

Scientific Article

Development and Optimization of a Machine-Learning Prediction Model for Acute Desquamation After Breast Radiation Therapy in the Multicenter REQUITE Cohort



Mahmoud Aldrimli, PhD,^a Sarah Osman, PhD,^b Diana Grishchuck, MSc,^c Samuel Ingram, MSc,^d Robert Lyon, PhD,^e Anil Mistry, MSc,^f Jorge Oliveira, PhD,^g Robert Samuel, MBChB,^h Leila E.A. Shelley, PhD,ⁱ Daniele Soria, PhD,^j Miriam V. Dwek, PhD,^k Miguel E. Aguado-Barrera, MD, PhD,^{l,m} David Azria, MD,ⁿ Jenny Chang-Claude, PhD,^{o,p} Alison Dunning, PhD,^q Alexandra Giraldo, MD,^r Sheryl Green, MD,^s Sara Gutiérrez-Enríquez, PhD,^t Carsten Herskind, PhD,^u Hans van Hulle, MD,^v Maarten Lambrecht, MD,^w Laura Lozza, MD,^x Tiziana Rancati, MSc,^y Victoria Reyes, MD,^r Barry S. Rosenstein, PhD,^z Dirk de Ruyscher, MD,^{aa} Maria C. de Santis, MD,^x Petra Seibold, PhD,^o Elena Sperk, MD,^u R. Paul Symonds, MD,^{bb} Hilary Stobart,^{cc}

Sources of support: This research collaboration was formed by the UK Radiotherapy Machine Learning Network (RTML) funded through the Advanced Radiotherapy Challenge+ by the Science and Technology Facilities Council (STFC). The REQUITE study received funding from the European Union's 7th Framework Programme for research, technological development and demonstration under grant agreement no. 601826. The research was supported by the Quintin Hogg Trust research awards award no.165435391. The workshops were hosted by the University of Manchester and the Health and Innovation Ecosystem at the University of Westminster. Dr Alison M. Dunning was supported by Cancer Research-UK C8197/A16565. Dr Sara Gutiérrez-Enríquez is supported by the ISCIII Miguel Servet II Program (CP16/00034). Dr Tim Rattay is currently an NIHR Clinical Lecturer (CL 2017-11-002). He was previously funded by a National Institute of Health Research (NIHR) Doctoral Research Fellowship (DRF 2014-07-079). This publication presents independent research funded by the NIHR. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health. Dr Leila Shelley reports grants from Chief Scientist Office (CSO) Scotland grant (TCS/17/26 - CSO Award). Dr Petra Seibold is supported by the ERA-Net ERAPerMed / German Federal Ministry of Education and Research (BMBF) as well as German Federal Office for Radiation Protection (BfS). Dr Elena Sperk was previously supported by the Ministry of Science and Arts of the State of Baden-Württemberg (2017-19) through the Brigitte-Schlieben-Lange-Programme. Dr Ana Vega is supported by Spanish Instituto de Salud Carlos III (ISCIII) funding, an initiative of the Spanish Ministry of Economy and Innovation partially supported by European Regional Development FEDER Funds (INT15/00070, INT16/00154, INT17/00133, INT20/00071, PI19/01424, PI16/00046, PI13/02030, PI10/00164), and through the Autonomous Government of Galicia (Consolidation and structuring program: IN607B). Prof Catharine West is supported by Cancer Research UK (C1094/A18504, C147/A25254) and by the NIHR Manchester Biomedical Research Centre.

Disclosures: Prof David Azria: has been involved in the creation of the start-up NovaGray in 2015. Prof Dirk de Ruyscher: none related to the current manuscript. Outside the current manuscript: advisory board of Astra Zeneca, Bristol-Myers-Squibb, Roche/ Genentech, Merck/ Pfizer, Celgene, Noxxon, Mologen and has received investigator initiated grants from Bristol-Myers-Squibb, Boehringer Ingelheim and Astra-Zeneca. Dr Elena Sperk: none related to the current manuscript. Outside the current manuscript: General speakers bureau Zeiss Meditec, travel support Zeiss Meditec. The other authors have no conflict of interests to disclose.

Data sharing statement: Research data are stored in an institutional repository and will be shared upon request to the corresponding author and the REQUITE consortium.

¹ Members of the REQUITE consortium are listed in the Acknowledgments.

*Corresponding author; E-mail: tr104@le.ac.uk

<https://doi.org/10.1016/j.adro.2021.100890>

2452-1094/© 2022 The Authors. Published by Elsevier Inc. on behalf of American Society for Radiation Oncology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Begoña Taboada-Valadares, MD,^{i,dd} Christopher J. Talbot, PhD,^{ee}
 Vincent J.L. Vakaet, MD,^v Ana Vega, PhD,^{l,m,ff} Liv Veldeman, MD, PhD,^v
 Marlon R. Veldwijk, PhD,^u Adam Webb, PhD,^{ee} Caroline Weltens, MD,^w
 Catharine M. West, PhD,^{gg} Thierry J. Chausalet, PhD,^a and
 Tim Rattay, MBChB, PhD^{bb,*}, REQUITE consortium¹

^aHealth Innovation Ecosystem, University of Westminster, London, United Kingdom; ^bPatrick G. Johnston Centre for Cancer Research, Queen's University Belfast, Belfast, United Kingdom; ^cImperial College Healthcare NHS Trust, London, United Kingdom; ^dDivision of Cancer Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, United Kingdom; ^eDepartment of Computer Science, Edge Hill University, Ormskirk, Lancashire, United Kingdom; ^fGuy's and St. Thomas' NHS Foundation Trust, London, United Kingdom; ^gMirada Medical, Oxford, United Kingdom; ^hUniversity of Leeds, Leeds Cancer Centre, St. James's University Hospital, Leeds, United Kingdom; ⁱEdinburgh Cancer Centre, Western General Hospital, Edinburgh, United Kingdom; ^jSchool of Computing, University of Kent, Canterbury, United Kingdom; ^kSchool of Life Sciences, University of Westminster, London, United Kingdom; ^lFundación Pública Galega Medicina Xenomica, Santiago de Compostela, Spain; ^mInstituto de Investigación Sanitaria de Santiago (IDIS), Servicio Galego de Saúde (SERGAS), Santiago de Compostela, Spain; ⁿUniversity of Montpellier, Montpellier, France; ^oDivision of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany; ^pUKE University Cancer Center Hamburg, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; ^qCentre for Cancer Genetic Epidemiology, University of Cambridge, Strangeways Research Laboratory, Worts Causeway, Cambridge, United Kingdom; ^rRadiation Oncology Department, Vall d'Hebron Hospital Universitari, Vall d'Hebron Hospital Campus, Barcelona, Spain; ^sDepartment of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, New York; ^tHereditary Cancer Genetics Group, Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron Hospital Campus, Barcelona, Spain; ^uDepartment of Radiation Oncology, Universitätsmedizin Mannheim, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany; ^vDepartment of Human Structure and Repair, Ghent University, Ghent, Belgium; ^wDepartment of Radiation Oncology, University Hospital, Leuven, Belgium; ^xDepartment of Radiation Oncology 1, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; ^yProstate Cancer Program, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; ^zIcahn School of Medicine at Mount Sinai, New York, New York; ^{aa}Maastricht University Medical Center, Department of Radiation Oncology (Maastro), GROW, Maastricht, The Netherlands; ^{bb}Cancer Research Centre, University of Leicester, Leicester, United Kingdom; ^{cc}Independent Cancer Patients' Voice, London, United Kingdom; ^{dd}Department of Radiation Oncology, Complejo Hospitalario Universitario de Santiago, Servicio Galego de Saúde (SERGAS), Santiago de Compostela, Spain; ^{ee}Department of Genetics and Genome Biology, University of Leicester, Leicester, United Kingdom; ^{ff}Biomedical Network on Rare Diseases (CIBERER), Madrid, Spain; ^{gg}University of Manchester, Christie Hospital, Manchester, United Kingdom

Received October 11, 2021; accepted December 6, 2021

Abstract

Purpose: Some patients with breast cancer treated by surgery and radiation therapy experience clinically significant toxicity, which may adversely affect cosmesis and quality of life. There is a paucity of validated clinical prediction models for radiation toxicity. We used machine learning (ML) algorithms to develop and optimise a clinical prediction model for acute breast desquamation after whole breast external beam radiation therapy in the prospective multicenter REQUITE cohort study.

