

Predicting Radiation Esophagitis in Patients Undergoing Synchronous Boost Radiotherapy Post-Breast-Conserving Surgery

Dose-Response:
An International Journal
Vol. 23(2): 1–12
© The Author(s) 2025
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/15593258251335802
journals.sagepub.com/home/dos



Huai-wen Zhang^{1,*}, Yi-ren Wang^{2,*}, Jingao Li^{1,*}, Wei Huang^{3,*}, Bin Xu¹, Hao-wen Pang⁴, and Chun-ling Jiang^{1,5}

Abstract

This study constructed a predictive model for occurrence of radiation esophagitis during breast-cancer radiotherapy. 308 breast-cancer patients were analyzed. Lasso regression identified crucial variables that were further integrated into a radiation esophagitis risk score, which was used to segregate patients into high- and low-risk groups. A nomogram model was designed for clinical applicability. Training and validations were performed to assess robustness and generalizability of proposed models, employing C-index, AUCs, calibration curves, and decision curves. SHAP algorithm was used for model interpretation, offering insights into the major contributory factors. Seven significant variables were identified by Lasso regression. C-indexes of nomograms of individual clinical variables and risk score were 0.795 and 0.784, respectively, exhibiting strong predictive ability. In internal validation, AUCs for risk score, nomogram, and logistic models were 0.784, 0.795, and 0.812, respectively. Calibration curves showed a close fit between predicted and observed outcomes across models. Decision curve analysis indicated logistic model's superior clinical utility when the risk threshold was above 0.2. SHAP interpretation emphasized radiation dose, pruritus, molecular type, and hepatic dysfunction as top contributory factors for radiation esophagitis. Models based on interpretable machine learning offer an intuitive tool to assess risk of radiation esophagitis in breast-cancer radiotherapy.

Keywords

machine learning, normal tissue complication, radiotherapy, radiation esophagitis, breast cancer

Received: 4 October 2024; accepted: 1 April 2025

¹ Department of Radiation Oncology, Jiangxi Cancer Hospital & Institute, Jiangxi Clinical Research Center for Cancer, The Second Affiliated Hospital of Nanchang Medical College, Nanchang, China

² School of Nursing, Southwest Medical University, Luzhou, China

³ Department of Radiation Oncology, Shandong Cancer Hospital and Institute, Shandong First Medical University and Shandong Academy of Medical Sciences, Jinan, China

⁴ Department of Oncology, The Affiliated Hospital of Southwest Medical University, Luzhou, China

⁵ Key Laboratory of Personalized Diagnosis and Treatment of Nasopharyngeal Carcinoma, Medical College of Nanchang University, China

*These authors contributed equally to this work.

Corresponding Authors:

Bin Xu, Department of Radiation Oncology, Jiangxi Cancer Hospital & Institute, The Second Affiliated Hospital of Nanchang Medical College, Beijing East Road street No 519, Nanchang 330029, China.

Email: 469996535@qq.com

Hao-wen Pang, Department of Oncology, The Affiliated Hospital of Southwest Medical University, taiping street No25#, luzhou 646000, China.

Email: haowenpang@foxmail.com

Chun-ling Jiang, Department of Radiation Oncology, Jiangxi Cancer Hospital & Institute, The Second Affiliated Hospital of Nanchang Medical College, Key Laboratory of Personalized Diagnosis and Treatment of Nasopharyngeal Carcinoma, Medical College of Nanchang University, Beijing East Road street No 519, Nanchang 330029, PR China.

Email: jclil2002@163.com



Creative Commons Non Commercial CC BY-NC: This article is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 License (<https://creativecommons.org/licenses/by-nc/4.0/>) which permits non-commercial use, reproduction and distribution of the work without further permission provided the original work is attributed as specified on the SAGE and

Open Access pages (<https://us.sagepub.com/en-us/nam/open-access-at-sage>).

Highlight

1. The study uniquely employs SHAP, revealing deeper insights into the key contributors of radiation-induced esophagitis, bridging algorithmic understanding with clinical implications.
2. Identified several previously unreported potential factors in radiation esophagitis, paving new avenues for targeted preventive strategies.
3. Provides a basis for individualized prevention in patients undergoing radiation therapy, enhancing the precision and effectiveness of interventions to mitigate the risk of radiation-induced esophagitis.

Introduction

Breast cancer remains one of the most prevalent malignancies affecting women worldwide.¹ Radiation therapy is a common and effective treatment modality for breast cancer.^{2,3} Radiation esophagitis is a common and complication of breast cancer radiotherapy, affecting up to 50% of patients.⁴ Most symptoms of esophageal inflammation are mild, but severe ones can cause pain and difficulty swallowing, and malnutrition, affecting the patient's quality of life and treatment compliance. Prevention and management of radiation esophagitis in a timely and effective way can effectively reduce the severity symptoms in patients. Identifying factors in radiation esophagitis during breast cancer radiotherapy in advance is crucial for enhancing patient quality of life and radiotherapy effectiveness, and reducing treatment costs and potentially increase survival rates.⁵

Several factors have been associated with the risk of radiation esophagitis, such as radiation dose, volume, and technique, as well as patient characteristics, comorbidities, and genetic factors.^{6,7} However, the existing predictive models for radiation esophagitis are mostly based on conventional statistical methods, which have limitations in handling complex and high-dimensional data, and may not capture the nonlinear interactions and heterogeneity among variables.⁸ Moreover, most of the models are not easily interpretable or applicable in clinical practice, and have not been validated in independent cohorts.⁹ Clinical medical data based on EHR and HIS systems emphasize the need for improved predictive models due to their complex composition of nonlinear relationships and high-dimensional and multimodal nature.

Machine Learning is a branch of Artificial Intelligence that can learn and make predictions or decisions from non-linear and high-dimensional data compared to traditional statistical methods, and is particularly applicable to complex problems in the medical field.¹⁰ Machine learning techniques have been increasingly applied to various fields of medicine, including oncology, and have shown promising results in improving diagnosis, prognosis, and treatment outcomes.^{11,12} However, machine learning models are often considered as “black

boxes”, meaning that their internal logic and reasoning are not transparent or understandable to human users.^{13–16} This poses challenges for the trustworthiness, accountability, and ethical use of machine learning in clinical settings.

Interpretable machine learning is an emerging field that aims to provide explanations for the predictions or decisions made by machine learning models, and to enhance their interpretability, transparency, and fairness.^{17,18} Interpretable machine learning can help clinicians and patients to understand the rationale behind the machine learning models, to assess their validity and reliability, and to identify the most influential factors for the outcomes of interest.¹⁹ For instance, the weight of the output variable's contribution to the model or the effect of different variables on the prediction results in a single sample condition.²⁰ Interpretable machine learning can also facilitate the communication and implementation of machine learning models in clinical practice, and foster the collaboration and feedback between machine learning researchers and clinicians.²¹ This interpretable modeling enhances the possibility of translating predictive models from theory into programs or web tools that can be integrated into EHR systems, with the potential for clinical applications.

