Asian Journal of Current Research in Clinical Cancer
ISSN: 3062-4444
2021, Volume 1, Issue 2, Page No: 113-130

Available online at: www.galaxypub.co/page/journals

Galaxy Publication

Galaxy Publication

Explainable Multimodal Machine Learning Integrating Radiomics and Pathomics for Prediction of Pathologic Complete Response in Esophageal Squamous Cell Carcinoma Treated With Neoadjuvant Chemoimmunotherapy

Lucia C. Nielsen[1], Rafael Garcia[1], Juan Singh[1], Jonas Santos[1]*, Camila Smith[1]

[1]Department of Translational Oncology, School of Medicine, University of Toronto, Toronto, Canada.

***E-mail** ✉ jsantos@hotmail.com

Received: 17 September 2021; **Revised:** 04 December 2021; **Accepted:** 05 December 2021

ABSTRACT

Determining, before surgery, which patients with esophageal squamous cell carcinoma (ESCC) will achieve a pathological complete response (pCR) after neoadjuvant chemoimmunotherapy (nCIT) remains a major unmet clinical need. Reliable prediction of pCR could enable risk-adapted treatment strategies and avoid unnecessary surgical intervention. The objective of this study was to design and independently validate a transparent multimodal learning system that jointly leverages radiological and histopathological imaging data to estimate pCR. We retrospectively collected data from 335 patients with ESCC treated with nCIT followed by surgical resection at three tertiary centers. One institution contributed cases that were split into model development (n=181) and internal validation (n=115) cohorts, whereas patients from the remaining centers constituted an external validation cohort (n=39). Quantitative features were extracted from contrast-enhanced CT scans and H&E-stained whole-slide images to construct radiomics-only and pathomics-only classifiers. Two strategies were implemented to integrate these modalities: a feature-level intermediate fusion approach and a prediction-level late fusion approach. Model discrimination and classification performance were evaluated using AUC, accuracy, sensitivity, specificity, and F1 score. Survival differences were explored according to both histologically confirmed and model-inferred pCR status. Model transparency was enforced through the use of interpretable feature definitions and explainable decision mechanisms. The intermediate fusion strategy demonstrated consistently superior performance compared with unimodal models and late fusion across all cohorts. In the development, internal validation, and external validation cohorts, the intermediate fusion model achieved AUC values of 0.97, 0.78, and 0.76, respectively, with corresponding accuracy values of 0.93, 0.87, and 0.77. Both true pCR and predicted pCR groups exhibited distinct overall survival trends in exploratory analyses. Importantly, the model relied on explicitly defined radiological and histomorphological attributes, and its predictions were accompanied by case-specific and population-level explanatory visualizations that clarified the underlying decision logic. A clinician-oriented graphical interface was also implemented to support real-world application. This study presents a clinically interpretable radiopathomics-based prediction framework capable of estimating pCR following neoadjuvant chemoimmunotherapy in ESCC using routinely available imaging data. The proposed approach may assist clinicians in tailoring post-treatment management, particularly when weighing active surveillance against immediate surgery.

Keywords: Multimodal machine, Radiomics, Pathomics, Pathologic, Esophageal squamous cell carcinoma

How to Cite This Article: Nielsen LC, Garcia R, Singh J, Santos J, Smith C. Explainable Multimodal Machine Learning Integrating Radiomics and Pathomics for Prediction of Pathologic Complete Response in Esophageal Squamous Cell Carcinoma Treated with Neoadjuvant Chemoimmunotherapy. Asian J Curr Res Clin Cancer. 2021;1(1):113-30. https://doi.org/10.51847/sVlPwgIlat

Introduction

Esophageal squamous cell carcinoma (ESCC) continues to represent a major global health burden due to its high incidence and aggressive biological behavior [1]. For patients with locally advanced disease, neoadjuvant chemoradiotherapy (nCRT) followed by surgical resection remains the established standard of care [2, 3]. In recent years, however, emerging clinical evidence has positioned neoadjuvant chemoimmunotherapy (nCIT) combined

with surgery as a compelling alternative. Clinical trials have reported R0 resection rates between 80.9% and 98.0%, alongside pathologic complete response (pCR) rates ranging from 16.7% to 50.0% [4, 5]. Moreover, a prospective comparison suggested that nCIT may confer improved 2-year overall survival (OS) and disease-free survival (DFS) relative to nCRT, despite comparable pCR rates (22.9% vs 25.9%) [6]. Importantly, attainment of pCR has been consistently associated with favorable long-term outcomes and may support nonoperative "watch-and-wait" strategies, thereby preserving organ function and improving patient quality of life [4, 7, 8]. These considerations underscore the importance of accurately identifying pCR prior to surgery in patients undergoing nCIT.

