Prediction of skin dose in low-kV intraoperative radiotherapy using machine learning models trained on results of *in vivo* dosimetry

Michele Avanzo,[a] and Giovanni Pirrone
Division of Medical Physics, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, 33081 Aviano, PN, Italy

Mario Mileto, and Samuele Massarut
Department of Breast Surgery, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, 33081 Aviano, PN, Italy

Joseph Stancanello, and Milad Baradaran-Ghahfarokhi
Division of Medical Physics, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, 33081 Aviano, PN, Italy

Alexandra Rink
Department of Radiation Physics, Princess Margaret Cancer Centre, ON M5G 2M9, Canada

Loredana Barresi
Division of Medical Physics, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, 33081 Aviano, PN, Italy

Lorenzo Vinante
Radiation Oncology, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, 33081 Aviano, PN, Italy

Erica Piccoli
Department of Breast Surgery, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, 33081 Aviano, PN, Italy

Marco Trovo
Department of Radiation Oncology, Udine General Hospital, 33100 Udine, UD, Italy

Issam El Naqa
Department of Radiation Oncology, Physics Division, University of Michigan, Ann Arbor, MI 48103-493, USA

Giovanna Sartor
Division of Medical Physics, Centro di Riferimento Oncologico di Aviano (CRO) IRCCS, 33081 Aviano, PN, Italy

(Received 27 August 2018; revised 26 November 2018; accepted for publication 1 January 2019; published 25 January 2019)

Purpose: The purpose of this study was to implement a machine learning model to predict skin dose from targeted intraoperative (TARGIT) treatment resulting in timely adoption of strategies to limit excessive skin dose.

Methods: A total of 283 patients affected by invasive breast carcinoma underwent TARGIT with a prescribed dose of 6 Gy at 1 cm, after lumpectomy. Radiochromic films were used to measure the dose to the skin for each patient. Univariate statistical analysis was performed to identify correlation of physical and patient variables with measured dose. After feature selection of predictors of *in vivo* skin dose, machine learning models stepwise linear regression (SLR), support vector regression (SVR), ensemble with bagging or boosting, and feed forward neural networks were trained on results of *in vivo* dosimetry to derive models to predict skin dose. Models were evaluated by tenfold cross validation and ranked according to root mean square error (RMSE) and adjusted correlation coefficient of true vs predicted values (adj-R^2).

Results: The predictors correlated with *in vivo* dosimetry were the distance of skin from source, depth-dose in water at depth of the applicator in the breast, use of a replacement source, and irradiation time. The best performing model was SVR, which scored RMSE and adj-R^2, equal to 0.746 [95% confidence intervals (CI), 95% CI 0.737,0.756] and 0.481 (95% CI 0.468,0.494), respectively, on the tenfold cross validation.

Conclusion: The model trained on results of *in vivo* dosimetry can be used to predict skin dose during setup of patient for TARGIT and this allows for timely adoption of strategies to prevent of excessive skin dose. [https://doi.org/10.1002/mp.13379]

Key words: breast, cancer, *in vivo*, intraoperative, IORT, machine learning, radiochromic

1. INTRODUCTION

An intraoperative radiotherapy (IORT) technique called TARGIT (targeted intraoperative radiotherapy) has been developed based on the Intrabeam system (Carl Zeiss, Oberkochen, Germany). In this technique, a point source, emitting low energy x rays of 50 kVp coupled with a spherical applicator, is inserted into the surgical bed. The irradiation is administered soon after the primary tumor resection during the same operative session. The target tissue is the breast volume surrounding

the excised tumor, wrapped around the radiotherapy source soon after the primary surgery. Two multicenter prospective randomized trials, Targit-A and Targit-B, are currently testing the clinical efficacy of TARGIT, as partial breast technique in selected early-stage low-risk breast cancer patients, and as a boost to the tumor bed before conventional whole breast irradiation (WBRT) for high risk patients, respectively.[1]

The skin represents the main organ at risk in TARGIT, because of its proximity to the source. A few cases of dermatitis and skin necrosis have been reported in early reports[2,3] as well as more recently[4] on this technique. No complications have been reported to other organs such as rib cage, lungs, and heart, receiving a lower dose of radiation because of their larger distance from the radiation source and steep dose fall-off. Safety levels for skin effects as low as 1 Gy[5] or 2 Gy[6] have been recommended to the skin after single fraction low kV irradiation. A dose of 6 Gy has been identified as a reasonably threshold for transient skin injury. One case of grade 2 dermatitis was reported after IORT delivered as a boost prior to WBRT where dose measured by *in vivo* dosimetry was 8.42 Gy.[7]

