npj | precision oncology

Article

Published in partnership with The Hormel Institute, University of Minnesota

https://doi.org/10.1038/s41698-025-01134-x

Quantitative ultrasound imaging for predicting response and guiding personalized neoadjuvant chemotherapy in breast cancer: randomized phase 2 clinical trial results

Daniel Moore-Palhares[1,2,3,13], David Alberico[3,13], Adrian Wai Chan[1,2,3,13], Daniel DiCenzo[3], Lakshmanan Sannachi[3], Archya Dasgupta[1,2,3], Joyce Yip[3], Maria Lourdes Anzola Pena[3], Sonal Gandhi[4,5], Rossanna Pezo[4,5], Andrea Eisen[4,5], Katarzyna J. Jerzak[4,5], Carlos A. Carmona Gonzalez[4,5], Ellen Warner[4], Frances C. Wright[6,7], Nicole Look-Hong[6,7], Amanda Roberts[6,7], Ali Sadeghi-Naini[1,3,8,9], Belinda Curpen[10,11], Mia Skarpathiotakis[10,11], Carrie Betel[10,11], Michael C. Kolios[12], Maureen Trudeau[4,5] & Gregory J. Czarnota[1,2,3,8] ✉

Quantitative ultrasound (QUS) detects early tumor microstructural changes during neoadjuvant chemotherapy (NAC), enabling personalized treatment adaptation. This study assessed the accuracy of machine learning models using serial QUS data to predict treatment response and evaluated their feasibility for guiding treatment personalization. This single-center, phase 2 randomized controlled trial (clinicaltrials.gov NCT04050228, Dec/2019) enrolled stage II–III breast cancer patients planned for standard NAC. QUS imaging was performed at baseline and week 4, with the latter used for response prediction. Patients were randomized 1:1 to standard or experimental arms, stratified by hormone receptor status. In the standard arm, oncologists were blinded to QUS results. In the experimental arm, predictions were disclosed to allow treatment modification at week 4. Final response was determined histopathologically (>30% tumor reduction or <5% cellularity). Between June 2018 and September 2023, 146 patients were enrolled, and 120 randomized (standard: 57, experimental: 63). Response rates were 93.0% (standard) and 96.8% (experimental). The model achieved 92% accuracy, 83% sensitivity, 93% specificity, and 99% positive predictive value. In the experimental arm, 8/63 patients were predicted non-responders, with 4 undergoing treatment modification. QUS-based machine learning enables accurate early response prediction and supports adaptive treatment strategies in future trials.

Neoadjuvant chemotherapy is the standard treatment approach for patients with locally advanced and inoperable breast cancer and is the preferred option for those with human epidermal growth factor receptor 2 (HER2) positive or 'triple negative' operable breast cancer, as they are known to have an increased risk of distant recurrence[1–3]. One benefit of neoadjuvant treatment is tumour down-staging, which can improve operability, facilitate breast conservation, and allows for the de-intensification of axillary treatment in cases where there is a response in involved lymph nodes. In addition, neoadjuvant chemotherapy allows for the in vivo assessment of treatment response, enabling the identification of patients who have residual disease after systemic therapy, which guides further adjuvant treatment decisions. This is particularly important, as randomized controlled trials

A full list of affiliations appears at the end of the paper. ✉e-mail: gregory.czarnota@sunnybrook.ca

The Hormel Institute
University of Minnesota

have demonstrated that intensified adjuvant systemic therapy in patients with residual disease significantly reduces the risk of relapse and improves survival, especially in those with triple-negative breast cancer (TNBC) or HER2-positive breast cancer[4,5].

While neoadjuvant treatment provides valuable prognostic information at the individual patient level, it may take multiple treatment cycles to identify patients who are not responding to chemotherapy, and decisions regarding treatment intensification are typically deferred until after chemotherapy completion and tumour resection. Currently, no clinical, radiological, or pathological test can reliably predict the response of an individual specific tumour before or shortly after treatment begins. However, a test capable of predicting response very early in treatment course could enable personalized adjustments, such as sparing patients from ineffective therapy and facilitating a timely switch to more effective treatment.

Serial quantitative ultrasound (QUS) imaging emerges as an appealing modality for monitoring tumour response due to its ability to non-invasively detect early changes in tissue microstructural properties (such as arrangement, density, and elasticity) following chemotherapy administration[6–13]. These changes can serve as early indicators of response or resistance to therapy. Previously, a prospective study was conducted in which up to 200 breast cancer patients undergoing neoadjuvant treatment were scanned with serial ultrasound imaging before and during chemotherapy[6,7]. Utilizing QUS texture analysis features extracted from ultrasound image radiofrequency (RF) data of the primary tumour, machine learning algorithms were developed that predicted the response to treatment within the first month of neoadjuvant chemotherapy with 90% accuracy[6,7].

In the current phase 2 clinical trial, these machine learning algorithms were used to predict tumour response at the 4th week of neoadjuvant treatment[14]. Patients were randomized to have their response prediction results disclosed (experimental arm) or not disclosed (observational arm) to the treating medical oncologist, with modification of neoadjuvant treatment allowed in the experimental arm at the medical oncologist's discretion. The aim of this study was to assess the accuracy of the prediction model and the feasibility of using these machine learning models in clinical practice to adapt treatment management early in the treatment course.

Results

Studied population and response rate

A total of 188 patients were screened for study eligibility, with 146 patients ultimately enrolled in the study and randomized evenly into two arms, with 73 patients in each, between June 2018 and September 2023. Sixteen patients in the observational arm and ten in the experimental arm were excluded, primarily due to the unavailability of complete image data, leaving 57 patients in the observational arm and 63 patients in the experimental arm for analysis. Patients were followed until February 2024. The CONSORT diagram is shown in Fig. 1.

Baseline characteristics were similar in both arms (Table 1, Supplementary Tables 1 and 2). The median age in both arms was 50 years (24–80). The median primary tumour size was 3.7 cm (1.7–12.0 cm) in the observational arm and 4.7 cm (1.8–11.2 cm) in the experimental arm. The most common molecular subtype was estrogen receptor (ER)-positive, progesterone receptor (PR)-positive, and HER2-negative, representing 56.1% of the patients in the observational arm and 46% in the experimental arm. The most common neoadjuvant treatment regimen among all patients analysed was dose-dense AC-T (±H) (63%), followed by FEC-D (±H) (28%) and KEYNOTE 522 (6%). The response rate was 93% in the observational arm and 97% in the experimental arm. When the response was assessed according to RECIST 1.1 criteria, the rates of complete response, partial response, stable disease, and progressive disease were 30%, 49%, 14%, and 7% in the observational arm and 27%, 57%, 13%, and 3% in the experimental arm, respectively.

Representative QUS results are presented in Fig. 2. The figure demonstrates typical changes in select QUS parameters colour-coded and overlaid on B-mode ultrasound images for a responding patient and a typical non-responding patient. Results presented include those from before treatment and at 4 weeks after the start of treatment. As is visually evident, there were obvious changes in the parametric images for the responding patient, whereas the non-responding patient results were invariant. Supplementary Figs. 1 and 2 present a gallery of

Fig. 1 | Trial consort diagram.

