Sheffield Hallam University

Quantitative thermal imaging biomarkers to detect acute skin toxicity from breast radiation therapy using supervised machine learning

SAEDNIA, Khadijeh, TABBARAH, Sami, LAGREE, Andrew, WU, Tina, KLEIN, Jonathan, GARCIA, Eduardo, HALL, Michael, CHOW, Edward, RAKOVITCH, Eileen, CHILDS, Charmaine <http://orcid.org/0000-0002-1558-5633>, SADEGHI-NAINI, Ali and TRAN, William

Available from Sheffield Hallam University Research Archive (SHURA) at:

http://shura.shu.ac.uk/25736/

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

Published version

SAEDNIA, Khadijeh, TABBARAH, Sami, LAGREE, Andrew, WU, Tina, KLEIN, Jonathan, GARCIA, Eduardo, HALL, Michael, CHOW, Edward, RAKOVITCH, Eileen, CHILDS, Charmaine, SADEGHI-NAINI, Ali and TRAN, William (2020). Quantitative thermal imaging biomarkers to detect acute skin toxicity from breast radiation therapy using supervised machine learning. International Journal of Radiation Oncology*Biology*Physics, 106 (5), 1071-1083.

Copyright and re-use policy

See http://shura.shu.ac.uk/information.html

Sheffield Hallam University Research Archive
http://shura.shu.ac.uk

Quantitative Thermal Imaging Biomarkers to Detect Acute Skin Toxicity from Breast Radiotherapy Using Supervised Machine Learning

Khadijeh Saednia MSc[1,3,4], Sami Tabbarah BSc MSc[1,8], Andrew Lagree BSc MSc[1], Tina Wu[1], Jonathan Klein MD MSc FRCPC DABR[7], Eduardo Garcia BSc[1], Michael Hall BSc[1], Edward Chow MSc PhD MBBS FRCPC [1,2], Eileen Rakovitch MSc MD FRCPC[1,2], Charmaine Childs PhD[6], Ali Sadeghi-Naini PhD[1,3,4], William T. Tran MRT(T) MSc PhD[1,2,3,5,6,8]

[1.] Department of Radiation Oncology, Sunnybrook Health Sciences Centre, Toronto, Canada.
[2.] Department of Radiation Oncology, University of Toronto, Toronto, Canada
[3.] Department of Electrical Engineering and Computer Science, York University, Toronto, Canada.
[4.] Physical Sciences Platform, Sunnybrook Research Institute, Toronto, Canada.
[5.] Department of Biomedical Physics, Ryerson University, Toronto, Canada.
[6.] Department of Radiotherapy & Oncology, Sheffield Hallam University, Sheffield, United Kingdom.
[7.] Department of Radiation Oncology, Albert Einstein College of Medicine, New York, USA.
[8.] Evaluative Clinical Sciences Platform, Sunnybrook Research Institute, Toronto, Canada.

Key Words: Thermal Imaging, Breast Cancer, Radiation Therapy, Machine Learning, Texture Analysis.

Corresponding Author:

Dr. William T. Tran, MRT(T), MSc, PhD
Department of Radiation Oncology
Sunnybrook Health Sciences Centre
University of Toronto
2075 Bayview Avenue, RM TB 097
Toronto, Ontario, Canada, M4N3M5
Tel: 416 480 6100 x 3746
Email: william.tran@sunnybrook.ca

Authors Responsible for Statistical Analysis:

Khadijeh Saednia, MSc
Dept. of Electrical Engineering and Computer Science
York University
116-124 Campus Walk
Toronto, Ontario, Canada, M3J 1P3
Tel: 437 989 1371
Email: sh.saednia@gmail.com

Dr. Ali Sadeghi-Naini, PhD
Dept. of Electrical Engineering and Computer Science
York University
116-124 Campus Walk
Toronto, Ontario, Canada, M3J 1P3
Tel: 416 480 6100
Email: asn@yorku.ca

Conflict of Interest:

The authors declare that there are no conflicts of interests to disclose.

Funding Statement:

Funding for this study was provided by the Terry Fox Research Institute (Grant number: 1083), the Kavelman Fonn Foundation and the Women's Health Golf Classic Fund. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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.

This project was partly funded by the Terry Fox Research Institute and we thank them for their support.

ABSTRACT

Purpose: Radiation-induced dermatitis is a common side effect of breast radiotherapy (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 using supervised machine-learning to develop a predictive model for radiation dermatitis.