As skin dose can be critical for IORT, many studies focused on developing techniques to measure *in vivo* dose to the skin using different types of dosimeter, including radiochromic films,[8] Thermoluminescent dosimeters (TLDs),[9,10] and optically stimulated luminescent dosimeters (OSLDs).[11] *In vivo* measurements of dose are also essential as they help identifying systematic and random errors in treatment delivery.[12,13] Currently, individual pretreatment calculation of skin dose is challenged by the lack of treatment planning, mainly due to unsolved difficulties in installing useful in-room imaging systems, one of the major limitations of IORT techniques.[14]

In the present study, we want to implement a model to estimate dose to the skin in TARGIT before beginning of the IORT treatment. This tool would allow timely adoption of strategies to prevent excessive skin dose, such as placing a saline solution–soaked gauze as a spacer around the applicator, in order to increase source to skin distance. For this purpose, we use statistical and/or machine learning algorithms able to infer a hypothesis (the function/model), to predict the labels (skin dose) of out-of-sample observations.[15,16] With the goal of achieving the best possible accuracy, the models are trained on data from *in vivo* skin dosimetry performed with an established technique on a large cohort of patients during more than 4 yr of TARGIT practice at our center.

2. METHODS AND MATERIALS

2.A. Patient data and follow-up

From October 2013 to March 2018, 283 patients with invasive breast carcinoma underwent TARGIT after lumpectomy. Patients and treatments relevant data are summarized in Table I.

Patients were evaluated by expert breast surgeons during the 3 days after lumpectomy and then weekly until the complete wound healing. Incidences of acute toxicities, and in particular acute skin reactions, were collected during the first months after IORT. Subsequent clinical follow-up was scheduled every 6 months during the first 3 yr after IORT and then yearly. Bilateral mammography and ultrasonography were performed annually and late toxicities were evaluated during the follow-up visits.

TABLE I. Overview of patient and *in vivo* dosimetry data. For dichotomic variables, the number (percentage) of patients, and for continuous variables, the average value with 95% confidence intervals (CI) is specified.

Characteristics (N = 283)	Data
Tumor site (inner/central/outer):	
Inner	54 (19.1%)
Central	53 (18.7%)
Outer	176 (62.2%)
Tumor site (inferior/central/superior)	
Inferior	43 (15.2%)
Central	77 (27.2%)
Superior	163 (57.6%)
Prescribed dose	
6 Gy	262 (92.6%)
5 Gy	21 (7.4%)
Intrabeam source used, serial number	
Source 1, #338	111 (39.2%)
Source 2, #321	127 (44.9%)
Temporary replacement	45 (15.9%)
Applicator diameter (3.5/4/4.5/5 cm)	
3.5 cm	6 (2.1%)
4.0 cm	40 (14.1%)
4.5 cm	69 (24.4%)
5.0 cm	168 (59.4%)
Age [mean (95% CI)]	65 (34–85)
Depth of applicator [mean (95% CI)] (cm)	1.4 (0.8–2.5)
Distance film applicator [mean (95% CI)] (cm)	3.1 (1.5–4.5)
Measured dose [mean (95% CI)] (Gy)	3.21 (1.53–5.38)
Relative standard deviation of measured dose [mean (95% CI)] (%)	5.2 (2.1–9.9)

2.B. *In vivo* dosimetry

During surgery, a wide local excision was carried out to remove the tumor. A spherical applicator with the proper size was chosen based on the excision cavity. A purse string suture was then applied deeply to the whole cavity edges.[17] The treatment delivery time was calculated by the Medical Physics staff using the Intrabeam Treatment Software, in order to deliver a prescribed dose of 5 or 6 Gy to 1 cm from the applicator surface in water, with 50 kVp, 40 μA x rays. The source was attached to the applicator and the gantry. Once the applicator was in place, the purse string was tightened carefully so that the breast tissue wrapped around the applicator. In order to prevent excessive irradiation of the skin, the edges were kept at least 1 cm away from the applicator shaft.[3]

Gafchromic EBT3 (Ashland Special Ingredients, Bridgewater, NJ, USA) films were used to measure dose to the skin on patients who underwent TARGIT. Because the sensitive layer of the film is 0.125 mm from the surface of the patient, correction for the effective point of measurement is negligible for radiochromic films,[18] and it can be safely assumed that EBT3 measures directly the skin dose. Before the surgical procedure, pieces of radiochromic films were wrapped by a nurse of the surgical staff in a thin sterile envelope. Films were calibrated in air using the Intrabeam with spherical applicators, following a previously established procedure.[8]

Two sheets of tungsten-impregnated rubber (0.1 mm lead equivalent; Carl Zeiss Surgical) were placed over the wound, thus covering also the films, to reduce the amount of stray radiation in the operating room during irradiation. Once the treatment finished, the purse string was cut, and the applicator and the shield were removed.[17] Films were collected and scanned using an Epson Expression 1680 (Epson-Seiko Corporation, Japan) flatbed scanner, and film images were converted into dose matrices according to a multichannel scanning protocol.[19–21] For measurement of dose and its uncertainty, average dose and standard deviation were calculated inside a region of interest (ROI) consisting of a 2×2 mm^2 square around the highest dose reading on the piece of film and positioned at least 2 mm from the film edges.

