INTERNATIONAL JOURNAL OF
RADIATION ONCOLOGY • BIOLOGY • PHYSICS

www.redjournal.org

PHYSICS CONTRIBUTION

Prediction of Side Effects from Breast Radiation Therapy—Integration of Clinical and Genomic Data

Abeer Al Janapy, MSc,[a] Adam J. Webb, PhD,[a] Harkeran K. Jandu, PhD,[a] Miguel E. Aguado-Barrera, MD, PhD,[b] Ester Aguado-Flor, PhD,[c] David Azria, MD,[d] Celine Bourgier, MD, PhD,[d] Renée Bultijnck, PhD,[e] Ananya Choudhury, MA, PhD, MRCP, FRCR,[f] Dirk KM De Ruysscher, MD,[g] Maria Carmen De Santis, MD,[h] Alison M. Dunning, PhD,[i] Olivia Fuentes-Ríos, PhD,[b] Carlotta Giandini, MD,[h] Antonio Gómez-Caamaño, MD, PhD,[j] Sara Gutiérrez-Enríquez, PhD,[c] Eliana La Rocca,[k] Kerstie Johnson, MBChB, MD,[l] Christel Monten, PhD,[e] Tiziana Rancati, MSc,[m] Victoria Reyes, MD,[n] Barry S. Rosenstein, PhD,[o,p] Petra Seibold, PhD,[q] Elena Sperk, MD,[r] Hilary Stobart, BSc,[s] Begoña Taboada-Valladares, MD,[t] Ana Vega, PhD,[u] Liv Veldeman, MD, PhD,[v] Marlon R. Veldwijk, PhD,[w] Catharine M.L. West, PhD,[x] Tim C.D. Lucas, PhD,[y] Tim Rattay, MBChB, PhD,[z] and Christopher J. Talbot, PhD[a]

[a]Division of Genetics and Genome Biology, College of Life Sciences, University of Leicester, University Road, Leicester, United Kingdom; [b]Instituto de Investigación Sanitaria de Santiago de Compostela (IDIS), Santiago, Spain; [c]Hereditary Cancer Genetics Group, Vall d'Hebron Institute of Oncology (VHIO), Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain; [d]Department of Radiation Oncology, Montpellier Cancer Institute, Université Montpellier, Inserm U1194, Montpellier, France; [e]Department of Human Structure and Repair, Ghent University, Ghent, Belgium; [f]University of Manchester and The Christie NHS Foundation Trust, Wilmslow Road, Manchester, United Kingdom; [g]Department Of Radiation Oncology (Maastro), Maastricht University Medical Center+, GROW School, Maastricht, The Netherlands; [h]Radiation Oncology Unit, Fondazione IRCCS Istituto Nazionale Tumori, Via Giacomo Venezian 1, Milan, Italy; [i]Centre for Cancer Genetic Epidemiology, Department of Oncology, University of Cambridge, Cambridge, United Kingdom; [j]Instituto de Investigación Sanitaria de Santiago de Compostela, Santiago, Spain; [k]Department of Radiation Oncology, Azienda Ospedaliera Universitaria Integrata di Verona, Verona, Italy; [l]Clinical Oncology, City Hospital, Nottingham University Hospitals NHS Trust, Nottingham, United Kingdom; [m]Data Science Unit, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy; [n]Radiation Oncology Department, Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus, Barcelona, Spain; [o]Department of Radiation Oncology, Icahn School of Medicine at Mount Sinai, New York, New York; [p]Department of Genetics

Corresponding author: Christopher J. Talbot, PhD.; E-mail: cjt14@leicester.ac.uk

Author Responsible for Statistical Analysis: Abeer Al Janapy was responsible for statistical analyses.

Authors' contribution statement: C.J.T., T.R., and A.A.J. conceived and designed the study; A.A.J. wrote the original draft of the manuscript under the supervision and guidance of C.J.T., T.R., and T.C.D.L., who performed the investigation of the machine learning models; H.K.J. provided the genetic data results, and A.J.W. was responsible for clinical data curation for the REQUITE study; A.A.J. performed the formal analysis of the data. A.A.J., C.J.T., T.R. and T.C.D.L. contributed to the interpretation of the results; C.M.W. served as the lead investigator and C.J.T. as the deputy lead investigator of the REQUITE study; D.A., C.B., R.B., M.P.F.-J., S.G., C.M., V.R., T.R., D.D.R., M.C.d.S., E.S., B.T.-V., L.V., and C.W. were responsible for patient recruitment across study centers; H.S. served as the patient advocate. PS provided project administration and study management. All authors critically reviewed, commented on, and approved the final version of the manuscript.

Disclosures: none. REQUITE received funding from the European Union's Seventh Framework Programme for research, technological development, and demonstration under grant agreement No. 601826; A.A.J. was supported by the Wellcome Trust Genetic Epidemiology and Public Health Genomics Doctoral Training Partnership (grant No. 218505/Z/19/Z) and the HOPE Clinical Trials Unit and Leicester Experimental Medicine Centre; T.R. was previously funded by a National Institute for Health Research (NIHR) Doctoral Research Fellowship (DRF-2014-07-079) and an Academic Clinical Lectureship (CL 2017-11-002); T.R. and C.T. were supported by the NIHR Leicester Biomedical Research Centre and by Horizon Europe PRE-ACT project (Grant Agreement number: 101057746); C.W., A.C., and R.E. were supported by the NIHR Manchester Biomedical Research Centre, and C.W. is supported by Cancer Research UK (C1094/A18504, C147/A25254); P.S. was supported by ERA PerMed JCT2018 funding (ERAPERMED2018-244; BMBF No. 01KU1912) and BfS funding (No. 3619S42261); A.V. was supported by funding from the Spanish Instituto de Salud Carlos III (ISCIII), an initiative of the Spanish Ministry of Economy and Innovation partially supported by European Regional Development Fund (FEDER) funds (PI22/00589, PI19/01424, PI16/00046, PI13/02030, PI10/00164, INT20/00071, INT17/00133, INT16/00154, and INT15/00070); the Autonomous Government of Galicia (Consolidation and Structuring Program, IN607B); ERAPerMed JTC2018 funding (ERAPERMED2018-244); and the AECC (PRYES211091VEGA); S.G.-E. was supported by ERAPerMed JTC2018 funding (ERAPERMED2018-244 and

0360-3016/$ - see front matter © 2026 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)
https://doi.org/10.1016/j.ijrobp.2026.08.026

and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, New York; [q]German Cancer Research Center (DKFZ), Heidelberg, Germany; [r]Mannheim Cancer Center, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany; [s]Independent Cancer Patients' Voice, London, United Kingdom; [t]Department of Radiation Oncology, Complexo Hospitalario Universitario de Santiago, SERGAS, Santiago de Compostela, Spain; [u]Fundación Pública Galega de Medicina Xenómica (FPGMX), Servicio Galego de Saúde (SERGAS), Santiago de Compostela, Spain; [v]Department of Radiation Oncology, Ghent University Hospital, Ghent, Belgium; [w]Department of Radiation Oncology, Medical Faculty Mannheim Heidelberg University, Mannheim, Germany; [x]Division of Cancer Sciences, The University of Manchester, Manchester Academic Health Science Centre, Christie Hospital, Manchester, United Kingdom; [y]Department of Population Health Sciences, University Road, Leicester, United Kingdom; and [z]Leicester Cancer Research Centre, University of Leicester, Clinical Sciences Building, Leicester Royal Infirmary, Leicester, United Kingdom

Received Mar 10, 2026; Revised Jul 24, 2026; Accepted for publication Aug 1, 2026

Purpose: To compare the ability of different machine learning models to predict the risk of side effects in patients with breast cancer undergoing radiation therapy.

Methods and Materials: Data from the multicenter REQUITE cohort (n = 2067) were analyzed retrospectively. Side effects (8 endpoints) were assessed 24 months post radiation therapy. Clinical, treatment, and genetic data were available. Predictive performance of 12 machine learning models was assessed using area under the receiver operating characteristic curve (AUC-ROC) and area under the precision-recall curve, with n-repeated k-fold cross-validation (CV).

Results: Random forest achieved the best performance using clinical data alone (mean AUC-ROC = 0.84, 5 × repeated 10-fold CV). Combining clinical and single nucleotide polymorphism data yielded the highest overall performance with logistic regression (mean AUC-ROC = 0.92, 5 × repeated 10-fold CV), as the highest followed by ensemble tree classifiers (mean AUC-ROC = 0.91, 5 × repeated 10-fold CV). The best prediction was obtained for arm lymphedema, followed by breast edema and nipple retraction.

