Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

# Value of 18F-FDG PET/CT radiomics combined with clinical features for predicting Ki-67 expression in colorectal cancer

**ShuangronD Tian**
Shehong Hospital of Traditional Chinese Medicine

**yu zhang**
zyu0206@163.com

Shehong Hospital of Traditional Chinese Medicine https://orcid.org/0009-0006-2893-3672

**Hu Wang**
Shehong Hospital of Traditional Chinese Medicine

**Research Article**

**Keywords:** colorectal cancer, 18F-FDG PET/CT, radiomics, Ki-67, machine learning

**Posted Date:** November 13th, 2026

**DOI:** https://doi.org/10.21203/rs.3.rs-11577691/v1

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

# Value of 18F-FDG PET/CT radiomics combined with clinical features for predicting Ki-67 expression in colorectal cancer

Shuangrong Tian, Yu Zhang*, Hu Wang

Medical Imaging Center, Shehong Traditional Chinese Medicine Hospital, Shehong 629200, Sichuan Province, China

*Correspondence: Yu Zhang, E-mail: 1287515242@qq.com

## Abstract

Objective: To evaluate the value of 18F-FDG PET/CT radiomics combined with clinical features for the preoperative prediction of Ki-67 expression in colorectal cancer (CRC). Methods: A total of 117 patients with pathologically confirmed CRC who underwent pretreatment 18F-FDG PET/CT between June 2016 and October 2024 were retrospectively included. Patients were classified into high- and low-expression groups using a 50% Ki-67 cutoff and were randomly divided into a training cohort and an internal validation cohort at a 7:3 ratio. Radiomic features were extracted from CT and PET images and reduced by standardization, univariate screening, correlation analysis, and least absolute shrinkage and selection operator (LASSO) regression. Logistic regression (LR), support vector machine (SVM), k-nearest neighbors (KNN), ExtraTrees, LightGBM, and multilayer perceptron (MLP) classifiers were compared. Independent clinical predictors were identified to construct a clinical model and a combined clinical-radiomics nomogram. Model performance was assessed using receiver operating characteristic (ROC) analysis, the DeLong test, and decision curve analysis (DCA). Results: Six CT features, four PET features, and ten combined PET/CT features were retained. TNM stage, SUVmax, and SUVmean were independent predictors of high Ki-67 expression. The LightGBM-based PET/CT radiomics model achieved AUCs of 0.943 and 0.911 in the training and internal validation cohorts, respectively, with an accuracy of 0.861 in the validation cohort. The combined nomogram achieved AUCs of 0.951 and 0.913, with validation-cohort accuracy, sensitivity, specificity, and F1 score of 0.917, 0.920, 0.909, and 0.939, respectively. The AUC difference between the combined model and the PET/CT radiomics model was not statistically significant, whereas DCA showed a higher net benefit for the combined model within part of the clinically relevant threshold range. Conclusion: Multimodal 18F-FDG PET/CT radiomics characterizes CRC from complementary morphologic and metabolic perspectives. Combining radiomics with TNM stage and conventional metabolic parameters may improve preoperative identification of high Ki-67 expression. External validation is required before clinical implementation.

*Keywords: colorectal cancer; 18F-FDG PET/CT; radiomics; Ki-67; machine learning*

# 1 Introduction

Colorectal cancer (CRC) is one of the most common malignant tumors of the digestive system in China, and its disease burden continues to increase [1]. Clinical staging reflects the anatomic extent of the tumor; however, patients at the same stage may differ substantially in treatment response and long-term outcome, indicating that TNM staging alone cannot fully capture tumor biological heterogeneity. Biomarker-assisted risk stratification has therefore become an important direction in precision oncology [2]. Ki-67 is a classic marker of cell proliferation, and its expression level partly reflects the proliferative status of the tumor [3]. Previous studies have shown that Ki-67 expression is associated with clinicopathological features and prognosis in CRC, and higher expression may indicate a worse outcome [4-5]. Accurate assessment of Ki-67 expression before treatment decisions is therefore of practical value for supplementing staging information and identifying patients with more aggressive biological behavior.

Currently, Ki-67 is mainly assessed by immunohistochemistry on biopsy or surgical specimens. Although this approach provides direct pathological evidence, it is limited by sampling scope and can hardly reflect intratumoral spatial heterogeneity, which is itself an important basis of disease progression and treatment response [6]. Radiomics converts routine medical images into quantifiable high-dimensional data and, through feature selection and machine learning, extracts phenotypic information imperceptible to the human eye, providing a new noninvasive approach for assessing tumor heterogeneity and building individualized prediction models [7-9]. 18F-FDG PET/CT provides both glucose metabolic and anatomic information, and PET radiomics has shown value in predicting prognosis and treatment response in CRC [10-11]. However, PET/CT radiomics studies targeting Ki-67 expression remain scarce, particularly studies that compare PET, CT, and fused features within the same framework and further evaluate the incremental value of combining them with clinical variables.

Based on these considerations, we retrospectively analyzed pretreatment 18F-FDG PET/CT images of patients with CRC, established CT, PET, and fused PET/CT radiomics models, and compared the predictive performance of six machine learning classifiers. We further identified clinical predictors independently associated with Ki-67 expression and constructed a clinical model and a combined clinical-radiomics nomogram. The focus was not only on discriminative performance but also on whether multimodal imaging information and conventional clinical variables are complementary, so as to provide a verifiable imaging basis for preoperative noninvasive risk stratification of high Ki-67 expression.

# 2 Materials and Methods

## 2.1 Patients

Patients with CRC who underwent 18F-FDG PET/CT at the Department of Nuclear Medicine, Inner Mongolia Medical University, between June 2016 and October 2024 were retrospectively reviewed. Of 683 initially screened patients, 117 were finally included according to the following criteria.

Inclusion criteria: (1) histopathologically confirmed colonic or rectal adenocarcinoma; (2) 18F-FDG PET/CT performed before treatment; (3) Ki-67 immunohistochemical detection available; and (4) complete clinicopathological data.

Exclusion criteria: (1) prior treatment before PET/CT; (2) history of other malignant tumors; and (3) poor image quality precluding lesion delineation.

## 2.2 Clinical data

Clinical data, including sex, age, body mass index (BMI), tumor location, TNM stage, and Ki-67 expression level, were obtained from the electronic medical record system. Referring to previous multicenter radiomics studies, a cutoff of 50% was used [12]: Ki-67 ≥ 50% was defined as high expression and < 50% as low expression. PET metabolic parameters included SUVmax, SUVmean, metabolic tumor volume (MTV), and total lesion glycolysis (TLG). These parameters were independently measured by two nuclear medicine physicians, and the mean values were used for analysis.

## 2.3 18F-FDG PET/CT imaging

Imaging was performed on a Siemens Biograph MCT PET/CT scanner with a 64-slice spiral CT. 18F-FDG was produced by a Sumitomo HM-20 medical cyclotron with radiochemical purity > 95%. All patients followed a uniform protocol: fasting for at least 6 h, blood glucose < 11.1 mmol/L, intravenous injection of 18F-FDG at 0.10–0.15 mCi/kg body weight, and a rest period of 60 ± 5 min before whole-body PET/CT acquisition.

## 2.4 Region-of-interest segmentation

Accurate region-of-interest (ROI) segmentation is the basis of radiomic feature extraction and model construction. Display parameters were standardized before segmentation, with a CT window width of 350 HU and window level of 50 HU. A threshold of 40% of SUVmax was used for semi-automatic PET segmentation. An experienced nuclear medicine physician delineated the three-dimensional ROI of the primary lesion slice by slice on CT and PET images, and a senior physician reviewed all contours. Disagreements were resolved by discussion. An example of ROI segmentation in a patient with left-sided colon cancer is shown in Figure 1.

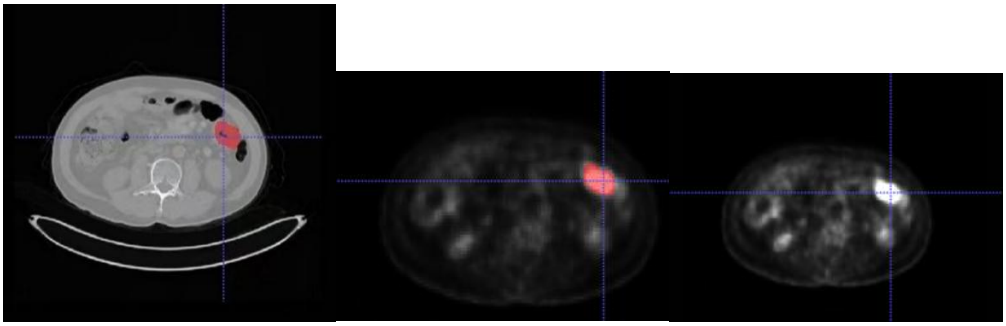


**Figure 1 Example of ROI segmentation in a patient with left-sided colon cancer.**

(A) CT image with ROI; (B) PET image with ROI; (C) fused PET/CT image.

## 2.5 Radiomic feature extraction

Following the Image Biomarker Standardisation Initiative, the original CT and PET images and ROIs were imported into the open-source PyRadiomics package for feature extraction. Seven feature classes were extracted: first-order statistics, shape features, gray-level co-occurrence matrix (GLCM), gray-level size-zone matrix (GLSZM), gray-level dependence matrix (GLDM), gray-level run-length matrix (GLRLM), and neighboring gray-tone difference matrix (NGTDM).