Methods and Materials: Using demographic and treatment-related features ($m = 122$) from patients ($n = 2058$) at 26 centers, we trained 8 ML algorithms with 10-fold cross-validation in a 50:50 random-split data set with class stratification to predict acute breast desquamation. Based on performance in the validation data set, the logistic model tree, random forest, and naïve Bayes models were taken forward to cost-sensitive learning optimisation.

Results: One hundred and ninety-two patients experienced acute desquamation. Resampling and cost-sensitive learning optimisation facilitated an improvement in classification performance. Based on maximising sensitivity (true positives), the “hero” model was the cost-sensitive random forest algorithm with a false-negative: false-positive misclassification penalty of 90:1 containing $m = 114$ predictive features. Model sensitivity and specificity were 0.77 and 0.66, respectively, with an area under the curve of 0.77 in the validation cohort.

Conclusions: ML algorithms with resampling and cost-sensitive learning generated clinically valid prediction models for acute desquamation using patient demographic and treatment features. Further external validation and inclusion of genomic markers in ML prediction models are worthwhile, to identify patients at increased risk of toxicity who may benefit from supportive intervention or even a change in treatment plan.

© 2022 The Authors. Published by Elsevier Inc. on behalf of American Society for Radiation Oncology. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Introduction

Radiation therapy is recommended for all patients with breast cancer who have a local excision and after mastectomy in high-risk patients.¹ Over 70% of patients with breast cancer receive radiation therapy, which reduces local recurrence rates and increases long-term survival.² As survival from breast cancer continues to improve,³ quality of life and survivorship have become increasingly important research priorities.⁴ Risk of radiation toxicity can be estimated from empirical dosimetric models based on the dose to the target organ and surrounding tissue.⁵ However, there is considerable variation between individual patient normal tissue reaction to radiation therapy and the extent to which they develop toxicity.⁶ Acute toxicity (<90 days from starting treatment) includes breast erythema and desquamation (skin loss).⁷ In a minority of patients, desquamation can cause substantial patient morbidity, worsen the cosmetic outcome after surgery, and affect quality of life.⁸ It can even result in the interruption of radiation therapy or a dose reduction, potentially increasing the risk of local recurrence.

Several studies have examined the association between acute breast radiation toxicity and clinical or treatment risk factors.^{9–18} Nevertheless, statistical models have had limited success to date in predicting individual patient toxicity risk,¹⁹ and there is a paucity of validated prediction models for acute breast radiation toxicity. It is hypothesized that earlier prediction models failed to validate because they did not include sufficient variables to capture the variety of scenarios that occur among individual patients and individual treatment settings. Recent studies have demonstrated the capability of machine learning (ML) to develop predictive models for radiation toxicities in different cancers,^{20,21} including a thermal image-based random forest (RF) classifier for radiation dermatitis (skin erythema) after the first week of radiation therapy.²² Another recently published abstract describes how RF, gradient boosted decision tree, and logistic regression models were trained and validated on treatment planning and patient data comprising 230 variables including toxicity symptoms from patients at 5 collaborating U.S. centers to predict moist desquamation and Common Terminology Criteria for Adverse Events (CTCAE) grade ≥ 2 radiation dermatitis.²³

For cancers with generally good local tumor control such as breast cancer, it is hypothesized that if a patient's individual risk of radiation toxicity could be estimated at the time of diagnosis, this could inform discussions about risks and benefits and allow treatment plans to be personalized for high-risk patients to minimize toxicity. Clinicians are particularly interested in models that include readily available clinical and treatment variables, which would allow toxicity risk to be estimated before treatment is planned. It is also important to predict toxicities that are sufficiently significant to warrant increased supportive

intervention or treatment de-escalation. To that extent, a logistic regression model for acute breast desquamation after external beam radiation therapy (EBRT) recently developed in 3 combined radiation therapy cohorts failed to validate externally in the multicenter international REQUITE cohort.²⁴ Therefore, the aim of this study was to use ML algorithms to develop and optimise a prediction model for acute breast desquamation after EBRT in the REQUITE breast cancer cohort.

Methods and Materials

This was a TRIPOD (transparent reporting of a multi-variable prediction model for individual prognosis or diagnosis) type 2a study using a single data set with a random split sample for development and validation.²⁵ The full study design is shown in [Figure 1](#).

Study cohort and participants

REQUITE is an international, prospective cohort study that recruited patients with cancer before radiation therapy in 26 hospitals from 8 countries between April 2014 and March 2017 with unified standardized data collection.²⁶ Patient baseline characteristics and methodology have been described in detail elsewhere.²⁷ The present study used data from the breast cancer cohort ($n = 2069$). All patients were treated with breast-conserving surgery followed by whole breast EBRT according to local protocol. Partial breast irradiation and brachytherapy were excluded. Patients were assessed at the start and end of radiation therapy, and annually thereafter. Data collected at the start and at the end of radiation treatment were used to document acute toxicity. All patients gave written informed consent. The study was approved by local ethics committees in participating countries and registered at www.controlled-trials.com (ISRCTN 98496463).

Endpoint definition

Toxicity in REQUITE was scored by treating physicians using CTCAE v4.0.²⁸ CTCAE v4.0 has separate scales for radiation dermatitis (erythema) and skin ulceration (skin loss). The primary endpoint of this study was acute desquamation (skin loss or moist desquamation) occurring by the end of radiation therapy, defined as either CTCAE grade ≥ 3 radiation dermatitis (moist desquamation) or CTCAE grade ≥ 1 skin ulceration, implying that skin integrity was broken over the breast or in the inframammary fold. Patients with high baseline scores