Conclusions: In this large multicenter cohort, radiation therapy side effect prediction was highly endpoint dependent, with lymphedema emerging as a particularly robust outcome. When combining clinical and single nucleotide polymorphism data, logistic regression performed comparably to more complex approaches, while offering greater transparency. External validation and cost−benefit evaluation are required, but our findings provide a comparative benchmark and support the feasibility of targeted prediction of breast radiation therapy side effects. © 2026 The Authors. Published by Elsevier Inc. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

Introduction

Long-term side effects of radiation therapy can manifest months or even years following treatment completion and may negatively affect the quality of life of breast cancer survivors.[1,2] These side effects, including radiodermatitis, fibrosis, lymphedema, breast atrophy, and telangiectasia, vary substantially between individuals, reflecting differences in patient- and treatment-related factors. The ability to identify patients at increased risk before treatment could

SLT011/18/00005) and the Government of Catalonia (2021SGR01112); T. R. was partially funded by Fondazione Italo Monzino.

Data Sharing Statement: Funding for the 5-year REQUITE project ended on September 30, 2018. REQUITE does not benefit financially from supplying data and/or samples to researchers, but does make a charge to cover its costs and support continued maintenance of the database and biobank beyond the ending of the funding period. To facilitate this continued access to researchers, the REQUITE Steering Committee approved a tiered cost recovery model for access to data and/ or samples. Contact REQUITE (requite@leicester.ac.uk) for more pricing information. The machine learning workflow, feature selection pipeline, and model evaluation scripts used in this study are available through GitHub at: https://github.com/Abeersuhail/Prediction-of-Radiotherapy-Side-effct.git.

Acknowledgments—We thank all patients who participated in the REQUITE study and all REQUITE staff involved at the following sites: 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. ICM Montpellier, CHU Nîmes (Department of Radiation Oncology, ICM and CHU Nîmes, Montpellier and Nîmes, France). Zentrum für Strahlentherapie Freiburg (Dr Petra Stegmaier); Städtisches Klinikum Karlsruhe (Dr Bernhard Neu); ViDia Christliche Kliniken Karlsruhe (Prof Johannes Claßen); Klinikum der Stadt Ludwigshafen GmbH (PD Dr Thomas Schnabel); Universitätsklinikum Mannheim: Anette Kipke, Stefanie Kolb, Anke Keller, and Christiane Zimmermann; Strahlentherapie Speyer (Dr Jörg Schäfer). The researchers at DKFZ also thank Anusha Müller, Irmgard Helmbold, Thomas Heger, and Sabine Behrens. Fondazione IRCCS Istituto Nazionale dei Tumori, Milano; Candiolo Cancer Institute − FPO, IRCCS. The Netherlands: Sylvie Canisius at Maastro Clinics, Maastricht. Complexo Hospitalario Universitario de Santiago; Barcelona: Meritxel Mollà and Victoria Reyes at Vall d'Hebron Hospital Universitari, Vall d'Hebron Barcelona Hospital Campus; VHIO acknowledges the Cellex Foundation for providing research facilities and the CERCA Programme/Generalitat de Catalunya for institutional support. University Hospitals of Leicester NHS Trust: Drs. Donna Appleton, Frances Kenny, Jaroslaw Krupa, Monika Kaushik, Kelly V Lambert, Simon Pilgrim, Sheila Shokuhi, Kalliope Valassiadou, Luis Aznar-Garcia, Kerstie Johnson, Kiran Kancherla, and Kufre Sampson. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care. Manchester: J.R. at The Christie NHS Foundation Trust; Derby Hospitals NHS Foundation Trust; Nottingham University Hospitals NHS Trust; and Salford Royal NHS Trust. USA: Mount Sinai Hospital, New York; and Memorial Sloan Kettering Cancer Center, New York.

Supplementary material associated with this article can be found in the online version at doi:10.1016/j.ijrobp.2026.08.026.

support personalized therapeutic decision-making and the mitigation of side effects. Machine learning−based prediction models offer a promising framework for integrating clinical and genetic data to enable such individualized risk assessment.[3,4] A recent systematic review and meta-analysis synthesized the existing literature on machine learning −based prediction of radiation therapy-induced side effects, reporting overall moderate predictive performance and substantial heterogeneity across studies, model types, endpoints, and validation strategies. The authors highlighted several persistent limitations in the field, including small sample sizes, retrospective study designs, limited external validation, and a lack of disease- and endpoint-specific comparative analyses.[5] Existing clinical prediction models primarily incorporate a wide range of clinical variables, including the patient's age, body mass index, smoking habits, details about surgical procedures, and treatment-related variables, including radiation dose and the precise schedule of treatment fractions.[6,7] In recent years, the role of genetic susceptibility has gained significant attention as a key factor affecting the risk of side effects. Researchers suggested that single nucleotide polymorphisms (SNPs), including those found within the pathways responsible for DNA damage repair and fibrogenesis, might serve as potential biomarkers for assessing how sensitive normal tissues are to radiation exposure.[8,9] Yet, most research focused on a limited selection of candidate genes, which limited reproducibility. While extensive initiatives like the Radiogenomics Consortium[10] have discovered new and validated existing genetic predictors using genome-wide association studies (GWAS),[11] how these SNPs can be integrated into predictive models still requires significant attention and exploration.[12]

Previous research showed that including genetic data can significantly boost the predictive accuracy for specific radiation therapy-related adverse effects.[13,14] Deep learning approaches have also been applied within the REQUITE cohort to validate genetic risk factors for late radiation therapy side effects in prostate cancer, demonstrating the feasibility of SNP-based modeling for long-term outcomes.[15] In parallel, extensive comparisons of machine learning algorithms have previously been conducted to predict acute radiation therapy side effects within the REQUITE cohort, demonstrating moderate predictive performance and the importance of methodological considerations such as class imbalance.[16] However, there has been limited systematic evaluation of multiple machine learning algorithms applied to integrated clinical and genetic data for the prediction of long-term radiation therapy side effects in large, multicenter cohorts. This issue is timely because the first clinical trial based on radiation therapy prediction models, PRE-ACT-01, has commenced recruitment,[17] with more such trials likely soon.

The primary aim of this research paper is to evaluate the performance of twelve machine learning algorithms for their ability to predict radiation therapy side effects 2 years after treatment, using a range of clinical data, SNP data, and a combination of both types of data.

Methods and Materials

Data collection and preprocessing

The clinical data were obtained from the REQUITE data set, which provides a rich source of longitudinal data, utilizing 2-year follow-up information on 2067 breast cancer cases, containing >500 features such as age, height, weight, smoking status, surgical details, and radiation therapy parameters.[18] The data set includes patients from multiple international centers, with harmonized prospective data collection; however, variability in treatment practices across centers is expected. Radiation therapy-related variables were included to capture this heterogeneity, including fractionation (≤2.2 Gy vs >2.2 Gy per fraction), use of intensity modulated radiation therapy, treatment position (prone vs supine), and boost delivery (Table E3). To account for inter-center variability, the treatment site (site ID) was included as a predictor variable in the modeling process. The outcome variables in this analysis are long-term side effects: arm lymphedema, telangiectasia in the tumor area, breast edema, breast skin hyperpigmentation, induration in the tumor bed, induration outside the tumor area, nipple retraction, and breast atrophy. These side effects were originally graded on the Common Terminology Criteria for Adverse Events version 4.0 scale: 0 = none, 1 = mild, 2 = moderate, 3 = severe, 4 = life-threatening, 5 = death. No grade 4 or 5 events were reported, and we simplified our analysis by categorizing outcomes as "no side effects" (grade 0) and "any side effects" (grade ≥1), as detailed in Table 1. This decision was driven by the low prevalence of severe effects (grade ≥2; <10% for most outcomes). This approach is commonly used in radiogenomic studies to address class imbalance and enable stable model training, although it may reduce clinical granularity.[12,19] Importantly, all baseline endpoints graded before radiation therapy were retained and considered candidate features for subsequent feature selection. Then, the data set was cleaned and formatted to ensure suitability for analysis, including handling missing values, formatting variables, normalizing data, and ensuring consistency across other follow-up timepoints.

