Radiology: Imaging Cancer

ORIGINAL RESEARCH

Quantitative US Delta Radiomics to Predict Radiation Response in Individuals with Head and Neck Squamous Cell Carcinoma

*Laurentius Oscar Osapoetra, PhD** • *Archya Dasgupta, MD** • *Daniel DiCenzo, MSc* • *Kashuf Fatima, BSc* • *Karina Quiaoit, MSc* • *Murtuza Saifuddin, MD* • *Irene Karam, MD* • *Ian Poon, MD, FRCPC* • *Zain Husain, MD* • *William T. Tran, PhD* • *Lakshmanan Sannachi, PhD* • *Gregory J. Czarnota, MD, PhD*

From the Departments of Radiation Oncology (L.O.O., A.D., I.K., I.P., Z.H., W.T.T., G.J.C.), Medical Oncology (W.T.T.), and Medicine (W.T.T.), Sunnybrook Health Sciences Centre, 2075 Bayview Ave, Toronto, ON, Canada M4N 3M5; Departments of Radiation Oncology (L.O.O., A.D., I.K., I.P., Z.H., W.T.T., G.J.C.) and Medical Biophysics (G.J.C.), University of Toronto, Toronto, Canada; and Departments of Physical Sciences (L.O.O., A.D., D.D., K.F., K.Q., M.S., L.S., G.J.C.) and Evaluative Clinical Sciences (W.T.T.), Sunnybrook Research Institute, Toronto, Canada. Received March 21, 2023; revision requested May 12; revision received November 24; accepted January 17, 2024. **Address correspondence to** G.J.C. (email: *gregory.czarnota@sunnybrook.ca*).

* L.O.O. and A.D. contributed equally to this work.

Supported by a Terry Fox Foundation Program Project Grant from the Hecht Foundation (grant 1083) and the Natural Sciences and Engineering Research Council of Canada. The funding agencies had no role in the study design, study methods, study results, or the preparation of the article.

Conflicts of interest are listed at the end of this article.

Radiology: Imaging Cancer 2024; 6(2):e230029 • https://doi.org/10.1148/rycan.230029 • Content codes: HN OI US

Purpose: To investigate the role of quantitative US (QUS) radiomics data obtained after the 1st week of radiation therapy (RT) in predicting treatment response in individuals with head and neck squamous cell carcinoma (HNSCC).

Materials and Methods: This prospective study included 55 participants (21 with complete response [median age, 65 years {IQR: 47–80 years}, 20 male, one female; and 34 with incomplete response [median age, 59 years {IQR: 39–79 years}, 33 male, one female) with bulky node-positive HNSCC treated with curative-intent RT from January 2015 to October 2019. All participants received 70 Gy of radiation in 33–35 fractions over 6–7 weeks. US radiofrequency data from metastatic lymph nodes were acquired prior to and after 1 week of RT. QUS analysis resulted in five spectral maps from which mean values were extracted. We applied a gray-level co-occurrence matrix technique for textural analysis, leading to 20 QUS texture and 80 texture-derivative parameters. The response 3 months after RT was used as the end point. Model building and evaluation utilized nested leave-one-out cross-validation.

Results: Five delta (Δ) parameters had statistically significant differences ($P < .05$). The support vector machines classifier achieved a sensitivity of 71% (15 of 21), a specificity of 76% (26 of 34), a balanced accuracy of 74%, and an area under the receiver operating characteristic curve of 0.77 on the test set. For all the classifiers, the performance improved after the 1st week of treatment.

Conclusion: A QUS Δ-radiomics model using data obtained after the 1st week of RT from individuals with HNSCC predicted response 3 months after treatment completion with reasonable accuracy.

Clinicaltrials.gov registration no. NCT03908684

Supplemental material is available for this article.

Radiomics is an emerging field in medicine and oncology involving quantitative imaging analysis. While imaging has traditionally been used for diagnostic purposes, the introduction of computer vision and advanced machine learning classifier algorithms has enabled the development of noninvasive imaging biomarkers (1,2). High-dimensional image analysis can be undertaken with different imaging modalities such as US, CT, MRI, and PET, which are performed during various stages of standard oncologic practice. Quantitative imaging radiomics features are linked with a diverse spectrum of clinical end points, including histopathologic or molecular characterization, prognostication, and determination of treatment response (3). Quantitative US (QUS) techniques provide quantitative parameters that reflect the microstructures of tissues (4). QUS is more robust than qualitative imaging, as the former provides objective operator- and system-independent characterization of the tissues for diagnostic and prognostic purposes. QUS spectroscopy is a type of QUS technique that converts radiofrequency data into frequency spectra using fast-Fourier transform, which retains more detailed microstructural information compared with conventional sonography (5,6). QUS spectroscopy is sensitive to changes in the cellular characteristics of tissue microstructure and their associated mechanical properties, which have been linked to differential clinical outcomes (7–10). Furthermore, QUS spectroscopy has been demonstrated as a reliable marker for detecting treatment-related changes at a cellular level arising from nuclear fragmentation, apoptosis, and other events related to cell death, leading to changes in US scatterer properties (11–15). Clinical studies have demonstrated the utility of QUS radiomics in the prediction of response to neoadjuvant chemotherapy in patients with

Abbreviations

AAC = average acoustic concentration, AUC = area under the receiver operating characteristic curve, CR = complete response, CV = cross-validation, GLCM = gray-level co-occurrence matrix, HNSCC = head and neck squamous cell carcinoma, IR = incomplete response, LOO = leave one out, MBF = midband fit, QUS = quantitative US, RT = radiation therapy, SVM-RBF = support vector machines–radial basis function

Summary

A multivariable quantitative US delta-radiomics model using data obtained after 1 week of radiation therapy from individuals with head and neck squamous cell carcinoma predicted the long-term response with reasonable accuracy.

Key Points

- In participants with bulky node-positive head and neck squamous cell carcinoma, five quantitative US delta-radiomics parameters demonstrated statistically significant differences ($P < .05$) between complete and incomplete response groups.
- The developed multivariable classification model had a balanced accuracy of 74% and an area under the receiver operating characteristic curve of 0.77, evaluated using a nested leave-one-out cross-validation approach.

Keywords

Computer-Aided Diagnosis (CAD), Ultrasound, Radiation Therapy/Oncology, Head/Neck, Radiomics, Quantitative US, Radiotherapy, Head and Neck Squamous Cell Carcinoma, Machine Learning

locally advanced breast cancer or radiation therapy (RT) response in head and neck squamous cell carcinoma (HNSCC).

HNSCC commonly arises from the epithelial lining of the upper aerodigestive system, accounting for major cancer-related morbidity and mortality globally (16,17). Radical RT with concurrent chemotherapy is the standard of care for patients with locally advanced oropharynx, larynx, and hypopharyngeal HNSCC. Fractionated RT is typically delivered over 6 to 7 weeks, and the optimal time for response assessment is approximately 12 weeks following treatment completion, using clinical examination and morphologic and metabolic imaging (18). However, 12 weeks after treatment is somewhat late for changing the treatment plan, as the treatment has already been administered. Consequently, the ability to predict treatment response should ideally be as early as possible, either before or during treatment, allowing for timely treatment adjustments when necessary. This is supported by findings showing that radiation-induced cellular events are known to be initiated from the 1st day of RT (19,20). Determination of treatment response in real time during RT can provide a valuable treatment window including treatment escalation and de-escalation strategies to achieve an optimal balance between survival and treatment-induced toxicities (21). Treatment escalation strategies could involve intensifying chemotherapy regimens, increasing the prescribed dose of radiation, or prescribing different chemotherapy regimens. On the other hand, de-escalation strategies include the reduction of radiation dose for patients expected to respond very well to treatment. There are ongoing clinical trials that attempt to de-escalate radiation treatment for patients with human papillomavirus oropharyngeal squamous cell carcinoma given their very good response to radical treatment, achieving reduced toxicity while maintaining good local control. Radiomics analysis on QUS spectral parametric images can enhance response prediction in addition to the well-established human papillomavirus–positive status (22).