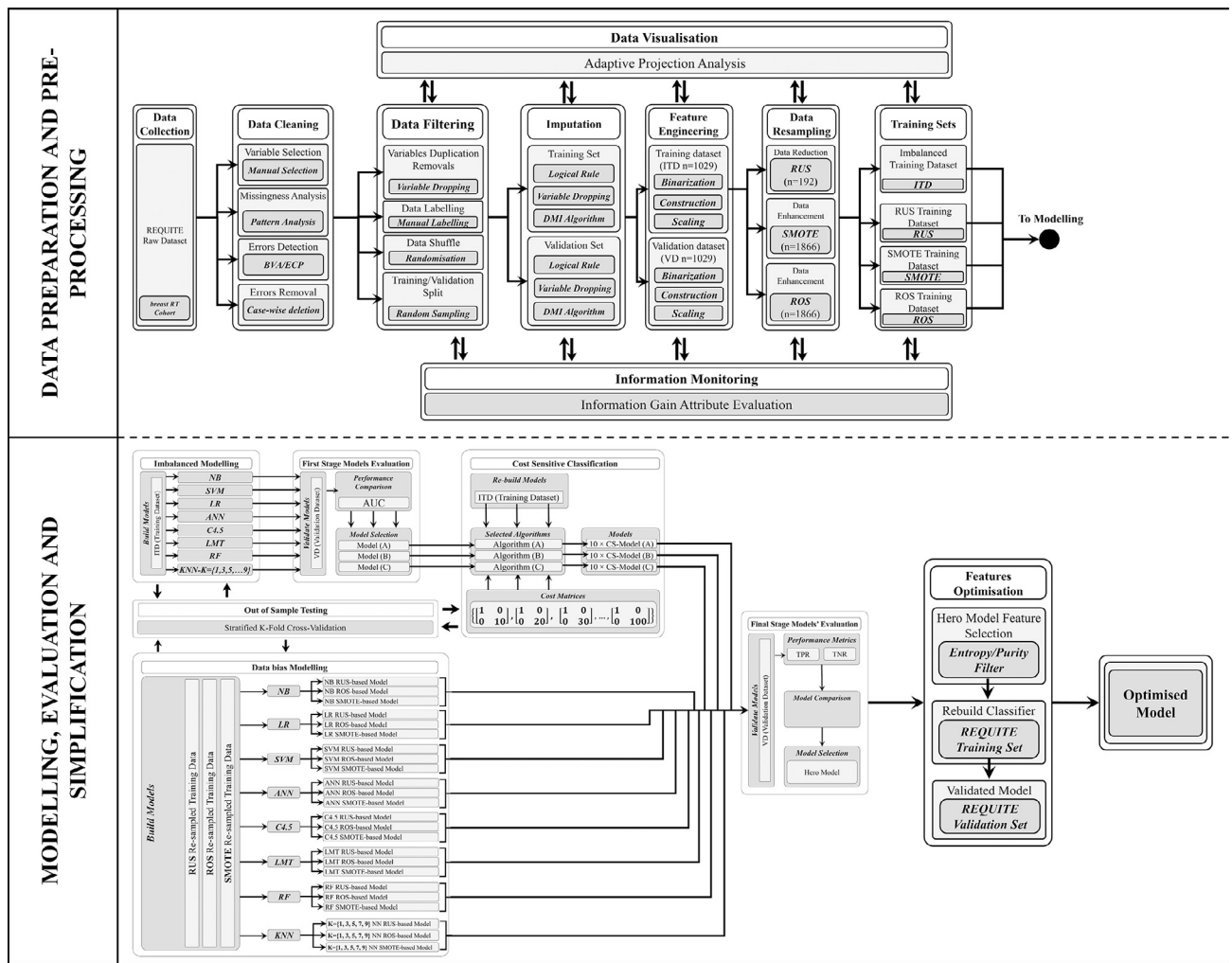


Fig. 1 Diagram depicting the overall study design showing data preprocessing, splitting, and imputation at the top, and model development and optimization at the bottom. Abbreviations: ANN = artificial neural network; BVA = boundary value analysis; C4.5 = C4.5 decision tree; CS = cost sensitive optimisation; DMI = decision-tree-based missing-value imputation; ECP = equivalence class partitioning; ITD = imbalanced training data set; KNN = K-nearest neighbor; LMT = logistic model tree; LR = logistic regression; NB = naive Bayes; RF = random forest; ROS = random over-sampling; RUS = random under-sampling; SMOTE = synthetic minority oversampling technique; SVM = support vector mechanism; VD = validation data set.

(defined as CTCAE grade >1 radiation dermatitis or CTCAE grade >0 skin ulceration) were excluded from the analysis, as this would not be attributable to the effect of radiation therapy.

Variable selection, imputation, and preprocessing

The raw REQUIRE data set (n = 2069) contained m = 204 variables (features) relating to patient baseline characteristics, comorbidities, cotreatments, and radiation therapy. Variables were initially checked for plausibility using domain expertise by physicians and radiation

therapy physicists, and m = 136 variables remained. Boundary value analysis and equivalence class partitioning techniques²⁹ were used for correcting or removing corrupt or inaccurate records from the data set. After variable-dropping (m = 13 with >37% missing values at random compared with observed values in the remaining variables) and case-wise deletion (n = 11 with missing class endpoint observations),³⁰ the final data set for modeling had m = 123 variables including the endpoint variable and n = 2058 patient records.

The final data set was randomly shuffled and split 50:50 into training and test sets with class stratification, yielding imbalanced training (ITD, n = 1029) and validation (VD, n = 1029) data sets. Each data set was imputed

independently with the ML decision-tree-based missing-value imputation technique³¹ to enhance best expectations of missing values. By carrying out the imputation of the data sets separately, the ITD and VD remained completely independent and perfectly isolated. Information levels were monitored in each data set pre- and post-imputation with information gain attribute evaluation³² (see Fig. E4). The information gain of a feature is defined as the expected reduction of entropy (uncertainty within the data set) when partitioning the data; in other words, by how much the prediction of the endpoint/class would improve if the data were split using just that feature. The more plausible the pattern of information gain among data sets, the less bias is introduced in modeling. The evaluation of information worth is affected by the number of records and the 50:50 training-test split allowed for a fair information bias comparison between training and validation data sets.

The final set of $m = 123$ features consisted of 106 raw variables. Sixteen additional features were constructed to account for the vast number of possible combinations of chemotherapeutic agents received by some patients before radiation therapy. The adjuvant chemotherapy regimens were binarized based on their generic drug names (not shown). To adjust for different radiation therapy regimens, dose was calculated as the biologically effective dose (BED). BED is the product of the number of fractions (n), dose per fraction (d), and a factor determined by the dose and α/β ratio (10 Gy) for desquamation (acute toxicity):

$$BED = n d \left(1 + \frac{d}{\alpha/\beta} \right)$$

The endpoint definition (acute desquamation = *Desq*) was used to label the patients to create a binary class variable. All numeric features ($m = 63$) were normalised with z score standardization.³³

Resampling

Although it is a clinically significant side effect from breast radiation therapy, only a small proportion of patients suffer from acute desquamation, an issue known as “class imbalance.”^{9,18} Both ITD and VD in this study were equally imbalanced ($Desq^+ = 96$, $Desq^- = 933$). To address the issue of class imbalance, 3 resampling techniques were applied to the training data to obtain equal proportions of records in each class: random under-sampling (RUS) ($n = 192$, $Desq^+ = 96$, $Desq^- = 96$),³⁴ random over-sampling (ROS) ($n = 1866$, $Desq^+ = 933$, $Desq^- = 933$),³⁵ and the synthetic minority oversampling technique (SMOTE) ($n = 1866$, $Desq^+ = 933$, $Desq^- = 933$).³⁶ The effect of resampling techniques on the training data set was monitored with a multidimensional adaptive

projection analysis into a 3-dimensional point cloud. Adaptive projection analysis³⁷ is a multidimensional tool to visualise the classes that can be separated, any outliers or sources of error in the classification algorithms, and the existence of clusters in the data (see Fig. E5).

Modeling

Eight different ML algorithms were used to build binary classification models to predict acute desquamation in patients undergoing breast-conserving surgery and adjuvant whole breast EBRT. They were trained in the ITD (imbalanced modeling, Fig. 1) as well as in the 3 resampled data sets (RUS, ROS, SMOTE; data-bias modeling, Fig. 1) with 10-fold cross-validation to reduce overfitting,³⁸ and then each was tested in the VD (see Fig. 1). The ML algorithms were discretized naïve Bayes (NB), logistic regression with ridge estimator,³⁹ artificial neural networks with a multilayer perceptron architecture,⁴⁰ support vector machine with polynomial kernel and logistic calibrator, K-nearest neighbour⁴¹ with $K = 1,3,5,7,9$, decision trees (C4.5),³² logistic model tree (LMT),⁴² and RF.⁴³

Model performance was assessed using the area under the curve (AUC). The models with the highest AUC in the VD were taken forward for cost-sensitive learning optimisation. Cost-sensitive classification addresses the issue of class imbalance by imposing penalties (costs) for the misclassification of the positive cases (ie, making a false negative [FN] prediction). In this study, the cost for a FN prediction was not linked to a monetary value, instead a 10-step incremental inverse class distribution cost was used.⁴⁴ The ITD has a $96:933 \cong 1:10$ ratio of examples in the positive class to examples in the negative class. This ratio is inverted to penalize FN with a 10-step incrementation at an initial cost $\chi : 1$ of 10:1 increasing to 100:1. The cost is applied in the form of Charles Elkan’s explicit cost matrix notation⁴⁵:

$$\begin{aligned} \text{Cost Matrix combinations} & \begin{bmatrix} \text{FP}(1) & \text{TN}(0) \\ \text{TP}(0) & \text{FN}(\chi) \end{bmatrix} \\ & = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 10 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 20 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 30 \end{bmatrix}, \dots, \begin{bmatrix} 1 & 0 \\ 0 & 100 \end{bmatrix} \right\} \end{aligned}$$

