

ORIGINAL ARTICLE

A radiomics-based artificial intelligence model to assess the risk of relapse in localized colon cancer

C. Prieto-de-la-Lastra^{1,2}, J. A. Carbonell-Asins³, A. Bueno^{4,5}, A. Gómez-Alderete⁶, M. Busto⁷, A. B. Alcolado-Jaramillo⁸, A. Jimenez-Pastor⁵, X. Monzonís⁹, A. Cuñat⁸, C. Montagut^{6,10}, P. Moreno-Ruiz⁵, M. Huerta^{4,10}, D. Roda^{4,10}, F. Gimeno-Valiente⁴, A. Estepa-Fernández⁵, F. Bellvís-Bataller⁵, A. Fuster-Matanzo⁵, J. Gibert⁹, S. Roselló^{4,10}, C. Martinez-Ciarpaglini^{10,11}, J. Vidal^{6,10}, Á. Alberich-Bayarri^{5*}, A. Cervantes^{4,10*} & N. Tarazona^{4,10*}

¹Quantitative Imaging Biomarkers in Medicine, Quibim S.L., Madrid; ²Medical Image Analysis and Biometry Laboratory, Universidad Rey Juan Carlos, Madrid; ³Biostatistics Unit, INCLIVA Biomedical Research Institute, University of Valencia, Valencia; ⁴Department of Medical Oncology, Hospital Clínico Universitario, INCLIVA Biomedical Research Institute, University of Valencia, Valencia; ⁵Quantitative Imaging Biomarkers in Medicine, Quibim S.L., Valencia; ⁶Department of Medical Oncology, Hospital del Mar Research Institute, Universitat Pompeu Fabra, Barcelona; ⁷Department of Radiology, Hospital del Mar Research Institute-IMIM, Universitat Pompeu Fabra, Barcelona; ⁸Department of Radiology, Hospital Clínico Universitario, INCLIVA Biomedical Research Institute, Valencia; ⁹Hospital del Mar Research Institute, Barcelona; ¹⁰CIBERONC, Instituto de Salud Carlos III, Madrid; ¹¹Department of Pathology, Hospital Clínico Universitario, INCLIVA Biomedical Research Institute, University of Valencia, Valencia, Spain



Available online xxx

Background: Accurately estimating relapse risk in localized colon cancer (LCC) remains a challenge, as clinicopathological staging often fails to differentiate patients with a higher likelihood of recurrence. There is a need for novel tools to improve patient selection for post-operative chemotherapy. Radiomics has emerged as a powerful, noninvasive approach that may enhance clinical decision making.

Methods: This retrospective study selected consecutive stage II and III LCC patients operated with curative intent from 2015 to 2017 in two academic institutions. Patients were assigned to either a training cohort made up of 80% of them or a test cohort, to further validate the initial findings. Penalized Cox proportional hazards and gradient boosted algorithms were designed to estimate time to relapse following a five-fold cross-validation process. Three models were assessed: (i) based only on clinical and pathological features, (ii) on radiomic features alone, and (iii) including clinical/pathological and radiomic variables. A new ‘Risk Classification’ score was generated based on the best risk assessment.

Results: A total of 278 patients were included in both cohorts. The Cox model trained with clinical and imaging variables showed the highest prognostic power, with a C-index of 0.68 and a mean cumulative dynamic area under the curve (AUC) of 0.69 on the test set. Feature screening identified 20 variables, including clinical data, radiomics features, and fractal features. SHapley Additive exPlanations (SHAP) analysis highlighted factors related to geometry, vascular invasion, and tumor stage as significant variables related to relapse. The new ‘Risk Classification’ score was able to identify patients with high risk of relapse both in univariable [hazard ratio (HR) 14.22, 95% confidence interval (CI) 1.91-106.08, $P = 0.010$] and multivariable (HR 11.74, 95% CI, 1.54-89.34, $P = 0.017$) models.

Conclusions: Risk analysis revealed the new ‘Risk Classification’ variable as the one with the highest prognostic power compared with the ones currently used. Our findings suggest the potential for improved time-to-relapse estimation, enabling better patient stratification.

Key words: localized colon cancer, imaging biomarkers, machine learning, prognostic biomarkers, radiomics, artificial intelligence

*Correspondence to: Prof. Andrés Cervantes and Dr Noelia Tarazona, Department of Medical Oncology, Hospital Clínico Universitario, INCLIVA Biomedical Research Institute, Av. Blasco Ibañez 17, 46010 Valencia, Spain. Tel: +34-961973543

E-mails: andres.cervantes@uv.es (A. Cervantes). or noetalla@incliva.es (N. Tarazona).