In this study, we explored the potential of a QUS radiomics model for early prediction of treatment response in metastatic lymph nodes. Our analysis involved the inclusion of higher-dimensional (second-order texture) features and the application of advanced machine learning classifiers. We hypothesized that QUS spectral parametric imaging can help determine the cellular events and changes in acoustic properties during the early phase of RT, which can be correlated with the long-term response to RT as detected with standard imaging modalities.

Materials and Methods

Participant Selection

The institutional research ethics board (project identification no. SUN-3047) approved this prospective study conducted in a single institution and registered with ClinicalTrials.gov (registration no. NCT03908684). The study was conducted following good clinical practice according to the Declaration of Helsinki. All participants provided written informed consent for their participation. Study accrual was carried out between January 2015 and October 2019, with the final analysis performed in 2020.

All participants in this study have been previously reported (10). This prior study dealt with a pretreatment model, whereas in this study we report both pretreatment and early-in-treatment models. In addition, we have refined our model building and evaluation strategies to conform to standards in statistical learning. Participants diagnosed with biopsy-proven HNSCC with a primary site involving the oropharynx, hypopharynx, or larynx or carcinoma of unknown primary with bulky metastatic lymph node in the neck treated with curative-intent RT were considered eligible for this study. Participants with a history of prior interventions (surgery, systemic therapy, RT) for HNSCC, a history of RT in the head and neck region, or severe medical comorbidities resulting in limited life expectancy were excluded from the study. Additionally, participants with metastatic disease beyond the regional lymph node, nasopharyngeal carcinoma, or carcinoma of unknown primary with suspected nasopharyngeal carcinoma origin (Epstein-Barr positivity or histologic findings) were also excluded.

Treatment Protocols

All participants received 70 Gy of radiation in 33–35 fractions over 6–7 weeks to the high-risk volumes, using intensity-modulated RT with image guidance following standard institutional practice. Use of concurrent systemic therapy was decided by the oncologist. Participation in the study did not influence treatment decisions. Following treatment completion, response assessment was performed with clinical examination, endoscopy, and imaging with CT, MRI, or PET, as decided by the treating oncologists. Participants with complete resolution of the primary disease and lymph nodes

smaller than 1 cm (without any high-risk radiologic features suspicious for disease involvement) or complete metabolic response at PET were considered as having complete response (CR) and all others as having incomplete response (IR). The latter includes those with partial response, stable disease, or progressive disease. Those with IR were followed up with serial imaging, histopathologic examination, or surgical intervention as decided by the multidisciplinary team. In the current study, we considered the end point as treatment response at 3 months, which was labeled as CR or IR. Three of the authors (I.K., I.P., and Z.H.), serving as treating radiation oncologists, provided the ground truth of response. They identified head and neck nodal targets, planned treatment, and followed target response. Each of them has 15–30 years of oncology experience in treating head and neck cancers. They were completely blinded to the QUS results when determining the response as CR or IR. We subsequently used these radiologist labels as the true labels for supervised learning.

Data Acquisition

The largest metastatic neck lymph node was targeted for QUS imaging as decided by the radiation oncologist, guided by imaging and clinical examination. We acquired US radiofrequency data using a Sonix RP (Ultrasonix) or Elekta imaging system prior to RT initiation (preferably within 24 hours of starting, up to 7 days prior to RT initiation allowed) and after 1 week into treatment in a 4-week RT treatment period. The systems were equipped with linear-array transducers (Sonix RP: L14–5/60, Elekta: 4DL14–5/39) of 6.5-MHz transmit frequency and 3–8-MHz bandwidth. The radiofrequency signal was digitized using a 40-MHz sampling frequency. A research sonographer with more than 5 years of experience in head and neck US imaging performed the scan. We collected radiofrequency images in a panoramic scan, where the sonographer manually moved the transducer in the elevation direction to acquire radiofrequency images traversing the head and neck nodes. Each radiofrequency image consisted of 512 beam-formed A-lines, covering a 6-cm lateral field of view and a 4-cm depth. The transmit foci were 1.75 cm and 2.50 cm for Sonix RP and Elekta, respectively. From the acquired radiofrequency images, we determined sections containing head and neck nodes. From this set of elevation sections, we randomly sampled three to five representative sections at approximately regular intervals for QUS analysis. Manual contouring of the target lymph node was performed on these B-mode US images to acquire the region of interest. One of the authors (A.D.) manually contoured the involved head and neck nodes under the supervision of the three treating radiation oncologists (I.K., I.P., and Z.H.). QUS spectroscopy, texture, and texture-derivate analyses were subsequently performed on the selected regions from the target lymph node. We acquired these representative features before treatment initiation and after 1 week into 6–7 weeks of radiation treatment. We selected the 1-week time point, as we consider it sufficient to assess early response. A similar QUS-based radiomics study that assessed the early response of locally advanced breast tumors to neoadjuvant chemotherapy also used the 1-week time point (6,13).

QUS Spectral Parametric Imaging

A sliding window analysis with a 2 × 2-mm kernel was used to create parametric images for each QUS spectral parameter. We selected the kernel size to include five to 10 acoustic wavelengths for reliable spectral estimation. A 94% overlap was used between adjacent windows in both axial and lateral directions. Prior to spectral analysis, individual radiofrequency scan lines within the kernel were gated using a Hanning function. The power spectrum of the sample was estimated using a fast-Fourier transform technique. An average power spectrum was acquired by averaging several independent adjacent spectra from individual radiofrequency scan lines within the window. A normalization procedure was applied using a reference phantom technique (23,24). The reference phantom was composed of 5–30-μm glass beads embedded in a homogeneous medium of oil droplets immersed in gelatin. The measured attenuation coefficient and speed of sound of the phantom were 0.8 dB/cm/MHz and 1540 m/sec, respectively (the University of Wisconsin, Department of Medical Physics, Madison). We performed an attenuation correction to account for US losses along the propagation path. We assumed an attenuation coefficient of 1 dB/cm/MHz for the overlying head and neck tissues (24,25). The attenuation coefficient estimate for the target lymph node was determined using a spectral difference method. Spectral normalization and attenuation compensation allow parameterization of the tissue microstructures component of the radiofrequency spectra. Spectral and lymph node scattering parameters were estimated: midband fit (MBF), spectral slope, spectral intercept, average scatterer diameter, and average acoustic concentration (AAC). QUS spectral parametric images provide instrument- and operator-independent representations of head and neck cancer nodes that are useful for assessing treatment response. The weighted average values from all the sections for individual spectral parameters were used as first-order imaging features.

Texture Parameters

Quantitative color-coded parametric maps were generated based on the individual values of the spectral parameters from the individual subregions of interest. A two-dimensional texture analysis using the gray-level co-occurrence matrix (GLCM) method was performed on the parametric maps with the gray level intensity linearly scaled into 16 discrete values (26). The GLCM matrices were created from each QUS parametric map at interpixel distances of 1, 2, 3, 4, and 5 pixels and at four angular directions of 0°, 45°, 90°, and 135°. Four textural features of contrast, correlation, energy, and homogeneity were extracted and subsequently averaged over distances and angular directions to generate second-order imaging features (QUS-Tex^1). Therefore, five spectral parametric maps led to the generation of 20 QUS-Tex^1 features.