2.C. Statistical analysis

For each patient, age, laterality of breast cancer, and quadrant of breast (inner/midline/outer and upper/medial/lower) where the source was introduced were registered. Before placing the dosimeter, the point where the spherical applicator is at minimum depth under the patient skin was found by palpation. Then, the minimum depth, d_{min}, of the spherical applicator was measured by insertion of a needle and measurement of the extent of needle insertion. The film was placed at the same position on the skin from which d_{min} was measured [Supplementary Information Figs. S1(a)–S1(d)]. The distance from the point of measurement to the point of insertion of the source along the skin was also collected using a ruler. The value of depth–dose curve at d_{min}, $D(d_{min})$ calculated by the Intrabeam software was also recorded, as it was previously postulated to be related to measured skin dose.[10,22]

The prescribed dose, diameter of the spherical applicator, and irradiation time were recorded among the variables related to treatment. Our Intrabeam system has two x-ray sources (serial numbers 321 and 338, respectively). When one of the two sources was not available because of maintenance, repair, or recalibrations, Zeiss provided a temporary replacement. It was also then recorded if source 1, 2 or a replacement source was used. Univariate statistical analysis was performed to find correlation of variables with measured dose, using the Spearman rank correlation test for ordinal variables, while the Kruskal–Wallis rank test was used for categorical variables.

2.D. Model implementation

Before training models, outliers were removed by performing a nonlinear multivariate regression using support vector machine (SVM), which was preferred because of its capability to deal with linear and nonlinear functional mappings. The data points whose residuals were three times higher than the standard deviation (larger than 99% confidence level) of the residuals of the entire dataset, were considered as possibly caused by incorrect measure of distance and/or depth, and were removed from analysis as outliers. Then, the least absolute shrinkage and selection operator (LASSO)[23] was applied in order to eliminate redundant variables that are likely not to be related to the measured dose.

First, an univariable linear regression was attempted using the most relevant predictor, $D(d_{min})$ with the aim to investigate if *in vivo* dosimetry can be predicted also using a unimetric model. Then, the following statistical and machine learning algorithms were trained on the dataset (predictors and *in vivo* dose from patients) to build multimetric models able to predict *in vivo* dose:

1. **Stepwise linear regression (SLR),** which results in a linear multimetric model; terms from a linear or generalized linear model are removed or added in order to find the subset of variables in the dataset resulting in the smallest model with lowest prediction error.[24]
2. **SVR** is the regression version of SVM, a supervised machine learning technique which, by means of a kernel function, projects the data into a higher dimensional feature space and can perform linear regression in this high-dimensional feature space. SVR tolerates errors that are inferior to a margin, thus providing a trade-off between goodness of fit and model robustness.[25]
3. **Ensemble machine learning (EML)** uses multiple learners aggregated into a single learner. We used for weak learners Decision Trees, a popular concept in machine learning.[15] In EML **with bagging (EML/bag)**, a group of decision trees is trained, and the algorithm randomizes training samples by resampling with replacements; second, at each branching step, it chooses an attribute to split among a randomly selected subset of attributes. After a bag of trees is trained, prediction is made for all the individual trees and the most frequent class selected by the trees is taken as a final result.[26]
4. **EML** with boosting (EML/boost), where decision trees are trained sequentially on subsets containing the data that were misclassified in previous training steps, then they are feedback in order to boost performance.[27]
5. **Feed-forward neural network (NN)** using Bayesian regularization training algorithm. Neural networks are described as parallel-distributed computational models that consist of many nonlinear elements arranged in patterns that imitate biological nervous systems.[28]

In order to compare the performance of the different models, we used the root mean square error (RMSE) which describes the difference among modeled and measured dose, and the adjusted correlation coefficient (adj-R^2) of predicted and measured data, which reflects the goodness of fit of the model to the population taking into account the sample size and the number of predictors used.[29]

2.E. Cross validation and analysis of sample size

In order to test out-of-sample accuracy, tenfold cross validation, where the 267 patients were randomly partitioned into ten subsamples of nearly equal size, was performed. Nine subsamples were used to train the model, which was validated on the tenth subsample. Then, the training and validation subsamples were rotated, so that all patients are used once in the validation subsample. RMSE and adj-R^2 were calculated from results in the validation subsamples ($RMSE_v$ and adj-R^2_v). The whole process was repeated 100 times in order to derive average values and 95% confidence intervals (95% CI) of $RMSE_v$ and adj-R^2_v.

In order to determine if the dataset was large enough to build a multivariable model, the SVR was performed on randomly chosen samples of increasing number of patients. The procedure was repeated 1000 times for each sample size. The *P*-value of a *t*-test (zero vs nonzero coefficient in the SVR) was performed for the three most predictive variables: D(d_{min}), distance, and use of a replacement source. The sample size was considered large enough to build a multivariable model when all these three variables reached a *P*-value under the 0.05 significance level.