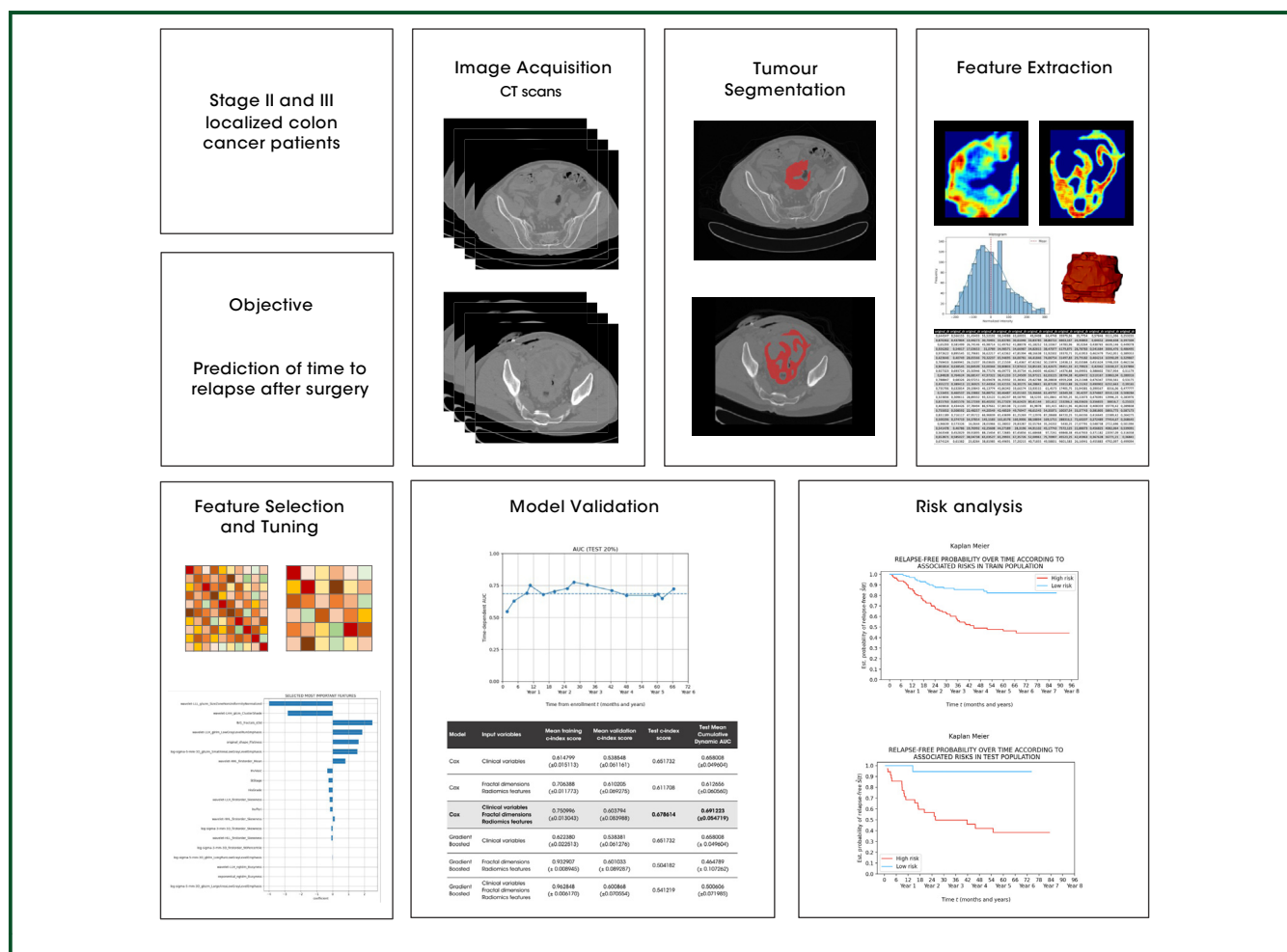
*Co-corresponding authors.

2059-7029/© 2025 The Author(s). Published by Elsevier Ltd on behalf of European Society for Medical Oncology. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

INTRODUCTION

Colorectal cancer is one of the most prevalent cancers, ranking among the top five leading causes of cancer-related death worldwide.^{1,2} Despite advances in treatment and early detection, up to 30%-50% of patients with localized colon cancer (LCC) will eventually relapse.³ The unpredictability of relapse shows an urgent need for innovative prognostic tools capable of swiftly and accurately identifying patients up front with a high risk of relapse.

GRAPHICAL ABSTRACT



Current clinical management relies mainly on the TNM (tumor–node–metastasis) staging system to guide treatment decisions and prognosis for LCC.^{4–6} While stage III LCC patients are globally recognized to benefit from chemotherapy, such an advantage is still contentious in stage II.^{3,7} Furthermore, overall survival rate varies significantly for stage III patients.^{8,9} Risk estimation in LCC is challenging due to the heterogeneous nature of this condition, making a treatment decision a complex task.^{3,6,7}

The value of radiomics in the nonmetastatic colon cancer setting has been already suggested by other authors, with classification of patients between low and high risk of development of colon carcinoma,¹⁰ as well as estimation of the different patient outcomes including survival,^{11–13} response to therapy,¹⁴ disease relapse,¹⁵ lymph node metastasis,^{16,17} microsatellite instability,¹⁸ and recurrence risk.¹⁹ These studies support the idea that radiomics could be a helpful tool in improving risk assessment in colon cancer patients.

In this context, radiomics offers a promising solution for addressing the heterogeneous nature of nonmetastatic LCC and the limitations of current prognostic tools. Radiomics facilitates the extraction of quantitative features from medical images, aiming to transform qualitative imaging data into

actionable insights. This noninvasive approach, together with artificial intelligence (AI) techniques, enables the identification of patterns that relate radiomics features with clinical endpoints. Recent studies have demonstrated the power of predictive models including radiomics features to better estimate survival, treatment response, and recurrence risk of colon cancer in a noninvasive manner.^{13–15,19,20}

In this study, we aim to develop and validate AI-based models for early estimation of time to relapse and differentiation between low- and high-risk stage II and stage III LCC patients by combining clinical real-world data with an imaging biomarker panel of radiomics features extracted from computed tomography (CT) scans acquired before surgery. By developing and validating this AI-based algorithm, our objective is to provide clinicians with robust tools for guiding their treatment decisions, potentially improving patient outcomes.

MATERIALS AND METHODS

Study design

A retrospective real-world data multicenter study was conducted. The subject cohort included stage II and III

consecutive localized colon cancer patients with a CT examination before surgery, acquired from 2015 to 2017. Patients may or may not have received adjuvant chemotherapy after clinical assessment, consistent with standard clinical practice. The inclusion criteria were as follows: (i) patients >18 years old with a stage II or stage III colon cancer diagnosis and histological confirmation, who underwent surgery with curative intent, (ii) availability of a detailed pathology report indicating surgical staging, (iii) availability of abdomen–pelvic CT scan acquired within 1 month before surgery, and (iv) availability of clinical data on follow-up. Patients from two different centers (INCLIVA Biomedical Research Institute and IMIM Hospital del Mar Research Institute) who met the inclusion criteria were included. The dataset was obtained retrospectively as part of an observational study that received approval from the ethics committee at each institution. Consequently, informed consent collection was waived.

Clinical data were collected for each subject, including tumor location, stage, histological grade, vascular invasion, perineural invasion, colonic obstruction and perforation and the number of retrieved lymph nodes at the surgical specimen.

For the development of the algorithm, the whole database was divided into 80% for training and validation, and 20% for external test, as shown in Figure 1A. The partition consisted of a stratified split based on clinical and pathological categorical variables and on the occurrence of relapse, to keep the variables balanced among groups.