The third-order imaging features represented the texture derivatives (QUS-Tex^1-Tex^2) (27). The texture-derivative analysis was done by creating intermediate texture-encoded maps using sliding window analysis with a 15 × 15-pixel window, with each pixel in these maps representing the quantification of local textures within the concerned window. A second-pass GLCM

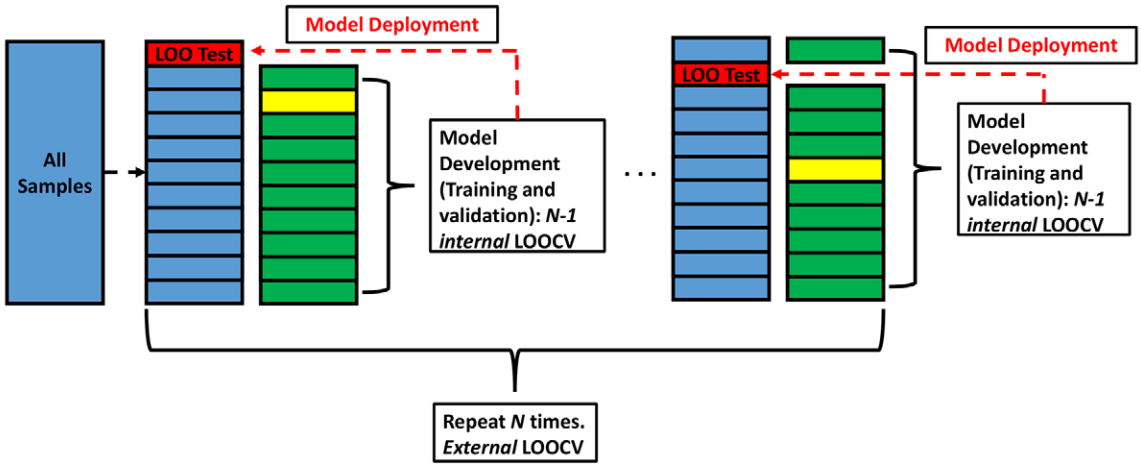


Figure 1: Model building and evaluation strategy. We created *n* external leave-one-out cross-validation (LOOCV) partitions from all samples. For each fold, we have *n*-1 samples for model development and a single LOO test sample, which we kept hidden for final model evaluation. From the *n*-1 samples, we created *n*-1 internal LOOCV partitions, where we fitted a classifier model on the *n*-2 samples and evaluated its performance on an out-of-sample LOO validation sample. We fitted *n*-1 models each time on different *n*-2 samples and eventually averaged the prediction score from all the *n*-1 LOO test samples. We chose the final model as one that resulted in the highest average validation performance. The selected model then predicts the output of the previously hidden LOO test sample. This provides an objective assessment of model performance in production, when predicting new unseen samples. This strategy is robust against overfitting especially when dealing with data in the limited sample regimen, as indicated by Vabalas et al (40). This strategy effectively uses *n*-2 training samples (highlighted in green) for fitting a classifier, one validation sample (highlighted in yellow) for selecting a model and its hyperparameters, and one test sample (highlighted in red) for assessment of model generalization beyond the development samples. Model development that includes feature standardization, filter-based feature selection, data balancing, and wrapper-based feature selection utilized *n*-2 training samples. Model selection and hyperparameter optimization were based on the average LOO validation set performance. Subsequently, we tested the final selected model on each LOO test sample. We aggregated the prediction scores from all LOO test samples to produce the final test confusion matrix, from which classification metrics can be derived.

analysis was performed on the texture maps, resulting in 80 QUS-Tex^1-Tex^2 features.

The weighted average measures of the features were used for building models that predict the response. A total set of 105 QUS radiomics features (five spectral, 20 QUS-Tex^1, 80 QUS-Tex^1-Tex^2) were acquired before starting RT and after 1 week of treatment. The segmentation, feature extraction, and texture analysis were done using MATLAB 2019b (MathWorks).

Prediction Rule Training and Evaluation

We developed a prediction rule that classifies observations into one of the two responses: [0, 1] for CR and IR, respectively. Figure 1 illustrates the model building and evaluation strategy. The preprocessing stage includes data partitioning for nested cross-validation (CV) and feature scaling. We divided the whole data set into outer leave-one-out (LOO) folds (n = 55). In each fold, the LOO observation was reserved in a "vault," serving as test data. Model development and validation were performed on the remaining n-1 observations. We performed feature standardization on training data by using the RobustScaler class (28). Test data were scaled using the same scaling computed from training data. A subset of 50 "good" features was selected based on minimum redundancy maximum relevance criterion (29). Subsequently, data balancing was performed on reduced training data by using a synthetic minority oversampling technique (30).

In each outer LOO fold, we constructed inner LOO folds to train and validate the model. Classification models were fitted on balanced reduced training data. Model performance was evaluated on validation data by using LOOCV. We selected the best combination of three features by using forward-sequential feature selection based on a balanced accuracy metric. Having selected a final model, its generalization performance was further evaluated on LOO test data. These predictions were accumulated over all outer LOO folds to obtain a CV estimate of prediction error.

We compared the performance of different estimators that include linear discriminant analysis, k-nearest neighbors, support vector machines–radial basis function (SVM-RBF), and shallow artificial neural network classifiers. We performed a grid search (GridSearchCV function in Scikit-learn) on training data to find the optimum set of parameters for each estimator that requires externally set parameters. Details on the classifiers have been provided extensively elsewhere (31). Appendix S1 details the hyperparameters of the machine learning classifiers.

We implemented machine learning routines in Python using the Scikit-learn library (28).

Statistical Analysis

We performed statistical analysis using an open source SciPy software package. We checked for any statistically significant differences between the two groups of responders by using either a two-sample t test for normally distributed data or a Mann-Whitney U test for nonnormally distributed data. The Shapiro-Wilk test was used to assess normality. For a statistical test, the threshold for significance was a P value of .05 or less.

We quantified model performance using sensitivity, specificity, balanced accuracy, precision, negative predictive value, and area under the receiver operating characteristic curve

Table 1: Clinical Characteristics and Treatment Outcomes for the Two Response Groups

Parameter	Complete Response (n = 21)	Incomplete Response (n = 34)
Age (y)*	65 (47–80)	59 (39–79)
Sex		
Male	20	33
Female	1	1
Primary site		
Oropharynx	14	25
Larynx	3	3
Hypopharynx	2	1
CUP	2	5
T stage		
T0	2	5
T1	7	5
T2	4	11
T3	2	5
T4	6	8
N stage		
N1	10	12
N2	10	14
N3	1	8
Concurrent therapy		
Cisplatin	16	25
Carboplatin	1	2
Cisplatin, followed by carboplatin	0	1
Cetuximab	1	1
None	3	5
2-year RFS rate (%)	100	59
2-year OS rate (%)	100	82
Node size after 1 week into treatment (mm)†	50.4 ± 10.5	55.0 ± 12.1

Note.—Unless otherwise specified, data are numbers of participants. CUP = carcinoma of unknown primary, OS = overall survival, RFS = recurrence-free survival.
* Values are medians, with ranges in parentheses.
† Values are means ± SDs.

(AUC) metrics. Validation and test performances are presented as described below.

We performed survival analysis using the Kaplan-Meier method. The date of starting RT was considered to be the start of survival time.

Results

Clinical Characteristics and Treatment Outcomes

A total of 55 participants were included in the study, with 21 complete responders (median age, 65 years [IQR: 47–80 years]) and 34 incomplete responders (median age, 59 years [IQR: 39–79 years]). The clinical and treatment characteristics are presented in Table 1 and Table S1. To summarize, 53 participants were male and two were female; the primary tumor sites were the oropharynx (n = 39), larynx (n = 6), hypopharynx (n = 3), and carcinoma of unknown primary (n = 7). Human papillomavirus p16 immunostaining results were available in 42 participants, of which 36 had positive findings. A number of participants were concurrently administered systemic therapy, with 45 participants undergoing concurrent chemotherapy. On the other hand, two participants were treated with cetuximab for concurrent immunotherapy. We found no evidence of a difference in certain clinical characteristics, including age, sex, primary tumor site, primary tumor, and nodal stages among the response groups. With a median follow-up of 32 months (IQR: 18–47 months), 2-year recurrence-free survival rate for complete responders and incomplete responders was 100% and 59%, respectively (P < .01). The 2-year overall survival rate for the participants with CR and IR was 100% and 82%, respectively (P = .03). A per-participant summary is provided in Table S1.

