

An interpretable machine-learning model to predict moderate-to-severe acute radiation esophagitis following regional nodal irradiation for breast cancer

BIN HE^{1,2*}, ZHONGCHUN QI^{2*}, HENGDI ZHANG³, LINLING LI⁴, HUA LI²,
XINYUN ZOU², YU TAN⁴, LING ZHANG² and FANGYUAN KONG²

¹Department of Oncology and Radiotherapy, Chengdu Qingbaijiang District People's Hospital, Chengdu, Sichuan 610300, P.R. China; ²Department of Oncology, The General Hospital of Western Theater Command of The Chinese People's Liberation Army, Chengdu, Sichuan 610083, P.R. China; ³Department of Ophthalmology, The General Hospital of Western Theater Command of The Chinese People's Liberation Army, Chengdu, Sichuan 610083, P.R. China; ⁴School of Clinical Medicine, Chengdu Medical College, Chengdu, Sichuan 610500, P.R. China

Received January 27, 2026; Accepted September 16, 2026

DOI: 10.3892/ol.2026.15909

Abstract. Acute radiation esophagitis (ARE) is a clinically relevant toxicity occurring in patients with breast cancer receiving post-operative regional nodal irradiation. The present study aimed to develop and evaluate an interpretable machine-learning (ML) model for grade ≥ 2 ARE. A total of 324 patients were allocated by outcome-stratified random sampling (7:3) to a training set (n=228) and a held-out test set (n=96). A total of six consensus predictors were retained: Pre-existing pruritus, histological grade, body mass index (BMI), diabetes, tumor laterality and the relative esophageal volume receiving ≥ 30 Gy (RV_V30). In total, 10 ML algorithms were evaluated. Bootstrap internal validation (1,000 resamples) revealed that logistic regression (LR) had the lowest mean optimism among the three preferred candidates (0.027). Final model selection considered bootstrap stability and corrected performance together with held-out discrimination and calibration, model complexity and clinical interpretability. In the held-out test set, LR achieved an area under the curve value of 0.797 (95% confidence interval, 0.706-0.889), a Brier score of 0.179, a calibration slope of 1.258 and a calibration-in-the-large of -0.056. Decision curve analysis revealed positive net benefit

relative to treat-all and treat-none strategies across threshold probabilities of 10-66%. SHapley Additive exPlanations identified histological grade, RV_V30 and BMI as the most influential predictors. The final model was implemented as a nomogram and web-based calculator. Overall, the present study demonstrates that this six-feature clinical-dosimetric model may support individualized risk stratification and radiotherapy planning, but prospective multicenter external validation is required before routine clinical use.

Introduction

Breast cancer continues to pose a notable global health burden, representing one of the most prevalent malignancies affecting women worldwide. According to GLOBOCAN 2024, female breast cancer accounted for an ~2.43 million new cases and 693,660 mortalities worldwide in 2024, and was the most commonly diagnosed cancer and the leading cause of cancer-associated mortality among women globally (1). For patients with lymph node metastasis or other high-risk factors, post-operative radiotherapy is important for reducing local recurrence and breast cancer-specific mortality (2,3). The increasing adoption of regional nodal irradiation (RNI) has further improved oncological outcomes (4); however, it also elevates the risk of developing toxicities to adjacent organs.

The esophagus, due to its anatomical proximity to the supraclavicular and internal mammary nodal regions, often receives incidental radiation during RNI, resulting in acute radiation esophagitis (ARE). It has been reported that symptomatic ARE occurs in 30-50% of patients with breast cancer receiving RNI (5). ARE typically manifests as odynophagia, dysphagia and retrosternal pain, which, in severe cases, may lead to malnutrition, dehydration and treatment interruptions, thereby substantially impairing the quality of life of patients. In the era of precision radiotherapy, the early identification of high-risk patients is key for improving quality of life, maintaining treatment efficacy and reducing treatment-related morbidity (6).

Correspondence to: Professor Ling Zhang or Professor Fangyuan Kong, Department of Oncology, The General Hospital of Western Theater Command of The Chinese People's Liberation Army, 270 Tianhui Road, Rongdu Avenue, Jinniu, Chengdu, Sichuan 610083, P.R. China
E-mail: zhangling71@163.com
E-mail: 13540212456@163.com

*Contributed equally

Key words: breast cancer, radiation-induced esophagitis, regional lymph node radiotherapy, machine-learning, predictive model

Previous studies have identified numerous factors associated with ARE, including esophageal dose-volume parameters, treatment techniques and patient characteristics. Mean esophageal dose (Dmean), V10, V20, and absolute volumes (AV) of an organ at risk receiving 30 or 35 Gy (AV30 and AV35) have been shown to be associated with the incidence and severity of esophagitis (7-9). Treatment techniques may also influence esophageal exposure during breast RNI (8,10), and tumor laterality has been found to be associated with the risk of developing ARE (11). However, conventional prediction approaches may not adequately capture interactions among clinical and dosimetric variables, and clinically interpretable risk tools remain limited.

Cancer is biologically heterogeneous, with recent studies identifying prognostic molecular markers, including endoplasmic reticulum stress-associated proteins, AIMP2 and transcription factors associated with disease-free survival in triple-negative breast cancer (12-14). Although these molecular studies did not evaluate radiotherapy toxicity, they reinforce the broader rationale for individualized risk assessment. In parallel, artificial intelligence (AI) is increasingly being explored in radiotherapy and immunotherapy, with emphasis on clinically validated and interpretable models (15).

Machine-learning (ML) provides a framework for analyzing complex, high-dimensional data and may capture non-linear associations that are difficult to represent with conventional approaches (16,17). This flexibility has motivated the use of ML for clinical outcome prediction (18,19). Interpretability methods, such as SHapley Additive exPlanations (SHAP) can further decompose model predictions into feature-level contributions and may improve transparency in clinical decision support (20).

Therefore, the present study aimed to develop and evaluate an interpretable ML-based model for predicting grade ≥ 2 ARE in patients with breast cancer receiving RNI. In the present study, univariable and multivariable analysis, least absolute shrinkage and selection operator (LASSO) regression and the Boruta algorithm for feature selection were used, followed by the comparison of 10 ML algorithms. Bootstrap resampling was used to quantify optimism and model stability. The final model was selected by considering internal validation, held-out performance, calibration, parsimony and clinical interpretability. The present study was reported in accordance with the transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD)+AI statement (21).

Materials and methods

Study design. The present single-center retrospective observational study included patients with breast cancer who received post-operative RNI at The General Hospital of Western Theater Command of the Chinese People's Liberation Army (Chengdu, China) between January 2016 and June 2025. Patient data were retrieved between July and September 2025, and statistical analyses were conducted between September and December 2025. The study protocol received approval from the Institutional Review Board of The General Hospital of Western Theater Command of the Chinese People's Liberation Army (approval no. 2025EC11-ky019) and was conducted in

accordance with the principles of the amended Declaration of Helsinki. Due to the retrospective design, the Institutional Review Board granted a waiver for written informed consent. All patient identifiers were anonymized during the analysis.

Patient selection. A total of 324 patients with breast cancer who received adjuvant RNI between January 2016 and June 2025 were retrospectively enrolled. The inclusion criteria were as follows: i) Female patients with pathologically confirmed breast cancer; ii) those who underwent mastectomy or breast-conserving surgery; iii) those whose post-operative radiotherapy target volume included the supraclavicular/intraclavicular and/or internal mammary lymph node regions; iv) those whose radiotherapy technique was intensity-modulated radiotherapy (IMRT) alone or IMRT combined with electron beams; and v) those with an availability of complete clinical, laboratory and radiotherapy planning data. The key exclusion criteria included the following: i) The receipt of concurrent chemotherapy; ii) a prior history of thoracic radiotherapy; iii) pre-existing esophageal disease or swallowing dysfunction; and iv) a diagnosis of other active malignancies or notable missing key data ($>10\%$ for any variable).

Treatment and dosimetric parameters. All patients received radiotherapy within 8 weeks post-operatively (if adjuvant chemotherapy was not planned) or upon the completion of adjuvant chemotherapy. Patients were positioned using computed tomography (CT) simulation scans and immobilized with thermoplastic body masks. For patients who underwent mastectomy, a 5-mm-thick silicone sheet was placed on the chest wall surface during radiotherapy. The irradiation fields encompassed the whole breast/chest wall, supraclavicular/intraclavicular fossa and level II axillary lymph nodes.

For the present study, the esophagus was contoured from the inferior border of the cricoid cartilage to the inferior border of the aortic arch (Fig. S1). A representative intensity-modulated radiotherapy plan and dose distribution is presented in Fig. S2. Other organs at risk, including the spinal cord, thyroid gland, contralateral breast, lungs and heart, were also delineated. Patients received IMRT with conventional or hypofractionated regimens. As the total dose varied with fractionation and the use of a tumor-bed boost following breast-conserving surgery, the total dose was analyzed as a continuous variable. Dose-volume histograms were used to extract the Dmean, maximum dose (Dmax), total esophageal volume, relative volumes receiving 10-40 Gy (RV_V10-RV_V40) and AV receiving 10-40 Gy (AV_V10-AV_V40). Relative dose-volume metrics are expressed as the percentage of the contoured esophageal volume receiving at least the specified dose. Concurrent targeted therapy was permitted for patients with HER2-positive disease, whereas concurrent chemotherapy was not permitted. Baseline pruritus was defined as pruritus documented prior to radiotherapy initiation, either at CT simulation or at the pretreatment visit within 7 days and was recorded as present or absent. Pruritus arising during radiotherapy was not included in the prediction model to avoid post-baseline information leakage.

Missing data. Missing data was assessed for all candidate predictors and was found to account for $<5\%$. Little's test was used to assess whether the missingness pattern was

compatible with missing completely at random (MCAR). The null hypothesis was that the data were MCAR, and a two-sided $P > 0.05$ was interpreted as insufficient evidence to reject this hypothesis. Given the low overall missingness and the MCAR result ($P > 0.05$), complete-case analysis was used for model development and evaluation.

Follow-up and toxicity evaluation. ARE was graded according to the Radiation Therapy Oncology Group acute radiation morbidity scoring criteria (22). Assessments were performed during radiotherapy and at 1 week, 2 weeks and 3 months following the completion of treatment. At least two radiation oncologists reviewed the ARE status. Disagreements were adjudicated by a senior oncologist. The highest recorded ARE grade during the assessment period was used for analysis. The primary endpoint was the occurrence of grade ≥ 2 ARE at any assessment during this predefined period.

