Translational Oncology
www.transonc.com

Volume 12 Number 10 October 2019 pp. 1271–1281 1271

Breast Cancer Treatment Response Monitoring Using Quantitative Ultrasound and Texture Analysis: Comparative Analysis of Analytical Models

Lakshmanan Sannachi[*,†], Mehrdad Gangeh[*,†], Hadi Tadayyon[*,†], Sonal Gandhi[‡], Frances C. Wright[§], Elzbieta Slodkowska[¶], Belinda Curpen[#], Ali Sadeghi-Naini[**], William Tran[*,†] and Gregory J. Czarnota[*,†,††]

[*]Physical Sciences, Sunnybrook Research Institute, Sunnybrook Health Sciences Centre, Toronto, ON, Canada; [†]Department of Radiation Oncology, Odette Cancer Centre, Sunnybrook Health Sciences Centre, Toronto, ON, Canada; [‡]Medical Oncology, Department of Medicine, Sunnybrook Health Sciences Centre, Toronto, ON, Canada; [§]Surgical Oncology, Department of General Surgery, Sunnybrook Health Sciences Centre, Toronto, ON, Canada; [¶]Department of Anatomic Pathology, Sunnybrook Health Sciences Centre, Toronto, ON, Canada; [#]Department of Medical Imaging, Sunnybrook Health Sciences Centre, Toronto, ON, Canada; [**]Department of Electrical Engineering & Computer Science, York University, Toronto, ON, Canada; [††]Department of Medical Biophysics, Faculty of Medicine, University of Toronto, Toronto, ON, Canada

Abstract

PURPOSE: The purpose of this study was to develop computational algorithms to best determine tumor responses early after the start of neoadjuvant chemotherapy, based on quantitative ultrasound (QUS) and textural analysis in patients with locally advanced breast cancer (LABC). *METHODS:* A total of 100 LABC patients treated with neoadjuvant chemotherapy were included in this study. Breast tumors were scanned with a clinical ultrasound system prior to treatment, during the first, fourth and eighth weeks of treatment, and prior to surgery. QUS parameters were calculated from ultrasound radio frequency data within tumor regions. Texture features were extracted from each QUS parametric map. Patients were classified into two groups based on identified clinical/pathological response: responders and non-responders. In order to differentiate treatment responders, three multi-feature response classification algorithms, namely a linear discriminant, a k-nearest-neighbor and a nonlinear support vector machine classifier were compared. *RESULTS:* All algorithms distinguished responders and non-responders with accuracies ranging between 68% and 92%. In particular, support vector machine performed the best in differentiating responders from non-responders with accuracies of 78%, 90% and 92% at weeks 1, 4 and 8 after the start of treatment, respectively. The most relevant features in separating the two response groups at early stages (weeks 1and 4) were texture features and at a later stage (week 8) were mean QUS parameters, particularly ultrasound backscatter intensity-based parameters. *CONCLUSION:* An early stage treatment response prediction model developed

Address all correspondence to: Dr. Gregory J. Czarnota, Department of Radiation Oncology, Sunnybrook Health Sciences Centre, 2075, Bayview Avenue, Toronto, Ontario M4N 3M5. E-mail: ;Gregory.Czarnota@sunnybrook.ca
Received 5 February 2019; Revised 12 June 2019; Accepted 17 June 2019

© 2019 Published by Elsevier Inc. on behalf of Neoplasia Press, Inc. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).
1936-5233/19
https://doi.org/10.1016/j.tranon.2019.06.004

by quantitative ultrasound and texture analysis combined with modern computational methods permits offering effective alternatives to standard treatment for refractory patients.

Translational Oncology (2019) 12, 1271–1281

Introduction

Locally advanced breast cancer (LABC) is an aggressive form of breast cancer that comprises a wide range of clinical presentations including T3/T4 disease, tumor larger than 5 cm or extensive axillary lymph node involvement. Due to rapid disease progression and a high risk for metastatic spread, patients with LABC typically have a poorer long-term survival rate compared to early stage breast cancer patients [1]. The standard treatment for patients with LABC is neoadjuvant chemotherapy (NAC), followed by surgery and, if required, radiation [2]. However, variable tumor responses have been observed in patients receiving NAC with only up to 5−10% of patients achieving pathological complete response to NAC. Evidence also suggests that the pathological response of a tumor to NAC correlates to long-term disease-free survival (DFS) and overall survival (OS) [3,4]. Measuring tumor response at early stages of NAC may help guide treatments for potentially improved DFS. LABC treatment response is typically evaluated at the conclusion of treatment based on pathology, commonly using a Miller-Payne (MP) score calculated from changes in tumor cellularity comparing pre-treatment biopsy to post-treatment surgical specimens. This is done months after treatment [5]. Therefore, the early detection of treatment response of breast tumors is very important for guiding cancer therapy decisions based on individual patient responses.

Functional imaging techniques that can detect changes in tumor micro-structure and physiology in response to treatment could provide early assessments of therapy response. In this context, a number of imaging modalities, including positron emission tomography (PET) and magnetic resonance imaging (MRI) have recently been demonstrated capable of evaluating cancer treatment response within weeks of treatment initiation. Specifically, [18F]fluorodeoxyglucose PET has been demonstrated to detect the pathological response of breast cancer as early as after the first cycle of neo-adjuvant chemotherapy [6]. Several functional magnetic resonance imaging techniques such as dynamic contrast-enhanced (DCI)-MRI, diffusion weighted (DW)-MRI, blood oxygenation level-dependent (BOLD)-MRI, and MR elastography have been used to characterize breast cancer and its response to chemotherapy early or after a full course of treatment based on changes in tumor microvasculature, cellularity, hypoxia, metabolism, oxygenation and stiffness [7−11]. These modalities are often costly and require contrast agents to monitor tumor response to treatment.

Previous studies have demonstrated the potential of quantitative ultrasound (QUS) techniques to characterize various tissue types, classify tissue abnormalities, and differentiate tumor types [12,13]. This technique uses variation in the acoustic property within tissues to characterize microstructural features. In preclinical studies, QUS techniques have been used in the detection of tumor response to cancer therapies such as chemotherapy, photodynamic therapy, X-ray radiation therapy, and ultrasonically-stimulated, anti-vascular microbubble treatments, or combinatorial treatments [14−17]. Similarly, in a pilot clinical study with limited numbers of patients, QUS techniques were used to differentiate pathological responders (n = 23) from non-responders (n = 7) in an LABC population treated with NAC. The sensitivity and specificity were 77% and 86%, respectively, at week 1 and 83% and 100% at week 4, respectively, after the start of treatment [18,19]. In those studies, quantitative ultrasound parameters, mid-band fit (MBF), spectral slope (SS) and 0-MHz intercept (SI), average scatterer diameter (ASD), and average acoustic concentration (AAC) exhibited a strong correlation with tumor response. These quantitative ultrasound parameters reflect tissue micro-structural properties, such as scatterer size, shape and organization in addition to elastic properties. A recent study combining QUS parameter and texture features such as contrast, correlation, energy and homogeneity has reported an improvement in accuracy in differentiating responder and non-responder groups [20]. The sensitivity and specificity were reported as high as 100% and 93%, respectively, 1 week after NAC initiation. These texture parameters quantify the spatial relationship between neighboring regions with respect to acoustic properties. In those studies, classification analysis was performed on estimated parameters using a linear discriminant classifier.