Feature Analysis

Figure 2 presents representative B-mode US images and parametric images of average scatterer diameter, AAC, MBF, spectral slope, and spectral intercept for one participant each from the CR and IR groups. The changes within the region of interest (lymph node) can be appreciated at visual evaluation after 1 week of RT, suggesting the differences in scatterer properties induced by the treatment. At the end of 1 week, five features demonstrated statistically significant differences between the two response groups. These include one spectral (change in the mean value of the AAC map), one texture (change in the energy texture of the AAC map), and three texture-derivative (change in the correlation texture of the contrast map of MBF, change in the contrast texture of the contrast map of MBF, change in the correlation texture of the homogeneity map of MBF) features (Table 2). A diagrammatic representation of the normalized feature values for all the participants from pretreatment to week 1 is shown in Figure 3.

Classification Performance

Performances using the different machine learning classifiers before treatment and after 1 week of RT are summarized in Table 3. Before the treatment initiation, performances were roughly equivalent independent of the classifier used. The best performances were obtained using the SVM classifier, with sensitivity, specificity, balanced accuracy, and AUC of 82%, 80%, 81%, and 0.82, respectively. For all the classifiers, the

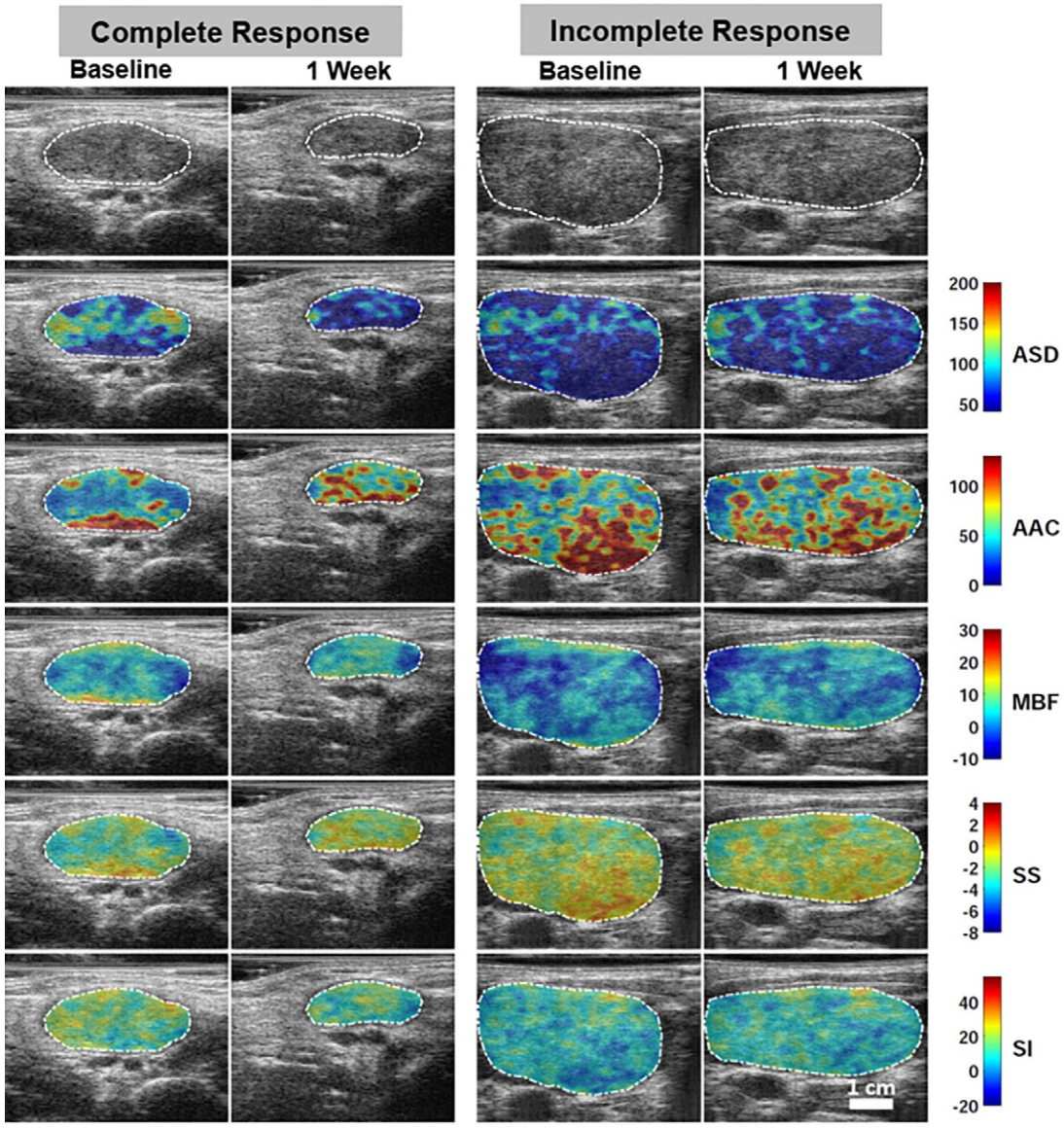


Figure 2: Representative B-mode US and quantitative US (QUS) spectral parametric images of average scatterer diameter (ASD), average acoustic concentration (AAC), midband fit (MBF), spectral slope (SS), and spectral intercept (SI) in one participant (a 74-year-old woman) with complete response (left two columns) and one participant (a 61-year-old man) with incomplete response (right two columns) acquired at baseline (before radiation therapy) and after week 1 of radiation therapy. QUS parametric images include the largest involved cervical lymph node (central region bounded by closed dotted white curve). The color bar range is 160 μm for ASD, 130 dB/cm^3 for AAC, 40 dB for MBF, 12 dB/MHz for SS, and 75 dB for SI. The scale bar represents 1 cm.

performance increased when the delta (Δ) features were included after 1 week of RT. On a test set with limited data after the 1st week of RT, the best-performing classification models using baseline features achieved sensitivity, specificity, balanced accuracy, and AUC values of 67% (14 of 21), 71% (24 of 34), 69%, and 0.73 (95% CI: 0.58, 0.87), respectively. The best-performing classification models using changes in features from week 0 to week 1 achieved sensitivity, specificity, balanced accuracy, and AUC values of 71% (15 of 21), 76% (26 of 34), 74%, and 0.77 (95% CI: 0.63, 0.90), respectively. Models developed using linear discriminant analysis and SVM-RBF classifiers performed comparably at both baseline and week 1. Models developed using imaging features early into treatment performed better in predicting head and neck nodal response to radiation treatment compared with their baseline model counterparts. Figure 4A represents the two response groups, using a set of three Δ features, suggesting visually evident different values between participants with CR and IR. Figures 4B and 4C show the receiver operating characteristic plot using the four machine learning classifiers on the validation and test sets, respectively.