Feature selection and model construction. Patients were divided into a training set (70%) and a held-out test set (30%) by outcome-stratified random sampling using a fixed seed of 52. Feature selection and model development were performed primarily in the training set. No post-enrollment imputation was required for the complete-case dataset used for model development. For algorithms requiring scaled inputs, centering and scaling parameters were estimated from the training data and then applied unchanged to the held-out data.

Candidate predictors were screened in the training set using three complementary methods: Univariable analysis followed by multivariable logistic regression (LR), LASSO regression and the Boruta algorithm (23-25). For LASSO, a fixed random seed of 43 was used with a binomial model, $\alpha = 1$, a maximum of 100,000 iterations and 10-fold cross-validation based on binomial deviance; λ_1 standard error (λ_1 SE) was selected as the penalty parameter. For Boruta, a fixed random seed of 123 and $\text{maxRuns} = 500$ were used, followed by TentativeRoughFix for tentative attributes. The final consensus feature set comprised variables retained by > 2 of the three selection approaches: Pre-existing pruritus, histological grade, BMI, diabetes, tumor laterality and esophageal RV_V30.

In total, 10 supervised ML algorithms were evaluated: LR, support vector machine (SVM), gradient boosting machine (GBM), neural network (NN), random forest (RF), extreme gradient boosting (XGBoost), K-Nearest Neighbors (KNN), AdaBoost, light gradient boosting machine (LightGBM) and categorical boosting (CatBoost). For models implemented through the caret framework, a fixed set of outcome-stratified five-fold cross-validation folds ($\text{seed} = 520$) was used for tuning and comparison. XGBoost was evaluated using its native R interface with fixed hyperparameters over the same training folds, whereas LightGBM and CatBoost were fitted using their native R interfaces. Model-specific tuning grids, fixed settings, resampling details and available random-seed information are reported in Table SI.

Internal validation was used to assess model stability and identify preferred candidate models. For the three pre-selected models, bootstrap resampling ($B = 1,000$) was performed within the training set to quantify optimism. In each bootstrap sample, the apparent area under the curve (AUC) was compared with performance on the original training cohort,

and the mean optimism was subtracted from the apparent AUC of the full training-set model to obtain the optimism-corrected AUC. Final model selection considered bootstrap optimism and corrected performance together with held-out test-set discrimination and calibration, model complexity and clinical interpretability. For threshold-dependent classification metrics, the probability threshold was determined in the training set using the maximum Youden index and then fixed for application to the held-out test set; no threshold optimization was performed in the test set. Held-out discrimination was summarized by AUC with DeLong 95% confidence intervals (CIs). Calibration was assessed using the Brier score, calibration slope and calibration-in-the-large (CITL), and clinical utility was evaluated by decision curve analysis. Pairwise AUC comparisons with LR used DeLong's test.

Model interpretation was performed using SHAP. Training data were used as the SHAP background distribution, and global importance, beeswarm and dependence plots were generated for the final LR model. Dependence plots were prespecified for RV_V30 and BMI to illustrate the direction of the two continuous predictors with high global importance. Continuous predictor values in the dependence plots are displayed on the standardized model-input scale. For clinical implementation, a static nomogram was constructed from the final LR coefficients using the rms package (<https://cran.r-project.org/package=rms>), and a web-based dynamic calculator was generated using DynNom (<https://cran.r-project.org/package=DynNom>) and shiny (<https://cran.r-project.org/package=shiny>). The nomogram and web calculator were based on the same final LR model.

Statistical analysis. Continuous variables are presented as the mean $\pm$ standard deviation when normally distributed and as median (interquartile range) otherwise; categorical variables are presented as frequencies and percentages. Binary variables were coded according to their prespecified categories, and multilevel categorical variables were represented using indicator variables as required. Between-group comparisons were performed using the independent-samples t-test for normally distributed continuous variables and the Mann-Whitney U test for non-normally distributed continuous variables. χ^2 or Fisher's exact tests were used for categorical variables. All tests were two-sided, and $P < 0.05$ was considered to indicate a statistically significant difference.

Statistical analyses were performed using R software (version 4.5.2; Posit Software, PBC). The glmnet and Boruta packages were used for LASSO and Boruta feature selection, respectively (23-25). The caret framework was used for LR, SVM, GBM, NN, RF, KNN and AdaBoost. XGBoost, LightGBM and CatBoost were fitted with their native R interfaces in the final analysis. Model-specific tuning grids, fixed settings, resampling details and available random-seed information are summarized in Table SI.

Results

Baseline characteristics. A total of 324 patients with breast cancer who received post-operative RNI were included. Grade ≥ 2 ARE occurred in 35.8% (116/324) of the patients. The

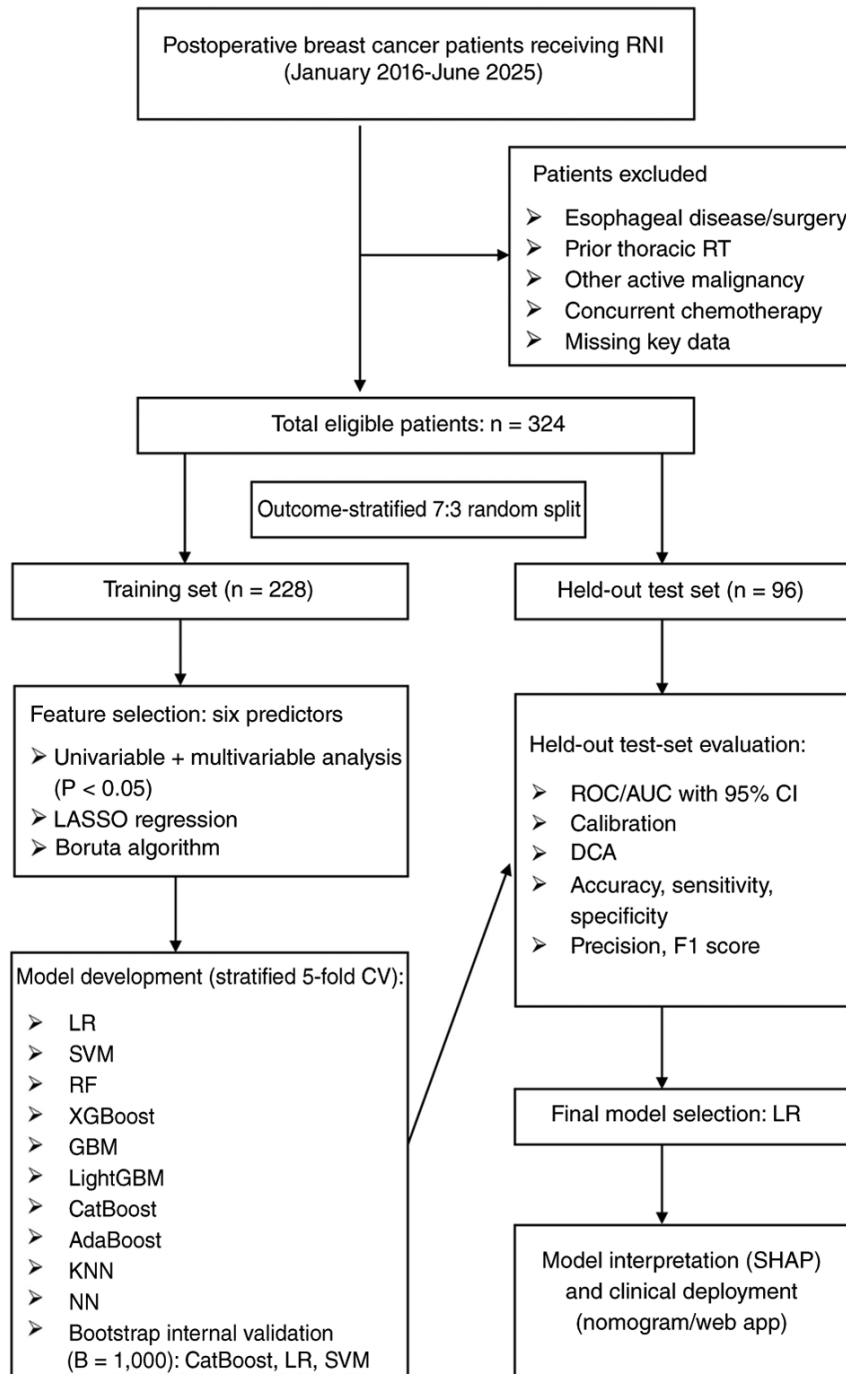


Figure 1. Study flowchart for development and evaluation of the prediction model for grade ≥ 2 acute radiation esophagitis. Eligible patients were divided by outcome-stratified 7:3 random sampling into a training set ($n=228$) and a held-out test set ($n=96$). Feature selection, five-fold CV and bootstrap internal validation were performed in the model-development workflow. Held-out performance was assessed using discrimination, calibration, threshold-dependent classification metrics and DCA. RNI, regional nodal irradiation; CV, cross-validation; DCA, decision curve analysis; RT, radiotherapy; LASSO, least absolute shrinkage and selection operator; ROC, receiver operating characteristic; AUC, area under the curve; LR, logistic regression; SVM, support vector machine; GBM, gradient boosting machine; NN, neural network; RF, random forest; XGBoost, extreme gradient boosting; KNN, K-Nearest Neighbors; LightGBM, light gradient boosting machine; CatBoost, categorical boosting; SHAP, SHapley Additive exPlanations.

cohort was divided by outcome-stratified random sampling into a training set ($n=228$; 70%) and a held-out test set ($n=96$; 30%). The study workflow is illustrated in Fig. 1.

In accordance with TRIPOD+AI (21), baseline characteristics are reported for both the training and held-out test sets (Table I). The two subsets were generally comparable for the majority of the variables. However, internal mammary node irradiation and several low-dose esophageal dose-volume

histogram (DVH) parameters, including RV_V10, RV_V15, AV_V10 and AV_V15, exhibited significant imbalances. None of these imbalanced variables was retained among the six consensus predictors.

Feature selection and predictor identification. In total, three feature-selection approaches were applied in the training set: Univariable analysis followed by multivariable logistic

Table I. Baseline characteristics of the training and held-out test sets.

