Prediction of Radiation-induced Lung Injury in Postoperative Breast Cancer Patients with Intensity-modulated Radiotherapy based on Radiomics and Dosiomics

Li Mo, Ling Lin, Liuke Liang, Qingguo Fu, Suning Huang, Haiming Yang*
Department of Radiotherapy, Guangxi Medical University Cancer Hospital, Nanning, Guangxi, China
**Corresponding Author.*

Abstract: This study aims to establish a prediction model for radiation-induced lung injury in postoperative breast cancer patients with intensity-modulated radiotherapy (IMRT) by combining CT-based radiomics with dosiomics. A retrospective analysis was conducted on 79 female breast cancer patients diagnosed and treated at the Affiliated Tumor Hospital of Guangxi Medical University from September 2019 to October 2022. CT images, treatment plans, and clinical follow-up data of these patients were collected. The affected lung was taken as the region of interest (ROI), from which 3384 radiomic and dose radiomic features were extracted. 25 prediction models were constructed using 5 feature extraction methods and 5 classifiers. The 'fsv-Logistic Regression model' emerged as the optimal combination, with the training set AUC values of 0.822, 0.844and 0.898, and validation set AUC values of 0.752, 0.752and 0.914. The combined model demonstrated superior predictive performance compared to individual radiomics or dosiomics models. The results of this study show that, CT-based radiomic features combined with dosiomics enhance the predictive efficacy of radiation-induced lung injury in postoperative breast cancer patients with IMRT.

Keywords: Breast Cancer; Radiation-Induced Lung Injury; Radiomics; Dosiomics

1. Introduction

In 2020, there were 2.2614 million new cases of breast cancer globally, surpassing lung cancer as the most prevalent cancer type worldwide. The incidence of breast cancer in China is rising faster than the global average. Radiotherapy, as a vital component of comprehensive breast cancer treatment, can reduce recurrence, metastasis, and mortality rates of breast cancer and improve local control rates post-surgery. With the continuous improvement and widespread application of intensity-modulated radiotherapy (IMRT), while the dose to the breast target area is significantly increased, the lung tissue also receives a certain dose, leading to varying degrees of radiation-induced lung injury [1]. This affects the quality of life of patients, and has an adverse impact on the treatment outcome [2]. However, there are few definitive indicators to predict the occurrence of radiation-induced lung injury in breast cancer radiotherapy patients.

In recent years, with the machine learning and radiomics technologies are growing, there has been an increasing number of studies on predicting the therapeutic effects and adverse reactions of various diseases using medical imaging features. CT simulation positioning for IMRT radiotherapy preparation is valuable in predicting radiation-induced lung injury caused by IMRT in breast cancer [3, 4]. Additionally, dosimetrics is a method of automatically extracting quantitative features from dose distribution matrices. Dosimetric quantitative features include both one-dimensional DVH features and three-dimensional dose distribution features. Compared to traditional one-dimensional DVH dosimetric features, three-dimensional dose distribution features can reflect deeper spatial information and more effectively describe the impact of dosage on the human body [5]. This study aims to extract pre-radiotherapy positioning CT image features and combine them with three-dimensional dose distribution features in a multimodal fusion radiomics approach to analyze and build a predictive model for breast cancer patients post-surgery undergoing IMRT.

2. Materials and Methods

2.1 Study Subjects

A retrospective collection was conducted of 79 female breast cancer patients who were clinically and pathologically diagnosed from September 2019 to October 2022 at the Affiliated Tumor Hospital of Guangxi Medical University. The age range was 25 to 72 years. All patients had no history of smoking, no tumor recurrence or distant metastasis, no concurrent cardiac or pulmonary diseases, and were staged according to the 7th edition of the AJCC staging system: Stage I (3 cases), Stage II (38 cases), Stage III (38 cases). Inclusion criteria: (1) All patients received intensity-modulated radiotherapy (IMRT) for the first time, with a Karnofsky performance status (KPS) score ≥ 70; (2) Successfully completed the entire radiotherapy process, with complete IMRT planning parameters available; (3) Follow-up time > 3 months.

