Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Assessing treatment response in triple-negative breast cancer from quantitative image analysis in perfusion magnetic resonance imaging

Imon Banerjee
Sadhika Malladi
Daniela Lee
Adrien Depeursinge
Melinda Telli
Jafi Lipson
Daniel Golden
Daniel L. Rubin

SPIE.

Imon Banerjee, Sadhika Malladi, Daniela Lee, Adrien Depeursinge, Melinda Telli, Jafi Lipson, Daniel Golden, Daniel L. Rubin, "Assessing treatment response in triple-negative breast cancer from quantitative image analysis in perfusion magnetic resonance imaging," *J. Med. Imag.* **5**(1), 011008 (2017), doi: 10.1117/1.JMI.5.1.011008.

Journal of Medical Imaging 5(1), 011008 (Jan–Mar 2018)

Assessing treatment response in triple-negative breast cancer from quantitative image analysis in perfusion magnetic resonance imaging

Imon Banerjee,[a,*] Sadhika Malladi,[b] Daniela Lee,[c] Adrien Depeursinge,[d] Melinda Telli,[e] Jafi Lipson,[a] Daniel Golden,[f] and Daniel L. Rubin[a]
[a]Stanford University, Department of Radiology, Stanford, California, United States
[b]Massachusetts Institute of Technology, Department of Mathematics, Cambridge, Massachusetts, United States
[c]Yale University, Department of Ecology and Evolutionary Biology, New Haven, Connecticut, United States
[d]University of Applied Sciences Western Switzerland (HES-SO), Department Institute of Information Systems, Sierre, Switzerland
[e]Stanford University, Department of Medicine (Oncology), Stanford, California, United States
[f]Arterys Inc., San Francisco, California, United States

Abstract. Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) is sensitive but not specific to determining treatment response in early stage triple-negative breast cancer (TNBC) patients. We propose an efficient computerized technique for assessing treatment response, specifically the residual tumor (RT) status and pathological complete response (pCR), in response to neoadjuvant chemotherapy. The proposed approach is based on Riesz wavelet analysis of pharmacokinetic maps derived from noninvasive DCE-MRI scans, obtained before and after treatment. We compared the performance of Riesz features with the traditional gray level co-occurrence matrices and a comprehensive characterization of the lesion that includes a wide range of quantitative features (e.g., shape and boundary). We investigated a set of predictive models (~96) incorporating distinct combinations of quantitative characterizations and statistical models at different time points of the treatment and some area under the receiver operating characteristic curve (AUC) values we reported are above 0.8. The most efficient models are based on first-order statistics and Riesz wavelets, which predicted RT with an AUC value of 0.85 and pCR with an AUC value of 0.83, improving results reported in a previous study by ~13%. Our findings suggest that Riesz texture analysis of TNBC lesions can be considered a potential framework for optimizing TNBC patient care. [DOI: 10.1117/1.JMI.5.1.011008]

Keywords: triple-negative breast cancer; neoadjuvant chemotherapy; image features; machine learning; quantitative imaging.

Paper 17187SSR received Jun. 23, 2017; accepted for publication Oct. 16, 2017; published online Nov. 2, 2017.

1 Introduction

Between 15% and 20% of all diagnosed breast cancers are a unique subtype called triple-negative breast cancer (TNBC), characterized by the absence of estrogen, progesterone, and HER2/neu receptors.[1] Compared with other breast cancer subtypes, TNBC is less likely to be detected by an annual mammogram and has a higher risk of distant relapse.[2,3] Unlike patients with hormone receptor or HER2/neu-positive breast tumors, TNBC patients have limited targeted treatment options. Combination chemotherapy in the early stage setting is the standard of care for TNBC. This therapy is often administered prior to surgery and is referred to as neoadjuvant chemotherapy (NAC).[4] Achieving a pathologic complete response (pCR) to NAC is associated with excellent long-term outcomes.[4–6] NAC combinations targeting DNA repair defects in TNBC have recently been investigated.[7,8] Finding biomarkers that can predict therapeutic response to these and other breast cancer treatments is crucial for identifying patients who will benefit.

Radiological imaging plays a crucial role in the diagnosis of cancer patients. Dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI), in particular, has shown promise in diagnosing and characterizing breast cancer as it provides quantitative information about the tumor and its microenvironment.[9–11] This imaging procedure involves the intravenous injection of gadolinium-based contrast agent to highlight the vascularity of the tumor.[12] DCE-MRI does not use ionizing radiation, allowing longitudinal imaging to capture change in tumor vasculature during treatment, without being limited by patient dose considerations.[13]

Currently, biological variability in tumors is recognized to be important in differentiating treatment outcomes.[14] However, this variability is difficult to measure using random tumor sampling or biopsies because the entire tumor must be considered to comprehensively measure biological variations.[15] Therefore, the noninvasive assessment of treatment response using texture analysis may provide clinical benefit. For example, MRI texture analysis has shown potential in predicting treatment responses for rectal cancer, allowing the potential development of personalized treatment.[16]

DCE-MRI produces pharmacokinetic maps that measure the kinetic properties of the contrast agent in the tumor. These kinetic parameters of the lesion have been shown to correlate with NAC response in breast cancer, especially when used in conjunction with quantitative measurements of tumor heterogeneity, which was shown to be associated with breast lesion

*Address all correspondence to: Imon Banerjee, E-mail: imonb@stanford.edu

2329-4302/2017/$25.00 © 2017 SPIE

malignancy and disease progression.[17–19] Heterogeneity reflects the biological variability of the tumor tissue by characterizing necrosis, angiogenesis, and various cell growth patterns.

The pattern of lesion heterogeneity can be determined by analyzing the texture that characterizes the local spatial organization of pixels inside the tumor region.[20] While tumor heterogeneity itself is generally associated with a worse prognosis, heterogeneity measures can gauge inter- or intratumoral texture to improve the performance.[19] Over the last decade, most commonly applied techniques for analyzing texture have been statistics-based. For example, gray level co-occurrence matrices (GLCM) and runlength matrices have been widely used second-order statistical method for texture analysis.[21] In particular, the GLCM features have been shown to correlate closely with many aspects of breast cancer, such as lesion malignancy and the presence of heterogeneous enhancement.[22] Moreover, there exist other prior studies[23,24] that exploit traditional GLCM features to predict the response of breast cancer patients to the chemotherapy. Previous research has provided promising evidence that GLCM features extracted from pharmacokinetic maps can also be used to develop a model that identifies treatment responders and nonresponders in TNBC patients.[25] However, the performance of the reported GLCM framework depends on the arbitrary choice of scales and orientations as well as a necessary gray level reduction, which entail the risk of losing precious information contained in the full bit depth of the original image.

We proposed characterizing the TNBC lesion using a multiscale, translation, and rotation-covariant texture analysis framework based on the Riesz wavelets,[26] which showed promising results for interpreting high-resolution computed tomography scan images of interstitial lung diseases and liver lesions.[27] Its balanced performance among various lung and liver texture features demonstrates an adaptability that suggested its potential suitability for breast lesion texture analysis in TNBC patients. Additionally, to complement the texture features, we sought to incorporate first-order statistics, which deal with the gray-level frequency distribution within the region of interest (ROI) and are dependent on single pixel intensities rather than comparisons with neighboring pixels.[21] Our ultimate goal is to create a statistically valid framework that can predict the response of patients to NAC in newly diagnosed early stage TNBC.