Characteristics	Training (n=228)	Held-out test set (n=96)	P-value
Hepatic dysfunction			0.617
No	198 (86.84)	86 (89.58)	
Yes	30 (13.16)	10 (10.42)	
Pruritus			0.200
No	154 (67.54)	57 (59.38)	
Yes	74 (32.46)	39 (40.62)	
Hypertension			>0.999
No	191 (83.77)	81 (84.38)	
Yes	37 (16.23)	15 (15.62)	
Diabetes			>0.999
No	210 (92.11)	89 (92.71)	
Yes	18 (7.89)	7 (7.29)	
Family history			0.919
No	223 (97.81)	93 (96.88)	
Yes	5 (2.19)	3 (3.12)	
Tumor laterality			0.504
Left	132 (57.89)	51 (53.12)	
Right	96 (42.11)	45 (46.88)	
T stage			0.232
T1	49 (21.49)	17 (17.71)	
T2	147 (64.47)	61 (63.54)	
T3	22 (9.65)	16 (16.67)	
T4	10 (4.39)	2 (2.08)	
N stage			0.416
N0	21 (9.21)	5 (5.21)	
N1	97 (42.54)	49 (51.04)	
N2	56 (24.56)	20 (20.83)	
N3	54 (23.68)	22 (22.92)	
Lymphatic vessel infiltration			0.241
No	96 (42.11)	33 (34.38)	
Yes	132 (57.89)	63 (65.62)	
Histological grade			0.294
Grade 1-2	151 (66.23)	70 (72.92)	
Grade 3	77 (33.77)	26 (27.08)	
Molecular type			0.502
Luminal A	26 (11.40)	14 (14.58)	
Luminal B	102 (44.74)	48 (50.00)	
HER-2 positive	55 (24.12)	17 (17.71)	
Triple negative	45 (19.74)	17 (17.71)	
Surgical modality			0.776
Modified radical mastectomy	210 (92.11)	90 (93.75)	
Breast-conserving surgery	18 (7.89)	6 (6.25)	
Axillary lymph node dissection			0.827
No	10 (4.39)	3 (3.12)	
Yes	218 (95.61)	93 (96.88)	

Table I. Continued.

Characteristics	Training (n=228)	Held-out test set (n=96)	P-value
Targeted therapy			0.598
No	153 (67.11)	68 (70.83)	
Yes	75 (32.89)	28 (29.17)	
Endocrine therapy			0.689
No	83 (36.40)	32 (33.33)	
Yes	145 (63.60)	64 (66.67)	
Chemotherapy timing			0.430
Adjuvant chemotherapy	136 (59.65)	52 (54.17)	
Neoadjuvant therapy	92 (40.35)	44 (45.83)	
Tumor bed boost			0.881
No	211 (92.54)	90 (93.75)	
Yes	17 (7.46)	6 (6.25)	
Internal mammary lymph nodes irradiation			0.009
No	223 (97.81)	87 (90.62)	
Yes	5 (2.19)	9 (9.38)	
Radiotherapy technique			0.359
Hybrid IMRT and electron	81 (35.53)	40 (41.67)	
IMRT	147 (64.47)	56 (58.33)	
Age, years	54.21 (9.81)	53.34 (9.41)	0.461
Body mass index, kg/m ²	24.16 (3.48)	24.50 (3.33)	0.411
Lymph nodes, n removed	13.09 (7.31)	12.89 (6.69)	0.812
Positive lymph nodes, n	4.57 (6.64)	4.39 (4.98)	0.802
Neutrophil, x10 ⁹ /l	3.20 (1.51)	2.97 (1.42)	0.207
Lymphocyte, x10 ⁹ /l	1.28 (0.49)	1.27 (0.47)	0.830
Monocyte, x10 ⁹ /l	0.38 (0.22)	0.36 (0.15)	0.465
Platelet, x10 ⁹ /l	192.66 (62.01)	186.44 (63.12) (62.01)	0.412
Albumin, g/l	44.19 (3.66)	43.61 (3.43)	0.180
Neutrophil-to-lymphocyte ratio	2.74 (1.42)	2.62 (1.65)	0.516
Platelet-to-lymphocyte ratio	168.50 (78.72)	168.92 (88.75)	0.966
Lymphocyte-to-monocyte ratio	4.31 (6.12)	4.66 (5.16)	0.624
Systemic inflammation response index	1.09 (1.01)	0.99 (0.93)	0.441
Systemic immune-inflammation index	526.13 (302.12)	500.42 (360.49)	0.510
Total dose, Gy	50.34 (3.06)	50.24 (2.61)	0.761

Table I. Continued.

Characteristics	Training (n=228)	Held-out test set (n=96)	P-value
Radiotherapy days,	24.61 (2.77)	24.61 (2.30)	0.999
Esophageal volume, cm ³	12.50 (6.39)	12.43 (6.30)	0.931
Maximum esophageal dose, Gy	49.87 (6.80)	49.98 (6.94)	0.892
Mean esophageal dose, Gy	16.22 (7.40)	17.40 (7.46)	0.190
RV_V10, %	46.79 (21.74)	53.33 (24.52)	0.018
RV_V15, %	39.42 (19.97)	44.76 (22.18)	0.034
RV_V20, %	33.89 (18.77)	37.63 (20.29)	0.111
RV_V25, %	29.03 (17.38)	31.36 (18.89)	0.284
RV_V30, %	24.14 (16.36)	25.82 (17.87)	0.413
RV_V35, %	18.93 (15.17)	19.77 (15.59)	0.651
RV_V40, %	13.95 (13.35)	14.36 (13.33)	0.798
AV_V10, cm ³	5.70 (3.94)	6.74 (5.12)	0.049
AV_V15, cm ³	4.69 (2.97)	5.51 (3.79)	0.038
AV_V20, cm ³	3.96 (2.51)	4.46 (2.76)	0.109
AV_V25, cm ³	3.33 (2.19)	3.59 (2.18)	0.335
AV_V30, cm ³	2.73 (1.96)	2.88 (1.93)	0.530
AV_V35, cm ³	2.10 (1.69)	2.18 (1.69)	0.708
AV_V40, cm ³	1.55 (1.46)	1.56 (1.40)	0.948

Data are presented as n (%) for categorical variables and mean (SD) for continuous variables. P-values compare the training and held-out test sets. IMRT, intensity-modulated radiotherapy; RV, relative volume (%); AV, absolute volume (cm³); SD, standard deviation.

regression, LASSO with 10-fold cross-validation (λ 1SE=0.069) and Boruta. The corresponding reproducibility settings are described in the aforementioned section in materials and methods.

Multivariable LR was used for association screening in the training set; the resulting coefficients were interpreted as predictive associations rather than causal effects. LASSO retained five variables with non-zero coefficients: Pre-existing pruritus, tumor laterality, histological grade, BMI and diabetes. Boruta retained 10 features: Pre-existing pruritus, tumor laterality, histological grade, BMI, RV_V25, RV_V30, RV_V40, AV_V30, AV_V35 and AV_V40. The results of univariable and multivariable LR analyses for the candidate predictors are presented in Table II.

After integrating the three selection approaches and accounting for collinearity among related DVH and inflammatory variables, six consensus predictors were retained: Pre-existing pruritus, histological grade, BMI, diabetes, tumor laterality and esophageal RV_V30. In the present study, RV_V30 denotes the percentage of the contoured esophageal volume receiving ≥ 30 Gy; it emerged from the multi-strategy feature-selection process among the candidate DVH metrics.

The selection process is illustrated in Fig. 2, and univariable discrimination and correlation analyses are provided in Fig. S3.

Model development and performance evaluation. Using the six consensus predictors, 10 ML models were developed. The models implemented in caret were evaluated using the same outcome-stratified five-fold cross-validation folds within the training set. Based on internal performance, stability considerations and parsimony, CatBoost, LR and SVM were carried forward for 1,000-resample bootstrap internal validation.

The bootstrap internal-validation results are presented in Table III. The optimism-corrected AUC values were 0.804 for CatBoost, 0.744 for SVM and 0.715 for logistic regression. The corresponding mean optimism values were 0.113, 0.089 and 0.027, respectively. Thus, LR exhibited substantially lower bootstrap optimism than the two more complex pre-selected models, although its optimism-corrected AUC was not the highest.

Threshold-dependent held-out test-set metrics are summarized in Table IV. Using the threshold determined in the training set (0.396), LR achieved an accuracy of 0.729, sensitivity of 0.735, a specificity of 0.726, a precision of 0.595 and a F1 score of 0.658. Performance profiles varied across the other algorithms, and no single model was uniformly superior across all threshold-dependent metrics.

The discrimination and calibration results for all 10 models are reported in Table V. LR achieved the highest held-out AUC value of 0.797 (95% CI, 0.706-0.889), followed by SVM (0.791; 95% CI, 0.699-0.883) and RF (0.786; 95% CI, 0.693-0.878). CatBoost achieved a held-out AUC value of 0.762 (95% CI, 0.659-0.865), as shown in Fig. 3A. In the final 10-model comparison, CatBoost exhibited a training-cohort AUC of 0.923 (95% CI, 0.887-0.959; Fig. S4A). The held-out set was inspected during candidate-model evaluation and served as the early-stopping monitor for LightGBM. Accordingly, these estimates are internal test-set results and should not be interpreted as independent validation performance. For LR, the Brier score was 0.179, the calibration slope was 1.258 and CITL was -0.056. These values indicate generally acceptable calibration with a modest deviation from the ideal slope of 1. Decision curve analysis revealed that LR provided greater net benefit than both treat-all and treat-none strategies across threshold probabilities of 10-66% (Fig. 3B). Considering bootstrap stability, held-out discrimination and calibration, parsimony and clinical interpretability, LR was selected as the final model. The test-set receiver operating characteristic, decision-curve and calibration results are illustrated in Fig. 3, and the corresponding training-cohort plots are provided in Fig. S4.