## 2.6 Feature selection and model construction

### 2.6.1 Feature selection

Radiomic feature selection reduces computational complexity, improves model performance, and lowers the risk of overfitting. The following workflow was applied:

- Z-score standardization was applied to the raw data, transforming each feature to a mean of 0 and a standard deviation of 1 to prevent overfitting.
- Independent-samples t tests were used for preliminary screening, and features with $P < 0.05$ were retained.
- Pearson correlation analysis was then performed; when the correlation coefficient between two features exceeded 0.9, only one was retained to reduce redundancy.
- LASSO regression was used for dimensionality reduction, with the optimal penalty coefficient $\lambda$ determined by 10-fold cross-validation. LASSO compresses the coefficients of low-contribution features to zero through L1 regularization. The features with nonzero coefficients were retained and used to calculate the radiomics score (RS) for subsequent model construction.

### 2.6.2 Model construction

Because patients with low Ki-67 expression accounted for a smaller proportion, synthetic minority oversampling technique (SMOTE) was applied to the training set only to alleviate class imbalance during classifier training; the internal validation set remained untouched. The selected radiomic features were separately input into six classifiers—SVM, LR, KNN, ExtraTrees, LightGBM, and MLP—to construct CT, PET, and PET/CT radiomics models, and model performance was evaluated in the internal validation cohort. In addition, univariate and multivariate logistic regression analyses of clinical features were performed to identify independent predictors for the clinical model, and the clinical factors were combined with the PET/CT radiomics score to build a nomogram.

## 2.7 Model validation and evaluation

For each patient in the internal validation cohort, the predicted probability of high Ki-67 expression was calculated, and ROC and decision curves were plotted. Model performance was comprehensively evaluated using the area under the curve (AUC), accuracy, precision, sensitivity, specificity, and F1 score. Accuracy is the proportion of correctly classified samples; precision is the proportion of true positives among predicted positives; sensitivity is the proportion of correctly identified positives; specificity is the proportion of correctly identified negatives; and the F1 score is the harmonic mean of precision and sensitivity.

## 2.8 Statistical analysis

Statistical analyses were performed with SPSS Statistics 25.0 and Python 3.11.5. Normally distributed continuous variables are presented as mean ± standard deviation, non-normally distributed variables as median with interquartile range, and categorical variables as counts with percentages. Independent-samples t tests were used for normally distributed continuous variables with homogeneous variance, Mann-Whitney U tests for non-normally distributed variables, and chi-square tests or Fisher exact tests as appropriate for categorical variables. Discriminative performance was evaluated by ROC analysis and AUC, and differences between AUCs were compared with the DeLong test. Two-sided $P < 0.05$ was considered statistically significant.

# 3 Results

## 3.1 Baseline characteristics

A total of 117 patients with CRC were included, comprising 80 colon cancers and 37 rectal cancers. High Ki-67 expression was present in 86 patients and low expression in 31. After 7:3 random division, the training cohort comprised 81 patients (61 high, 20 low expression) and the internal validation cohort 36 patients (25 high, 11 low expression). Baseline characteristics are shown in Table 1. TNM stage differed significantly between expression groups in both cohorts. SUVmax, SUVmean, MTV, and TLG differed significantly in the training cohort but not in the validation cohort.

**Table 1 Baseline clinical characteristics of the training and validation cohorts**

| Characteristic | High (n=61) | Low (n=20) | P | High (n=25) | Low (n=11) | P |
|---|---|---|---|---|---|---|
| Sex | | | 1.000 | | | 0.364 |
| Male | 35 (57.38) | 12 (60.00) | | 17 (68.00) | 5 (45.45) | |
| Female | 26 (42.62) | 8 (40.00) | | 8 (32.00) | 6 (54.55) | |
| Age (years) | 67.54±10.71 | 65.00±8.55 | 0.192 | 63.84±13.26 | 57.55±13.37 | 0.199 |
| BMI (kg/m2) | 23.00±3.34 | 22.57±2.86 | 0.614 | 24.18±3.69 | 23.99±4.02 | 0.891 |
| Tumor location | | | 0.926 | | | 0.696 |
| Right colon | 21 (34.43) | 6 (30.00) | | 5 (20.00) | 1 (9.09) | |
| Left colon | 22 (36.07) | 8 (40.00) | | 11 (44.00) | 6 (54.55) | |
| Rectum | 18 (29.51) | 6 (30.00) | | 9 (36.00) | 4 (36.36) | |
| TNM stage | | | 0.002 | | | 0.006 |
| II | 6 (9.84) | 9 (45.00) | | 2 (8.00) | 6 (54.55) | |
| III | 21 (34.43) | 3 (15.00) | | 9 (36.00) | 1 (9.09) | |
| IV | 34 (55.74) | 8 (40.00) | | 14 (56.00) | 4 (36.36) | |
| SUVmax | 18.57±12.64 | 10.90±4.18 | <0.001 | 14.25±5.82 | 12.63±5.02 | 0.398 |
| SUVmean | 9.20±3.10 | 7.80±2.91 | <0.01 | 6.80±3.46 | 8.80±2.17 | 0.150 |
| MTV (cm3) | 28.42±31.54 | 17.04±14.92 | <0.01 | 13.19±13.03 | 13.75±11.11 | 0.744 |
| TLG (g) | 261.44±340.57 | 132.91±130.25 | <0.01 | 89.69±80.24 | 121.97±115.35 | 0.520 |

Note: Training cohort on the left (columns 2–4), validation cohort on the right (columns 5–7). Continuous data are mean ± SD; categorical data are n (%).

## 3.2 Univariate and multivariate analyses of clinical features

The results of univariate and multivariate logistic regression are shown in Table 2. TNM stage, SUVmax, and SUVmean remained independently associated with high Ki-67 expression in multivariate analysis and were therefore included in the clinical model. The associations of the other variables weakened after multivariable adjustment.

**Table 2 Univariate and multivariate logistic regression of clinical features for Ki-67 expression**

| Characteristic | Univariate OR | 95% CI | P | Multivariate OR | 95% CI | P |
|---|---|---|---|---|---|---|
| TLG (g) | 1.004 | 1.002–1.006 | <0.01 | 1.001 | 0.993–1.009 | 0.812 |
| Age (years) | 1.016 | 1.011–1.022 | <0.01 | 1.003 | 0.973–1.034 | 0.860 |
| MTV (cm3) | 1.042 | 1.023–1.061 | <0.01 | 1.009 | 0.939–1.083 | 0.843 |
| BMI (kg/m2) | 1.045 | 1.029–1.060 | <0.01 | 0.967 | 0.891–1.050 | 0.504 |
| SUVmax | 1.050 | 1.015–1.085 | 0.005 | 1.120 | 1.025–1.225 | 0.013 |
| SUVmean | 1.115 | 1.070–1.162 | <0.01 | 1.175 | 1.045–1.321 | 0.017 |
| TNM stage | 1.414 | 1.270–1.575 | <0.01 | 1.857 | 1.190–2.898 | 0.022 |
| Tumor location | 1.812 | 1.384–2.373 | <0.01 | 0.804 | 0.481–1.344 | 0.486 |
| Sex | 2.429 | 1.441–4.096 | 0.005 | 0.656 | 0.295–1.458 | 0.385 |

## 3.3 Radiomics model based on CT features

### 3.3.1 Feature selection

A total of 1834 features were extracted from the CT images of the training cohort. After t-test screening ($P \leq 0.05$), 43 features were retained; after Pearson correlation filtering ($|r| > 0.9$), 17 features remained; and after LASSO-logistic regression, 6 features associated with Ki-67 expression were finally selected. The feature selection process is shown in Figure 2 and the retained features in Table 3.

The CT radiomics score (CT RS) of each patient was calculated from the LASSO coefficients as follows: CT RS = 0.7531 + 0.008408 × exponential_CTfirstorder_CTRootMeanSquared + 0.004455 × exponential_CTglcm_CTSumEntropy + 0.000752 × exponential_CTglszm_CTGrayLevelVariance + 0.025637 × exponential_CTglszm_CTSizeZoneNonUniformity + 0.051792 × square_CTfirstorder_CTMaximum + 0.067607 × wavelet_CTHLL_CTfirstorder_CTSkewness.

