International Journal of Radiation Oncology biology • physics

www.redjournal.org

Physics Contribution

Quantitative Thermal Imaging Biomarkers to Detect Acute Skin Toxicity From Breast Radiation Therapy Using Supervised Machine Learning

Khadijeh Saednia, MSc,[*,†] Sami Tabbarah, BSc, MSc,[*,‡] Andrew Lagree, BSc, MSc,[*] Tina Wu, BSc,[*] Jonathan Klein, MD, MSc, FRCPC, DABR,[§] Eduardo Garcia, BSc,[*] Michael Hall, BSc,[*] Edward Chow, MSc, PhD, MBBS, FRCPC,[*,‖] Eileen Rakovitch, MSc, MD, FRCPC,[*,‖] Charmaine Childs, PhD,[#] Ali Sadeghi-Naini, PhD,[*,†,**] and William T. Tran, MRT(T), MSc, PhD[*,‡,‖,#,††]

[*]Department of Radiation Oncology, Sunnybrook Health Sciences Centre, Toronto, Canada; [†]Department of Electrical Engineering and Computer Science, York University, Toronto, Canada; [‡]Evaluative Clinical Sciences Platform, Sunnybrook Research Institute, Toronto, Canada; [§]Department of Radiation Oncology, Albert Einstein College of Medicine, New York City, New York; [‖]Department of Radiation Oncology, University of Toronto, Toronto, Canada; [#]Department of Radiotherapy & Oncology, Sheffield Hallam University, Sheffield, United Kingdom; [**]Physical Sciences Platform, Sunnybrook Research Institute, Toronto, Canada; and [††]Department of Biomedical Physics, Ryerson University, Toronto, Canada

Received Jul 11, 2019. Accepted for publication Dec 24, 2019.

Summary

Quantitative thermal imaging can be used to detect physiological changes associated with radiation-induced dermatitis and can be integrated with machine learning

Purpose: Radiation-induced dermatitis is a common side effect of breast radiation therapy (RT). Current methods to evaluate breast skin toxicity include clinical examination, visual inspection, and patient-reported symptoms. Physiological changes associated with radiation-induced dermatitis, such as inflammation, may also increase body-surface temperature, which can be detected by thermal imaging. Quantitative thermal imaging markers were identified and used in supervised machine learning to develop a predictive model for radiation dermatitis.

Corresponding author: William T. Tran, MRT(T), MSc, PhD; E-mail: william.tran@sunnybrook.ca

Funding for this study was provided by the Terry Fox Foundation, Canada (grant number: 1083), the Kavelman Fonn Foundation, the Women's Health Golf Classic Fund, and the Natural Sciences and Engineering Research Council of Canada (grant number RGPIN-2016-06472). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Disclosures: The authors declare that there are no conflicts of interests to disclose. Dr Eileen Rakovitch reports grants from Genomic Health Inc outside the submitted work.

Supplementary material for this article can be found at https://doi.org/10.1016/j.ijrobp.2019.12.032.

Acknowledgments—We would like to acknowledge the Department of Radiation Oncology at the Odette Cancer Centre and Sunnybrook Health Science Centre, particularly Senior Manager Mr. Stephen Russell, for their support of this project. We wish to also thank Dr. Calvin Law, Ms. Angela Leahey, Ms. Jan Stewart, and Ms. Brianna Law for their research support. This project was partly funded by the Terry Fox Foundation and we thank them for their support.

https://doi.org/10.1016/j.ijrobp.2019.12.032

frameworks to develop a predictive tool for skin-toxicity at early treatment times. Ninety patients treated for whole-breast radiotherapy were recruited for the study. Thermal images of the treated breast were taken at four intervals. Machine learning models demonstrated early thermal signals associated with skin-toxicity after the fifth radiotherapy fraction with high prediction accuracy.

Methods and Materials: Ninety patients treated for adjuvant whole-breast RT (4250 cGy/f_x = 16) were recruited for the study. Thermal images of the treated breast were taken at 4 intervals: before RT, then weekly at f_x = 5, f_x = 10, and f_x = 15. Parametric thermograms were analyzed and yielded 26 thermal-based features that included surface temperature (°C) and texture parameters obtained from (1) gray-level co-occurrence matrix, (2) gray-level run-length matrix, and (3) neighborhood gray-tone difference matrix. Skin toxicity was evaluated at the end of RT using the Common Terminology Criteria for Adverse Events (CTCAE) guidelines (Ver.5). Binary group classes were labeled according to a CTCAE cut-off score of ≥2, and thermal features obtained at f_x = 5 were used for supervised machine learning to predict skin toxicity. The data set was partitioned for model training, independent testing, and validation. Fifteen patients (∼17% of the whole data set) were randomly selected as an unseen test data set, and 75 patients (∼83% of the whole data set) were used for training and validation of the model. A random forest classifier with leave-1-patient-out cross-validation was employed for modeling single and hybrid parameters. The model performance was reported using receiver operating characteristic analysis on patients from an independent test set.

Results: Thirty-seven patients presented with adverse skin effects, denoted by a CTCAE score ≥2, and had significantly higher local increases in skin temperature, reaching 36.06°C at f_x = 10 (P = .029). However, machine-learning models demonstrated early thermal signals associated with skin toxicity after the fifth RT fraction. The cross-validated model showed high prediction accuracy on the independent test data (test accuracy = 0.87) at f_x = 5 for predicting skin toxicity at the end of RT.

Conclusions: Early thermal markers after 5 fractions of RT are predictive of radiation-induced skin toxicity in breast RT.

Introduction

Radiation therapy (RT) uses ionizing radiation to target residual cancer cells of the breast and induce cellular death. This results in the reduction of locoregional and distant cancer recurrence after lumpectomy (15.7%; 95% confidence interval, 13.7-17.7; $P < .00001$)[1] or mastectomy (rate ratio = 0.68; 95% confidence interval, 0.57-0.82; P = .00006).[2] However, RT is associated with dermatologic risks such as erythema, which affects approximately 90% of treated patients.[3,4] The skin is a highly proliferative organ and is therefore susceptible to radiation damage and toxicity.[5] Cellular apoptotic and necrotic events are initiated in the skin from repeated and fractionated exposure to radiation.[5,6] Cellular death leads to the recruitment of cytokines, prompting an inflammatory response (acute dermatitis) that stimulates the transendothelial migration of immune cells to the target site.[5,6] Consequently, blood vessels dilate causing an increase in blood volume and the damaged irradiated skin clinically presents as erythema and may lead to more severe effects such as desquamation.[3] As a result, patients undergoing RT are carefully monitored using standard assessment tools such as the Common Terminology Criteria for Adverse Events (CTCAE) system to manage toxicity burden.[7] Major challenges associated with visual inspection of the breast to evaluate skin-related symptoms include low sensitivity and specificity for detecting early signs of dermatitis in addition to differentiating the degree of severe skin toxicity (ie, CTCAE ≥ 2). This is caused by practitioner bias, underreporting by patients, and clinician expertise that may lead to variability in grading skin toxicity.[7-9] Topographic imaging modalities, such as quantitative thermal imaging (QTI), have the potential to overcome these challenges and objectively measure the changes in surface skin temperature associated with radiation-induced skin dermatitis. Exploiting QTI and machine-learning frameworks (ie, thermoradiomics) may yield actionable insight into symptom management during RT.