3. RESULTS

3.A. Patients results

The mean measured skin dose was 3.21 Gy, (95% CI 1.53–5.38 Gy). No acute radiation injury to the skin, necrosis, skin breakdown or delayed wound healing was observed during the patients' follow-up. None of the patients showed evidence of significant clinical complications. No late toxicity (i.e., hyperpigmentation, telangiectasia, ulceration) was found. In patients who received WBRT after TARGIT, no additional risk of skin toxicity was detected in the skin area close to the lumpectomy.

3.B. Statistical analysis

The results of the univariate statistical analysis are shown in Table II. The *in vivo* measured dose was strongly ($P < 0.0001$) correlated with D(d_{min}), distance of film from the applicator, and applicator diameter. Significant but weaker correlations ($P < 0.05$) were found with prescribed dose, irradiation time, and Intrabeam source used. *In vivo* dose was uncorrelated with tumor laterality, point of insertion of the source, and patient age.

TABLE II. Results of univariate statistical analysis of variables related to treatment and *in vivo* dosimetry. For ordinal variables, the *P*-value and correlation coefficient, r, from the Spearman rank correlation test are reported, while for categorical variables, the Kruskal–Wallis rank test's *P*-value is shown.

Variable	*P*-value	r
Prescribed dose (5 Gy or 6 Gy)	<0.05	0.1394
Irradiation time	<0.05	−0.1934
Distance of film from the applicator shaft	<0.0001	−0.2936
Depth of the applicator surface beneath the film (d_{min})	<0.0001	−0.5679
Calculated dose in water at d_{min}, D(d_{min})	<0.0001	0.5751
Spherical applicator diameter	<0.0001	−0.2564
Age of the patient (y)	0.48	−0.0424
Intrabeam source 1	0.387	
Intrabeam source 2	0.2667	
Replacement source	<0.05	
Breast laterality	0.27	
Inner/medial/outer insertion of source	0.14	
Superior/midline/inferior insertion of source	0.97	

3.C. Models development

A total of 16 measurements (5.6%) were removed from the analysis as outliers. In the feature selection, the following variables were selected, and then passed to the machine learning algorithms: prescribed dose, D(d_{min}), use of a replacement source, applicator diameter, laterality of breast (left and right), outer and superior position of the source, distance of the film, and d_{min}.

The plots of estimated vs measured dose from calibration and validation of machine learning methods are shown in Fig. 1. The performances in training and tenfold cross validation are compared in Table III, and the distributions of $RMSE_v$ and adj-R^2_v from the tenfold cross validation are shown in Fig 2.

The univariable linear regression of *in vivo* dose vs D(d_{min}) had a lower $RMSE_v$ and Adj-R^2_t (Wilcoxon Rank sum test $P \ll 0.05$) than multimetric machine learning models in the cross validation, and was excluded in the remainder of the analysis. The machine learning model with the best scores in the tenfold cross validation, lowest $RMSE_v$ and highest adj-R^2_v, was the SVR. The $RMSE_v$ and adj-R^2_v were significantly different among all models (Wilcoxon Rank sum test, $P \ll 0.05$). Therefore, SVR was considered as our best performing model ($RMSE_v = 0.746$, adj-$R^2_v = 0.481$), followed by SLR. The learning curve power analysis (Fig. 3) shows that the *P*-values for inclusion of the three most predictive variables in the SVR model were all below 0.05 for sample size larger than 91 patients. Therefore, the dataset of 267 patients used in the analysis was considered largely adequate to build multivariable models predictive of *in vivo* dose.

4. DISCUSSION

Studies reported results of *in vivo* dosimetry in patients who received TARGIT as soon as the skin was recognized as