Discussion

RT is the primary treatment modality in patients with HNSCC having primary sites in the oropharynx, hypopharynx, and larynx (17). Although treatment-related cell death is expected to start from the initiation of RT, the final response is manifested from cumulative cell death. The standard treat-

Table 2: Delta Features after Standardization Preprocessing with Statistically Significant Differences between Participants with Complete Response and Incomplete Response

Feature	Complete Response	Incomplete Response	*P* Value
ΔMBF-CON-COR	−0.73 ± 0.17	0.26 ± 0.16	<.001
ΔMBF-CON-CON	0.45 ± 0.22	−0.45 ± 0.15	.001
ΔMBF-HOM-COR	−0.63 ± 0.17	0.20 ± 0.17	.002
ΔAAC-ENE	−0.20 ± 0.25	0.21 ± 0.15	.047
ΔAAC	0.18 ± 0.17	−0.41 ± 0.18	.049

Note.—Unless otherwise noted, values are means ± standard errors of the means. ΔAAC = change in mean value of average acoustic concentration map, ΔAAC-ENE = change in energy texture of average acoustic concentration map, ΔMBF-CON-CON = change in contrast texture of contrast map of midband fit, ΔMBF-CON-COR = change in correlation texture of contrast map of midband fit, ΔMBF-HOM-COR = change in correlation texture of homogeneity map of midband fit.

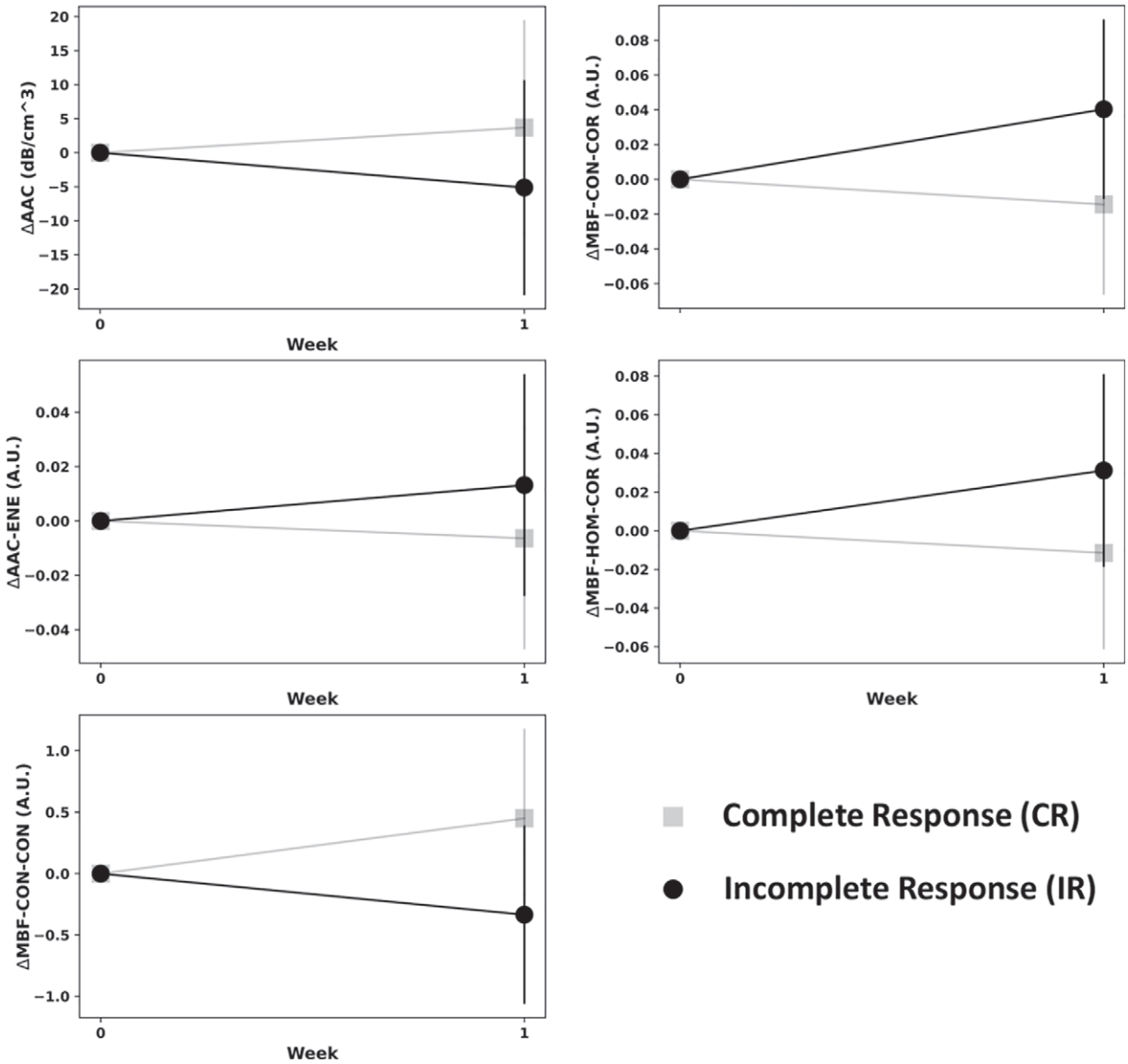


Figure 3: Changes of the mean feature values for parameters with statistically significantly different values between participants with complete response (gray) and incomplete response (black) after 1 week of radiation therapy. The feature estimates for the two response groups have been normalized to the same value before starting radiation therapy, and the relative changes in the mean parameter values with 95% CIs are indicated at a week 1 time point. The representative features include ΔAAC (change in the mean value of average acoustic concentration map), ΔAAC-ENE (change in the energy texture of average acoustic concentration map), ΔMBF-CON-CON (change in the contrast texture of contrast map of midband fit), ΔMBF-CON-COR (change in the correlation texture of contrast map of midband fit), and ΔMBF-HOM-COR (change in the correlation texture of homogeneity map of midband fit). A.U. = arbitrary units.

Table 3: Validation and Test Performances of Classifiers before Treatment and at Week 1 of Radiation Therapy

Classifier	Sensitivity (%)	Specificity (%)	Balanced Accuracy (%)	AUC*	PPV (%)	NPV (%)
Validation Performance						
Baseline						
LDA	76 ± 5	79 ± 3	77 ± 3	0.76 ± 0.02	69 ± 4	84 ± 3
KNN	82 ± 7	77 ± 5	79 ± 4	0.79 ± 0.04	69 ± 4	88 ± 4
SVM-RBF	82 ± 7	77 ± 7	79 ± 2	0.79 ± 0.03	69 ± 5	88 ± 4
ANN	79 ± 7	80 ± 4	79 ± 4	0.83 ± 0.05	71 ± 4	86 ± 4
Week 1						
LDA	79 ± 4	80 ± 3	80 ± 2	0.81 ± 0.02	71 ± 3	86 ± 2
KNN	83 ± 6	80 ± 6	81 ± 4	0.81 ± 0.04	72 ± 6	88 ± 4
SVM-RBF	82 ± 4	80 ± 4	81 ± 2	0.82 ± 0.02	72 ± 4	88 ± 2
ANN	82 ± 7	80 ± 5	81 ± 3	0.83 ± 0.05	72 ± 4	88 ± 4
Test Performance						
Baseline						
LDA	67 (14/21)	71 (24/34)	69	0.71 (0.56, 0.86)	58 (14/24)	77 (24/31)
KNN	71 (15/21)	65 (22/34)	68	0.65 (0.50, 0.81)	56 (15/27)	79 (22/28)
SVM-RBF	67 (14/21)	71 (24/34)	69	0.73 (0.58, 0.87)	58 (14/24)	77 (24/31)
ANN	43 (9/21)	62 (21/34)	52	0.59 (0.44, 0.75)	41 (9/22)	64 (21/33)
Week 1						
LDA	71 (15/21)	76 (26/34)	74	0.75 (0.61, 0.89)	65 (15/23)	81 (26/32)
KNN	71 (15/21)	62 (21/34)	67	0.69 (0.54, 0.84)	54 (15/28)	78 (21/27)
SVM-Linear	71 (15/21)	76 (26/34)	74	0.77 (0.63, 0.90)	65 (15/23)	81 (26/32)
ANN	52 (11/21)	71 (24/34)	61	0.67 (0.52, 0.82)	52 (11/21)	71 (24/34)