2.2 CT Image Collection, Target Delineation, and Radiotherapy Implementation

CT Simulation and Target Delineation: Patients were positioned comfortably and fixed for treatment positioning. Continuous scanning with a 5mm slice thickness was performed using a GE CT simulator in a calm breathing state. The range is from the neck to the whole liver. Imaging information was transmitted to the Philips Pinnacle9.0 treatment planning system via LAN. The chief physician delineated the Clinical Target Volume (CTV) and Organs at Risk (OAR) on the patient's positioning CT, including the spinal cord, heart, ipsilateral lung, contralateral lung, and contralateral breast. The Planned Target Volume (PTV) was obtained by expanding the CTV: no expansion in the skin direction, 1.0cm expansion towards the head, feet, medial to sternum, and lateral to axilla, and 0.5cm expansion from the chest wall towards the lung. The radiotherapy plan was set to encompass 100% of the target volume with 95% of the target dose, with a dose calculation grid spacing of 3mm×3mm×3mm. The dose limit for the ipsilateral lung was V20<30%, V30<20%, V5<60%, and could be adjusted according to the patient's condition. The physicist designed the IMRT plan according to the treatment sheet requirements, and the plan was executed after review and approval by the chief physician.

2.3 Diagnosis and Assessment of Radiation-Induced Lung Injury

Clinically, radiation-induced lung injury is diagnosed mainly through imaging examinations such as chest CT. The diagnosis is made by observing inflammatory changes in lung tissue within the irradiation field in patients with symptoms of pneumonia (such as cough, fever, difficulty breathing, etc.), physical signs, and chest X-ray or lung CT findings, or in patients without significant clinical symptoms but with imaging signs of inflammatory changes in the lungs on chest X-ray or lung CT, excluding lung metastases, tuberculosis, etc. Regular follow-ups were conducted for 79 patients, and the acute and chronic radiation-induced lung injuries were evaluated according to the RTOG 0-5 grading standards [6]. Acute and chronic radiation pneumonitis was defined within 1-6 months after radiotherapy and after 6 months, respectively.

Of the 79 postoperative breast cancer patients who successfully completed the planned radiotherapy dose, 55 patients did not develop lung injury and 24 patients developed lung injury, resulting in a radiation-induced lung injury incidence of 30%, all of which were Grade 1 acute radiation-induced lung injury.

2.4 Extraction of Radiomic Features

In this study, the feature extraction and model construction were performed using a self-developed Matlab platform. The region of interest (ROI) for extraction parameter settings was the ipsilateral lung (i.e., the side where the tumor tissue was excised). All CT images and dose distribution data were resampled (2.5mm×2.5mm×2.5mm), and texture features and wavelet transform features were extracted [7]. A total of 3384 features were extracted per case, including Gray Level Co-occurrence Matrix (GLCM, n=24), Gray Level Run Length Matrix (GLRLM, n=16), Gray Level Size Zone Matrix (GLSZM, n=16), Neighbouring Gray Tone Difference Matrix (NGTDM, n=5), Gray Level Dependence Matrix (GLDZM, n=14), and Neighbouring

Gray Level Dependence Matrix (NGLDM, n=5).

2.5 Feature Selection and Model Construction

The data were randomly divided into a training set and a validation set in a 2:1 ratio. For the training set, a t-test was first used for preliminary feature screening, with $P < 0.05$ considered for inclusion. Then, five feature selection algorithms and five classifiers were selected for this study. A model consisted of one feature extraction method and one classifier [8], forming a total of 25 different predictive models. The 18 feature extraction methods were: relieff, fsv, laplacian, L0, lasso [9]. The five different classifiers were: Logistic Regression, SVM, KNN, Decision Tree, Random Forest. For both CT image and dose distribution features, the top 10 features in terms of weight ranking were selected. Simultaneously, multimodal features of radiomics and dosimetrics were fused, and the top ten features in weight ranking were selected for joint model modeling.