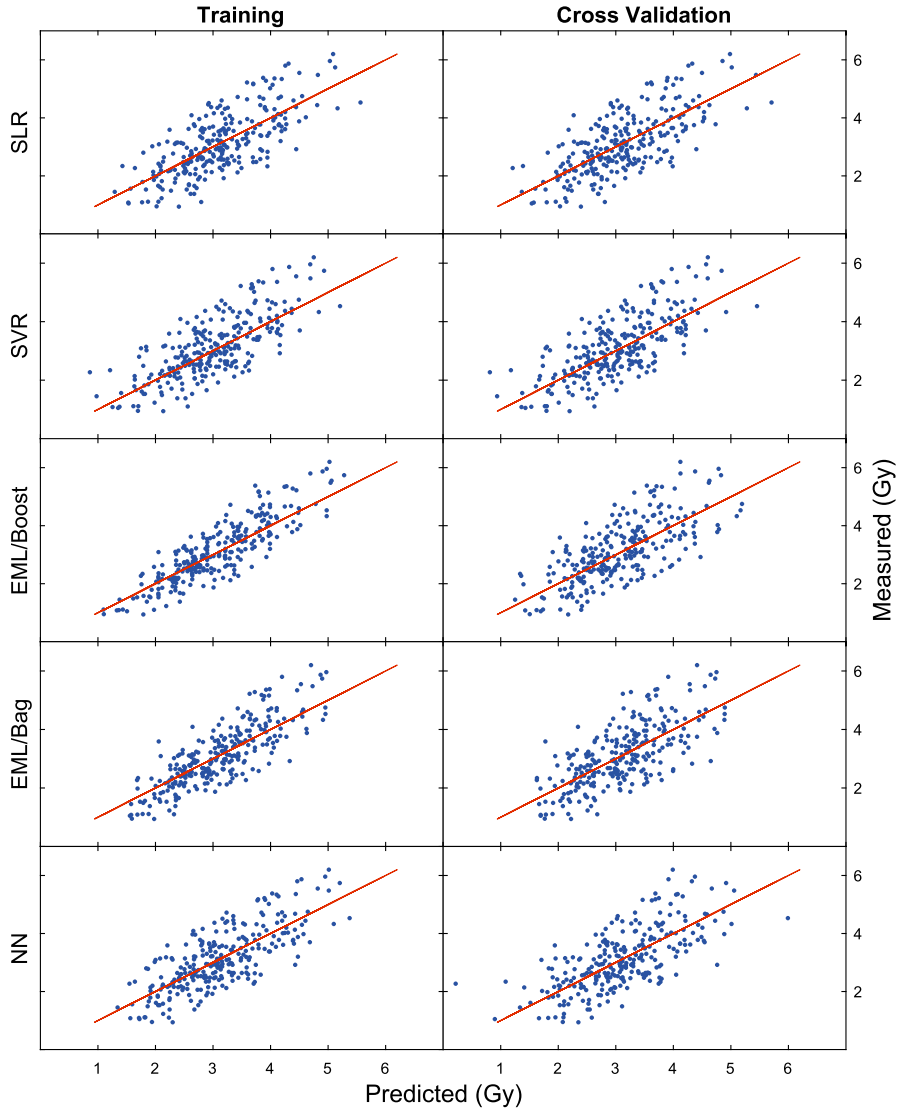

FIG. 1. Calibration and cross validation of machine learning models. [Color figure can be viewed at wileyonlinelibrary.com]

TABLE III. Assessment of machine learning models with different methods. $RMSE_t$, Adj-R^2_t, $RMSE_v$, Adj-R^2_v, are root mean square error, and adjusted R^2 in the training and validation datasets with 95% confidence intervals (CI).

Learning method	$RMSE_t$	Adj-R^2_t	$RMSE_V$ (95% CI)	Adj-R^2_V (95% CI)
SLR	0.725	0.51	0.751 (0.74,0.762	0.474 (0.459,0.489)
SVR	0.714	0.525	0.746 (0.737,0.756)	0.481 (0.468,0.494)
EML Boost	0.536	0.732	0.767 (0.747,0.79)	0.451 (0.417,0.479)
EML Bag	0.611	0.652	0.760 (0.749,0.770)	0.461 (0.447,0.476)
NN	0.683	0.564	0.765 (0.742,0.795)	0.453 (0.41,0.487)
Linear regression on $D(d_{min})$	0.788	0.440	0.791 (0.788,0.797)	0.439 (0.408,0.467)

the main organ at risk during this procedure. *In vivo* dosimetry in a series of 72 patients resulted in average measured dose of 2.9 Gy, with 11% of readings ≥6 Gy.[10] Measurements with TLD on the skin at distances of 5 and 15 mm from the incision on 57 patients resulted in maximum dose of 2.93 Gy on average.[9] The distribution of skin doses measured in our patient dataset is in agreement with the previous studies.

It is no surprise that, in our results, the variable with the strongest correlation was $D(d_{min})$ in water, as it describes the dependency of dose from the applicator diameter, depth of the applicator, the prescribed dose, and the amount of tissue between the skin and the applicator. However, the linear regression model based on $D(d_{min})$ resulted in poor predictive capability compared to other models, thus showing the advantage of using multimetric machine learning to predict skin dose.

The univariate statistical analysis confirmed the dependencies of *in vivo* measured skin dose on distance from the measured point from the applicator shaft, in agreement with previous reports.[9,10] The distance from the applicator shaft was an independent predictor of skin dose possibly