Statistical analysis

Distributions of clinical variables between the training and test sets were compared using the Mann–Whitney–Wilcoxon test for the numerical variables and the chi-squared statistic for the categorical variables. Moreover, the time to the follow-up date—in the case of nonrelapsing or censored patients—and the time to relapse—in case of relapsing patients—after tumor resection surgery was recorded, and a Kaplan–Meier analysis was carried out. To reduce bias, the log-rank test was used to check statistical

similarity between training and test groups according to relapse.

Data acquisition and annotations

Scanners from four manufacturers were used for image acquisitions: 53 patients were examined in 4 different scanner models from GE Medical Systems, 3 exams were acquired in 2 types of Philips models, 104 exams were carried out in 4 different Siemens models, and 118 exams were made in 4 different scanner models from Canon (Supplementary Table S1, available at <https://doi.org/10.1016/j.esmoop.2025.105495>).

Each patient CT exam in the training and the testing group was manually segmented slice by slice and supervised by six dedicated radiologists with >12 years of experience. All the annotations were carried out on the transverse plane of the original acquisitions using ITK-SNAP v3.8.0.

Data preprocessing

All the images were resampled to have isotropic voxels of $1 \times 1 \times 1 \text{ mm}^3$ through bicubic spline interpolation. Moreover, a z-score normalization was applied to the volumes before the radiomics analysis. The corresponding masks of the tumors were also converted to isovoxels of $1 \times 1 \times 1 \text{ mm}^3$ by applying nearest neighbor interpolation.

Feature extraction

3D radiomics analysis was carried out extracting low- and high-order statistics using QP-Insights® platform (Quibim S.L., Valencia, Spain). In total, 1379 radiomic features were extracted from each region of interest, including shape and volume ($n = 14$), first-order ($n = 18$), second-order ($n = 73$), and higher order features (14 filters). In the case of some first-order features that involve square operations, after clipping CT intensities to -1024 HU , a voxel shift of $+1024$ was added to grey level intensities to prevent negative values from being squared. Second-order variables included features derived from the Gray Level Co-occurrence Matrix (GLCM), Gray Level Size Zone Matrix

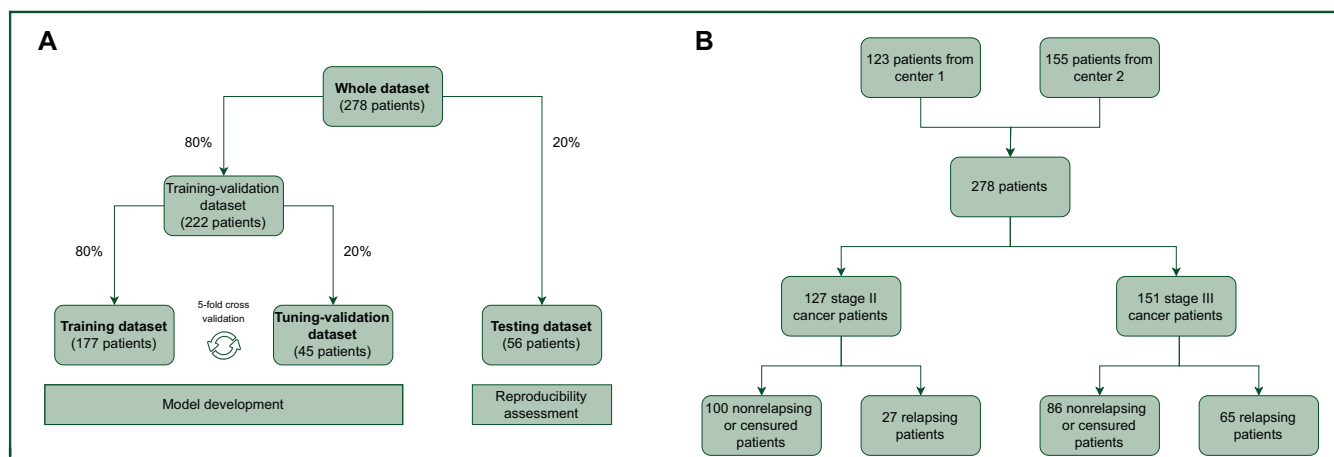


Figure 1. (A) Patient cohort; (B) datasets for training, validation, and testing of the models.

(GLSZM), Gray Level Run Length Matrix (GLRLM), Neighboring Gray Tone Difference Matrix (NGTDM) and Gray Level Dependence Matrix (GLDM). To calculate higher order features, statistical filters were applied to the original image: square, exponential, logarithm, Haar wavelet with eight decompositions—all possible combinations of applying either a high- or a low-pass filter in each of the three dimensions—and Laplacian of Gaussian (LoG) filter with sigma values equal to 0.5, 3, and 5.

In addition, the 2D and 3D fractal dimensions were calculated from the CT exams. Fractal geometries are useful to understand patterns of irregularity changes in the tissue. An in-house Python-based algorithm was developed for that purpose. Categorical clinical and pathological features were converted to one-hot encoding to be processed. Once all numeric variables were extracted from the imaging exams, a z-score standardization was applied to avoid biases in variable ranges and distributions. Thus, the whole database was composed of 278 subjects with 1388 variables (1379 radiomics, 2 fractal, and 7 clinical and pathological variables) for each patient.

Model training and tuning

Two different methodologies were explored to find the best strategy for time to relapse and relapse risk prognosis. Thus, a penalized Cox proportional hazards model and a gradient boosted method were implemented to estimate the time to relapse after surgery. In addition, a five-fold cross-validation with hyperparameter tuning strategy based on grid search was implemented in both types of model.

Three models were developed to train and evaluate each assessment method: the first model exclusively received clinical and pathological variables as input, the second model was trained with imaging-based (i.e. radiomics and fractal) features only, and the third one included the combination of imaging and clinical variables as input to the model.

The whole database was divided into 80% for training (222 patients) and validation, and 20% for external testing (56 patients), as shown in Figure 1. Each fold of the cross-validation consisted of several steps. Firstly, highly correlated features were removed by means of the Pearson correlation coefficient. Then, the Minimum Redundancy—Maximum Relevance (mRMR) feature selection method^{21,22} was applied to select the most representative features. Finally, two methods were evaluated for time-to-relapse estimations: Cox proportional hazards²³⁻²⁵ and gradient boosted model.^{26,27} In each fold of the cross-validation, 80% of the training set was used for training while the remaining 20% was used as a validation or tuning set (Figure 1). At each step, different hyperparameters were evaluated to choose those offering the best performance, therefore maximizing the C-index on the validation dataset. [Supplementary Table S2](https://doi.org/10.1016/j.esmoop.2025.105495), available at <https://doi.org/10.1016/j.esmoop.2025.105495>, represents a summary of the hyperparameters that were evaluated and adjusted during the cross-validation to obtain a final model to further test on an external dataset.