In this study, we aimed to develop and validate a predictive model for the occurrence of radiation esophagitis during breast cancer radiotherapy using interpretable machine learning techniques. Aims to facilitate early detection of radiation esophagitis, thereby assisting clinicians in decision-making and intervention. We used a cohort of breast cancer patients who received radiotherapy and collected various clinical and dosimetric variables that may affect the risk of radiation esophagitis. We applied lasso regression to select the most relevant variables and constructed a risk score based on their coefficients. We also developed a nomogram model that integrated the risk score and individual clinical variables. We evaluated the performance and clinical utility of the models using various metrics and methods, such as C-index, AUC, calibration curve, and decision curve. We used the SHAP algorithm to interpret the models and to identify the major contributory factors for radiation esophagitis. To our knowledge, this is the first study to use interpretable machine learning to predict radiation esophagitis in breast cancer radiotherapy.

Methods

Patients

A total of 308 breast cancer patients between January, 2016, and December, 2020, were retrospectively analyzed. The median age was 45.1 (24–72) years. The inclusion criteria comprised breast cancer patients at stages 0 to III who underwent breast-conserving surgery followed by radiation therapy. The exclusion criteria were as follows: patients with existing esophageal diseases, patients who have undergone esophageal surgery, patients who are undergoing or have

undergone esophageal radiotherapy, patients with severe systemic diseases.

The prescribed doses included 59.4-59.94 Gy at 2.20-2.22 Gy/fraction to original tumor bed and 49.95 Gy at 1.85 Gy/fraction to subclinical lesions or lymphatic drainage area in 27 fractions.

Selection of the Target Variable

The primary objective of our research was to predict the onset of acute radiation esophagitis during radiotherapy. Throughout the course of radiation therapy, the patient's condition of acute radiation esophagitis is evaluated by more than two physicians according to the RTOG standards, and the highest grade of esophagitis is recorded. RTOG grading ≥ 1 was diagnosed as radiation esophagitis.²²

K-Nearest Neighbors Imputation

The K-nearest neighbors imputation method was used to address the missing values in our data. KNN imputation is a widely used and effective method in clinical prediction model, especially when the dataset is relatively small and the missing data mechanism can be reasonably assumed to be missing at random.²⁰ This method relies on a distance metric, identifying the 'k' closest samples in the training dataset to the point requiring imputation and then filling in the missing value using the mode (for categorical variables) and the mean (for continuous variables) of these 'k' neighbors.²³ The Euclidean distance formula was employed to determine the proximity between data points. In an n-dimensional space, given two points p and q , the Euclidean distance, d , is calculated as follows:

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

The optimal number of neighbors 'k' was determined using a grid search algorithm.

Variable Selection Based on Lasso

Least Absolute Shrinkage and Selection Operator (Lasso) regression is a L1 regularization method that improves the predictive accuracy of statistical models due to its ability to perform both variable selection and regularization.²⁴ In addition, lasso regression is able to assign coefficients to each variable thus enhancing the interpretability of model decisions and is therefore used for variable selection in most clinical prediction models as well as in this study.²⁵ Lasso regression was performed using the "glmnet" package in R (Version 4.2.1).²⁶ A range of lambda (λ) values were tested to identify the optimal regularization parameter through k-fold cross-validation. The λ that resulted in the smallest mean cross-validated error was chosen as the optimal value.

The choice of the optimal λ value plays a crucial role in balancing model complexity and predictive performance. A smaller λ allows the model to fit the data more closely, potentially leading to overfitting, where the model captures noise or random fluctuations in the training data rather than the underlying signal. This can result in poorer generalization to new, unseen data.²⁷ On the other hand, a larger λ value applies a stronger regularization penalty, shrinking the coefficients of less important variables towards zero. While this helps reduce overfitting and improves the model's ability to generalize, it may also cause underfitting if the regularization is too strong, removing potentially useful variables from the model.²⁸ Thus, selecting the optimal λ involves finding a balance between underfitting and overfitting.

Calculation of Radiation Esophagitis Risk Score and Construction of a Prediction Model

To evaluate the individual risk of radiation esophagitis for patients, a risk score was calculated for each patient based on the non-zero coefficients derived from Lasso regression. The formula for the risk score is as follows:

$$\text{RadiationEsophagitisRiskScore} = \sum_{i=1}^n (\text{Coefficient of Variable} \times \text{Value of Variable})$$

Subsequently, combining the risk score with individual variables obtained from Lasso regression, univariate regression was followed by multivariate regression to construct and visualize a nomogram model.

In addition to the nomogram, a logistic regression model was constructed to predict the occurrence of radiation esophagitis. The logistic model provides a probability score ranging between 0 and 1²⁹ with a formula as follows:

$$P(Y = 1) = \frac{e^{(\beta_0 + \sum \beta_i X_i)}}{1 + e^{(\beta_0 + \sum \beta_i X_i)}}$$

where $P(Y = 1)$ is the probability of developing radiation esophagitis, β_0 is the intercept, and β_i represents the coefficient of the i -th predictor variable X_i .

To ensure the reliability and precision of both the nomogram and logistic models, training and internal validation was conducted using a bootstrap method with 1000 resamples. Specifically, for each resample, a new training set was generated by randomly selecting patients (with replacement) from the original training dataset. This allows some patients to be selected multiple times while others may be left out of the training set. The remaining patients, those not selected in each resample, were used as the validation. This process was repeated 1000 times to assess the robustness of the model and generate a distribution of model performance metrics. To ensure the reliability and accuracy of the nomogram and

logistic models, internal validation was performed using bootstrap resampling on the training dataset ($n = 201$). The test dataset ($n = 107$) served as an independent dataset for validation.

Statistical Analysis

Continuous variables are presented as medians with interquartile range and were compared using the Mann-Whitney U test. Categorical variables are expressed as counts and percentages and were compared using the chi-squared test. A two-sided P value of < 0.05 was considered to be statistically significant. The Lasso regression algorithm was performed using the glmnet package in R, whereas the logistic regression algorithm was conducted through the Scikit-learn package within the Python (Version 3.9). Model interpretation was executed using the fastshap and DALEX packages.

Results

Baseline Analysis

The baseline characteristics of the 308 participants (Table 1) were categorized based on the occurrence of radiation esophagitis. Upon closer examination of the data, distinct differences in several variables emerged between patients who developed radiation esophagitis and those who did not (Table 1).

Variables Selected by Lasso Regression

For feature selection, 19 clinical variables were incorporated into the Lasso regression model, with the optimal model parameters being automatically identified through a grid search algorithm. We subsequently employed seven-fold cross-validation for model training and chose lambda.min as the penalty parameter. This process yielded seven important variables with corresponding non-zero coefficients, which were then incorporated into the subsequent calculation of radiation esophagitis risk score and construction of the nomogram (Figure 1).

Risk-Score Calculation and Nomogram Model Construction

Using the seven important variables identified by Lasso regression, we formulated a radiation esophagitis risk score. Each patient's risk score was calculated based on the weighted sum of these variables; the weights were determined by the respective coefficients derived from Lasso regression. The formula of risk score is as follows:

Subsequently, we calculated the risk score for each patient. Based on the median risk score, patients were divided into a high group and a low group. The high group represents those with a higher incidence of radiation esophagitis. The distribution of risk scores for all patients can be seen in Figure 2A.