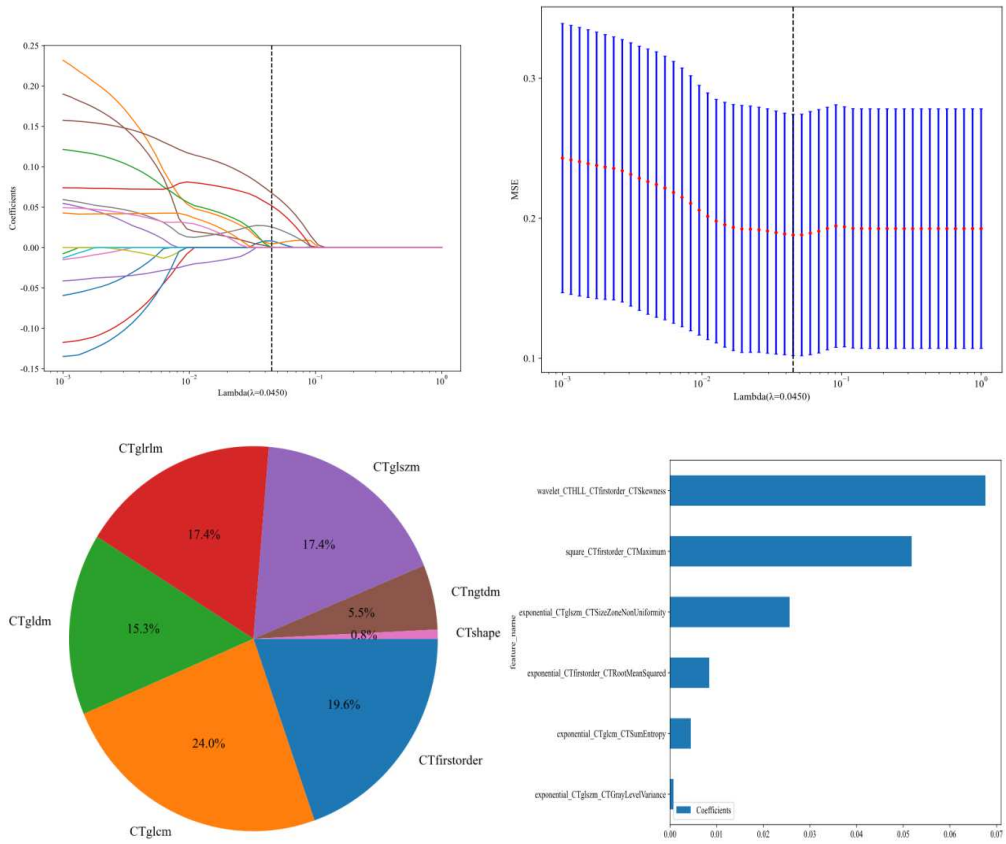



**Figure 2 Radiomic feature selection based on CT images.**

(A) LASSO coefficient paths; (B) cross-validation error with optimal λ = 0.0450; (C) composition of initial features; (D) the six retained features and their weights.

**Table 3 Retained CT radiomic features and corresponding coefficients**

| Feature | Coefficient |
|---|---|
| wavelet_CTHLL_CTfirstorder_CTSkewness | 0.067607 |
| square_CTfirstorder_CTMaximum | 0.051792 |
| exponential_CTfirstorder_CTRootMeanSquared | 0.008408 |
| exponential_CTglcm_CTSumEntropy | 0.004455 |

| | |
|---|---|
| exponential_CTglszm_CTGrayLevelVariance | 0.000752 |
| exponential_CTglszm_CTSizeZoneNonUniformity | 0.025637 |

### 3.3.2 Model performance

Six classifiers were built on the CT radiomic features (Figure 3, Table 4). KNN achieved the highest training AUC but dropped to 0.636 in the validation cohort with an accuracy of only 0.306, indicating poor stability. SVM performed well in training but likewise declined in validation. LightGBM achieved training and validation AUCs of 0.863 and 0.775 with relatively balanced validation accuracy, precision, and specificity (0.639, 0.929, and 0.909). LightGBM was therefore used to construct the CT RS model.

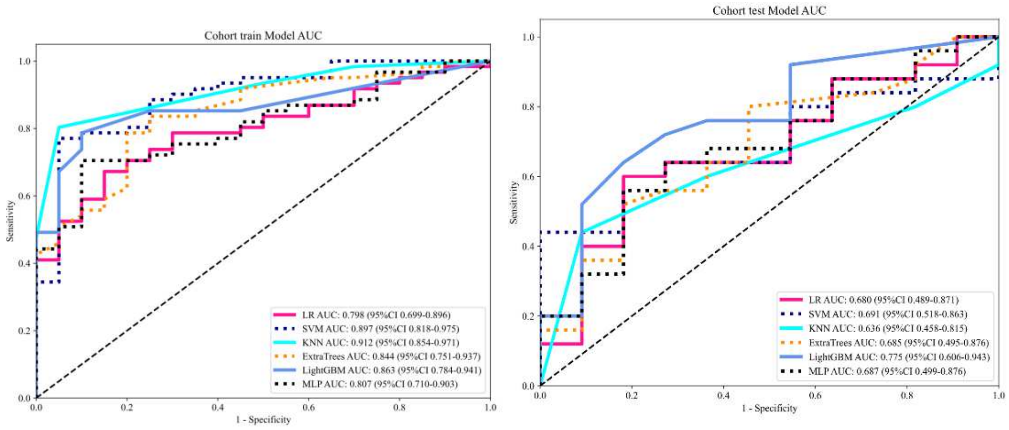



**Figure 3 ROC curves of different machine learning models based on CT features.**

(A) Training cohort; (B) internal validation cohort.

**Table 4 Diagnostic performance of CT-feature models in the training and validation cohorts**

| Model | AUC (95% CI) | Accuracy | Precision | Sensitivity | Specificity | F1 |
|---|---|---|---|---|---|---|
| LR-Training | 0.798 (0.699–0.896) | 0.704 | 0.930 | 0.656 | 0.850 | 0.769 |
| LR-Validation | 0.680 (0.489–0.871) | 0.639 | 0.875 | 0.560 | 0.818 | 0.683 |
| SVM-Training | 0.897 (0.818–0.975) | 0.802 | 0.979 | 0.754 | 0.950 | 0.852 |
| SVM-Validation | 0.691 (0.519–0.853) | 0.583 | 1.000 | 0.400 | 1.000 | 0.571 |
| KNN-Training | 0.912 (0.854–0.971) | 0.605 | 1.000 | 0.475 | 1.000 | 0.644 |
| KNN-Validation | 0.636 (0.458–0.815) | 0.306 | 0.000 | 0.000 | 1.000 | 0.000 |
| ExtraTrees-Training | 0.844 (0.751–0.937) | 0.753 | 0.918 | 0.738 | 0.800 | 0.818 |
| ExtraTrees-Validation | 0.685 (0.495–0.876) | 0.694 | 0.792 | 0.760 | 0.545 | 0.776 |
| LightGBM-Training | 0.863 (0.785–0.941) | 0.790 | 0.958 | 0.754 | 0.900 | 0.844 |
| LightGBM-Validation | 0.775 (0.606–0.943) | 0.639 | 0.929 | 0.520 | 0.909 | 0.667 |
| MLP-Training | 0.807 (0.710–0.903) | 0.741 | 0.955 | 0.689 | 0.900 | 0.769 |
| MLP-Validation | 0.687 (0.499–0.876) | 0.611 | 0.867 | 0.520 | 0.818 | 0.683 |

## 3.4 Radiomics model based on PET features

### 3.4.1 Feature selection

A total of 2016 features were extracted from the PET images of the training cohort. After the same selection workflow, 4 features associated with Ki-67 expression were finally retained (Figure 4, Table 5).

The PET radiomics score (PET RS) was calculated as: PET RS = 0.7531 − 0.074300 × lbp_PET3D_PETm2_PETfirstorder_PETKurtosis − 0.003722 × logarithm_PETfirstorder_PETSkewness + 0.026056 × wavelet_PETHLH_PETfirstorder_PETMean − 0.028711 × wavelet_PETLHL_PETfirstorder_PETMean.

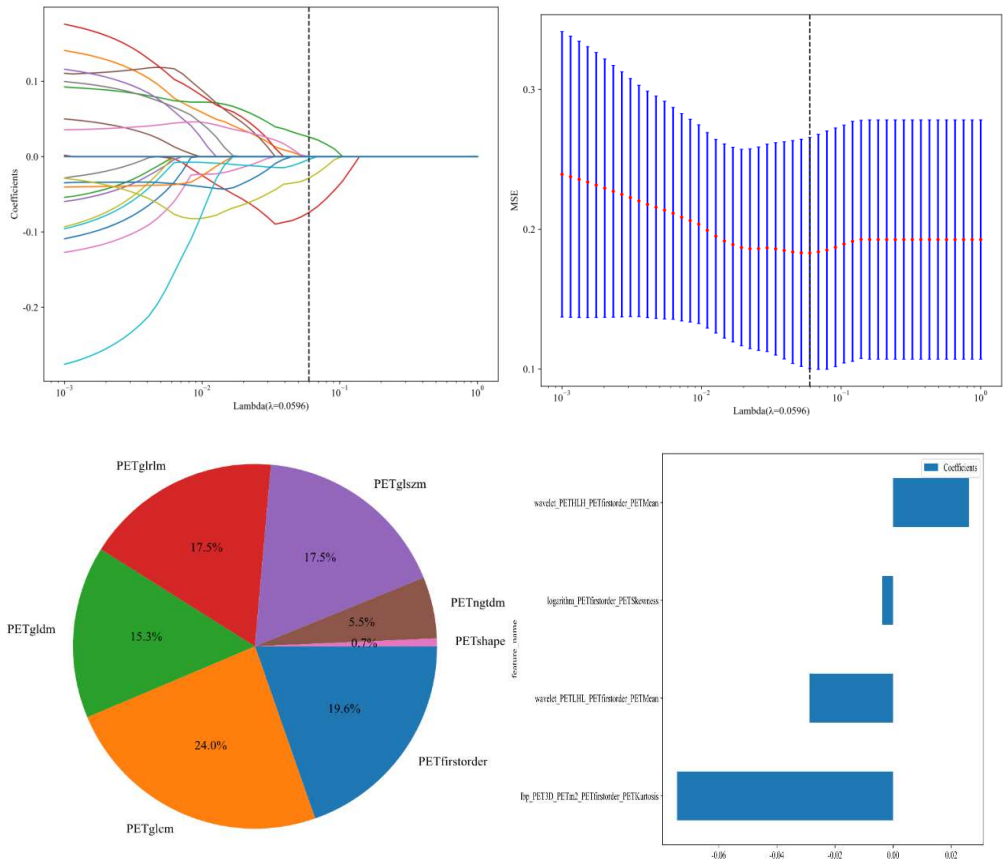



**Figure 4 Radiomic feature selection based on PET images.**

(A) LASSO coefficient paths; (B) cross-validation error with optimal λ = 0.0596; (C) composition of initial features; (D) the four retained features and their weights.

