Early Prediction of Breast Cancer Therapy Response to Neoadjuvant Chemotherapy through Texture Analysis of DCE-MRI

Guillaume Thibault, Alina Tudorica, Aneela Afzal, Stephen Y-C Chui, Arpana Naik, Megan L Troxell, Kathleen A Kemmer, Karen Y Oh, Nicole Roy, Megan L Holtorf, Wei Huang, and Xubo Song

Oregon Heath & Science University (OHSU)

Abstract. This paper presents the initial demonstration that changes in 3D texture features of dynamic contrast enhanced magnetic resonance imaging (DCE-MRI) based parametric maps in the early phase of neoadjuvant (preoperative) chemotherapy (NAC) can be effective predictors of eventual breast cancer residual cancer burden (RCB) score, and thus, pathologic response to NAC. The use of powerful 3D statistical matrices produced high RCB score correlations and allowed automatic determination of parametric maps whose feature changes are most likely to be useful for predicting tumor therapy response.

1 Introduction

Approximately $290,000$ new breast cancer cases will be diagnosed in USA alone in 2015 [1]. An estimated one in eight women (12%) will be diagnosed with breast cancer during their lifetime, making breast cancer the second most common cancer among women, and the second leading cause of cancer-related death. There are evidences indicating that the earlier the detection, treatment and follow up, the better the disease-free survival (DFS) rate [2]. Recently it has been reported [3] reported that MRI can detect breast cancer that is missed by other techniques (X-ray mammography, clinical examination, ultra-sounds, etc.). Consequently, the American Cancer Society published guidelines for breast cancer screening in high risk population with MRI as an adjunct, because DCE-MRI has been shown to have 85% to 100% sensitivity and 35% to 85% specificity for tumor detection [4]. DCE-MRI detects angiogenesis and provides high resolution 4D imaging of the entire breast, allowing to: (a) show variations in contrast that provide information about the tissue vascularity (poorly formed high density of blood vessels are sign of malignant tumors [3]), (b) visualize the 3D inner tumor texture (increased heterogeneity in malignant tumors), and (c) distinguish between tumor types. These features make the DCE-MRI an in vivo virtual biopsy, ideal for detecting, visualizing and characterizing the tumors and its changes during different treatments, and finally to assess the tumor response to treatment [5, 6].

For patients with locally advanced breast cancer (LABC), NAC has been established as a standard of care. In patients with inoperable LABC, NAC has been shown to improve both DFS and overall survival [7]. Patients with operable tumors may also choose to receive NAC. Together with radiation therapy, NAC facilitates breast conservation surgery and leads to overall survival and DFS rates similar to those obtained by mastectomy [8, 9]. Unfortunately, patients undergoing NAC do not always respond. Early identification (after the first NAC cycle) of non-responders (NRs) can make it possible to spare them from ineffective and/or toxic treatments, allow quick adoption of alternative treatment regimen, improve clinical management, and better patient stratification for trials of novel therapies.

In order to assess and predict tumor response to NAC using noninvasive DCE-MRI method, automatic methods have been developed, which typically consist of two modules: i) extraction of quantitative features (surrogate imaging biomarkers) from various DCE-MRI parametric maps (detailed in section 2), and ii) a classification by combining the extracted features to build a prediction model. Early DCE-MRI studies focused on changes in tumor size tumor size, volume, and morphology (longest diameters) to evaluate treatment response. Unfortunately, for most patients the tumor size shrinks after at least 2 NAC cycles [10], making it unsuitable for early prediction. More recently, studies have focused on measuring and tracking changes in pharmacokinetics features[1] captured through quantitative pharmacokinetic modeling of DCE-MRI time-course data [11]. However, not all studies have shown that quantitative DCE-MRI provides results superior to those obtained by the simple volume changes [6]. Moreover, measuring changes in pharmacokinetic parameters cannot capture the spatial tissue heterogeneity, and therefore discards a significant amount of valuable information. Recently, Teruel et al. [12] have presented a detailed analysis of 16 textural statistical features capable of predicting early tumor response to NAC, and pointed out three promising features. Unfortunately, this study uses a 2D statistical texture description and then does not take advantage of all the 3D information provided by the 3D nature of breast DCE-MRI data acquisition.