QTI has previously been used to identify temperature changes associated with alterations in blood flow and disease presentation.[10-12] Within the RT setting, Maillot et al[13] investigated the use of thermography to quantitatively evaluate skin toxicity. Patients who were classified using the CTCAE and Radiation Therapy Oncology Group criteria demonstrated a high-grade of skin toxicity (≥2) that was also associated with an increase in the average local temperature ($P < .05$). Furthermore, Maillot et al[13] found that thermography-derived temperature features recorded a week before the clinical presentation of skin toxicity had a predictive value of 70%. Their study also demonstrated a higher incidence of high-grade radiation-induced dermatitis after 10 to 15 RT fractions. Other QTI applications for breast cancer include using first-order temperature features and second-order features such as thermogram-texture parameters for detecting breast lesions. Milosevic et al tested

the feasibility of thermal breast imaging to screen for malignancies. The study exploited machine-learning classification algorithms to distinguish thermal features associated with benign versus malignant masses in the breast. Second-order features included gray-level co-occurrence matrix (GLCM) features, and the results of the study demonstrated a diagnostic accuracy of 92.5%.[14] Potential applications and advantages of employing image guided decision support tools, such as QTI, include early intervention and preventive therapeutics that could mitigate radiation-induced skin toxicity. Current treatment strategies for managing skin toxicity include the use of glucocorticosteroids to inhibit the inflammatory response that mediates acute skin dermatitis.[15,16] Early thermoradiomic markers for skin toxicity would potentially allow radiation oncologists to target patients for prophylactic corticosteroid use, which has been shown to reduce the incidence of radiation-induced dermatitis.[17]

Although the capability of thermal imaging to monitor the occurrence of disease has been demonstrated in previous studies, the potential of using temperature-based and textural features as imaging biomarkers for radiation-induced dermatitis remains unclear and warrants further investigation.[12,13] Here, we investigated QTI and machine-learning frameworks (ie, thermoradiomics) to develop a predictive tool for radiation-induced skin toxicity at early treatment time intervals. This study aimed to measure thermal characteristics of the irradiated skin in patients with breast cancer. We hypothesized that radiation-induced skin toxicity is associated with an increase in skin-surface temperature and changes in thermoradiomic parameters.

Methods and Materials

Patient recruitment criteria and radiation treatment parameters

This study was carried out at a single academic health center and approved by the institutional research ethics board. Participants signed an informed consent form before enrollment. Patients were included in the study based on the following inclusion criteria: confirmed diagnosis of invasive breast cancer or ductal carcinoma in situ, age (18+), and undergoing adjuvant hypofractionated RT to the whole breast or chestwall (4250 cGy/f_x = 16). Patients were excluded from the study if their planned treatment position was prone, undergoing expander-implant breast reconstruction, or had a history of inflammatory breast cancer.[18,19] Ninety patients were prospectively recruited to the study, and clinical characteristics are presented in Table 1. All patients completed a full whole-breast or chestwall treatment course, that is, 16 fractions of adjuvant RT, and received standard skin management strategies consisting of saline rinses only. As part of the study protocol, patients were not prescribed topical corticosteroids (eg, hydrocortisone) or topical antibiotics (eg, bacitracin/polymyxin B or silver sulfadiazine cream) during their treatment course. Figure 1 summarizes the study workflow and methods.

Data acquisition: Quantitative thermal imaging and clinical information

Thermal infrared imaging data were captured using a forward-looking infrared (FLIR) E53 advanced thermal imaging device (FLIR, Wilsonville, USA). Images were acquired from the ipsilateral and contralateral breasts at the following time intervals: baseline (before RT) and after every fifth RT fraction (f_x = 5, f_x = 10, and f_x = 15). Imaging settings were kept constant throughout the time series for each patient. A fixed distance of 2 m was used between the patient and camera. Thermograms were reconstructed at a resolution of 240 × 180 pixels. The instantaneous field of view, which comprises the pixel size on the measurement surface (ie, breast or chestwall) was 3.54 mm. Thermal images were captured using a fixed emissivity setting of $\varepsilon = 0.98$.

Long wave infrared thermography was carried out in a designated examination room. The room was controlled for ambient temperature and air flow during examination. Patients were positioned, while standing, with their arm behind their head, exposing the axilla, midline, and inframammary folds. All study participants were evaluated for radiation-induced skin toxicity using the CTCAE Ver. 5 guidelines, and information was recorded in the electronic medical record. The final CTCAE assessment (ie, end of the 16th fraction) was used for ground truth labeling in machine-learning models (described in a later section). Other clinical and demographic information was collected from the electronic medical records and included the following variables: age, cancer diagnosis, clinicopathologic characteristics of the tumor, surgery details, and RT treatment information. Patient ethnicity information was collected through clinical reports and breast cup size was measured as per the standardized North American scale (Table 1).

Other clinical information associated with risk factors for radiation-induced dermatitis were collected from the patient's electronic medical record and recorded in the patient's case report forms. Table 1 includes information on those variables, which includes the administration of adjuvant chemotherapy (y/n), type of adjuvant chemotherapy, local treatment (whole breast only) versus locoregional irradiation (4-field technique involving the chestwall and regional lymph nodes), the radiation dose-to-skin volume (cGy), and menopausal status.[20-22]

Image processing and feature extraction

Data preprocessing was performed to construct parametric thermal images using FLIR's proprietary software development kit (FileReader). Thermal images were normalized using the software development kit before segmentation

Table 1 Demographic and clinical information of the patients (n = 90) involved in the study

Demographic and clinical characteristics	CTCAE ≥ 2 (n = 37) (n)	CTCAE ≥ 2 (n = 37) %	CTCAE ≤ 1 (n = 53) (n)	CTCAE ≤ 1 (n = 53) %	*P* values
Age (years)					.87
Mean age (±SD)	59 (±12)		59 (±12)		
Maximum age	85	—	88	—	
Minimum age	32	—	28	—	
Race					.84
Asian	5	13.5%	14	26.4%	
Black	2	5.4%	0	0%	
Caucasian	26	70.3%	33	62.3%	
Hispanic	1	2.7%	5	9.4%	
Other	3	8.1%	1	1.9%	
Cup size*					.22
A	5	13.5%	11	20.8%	
B	10	27%	6	11.3%	
C	6	16.2%	9	17%	
D	8	21.6%	11	20.8%	
E	2	5.4%	10	18.9%	
Not applicable	6	16.2%	6	11.3%	
Histologic type					.20
IDC	22	59.5%	34	64.2%	
DCIS	12	32.4%	10	18.9%	
ILC	2	5.4%	7	13.2%	
IMC	1	2.7%	0	0%	
Not evaluated	0	0%	2	3.8%	
Tumor grade†					.30
Low	4	10.8%	11	20.8%	
Intermediate	22	59.5%	23	43.4%	
High	11	29.7%	19	35.8%	
Molecular subtype					
ER status:					.46
Positive	29	78.4%	40	75.5%	
Negative	6	16.2%	6	11.3%	
Undefined	2	5.4%	7	13.2%	
PR status:					.46
Positive	24	64.9%	33	62.3%	
Negative	11	29.7%	13	24.5%	
Undefined	2	5.3%	7	13.2%	
HER2 status:					.41
Positive	2	5.4%	6	11.3%	
Negative	33	89.2%	41	77.4%	
Undefined	2	5.4%	6	11.3%	
Tumor size (cm)‡					.83
Average size (±SD)	2.2 (±2.2)	—	2.2 (±1.3)	—	
Maximum size	12.7	—	6.5	—	
Minimum size	0.2	—	0.1	—	
Tumor laterality					1.00
Left	18	48.6%	26	49.1%	
Right	19	51.4%	27	50.9%	
Surgical procedure					.76
Mastectomy	6	16.2%	7	13.2%	
Lumpectomy	31	83.8%	46	86.8%	
Menopausal status					.26
Pre/perimenopausal	9	24.3%	8	15.1%	
Postmenopausal	22	59.5%	41	77.4%	
Unknown	6	16.2%	4	7.5%	

(*continued on next page*)

Table 1 (*continued*)

Demographic and clinical characteristics	CTCAE ≥ 2 (n = 37) (n)	CTCAE ≥ 2 (n = 37) %	CTCAE ≤ 1 (n = 53) (n)	CTCAE ≤ 1 (n = 53) %	*P* values
Adjuvant chemotherapy					.65
Yes	14	37.8%	17	32.1%	
No	22	59.5%	36	67.9%	
Unknown	1	2.7%	0	0%	
Adjuvant chemotherapy type					.87
AC-T	6	42.9%	6	35.3%	
FEC-D	3	21.4%	2	11.8%	
TC	2	14.3%	4	23.5%	
PACL	1	7.1%	1	5.9%	
Unknown	2	14.3%	4	23.5%	
Local vs locoregional irradiation					.50
Locoregional (4-field)	14	37.8%	16	30.2%	
Local (whole breast only)	23	62.2%	37	69.8%	
Radiation dose to skin volume (cGy)					.17
Average dose (±SD)	2313 (±591)	—	2182 (±253)	—	
Maximum dose	4500	—	3405	—	
Minimum dose	2128	—	2000	—	

Abbreviations: CTCAE = Common Terminology Criteria for Adverse Events; DCIS = ductal carcinoma in situ; ER = estrogen receptor; HER2 = human epidermal growth factor receptor 2; IDC = invasive ductal carcinoma; ILC = invasive lobular carcinoma; IMC = invasive mammary carcinoma; PR = progesterone receptor; SD = standard deviation.