**Table 5 Retained PET radiomic features and corresponding coefficients**

| Feature | Coefficient |
|---|---|
| lbp_PET3D_PETm2_PETfirstorder_PETKurtosis | −0.074300 |
| logarithm_PETfirstorder_PETSkewness | −0.003722 |
| wavelet_PETHLH_PETfirstorder_PETMean | 0.026056 |
| wavelet_PETLHL_PETfirstorder_PETMean | −0.028711 |

### 3.4.2 Model performance

Six classifiers were built on the PET radiomic features (Figure 5, Table 6). ExtraTrees and SVM achieved high training AUCs but declined markedly in validation. LightGBM achieved training and validation AUCs of 0.872 and 0.727, with validation accuracy, precision, sensitivity, specificity, and F1 of 0.722, 0.857, 0.720, 0.727, and 0.783. Considering both discrimination and stability, LightGBM was used to construct the PET RS model.

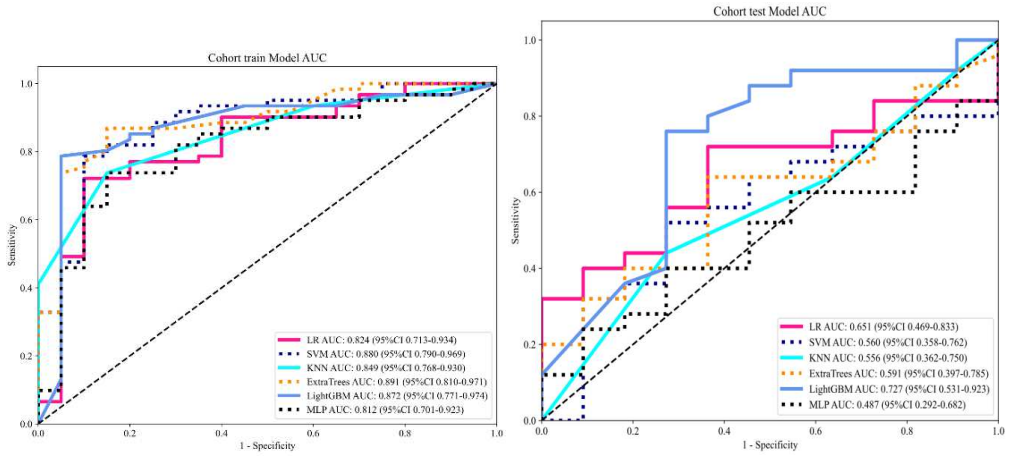



**Figure 5 ROC curves of different machine learning models based on PET features.**

(A) Training cohort; (B) internal validation cohort.

**Table 6 Diagnostic performance of PET-feature models in the training and validation cohorts**

| Model | AUC (95% CI) | Accuracy | Precision | Sensitivity | Specificity | F1 |
|---|---|---|---|---|---|---|
| LR-Training | 0.824 (0.713–0.935) | 0.753 | 0.956 | 0.705 | 0.900 | 0.811 |
| LR-Validation | 0.651 (0.469–0.833) | 0.667 | 0.810 | 0.680 | 0.636 | 0.739 |
| SVM-Training | 0.880 (0.790–0.969) | 0.802 | 0.959 | 0.770 | 0.900 | 0.855 |
| SVM-Validation | 0.560 (0.358–0.762) | 0.556 | 0.800 | 0.480 | 0.727 | 0.600 |
| KNN-Training | 0.849 (0.768–0.930) | 0.556 | 1.000 | 0.410 | 1.000 | 0.581 |
| KNN-Validation | 0.556 (0.362–0.750) | 0.306 | 0.000 | 0.000 | 1.000 | 0.000 |
| ExtraTrees-Training | 0.891 (0.810–0.971) | 0.852 | 0.945 | 0.852 | 0.850 | 0.897 |
| ExtraTrees-Validation | 0.591 (0.397–0.785) | 0.611 | 0.789 | 0.600 | 0.636 | 0.682 |
| LightGBM-Training | 0.872 (0.771–0.974) | 0.790 | 0.978 | 0.738 | 0.950 | 0.841 |
| LightGBM-Validation | 0.727 (0.531–0.924) | 0.722 | 0.857 | 0.720 | 0.727 | 0.783 |
| MLP-Training | 0.812 (0.701–0.923) | 0.753 | 0.936 | 0.721 | 0.850 | 0.815 |
| MLP-Validation | 0.487 (0.292–0.682) | 0.417 | 0.833 | 0.200 | 0.909 | 0.323 |

## 3.5 Radiomics model based on fused PET/CT features

### 3.5.1 Feature selection

The PET and CT features were fused and processed through the same selection workflow, yielding 10 features for model construction (Figure 6, Table 7).

The PET/CT radiomics score (PET/CT RS) was calculated as: PET/CT RS = 0.7531 + 0.013131 × gradient_PETglszm_PETSmallAreaEmphasis − 0.068530 × lbp_PET3D_PETm2_PETfirstorder_PETKurtosis − 0.013586 × logarithm_PETfirstorder_PETSkewness + 0.027223 × wavelet_PETHLH_PETfirstorder_PETMean + 0.009036 × wavelet_PETLHH_PETfirstorder_PETMedian − 0.019933 × wavelet_PETLHL_PETfirstorder_PETMean + 0.02176 × exponential_CTfirstorder_CTRootMeanSquared + 0.000865 × exponential_CTglszm_CTSizeZoneNonUniformity + 0.048740 × square_CTfirstorder_CTMaximum + 0.054845 × wavelet_CTHLL_CTfirstorder_CTSkewness.

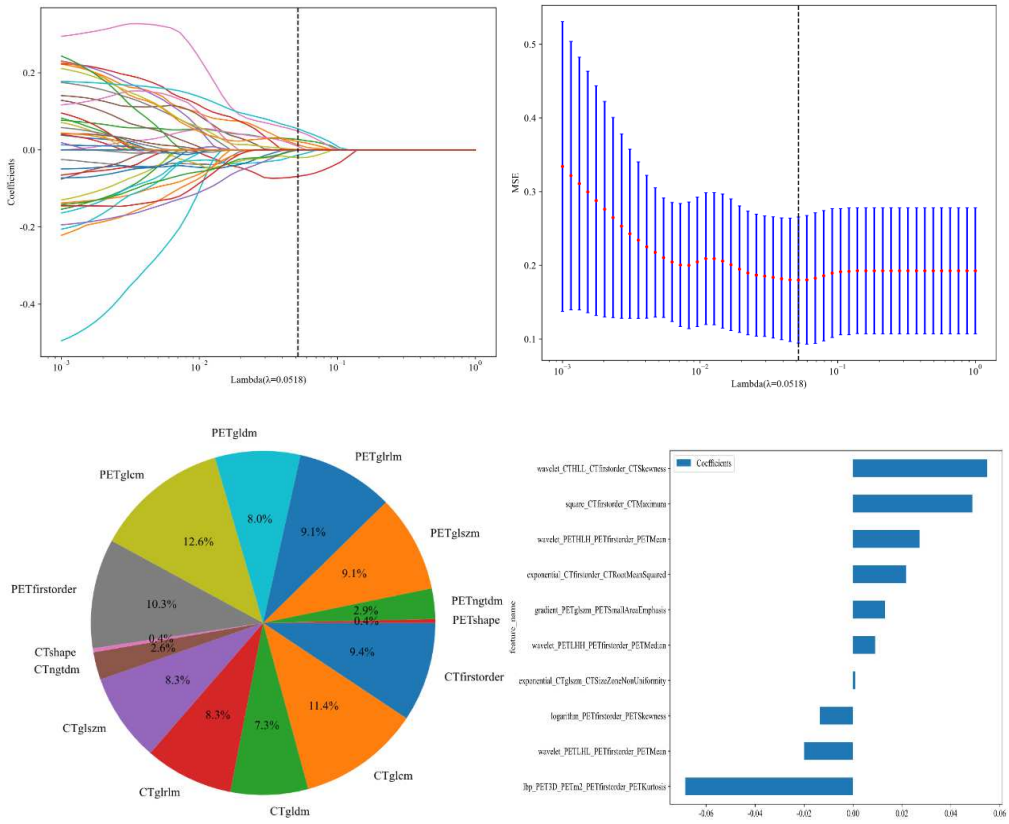



**Figure 6 Radiomic feature selection based on fused PET/CT features.**

(A) LASSO coefficient paths; (B) cross-validation; (C) composition of initial features; (D) the ten retained features.

