$$\mathcal{P}(y|\mathcal{GMM}_c) = \sum_{r=1}^M w_r g(y|\mu_r, \sigma_r^2) \quad (14)$$

where  $\mathcal{GMM}_c$  represents the Gaussian mixture model for class  $c$  (e.g., tumor and nontumor pixels),  $w_r$  is the weight of the  $r$ th Gaussian component specified by the respective mean and variance parameters,  $\mu_r$  and  $\sigma_r^2$ .  $g(y|\mu_r, \sigma_r^2)$  represents the Gaussian pdf with the respective parameters as follows:

$$g(\mu_r, \sigma_r^2) = \frac{1}{\sqrt{2\pi}\sigma_r} e^{-\frac{(y-\mu_r)^2}{2\sigma_r^2}}. \quad (15)$$

At training step we need to learn the following parameters for class  $c$ :

$$\theta_c = \{w_r, \mu_r, \sigma_r^2\}; \quad r = 1, \dots, M. \quad (16)$$

There are several techniques available for estimating the parameters of a  $\mathcal{GMM}$  [35]. A popular approach for estimating these parameters is through an iterative maximum likelihood estimation using the expectation maximization (EM) algorithm [36]. For learning the node potentials for our multichannel MRF we employed the EM algorithm based method (Matlab statistics toolbox<sup>2</sup>) and used  $M = 5$  Gaussian components. To enable maximal usage of the training data we used a leave-one-out cross validation strategy for computing node potentials. For every test image, the rest of the data was considered as training set and node potentials were computed using the training set.

## IV. SEGMENTATION EXPERIMENTS

### A. Dataset

All experiments presented in this paper were conducted on DCE breast MR images of 60 women diagnosed with breast cancer. The dataset consists of bilateral breast MRI sagittal

<sup>1</sup><http://www.cs.sfu.ca/~mori/research/superpixels/>

<sup>2</sup><http://www.mathworks.com/help/toolbox/stats/>

scans collected at our institution from 2007–2010. The ages of the women at the time of the imaging ranged from 37 to 74 years with a mean age of 55.5 years. Of the women analyzed, 93% had their primary tumor classified as T1 i.e., the tumor was 2.0 cm or less in the greatest dimension ( $GD$ ). The rest of the primary tumors were T2 ( $2\text{ cm} < GD \leq 5\text{ cm}$ ). Within the T1 tumors the distribution was as follows: 8% ( $0.1\text{ cm} < GD \leq 0.5\text{ cm}$ ); 37% ( $0.5\text{ cm} < GD \leq 1\text{ cm}$ ); 45% ( $1\text{ cm} < GD \leq 2\text{ cm}$ ). The imaging protocol is as follows: The women were imaged prone in a 1.5T scanner (GE LX echo, GE Healthcare, or Siemens Sonata, Siemens); matrix size:  $512 \times 512$ ; slice thickness: 2.4–4.4 mm; flip angle:  $25^\circ$  or  $30^\circ$ . The images were collected before and after the administration of gadodiamide (Omniscan) or gadobenate dimeglumine (MultiHance) contrast agents. Dynamic contrast enhanced images were acquired at 90-s intervals for three postcontrast time points.

The women in the dataset had estrogen receptor positive (ER+), node negative tumors, which were analyzed with the Oncotype DX prognostic gene expression assay [18]. Oncotype DX is a validated reverse-transcriptase-polymerase-chain-reaction (RT-PCR) assay (developed by Genomic Health Inc.) that measures the expression of 21 genes in RNA from formalin-fixed paraffin-embedded (FFPE) tumor tissue samples from the primary breast cancer [37]. The final outcome of the Oncotype DX assay is a continuous recurrence score that predicts the likelihood of breast cancer recurrence in 10 years after the treatment (risk: low  $\leq 17\%$ ,  $18\% < \text{medium} < 30\%$ , high  $\geq 31\%$ ). It has been shown that the benefit of adjuvant chemotherapy regimen starts becoming significant only in patients with an Oncotype score greater than 30 [37]. To learn feature statistics for distinguishing the breast tumor and non-tumor area of an image, a fellowship-trained board-certified breast imaging radiologist delineated the lesion boundaries using the software ITK-SNAP [38]. The lesion boundaries were defined on a central representative slice selected by the radiologist, as usually assessed in standard clinical practice.

### B. Segmentation Results

Using the strategy described in Section III-C we computed pairwise conditional mutual information between all possible pairs of the following six features:  $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{PE}, \mathcal{WIS}, \mathcal{WOS}$ . A visualization of the mutual information matrix that resulted from this pairwise computation is shown in Fig. 4. The feature pair  $(\mathcal{E}_1, \mathcal{WOS})$  had the least conditional mutual information (0.02) and were selected as the feature maps for our MRF. The mutual information was computed in a leave-one-out sense and the same features were consistently selected for all leave-one-out folds.

The input to the inference engine is the superpixelized version of each feature map. For every superpixel the engine outputs a probability estimate that the superpixel belongs to a tumor region. These inference probabilities are calculated using (7). In Fig. 5 we show a comparison between single channel and multichannel MRF segmentation. The multichannel MRF is based on the feature pair:  $(\mathcal{E}_1, \mathcal{WOS})$ . For single channel MRF we tested each of the individual features  $(\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{PE}, \mathcal{WIS}, \mathcal{WOS})$ ,

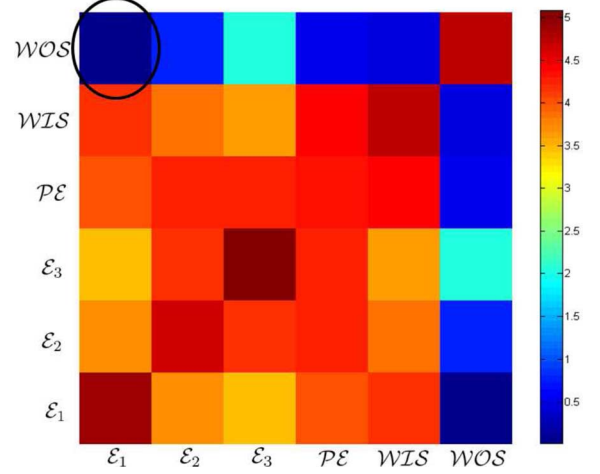


Fig. 4. Visualization of the mutual information matrix generated by computing conditional mutual information between feature pairs. The encircled cell corresponds to the feature pair with the least conditional mutual information.