Note.—Unless otherwise noted, data are presented as means ± SDs across leave-one-out folds in the development set (validation performance) and percentages, with numerators and denominators in parentheses, in the test set (test performance). ANN = artificial neural network, AUC = area under the receiver operating characteristic curve, KNN = k-nearest neighbor, LDA = linear discriminant analysis, NPV = negative predictive value, PPV = positive predictive value, SVM-RBF = support vector machines–radial basis function.
* Data in parentheses are 95% CIs.

ment includes a fractionated course of RT over several weeks, with response evaluation carried out typically 3 months after treatment completion. Developing an appropriate biomarker to detect treatment response during the early phase of RT in real time forms a strategy that could be used to adopt personalized radiation regimens. QUS is an inexpensive portable imaging modality able to help detect the underlying tissue microstructural elastic properties. In the current study, the work has demonstrated the ability of a Δ-radiomics model from QUS imaging of metastatic lymph nodes after 1 week of treatment to predict treatment response after 3 months of RT completion.

In the past decade, substantial research has been directed toward the development of noninvasive biomarkers by using quantitative analytics of imaging aided by artificial intelligence (3). For patients with head and neck malignancies, radiomics analysis has been used more commonly with imaging modalities like MRI, CT, and PET and less commonly with US, with clinical end points like prognostication, molecular characteristics, and survival prediction (32,33). In a multicentric study, CT radiomics features were able to predict human papillomavirus status in patients with oropharyngeal squamous cell carcinoma (34). Several studies have demonstrated the ability of quantitative analysis of CT and PET imaging features to help predict recurrence and survival in patients with head and neck cancer (35–37). Similarly, MRI radiomics have demonstrated value in the prognostication of patients with HNSCC, with a major share of such studies performed in nasopharyngeal cancer (38,39). Although survival outcomes or probabilities of recurrence are commonly studied end points in radiomics studies

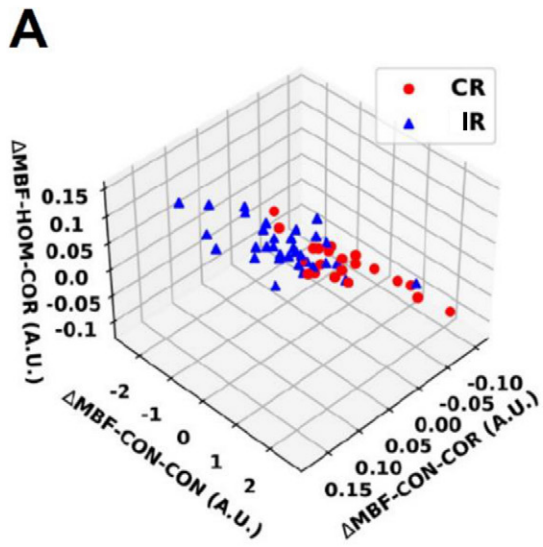
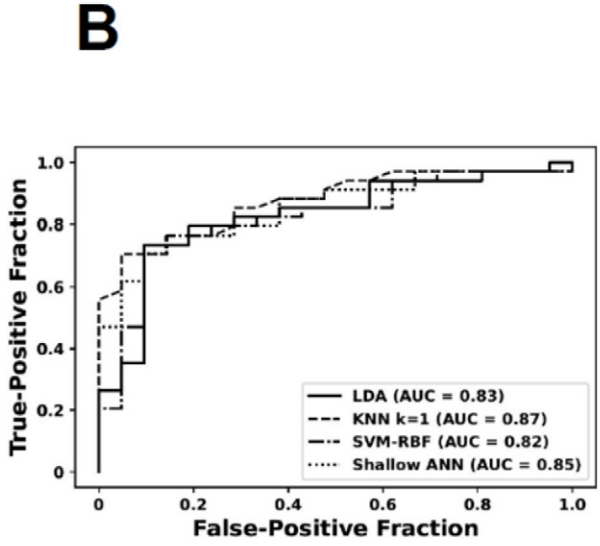
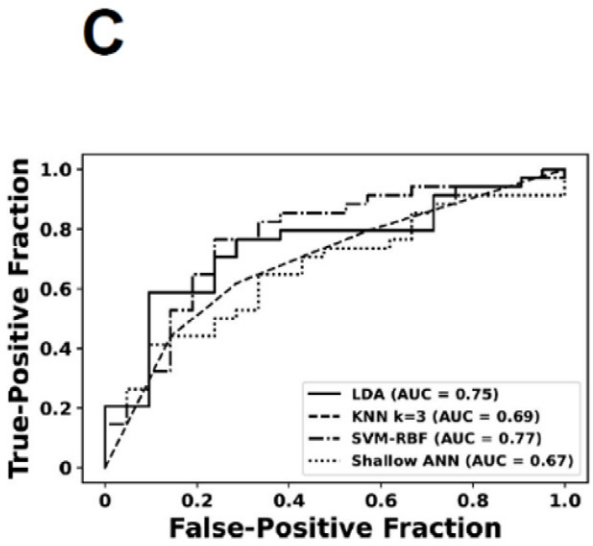


Figure 4: **(A)** Scatterplots in three-dimensional plane using three delta quantitative US features: ΔMBF-CON-COR (change in the correlation texture of contrast map of midband fit), ΔMBF-CON-CON (change in the contrast texture of contrast map of midband fit), and ΔMBF-HOM-COR (change in the correlation texture of homogeneity map of midband fit). Red circles show the participants with complete response (CR), while blue triangles represent those with incomplete response (IR). **(B, C)** Receiver operating characteristic plots for the radiomics models using different machine learning classifiers for the validation and test set, respectively. The classifiers included LDA, KNN (*k* = 3), SVM-RBF, and a shallow ANN. ANN = artificial neural network, A.U. = arbitrary units, AUC = area under the receiver operating characteristic curve, KNN = k-nearest neighbors, LDA = linear discriminant analysis, SVM-RBF = support vector machine–radial basis function.





involving HNSCC, quantitative imaging for the evaluation of response to RT has been seldom reported.

QUS spectroscopy transforms time-domain radiofrequency data into its spectral-domain representation, enabling the extraction of quantitative parameters that provide more detailed microstructural information compared with standard sonography. With direct analysis of the normalized power spectrum obtained from the target region, the biophysical properties of the tissues provide enriched information that can be strongly linked with clinical outcomes. The mechanical properties of the tissue are dependent on several factors like cell size, shape, diameter, and arrangement, which are represented by different QUS parameters (4,6). Earlier studies have shown the ability of QUS features obtained before initiation of treatment in predicting the response to neoadjuvant chemotherapy in locally advanced breast cancer or response to RT for patients with HNSCC (7,8,10). Also, pretreatment QUS radiomics models have been able to predict the risk of recurrence in patients with breast and head and neck malignancies (9,15). The ability of baseline QUS features to determine the clinical course before the initiation of treatment provides evidence for the relation of different tissue microstructural elastic properties with biologic behavior. With the initiation of cancer-directed treatment like chemotherapy or RT, changes induced by cell death start within the tumor microenvironment, leading to morphologic changes such as cell fragmentation, pyknosis, and apoptosis. Preclinical studies have demonstrated the ability of QUS parameters to effectively determine the changes in the physical properties of the cellular and nuclear architecture induced by treatment results in changes in the scatterer properties (12). In clinical studies involving breast and head and neck malignancies, QUS features obtained during the early course were shown to improve classification performances of the radiomics model (13,14).