Model interpretation. SHAP analysis was used to interpret the final LR model (Fig. 4A-D). Histological grade exhibited the highest global importance, followed by RV_V30 and BMI; pre-existing pruritus, tumor laterality and diabetes also contributed to model predictions (Fig. 4A). The beeswarm plot exhibited positive SHAP contributions for a higher histological grade, higher RV_V30 and a higher BMI (Fig. 4B). Left-sided laterality also tended to contribute toward a higher predicted risk, whereas the encoded higher

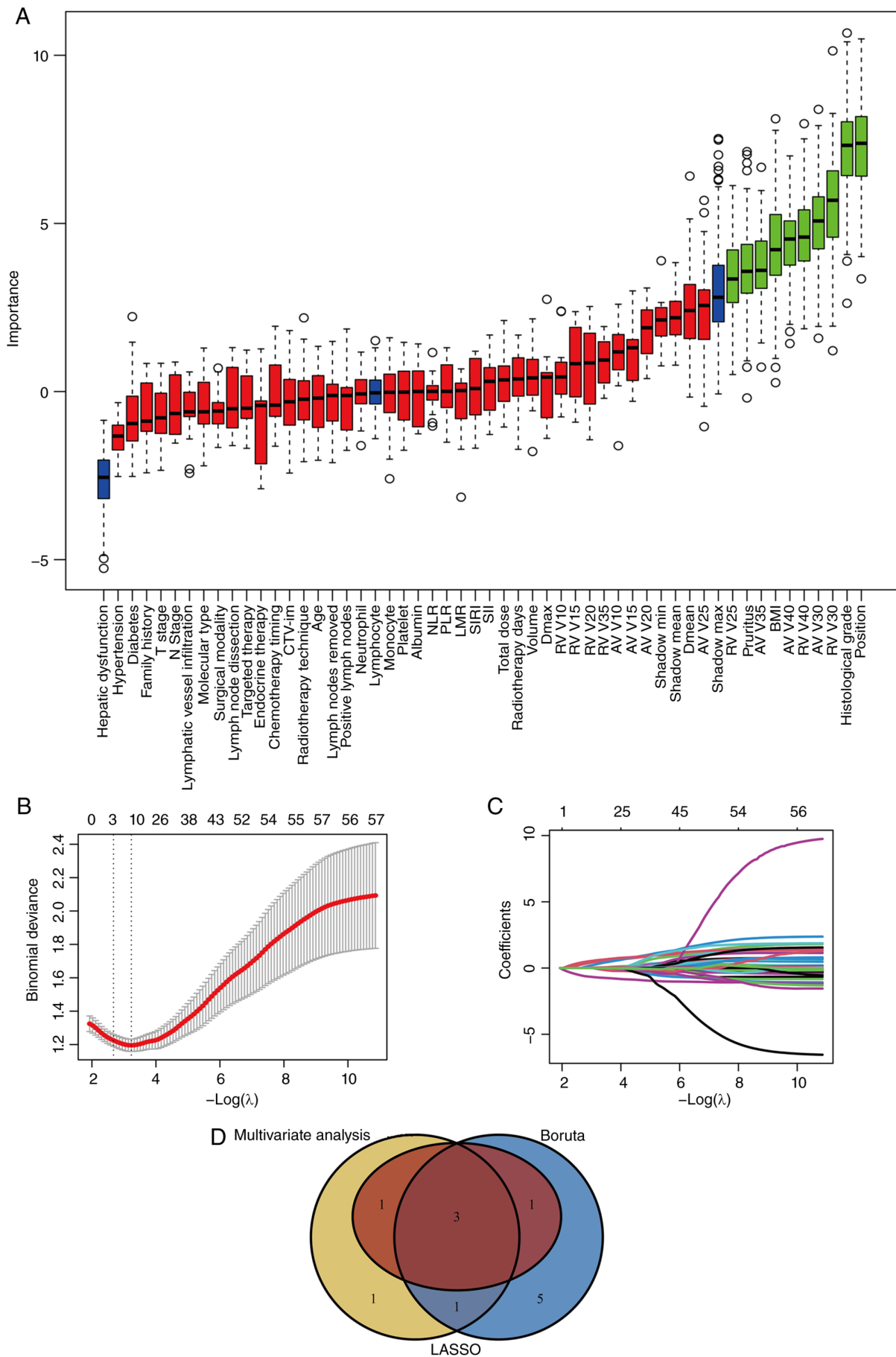


Figure 2. Feature selection and identification of the six consensus predictors. (A) Variable-importance ranking from the Boruta algorithm. (B) A 10-fold cross-validation curve used to select the LASSO penalty parameter. (C) LASSO coefficient profiles across the penalty path. (D) Intersection of variables identified by multivariable logistic regression, LASSO and Boruta. LASSO, least absolute shrinkage and selection operator.

Table II. Univariable and multivariable logistic regression analyses of candidate predictors in the training set.

Variable	No (n=146)	Yes (n=82)	OR (univariable)	OR (multivariable)
Hepatic dysfunction				
No	127 (87.0%)	71 (86.6%)		
Yes	19 (13.0%)	11 (13.4%)	1.04 (0.47-2.30; P=0.931)	
Pruritus				
No	107 (73.3%)	47 (57.3%)		
Yes	39 (26.7%)	35 (42.7%)	2.04 (1.15-3.62; P=0.014)	2.02 (1.03-3.98; P=0.042)
Hypertension				
No	123 (84.2%)	68 (82.9%)		
Yes	23 (15.8%)	14 (17.1%)	1.10 (0.53-2.28; P=0.795)	
Diabetes				
No	140 (95.9)	70 (85.4%)		
Yes	6 (4.1 %)	12 (14.6%)	4.00 (1.44-11.11; P=0.008)	3.80 (1.14-12.61; P=0.029)
Family history				
No	142 (97.3%)	81 (98.8%)		
Yes	4 (2.7%)	1 (1.2%)	0.44 (0.05-3.99; P=0.464)	
Tumor laterality				
Left	73 (50.0%)	59 (72.0%)		
Right	73 (50.0%)	23 (28.0%)	0.39 (0.22-0.70; P=0.002)	0.64 (0.31-1.32; P=0.228)
T stage				
T1	33 (22.6%)	16 (19.5%)		
T2	94 (64.4%)	53 (64.6%)	1.16 (0.59-2.31; P=0.666)	
T3	14 (9.6%)	8 (9.8%)	1.18 (0.41-3.38; P=0.760)	
T4	5 (3.4%)	5 (6.1%)	2.06 (0.52-8.16; P=0.302)	
N stage				
N0	15 (10.3%)	6 (7.3%)		
N1	65 (44.5%)	32 (39.0%)	1.23 (0.44-3.47; P=0.695)	
N2	35 (24%)	21 (25.6%)	1.50 (0.50-4.46; P=0.466)	
N3	31 (21.2%)	23 (28.0%)	1.85 (0.62-5.51; P=0.267)	
Lymphatic vessel infiltration				
No	66 (45.2%)	30 (36.6%)		
Yes	80 (54.8%)	52 (63.4%)	1.43 (0.82-2.49; P=0.207)	
Histological grade				
Grade 1-2	111 (76.0%)	40 (48.8%)		
Grade 3	35 (24.0%)	42 (51.2%)	3.33 (1.87-5.92; P<0.001)	3.49 (1.79-6.83; P<0.001)
Molecular type				
Luminal A	19 (13.0%)	7 (8.5%)		
Luminal B	60 (41.1%)	42 (51.2%)	1.90 (0.73-4.92; P=0.186)	
HER-2 positive	39 (26.7%)	16 (19.5%)	1.11 (0.39-3.16; P=0.840)	
Triple negative	28 (19.2%)	17 (20.7%)	1.65 (0.57-4.74; P=0.354)	
Surgical modality				
Modified radical mastectomy	135 (92.5%)	75 (91.5%)		
Breast-conserving surgery	11 (7.5%)	7 (8.5%)	1.15 (0.43-3.08; P=0.788)	
Axillary lymph node dissection				
No	6 (4.1 %)	4 (4.9%)		
Yes	140 (95.9%)	78 (95.1%)	0.84 (0.23-3.05; P=0.786)	
Targeted therapy				
No	93 (63.7%)	60 (73.2%)		
Yes	53 (36.3%)	22 (26.8%)	0.64 (0.36-1.16; P=0.145)	

Table II. Continued.

Variable	No (n=146)	Yes (n=82)	OR (univariable)	OR (multivariable)
Endocrine therapy				
No	56 (38.4%)	27 (32.9%)		
Yes	90 (61.6%)	55 (67.1%)	1.27 (0.72-2.24; P=0.414)	
Chemotherapy timing				
Adjuvant chemotherapy	83 (56.8%)	53 (64.6%)		
Neoadjuvant therapy	63 (43.2%)	29 (35.4%)	0.72 (0.41-1.26; P=0.251)	
Tumor bed boost				
No	136 (93.2%)	75 (91.5%)		
Yes	10 (6.8%)	7 (8.5%)	1.27 (0.46-3.47; P=0.642)	
Internal mammary lymph node irradiation				
No	143 (97.9%)	80 (97.6)		
Yes	3 (2.1%)	2 (2.4%)	1.19 (0.20-7.28; P=0.849)	
Radiotherapy technique				
Hybrid IMRT and electron	49 (33.6%)	32 (39%)		
IMRT	97 (66.4%)	50 (61%)	0.79 (0.45-1.38; P=0.409)	
Age, years	53.6±9.9	55.2±9.6	1.02 (0.99-1.05; P=0.241)	
BMI, kg/m ²	23.7±3.5	24.9±3.3	1.10 (1.02-1.19; P=0.017)	1.12 (1.02-1.23; P=0.014)
Lymph nodes removed, n	13.2±8.1	12.9±5.7	0.99 (0.96-1.03; P=0.769)	
Positive lymph nodes, n	4.4±7.2	4.8±5.5	1.01 (0.97-1.05; P=0.649)	
Neutrophil, ×10 ⁹ /l	3.1±1.3	3.3±1.9	1.10 (0.92-1.31; P=0.307)	
Lymphocyte, ×10 ⁹ /l	1.3±0.5	1.3±0.5	1.11 (0.64-1.92; P=0.720)	
Monocyte, ×10 ⁹ /l	0.4±0.2	0.4±0.2	0.52 (0.13-2.10; P=0.362)	
Platelet, ×10 ⁹ /l	195.3±62.4	187.9±61.4	1.00 (0.99-1.00; P=0.387)	
Albumin, g/l	44.3±3.9	44.0±3.3	0.97 (0.90-1.05; P=0.506)	
Neutrophil-to-lymphocyte ratio	2.7±1.4	2.7±1.4	0.99 (0.82-1.20; P=0.939)	
Platelet-to-lymphocyte ratio	173.1±81.5	160.3±73.2	1.00 (0.99-1.00; P=0.242)	
Lymphocyte-to-monocyte ratio	3.9±1.9	5.1±9.9	1.05 (0.95-1.17; P=0.352)	
Systemic inflammation response index	1.1±0.9	1.1±1.2	1.01 (0.78-1.32; P=0.929)	
Systemic immune-inflammation index	532.8±304.6	514.3±299.1	1.00 (1.00-1.00; P=0.658)	
Total dose, Gy	50.3±3.0	50.4±3.1	1.02 (0.93-1.11; P=0.712)	
Radiotherapy days, days	24.5±2.9	24.9±2.5	1.05 (0.95-1.17; P=0.330)	
Esophageal volume, cm ³	12.8±6.8	12.0±5.5	0.98 (0.94-1.03; P=0.401)	
Maximum esophageal dose, Gy	49.0±7.1	51.4±5.9	1.06 (1.01-1.11; P=0.014)	1.07 (0.99-1.15; P=0.074)
Mean esophageal dose, Gy	15.3±7.2	17.9±7.5	1.05 (1.01-1.09; P=0.013)	0.91 (0.75-1.11; P=0.367)
RV_V10, %	45.4±21.1	49.2±22.7	1.01 (1.00-1.02; P=0.209)	
RV_V15, %	37.3±18.9	43.2±21.4	1.01 (1.00-1.03; P=0.034)	1.01 (0.92-1.11; P=0.812)
RV_V20, %	31.5±17.9	38.2±19.6	1.02 (1.00-1.03; P=0.011)	1.04 (0.89-1.22; P=0.584)
RV_V25, %	26.6±16.9	33.4±17.5	1.02 (1.01-1.04; P=0.005)	0.92 (0.75-1.13; P=0.423)
RV_V30, %	21.7±16.1	28.5±16.0	1.03 (1.01-1.04; P=0.003)	1.36 (1.06-1.74; P=0.016)
RV_V35, %	16.8±14.7	22.7±15.4	1.03 (1.01-1.04; P=0.006)	0.73 (0.55-0.96; P=0.024)
RV_V40, %	12.2±12.5	17.1±14.2	1.03 (1.01-1.05; P=0.008)	1.09 (0.91-1.29; P=0.343)
AV_V10, cm ³	5.7±4.3	5.6±3.3	0.99 (0.93-1.07; P=0.859)	
AV_V15, cm ³	4.6±3.1	4.9±2.7	1.04 (0.95-1.13; P=0.444)	
AV_V20, cm ³	3.8±2.6	4.3±2.3	1.08 (0.97-1.20; P=0.172)	
AV_V25, cm ³	3.1±2.3	3.7±1.9	1.12 (0.99-1.27; P=0.077)	
AV_V30, cm ³	2.5±2.1	3.1±1.7	1.17 (1.02-1.35; P=0.028)	0.29 (0.07-1.19; P=0.086)