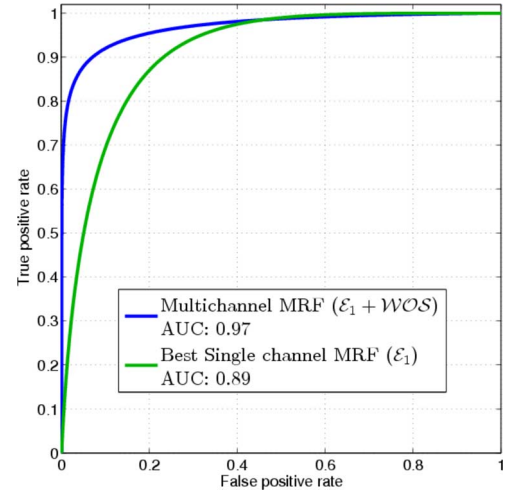


Fig. 5. ROC comparison for breast tumor segmentation between multichannel and single channel MRF. The multichannel MRF using the feature pair  $(\mathcal{E}_1, \mathcal{WOS})$  gives an AUC of 0.97 under the ROC curve as compared to 0.89 for single channel MRF based on the best individual feature  $(\mathcal{E}_1)$ .

and show the ROC for  $\mathcal{E}_1$  that gives the highest AUC. The inference probabilities are computed in a leave-one-out fashion i.e., the node potentials used in (7) are based on all images but the test image. The manual segmentation provided by the radiologist was used as ground-truth to assess the pixelwise agreement for ROC analysis. The ROC scores were derived from the inference probabilities. The multichannel MRF based method gives an AUC of 0.97 outperforming the best single channel MRF that gives an AUC of 0.89.

We compare our method against the following segmentation algorithms.

### C. Normalized Cuts

The normalized cuts algorithm [8] is a spectral graph theoretic method that is widely used for diverse segmentation applications [11]–[14]. It is an unsupervised segmentation technique that maximizes both the dissimilarity between different groups and the similarity within groups of pixels. The normalized cuts

algorithm [8] models an image as a graph  $G = (V, E)$  and then aims to find a partition of the graph to minimize the cut cost normalized by the cost of the total edge connections to all the nodes in the graph

$$Ncut(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(A, B)}{assoc(B, V)} \quad (17)$$

where  $cut(A, B)$  is the sum of edge costs that are being cut by the partition i.e.,  $cut(A, B) = \sum_{u \in A, v \in B} w(u, v)$ , and  $assoc(A, V) = \sum_{u \in A, t \in V} w(u, t)$  is the sum of costs for edge connections from nodes in  $A$  to all nodes in the graph. Minimizing normalized cut exactly is NP-complete but [8] proposed an approximate method that is able to solve the problem as a generalized eigenvalue system (see [8] for details). Usually the resultant eigenvectors are thresholded to obtain partitions of the image. However, in order to generate ROC curves we used the eigenvectors as the scores. We used the publicly available implementation of normalized cuts in our experiments [39]. We ran the normalized cuts algorithm using each of the six features and report the best performance in Fig. 6 that gives an AUC of 0.92 by using  $\mathcal{E}_1$  as an input.

#### D. Fuzzy C Means

Fuzzy C means (FCM) is a popular method for segmenting breast lesions from DCE-MRI images e.g., [15], [16]. FCM is an unsupervised learning technique [40] that aims to find a fuzzy partitioning of data points into clusters. If  $u_{ki}$  represents the extent of membership of the  $i$ th data point to the  $k$ th class, the FCM problem is to find all  $u_{ki}$  such that the following within group weighted squared Euclidean distance is minimized:

$$\arg \min_{\mathbf{U}} \sum_{k=1}^c \sum_{i=1}^{N_c} u_{ki}^b \|\mathbf{x}_i - \boldsymbol{\mu}_k\|^2 \quad (18)$$

where  $b$  is a weighting exponent on each fuzzy membership, and  $\boldsymbol{\mu}_k$  is the weighted mean for  $k$ th class given the membership probabilities  $u_{ki}$ . To ensure that  $u_{ki}$  represent probabilities the above optimization is subject to the following constraints:

$$\sum_{k=1}^c u_{ki} = 1, \quad \forall i; \quad 0 \leq u_{ki} \leq 1, \quad \forall k. \quad (19)$$

For our experiments we used Matlab's<sup>3</sup> standard implementation of the FCM algorithm with  $b = 2$ . As input to the FCM algorithm we tested all the individual features and pairs of features, and FCM performed the best when  $\mathcal{E}_1$  and  $\mathcal{WOS}$  were concatenated together to represent every pixel by a  $2 \times 1$  feature vector. Connected component analysis was done on the FCM output and the largest component was selected as the segmented lesion [16]. The FCM probabilities were used as scores for generating the ROC curve which gave an AUC of 0.88 (Fig. 6).

#### E. FCM on Enhancement Variance

In [17], the authors first compute pixelwise variance of relative enhancement of breast MRI images. They then run FCM

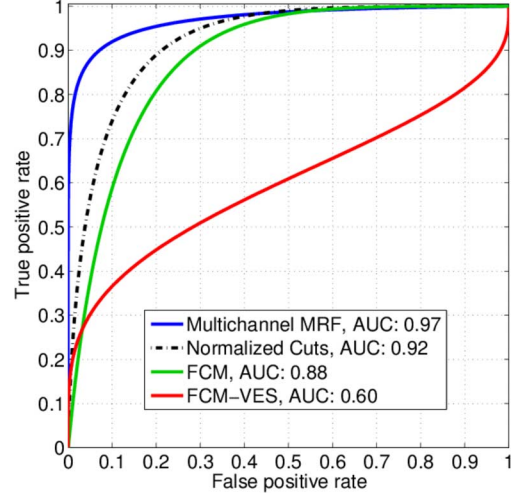


Fig. 6. ROC comparison for breast tumor segmentation. The multichannel MRF based segmentation method gives an AUC of 0.97 under the ROC curve as compared to 0.92, 0.88, and 0.60 for normalized cuts [8], FCM [15], [16] and FCM-VES [17], respectively.

on the pixelwise variance. This method gives an AUC of 0.60 as shown in Fig. 6.