2 Method

2.1 *Dataset*

2.1.1 *Patient population*

PrECOG 0105 was a multicenter phase II trial that examined the efficacy of gemcitabine, carboplatin, and iniparib as neoadjuvant therapy for patients with triple-negative and BRCA1/2 mutation-associated breast cancer.[28] Patients had stage I-IIIA newly diagnosed breast cancer and had received no prior therapy. The investigational therapy was given for four to six cycles prior to surgery, at which point pathologic response was assessed. Eighty patients received six cycles of therapy and thirteen received four cycles. Breast MRI was performed pre- and posttreatment per protocol. The cohort of patients used in this study is identical to that from a previous study.[25] The patients had provided written informed consent for their data to be used in this research, and our study was approved by our institutional review board. These patients had early stages of TNBC (Table 1). Seventy-five patients who had completed the PrECOG 0105 study were identified as being potentially eligible for the analysis. However, 15 were excluded because they had BRCA mutation-associated non-TNBC and had more than one dominant lesion (i.e., multifocal disease). Seven patients of the remaining 60 were later excluded due to poor

Table 1 Patient cohort information. The cohort used in this study was also used in a previous study.[25]

	Patients	Percentage
AJCC clinical stages		
IA	5	12.2
IIA	15	36.9
IIB	11	26.8
IIIA	10	24.4
AJCC grade		
I	0	0
II	6	14.6
III	35	85.4
Estrogen receptor status (> 5%)		
Positive	0	0
Negative	41	100
Progesterone receptor status (> 5%)		
Positive	0	0
Negative	41	100
HER2 status		
Positive	0	0
Negative	41	100
Germline BRCA1 status		
Positive	7	17.1
Negative	34	82.9
Germline BRCA2 status		
Positive	1	2.4
Negative	39	95.1
Variant of unknown significance	1	2.4
No. of treatment cycles at postchemotherapy point		
Four cycles	8	19.5
Six cycles	33	80.5
Magnetic field strength		
1.5 T	10	24.3
3 T	31	75.6

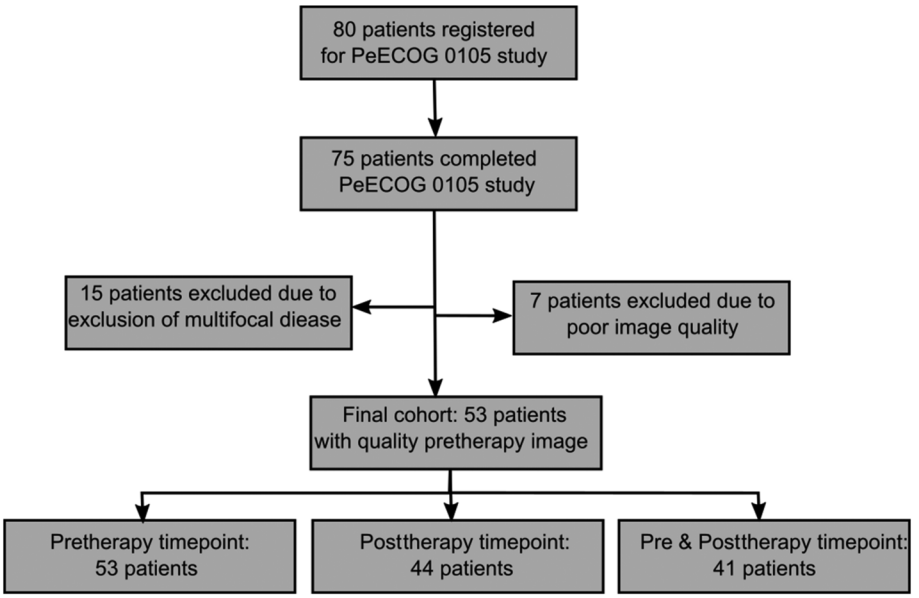


Fig. 1 Flowchart of the cohort selection.

quality imaging. Thus, when generating the predictive models, the 53 patients who had quality prechemotherapy images were used for models based on the prechemotherapy image only, 44 for models based on the postchemotherapy image only, and 41 patients for models including both time points. Figure 1 summarizes the cohort selection process for the predictive modeling as a flowchat diagram.

Patients were imaged at over 25 different imaging centers, each of which conformed to the standard-of-care protocols specific to the location. As such, the scans taken were with two field strengths, either 1.5 or 3 T. Typical slice thickness is between 3 to 5 mm, and the pixel size (field of view/matrix) is less than 1×1 mm. A dynamic T1 sequence was used. The DCE-MRI imaging protocol also varied between centers. All centers used one of the following gadolinium-based contrast agents: Magnevist, Multihance, or Omniscan. The agents were injected at a concentration of either 0.2 mL per kilogram or 20 mL for all patients. Certain centers injected the agent at a rate of 2 mL per second, while others did so at a rate of 1.5 mL per second. Postchemotherapy scans were taken after either four or six cycles of treatment, depending on the center at which the patient was treated. The time between pre- and postchemotherapy scans varied across the imaging centers, from ~3 months to 1 year. The maximum interval between pre- and posttherapy was <12 months. The scans were all done at baseline and just after the completion of NAC (12 to 18 weeks duration) prior to surgery.

2.1.2 *Kinetic maps*

A board certified radiologist circumscribed the breast lesion on a single axial slice with the largest diameter of the lesion. All analyses were performed on the two-dimensional (2-D) ROI. The lesions were delineated in both pre- and postchemotherapy images using a web-based, freely available quantitative imaging informatics platform—epad.[29] This ROI was used to select a subset of the pixel values in the six DCE-MRI kinetic maps used in this analysis. DCE-MRI kinetic maps include both empirical and pharmacokinetic parameters. The pharmacokinetic parameters consisted of the k_{ep}, K^{trans}, and v_e, which was derived from k_{ep} and K^{trans} values through $v_e = K^{trans}/k_{ep}$, shown in Fig. 2. The K^{trans} measures the volume transfer constant of the diffusion of contrast agent from the blood plasma to the extracellular extravascular space (EES), whereas the k_{ep} parameter looks at the rate constant of contrast agent from the EES to the blood plasma. Thus, v_e represents the EES volume per unit of tissue volume. The nonlinear least-squares curve fitting was used to estimate the values for k_{ep} and K^{trans}. Unlike the empirical maps, the pharmacokinetic maps used all available patient time points. The pharmacokinetic modeling is based on the Tofts model.[30]

The empirical parameters consisted of the wash-in and wash-out, which measure the rate of contrast flow into and out of the area of interest, as well as the area under the receiver operating characteristic curve (AUC) of pixel intensity versus