P values were calculated using Fisher exact test, Mann-Whitney *U* test, and *t* test as appropriate.

* Cup sizes follow the North American standard.

† Grading was performed through the Nottingham score.

‡ Tumor sizes were obtained through pathology reports.

and analysis, which used a nonuniformity correction process. Analytical software for segmentation and to extract first and second order QTI features was developed using MATLAB (The MathWorks, Natick, USA). The region of interest (ROI) was delineated manually by using a standard protocol. The ROI comprised the treatment field borders according to clinical and anatomic landmarks, that is, clavicle (superior border), 2 cm below inframammary fold (inferior border), midline sternum (medial border), and midaxillary line (lateral border). The radiation treatment field published from the radiation treatment planning system (Pinnacle, Philips Healthcare, Amsterdam, Netherlands) was used as a reference to the targeted irradiated area of the thermograms for each time interval. All ROIs were verified with our collaborating radiation oncologists with 5 to 30 years of experience (Fig. 2F).

First-order features included temperature (°C) measurements, which were calculated as an average value across the entire breast treatment area (Fig. 2C). Other first-order features included entropy, skewness, and kurtosis. All first-order features recorded are presented in Table E1 (available online at https://doi.org/10.1016/j.ijrobp.2019.12.032). The thermal images were also analyzed using second-order statistics to extract 25 textural features related to Haralick textures. For extracting textural features from the thermal images, the original thermograms were used without resampling, and the full range of gray-level intensities in each thermal image was quantized into 16 levels. Second-order texture features were extracted using MATLAB and adapted from open-source radiomics codes using the Pyradiomics platform.[23,24] The Pyradiomics platform has been used in previous radiomic studies and comprises standard methods for radiomics feature extraction and image analysis.[24] Second-order features were derived using a GLCM, which yielded attributes associated with the spatial relationship between pixel intensities.[25,26] Other second-order features were computed from a gray-level run-length matrix (GLRLM) and a gray-tone difference matrix (GTDM). Other available second-order feature matrices (eg, a gray-level size zone and a gray-level dependence matrix) were excluded owing to redundancy with the selected feature sets and sample size limitations, that is, to avoid "data fishing" that can potentially over fit the models (ie, the total number of selected features within a given model should be limited to 1/10 of the sample size).[27,28]

Overall, there were 4 Haralick texture GLCM features ($F_{GLCM} = 4$), 5 GTDM features ($F_{GTDM} = 5$), and 16 GLRLM attributes ($F_{GLRLM} = 16$). Gray-tone texture features were calculated based on a gray-scale of 16 tones ($N_g = 16$). The displacement vector (d) and offset angle (θ), relative to the central pixel, were constrained to $d = 1$ and $\theta = 0°, 45°, 90°$, and 135°, respectively. Directional matrices (GLCM and GLRLM) were summated into a global matrix and normalized before feature extraction. The texture equations (GLCM, GTDM, GLRLM) and the

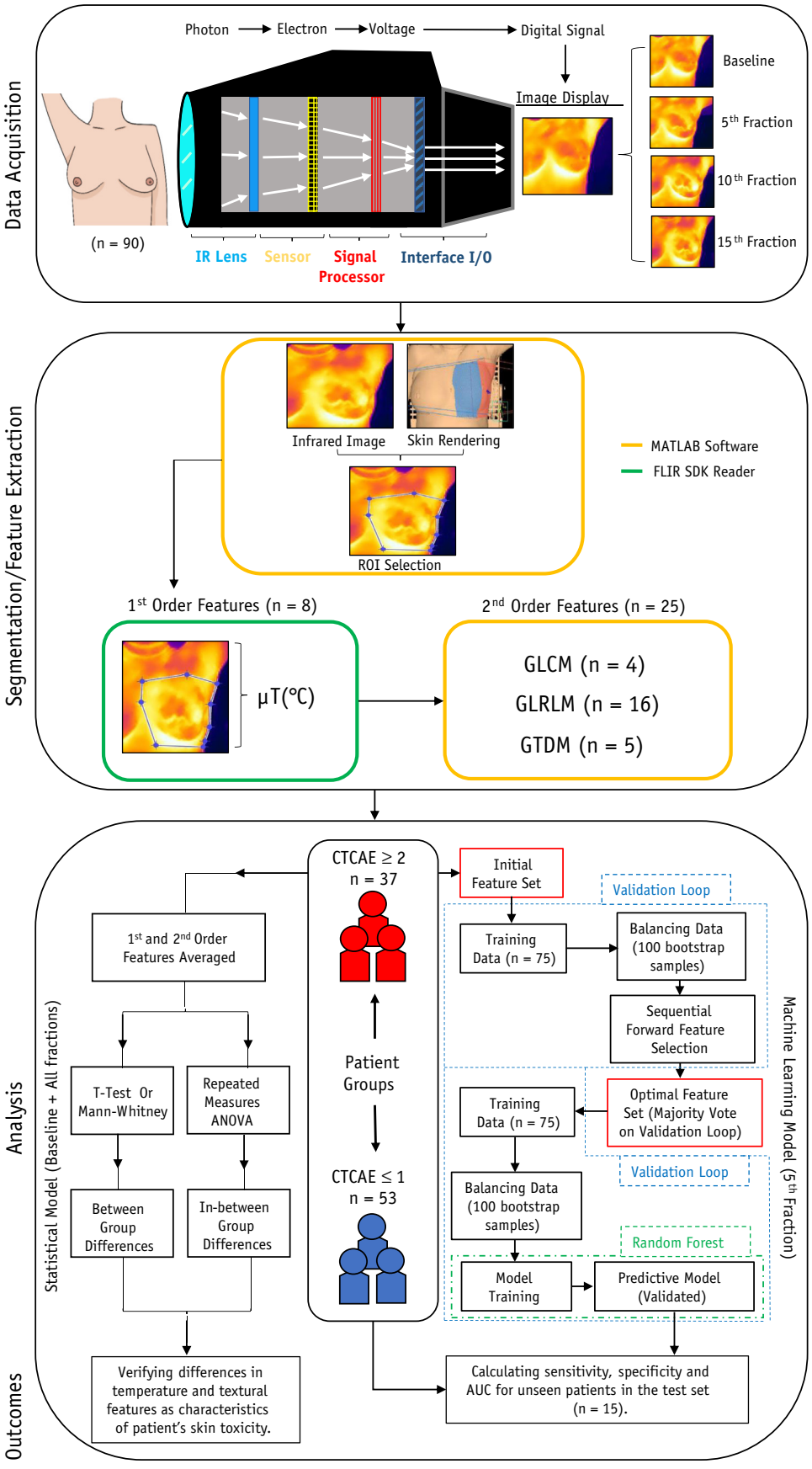

Fig. 1. Schematic demonstrating the workflow and methodology of the study.