#### F. Performance Comparison

Fig. 6 presents the comparison of all the above approaches against our method. Our segmentation algorithm yields an AUC of 0.97 under the ROC curve for breast tumor segmentation compared to 0.92, 0.89, 0.88, and 0.60 for the normalized cuts algorithm, single channel MRF, FCM, and FCM on enhancement variance respectively. In Fig. 7, we show representative segmentation results of our algorithm (row 7), FCM [15] (row 3), FCM-VES [17], (row 4), normalized cuts algorithm (row 5), and the best single channel MRF (row 6).

### V. CLASSIFICATION EXPERIMENTS

In this section we investigate if better quality segmentation leads to better feature extraction and tumor characterization. In the specific application presented here we explore the role of segmentation in improved prediction of breast cancer prognosis as assessed by a validated gene expression assay (Oncotype DX). Although the outcome of the Oncotype assay is a continuous score, it has been shown that the benefit of a chemotherapy regimen starts becoming significant only in patients with an Oncotype score greater than 30, which is the high recurrence risk category [37]. As such, from a clinical perspective, the decision of low versus high recurrence risk is very important to choose women who are expected to benefit the most from adjuvant chemotherapy [41], [42]. Recent data indicate that in addition to distinguishing between benign and malignant findings, breast MRI kinetic tumor features and contrast enhancement patterns, are also significantly associated with histopathological findings that are indicative of subsequent prognosis and probability of relapse [43], [44]. To explore the effect of better segmentation we focus on the task of binary prediction of Oncotype scores as high or low, where we consider patients with scores greater than 30 in the high risk category.

<sup>3</sup>Matlab version 7.13, R2011b



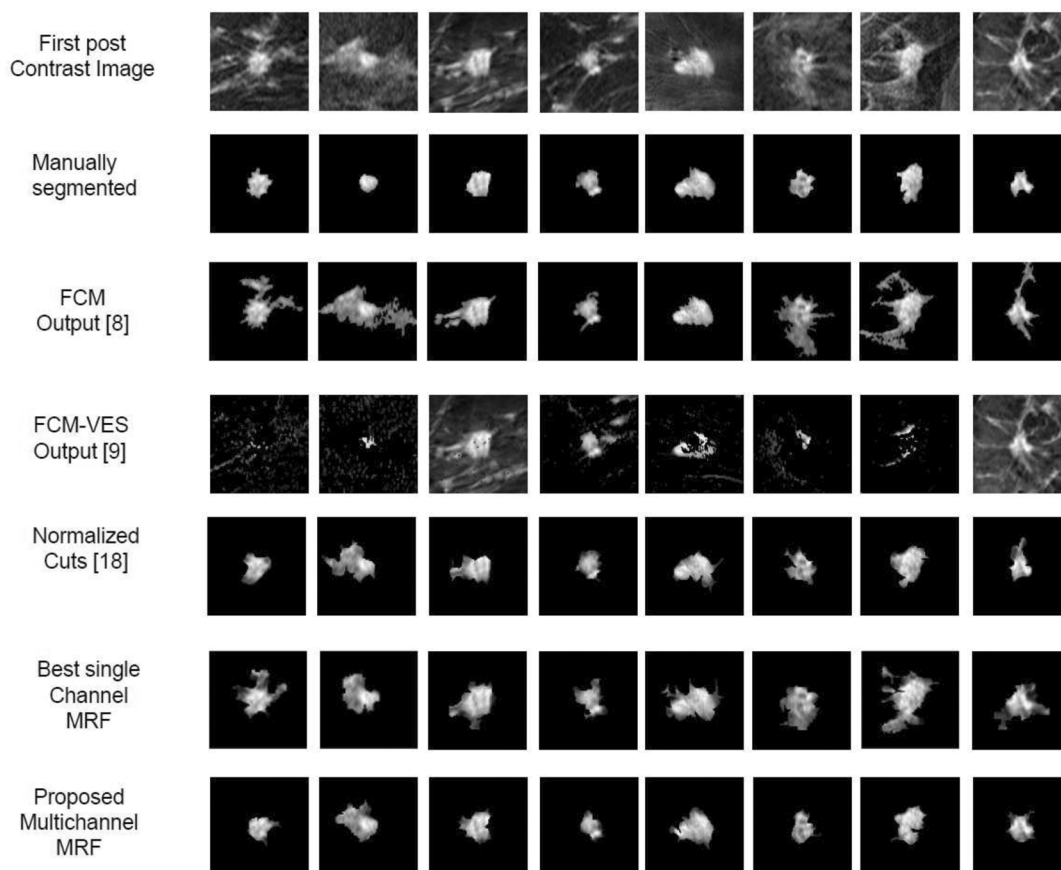


Fig. 7. Qualitative comparison of segmentation results as overlaid on the first postcontrast image. Results of multichannel MRF approach (Row 7) are qualitatively more similar to the radiologist's provided ground truth (Row 2) as compared to traditional approaches (Row 3, 4), the normalized cuts algorithm (Row 5), and the best single channel MRF (Row 6).

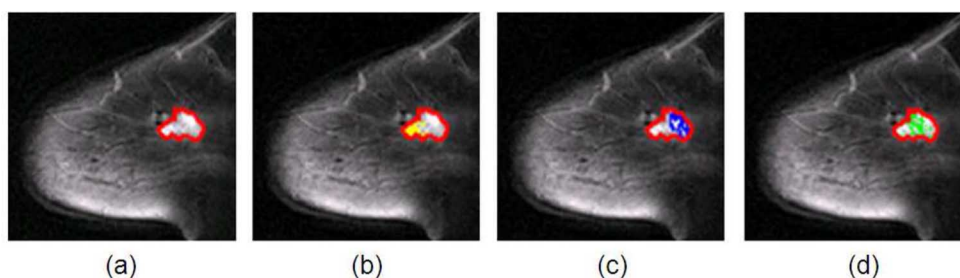


Fig. 8. Illustration of kinetic partitioning based on the TTP pixel-wise values. (a) Segmented lesion. (b) Set 1 pixels (yellow). (c) Set 2 pixels (blue). (d) Set three pixels (green). Set  $i$  consists of pixels that have their peak enhancement at the  $i$ th postcontrast time point.