We constructed a nomogram model using the risk score and an individual clinical variable nomogram with the clinical variables selected by Lasso regression. This graphical tool offers clinicians an intuitive interface, allowing easy evaluation of the risk of radiation esophagitis in individual patients. By simply plotting the values of each variable on the nomogram, a cumulative score can be obtained, which corresponds to a specific probability of developing radiation esophagitis (Figure 2B and C). The logistic regression model was also constructed and compared the predictive capabilities of the risk-score nomogram and clinical variable nomogram.

Model Evaluation and Validation

Through 1000 bootstrap resampling, we performed internal validation of the risk-score nomogram, individual clinical variable nomogram, and logistic model. The risk-score nomogram and clinical variable nomogram had C-indexes of 0.784 and 0.795, respectively, indicating their robust predictive capability and distinction. The variance inflation factor (VIF) of the nomogram was 1, suggesting that the model is free from multicollinearity. In the training set, the areas under the curve (AUCs) for the risk-score nomogram, clinical variable nomogram, and logistic model were ($AUC = 0.841$, 95% CI = 0.794-0.871), ($AUC = 0.873$, 95% CI = 0.822-0.910), and ($AUC = 0.879$, 95% CI = 0.819-0.921), respectively. Subsequent independent validation was performed using an independent dataset, with the predictive capabilities of the three models presented in Figure 3A-C. The results showed that all three prediction models exhibited good generalizability, with the logistic model attaining the highest AUC. Calibration curves for the three models in the independent validation set were plotted to examine the relationship between actual probabilities and observed outcomes (Figure 3D-F). All three models displayed an promising results, with the predicted probabilities of radiation esophagitis closely matching the observed rates across the prediction range. Finally, clinical decision curves were depicted in Figure 3G. The Decision Curve Analysis(DCA) results indicated that when the risk threshold is above 0.2, the net clinical benefits predicted by the three models exceed both the "All" and "None" scenarios. Notably, the logistic model exhibited the best performance, suggesting its superior clinical applicability.

$$\begin{aligned} \text{RadiationEsophagitisRiskScore} = & (\text{AxillaryLymphNodeDissection} \times 0.00401245 + \text{PathologicalType} \times 0.53204249 \\ & + \text{Chemotherapy} \times 0.21195632 + \text{RadiationDose} \times 2.644223 + \text{RadiationDays} \times 0.0048358 \\ & - 1.117188 \times \text{HepaticDysfunction} + \text{Pruritus} \times 1.2208466) \end{aligned}$$

Table 1. Baseline Information.

Radiation Esophagitis	No (n = 162)	Yes (n = 146)	P value
Age, median (IQR)	44 (39, 49.75)	46 (40, 50.75)	0.291
Menstruation, n (%)			0.951
Yes	116 (37.7%)	105 (34.1%)	
No	46 (14.9%)	41 (13.3%)	
Breast cancer position, n (%)			0.456
Right	79 (25.6%)	65 (21.1%)	
Left	83 (26.9%)	81 (26.3%)	
Axillary lymph node dissection, n (%)			0.185
No	103 (33.4%)	82 (26.6%)	
Yes	59 (19.2%)	64 (20.8%)	
T Stage, n (%)			0.921
Tis	7 (2.3%)	5 (1.6%)	
T1	98 (31.8%)	89 (28.9%)	
T2	57 (18.5%)	52 (16.9%)	
N Stage, n (%)			0.721
N0	114 (37%)	100 (32.5%)	
N1	48 (15.6%)	46 (14.9%)	
Histological grade, n (%)			0.921
3	61 (19.8%)	56 (18.2%)	
1	18 (5.8%)	18 (5.8%)	
2	83 (26.9%)	72 (23.4%)	
Pathological type, n (%)			0.104
Carcinoma in situ	7 (2.3%)	2 (0.6%)	
Invasive carcinoma	155 (50.3%)	142 (46.1%)	
Papillary lesion	0 (0%)	2 (0.6%)	
Molecular type, n (%)			0.006
Luminal A	42 (13.6%)	24 (7.8%)	
Luminal B	76 (24.7%)	87 (28.2%)	
HER-2 over-expression	18 (5.8%)	5 (1.6%)	
Triple-negative	26 (8.4%)	30 (9.7%)	
Targeted therapy, n (%)			0.307
No	136 (44.2%)	116 (37.7%)	
Yes	26 (8.4%)	30 (9.7%)	
Endocrine therapy, n (%)			0.435
Yes	118 (38.3%)	112 (36.4%)	
No	44 (14.3%)	34 (11%)	
Chemotherapy, n (%)			0.248
No	29 (9.4%)	19 (6.2%)	
Yes	133 (43.3%)	126 (41%)	
Radiation dose, n (%)			<0.001
59.4 Gy	160 (51.9%)	110 (35.7%)	
59.94 Gy	2 (0.6%)	36 (11.7%)	
Radiotherapy range, n (%)			0.953
Whole breast	106 (34.4%)	96 (31.2%)	
Whole breast + Clavicle	56 (18.2%)	50 (16.2%)	
Radiotherapy days, median (IQR)	38 (37, 40)	38 (37, 40)	0.972
Hepatic dysfunction, n (%)			0.003
No	145 (47.1%)	143 (46.4%)	
Yes	17 (5.5%)	3 (1%)	
Myelosuppression, n (%)			0.788
No	122 (39.6%)	108 (35.1%)	
Yes	40 (13%)	38 (12.3%)	

(continued)

Table I. (continued)

Radiation Esophagitis	No (n = 162)	Yes (n = 146)	P value
Pruritus, n (%)			<0.001
No	125 (40.6%)	73 (23.7%)	
Yes	37 (12%)	73 (23.7%)	
Late skin reactions, n (%)			0.545
No	137 (44.5%)	127 (41.2%)	
Yes	25 (8.1%)	19 (6.2%)	

Model Interpretation Based on SHAP

Interpretation of the optimal model was conducted with the SHAP algorithm. Initially, a global SHAP interpretation of feature importance was performed using the DELAX package in R (Figure 4). For categorical variables, SHAP results revealed the top five risk factors contributing to the occurrence of radiation esophagitis: radiation dose, pruritus, molecular type, hepatic dysfunction, and breast cancer position (Figure 4A). For continuous variables, age and the duration of radiation therapy showed a positive correlation with the occurrence of radiation esophagitis (Figure 4B).

Figure 5, generated through the Scikit-learn and fastshap packages in Python, illustrates the influence of each variable on the model's output. All of the patient attributions to the outcome were plotted, with red points indicating high-risk values, and blue points indicating low-risk values. Pruritus, a radiation dose of 59.94 Gy, and the triple-negative breast cancer subtype were significant high-risk factors leading to radiation esophagitis.