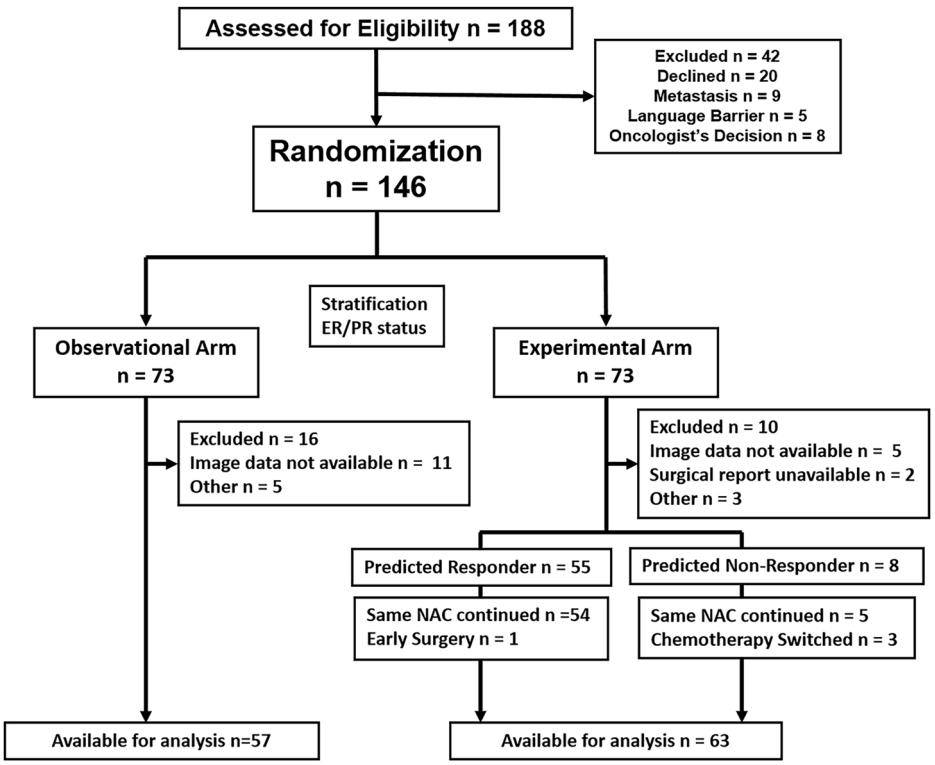


Table 1 | Patient, disease, and treatment characteristics for patients in observational and experimental arms

Characteristic	Observational arm ($n = 57$)	Experimental arm ($n = 63$)	Total ($n = 120$)
Age (years)			
Median (Range)	51 (29–73)	50 (24–80)	50 (24–80)
Side			
Right	31 (54.4%)	31 (49.2%)	62 (51.7%)
Left	25 (43.9%)	32 (50.8%)	57 (47.5%)
Bilateral	1 (1.8%)	0 (0.0%)	1 (0.8%)
Initial tumour size (cm)			
Median (Range)	3.7 (1.7–12.0)	4.7 (1.8–11.2)	4.0 (1.7–12.0)
Molecular markers			
ER/PR + HER-	32 (56.1%)	29 (46.0%)	61 (50.8%)
ER/PR + HER+	6 (10.5%)	7 (11.1%)	13 (10.8%)
ER/PR- HER+	8 (14.0%)	9 (14.3%)	17 (14.2%)
ER/PR/HER -	11 (19.3%)	18 (28.6%)	29 (24.2%)
Histological type			
IDC	50 (87.7%)	58 (92.1%)	108 (90.0%)
ILC	3 (5.3%)	1 (1.6%)	4 (3.3%)
Other	4 (7.0%)	4 (6.3%)	8 (6.7%)
Chemotherapy			
AC-T	26 (45.6%)	31 (49.2%)	57 (47.5%)
AC-TH	9 (15.8%)	9 (14.3%)	18 (15.0%)
FEC-D	10 (17.5%)	13 (20.6%)	23 (19.2%)
FEC-DH	4 (7.0%)	6 (9.5%)	10 (8.3%)
KEYNOTE-522	4 (7.0%)	3 (4.8%)	7 (5.8%)
Other	4 (7.0%)	1 (1.6%)	5 (4.2%)
Treatment response			
Responder	53 (93.0%)	61 (96.8%)	114 (95.0%)
Non-Responder	4 (7.0%)	2 (3.2%)	6 (5.0%)

additional patient data for additional representative responding and non-responding patients.

Response prediction

Figure 3 presents classification results for all patients showing their Class Score, representing their response-prediction score and the associated hyperplane distance. The hyperplane distance is the distance from a patient's score and the mean of the associated training data for a responder or non-responder. Figure 3 highlights with colour-coding those patients who had a "weak" clinical tumour response—in which either there was a diminishment in tumour size but not necessarily cellularity or, more commonly, there was a diminishment in cellularity to less than 10% but not necessarily a response-related (>30%) diminishment in tumour size. Many of these patients have class scores close to an indeterminate score—the threshold at which responders cannot be well differentiated from non-responders (Fig. 3).

A breakdown of response predictions by patient group and study arm is presented in Table 2. The performance metrics of the prediction model are summarized in Table 3, with an accuracy of 92%, sensitivity of 93%, specificity of 83%, positive predictive value of 99%, and negative predictive value of 38% in all but switched patients or performance metrics of accuracy of 93%, sensitivity of 93%, specificity of 100%, positive predictive value of 100%, and negative predictive value of 50% (Supplementary Fig. 3, Tables 2 and 3). Figure 4 delineates the anticipated and actual reactions of each patient, categorized by study arm and therapy modification, while Table 4 offers descriptions of the misclassified individuals. Simplified versions showing class scores and Sankey figures using only two patient classes are presented in Supplementary Figs. 4 and 5.

Misclassified patients

Data indicated that most misclassified patients had unusual tumour features such as tumour necrosis (1 patient), mucinous features (2) patients, or were patients who demonstrated a weak tumour response to treatment (5 patients). Figure 4 also presents the associated histopathology for patients and there did not appear to be any correlation between misclassification of results and histological subtype.

Treatment modification

Eight patients in the experimental arm were predicted to be non-responders, and ultimately, 4 (50%) had their treatment modified during the course of neoadjuvant chemotherapy. One patient prematurely switched from AC to T after 2 cycles and then underwent surgery after completing all 4 cycles of biweekly paclitaxel; another switched from AC to T after 2 cycles and subsequently underwent surgical resection after a single cycle of biweekly paclitaxel; and the third switched from AC to T after 3 cycles and underwent surgical resection after 11 cycles of weekly paclitaxel. On the final specimen, they were found to have a tumour response. In addition, one patient in the experimental arm was predicted to be a responder (but close to being indeterminate near the threshold for discriminating response and non-response), however, the medical oncologist decided to stop FEC-D prematurely based on the class score being close to indeterminate and proceed with early resection, and the patient achieved a response as previously predicted. The details are provided in Fig. 5 and Supplementary Table 3. In the observational arm, 7 patients were predicted to be non-responders. Of these, 4 were actual non-responders, and 3 were "weak" responders. In the experimental adapted arm, 3 patients that were predicted to be non-responders had their treatment modified, and either tumours responded ($n = 3$) or, in the case of a responder predicted to be near borderline indeterminate, had their tumour removed surgically early ($n = 1$) despite responding, with that patient then undergoing different adjuvant therapy afterwards. In the Experimental non-adapted arm, there were 5 patients predicted to be non-responders. One ($n = 1$) remained a non-responder, one ($n = 1$) was an actual "weak" responder. Three ($n = 3$) were actual responders, but their response prediction classification as non-responders was borderline.