### A. Features for Classification

Here, we propose to investigate a rich set of kinetic statistics derived from DCE MRI images for the task of detecting patients with high recurrence risk. We extract the kinetic statistics in a two phase process: In the first phase the time to peak (TTP) for every pixel within the segmented tumor is computed. In the second phase we cluster the pixels based on their TTP values. In the dataset under consideration there are three post-contrast time points. TTP thus can have three discrete values and the clustering step results in three groups of pixels. It has been suggested that the heterogeneities in the contrast agent uptake can be used for tumor characterization [20], [45]. One way

to interpret the kinetic heterogeneity is to divide the pixels into clusters of homogeneity. Therefore the idea behind clustering the pixels according to their TTP values is to group the pixels that show similar contrast uptake behavior such that statistics within these clusters of kinetic homogeneity could be explored. This step partitions the pixels into as many sets as the number of postcontrast time points, i.e., set  $i$  consists of pixels that achieve their peak enhancement at the  $i$ th postcontrast time point. Fig. 8 illustrates these partitions for an example case. In the second step, pixel-wise maps of the following features were computed: peak enhancement (PE), wash-in-slope (WIS), wash-out-slope (WOS), (Fig. 2). Based on these pixel groupings we compute partition-wise kinetic statistics as follows: Let  $\mathcal{M}$  be the pixel



TABLE I

COMPARISON OF SINGLE FEATURE CLASSIFICATION PERFORMANCE IN TERMS OF AREA UNDER THE ROC CURVE WHEN OUTPUTS OF VARIOUS SEGMENTATION STRATEGIES WERE FED INTO THE FEATURE EXTRACTION STAGE. IN COLUMN 1 FEATURE NAMES FOLLOW THE NOTATION OF (20)–(22). E.G.,  $\mu(1, PE)$  REPRESENTS THE MEAN FEATURE MAP VALUE OF PEAK-ENHANCEMENT FOR SET 1 PIXELS. COLUMNS 3–6 SHOW THE RESPECTIVE AUCs. IN EACH COLUMN THE BEST TWO VALUES ARE SHOWN IN BOLD

Feature	Multichannel MRF	Single Channel MRF	Normalized Cuts	FCM	FCM-VES
$\mathcal{P}(Set = 1 \mathcal{M})$	0.76	0.67	0.70	0.68	0.60
$\mathcal{P}(Set = 2 \mathcal{M})$	0.76	0.68	0.69	0.66	0.64
$\mathcal{P}(Set = 3 \mathcal{M})$	0.68	0.53	0.57	0.56	0.51
$\mu(1, \mathcal{PE})$	0.77	0.69	0.71	0.65	0.57
$\mu(2, \mathcal{PE})$	0.74	0.60	0.67	0.59	0.63
$\mu(3, \mathcal{PE})$	0.58	0.52	0.51	0.57	0.54
$\sigma^2(1, \mathcal{PE})$	0.75	0.67	0.69	0.67	0.55
$\sigma^2(2, \mathcal{PE})$	0.62	0.58	0.56	0.53	0.51
$\sigma^2(3, \mathcal{PE})$	0.59	0.52	0.51	0.57	0.55
$\mu(1, \mathcal{WIS})$	<b>0.82</b>	<b>0.71</b>	<b>0.73</b>	<b>0.69</b>	0.54
$\mu(2, \mathcal{WIS})$	0.78	<b>0.70</b>	0.72	<b>0.70</b>	0.60
$\mu(3, \mathcal{WIS})$	0.54	0.51	0.53	0.51	0.52
$\sigma^2(1, \mathcal{WIS})$	0.74	0.64	0.68	0.65	0.54
$\sigma^2(2, \mathcal{WIS})$	0.75	0.69	0.70	0.67	0.59
$\sigma^2(3, \mathcal{WIS})$	0.64	0.52	0.62	0.56	0.60
$\mu(1, \mathcal{WOS})$	<b>0.81</b>	0.68	<b>0.74</b>	0.68	<b>0.65</b>
$\mu(2, \mathcal{WOS})$	0.76	0.64	0.67	0.66	0.54
$\mu(3, \mathcal{WOS})$	0.74	0.69	0.70	0.67	0.60
$\sigma^2(1, \mathcal{WOS})$	0.75	0.70	0.71	0.63	0.57
$\sigma^2(2, \mathcal{WOS})$	0.74	0.66	0.69	0.66	0.52
$\sigma^2(3, \mathcal{WOS})$	0.79	0.69	0.72	0.64	<b>0.65</b>

partitioning such that  $\mathcal{M}_k$  represents the membership mapping of pixel  $k$  to its respective set. Given  $\mathcal{M}$  one can derive the following set-wise statistics.

- Posterior probability of observing set  $i$  given the partition

$$\mathcal{P}(Set = i|\mathcal{M}) = \frac{1}{N} \sum_{k=1}^N \delta(\mathcal{M}_k = i) \quad (20)$$

where  $\delta(\mathcal{M}_k = i)$  is an indicator function that equals 1 when  $\mathcal{M}_k = i$ , and zero otherwise.  $N$  is the total number of pixels. These  $N$  pixels may come from an arbitrarily shaped segmentation mask specifying the lesion.

- Mean value of feature map  $j$  for Set  $i$

$$\mu(i, j) = \frac{\sum_{k=1}^N f_j(k) \cdot \delta(\mathcal{M}_k = i)}{\sum_{k=1}^N \delta(\mathcal{M}_k = i)} \quad (21)$$

where  $f_j(k)$  is the value of the  $j$ th feature map for  $k$ th pixel, and the feature map can be either PE, WIS, or WOS.