The genetic data also came from the REQUITE cohort using GWAS results described in Jandu et al.[11] That analysis was restricted to patients of European ancestry to minimize potential bias related to population stratification and genotype imputation accuracy. Patients of non-European ancestry (n = 121) were excluded because of their limited representation within the cohort. Featured SNPs were selected based on genome-wide significance ($P< 5 \times 10^{-8}$) and suggestive significance ($P< 1 \times 10^{-5}$), ensuring that the genetic feature set in this study was derived independently of the current machine learning modeling process. Following SNP selection, genotype dosage data were extracted at the patient level and linked with phenotype data to produce a genetic data set containing the SNPs and dosage information for use in predictive modeling. To reduce redundancy

Table 1 Mapping of 2-year radiation therapy side effect grades (0-3) to binary categories

Side effects at 2 y	No toxicity n (%)	Any toxicity grade (≥1) n (%)	Grade 1 n (%)	Grade 2 n (%)	Grade 3 n (%)
Arm lymphedema	1628 (96.27)	63 (3.72)	54 (3.19)	9 (0.53)	0
Nipple retraction	1441 (86.29)	229 (13.71)	210 (12.57)	19 (1.13)	—
Breast edema	1469 (86.82)	223 (13.18)	193 (11.41)	30 (1.77)	0
Breast skin hyperpigmentation	1375 (81.31)	316 (18.68)	312 (18.45)	4 (0.24)	—
Breast atrophy	889 (52.73)	797 (47.27)	577 (34.22)	196 (11.63)	24 (1.42)
Induration outside tumor area	1452 (85.82)	240 (14.18)	215 (12.71)	22 (1.3)	3 (0.18)
Induration in tumor bed (scar tissue)	1025 (60.545)	668 (39.46)	559 (33.02)	101 (5.97)	8 (0.47)
Telangiectasia in (tumor area)	1614 (95.39)	76 (4.61)	73 (4.31)	5 (0.30)	—

(n) presents the number of patients, (%) percentage of the total population (%). Toxicities were graded according to the REQUITE protocol (0 = none, 1 = mild, 2 = moderate, 3 = severe), "No toxicity" refers to patients without any grade of the listed side effect, "Any Toxicity (≥1)" includes all patients experiencing grade ≥1 toxicity and dash (—) indicates that the category was not applicable for that side effect.

among genetic features, highly correlated SNPs ($r > 0.7$) were removed, serving as an approximation of linkage disequilibrium pruning.

Feature selection and machine learning models

To address the high dimensionality of the data set, feature selection methods were employed to enhance model interpretability and to identify the predictors most relevant to radiation therapy side effects in breast cancer patients. Feature selection steps were performed within each training fold during cross-validation (CV) to avoid data leakage and ensure an unbiased estimation of model performance.[20,21] In this study, the feature importance–based method using tree-based models[22,23] was used to identify the most informative predictors. Other techniques such as univariate selection with chi-squared,[24,25] autoencoder-based dimensionality reduction,[26] and recursive feature elimination[27] were also explored, but they did not yield a robust, relevant subset of predictors. Across these techniques, different feature set sizes were evaluated in preliminary analyses, and an optimal set of features was selected as a balance between predictive performance, interpretability, and data completeness (Fig. E2).

Twelve diverse machine learning models were implemented in Python using Jupyter Notebook and Google Colaboratory, with the scikit-learn and XGBoost machine learning packages. The models inlcluded: (1) tree-based methods—random forest, Extra Tree Classifier, Decision Tree, Adaptive Boosting (AdaBoost), Gradient Boosting, and eXtreme Gradient Boosting (XGBoost) classifiers; (2) linear and probabilistic models—logistic regression, Gaussian Naïve Bayes (GaussianNB), and Quadratic Discriminant Analysis; (3) a neural network—the multilayer perceptron (MLP) classifier as the deep learning model; (4) other models—support vector classifier (SVC), which finds an optimal hyperplane to separate classes in feature space, and K-nearest neighbors (KNN), which classifies samples based on the labels of the closest training points.

Hyperparameters were selected using a structured, model-specific approach, as detailed in Table E1. Grid search with CV was applied to selected models where hyperparameter optimization was expected to have a substantial effect on predictive performance, such as random forest, SVC and KNN. In particular, for the KNN model, the optimal number of neighbors (k) was determined by evaluating values from 1 to 20 using 10-fold stratified CV, with the mean area under the curve (AUC) as the selection criterion. This approach provided an unbiased estimate of generalization performance and guided the selection of hyperparameters that maximized discriminative ability.[21,28] For models with fewer tunable parameters or lower sensitivity (eg, GaussianNB, Quadratic Discriminant Analysis, and logistic regression), standard or literature-supported default settings were retained to balance interpretability, computational cost, and model robustness.[29,30] All models were evaluated using the same feature sets and CV framework to ensure fair comparison.

Model evaluation

Model performance was evaluated using repeated stratified k-fold CV to ensure robust estimation and reduce variance in performance metrics. Class imbalance was not explicitly addressed through resampling or weighting; instead, performance was assessed using metrics appropriate for imbalanced data. Discriminative performance was quantified using receiver operating characteristic (ROC) curves and the corresponding AUC-ROC. In addition, precision-recall (PR) curves and AUC-PR were computed, as these provide a more informative evaluation in the presence of skewed class distributions and emphasize the trade-off between precision (positive predictive value) and recall (sensitivity). All performance metrics are reported with 95% CI, providing a more informative evaluation in skewed class distributions.[31] To assess model robustness, feature stability across CV folds was evaluated using the Jaccard similarity index for selected endpoints.[32]

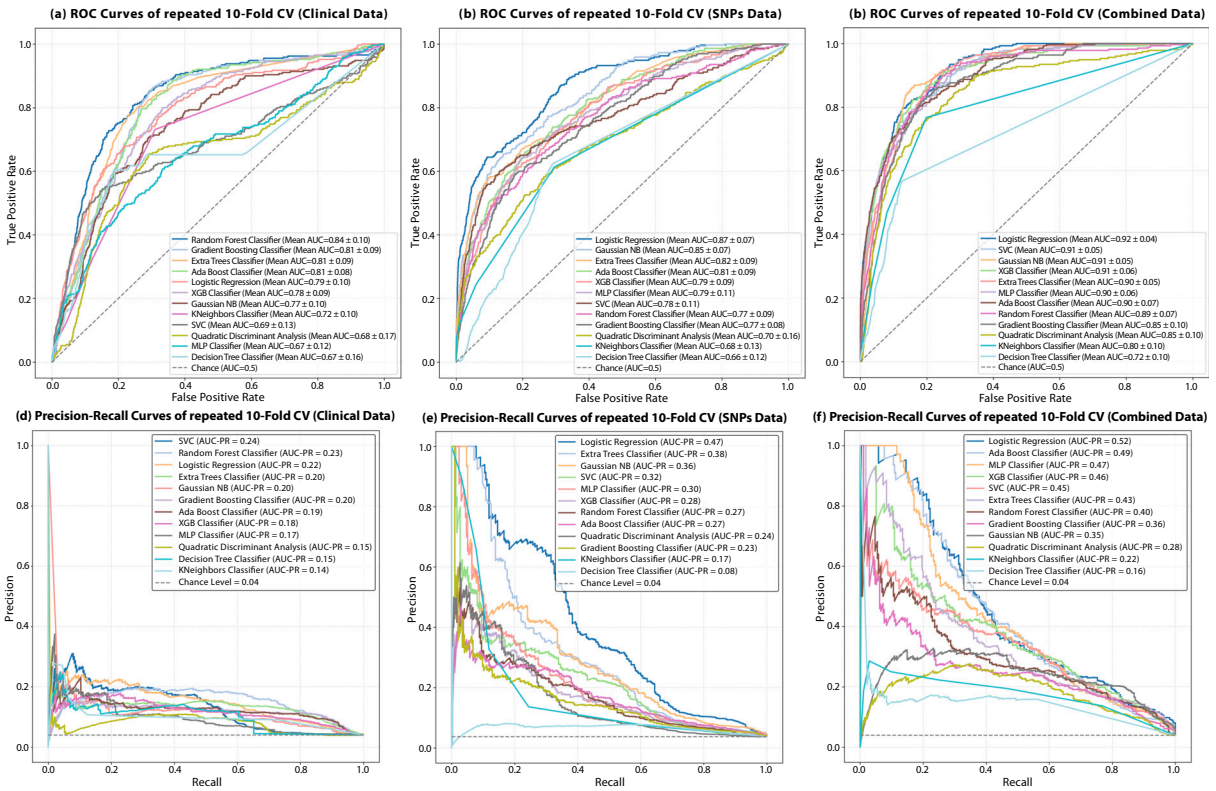


Fig. 1. Comparison of the repeated 10-fold CV performance for arm lymphedema prediction. Receiver operating characteristic (ROC) curves from n-repeated 10-fold CV using: (a) clinical features, (b) single nucleotide polymorphism (SNP) features, and (c) a combination of clinical and SNP features. Mean area under the ROC curve (AUC-ROC) ± SD for each machine learning classifier is shown in the panel legends. The diagonal line indicates chance-level performance (AUC = 0.5). (d-f) Precision-recall (PR) curves from n-repeated 10-fold CV using: (d) clinical features, (e) SNP features, and (f) a combination of clinical and SNP features. Mean area under the PR curve (AUC-PR) for each machine learning classifier is shown in the panel legends. The horizontal dashed line represents chance-level precision, which varies depending on the outcome distribution in each data set. *Abbreviation:* CV = cross-validation.