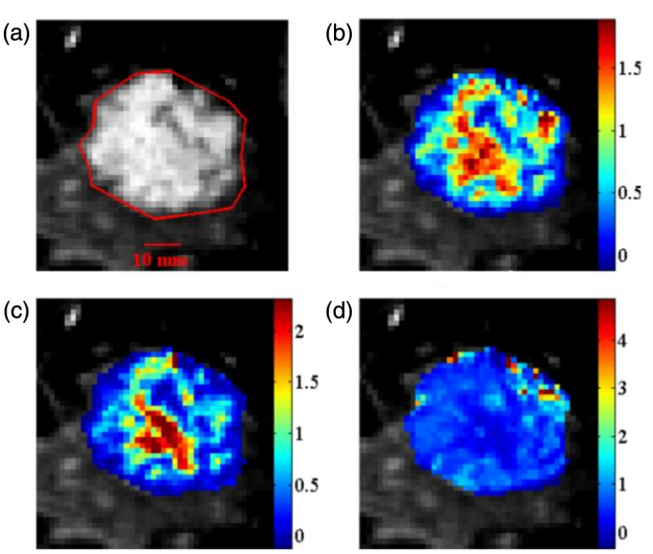


Fig. 2 Pharmacokinetic maps for a patient (id 104): (a) postcontrast MRI of pretherapy lesion ($t = 1.5$ min), (b) $K^{trans}(\text{min}^{-1})$, (c) $k_{ep}(\text{min}^{-1})$, and (d) v_e(unitless). Each image illustrates values (colored pixels) of a particular pharmacokinetic parameter.

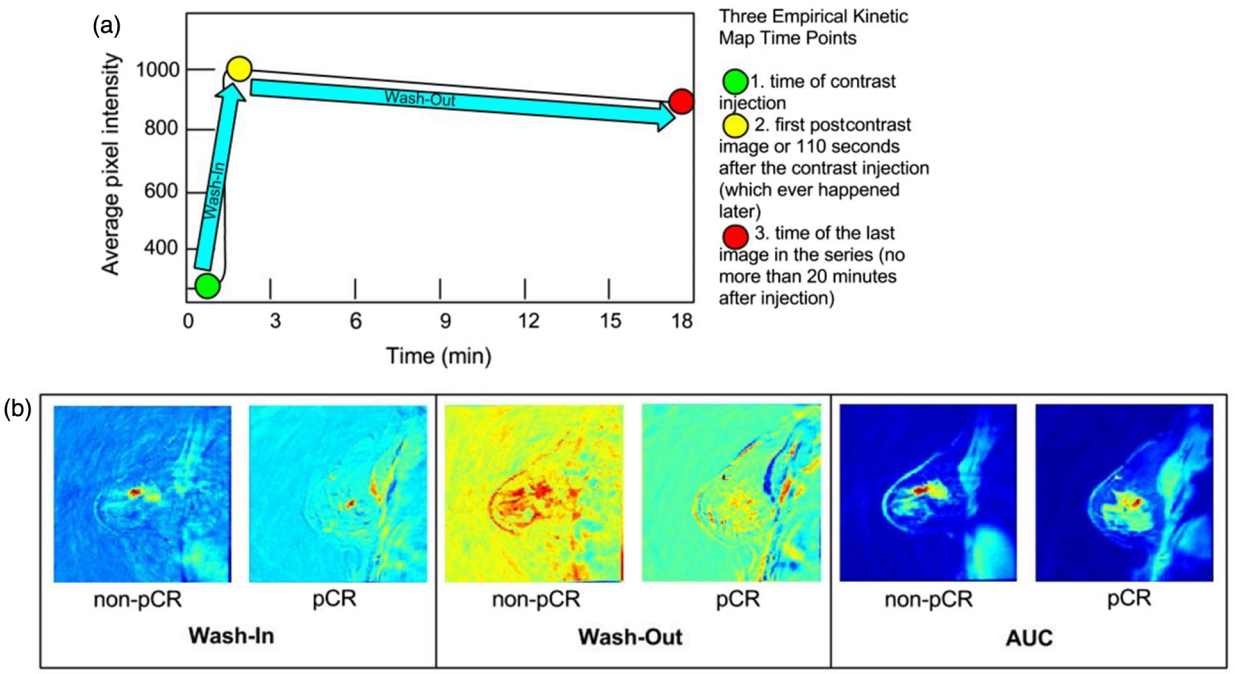


Fig. 3 Visualization of the kinetic maps with empirical parameters. (a) Average pixel intensity versus time showing dynamic sequence of contrast and the three time points used to derive the images. The slope of the wash-in and wash-out segments was used as the feature value. (b) Example kinetic maps of a patient who has not achieved pCR and of a patient who has achieved pCR.

time. The empirical kinetic maps had three time points: the time of contrast injection, the first postcontrast image or 110 s after the contrast injection (whichever happened later), and the time of the last image in the series (no more than 20 min after injection). As shown in Fig. 3, these time points captured the flow of contrast into and out of the lesion. Before the feature extraction, we applied a zero-center normalization technique to have zero mean and unit variance in all the channels of the kinetic maps.

2.2 *Feature Extraction*

2.2.1 *Combination of Riesz and first-order statistical features*

We propose characterizing the lesion texture from six distinct pharmacokinetic maps using the responses of Riesz wavelets.[26] The multiscale decomposition inherited from wavelets allows the separation of fine from coarse texture properties.[31] The Riesz R filterbank of order N is defined as $R(n^1, \widehat{n^2})f(\omega) = \sqrt{\frac{n^1+n^2}{n^1!n^2!}}\frac{(j\omega_1)^{n^1}(j\omega_2)^{n^2}}{\|\omega\|^{n^1+n^2}}\widehat{f(\omega)}$, for all combinations of (n^1, n^2) with $n^1 + n^2 = N$. For a 2-D image, the vector ω is composed of $\omega_{1,2}$ corresponding to the frequencies in the two image axes, and $\widehat{f(\omega)}$ denotes the Fourier transform of the image. For each scale, the Riesz transform decomposes image directions according to the N'th order partial image derivatives. For instance, the second-order Riesz transform corresponds qualitatively to second-order partial derivatives, as $\frac{\partial}{\partial x^2}$, $\frac{\partial}{\partial x \partial y}$, and $\frac{\partial}{\partial y^2}$, shown in Fig. 4.

The corresponding filterbank is steerable, which means that all possible image directions can be encoded from linear combinations of the responses of all components of the filterbank $R(\widehat{n1, n2})$. This property enables rotation covariance of the proposed texture descriptors with infinitesimal angular precision. The second-order transform—a similar function as the Hessian matrix—is known to be the most effective balance between complexity of the filterbank and accuracy.[32] The latter was also shown to best model image profiles of vascular structures.[33] Moreover, even-order transforms detect ridges in the image, where the pixel intensity peaks, which corresponds to the values in the kinetic maps. Therefore, the filters obtained with the second-order Riesz transform can quantify the presence of tubular structures (e.g., vessels) at multiple scales in an image.