**Table 7 Retained PET/CT radiomic features and corresponding coefficients**

| Feature | Coefficient |
|---|---|
| gradient_PETglszm_PETSmallAreaEmphasis | 0.013131 |
| lbp_PET3D_PETm2_PETfirstorder_PETKurtosis | −0.068530 |
| logarithm_PETfirstorder_PETSkewness | −0.013586 |
| wavelet_PETHLH_PETfirstorder_PETMean | 0.027223 |
| wavelet_PETLHH_PETfirstorder_PETMedian | 0.009036 |

| | |
|---|---|
| wavelet_PETLHL_PETfirstorder_PETMean | −0.019933 |
| exponential_CTfirstorder_CTRootMeanSquared | 0.02176 |
| exponential_CTglszm_CTSizeZoneNonUniformity | 0.000865 |
| square_CTfirstorder_CTMaximum | 0.048740 |
| wavelet_CTHLL_CTfirstorder_CTSkewness | 0.054845 |

### 3.5.2 Model performance

Six classifiers were built on the 10 fused PET/CT features (Figure 7, Table 8). LightGBM achieved the highest AUC in both the training and validation cohorts (0.943 and 0.911), with accuracies of 0.877 and 0.861 and consistently high precision, sensitivity, specificity, and F1. LightGBM was therefore selected as the final classifier for the PET/CT RS model.

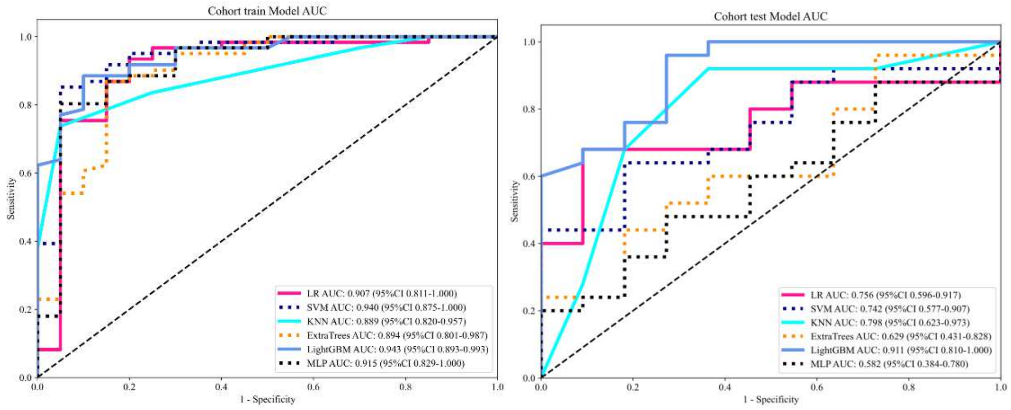



**Figure 7 ROC curves of different machine learning models based on fused PET/CT features.**

(A) Training cohort; (B) internal validation cohort.

**Table 8 Diagnostic performance of PET/CT-feature models in the training and validation cohorts**

| Model | AUC (95% CI) | Accuracy | Precision | Sensitivity | Specificity | F1 |
|---|---|---|---|---|---|---|
| LR-Training | 0.907 (0.811–1.000) | 0.889 | 0.933 | 0.918 | 0.800 | 0.926 |
| LR-Validation | 0.756 (0.596–0.917) | 0.722 | 0.941 | 0.640 | 0.909 | 0.762 |
| SVM-Training | 0.940 (0.875–1.000) | 0.864 | 0.981 | 0.836 | 0.950 | 0.903 |
| SVM-Validation | 0.742 (0.577–0.907) | 0.667 | 0.882 | 0.600 | 0.818 | 0.714 |
| KNN-Training | 0.889 (0.820–0.957) | 0.531 | 1.000 | 0.377 | 1.000 | 0.548 |
| KNN-Validation | 0.798 (0.623–0.973) | 0.722 | 0.895 | 0.680 | 0.818 | 0.773 |
| ExtraTrees-Training | 0.894 (0.801–0.987) | 0.852 | 0.945 | 0.852 | 0.850 | 0.897 |
| ExtraTrees-Validation | 0.629 (0.431–0.828) | 0.528 | 0.833 | 0.400 | 0.818 | 0.541 |
| LightGBM-Training | 0.943 (0.893–0.994) | 0.877 | 0.964 | 0.869 | 0.900 | 0.914 |
| LightGBM-Validation | 0.911 (0.811–1.000) | 0.861 | 0.885 | 0.920 | 0.727 | 0.902 |
| MLP-Training | 0.915 (0.829–1.000) | 0.827 | 0.980 | 0.787 | 0.950 | 0.873 |
| MLP-Validation | 0.582 (0.384–0.780) | 0.528 | 0.786 | 0.440 | 0.727 | 0.564 |

## 3.6 Comparison of the CT RS, PET RS, and PET/CT RS models

The performance of the three radiomics models is shown in Table 9 and Figure 8. The PET/CT RS model achieved training and validation AUCs of 0.943 and 0.911, higher than the corresponding CT RS and PET RS models, with balanced accuracy, sensitivity, and F1, suggesting complementary value of PET and CT information fusion.

DeLong tests were used to compare the AUCs of the three models (Figure 9). The difference between the PET/CT RS model and the CT RS model was statistically significant in both the training and validation cohorts, whereas the differences among the other models did not reach statistical significance.

**Table 9 Diagnostic performance of the CT RS, PET RS, and PET/CT RS models**

| Model | AUC (95% CI) | Accuracy | Precision | Sensitivity | Specificity | F1 |
|---|---|---|---|---|---|---|
| CT RS-Training | 0.863 (0.785–0.941) | 0.790 | 0.958 | 0.754 | 0.900 | 0.844 |
| CT RS-Validation | 0.775 (0.606–0.943) | 0.639 | 0.929 | 0.520 | 0.909 | 0.667 |
| PET RS-Training | 0.872 (0.771–0.974) | 0.790 | 0.978 | 0.738 | 0.950 | 0.841 |
| PET RS-Validation | 0.727 (0.531–0.924) | 0.722 | 0.857 | 0.720 | 0.727 | 0.783 |
| PET/CT RS-Training | 0.943 (0.893–0.994) | 0.877 | 0.964 | 0.869 | 0.900 | 0.914 |
| PET/CT RS-Validation | 0.911 (0.811–1.000) | 0.861 | 0.885 | 0.920 | 0.727 | 0.902 |

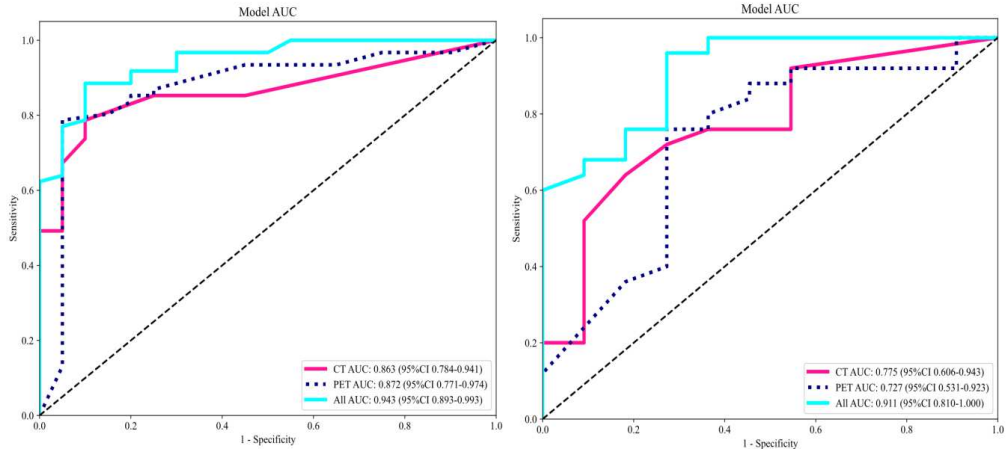



**Figure 8 ROC curves of the CT RS, PET RS, and PET/CT RS models.**

(A) Training cohort; (B) internal validation cohort.

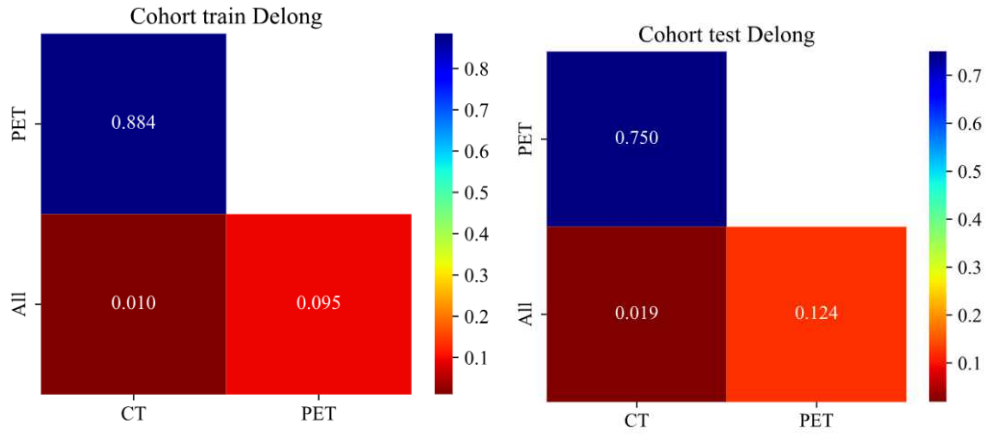



**Figure 9 DeLong test results for the CT RS, PET RS, and PET/CT RS models.**

(A) Training cohort; (B) internal validation cohort; All denotes the PET/CT RS model.