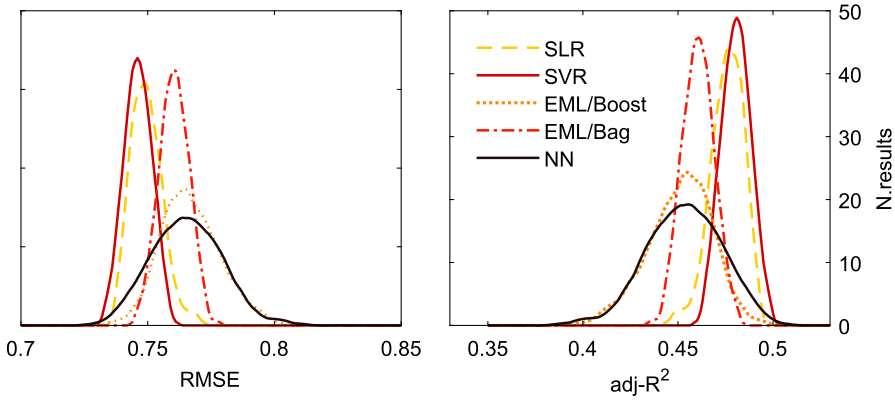

FIG. 2. Distribution of values of RMSE and adj-R^2 from tenfold cross validation repeated 100 times for multivariable machine learning models. [Color figure can be viewed at wileyonlinelibrary.com]

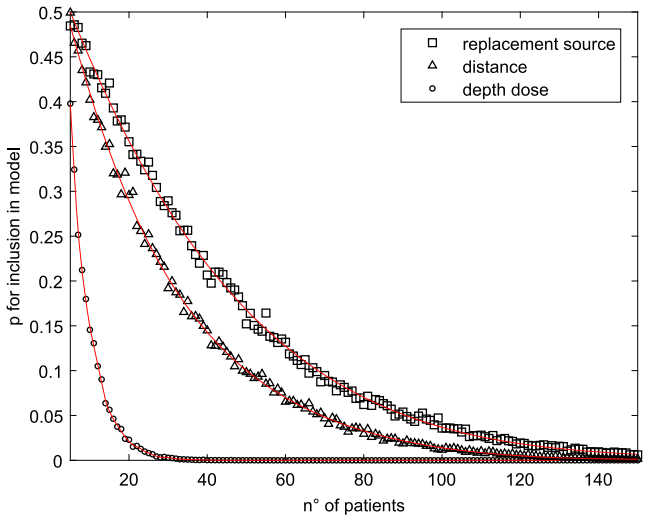

FIG. 3. Analysis of statistical power of the study by learning curve. The *P*-value of *t*-tests (zero vs nonzero coefficient) for inclusion of the three most predictive variables of *in vivo* dosimetry in the SVR model is plotted against the number of patients used for training. [Color figure can be viewed at wileyonlinelibrary.com]

because, for a fixed depth, an increase in distance implies a different geometry of irradiation, where the applicator is more oblique.

Monte Carlo simulation of a breast phantom with realistic tissue compositions and skin layers, demonstrated that *in vivo* dose depends on the size of applicator[22] as well as the amount of breast tissue between the applicator and the skin. In the study of Fogg et al.,[9] the ratios between average doses measured with TLDs at 5 mm and at 15 mm from the point of insertion were 1.47, 1.22, 1.24, and 1.11 for the 3.5, 4.0, 4.5, and 5.0 cm applicators, respectively. This effect has been explained with more penetrating beam spectrum with larger applicators,[22] but it should also be noted that the use of larger applicators is preferred on patients with larger breast size, as confirmed by correlation in our dataset among d_{min} and applicator diameter (Spearman correlation test $P \ll 0.0001$).As expected, in the univariate analysis, measured dose was also correlated with other variables related to the quantity of radiation emitted, such as the irradiation time and the prescribed dose which is specified, according to the TARGIT protocol, as the dose at 1 cm from the applicator surface in water. Seven units were used for spare sources, when one of the two sources was not available because of maintenance, repair or recalibrations, and the use of a replacement source was related to *in vivo* dose. Dosimetric and spectrum characteristics different from the other sources may explain this result. Differences in the output dose rate of different sources have been reported.[30] However, given that the dose rate is checked and, if necessary, readjusted before each treatment,[4] these are unlikely to affect measured dose. Small differences in the various structures involved in the generation of x rays, including the electron source, the beam deflector as well as the gold target, such as between manufactured and ideal target thickness, could change photon spectra and doses between different sources, thus explaining a change of dose.[22] In our results, the feature selection successfully removed those variables that are not correlated with measured dose (e.g., age) and those that suffer from collinearity and are redundant, such as d_{min}, which is related to $D(d_{min})$.

Among the multimetric machine learning models, SVR showed the best performance (lowest RMSE and highest adj-R^2) on the cross validation, and therefore, we consider it as the model of choice. This result is in agreement with other reports where SVM classifiers were among the best performing machine learning methods.[31] The second best method, SLR, has the advantage to provide a simple calculation to be used in an operating room. For this reason, the formula resulting from the SLR is shown in Table IV.

Our machine learning models were trained on a dataset of 283 patients, which is the largest published results of *in vivo* dosimetry in IORT. The plot of the model residuals shows that the *in vivo* dose calculated by considered models has uncertainties. A possible explanation for the residuals could be the manual measurements of d_{min} and distance, whose accuracy could be improved, for example, by using ultrasound imaging to measure the thickness of breast tissue between the applicator and the point of interest. Skin dose

TABLE IV. Estimated coefficients from the SLR modeling with standard error (SE) and *P*-value for inclusion in the model.