In this study, we have explored the best classification methods for discriminating responding patients from non-responding patients. Specifically, we investigate three classification algorithms based on changes in quantitative ultrasound and texture features early after neoadjuvant chemotherapy using modern computational approaches. Machine learning in biomedical imaging research is an emerging field with algorithms facilitating the handling of large data sets, the integration of multiple complex measurements, and the identification of statistically significant data. Typically, these algorithms are implemented in two stages, the first being the establishment of a classification algorithm using training data and, the second being the testing of the model with separate data. Finally, classification algorithm accuracy, sensitivity and specificity are reported. As a continuation of the previous work [20], in the work here, the patient sample size was expanded to a larger cohort of 100 LABC patients. Six quantitative ultrasound parameters as well as four texture features were determined for each patient at multiple times during treatment with chemotherapy. Three multi-feature classification algorithms for response evaluation were developed using linear discriminant analysis, k-nearest neighbor, and support vector machine classifiers in conjunction with a sequential forward feature selection method and a leave-one-patient-out scheme for cross-validation. The results of these analytical methods were compared in terms of their capability to best identify responding and non-responding patients at various times during their treatments.

Materials and Methods

Study Design

This study was approved by the Sunnybrook Research Institute research ethics board. One-hundred LABC patients with an eligible

age range of 18−85 years were enrolled with informed consent in this study. Before treatment, all patients were subjected to a core needle biopsy to confirm a cancer diagnosis and to determine tumor histological subtype and hormone receptor status (estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2)). Magnetic resonance images were acquired as part of clinical care before treatment in order to determine initial tumor size. Ultrasound data were collected from each patient at 5 specific times: before treatment, at weeks 1, 4, and 8 during treatment and finally at 1 week before surgery.

After surgery, surgical specimens were prepared onto a $5'' \times 7''$ whole-mount pathology slide and digitized using a confocal scanner (TISSUEscope™, Huron Technologies, Waterloo, ON). A board--certified pathologist examined the specimens and recorded the results in the patient's medical chart.

The clinical/pathological tumor response of each patient was determined at the end of their treatment using a modified response (MR) grading system. This system was based on RECIST criteria but took into account further response characterization based on pathology [MR Score 1: no diminishment in tumor size (cNR); MR2: up to 30% diminishment in tumor size (cNR); MR 3: between an estimated 30% and 90% reduction in tumor size (cPR); MR 4: a diminishment of more than 90% in tumor size (almost pCR); MR 5: no evident tumor and no malignant cells identifiable in sections from the site of the tumor; only vascular fibroelastic stroma remains, often containing macrophages; however, ductal carcinoma in situ may be present (pCR)]. This avoids confounding factors and prevents misclassification of response. For instance, scar tissue rarely results in a persistent mass with no tumor cellularity, and is considered a responder, when it would otherwise be classified as a non-responder.

A binary classification was investigated where a patient with an MR score of 3−5 was deemed to be a responder (R) and a patient with an MR score of 1−2 was deemed to be a non-responder (NR). All patients received anthracycline/taxane-based treatment lasting several months. Recurrence-free survival was determined based on a 6-year follow up timeframe, free of any local or distant cancer recurrence.

Quantitative Ultrasound and Texture Parameter Estimation

Ultrasound data were collected from tumor regions using a Sonix RP clinical research system (Analogic Medical Corp., Vancouver, Canada) equipped with a L14−5/60 transducer with a central frequency of 6.5 MHz and bandwidth of 3−8 MHz. Data were digitally collected with a sampling frequency of 40 MHz. From each breast tumor, frames were collected with intervals of 1 cm across the breast, with the transducer focus at the center of the tumor. The sector size for each image frame was 6 cm along the lateral direction and 4−6 cm along the axial direction.

From each ultrasound frame, MBF (mid-band fit), SS (spectral slope), SI (spectral intercept), ACE (attenuation co-efficient estimate), SAS (spacing among scatterers), ASD (acoustic scatterer diameter) and AAC (average acoustic-scatterer concentration) were determined using quantitative ultrasound methods. In this technique, tumor regions of interest were selected. Each region of interest was divided into window blocks of size $10\lambda \times 10\lambda$, where λ represents that ultrasound wavelength, with a 94% overlap in the axial and lateral directions. Tumor attenuation (ACE) was determined using a spectral difference method. In previous work, for more complex transducers, such as arrays, a reference phantom method was proposed to remove system dependent effects in backscatter property estimation [21]. This reference phantom method was used here to remove any clinical system dependences in quantitative ultrasound parameters. The reference phantom used in this study consisted of glass beads (5−30 μm) embedded in a homogeneous background of microscopic oil droplets in gelatin. The attenuation coefficient and speed of sound of the reference phantom were 0.786 dB/MHz/cm and 1540 m/s, respectively. The MBF, SS and SI were calculated using linear regression analysis of the normalized backscatter power spectrum over the −6 dB bandwidth of the transducer [22]. Spacing among scatterers (SAS) was determined using an autoregressive spectral analysis method by modeling the tumor echo signal as an autoregressive signal [23]. The ASD and AAC parameters were derived from the backscatter coefficient by comparing measured data with a theoretically derived backscatter coefficient using a spherical Gaussian scatterer model (SGM) [24]. Finally, color-coded parametric maps for each estimated quantitative ultrasound parameter were constructed by generating a spatial map of the parameter values computed over all window blocks. The mean values of quantitative ultrasound parameters were determined by averaging QUS parametric map values.

In addition to the mean values of QUS parameters, spatial distributions of QUS parameters in parametric maps were evaluated using a gray-level co-occurrence matrix (GLCM) [25] method. The GLCM represents the angular relationships and distances between neighboring pixels in parametric maps. Using GLCM methods, four texture features including contrast (CON), correlation (COR), homogeneity (HOM), and energy (ENE) were derived from each QUS parametric map. Here, the contrast feature represents location--dependent gray level variation of an image. The energy feature measures textural uniformity. The homogeneity feature measures the incidence of pixel pairs of different intensities. The correlation feature measures the linear dependency between neighboring pixels. In short, a total of 24 textural features (four texture features from each of the MBF, SS, SI, SAS, ASD and AAC parametric maps) were computed. The changes in mean of QUS parameters and texture features for each patient during the course of treatment were calculated using the corresponding parameters acquired at pre-treatment as the baseline. The algorithms to detect treatment response groups were developed by performing classification analysis on these 31 features (7 changes in mean of QUS parameters after treatment, and 24 changes in texture features after treatment).

Classification Model Algorithms

In order to differentiate treatment responders from non-responders, three multi-feature response classification algorithms, namely a linear discriminant analysis (LDA), a k-nearest-neighbor classifier (KNN), and a radial-basis-function support vector machine classifier (SVM-RBF) were compared. Linear discriminant analysis is a standard approach to classification problems. LDA assumes that each class probability density function is a normal density function with the same covariance for all classes. Then, a new data point is classified by determining a probability density function whose value is larger than the others.

A KNN classifier is based on comparing a given test sample with training samples that are similar to it. The training samples are described by n features in an n-dimensional feature space. Given an unknown sample, a k-nearest neighbor (KNN) classifier searches a feature space for the k training samples that are closest to the unknown sample. These k training samples are the k-nearest neighbors of the unknown sample. Closeness is defined in terms of distance metrics, such as Euclidean distance. In a two-class case, the number of nearest-neighbor, k, is set to

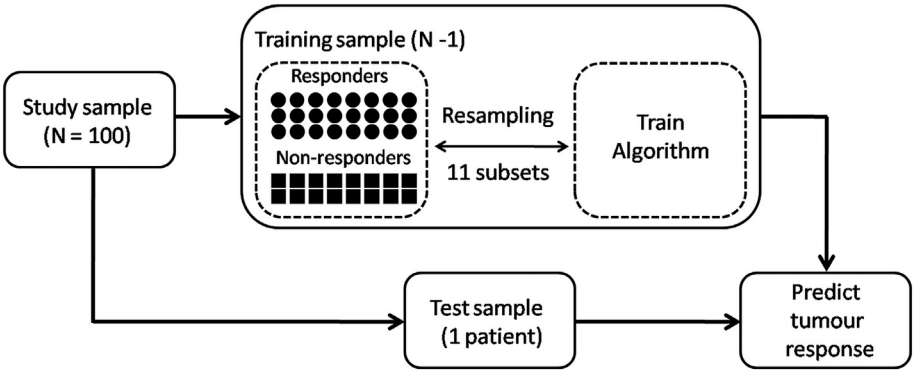

Figure 1. Flow diagram of computational algorithm training and testing process. The group imbalance problem was addressed through a resampling step, which entailed down sampling the majority group (responders) followed by algorithm training. This process was repeated over 11 iterations and the algorithm predicted tumor response by majority voting over 11 iterations or subset models. A leave-one-out cross-validation approach was used to separate algorithm training and testing data.

an odd number to avoid tied classes. The optimal k value is tuned to obtain the best classification result.