Testing and model evaluation

Images from 56 different patients were used for external validation of both models (Figure 1). Preprocessing and feature extraction were carried out in the same way as for the training data. After the training and hyperparameter tuning, the best Cox proportional hazards model and gradient boosted models trained on each subset of features (i.e. clinical/pathological data only, imaging data only, or both sets of data together) were evaluated independently on the testing data. To evaluate the performance of the models, the C-index score and the cumulative dynamic area under the curve (AUC)²⁸ were measured and inspected. C-index is a measurement of rank correlation between predicted risk scores and observed time points in which a value of 1.0 would indicate perfect prediction, while a random ordering of risk scores would have a C-index of 0.5. Furthermore, SHAP values of the best model were calculated for interpretability and explicability of the algorithm purposes.²⁹ The model with best performing metrics on the left-out data from the cross-validation was selected for time-to-relapse prediction on the test set.

Development and implementation of a novel risk score

A new risk score was elaborated. This way, the relapse risk for each patient was predicted by the selected model, and a risk cut-off was calculated to divide the population into high and low risk of relapse. For this purpose, the optimal threshold for maximum distance division was established as the median risk. This threshold was calculated using the training set after risk score prediction by the best selected model. The risk threshold was then applied on the test set, and the log-rank test was implemented to prove significant differences between classified risk groups. Therefore, a new dichotomous variable—‘Risk Classification’—that divided patients into low- and high-risk subjects was generated. In addition, the prognostic power and the impact of this new variable on the clinical routine were also evaluated through two different methodologies: Cox regression with elastic net penalty and Cox regression with forward—backward stepwise variable selection using Akaike’s information criterion (AIC).³⁰ Clinical variables together with the new Risk Classification variable served as input to these models, which were developed using the test set. All the developments were done using Python v3.9.

Ethics approval and consent to participate

This work was authorized by the local ethics committees of the Valencia Clinic University Hospital (Spain) (2023/088) and of Hospital del Mar Research Institute (IMIM) and conducted following the Declaration of Helsinki principle. Informed consent collection was waived as the dataset was obtained retrospectively as part of an observational study.

RESULTS

Patient characteristics

A total of 278 stage II and stage III LCC patients were eligible for inclusion as they had a CT scan acquired before surgery that passed the quality check and clinical data available, and they met the other criteria detailed in the Methods section. Of them, 108 (38.8%) were females and the median age was 78 years (range 41–103 years).

Of the 278 patients included in the study, 127 patients (45.7%) presented with stage II disease, and 151 patients (54.3%) had stage III LCC. At median follow-up of 42.2 months, 92 patients (33.1%) from the whole cohort experienced a relapse. Specifically, relapses occurred in 27 stage II patients (29.3%), while the remaining 65 patients (70.7%) presented with stage III disease (Figure 1B).

A summary of patient clinical and pathological characteristics in both the training and test cohorts is shown in Table 1. No significative differences were found between the training and test cohorts. In addition, no statistical differences were found in time to relapse between both groups (log rank test, $P = 0.22$). The Kaplan–Meier curve of relapse-free survival for the whole population is depicted in Supplementary Figure S1, available at <https://doi.org/10.1016/j.esmoop.2025.105495>.

Prognostic models and model selection

A Cox proportional hazards model and a gradient boosted model were implemented. Moreover, three different models were evaluated considering clinical–pathological variables only, radiomic features alone, or the combination of both. Results showed a better performance of the penalized Cox proportional hazards models over the gradient boosted model. The Cox model, including clinical/pathological variables, as well as fractal and radiomic features, performed

better than the other two with a C-index score of 0.678 and a mean cumulative dynamic AUC of 0.691 (0.05) in the test set. Furthermore, the value of the area under the cumulative dynamic curve was maintained at ~ 0.7 over time, as depicted in Supplementary Figure S2, available at <https://doi.org/10.1016/j.esmoop.2025.105495>. The C-index and the mean dynamic cumulative AUC of each model are detailed in Supplementary Table S3, available at <https://doi.org/10.1016/j.esmoop.2025.105495>. Therefore, this Cox model was selected for risk and time-to-relapse prognosis. The threshold employed for variable reduction through the Pearson correlation method was 0.9. The regularization term α that was chosen for Cox training in the hyperparameter tuning was 0.0001. Thus, the 20 most significant variables were selected using the Pearson reductor and mRMR methods (Supplementary Table S4, available at <https://doi.org/10.1016/j.esmoop.2025.105495>). Four clinical variables, 15 radiomics features, and 1 fractal feature were selected. In addition, a SHAP analysis was executed to provide interpretability of the model. Figure 2 represents the SHAP values of the 20 parameters ordered by their impact on the model for the test set, being the 3D fractal dimension the most powerful feature followed by some higher order variables. In the plot, larger SHAP values (to the right) are associated with worse prognosis, and variable colors are correlated to feature values. Thus, the presence of vascular or perineural invasion, a higher stage (III versus II), or a high histological grade (red colors) are associated with worse prognosis (to the right), as was expected. In a similar way, the higher the value of the 3D fractal feature, the higher the resulting SHAP value and, therefore, the associated risk of relapse also increases.

Novel risk score

To be able to estimate the probability of relapse of new patients, a risk score was developed. The threshold to classify patients into high or low risk was defined by the median risk of relapse of the training set. The risk score resulted in a risk cut-off of -0.49 . Patients with associated risk above the threshold were classified as high-risk patients while patients with estimated risk under the cut-off were categorized as low-risk patients.

Differences in time to relapse were observed between patients with high risk versus patients with low risk (log-rank test, $P < 0.001$), as shown in Figure 3. In brackets, the total considered patients per year is indicated, excluding the patients that relapse in previous years and those with lost follow-up. The distribution of variables according to risk group was compared for the training and test sets (Supplementary Table S5, available at <https://doi.org/10.1016/j.esmoop.2025.105495>). Therefore, the risk analysis provided a new dichotomous variable—named as ‘Risk Classification’—that classified patients as low or high risk. The added value that this new variable may provide to clinical practice was then evaluated by comparing its prognostic power to the power of the clinical variables that are already considered in clinical practice.