AUC, sensitivity (true positive rate [TPR]), and specificity (true negative rate [TNR]) were used to compare and interpret the final models’ performance including those developed in the resampled data sets (see bottom half of Fig. 1). Final model selection was based on performance in the VD in terms of AUC and the clinicians’ trade-off maximizing both TPR and TNR. The selected model was further optimized using the mean decrease impurity entropy filter to select fewer features and simplify the “hero” model.⁴⁶ All ML algorithms were implemented in the Waikato environment for knowledge

Table 1 Summary study characteristics of eligible patients from the REQUITE patient cohort

REQUITE breast cancer cohort	
Eligible patients	2059
Location	Western Europe, United States
Study design	Prospective cohort
Recruitment year (range)	2014-2016
Treatment year (range)	2014-2016
Toxicity assessment scale	CTCAE v4.0
Toxicity assessment time points	Start-of-RT End-of-RT
Age (median, range)	58 (23-90)
Whole breast dose (Gy, median, range)	50 (28.5-56)
Whole breast fractions (median, range)	25 (5-31)
Hypofractionated regimen (proportion of patients)	47.9%
IMRT, simple field-in-field	39.7%
IMRT, complex modulated	9.8%
RT to axilla	11.9%
RT to supraclavicular fossa	12.8%
Boost	67.8%
BMI ≥ 25	54.0%
Smoker (current or previous)	42.7%
Chemotherapy	31.0%
Diabetes	6.1%
Hypertension	28.0%
Cardiovascular disease	6.9%
Toxicity (end of treatment)	
Ulceration	
Grade 0	1868 (91.2%)
Grade ≥ 1	181 (8.8%)
Dermatitis	
Grade 0	257 (12.5%)
Grade 1	1288 (62.6%)

*(continued on next page)***Table 1 (Continued)**

REQUITE breast cancer cohort	
Grade 2	462 (22.4%)
Grade 3	28 (1.4%)
Acute desquamation	
Ulceration $\geq G1$ or dermatitis $\geq G3$	192 (9.3%)
Abbreviations: BMI = body mass index; CTCAE = Common Terminology Criteria for Adverse Events; IMRT = intensity modulated radiation therapy; RT = radiation therapy.	

analysis 3.8.3 (with the default models' parameters settings),⁴⁷ with the C4.5 decision tree using the J48⁴⁸ implementation, K-nearest neighbor using the IBK (instance-bases learning with parameter k) implementation, and support vector machine using the SMO (sequential minimal optimization)⁴⁹ implementation.

Results

Table 1 shows the main patient and treatment demographics for eligible patients. Median patient age was 58 years (range, 23-80 years). Patients were treated with a median breast dose of 50 Gy (28.5-56 Gy) in 25 fractions (range, 5-31) according to local protocol. In terms of important demographic features, 54.0% of patients had a body mass index ≥ 25 , 42.7% were previous or current smokers, 31.0% had also undergone chemotherapy, 6.1% had diabetes, and 28.0% and 6.9% had hypertension and cardiovascular disease, respectively. About half of the patients were treated with intensity modulated radiation therapy, with a lower proportion in France and none at Italian or U.S. centers. The majority of patients received a tumor-bed boost (64%), ranging from less than 20% at the French, Italian, and Spanish centers to over 80% at the Belgian centers, given either simultaneously ($n = 257$) or sequentially ($n = 1138$). Patients with invasive breast cancer in Belgium and the United Kingdom were treated using the Standardisation of Breast Radiation therapy Trial B (START-B) hypofractionated regimen (40 Gy in 15 fractions). In terms of regional nodal irradiation, axillary nodes were treated in 11.9% and the supraclavicular fossa was treated in 12.8% of patients, respectively. Detailed characteristics of the REQUITE patient cohorts have previously been described elsewhere.²⁷

Table 2 lists the performance of 12 ML classifiers using 8 different algorithms in terms of each model's AUC, TPR (sensitivity), and TNR (specificity). Accuracy was biased strongly toward the majority negative class (*Desq⁻*) as shown by consistently high TNRs and low TPRs across all models, likely due to class imbalance in

Table 2 Model performance with imputed imbalanced training data set DMI(ITD) and validation data set DMI(VD)

Classifier	Training in ITD (n = 1029)			Validation in VD (n = 1029)			Rank
	Specificity (TNR)	Sensitivity (TPR)	AUC	Specificity (TNR)	Sensitivity (TPR)	AUC	
(K = 1) NN	0.908	0.167	0.548	0.923	0.292	0.607	9
(K = 3) NN	0.975	0.094	0.601	0.979	0.125	0.627	8
(K = 5) NN	0.985	0.042	0.624	0.989	0.063	0.651	6
(K = 7) NN	0.996	0.031	0.648	0.998	0.052	0.644	7
(K = 9) NN	0.999	0.031	0.660	0.999	0.042	0.665	5
ANN	0.945	0.198	0.694	0.953	0.177	0.676	4
C4.5	0.985	0.083	0.575	0.979	0.125	0.496	12
LMT	0.996	0.010	0.578	0.995	0.042	0.746	1
LR	0.910	0.188	0.567	0.959	0.135	0.596	10
NB	0.810	0.438	0.697	0.833	0.500	0.737	3
SVM	0.966	0.156	0.561	0.976	0.146	0.561	11
RF	0.998	0.021	0.725	0.999	0.010	0.742	2

Abbreviations: ANN = artificial neural network; AUC = area under the curve; C4.5 = decision tree; DMI = decision-tree based missing value imputation; ITD = imbalanced training; KNN = K-nearest neighbor; LMT = logistic model tree; LR = logistic regression; NB = naïve Bayes; RF = random forest; SVM = support vector machine; TNR = true negative rate; TPR = true positive rate; VD = validation.

the ITD. The 3 best-performing classifiers in terms of AUC in the VD were LMT, RF, and NB with 0.75, 0.74, and 0.74, respectively. These were selected for cost-sensitive learning optimisation with incremental penalty rising in 10 steps from 10 to 100. All 12 ML classifiers listed in Table 1 were also applied to the 3 resampled training data sets (RUS, ROS, and SMOTE).

Figure 2 shows radar charts plotting sensitivity (TPR) and specificity (TNR) in the VD for a total of 66 models in the resampled training data (Fig. 2A-C) and after applying cost-sensitive penalties to the 3 best performing classifiers (Fig. 2D-F). Resampling improved sensitivity across all classifiers, with RUS (Fig. 2A) achieving the least variance between specificity and sensitivity on validation. For the cost-sensitive classifiers, the incremental penalty skewed the correct classification toward the true positives and models with higher penalty showed higher sensitivity (TPR). NB model sensitivity ranged from 0.50 in the unpenalized model to 0.77 for a penalty of 100. The largest improvement in sensitivity was achieved for the RF classifier, ranging from 0.01 for the unpenalized model to 0.79 at penalty of 100. LMT sensitivity improved from 0.04 without a penalty to 0.65 with a penalty of 100. Specificity (TNR) decreased for all 3 cost-sensitive classifiers because the number of predicted false-positives increased with each incremental penalty.