Although the clinical value of pCR prediction is well recognized, reliable biomarkers capable of forecasting response to nCIT remain limited. Commonly investigated tissue-based markers, such as microsatellite instability [9, 10], programmed cell death ligand-1 (PD-L1) expression [11, 12], and tumor mutational burden (TMB) [13–15], demonstrate suboptimal predictive performance and are often constrained by high costs, technical demands, and limited accessibility. As a result, there is a pressing need for practical, reproducible, and economical approaches to response prediction that can be readily implemented in routine clinical workflows.

Medical imaging constitutes a rich source of both macroscopic and microscopic information and is particularly amenable to artificial intelligence (AI)-driven analysis. Radiologic modalities such as contrast-enhanced computed tomography capture tumor-scale characteristics, while histopathological whole-slide images (WSIs) stained with hematoxylin and eosin provide detailed insights into cellular morphology and tissue architecture. These complementary data streams offer the potential for enhanced predictive performance through multimodal integration [16]. Radiomics and pathomics techniques enable the extraction of high-dimensional quantitative descriptors of tumor phenotype and microenvironment and have demonstrated predictive utility across multiple cancer types, including ESCC [17–19]. Building on prior evidence supporting radiomics-based prediction of pCR after nCIT [20], as well as studies highlighting the prognostic relevance of nuclear morphology and textural patterns [21, 22], the combined use of radiomics and pathomics features represents a logical strategy for improving preoperative response assessment in ESCC.

However, successful clinical translation of multimodal AI models depends not only on predictive accuracy but also on transparency and interpretability. In this work, interpretability was explicitly incorporated as a core design principle, addressed at both the model and feature levels. At the model level, we favored algorithms with traceable decision-making processes and robust post hoc explanation techniques, such as Shapley-value–based attribution, to enable visualization of feature contributions at both the individual and population levels while reducing the opacity associated with deep learning approaches [23]. At the feature level, we prioritized radiomics features with clear mathematical definitions and pathomics descriptors grounded in interpretable histomorphological characteristics—such as nuclear size, shape, perimeter, and chromatin texture—to enhance clinical interpretability and facilitate communication with domain experts.

Accordingly, the present study aimed to construct an interpretable multimodal machine learning framework for preoperative prediction of pCR to nCIT in ESCC using multicenter data. We systematically evaluated multiple machine learning algorithms and multimodal fusion strategies to integrate CT-based radiomics with WSI-derived pathomics features, while formalizing interpretability at both the algorithmic and feature-definition levels. To support clinical adoption and usability, we further designed case-level and cohort-level explanatory outputs, including contribution-based visualizations, and developed a user-facing software prototype to demonstrate potential integration into clinical workflows.

Materials and Methods

Owing to the retrospective nature of the study, the requirement for informed consent was waived. All procedures were conducted in accordance with the Declaration of Helsinki and complied with established methodological recommendations for radiomics research [24]. To enhance methodological transparency and rigor, study conduct was evaluated using a previously proposed 12-item methodological quality checklist [25]. An overview of the study design and analytical workflow is illustrated in **Figure 1.**