Table 1. Patient characteristics in training and test sets			
Variable	Training set (N = 222), n (%)	Test set (N = 56), n (%)	P value
Tumor location			1.00
Right	115 (51.80)	29 (51.79)	
Left	107 (48.20)	27 (48.21)	
Stage			0.98
II	102 (45.95)	25 (44.64)	
III	120 (54.05)	31 (55.36)	
Histological grade			1.00
Low	191 (86.04)	48 (85.71)	
High	31 (13.96)	8 (14.29)	
Vascular invasion			0.52
No	124 (55.86)	28 (50.00)	
Yes	98 (44.14)	28 (50.00)	
Perineural invasion			0.09
No	155 (69.82)	46 (82.14)	
Yes	67 (30.18)	10 (17.86)	
Colonic obstruction and perforation			0.86
No	210 (94.59)	52 (92.86)	
Yes	12 (5.41)	4 (7.14)	
Number of retrieved lymph nodes	22.49 (0.37)	25.23 (12.97)	0.18

The table presents categories and distributions of clinical features, and results from the Mann–Whitney–Wilcoxon and chi-square tests.

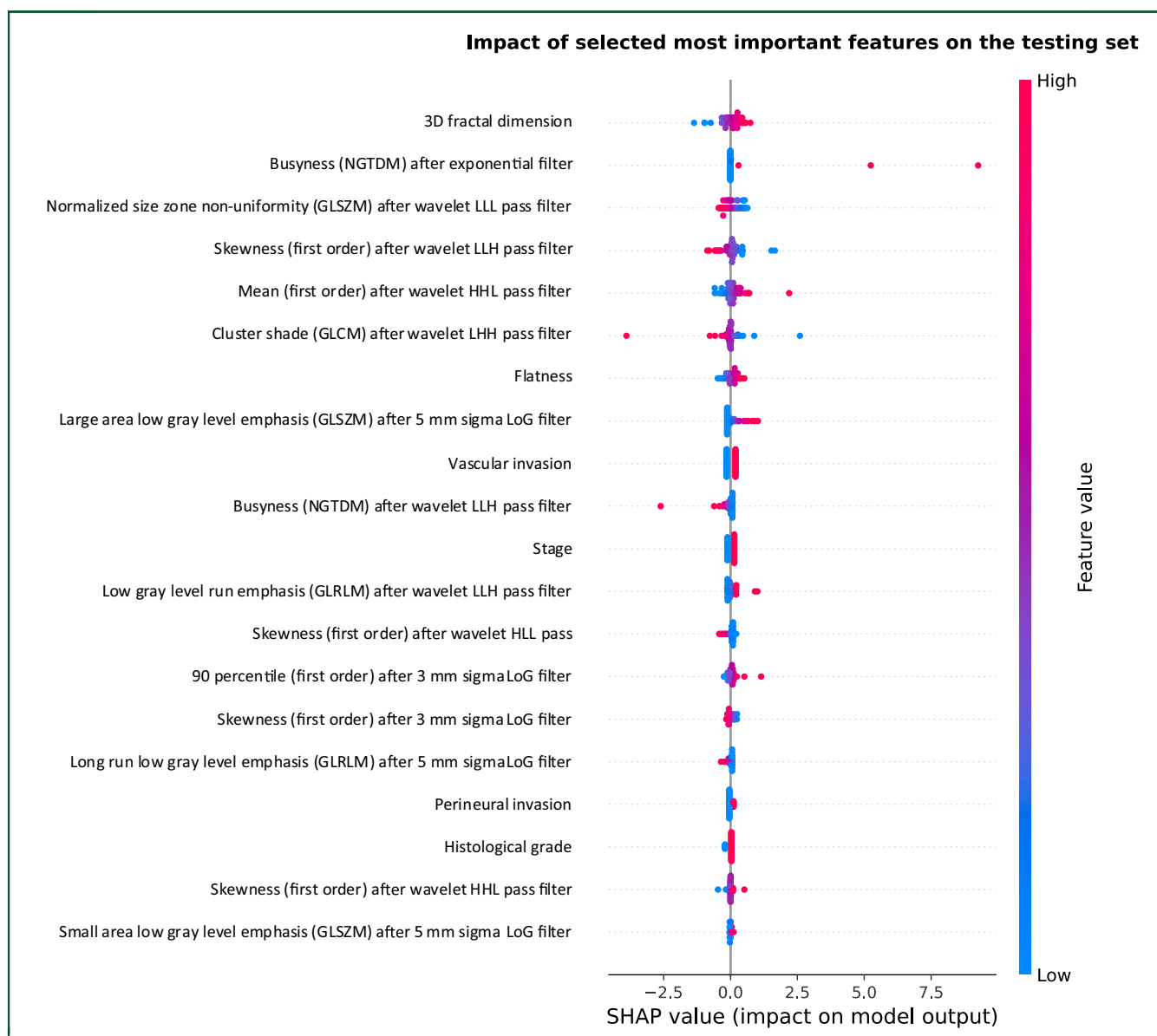


Figure 2. SHAP (SHapley Additive exPlanations) values from model testing set.

GLCM, Gray Level Co-occurrence Matrix; GLSZM, Gray Level Size Zone Matrix; HHL, high-high-low pass filter; HLL, high-low-low pass filter; LHH, low-high-high pass filter; LLH, low-low-high pass filter; LLL, low-low-low pass filter; LoG, Laplacian of Gaussian; NGTDM, Neighboring Gray Tone Difference Matrix.

On the one hand, Cox analysis was carried out by varying the elastic net penalty α . The impact of each variable was inspected depending on the penalty (Supplementary Figure S3, available at <https://doi.org/10.1016/j.esmoop.2025.105495>). Results proved that the variable with the largest Cox coefficient (absolute value) across different penalties was the risk score ('Risk Classification'), highlighting that this was the feature with the highest impact on the prediction of relapse. On the other hand, a Cox regression model with forward–backward stepwise feature selection was applied to select best model according to AIC. Both univariable and multivariable analyses were carried out. Table 2 shows the hazard ratio, confidence intervals, and *P* values for each variable, representing their impact on the relapse prediction. The analyses outputted the risk score ('Risk Classification') variable as the one with highest

impact in both cases, overcoming the power of every clinical variable.