In the main analysis (Figs. 1-4), models were trained separately on clinical, SNP, and combined data sets using all available samples to maximize case inclusion. While assessing the added predictive value of SNPs, all data sets were aligned by common patient identifiers (IDs) to ensure that models trained on clinical, SNP, and combined feature sets were evaluated on the same subset of subjects. This alignment maintained methodological consistency and satisfied the assumptions required for paired statistical tests (paired *t* tests) and effect size measures (Cohen's *d*) to compare model performance across data sets. By using these tests, we compared the mean AUC of the combined model with that of the clinical model, as well as with the SNP model. *P* values of <.05 were considered statistically significant, and effect sizes were interpreted using conventional thresholds (0.2 = small; 0.5 = medium; 0.8 = large).

Gene and pathway analysis

Genes associated with each SNP were identified using snpXplorer batch annotation with Whole Blood GTEx data (https://snpxplorer.net/annotation/). Pathway analysis was carried out for each clinical endpoint using Reactome without interactors (https://reactome.org/PathwayBrowser/#TOOL=AT). Significant enrichment of pathways is judged by the false discovery rate (FDR) without correction for the number of endpoints.

Results

Distribution of 2-year radiation therapy side effects

The distribution of radiation therapy side effects at 2 years is shown in Table 1. Across the cohort, the most common events were breast atrophy (47.3%) and induration in the tumor area (39.5%), followed by breast skin hyperpigmentation (18.7%), nipple retraction (13.7%), breast edema (13.2%), and induration outside the tumor area (14.2%). Less frequent side effects included arm lymphedema (3.7%) and telangiectasia in the tumor area (4.6%). Across

endpoints, most cases were grade 1, whereas grade 3 was rare (<2%), observed only in breast atrophy and induration, both within and outside the tumor. For predictive modeling, outcomes were collapsed into 2 categories: no side effects (grade 0) versus any side effects (grades ≥1) to ensure adequate sample sizes and consistent comparison across side effects.

Feature importance and selection

Different feature subset sizes (20, 30, and 40 features) were evaluated during model development. The top 20 features provided the most consistent predictive performance within this data set and were therefore retained for subsequent analyses. We acknowledge that the optimal number of features may vary in external data sets. The top recurrent features across multiple side effects in the clinical data were age, smoking, mean heart radiation dose, number of nodes examined, and tumor size. When clinical and SNP features were combined, genetic variants (eg, rs74711328, rs145201352, rs146498859, rs7198421, and rs73478113) appeared among the most frequently selected predictors. The most important features for all 8 radiation therapy side effects across the 3 data sets are shown in Figure E1.

Prediction of arm lymphedema and breast atrophy

Model performance for arm lymphedema side effects was evaluated using ROC and PR curves. Prediction models were developed using clinical, SNP, and combined data sets and assessed using the mean AUC from repeated 10-fold CV, as illustrated in Figure 1a-c.

Figure 1a shows that ensemble methods—random forest, Gradient Boosting, and Extra Tree Classifier consistently achieved good performance, with the AUC around 0.84. In Figure 1b, some mean AUC values improved, with logistic regression achieving the highest AUC of 0.87, whereas the mean AUC of other models—random forest and KNN—decreased. The best result came from the combined data set in Figure 1c, where logistic regression and SVC models achieved AUCs >0.90. Based on these ROC curves, we confirm that integrating both sources of data leads to a more accurate prediction of arm lymphedema.

Given the class imbalance in arm lymphedema, we also evaluated models using PR curves (Fig. 1). Figure 1d-f presents the PR curves for the same machine learning models with 5 times repeated 10-fold CV, providing complementary insight into their ability to identify patients at risk of developing arm lymphedema. Figure 1d shows that most models achieved relatively low performance but much better than the chance-level, with AUC-PR values ranging from 0.14 to 0.24. The SVC obtained the highest AUC-PR (0.24), followed by random forest (0.23) and logistic regression (0.22). Most models performed only slightly better than chance (0.04), indicating limited predictive ability when using clinic variables alone. In Figure 1e, the models performed better compared with clinical data, with AUC-PR values ranging from 0.08 to 0.47. Logistic regression achieved the best result (AUC-PR = 0.47), followed by Extra Tree Classifier (0.38) and GaussianNB (0.36). This suggests that SNP data contain stronger discriminative information for arm lymphedema prediction compared to clinical variables. Integrating both the clinical and SNP data in Figure 1f improved overall model performance, with AUC-PR values ranging from 0.16 to 0.52. Logistic regression again achieved the highest performance (AUC-PR = 0.52), followed closely by AdaBoost (0.49) and MLP Classifier (0.47). These findings highlight that combining clinical and genetic features enhances predictive accuracy for arm lymphedema, outperforming either data source alone.

We next examined model performance for another clinically relevant endpoint, breast atrophy, using both ROC and PR curve analyses. In this case, the class distribution was more balanced, as shown in Table 1. Both metrics produced largely similar results, providing a reliable assessment of model performance. Figure 2 presents the ROC curves and PR curves for breast atrophy prediction using the same repeated 10-fold CV framework based on clinical data, SNPs, and their combination.

Figure 2 revealed a different pattern compared with the arm lymphedema ROC curve. Using clinical data alone in Figure 2a, the best-performing models achieved mean AUCs of ∼0.72, whereas models trained on SNP data alone showed reduced predictive performance, with AUCs around 0.60 across most machine learning algorithms (Fig. 2b). The combined data gave the best AUCs, reaching a mean AUC of 0.76 for ensemble models (Fig. 2c). Figure 2d-f presents the corresponding PR curves, which followed a similar trend. The mean AUC-PR values ranged from 0.53 to 0.73, again showing some improved performance for the combined data set. Both ROC and PR curves indicate that incorporating genetic markers adds some predictive value for breast atrophy side effects, but less compared with arm lymphedema (Fig. 2).

Comparison of machine learning across radiation therapy side effects

Across 12 machine learning models, the mean AUC values were obtained from repeated 10-fold CV as reported in the heatmaps, which summarize predictive performance across all side effects (Fig. 3). Figure 3a shows that random forest, Extra Tree Classifier, and gradient boosting consistently achieved the highest AUC values across most side effects. Whereas random forest reached an AUC of 0.84 for arm lymphedema and ∼0.70 for several other outcomes. In contrast, MLP Classifier and decision tree models exhibited comparatively poor performance, with AUC values frequently <0.65. Among these side effects, telangiectasia and induration outside the tumor were particularly difficult to predict, with all models achieving relatively low AUCs.