Coefficient of SLR model	Estimate	SE	*P*-value
(Intercept)	1,986	0,209	$1.2{\cdot}10^{-18}$
Distance	−0,246	0,048	$4.8{\cdot}10^{-7}$
$D(d_{min})$	0,408	0,027	$6.46{\cdot}10^{-37}$
Use of a replacement source	−0,555	0,124	$1.16{\cdot}10^{-5}$

may also vary because of variation in the compositions of the treated breast. Treatments are usually performed covering the patient surface over the applicator with a lead-rubber protection sheet whose backscatter could increase skin dose, however, to a relatively small amount.[14] As all our treatments were performed with the protection sheet, it was not possible to verify how much the skin dose is dependent on its use.

By using the model recommended by this study, it is possible to anticipate measured *in vivo* dose from TARGIT in the operating room. This work shows that the machine learning approach can be applied to *in vivo* dosimetry, as successfully done in other areas of radiotherapy workflow, including error detection and prevention,[16] treatment dose planning,[32] and verification.[33,34] It also suggests that the results of *in vivo* dosimetry should be included among other categories of patient data in the electronic patient records in order to be available for automated data extraction with the aim of generating or improving prediction models for patient dose.

A word of caution is needed. The availability of the models provided here to predict skin dose should not prevent measured *in vivo* dosimetry to the skin, which remains fundamental as an independent quality check of radiotherapy identifying errors in delivery, for example, caused by wrong calibration of the machine. *In vivo* dosimetry still holds as an effective method to validate the introduction and any modification to the delivery of this intraoperative technique by measuring dose to the skin and other critical structures as required.

5. CONCLUSION

This work shows the potential of extending the machine learning approach to *in vivo* dosimetry in IORT. Models for the prediction of *in vivo* dose to the skin in TARGIT were derived from measurements of skin dose performed for over 4 yr at our center with a consistent technique. Of the machine learning models used, SVR was the best performing model and can be used to predict and evaluate dose to the skin before treatment delivery and to adopt measures to minimize excessive dose to the skin.

FUNDING

The present research was in part supported by CRUP – Friuli Exchange Program (grant CUP J38C13001310007) and in part by 5 × 1000 (fund J32F16001310007).

CONFLICT OF INTEREST

The authors have no conflicts to disclose.

[a]Author to whom correspondence should be addressed. Electronic mail: mavanzo@cro.it; Telephone: +39 0434659175.

REFERENCES

1. Vaidya JS, Baum M, Tobias JS, et al. Long-term results of targeted intraoperative radiotherapy (Targit) boost during breast-conserving surgery. *Int J Radiat Oncol Biol Phys*. 2011;81:1091–1097.
2. Joseph DJ, Bydder S, Jackson LR, et al. Prospective trial of intraoperative radiation treatment for breast cancer. *ANZ J Surg*. 2004;74:1043–1048.
3. Chua BH, Henderson MA, Milner AD. Intraoperative radiotherapy in women with early breast cancer treated by breast-conserving therapy. *ANZ J Surg*. 2010;81:65–69.
4. Maoz Z, Ayelet S, Michelle L, et al. Short-term complications of intraoperative radiotherapy for early breast cancer. *J Surg Oncol*. 2016;113:370–373.
5. Shope TB. Radiation-induced skin injuries from fluoroscopy. *Radiographics*. 1996;16:1195–1199.
6. Geleijns J, Wondergem J. X-ray imaging and the skin: radiation biology, patient dosimetry and observed effects. *Radiat Prot Dosimetry*. 2005;114:121–125.
7. Lee JJB, Choi J, Ahn SG, et al. In vivo dosimetry and acute toxicity in breast cancer patients undergoing intraoperative radiotherapy as boost. *Radiat Oncol J*. 2017;35:121–128.
8. Avanzo M, Rink A, Dassie A, et al. In vivo dosimetry with radiochromic films in low-voltage intraoperative radiotherapy of the breast. *Med Phys*. 2012;39:2359–2368.
9. Fogg P, Das KR, Kron T, Fox C, Chua B, Hagekyriakou J. Thermoluminescence dosimetry for skin dose assessment during intraoperative radiotherapy for early breast cancer. *Austr Phys Eng Sci Med*. 2010;33:211–214.
10. Eaton DJ, Best B, Brew-Graves C, et al. In vivo dosimetry for single-fraction targeted intraoperative radiotherapy (TARGIT) for breast cancer. *Int J Radiat Oncol Biol Phys*. 2012;82:e819–e824.
11. Price C, Pederson A, Frazier C, Duttenhaver J. In vivo dosimetry with optically stimulated dosimeters and RTQA2 radiochromic film for intraoperative radiotherapy of the breast. *Med Phys*. 2013;40:091716.
12. Higgins PD, Alaei P, Gerbi BJ, Dusenbery KE. In vivo diode dosimetry for routine quality assurance in IMRT. *Med Phys*. 2003;30:3118–3123.
13. Essers M, Mijnheer BJ. In vivo dosimetry during external photon beam radiotherapy. *Int J Radiat Oncol Biol Phys*. 1999;43:245–259.
14. Hensley FW. Present state and issues in IORT Physics. *Radiat Oncol*. 2017;12:37.
15. Avanzo M, Stancanello J, El Naqa I. Beyond imaging: the promise of radiomics. *Phys Med*. 2017;38:122–139.
16. Feng M, Valdes G, Dixit N, Solberg TD. Machine learning in radiation oncology: opportunities, requirements, and needs. *Front Oncol*. 2018;8:110.
17. Vaidya JS, Baum M, Tobias JS, Morgan S, D'Souza D. The novel technique of delivering targeted intraoperative radiotherapy (Targit) for early breast cancer. *Eur J Surg Oncol*. 2002;28:447–454.
18. Reynolds TA, Higgins P. Surface dose measurements with commonly used detectors: a consistent thickness correction method. *J Appl Clin Med Phys*. 2015;16:358–366.
19. Devic S, Tomic N, Soares CG, Podgorsak EB. Optimizing the dynamic range extension of a radiochromic film dosimetry system. *Med Phys*. 2009;36:429–437.
20. Micke A, Lewis DF, Yu X. Multichannel film dosimetry with nonuniformity correction. *Med Phys*. 2011;38:2523–2534.
21. Lewis D, Micke A, Yu X, Chan MF. An efficient protocol for radiochromic film dosimetry combining calibration and measurement in a single scan. *Med Phys*. 2012;39:6339–6350.
22. Moradi F, Ung NM, Khandaker MU, , et al. Monte Carlo skin dose simulation in intraoperative radiotherapy of breast cancer using spherical applicators. *Phys Med Biol*. 2017;62:6550–6566.