2.6 Statistical Methods

The chi-square test and independent sample t-test were used for statistical testing of clinical information. $P<0.05$ was statistical differences. The main indicator for evaluating the predictive ability of the constructed model for radiation-induced lung injury was the area under the receiver operating characteristic curve (AUC).

3. Results

(1) According to the inclusion criteria, 79 patients were included in the analysis, among which 24 cases (30%) had ≥ grade 1 injury. The general patient data is shown in Table 1.

Table1. 79 General Clinical Data of Patients after Breast Cancer

Index	Non-Radiation-induced lung injury(n=55)	With-radiation-induced lung injury(n=24)	Rate (%)	X^2	P
Age (years)				1.027	0.311
<50	32	11	25.6		
≥50	23	13	36.1		
Location				1.206	0.272
Left side	28	9	24.3		
Right side	29	15	35.7		
Clinical Stages				6.397	0.011
Stage I	1	2	66.7		
Stage II	22	16	42.1		
Stage III	32	6	15.8		
Chemotherapy				0.032	0.858
Yes	53	23	30.3		
No	1	1	50.0		
Targeted therapy				0.508	0.476
Yes	25	13	34.2		
No	30	11	26.8		
Endocrinetherapy				1.914	0.167
Yes	35	10	22.2		
No	20	14	41.2		

(2) The training set data were initially screened using a t-test, with features having $P < 0.05$ being included, and the rest were excluded. As a result, 569 features were selected. After a second round of feature selection, the joint model used feature extraction methods to select the optimal top 10 feature subset for input into the classifier for training. The training results of the model were represented as a heat map of AUC values (Figure 1). Figure 1 shows the AUC values corresponding to the 5 feature algorithms and 5 classifiers. Among the 25 models constructed, the highest AUC value was 0.914, corresponding to the combination of 'fsv-Logistics Regression' (considered the optimal model).

(3) Among the 53 patients in the training set, 19 (35.8%) developed radiation-induced lung injury, while in the validation set of 26 patients, 5 (19.2%) developed radiation-induced lung

injury. The ACC and AUC of the models constructed with CT radiomic features, dose distribution features, and combined radiomics-dosimetrics features were as follows: The ACC of the training set was 0.774, 0.774, and 0.811, and in the validation set, the ACC was 0.615, 0.654, and 0.923. The AUC of the training set was 0.822, 0.844, and 0.898, and in the validation set, the AUC was 0.752, 0.752, and 0.914. It is evident that the fusion model's AUC showed significant improvement over the individual feature models, as shown in Figure 2.

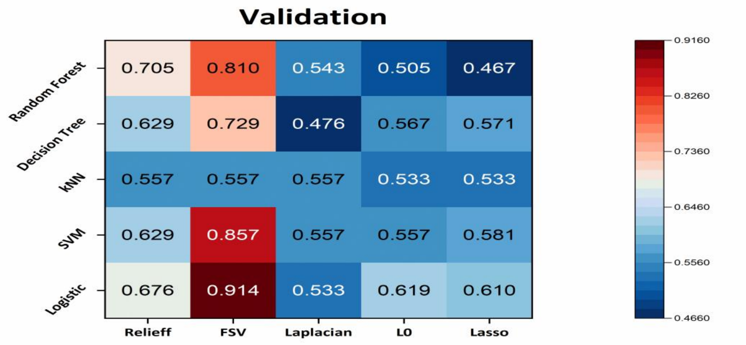

Figure 1. AUC of the Models for Feature Extraction Methods and Classifiers Combination