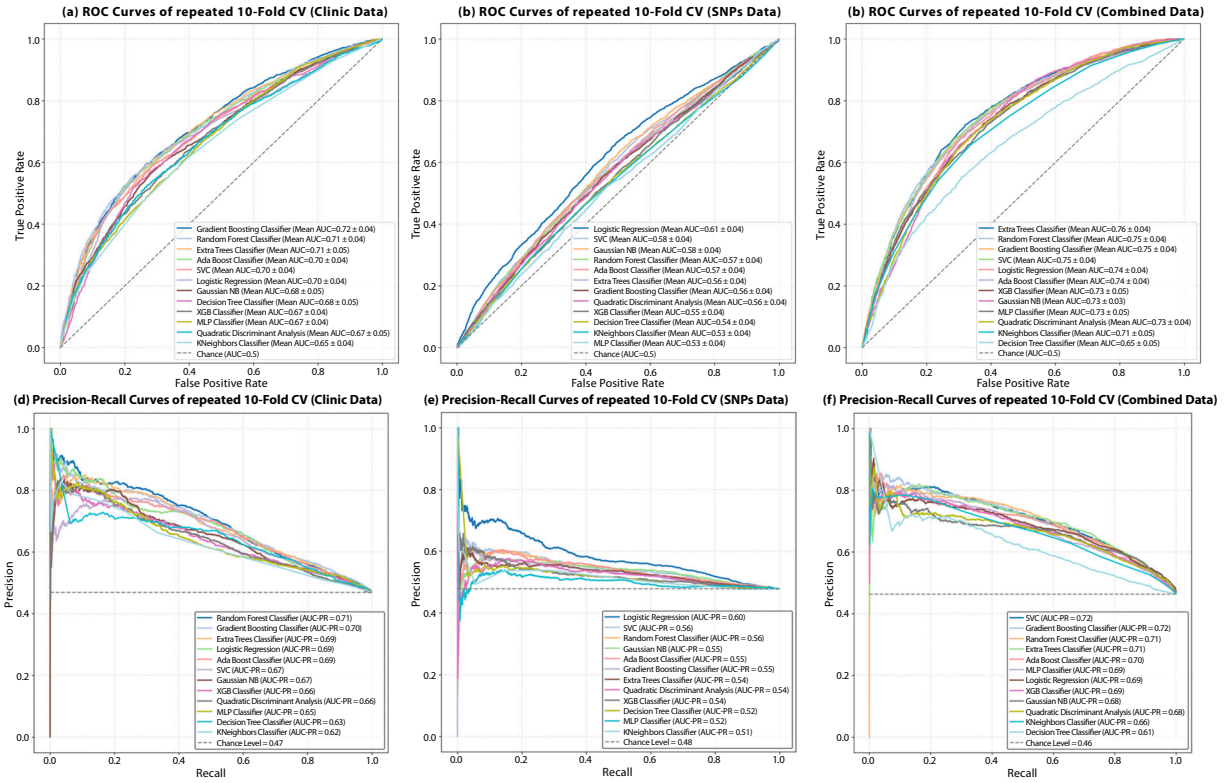


Fig. 2. Comparison of the repeated 10-fold CV performance for breast atrophy prediction. Receiver operating characteristic (ROC) curves from n-repeated 10-fold CV using: clinical features in panel (a), single nucleotide polymorphism (SNP) features in panel (b), and a combination of both in panel (c). Mean area under the ROC curve (AUC-ROC) ± SD for each machine learning classifier is shown in the legends. The diagonal line represents chance-level performance (AUC = 0.5). (d-f) Precision-recall (PR) curves from n-repeated 10-fold CV using clinical features in panel (d), SNPs in panel (e), and a combination of both in panel (f). The mean area under the PR curve (AUC-PR) for each machine learning classifier is shown in the legend. The horizontal dashed line in each panel indicates the chance-level precision, which varies according to the outcome distribution in each data set. *Abbreviation:* CV = cross-validation.

Although in Figure 3b, the overall performance trends remain similar to Figure 3a, the AUC scores are generally slightly higher for arm lymphedema, breast edema, and breast skin hyperpigmentation. Ensemble models like random forests and gradient boosting outperformed others, with AUCs > 0.75 for some predictable side effects, such as arm lymphedema and skin hyperpigmentation. However, the prediction of breast atrophy, induration in the tumor region, and telangiectasia was consistently weak, suggesting that SNP-based features may be insufficient for modeling these side effects correctly.

The strongest overall model performance was observed in Figure 3c, which presents the combined clinical and SNP data sets. In this setting, the random forest model achieved the highest AUC (0.90) for arm lymphedema and maintained strong performance across most outcomes. Extra Tree Classifier and gradient boosting also showed consistently high AUCs, confirming their effectiveness in modeling radiation therapy side effects. In contrast, MLP Classifier and Decision Tree underperformed, while induration inside the tumor and telangiectasia remained poorly predicted across all outcomes. Overall, the combined analysis confirms that arm lymphedema, skin hyperpigmentation, breast edema, nipple retraction, breast atrophy, and induration outside the tumor area are the most predictable outcomes, whereas induration inside the tumor and telangiectasia are less accurately predicted, regardless of model choice.

We next examined the performance of individual models across all side effects calculated with the standard error (Fig. 4). Among the 12 algorithms, random forest and logistic regression consistently achieved the highest predictive accuracy. Figure 4a shows that the random forest model performed consistently well across nearly all radiation therapy side effects. The highest AUC was observed for arm lymphedema, reaching 0.90 when both clinical and SNP features are used together, followed closely by strong performance on nipple retraction, breast edema, and skin hyperpigmentation. Notably, random forest could still find predictive value in the SNP-only feature set, which showed moderate performance on its own but better than expected for some outcomes—showing the model's ability to handle complex, nonlinear interactions.

Figure 4b illustrates the performance of the logistic regression model, which shows a different pattern compared

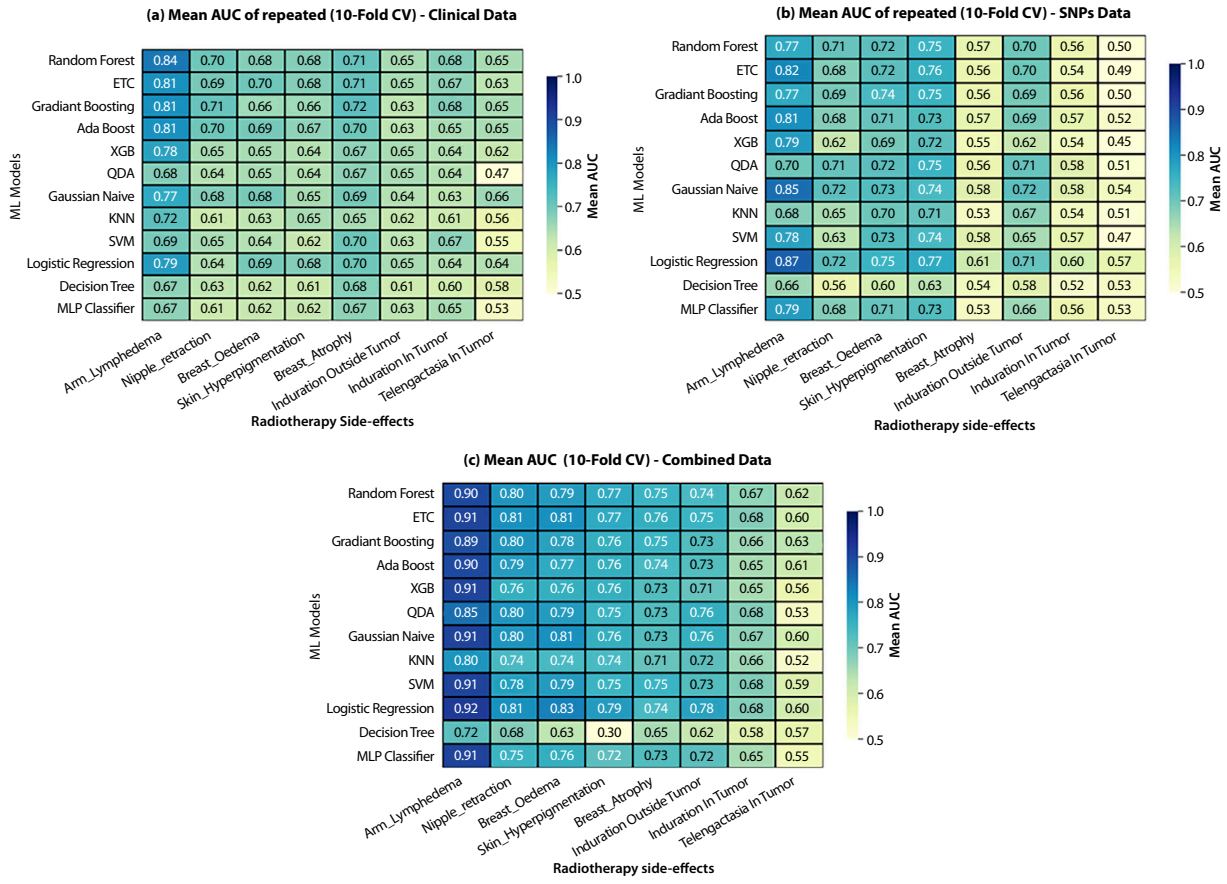


Fig. 3. Heatmaps of mean predictive performance across 12 machine learning models for 8 radiation therapy side effects. Heatmaps show the mean area under the curve (AUC) from repeated 10-fold CV for 3 types of input data: (a) clinical features, (b) single nucleotide polymorphisms (SNPs), and (c) combined clinical and SNP features. Rows represent machine learning models, and columns represent different radiation therapy-related side effects. Higher mean AUC values (darker colors) indicate better classification performance. *Abbreviation:* CV = cross-validation.

with random forest. Although overall performance is lower, combining clinical and SNP data again yielded the best results across most outcomes. Logistic regression achieved its highest AUC of 0.92 for arm lymphedema, outperforming random forest for this endpoint. However, when using clinical data alone, logistic regression performed weaker in most side effects, with the highest AUC nearly 0.78. With SNPs alone, logistic regression performed well, with AUC reaching 0.87 for the arm lymphedema side effect.