In the current study, we used a Δ-radiomics model incorporating QUS features, obtained 1 week after RT, from metastatic neck nodes in participants with HNSCC scheduled for a 7-week course of RT treatment. Five of the imaging features were observed to have a significant difference in values in the

two response groups (CR and IR) after the 1st week of treatment. The changes in AAC, one texture feature of AAC (AAC-energy), and three MBF texture-derivative features turned out to have different trajectories after treatment initiation. The AAC parameter is influenced by the scatterer number density and organization, while the MBF is dependent upon scatterer size, shape, number, and organization (4,6). In an earlier study, we demonstrated the improved performance of a QUS radiomics model in a limited number of individuals (n = 36) using feature values after the 1st and 4th week of RT (14). The current study included a larger number of individuals (n = 55), and higher-dimensional texture features in the form of texture derivatives were extracted. Previous studies showed the improvement of classification indexes with the inclusion of texture-derivative features as opposed to QUS spectral and texture features alone, suggesting better representation of tumor heterogeneity with the use of higher-order imaging features (8,27). Univariate assessment of features for any statistically significant difference necessitates correction to mitigate the increased risk of false-positive findings, especially when a large number of features are involved. Still, univariate assessment of feature quality does not necessarily predict the performance of multivariate models. A combination of features with less discriminating power can still produce a reasonable generalization performance.

When dealing with small biomedical data sets of cancer images from the real world, it is imperative to implement an appropriate model building and evaluation strategy (40). Here, we employed a nested LOOCV analysis, allowing the separation of observations for model development and testing. As we acquire more data, we will be able to implement development-test partitioning that leaves out 10%–30% of samples for testing. Tumors are heterogeneous and vary from patient to patient in terms of certain structural features and often have necrotic areas or areas of variable histologic features. In the limited data regimen, the implementation of a holdout set of 10%–30% would not be appropriate due to the effect of data granularity that can make certain important characteristics underrepresented in the development set. A nested LOO analysis guarantees that most of the limited available samples are used for model building, while at the same time preserving the strict requirement of out-of-sample data for the objective evaluation of model performance.

On the other hand, in the large data regimen, such as when n is approximately several hundred samples, the use of LOOCV will require excessive computational resources, as several hundred LOO partitions require evaluation. In this regimen, a holdout validation approach is best, as data granularity effect will be less prominent. Subsequently, it is necessary to test the feasibility of the framework on a multicentric data set. Previous studies have demonstrated the robustness of the QUS spectroscopy and texture framework to US system variations and multi-institutional data (41,7,13).

This study demonstrates that changes in radiomics features early in the treatment process can be used to construct a response-predictive model with good generalization performance. Our results should encourage further studies to evaluate the proposed framework on larger patient cohorts. The response-predictive model was built by representing the characteristics of QUS spectral parametric images of the head and neck nodes with first-order mean and GLCM-based texture analyses. Future work will involve analyzing a full set of radiomics features, including features beyond first-order mean and GLCM-based textures, such as other first-order statistics features, morphologic features, and different methods for quantifying image texture. As we collect more samples, we will be able to apply more advanced models, such as convolutional neural networks and vision transformer models, as generic feature extractors to directly learn important characteristics from QUS spectral parametric images of the head and neck nodes are essential for predicting treatment response. We believe that both traditional machine learning with a comprehensive set of radiomics features and deep learning approaches can lead to even better generalization performance.

A recent study has demonstrated the use of contrast-enhanced US for characterizing breast lesions (42). Contrast-enhanced harmonic and subharmonic imaging allow the analysis of vascular heterogeneity and perfusion, which are useful for distinguishing malignant from benign breast tumors (42). A similar strategy can potentially be implemented for assessing treatment responses in head and neck nodes. The characterization of node vasculature can add a distinct dimension in representing head and neck nodes, potentially benefiting treatment assessment.

The ability to detect tumor response to real-time RT early in the course of treatment provides a window of opportunity for personalized RT. Treatment modifications can be pursued on an individual basis, depending on the anticipated response. This may involve treatment escalation for refractory disease or treatment de-escalation for radiosensitive disease, aiming to strike an optimal balance between cure and treatment-related toxicity.

The current study had several limitations, including the relatively small cohort size and the single-institutional nature of the data. The feasibility of the QUS spectroscopy framework has been demonstrated for the diagnosis and prognosis evaluation of different tumors in a larger cohort (43,44,8). However, we still need to perform a similar large-scale analysis for the QUS-based radiomics study of head and neck cancers. In this preliminary study, we have demonstrated that the QUS spectral and texture representation of head and neck cancer can be utilized to build a response prediction model with a balanced accuracy of 74% and an AUC of 0.77 evaluated on the test data.

In conclusion, this prospective observational study demonstrated the potential of a Δ-radiomics model, utilizing QUS spectral and textural features from metastatic neck nodes in participants with HNSCC, collected after the 1st week of radical RT, for predicting treatment response. Notably, the AAC and MBF QUS parameters exhibited significant changes after the 1st week of RT in participants with either a CR or an IR 3 months after treatment completion. These promising results have spurred further ongoing research, aiming to involve a larger patient cohort. This expanded data set will enable the development of a more robust radiomics model, laying the foundation for future interventional studies. While the primary tumor was not imaged in this study because of its deep-seated location and potential air-tissue interface influences on QUS parameters, the

higher-dimensional radiomics features extracted from the lymph node were effective for predicting treatment responses.

Acknowledgments: We would like to thank all the patients for their participation in the study. Our sincere gratitude to the physicians and other health care staff for their support in patient care. We express our regard to the Terry Fox Foundation and the Natural Sciences and Engineering Research Council of Canada for funding support. We would like to thank Daniel Palhares, MD, MSc, for fruitful discussions.

Author contributions: Guarantors of integrity of entire study, **L.O.O., A.D., M.S., I.K., W.T.T., G.J.C.**; study concepts/study design or data acquisition or data analysis/interpretation, all authors; manuscript drafting or manuscript revision for important intellectual content, all authors; approval of final version of submitted manuscript, all authors; agrees to ensure any questions related to the work are appropriately resolved, all authors; literature research, **L.O.O., A.D., K.F., M.S., W.T.T., L.S., G.J.C.**; clinical studies, **L.O.O., A.D., D.D., K.Q., M.S., I.K., I.P., W.T.T., L.S., G.J.C.**; statistical analysis, **L.O.O., A.D., K.F., K.Q., M.S., L.S., G.J.C.**; and manuscript editing, **L.O.O., A.D., K.F., M.S., I.K., I.P., Z.H., W.T.T., L.S., G.J.C.**

Data sharing: Data generated or analyzed during the study are available from the corresponding author by request.

Disclosures of conflicts of interest: L.O.O. No relevant relationships. **A.D.** No relevant relationships. **D.D.** Worked in the lab and assisted with study materials. **K.F.** No relevant relationships. **M.S.** No relevant relationships. **I.K.** No relevant relationships. **I.P.** No relevant relationships. **Z.H.** Honoraria and travel support from the American Head and Neck Society, the American Society of Clinical Oncology, the American Society for Radiation Oncology, and the Society for Immunotherapy of Cancer for attending the Multidisciplinary Head and Neck Cancer Symposium. **W.T.T.** No relevant relationships. **L.S.** No relevant relationships. **G.J.C.** No relevant relationships.