In this study, the second-order filterbanks were built at four dyadic scales, creating a total of twelve overall filters. The convolution of the 12 filters with the image resulted in 12 subbands (i.e., Riesz coefficients), which indicate the image's response to the filter (e.g., the presence of tubular structures at a particular scale). With 12 filters, 12 coefficient matrices were generated. We isolated the coefficients from the ROI, where the kinetic map was generated. To combine the numerous Riesz measurements into a single scalar value representing the feature value for a particular filter, we measured the average energy of the

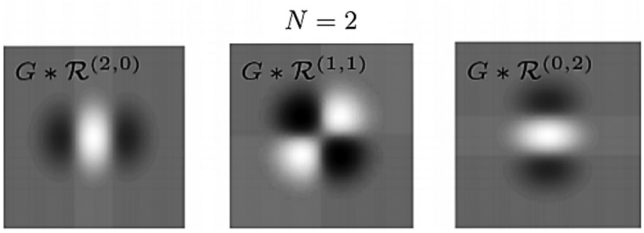


Fig. 4 The second-order Riesz filterbank. Filters in the bank were combined with each other and convolved with the image to compute Riesz features.

coefficients in the ROI, defined by the sum of the squares of the latter. This energy ultimately defines the scalar value of the Riesz feature. In total, we computed 72 Riesz energy values for each lesion image.

As with first-order statistics, we incorporated the average intensity value over each of the six kinetic maps mentioned above, as well as the area of the ROI. These measurements may complement the texture features by providing information about lesion size and global enhancement.

2.2.2 *Combination of gray level co-occurrence matrices and first-order statistical features*

The GLCM texture features, which are the standard way to characterize intratumor heterogeneity, have been investigated to predict TNBC response in a previous study.[25] To obtain a fair comparison with the Riesz features, we chose to retain the same configuration of GLCM computation as applied in the earlier study, i.e., interpixel distance of 1 when using an 8-pixel connected neighborhood. Additionally, the analysis of texture at such a low spatial level may be able to capture scalar measures of the microtexture properties of a lesion. We quantized the amplitudes of the kinetic maps into bins of eight gray levels between the 1st and 99th percentiles of their full range and resized them using the Lanczos-3 kernel 58 to a common resolution of 1.5 mm/pixel the coarsest resolution. Thus, GLCM was performed on a fixed distance level to avoid bias introduced by different imaging protocols used by distinct institutions.

Of the original 14 scalar GLCM features, we chose to incorporate the four features: angular second moment (energy), contrast, correlation, and inverse difference moment (homogeneity).[21] The four GLCM features were computed for each of the four directions (0, 45, 90, and 135 deg), and finally we summarized each feature value by computing the average on the four directions. With four different GLCM features and six kinetic maps (three with empirical parameters and three with pharmacokinetic parameters, specified in Sec. 2.1.2), a total of 24 GLCM features were extracted for each lesion image.

We combined the same first-order statistical features—average intensity value over each of the six kinetic maps and the area of the ROI, with the GLCM features.

2.2.3 *Rich set of quantitative image features*

We also studied a broad and overlapping range of features that can capture not only texture but also diverse characteristics of the lesion, and finally evaluated their performance in a discriminative model. Among state-of-the-art imaging features, we select a comprehensive set of quantitative imaging features that are potentially relevant in delineating an exhaustive characterization of a lesion and the surrounding normal tissue from the perfusion image maps. Table 2 summarizes the quantitative features that we computed in this study. The features are selected in such a way that they are either capturing the characteristics commonly visible to the human observers (e.g., intensity-based, texture-based features), or they may contain some discriminative signals for characterizing lesion properties (e.g., lesion shape). In total, we extract 442-dimensional feature vectors from each pharmacokinetic map separately. Finally, we concatenate the feature vectors extracted from 6 maps into 2652-dimensional feature vectors.

Table 2 Quantitative image features included in this study.

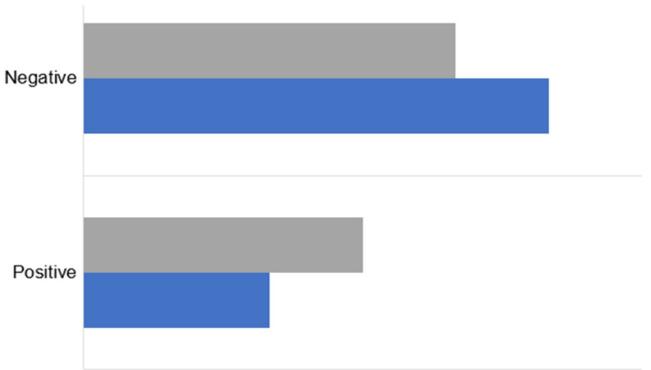
Type	Name (dimension)	Represents
Intensity	Intensity median inside lesion (1)	Quantify first-order intensity distribution within the tumor
	Entropy inside lesion (1)	
	Proportion of pixels with intensity larger than predefined threshold (1)	
	Intensity different between lesion and its peripheral zone (1)	
	No of pixels in different Hist. bins, (20)	
Texture	Daubechies features (324)	Capture occurrence of gray level pattern within the tumor
	Haar wavelets Haar (1)	
	Local binary pattern (12)	
Shape	Compactness (1)	Describe the morphology of the tumor and its boundary
	Eccentricity (1)	
	Roughness (1)	
	Local area integral invariant (15)	
	Radial distance signatures (2)	
Edge-based features	Edge sharpness (60)	Quantify edge sharpness along the tumor boundary
	Histogram on edge—EdgeHist (1)	

2.3 *Predictive Models*

The next step is to design an efficient machine learning model that can exploit the quantitative image descriptors for the prediction of clinical outcomes. Our objective is to predict two different clinical outcomes: (1) pCR, which is the absence of residual lymph metastases and residual invasive breast tumors after neoadjuvant therapy and (2) RT, which is the presence of a residual tumor (RT) with/without lymp node metastases after neoadjuvant therapy; these endpoints can be predicted separately, as they are not perfect inverses of each other. These responses are well-distributed among the patients in the cohort (see Fig. 5). The pCR and RT statuses are determined based on histopathology of the respected tumor bed after therapy.

Two supervised learning approaches are investigated: (i) least absolute shrinkage and selection operator—Lasso[34] (l_1 norm) and (ii) support vector machine—SVM[35] (l_2 norm). Our primary interest to test the performance of Lasso is to make a valid comparison with the state-of-the-art study.[25] But, the consideration of Riesz features has significantly increased the dimensionality of our feature vector, and when the dimensionality of the feature vector p is much larger than the number of training samples n, Lasso selects at most n variable before it saturates.

Therefore, in addition to Lasso regression, we used SVM to create an alternative predictive model, where the model complexity is characterized by the number of support vectors, rather than the dimension of the feature space. SVMs learn linear



	Positive	Negative
pCR	20	21
RT	19	22

Fig. 5 Distribution of pCR and RT responses in the cohort.

separating hyperplanes that maximize the margin between cloud points from two opposed classes (i.e., responders versus nonresponders). Using the kernel trick, SVMs are also capable of creating a curved delineation between responders and nonresponders by mapping the feature vectors in a higher dimensional space using a nonlinear transform. For this study, we designed a nonlinear classifier using SVM with the Gaussian radial basis function kernel, which is a normalized polynomial kernel of infinite degree. The parameters of the Gaussian kernel function (i.e., size of the Gaussian kernel) were optimized using a grid search with cross validation on the training set to maximize the value of AUC. The regions of the grid yielding the best performance were found to be consistent between the training and testing set, which suggests that the trained models generalize well and would likely provide consistent results for new, unseen patients. Thus, the models were not over fit due to the range of hyperparameters tested and the observed consistency in the grids.