DISCUSSION

This study addresses a critical need in the management of LCC, particularly for stage II and III patients, who face a risk of relapse despite the use of post-operative chemotherapy. With colorectal cancer ranking among the top causes of cancer-related mortality worldwide^{1,2} and current prognostic tools showing limited effectiveness, exploring novel approaches such as radiomics has become increasingly relevant.^{3,4,6} Addressing this, we employed radiomics, a noninvasive imaging-based technique, to extract quantitative features from presurgery CT scans and develop novel assessment models for better estimation of time to relapse in LCC patients.

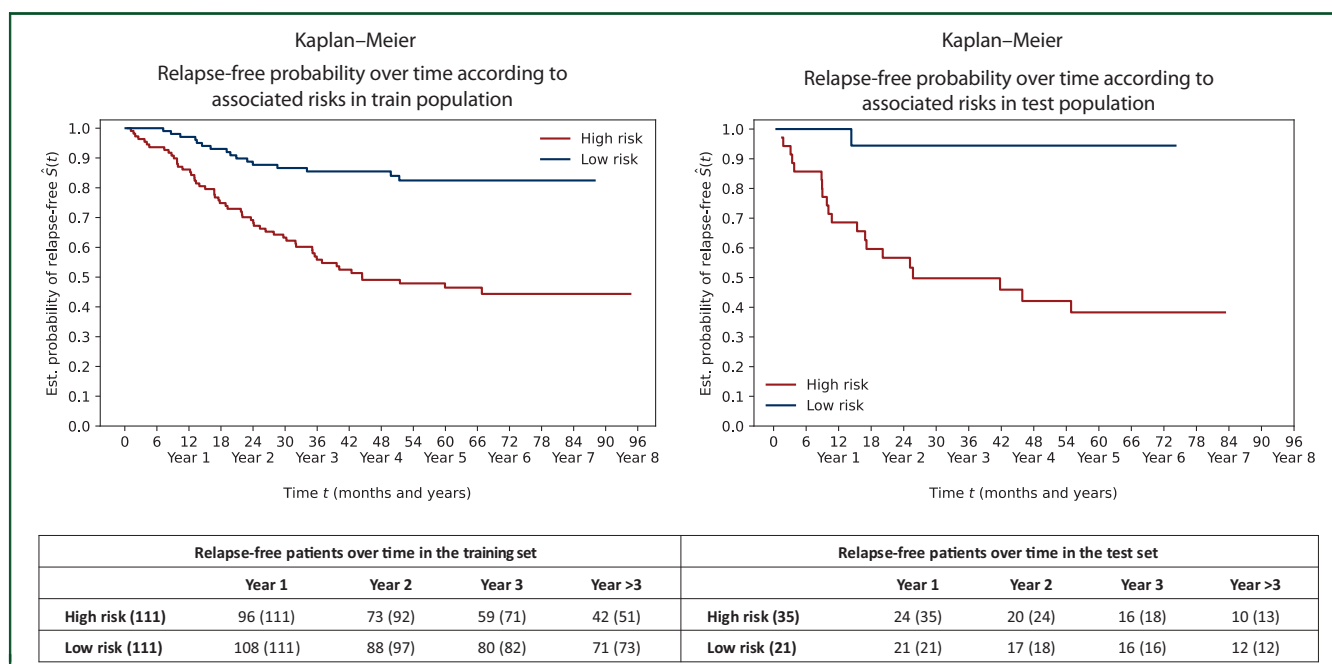


Figure 3. Kaplan–Meier analysis of training and test set divided by risk, and number of patients that are relapse free over the years (in brackets, the total of considered patients per year).

In line with previous studies, our findings demonstrate that models integrating both radiomics features and clinical data outperform those relying only on conventional clinical–pathological variables,^{31,32} offering superior prognostic accuracy. The Cox proportional hazards model trained on imaging and clinicopathological features showed better overall performance than other evaluated models and methodologies. This may be attributable to the sample size, as Cox models are generally more robust with smaller

cohorts, while gradient boosted models tend to perform better with larger datasets. Consequently, gradient boosted models exhibited overfitting during training, which was not matched by improved performance on the validation set. Moreover, Cox models allow interpretation of results through hazard ratios, enhancing their clinical applicability. This was evidenced by a C-index of 0.68 on the external test set compared with the results from the clinical–pathological variables alone model (0.65) and the

Table 2. Variables analysis in relapse-free survival Cox regression

Variable		HR (univariable)	HR (multivariable)
Location	<i>n</i> (%)		
Left	27 (48.2)	—	—
Right	29 (51.8)	0.85 (0.36-1.99, <i>P</i> = 0.702)	0.49 (0.20-1.18, <i>P</i> = 0.113)
Stage			
II	25 (44.6)	—	—
III	31 (55.4)	3.85 (1.29-11.44, <i>P</i> = 0.015)	4.40 (1.44-13.38, <i>P</i> = 0.009)
Number of retrieved lymph nodes			
Mean (SD)	25.2 (13.1)	0.97 (0.93-1.01, <i>P</i> = 0.209)	0.96 (0.92-1.01, <i>P</i> = 0.085)
Histological grade			
High	8 (14.3)	—	—
Low	48 (85.7)	0.96 (0.28-3.26, <i>P</i> = 0.949)	—
Vascular invasion			
No	28 (50.0)	—	—
Yes	28 (50.0)	1.50 (0.62-3.63, <i>P</i> = 0.365)	—
Perineural invasion			
No	46 (82.1)	—	—
Yes	10 (17.9)	2.72 (1.09-6.77, <i>P</i> = 0.032)	2.68 (1.01-7.13, <i>P</i> = 0.049)
Colonic obstruction and perforation			
No	52 (92.9)	—	—
Yes	4 (7.1)	1.35 (0.31-5.83, <i>P</i> = 0.686)	—
Risk classification			
0	21 (37.5)	—	—
1	35 (62.5)	14.22 (1.91-106.08, <i>P</i> = 0.010)	11.74 (1.54-89.34, <i>P</i> = 0.017)

In bold, hazard ratios (HR) from the Risk Classification variable, representing the new risk score as the feature with highest impact on time-to-relapse prediction. Multivariable model was selected according to Akaike's information criterion using all variables included in univariable analyses.

radiomic-only features model (0.61). Furthermore, the mean cumulative dynamic AUC showed a value of ~ 0.7 over time, which was higher than other competing models. Importantly, we also developed a novel dichotomous risk score that effectively stratifies patients into low- and high-risk groups. This score, generated based on the best risk assessment obtained, emerged as the strongest predictor, with a hazard ratio higher than any other clinical–pathological variable. Besides, it remained independently significant after adjustment for potential confounders, and its prognostic value was consistent across both training and test sets, supporting its potential utility as a tool for patient stratification in clinical practice.