Prediction class scores versus size changes

A comparison was made between class scores for patients and early size changes (assessed at week 4) of chemotherapy contrasting sizes at that point with those immediately preceding the initiation of chemotherapy. That data suggested that responders (both complete responders and partial responders) generally exhibited a reduction in the tumour size as well as a predictive class score. Most non-responders demonstrated an initial diminishment in tumour size similar to responders (Fig. 6). The latter changed during therapy, as indicated by comparing tumour size changes many months later with sizes immediately before surgery compared to tumour sizes measured immediately before starting chemotherapy (Supplementary Fig. 6). Tumour size changes were mostly significant between different groups pre-operatively and amongst fewer groups at week 4 after starting chemotherapy (data provided in Tables 5 and 6 and Supplementary Tables 4 and 5). At 4 weeks after starting chemotherapy the Spearman correlation coefficient was 0.17, indicating a weak correlation between tumour size change and the computed class score but was not significant with a p-value of 0.07. After chemotherapy, a Spearman correlation coefficient for the relationship between tumour size change and quantitative ultrasound class score was 0.07 with a p-value of 0.48. In order to better compare QUS versus size changes alone in terms of predictive correlations, a group-wise analysis was conducted. At week 4 no size changes for different response groups (Table 5, supplementary table 4) were statistically significant. However, for QUS predictors class scores for certain comparisons

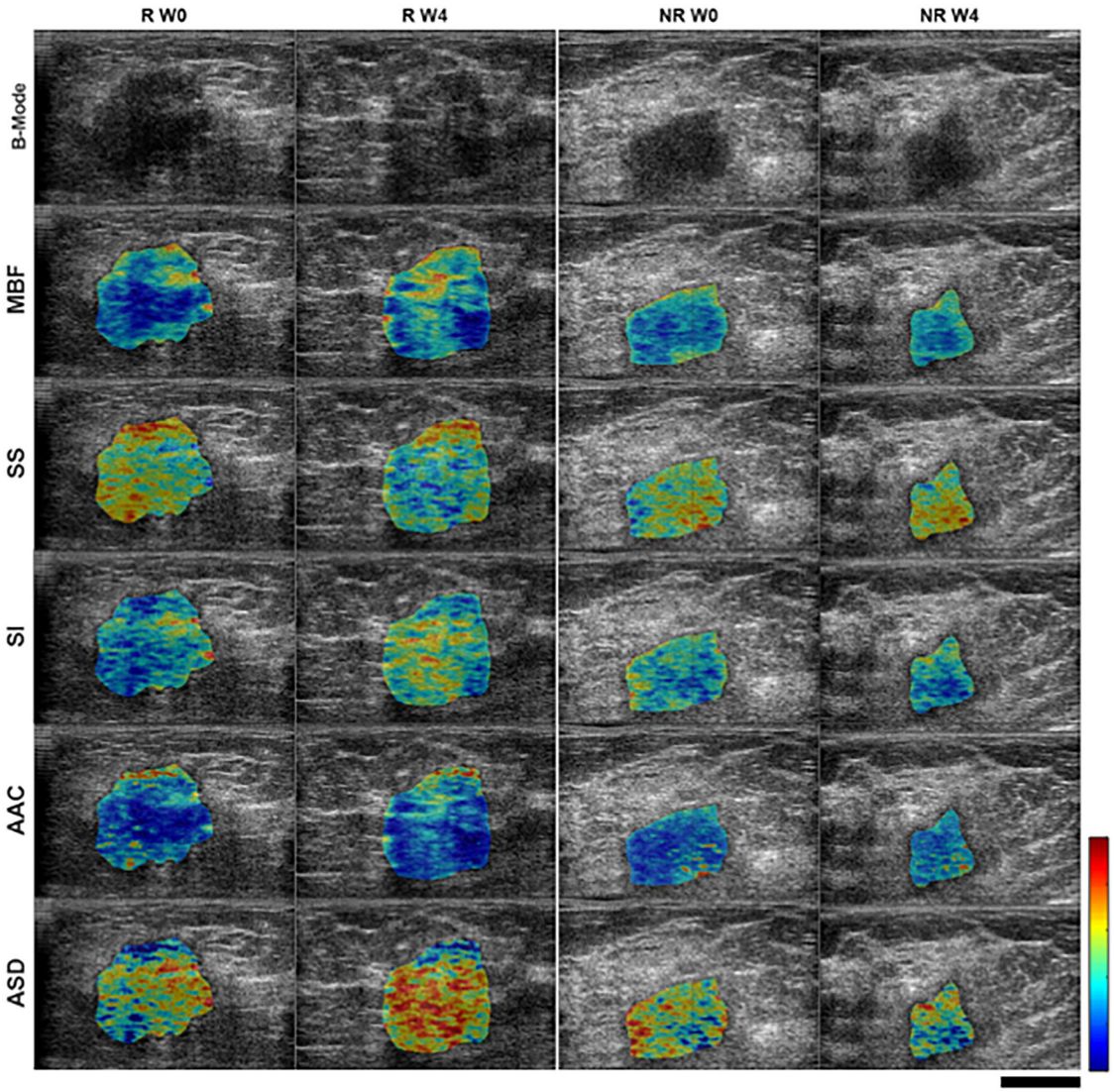


Fig. 2 | B-mode imaging and corresponding QUS-parametric maps (colour coded) at different experimental times (pre-treatment or week 0 (W0), and week 4 of neoadjuvant chemotherapy (w4)) for one representative patient from responder (R) and non-responder (NR) groups. The scale bar represents a length of 2 cm. The colour bar indicates the range of parametric colour overlay: MBF range was from −9.6 dB to 34.0 dB, SS range was from −5.7 dB/MHz to 1.6 dB/MHz, SI range from −7.3 dB to 49.0 dB, AAC range was from 20.2 db/cm^3 to 81.6 db/cm^3, ASD range was from 40 to 171 μm. MBF mid-band fit, SS spectral slope, SI spectral intercept, AAC average acoustic concentration, ASD average spectral diameter.

were statistically significantly different, indicating quantitative ultrasound-based class score were predictive of response, not size changes. Only pre-operatively, many months later, did size changes become more evident and were statistically significantly different in several comparisons between response groups (Table 6, Supplementary Table 5).

Discussion

This phase 2 randomized controlled trial demonstrated the feasibility of using QUS-guided radiomics to predict breast cancer response to neoadjuvant chemotherapy. Our machine-learning algorithms were highly accurate (92%), sensitive (93%), and specific (83%). These findings indicated that serial QUS, combined with radiomic-based machine learning, can be used as a valuable non-invasive tool for monitoring therapeutic effects early during treatment and predicting at the start of neoadjuvant chemotherapy course which patients will respond to treatment.

Over the past decade, the field of radiomics, which involves the quantitative analysis of medical images, has shown great promise in developing biomarkers for prognosis and prediction in oncology. This study utilized a response-monitoring model based on a Support Vector Machine with a Radial Basis Function (SVM-RBF) algorithm to monitor early response and predict ultimate treatment responses. The classifier was built using four texture features from QUS parametric images acquired four weeks after the start of 4–5 months of neoadjuvant chemotherapy from a separate cohort of up to 200 patients, separate from the patients enrolled in this study[7,15,16]. This algorithm had previously demonstrated a cross-validated accuracy of 90%[7], and the performance using the entirely independent separate dataset in the work here achieved an accuracy of 92%.

This study is the first to investigate using a QUS radiomics-guided approach in a relatively large cohort of patients for personalizing neoadjuvant treatment. Traditionally, metabolic imaging techniques like positron emission tomography (PET) have been used to predict early responses to neoadjuvant chemotherapy. For instance, in the AVATAXHER phase 2 randomized trial, fluorodeoxyglucose (FDG) scans were conducted before the second cycle of neoadjuvant chemotherapy in patients with HER2-positive breast cancer[17]. PET-identified non-responders were then randomly assigned to either intensify therapy with the addition of bevacizumab or to continue with the existing docetaxel and trastuzumab regimen. The addition of bevacizumab led to higher pathological complete response rates