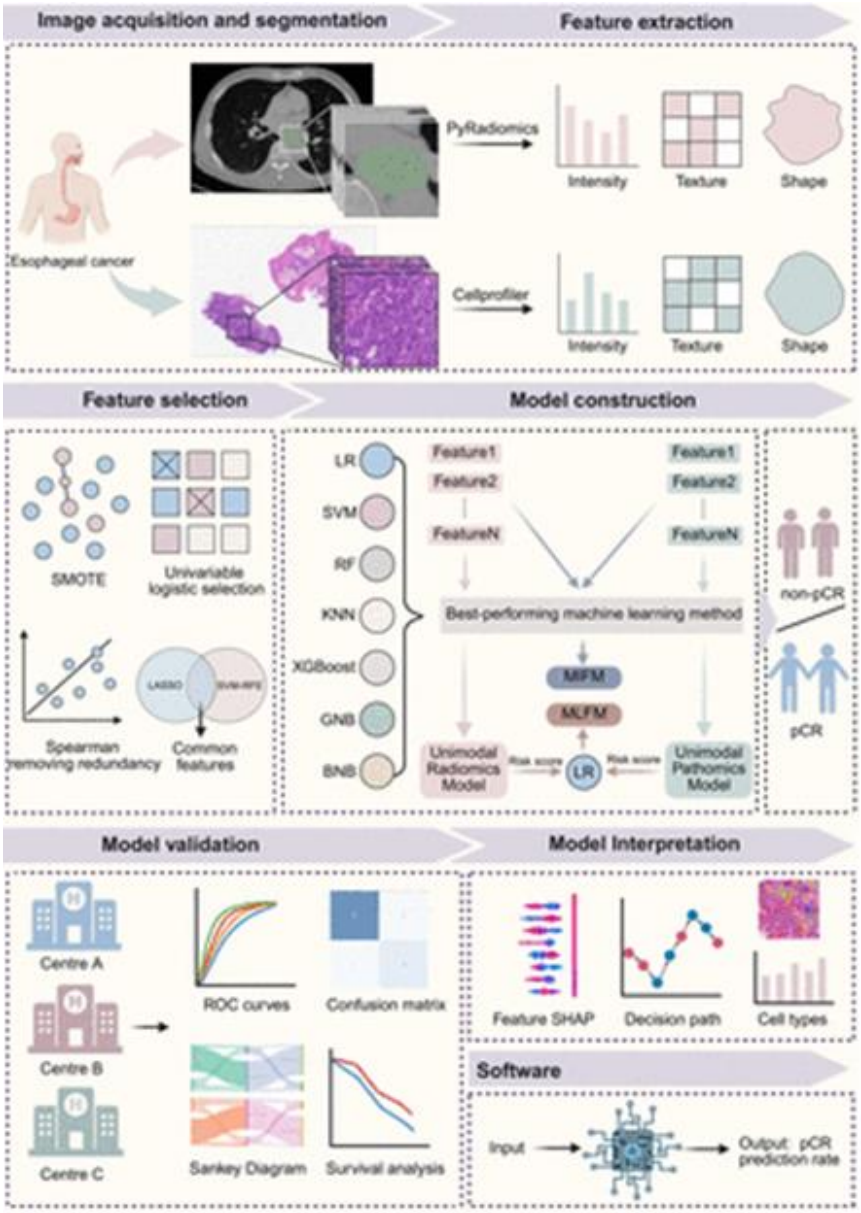


Figure 1. Overview of the research workflow.

Pre-treatment contrast-enhanced CT scans and digitized H&E-stained whole-slide images (WSIs) were collected from 335 patients diagnosed with esophageal squamous cell carcinoma (ESCC) at three different medical centers. Tumor boundaries were hand-drawn on the CT slices, and the most tumor-dense areas were carefully picked from each WSI. Quantitative radiomics features were obtained using PyRadiomics, and pathomics features were calculated with CellProfiler; both feature sets were then filtered for quality and relevance. Four separate prediction models were constructed: a radiomics-only model, a pathomics-only model, an intermediate multimodal fusion model (MIFM), and a late multimodal fusion model (MLFM). These were tested on the training group, internal validation group, and independent external group using receiver operating characteristic (ROC) curves, confusion matrices, Sankey flow diagrams for reclassification, and survival outcome comparisons. To ensure transparency, we applied SHAP value analysis, visualized decision pathways for individual patients, and quantified different cell populations. A simple browser-based interface was built that takes CT images with drawn regions of interest (ROIs) and CellProfiler results as input and returns the estimated probability of pathologic complete response (pCR) for each patient. BNB, Bernoulli Naïve Bayes; ESCC, esophageal squamous cell carcinoma; GNB, Gaussian Naïve Bayes; KNN, k-nearest neighbors; LASSO, Least Absolute Shrinkage and Selection Operator; LR, logistic regression; MIFM, multimodal intermediate fusion model; MLFM, multimodal late fusion model; pCR, pathologic complete response; RF, random forest; ROC, receiver operating characteristic; ROI, region of

interest; SHAP, SHapley Additive exPlanations; SMOTE, Synthetic Minority Over-sampling Technique; SVM-RFE, Support Vector Machines-Recursive Feature Elimination; XGBoost, eXtreme Gradient Boosting; WSIs, whole-slide images.