- Variance of feature map  $j$  for Set  $i$

$$\sigma^2(i, j) = \frac{\sum_{k=1}^N (f_j(k) - \mu(i, j))^2 \cdot \delta(\mathcal{M}_k = i)}{\sum_{k=1}^N \delta(\mathcal{M}_k = i)}. \quad (22)$$

Based on the above definitions,  $m$  pixel partitions and  $n$  feature maps would result into a total of  $m(2n + 1)$  features. For this dataset we had three postcontrast time points and based on the TTP values the pixels were partitioned into three sets ( $m = 3$ ). Within these partitions we computed statistics for three feature

maps ( $n = 3$ ) i.e.,  $\mathcal{PE}$ ,  $\mathcal{WIS}$ , and  $\mathcal{WOS}$ . This resulted in  $m(2n + 1) = 21$  kinetic statistic features.

### B. Single Feature Classifiers

To assess the classification performance of individual features we learned single feature classifiers to distinguish between the categories of low and high recurrence risk. For each feature we learned a leave-one-out binary decision stump classifier [46] using the individual feature values as scores for generating ROCs to evaluate the univariate performance of each feature. Table I shows the AUCs for the ROCs corresponding to each feature classifier. When features extracted from the output of the multichannel MRF algorithm were used for classification, 15 of the 21 features give AUCs greater or equal to 0.74 with an average AUC of 0.76. In Table I we also present comparison with single feature classification performance when features were extracted from the outputs of other segmentation techniques. For all the features, AUCs using the multichannel MRF based segmentation are higher than that for other approaches.

### C. Multi-Feature Classifier

Using the features described in Section V-A we followed a leave-one-out sequential forward feature selection scheme. For each leave-one-out loop the feature selection was done as follows: Starting with a single feature, we sequentially added features, learned a linear support vector machine (SVM) on the training set corresponding to the current cross-validation fold, and kept adding the features till the incremental classification improvement dropped below a threshold ( $10^{-4}$ ) as compared to

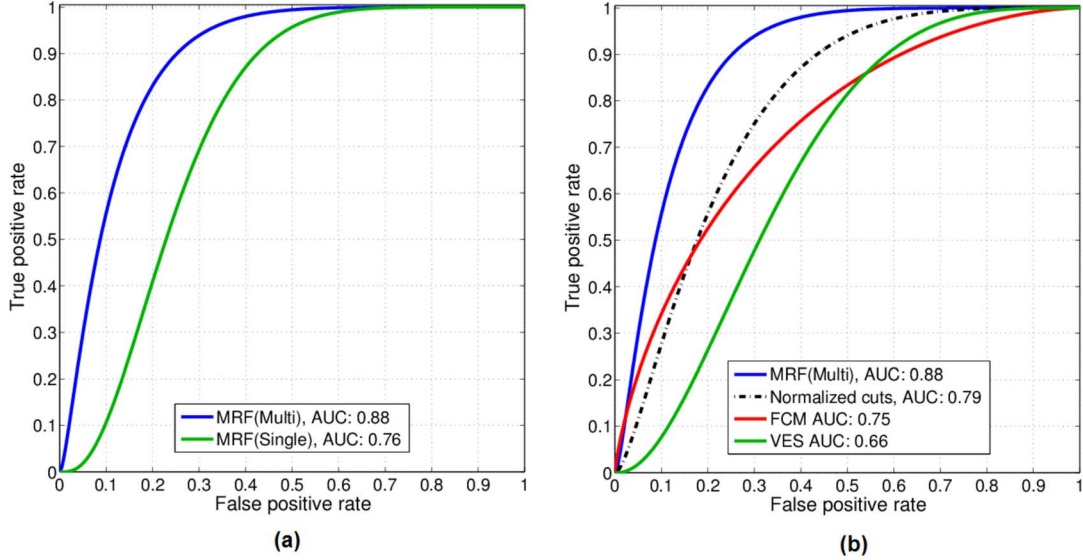


Fig. 9. ROC Curves for the multi-feature SVM for classifying high versus low breast cancer recurrence risk. We get the highest AUC when the output of multi-channel MRF segmentation method is fed to the feature extraction stage.

TABLE II

SELECTION FREQUENCIES OF THE KINETIC STATISTIC FEATURES OVER ALL THE CLASSIFICATION LEAVE-ONE-OUT LOOPS. FOR EVERY LEAVE-ONE-OUT LOOP A SEQUENTIAL FORWARD FEATURE SELECTION SCHEME WAS FOLLOWED. IN COLUMN 1 FEATURE NAMES FOLLOW THE NOTATION OF (20)(22). E.G.,  $\mu(1, PE)$  REPRESENTS THE MEAN FEATURE MAP VALUE OF PEAK-ENHANCEMENT FOR SET 1 PIXELS. COLUMN 2 SHOWS THE SELECTION FREQUENCY OF EACH FEATURE OVER ALL CROSS VALIDATION FOLDS

Feature	Selection Frequency
$\mathcal{P}(Set = 1 \mathcal{M})$	60/60 (100.00%)
$\mathcal{P}(Set = 2 \mathcal{M})$	60/60 (100.00%)
$\mathcal{P}(Set = 3 \mathcal{M})$	52/60 (86.67%)
$\mu(1, \mathcal{PE})$	60/60 (100.00%)
$\mu(2, \mathcal{PE})$	57/60 (95.00%)
$\mu(3, \mathcal{PE})$	43/60 (71.67%)
$\sigma^2(1, \mathcal{PE})$	54/60 (90.00%)
$\sigma^2(2, \mathcal{PE})$	47/60 (78.33%)
$\sigma^2(3, \mathcal{PE})$	42/60 (70.00%)
$\mu(1, \mathcal{WIS})$	58/60 (96.67%)
$\mu(2, \mathcal{WIS})$	60/60 (100.00%)
$\mu(3, \mathcal{WIS})$	39/60 (65.00%)
$\sigma^2(1, \mathcal{WIS})$	56/60 (93.33%)
$\sigma^2(2, \mathcal{WIS})$	60/60 (100.00%)
$\sigma^2(3, \mathcal{WIS})$	46/60 (76.67%)
$\mu(1, \mathcal{WOS})$	60/60 (100.00%)
$\mu(2, \mathcal{WOS})$	60/60 (100.00%)
$\mu(3, \mathcal{WOS})$	52/60 (86.67%)
$\sigma^2(1, \mathcal{WOS})$	60/60 (100.00%)
$\sigma^2(2, \mathcal{WOS})$	55/60 (91.67%)
$\sigma^2(3, \mathcal{WOS})$	53/60 (88.33%)

the previous feature set. This SVM classifier was tested on the left out example and the process was repeated for all cross-validation folds. Table II shows the frequency of selection for each of the features across all leave-one-out loops.