Model selection and feature filtering

Figure 3 shows 2 conditions (finishing lines) selected to maximize accuracy, that is, maximizing both TPR

and TNR (the clinicians' trade-off), with lower and upper threshold values of 0.63 and 0.70, respectively, which were crossed by 5 classifiers: cost-sensitive RF (CS-RF) with an FN:false positive (FP) 90:1 penalty (TNR = 0.65, TPR = 0.77, AUC = 0.76); RUS-RF (TNR = 0.65, TPR = 0.74, AUC = 0.74); cost-sensitive NB with an FN:FP 60:1 penalty (TNR = 0.64, TPR = 0.70, AUC = 0.72); CS-RF with an FN:FP 80:1 penalty (TNR = 0.70, TPR = 0.65, AUC = 0.75); and cost-sensitive NB with an FN:FP 20:1 penalty (TNR = 0.70, TPR = 0.63, AUC = 0.73). As maximizing sensitivity (TPR) was most important, the best performing "hero" model was the CS-RF classifier with an FN:FP penalty of 90:1. This model exceeded others for sensitivity and AUC performance while maintaining moderate specificity.

The hero CS-RF (90:1) model had $m = 122$ features. Eight features were estimated to have zero importance including features about presence/absence of systemic lupus erythematosus and other collagen vascular diseases and use of pertuzumab, eribulin, and amiodarone therapy. In a final step, these features were removed and the model was rebuilt and revalidated in the VD. Table 3 lists the features included in the final hero CS-RF classifier by order of importance. In descending order, the top 10 features were duration of other lipid-lowering drug use, type of surgery (wide local excision vs quadrantectomy), use of radiation therapy bolus, use of chemotherapy, use of boost, radiation therapy photon dose (MV), use of epirubicin therapy, hypertension, bra band size, and side of radiation therapy. Performance of the optimized hero final model in the VD improved slightly in terms of

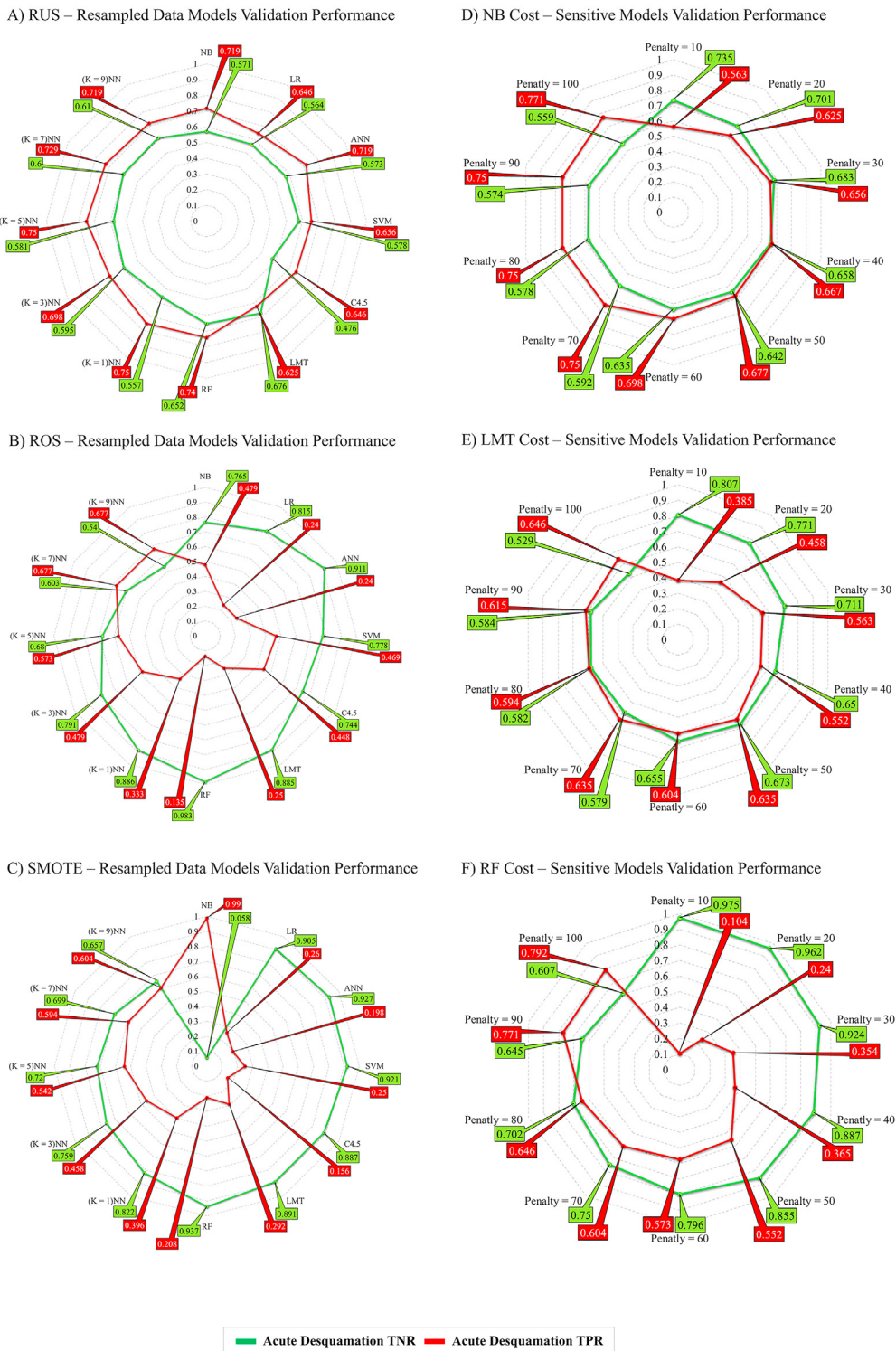


Fig. 2 Radar charts plotting sensitivity (TPR) and specificity (TNR) in the validation data set for all ML models developed with the RUS, ROS, and SMOTE resampled training data and after applying cost-sensitive learning to the 3 best-performing ML models (RF, NB, and LMT). Abbreviations: LMT = logistic model tree; ML = machine learning; NB = naïve Bayes; RF = random forest; ROS = random over-sampling; RUS = random under-sampling; SMOTE = synthetic minority over-sampling technique; TNR = true negative rate; TPR = true positive rate.