A kernel-based support vector machine is an algorithm that uses nonlinear mapping to transfer the original training data to a higher-dimensional feature space where it searches for a linear, optimal hyperplane that has a maximum margin in separating two classes [26]. In this study, a support vector machine with a radial-basis-function kernel (SVM-RBF) was used to separate the two response groups. In the SVM-RBF algorithm, two classifier parameters are tuned (with C, the penalty for misclassification, and γ, the width of a radial basis function). C is important for good generalization performance as it controls regularization, which counteracts overfitting of the training data. In the work here, the optimal C and γ were selected by a grid search in the ranges of $C = 2^{1}$ to 2^{15} and $\gamma = 2^{-15}$ to 2^{5}, respectively, using the LIBSVM-3.22 package for SVM analyses [27].

The work here had a larger number of samples in one class (e.g. responders) than in the other class (e.g. non-responders). This class imbalance problem was circumvented by subsampling the original data into 11 subsets, such that each subset had an equal number of R and NR population members. This also ensured that all patients in the responder group were selected at least once over all subsets. The feature selection was performed based on a sequential-forward selection (SFS) method that learns which features are most informative at each time, choosing the next features based on already selected features and the internal belief of the classifier. In this study, a leave-one-out cross validation approach was used. It involves training the computational algorithm with all subjects except one while the 'left-out' subject is used for algorithm testing. This interactive process is repeated until all subjects are left out for algorithm testing at least once (Figure 1). Cross-validation was used to establish the generalization ability of an algorithm to new or previously unseen subjects. The validity of the algorithm in predicting treatment response group was evaluated using prediction accuracy. Receiver operating characteristic curves (ROC) and the corresponding area under curve were (AUC) computed.

Statistical Analysis

Statistical tests were used to compare response groups in terms of mean QUS and texture-based parameters. To determine the type of statistical test to use to compare the groups, a Shapiro–Wilk normality test was performed on each feature data set to determine whether it followed a normal distribution. An unpaired *t*-test was completed for the data that passed the normality test, otherwise, a Mann–Whitney unpaired test was used. Recurrence-free survivals for the two response populations were created by the Kaplan–Meier method to clarify the time-dependent cumulative survival rate, and the curves were compared using a log-rank test. A value of $P < .05$ was considered to determine statistical significance.

SVM-RBF Classification Model Validation

In order to test the robustness of the best performing classification algorithm, the SVM-RBF model was validated using a new patient data set acquired with two different clinical ultrasound systems: (i) Ultrasonix - RP (L14–5/60, Analogic Medical Corp., Vancouver, Canada) equipped with L14–5/60 linear array transducer with a central frequency of 6.5 MHz and bandwidth of 3–8 MHz, and (ii) GE – LOGIQ E9 (GE Healthcare, Milwaukee, Wisconsin, USA) with 9 L-D linear array transducers, with a center frequency of 6 MHz and bandwidth of 3–8.5 MHz. Twenty-four LABC patients were included in the model validation study. Out of 24 patients, 20 were clinical-pathological responders and 4 were non-responders. Their mean age was 52 ± 11 years, and the mean tumor sizes before and after treatment were 5.7 ± 2.6 and 2.8 ± 3.6 cm, respectively. Among the 24 patients, 88% had invasive ductal carcinoma. Ultrasound data were collected from these patients before treatment and at weeks 1, 4, and 8 during treatment. Quantitative ultrasound parameters and texture parameters were determined from the ultrasound data as above.

Results

LABC Patient Characteristics

The clinical and pathological characteristics of the study patients are summarized in Table 1. The average age of patients was 49 ± 11 years (range: 29–82 years). The average tumor size along the longest axis before treatment was 5.9 ± 2.8 cm (range: 1.6–14 cm). Among patients, 90% had invasive ductal carcinoma and 5% had lobular carcinoma. Seven had grade I tumors, 46 had grade II tumors, and 20 had grade III tumors. Also, 37% of patients were triple negative in terms of molecular subtypes (ER/PR/Her2). Almost all patients (88%) received combined anthracycline-taxane neoadjuvant treatment plans. Out of 100 patients, 81 were responders and 19 were non-responders based on the modified response score. The average

Table 1. Clinical and pathologic characteristics of LABC patients receiving neo-adjuvant chemotherapy

Characteristics	R (N = 81)	NR (N = 19)	All (N = 100)
Age (year)	50 ± 10	49 ± 11	49 ± 11
Menopause			
Postmenopausal (%)	32.1	31.5	32
Premenopausal (%)	56.8	63.2	58
Perimenopausal (%)	6.2	0	5
Unknown (%)	4.9	5.3	5
Initial tumor size (cm)	5.7 ± 2.7	6.5 ± 3.4	5.9 ± 2.8
Histology			
IDC (%)	91.4	84.2	90
ILC (%)	3.7	10.5	5
IMC (%)	4.9	5.3	5
Grade			
I (%)	7.4	5.3	7
II (%)	44.4	52.6	46
III (%)	16.1	36.8	20
Unknown (%)	32.1	5.3	27
ER/PR/Her2			
Triple negative (%)	25.9	31.6	37
Non-triple negative (%)	74.1	68.4	63
Treatment			
ACT (%)	59.3	63.2	60
FECD (%)	28.4	26.3	28
Others (%)	12.3	10.5	12
Residual tumor size (cm)	2.1 ± 2.7	7.4 ± 5.5	3.1 ± 3.9

Abbreviations: NAC, neoadjuvant chemotherapy; R, responder; NR, non-responder; IDC, invasive ductal carcinoma; IMC, invasive mammary carcinoma; ILC, invasive lobular carcinoma; ER, estrogen receptor; PR, progesterone receptor; HER2, human epidermal growth factor receptor 2; ACT, adriamycin, cytoxan, and paclitaxel; FECD, 5-fluourouracil, epirubicin, cyclophosphamide and docetaxel.

tumor sizes of responders and non-responders after treatment were 2.1 ± 2.7 and 7.4 ± 5.5 cm, respectively.