Methods: Ninety patients treated for adjuvant whole-breast radiotherapy (4250 Gy/f_x=16) were recruited to the study. Thermal images of the treated breast were taken at four intervals: prior to RT, then weekly, at f_x=5, f_x=10, and f_x=15. Parametric thermograms were analyzed and yielded 26 thermal-based features which included surface temperature (°C) and texture parameters obtained from 1) grey-level co-occurrence matrix (GLCM), 2) grey-level run-length matrix (GLRLM) and 3) neighborhood grey-tone difference matrix (GTDM). 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 labelled 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 dataset was partitioned for model training, independent testing, and validation. Fifteen patients (~ 17% of the whole dataset) were randomly selected as an unseen test dataset, and 75 patients (~ 83% of the whole dataset) were used for training and validation of the model. A random forest classifier with leave-one-patient-out cross-validation was employed for modelling 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=0.029*). 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 (Acc) on the independent test data (test Acc=0.87) at f_x=5 for predicting the skin toxicity at the end of RT.

Conclusion: Early thermal markers after five fractions of RT are predictive of radiation-induced skin toxicity in breast radiotherapy.

INTRODUCTION

Radiation therapy (RT) utilizes 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 following lumpectomy (15.7%; 95% CI: 13.7-17.7; $p<0.00001$) (1) or mastectomy (11.5%; 95% CI: 0.57-0.82; $p=0.00006$) (2). However, RT is associated with dermatological 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 increases in blood volume; while the damaged irradiated skin clinically presents as erythema and 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, as well as differentiating the degree of severe skin toxicity (i.e. CTCAE≥2). This is caused by practitioner bias, under-reporting by patients, and clinician expertise that may lead to the variabilities in grading skin toxicity (7–9). Topographical 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 (i.e. thermoradiomics) may yield *actionable insight* into symptom management during radiotherapy.

Quantitative thermal imaging 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., investigated the use of thermography to quantitatively evaluate skin toxicity and found that patients that were classified using the CTCAE, and Radiotherapy Oncology Group (RTOG) criteria, demonstrated a high grade of skin toxicity ($\geq$2) that was also associated with an increase in the average local temperature ($p<0.05$) (13). Furthermore, Maillot, et al. 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 – 15 RT fractions (13). 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 grey-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 (i.e. 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 breast cancer patients and we hypothesized that radiation-induced skin toxicity is associated with an increase in skin-surface temperature and changes in thermoradiomic parameters.

METHODS

Patient Recruitment Criteria and Radiation Treatment Parameters

This study was carried out at a single academic health centre and approved by the institutional research ethics board. Participants signed an informed consent form prior to enrolment. 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 radiotherapy to the whole breast or chestwall (4250 Gy/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, i.e. 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 (e.g. hydrocortisone) or topical antibiotics (e.g. 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

Infrared (IR, thermal) imaging data was 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 (prior to 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 x180 pixels. The instantaneous field of view (IFOV), which comprises the pixel size on the measurement surface (i.e. breast or chestwall) was 3.54 mm. Thermal images were captured using a fixed emissivity setting of ε=0.98.

Long wave infrared thermography was carried out in a designated exam 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 (i.e. end of the 16^{th} fraction) was used for ground truth labelling in machine learning models (described below). Other clinical and demographic information were collected from the electronic medical records, and included the following variables: age, cancer diagnosis, clinicopathological characteristics of the tumor, surgery details and radiotherapy treatment information. Patient ethnicity 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 patients' electronic medical record and recorded in the patients' case report forms (CRFs). **Table 1** includes information on those variables, which include the administration of adjuvant chemotherapy (y/n), type of adjuvant chemotherapy, local treatment (whole breast only) versus locoregional irradiation (four-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; FLIR, Wilsonville, USA). Thermal images were normalized using the software development kit prior to segmentation and analysis, which used a non-uniformity correction (NUC) 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 anatomical landmarks, i.e. clavicle (superior border), 2 cm below inframammary fold (inferior border), midline sternum (medial border) and mid-axillary 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 5-30 years of experience (**Figure 2F)**.