Both machine learning models are designed to make separate predictions with same samples before and after completion of treatment, as well as with the two time points combined. On total, we created 90 predictive models using different combinations of intensity and texture features, machine learning models, and treatment time points. The models are able to not only predict response using the prechemotherapy features but also refine these predictions during the treatment process by including the postchemotherapy features. Models involving the pre- and postchemotherapy features did not include a percent change or direct comparison of the two.

In addition to the texture-based models, we also built rich-feature prediction models (pCR and RT) by considering the large set of quantitative features (2652 dimensions) that include not only texture but also shape, boundary, and other characterizations of the lesion (see Table 2). For this kind of situation where $p \gg n$, Lasso is not the ideal method because of the nature of the convex optimization problem and it can only select at most n variables out of p candidates.[36] Elastic net,[37] which is a convex combination of the Lasso and ridge penalty, can deal with a large feature vector, small sample size problem and generate a sparse model with good prediction accuracy. Thus, we applied elastic net with the same cross-validation configuration (10-folds and 20 Monte-Carlo repetition) for the high-dimensional quantitative feature vector.

3 Results

3.1 *Model Performance*

We applied 10-fold cross validation with 20 Monte-Carlo repetitions (MCCV), and the results were averaged. As pointed out in the literature,[38] the MCCV strategy is particularly well-suited for small data sets, which simultaneously optimize the complexity of models, decrease the risk of over fitting, and provide a realistic estimation of prediction errors when the model is applied to new cases. We evaluated the performance of the predictive models using the receiver operating characteristic curves, indicating specificity as related to sensitivity. We show the performance of first-order plus texture-based models and rich-feature models separately in two different tables. Table 3 presents the accuracy of 90 models that are created by considering five different combinations of intensity and texture features for each time point. The performance of rich-feature models that consider an extensive characterization of the lesion (see Table 2) for each time point is summarized in Table 4. In Tables 3 and 4, the accuracy is reported in terms of the averaged AUC value obtained from the Monte-Carlo repetitions.

In Table 3, RT indicates model performance in predicting residual tumor outcome in patients, and pCR indicates model performance in predicting pCR status of patients. To further contextualize the results, N-feature random model performance is recorded in this table, along with textural model performance. RM indicates the random model generated with N features, where N is the number of features in each of the textural measurements involved multiplied by the number of patients in the incorporated time point. We consider N-feature random models partially for over fitting, especially in the Riesz SVM models. Because more features than patients were included in the models, over fitting is a risk. However, the lower performance of the N-feature models (RM) suggests that over fitting was not an issue for the learning.

Notably, the SVM analysis yielded a higher performance than the Lasso regression for the majority of predictive models tested, shown in Table 3. Moreover, the Riesz features-based models outperformed the GLCM models (specially for pre- and posttreatment), suggesting that the multiscale description of texture by Riesz features correlates more closely with pCR and RT status than GLCM features.

From the lower AUC value resulting from the rich-feature model that applied the elastic net regularization on a comprehensive set of quantitative features (see Table 4), we can see that incorporation of higher order characterization of the lesion deteriorates the model performance in this case study. Moreover, significant boosts in prediction accuracy of texture-based models over the rich-feature models show that texture can be considered the most discriminative feature to characterize treatment response of the lesion from the kinetic maps.

The best model created in our study—a concatenation of Riesz and first-order statistical features over pre- and post-chemotherapy for predicting RT—can predict treatment outcomes for more than four of every five patients, shown in Table 3. This shows that Riesz multiscale rotation-covariant texture analysis can be helpful in assessing midtreatment options for TNBC patients. Though the best performing model includes both pre- and posttreatment data for prediction, the posttreatment data are at the very early posttreatment stage; thus, if the model predicts that treatment will not work, there is still time to change strategies, so the method is still potentially useful.

Table 3 The performance of first-order and texture-based models in terms of averaged AUC value. The best performance for predicting RT and pCR with scans from each time point is mentioned in bold font style.

Model name	Lasso (RM)	SVM (RM)	Lasso (RT)	SVM (RT)	Lasso (pCR)	SVM (pCR)
			Prechemotherapy only			
Only first order	Failed	0.20 ± 0.08	0.21 ± 0.02	0.30 ± 0.03	0.27 ± 0.04	0.34 ± 0.08
GLCM	Failed	0.42 ± 0.01	0.27 ± 0.04	0.56 ± 0.03	0.64 ± 0.02	$\mathbf{0.74 \pm 0.01}$
GLCM + first order	0.57 ± 0.03	$\mathbf{0.51 \pm 0.05}$	0.24 ± 0.05	0.63 ± 0.02	0.67 ± 0.04	0.73 ± 0.07
Riesz	Failed	0.48 ± 0.01	0.66 ± 0.03	$\mathbf{0.64 \pm 0.02}$	0.66 ± 0.01	0.69 ± 0.02
Riesz + first order	Failed	0.49 ± 0.03	$\mathbf{0.66 \pm 0.03}$	0.61 ± 0.02	0.69 ± 0.03	0.70 ± 0.01
			Postchemotherapy only			
Only first order	Failed	0.32 ± 0.04	0.21 ± 0.03	0.20 ± 0.04	0.20 ± 0.08	0.30 ± 0.04
GLCM	Failed	$\mathbf{0.61 \pm 0.03}$	0.70 ± 0.01	$\mathbf{0.77 \pm 0.04}$	0.64 ± 0.01	0.71 ± 0.05
GLCM + first order	Failed	0.57 ± 0.04	$\mathbf{0.76 \pm 0.03}$	0.73 ± 0.01	0.61 ± 0.03	0.77 ± 0.04
Riesz	Failed	0.50 ± 0.03	0.66 ± 0.03	0.70 ± 0.05	0.63 ± 0.03	0.78 ± 0.03
Riesz + first order	Failed	0.55 ± 0.01	0.71 ± 0.01	0.73 ± 0.03	$\mathbf{0.68 \pm 0.03}$	$\mathbf{0.80 \pm 0.01}$
			Pre- and postchemotherapy			
Only first order	Failed	0.28 ± 0.02	0.31 ± 0.02	0.42 ± 0.05	0.25 ± 0.03	0.34 ± 0.04
GLCM	Failed	0.60 ± 0.04	0.63 ± 0.03	0.71 ± 0.03	0.70 ± 0.04	0.73 ± 0.05
GLCM + first order	0.51 ± 0.06	0.70 ± 0.05	0.76 ± 0.05	0.80 ± 0.01	0.69 ± 0.03	0.80 ± 0.05
Riesz	Failed	0.66 ± 0.01	0.58 ± 0.03	0.80 ± 0.03	0.54 ± 0.05	0.82 ± 0.03
Riesz + first order	0.58 ± 0.02	$\mathbf{0.72 \pm 0.05}$	$\mathbf{0.77 \pm 0.01}$	$\mathbf{0.85 \pm 0.07}$	$\mathbf{0.72 \pm 0.02}$	$\mathbf{0.83 \pm 0.01}$

Table 4 The performance (AUC values) of the models considering a rich set of quantitative features (see Table 2).