Quantitative Ultrasound and Texture Parameters

Parameters determined from ultrasound radio frequency signals were MBF, SS, SI, ACE, SAS, ASD and AAC. Representative ultrasound B-mode images and MBF parametric images corresponding to responding and non-responding patients acquired prior to chemotherapy treatment, and after 4 weeks of treatment, are displayed in Figure 2 along with high magnification H&E images of corresponding histopathology. MBF parametric images demonstrated increases in the ultrasound backscatter power and changes in associated textural patterns within the tumor region for treatment responding patients after the start of treatment by weeks 1−4. In contrast, non-responder MBF parametric images exhibited no significant changes in their mean value and textural patterns. Histopathology analysis revealed no residual disease, with apparent degradation of tumor cells in responding patients. In contrast, there were large deposits of residual disease and viable-appearing cells in non-responders. These changes in tumor macro- and microstructure were reflected in estimated ultrasound parameters. Statistical analysis using unpaired t-tests was performed in order to compare changes in mean QUS values and texture parameters between responders and non-responders acquired at weeks 1, 4, and 8 during treatment. None of the changes in mean QUS and texture parameters alone demonstrated significant differences between R and NR groups at weeks 1 and 4. During treatment at week 8, the parameter values of ΔMBF ($P = .0004$), ΔSI ($P = .001$), ΔAAC ($P = .0004$), ΔMBF-ENE ($P = .007$), ΔMBF-HOM ($P = .008$), ΔASD-CON ($P = .031$) and ΔAAC-CON ($P = .015$) were significantly different between R and NR.

Comparison of Classification Algorithms

First, correlation between estimated 31 features (7 changes in mean of QUS parameters after treatment, and 24 changes in texture features after treatment) were investigated. None of the two features were highly ($R^2 \geq 0.9$) correlated to each other. In order to differentiate responder and non-responder patient groups, a multi-feature classification analysis was performed on changes in mean quantitative ultrasound and texture parameters using LDA, KNN and SVM-RBF classifiers. Classification performance between classifiers was compared using paired t-tests. First, the average classification accuracy for each sample across cross-validation folds was computed. Those sample-specific average performances using the different classification algorithms were then statistically compared. The classification results obtained using LDA, KNN and SVM-RBF algorithms, and where there were statistically significant differences in performance, are displayed in Figure 3. The associated receiver-operating-characteristic (ROC) curve and area under curve values are presented in Figure 4. All algorithms differentiated responder and non-responders with prediction accuracy ranges of 63−80%, 68−90%, and 88−92%, at weeks 1, 4, and 8 after the start of therapy, respectively. In particular, at week 1, both KNN and SVM-RBF algorithms identified response groups correctly with accuracies of 79% and 77%, respectively. Performance improved at later times. At week 4, the SVM-RBF algorithm performed very well in predicting the two response groups with sensitivity, specificity, and accuracy of 90%, 89% and 90%, respectively. At week 8, all classifiers demonstrated a sensitivity range of 84−90%, a specificity range of 85−95%, and an accuracy range of 88−92%.

Overall, the SVM-RBF algorithm performed very well in differentiating responder and non-responder patients at all times. The accuracies in separating the two response groups using the SVM-RBF-based algorithms were 78%, 90%, and 92%, at week 1, 4, and 8, respectively. The most relevant features in separating the two response groups were changes in texture features at early stages (week 1 and 4) and change in mean quantitative ultrasound parameters, particularly ultrasound backscatter intensity-related parameters, at a later stage (week 8) (Table 2). Hyperplanes of decision-making, defined by the SVM-RBF kernel using one of the week 1, 4, and 8 subsets in three-dimensional feature space, are displayed in Figure 5. Visual representation of the data indicates visible separability of the response groups.

Figure 6 displays results of recurrence-free survival (RFS) analyses using QUS-based predictions of outcome early during treatment. The plots present the 6-year survival curves calculated for responding and non-responding patients classified using the SVM-RBF model based on changes in mean QUS and texture parameters at weeks 1, 4 and 8 after the start of treatment, compared to those at post-treatment based on a modified response score. A statistically significant difference between response groups was identified between the survival curves generated from the patients' pathological response ($P < .0001$). Similarly, a statistically significant difference was observed between the survival curves predicted using the SVM-RBF model based on optimal feature sets at week 1 ($P = .026$), week 4 ($P = .0004$) and week 8 ($P = .0158$) of treatment.

SVM-RBF Classification Model Validation

In order to test the robustness of the best performing classification algorithm, the SVM-RBF model was tested with an independent data

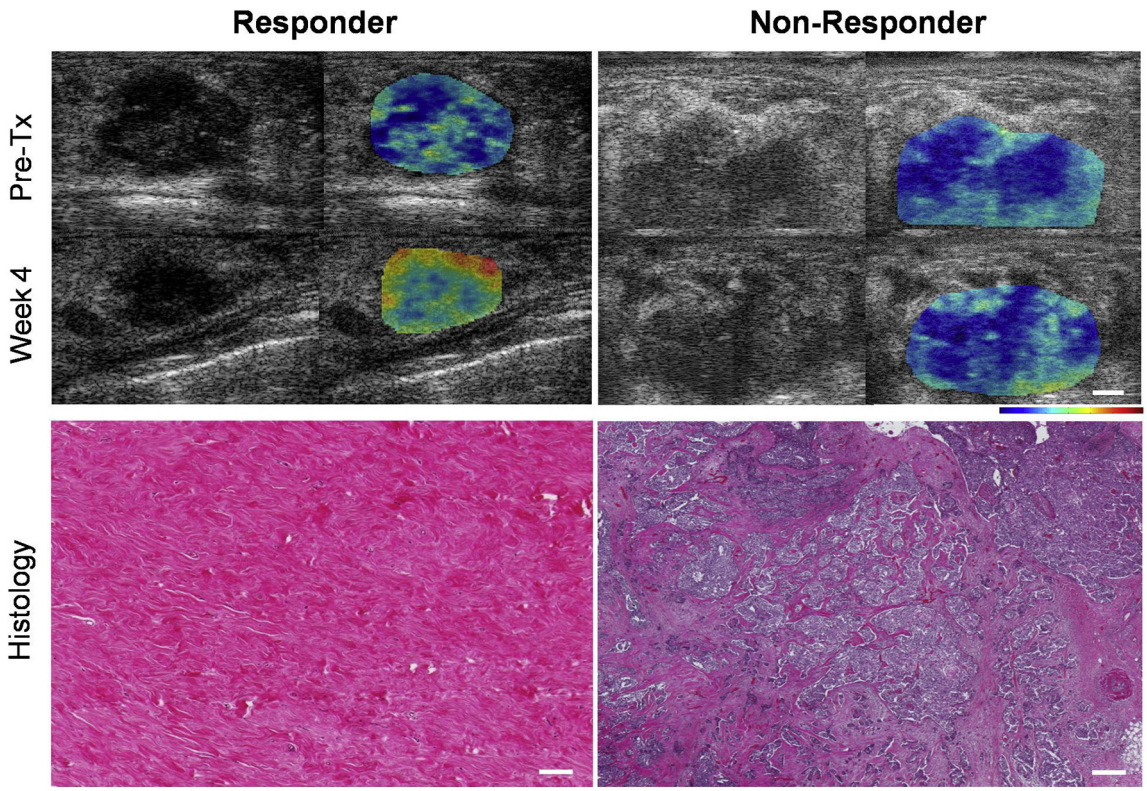

Figure 2. Representative B-mode and mid-band fit parametric images from a responder and non-responder before the start of NAC (Pre-Tx) and after 4 weeks of treatment (top). Representative high magnification light microscope images of whole mount histopathology from responder and non-responder patient samples after treatment completion (bottom). The scale bar in the ultrasound images represents 5 mm. The color bar represents the scale for the mid-band fit parameter of −16 to 18 dB (left to right). The scale bar in the histology images represents 200 microns. Pre-Tx indicates prior to treatment; week 4 indicates images obtained from 4 weeks after the start of treatment with chemotherapy.

set of patients acquired using two different clinical ultrasound systems. The SVM-RBF algorithm predicted treatment responses from independent patients' data acquired on the Ultrasonix system (used above) with accuracies of 82% (sensitivity: 87% and specificity: 50%), 78% (sensitivity: 80% and specificity: 67%), and 88% (sensitivity: 87% and specificity: 100%) at weeks 1, 4, and 8, respectively. For data acquired using the GE clinical system (GE Healthcare, Milwaukee, Wisconsin, USA), the accuracies were 72% (sensitivity: 73% and specificity: 50%), 81% (sensitivity: 84% and specificity: 67%), and 93% (sensitivity: 93% and specificity: 100%) at weeks 1, 4, and 8, respectively. The McNemar test was performed to compare misclassified patients from the Ultrasonix and GE data sets. The *p*-values obtained from the McNemar test (0.62, 0.75, and 0.97 at week 1, 4, and 8, respectively) revealed no significant differences between the classification results obtained from the two clinical systems.