Decision curve analysis was further used to evaluate the potential clinical benefit of the three models (Figure 10). All models yielded positive net benefit within part of the threshold probability range, and the PET/CT RS model showed the most stable overall performance.

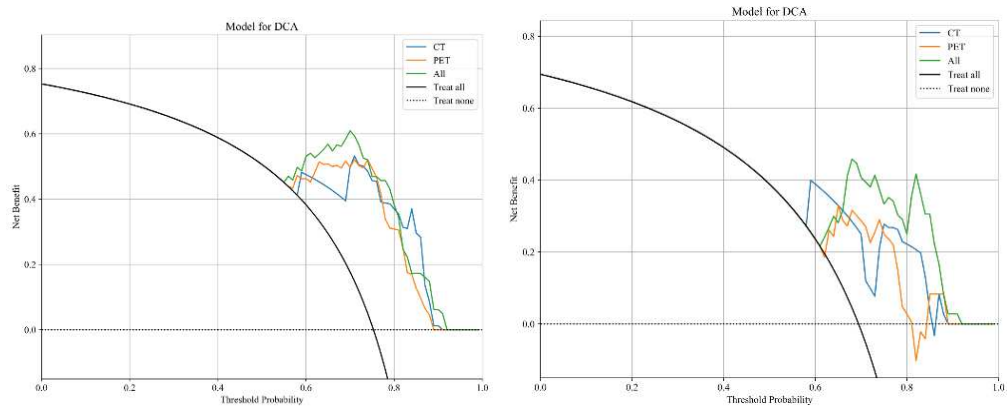



**Figure 10 Decision curves of the CT RS, PET RS, and PET/CT RS models.**

(A) Training cohort; (B) internal validation cohort; All denotes the PET/CT RS model.

## 3.7 Clinical model and combined clinical-radiomics nomogram

Based on the multivariate logistic regression results, TNM stage, SUVmax, and SUVmean were included in the clinical model. Six classifiers were compared (Figure 11, Table 10). LightGBM achieved training and validation AUCs of 0.856 and 0.755, with relatively high validation performance and close training-validation agreement, and was therefore used to construct the final clinical model.

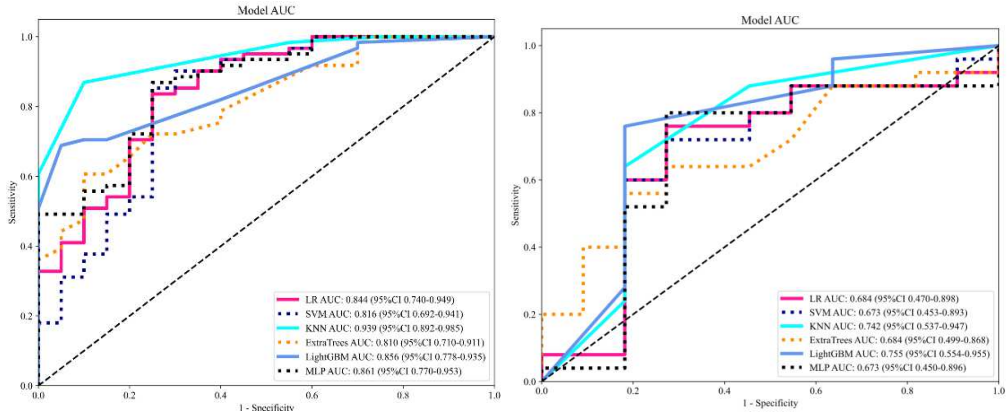



**Figure 11 ROC curves of different machine learning models based on clinical features.**

(A) Training cohort; (B) internal validation cohort.

**Table 10 Diagnostic performance of clinical-feature models in the training and validation cohorts**

| Model | AUC (95% CI) | Accuracy | Precision | Sensitivity | Specificity | F1 |
|---|---|---|---|---|---|---|
| LR-Training | 0.844 (0.740–0.949) | 0.802 | 0.909 | 0.820 | 0.750 | 0.862 |
| LR-Validation | 0.684 (0.470–0.898) | 0.722 | 0.857 | 0.720 | 0.727 | 0.783 |
| SVM-Training | 0.816 (0.692–0.941) | 0.815 | 0.911 | 0.836 | 0.750 | 0.872 |
| SVM-Validation | 0.673 (0.453–0.893) | 0.694 | 0.850 | 0.680 | 0.727 | 0.756 |
| KNN-Training | 0.939 (0.893–0.986) | 0.704 | 1.000 | 0.607 | 1.000 | 0.755 |
| KNN-Validation | 0.742 (0.537–0.947) | 0.417 | 0.750 | 0.240 | 0.818 | 0.364 |
| ExtraTrees-Training | 0.810 (0.710–0.911) | 0.667 | 0.947 | 0.590 | 0.900 | 0.727 |
| ExtraTrees-Validation | 0.684 (0.499–0.868) | 0.583 | 0.857 | 0.480 | 0.818 | 0.615 |
| LightGBM-Training | 0.856 (0.778–0.935) | 0.630 | 1.000 | 0.508 | 1.000 | 0.674 |
| LightGBM-Validation | 0.755 (0.554–0.955) | 0.611 | 0.867 | 0.520 | 0.818 | 0.650 |
| MLP-Training | 0.861 (0.770–0.953) | 0.827 | 0.912 | 0.852 | 0.750 | 0.881 |
| MLP-Validation | 0.673 (0.450–0.896) | 0.750 | 0.864 | 0.760 | 0.727 | 0.809 |

The PET/CT RS was further combined with TNM stage, SUVmax, and SUVmean to build a clinical-radiomics nomogram (Figure 12). The PET/CT RS carried a high weight in the nomogram, indicating an important contribution of radiomic information to the combined prediction.

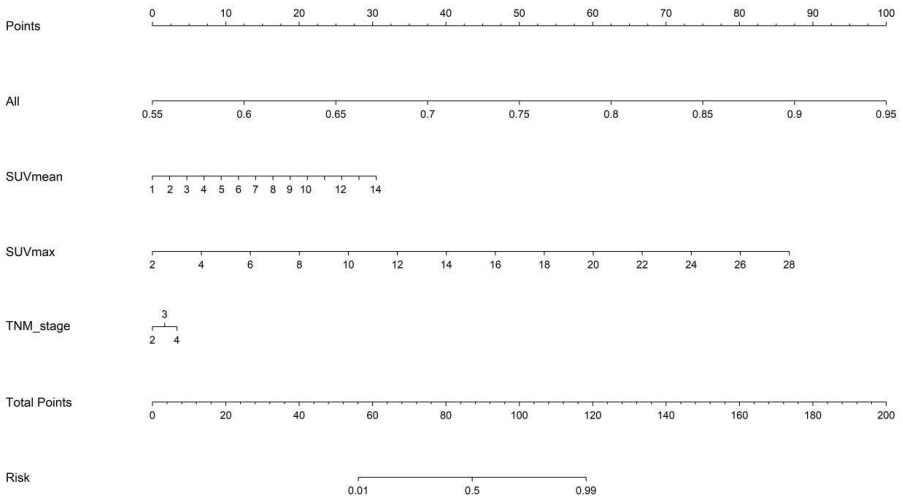



**Figure 12 Clinical-radiomics nomogram combining PET/CT RS, TNM stage, SUVmax, and SUVmean.**

All denotes the PET/CT RS; TNM_stage denotes TNM stage.

The ROC curves of the clinical model, PET/CT RS model, and combined nomogram are shown in Figure 13, and performance metrics in Table 11. The training AUCs of the three models were 0.856, 0.943, and 0.951, and the validation AUCs were 0.755, 0.911, and 0.913, respectively. The combined model achieved validation accuracy, sensitivity, specificity, and F1 of 0.917, 0.920, 0.909, and 0.939. DeLong tests showed that the AUC differences among the three models did not reach statistical significance (Figure 14). Decision curves are shown in Figure 15: the combined model yielded higher net benefit within part of the threshold range, suggesting complementarity between clinical features and PET/CT radiomic features.