First-order features included temperature (°C) measurements, which were calculated as an average value across the entire breast treatment area (**Figure 2C**). Other first-order features included entropy, skewness and kurtosis. All first-order features recorded are presented in **Supplementary Table 1.** 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 (The MathWorks, Natick USA), and adapted from open-source radiomics codes using the *Pyradiomics* platform (23,24). The *Pyradiomics* platform

has been used in previous radiomic studies, and comprise standard methods for radiomics feature extraction and image analysis (24). Second-order features were derived using a grey-level co-occurrence matrix (GLCM), which yielded attributes associated with the spatial relationship between pixel intensities (25,26). Other second-order features were computed from a grey-level run-length matrix (GLRLM) and a grey-tone difference matrix (GTDM). Other available second-order feature matrices (e.g. a grey-level size zone [GLSZM] and a grey-level dependence matrix [GLDM]) were excluded due to redundancy with the selected feature sets, and sample size limitations, i.e., to avoid "data fishing" that can potentially overfit the models (i.e. 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 four Haralick texture GLCM features ($F_{GLCM=4}$), five GTDM features ($F_{GTDM=5}$), and 16 GLRLM attributes ($F_{GLRLM=16}$). Grey-tone texture features were calculated based on a grey-scale of 16-tones (N_g=16). The displacement vector (d) and offset angle (θ), relative to the central pixel was constrained to, d=1, and θ = 0°, 45°, 90°, 135°, respectively. Directional matrices (GLCM and GLRLM) were summated into a global matrix and normalized prior to feature extraction. The texture equations (GLCM, GTDM, GLRLM) and the textural features (F=25) are described in **Supplementary Tables 2, 3 and 4**.

Statistical Analysis

Descriptive statistics were calculated for differences in temperature and textural features between CTCAE≥2 and CTCAE≤1 patient classes using SPSS V. 24 (IBM Corp., Armonk, NY, USA). This was calculated for the ipsilateral (irradiated) side and contralateral (non-irradiated) side. A Shapiro-Wilk test was used to test for normality violations. Averages were calculated

(**Figure 2C**) and compared between groups using both parametric (unpaired, 2-sided independent t-test) and non-parametric (Mann-Whitney) statistical tests based on normal versus non-normal data distributions, respectively. A repeated-measures analysis of variance (ANOVA) was carried out to determine the significance of temporal changes in features. A Bonferroni correction for multiple comparisons was carried out, as well as random sampling with replacement (i.e. bootstrapping) over 1000 repetitions. Group comparisons for clinical and demographic data used a Fisher's exact test to compare categorical variables **(Table 1).** *P* values < 0.05 were considered significant.

Dimensionality Reduction and Machine Learning Modelling

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 10^{th} RT fraction; therefore, the objective was to test if QTI and texture parameters obtained from earlier time intervals (i.e., 5^{th} 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 dataset was divided into two sets for training and independent testing of the model. Fifteen patients (~ 17% of the whole dataset) were randomly selected as an unseen test dataset, and 75 patients (~ 83% of the whole dataset) were used for training and validation of the model. Feature selection was performed using a sequential forward feature selection (SFS) approach. The leave-one-patient-out (LOPO) cross-validated area under the receiver operating characteristic (ROC) curve (area under the curve, AUC) was used as the criteria for feature selection. In the first two experiments, the first-order temperature features alone, were used as the initial feature set. In experiments

three and four, all first-order and texture features were included in the initial feature set. Since the first initial feature set only included 8 first-order features, no feature reduction was applied or required prior to 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 inter-feature correlation. Specifically, the correlation between each two features was calculated. The features with high inter-feature 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 two classes (45 negative versus 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 modelled, which have been previously shown to predict radiation-induced skin toxicity: adjuvant chemotherapy (y/n), type of adjuvant chemotherapy, local (whole breast only) versus locoregional (four-field technique) irradiation, radiation dose to skin volume (cGy), menopausal status, and cup size. The clinical model was trained and subsequently evaluated using the same training and testing sets (subjects) that were utilized 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 ninety patients enrolled in this study, 37 (41%) presented with bright tender erythema and/or desquamation (CTCAE≥2) at the end of their treatment. The difference in reported skin toxicity between CTCAE≥2 and CTCAE≤1 groups is represented in **Supplementary Figure A**. Of the patients that presented with skin toxicity (CTCAE≥2), 70.3% were Caucasian, 13.5% Asian, and 5.4% were Black (**Table 1**). Further demographic and clinical features such as cancer histological type, tumor grade, and molecular subtype are presented in **Table 1**.