In this paper, we investigate the predictive value of 3D statistical texture features for early prediction of breast cancer response to NAC using DCE-MRI pharmacokinetic parametric map (Section 2). These statistical features have been widely used for texture analysis. We extract these texture features in 3D in order to take advantage of all the available information. With these features, we seek to predict the residual cancer burden (RCB) score measured pathologically from the surgical specimens [15]. The RCB score provides the extent of residual disease after NAC completion. For example, a RCB score of 0 means a patient achieving pathologic complete response (pCR), and the higher the RCB value, the more severe the residual disease. The following sections present DCE-MRI data collection and analysis (Section 2), a brief description of the extracted texture features (Section 3), and then the results (Section 4).

---

[1] Dynamic variation of textural features of tumors at each time point during contrast material uptake, and then plotted as a time function

2   Breast DCE-MRI Data Collection and Analysis

Twenty eight women with LABC who underwent $6 - 8$ NAC cycles as standard of care consented to research DCE-MRI studies performed at Visit 1 (V_1) - before NAC, V_2 - after first NAC cycle, V_3 at NAC midpoint, and V_4 - after NAC completion but before surgery. Axial bilateral DCE-MRI images with fat-saturation and full breast coverage were acquired with a 3D gradient echo-based TWIST sequence using a 3T Siemens scanner. DCE-MRI acquisition parameters included $10°$ flip angle, 2.9/6.2 ms TE/TR, a parallel imaging acceleration factor of two, 30-34 cm FOV, 320×320 in-plane matrix size, and 1.4 mm slice thickness. The total acquisition time was 1̃0 minutes for 32-34 image volume sets with 18-20 seconds temporal resolution. Gadolinium contrast agent (Prohance) IV injection (0.1 mmol/kg at 2 mL/s) was timed to start following acquisitions of two baseline image volumes. Tumor region of interest (ROIs) were drawn by experienced radiologists using post contrast DCE images. The pixel-by-pixel (within the ROI) DCE time-course data were subjected to both the SM (Standard Tofts Model [13]) and SSM (Shutter-Speed Model [14]) pharmacokinetic analyses to obtain tumor parametric maps (see figure 1) of K^{trans} (contrast agent transfer rate constant), v_e (extravascular and extracellular volume fraction), k_{ep} ($= K^{trans}/v_e$, contrast agent intravasation rate constant), τ_i (mean intracellular water lifetime, SSM only), and $\Delta K^{trans} = K^{trans}(SSM) - K^{trans}(SM)$. ΔK^{trans} is a measure of water exchange effects on K^{trans} quantification as the SSM takes into account while the SM ignores the effects of intercompartmental water exchange kinetics. For each patient pathologic response to NAC and RCB score were determined by pathological analysis of post-therapy resection specimens and comparison with pre-therapy core biopsy specimens using previously published methods [15]. A total of 5 patients were pCRs, while the others were non-pCRs with a RCB range of $0.433 - 3.707$.

3   Characterization: 3D Statistical Matrices

Let $f : \begin{cases} E \to \mathcal{T} \\ \mathbf{x} \mapsto f(\mathbf{x}) \end{cases}$ be a gray-levels image of dimension $w \times h$, where $E \subset \mathbf{Z}^2$ is the pixels support space and the image intensities are discrete values which range in a closed set $\mathcal{T} = \{t_1, t_2, ..., t_N\}$, $\Delta t = t_{i+1} - t_i$, e.g., for a 8 bits image $t_1 = 1$, $N = 256$ and $\Delta t = 1$. Assume that the image f is segmented into its J flat zones $R_j[f]$ (i.e., connected regions of constant value): $E = \cup_{j=1}^{J} R_j[f]$, $\cap_{j=1}^{J} R_j[f] = \emptyset$. Each region size (surface area) is $s(j) = |R_j[f]|$ (|.| is the cardinal). Hence, we consider that each zone $R_j[f]$ has an associated constant gray-level intensity.