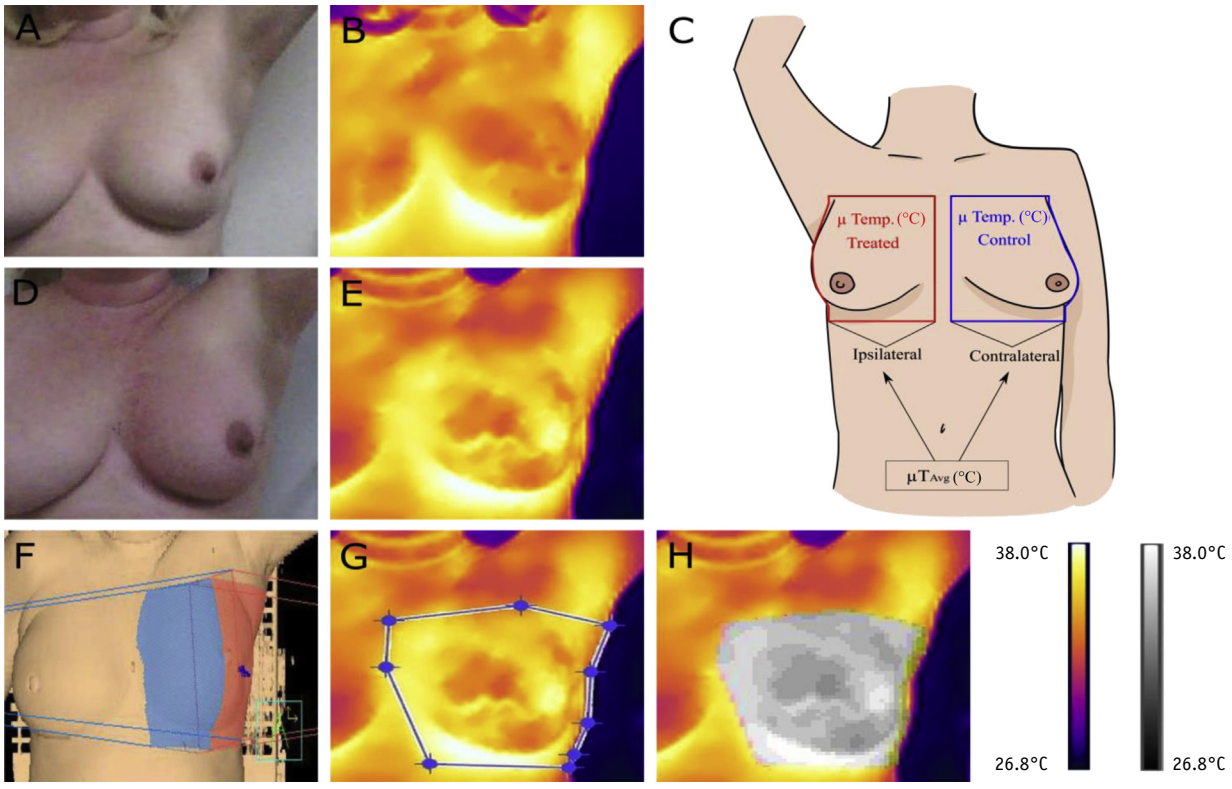

Fig. 2. (A) Digital image of a lumpectomy patient at baseline. (B) Thermal image representation of image (A). (C) A schematic illustrating the area of measurement for the mean temperature values in both ipsilateral and contralateral breasts. (D) Digital image at the end of a 4250 cGy radiation therapy (RT) regimen. (E) Thermal image representation of image (D). (F) Skin rendition highlighting the area receiving radiation. (G) Region of interest (ROI) selection based on the target area outlined in the skin rendition. (H) The gray-level representation of the selected ROI. Temperature scale bar: (B, E, and G); gray level scale bar: (H). *Abbreviations:* Temp. = temperature; μ = mean; μT = mean temperature; Avg = average. (A color version of this figure is available at https://doi.org/10.1016/j.ijrobp.2019.12.032.)

textural features (F = 25) are described in Tables E2 through E4 (available online at https://doi.org/10.1016/j.ijrobp.2019.12.032).

Statistical analysis

Descriptive statistics were calculated for differences in temperature and textural features between CTCAE $\geq$ 2 and CTCAE $\leq$ 1 patient classes using SPSS V. 24 (IBM Corp, Armonk, NY). This was calculated for the ipsilateral (irradiated) side and contralateral (nonirradiated) side. A Shapiro-Wilk test was used to test for normality violations. Averages were calculated (Fig. 2C) and compared between groups using both parametric (unpaired, 2-sided independent *t* test) and nonparametric (Mann-Whitney) statistical tests based on normal versus nonnormal data distributions, respectively. A repeated-measures analysis of variance was carried out to determine the significance of temporal changes in features. A Bonferroni correction for multiple comparisons was carried out, as was random sampling with replacement (ie, bootstrapping) over 1000 repetitions. Group comparisons for clinical and demographic data used a Fisher exact test to compare categorical variables (Table 1). *P* values <.05 were considered significant.

Dimensionality reduction and machine-learning modeling

For thermoradiomic markers, the relative changes from the baseline value of all first- and second-order features were calculated for each subject and class at all time intervals. Skin toxicity is typically observed after the 10th RT fraction; therefore, the objective was to test whether QTI and texture parameters obtained from earlier time intervals (ie, fifth RT fraction) demonstrated early prediction capabilities to severe skin toxicity. Several machine-learning classification experiments were carried out to yield various predictive models. First, the data set was divided into 2 sets for training and independent testing of the model. Fifteen patients (~17% of the whole data set) were randomly selected as an unseen test data set, and 75 patients (~83% of the whole data set) were used for training and validation of the model. Feature selection was performed using a sequential forward feature selection approach. The leave-1-patient-out (LOPO) cross-validated area under the receiver