The ROC for the multi-feature classifiers when the output of different segmentation strategies were fed into the classification

stage is shown in Fig. 9. The multi-feature classifier was able to distinguish between low and high recurrence risk patients with an  $AUC = 0.88$  under the ROC curve when the proposed multi-channel MRF segmentation strategy was employed, which was higher than that for other segmentation algorithms [Fig. 9(a) and (b)] which give lower AUCs of 0.79, 0.76, 0.75, 0.66 for normalized cuts, single channel MRF, FCM, and FCM-VES, respectively.

## VI. DISCUSSION

In Section IV-B, we derive features for our multi-channel MRF that allow us to invoke conditional independence conditions by computing pairwise conditional mutual information between all possible pairs of features ( $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{PE}, \mathcal{WIS}, \mathcal{WOS}$ ). Although data-driven, the selection of the first postcontrast enhancement ( $\mathcal{E}_1$ ) and wash-out slope ( $\mathcal{WOS}$ ) features as different observation channels of the multichannel MRF deserves attention. These two features capture different contrast agent uptake properties, each depending on different directions of vascular permeability e.g.,  $\mathcal{E}_1$  represents the initial uptake of the contrast agent while  $\mathcal{WOS}$  captures the drop in the uptake after the peak enhancement. As such they are less correlated with one another. The rest of the kinetic features are essentially derived from one another and are not expected to contain much uncorrelated information as is evident from the mutual information matrix of Fig. 4. As mentioned in (13) the conditional mutual information is conditioned on  $X$  i.e., the class of the node as to whether it belongs to tumor or background tissue. As such the selection of  $\mathcal{E}_1$  and  $\mathcal{WOS}$  as good features for segmentation is also noteworthy as they capture properties of the contrast uptake in postcontrast MR images, which are shown to be different for background and tumor tissues [26], [15]. Although we use cross validation for generalization, and  $\mathcal{E}_1, \mathcal{WOS}$  do seem to represent good features for segmentation, their applicability for other cancer types requires more extensive validation.

In this study, we have investigated pairs of features for the multichannel MRF model. In principle, the multichannel MRF framework is applicable to any number of channels as long as the features satisfy conditional independence criterion. For the feature set currently in question ( $\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3, \mathcal{PE}, \mathcal{WIS}, \mathcal{WOS}$ ), since only one pair satisfied the empirical test of conditional independence, adding more features from the same feature set would have essentially included features which already had significant mutual information with the selected pair ( $\mathcal{E}_1, \mathcal{WOS}$ ). In the future, however, if more elaborate feature sets are explored, the possibility of experimenting with more channels remains open. We should also note that in Fig. 4, the diagonal of the mutual information matrix varies for different features, unlike what one would expect for a correlation matrix, where all diagonal values are same and equal to unity. The reason for this is that the mutual information of a random variable with itself represents the entropy of the variable and hence would vary from variable to variable [30].

It is worth mentioning that the work presented in [47] comes closest in concept to ours in terms of using multiple features for MRF. However, they assume *a priori* independence between features, whereas in our current work we present a principled method to derive conditionally independent features to arrive at appropriate features for the multichannel MRF. Moreover, [47] solves the MRF using an expectation maximization scheme, whereas we present how inference methods like loopy BP could be extended to multichannel MRFs. As mentioned in Section II-A, the BP algorithm is equivalent to minimizing the free energy of the graphical model, thereby arriving at better approximation of the distribution that needs to be recovered.

We should note that while learning the node potentials for our multichannel MRF we are modeling the features as a mixture of Gaussian PDFs. The features in and of themselves may not necessarily be Gaussians. However, it has been shown that the Gaussian mixture models provide good practical approximations to non-Gaussian noise processes [48]. That said, approximating the features with other distributions is a potential direction for future research.

We have also explored the impact of improved segmentation on the characterization and classification of breast cancer tumors. Our multichannel MRF segmentation method gives a much lower pixelwise false positive rate for the same true positive rate as depicted in the ROC curves of Fig. 6. This perhaps explains the improvement in classifying the Oncotype recurrence risk categories when tumors segmented by our algorithm are used for feature extraction. From a clinical perspective, the minimally-invasive estimation of recurrence risk from DCE-MRI features is an important question as it could become a useful decision tool for routing candidate patients to appropriate treatment options [49]. Our results suggest that improved segmentation could play a significant role in assisting this clinical decision. Moreover, the DCE-MRI features are currently extracted from a representative 2-D slice of the primary lesion. Kinetic partitioning of the entire 3-D volume of the lesion could potentially lead to richer statistics that may further improve the prediction of breast cancer recurrence in future work. Also, our current work focuses on accurately segmenting the tumor given a coarsely defined rectangular ROI. Adapting

the existing method for a fully automated tumor detection is a potential topic for future research.