Discussion and Conclusion

In this study, we present the result of a clinical investigation of 100 locally advanced breast cancer patients receiving NAC, whose tumor responses were monitored using quantitative ultrasound and texture analysis techniques with advanced computational algorithms. All algorithms evaluated in this study achieved greater than 60% accuracy in distinguishing non-responding from responding patients. Specifically, the SVM-RBF algorithm achieved the best prediction accuracy at all scan times. The prediction accuracy, sensitivity and specificity of the algorithm in characterizing tumor response were determined through a robust cross-validation approach and the results tested using an independent set of patient data. Previous preclinical studies demonstrated that the pathological changes (nuclear aggregation, fragmentation, and condensation) that occur with the cell death process can modify ultrasonic backscatter properties of tissue at both low and high ultrasound frequency [14,17,28]. This has been further evaluated in this study, in which ultrasound backscatter characteristics were quantified using QUS parameters including MBF, SI, ASD and AAC.

In order to understand the relationship between QUS-derived indicators of tumor microstructure and treatment response, the correlation of change in estimated QUS and texture features with treatment response was investigated. Statistical tests revealed no significant differences between the QUS and texture parameters estimated from responders and non-responders during treatment at weeks 1 and 4 on a single parameter basis. The smallest statistically significant *p*-values observed were with the ΔSS-COR ($P = .0711$) and ΔASD-COR ($P = .087$) parameters at weeks 1 and 4, respectively. The range of scatterer sizes estimated from tumor ultrasound data for the frequency bandwidth of 4−9 MHz was 80−160 μm and those values were comparable with lobule diameters observed from histopathology images. At week 8 after the start of treatment, statistically significant differences were observed mostly in change in backscatter intensity-related parameters such as MBF, SI and AAC and their corresponding texture parameters. These backscatter intensity parameters are strongly related to scatterer number density and scatterer elastic properties. The results indicate that changes in the texture features of QUS parametric maps become apparent at early stages (weeks 1 and 4) after the start of treatment,

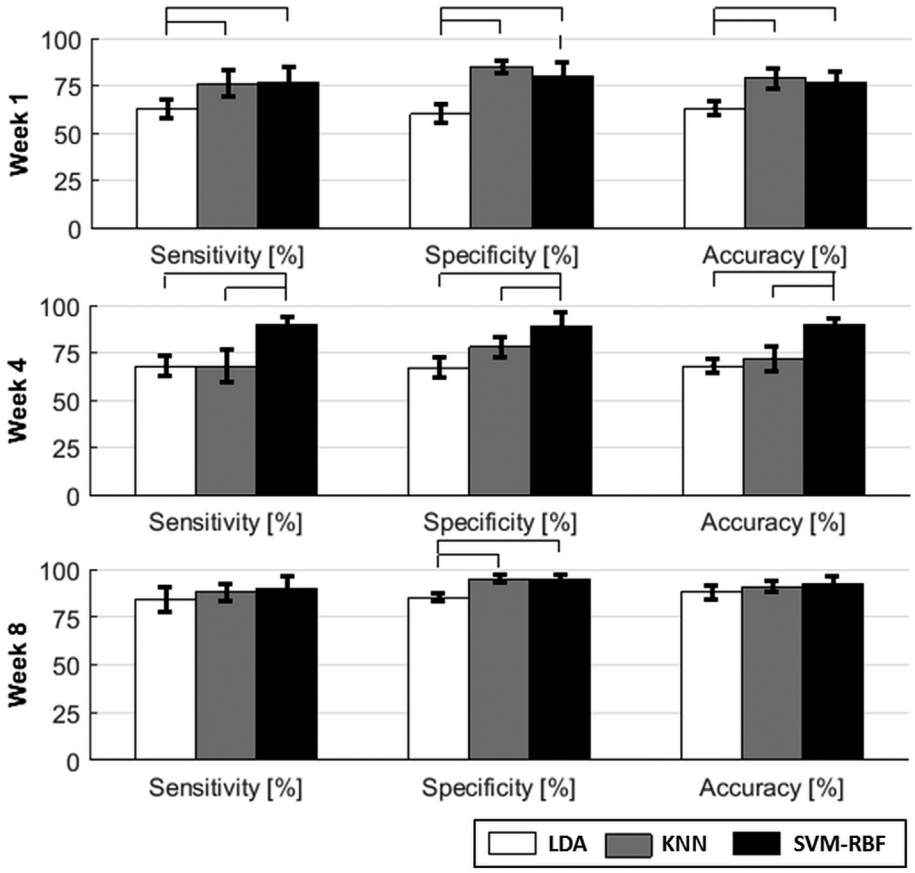

Figure 3. Average classification sensitivity, specificity and accuracy percentages over 11 iterations for LDA, KNN, and SVM-RBF classifiers in differentiating responders and non-responders at weeks 1, 4, and 8 after treatment. Leave-one-subject-out analysis was used for classification. The horizontal connection lines above the bars indicate significant differences between classifiers ($P < .05$) using a paired t-test.

and subsequently result later in dominant changes in the mean values of these QUS parametric maps.

Three multi-feature discriminant algorithms were developed using LDA, KNN, and SVM-RBF classifiers based on changes in QUS and texture features in order to improve the accuracy of response-related group classification. A sequential feature selection method was applied to balanced subsets of training data to obtain optimal feature sets for tumor response classification. Three to four features for each time and method were selected (Table 2). The features selected to differentiate two response classes at weeks 1 and 4 were all texture features, and those selected at week 8 were all mean QUS values, especially ultrasound backscatter intensity parameters. Using an SVM-RBF classifier with a small number of features, the patient responses were classified with cross-validated accuracies of 78%, 90%, and 92% at weeks 1, 4, and 8, respectively. This finding suggests that the development of responses in tumor cells to neoadjuvant chemotherapy is a gradual process. As a working model, at an early response stage (week 1 and 4), tumor microstructure arrangements such as the spacing between lobules can be affected heterogeneously. At a later stage (week 8), in addition to tumor microstructural arrangements, macrostructures such as lobule size, shape and their elastic properties can be affected, and finally, replaced with collagen

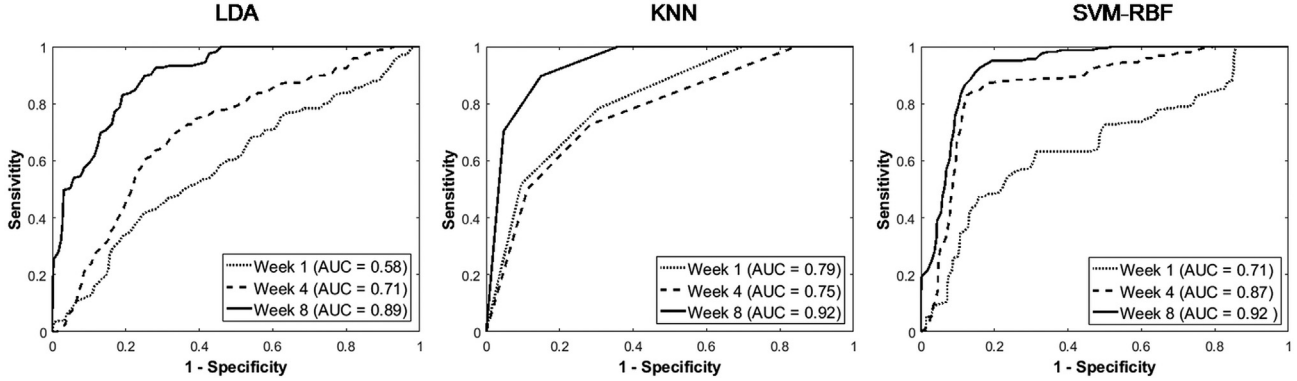

Figure 4. Receiver operator characteristic curves for early tumor response prediction models using LDA, KNN and SVM-RBF obtained by averaging over 11 iterations. Overall, the SVM-RBF algorithm performed best as compared to LDA and KNN.