References

1. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer 2012;48(4):441–446.
2. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. Radiology 2016;278(2):563–577.
3. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol 2017;14(12):749–762.
4. Mamou J, Oelze ML, eds. Quantitative ultrasound in soft tissues. advances in experimental medicine and biology series. 2nd ed. Springer, 2023.
5. Czarnota GJ, Kolios MC, Abraham 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 1999;81(3):520–527.
6. Sannachi L, Tadayyon H, Sadeghi-Naini A, et al. Non-invasive evaluation of breast cancer response to chemotherapy using quantitative ultrasonic backscatter parameters. Med Image Anal 2015;20(1):224–236.
7. DiCenzo D, Quiaoit K, Fatima K, et al. Quantitative ultrasound radiomics in predicting response to neoadjuvant chemotherapy in patients with locally advanced breast cancer: Results from multi-institutional study. Cancer Med 2020;9(16):5798–5806.
8. Dasgupta A, Brade S, Sannachi L, et al. Quantitative ultrasound radiomics using texture derivatives in prediction of treatment response to neoadjuvant chemotherapy for locally advanced breast cancer. Oncotarget 2020;11(42):3782–3792.
9. Dasgupta A, Fatima K, DiCenzo D, et al. Quantitative ultrasound radiomics in predicting recurrence for patients with node-positive head-neck squamous cell carcinoma treated with radical radiotherapy. Cancer Med 2021;10(8):2579–2589.
10. Osapoetra LO, Dasgupta A, DiCenzo D, et al. Assessment of clinical radiosensitivity in patients with head-neck squamous cell carcinoma from pre-treatment quantitative ultrasound radiomics. Sci Rep 2021;11(1):6117.
11. Kolios MC, Czarnota GJ, Lee M, Hunt JW, Sherar MD. Ultrasonic spectral parameter characterization of apoptosis. Ultrasound Med Biol 2002;28(5):589–597.
12. Vlad RM, Brand S, Giles A, Kolios MC, Czarnota GJ. Quantitative ultrasound characterization of responses to radiotherapy in cancer mouse models. Clin Cancer Res 2009;15(6):2067–2075.
13. Quiaoit K, DiCenzo D, Fatima K, et al. Quantitative ultrasound radiomics for therapy response monitoring in patients with locally advanced breast cancer: Multi-institutional study results. PLoS One 2020;15(7):e0236182.
14. Tran WT, Suraweera H, Quiaoit K, et al. Quantitative ultrasound delta-radiomics during radiotherapy for monitoring treatment responses in head and neck malignancies. Future Sci OA 2020;6(9):FSO624.
15. Fatima K, Dasgupta A, DiCenzo D, et al. Ultrasound delta-radiomics during radiotherapy to predict recurrence in patients with head and neck squamous cell carcinoma. Clin Transl Radiat Oncol 2021;28:62–70.
16. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 2018;68(6):394–424. [Published correction appears in CA Cancer J Clin 2020;70(4):313.]
17. Chow LQM. Head and neck cancer. N Engl J Med 2020;382(1):60–72.
18. Mehanna H, McConkey CC, Rahman JK, et al. PET-NECK: a multicentre randomised Phase III non-inferiority trial comparing a positron emission tomography-computerised tomography-guided watch-and-wait policy with planned neck dissection in the management of locally advanced (N2/N3) nodal metastases in patients with squamous cell head and neck cancer. Health Technol Assess 2017;21(17):1–122.
19. Eriksson D, Stigbrand T. Radiation-induced cell death mechanisms. Tumour Biol 2010;31(4):363–372.
20. Sia J, Szmyd R, Hau E, Gee HE. Molecular mechanisms of radiation-induced cancer cell death: a primer. Front Cell Dev Biol 2020;8:41.
21. Caudell JJ, Torres-Roca JF, Gillies RJ, et al. The future of personalised radiotherapy for head and neck cancer. Lancet Oncol 2017;18(5):e266–e273.
22. Rühle A, Grosu AL, Nicolay NH. De-escalation strategies of (chemo)radiation for head-and-neck squamous cell cancers-HPV and beyond. Cancers (Basel) 2021;13(9):2204.
23. Yao LX, Zagzebski JA, Madsen EL. Backscatter coefficient measurements using a reference phantom to extract depth-dependent instrumentation factors. Ultrason Imaging 1990;12(1):58–70.
24. Labyed Y, Bigelow TA, McFarlin BL. Estimate of the attenuation coefficient using a clinical array transducer for the detection of cervical ripening in human pregnancy. Ultrasonics 2011;51(1):34–39.
25. Labyed Y, Bigelow TA. Estimating the total ultrasound attenuation along the propagation path by using a reference phantom. J Acoust Soc Am 2010;128(5):3232–3238.
26. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. IEEE Trans Syst Man Cybern 1973;SMC-3(6):610–621.
27. Osapoetra LO, Sannachi L, Quiaoit K, et al. *A priori* prediction of response in multicentre locally advanced breast cancer (LABC) patients using quantitative ultrasound and derivative texture methods. Oncotarget 2021;12(2):81–94.
28. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: machine learning in Python. J Mach Learn Res 2011;12:2825–2830.
29. Peng H, Long F, Ding C. Feature selection based on mutual information: criteria of max-dependency, max-relevance, and min-redundancy. IEEE Trans Pattern Anal Mach Intell 2005;27(8):1226–1238.
30. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. J Artif Intell Res 2002;16:321–357.
31. Hastie T, Tibshirani R, Friedman JH. The elements of statistical learning: data mining, inference, and prediction. 2nd ed. New York, NY: Springer, 2009.
32. Wong AJ, Kanwar A, Mohamed AS, Fuller CD. Radiomics in head and neck cancer: from exploration to application. Transl Cancer Res 2016;5(4):371–382.
33. van Dijk LV, Fuller CD. Artificial Intelligence and Radiomics in Head and Neck Cancer Care: Opportunities, Mechanics, and Challenges. American Society of Clinical Oncology Educational Book, Volume 41, Number 41. American Society of Clinical Oncology, 2021; 1–11.
34. Leijenaar RT, Bogowicz M, Jochems A, et al. Development and validation of a radiomic signature to predict HPV (p16) status from standard CT imaging: a multicenter study. Br J Radiol 2018;91(1086):20170498.
35. Parmar C, Grossmann P, Rietveld D, Rietbergen MM, Lambin P, Aerts HJWL. Radiomic machine-learning classifiers for prognostic biomarkers of head and neck cancer. Front Oncol 2015;5:272.
36. Vallières M, Kay-Rivest E, Perrin LJ, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Sci Rep 2017;7(1):10117.
37. Bogowicz M, Riesterer O, Stark LS, et al. Comparison of PET and CT radiomics for prediction of local tumor control in head and neck squamous cell carcinoma. Acta Oncol 2017;56(11):1531–1536.
38. Jethanandani A, Lin TA, Volpe S, et al. Exploring applications of radiomics in magnetic resonance imaging of head and neck cancer: a systematic review. Front Oncol 2018;8:131.
39. Mes SW, van Velden FHP, Peltenburg B, et al. Outcome prediction of head and neck squamous cell carcinoma by MRI radiomic signatures. Eur Radiol 2020;30(11):6311–6321.
40. Vabalas A, Gowen E, Poliakoff E, Casson AJ. Machine learning algorithm validation with a limited sample size. PLoS One 2019;14(11):e0224365.

41. Sannachi L, Gangeh M, Naini AS, et al. Quantitative ultrasound monitoring of breast tumour response to neoadjuvant chemotherapy: comparison of results among clinical scanners. Ultrasound Med Biol 2020;46(5):1142–1157.
42. Forsberg F, Piccoli CW, Sridharan A, et al. 3D harmonic and subharmonic imaging for characterizing breast lesions: a multi-center clinical trial. J Ultrasound Med 2022;41(7):1667–1675.
43. Osapoetra LO, Sannachi L, DiCenzo D, Quiaoit K, Fatima K, Czarnota GJ. Breast lesion characterization using Quantitative Ultrasound (QUS) and derivative texture methods. Transl Oncol 2020;13(10):100827.
44. Osapoetra LO, Chan W, Tran W, Kolios MC, Czarnota GJ. Comparison of methods for texture analysis of QUS parametric images in the characterization of breast lesions. PLoS One 2020;15(12):e0244965.