Our results are in line with earlier efforts to improve recurrence risk stratification beyond classical staging systems. Thus, in 2018, Pagès et al.³³ proposed Immunoscore, a new quantitative immune-based classifier based on the evaluation of the density of CD3+ and CD8+ T lymphocytes in the tumor core and invasive margin. Immunoscore reflects the host immune response to the tumor and demonstrated superior prognostic accuracy compared with TNM staging, microsatellite instability (MSI) status, and other clinicopathological markers. Despite its clinical promise, it is worth noting that the Immunoscore requires postsurgical histological analysis and immunohistochemical quantification of T-cell infiltration, limiting its applicability in preoperative decision making. In contrast, our risk score, based on radiomic features, offers a noninvasive alternative capable of effectively stratifying relapse risk before surgery and is potentially easier to implement in clinical practice.

In parallel, serial circulating tumor DNA (ctDNA) analysis has emerged as a surveillance strategy for patients with resected LCC, as its detection has been independently associated with higher risk of recurrence.^{34,35} However, compared with CT imaging—already part of standard patient management—ctDNA entails higher cost, invasiveness, and more limited accessibility. Moreover, the sensitivity and specificity for the determination of ctDNA is far from optimal, requiring better diagnostic tools and further validation in clinical trials. On the other hand, radiomic analysis of these routinely acquired CT scans offers an opportunity to extract prognostically relevant information before surgery. Indeed, several prior studies have demonstrated the potential of radiomics in colon cancer, particularly showing promising prognostic performance for recurrence with C-indices ranging from 0.68 to 0.78 and AUCs reaching up to 0.91 in integrated nomograms during validation.^{11–13,15,19}

Regarding feature selection and its contribution to model performance, our analysis highlighted the 3D fractal dimension, which is related to tumor geometry, as the most powerful variable. Beyond this feature, significant variables overall revealed the impact of the tumor morphology on relapse predictions and included several high-order features related to the texture and homogeneity/heterogeneity of the tissue. These findings are consistent with prior studies where radiomic features such as wavelet-based textures, GLDM-derived measures of local uniformity, and

GLSZM-derived metrics of small area emphasis have been associated with relapse risk, all capturing aspects of spatial disorganization and structural complexity within the tumor.^{13,15,19} Moreover, pathological variables such as tumor stage, histological grade, vascular invasion, or perineural invasion also stood out for their relapse prediction power. These prognostic variables highlight the complex interplay between tumor phenotype and LCC patient outcomes.

Despite the promising results of our study, several limitations should be acknowledged. The limited reliability of AI-based models, mainly due to the risk of overfitting and the challenges associated with interpreting radiomics-based models, must be considered. To strengthen model reliability and interpretability, several strategies were implemented. Specifically, to mitigate the risk of overfitting, feature selection and cross-validation techniques were carefully applied. In parallel, to avoid the ‘black box’ effect—referring to the limited interpretability of the internal decision-making processes of AI-based models—SHAP analysis and Cox regression models, with elastic net penalty and with backward stepwise feature selection, were used to assess the contribution of each individual variable. These strategies helped improve both the robustness and transparency of our models, facilitating a better understanding of the mechanisms driving relapse risk. Additionally, while the retrospective nature of the study and the relatively small sample size may still limit the generalizability of our findings, the inclusion of patients from two different centers already represents a step forward compared with most previous studies conducted in a single-center setting. Future research should aim to validate these models in larger and more diverse multicenter cohorts to further enhance their clinical applicability.

In conclusion, our study shows that the integration of an AI-based analysis of CT scans with conventional clinical–pathological data holds some promise for improving relapse prediction in stage II and III LCC patients. The proposed novel risk score, integrating both radiomic and clinical features, demonstrated the strongest relapse estimation capability. These findings may have future implications for further patient stratification and planning, providing clinicians with a new valuable tool to help them in patient management decisions.

FUNDING

This work was supported by grants from the Instituto de Salud Carlos III [grant numbers PI21/00689 and PI24/00103 to NT and AC, grant number PI21/00041 to CM]; the Fundación Científica de la Asociación Española contra el Cáncer [grant number GCAEC20030CERV to CMC, AC, and NT]; the Fundación Mutua Madrileña [grant number AP-16019/2024 to NT]; the Sociedad Española de Oncología Médica [reference Gladiator to NT]; the Gilead Sciences Juan Rodés contracts [contract number JR20/00005 to NT]; and the Consellería de Cultura, Educación y Ciencia de la Generalitat Valenciana [grant number APOSTD/2021/168 to FG-V; grant number APOTIP/2021/007 to AB].

DISCLOSURE

AC reports receiving institutional research funding from AbbVie, Bayer, BeiGene, Bristol Myers Squibb, Genentech, HiFiBio Therapeutics, invoX Pharma, Merck Serono, Merck Sharp & Dohme, Natera, Novartis, Roche, and Servier and serving on the advisory board or receiving speaker fees paid to his institution from AbbVie, Agenus, Arcus Biosciences, GSK, and Roche. CM reports receipt of honoraria for participation in advisory board from Biocartis, Merck Serono; receipt of honoraria as invited speaker from Amgen, Guardant Health, Merck Serono, Pierre Fabre, and Roche; and receipt of royalties for licensing fees administered by Institut Investigació Hospital del Mar. AA-B is the chief executive officer (CEO) and a shareholder of Quibim. AE-F, AF-M, AJ-P, CP, PM-R, and FB-B are current employees of Quibim. NT declares advisory board or speaker fees from Merck, Servier, Pfizer, Grifols, Natera, and Guardant Health and research support from GSK. All remaining authors have declared no conflicts of interest.

DATA SHARING

The data sets used and/or analyzed during the current study are available from the corresponding author upon request.