Table II. Continued.

Variable	No (n=146)	Yes (n=82)	OR (univariable)	OR (multivariable)
AV_V35, cm ³	1.9±1.7	2.5±1.7	1.23 (1.05-1.44; P=0.013)	3.93 (0.29-52.43; P=0.300)
AV_V40, cm ³	1.4±1.4	1.9±1.5	1.27 (1.06-1.53; P=0.012)	0.90 (0.15-5.59; P=0.914)

Outcome: Grade ≥ 2 acute radiation esophagitis. Data are presented as n (%) for categorical variables and mean $\pm$ SD for continuous variables. Blank cells indicate variables that were not retained in the final multivariable model. OR, odds ratio; CI, confidence interval; IMRT, intensity-modulated radiotherapy; RV, relative volume; AV, absolute volume; SD, standard deviation.

Table III. Bootstrap internal validation (B=1,000) for the three pre-selected models in the training cohort.

Model	Apparent AUC	Mean optimism (SD)	Optimism-corrected AUC (95% CI)	95% percentile interval for bootstrap-refit AUC on original training set
Categorical boosting	0.917	0.113 (0.018)	0.804 (0.764-0.838)	0.822-0.886
Support vector machine	0.833	0.089 (0.025)	0.744 (0.698-0.795)	0.745-0.815
Logistic regression	0.741	0.027 (0.034)	0.715 (0.650-0.778)	0.708-0.741

Optimism-corrected AUC was computed as the apparent AUC minus the mean optimism across bootstrap samples. The 95% CI was derived from the percentile distribution of corrected AUC values across bootstrap resamples (B=1,000). Bootstrap refit AUC was obtained by fitting the model in each bootstrap sample and evaluating it on the original training data ('test-on-original'); the 95% CI was computed using percentile intervals. AUC, area under the receiver operating characteristic curve; CI, confidence interval; SD, standard deviation.

Table IV. Threshold-dependent classification metrics for the 10 machine-learning models in the held-out test set.

Model	Threshold	Accuracy	Sensitivity	Specificity	Precision	F1 score
Logistic regression	0.396	0.729	0.735	0.726	0.595	0.658
Support vector machine	0.331	0.729	0.794	0.694	0.587	0.675
Gradient boosting machine	0.346	0.688	0.676	0.694	0.548	0.605
Neural network	0.386	0.677	0.735	0.645	0.532	0.617
Random forest	0.301	0.708	0.794	0.661	0.562	0.659
Extreme gradient boosting	0.370	0.719	0.794	0.677	0.574	0.667
K-nearest neighbors	0.441	0.646	0.559	0.694	0.500	0.528
AdaBoost	0.437	0.719	0.765	0.694	0.578	0.658
Light gradient boosting machine	0.366	0.719	0.824	0.661	0.571	0.675
Categorical boosting	0.361	0.719	0.853	0.645	0.569	0.682

Thresholds were determined in the training set using the maximum Youden index and were then fixed for application to the held-out test set; no threshold optimization was performed in the held-out test set.

value for laterality (right) tended to contribute negatively. Dependence plots for RV_V30 and BMI further illustrated the direction and distribution of these continuous feature effects (Fig. 4C and D).

Visual nomogram and clinical deployment. To support clinical interpretation, the final LR model was presented as a static nomogram (Fig. 5A). The nomogram incorporates the six consensus predictors: Pre-existing pruritus, histological grade, BMI, diabetes, tumor laterality and esophageal RV_V30. Each predictor contributes points according to its regression

coefficient, and the summed score is mapped to the predicted probability of grade ≥ 2 ARE.

An open-access web-based calculator was also developed for point-of-care risk estimation (Fig. 5B). The calculator uses the same final LR model as the static nomogram. A numerical consistency check confirmed concordant predicted probabilities between the nomogram implementation and the final LR model (correlation=1.000; maximum absolute difference, 1.16×10^{-8}). The web application is accessible at: <https://hebin7290.shinyapps.io/Radiation-esophagitis/>. As an illustrative example, a hypothetical patient with pre-existing

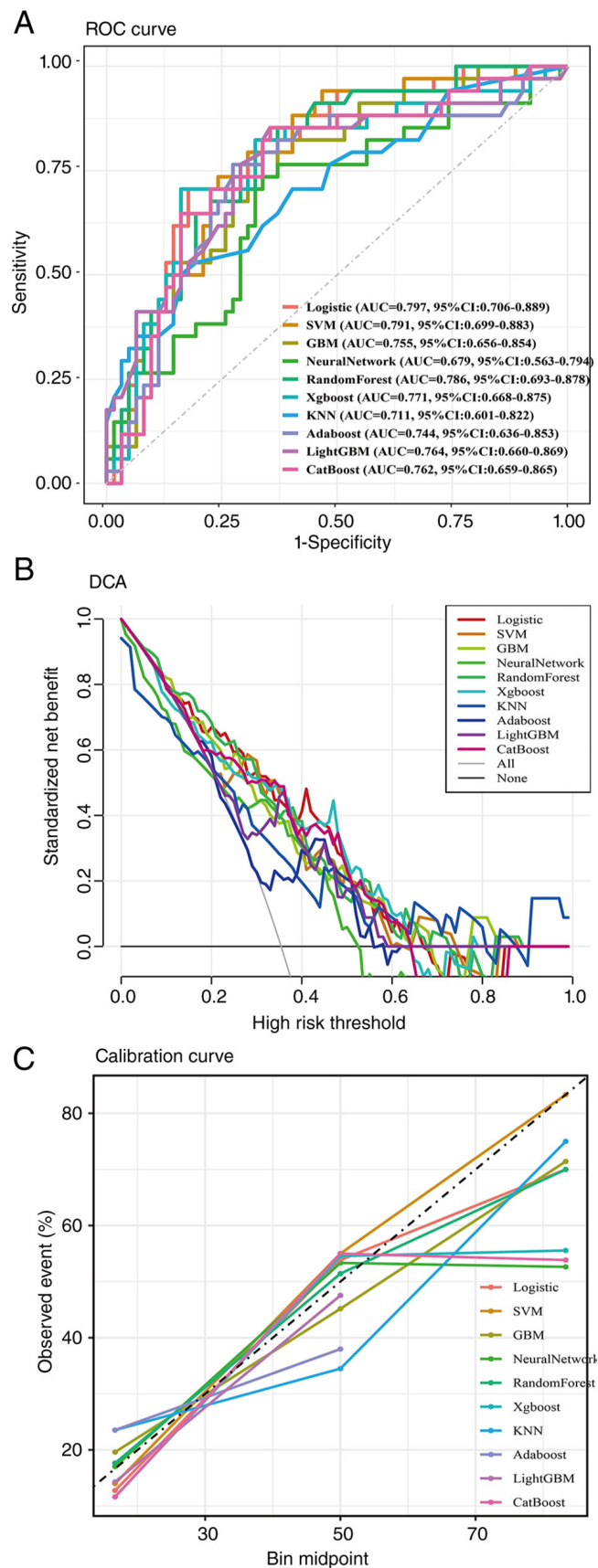


Figure 3. Performance of the prediction models in the held-out test set. (A) ROC curves for the 10 machine-learning algorithms; LR achieved an AUC of 0.797 (95% CI, 0.706-0.889). CatBoost achieved an AUC of 0.762 (95% CI, 0.659-0.865). (B) DCA calculated from the established predicted risks. LR showed greater net benefit than treat-all and treat-none strategies across threshold probabilities of 10-66%. (C) Calibration curves for the 10 models. For LR, the Brier score was 0.179, calibration slope was 1.258, and calibration-in-the-large was -0.056. ROC, receiver operating characteristic; AUC, area under the curve; CI, confidence interval; LR, logistic regression; DCA, decision curve analysis; LR, logistic regression; SVM, support vector machine; GBM, gradient boosting machine; XGBoost, extreme gradient boosting; KNN, K-Nearest Neighbors; LightGBM, light gradient boosting machine; CatBoost, categorical boosting; CI, confidence interval.

Table V. Discrimination and calibration performance of the 10 machine-learning models in the held-out test set.

Model	AUC (95% CI, DeLong)	Brier score	Calibration slope	CITL
Logistic regression	0.797 (0.706-0.889)	0.179	1.258	-0.056
Support vector machine	0.791 (0.699-0.883)	0.188	1.530	-0.114
Random forest	0.786 (0.693-0.878)	0.181	0.936	0.153
Extreme gradient boosting	0.771 (0.668-0.875)	0.184	1.044	-0.090
Light gradient boosting machine	0.764 (0.660-0.869)	0.199	2.773	-0.068
Categorical boosting	0.762 (0.659-0.865)	0.189	0.894	-0.184
Gradient boosting machine	0.755 (0.656-0.854)	0.189	0.898	-0.021
AdaBoost	0.744 (0.636-0.853)	0.210	1.676	-0.285
K-nearest neighbors	0.711 (0.601-0.822)	0.211	0.165	0.086
Neural network	0.679 (0.563-0.794)	0.232	0.216	-0.313

AUC 95% CIs were calculated using DeLong's method. Lower Brier scores indicate better overall prediction accuracy. The ideal calibration slope is 1 and the ideal CITL is 0; a slope >1 indicates under-dispersed predicted risks, whereas a slope <1 indicates predictions that are too extreme. AUC, area under the receiver operating characteristic curve; CI, confidence interval; CITL, calibration-in-the-large.

pruritus, grade 3 histology, BMI 25 kg/m², no diabetes, a left-sided tumor and RV_V30 of 20% would have a predicted probability of grade ≥ 2 ARE of ~ 0.62 .