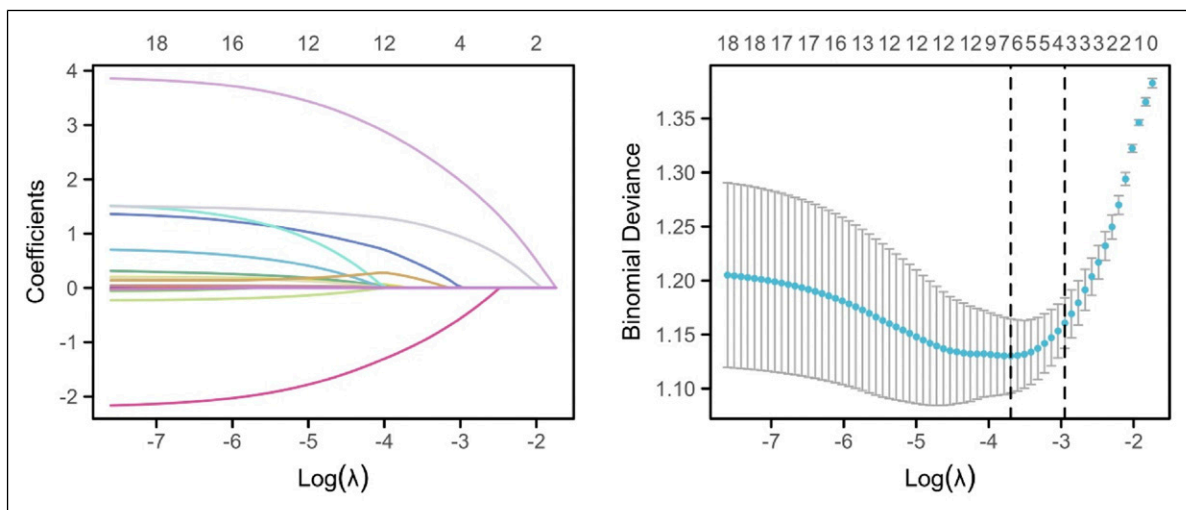
Moreover, we randomly selected one sample from each group with and without radiation esophagitis to further validate the model's interpretability. Figure 6A displays a patient with radiation esophagitis, for which the SHAP predicted score is high (3.48). The top three factors elevating the risk of radiation esophagitis are a higher dose of radiation treatment (radiation

dose = 2 = 59.94 Gy), pruritus (pruritus = 1), and the breast cancer lesion being on the left side (breast cancer position = 1). Figure 6B represents a patient without radiation esophagitis. The SHAP predicted score was low (0.20), with the primary risk factors for increased radiation esophagitis being prolonged radiation treatment (radiation days = 60), the breast cancer lesion being on the left side (breast cancer position = 1), and receiving endocrine therapy (endocrine therapy = 1). Factors that decrease the risk of radiation esophagitis include the absence of pruritus (pruritus = 0) and receiving a lower dose of radiation treatment (radiation dose = 1 = 59.4 Gy).

Discussion

In this study, we developed three predictive models for radiation esophagitis, all of which demonstrated robust generalizability in both the training set and an external validation dataset. Notably, the logistic model outperformed the other two models (risk score and nomogram) in the independent validation set, with an AUC of 0.812 (95% CI = 0.766-0.858) compared to 0.784 (95% CI = 0.733-0.835) and 0.795 (95% CI = 0.745-0.844), respectively.

To further understand the internal workings of the models, we employed interpretable machine learning algorithms. Specifically, we used SHAP values to gain insights into the contribution and importance of each individual variable in predicting radiation

**Figure 1.** Lasso Regression Variable Trajectories and Coefficient Selection.

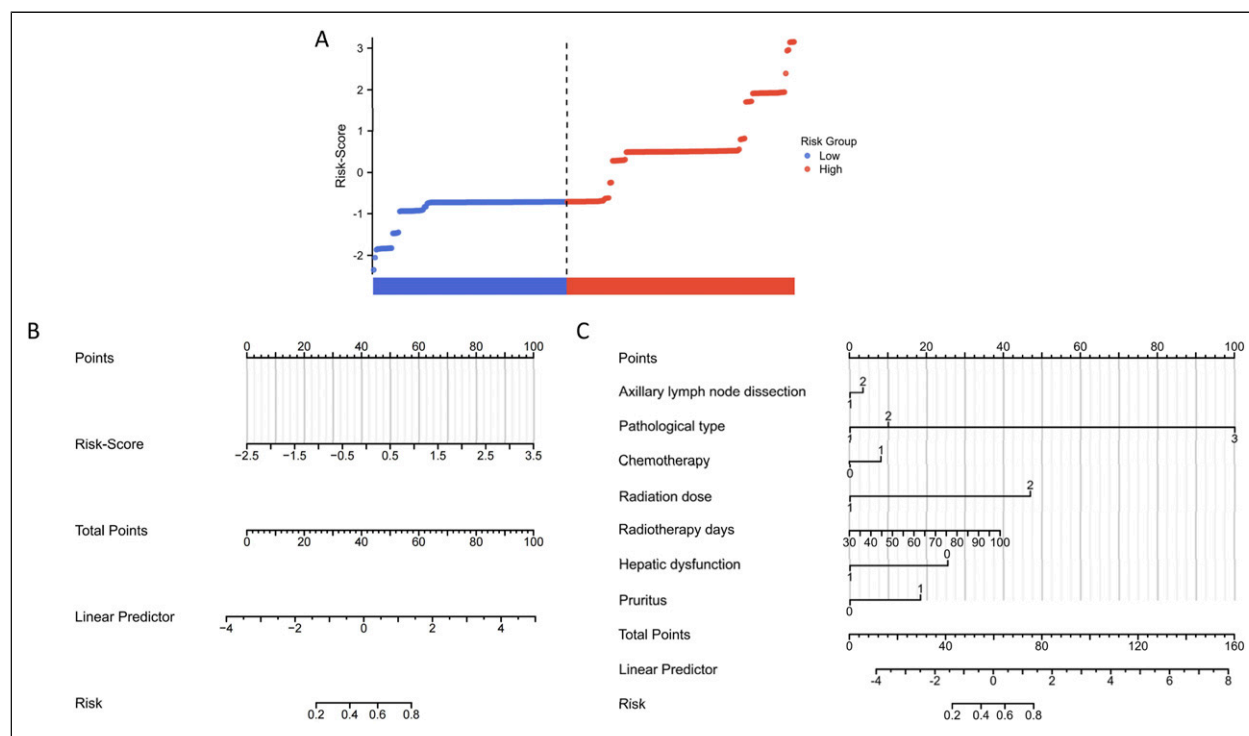


Figure 2. Risk-Score Calculation and Nomogram Construction. (A) Distribution of all Patients Based on Risk Score. (B) Risk-Score Nomogram. (C) Individual Clinical Variable Nomogram.

esophagitis. This interpretative capability is particularly crucial for clinical applications, as it allows physicians to better interpret the model's predictions based on both the unique circumstances of each patient and the overarching context.