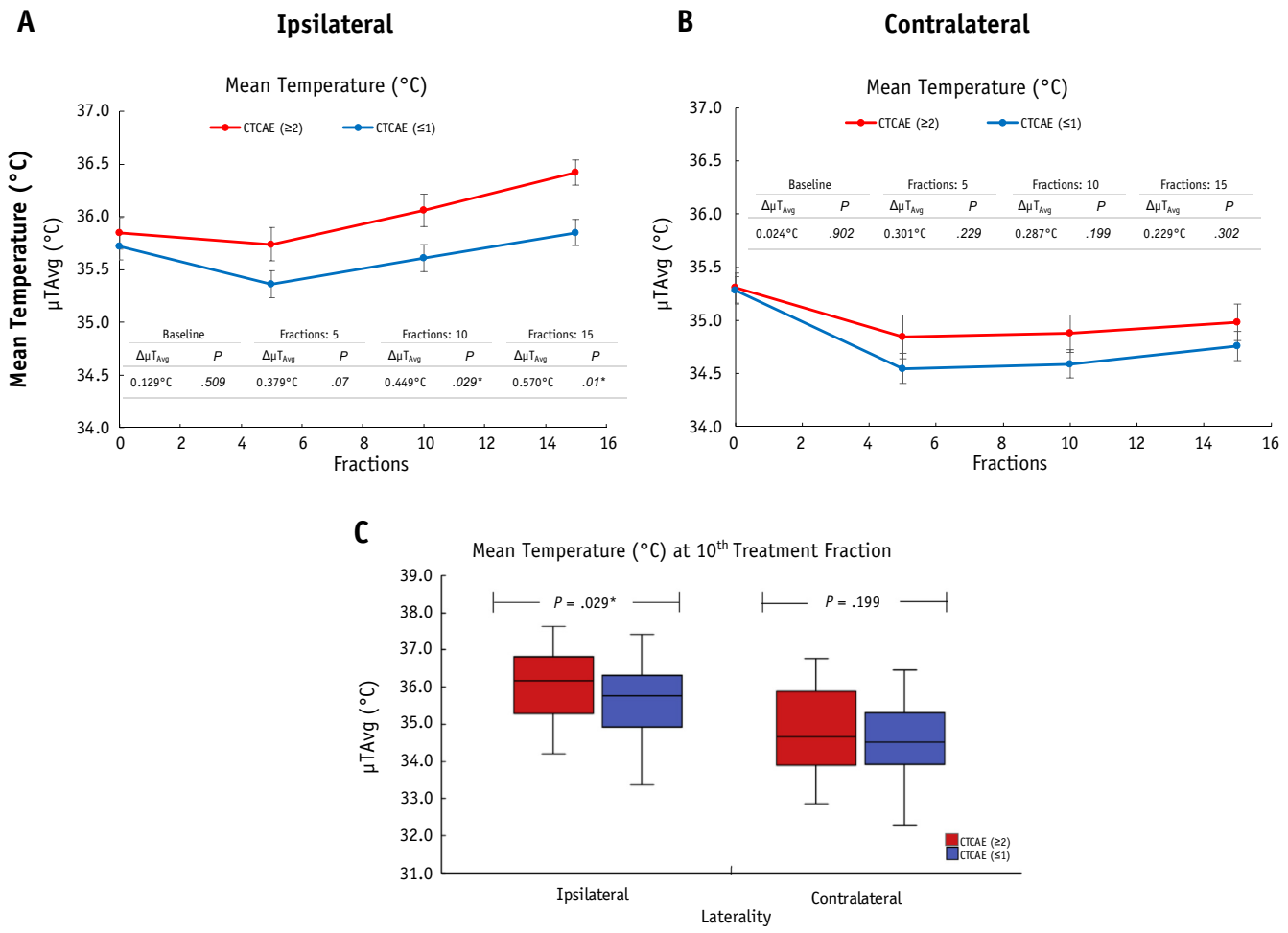

Fig. 3. (A, B) Comparison of ipsilateral and contralateral mean temperature averages between patients evaluated with a Common Terminology Criteria for Adverse Events (CTCAE) score of either ≥ 2 or ≤ 1 at baseline and at every fifth RT fraction. (C) Sample distribution comparison of mean temperature values between CTCAE ≥ 2 and CTCAE ≤ 1 groups after 10 fractions of radiation therapy (RT) for both ipsilateral and contralateral sides. Mean temperature ($\Delta\mu T_{Avg}$) value differences between CTCAE ≤ 1 and CTCAE ≥ 2 patient groups. *$P < .05$, **$P \leq .01$, based on independent t test. Error bars represent standard error of the mean.

operating characteristic (ROC) curve (area under the curve, AUC) was used as the criteria for feature selection. In the first 2 experiments, the first-order temperature features alone were used as the initial feature set. In experiments 3 and 4, all first-order and texture features were included in the initial feature set. Because the first initial feature set only included 8 first-order features, no feature reduction was applied or required before feature selection. The second initial feature set included 33 features (8 first-order and 25 texture features) and the redundant features were eliminated based on interfeature correlation. Specifically, the correlation between 2 features was calculated. The features with high interfeature correlation ($r^2 > 0.70$) were selected for analysis, and the retention criterion was based on the feature that yielded a higher AUC in the training set. Bootstrapping was used on the training data to improve the generalization performance of the trained classifier on unseen data.[29] Specifically, the classifier was trained using 100 bootstrap samples for each fold of the data during LOPO cross-validation. Before each bootstrap sampling, the majority class (negative) was randomly down sampled to compensate for the imbalance of data between the 2 classes (45 negative vs 30 positive cases). The optimal feature set was selected using a majority vote on the selected features for all folds of the data. The sensitivity, specificity, and accuracy of the trained model were calculated on the unseen test data and used in addition to the AUC to evaluate the efficacy of the optimal feature set to predict skin toxicity.[30]

Machine-learning classification experiments were repeated using clinical features alone to develop a baseline clinical model. Six clinical features were modeled that have been previously shown to predict radiation-induced skin toxicity: adjuvant chemotherapy (y/n), type of adjuvant chemotherapy, local (whole breast only) versus locoregional (4-field technique) irradiation, radiation dose-to-skin volume (cGy), menopausal status, and cup size. The clinical model was trained and subsequently evaluated using

Table 2 Prediction of skin toxicity using the optimal feature sets (relative change from the baseline at $f_x = 5$)

Experiment	Optimal feature set	Tr Acc	Te Acc	Te AUC	Te Sen	Te Spec
One	Max. temp (°C), median temp (°C), mean temp (°C), skewness, kurtosis	0.79	0.73	0.90	0.86	0.63
Two	Median temp (°C), mean temp (°C), skewness, kurtosis	0.76	0.73	0.84	0.71	0.75
Three	Mean temp (°C), GLCM-COR, GLCM-ENE, GLCM-HOM, GLRLM-GLN	**0.91**	**0.87**	**0.98**	0.86	**0.88**
Four	Mean Temp (°C), GLCM-COR, GLCM-ENE, GLRLM-LRLGLE	0.86	0.80	0.93	**1.0**	0.63

Abbreviations: Acc = accuracy; AUC = area under the curve; COR = correlation; ENE = energy; GLCM = gray level co-occurrence matrix; GLN = gray level nonuniformity; GLRLM = gray level run length matrix; HOM = homogeneity; LRLGLE = long run low gray level emphasis; Sen = sensitivity; Spec = specificity; Te = test; Tr = train.

Bolded values denote the highest value among the feature set for predictive accuracy and AUC values.

the same training and testing sets (subjects) that were used to develop the thermoradiomic model. The baseline clinical model was used to compare the performances between clinical features alone versus thermoradiomic predictors.

Results

Study participant demographics and outcomes

Of the 90 patients enrolled in this study, 37 (41%) presented with bright tender erythema or desquamation (CTCAE $\geq$ 2) at the end of their treatment. The difference in reported skin toxicity between CTCAE $\geq$ 2 and CTCAE $\leq$ 1 groups is represented in Figure E1 (available online at https://doi.org/10.1016/j.ijrobp.2019.12.032). Of the patients who presented with skin toxicity (CTCAE $\geq$ 2), 70.3% were Caucasian, 13.5% Asian, and 5.4% were Black (Table 1). Further demographic and clinical features such as cancer histologic type, tumor grade, and molecular subtype are presented in Table 1.

Temperature measurements of the treated breast

Significant differences ($P < .05$) in skin-surface temperatures (mean value and central tendency measures) were observed between the CTCAE $\geq$ 2 and CTCAE $\leq$ 1 classes, at fraction intervals $f_x = 10$ and $f_x = 15$ (Fig. 3A, Fig. E2, available online at https://doi.org/10.1016/j.ijrobp.2019.12.032). The CTCAE $\geq$ 2 patients demonstrated an increase in mean skin temperature, reaching 36.42°C at the end of treatment with a significant temperature increase of 0.58°C (±0.172°C) from baseline ($P < .01$), whereas the CTCAE $\leq$ 1 group had an insignificant temperature increase of 0.13°C (±0.133°C) compared with the baseline measurements ($P > .05$). Figure 3C illustrates a significant difference in mean temperature distributions of 0.45°C (±0.202°C) ($P = .029$) on the ipsilateral side at the 10th treatment fraction between CTCAE $\geq$ 2 and CTCAE $\leq$ 1 classes, with the CTCAE $\geq$ 2 patients demonstrating higher breast surface temperature.

Thermoradiomic markers using textural features of the treated breast

Most of the textural feature distributions did not demonstrate any statistical significance (Figs. E3-E5, available online at https://doi.org/10.1016/j.ijrobp.2019.12.032). However, texture analyses identified the GLRLM-short run emphasis (SRE) as being significantly different between the 2 patient groups (Fig. E4, available online at https://doi.org/10.1016/j.ijrobp.2019.12.032). Thermal measurements of CTCAE $\geq$ 2 patients exhibited higher SRE values than the CTCAE $\leq$ 1 group at $f_x = 10$ ($P = .033$). However, this effect demonstrates insignificant differences by $f_x = 15$ ($P = .69$)

Machine-learning predictive models using clinical variables

The performance of the clinical model demonstrates a prediction accuracy (Acc) of 67%, sensitivity of 75%, and specificity of 57%, using the following 5 clinical features selected through a forward feature selection: adjuvant chemotherapy (y/n), type of adjuvant chemotherapy, radiation dose-to-skin volume (cGy), menopausal status, and cup size. In a second experiment, the forward feature selection algorithm was used to select 4 clinical features (instead of 5). The selected features included adjuvant chemotherapy (y/n), radiation dose-to-skin volume (cGy), menopausal status, and cup size. The accuracy of the model in this experiment was 60%, with a sensitivity of 62.5% and a specificity of 57%.