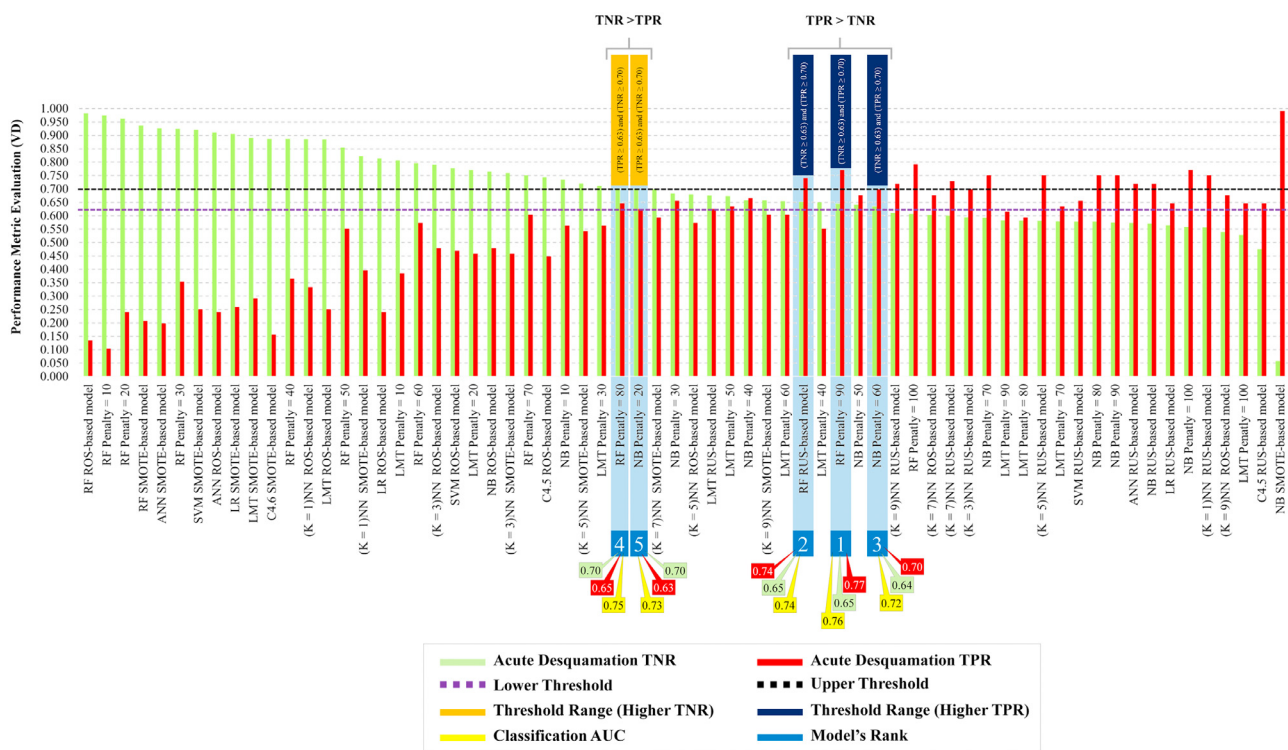


Fig. 3 Trade-off threshold lines are shown for sensitivity (TPR) and specificity (TNR) at 0.63 and 0.70, respectively. Five models cross both threshold lines and their TPR, TNR, and AUC values are shown at the bottom. Two out of 5 models have a higher TNR than TPR and 3 out of the 5 models have a higher TPR than TNR. The “hero” model (no. 1) was the cost-sensitive random forest algorithm with a penalty of 90:1. *Abbreviations:* AUC = area under the curve; TNR = true negative rate; TPR = true positive rate.

specificity (TNR = 0.66) and AUC (0.77) while sensitivity (TPR) remained unchanged.

Discussion

A recently published logistic regression model for acute breast desquamation after adjuvant external beam breast radiation therapy developed in 3 combined external breast radiation therapy cohorts failed to validate in the multicenter REQUITE cohort.²⁴ The aim of this study was to use ML algorithms to develop and optimise a prediction model for acute desquamation in the REQUITE breast cancer cohort. ML techniques have previously been used to predict acute skin toxicity during breast radiation therapy.^{22,23} We elected to predict the occurrence of acute desquamation rather than dermatitis (skin erythema) because it can cause clinically significant patient morbidity and can worsen the cosmetic outcome after breast surgery. This accounts for the lower proportion of cases with skin toxicity reported in our study versus the study by Saednia et al²² (0.09 vs 0.38), although the proportion of cases was similar to those with moist desquamation in the abstract published by Reddy et al.²³

Predicting cases of clinically significant radiation toxicity such as acute desquamation remains challenging for both parametric statistical and ML models due to the issue of class imbalance leading to high FN rates, that is, poor sensitivity. In this study, a combination of resampling techniques and cost-sensitive learning was used to try and improve predictive performance. RUS and cost-sensitive optimisation contributed the most to optimal performance across the different ML algorithms. Of 66 models tested, 5 fulfilled prespecified criteria for maximizing both TPR and TNR. On the basis of highest TPR, the hero model was the CS-RF, with an FN:FP misclassification penalty of 90:1. Given that our modeling used somewhat fewer features and had a multicenter patient sample with diverse radiation treatment regimens, it is reassuring that its AUC of 0.77 in the VD is similar to the range of AUCs reported in the abstract by Reddy et al.²³

Our initial models for acute desquamation included 122 features. Information gain (IG) represents the amount of information gained about a random variable or signal from observing another random variable. After the randomized and stratified training/validation data split, only a few variables in the VD had a different IG to discriminate between the positive and negative cases compared with the ITD. Zero IG does not negate the feature’s worth as this

Table 3 Features in the “hero” optimized cost-sensitive RF classifier ranked by importance

Model's feature	MDI	Model's feature	MDI
other_lipid_lowering_drugs_duration_yrs	0.52	alcohol_current_consumption	0.2
surgery_type	0.41	smoking_time_since_quitting_yrs	0.2
radio_bolus	0.4	radio_imrt	0.19
chemotherapy	0.36	radio_photon_boostdose_Gy	0.19
boost	0.35	other_antihypertensive_drug	0.19
radio_photon_dose_MV	0.34	household_members	0.19
epirubicin_chemo_drug	0.34	radio_breast_fractions_dose_per_fraction_Gy	0.19
blood_pressure	0.33	radio_elec_boost_field_y_cm	0.19
Bra_band_size	0.3	radio_photon_2nd	0.19
radio_treated_breast	0.3	bra_cup_size	0.19
tumour_size_mm	0.29	radio_breast_fractions	0.19
paclitaxel_chemo_drug	0.29	n_stage	0.18
grade_invasive	0.28	hypertension_duration_yrs	0.18
breast_separation	0.28	radio_supraclavicular_fossa	0.18
smoking	0.27	education_profession	0.18
radio_elec_energy_MeV	0.27	radio_axillary_levels	0.18
BED_boost	0.27	hypertension	0.18
docetaxel_chemo_drug	0.27	radio_photon_boost_fractions_per_week	0.17
BED_Total	0.27	smoker	0.17
radio_elec_boost_dose_Gy	0.27	depression	0.17
On_tamoxifen	0.26	menopausal_status	0.17
radio_heart_mean_dose_Gy	0.26	radio_boost_diameter_cm	0.16
t_stage	0.26	5-fluorouracil (5-FU)_chemo_drug	0.16
radio_hot_spots_107	0.25	radio_photon_boost_dose_per_fraction_Gy	0.16
BED_Breast	0.25	antidepressant_duration_yrs	0.16
tobacco_products_per_day	0.25	radio_breast_fractions_per_week	0.15
age_at_radiotherapy_start_yrs	0.25	radio_boost_type	0.15
radio_breast_ct_volume_cm3	0.25	Carboplatin_chemo_drug	0.15
hormone_replacement_therapy	0.24	radio_boost_sequence	0.15
radio_photon_boost_volume_cm3	0.24	radio_photon_boost_fractions	0.15
antidepressant	0.24	household_income	0.15
height_cm	0.24	methotrexate_chemo_drug	0.15
radio_photon_2nd_energy_MV	0.24	other_lipid_lowering_drugs	0.14
radio_ipsilateral_lung_mean_Gy	0.24	radio_photon_energy_MV or kV	0.14
alcohol_previous_consumption	0.24	ace_inhibitor	0.13
radio_photon_2nd_dose_fractions_per_week	0.23	analgesics_duration_yrs	0.13
radio_skin_max_dose_Gy	0.23	radio_photon_2nd_dose_per_fraction_Gy	0.13
histology	0.23	antidiabetic_duration_yrs	0.13
monopause_age_yrs	0.23	depression_duration_yrs	0.13
other_antihypertensive_drug_duration_yrs	0.23	on_statin_duration_yrs	0.12
weight_at_cancer_diagnosis_kg	0.23	antidiabetic	0.12

(continued on next page)

Table 3 (Continued)

Model's feature	MDI	Model's feature	MDI
tobacco_product	0.23	diabetes	0.11
cyclophosphamide_chemo_drug	0.22	ace_inhibitor_duration_yrs	0.11
combined_chemo_drugs	0.22	on_statin	0.11
boost_frac	0.22	doxorubicin_chemo_drug	0.11
analgesics	0.22	history_of_heart_disease	0.09
breast_cancer_family_history_1st_degree	0.22	radio_axillary_other	0.09
smoking_duration_yrs	0.21	ethnicity	0.09
radio_photon_boostdose_precise_Gy	0.21	radio_interrupted	0.08
radio_elec_boost_field_x_cm	0.21	pegfilgrastim_chemo_drug	0.07
radio_photon_2nd_fractions	0.21	history_of_heart_disease_duration_yrs	0.06
radio_boost_fractions	0.21	radiotherapy_toxicity_family_history	0.06
alcohol_intake	0.21	diabetes_duration_yrs	0.05
radio_type_imrt	0.21	radio_interrupted_days	0.05
radio_treatment_pos	0.21	trastuzumab_chemo_drug	0.04
radio_breast_dose_Gy	0.2	other_collagen_vascular_disease	0.03
rheumatoid_arthritis_duration_yrs	0.2	rheumatoid_arthritis	0.02