One potential explanation for the occurrence of radiation esophagitis in patients undergoing synchronous boost radiotherapy post-breast-conserving surgery in this study could be attributed to the lack of consideration given to the esophagus as an organ at risk for dose limitation during the design of radiotherapy plans, as the tumor target area was situated slightly distant from the patients' esophagus. Our study revealed that radiation dosage and the prevalence of pruritus were the most significant variables. Most patients without esophagitis were administered a dose of 59.4 Gy, whereas a significant fraction of patients with esophagitis received 59.94 Gy (35.7%, $P < .001$). The two dosage values in this study were 59.94 Gy and 59.4 Gy, both used for simultaneous boost intensity-modulated radiation therapy. The mere difference of 0.54 Gy in radiation dosage led to significant differences in patient outcomes, suggesting that beyond a certain dosage threshold, the likelihood of a patient developing radiation esophagitis markedly increases. This suggests that during design the radiotherapy treatment plans for these patients, the esophagus must be fully regarded as the organ at risk for dose limits, so as to minimize the occurrence of radiation esophagitis.

The SHAP analysis of individual samples revealed that the location of breast cancer lesions on the left side is a significant contributing factor that increases the risk of developing

radiation esophagitis. This is consistent with the findings of Wang et al. That left breast cancer and internal breast irradiation are risk factors for radiation esophagitis.³⁰ In addition, This finding maybe also related to the slightly higher exposure dose of esophageal Dmax in patients with left breast compared with that in patients with right breast.^{4,5,31} Such considerations might account for the increased vulnerability of patients with left-sided lesions to esophageal inflammation post-radiation. Furthermore, this observation underscores the importance of meticulous planning and precision during radiation therapy, particularly for patients with left-sided breast cancer. It suggests the need for advanced techniques, perhaps incorporating real-time imaging or adaptive radiotherapy, to ensure minimal radiation spillover to the esophagus.³²

For continuous variables, our single-sample SHAP analysis revealed that the number of days a patient undergoes radiation therapy is most significant for increasing the risk of radiation esophagitis. This is an intuitive finding as prolonged exposure to radiation, even if each individual dose is considered safe, can culminate in cumulative tissue damage, making the esophagus progressively more susceptible to inflammation.³³ The esophagus, being a radiosensitive organ, may exhibit a dose-response relationship with radiation exposure. As the number of treatment days increases, the tissue might not have adequate time for repair and recovery between sessions.³⁴ As a result, subsequent doses compound pre-existing damage, thereby increasing the propensity for radiation-induced inflammation and subsequent esophagitis.

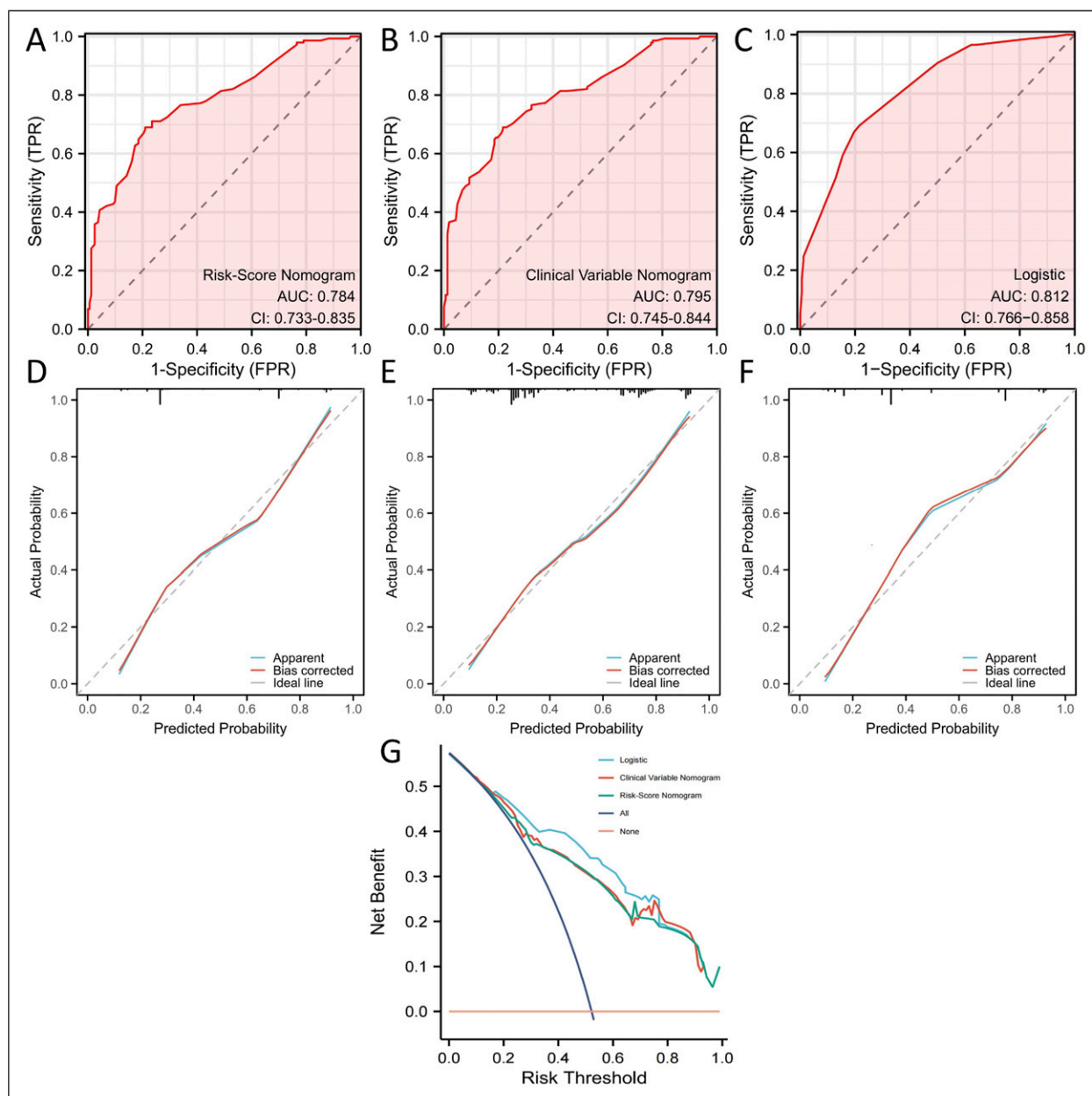


Figure 3. Model Evaluation in Validation Dataset. (A and D) Risk-Score Nomogram. (B and E) Individual Clinical Variable Nomogram. (C and F) Logistic Model. (G) Decision Curve Analysis.