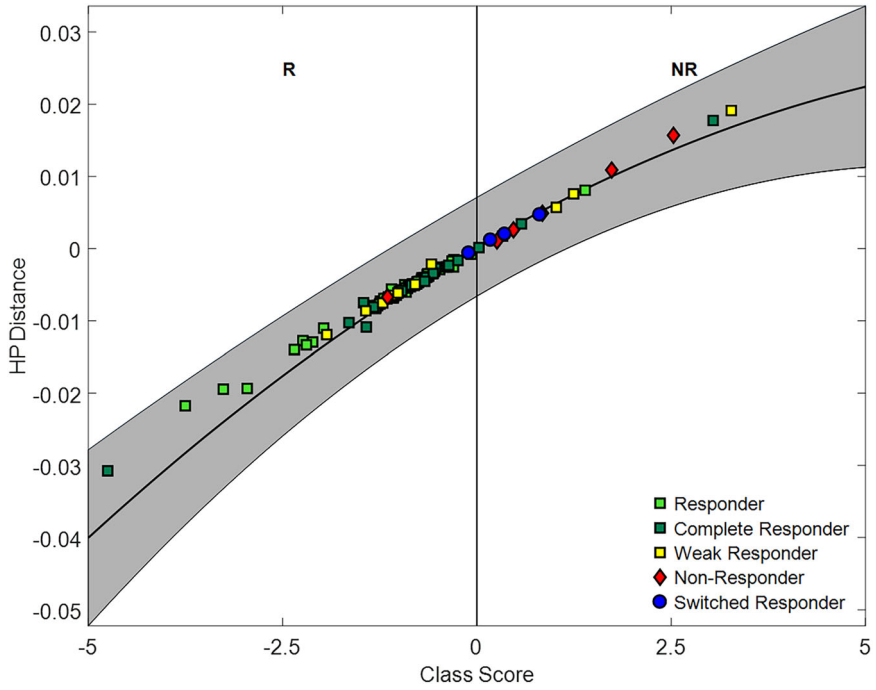


Fig. 3 | Detailed class score plot. Individual patient predictions based on predictor class scores at week 4 for patient response. R indicates the zone (negative class score) for predicted response, and NR indicates the zone (+ve class score) for non-response. Responder, complete responder and non-responder are defined as per RECIST criteria. Weak responders were borderline in regards to size changes (±5%) or had significantly diminished cellularity (5% or less) but did not meet size criteria.

Table 2 | Patient predicted and actual responses to neoadjuvant chemotherapy

Response		Study arm		
Predicted	**Actual**	**Observational**	**Experimental—non-adapted**	**Experimental—adapted**
Non-Responder	Non-Responder	4	1	0
Non-Responder	Responder	4	4	3
Responder	Responder	49	53	1
Responder	Non-Responder	0	1	0

True Positives: 102; False Negatives: 8; True Negatives: 5; False Positives: 1.

Table 3 | Classifier performance of QUS-radiomics model at week 4 of neoadjuvant chemotherapy for observational patients only (*n* = 56) and for entire population except switched patients (*N* = 116)

Parameter	Observational patients (*N* = 56)		Entire population except switched patients (*N* = 116)	
	Value	**95% CI**	**Value**	**95% CI**
Sensitivity	0.93	(0.81, 0.97)	0.83	(0.42, 0.98)
Specificity	1.00	(0.40, 1.00)	0.93	(0.86, 0.96)
Positive predictive value	1.00	(0.93, 1.00)	0.99	(0.94, 1.00)
Negative predictive value	0.50	(0.28, 0.72)	0.38	(0.18, 0.64)
Accuracy	0.93	(0.83, 0.98)	0.92	(0.86, 0.96)

Confidence intervals were calculated using the Agresti-Coull interval.
CI Confidence interval.

(44% versus 24%), highlighting the potential for personalized early treatment intensification. Compared to PET, QUS offers the benefits of being a lower-cost, faster, and portable imaging modality with fewer technical challenges, making it a promising option for implementation in clinical practice. However, PET can be more readily used for whole-body imaging of responses to treatment. Other modalities, including CT and MRI, have also been investigated for therapy response prediction and monitoring[18].

The results observed in this trial reflect QUS's capability to identify and quantify early microscopic changes in tissue microstructural properties following chemotherapy administration[6–13]. Most patients in the development cohort[7] and this clinical trial responded to chemotherapy, allowing our models to be extensively trained to identify responders, for the continuance of chemotherapy. As a consequence, the model correctly identified 102 of the 103 patients predicted to be responders in this trial, demonstrating that the methodology rarely misclassifies responders as non-responders. However, due to the small number of non-responders in both cohorts, the model was less effective at identifying non-responders. Consequently, the current study correctly identified only 5 of the 13 patients predicted to be non-responders, yielding a negative predictive value of 38%. The majority of misclassified non-responders were "weak" responders with a predictive class score close to indeterminate. In addition, the predictions were based on ultrasound data obtained during the first phase of chemotherapy. Most patients who responded during their first phase of chemotherapy (anthracycline-based) continued to respond during their second phase (taxane-based). There, however, may be a subset of non-responders who do not respond during their first phase but do so during their second chemotherapy phase. This may be responsible for the negative-predictive value results and highlights the limitations of binary classification models in complex clinical scenarios, where some patients show mixed or incomplete responses to therapy.

Additional research conducted outside this trial has enabled the differentiation of patients exhibiting a partial response from those with a complete response at 4 weeks after the start of therapy. It may be possible,

Fig. 4 | Trial Sankey diagram. Sankey diagram for predicted response at week 4 using QUS-radiomics model with the final response on an individual patient basis. Individual tiles indicate response type (actual (A) or predicted (P)) and tumour histopathological type. In the experimental arm, 59/63 patients were not adapted based on information provided to their oncologist, whereas 4/63 were adapted. Colour tiles in the first two columns indicate complete responders, responders, and non-responders as per RECIST. Weak responders were borderline in regards to size changes (±5%) or had significantly diminished cellularity (5% or less) but did not meet size criteria. The third column indicates tumour histopathological type (invasive ductal carcinoma, mucinous type, invasive lobular carcinoma, ductal carcinoma in situ in over 50% of tumour volume, micropapillary type and other).

Table 4 | Summary of misclassified patients(a)

Patient number	Arm	W4 Prediction	Response	Notes
O-47	O	NR	R	Complete response. Prediction accuracy 54.5%, prediction class score close to zero.
O-48	O	NR	R	Size reduction 26%, post-treatment cellularity 16%. Weak response.
O-53	O	NR	R	Sized reduction 42%, post-treatment cellularity 70%. Weak response. Post-operative note indicated residual intraductal extension.
O-55	O	NR	R	Size reduction 20%, post-treatment cellularity 15%. Weak response.
EAN-8	E	NR	R	Complete response.
EAN-39	E	NR	R	Size reduction 32%, post-treatment cellularity 23.3%. Weak response.
EAN-43	E	NR	R	Prediction accuracy 81.8%. Extensive tumour necrosis.
EAN-47	E	NR	R	Mucinous. Large initial size (4.9 cm). Size reduction 35%, post-treatment cellularity 11.8%. Weak response.
EAN-56	E	R	NR	Large initial size (5.5 cm). Size reduction 22%, post-treatment cellularity 8.5%. Weak response. Mixed ductal and lobular carcinoma. Examination of images suggests that ductal regions had a greater response than lobular regions.
EAA-57	E	R	R*	Mucinous. Prediction accuracy 83%, prediction class score close to zero.

[a]To generate robust predictions, eleven models were trained using balanced subsets of the training dataset. For each independent patient, treatment response was predicted using a majority voting strategy across the outputs of the eleven models. Therefore, the individual prediction accuracy was determined by the proportion of models (out of eleven) that agreed with the majority-predicted response.

with continued research and scanning of additional patients, may enable the differentiation between responders from those exhibiting a poor or delayed response to neoadjuvant chemotherapy.