Abbreviations: BED = biologically effective dose; IMRT = intensity modulated radiation therapy; MDI = mean decrease impurity; MeV = mega electron volt; MV = mega volt; RF = random forest.
Feature importance is calculated as the decrease in node impurity weighted by the probability of reaching that node. The node probability can be calculated by the number of samples that reach the node, divided by the total number of samples. The higher the value, the more important the feature.

depends on the ML algorithm used, and any given feature could climb up the ranking in terms of IG if additional observations were added to the same data set. Hence, we included all 122 features in the modeling process. The 10 most important features in the final hero model included some that might be expected to predict breast radiation toxicity, such as use of radiation therapy bolus, chemotherapy, boost, radiation therapy dose, and bra size. Interestingly, the most important feature (use of lipid-lowering drugs) is not usually included in parametric statistical models for radiation toxicity, although HMG-CoA reductase inhibitors (statins) have previously been proposed as radioprotective agents.⁵⁰ Yet unlike traditional statistical probability modeling, feature importance should only be interpreted within the context of the ML prediction model but not outside.

Study limitations

Despite the rigorous error detection in the data preprocessing phase, we cannot exclude errors occurring due to manual recording during data collection. According to the REQUITE study protocol, patients were assessed at the start and end of treatment and annually thereafter. This may have missed cases of acute desquamation as acute radiation toxicity is known to peak up to 2 weeks

after the end of treatment. Although we incorporated differences in radiation therapy techniques by including all available recorded treatment parameters in the analysis, this may not fully account for variability in treatment plans between participating centers or treating physicians. Similarly, variable transformation or feature engineering (eg, calculating the BED, binarization of chemotherapy drugs) could have led to the creation of a new feature that is less powerful and suppresses important information inferred by its raw components. In modeling the radiation therapy dose variable, alternatives such as a categorical variable divided by type of radiation therapy regimen could have been used (eg, hypo- vs standard fractionation). Variable aggregation could have led to model overfitting due to misleading combined features and may show false significance or insignificance in the analysis.⁵¹ Although the resampling techniques used in this study have advantages in their simplicity and transportability, other remedies to address imbalanced data, such as ensemble learning (which is implemented at the algorithmic level), could be used to improve model performance.⁵² Cost-sensitive learning was selected to penalize false negatives. However, its application depends on the clinical situation. For example, if a model was designed to allocate patients to a toxicity-lowering radiation therapy regimen that might affect tumor control, then FPs may need to have a higher cost than FNs. This study used the

impurity-based ranking mean decrease impurity filter to simplify the final model with a known performance, but it is important to keep in mind that feature selection based on impurity reduction is generally biased toward preferring variables with more categories.⁵³

Conclusion

Application of ML algorithms with resampling and cost-sensitive learning resulted in valid prediction models for acute desquamation after whole breast EBRT using clinical and treatment features. After optimisation, the best model was able to classify patients with acceptable performance in the validation cohort (AUC = 0.77). Before they can be used in clinical practice, further optimization of ML prediction models, including genomic markers, is required, and the models should be validated in external cohorts. This approach could help identify breast cancer patients at increased risk of toxicity to inform discussions about risks and benefits and allow treatment plans to be personalized with the aim of minimizing toxicity or offering the patient increased supportive management during treatment.

Acknowledgments

We sincerely thank all patients who participated in the REQUITE study and all REQUITE staff involved at the following hospitals:

Belgium: Prof. Wilfried de Neve, Dr. Christel Monten, Dr. Pieter Deseyne, Giselle Post and Renée Bultijnck at Ghent University Hospital; Gilles Defraene, Rita Aerts, Soumia Arredouani, Maarten Lambrecht at KU Leuven.

France: ICM Montpellier, CHU Nîmes (Department of Radiation Oncology, CHU Nîmes, Nîmes, France).

Germany: Praxis für Strahlentherapie an der Stadtklinik Baden-Baden (Dr. Thomas Blaschke); Klinikum Darmstadt GmbH (PD Dr. Christian Weiß); Zentrum für Strahlentherapie Freiburg (Dr. Petra Stegmaier); ViDia Christliche Kliniken Karlsruhe (Prof. Dr. Johannes Claßen); Klinikum der Stadt Ludwigshafen gGmbH (PD Dr. Thomas Schnabel); Universitätsklinikum Mannheim; Strahlentherapie Speyer (Dr. Jörg Schäfer). The researchers at DKFZ also thank Anusha Müller, Irmgard Helmbold and Dr Sabine Behrens.

Italy: Fondazione IRCCS Istituto Nazionale dei Tumori, Milano.

Spain: Santiago: Complejo Hospitalario Universitario de Santiago. Barcelona: The authors acknowledge the Cellex Foundation for providing research facilities and equipment and thank the CERCA Programme/ Generalitat de Catalunya for institutional support.

UK: Leicester: University Hospitals Leicester (Drs. Donna Appleton, Frances Kenny, Jaroslaw Krupa,

Monika Kaushik, Kelly V Lambert, Simon M Pilgrim, Sheila Shokuhi, Kalliope Valassiadou, Luis Aznar-Garcia, Kerstie Johnson, Kiran Kancharla, and Kufre Sampson); support from the HOPE Clinical Trials Unit and Leicester Experimental Cancer Medicine Centre (ECMC), Theresa Beaver, Kaitlin Duckworth, and Sara Barrows.

USA: Mount Sinai Hospital, New York.

Supplementary materials

Supplementary material associated with this article can be found in the online version at [doi:10.1016/j.adro.2021.100890](https://doi.org/10.1016/j.adro.2021.100890).

References

1. *Early and Locally Advanced Breast Cancer: Diagnosis and Management*. London: National Institute for Health and Clinical Excellence; 2018.
2. 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 randomised trials. *The Lancet*. 2011;378:1707–1716.
3. Office for National Statistics. *Cancer Survival in England: Patients Diagnosed Between 2010 and 2014 and Followed Up to 2015*. London: Office for National Statistics; 2016.
4. National Cancer Institute. *Priorities for Research on Cancer Survivorship*. London: Department of Health; 2010.
5. Emami B, Lyman J, Brown A, et al. Tolerance of normal tissue to therapeutic irradiation. *Int J Radiat Oncol Biol Phys*. 1991;21:109–122.
6. Bentzen SM, Overgaard J. Patient-to-patient variability in the expression of radiation-induced normal tissue injury. *Sem Radiat Oncol*. 1994;4:68–80.
7. Knobf MT, Sun Y. A longitudinal study of symptoms and self-care activities in women treated with primary radiotherapy for breast cancer. *Cancer Nurs*. 2005;28:210–218.
8. Rochlin DH, Jeong AR, Goldberg L, et al. Postmastectomy radiation therapy and immediate autologous breast reconstruction: Integrating perspectives from surgical oncology, radiation oncology, and plastic and reconstructive surgery. *J Surg Oncol*. 2015;111:251–257.
9. Twardella D, Popanda O, Helmbold I, et al. Personal characteristics, therapy modalities and individual DNA repair capacity as predictive factors of acute skin toxicity in an unselected cohort of breast cancer patients receiving radiotherapy. *Radiother Oncol*. 2003;69:145–153.
10. Back M. Impact of radiation therapy on acute toxicity in breast conservation therapy for early breast cancer. *Clin Oncol*. 2004;16:12–16.
11. Deantonio L, Gambaro G, Beldi D, et al. Hypofractionated radiotherapy after conservative surgery for breast cancer: Analysis of acute and late toxicity. *Radiat Oncol*. 2010;5:112.
12. Barnett GC, Wilkinson JS, Moody AM, et al. The Cambridge Breast Intensity-Modulated Radiotherapy trial: Patient- and treatment-related factors that influence late toxicity. *Clin Oncol*. 2011;23:662–673.
13. Kraus-Tiefenbacher U, Sfantizky A, Welzel G, et al. Factors of influence on acute skin toxicity of breast cancer patients treated with standard three-dimensional conformal radiotherapy (3D-CRT) after breast conserving surgery (BCS). *Radiat Oncol*. 2012;7:217.
14. Terrazzino S, Mattina PL, Masini L, et al. Common variants of eNOS and XRCC1 genes may predict acute skin toxicity in breast