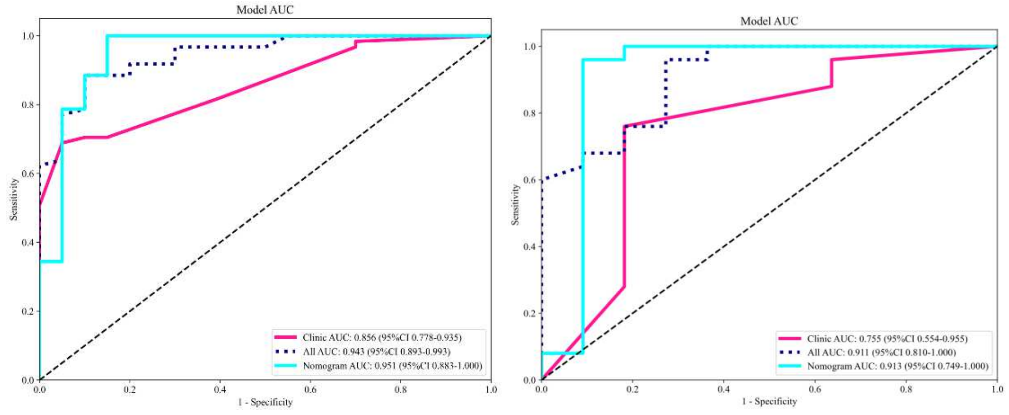



**Figure 13 ROC curves of the clinical model, PET/CT RS model, and combined nomogram.**

(A) Training cohort; (B) internal validation cohort; Clinic denotes the clinical model; All denotes the PET/CT RS model; Nomogram denotes the combined model.

**Table 11 Diagnostic performance of the clinical model, PET/CT RS model, and combined nomogram**

| Model | AUC (95% | Accuracy | Precision | Sensitivity | Specificity | F1 |
|---|---|---|---|---|---|---|

| | CI) | | | | | |
|---|---|---|---|---|---|---|
| Clinical-Training | 0.856 (0.778–0.935) | 0.630 | 1.000 | 0.508 | 1.000 | 0.674 |
| Clinical-Validation | 0.755 (0.554–0.955) | 0.611 | 0.867 | 0.520 | 0.818 | 0.650 |
| PET/CT RS-Training | 0.943 (0.893–0.994) | 0.877 | 0.964 | 0.869 | 0.900 | 0.914 |
| PET/CT RS-Validation | 0.911 (0.811–1.000) | 0.861 | 0.885 | 0.920 | 0.727 | 0.902 |
| Nomogram-Training | 0.951 (0.883–1.000) | 0.951 | 0.885 | 0.984 | 0.850 | 0.902 |
| Nomogram-Validation | 0.913 (0.749–1.000) | 0.917 | 0.958 | 0.920 | 0.909 | 0.939 |

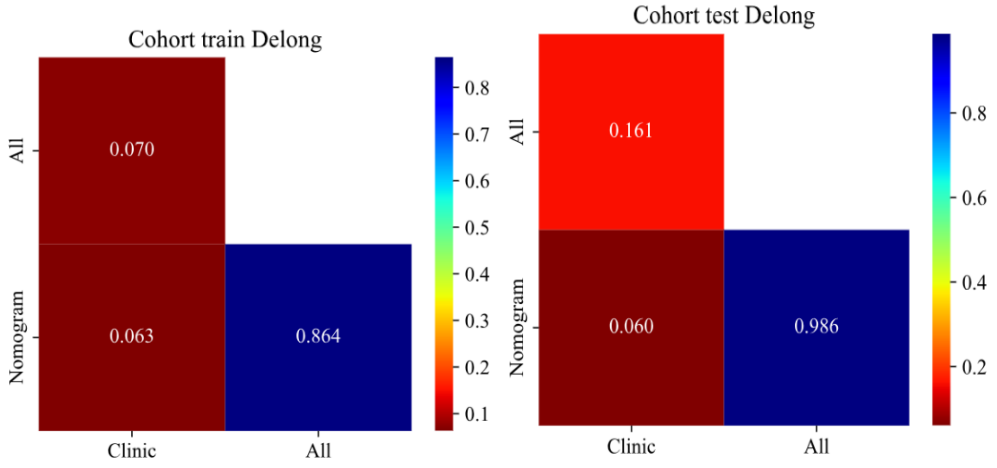



**Figure 14 DeLong test results for the clinical model, PET/CT RS model, and combined nomogram.**

(A) Training cohort; (B) internal validation cohort; Clinic denotes the clinical model; All denotes the PET/CT RS model; Nomogram denotes the combined model.

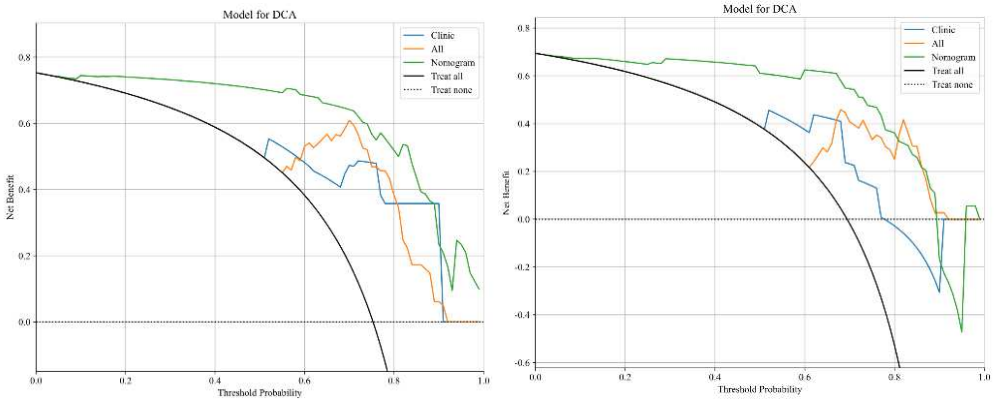



**Figure 15 Decision curves of the clinical model, PET/CT RS model, and combined nomogram.**

(A) Training cohort; (B) internal validation cohort.

# 4 Discussion

This study focused on the preoperative noninvasive identification of high Ki-67 expression in CRC. The main findings were that TNM stage, SUVmax, and SUVmean were independently associated with Ki-67 expression, whereas age, sex, BMI, tumor location, MTV, and TLG did not retain independent associations after multivariable adjustment. From a biological perspective, elevated Ki-67 indicates an increased proportion of proliferating cells, making higher expression in more advanced stages plausible; our results are consistent with the biological behavior of tumor proliferation [13]. Evidence regarding the correlation between PET metabolic parameters and Ki-67 expression has not been consistent. Song et al. found in 358 patients with CRC that the metabolic parameters SUVmax, SUVmean, MTV, and TLG of the primary lesion could not predict Ki-67 expression, with no statistically significant differences between groups [14]. In contrast, Riedl et al. and Güreşci et al. reported significant correlations between SUVmax and high Ki-67 expression [15-16]. In the present study, SUVmax and SUVmean remained statistically significant after multivariate analysis, suggesting that glucose metabolic activity may reflect tumor proliferation at different levels. The inconsistency among studies may be related to differences in case composition, lesion segmentation, measurement of metabolic parameters, and Ki-67 cutoff values; conventional PET parameters are therefore better suited as complementary information for radiomics models rather than as standalone surrogates for pathological detection.

Regarding the radiomics models, we compared six classifiers within the same cohort, and the LightGBM-based PET/CT RS model achieved an AUC of 0.911 in the internal validation set. LightGBM can handle high-dimensional features and capture nonlinear relationships, and its algorithmic basis has been systematically described [17]. After further combining the PET/CT RS with TNM stage, SUVmax, and SUVmean, the nomogram achieved a validation AUC of 0.913 and an accuracy of 0.917, suggesting that anatomic morphologic information and glucose metabolic information are complementary in characterizing tumor phenotype. Similar multi-source information fusion strategies have been applied to predicting treatment response in rectal cancer [18], perineural invasion [19], and microsatellite instability status [20]. Notably, these two conclusions refer to different levels of comparison: the superiority of fused PET/CT features over single modality features was statistically supported, whereas the AUC difference between the combined nomogram and the PET/CT RS model did not reach statistical significance, and the limited validation sample size (n = 36) provides insufficient power to detect such a small difference. Therefore, the value of the combined model should be understood not as a proven gain in discriminative performance, but as the maintenance of an equivalent discrimination level with enhanced interpretability and decision transparency through the inclusion of readily available clinical variables; the higher net benefit of the combined model within part of the threshold range in decision curve analysis also supports this interpretation.