In addition, age was positively correlated with the risk of radiation esophagitis in our study. However, the relationship between age and adverse reactions in normal tissues remains a contentious issue in recent research.³⁵ Aging is accompanied by a myriad of physiological changes, including a decrease in tissue regeneration capacity, alterations in cellular repair mechanisms, and a general decline in organ function.³⁶ Aging may lead to the weakening of the immune system and the decrease in the tolerance of the patient's body tissue, making the body more sensitive to the side effects of radiotherapy. These changes might render the elderly more susceptible to radiation-induced damage.^{37,38} Although some studies have

suggested that older patients might exhibit an increased vulnerability to radiation due to compromised DNA repair processes and reduced cellular turnover, others have not found significant age-dependent differences in radiation responses.³⁹ It is also plausible that other age-associated factors, such as comorbidities or concurrent medications, could indirectly influence radiation tolerance and subsequently radiation-induced esophagitis.⁴⁰

In the SHAP interpretation model of our study, several variables exhibited substantial contributory values, including whether the patient underwent endocrine therapy, the molecular type of breast cancer, and the presence of hepatic

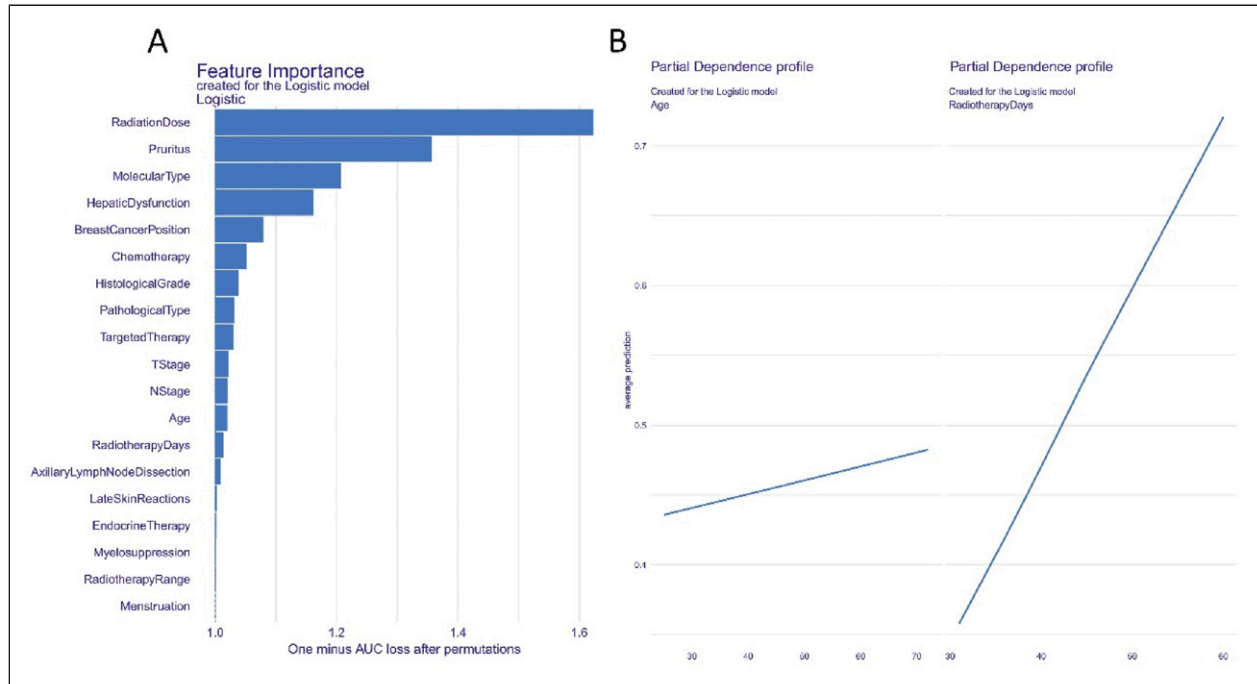


Figure 4. Global Average Absolute SHAP Values Algorithm Based on the DALEX Package. (A) Ranking of Each Variable's Contribution to the Model. (B) The Relationship Between Continuous Variables and Predicted Outcome.

dysfunction. Hepatic dysfunction appeared to be more common among patients without esophagitis, suggesting its potential protective role against the development of radiation esophagitis ($P = .003$). Pruritus stood out distinctly, with a

substantially higher prevalence in patients with radiation esophagitis ($P < .001$). Furthermore, molecular type variations in tumors were evident between the groups ($P = .006$). To date, no research has explored the mechanisms underlying the

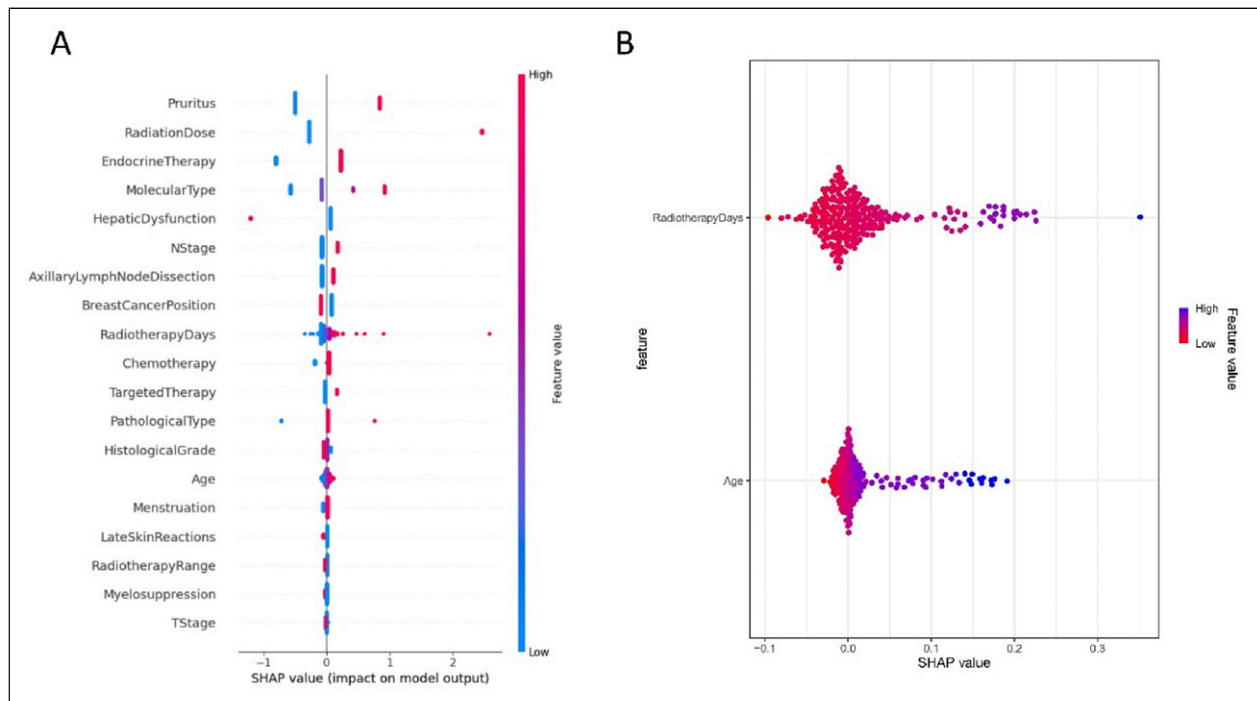


Figure 5. Variable Attributes Explained by SHAP. (A) Variable Attributes in SHAP. Each Line Represents a Variable, With the Horizontal Axis Representing the SHAP Value. Red and Blue Dots Indicate Higher and Lower Risk Feature Values, Respectively. (B) An Enlarged View of the Scatter Plot for Continuous Variables.