Patient recruitment

We retrospectively gathered data on all patients with biopsy-confirmed ESCC who completed neoadjuvant chemoimmunotherapy (nCIT) and then underwent surgery with curative intent at three major academic hospitals: Zhejiang Cancer Hospital, Renmin Hospital of Wuhan University, and Tianjin Medical University Cancer Institute and Hospital, spanning July 2019 to July 2023 (total 335 patients). At Zhejiang Cancer Hospital (296 patients), cases were randomly assigned in a 6:4 ratio to either the training dataset or an independent internal test set (Test-set-1). The external test set (Test-set-2, 39 patients) came from Renmin Hospital of Wuhan University (July 2020–September 2023, 22 patients) and Tianjin Medical University Cancer Institute and Hospital (June 2020–February 2022, 17 patients). Every patient had a contrast-enhanced chest CT performed no more than 14 days before starting nCIT, and H&E-stained whole-slide images were created from endoscopic biopsy tissue taken within 7 days of the CT scan. Full eligibility and exclusion rules, along with a detailed patient selection diagram, appear in **Figure 1.**

Treatment regimen and pathology assessment

Each patient received a minimum of one cycle combining neoadjuvant immunotherapy with chemotherapy. Immunotherapy used standard 200 mg doses every 3 weeks of various PD-1 or PD-L1 inhibitors (tislelizumab, sintilimab, durvalumab, envafolimab, pembrolizumab, camrelizumab, or nivolumab). Chemotherapy followed platinum-based doublets: TC schedule (repeated every 3 weeks): 1–4 cycles of nab-paclitaxel 260 mg/m² or paclitaxel 135–175 mg/m² on day 1, combined with carboplatin AUC 5 mg/mL/min on day 1; TP schedule (repeated every 3 weeks): 1–4 cycles of nab-paclitaxel 260 mg/m² or paclitaxel 175 mg/m² on day 1, combined with cisplatin 75 mg/m² on day 1.

Surgery (esophagectomy with curative aim) took place 4–8 weeks after the last nCIT cycle. The choice of open or minimally invasive technique and the extent of lymph node removal (two-field or three-field) depended on tumor location and the operating surgeon's judgment.

Pathologists with extensive experience examined the surgical specimens, with a senior esophageal cancer specialist providing final confirmation. Tumor regression was graded using the College of American Pathologists Esophageal Carcinoma Protocol [26]: TRG 0 = no remaining cancer cells (complete response); TRG 1 = only isolated cells or tiny clusters; TRG 2 = noticeable regression but more than minimal residual cancer; TRG 3 = abundant residual tumor with little or no regression

Pathologic complete response (pCR) was defined exclusively as TRG 0 in the primary tumor site; all other grades (TRG 1–3) were grouped as non-pCR. This binary outcome was the main target for building and testing the prediction models.

Image acquisition and tumor outlining

Two radiologists (HS and XW), each with more than 3 years of experience, manually outlined the primary esophageal tumors on contrast-enhanced CT images to define regions of interest (ROIs). They were unaware of pathology results or any model predictions during this step. A senior radiologist (YJ) with over 25 years of experience reviewed every contour, made adjustments when required, and resolved any differences through discussion. The final agreed-upon ROIs became the standard for radiomics analysis. All outlining was done in 3D Slicer software (version 5.1.0) [27].

H&E-stained slides were fixed in formalin, embedded in paraffin, scanned at 20× magnification, and converted into whole-slide images (WSIs). A thoracic pathologist (BQ) with 3 years of experience, blinded to clinical and outcome data, identified five representative tumor-rich fields of view on each WSI. These fields were cropped into 512×512-pixel patches and saved as PNG files. Every patch passed a strict visual quality check to remove any with staining issues, folds, chatter, sparse tissue, bubbles, ink marks, or other artifacts.