Table 2. Optimal features selected for tumor response classification using a LDA, KNN and SVM-RBF classifier over 11 iterations at week 1, 4 and 8

Classifier	Week 1	Week 4	Week 8
LDA	Δ MBF-COR	Δ SI	Δ MBF
	Δ MBF-ENE	Δ SS-ENE	Δ ACE
		Δ SS-HOM	
		Δ ASD-COR	
KNN	Δ SS-HOM	Δ ASD-ENE	Δ AAC
			Δ SI
SVM-RBF	Δ SS-COR	Δ SS-CON	Δ MBF
	Δ MBF-ENE	Δ ASD-COR	Δ SAS
	Δ SAS-HOM	Δ SI-HOM	Δ SI
		Δ SI-ENE	

and fibrotic deposition. These findings, and the link to the long-term clinical outcome of these patients suggests that the QUS parameters and texture features investigated in this study have potential for the early prediction of treatment response and long-term survival of patients undergoing neoadjuvant chemotherapy.

Previous studies confirm that pathological response is a prognostic indicator for long-term, disease-free and overall survival [3,4]. This was confirmed in this current study. Quantitative ultrasound biomarkers acquired from cancer patients after the start of treatment indicated links to patient outcomes in terms of progression-free survival. The classification models developed based on combined quantitative ultrasound and texture parameters obtained during weeks 1, 4, and 8 after the start of treatment could differentiate between patient outcomes with good agreement with those based on histopathology, obtained months later after surgery. Such an early insight into patient outcomes could facilitate the decision of changing ineffective therapy to a more effective therapy for treatment-refractory patients or even an earlier shift to salvage therapy, before missing the "therapeutic window" for benefit. Several previous studies have demonstrated significant differences in chemotherapy response rates and survivals of patients with breast cancers of different molecular subtypes including HER2+, triple negative, and ER and/or PR + with HER2- status [4,29,30]. A six-year disease-free survival rate was calculated for patients with these molecular subtypes of breast cancer and was found to be 0.92, 0.82 and 0.84, respectively. However, there was no statistical difference detected between subtypes ($P = .233$). Most likely, a larger cohort of patients with a high enough number of cases for each subtype is required for such a multi-class classification study to achieve a more accurate cross--validated evaluation.

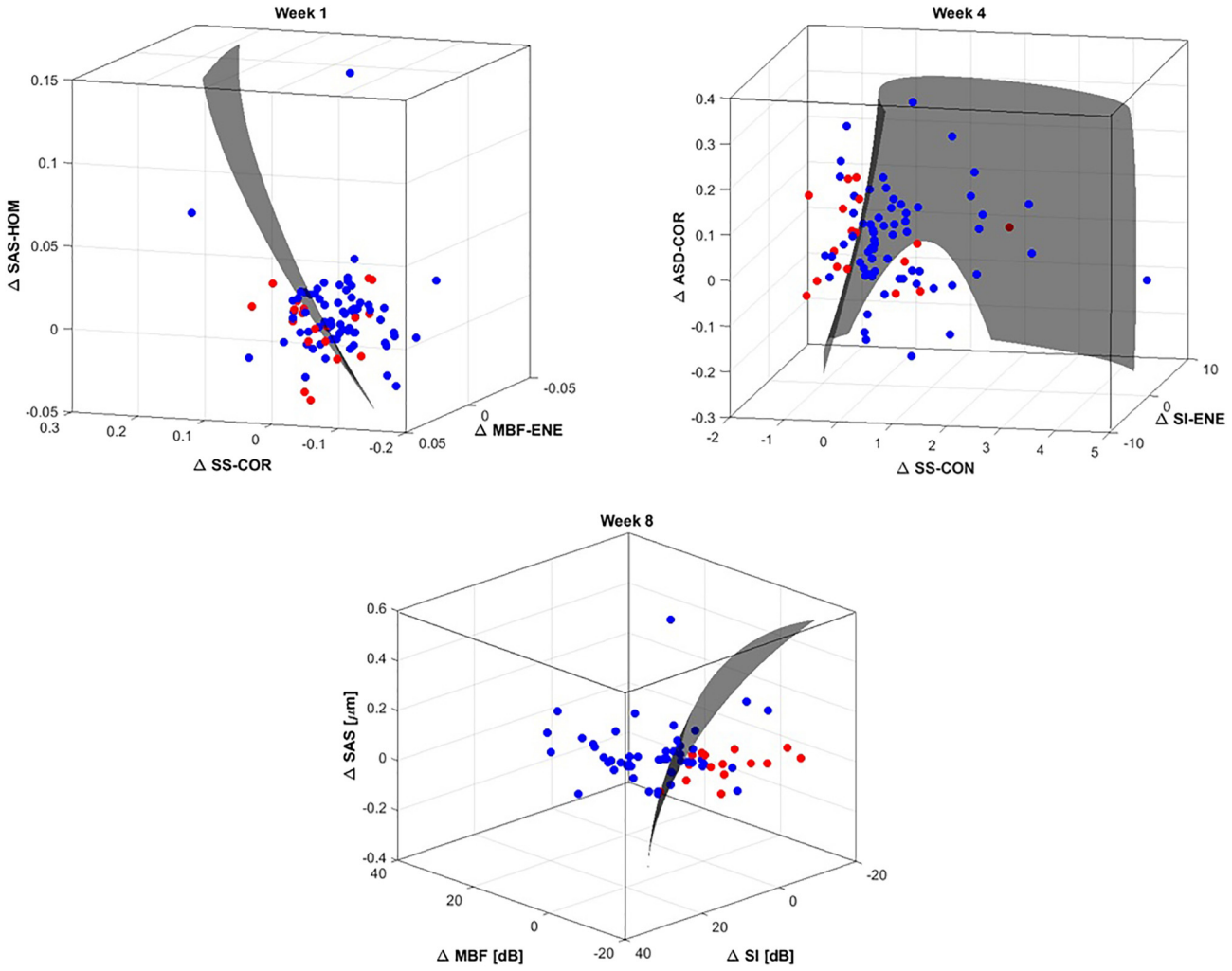

Figure 5. Hyperplanes of decision defined by an SVM-RBF classifier using one of the week 1, 4 and 8 subsets in three-dimensional feature space. Responders and non-responders are represented by blue and red dots, respectively. Four features were selected to develop a tumor response prediction model from the week 4 dataset. For display purposes, the three best features were used in this plot.