Statistical comparison of clinical, SNP, and combined models

Table 2 presents statistical comparisons of AUCs obtained from repeated 10-fold CV using paired *t* tests and Cohen's d effect sizes between models trained with clinical data alone, SNP data alone, and the combined data set. The mean scores of AUC-PR are also reported in Table E2.

Across most side effects, the combined models significantly outperformed models using clinical or SNP features alone. The mean AUCs in this table differ slightly in some side effects from those in Figures 1−4 because here the data sets for each side effect were aligned to include the same patients, as required for the paired statistical test. Table 2 indicates that the improvements were not only statistically significant ($P \leq .00001$ in nearly all comparisons) but also of large practical magnitude, as indicated by Cohen's d values of >1 for most endpoints. In breast atrophy, the combined model showed a significant improvement over the clinical model ($P \leq .00001$, Cohen's $d = 1.14$). Similarly, skin hyperpigmentation demonstrated strong gains compared with the clinical (Cohen's $d = 1.93$). For induration outside the tumor, the improvement was moderate. Both induration and telangiectasia showed no significant improvement over the clinical model ($P = .58$, Cohen's $d = 0.08$ and $P = .89$, Cohen's $d = -0.02$, respectively).

Table 2 shows the combined model for arm lymphedema achieved the highest performance (AUC = 0.93; 95% CI, 0.92-0.94), significantly improving upon the clinical-only model (AUC = 0.78; 95% CI, 0.75-0.81; $P < .001$; Cohen's $d = 1.44$) and the SNPs-only model (AUC = 0.89; 95% CI, 0.87-0.91; $P < .001$; Cohen's $d = 0.80$). For breast skin hyperpigmentation, the improvement was particularly large

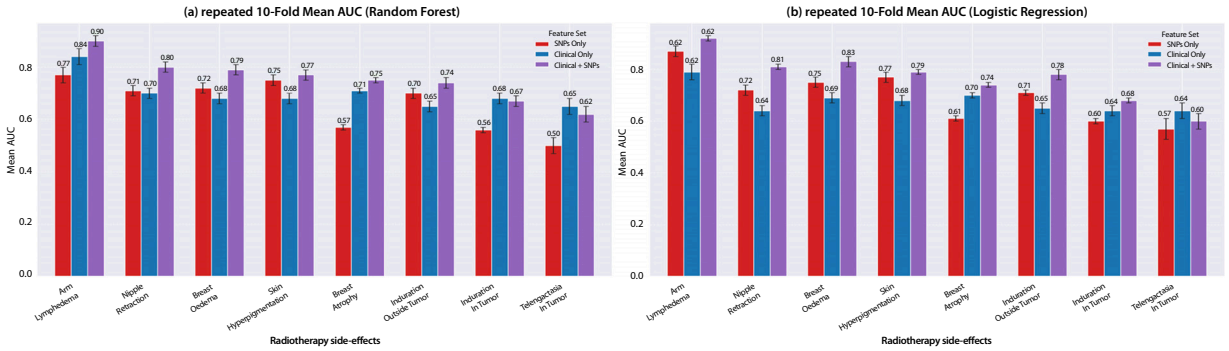


Fig. 4. Mean area under the curve (AUC) of repeated 10-fold CV of eight side effects using random forest and logistic regression models with standard error. Panel (a) presents results from the random forest model, whereas panel (b) shows those from logistic regression. Red bars represent single nucleotide polymorphism (SNP) features, blue bars refer to the clinical data, and purple bars represent features that came from combining both. Error bars indicate standard deviations across folds. *Abbreviation:* CV = cross-validation.

(Cohen's d = 1.86), indicating a strong contribution of SNPs to predictive accuracy. Improvements were high for breast edema (combined AUC = 0.82 vs clinical AUC = 0.67; Cohen's d = 1.78) and breast nipple retraction, where Cohen's d= 2.03, which means the improvement is double the pooled SD. Yet, for telangiectasia inside the tumor, neither SNPs nor combined models provided a meaningful improvement (combined AUC = 0.616 vs clinical AUC = 0.620; P> .7; Cohen's d= 0).

This contrast highlights 2 important points. First, model choice matters: random forest, as a nonlinear ensemble method, is more effective at capturing the complexity in genetic and clinical data. Second, integrating clinical and genetic data consistently improves performance significantly across both models, reinforcing the value of a multimodal approach in predicting radiation therapy side effects.

Feature stability and pathway analysis

Feature stability analysis demonstrated high consistency for SNP-only models (Jaccard index = 1.00) and clinical models (0.92). For combined clinical and SNP models, stability was moderate (0.61), likely reflecting the greater complexity of the combined feature space. Similar trends were observed for arm lymphedema (SNP: 0.95; clinical: 0.73; combined: 0.64). Detailed feature selection frequencies across CV folds are provided in Tables E4 and E5.

A pathway analysis for genes associated with the SNPs included in the prediction models for each clinical endpoint was carried out (Table E6). These analyses show significant pathways for arm lymphedema (FDR = 1.3×10^{-3}; regulation of cadherin 19 (CDH19) and type II classical cadherin expression and function), nipple retraction (FDR = 1.9×10^{-3}; formation of the cornified envelope and keratinization), and induration in the tumor bed (FDR = 1.7×10^{-3}; development of pancreatic precursor cells) and breast edema (FDR = 5.5×10^{-4}; Ionotropic activity of kainite receptors). No significant pathways were identified for breast atrophy, skin hyperpigmentation, and induration outside the tumor bed.

Discussion

Our study evaluated the performance of 12 machine learning approaches to predict breast radiation therapy-related effects using clinical and/or genetic data. Predictive performance varied across the eight endpoints studied, with some outcomes, such as arm lymphedema, having relatively high discrimination. However, this finding should be interpreted with caution given the low event rate and the potential impact of class imbalance on performance metrics. Our results compare favorably with those reported in the recent systematic review and meta-analysis, which found pooled discrimination metrics in the moderate range across a heterogeneous set of studies and cancer sites.[5]

An important observation from our study is that simpler and clinically interpretable models, such as logistic regression, consistently performed as well as or better than more complex algorithms, such as tree-based and ensemble methods, across endpoints. Each endpoint was trained and tested independently using 12 different machine learning algorithms with repeated 10-fold CV. This is particularly relevant in a clinical context, where model transparency and interpretability are essential. Both clinical and genetic factors contributed independent predictive value, and their combination improved model performance for several endpoints. While improvements were modest, they were consistent across several endpoints and evaluation metrics. The endpoint-specific performance pattern suggests selective, rather than indiscriminate, model fitting. However, these findings should be interpreted within the context of the binary side effects definitions used in this study.

Previous studies anticipated the integration of SNPs into multimodal models of radiation adverse effects

Table 2 Mean AUC values with 95% CI and statistical comparisons of clinical, SNP, and combined models for 2-year RT side effects