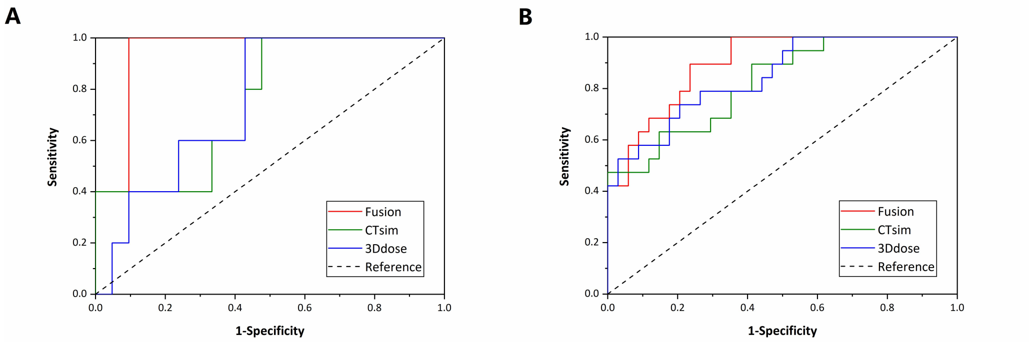

Figure 2. ROC Curves for the Predictors of Symptomatic Radiation Pneumonitis. (A) Training Set; (B) Validation Set

(4) The details of the features extracted by the fusion model are shown in Table 2. Among them, the coefficient for Wavelet-LHL+ GLSZM did not contribute to the model due to the empty estimate value. The features Wavelet-HHH+ GLSZM, Wavelet-LHL+ GLDZM, Origin+GLRLM, and Origin+NGLDM showed significant contributions to the model prediction ($p \leq 0.05$).

Table 2. Selected Top10 Fertures

Selected features	Estimate	tState	pValue
(Intercept)	4.0115	0.7308	0.4649
Wavelet-HLH, NGLDM, Fngl_ldhge	-0.0197	-2.2332	0.0255*
Wavelet-HHH,GLSZM, Fszm_glnu	0.0042	1.2148	0.2244
Wavelet-LHL,GLSZM, Fszm_hgze	0	NaN	NaN
Wavelet-LHL,GLDZM, Fdzm_hgze	-0.0067	-2.5075	0.0122*
Origin,GLRLM, Frlm_lrhge	0.0068	1.9703	0.0488*
Origin,NGLDM, Fngl_hdhge	-0.00003	-0.1755	0.8607
Wavelet-LHH,GLRLM, Frlm_glnu	0.00024	0.5107	0.6095
Wavelet-LHH,GLRLM, Frlm_glnu	-0.00032	-0.6403	0.5220
Origin,NGLDM, Fngl_hdhge	-0.OO077	-2.2077	0.0273*
Wavelet-LLL,GLRLM, Frlm_lrhge	-0.00028	-0.1396	0.8890

4. Conclusion

The incidence of radiation-induced lung injury in patients undergoing radiotherapy after breast cancer surgery is about 3.7%. Therefore, in clinical practice, early identification of risk and protective factors for radiation-induced lung injury, and intervention and prevention based on this, can improve the quality of life for patients. This study retrospectively collected relevant clinical data, CT radiomic data, and dose distribution data from the radiotherapy plan of patients undergoing IMRT after breast cancer surgery, to research the prediction of radiation-induced lung injury, an adverse reaction caused by treatment. It was found that the predictive efficacy of the model constructed from radiomic features, is not significantly different from that in the dosimetry prediction model. However, the predictive efficacy of the combined prediction model is significantly better than that of the single prediction. This suggests that radiomic features have the potential to independently predict radiation-induced lung injury, and combining them with dose distribution factors can enhance clinical benefits for patients.

The features with strong correlations selected in the fusion model of this study include wavelet features and texture features. Among the 8 types of wavelet features obtained through wavelet transformation, the ones labeled as ‘HHH, HLH, LLL, LHH, LHL’ demonstrated significant advantages. Among the texture features selected in this study, GLSZM and GLDZM features describe the regional consistency of dose distribution texture, GLRLM reflects two-dimensional or three-dimensional spatial distribution information, and NGLDM describes the heterogeneity and dependency uniformity of dose distribution. Therefore, this study considers that factors related to radiation-induced lung injury are mainly associated with local dose changes in the ipsilateral lung tissue. It is evident that wavelet transformation, considering time-frequency changes, can more accurately describe the separation of local signal characteristics and are therefore more likely to be selected, while texture features provide information about the spatial relationships of dose distribution and can be used for predicting radiotherapy response.