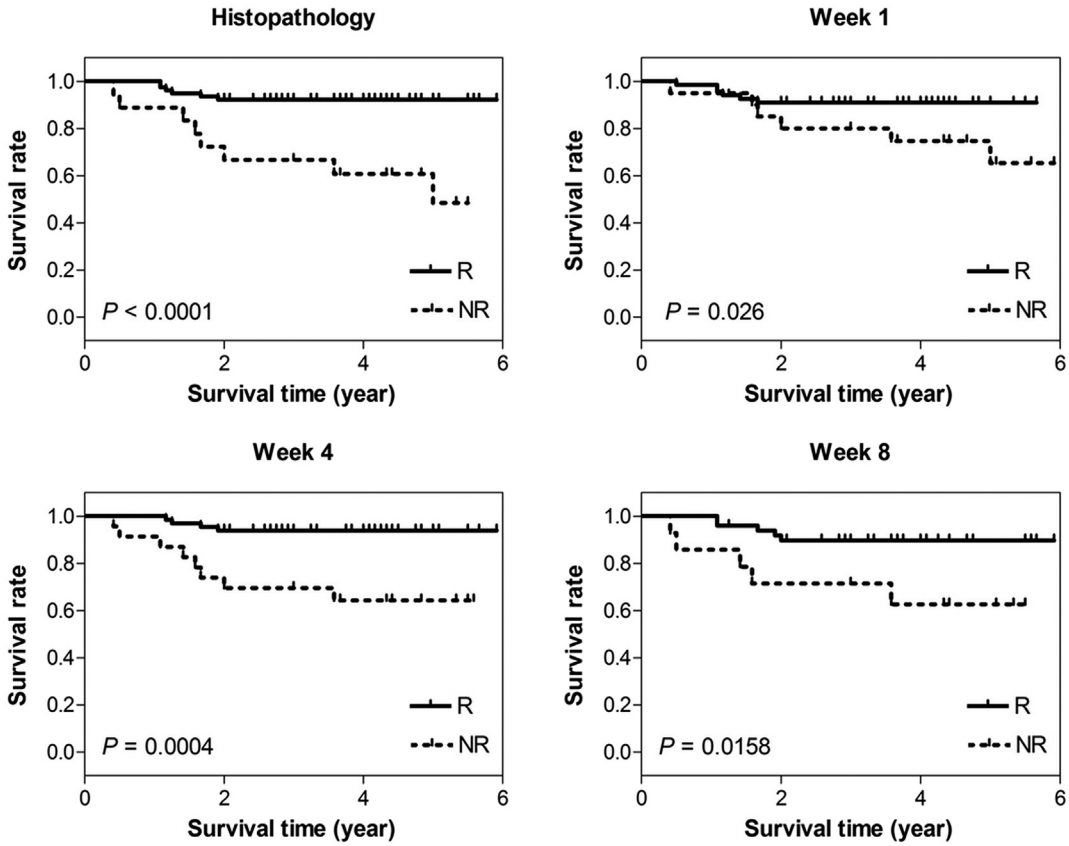

Figure 6. Recurrence-free survival curves for chemotherapy treatment responders and non-responders. Patients were differentiated based on a RECIST score determined from clinical data with validation from histopathology images at post-treatment, and also based on QUS and texture parameters at weeks 1, 4, and 8 using the SVM-RBF algorithm.

Several other functional imaging modalities, including PET, and MRI, are capable of evaluating cancer treatment response within weeks of treatment initiation. These modalities are often costly and require contrast agents to monitor tumor response to treatment. In contrast, the QUS techniques used in this study depend on internal contrast alterations caused by differential scattering due to changes in the acoustic impedance of tumor cells when they respond to treatment. Compared to previous QUS-based tumor response monitoring studies [18–20], the number of patients used in this study was relatively large and differences within tumor response group sample sizes were balanced by randomly subsampling the total population into several subsets before performing classification analysis. To avoid the curse of dimensionality, the number of selected features was less than the number of samples in the training set divided by a factor of 10 based on rule of thumb [31]. Finally, the discriminant model was cross-validated to avoid overfitting.

When comparing all types of classifiers, nonlinear classifiers, such KNN and SVM-RBF performed better than the linear classifier LDA method. This could mean that the true distributions, and consequently the tumor response decision boundaries, were nonlinear or that the data for each group were not easy to differentiate using a simple linear classifier. The accuracy of differentiating binary response groups at a very early stage (week 1) with the KNN classifier was slightly higher than with the SVM-RBF, but not significantly different. However, the overall performance of the SVM-RBF algorithm was best in differentiating response groups at all data acquisition times. This was confirmed in independent testing with an additional separate patient cohort and using data obtained from two different ultrasound systems. One drawback of the SVM-RBF classifier compared to other classifiers is that it requires a more involved computational training process to select the best classifier parameters C and γ using a grid search. However, once a classification model is developed with the SVM-RBF, it is easy to implement in clinical systems. It does not require much space to store the model like a KNN system, which requires an N-dimensional feature space derived from observed data to detect an unknown sample response. Additionally, the SVM classifier is very robust and performs well with outliers since it only uses the most relevant points to define a decision boundary between two groups. Here, the robustness of a model developed by the SVM classifier was demonstrated by comparing the classification results obtained from two different clinical system data sets.

One important question is how this early response prediction tool can be applied in a clinical setting in order to optimize individual patient treatments. For example, it is possible that individual, patient-based, early tumor response could be used to guide changes from ineffective treatments to more effective treatments. For instance, several studies have indicated that patients who do not respond to initial chemotherapy may benefit from additional systemic chemotherapy, preoperative radiation or surgery [32–34], in the case where there is insufficient intra-treatment response. Specifically, in a response-guided neoadjuvant chemotherapy study for breast cancer [33], patients were treated with 2 cycles of docetaxel, doxorubicin, and cyclophosphamide (TAC). The study randomly assigned early non-responders to 4 cycles of TAC, or as an alternative, vinorelbine and capecitabine (NX) before surgery. The results of that study indicated that disease-free survival was greater in early non-responders receiving TAC-NX than in those receiving 6 cycles of TAC. In the

Aberdeen study [32], LABC patients underwent 4 cycles of CVAP (cyclophosphamide / vincristine / doxorubicin / prednisone) chemotherapy treatment. Early and complete responders were randomly assigned to further 4 cycles of CVAP or 4 cycles of docetaxel. In that study, it was reported that patients receiving docetaxel had a significantly higher clinical response rate (94%) than patients receiving CVAP (64%). These studies highlight the need to develop novel tumor response detection algorithms, like the one presented in this study, to facilitate clinical treatment changes. Further work will focus on validating the algorithms presented here with larger population cohorts from multiple different centers.

In conclusion, we report classification algorithms that can detect early tumor response to neoadjuvant chemotherapy with high accuracy, using quantitative ultrasound and texture analysis techniques with advanced computational algorithms. The results were validated in a second test sample set of patient data obtained from two independent ultrasound systems. The results presented in this study imply that these classification models, developed based on quantitative ultrasound biomarkers as early survival-linked surrogates of response to cancer-targeting therapies, could facilitate switching an inefficient treatment regimen to a more effective one on an individual patient-basis, early after starting treatment.

Funding

This project was funded by the Terry Fox Foundation, the Natural Sciences and Engineering Research Council of Canada, and the Canadian Institutes of Health Research. The authors disclose no conflict of interest. Clinical trial number: clinicaltrials.gov NCT00437879. GJC holds the University of Toronto James and Mary Davie Chair in Breast Cancer Imaging and Ablations.

Acknowledgments

The authors would like to thank Karina Quiaoit, Elyse Watkins, and Harini Suraweera for their help with collecting and organizing the clinical data.

Declaration of Competing Interest

The authors declare that they have no competing interests.

References

[1] Giordano SH (2003). Update on locally advanced breast cancer. *Oncologist*, 521–530.

[2] Scholl SM (1994). Fourquet a, Asselain B, Pierga JY, Vilcoq JR, Durand JC, et al. Neoadjuvant versus adjuvant chemotherapy in premenopausal patients with tumours considered too large for breast conserving surgery: preliminary results of a randomised trial: S6. *Eur J Cancer* **30A**, 645–652.