At the feature level, the radiomic features most strongly associated with Ki-67 expression in this study were lbp_PET3D_PETm2_PETfirstorder_PETKurtosis from PET images and square_CTfirstorder_CTMaximum and wavelet_CTHLL_CTfirstorder_CTSkewness from CT images. The former is the kurtosis of three-dimensional PET images based on the local binary pattern (LBP). Kurtosis describes the peakedness of the gray-value distribution: higher values indicate more extreme observations, whereas lower values indicate that the data are concentrated around the mean. Our results showed a negative correlation between the kurtosis feature and Ki-67 expression, suggesting

that in tumors with higher proliferative activity, metabolic activity tends to be homogeneous and regional heterogeneity is correspondingly reduced. A possible explanation is that, to meet the demands of rapid division and proliferation, highly proliferative tumors systematically activate multiple metabolic pathways to secure energy supply and material reserves, making the overall metabolic process more coordinated and ultimately manifesting as a more homogeneous signal distribution on PET images [21]. Studies of lung adenocarcinoma likewise found an inverse association between kurtosis features and Ki-67 expression, further supporting the link between metabolic distribution patterns and proliferative activity [22-23]. In contrast, the CT features square_CTfirstorder_CTMaximum and wavelet_CTHLL_CTfirstorder_CTSkewness were both positively correlated with Ki-67 expression. The former is the squared transformation of the maximum first-order CT value; a higher maximum pixel value may correspond to regions of more active proliferation, because increased cell number during proliferation raises tissue density and thus CT signal intensity [24]. The latter is a first-order skewness feature extracted from the HLL subband of the wavelet transform, describing the asymmetry of the gray-value distribution at a specific scale and direction; positive skewness suggests abnormal high-signal regions within the tumor, which may correspond to sites of concentrated cell proliferation. Liu et al. [25] similarly showed in rectal cancer that the primary lesions with high Ki-67 expression had significantly higher pixel intensities on DCE-MRI, consistent with our findings. Moreover, CT studies of lung adenocarcinoma indicated that elevated Ki-67 expression is accompanied by increased cell density and disordered cell arrangement, manifesting as heterogeneous density on images [26]. It can therefore be inferred that highly proliferative tumors present two coexisting CT phenotypes—focal density elevation and heterogeneous distribution—reflecting densification of proliferative zones and spatial heterogeneity of proliferation, respectively, which are not contradictory. It should be emphasized that these interpretations remain mechanistic speculations. Radiomic features do not correspond one-to-one to specific biological processes; feature values are substantially affected by scanning parameters, reconstruction algorithms, and segmentation methods, and the biological meaning of some features lacks consensus [27]. These features are therefore better positioned as reproducibly verifiable imaging phenotypes rather than directly attributable biomarkers, and their clinical translation depends on larger-sample, multicenter, and prospective confirmation.

Several limitations should be acknowledged. First, this was a single-center retrospective study with only 117 patients and an imbalanced expression distribution; the randomly divided validation set constitutes internal validation only and cannot replace independent external validation, so the reported performance may be optimistically biased. Second, the 50% cutoff for Ki-67 has supporting evidence from previous studies, but different pathological evaluation procedures and threshold choices may affect the applicability of the model. Third, ROI delineation relied mainly on manual contouring and review; inter-observer feature robustness has not been systematically quantified, and the nomogram lacks calibration validation in an independent cohort. In addition, SMOTE was applied only to the training set, but oversampling may still introduce optimism related to synthetic samples, which requires evaluation in larger cohorts. Future studies should conduct multicenter prospective validation under unified acquisition and reconstruction protocols, include segmentation consistency and resampling stability analyses, and evaluate discrimination, calibration, and decision benefit in independent external test sets. Interpretable machine learning methods could further clarify the contribution of key features and define the practical boundaries of the model in preoperative risk stratification, biopsy sampling optimization, and treatment response evaluation.

## 5 Conclusion

This study demonstrates that multimodal 18F-FDG PET/CT radiomics can complement conventional clinical information from both anatomic and glucose metabolic dimensions. The LightGBM-based PET/CT RS model showed good capability for identifying high Ki-67 expression in internal validation, and the combined nomogram further improved accuracy and decision net benefit. This approach provides a feasible strategy for preoperative noninvasive risk stratification of Ki-67 expression in CRC, but larger-sample, multicenter, and independent external validation of its stability, calibration, and generalizability is required before clinical implementation.

## Declarations

Ethics approval and consent to participate: This retrospective study was approved by the institutional ethics committee, and the requirement for informed consent was waived.

Consent for publication: Not applicable.

Availability of data and materials: The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Competing interests: The authors declare no competing interests.

Funding: Not applicable.

Authors' contributions: ST: data collection, formal analysis, writing—original draft. YZ: conceptualization, methodology, writing—review and editing, supervision. HW: image interpretation, data curation. All authors read and approved the final manuscript.

Acknowledgements: Not applicable.

# References


1. Teng Y, Zhang X, Xia C, et al. Comparison of cancer incidence, mortality and prevalence between China and the world: an interpretation of GLOBOCAN 2022 data [in Chinese]. Zhonghua Zhong Liu Fang Za Zhi (Chin J Cancer Prev Treat). 2024;31(23):1413-1420.

2. Sveen A, Kopetz S, Lothe RA. Biomarker-guided therapy for colorectal cancer: strength in complexity. Nat Rev Clin Oncol. 2020;17(1):11-32.

3. Menon SS, Guruvayoorappan C, Sakthivel KM, et al. Ki-67 protein as a tumour proliferation marker. Clin Chim Acta. 2019;491:39-45.

4. Liang J, Zhong L, Guo H. Meta-analysis of the effect of Ki-67 expression on the prognosis and clinicopathological features of colorectal cancer [in Chinese]. J Mudanjiang Med Univ. 2021;42(1):14-21.

5. Luo ZW, Zhu MG, Zhang ZQ, et al. Increased expression of Ki-67 is a poor prognostic marker for colorectal cancer patients: a meta-analysis. BMC Cancer. 2019;19(1):123.

6. Dagogo-Jack I, Shaw AT. Tumour heterogeneity and resistance to cancer therapies. Nat Rev Clin Oncol. 2018;15(2):81-94.

7. Gillies RJ, Kinahan PE, Hricak H. Radiomics: images are more than pictures, they are data. Radiology. 2016;278(2):563-577.

8. Lambin P, Rios-Velazquez E, Leijenaar R, et al. Radiomics: extracting more information from medical images using advanced feature analysis. Eur J Cancer. 2012;48(4):441-446.

9. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749-762.

10. Kang J, Lee JH, Lee HS, et al. Radiomics features of 18F-fluorodeoxyglucose positron-emission tomography as a novel prognostic signature in colorectal cancer. Cancers. 2021;13(3):392.

11. Zhou S, Sun D, Mao W, et al. Deep radiomics-based fusion model for prediction of bevacizumab treatment response and outcome in patients with colorectal cancer liver metastases: a multicentre cohort study. EClinicalMedicine. 2023;65:102271.

12. Yao X, Ao W, Zhu X, et al. A novel radiomics based on multi-parametric magnetic resonance imaging for predicting Ki-67 expression in rectal cancer: a multicenter study. BMC Med Imaging. 2023;23(1):168.

13. Tong G, Zhang G, Liu J, et al. Cutoff of 25% for Ki67 expression is a good classification tool for prognosis in colorectal cancer in the AJCC-8 stratification. Oncol Rep. 2020;43(4):1187-1198.

14. Song J, Li Z, Wei M, et al. Relationship between 18F-FDG PET/CT metabolic parameters and clinicopathological features of colorectal cancer [in Chinese]. Zhong Liu Fang Zhi Yan Jiu (Cancer Res Prev Treat). 2019;46(11):1006-1012.

15. Riedl CC, Akhurst T, Larson S, et al. 18F-FDG PET scanning correlates with tissue markers of poor prognosis and predicts mortality for patients after liver resection for colorectal metastases. J Nucl Med. 2007;48(5):771-775.

16. Güreşci S, Özmen Ö, Uzman M, et al. Correlation of Ki-67 proliferation index and 18F-fluorodeoxyglucose uptake in colorectal incidental lesions detected by positron emission tomography-computed tomography. Turk J Med Sci. 2016;46(4):1182-1187.

17. Ke G, Meng Q, Finley T, et al. LightGBM: a highly efficient gradient boosting decision tree. Adv Neural Inf Process Syst. 2017;30:3146-3154.

18. Jiang H, Guo W, Yu Z, et al. A comprehensive prediction model based on MRI radiomics and clinical factors to predict tumor response after neoadjuvant chemoradiotherapy in rectal cancer. Acad Radiol. 2023;30(Suppl 1):S185-S198.

19. Liu Y, Sun BJT, Zhang C, et al. Preoperative prediction of perineural invasion of rectal cancer based on a magnetic resonance imaging radiomics model: a dual-center study. World J Gastroenterol. 2024;30(16):2233-2248.

20. Ying M, Pan J, Lu G, et al. Development and validation of a radiomics-based nomogram for the preoperative prediction of microsatellite instability in colorectal cancer. BMC Cancer. 2022;22(1):524.

21. Li Z, Sun C, Qin Z. Metabolic reprogramming of cancer-associated fibroblasts and its effect on cancer cell reprogramming. Theranostics. 2021;11(17):8322-8336.

22. Luo J, Li J, Chen J, et al. Multicenter 18F-FDG PET/CT radiomics and clinical features for predicting Ki-67 expression in lung adenocarcinoma [in Chinese]. J Clin Radiol. 2025;44(3):450-455.

23. Tang M. Predictive value of 18F-FDG PET/CT radiomics for EGFR mutation in lung adenocarcinoma [master's thesis, in Chinese]. Zhejiang Chinese Medical University; 2024.

24. Chen S, Chen W, Wang L, et al. Multiple quantitative dual-energy CT parameters can assess proliferative activity of tumor cells in high-grade serous ovarian carcinoma [in Chinese]. J Mol Imaging. 2023;46(4):661-668.

25. Liu Y, Chen A, Wu J, et al. Value of DCE-MRI texture analysis in predicting Ki-67 expression status of rectal cancer [in Chinese]. J China Clin Med Imaging. 2022;33(4):273-278, 290.

26. Yan J, Wang H, He H, et al. Correlation between Ki-67 expression and MSCT findings in different pathological types of lung adenocarcinoma [in Chinese]. Zhejiang Med J. 2020;42(9):898-901.

27. Ana BM, Javaid U, Gilmer V, et al. Artificial intelligence and machine learning for medical imaging: a technology review. Phys Med. 2021;83:242-256.