Temperature Measurements of the Treated Breast

Significant differences ($p<0.05$) in skin-surface temperatures (mean value and central tendency measures) were observed between the CTCAE≥2 and CTCAE≤1 classes, at fraction intervals f_x=10 and f_x=15 **(Figure 3A, Supplementary Figure B)**. The CTCAE≥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 ($p<0.01$), whereas the CTCAE≤1 group had an insignificant temperature increase of 0.13°C(±0.133°C) compared to the baseline measurements ($p>0.05$). **Figure 3C** illustrates a significant difference in mean temperature distributions of 0.45°C (±0.202°C) ($p=0.029$) on the ipsilateral side at the 10^{th} treatment fraction between CTCAE≥2 and CTCAE≤1 classes, with the CTCAE≥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 (**Supplementary Figures C, D, & E**). However, texture analyses identified the GLRLM-Short Run Emphasis (SRE) as being significantly different between the two patient groups **(Supplementary Figure D)**. Thermal measurements of CTCAE≥2 patients exhibited higher SRE values than the CTCAE≤1 group at f_x=10 (*p=0.033*). However, this effect demonstrates insignificant differences by f_x=15 (*p=0.69*)

Machine Learning Predictive Models Using Clinical Variables

The performance of the clinical model demonstrates a prediction accuracy (Acc) of 67%, sensitivity (Sen) of 75% and specificity (Spec) of 57%, using the following five 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 four clinical features (instead of five). 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 one 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 one (five selected features from first-order temperature features) demonstrated test AUC of 0.90 and test accuracy

of 73%; and experiment three (five 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 three was trained particularly well in classifying the patients in our overall analysis (train Acc = 91%; Sen = 0.86; Spec = 0.88). **Figure 4** displays the ROC curves for experiments one and three.

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 to those who demonstrated a CTCAE≤1 score. The temperature differences between patient groups were most evident at the 10^{th} fraction of radiotherapy. 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 10^{th} 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) represent a finer/smoother image texture within the thermograms. Clinically, this corresponds to unremarkable dermatological 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 10-14 days of initiating RT (33). Early thermal parameters from the 5^{th} 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; i.e., (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 three 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 radiotherapy. 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 (i.e. carrying out inferential statistical analysis), the machine learning algorithm utilized a non-linear 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 were carried out and the results suggest that thermoradiomic markers demonstrate superior early-predictors of radiation-induced skin toxicity compared to 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≥2).

In comparison to other investigations, Templeton et al. measured radiation-induced dermatitis in mice using three-dimensional thermal tomography. 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≥2 (i.e. high-grade dermatitis) showed a very significant increase in the average skin-surface temperature over the course of radiotherapy ($p<0.001$) (13). Here, our novel approach incorporated textural features from GLCM, GTDM, and GLRLM analyses to evaluate and predict dermatitis in breast cancer patients. 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 non-invasively 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), due to readily available technology to radiation oncology clinics, as well as relaying practical and intuitive information about the macroscopic changes of the skin during RT (43).

Radiation therapy remains a crucial component in the post-operative management of breast cancer. The associated side effects from treatment may affect patients' 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 two weeks of treatment (44). In our patient cohort, 41.1% of patients had a CTCAE ≥ 2. Therefore, it 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 and has several advantages such as portability, relatively low cost compared to 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% ($p<0.0001$) and improved post-radiation patient satisfaction (45). Thermal imaging may also provide a method to validate the efficacy of pharmacological agents to manage skin toxicity (16,48). For instance, while glucocorticosteroids are successful in treating radiation-induced skin toxicity, their anti-inflammatory 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). Quantitative thermal imaging 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 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 low sample size, which affects the framework of the prediction model. First, small sample sets limit the approach for group classification (i.e. sufficient samples and distributions are required between classes). Second, model testing and validation in small sample sets have a greater risk of yielding an overfitted 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, while 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 prior to imaging. We attempted to reduce this effect by instructing patients to change into standard hospital gowns prior to 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 anatomical changes (i.e. changes in the size of the breast), as well as 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 (H&N) radiation oncology. Severe skin toxicity, characterized as confluent moist desquamation (i.e. grade 3) presents in approximately 23% of H&N patients (52); thus, early thermoradiomic

markers in this setting may significantly improve management strategies at the beginning of the seven-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 to current practices are limited by the available imaging tools that can objectively measure skin toxicity. As a result, visual inspection and patient-reported symptoms remains the primary method, but may 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 radiotherapy 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, Taylor C, Arriagada R, Clarke M, 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 randomised trials. Lancet. 2011;378(9804):1707–16.