The factors influencing radiation-induced lung injury after breast cancer surgery are numerous, and selecting the more relevant factors is important. Most existing studies focus on the analysis of single factors in medical statistics or the design and analysis of a single model. This study shows that a combination of multiple model was adopted, which is combinated with radiomics and dosimetrics. The AUC values for radiomics, dosimetrics and fusion model are 0.752, 0.752, and 0.914, respectively. It is evident that compared to independent models, the AUC value of the fusion model has improved, indicating an enhancement in the predictive performance of the fusion model. In this study, the best predictive combination in the fusion model is 'fsv-Logistic Regression', showing that such a combination of feature selection and classifiers could better match the data of CT sim and dose in the breast cance. This finding also indicates that different classifiers have varying discriminative abilities, and different feature extraction methods choose different optimal feature subsets to input into classifiers. This process filters out the most frequently selected key features, which are the most likely critical factors. Such a combination of selection not only identify effective models but also selects good features.

The limitations of this study mainly include: (1) The sample size is too small , which affects the stability of the predictive model and needs further improvement; (2) The generalizability of the model still needs to be confirmed through external validation; (3) The lack of inclusion of clinical treatment method features, such as chemotherapy drugs, immunotherapy drugs, endocrine therapy drugs, etc., and the lack of quantification of the impact of clinical treatment on radiation-induced lung injury.

In summary, this study finds that CT radiomics combined with dosimetry features can effectively predict symptomatic radiation-induced lung injury after intensity-modulated radiotherapy for breast cancer. In future research, it is necessary to expand the sample size and include factors highly related to clinical treatment methods, in order to obtain a more stable and accurate predictive model.

References

[1] Wen G, Tan YT, Lan XW, et al. New clinical features and dosimetric predictor identification for symptomatic radiation pneumonitis after tangential irradiation in breast cancer patients. Cancer, 2017, 8(18): 3795-3802.

[2] Liu Jing, Zhou Dong, He Jiabei, et al.

Research Progress on Factors Affecting Radiation Pneumonia in Breast Cancer Radiotherapy. Journal of Dalian Medical University, 2022, 44(5): 440-445.
[3] Gillies RJ, Kinahan PE, Hricak H. Radiomics images are more than pictures, they are data. Radiology, 2016, 278(2): 563-577.
[4] Van Timmeren JE, Leijenaar RT, Van Elmpt W, et al. Survival prediction of non-small cell lung cancer patients using radiomics analyses of cone-beam CT images. Radiother Oncol, 2017, 123(3): 363-369.
[5] Rossi L, Bijman R, Schillemans W, et al. Texture analysis of 3D dose distributions for predictive modelling of toxicity rates in radiotherapy. Radiother Oncol, 2018, 129(3): 548-553.
[6] CoxJ D, Stetz J, Pajak TF. Toxicity criteria of the Radiation Therapy Oncology Group (RT0G) and the European Organization for Research and Treatment of Cancer (EORTC). Int J Radiat Oncol Biol Phys. 1995, 31(5): 1341-1346.
[7] Peng Lin, Peng-Fei Yang, Shi Chen, et al. A Delta-radiomics model for preoperative evaluation of Neoadjuvant chemotherapy response in high-grade osteosarcoma. Cancer Imaging, 2020, 20(7): 1-12.
[8] Zhou Lu, Wang Linjing, Zhang Guoqian, et al. Prediction of radiation pneumonitis based on radiomics and dosiomics. Chinese Journal of Medical Physics, 2023, 40(7): 808-813.
[9] Roffo G, Melzi S, Castellani U, et al. Infinite Latent Feature Selection: A Probabilistic Latent Graph-Based Ranking Approach. IEEE, 2017.