In the experimental arm of this study, clinicians had the discretion to adapt treatment based on the prediction results. Among the 8 predicted non-responders in the experimental arm, only 4 patients (50%) had their treatment modified based on the prediction, while the other 4 continued with the original chemotherapy regimen. Of those 4 non-adapted patients, 3 were found to have a tumour response on the final specimen. Treatment modifications primarily included premature interruption of chemotherapy and early surgical resection rather than treatment intensification with additional chemotherapy drugs or a change to a different regimen. Due to the small number of patients who had their treatment adapted based on the prediction, and also the fact that 3 out of 4 non-adapted patients ultimately responded to treatment, the study here did not focus nor could determine whether modifying the chemotherapy regimen guided by the prediction

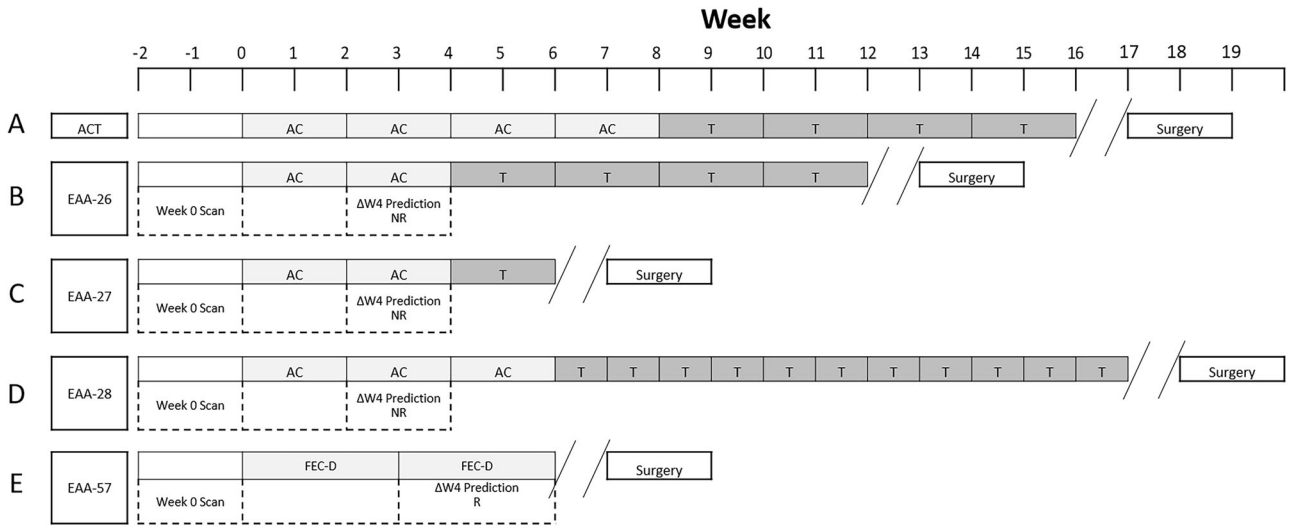
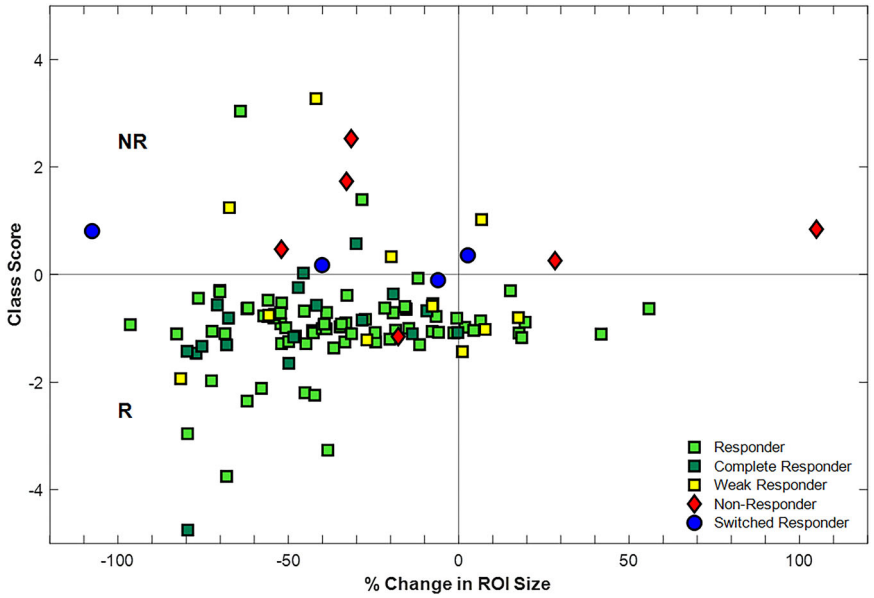


Fig. 5 | Clinical histories of adapted patients. A Schematic diagram for the administration of standard AC-T chemotherapy. Weeks are shown from left to right. Typical durations are illustrated. **B**–**E** The four patients in the Experimental Arm that were adapted (EAA-26, EAA-27, EAA-28, and EAA-57) in Fig. 4. In (**B**), AC was shortened to move to T; in (**C**), the AC and T were shortened to move to surgery; in (**D**), AC was shortened and T was intensified; and in (**E**), FEC-D was stopped early due to prediction and full clinical response.

Fig. 6 | ROI size change, (W4 – W0)/W0, versus W4 Prediction Class Score. Size changes at Week 4 versus Ultrasound Class Score. This plot presents tumour size change between week 4 and week 0 per tumour versus predictive class score for response. Responder, complete responder and non-responder are defined as per RECIST criteria. Weak responders were borderline in regards to size changes (±5%) or had significantly diminished cellularity (5% or less) but did not meet size criteria.



model could improve clinical outcomes (i.e., disease-free survival, overall survival). Nevertheless, the delivery of chemotherapy that has cytotoxic efficacy is the fundamental tenet on which success in medical oncology is based. That stated, for the 4 patients 3 are alive with survivals of 4.5, 6, and 6 years. One patient did succumb to widespread metastasis after 2 years. Due to the lack of direct comparison of outcomes between patients predicted to be non-responders who underwent treatment modification versus those who did not, the actual causal benefit of QUS-guided treatment modification on long-term patient outcomes (e.g., disease-free survival, overall survival) cannot be clearly quantified at this time.

In the GeparTrio trial[19], patients whose tumours showed less than 50% shrinkage by conventional B-mode ultrasound after two cycles of neoadjuvant chemotherapy (docetaxel, doxorubicin, and cyclophosphamide) were randomized to either continue the same treatment or switch to an alternative regimen with vinorelbine and capecitabine. While this adaptive strategy did not lead to higher pathological complete response rates, the negative results may be attributed to uncertainties regarding the best timing for assessing response, the most appropriate imaging technique, and the optimal intensification regimen. Despite this, early treatment adaptation remains a promising concept that deserves further exploration. Therefore, future QUS-guided adapted studies could focus on pre-specifying an intensified chemotherapy regimen (duration or dose) for those predicted to be non-responders rather than leaving treatment modifications at the physician's discretion. This approach would be important for comparing the impact of QUS-guided treatment changes, assessing success in converting non-responders into responders, and determining whether it provides disease-free survival benefits with the intensified regimen.

The study here also indicated that quantitative-ultrasound-based measures of response occur in advance of any size-based changes which become evident mostly pre-operatively after the completion of

Table 5 | Pair-wise statistical tests of differences for size differences (week 4 − pre-treatment), and class score difference

Groups	% Size change *p*-value	Significant at 95%?	Class score *p*-value	Significant at 95%?
NR vs R	0.0977	–	0.00425	Yes
NR vs CR	0.0667	–	0.00693	Yes
NR vs WR	0.580	–	0.209	–
NR vs S	0.594	–	0.337	–
R vs CR	0.126	–	0.996	–
R vs WR	0.243	–	0.152	–
R vs S	0.739	–	0.00173	Yes
CR vs WR	0.056	–	0.271	–
CR vs S	0.371	–	0.00733	Yes
WR vs S	0.744	–	0.327	–

Table 6 | Pair-wise statistical tests of differences for size differences (post-treatment − pre-treatment) and class score difference

Groups	% Size change *p*-value	Significant at 95%?	Class score *p*-value	Significant at 95%?
NR vs R	7.31E-4	Yes	0.00425	Yes
NR vs CR	8.98E-6	Yes	0.00693	Yes
NR vs WR	0.0104	Yes	0.209	–
NR vs S	0.110	–	0.337	–
R vs CR	3.17E-10	Yes	0.996	–
R vs WR	0.249	–	0.152	–
R vs S	0.456	–	0.00173	Yes
CR vs WR	7.31E-7	Yes	0.271	–
CR vs S	3.87E-5	Yes	0.00733	Yes
WR vs S	0.948	–	0.327	–

chemotherapy. The GeparTrio trial[19] did demonstrate that after two cycles of chemotherapy (8 weeks for patients here) a prediction of complete response by simple measurement of size reduction by ultrasound. In the work here, class scores were predictive for responders and complete responders compared to non-responders at 4 weeks into several months of chemotherapy, whereas size-based measures were not. Specifically, at week 4 results indicated a statistically significant concordance with QUS class score and response (for NR vs R, and NR vs CR, R vs S, and CR vs S), but no concordance was observed with size change that was statistically significant for all groups. Comparison of data (Fig. 6 versus Supplementary Fig. 6) indicates how size changes compared between weeks 4 and after chemotherapy versus class score. Size changes at later times during chemotherapy from the study here are part of a separate analysis for the future. We anticipate that size changes will become more prominent and predictive beyond 1 cycle of chemotherapy as per the GeparTrio study[19].