(a) Random forest performance and statistical comparisons										
2-y RT side effects	**clinic AUC mean**	**clinic AUC CI**	**SNPs AUC mean**	**SNPs AUC CI**	**Combined AUC mean**	**combined AUC CI**	***P* value: clinic vs combined AUC**	**Cohen's *d* clinic vs combined AUC**	***P* value: SNPs vs combined AUC**	**Cohen's *d* SNPs vs combined AUC**
Arm lymphedema	0.84	(0.82-0.86)	0.79	(0.76-0.81)	0.90	(0.88-0.92)	0	0.62	≤.00001	1.34
Breast atrophy	0.71	(0.70- 0.72)	0.55	(0.54-0.57)	0.75	(0.73-0.76)	≤.00001	1.14	≤.00001	2.84
Skin hyperpigmentation	0.66	(0.65-0.68)	0.74	(0.72-0.76)	0.78	(0.77-0.79)	≤.00001	1.93	≤.00001	0.88
Breast edema	0.69	(0.67-0.71)	0.73	(0.71-0.74)	0.78	(0.77-0.80)	≤.00001	1.16	≤.00001	1.12
Breast nipple retraction	0.70	(0.67-0.72)	0.66	(0.64-0.69)	0.79	(0.77-0.80)	≤.00001	1.28	≤.00001	1.29
Induration outside tumor area	0.62	(0.60-0.64)	0.69	(0.67-0.72)	0.73	(0.71-0.75)	≤.00001	1.24	0	0.51
Induration in tumor bed (scar tissue)	0.66	(0.65-0.68)	0.56	(0.54-0.57)	0.66	(0.65-0.68)	.58	0.08	≤.00001	1.96
Telangiectasia in (tumor area)	0.64	(0.61-0.70)	0.48	(0.45-0.51)	0.64	(0.61-0.67)	.89	-0.02	≤.00001	1.01
(b) Logistic regression performance and statistical comparisons										
2-y RT side effects	clinic AUC mean	clinic AUC CI	SNPs AUC mean	SNPs AUC CI	combined AUC mean	combined AUC CI	*P* clinic vs combined AUC	Cohen's d clinic vs combined AUC	*P* SNPs vs combined AUC	Cohen's d SNPs vs combined AUC
Arm lymphedema	0.78	(0.75-0.81)	0.89	(0.87-0.91)	0.93	(0.92-0.94)	≤.00001	1.44	≤.00001	0.80
Breast atrophy	0.70	(0.68-0.71)	0.60	(0.59-0.62)	0.75	(0.73-0.76)	≤.00001	1.44	≤.00001	2.75
Skin hyperpigmentation	0.68	(0.66-0.69)	0.77	(0.75-0.78)	0.80	(0.79-0.81)	≤.00001	1.86	≤.00001	0.94
Breast edema	0.70	(0.67-0.72)	0.76	(0.74-0.78)	0.82	(0.81-0.83)	≤.00001	1.78	≤.00001	1.13
Breast nipple retraction	0.64	(0.61-0.66)	0.70	(0.68-0.72)	0.80	(0.78-0.82)	≤.00001	2.03	≤.00001	1.45
Induration outside tumor area	0.63	(0.61-0.65)	0.72	(0.70-0.74)	0.77	(0.75-0.78)	≤.00001	1.81	≤.00001	0.87
Induration in tumor bed (scar tissue)	0.63	(0.62-0.65)	0.59	(0.57-0.60)	0.67	(0.66-0.68)	≤.00001	0.86	≤.00001	1.64
Telangiectasia in (tumor area)	0.62	(0.59-0.65)	0.56	(0.53-0.58)	0.62	(0.58-0.65)	0.77	-0.04	0	0.46

Abbreviations: AUC = area under the curve; SNP = single nucleotide polymorphism.

(a) Results obtained using the random forest classifier; (b) results obtained using logistic regression. For each outcome, model discrimination was evaluated using the mean of AUC-ROC computed from repeated 10-fold cross-validation. Paired *t* test comparisons between models were performed with corresponding *P* values reported for "Clinical vs Combined" and "SNPs vs Combined" comparisons. Effect sizes were quantified using Cohen's *d* to indicate the magnitude of improvement (small ≈ 0.2, medium ≈ 0.5, large ≥ 0.8). AUC values near 0.5 represent chance-level discrimination, while higher values indicate better model performance.

incorporating clinical risk factors.[11,12,14] Training such models in large, well-characterized cohorts provides an excellent opportunity to evaluate and refine radiogenomic risk prediction approaches. In this study, we used the REQUITE data set, which contains comprehensive longitudinal data on radiation therapy-related side effects, along with detailed clinical and treatment information.

We deliberately evaluated 12 machine learning algorithms to provide a comparative benchmark, rather than aiming to identify a single optimal model for clinical deployment.[33] Our approach allowed us to assess the consistency of predictive performance across fundamentally different learning paradigms[34,35] and to examine whether increasing model complexity yields meaningful gains over simpler, interpretable approaches. By comparing models trained on clinical data, genetic data, and their combination, we further assessed the relative contribution of each data type and the value of data integration for side effects prediction.[14,33]

Considering the eight side effects endpoints, the impact of incorporating SNP features was highly endpoint dependent. Mean predictive performance improved for 7 side effects: arm lymphedema, breast atrophy, breast edema, breast skin hyperpigmentation, induration outside the tumor, induration in the tumor area and nipple retraction. PR analysis was included as it provides a more informative evaluation in the presence of imbalanced outcomes. Although the mean AUC-PR for the combined data set was lower (~52%) compared with the mean AUC-ROC (92%), this value remains substantially higher than random guessing, given that the baseline chance-level for arm lymphedema is around 4%, which represents a 13-fold improvement over the baseline. In breast atrophy, the combined data set clearly highlights the synergistic advantage of integrating clinical and genetic information, with classifiers showing consistently better performance than when using clinical or SNP data alone. Notably, the AUC-ROC values for the best-performing classifiers reach 0.76, whereas the corresponding AUC-PR values reach 0.72. The relatively close values of AUC-ROC and AUC-PR can be explained by the balanced classes of the atrophy side effect in the data set. In contrast, for certain endpoints such as telangiectasia in the tumor area and induration in the tumor bed (fibrosis), predictive performance remained limited regardless of data type, suggesting that these outcomes may be driven by factors not captured in the available clinical or genetic variables.

Consistent with previous radiogenomic studies, this study highlights the importance of identifying additional risk-associated SNPs and rare variants that may contribute to inter-individual differences in radiation response.[12] In line with these observations, our results indicate that SNP-based models alone provide good predictive performance for certain side effects, such as arm lymphedema, breast edema and breast skin hyperpigmentation, even in the absence of clinical variables. These results reinforce the need to continue searching for genetic variants associated with radiation response, while also underscoring the importance of careful endpoint selection when integrating genetic data into predictive models.

Among the 12 models evaluated, random forest and logistic regression consistently provided the highest predictive accuracy. The two are complementary strategies: ensemble tree-based methods capable of modeling complex feature interactions, and simpler linear models that remain robust in smaller data sets or when predictive signals are more additive in nature.[36,37] Notably, logistic regression, though a less complex model, achieved its highest AUC (0.92) for arm lymphedema, highlighting that even additive effects of SNPs can meaningfully enhance prediction when relevant variants are included. This suggests that model interpretability and feature selection remain crucial in radiogenomic analyses—complex models are not always necessary if key predictors are biologically relevant.[38,39]

The question of whether clinical, genetic, or combined data are necessary for the prediction of side effects is an important question for clinical implementation in future trials. Clinical data and treatment-related data are easy and inexpensive to capture using electronic case report forms (eCRFs) in a trial, whereas at the moment, germline SNP genotyping adds an additional step and cost to a trial. Although the cost of genotyping is relatively modest in the context of a clinical trial (<$100 per patient), there may also be logistical implications and a cost for shipping samples to a central genotyping facility. For successful clinical implementation, one important limitation is the short time interval available for genotyping, which must occur between the initial patient consultation and radiation therapy planning, often a period between 7 and 14 days. These issues mean that the increase in predictive value from genetic variables must be clinically significant and, in the context of a clinical trial, justified by a power calculation showing lower patient recruitment is needed. These issues cannot be resolved by inspection of the AUCs alone, but the size of the increase in predictive value, as reported here, will be an important component of the decision.

Despite its strengths and novelty, this study has limitations. Although the REQUITE cohort provides a large and well-characterized data set with ≤96 months of follow-up, external validation is required to confirm generalizability. Although treatment-related variables such as fractionation, radiation therapy dose, intensity modulated radiation therapy use, and nodal involvement were included in the analysis, more detailed dosimetric parameters (eg, dose-volume metrics) were not consistently available across centers despite harmonized prospective data collection. As a result, these were not comprehensively incorporated into the modeling, which may limit the ability to fully capture treatment-related and center-specific variability in side effect outcomes.

Moreover, the SNPs analyzed represent only those identified from previous GWAS results based on 2-year side effect outcomes,[11] which was conducted within the same data set, although patients of non-European origin were excluded because the analysis was restricted to patients of European

ancestry, which may limit the generalizability of these SNPs to more diverse populations. Expanding our analyses to cumulative polygenic risk scores, epigenetic markers, and transcriptomic profiles may yield more comprehensive models. Nevertheless, the pathway analyses identified 2 highly plausible, enriched biological pathways: first, regulation of cadherins in arm lymphedema, which accords with evidence of the role particularly of VE-cadherin in maintaining lymphatic endothelial cell junctions. Second, formation of the cornified envelope and keratinization genes were associated with nipple retraction.

Lastly, this approach may hide differences between mild and severe side effects because it treats side effects as a simple yes-or-no outcome, defined according to clinically significant cut-offs determined by side effects grade. Although this simplifies modeling, it may obscure clinically meaningful differences in side effects severity. Future work should explore ordinal or multi-class models, as well as time-to-event approaches, to better capture side effects dynamics.