Feature calculation and refinement

Radiomics features (1,094 total) were extracted from the finalized CT ROIs using PyRadiomics [28] (version 3.0.1). These included shape and volume measurements, first-order intensity statistics, multiple texture matrices

(co-occurrence, size zone, run length, dependence, neighboring gray-tone difference), and wavelet-filtered versions.

Pathomics features (4,892 total) were computed from the H&E WSIs with CellProfiler [29] (version 4.2.8) through an automated workflow that captured intensity patterns, spatial arrangements, cell and tissue shapes, texture properties, and fractional area measurements **(Figure 2**). The five patches per case were averaged to produce one set of slide-level features per patient.

Feature selection occurred only in the training set and was performed separately for each modality. The steps were: Address class imbalance using Synthetic Minority Over-sampling Technique (SMOTE); Normalize all features with Z-score transformation; Apply univariate logistic regression to keep features with $p < 0.05$ (radiomics) or $p < 0.01$ (pathomics); Remove redundant pairs by calculating Spearman correlation (discard one when $|\rho| > 0.85$, keeping the feature with stronger outcome association); Run two independent selectors—LASSO with 10-fold cross-validation and SVM-RFE—and use only the overlapping features for final model training.

Model validation and construction

For single-modality modeling, seven supervised machine learning approaches were systematically assessed for both the radiomics and pathomics feature spaces, including logistic regression, Gaussian and Bernoulli Naïve Bayes classifiers, support vector machines, random forest ensembles, k-nearest neighbors, and eXtreme Gradient Boosting (XGBoost). Hyperparameters for each algorithm were optimized via grid search coupled with fivefold cross-validation. To ensure experimental reproducibility, a fixed random seed was consistently applied during the tuning process.

Two distinct strategies were explored to integrate radiomics and pathomics information in the multimodal setting. In the intermediate fusion framework (MIFM), features derived from both modalities were concatenated to form a unified feature vector, which was then used to train a model using the algorithm that demonstrated the strongest performance during unimodal evaluation. In contrast, the late fusion framework (MLFM) combined modalities at the prediction level: the optimal radiomics and pathomics models identified during unimodal screening were first independently trained, after which their output probabilities were integrated using a logistic regression meta-classifier. Overall, four categories of predictive models were constructed: radiomics-only, pathomics-only, multimodal intermediate fusion, and multimodal late fusion.

Model generalizability was assessed using two independent validation cohorts (Test-set-1 and Test-set-2), without any additional parameter tuning. Validation data underwent identical preprocessing steps as the development data; however, no resampling strategies (such as SMOTE) or further normalization procedures were applied to the test cohorts. Model discrimination and classification performance were evaluated using the area under the receiver operating characteristic curve (AUC), accuracy, sensitivity, specificity, and F1 score. Confidence intervals were estimated using bootstrap resampling with 1,000 iterations for each validation cohort, and 95% confidence intervals were reported for all metrics. Receiver operating characteristic and precision–recall curves were generated to facilitate visual comparison among models. For the highest-performing model, decision curve analysis (DCA) was additionally performed to evaluate potential clinical benefit across a range of threshold probabilities.

Model interpretation

To clarify how the integrated model arrived at its predictions, we adopted SHapley Additive exPlanations (SHAP) as a quantitative interpretability strategy. This approach estimates the marginal impact of each radiomics and pathomics variable on the predicted likelihood of pathological complete response, both for individual patients and across the study population. The Shapley value is defined as follows:

$$\emptyset_j = \sum S \subseteq N\backslash\{j\} \frac{|S|!\,(N-|S|-1)!}{N!}(\nu(S\cup\{j\})-\nu(S)) \quad (1)$$

where: $\nu(S\cup\{j\})-\nu(S)$ denotes the incremental contribution of feature j within coalition S; $\sum[S\subseteq N\{j\}]$ represents the summation across all possible feature combinations; and $\frac{|S|!(N-|S|-1)!}{N!}$ corresponds to the weighting coefficient assigned to each coalition. At the individual level, SHAP visual outputs illustrate how specific feature values drive the predicted probability above or below the model's baseline estimate. At the cohort level,

aggregated SHAP analyses order features by their overall contribution and reveal whether their influence predominantly favors or opposes pCR.