## VII. CONCLUDING REMARKS

In this paper, we have presented a methodological framework for a multichannel extension of MRFs for breast tumor segmentation from DCE-MRI images. Using conditional independence we have shown that loopy belief propagation can solve a multichannel MRF as a single channel MRF. We have presented a method to search for optimal MRF features that satisfy conditional independence conditions by employing pairwise conditional mutual information. Using this framework we incorporate a kinetic observation model derived from DCE breast MR images into the multichannel MRF and demonstrate superior segmentation results (AUC = 0.97). Our multichannel MRF outperforms single channel MRF (AUC: 0.89) employing the best individual feature from the same pool of features as used by the multichannel MRF. Our approach also performs significantly better as compared to the well known normalized cuts segmentation algorithm and commonly used previous methods for breast tumor segmentation (AUCs: 0.92, 0.88, 0.60). Moreover, we also demonstrate that our improved segmentation leads to better feature extraction and tumor characterization with an application on predicting breast cancer prognosis as determined by a validated tumor gene expression assay. Specifically, when tumors segmented by our algorithm were fed into an SVM classifier, it was able to distinguish between women with high and low breast cancer recurrence risk with an AUC of 0.88 under the ROC as compared to AUCs of 0.79, 0.76, 0.75, and 0.66 for other segmentation methods. The decision of low versus high recurrence risk is very important from a clinical perspective as it could enable to select patients that are expected to benefit the most from adjuvant chemotherapy [50], [41], [42]. We believe that the proposed methodological framework could also be extended to other imaging modalities and organs as well.

## REFERENCES

- [1] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active contour models," *Int. J. Comput. Vis.*, vol. 1, no. 4, pp. 321–331, 1988.
- [2] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models—Their training and application," *Comput. Vis. Image Underst.* vol. 61, pp. 38–59, Jan. 1995 [Online]. Available: <http://dl.acm.org/citation.cfm?id=206543.206547>
- [3] C. Tauber, H. Batatia, and A. Ayache, "Quasi-automatic initialization for parametric active contours," *Pattern Recognit. Lett.*, vol. 31, no. 1, pp. 83–90, 2010.
- [4] S. Saha and S. Bandyopadhyay, "MRI brain image segmentation by fuzzy symmetry based genetic clustering technique," in *IEEE Congr. Evol. Comput.*, 2007, pp. 4418–4424.
- [5] O. K. Yoon, D. M. Kwak, and K. H. P. D. W. Kim, "MR brain image segmentation using fuzzy clustering," in *Fuzzy Syst. Conf. Proc.*, 1999, vol. 2, pp. 22–25.
- [6] D. Q. Zhang and S. C. Chen, "A novel kernelized fuzzy c-means algorithm with application in medical image segmentation," *Artif. Intell. Med.*, vol. 32, pp. 37–50, 2004.
- [7] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain MR images through a hidden Markov random field model and the expectationmaximization algorithm," *IEEE Trans. Med. Imag.*, vol. 20, no. 1, pp. 45–57, Jan. 2001.
- [8] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 8, pp. 888–905, Aug. 2000.
- [9] Y. Weiss, "Segmentation using eigenvectors: A unifying view," in *Proc. Int. Conf. Comput. Vis.*, 1999, pp. 975–982.