Statistical matrices have been extensively used in texture characterization. The best-known of which is the gray level Co-Occurrence Matrix (GLCM), often used with Haralick's features [16]. The GLCM represents the texture by the second order statistics: co-occurring values distribution at a given offset. In 2D
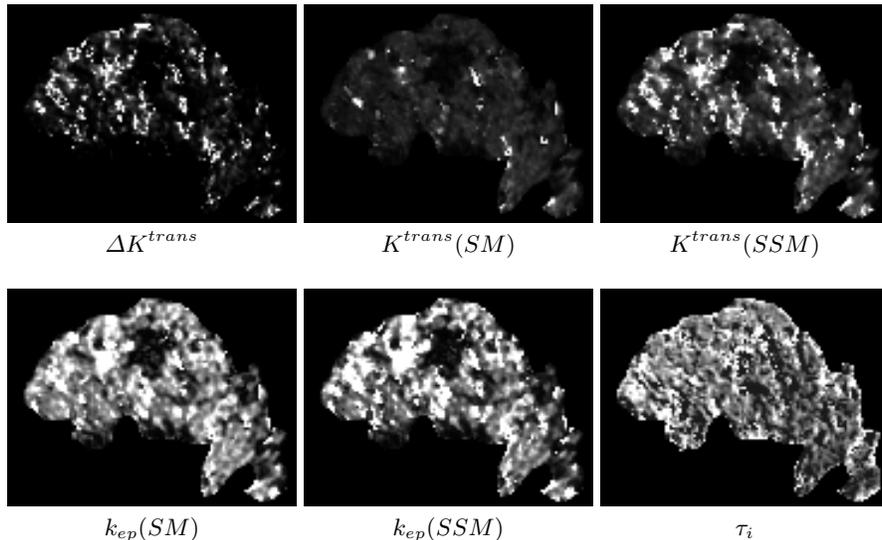
<div align="center">

ΔK^{trans} $K^{trans}(SM)$ $K^{trans}(SSM)$

$k_{ep}(SM)$ $k_{ep}(SSM)$ τ_i

</div>

Fig. 1. Examples of parametric maps.

and for an offset $\Delta = (\Delta_x, \Delta_y)$, the GLCM is defined as:

$$GLCM_{f,\Delta}(i,j) = \sum_{x=1}^{w} \sum_{y=1}^{h} \begin{cases} 1, \ if \ f(x,y) = i \ and \ f(x + \Delta_x, y + \Delta_y) = j \\ 0, \ otherwise \end{cases}$$

By design, the GLCM is dependent on the offset and therefore is not rotation invariant (see figure 2 a). This is addressed by computing GLCM in four different directions: $\theta_{0°} = (0,1)$, $\theta_{45°} = (1,1)$, $\theta_{90°} = (1,0)$ and $\theta_{135°} = (-1,1)$. The amount of information extracted depends on the number of offset directions and distances. Typically, a large number of offsets are needed to extract all the useful information and this is the main drawback of this approach.

Another classical technique is the gray level Run Length Matrix (RLM) [17], which has been extensively developed for texture classification. The RLM extracts higher order statistical features: the matrix element $RLM_{f,\theta}(g,l)$ gives the gray level g and length l runs (i.e., collinear pixels with the same intensity in the direction θ) total number (see figure 2 c). This method is particularly effective for periodic textures and completes the information provided by the GLCM. Extracted features from the RLM are moments of order from -2 to 2. As GLCM, RLM requires computation in many directions in order to achieve the rotation invariance.