Beyond global attribution, we reconstructed patient-specific inference trajectories from the trained classifier to expose the internal decision logic in a sequential manner. For each case, these trajectories enumerate the ordered decision conditions evaluated by the model, explicitly identifying the contributing feature, its threshold value, and the magnitude of its incremental effect on the prediction (expressed in log-odds or probability units). The cumulative effect of these steps yields the final estimated pCR probability. This visualization enables direct identification of the features and value ranges that most decisively shifted a given prediction toward pCR or non-pCR, thereby bridging population-level importance with individualized reasoning.

To further investigate histopathological correlates underlying model predictions, we analyzed the tumor microenvironment (TME) in H&E-stained whole-slide images using Hover-Net [30], an open-source deep learning framework for nuclear segmentation and coarse cell classification. Within the analyzed image regions, nuclei were categorized into tumor cells, lymphocytes, connective tissue cells, necrotic cells, and an additional miscellaneous group. The relative abundance of these cell populations was subsequently compared according to pathological response (observed pCR versus observed non-pCR) and, independently, according to model-based stratification (predicted pCR versus predicted non-pCR).

Statistical analysis

Baseline clinical characteristics were analyzed using SPSS software (version 27). Categorical variables were compared using Pearson's chi-square test or likelihood-ratio test, as appropriate, while continuous variables were evaluated using analysis of variance (ANOVA) or the Kruskal–Wallis H test. The predictive relevance of clinical covariates was assessed through univariable logistic regression analysis. A two-sided p value below 0.05 was considered statistically significant.

Overall survival comparisons were conducted between patients with observed pCR and non-pCR, as well as between groups stratified by model-predicted response status. Survival curves were generated using the Kaplan–Meier method, with group differences evaluated using the log-rank test based on a prespecified probability threshold. Hazard ratios (HRs) and corresponding 95% confidence intervals were estimated using Cox proportional hazards regression models.

Survival analyses were performed in R software (version 4.4.2) using the "survival" package (version 3.8.3), with visualization facilitated by the "survminer" package (version 0.5.0). Machine learning analyses were implemented in Python (version 3.13.1) using the "scikit-learn" library (version 1.6.1) and the "xgboost" package (version 3.0.0).

Results and Discussion

Patient characteristics

Baseline demographic and clinicopathological characteristics stratified by pathological response are summarized in **Table 1**. Among the 335 included patients, 77 (22.99%) achieved pathological complete response, whereas 258 (77.01%) were classified as non-pCR. Comparisons between groups revealed no significant differences across most clinical variables, with the exception of smoking history, number of nCIT cycles administered, and the number of suspicious lymph nodes (s-LNs), which differed significantly between response groups ($p < 0.05$).

Table 1. Clinical and demographic profile of patients included across all cohorts

Characteristic	P value	Overall (N=335)	Test set-2 (n=39)	Test set-1 (n=115)	Training set (n=181)
Sex	0.189				
Male		312 (93.13%)	34 (87.18%)	110 (95.65%)	168 (92.82%)
Female		23 (6.87%)	5 (12.82%)	5 (4.35%)	13 (7.18%)
Age (median [range])	0.825	64 [44–82]	62 [48–76]	65 [46–77]	64 [44–82]
Drinking status	0.047*				
Current or former		239 (71.34%)	22 (56.41%)	80 (69.57%)	137 (75.69%)
Never		96 (28.66%)	17 (43.59%)	35 (30.43%)	44 (24.31%)
Smoking status	0.286				