High-accuracy predictive model for toxicity outcomes using thermoradiomic biomarkers

Table 2 presents the results of the skin toxicity prediction 1 week after the start of RT (at $f_x = 5$) using a select number of feature subsets. The following experiments demonstrated the most optimal outcome within their respective feature subsets: Experiment 1 (5 selected features from first-order temperature features) demonstrated test AUC of 0.90 and

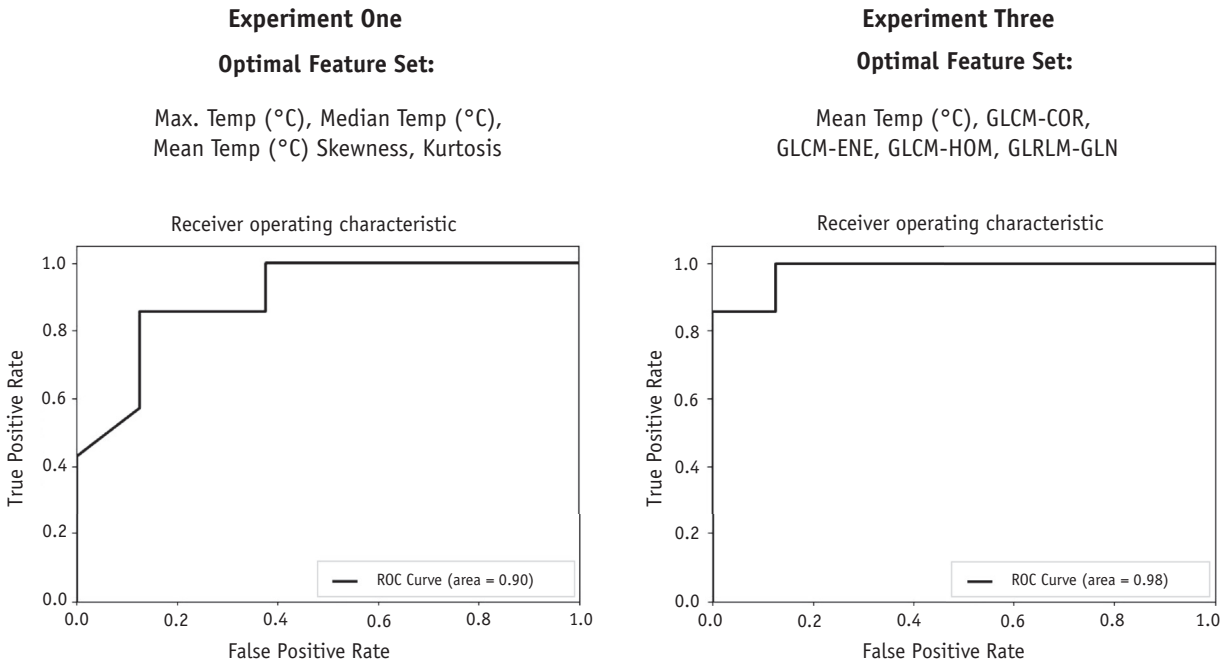

Fig. 4. Test receiver operating characteristic (ROC) curves for 2 representative experiments.

test accuracy of 73%, and experiment 3 (5 selected features from all first-order temperature and texture features) demonstrated test AUC of 0.98 and test accuracy of 87%. The model based on experiment 3 was trained particularly well in classifying the patients in our overall analysis (train Acc = 91%; sensitivity = 0.86; specificity = 0.88). Figure 4 displays the ROC curves for experiments 1 and 3.

Discussion

The aim of this study was to investigate the use of QTI biomarkers for radiation-induced skin toxicity in breast cancer. Our results demonstrate that patients who presented with a CTCAE ≥ 2, as evaluated at the time of their last RT fraction, exhibited higher skin-surface temperature values *during* treatment compared with those who demonstrated a CTCAE ≤ 1 score. The temperature differences between patient groups were most evident at the 10th fraction of RT. Moreover, using inferential statistical analyses alone, the QTI-texture features of the ipsilateral breast such as the GLRLM-SRE revealed a significant difference between the patient groups after the 10th radiation fraction. The CTCAE ≤ 1 patient group showed higher GLRLM-SRE average values. GLRLM textural features quantitate the length/number of homogenic pixels, and the GLRLM-SRE is indicative of how many short lengths of homogenic pixels are within the matrix.[31,32] Within this framework, we posit that a high GLRLM-SRE value in low-grade patients (CTCAE ≤ 1) represents a finer/smoother image texture within the thermograms. Clinically, this corresponds to unremarkable dermatologic changes and temperature variances on the breast skin surface.[31] In contrast, patients who demonstrated a CTCAE ≥ 2 had heterogeneous thermal maps of the skin, which may represent increased temperatures in regions of the breast that are at higher risk for dermatitis, such as the inframammary fold and axilla.

Using machine learning, we report early thermoradiomic signatures of acute skin toxicity, which is typically observed after 10 to 14 days of initiating RT.[33] Early thermal parameters from the 5th RT fraction were used in machine-learning models to classify patients and to test the accuracy of predicting symptom-based endpoints from selected QTI and textural hybrid feature sets.[34,35] These hybrid feature sets, in conjunction with a nonlinear classification model and bootstrapping, yielded high classification accuracy in a multidimensional space. We carried out several experiments using the random forest method to compute individual textural features and temperature parameters into optimized sets that are associated with binary outcomes, that is, CTCAE ≥ 2 versus CTCAE ≤ 1. Figure 4 displays the experiment containing only first-order feature sets and an experiment with both first-order and texture feature sets, where experiment 3 had the highest prediction accuracy and area under the ROC curve (test Acc = 0.87, test AUC = 0.98). The results presented in Figure 4 and Table 2 demonstrate the significance of using texture features in conjunction with the mean temperature parameter to predict skin toxicity after breast RT. In comparison, the first-order temperature features alone did not demonstrate a high prediction accuracy (test Acc = 0.73). Despite individual QTI features that showed insignificant differences between groups based on a Gaussian distribution (ie, carrying out inferential statistical analysis), the

machine-learning algorithm used a nonlinear classifier to assess the predictive performance of the combined features within a multidimensional space. The forward feature selection method yielded optimal complementary features based on the relative distances of attributes within the feature space.[36] We tested clinical features alone to develop a baseline machine-learning model. A comparison between clinical models versus thermoradiomic models was carried out, and the results suggest that thermoradiomic markers demonstrate superior early-predictors of radiation-induced skin toxicity compared with using clinical features alone. Within these frameworks, we propose that QTI may be used as a clinical tool in radiation oncology; specifically, that measuring the breast surface temperature and extracting the associated texture features may serve as possible predictive biomarkers for severe radiation-induced skin toxicity (CTCAE $\geq$ 2).