Recently Thibault et al. [18, 19] introduced the gray level Size Zone Matrix (SZM) original notion, as an alternative to the joint RLM distribution. The SZM is based on each flat zone size/intensity co-occurrences, so provides a statistical representation by the bivariate conditional probability density function estimation of the image distribution values. In this method, the matrix value

$SZM_f(s, g)$ is equal to the size zones s and gray level g total number in f (see figure 2 d). In this matrix, the more homogeneous the texture (large flat zones with closed gray levels), the wider and flatter the matrix. From this statistical matrix representation, we can calculate all the second-order moments as compact texture features [17], plus two features which are specific weighted variances [18]. This matrix is not dependent on a given parameter (contrary to both GLCM and RLM which are respectively dependent of the offset Δ and the orientation θ), so does not require computation in several directions, and thus is rotation and translation invariant. However it requires a flat zone labeling which is time consuming. The connectivity type used for labeling modifies the matrix but does not impact the classification performances [19].

Remark RLM and GLCM are appropriate for periodic textures whereas the SZM is typically adapted to describe heterogeneous non periodic textures. In addition, due to the intrinsic segmentation, texture description in SZM is more regional than the point-wise-based GLCM representation. But by design all these matrices are sensitive to noise. Indeed, a pixel gray level variation of Δv, involves: a different increment located at a distance Δv in the GLCM; potentially a shorter run (because cut), plus a singleton (run of length 1) and a new run (cut run remaining part), in the RLM; a singleton (flat zone of size 1) and a zone reduced of 1 for the SZM. In order to improve statistical matrices noise robustness, the texture gray levels number is reduced before matrix filling. Different methods exist to reduce the gray levels number to N possible values, and the most commonly used approaches are:

– the application of a linear or nonlinear function. For this method, a histogram spreading is first performed, and then a function is applied. Most of the time the function is linear (so a simple division is performed), but a *log* or other functions can be used.
– the generation of a cumulated histogram in order to separate the pixels distribution into N bins containing approximately the same pixels number.
– the application of a clustering algorithm with N clusters.

The classification performances can be very sensitive to the algorithm used, so it is generally recommended to test some or all of them, with different gray level quantizations.

4   Method and Results

In order to evaluate the feasibility of texture analysis for early prediction of tumor response to NAC, we focus on the differences between V_1 (before NAC) and V_2 (after the first NAC cycle). Consequently at V_1 and V_2, each parametric map is characterized with the statistical features described above, computed using two gray level reduction algorithms (linear and histogram) and five gray level quantization (dyadic values from 8 to 256), which generates hundreds of features. Due to the relatively small sample size (a cohort of 28 tumors), using

1	2	3	4
1	3	4	4
3	2	2	2
4	1	4	1

	Cooccurrences			
$g\backslash g$	1	2	3	4
1	0	1	1	2
2	1	4	2	0
3	1	2	0	2
4	2	0	2	2

	Run length		
$g\backslash l$	1	2	3
1	4	0	0
2	1	0	1
3	3	0	0
4	3	1	0

	Size zone		
$g\backslash s$	1	2	3
1	2	1	0
2	1	0	1
3	0	0	1
4	2	0	1

(a)   (b)   (c)   (d)

Fig. 2. Matrices filling example for a 4×4 texture with 4 gray levels (a), GLCM with $\Delta = (1,0)$ (b), RLM with $\theta = 0°$ (c), and SZM (d).

all the features simultaneously to train a classifier will likely lead to overfitting. Consequently, we only use three features at a time, selected from the same features pool and parametric map (see algorithm 1). Then we generate a prediction model with linear regression (LR) and leave-one-out protocol (LOO) using the triple features. Finally each model is validated using three different correlations measures: Pearson (linear), Spearman (rank) and Kendall (rank).