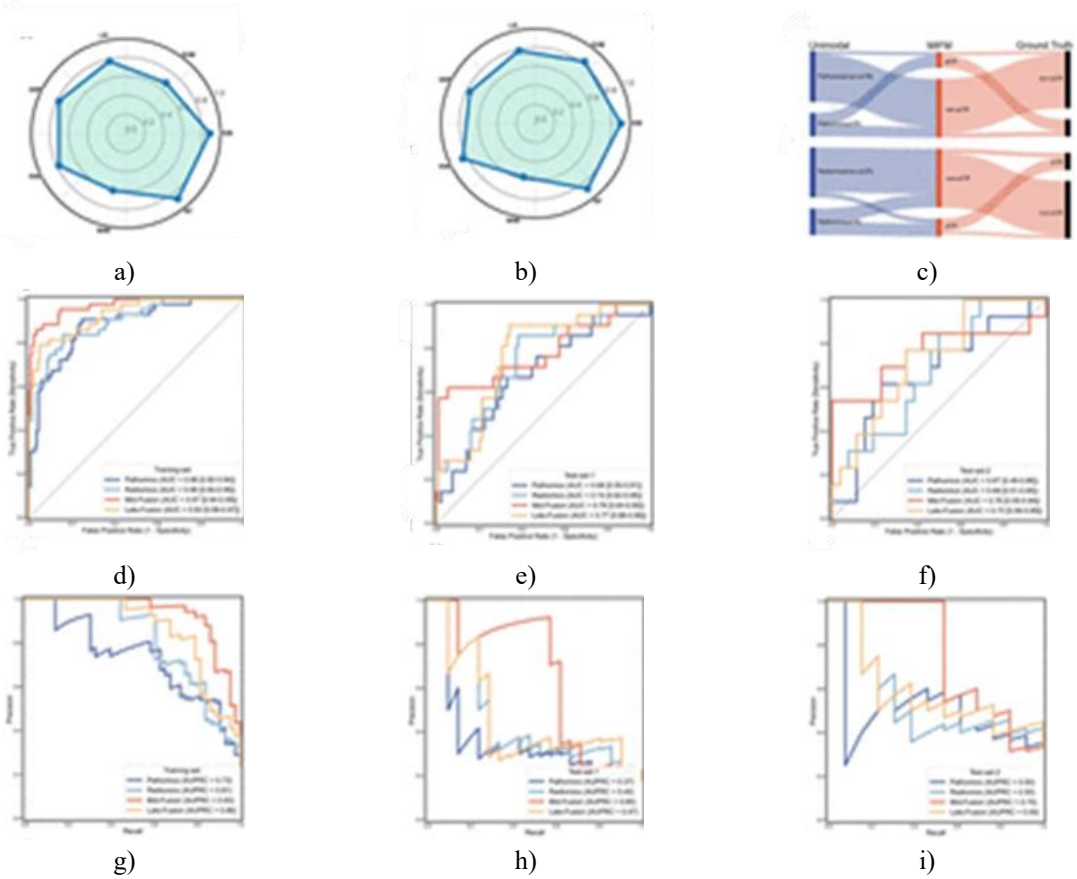
Current or former		220 (65.67%)	27 (69.23%)	69 (60.00%)	124 (68.51%)
Never		115 (34.33%)	12 (30.77%)	46 (40.00%)	57 (31.49%)
Tumor location	0.251				
Lower		120 (35.82%)	17 (43.59%)	47 (40.87%)	56 (30.94%)
Middle		169 (50.45%)	19 (48.72%)	51 (44.35%)	99 (54.70%)
Upper		46 (13.73%)	3 (7.69%)	17 (14.78%)	26 (14.36%)
ECOG performance status	0.269				
0		146 (43.58%)	21 (53.85%)	44 (38.26%)	81 (44.75%)
1		184 (54.93%)	18 (46.15%)	70 (60.87%)	96 (53.04%)
2		5 (1.49%)	0 (0%)	1 (0.87%)	4 (2.21%)
Clinical T stage (cT)	0.127				
1		3 (0.90%)	0 (0%)	1 (0.87%)	2 (1.10%)
2		50 (14.92%)	2 (5.13%)	19 (16.52%)	29 (16.02%)
3		267 (79.70%)	32 (82.05%)	90 (78.26%)	145 (80.11%)
4a		15 (4.48%)	5 (12.82%)	5 (4.35%)	5 (2.77%)
Clinical TNM stage (AJCC 8th edition)	0.475				
I		3 (0.90%)	0 (0%)	1 (0.87%)	2 (1.10%)
II		74 (22.09%)	8 (20.52%)	26 (22.61%)	40 (22.10%)
III		228 (68.06%)	23 (58.97%)	78 (67.82%)	127 (70.16%)
IVA		30 (8.95%)	8 (20.51%)	10 (8.70%)	12 (6.64%)
Clinical N stage (cN)	0.707				
0		47 (14.03%)	7 (17.95%)	16 (13.92%)	24 (13.26%)
1		170 (50.75%)	15 (38.46%)	60 (52.17%)	95 (52.49%)
2		105 (31.34%)	14 (35.90%)	35 (30.43%)	56 (30.94%)
3		13 (3.88%)	3 (7.69%)	4 (3.48%)	6 (3.31%)
Neoadjuvant immunotherapy cycles	<0.001*				
≤2		271 (80.90%)	22 (56.41%)	96 (83.48%)	153 (84.53%)
>2		64 (19.10%)	17 (43.59%)	19 (16.52%)	28 (15.47%)
Immunotherapy regimen	0.199				
PD-L1 inhibitor		37 (11.04%)	1 (2.56%)	14 (12.17%)	22 (12.15%)
PD-1 inhibitor		298 (88.96%)	38 (97.44%)	101 (87.83%)	159 (87.85%)
R0 resection	0.556				
Yes		314 (93.73%)	38 (97.44%)	108 (93.91%)	168 (92.82%)
No		21 (6.27%)	1 (2.56%)	7 (6.09%)	13 (7.18%)
Lymphadenectomy extent	0.243				
Two-field		36 (10.75%)	7 (17.95%)	13 (11.30%)	16 (8.84%)
Three-field		299 (89.25%)	32 (82.05%)	102 (88.70%)	165 (91.16%)
Surgical approach	0.103				
Minimally invasive		310 (92.54%)	33 (84.62%)	106 (92.17%)	171 (94.48%)
Open		25 (7.46%)	6 (15.38%)	9 (7.83%)	10 (5.52%)
Pathological complete response (tumor pCR)	0.144				
Yes		77 (22.99%)	13 (33.33%)	21 (18.26%)	43 (23.76%)
No		258 (77.01%)	26 (66.67%)	94 (81.74%)	138 (76.24%)
Pathological N stage (ypN)	0.465				
0		189 (56.42%)	18 (46.15%)	67 (58.26%)	104 (57.46%)
1		91 (27.16%)	15 (38.46%)	32 (27.83%)	44 (24.31%)
2		41 (12.24%)	5 (12.83%)	10 (8.70%)	26 (14.36%)
3		14 (4.18%)	1 (2.56%)	6 (5.21%)	7 (3.87%)
Pathological T stage (ypT)	0.052				