23. Coroller TP, Grossmann P, Hou Y, et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. *Radiother Oncol*. 2015;114:345–350.
24. Franke GR. Stepwise regression. In: Sheth J, Malhotra N, eds. Wiley International Encyclopedia of Marketing. John Wiley & Sons, Inc.; 2010.
25. Chen S, Zhou S, Yin FF, Marks LB, Das SK. Investigation of the support vector machine algorithm to predict lung radiation-induced pneumonitis. *Med Phys*. 2007;34:3808–3814.
26. Yang D, Rao G, Martinez J, Veeraraghavan A, Rao A. Evaluation of tumor-derived MRI-texture features for discrimination of molecular subtypes and prediction of 12-month survival status in glioblastoma. *Med Phys*. 2015;42:6725–6735.
27. Polikar R. Ensemble based systems in decision making. *IEEE Circ Syst Mag*. 2006;6:21–45.
28. Lee S, El Naqa I. Machine learning methodology. In: El Naqa I, Li R, Murphy MJ, eds. *Machine Learning in Radiation Oncology: Theory and Applications*. Cham, Switzerland: Springer International Publishing; 2015:21–39.
29. Montgomery DC, Runger GC. *Applied Statistics and Probability for Engineers*, 7th edn. Hoboken, NJ: Wiley; 2018.
30. Armoogum KS, Parry JM, Souliman SK, Sutton DG, Mackay CD. Functional intercomparison of intraoperative radiotherapy equipment – photon radiosurgery system. *Radiat Oncol*. 2007;2:11.
31. Fernàndez-Delgado M, Cernadas E, Barro S, Amorim D. Do we need hundreds of classifiers to solve real world classification problems? *J Mach Learn Res*. 2014;15:3133–3181.
32. Zhu X, Ge Y, Li T, Thongphiew D, Yin FF, Wu QJ. A planning quality evaluation tool for prostate adaptive IMRT based on machine learning. *Med Phys*. 2011;38:719–726.
33. Valdes G, Solberg TD, Heskel M, Ungar L, Simone CB. Using machine learning to predict radiation pneumonitis in patients with stage I non-small cell lung cancer treated with stereotactic body radiation therapy. *Phys Med Biol*. 2016;61:6105–6120.
34. Valdes G, Chan MF, Lim SB, Scheuermann R, Deasy JO, Solberg TD. IMRT QA using machine learning: a multi-institutional validation. *J Appl Clin Med Phys*. 2017;18:279–284.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1: (a–d) Positioning of radiochromic films on the patient's skin for *in vivo* dosimetry. The point where the spherical applicator is at minimum depth under the patient skin is found by palpation (a). Then the minimum depth, d_{min} of the spherical applicator is measured by insertion of a needle (b) and measurement of the extent of needle insertion (c). The film is placed at the same position on the skin from which d_{min} was measured (d).