Data: Let $F^I_{V_i,P_k}$ the Ith feature extracted at V_i on the P_k map.
foreach *parametric map P_k* **do**
    foreach *set of features S extracted from P_k* **do**
        foreach *triplet of features $(F^{I_1}_{V_i,P_k,}, F^{I_2}_{V_i,P_k}, F^{I_3}_{V_i,P_k}) \in S$* **do**
            $F_1 \leftarrow F^{I_1}_{V_1,P_k} - F^{I_1}_{V_2,P_k}$;
            $F_2 \leftarrow F^{I_2}_{V_1,P_k} - F^{I_2}_{V_2,P_k}$;
            $F_3 \leftarrow F^{I_3}_{V_1,P_k} - F^{I_3}_{V_2,P_k}$;
            RCBpred $\leftarrow LR(F_1, F_2, F_3) + LOO$;
            Correlations computation between RCBpred and RCB ;

Algorithm 1: Prediction models elaboration.

By applying the algorithm 1, we have generated and tested approximately $140,000$ prediction models. Among them, 779 models have strong correlations, i.e. all three correlation measures are greater than 0.7. The Figure $3abc$ shows the correlation scores distribution among all good models. We can observe that the median/average correlations for Pearson, Spearman and Kendall are approximately 0.8, 0.89 and 0.72 respectively. These high values demonstrate the models efficacy to predict the RCB score after only one NAC cycle, using only three statistical texture features at a time. The first distribution (Pearson) shows that all the models predictions are linearly correlated with the RCB scores, which is difficult to achieve using a simple linear regression. These results are confirmed by the Spearman correlations distribution, which are around 0.9, and therefore demonstrates the models predictions consistency with the RCB scores. Figure $3d$ lists for each of the three texture analysis methods the parameter map with the highest correlation coefficients.

Half of the models were built using the GLCM, but we can observe on figure

$3d$ that each statistical matrix can generate at least one highly efficient models with all correlations around 0.9. An other interesting result is the distribution of the parametric map used to generates the prediction models. We can observe on figure $3e$ that 80% of the models were generated using the SSM parametric maps.
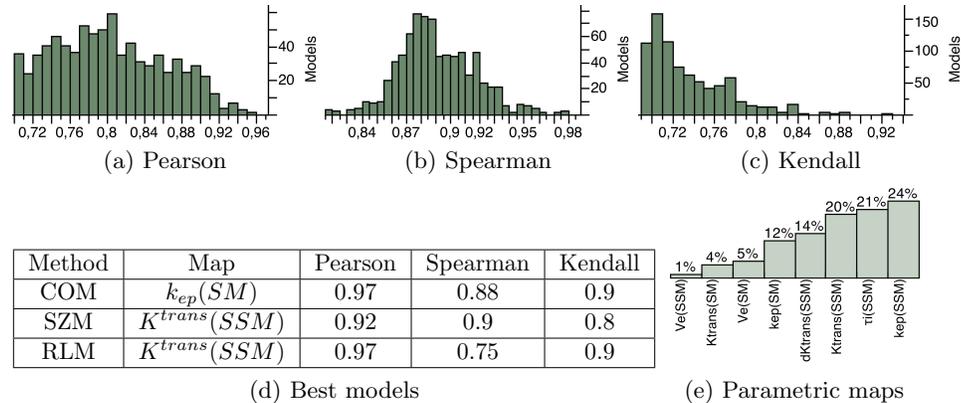

(a) Pearson (b) Spearman (c) Kendall

Method	Map	Pearson	Spearman	Kendall
COM	$k_{ep}(SM)$	0.97	0.88	0.9
SZM	$K^{trans}(SSM)$	0.92	0.9	0.8
RLM	$K^{trans}(SSM)$	0.97	0.75	0.9

(d) Best models

(e) Parametric maps

Fig. 3. The correlation scores (a-c) and parametric maps (e) distribution among all good models, and the best correlation scores for each statistical matrix (d).

5 Conclusions and Perspectives

This paper presents the initial result in using 3D characterization of DCE-MRI heterogeneity imaging biomarker of LABC tumors for early prediction of breast tumor response to NAC. The texture features were used with a linear regression in order to build prediction models. A total of 779 models have high correlations and can predict the RCB score. These results are promising and suggest that it is possible to predict the eventual RCB score using quantitative imaging features at baseline and after the first NAC cycle. Moreover, the results indicate that the SSM parametric maps seem to be more sensitive to therapy induced tumoral biological changes and thus provide better prediction of therapy response than the SM counterparts. Therefore, the SSM may be the pharmacokinetic model of choice for quantitative analysis of DCE-MRI data.