Model name	Elastic net (RT)	Elastic net (pCR)
Prechemotherapy only	0.23 ± 0.05	0.23 ± 0.05
Postchemotherapy only	0.26 ± 0.04	0.24 ± 0.05
Pre- and postchemotherapy	0.57 ± 0.04	0.46 ± 0.04

The AUC of our highest performing model was 0.85, and it predicted RT status, an improvement of ~13%. Also, the best models using the postchemotherapy time point showed an improvement of ~10% for predicting pCR status with an AUC of 0.80 and ~4% for predicting RT status with an AUC of 0.77. In addition, the estimated performance of the Lasso regression used with GLCM and first-order features was equivalent to, and thus validates, previous findings.[25]

3.2 *Feature Selection*

It can be seen from Table 3 that, in most cases, the model with Riesz feature outperforms the GLCM models. However, there are a few instances where the GLCM models resulted in better AUC values. Thus, we ran an automatic feature selection experiment by taking into account both Riesz and GLCM features computed from all six image sequences and incorporate them into a single model. We chose not to consider the Lasso feature selection results because the pairwise correlation among the GLCM group variables/Riesz group variables is quite high. In that case, the Lasso tends to select only one variable from the group and does not care which one is selected. Moreover, for the comparison study, we have to incorporate the whole set of feature vectors [GLCM (24) and Riesz (72)] for six different kinetic maps, which increase the dimension rapidly, and, as mentioned before, elastic net[37] is particularly useful in this situation and with select groups of correlated variables.

Figure 6 shows a digest version of the 10-fold cross validated elastic net feature selection results separately for both pre- and posttreatment datasets. Figures 6(a) and 6(c) show the coefficients values of the features against varying lambda values, where the different colored lines represent different features. Based on cross validation, the optimal solution is picked at the minimum mean square error (MinMSE) and Figs. 6(b) and 6(d) represent group-wise coefficient values of quantitative features at MinMSE, where the different colored bars represent different features computed from different kinetic maps. For the pretreatment dataset, the dominant features at MinMSE are

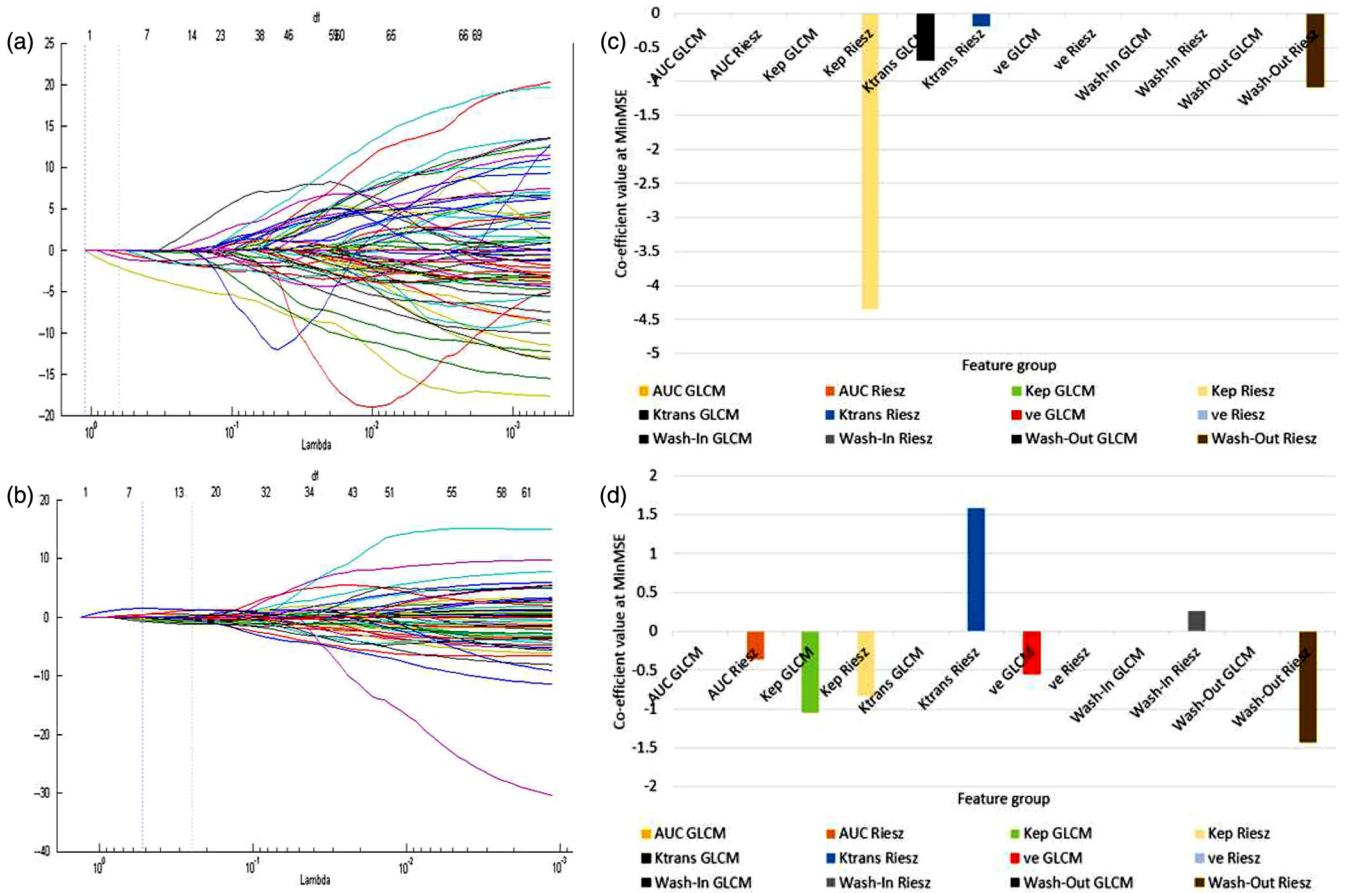


Fig. 6 Elastic net feature selection outcome. (1) considering GLCM and Riesz features calculated from pretreatment kinetic maps (a) trace plot of coefficient fit by elastic net, (b) bar plot represents feature-group-wise coefficient values at MinMSE; (2) considering GLCM and Riesz features calculated from posttreatment kinetic maps, (c) trace plot of coefficient fit by elastic net, and (d) bar plot represents feature-group-wise coefficient values at MinMSE.

mostly selected from the Riesz feature group calculated from wash-out, K^{trans}, and k_{ep} maps. Only the GLCM feature group of the K^{trans} map has nonzero coefficient values [see Fig. 6(b)]. Similarly, for the posttreatment dataset, five Riesz feature groups (AUC, wash-in, wash-out, K^{trans}, and k_{ep}) and only two GLCM feature groups (v_e and k_{ep}) are selected as the most dominant features. In both cases, Riesz features mostly have the higher coefficient values (β). Again, the feature selection results for both the pre- and posttreatment datasets prove that, for all parametric images, Riesz features are more dominant than GLCM, even when they are combined to build the model.