Importantly, the primary objective of this study was not to develop a clinically deployable prediction tool, but to provide a systematic comparison of machine learning approaches for side effects prediction. In this context, the findings offer a benchmark for future model development.

Conclusions

Radiation therapy side effect prediction in breast cancer is highly dependent on the specific endpoint and modeling strategy used. By applying 12 machine learning approaches across eight outcomes in a large multicenter cohort, this study identifies lymphedema as a particularly robust and predictable endpoint and demonstrates that simpler, clinically interpretable models can perform as well as or better than more complex algorithms. Although integration of genetic data provided additional predictive value for selected endpoints, its benefit was not uniform, showing the need for targeted rather than routine incorporation of genetic features. These findings provide a comparative framework to guide future model selection, data integration, and endpoint definition. External validation and evaluation of clinical utility across diverse populations and demographic groups are required before these approaches can be translated into clinical practice. This ensures fairness, robustness, and broader clinical applicability.

References

1. Barazzuol L, Coppes RP, van Luijk P. Prevention and treatment of radiotherapy-induced side effects. *Mol Onc* 2020;14:1538-1554.
2. Dilalla V, Chaput G, Williams T, Sultanem K. Radiotherapy side effects: Integrating a survivorship clinical lens to better serve patients. *Curr Oncol* 2020;27:107-112.
3. Khan M, Shiwlani A, Qayyum MU, Mannan A, Sherani K, Hussian HK. AI-powered healthcare revolution: An extensive examination of innovative methods in cancer treatment. *BULLET : Jurnal Multidisiplin Ilmu* 2024;3:87-98.
4. Patil S, Shankar H. Transforming healthcare: Harnessing the power of AI in the modern era. *Int J Multidiscipl Sci Arts* 2023;2:60-70.
5. Ugwu GC, Jalali F, Liu G, Li G, Langendijk JA, Alizadeh BZ. The actual performance of ML/AI models in predicting radiation-induced toxicity in head and neck cancer: a systematic review and meta-analysis. *Radiother Oncol* 2026;216:111350.
6. Taylor CW, Kirby AM. Cardiac side-effects from breast cancer radiotherapy. *Clin Oncol* 2015;27:621-629.
7. Rammant E, Deman E, Fonteyne V, et al. Using machine learning to predict patient-reported symptom clusters in prostate cancer patients receiving radiotherapy. *Health Qual Life Outcomes* 2025;24:3.
8. Andreassen CN, Rosenstein BS, Kerns SL, et al. Individual patient data meta-analysis shows a significant association between the ATM rs1801516 SNP and toxicity after radiotherapy in 5456 breast and prostate cancer patients. *Radiother Oncol* 2016;121:431-439.
9. Dorling L, Barnett GC, Michailidou K, et al. Patients with a high polygenic risk of breast cancer do not have an increased risk of radiotherapy toxicity. *Clin Cancer Res* 2016;22:1413-1420.
10. 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.
11. Jandu HK, Veal CD, Fachal L, et al. Genome-wide association study of treatment-related toxicity two years following radiotherapy for breast cancer. *Radiother Oncol* 2023;187:109806.
12. Kerns SL, Ostrer H, Rosenstein BS. Radiogenomics: using genetics to identify cancer patients at risk for development of adverse effects following radiotherapy. *Cancer Discov* 2014;4:155-165.
13. Kerns SL, Fachal L, Dorling L, et al. Radiogenomics Consortium genome-wide association study meta-analysis of late toxicity after prostate cancer radiotherapy. *J Natl Cancer Inst* 2020;112:179-190.
14. Aguado-Flor E, Reyes VM, Navarro V, et al. Integrating genetic polymorphisms and clinical data to develop predictive models for skin toxicity in breast cancer radiation therapy. *Breast* 2025;82:104506.
15. Massi MC, Gasperoni F, Ieva F, et al. A deep learning approach validates genetic risk factors for late toxicity after prostate cancer radiotherapy in a REQUITE multi-national cohort. *Front Oncol* 2020;10:541281.
16. Aldraimli M, Soria D, Grishchuck D, et al. A data science approach for early-stage prediction of patient's susceptibility to acute side effects of advanced radiotherapy. *Comput Biol Med* 2021;135:104624.
17. UNICANCER. Multicenter randomized pivotal stage clinical investigation assessing the communication of an individualized AI-based risk prediction of arm lymphedema to breast cancer patients with an indication for regional lymph node irradiation and their physician as part of treatment shared-decision. clinicaltrials.gov; 2025.
18. Seibold P, Webb A, Aguado-Barrera ME, et al. REQUITE: A prospective multicentre cohort study of patients undergoing radiotherapy for breast, lung or prostate cancer. *Radiother Oncol* 2019;138:59-67.
19. Oh JH, Kerns S, Ostrer H, Powell SN, Rosenstein B, Deasy JO. Computational methods using genome-wide association studies to predict radiotherapy complications and to identify correlative molecular processes. *Sci Rep* 2017;7:43381.
20. Demircioğlu A. Measuring the bias of incorrect application of feature selection when using cross-validation in radiomics. *Insights Imaging* 2021;12:172.
21. Wilimitis D, Walsh CG. Practical considerations and applied examples of cross-validation for model development and evaluation in health care: Tutorial. *JMIR AI* 2023;2:e49023.
22. Breiman L. Random forests. *Mach Learn.* 2001;45:5-32.
23. Geurts P, Ernst D, Wehenkel L. Extremely randomized trees. *Mach Learn* 2006;63:3-42.
24. Edwards AWF, et al. R.A. Fischer, statistical methods for research workers, first edition (1925). In: Grattan-Guinness I, Cooke R, Corry L, eds. *Landmark Writings in Western Mathematics 1640-1940*. Elsevier Science; 2005:856-870.

ARTICLE IN PRESS

25. Pearson KX. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *Lond Edinb Dublin Philos Mag J Sci.* 1900;50:157-175.
26. Hinton GE, Salakhutdinov RR. Reducing the dimensionality of data with neural networks. *Science* 2006;313:504-507.
27. Guyon I, Weston J, Barnhill S, Vapnik V. Gene selection for cancer classification using support vector machines. *Mach Learn* 2002;46:389-422.
28. Tsamardinos I, Rakhshani A, Lagani V. Performance-estimation properties of cross-validation-based protocols with simultaneous hyper-parameter optimization. In: Likas A, Blekas K, Kalles D, eds. *Artificial Intelligence: Methods and Applications*. *Artificial Intelligence: Methods and Applications*. 8445, Springer International Publishing; 2014:1-14. Lecture Notes in Computer Science.
29. Luo Y, Tseng H-H, Cui S, Wei L, Ten Haken RK, El Naqa I. Balancing accuracy and interpretability of machine learning approaches for radiation treatment outcomes modeling. *BJR Open* 2019;1:20190021.
30. Sousa S, Paredes S, Rocha T, Henriques J, Sousa J, Gonçalves L. Machine learning models' assessment: trust and performance. *Med Biol Eng Comput* 2024;62:3397-3410.
31. Saito T, Rehmsmeier M. The precision-recall plot is more informative than the ROC plot when evaluating binary classifiers on imbalanced datasets. *PLOS One* 2015;10:e0118432.
32. Kalousis A, Prados J, Hilario M. Stability of feature selection algorithms: a study on high-dimensional spaces. *Knowl Inf Syst* 2007;12:95-116.
33. Rattay T, Bologna G, Bombezin-Domino A, et al. Development of an explainable AI prediction model for arm lymphoedema following breast cancer surgery and radiotherapy. *Eur J Cancer* 2024;200.
34. Sayin B, Yang J, Chen X, Passerini A, Casati F. Rethinking and recomputing the value of machine learning models. *Artif Intell Rev.* 2025;58:238.
35. Tousi A, Luján M. Comparative analysis of machine learning models for performance prediction of the SPEC benchmarks. *IEEE Access* 2022;10:11994-12011.
36. Taha K. Tree-based ensemble learning models for protein-protein interactions detection: A review and experimental evaluation. *BioData Min* 2025;19:1.
37. Perez BC, Bink MCAM, Svenson KL, Churchill GA, Calus MPL. Prediction performance of linear models and gradient boosting machine on complex phenotypes in outbred mice. *G3 (Bethesda)* 2022;12:jkac039.
38. Chen X, Ishwaran H. Random forests for genomic data analysis. *Genomics* 2012;99:323-329.
39. Prencipe B, Delprete C, Garolla E, et al. An explainable radiogenomic framework to predict mutational status of KRAS and EGFR in lung adenocarcinoma patients. *Bioengineering (Basel)* 2023;10:747.