2. McGale P, Taylor C, Correa C, Cutter D, Duane F, Ewertz M, 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 randomised trials. Lancet. 2014;383(9935):2127–35.

3. Tesselaar E, Flejmer AM, Farnebo S, Dasu A. Changes in skin microcirculation during radiation therapy for breast cancer. Acta Oncol (Madr). 2017;56(8):1072–80.

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(3):e357–62.

5. Ryan JL. Ionizing Radiation: The Good, the Bad, and the Ugly. J Invest Dermatol [Internet]. 2012 Mar;132(3):985–93. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0022202X15356608

6. Wei J, Meng L, Hou X, Qu C, Wang B, Xin Y, et al. Radiation-induced skin reactions: mechanism and treatment. Cancer Manag Res. 2019;11:167–77.

7. Huang CJ, Hou MF, Luo KH, Wei SY, Huang MY, Su SJ, et al. RTOG, CTCAE and WHO criteria for acute radiation dermatitis correlate with cutaneous blood flow measurements. Breast. 2015;24(3):230–6.

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(SUPPL.1):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(1):34–8.

10. Owen R, Ramlakhan S. Infrared thermography in paediatrics: a narrative review of clinical use. BMJ Paediatr Open. 2017;1(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):1–9.

12. Martini G. Juvenile-onset localized scleroderma activity detection by infrared thermography. Rheumatology. 2002 Oct;41(10):1178–82.

13. Maillot O, Leduc N, Atallah V, Escarmant P, Petit A, Belhomme S, et al. Evaluation of acute skin toxicity of breast radiotherapy using thermography: Results of a prospective single-centre trial. Cancer/Radiothérapie [Internet]. 2018 Apr; Available from: http://linkinghub.elsevier.com/retrieve/pii/S1278321818300477

14. Thermography based breast cancer detection using texture features and minimum variance quantization. EXCLI J [Internet]. 2014 Nov [cited 2019 Apr 25];13:1204–15. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4464488/

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 randomised study of betamethasone versus two moisturizing creams.

Radiother Oncol. 2013;108(2):287–92.

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(4):736–8.

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(1):50–3.

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. 2009;(10):669–74.

19. Mulliez T, Veldeman L, Greveling A Van, Speleers B, Sadeghi S, Berwouts D, et al. Hypofractionated whole breast irradiation for patients with large breasts : A randomized trial comparing prone and supine positions. Radiother Oncol. 2013;108(2):203–8.

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 Sep;10.

21. Lee TF, Sung KC, Chao PJ, Huang YJ, Lan JH, Wu HY, 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 Jul;13(7).

22. Parekh A, Dholakia AD, Zabranksy DJ, Asrari F, Camp M, Habibi M, 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(1):8–15.

23. van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res [Internet]. 2017 Nov 1;77(21):e104–7. Available from: http://cancerres.aacrjournals.org/lookup/doi/10.1158/0008-5472.CAN-17-0339

24. Welch ML, McIntosh C, Haibe-Kains B, Milosevic MF, Wee L, Dekker A, et al. Vulnerabilities of radiomic signature development: The need for safeguards. Radiother Oncol [Internet]. 2019 Jan;130:2–9. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0167814018335515

25. Haralick RM, Shanmugam K, Dinstein I. Textural Features for Image Classification. IEEE Trans Syst Man Cybern. 1973 Nov;3(6):610–21.

26. Albregtsen F. Statistical Texture Measures Computed from Gray Level Coocurrence Matrices. Image Processing Laboratory, Department of Informatics, University of Oslo. 2008.

27. Thibault G, Fertil B, Navarro C, Pereira S, Cau P, Levy N, et al. Shape and texture indexes application to cell nuclei classification. Int J Pattern Recognit Artif Intell. 2013;27(1).

28. Sun C, Wee WG. Neighboring gray level dependence matrix for texture classification. Comput Vision, Graph Image Process. 1983;23(3):341–52.

29. Chernick MR, Murthy VK, Nealy CD. Application of bootstrap and other resampling techniques: Evaluation of classifier performance. Pattern Recognit Lett. 1985;3(3):167–

78.

30. He H, Garcia E. Learning from imbalanced data. IEEE Trans Knowl Data. 2009;21(9):9.

31. Xu D-H, Kurani AS, Furst JD, Raicu DS. Run-Length Encoding for Volumetric Texture. Heart. 2004;27(25):452–8.

32. Kolossváry M, Kellermayer M, Merkely B, Maurovich-Horvat P. Cardiac Computed Tomography Radiomics. J Thorac Imaging. 2018;33(1):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(3):143–53.