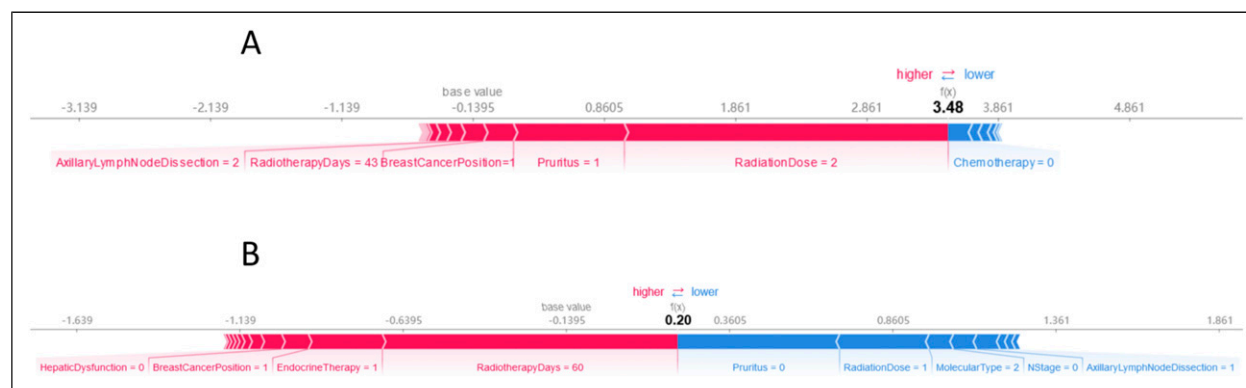


Figure 6. Model Output for a Single Sample Prediction Based on SHAP Values. The Baseline Value of the Model, also Known as the Starting Point of the Prediction, is $f(x)$ = Base Value. In This Binary Classification Study, it Represents the Average Output Probability of the Model. Each Variable's Contribution is Represented as Either a Red or Blue Block, which Indicates the Positive or Negative Impact, Respectively, of the Variable on the Prediction. The Width of Each Block Signifies the Magnitude of that Variable's Influence (ie, the Absolute Value of the SHAP Value). (A) A 47-Year-old Woman Diagnosed With Radiation Esophagitis. (B) A 49-Year-old Woman Without Radiation Esophagitis.

correlation between these variables and radiation-induced esophagitis. The associations we uncovered, especially regarding endocrine therapy, molecular type of breast cancer, and hepatic dysfunction, emphasize areas that are relatively uncharted in the context of radiation-induced esophagitis. Consequently, future research inspired by our findings could lead to more nuanced preventive measures and therapeutic strategies for those at risk of developing radiation-induced esophagitis. This study offers the possibility of model-assisted early detection of radiation esophagitis and facilitates decision-making assistance for clinicians. In clinical practice, a patient's clinical variables can be inputted, computed using the model developed in this study, and the final output predicted for the purpose of identification of radiation esophagitis.

This study has some limitations. First, due to its retrospective nature, we could only conduct a retrospective analysis on past samples. The deeper interpretative capability of the model must still be verified through prospective research, which is our next focus. In addition, although our research achieved good performance, the sample size was limited, potentially constraining our ability to explore certain associations. Furthermore, some crucial covariates might not have been considered or measured, which could affect the comprehensiveness and accuracy of the results. Therefore, its generalization ability may have some limitations. In the future, larger-scale and prospective studies are necessary to validate our findings. In addition, multi-omics and biophysical variables will be collected for the construction of a multimodal predictive model with the aim of improving the predictive performance for radiation esophagitis. While we have made every effort to ensure the accuracy of our findings, they should be interpreted with caution. Further studies are needed to validate these results and to explore the potential mechanisms underlying the observed associations. We hope that our study will provide a foundation for future research in this area.

Conclusion

This study developed and validated a predictive model for radiation esophagitis in breast cancer radiotherapy using machine learning and SHAP interpretation. The model showed good performance and clinical utility in both training and validation cohorts. The model also identified the main factors associated with radiation esophagitis, such as radiation dose, pruritus, molecular type, and hepatic dysfunction. This model could help clinicians to assess the risk of radiation esophagitis and tailor the radiotherapy plan accordingly.

ORCID iD

Huai-wen Zhang  <https://orcid.org/0000-0001-6259-6426>

Statements and Declarations

Ethics Approval

According to the ethical guide-lines of the Helsinki Declaration and was approved by the institutional review board of Jiang-xi Cancer Hospital. The study was performed in accordance with the Declaration of Helsinki and approved by the Ethics Committee of Jiang-xi Cancer Hospital (ethics no: 2023KY165).

Informed Consent

Written informed consents were obtained from all patients prior to treatment. Informed consent forms were signed were acquired from all patients in the study.

Author Contributions

Contributions: (I) Conception and design: Huaiwen Zhang, Bin Xu, Chun-ling Jiang; (II) Administrative support: Jingao Li, Haowen Pang, Huaiwen Zhang, Wei Huang; (III) Provision of study materials or patients: Huai-wen Zhang, Yi-ren Wang; (IV) Collection and assembly of data: Huai-wen Zhang, Chun-ling Jiang; (V) Data analysis and interpretation: Huaiwen Zhang, Yi-ren Wang, Haowen

Pang, Chun-ling Jiang; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by the National Natural Science Foundation of China (No. 82260607, 81760547), the Science and Technology Project of Jiangxi Province (No. 20212BAB206065, 20242BAB26138), and outstanding youth of Jiangxi Cancer hospital (No. 2021DYS02), High-level and high-skill leading talents of Jiangxi Province (GCSG2002001). The Sichuan Provincial Medical Research Project Plan (No. S21004), the Key-funded Project of the National College Student Innovation and Entrepreneurship Training Program (No. 202310632001), the National College Student Innovation and Entrepreneurship Training Program (No. 202310632028), the National College Student Innovation and Entrepreneurship Training Program (No. 202310632036).

Conflicting Interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Data Availability Statement

All data generated and analyzed during this study are included in this published article.