4 Discussion

4.1 *Contribution*

In this exploratory study, our main research contribution is twofold. First, the performance of the Riesz texture analysis framework is thoroughly compared with traditional GLCM to determine if the multiscale rotation-covariant characterization of DCE-MRI-derived lesion kinematic maps can improve the prediction of treatment response in early stage TNBC patients. To build an efficient prediction model, nonlinear (SVM with Gaussian kernel) and logistic (Lasso) regression methods have been tested in parallel with the same set of predictors. Second, rather than assuming texture is the only informative characterization, we also extracted a wide variety of quantitative image features from the kinematic maps to facilitate higher order characterization of the TNBC lesion and evaluate their combined performance in the prediction of treatment response. Among all the characterizations that we studied, our experiments demonstrate that the combination of Riesz texture and first-order statistical feature vectors used in an optimized SVM model (with Gaussian kernel) can most efficiently predict the treatment response. The optimized models achieved 0.85 AUC for RT prediction and 0.83 AUC for pCR prediction, and the improvement over the results in previous studies with GLCM[25] is at most 13% (see Sec. 4).

4.2 *Significance*

The results of this study suggest that quantitative analysis of kinetic imaging using Riesz wavelets can be useful for predicting whether the patient will respond to NAC (specifically, PARP inhibitor therapy). The main strength of our work is that we statistically evaluated a large set of predictive models (~96) incorporating distinct combinations of quantitative lesion characterization at different time points of the treatment, and some AUC values we reported are above 0.8. Generally, the models that included Riesz features performed better, as

evidenced by higher AUC values. This improved performance suggests that the use of Riesz features can provide better prediction for TNBC treatment response than traditional GLCM texture features as well as other popular quantitative image features. Moreover, our study incorporated assessment both before and after chemotherapy, which had better model performance than those without both time points. Therefore, the combination of the different types of image features and time points appear to provide complementary information for predicting treatment response. SVMs generally yielded models with greater performance than Lasso regression. This may be because SVMs use the Gaussian kernel, which allows for a nonlinear decision boundary. On the other hand, there is a risk of overfitting, and, though we evaluated our models using cross validation, future studies to confirm our results in large independent datasets will be needed.

4.3 *Limitations and Future Work*

There are several limitations in our work. One limitation is that the clinical trial from which patient data were obtained for our study were obtained from different centers that used a variety of imaging protocols or varied scan parameters. Despite these variations in imaging technique, the prediction results of our models were good, though their performance would be expected to be better if centers adopt standard procedures in DCE-MRI imaging. In addition, the breast coils that are used during the acquisitions are heterogeneous, and reproducibility of images could be a confounding factor. This was beyond the scope of our study, but it is a factor that would be worthwhile investigating in future studies. Moreover, another future study could be performed using the empirical parameters and the pharmacokinetic parameters extracted from a single center (or standardized) multiphasic contrast-enhanced MRI.

Another limitation is that our best result was using models that included both pre- and posttreatment data. The need for posttreatment data limits the utility of our method for predicting the best treatment prior to commencing that treatment. It is possible that including additional pretreatment imaging time points could provide additional information about the biology of the cancer that could improve the performance of our prediction models, though a limited number of pretreatment scans are generally obtained. Alternatively, incorporating the postchemotherapy time point in between the treatment cycles can be useful for refining predictions and monitoring patient response. Expanding our texture analysis methods to a three-dimensional (3-D) assessment of intra- and intertumoral heterogeneity could allow for more comprehensive characterizations of cancer lesions. However, the MRIs were not isotropic; thus, operating in 3-D would be challenging due to the poor Z-resolution.

A final limitation of our work is that the number of patients studied was small, though all patients were enrolled on a prospective clinical trial and received the same combination therapy. In the future, we will study our methods on a larger cohort of patients receiving other types of NAC treatment. Moreover, we did not include a rule to exclude patients predicted to have achieved pCR from those predicted to have an RT. Patients with pCR status by definition have no RT, so adding such a rule would relate the predictions of the two models. We hope to implement this functionality in the future and expect it will improve model performance. In the future, we will also explore the inclusion of features from complementary imaging modalities, such as mammography, to see if quantitative analysis of information from those modalities improves prediction performance.

Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

Acknowledgments

This work was supported in part by grants from the National Cancer Institute, National Institutes of Health, U01CA142555, 1U01CA190214, and 1U01CA187947.

References

1. K. R. Bauer et al., "Descriptive analysis of estrogen receptor (er)-negative, progesterone receptor (pr)-negative, and her2-negative invasive breast cancer, the so-called triple-negative phenotype," *Cancer* **109**(9), 1721–1728 (2007).
2. K. Collett, R. Skjaerven, and B. Maehle, "The prognostic contribution of estrogen and progesterone receptor status to a modified version of the Nottingham Prognostic Index," *Breast Cancer Res. Treat.* **48**(1), 1–9 (1998).
3. R. Dent et al., "Triple-negative breast cancer: clinical features and patterns of recurrence," *Clin. Cancer Res.* **13**(15), 4429–4434 (2007).
4. C. Liedtke et al., "Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer," *J. Clin. Oncol.* **26**(8), 1275–1281 (2008).
5. A. Ring et al., "Oestrogen receptor status, pathological complete response and prognosis in patients receiving neoadjuvant chemotherapy for early breast cancer," *Br. J. Cancer* **91**(12), 2012–2017 (2004).
6. G. von Minckwitz et al., "Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes," *J. Clin. Oncol.* **30**(15), 1796–1804 (2012).
7. A. A. Rodriguez et al., "DNA repair signature is associated with anthracycline response in triple negative breast cancer patients," *Breast Cancer Res. Treat.* **123**(1), 189–196 (2010).
8. D. P. Silver et al., "Efficacy of neoadjuvant cisplatin in triple-negative breast cancer," *J. Clin. Oncol.* **28**(7), 1145–1153 (2010).
9. M.-L. W. Ah-See et al., "Early changes in functional dynamic magnetic resonance imaging predict for pathologic response to neoadjuvant chemotherapy in primary breast cancer," *Clin. Cancer Res.* **14**(20), 6580–6589 (2008).
10. L. J. Wilmes et al., "Ag-013736, a novel inhibitor of VEGF receptor tyrosine kinases, inhibits breast cancer growth and decreases vascular permeability as detected by dynamic contrast-enhanced magnetic resonance imaging," *Magn. Reson. Imaging* **25**(3), 319–327 (2007).
11. L. W. Turnbull, "Dynamic contrast-enhanced MRI in the diagnosis and management of breast cancer," *NMR Biomed.* **22**(1), 28–39 (2009).
12. N. Hylton, "Dynamic contrast-enhanced magnetic resonance imaging as an imaging biomarker," *J. Clin. Oncol.* **24**(20), 3293–3298 (2006).
13. S. Rehman and G. C. Jayson, "Molecular imaging of antiangiogenic agents," *Oncologist* **10**(2), 92–103 (2005).
14. R. A. Gatenby, O. Grove, and R. J. Gillies, "Quantitative imaging in cancer evolution and ecology," *Radiology* **269**(1), 8–14 (2013).
15. F. Davnall et al., "Assessment of tumor heterogeneity: an emerging imaging tool for clinical practice?," *Insights Imaging* **3**(6), 573–589 (2012).
16. C. N. De Cecco et al., "Texture analysis as imaging biomarker of tumoral response to neoadjuvant chemoradiotherapy in rectal cancer patients studied with 3-T magnetic resonance," *Invest. Radiol.* **50**(4), 239–245 (2015).
17. H. J. Yu et al., "MRI measurements of tumor size and pharmacokinetic parameters as early predictors of response in breast cancer patients undergoing neoadjuvant anthracycline chemotherapy," *J. Magn. Reson. Imaging* **26**(3), 615–623 (2007).
18. K. Holli et al., "Characterization of breast cancer types by texture analysis of magnetic resonance images," *Acad. Radiol.* **17**(2), 135–141 (2010).