Discussion

The present study developed an interpretable clinical-dosimetric model for grade ≥ 2 ARE following RNI in patients with breast cancer. A total of six predictors captured baseline symptoms, tumor characteristics, metabolic status, anatomy and esophageal dose exposure. Bootstrap internal validation revealed that the more flexible pre-selected models had greater optimism, whereas LR had the lowest mean optimism. Final model selection therefore considered stability together with held-out discrimination and calibration, model complexity and interpretability rather than relying on a single performance metric. The held-out AUC value of LR was 0.797, with a Brier score of 0.179 and modest calibration deviation. These findings support LR as a clinically transparent candidate for further external validation. The selected variables should be interpreted as predictive markers rather than causal determinants of ARE.

Histological grade exhibited the strongest global contribution in the SHAP analysis and was associated with the risk of developing ARE in the predictive model. This association should not be interpreted as evidence that tumor grade directly alters esophageal radiosensitivity. High-grade breast cancers can exhibit hypoxia-related and inflammatory phenotypes (26), and hypoxia and inflammatory signaling can interact through pathways, including hypoxia-inducible factor and NF- κ B (27). Systemic inflammatory indices have also been evaluated in radiation-esophagitis prediction in other thoracic malignancies (28-30). These observations provide a possible biological context; however, any association between tumor grade and host normal-tissue susceptibility remains hypothesis-generating and requires dedicated biological and prospective validation.

RV_V30 was retained as the principal dosimetric predictor, supporting the relevance of intermediate-dose esophageal

exposure. Prospective evidence in breast RNI has also identified clinically relevant esophageal dose-volume constraints. For example, Wang *et al* (11) reported RV_V30 and AV_V30 constraints in conventionally fractionated RNI, whereas Ji *et al* (9) reported associations for AV30 and AV35 in a smaller cohort. In the present study, RV_V30 was selected from a broader candidate set that included Dmean, Dmax and multiple relative and absolute dose-volume metrics. As RV_V30 is modifiable during planning, reducing intermediate-dose esophageal exposure while maintaining target coverage is a plausible optimization strategy. Since no universal RV_V30 cut-off was derived in the present study, this finding should be interpreted as supporting dose minimization rather than a fixed planning constraint.

BMI and diabetes were both retained in the final prediction model. Their contributions may reflect aspects of metabolic status, inflammation, vascular health or treatment geometry rather than a single causal pathway. Obesity is associated with chronic inflammation and oxidative stress (31), and diabetes can impair microvascular function (32). A higher BMI can also affect radiotherapy geometry and planning complexity (33). These mechanisms provide plausible context for the observed predictive associations but were not directly measured in the present study.

Pre-existing pruritus was retained as a predictive marker. By contrast to a formal patient-reported outcome measure, pruritus was retrospectively documented as a binary present/absent variable before radiotherapy. Previous research has linked normal-tissue reactions to inter-individual susceptibility (34), while taxane-related skin toxicity and neuroimmune mechanisms may contribute to persistent pruritus (35-37). A recent breast-radiotherapy study also identified pruritus as a contributor to esophagitis prediction using SHAP (38). Although these observations support further investigation, they do not establish pruritus as a mechanistic marker of radiosensitivity.

A possible explanation is that residual treatment-related symptoms may reflect a host state that also modifies tolerance

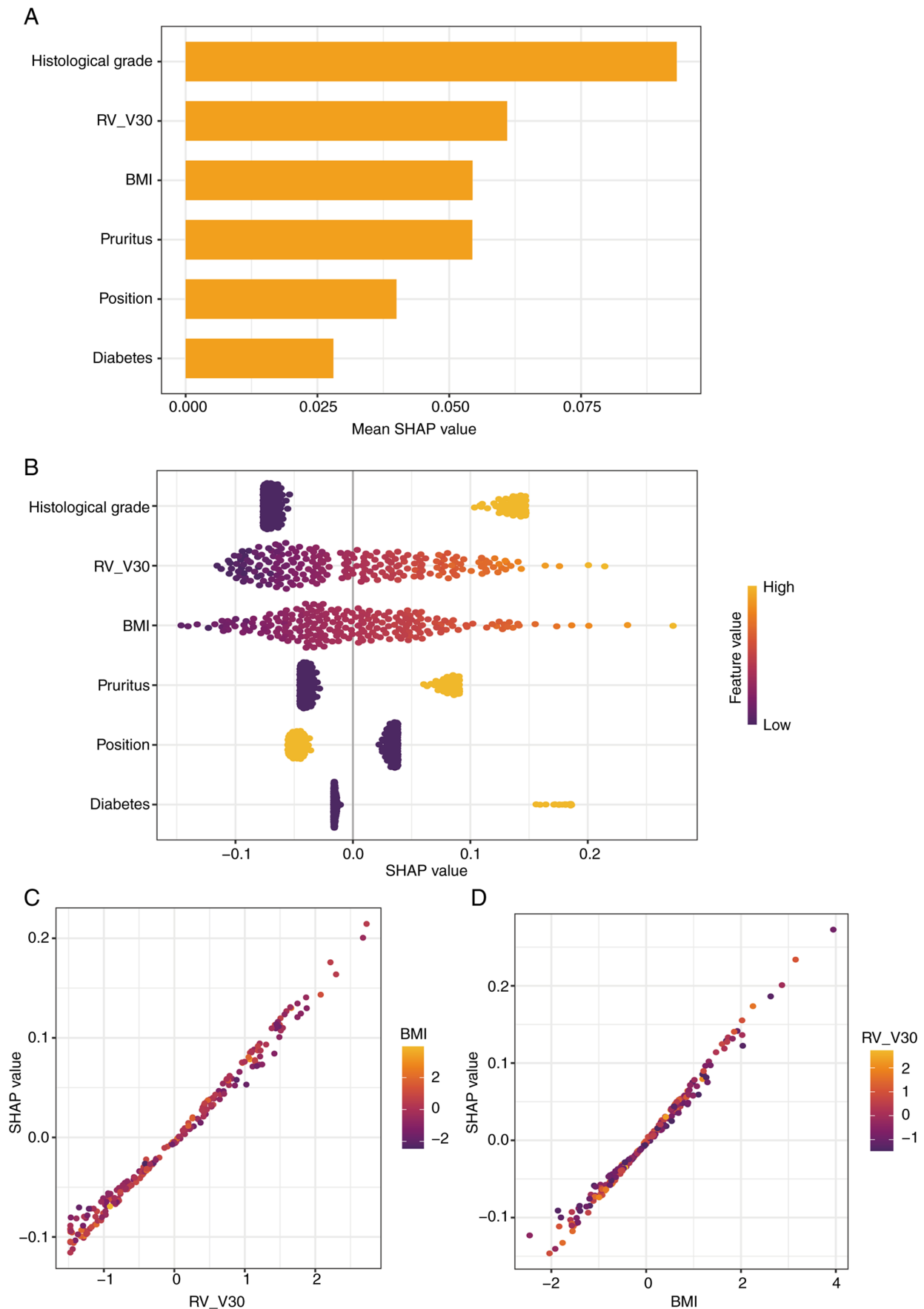


Figure 4. SHAP interpretation of the final Logistic Regression model. (A) Global feature-importance bar plot ranked by mean absolute SHAP value. (B) Beeswarm plot showing the direction and magnitude of feature contributions; color represents the encoded feature value from low to high. (C) SHAP dependence plot for RV_V30. (D) SHAP dependence plot for BMI. For tumor laterality, the higher encoded value corresponds to right and the lower value to left. In panels C and D, RV_V30 and BMI are displayed on the standardized model-input scale. SHAP, SHapley Additive exPlanations; BMI, body mass index; RV_V30, relative esophageal volume receiving ≥ 30 Gy.