References

1. Siegel RL, Miller KD, Wagle NS, Jemal A. Cancer statistics, 2023. *CA Cancer J Clin*. 2023;73(1):17-48.
2. Meattini I, Lambertini M, Desideri I, De Caluwé A, Kaidar-Person O, Livi L. Radiation therapy for young women with early breast cancer: current state of the art. *Crit Rev Oncol Hematol*. 2019;137:143-153.
3. Amin SSM, Faraj KA, Ali JS, Rahim HAH, Yarahmadi M. Prediction factors of radiation esophagitis in breast cancer patients undergoing supraclavicular radiotherapy. *J Med Phys*. 2023;48(1):38-42.
4. West K, Schneider M, Wright C, et al. Radiation-induced oesophagitis in breast cancer: factors influencing onset and severity for patients receiving supraclavicular nodal irradiation. *J Med Imaging Radiat Oncol*. 2020;64:113-119.
5. Dzul S, Ninia J, Jang H, Kim S, Dominello M. Predictors of acute radiation dermatitis and esophagitis in African American patients receiving whole-breast radiation therapy. *Pract Radiat Oncol*. 2022;12(1):52-59.
6. Guerra JL, Gomez D, Wei Q, et al. Association between single nucleotide polymorphisms of the transforming growth factor $\beta 1$ gene and the risk of severe radiation esophagitis in patients with lung cancer. *Radiation Oncol*. 2012;105(3):299-304.
7. Monti S, Xu T, Mohan R, Liao Z, Palma G, Cella L. Radiation-induced esophagitis in non-small-cell lung cancer patients: voxel-based analysis and NTCP modeling. *Cancers (Basel)*. 2022;14(7):1833.
8. Dinov ID. Methodological challenges and analytic opportunities for modeling and interpreting big healthcare data. *GigaScience*. 2016;5(1):12.
9. Zheng X, Guo W, Wang Y, et al. Multi-omics to predict acute radiation esophagitis in patients with lung cancer treated with intensity-modulated radiation therapy. *Eur J Med Res*. 2023;28(1):126.
10. Greener JG, Kandathil SM, Moffat L, Jones DT. A guide to machine learning for biologists. *Nat Rev Mol Cell Biol*. 2022;23(1):40-55.
11. Din NMU, Dar RA, Rasool M, Assad A. Breast cancer detection using deep learning: datasets, methods, and challenges ahead. *Comput Biol Med*. 2022;149:106073.
12. Sammut SJ, Crispin-Ortuzar M, Chin SF, et al. Multi-omic machine learning predictor of breast cancer therapy response. *Nature*. 2022;601(7894):623-629.
13. Srinivasu PN, Sandhya N, Jhaveri RH, Raut R. From blackbox to interpretable AI in healthcare: existing tools and case studies. *Mob Inf Syst*. 2022;2022:1-20.
14. Rösli E, Bozkurt S, Hernandez-Boussard T. Peeking into a black box, the fairness and generalizability of a MIMIC-III benchmarking model. *Sci Data*. 2022;9(1):24.
15. Rudin C. Why black box machine learning should be avoided for high-stakes decisions, in brief. *Nat Rev Methods Primers*. 2022;2(1):81.
16. Azodi CB, Tang J, Shiu SH. Opening the black box: interpretable machine learning for geneticists. *Trends Genet*. 2020;36(6):442-455.
17. Yang R, Xiong X, Wang H, Li W. Interpretable machine learning model to prediction EGFR mutation in lung cancer. *Front Oncol*. 2022;12:924144.
18. Lai Y, Lin P, Lin F, et al. Identification of immune microenvironment subtypes and signature genes for Alzheimer's disease diagnosis and risk prediction based on interpretable machine learning. *Front Immunol*. 2022;13:1046410.
19. Moncada-Torres A, van Maaren MC, Hendriks MP, Siesling S, Geleijnse G. Interpretable machine learning can outperform cox regression predictions and provide insights in breast cancer survival. *Sci Rep*. 2021;11(1):6968.
20. Wang Y, Zhang H, Wang H, et al. Development of a neoadjuvant chemotherapy efficacy prediction model for nasopharyngeal carcinoma integrating magnetic resonance radiomics and pathomics: a multi-center retrospective study. *BMC Cancer*. 2024;24(1):1-15.
21. Vrdoljak J, Boban Z, Barić D, et al. Applying interpretable machine learning models for detection of breast cancer lymph node metastasis in patients eligible for neoadjuvant treatment. *Cancers (Basel)*. 2023;15(3):634.
22. Huang EX, Bradley JD, El Naqa I, et al. Modeling the risk of radiation-induced acute esophagitis for combined Washington University and RTOG trial 93-11 lung cancer patients. *Int J Radiat Oncol Biol Phys*. 2012;82(5):1674-1679.
23. Lubis AR, Lubis M, Khowarizmi A. Optimization of distance formula in K-nearest neighbor method. *Bulletin EEI*. 2020;9(1):326-338.

24. Wen Z, Wang Y, Chen S, et al. Construction of a predictive model for postoperative hospitalization time in colorectal cancer patients based on interpretable machine learning algorithm: a prospective preliminary study. *Front Oncol.* 2024;14:1384931.
25. Kumar S, Attri SD, Singh KK. Comparison of Lasso and stepwise regression technique for wheat yield prediction. *J Agrometeorol.* 2019;21(2):188-192.
26. Engebretsen S, Bohlin J. Statistical predictions with glmnet. *Clin Epigenetics.* 2019;11(1):123.
27. Pak A, Rad AK, Nematollahi MJ, Mahmoudi M. Application of the Lasso regularisation technique in mitigating overfitting in air quality prediction models. *Sci Rep.* 2025;15(1):547.
28. Sahebalam H, Gholizadeh M, Hafezian SH. The effect of different approaches to determining the regularization parameter of bayesian LASSO on the accuracy of genomic prediction. *Mamm Genome.* 2024;36(1):331-345.
29. Domínguez-Almendros S, Benítez-Parejo N, Gonzalez-Ramirez AR. Logistic regression models. *Allergol Immunopathol.* 2011; 39(5):295-305.
30. Wang DQ, Zhang N, Dong LH, 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.* 2023;117(1):186-197.
31. Yaney A, Ayan AS, Pan X, et al. Dosimetric parameters associated with radiation-induced esophagitis in breast cancer patients undergoing regional nodal irradiation. *Radiother Oncol.* 2021;155:167-173.
32. Keall PJ, Nguyen DT, O'Brien R, et al. The first clinical implementation of real-time image-guided adaptive radiotherapy using a standard linear accelerator. *Radiother Oncol.* 2018;127(1):6-11.
33. Chapet O, Kong FM, Lee JS, Hayman JA, Ten Haken RK. Normal tissue complication probability modeling for acute esophagitis in patients treated with conformal radiation therapy for non-small cell lung cancer. *Radiother Oncol.* 2005; 77(2):176-181.
34. Wu YL, Li WF, Yang KB, et al. Long-Term evaluation and normal tissue complication probability (Ntcp) models for predicting radiation-induced optic neuropathy after intensity-modulated radiation therapy (IMRT) for nasopharyngeal carcinoma: a large retrospective study in China. *J Oncol.* 2022; 2022:3647462.
35. Konkol M, Śniatała P, Milecki P. Radiation-induced lung injury - what do we know in the era of modern radiotherapy? *Rep Pract Oncol Radiother.* 2022;27(3):552-565.
36. Guo J, Huang X, Dou L, et al. Aging and aging-related diseases: from molecular mechanisms to interventions and treatments. *Signal Transduct Target Ther.* 2022;7(1):391.
37. Sun Q, Chen Y, Li T, et al. Risk and prognosis of secondary esophagus cancer after radiotherapy for breast cancer. *Sci Rep.* 2023;13(1):3968.
38. Clarke M, Collins R, Darby S, et al. Effects of radiotherapy and of differences in the extent of surgery for early breast cancer on local recurrence and 15-year survival: an overview of the randomised trials. *Lancet.* 2005;366:2087-2106.
39. Vougioukalaki M, Demmers J, Vermeij WP, et al. Different responses to DNA damage determine ageing differences between organs. *Aging Cell.* 2022;21(4):e13562.
40. Akthar AS, Golden DW, Nanda R, et al. Early and severe radiation esophagitis associated with concurrent sirolimus. *J Clin Oncol.* 2016;34(9):e73-75.