We have examined the training data used for the model[7] with regards to HER2 status. For HER2+ patients, the overall model performance accuracy was 97% whereas in HER2- patients it was 87% (likely due to response to Herceptin in HER2+ anthracycline-taxane non-responding patients). In HER2+ patients, the sensitivity and specificity were 100% and 96%, respectively. In HER2- patients, the sensitivity and specificity were 86% and 87%, respectively. That was based on 31 HER2+ patients and 60 HER2- patients in the development cohort unadjusted for data balancing.

In the observational alone data here in the HER2+ patient alones there was an accuracy of 100% as 14/14 patients were predicted to be responders and were. In the HER2- patients in the observational arm, the accuracy of the test performance was 91% (sensitivity 90%, specificity 100%). This bias towards response in the HER2+ was similar to that in the training set (above and ref. 7) and is likely due to that for these patients Herceptin is efficacious.

In the study here, there were also differences between responders and complete responders versus weak responders where size changes or tumour cellularity changes were close to borderline regarding RECIST-based and pathological-based response measures. The predictor used here was for response (R: partial response + complete response) versus non-response (NR: progressive disease + stable disease). This was selected for study in order to facilitate the standard of care continuing in responding patients such that standard of care was not changed in these patients, and to permit consideration of modifications to chemotherapy in NR patients. Furthermore, this enabled investigation into whether modifying their treatment, such as accelerating surgery or adding a new agent, would be feasible and potentially improve outcomes. In the future, it should be able to discriminate using QUS methods CR from PR patients and consider PR patients (weaker responders) for dose intensification to convert them to PR patients. QUS Feature sets to differentiate these response types are under development for such a purpose[20].

Our work has several strengths, specifically the prediction models used were developed using a large number of prospectively imaged patients with QUS acquired at pre-established experimental times and a systematic methodology for acquiring and analysing images, which was consistently reproduced in this clinical trial[7]. The limitations of our work include that most patients were responders (in both development and validation cohorts), which limits the negative predictive value of our analysis. In addition, treatment adaptation was not pre-specified but left to the discretion of the medical oncologists. As a result, only one-third of patients predicted not to respond had their treatment modified, primarily involving the early cessation of chemotherapy rather than treatment intensification. Future work will involve further improving the current models by increasing the studied population and refining the model. Weak responders can be identified in the future for intervention. In addition, we aim to further refine models that can predict pathological complete response rather than just tumour reduction, as this has known prognostic value and clinical implications. And lastly, our data warrants external validation to ensure generalizability across multiple institutions.

In conclusion, this prospective phase 2 trial suggests that longitudinal QUS imaging-based radiomics, combined with machine learning, is accurate for predicting early treatment response to neoadjuvant chemotherapy. This research opens new opportunities for future clinical trial design focused on adapted neoadjuvant treatment courses. Future studies are needed to assess the generalizability of this data, enhance clinical applicability, and design the ideal treatment strategy for patients predicted to be non-responders.

Methods

Study population

This single-center, phase 2 randomized controlled trial included female patients over 18 years old with biopsy-proven clinical stage II-III breast cancer (according to the American Joint Committee on Cancer [AJCC] 7th edition). These patients were referred to the medical oncology team at Sunnybrook Health Sciences Centre (Toronto, Canada) and deemed suitable for neoadjuvant chemotherapy per standard of care. Eligibility criteria required normal blood counts, creatinine levels, liver function tests, and cardiac function. Exclusion criteria included inflammatory breast cancer, a history of connective tissue or dermatologic diseases involving the breast, and an Eastern Cooperative Oncology Group (ECOG) performance status of ≥3. This study received approval from the institutional ethics committee (Sunnybrook Research Institute, SUN 308-2017) and was registered on clinicaltrials.gov (NCT04050228) in December 2019, where the protocol is available (https://clinicaltrials.gov/study/NCT04050228). The sample size

of 120 patients was based on convenience. All participants provided written informed consent, and the study adhered to the Declaration of Helsinki. This study is reported as per the Consolidated Standards of Reporting Trials (CONSORT) statement.

Procedures and randomization

Participants were randomly assigned to either the observational or experimental arm before neoadjuvant treatment initiation using a 1:1 permuted block randomization method[21], with hormone receptor status as a stratification factor. Patients were stratified with block sizes of 4, 6, or 8 pregenerated using custom MATLAB code. The random allocation sequence was stored in a password-protected Excel spreadsheet on a secure server, accessible only to one research team member, while patients were assigned sequentially in strict recruitment order by the designated researcher. Patients and investigators were unblinded.

All patients were initially planned to receive a standard neoadjuvant chemotherapy regimen as deemed appropriate by the treating medical oncologist and underwent QUS imaging at four time points: within 7 days before neoadjuvant chemothearpy (baseline) and at the first, fourth, and eighth week during neoadjuvant chemotherapy. QUS response prediction techniques using ultrasound images acquired at week 4 of neoadjuvant chemotherapy were used to predict response or non-response to systemic therapy (as further detailed below). In the observational arm, medical oncologists remained blind to prediction results. In contrast, in the experimental arm, they received reports summarizing the week 4 prediction results, allowing for timely personalization of treatment based on clinical judgement. Treatment personalization was not predefined but left to the physician's discretion. This could include continuing chemotherapy where a response was indicated, modifying the regimen where a lack of response was observed, or discontinuing systemic therapy in favour of early surgical resection. Treatment changes were decided by the oncologist within 7 days of week 4 QUS data acquisition. Cases being considered for chemotherapy modification were discussed as indicated at an interdisciplinary tumour board with the whole medical oncology group in breast cancer. This trials defined no stopping rules.

The neoadjuvant chemotherapy regimens typically consisted of dose-dense AC-T (±H), FEC-D (±H), or the KEYNOTE 522[22] regimens as decided by their treating medical oncologist. Specifically, dose-dense AC-T chemotherapy typically consisted of doxorubicin 60 mg/m² and cyclophosphamide 600 mg/m² (AC) every two weeks for 4 cycles, followed by paclitaxel 175 mg/m² every two weeks (T) for 4 cycles. FEC-D included 5-FU 500 mg/m², epirubicin 100 mg/m², and cyclophosphamide 500 mg/m² (FEC) every 3 weeks for 3 cycles, followed by docetaxel 100 mg/m² every 3 weeks for 3 cycles. Patients with HER2+ tumours had trastuzumab administered concurrently with the first dose of taxane (paclitaxel for those receiving AC-TH or docetaxel for those undergoing FEC-DH) at an initial dose of 8 mg/kg IV and continued every three weeks with a subsequent dose of 6 mg/kg IV with docetaxel and 4 mg/kg IV every 2 weeks with paclitaxel. The KEYNOTE 522 regimen included 12 weeks of pembrolizumab (200 mg) every 3 weeks, paclitaxel (80 mg/m²) once weekly, and carboplatin (AUC 1.5) once weekly, followed by a subsequent phase of 4 cycles of pembrolizumab (200 mg), doxorubicin (60 mg/m²) plus cyclophosphamide (600 mg/m²) once every 3 weeks. Blood counts, liver function tests, and renal function tests were regularly monitored, and growth factor support was prescribed to aid bone marrow recovery, following standard institutional protocols. Subsequent to neoadjuvant chemotherapy and surgical resection, adjuvant radiation followed the standard of care. Adjuvant systemic therapy included hormone therapy for hormone receptor-positive tumours, trastuzumab for HER2+ tumours with complete pathological response, trastuzumab emtansine (TDM1) for HER2+ tumours without a pathological complete response, and adjuvant capecitabine for triple-negative tumours without a complete response, as recommended. Patients treated with the KEYNOTE 522 regimen underwent 9 cycles of adjuvant pembrolizumab given every 3 weeks.