In comparison to other investigations, Templeton et al measured radiation-induced dermatitis in mice using 3-dimensional thermal tomography.[37] Their results revealed an increase in the thermal effusivity, which was associated with high-grade skin dermatitis.[37,38] In the clinical setting, our findings are concordant with a previous study by Maillot et al, which tested thermography for monitoring and predicting skin toxicity in a prospective patient cohort (n = 64).[13] Patients in that study who demonstrated a CTCAE $\geq$ 2 (ie, high-grade dermatitis) showed a very significant increase in the average skin-surface temperature over the course of RT ($P <$ .001).[13] Here, our novel approach incorporated textural features from GLCM, GTDM, and GLRLM analyses to evaluate and predict dermatitis in patients with breast cancer. We also employed machine-learning classification to identify early signatures (at $f_x = 5$) of skin toxicity, which corresponded to the patients' CTCAE grade at the end of treatment. Other technologies have been used to noninvasively study radiation-induced skin toxicity. For example, laser Doppler flowmetry (LDF) has been shown to quantitatively monitor skin toxicity by measuring microscopic changes in blood flow associated with skin reactions.[39,40] Previous studies have demonstrated that the LDF microcirculation index values correspond with CTCAE scores and suggest that LDF may be used to monitor radiation-induced dermatitis.[41] Although there is interest in combining LDF with thermal imaging, QTI remains a more practical and economic imaging modality[39,42,43] owing to readily available technology in radiation oncology clinics and its ability to relay practical and intuitive information about the macroscopic changes of the skin during RT.[43]

RT remains a crucial component in the postoperative management of breast cancer. The associated side effects from treatment may affect patient quality of life. Particularly, severe skin toxicity is prevalent within this patient population and carries an increased risk of pain and discomfort. Approximately 61.9% of patients will develop CTCAE 2 toxicity, whereas 8.3% present with CTCAE 3 symptoms after 2 weeks of treatment.[44] In our patient cohort, 41.1% of patients had a CTCAE $\geq$ 2. Therefore, this represents a significant patient population that would potentially benefit from early detection and early intervention for symptom management.

Radiation-induced skin toxicity associated with RT may be better managed using thermography, which has several advantages such as portability and relatively low cost compared with other imaging devices, and may provide actionable biomarkers that may guide the administration of early-intervention therapeutics. Therapeutic options include Mepitel film for prophylaxis against the onset of skin toxicity. The mechanism of action of Mepitel film involves protecting the affected skin from external contamination and maintaining a moist environment to facilitate wound healing.[45-47] Herst et al demonstrated that Safe-tac-based Mepitel film prevented the occurrence of radiation-induced skin toxicity by 92% ($P <$ .0001) and improved postradiation patient satisfaction.[45] Thermal imaging may also provide a method to validate the efficacy of pharmacologic agents to manage skin toxicity.[16,48] For instance, although glucocorticosteroids are successful in treating radiation-induced skin toxicity, their antiinflammatory effects have been found to interfere with passive wound healing, which may compromise the structural integrity of the tissue in the long term.[48,49] QTI has potential uses as a decision-support tool. QTI-based biomarkers could steer symptom management decisions in radiation oncology; for example, avoiding unnecessary treatments for patients who demonstrate a low risk of developing skin symptoms. Conversely, for patients who have a high risk, there is an opportunity to develop a personalized thermography-guided approach for skin toxicity management and prevention.

The limitations of this study include small sample size, which affects the framework of the prediction model. First, small sample sets limit the approach for group classification (ie, sufficient samples and distributions are required between classes). Second, model testing and validation in small sample sets have a greater risk of yielding an over fitted prediction model. To address this problem, the predictive model was trained using a LOPO validation approach and subsequently evaluated using unseen data from an independent test set. Furthermore, although data collection conditions were controlled as best as possible, some inconsistencies in experimental conditions, such as heavy clothing attire, may have led to an increase in patients' skin-surface temperature before imaging. We attempted to reduce this effect by instructing patients to change into standard hospital gowns before imaging. Other limitations include interfraction ROI selection. Although we used clinical- and protocol-guided segmentation with reference to the radiation treatment plan, the region of interest may also fluctuate based on anatomic changes (ie, changes in the size of the breast) and positional differences of the patient at each time interval. Our study population was largely composed of Caucasian and Asian patients

(86.7%) with light skin pigmentation (Table 1) who tend to demonstrate less severe skin toxicity than patients with darker skin pigmentation.[50] Ethnicity is a known risk factor for radiation-induced skin toxicity, most notably, Black patients have a 73% greater risk of skin-dermatitis than other ethnic groups, and thus it is crucial that these methods be repeated in a patient population with greater diversity in ethnicity and skin phenotype.[50,51] In future work, thermoradiomic markers may also be useful in other cancer sites, such as head and neck radiation oncology. Severe skin toxicity, characterized as confluent moist desquamation (ie, grade 3) presents in approximately 23% of patients with head and neck cancer[52]; thus, early thermoradiomic markers in this setting may significantly improve management strategies at the beginning of the 7-week treatment course for this patient population.

In conclusion, QTI is a readily available technology and may potentially support clinical decisions in breast radiation oncology. Quantitative assessments of skin toxicity are useful to reduce diagnostic variability among health care providers and have the potential to validate early clinical management of skin-related side effects of treatment. Advances in current practices are limited by the available imaging tools that can objectively measure skin toxicity. As a result, visual inspection and patient-reported symptoms remain the primary method, but this can be subjective.[7,53,54] Thermal imaging has the potential to reduce these biases and may also complement current grading systems such as the CTCAE score. It could potentially help better define the grading scales within quantitative thermal boundaries. Image guided RT already plays an integral role in the clinic for treatment delivery. This study demonstrates the feasibility of additional image guided approaches; specifically, to use QTI as a clinical decision support tool for symptom management in the breast radiation oncology clinic.

References

1. Darby S, McGale P, Correa C, et al. Effect of radiotherapy after breast-conserving surgery on 10-year recurrence and 15-year breast cancer death: Meta-analysis of individual patient data for 10 801 women in 17 randomized trials. *Lancet* 2011;378:1707-1716.
2. McGale P, Taylor C, Correa C, et al. Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: Meta-analysis of individual patient data for 8135 women in 22 randomized trials. *Lancet* 2014;383:2127-2135.
3. Tesselaar E, Flejmer AM, Farnebo S, Dasu A. Changes in skin microcirculation during radiation therapy for breast cancer. *Acta Oncol (Madr)* 2017;56:1072-1080.
4. Chan RJ, Larsen E, Chan P. Re-examining the evidence in radiation dermatitis management literature: An overview and a critical appraisal of systematic reviews. *Int J Radiat Oncol Biol Phys* 2012;84:e357-362.
5. Ryan JL. Ionizing radiation: The good, the bad, and the ugly. *J Invest Dermatol* 2012;132:985-993.
6. Wei J, Meng L, Hou X, et al. Radiation-induced skin reactions: Mechanism and treatment. *Cancer Manag Res* 2019;11:167-177.
7. Huang CJ, Hou MF, Luo KH, et al. RTOG, CTCAE and WHO criteria for acute radiation dermatitis correlate with cutaneous blood flow measurements. *Breast* 2015;24:230-236.
8. Sprangers MAG. Response-shift bias: A challenge to the assessment of patients' quality of life in cancer clinical trials. *Cancer Treat Rev* 1996;22:55-62.
9. Murray CS, Rees JL. How robust are the dermatology life quality index and other self-reported subjective symptom scores when exposed to a range of experimental biases? *Acta Derm Venereol* 2010;90:34-38.
10. Owen R, Ramlakhan S. Infrared thermography in pediatrics: A narrative review of clinical use. *BMJ Paediatr Open* 2017;1:e000080.
11. Nagori A, Dhingra LS, Bhatnagar A, Lodha R. Predicting hemodynamic shock from thermal images using machine learning. *Sci Rep* 2019;9:1-9.
12. Martini G. Juvenile-onset localized scleroderma activity detection by infrared thermography. *Rheumatology* 2002;41:1178-1182.
13. Maillot O, Leduc N, Atallah V, et al. Evaluation of acute skin toxicity of breast radiotherapy using thermography: Results of a prospective single-center trial. *Cancer Radiother* 2018;22:205-210.
14. Milosevic M, Jankovic D, Peulic A. Thermography based breast cancer detection using texture features and minimum variance quantization. *EXCLI J* 2014;13:1204-1215.
15. Ulff E, Maroti M, Serup J, Falkmer U. A potent steroid cream is superior to emollients in reducing acute radiation dermatitis in breast cancer patients treated with adjuvant radiotherapy. A randomized study of betamethasone versus 2 moisturizing creams. *Radiother Oncol* 2013;108:287-292.
16. Freedman GM. Topical agents for radiation dermatitis in breast cancer: 50 shades of red or same old, same old? *Int J Radiat Oncol* 2014;90:736-738.
17. Ulff E, Maroti M, Serup J, Nilsson M, Falkmer U. Prophylactic treatment with a potent corticosteroid cream ameliorates radiodermatitis, independent of radiation schedule: A randomized double blinded study. *Radiother Oncol* 2017;122:50-53.
18. Piroth MD, Piroth DM, Pinkawa M, Woodruff SG, Holy R, Eble MJ. Immediate reconstruction with an expander/implant following ablatio mammae because of breast cancer side effects and cosmetic results after adjuvant chest wall radiotherapy. *Strahlenther Onkol* 2009;10:669-674.
19. Mulliez T, Veldeman L, Greveling A Van, et al. Hypofractionated whole breast irradiation for patients with large breasts: A randomized trial comparing prone and supine positions. *Radiother Oncol* 2013;108:203-208.
20. Chen MF, Chen WC, Lai CH, Hung CH, Liu KC, Cheng YH. Predictive factors of radiation-induced skin toxicity in breast cancer patients. *BMC Cancer* 2010;10.
21. Lee TF, Sung KC, Chao PJ, et al. Relationships among patient characteristics, irradiation treatment planning parameters, and treatment toxicity of acute radiation dermatitis after breast hybrid intensity modulation radiation therapy. *PLoS One* 2018;13:e0200192.
22. Parekh A, Dholakia AD, Zabranksy DJ, et al. Predictors of radiation-induced acute skin toxicity in breast cancer at a single institution: Role of fractionation and treatment volume. *Adv Radiat Oncol* 2018;3:8-15.
23. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res* 2017;77:e104-e107.
24. Welch ML, McIntosh C, Haibe-Kains B, et al. Vulnerabilities of radiomic signature development: The need for safeguards. *Radiother Oncol* 2019;130:2-9.
25. Haralick RM, Shanmugam K, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973;3:610-621.
26. Albregtsen F. Statistical texture measures computed from gray level coocurrence matrices. *Image Processing Laboratory, Department of Informatics, University of Oslo*. 2008. Available at: https://www.uio.no/studier/emner/matnat/ifi/INF4300/h08/undervisningsmateriale/glcm.pdf. Accessed July 13, 2019.