REFERENCES

- Arnold M, Sierra MS, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global patterns and trends in colorectal cancer incidence and mortality. *Gut*. 2017;66(4):683-691.
- Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. *CA Cancer J Clin*. 2023;73(3):233-254.
- Argilés G, Tabernero J, Labianca R, et al. Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2020;31(10):1291-1305.
- Brierley JD, Gospodarowicz MK, Wittekind C, editors. *TNM Classification of Malignant Tumours*. 8th ed. Hoboken, NJ: Wiley-Blackwell/ Union for International Cancer Control; 2017.
- Roth AD, Delorenzi M, Tejpar S, et al. Integrated analysis of molecular and clinical prognostic factors in stage II/III colon cancer. *J Natl Cancer Inst*. 2012;104(21):1635-1646.
- Weiser MR. AJCC 8th edition: colorectal cancer. *Ann Surg Oncol*. 2018;25(6):1454-1455.
- Baxter NN, Kennedy EB, Bergsland E, et al. Adjuvant therapy for stage II colon cancer: ASCO guideline update. *J Clin Oncol*. 2022;40(8):892-910.
- Webber C, Gospodarowicz M, Sobin LH, et al. Improving the TNM classification: findings from a 10-year continuous literature review. *Int J Cancer*. 2014;135(2):371-378.
- Hari DM, Leung AM, Lee JH, et al. AJCC Cancer Staging Manual 7th edition criteria for colon cancer: do the complex modifications improve prognostic assessment? *J Am Coll Surg*. 2013;217(2):181-190.
- Huang X, Cheng Z, Huang Y, et al. CT-based radiomics signature to discriminate high-grade from low-grade colorectal adenocarcinoma. *Acad Radiol*. 2018;25(10):1285-1297.
- Yao X, Sun C, Xiong F, et al. Radiomic signature-based nomogram to predict disease-free survival in stage II and III colon cancer. *Eur J Radiol*. 2020;131:109205.
- Dai W, Mo S, Han L, et al. Prognostic and predictive value of radiomics signatures in stage I-III colon cancer. *Clin Transl Med*. 2020;10(1):288-293.
- Lv L, Xin B, Hao Y, et al. Radiomic analysis for predicting prognosis of colorectal cancer from preoperative ¹⁸F-FDG PET/CT. *J Transl Med*. 2022;20(1):66.
- Caruso D, Zerunian M, Ciolina M, et al. Haralick's texture features for the prediction of response to therapy in colorectal cancer: a preliminary study. *Radiol Med*. 2018;123(3):161-167.
- Zhu H, Hu M, Ma Y, et al. Multi-center evaluation of machine learning-based radiomic model in predicting disease free survival and adjuvant chemotherapy benefit in stage II colorectal cancer patients. *Cancer Imaging*. 2023;23(1):74.
- Li M, Zhang J, Dan Y, et al. A clinical-radiomics nomogram for the preoperative prediction of lymph node metastasis in colorectal cancer. *J Transl Med*. 2020;18(1):46.
- Eresen A, Li Y, Yang J, et al. Preoperative assessment of lymph node metastasis in Colon Cancer patients using machine learning: a pilot study. *Cancer Imaging*. 2020;20(1):1-9.
- Golia Pernicka JS, Gagniere J, Chakraborty J, et al. Radiomics-based prediction of microsatellite instability in colorectal cancer at initial computed tomography evaluation. *Abdom Radiol (NY)*. 2019;44(11):3755-3763.
- Fan S, Cui X, Liu C, et al. CT-Based radiomics signature: a potential biomarker for predicting postoperative recurrence risk in stage II colorectal cancer. *Front Oncol*. 2021;11:644933.
- Caruso D, Polici M, Zerunian M, et al. Radiomic cancer hallmarks to identify high-risk patients in non-metastatic colon cancer. *Cancers*. 2022;14(14):3438.
- Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. *J Bioinform Comput Biol*. 2005;3(2):185-205.
- Ge G, Zhang J. Feature selection methods and predictive models in CT lung cancer radiomics. *J Appl Clin Med Phys*. 2023;24(1):e13869.
- Harrell FE, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982;247(18):2543-2546.
- Martínez-González MÁ, Alonso Á, Fidalgo JL. ¿Qué es una hazard ratio? Nociones de análisis de supervivencia. *Med Clin (Barc)*. 2008;131(2):65-72.
- George B, Seals S, Aban I. Survival analysis and regression models. *J Nucl Cardiol*. 2014;21(4):686-694.
- Friedman JH. Greedy function approximation: a gradient boosting machine. *Ann Statist*. 2001;29(5):1189-1232.
- Friedman JH. Stochastic gradient boosting. *Comput Stat Data Anal*. 2002;38(4):367-378.
- Lambert J, Chevret S. Summary measure of discrimination in survival models based on cumulative/dynamic time-dependent ROC curves. *Stat Methods Med Res*. 2016;25(5):2088-2102.
- Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Red Hook, NY: Curran Associates Inc; 2017. p. 4768-4777.
- Akaike H. A new look at the statistical model identification. *IEEE Trans Automat Contr*. 1974;19(6):716-723.
- Mo S, Zhou Z, Li Y, et al. Establishment and validation of a novel nomogram incorporating clinicopathological parameters into the TNM staging system to predict prognosis for stage II colorectal cancer. *Cancer Cell Int*. 2020;20(1):285.
- Hoshino N, Hasegawa S, Hida K, et al. Nomogram for predicting recurrence in stage II colorectal cancer. *Acta Oncol*. 2016;55(12):1414-1417.
- Pagès F, Mlecnik B, Marliot F, et al. International validation of the consensus Immunoscore for the classification of colon cancer: a prognostic and accuracy study. *Lancet*. 2018;391(10135):2128-2139.
- Henriksen TV, Tarazona N, Frydendahl A, et al. Circulating tumor DNA in stage III colorectal cancer, beyond minimal residual disease detection, toward assessment of adjuvant therapy efficacy and clinical behavior of recurrences. *Clin Cancer Res*. 2022;28(3):507-517.
- Martínez-Castedo B, Cambor DG, Martín-Arana J, et al. Minimal residual disease in colorectal cancer. Tumor-informed versus tumor-agnostic approaches: unraveling the optimal strategy. *Ann Oncol*. 2025;36(3):263-276.