Quantitative ultrasound data collection, feature extraction and response monitoring

Experienced sonographers used a Sonix RP clinical ultrasound system (Analogic Medical Corp.) with an L14-5W/60 linear array transducer (centre frequency 6.5 MHz, bandwidth range 3–8 MHz) to image the primary tumour volume at evenly spaced 5 mm intervals across the tumour mass. For each tumour, typically between four and six image frames were selected. For each frame, the sonographers created a region of interest (ROI) delineating the tumour boundary by hand. The ROI positions were verified by an expert breast radiologist and the principal investigator. For each ROI, a sliding window technique was employed to subdivide the tumour region in a series of 2 mm by 2 mm data blocks overlapped at 94%. The RF scan lines in each block were truncated using a Hanning window. A fast Fourier transform (FFT) was applied to each RF scan line, and the resulting ultrasound frequency spectra were averaged within each block. The averaged spectrum was normalized using a reference phantom technique[23].

Five QUS parameters (midband-fit, spectral slope, spectral intercept, average scatterer diameter, and average acoustic concentration) were then calculated for each data block and used to create a set of five parametric maps[7,8]. A grey-level co-occurrence matrix (GLCM) method[24] was used to extract four texture features (contrast, correlation, energy, and homogeneity) from each parametric map, and the set of QUS parameters and texture features obtained for the image frames selected for the tumour were then averaged across frames to obtain a set of average QUS features representative of the tumour.

A QUS classification model based on a support vector machine-radial basis function (SVM-RBF) algorithm developed from a training set of QUS features and texture features obtained from a cohort of over 200 breast tumours was used to classify patients as responders or non-responders at the 4th week of treatment[7]. The result was summarised in a response prediction report, which identified the patient as either a responder or non-responder, reported prediction accuracy according to majority decision rule, and included representative multi-feature parametric maps of the tumour QUS parameters. During the development of the classification model, eleven models were generated using balanced subsets of the training dataset. To predict the treatment response for an independent patient, a majority voting approach was applied across the predictions from these eleven models. The prediction accuracy of the individual patient was calculated based on how many times the majority response was predicted out of the eleven model predictions[7]. This was done here as well for model use and is reported an average accuracy on a per patient basis.

For model development, the feature selection method and the statistical test results comparing the two response groups were previously reported in an earlier publication, in which the treatment response prediction model was developed[7]. For clarity, a sequential feature selection method was applied to balanced subsets of the training data to identify the optimal set of features for classifying tumour responses. A total of 31 features were extracted from ultrasound data and top features when used in combination selected for classification. The features selected during the training process were ΔSS-con, ΔSI-ene, ΔSI-hom, and ΔASD-cor. Statistical analyses were performed on the training dataset, and the results were detailed[7]. To determine the appropriate statistical test for comparing the groups, we first assessed the normality of each feature dataset using the Shapiro–Wilk normality test. For features that followed a normal distribution, an unpaired t-test was used to compare the groups. For features that did not meet the normality assumption, a Mann–Whitney U test was employed instead.

The selected features are defined as follows along with their Chi2 score (as a relative ranking) and were independent parameters:

- ΔSS-con: Contrast texture parameter derived from the spectral slope map (related to heterogeneity (intensity variation) in scatterer size). Chi^2 score 2.07

- ΔASD-cor: Correlation texture parameter derived from the average scatter diameter map. This quantifies how similar the patterns of brightness or other characteristics are in neighbouring areas of an image. Chi^2 score 1.78

- ΔSI-ene: Energy texture parameter derived from the spectral intercept map (related to scatterer concentration variation). Chi^2 score 0.99

- ΔSI-hom: Homogeneity texture parameter derived from the spectral intercept map (related to scatter size homogeneity). Chi^2 score 0.73

(Here, Δ denotes the change in each parameter, calculated as the difference between the baseline ultrasound scan (week 0) and the scan acquired during treatment at week 4.)

Tumour response definition

Magnetic resonance imaging (MRI) was typically carried out as part of standard patient care before to the initiation of chemotherapy to evaluate tumour extent and serve as a reference for baseline tumour measurement. Response was subsequently assessed based on residual tumour size and cellularity on surgical specimen histopathology following completion of neoadjuvant treatment. Patients were categorized as responders (R) or non-responders (NR) using a modified response grading system based on histopathological evaluation following surgical resection[6,7]. The responder category included patients who demonstrated a reduction in the diameter of tumour by at least 30% from baseline imaging to the surgical specimen histopathology, or a decrease in cellularity to less than 5% in the tumour bed, or the complete disappearance of all target lesions (further subclassified as Complete Responder [CR]).

The responder category included patients who demonstrated a reduction in the diameter of tumour by at least 30% from baseline imaging to histopathology, or a decrease in cellularity to less than 5% in the tumour bed, or the complete disappearance of all target lesions (further subclassified as Complete Responder [CR]).

In cases (infrequent) where there was a diminishment in cellularity to less than 5% but not a diminishment in >30% of tumour size, responders [R] were subclassified as Weak Responders [WR]. The non-response category included patients with a tumour size reduction of less than 30% and cellularity (greater than 5%). This was done specifically to examine QUS class scores for these patients which often were close to indeterminate. Patients who were initially predicted to be non-responders but ultimately responded to therapy after having their treatment switched (adapted) were subclassified as Switched Responders (S).

Study outcomes

The primary endpoint of this phase 2 study was to assess the efficacy of QUS in predicting the response to neoadjuvant chemotherapy. The secondary endpoint was to evaluate the feasibility of using a radiomic-based approach to personalize systemic treatment adaptation early in the treatment course, including a qualitative description of how medical oncologists used the predicted results to personalize treatment.

Statistical analysis

Descriptive analysis was used to examine patient, disease, treatment-related factors, and response rates. MATLAB R2016a (Mathworks, Natick, MA, USA) was used for image preprocessing, ROI segmentation, feature extraction, and radiomics model development. Other statistical tests were performed using IBM SPSS version 22 (Amonk, NY, USA) with standard methods for calculating test performance applied to combined group patients.

Class score and tumour size analysis

Tumour size change was determined based on the size of ultrasound ROIs used to analyse baseline and week 4 images. Tumour size comparison for tumours post-treatment pre-operatively was carried out using tumour dimensions from clinical MRI reports before starting chemotherapy and upon completion of chemotherapy pre-operatively as that was available as used for RECIST-based size reporting. Pearson's coefficient with t-test was performed to assess the level of correlation for each of the data sets (not separated by group) of size change and ultrasound class score.

For further analysis, the Shapiro–Wilk normality test was used to assess whether the samples in each group followed a normal distribution for both tumour size change and prediction class score. None of the data was found to be normal, and hence group to group statistical comparisons were conducted using non-parametric testing (the Mann–Whitney test). This test has the null hypothesis that two independent groups of samples come from populations with the same median. Statistical comparison of both size change and prediction class score was conducted for each pair of groups.

Data availability

Data are stored in an institutional repository and will be made available on request to the office of the Vice-President, Research and Innovation, Sunnybrook Research Institute.