27. Thibault G, Fertil B, Navarro C, et al. Shape and texture indexes application to cell nuclei classification. *Int J Pattern Recognit Artif Intell* 2013;27:1357002.
28. Sun C, Wee WG. Neighboring gray level dependence matrix for texture classification. *Comput Vision, Graph Image Process* 1983;23:341-352.
29. Chernick MR, Murthy VK, Nealy CD. Application of bootstrap and other resampling techniques: Evaluation of classifier performance. *Pattern Recognit Lett* 1985;3:167-178.
30. He H, Garcia E. Learning from imbalanced data. *IEEE Trans Knowl Data* 2009;21:9.
31. Xu D-H, Kurani AS, Furst JD, Raicu DS. Run-length encoding for volumetric texture. *Heart* 2004;27:452-458.
32. Kolossváry M, Kellermayer M, Merkely B, Maurovich-Horvat P. Cardiac computed tomography radiomics. *J Thorac Imaging* 2018;33:26-34.
33. Porock D, Kristjanson L. Skin reactions during radiotherapy for breast cancer: The use and impact of topical agents and dressings. *Eur J Cancer Care (Engl)* 1999;8:143-153.
34. Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and machine learning forecasting methods: Concerns and ways forward. *PLoS One* 2018;13:1-26.
35. Bzdok D, Altman N, Krzywinski M. Points of significance: Statistics versus machine learning. *Nat Methods* 2018;15:233-234.
36. Lo A, Chernoff H, Zheng T, Lo SH. Why significant variables aren't automatically good predictors. *Proc Natl Acad Sci U S A* 2015;112:13892-13897.
37. Templeton A, Chu J, Sun M, et al. Thermal effusivity changes as a precursor to moist desquamation. *Radiat Res* 2012;178:295-303.
38. Chu J, Sun J, Templeton A, Yao R, Griem K. Thermal effusivity: A promising imaging biomarker to predict radiation-induced skin injuries. *Health Phys* 2012;103:204-209.
39. Petersen LJ. Direct comparison of laser Doppler flowmetry and laser Doppler imaging for assessment of experimentally induced inflammation in human skin. *Inflamm Res* 2013;62:1073-1078.
40. Micheels J, Aisbjorn B, Sorensen B. Laser Doppler flowmetry. A new noninvasive measurement of microcirculation in intensive care? *Resuscitation* 1984;12:31-39.
41. González Sanchis A, Brualla González L, Sánchez Carazo JL, et al. Evaluation of acute skin toxicity in breast radiotherapy with a new quantitative approach. *Radiother Oncol* 2017;122:54-59.
42. Burke-Smith A, Collier J, Jones I. A comparison of noninvasive imaging modalities: Infrared thermography, spectrophotometric intracutaneous analysis and laser Doppler imaging for the assessment of adult burns. *Burns* 2015;41:1695-1707.
43. Murray AK, Herrick AL, King TA. Laser Doppler imaging: A developing technique for application in the rheumatic diseases. *Rheumatology* 2004;43:1210-1218.
44. Drost L, Li N, Vesprini D, et al. Prospective study of breast radiation dermatitis. *Clin Breast Cancer* 2018;18:e789-e795.
45. Herst PM, Bennett NC, Sutherland AE, Peszynski RI, Paterson DB, Jasperse ML. Prophylactic use of Mepitel Film prevents radiation-induced moist desquamation in an intrapatient randomized controlled clinical trial of 78 breast cancer patients. *Radiother Oncol* 2014;110:137-143.
46. Wooding H, Yan J, Yuan L, et al. The effect of mepitel film on acute radiation-induced skin reactions in head and neck cancer patients: A feasibility study. *Br J Radiol* 2018;91:20170298.
47. Field CK, Kerstein MD. Overview of wound healing in a moist environment. *Am J Surg* 1994;167:2S-6S.
48. Yee C, Wang K, Asthana R, et al. Radiation-induced skin toxicity in breast cancer patients: A systematic review of randomized trials. *Clin Breast Cancer* 2018;18:e825-e840.
49. Grose R, Werner S, Kessler D, et al. A role for endogenous glucocorticoids in wound repair. *EMBO Rep* 2002;3:575-582.
50. Wright JL, Takita C, Reis IM, Zhao W, Lee E, Hu JJ. Racial variations in radiation-induced skin toxicity severity: Data from a prospective cohort receiving postmastectomy radiation. *Int J Radiat Oncol Biol Phys* 2014;90:335-343.
51. Schnur JB, Ouellette SC, Dilorenzo TA, Green S, Montgomery GH. A qualitative analysis of acute skin toxicity among breast cancer radiotherapy patients. *Psychooncology* 2011;20:260-268.
52. Meyer F, Fortin A, Wang CS, Liu G, Bairati I. Predictors of severe acute and late toxicities in patients with localized head and neck cancer treated with radiation therapy. *Int J Radiat Oncol* 2012;82:1454-1462.
53. D'haese S, Van Roy M, Bate T, Bijdekerke P, Vinh-Hung V. Management of skin reactions during radiotherapy in Flanders (Belgium): A study of nursing practice before and after the introduction of a skin care protocol. *Eur J Oncol Nurs* 2010;14:367-372.
54. Chan RJ, Webster J, Chung B, Marquart L, Ahmed M, Garantziotis S. Prevention and treatment of acute radiation-induced skin reactions: A systematic review and meta-analysis of randomized controlled trials. *BMC Cancer* 2014;14:1-19.