[3] Cleator SJ, Makris A, Ashley SE, Lal R, and Powles TJ (2005). Good clinical response of breast cancers to neoadjuvant chemoendocrine therapy is associated with improved overall survival. *Ann Oncol* **16**, 267–272.

[4] Andre F, Mazouni C, Liedtke C, Kau S-W, Frye D, and Green M, et al (2008). HER2 expression and efficacy of preoperative paclitaxel/FAC chemotherapy in breast cancer. *Breast Cancer Res Treat* **108**, 183–190.

[5] Ogston KN, Miller ID, Payne S, Hutcheon AW, Sarkar TK, and Smith I, et al (2003). A new histological grading system to assess response of breast cancers to primary chemotherapy: Prognostic significance and survival. *Breast* **12**, 320–327.

[6] Kim S-J, Kim S, Lee ES, Ro J, and Kang SH (2004). Predictive value of [18F] FDG PET for pathological response of breast cancer to neo-adjuvant chemotherapy. *Ann Oncol* **15**, 1352–1357.

[7] Belli P, Costantini M, Ierardi C, Bufi E, Amato D, and Mule' A, et al (2011). Diffusion-weighted imaging in evaluating the response to neoadjuvant breast cancer treatment. *Breast J* **17**, 610–619.

[8] O'Flynn E a M, DeSouza NM (2011). Functional magnetic resonance: biomarkers of response in breast cancer. *Breast Cancer Res* **13**, 204.

[9] Jiang L, Weatherall PT, McColl RW, Tripathy D, and Mason RP (2013). Blood oxygenation level-dependent (BOLD) contrast magnetic resonance imaging (MRI) for prediction of breast cancer chemotherapy response: A pilot study. *J Magn Reson Imaging* **37**, 1083–1092.

[10] Padhani AR, Yarnold JR, Regan J, and Husband JE (2002). Magnetic resonance imaging of induration in the irradiated breast. *Radiother Oncol* **64**, 157–162.

[11] Baek HM, Yu HJ, Chen JH, Nalcioglu O, and Su MY (2008). Quantitative correlation between (1)H MRS and dynamic contrast-enhanced MRI of human breast cancer. *Magn Reson Imaging* **26**, 523–531.

[12] Feleppa EJ, Liu T, Kalisz A, Shao MC, Fleshner N, and Reuter V, et al (1997). Ultrasonic spectral-parameter imaging of the prostate. *Int J Imaging Syst Technol* **8**, 11–25.

[13] Oelze ML, and O'Brien WD (2006). Application of three scattering models to characterization of solid tumours in mice. *Ultrason Imaging* **28**, 83–96.

[14] Banihashemi B, Vlad R, Debeljevic B, Giles A, Kolios MC, and Czarnota GJ (2008). Ultrasound imaging of apoptosis in tumour response: novel preclinical monitoring of photodynamic therapy effects. *Cancer Res* **68**, 8590–8596.

[15] Czarnota GJ, Karshafian R, Burns PN, Wong S, Al Mahrouki A, and Lee JW, et al (2012). Tumour radiation response enhancement by acoustical stimulation of the vasculature. *Proc Natl Acad Sci U S A* **109**, E2033–E2041.

[16] Sadeghi-Naini A, Papanicolau N, Falou O, Tadayyon H, Lee J, and Zubovits J, et al (2013). Low-frequency quantitative ultrasound imaging of cell death in vivo. *Med Phys* **40**:82901.

[17] Vlad RM, Brand S, Giles A, Kolios MC, and Czarnota GJ (2009). Quantitative ultrasound characterization of responses to radiotherapy in cancer mouse models. *Clin Cancer Res* **15**, 2067–2075.

[18] Sannachi L, Tadayyon H, Sadeghi-Naini A, Tran W, Gandhi S, and Wright F, et al (2015). Non-invasive evaluation of breast cancer response to chemotherapy using quantitative ultrasonic backscatter parameters. *Med Image Anal* **20**.

[19] Sadeghi-Naini A, Papanicolau N, Falou O, Zubovits J, Dent R, and Verma S, et al (2013). Quantitative Ultrasound Evaluation of Tumour Cell Death Response in Locally Advanced Breast Cancer Patients Receiving Chemotherapy. *Clin Cancer Res* **19**, 2163–2174.

[20] Sadeghi-Naini A, Sannachi L, Pritchard K, Trudeau M, Gandhi S, and Wright FC, et al (2014). Early prediction of therapy responses and outcomes in breast cancer patients using quantitative ultrasound spectral texture. *Oncotarget* **5**, 3497–3511.

[21] Yao LX, Zagzebski JA, and Madsen EL (1990). Backscatter coefficient measurements using a reference phantom to extract depth-dependent instrumentation factors. *Ultrason Imaging* **12**, 58–70.

[22] Lizzi FL, King DL, Rorke MC, Hui J, Ostromogilsky M, and Yaremko MM, et al (1988). Comparison of theoretical scattering results and ultrasonic data from clinical liver examinations. *Ultrasound Med Biol* **14**, 377–385.

[23] Wear KA, Wagner RF, Insana MF, and Hall TJ (1993). Application of autoregressive spectral analysis to cepstral estimation of mean scatterer spacing. *IEEE Trans Ultrason Ferroelectr Freq Control* **40**, 50–58.

[24] Insana MF, and Hall TJ (1990). Parametric ultrasound imaging from backscatter coefficient measurements: image formation and interpretation. *Ultrason Imaging* **12**, 245–267.

[25] Haralick RM, Shanmugam K, and Dinstein I (1973). Textural Features for Image Classification. *IEEE Trans Syst Man Cybern IEEE* **3**, 610–621.

[26] Cortes C, and Vapnik V (1995). Support-Vector Networks. *Mach Learn* **20**, 273–297.

[27] Chang C-C, and Lin C-J (2011). LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* **2**.

[28] Czarnota GJ, Kolios MC, Abraham J, Portnoy M, Ottensmeyer FP, and Hunt JW, et al (1999). Ultrasound imaging of apoptosis: high-resolution non-invasive monitoring of programmed cell death in vitro, in situ and in vivo. *Br J Cancer* **81**, 520–527.

[29] Liedtke C, Mazouni C, Hess KR, André F, Tordai A, and Mejia JA, et al (2008). Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. *J Clin Oncol* **26**, 1271–1281.

[30] Von Minckwitz G, Untch M, Blohmer JU, Costa SD, Eidtmann H, and Fasching PA, et al (2012). Definition and impact of pathologic complete response on prognosis after neoadjuvant chemotherapy in various intrinsic breast cancer subtypes. *J Clin Oncol* **30**, 1796–1804.

[31] Jain AK, and Chandrasekaran B (1982). 39 Dimensionality and sample size considerations in pattern recognition practice. *Handbook of Statistics* **2**, 835–855.

[32] Heys SD, Hutcheon AW, Sarkar TK, Ogston KN, Miller ID, and Payne S, et al (2002). Neoadjuvant docetaxel in breast cancer: 3-year survival results from the Aberdeen trial. *Clin Breast Cancer* **3**, S69–S74.

[33] von Minckwitz G, Blohmer JU, Costa SD, Denkert C, Eidtmann H, and Eiermann W, et al (2013). Response-guided neoadjuvant chemotherapy for breast cancer. *J Clin Oncol* **31**, 3623–3630.

[34] Masuda N, Lee S-J, Ohtani S, Im Y-H, Lee E-S, and Yokota I, et al (2017). Adjuvant Capecitabine for Breast Cancer after Preoperative Chemotherapy. *N Engl J Med* **376**, 2147–2159.