34. Makridakis S, Spiliotis E, Assimakopoulos V. Statistical and Machine Learning forecasting methods: Concerns and ways forward. PLoS One. 2018;13(3):1–26.

35. Bzdok D, Altman N, Krzywinski M. Points of Significance: Statistics versus machine learning. Nat Methods. 2018;15(4):233–4.

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(45):13892–7.

37. Templeton A, Chu J, Sun M, Yao R, Sun J, Coon A, et al. Thermal Effusivity Changes as a Precursor to Moist Desquamation. Radiat Res. 2012;178(4):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(2):204–9.

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(12):1073–8.

40. Micheels J, Aisbjorn B, Sorensen B. Laser doppler flowmetry. A new non-invasive measurement of microcirculation in intensive care? Resuscitation. 1984;12(1):31–9.

41. González Sanchis A, Brualla González L, Sánchez Carazo JL, Gordo Partearroyo JC, Esteve Martínez A, Vicedo González A, et al. Evaluation of acute skin toxicity in breast radiotherapy with a new quantitative approach. Radiother Oncol. 2017;122(1):54–9.

42. Burke-Smith A, Collier J, Jones I. A comparison of non-invasive imaging modalities: Infrared thermography, spectrophotometric intracutaneous analysis and laser Doppler imaging for the assessment of adult burns. Burns. 2015;41(8):1695–707.

43. Murray AK, Herrick AL, King TA. Laser Doppler imaging: A developing technique for application in the rheumatic diseases. Rheumatology. 2004;43(10):1210–8.

44. Drost L, Li N, Vesprini D, Sangha A, Lee J, Leung E, et al. Prospective Study of Breast Radiation Dermatitis. Clin Breast Cancer [Internet]. 2018 Mar; Available from: http://linkinghub.elsevier.com/retrieve/pii/S1526820918300065

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 intra-patient randomised controlled clinical trial of 78 breast cancer patients. Radiother Oncol. 2014;110(1):137–43.

46. Wooding H, Yan J, Yuan L, Chyou TY, Gao S, Ward I, 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(1081).

47. Field CK, Kerstein MD. Overview of wound healing in a moist environment. Am J Surg.

1994;167(1A):2S-6S.

48. Yee C, Wang K, Asthana R, Drost L, Lam H, Lee J, et al. Radiation-induced Skin Toxicity in Breast Cancer Patients: A Systematic Review of Randomized Trials. Clin Breast Cancer. 2018;18(5):e825–40.

49. Grose R, Werner S, Kessler D, Tuckermann J, Huggel K, Durka S, et al. A role for endogenous glucocorticoids in wound repair. EMBO Rep. 2002;3(6):575–82.

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(2):335–43.

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(3):260–8.

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 [Internet]. 2012 Mar;82(4):1454–62. Available from: https://linkinghub.elsevier.com/retrieve/pii/S0360301611005414

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(5):367–72.

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):1–19.

FIGURE CAPTIONS

Figure 1

Schematic demonstrating the workflow and methodology of the study.

Figure 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 4250Gy RT regimen. (E) Thermal image representation of image C. (F) Skin rendition highlighting the area receiving radiation. (G) ROI selection based on the target area outlined in the skin rendition. (H) The grey-level representation of the selected ROI. Temperature scale bar: Figures B, E & G, Grey level scale bar: Figure H; Abbreviations: Temp.= Temperature; μ= Mean; μT= Mean Temperature; Avg= Average.

Figure 3

(A – B). Comparison of ipsilateral and contralateral mean temperature averages between patients evaluated with a CTCAE score of either ≥ 2 or ≤ 1 at baseline and at every 5^{th} RT fraction. (C) Sample distribution comparison of mean temperature values between CTCAE$\geq$2 and CTCAE$\leq$1 groups after 10 fractions of RT for both ipsilateral and contralateral sides. Mean temperature ($\Delta\mu T_{Avg}$) value differences between CTCAE$\leq$1 and CTCAE$\geq$2 patient groups. * = $p<0.05$ ** = $P\leq 0.01$, based on independent t-test. Error bars represent standard error of the mean.

Figure 4

Test Receiver Operating Characteristic (ROC) curves for two representative experiments.