Received: 1 May 2025; Accepted: 28 September 2025;
Published online: 04 December 2025

References

1. Gradishar, W. J. et al. Breast cancer, version 3.2022, NCCN Clinical Practice Guidelines in Oncology. *J. Natl Compr. Cancer Netw.* **20**, 691–722 (2022).
2. Cardoso, F. et al. Early breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann. Oncol.* **30**, 1194–1220 (2019).
3. Cardoso, F. et al. 5th ESO-ESMO international consensus guidelines for advanced breast cancer (ABC 5). *Ann. Oncol.* **31**, 1623–1649 (2020).
4. Von Minckwitz, G. et al. Trastuzumab emtansine for residual invasive HER2-positive breast cancer. *N. Engl. J. Med.* **380**, 617–628 (2019).
5. Masuda, N. et al. Adjuvant capecitabine for breast cancer after preoperative chemotherapy. *N. Engl. J. Med.* **376**, 2147–2159 (2017).
6. Dasgupta, A. et al. Quantitative ultrasound radiomics using texture derivatives in prediction of treatment response to neo-adjuvant chemotherapy for locally advanced breast cancer. *Oncotarget* **11**, 3782–3792 (2020).
7. Sannachi, L. et al. Breast cancer treatment response monitoring using quantitative ultrasound and texture analysis: comparative analysis of analytical models. *Transl. Oncol.* **12**, 1271–1281 (2019).
8. Sannachi, L. et al. Non-invasive evaluation of breast cancer response to chemotherapy using quantitative ultrasonic backscatter parameters. *Med. Image Anal.* **20**, 224–236 (2015).
9. Sadeghi-Naini, A. et al. Early prediction of therapy responses and outcomes in breast cancer patients using quantitative ultrasound spectral texture. *Oncotarget* **5**, 3497–3511 (2014).
10. Sadeghi-Naini, A. et al. Quantitative ultrasound evaluation of tumor cell death response in locally advanced breast cancer patients receiving chemotherapy. *Clin. Cancer Res.* **19**, 2163–2174 (2013).
11. Banihashemi, B. et al. Ultrasound imaging of apoptosis in tumor response: novel preclinical monitoring of photodynamic therapy effects. *Cancer Res.* **68**, 8590–8596 (2008).
12. Vlad, R. M., Brand, S., Giles, A., Kolios, M. C. & Czarnota, G. J. Quantitative ultrasound characterization of responses to radiotherapy in cancer mouse models. *Clin. Cancer Res.* **15**, 2067–2075 (2009).
13. Czarnota, G. J. et al. 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 (1999).
14. Dasgupta, A. et al. Quantitative ultrasound radiomics guided adaptive neoadjuvant chemotherapy in breast cancer: early results from a randomized feasibility study. *Front Oncol.* **14**, 1273437 (2024).
15. Lambin, P. et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat. Rev. Clin. Oncol.* **14**, 749–762 (2017).
16. Lambin, P. et al. Radiomics: Extracting more information from medical images using advanced feature analysis. *Eur. J. Cancer* **48**, 441–446 (2012).
17. Coudert, B. et al. Use of [18F]-FDG PET to predict response to neoadjuvant trastuzumab and docetaxel in patients with HER2-

positive breast cancer, and addition of bevacizumab to neoadjuvant trastuzumab and docetaxel in [18F]-FDG PET-predicted non-responders (AVATAXHER): an open-label, randomised phase 2 trial. *Lancet Oncol.* **15**, 1493–1502 (2014).
18. Partridge, S. C. et al. Diffusion-weighted MRI findings predict pathologic response in neoadjuvant treatment of breast cancer: the ACRIN 6698 multicenter trial. *Radiology* **289**, 618–627 (2018).
19. Von Minckwitz, G. et al. Neoadjuvant vinorelbine-capecitabine versus docetaxel-doxorubicin-cyclophosphamide in early nonresponsive breast cancer: phase III randomized GeparTrio trial. *JNCI J. Natl Cancer Inst.* **100**, 542–551 (2008).
20. Sannachi, L. et al. Response monitoring of breast cancer patients receiving neoadjuvant chemotherapy using quantitative ultrasound, texture, and molecular features. *PLoS ONE* **13**, e0189634 (2018).
21. Broglio, K. Randomization in clinical trials: permuted blocks and stratification. *JAMA* **319**, 2223 (2018).
22. Cortes, J. et al. Pembrolizumab plus chemotherapy in advanced triple-negative breast cancer. *N. Engl. J. Med.* **387**, 217–226 (2022).
23. Yao, L. X., Zagzebski, J. A. & Madsen, E. L. Backscatter coefficient measurements using a reference phantom to extract depth-dependent instrumentation factors. *Ultrason Imaging* **12**, 58–70 (1990).
24. Haralick, R. M., Shanmugam, K. & Dinstein, I. Textural features for image classification. *IEEE Trans. Syst., Man, Cyber.* **3**, 610–621 (1973).

Acknowledgements

We express our gratitude to the Terry Fox Research Institute, Natural Sciences and Engineering Research Council of Canada, and Lotte & John Hecht Memorial Foundation for supporting this research. We extend our sincere appreciation to all the patients, family members, and participating staff from Sunnybrook Health Sciences Center and Sunnybrook Research Institute for their valuable contributions to this study.

Author contributions

Conceptualization: G.J.C. and M.T. Data curation: All authors. Formal analysis: D.M.P., D.A., M.L.A.P., J.Y., A.S.N. and G.J.C. Funding acquisition: G.J.C. Investigation: All authors. Methodology: All authors. Project administration: G.J.C. Resources: G.J.C. Software: L.S. and G.J.C. Supervision: G.J.C. Validation: All authors. Visualization: All authors. Writing—original draft: D.M.P., D.A., A.W.C., L.S. and G.J.C. Writing—review and editing: All authors. All the authors are in agreement and accountable for all the aspects of the work.

Competing interests

Sonal Gandhi received advisory board honorarium from Lily and AstraZeneca. Katarzyna Jerzak has been a consultant, speaker, or advisory board member for Amgen, AstraZeneca, Apo Biologix, Daiichi Sankyo, Eli Lilly, Esai, Genomic Health, Gilead Sciences, Knight Therapeutics, Merck, Myriad Genetics, Novartis, Organon, Pfizer, Roche, and Viatris; has received research funding from AstraZeneca, Eli Lilly, and Pfizer; has received support for attending meetings or travel from AstraZeneca and Daiichi Sankyo; and has received drug supply from Pfizer and Viatris for an investigator initiated clinical trial. Sunnybrook Research Institute hold patents related to the quantitative ultrasound methods discussed here. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41698-025-01134-x.

Correspondence and requests for materials should be addressed to Gregory J. Czarnota.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]Department of Radiation Oncology, Sunnybrook Health Sciences Centre, Toronto, Canada. [2]Department of Radiation Oncology, University of Toronto, Toronto, Canada. [3]Physical Sciences, Sunnybrook Research Institute, Toronto, Canada. [4]Division of Medical Oncology, Department of Medicine, Sunnybrook Health Sciences Centre, Toronto, Canada. [5]Department of Medicine, University of Toronto, Toronto, Canada. [6]Department of Surgical Oncology, Department of Surgery, Sunnybrook Health Sciences Centre, Toronto, Canada. [7]Department of Surgery, University of Toronto, Toronto, Canada. [8]Department of Medical Biophysics, University of Toronto, Toronto, Canada. [9]Department of Electrical Engineering and Computer Sciences, Lassonde School of Engineering, York University, Toronto, Canada. [10]Department of Medical Imaging, Sunnybrook Health Sciences Centre, Toronto, Canada. [11]Department of Medical Imaging, University of Toronto, Toronto, Canada. [12]Department of Physics, Toronto Metropolitan University, Toronto, Canada. [13]These authors contributed equally: Daniel Moore-Palhares, David Alberico, Adrian Wai Chan. ✉e-mail: gregory.czarnota@sunnybrook.ca