Our initial results need to be validated with a larger dataset in order to study response of various breast cancer molecular subtypes with statistical significance, and prevent overfitting. Moreover, we will estimate the predictive ability of each feature in order to fine tune a features selection algorithm and build better models, using non linear regression methods as neural networks, random forests and support vectors machine.

References

1. How many women get breast cancer? American Cancer Society, 2014.
2. Feig et al. American college of radiology guidelines for breast cancer screening. *American Journal of Roentgenology*, 171(1):29–33, 1998.
3. Saslow et al. American cancer society guidelines for breast screening with mri as an adjunct to mammography. *CA: A Cancer Journal for Clinicians*, 57(2):75–89, 2007.
4. Woods et al. Malignant-lesion segmentation using 4d co-occurrence texture analysis applied to dynamic contrast-enhanced magnetic resonance breast image data. *Journal of Magnetic Resonance Imaging*, 25:495–501, 2007.
5. Balu-Maestro et al. Imaging in evaluation of response to neoadjuvant breast cancer treatment benefits of mri. *Breast Cancer Research and Treatment*, 72(2):145–152, 2002.
6. Yu et al. Mri measurements of tumor size and pharmacokinetic parameters as early predictors of response in breast cancer patients undergoing neoadjuvant anthracycline chemotherapy. *Journal of Magnetic Resonance Imaging*, 26:615–623, 2007.
7. Jafri et al. Optimized breast mri functional tumor volume as a biomarker of recurrence-free survival following neoadjuvant chemotherapy. *Journal of Magnetic Resonance Imaging*, 40:476–482, 2014.
8. Khokher et al. "initial clinical response" to neoadjuvant chemotherapy: an in-vivo chemosensitivity test for efficacy in patients with advanced breast cancer. *Asian Pacific Journal of Cancer Prevention*, 12(4):939–946, 2011.
9. Dongfeng et al. Dynamic breast magnetic resonance imaging: Pretreatment prediction of tumor response to neoadjuvant chemotherapy. *Clinical Breast Cancer*, 12(2):94–101, 2012.
10. Martincich e al. Monitoring response to primary chemotherapy in breast cancer using dynamic contrast-enhanced magnetic resonance imaging. *Breast Cancer Research and Treatment*, 83:67–76, 2004.
11. Johansen et al. Predicting survival and early clinical response to primary chemotherapy for patients with locally advanced breast cancer using dce-mri. *Journal of Magnetic Resonance Imaging*, 29(6):1300–1307, 2009.
12. Teruel et al. *NMR in Biomedicine*, 27:887–896, 2014.
13. Tofts et al. Estimating kinetic parameters from dynamic contrast-enhanced t(1)-weighted mri of a diffusable tracer: standardized quantities and symbols. *Journal of Magnetic Resonance Imaging*, 10(3):223–232, 1999.
14. Yankeelov et al. Variation of the relaxographic "shutter-speed" for transcytolemmal water exchange affects the cr bolus-tracking curve shape. *Magnetic Resonance in Medicine*, 50(6):1151–1169, 2003.
15. Symmans et al. *Journal of Clinical Oncology*, 25(28):4414–4422, 2007.
16. Haralick et al. Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, 3(6):610–621, 1973.
17. Chu et al. Use of gray value distribution of run lengths for texture analysis. *Pattern Recognition Letters*, 11(6):415–419, 1990.
18. Thibault et al. Shape and texture indexes: Application to cell nuclei classification. *International Journal of Pattern Recognition and Artificial Intelligence (IJPRAI)*, 27(1), 2013.
19. Thibault et al. Advanced statistical matrices for texture characterization: Application to cell classification. *IEEE Transactions on Biomedical Engineering*, 61(3):630–637, 2014.