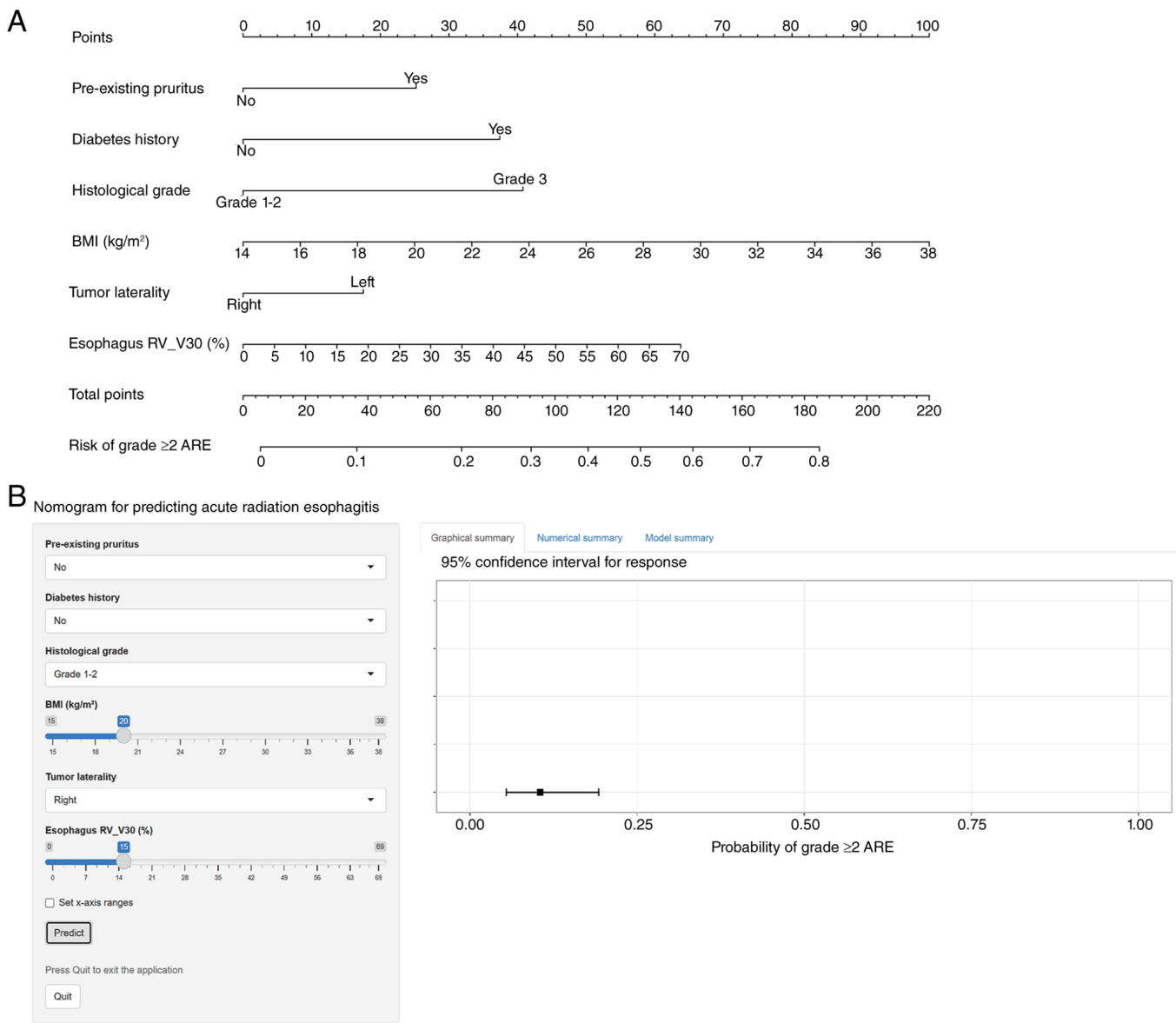


Figure 5. Visualization and clinical deployment of the final Logistic Regression model. (A) Static nomogram for predicting grade ≥ 2 acute radiation esophagitis from pre-existing pruritus, histological grade, BMI, diabetes, tumor laterality and esophageal RV_V30. (B) Web-based calculator using the same final Logistic Regression model for individualized risk estimation. BMI, body mass index; RV_V30, relative esophageal volume receiving ≥ 30 Gy.

to subsequent irradiation. Taxane-related cutaneous toxicity and radiation-induced mucosal injury involve inflammatory and epithelial-injury pathways (35,39-41); however, the data of the present study do not establish a shared mechanism or a causal association between pre-existing pruritus and esophageal radiosensitivity. This interpretation is therefore hypothesis-generating and requires prospective validation. The binary ascertainment used in the present retrospective cohort also did not capture pruritus severity, duration or patient-reported burden. Future studies using standardized grading and patient-reported assessment rather than a simple present/absent variable are warranted.

Tumor laterality was retained as a model predictor. In the current coding, right was the higher numeric category and had a negative model coefficient relative to left, consistent with a greater contribution from left-sided laterality in the SHAP analysis. This association may partly reflect treatment

geometry, particularly when internal mammary nodal irradiation places beam paths closer to the esophagus. Techniques, such as deep-inspiration breath-hold can modify organ-at-risk geometry and dose exposure (42). Laterality should therefore be regarded as an anatomical prediction marker rather than a causal determinant of ARE.

Published breast cancer RNI studies have largely emphasized dosimetric predictors, with heterogeneous sample sizes, fractionation schedules and validation strategies. Calibration and decision-analytic measures have not been reported consistently, which can make apparent model performance difficult to compare across cohorts.

Yaney *et al* (8) analyzed 531 conventionally fractionated RNI cases and proposed practical dose cut-points, including mean esophageal dose, V10 and V20. Wang *et al* (43) prospectively evaluated esophageal constraints in hypofractionated RNI and reported moderate discrimination for parsimonious

models. Ji *et al* (9) examined a smaller hypofractionated cohort and reported high apparent discrimination for AV30 and AV35. These studies illustrate both the clinical relevance of esophageal dosimetry and the sensitivity of reported performance to cohort size and validation design.

The results of the present study add a mixed clinical-dosimetric model that incorporates RV_V30, BMI, diabetes, laterality and a baseline symptom. The comparison of multiple algorithms also illustrates the potential for optimism in flexible learners when the sample size is modest. CatBoost, for example, had a high apparent AUC, but substantially greater bootstrap optimism than LR. A similar observation has been reported in radiotherapy-toxicity modeling outside the breast-RNI setting: In an externally validated head-and-neck dysphagia study, penalized LR using dose-volume metrics performed as well as support vector and RF models (external-validation AUC of 0.82), demonstrating that additional model complexity does not necessarily improve transportable performance (44). This does not imply that ensemble methods are intrinsically inferior; rather, higher-capacity models may be more vulnerable to optimism and calibration instability in limited datasets. Bootstrap methods are useful for quantifying this optimism, although internal validation does not replace external evaluation (45,46). When held-out discrimination is similar, calibration, stability, parsimony and interpretability are therefore key complementary selection criteria.

Internal validation was used to identify preferred candidates and quantify model optimism. Final model selection then considered bootstrap-corrected performance together with held-out test-set discrimination and calibration, model complexity and clinical interpretability. LR was selected as it provided a favorable balance across these criteria rather than because it was uniformly superior on every metric. The held-out test results should therefore be interpreted as internal held-out evaluation, not as external validation.

The model is intended to support pre-radiotherapy risk stratification rather than dictate treatment. Decision curve analysis revealed positive net benefit for LR over treat-all and treat-none strategies across threshold probabilities of 10-66%. In practice, in the event that the predicted risk of a patient exceeds a threshold judged clinically meaningful by the treating team, the plan could be reviewed for opportunities to reduce RV_V30, while maintaining target coverage, and the patient could receive closer symptom monitoring and earlier supportive care. The static nomogram and web-based calculator provide transparent tools for this risk estimate (Fig. 5).

The present study has several limitations which should be mentioned. First, its retrospective design may have introduced selection bias through eligibility criteria, complete-case inclusion and the availability of clinical and dosimetric records. Second, the single-center cohort and modest sample size may limit generalizability. Institutional differences in esophageal contouring, RNI technique, fractionation, treatment-planning systems, accelerators and supportive care may alter predictor distributions and model calibration. Third, calibration estimates from the held-out test set ($n=96$) remain imprecise. The calibration slope of 1.258 and CITL of -0.056 indicate modest deviation from ideal calibration. Fourth, pre-existing pruritus was recorded only as present or absent and did not

capture severity, duration or standardized patient-reported burden. Fifth, the held-out set was inspected during candidate-model evaluation and served as the early-stopping monitor for LightGBM. It therefore provided internal test-set evaluation rather than independent or external validation. Sixth, the present study did not formally compare the final clinical-dosimetric model with a clinical-only baseline model. The incremental value of RV_V30 therefore remains to be quantified. Prospective multicenter studies are thus required to validate the model across different patient populations, RNI techniques, fractionation schedules, planning systems and accelerators. Such studies should also evaluate recalibration and compare clinical-only and clinical-dosimetric models. This comparison should assess discrimination, calibration and decision-curve performance to quantify the incremental value of RV_V30.

In conclusion, an interpretable six-feature clinical-dosimetric model was developed and internally evaluated for grade ≥ 2 ARE after breast cancer RNI. The final LR model incorporates pre-existing pruritus, histological grade, BMI, diabetes, tumor laterality and esophageal RV_V30. It provided a favorable balance of discrimination, calibration, stability, parsimony and interpretability in the present cohort and was implemented as a nomogram and web-based calculator. Prospective multicenter external validation and, where necessary, recalibration are required before routine clinical implementation.

Acknowledgements

The authors wish to thank Dr Yong Diao (Department of Oncology, The General Hospital of Western Theater Command of the Chinese People's Liberation Army, Chengdu, China) for sharing clinical insights regarding patient management and toxicity assessment in radiotherapy. Their expertise contributed to the clinical perspective of the present study.

Funding

No funding was received.

Availability of data and materials

The data generated in the present study may be requested from the corresponding author.

Authors' contributions

BH and ZQ contributed equally to this work. BH and LZ conceived the study and designed the analytical framework. BH drafted the manuscript and coordinated the project. ZQ, LL and HL contributed to patient data collection, clinical variable assessment and model validation. XZ and YT performed the formal data analysis, developed the machine-learning models and software tools, and created the visualizations. HZ contributed to the acquisition and interpretation of clinical and radiotherapy data and critically revised the manuscript. FK contributed to study conception and design, interpretation of the results, critical revision of the manuscript and supervision of the study. LZ supervised

the study and provided critical methodological and clinical guidance. BH and ZQ confirm the authenticity of all the raw data. All authors read and approved the final version of the manuscript.

Ethics approval and consent to participate

The present study was approved by the Ethics Committee of The General Hospital of Western Theater Command (approval no. 2025EC11-ky019) and was performed in accordance with the ethical standards of the Declaration of Helsinki. Due to the retrospective nature of the study, the requirement for written informed consent was waived by the Ethics Committee of The General Hospital of Western Theater Command.

Patient consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Use of artificial intelligence tools

During the preparation of this work, artificial intelligence tools were used to improve the readability and language of the manuscript or to generate images, and subsequently, the authors revised and edited the content produced by the artificial intelligence tools as necessary, taking full responsibility for the ultimate content of the present manuscript.