- cancer patients receiving radiotherapy after breast conserving surgery. *Radiother Oncol*. 2012;103:199–205.
15. Sharp L, Johansson H, Hatschek T, Bergenmar M. Smoking as an independent risk factor for severe skin reactions due to adjuvant radiotherapy for breast cancer. *The Breast*. 2013;22:634–638.
 16. Tortorelli GDML, Barbarino R, Cicchetti S, et al. Standard or hypofractionated radiotherapy in the postoperative treatment of breast cancer: A retrospective analysis of acute skin toxicity and dose inhomogeneities. *BMC Cancer*. 2013;13:230.
 17. Ciammella P, Podgornii A, Galeandro M, et al. Toxicity and cosmetic outcome of hypofractionated whole-breast radiotherapy: Predictive clinical and dosimetric factors. *Radiat Oncol*. 2014;9:97.
 18. De Langhe S, Mulliez T, Veldeman L, et al. Factors modifying the risk for developing acute skin toxicity after whole-breast intensity modulated radiotherapy. *BMC Cancer*. 2014;14:711.
 19. Mbah C, Thierens H, Thas O, et al. Pitfalls in prediction modeling for normal tissue toxicity in radiation therapy: An illustration with the individual radiation sensitivity and mammary carcinoma risk factor investigation cohorts. *Int J Radiat Oncol Biol Phys*. 2016;95:1466–1476.
 20. Dean J, Wong K, Gay H, et al. Incorporating spatial dose metrics in machine learning-based normal tissue complication probability (NTCP) models of severe acute dysphagia resulting from head and neck radiotherapy. *Clin Transl Radiat Oncol*. 2018;8:27–39.
 21. Lee S, Kerns S, Ostrer H, Rosenstein B, Deasy JO, Oh JH. Machine learning on a genome-wide association study to predict late genitourinary toxicity after prostate radiation therapy. *Int J Radiat Oncol Biol Phys*. 2018;101:128–135.
 22. Saednia K, Tabbarah S, Lagree A, et al. Quantitative thermal imaging biomarkers to detect acute skin toxicity from breast radiation therapy using supervised machine learning. *Int J Radiat Oncol Biol Phys*. 2020;106:1071–1083.
 23. Reddy J, Lindsay WD, Berling CG, Ahern CA, Smith BD. Applying a machine learning approach to predict acute toxicities during radiation for breast cancer patients. *Int J Radiat Oncol Biol Phys*. 2018;102:S59.
 24. Rattay T, Seibold P, Aguado-Barrera ME, et al. External validation of a predictive model for acute skin radiation toxicity in the REQUITE breast cohort. *Front Oncol*. 2020;10:575909.
 25. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): The TRIPOD statement. *Br J Surg*. 2015;102:148–158.
 26. West C, Azria D, Chang-Claude J, et al. The REQUITE project: validating predictive models and biomarkers of radiotherapy toxicity to reduce side-effects and improve quality of life in cancer survivors. *Clin Oncol (R Coll Radiol)*. 2014;26:739–742.
 27. Seibold P, Webb A, Aguado-Barrera ME, Azria D, Bourcier C, Brengues M, et al. REQUITE. A prospective multicentre cohort study of patients undergoing radiotherapy for breast, lung or prostate cancer. *Radiotherapy and oncology: journal of the European Society for Therapeutic Radiology and Oncology*. 2019;138:59–67.
 28. National Cancer Institute. *Common Terminology Criteria for Adverse Events V4.0*. National Cancer Institute: NIH; 2009.
 29. Arnicane V. Complexity of equivalence class and boundary value testing methods. *Int J Comput Sci Inform Techn*. 2009;751:80–101.
 30. Liu WZ, White AP, Thompson SG, Bramer MA. Techniques for dealing with missing values in classification. In: Liu X, Cohen P, Berthold M, eds. *International Symposium on Intelligent Data Analysis*. Springer; 1997:527–536.
 31. Rahman G, Islam Z. A decision tree-based missing value imputation technique for data pre-processing. *Proc Ninth Australasian Data Min Conf*. 2011;121:41–50.
 32. Quinlan JR. Improved use of continuous attributes in C4. 5. *J Artif Intell Res*. 1996;4:77–90.
 33. Raschka S. About feature scaling and normalization and the effect of standardization for machine learning algorithms. *Political Leg Anthropology Rev*. 2014;30:67–89.
 34. Chawla NV. *Data mining for imbalanced datasets: An overview*. Data Mining And Knowledge Discovery Handbook. Boston, MA: Springer; 2009:875–886.
 35. Branco P, Torgo L, Ribeiro RP. A survey of predictive modeling on imbalanced domains. *ACM Comput Surv*. 2016;49:1–50.
 36. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic minority over-sampling technique. *J Artif Intell Res*. 2002;16:321–357.
 37. Faith J, Brockway M. Targeted projection pursuit tool for gene expression visualisation. *J Integrat Bioinform*. 2006;3:264–273.
 38. Fushiki T. Estimation of prediction error by using K-fold cross-validation. *Stat Comput*. 2011;21:137–146.
 39. Bielza C, Larrañaga P. Discrete Bayesian network classifiers: A survey. *ACM Comput Surv*. 2014;47:1–43.
 40. Ruck DW, Rogers SK, Kabrisky M, Oxley ME, Suter BW. The multi-layer perceptron as an approximation to a Bayes optimal discriminant function. *IEEE Trans Neural Net*. 1990;1:296–298.
 41. Cunningham P, Delany SJ. K-nearest neighbour classifiers. arXiv preprint arXiv:200404523. 2020.
 42. Landwehr N, Hall M, Frank E. Logistic model trees. *Mach Learn*. 2005;59:161–205.
 43. Ho TK. Random decision forests. In: *Proc Third Int Conf Doc Analysis and Recog*. 1995:278–282.
 44. Brownlee J. *Imbalanced Classification with Python: Better Metrics, Balance Skewed Classes, Cost-Sensitive Learning*. Machine Learning Mastery; 2020.
 45. Elkan C. The foundations of cost-sensitive learning. *Int Joint Conf Artif Intell*. 2001:973–978.
 46. Louppe G, Wehenkel L, Sutura A, Geurts P. Understanding variable importances in forests of randomized trees. *Adv Neur Inform Process Syst*. 2013;26:431–439.
 47. Garner SR. Weka: The waikato environment for knowledge analysis. In: *Proc New Zeal Comput Sci Res Student Conf*. 1995:57–64.
 48. Drazin S, Montag M. Decision tree analysis using weka. *Machine Learning-Project II*. 2012:1–3.
 49. Keerthi SS, Shevade SK, Bhattacharyya C, Murthy KRK. Improvements to Platt's SMO algorithm for SVM classifier design. *Neur Comput*. 2001;13:637–649.
 50. Fritz G, Henninger C, Huelsenbeck J. Potential use of HMG-CoA reductase inhibitors (statins) as radioprotective agents. *Br Med Bull*. 2011;97:17–26.
 51. Pyle D. *Data Preparation for Data Mining*. Burlington, MA: Morgan Kaufmann; 1999.
 52. More A. Survey of resampling techniques for improving classification performance in unbalanced datasets. arXiv preprint arXiv:160806048. 2016.
 53. Strobl C, Boulesteix AL, Zeileis A, Hothorn T. Bias in random forest variable importance measures: Illustrations, sources and a solution. *BMC Bioinform*. 2007;8:25.