- [10] A. Y. Ng, M. I. Jordan, and Y. Weiss, *Advances in Neural Information Processing Systems*. Cambridge, MA: MIT Press, 2001, pp. 849–856.
- [11] J. C. Gamio, S. Belongie, and S. Majumdar, “Normalized cuts in 3d for spinal MRI segmentation,” *IEEE Trans. Med. Imag.*, vol. 23, no. 1, pp. 36–44, Jan. 2004.
- [12] C. Fowlkes, S. Belongie, F. Chung, and J. Malik, “Spectral grouping using the Nyström method,” *IEEE Pattern Anal. Mach. Intell.*, vol. 26, no. 2, pp. 214–225, Feb. 2004.
- [13] J. Zhao, L. Zhang, W. Zheng, H. Tian, D.-M. Hao, and S.-H. Wu, J. He, X. Liu, E. Krupinski, and G. Xu, Eds., “Normalized cut segmentation of thyroid tumor image based on fractional derivatives,” in *Health Information Science*. New York: Springer, 2012, vol. 7231, LNCS, pp. 100–109.
- [14] P. Felzenszwalb and D. Huttenlocher, “Efficient graph-based image segmentation,” *Int. J. Comput. Vis.*, vol. 59, pp. 167–181, 2004.
- [15] N. Bhooshan, M. L. Giger, S. A. Jansen, H. Li, L. Lan, and G. M. Newstead, “Cancerous breast lesions on DCE MR images: Computerized characterization for image based prognostic markers,” *Radiology*, vol. 254, no. 3, pp. 680–690, 2010.
- [16] W. Chen, M. L. Giger, and U. Bick, “A fuzzy c-means (FCM)-based approach for computerized segmentation of breast lesions in dynamic contrast-enhanced MR images,” *Acad. Radiol.*, vol. 13, no. 1, pp. 63–72, Jan. 2006.
- [17] S. Lee, J. H. Kim, N. Cho, J. S. Park, Z. Yang, Y. S. Jung, and W. K. Moon, “Multilevel analysis of spatiotemporal association features for differentiation of tumor enhancement patterns in breast DCE-MRI,” *Med. Phys.*, vol. 37, p. 3940, 2010.
- [18] A. B. Ashraf, S. Gavenonis, D. Daye, C. Mies, M. D. Feldman, M. Rosen, and D. Kontos, “A multichannel markov random field approach for automated segmentation of breast cancer tumor in DCE-MRI data using kinetic observation model,” in *MICCAI (3)*, G. Fichtinger, A. L. Martel, and T. M. Peters, Eds. New York: Springer, 2011, vol. 6893, LNCS, pp. 546–553.
- [19] T. Meltzer, A. Globerson, and Y. Weiss, “Convergent message passing algorithms—A unifying view,” in *Proc. 28th Conf. Uncertainty Artif. Intell. (UAI)*, 2009.
- [20] N. Hylton, “Dynamic contrast-enhanced magnetic resonance imaging as an imaging biomarker,” *J. Clin. Oncol.*, vol. 24, pp. 3293–3298, 2006.
- [21] C. M. Bishop, *Pattern Recognition and Machine Learning (Information Science and Statistics)*. New York: Springer-Verlag, 2006.
- [22] P. Ravikumar, “Approximate inference, structure learning and feature estimation in Markov random fields: Thesis abstract,” *SIGKDD Explor. Newsl.*, vol. 10, no. 2, pp. 32–33, Dec. 2008.
- [23] D. M. Greig, B. T. Porteous, and A. H. Seheult, “Exact maximum a posteriori estimation for binary images,” *J. R. Stat. Soc. Ser. B (Methodological)*, vol. 51, no. 2, pp. 271–279, 1989.
- [24] J. S. Yedidia, W. T. Freeman, and Y. Weiss, “Constructing free energy approximations and generalized belief propagation algorithms,” *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2282–2312, Jul. 2005.
- [25] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [26] W. Chen, M. L. Giger, L. Lan, and U. Bick, “Computerized interpretation of breast MRI: Investigation of enhancement-variance dynamics,” *Med. Phys.*, vol. 31, no. 5, pp. 1076–1082, May 2004.
- [27] X. Ren and J. Malik, “Learning a classification model for segmentation,” in *ICCV’03: Proc. 9th IEEE Int. Conf. Comput. Vis.*, Washington, DC, 2003, pp. 10–17.
- [28] G. Mori, X. Ren, A. A. Efros, and J. Malik, “Recovering human body configurations: Combining segmentation and recognition,” in *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2004, vol. 2, pp. 326–333.
- [29] G. Mori, “Guiding model search using segmentation,” in *IEEE Int. Conf. Comput. Vis.*, 2005, vol. 2, pp. 1417–1423.
- [30] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [31] M. Rosenblatt, “Remarks on some nonparametric estimates of a density function,” *Ann. Mathe. Stat.*, vol. 27, pp. 832–837, 1956.
- [32] E. Parzen, “On estimation of a probability density function and mode,” *Ann. Math. Stat.*, vol. 33, pp. 1065–1076, 1962.
- [33] V. A. Epanechnikov, “Non-parametric estimation of a multivariate probability density,” *Theory Probabil. Appl.*, vol. 14, pp. 153–158, 1969.
- [34] Z. I. Botev, J. F. Grotowski, and D. P. Kroese, “Kernel density estimation via diffusion,” *Ann. Stat.*, vol. 38, no. 5, pp. 2916–2957, 2010.
- [35] G. J. McLachlan and K. E. Basford, *Mixture Models*. New York: Marcel Dekker, 1988.
- [36] A. Dempster, N. Laird, and D. Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *J. R. Stat. Soc.*, vol. 39, no. 1, pp. 1–38, 1977.
- [37] S. Paik, G. Tang, S. Shak, C. Kim, J. Baker, W. Kim, M. Cronin, F. L. Baehner, D. Watson, J. Bryant, J. P. Constantino, C. E. Geyer, D. L. Wickerham, and N. Wolmark, “Gene expression and benefit of chemotherapy in women with node-negative, estrogen receptor-positive breast cancer,” *J. Clin. Oncol.*, vol. 24, no. 23, pp. 3726–3734, 2006.
- [38] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, “User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability,” *Neuroimage*, vol. 31, no. 3, pp. 1116–1128, July 2006.
- [39] T. Cour and K. Shi, Normalized Cuts Matlab Code [Online]. Available: <http://www.cis.upenn.edu/~jshi/software/>
- [40] J. C. Bezdek and S. K. Pal, *Fuzzy Models for Pattern Recognition*. New York: IEEE Press, 1992.
- [41] S. Paik, “Development and clinical utility of a 21-gene recurrence score prognostic assay in patients with early breast cancer treated with tamoxifen,” *Oncologist*, vol. 12, pp. 631–635, 2007.
- [42] S. Paik, S. Shak, G. Tang, C. Kim, J. Baker, M. Cronin, F. L. Baehner, M. G. Walker, D. Watson, T. Park, W. Hiller, E. R. Fisher, D. L. Wickerham, J. Bryant, and N. Wolmark, “A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer,” *N. Eng. J. Med.*, vol. 351, pp. 2817–2826, 2004.
- [43] B. K. Szabó, P. Aspelin, M. K. Wiberg, T. Tot, and B. Boné, “Invasive breast cancer: Correlation of dynamic MR features with prognostic factors,” *Eur. Radiol.*, vol. 13, pp. 2425–2435, 2003.
- [44] R. Matsubayashi, Y. Matsuo, G. Edakuni, T. Satoh, O. Tokunaga, and S. Kudo, “Breast masses with peripheral rim enhancement on dynamic contrast-enhanced MR images: Correlation of MR findings with histologic features and expression of growth factors,” *Radiology*, vol. 217, pp. 841–848, 2000.
- [45] W. Chen, M. L. Giger, U. Bick, and G. M. Newstead, “Automatic identification and classification of characteristic kinetic curves of breast lesions on DCE-MRI,” *Med. Phys.*, vol. 33, no. 8, pp. 2878–87, Aug. 2006.
- [46] W. I. Ai and P. Langley, “Induction of one-level decision trees,” in *Proc. 9th Int. Conf. Mach. Learn.*, 1992, pp. 233–240.
- [47] H. Deng and D. Clausi, “Unsupervised image segmentation using a simple MRF model with a new implementation scheme,” in *Proc. 17th Int. Conf. Pattern Recognit.*, Aug. 2004, vol. 2, pp. 691–694.
- [48] R. Kozick and B. Sadler, “Maximum-likelihood array processing in non-Gaussian noise with Gaussian mixtures,” *IEEE Trans. Signal Process.*, vol. 48, no. 12, pp. 3520–3535, Dec. 2000.
- [49] S. Mook, M. K. Schmidt, E. J. Rutgers, A. O. van de Velde, O. Visser, S. M. Rutgers, N. Armstrong, L. J. van’t Veer, and P. M. Ravdin, “Calibration and discriminatory accuracy of prognosis calculation for breast cancer with the online adjuvant! program: A hospital-based retrospective cohort study,” *Lancet Oncol.*, vol. 10, pp. 1070–1076, 2009.
- [50] O. Ginsburg, P. Ghadirian, J. Lubinski, C. Cybulski, H. Lynch, S. Neuhausen, C. Kim-Sing, M. Robson, S. Domchek, C. Isaacs, J. Klijn, S. Armel, W. D. Foulkes, N. Tung, P. Moller, P. Sun, and S. A. Narod, “Smoking and the risk of breast cancer in BRCA1 and BRCA2 carriers: An update,” *Breast Cancer Res. Treat.*, vol. 114, pp. 127–135, 2009.