19. J. Parikh et al., "Changes in primary breast cancer heterogeneity may augment midtreatment MR imaging assessment of response to neoadjuvant chemotherapy," *Radiology* **272**(1), 100–112 (2014).
20. A. Depeursinge et al., "Three-dimensional solid texture analysis in biomedical imaging: review and opportunities," *Med. Image Anal.* **18**(1), 176–196 (2014).
21. G. Castellano et al., "Texture analysis of medical images," *Clin. Radiol.* **59**(12), 1061–1069 (2004).
22. N. Bhooshan et al., "Potential of computer-aided diagnosis of high spectral and spatial resolution (HiSS) MRI in the classification of breast lesions," *J. Magn. Reson. Imaging* **39**(1), 59–67 (2014).
23. A. Ahmed et al., "Texture analysis in assessment and prediction of chemotherapy response in breast cancer," *J. Magn. Reson. Imaging* **38**(1), 89–101 (2013).
24. S. Waugh et al., "Abstract p4-02-10: Response to neoadjuvant chemotherapy assessed by magnetic resonance imaging (MRI) texture analysis compared with residual cancer burden (RCB) and tumor subtype in primary breast cancer," *Cancer Res.* **76**(4 Suppl.), P4–02 (2016).
25. D. I. Golden et al., "Dynamic contrast-enhanced MRI-based biomarkers of therapeutic response in triple-negative breast cancer," *J. Am. Med. Inform. Assoc.* **20**(6), 1059–1066 (2013).
26. M. Unser and D. Van De Ville, "Wavelet steerability and the higher-order Riesz transform," *IEEE Trans. Image Process.* **19**(3), 636–652 (2010).
27. A. Depeursinge et al., "Predicting visual semantic descriptive terms from radiological image data: preliminary results with liver lesions in CT," *IEEE Trans. Med. Imaging* **33**(8), 1669–1676 (2014).
28. M. L. Telli et al., "Phase II study of gemcitabine, carboplatin, and iniparib as neoadjuvant therapy for triple-negative and BRCA1/2 mutation-associated breast cancer with assessment of a tumor-based measure of genomic instability: Precog 0105," *J. Clin. Oncol.* **33**(17), 1895–1901 (2015).
29. D. L. Rubin et al., "Automated tracking of quantitative assessments of tumor burden in clinical trials," *Transl. Oncol.* **7**(1), 23–35 (2014).
30. P. S. Tofts and A. G. Kermode, "Measurement of the blood-brain barrier permeability and leakage space using dynamic MR imaging. 1. fundamental concepts," *Magn. Reson. Med.* **17**(2), 357–367 (1991).
31. S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(7), 674–693 (1989).
32. A. Depeursinge et al., "Multiscale lung texture signature learning using the Riesz transform," in *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, pp. 517–524, Springer (2012).
33. A. F. Frangi et al., "Multiscale vessel enhancement filtering," in *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention*, pp. 130–137, Springer (1998).
34. R. Tibshirani, "Regression shrinkage and selection via the lasso: a retrospective," *J. R. Stat. Soc. B* **73**(3), 273–282 (2011).
35. C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.* **20**(3), 273–297 (1995).
36. B. Efron et al., "Least angle regression," *Ann. Stat.* **32**(2), 407–499 (2004).
37. H. Zou and T. Hastie, "Regularization and variable selection via the elastic net," *J. R. Stat. Soc. B* **67**(2), 301–320 (2005).
38. Q.-S. Xu and Y.-Z. Liang, "Monte Carlo cross validation," *Chemom. Intell. Lab. Syst.* **56**(1), 1–11 (2001).

Imon Banerjee received her PhD in computer science from the University of Genoa, Italy. During her PhD, she was awarded the Marie Curie European Fellowship and worked as an early-stage researcher at the National Research Council, Italy. She is a postdoctoral scholar in the Laboratory of Quantitative Imaging at Stanford University. Her research is focused on developing unstructural data analysis and big data mining techniques to support clinical diagnosis and treatment.

Sadhika Malladi is an undergraduate student at Massachusetts Institute of Technology (MIT) majoring in mathematics and computer science. She has worked on several projects in bioinformatics involving personalized prognosis predictions for cancer patients. More recently, her work focuses on machine learning algorithms and efficient optimizations of them.

Daniela Lee is currently a junior at Yale University majoring in ecology and evolutionary biology. During her internship at Stanford, she worked on the standardization of image data into AIM files as well as the analysis of image features to create a treatment response model to neoadjuvant chemotherapy. At Yale, her research interests involve studying phosphate homeostasis of Drosophila, specifically how phosphate transporter expression is regulated.

Adrien Depeursinge completed his PhD thesis on medical image analysis at the University Hospitals of Geneva. He did a postdoctoral fellowship in the Department of Radiology at Stanford University. He holds a joint position as a professor at the University of Applied Sciences Western Switzerland, Sierre, Switzerland, and a research associate in the Biomedical Imaging Group at EPFL. His research interests include *N*-dimensional texture analysis with control of image scales and directions as well as clinical workflows of image-based computer-aided diagnosis systems.

Melinda Telli received her medical degree from George Washington University with distinction and completed her residency in internal medicine at Stanford University. She is an assistant professor in the Division of Medical Oncology at Stanford University School of Medicine with a research focus on breast cancer. Her primary research focus is the development of innovative therapeutic strategies for the treatment of triple-negative and BRCA1/2 mutation-associated breast cancer.

Jafi Lipson received her MD degree from the University of California, San Francisco (UCSF) School of Medicine and completed a diagnostic radiology residency program at UCSF. She was a graduate of Harvard College. She is an investigator and faculty member in the Department of Radiology at Stanford University School of Medicine. Her research interests include mammographic density and breast cancer risk assessment, early breast cancer detection and extent of disease evaluation, neoadjuvant treatment response, and informatics applications in breast imaging.

Daniel Golden received his PhD in electrical engineering from Stanford, and during his postdoc there, he focused on using machine learning to predict outcomes and disease characteristics in cancer patients. He is the director of Machine Learning at Arterys, a startup focused on streamlining the practice of medical image interpretation and postprocessing. Before coming to Arterys, he led the machine learning team at CellScope, a San Francisco-based medical device startup.

Daniel L. Rubin is an associate professor of biomedical data science, radiology, and medicine (Biomedical Informatics Research) at Stanford University. His NIH-funded research program focuses on quantitative imaging, integrating imaging with clinical and molecular data, and mining these big data to discover imaging phenotypes that can predict disease biology, define disease subtypes, and personalize treatment. He has more than 240 scientific publications in biomedical imaging informatics and medical imaging.