References

- Sung H, Filho AM, Laversanne M, Ferlay J, Siegel RL, Soerjomataram I, Jemal A and Bray F: Global cancer statistics 2024: GLOBOCAN estimates of incidence and mortality worldwide for 34 cancers in 186 countries. *CA Cancer J Clin* 76: e70090, 2026.
- EBCTCG (Early Breast Cancer Trialists' Collaborative Group); McGale P, Taylor C, Correa C, Cutter D, Duane F, Ewertz M, Gray R, Mannu G, Peto R, *et al*: Effect of radiotherapy after mastectomy and axillary surgery on 10-year recurrence and 20-year breast cancer mortality: Meta-analysis of individual patient data for 8135 women in 22 randomised trials. *Lancet* 383: 2127-2135, 2014.
- Poortmans PM, Weltens C, Fortpied C, Kirkove C, Peignaux-Casasnovas K, Budach V, van der Leij F, Vonk E, Weidner N, Rivera S, *et al*: Internal mammary and medial supraclavicular lymph node chain irradiation in stage I-III breast cancer (EORTC 22922/10925): 15-year results of a randomised, phase 3 trial. *Lancet Oncol* 21: 1602-1610, 2020.
- Poortmans PM, Collette S, Kirkove C, Van Limbergen E, Budach V, Struikmans H, Collette L, Fourquet A, Maingon P, Valli M, *et al*: Internal mammary and medial supraclavicular irradiation in breast cancer. *N Engl J Med* 373: 317-327, 2015.
- West K, Schneider M, Wright C, Beldham-Collins R, Coburn N, Tiver K, Gebiski V and Stuart KE: Radiation-induced oesophagitis in breast cancer: Factors influencing onset and severity for patients receiving supraclavicular nodal irradiation. *J Med Imaging Radiat Oncol* 64: 113-119, 2020.
- Baker S and Fairchild A: Radiation-induced esophagitis in lung cancer. *Lung Cancer (Auckl)* 7: 119-127, 2016.
- Wang DQ, Zhang N, Dong LH, Zhong YH, Wu HF, Zhong QZ, Jin J, Hou XR, Jing H, Tang Y, *et al*: Dose-volume predictors for radiation esophagitis in patients with breast cancer undergoing hypofractionated regional nodal radiation therapy. *Int J Radiat Oncol Biol Phys* 117: 186-197, 2023.
- Yaney A, Ayan AS, Pan X, Jhawar S, Healy E, Beyer S, Lindsey K, Kuhn K, Tedrick K, White JR and Bazan JG: Dosimetric parameters associated with radiation-induced esophagitis in breast cancer patients undergoing regional nodal irradiation. *Radiother Oncol* 155: 167-173, 2021.
- Ji MC, Li ZJ, Li K, Wang YX, Yang B, Lv LL, Su Y, Zhang ZW, Huo ZC, Qi Q, *et al*: Dosimetric risk factors for radiation esophagitis in patients with breast cancer following regional nodal radiation. *World J Clin Cases* 12: 2995-3003, 2024.
- Bazan JG, Healy E, Beyer S, Kuhn K, DiCostanzo D, Smith TL, Jhawar S and White JR: Clinical effectiveness of an adaptive treatment planning algorithm for intensity modulated radiation therapy versus 3D conformal radiation therapy for node-positive breast cancer patients undergoing regional nodal irradiation/post-mastectomy radiation therapy. *Int J Radiat Oncol Biol Phys* 108: 1159-1171, 2020.
- Wang DQ, Shen DX, Han N, Lu YF, Yang WF, Tie J, Hou XR, Wang XH, Zhong YH, Yu XL, *et al*: Dosimetric and clinical predictors for radiation esophagitis in patients with breast cancer undergoing conventional fractionated regional nodal irradiation: A prospective study. *Int J Cancer* 158: 1055-1066, 2026.
- Shreya S, Pandey S, Ghosh DK, Puneet, Prasad SB, Grosset CF and Jain BP: Analysis of endoplasmic reticulum stress-associated proteins as prognostic markers in breast cancer. *Curr Genomics* 26: 621-632, 2025.
- Qiu J, Zhou T, Wang D, Hong W, Qian D, Meng X and Liu X: Pan-cancer analysis identifies AIMP2 as a potential biomarker for breast cancer. *Curr Genomics* 24: 307-329, 2023.
- Wang H, Hao R, Liu W, Zhang Y, Ma S, Lu Y, Hu J and Qi Y: Identification of transcription factors associated with the disease-free survival of triple-negative breast cancer through weighted gene co-expression network analysis. *CytoJournal* 21: 71, 2024.
- Shi Q, Xuan M, Lv C, Zhang Z, Geng X, Huang D and Hu X: Recent applications of artificial intelligence in cancer radiotherapy and immunotherapy: Current status and future directions. *Front Immunol* 17: 1776472, 2026.
- Shamout F, Zhu T and Clifton DA: Machine learning for clinical outcome prediction. *IEEE Rev Biomed Eng* 14: 116-126, 2021.
- Greener JG, Kandathil SM, Moffat L and Jones DT: A guide to machine learning for biologists. *Nat Rev Mol Cell Biol* 23: 40-55, 2022.
- Khene ZE, Bigot P, Doumerc N, Ouzaid I, Boissier R, Nouhaud FX, Albiges L, Bernhard JC, Ingels A, Borchicelli D, *et al*: Application of machine learning models to predict recurrence after surgical resection of nonmetastatic renal cell carcinoma. *Eur Urol Oncol* 6: 323-330, 2023.
- Yu YD, Lee KS, Kim JM, Ryu JH, Lee JG, Lee KW, Kim BW and Kim DS: Korean Organ Transplantation Registry Study Group: Artificial intelligence for predicting survival following deceased donor liver transplantation: Retrospective multi-center study. *Int J Surg* 105: 106838, 2022.
- Amann J, Vetter D, Blomberg SN, Christensen HC, Coffee M, Gerke S, Gilbert TK, Hagedorff T, Holm S, Livne M, *et al*: To explain or not to explain?—Artificial intelligence explainability in clinical decision support systems. *PLoS Digit Health* 1: e0000016, 2022.
- Collins GS, Moons KGM, Dhiman P, Riley RD, Beam AL, Van Calster B, Ghassemi M, Liu X, Reitsma JB, van Smeden M, *et al*: TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* 385: e078378, 2024.
- Cox JD, Stetz J and Pajak TF: Toxicity criteria of the radiation therapy oncology group (RTOG) and the European organization for research and treatment of cancer (EORTC). *Int J Radiat Oncol Biol Phys* 31: 1341-1346, 1995.
- Kursa MB and Rudnicki WR: Feature selection with the Boruta package. *J Stat Softw* 36: 1-13, 2010.
- Engelbrechtsen S and Bohlin J: Statistical predictions with glmnet. *Clin Epigenetics* 11: 123, 2019.
- Friedman J, Hastie T and Tibshirani R: Regularization paths for generalized linear models via coordinate descent. *J Stat Softw* 33: 1-22, 2010.
- Bos R, Zhong H, Hanrahan CF, Mommers EC, Semenza GL, Pinedo HM, Abeloff MD, Simons JW, van Diest PJ and van der Wall E: Levels of hypoxia-inducible factor-1 alpha during breast carcinogenesis. *J Natl Cancer Inst* 93: 309-314, 2001.
- D'Ignazio L, Batie M and Rocha S: Hypoxia and inflammation in cancer, focus on HIF and NF- κ B. *Biomedicines* 5: 21, 2017.

28. Zeng X, Ye L, Luo M, Zeng D and Chen Y: Prognostic value of pretreatment systemic immune-inflammation index in Chinese esophageal squamous cell carcinoma patients receiving radical radiotherapy: A meta-analysis. *Medicine (Baltimore)* 102: e34117, 2023.
29. Yu Y, Zheng H, Liu L, Li H, Zheng Q, Wang Z, Wu Y and Li J: Predicting severe radiation esophagitis in patients with locally advanced esophageal squamous cell carcinoma receiving definitive chemoradiotherapy: Construction and validation of a model based in the clinical and dosimetric parameters as well as inflammatory indexes. *Front Oncol* 11: 687035, 2021.
30. Qiu J, Ke D, Lin H, Yu Y, Zheng Q, Li H, Zheng H, Liu L and Li J: Using inflammatory indexes and clinical parameters to predict radiation esophagitis in patients with small-cell lung cancer undergoing chemoradiotherapy. *Front Oncol* 12: 898653, 2022.
31. Iyengar NM, Gucaip A, Dannenberg AJ and Hudis CA: Obesity and cancer mechanisms: Tumor microenvironment and inflammation. *J Clin Oncol* 34: 4270-4276, 2016.
32. Mehnati P, Baradaran B, Vahidian F and Nadiriazam S: Functional response difference between diabetic/normal cancerous patients to inflammatory cytokines and oxidative stresses after radiotherapy. *Rep Pract Oncol Radiother* 25: 730-737, 2020.
33. Winters E and Poole C: Challenges and impact of patient obesity in radiation therapy practice. *Radiography (Lond)* 26: e158-e163, 2020.
34. Barnett GC, West CML, Dunning AM, Elliott RM, Coles CE, Pharoah PDP and Burnet NG: Normal tissue reactions to radiotherapy: Towards tailoring treatment dose by genotype. *Nat Rev Cancer* 9: 134-142, 2009.
35. Sibaud V, Lebœuf NR, Roche H, Belum VR, Gladieff L, Deslandres M, Montastruc M, Eche A, Vigarios E, Dalenc F and Lacouture ME: Dermatological adverse events with taxane chemotherapy. *Eur J Dermatol* 26: 427-443, 2016.
36. Yosipovitch G and Bernhard JD: Clinical practice. Chronic pruritus. *N Engl J Med* 368: 1625-1634, 2013.
37. Agelopoulous K, Pereira MP, Wiegmann H and Ständer S: Cutaneous neuroimmune crosstalk in pruritus. *Trends Mol Med* 28: 452-462, 2022.
38. Zhang HW, Wang YR, Li J, Huang W, Xu B, Pang HW and Jiang CL: Predicting radiation esophagitis in patients undergoing synchronous boost radiotherapy post-breast-conserving surgery. *Dose Response* 23: 15593258251335802, 2025.
39. Sonis ST: The pathobiology of mucositis. *Nat Rev Cancer* 4: 277-284, 2004.
40. Bentzen SM: Preventing or reducing late side effects of radiation therapy: Radiobiology meets molecular pathology. *Nat Rev Cancer* 6: 702-713, 2006.
41. Koch A, Reinhardt P, Elicin O, Aebersold DM and Schanne DH: Predictive biomarkers of radiotherapy-related dermatitis, xerostomia, mucositis and dysphagia in head and neck cancer: A systematic review. *Radiother Oncol* 203: 110689, 2025.
42. Pandeli C, Smyth LML, David S and See AW: Dose reduction to organs at risk with deep-inspiration breath-hold during right breast radiotherapy: A treatment planning study. *Radiat Oncol* 14: 223, 2019.
43. Wang DQ, Yan L, Wu HF, Zhong YH, Huang XB, Jin J, Zhong QZ, Zhao LN, Wang XH, Jing H, *et al*: Prospective cohort study to validate esophageal dose constraints and predictive models for esophagitis in patients with breast cancer undergoing hypofractionated regional nodal radiotherapy. *Radiother Oncol* 212: 111120, 2025.
44. Dean JA, Wong KH, Gay H, Welsh LC, Jones AB, Schick U, Oh JH, Apte A, Newbold K, Bhide S, *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* 8: 27-39, 2018.
45. Steyerberg EW and Harrell FE Jr: Prediction models need appropriate internal, internal-external, and external validation. *J Clin Epidemiol* 69: 245-247, 2016.
46. Efron B and Tibshirani R: Improvements on cross-validation: The 632+ bootstrap method. *J Am Stat Assoc* 92: 548-560, 1997.



Copyright © 2026 He et al. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) License.