0		77 (22.99%)	13 (33.33%)	21 (18.26%)	43 (23.76%)
1		70 (20.90%)	13 (33.33%)	24 (20.87%)	33 (18.23%)
2		62 (18.51%)	7 (17.96%)	23 (20.00%)	32 (17.68%)
3		126 (37.60%)	6 (15.38%)	47 (40.87%)	73 (40.33%)
Pathological TNM stage (ypTNM, AJCC 8th edition)	0.550				
I		151 (45.07%)	16 (41.03%)	51 (44.35%)	83 (45.86%)
II		50 (14.93%)	3 (7.69%)	20 (17.39%)	27 (14.92%)
III		134 (40.00%)	20 (51.28%)	44 (38.26%)	71 (39.22%)
Survival time, days (median [range])	0.137	692 [96–1772]	381 [136–1240]	716 [96–1172]	727 [100–1661]
Number of dissected lymph nodes (median [range])	0.042*	23 [5–78]	23 [8–61]	22 [5–78]	24 [6–63]

- Values are presented as counts with corresponding percentages unless indicated otherwise.
- P values were derived from comparisons among the training cohort, Test-set-1, and Test-set-2.
- The s-LN count refers to the total number of lymph nodes excised during surgical resection.
- A two-sided P value <0.05 was used to denote statistical significance.
- AJCC: American Joint Committee on Cancer cN: node stage, clinical cT: tumor stage, clinical cTNM: Tumor-Node-Metastasis, clinical ECOG: Eastern Cooperative Oncology Group NCIT: chemoimmunotherapy, neoadjuvant pCR: complete response, pathologic PD-1: protein 1, programmed cell death PD-L1: ligand 1, programmed cell death s-LN number: lymph nodes dissected surgically, number of ypN: node stage, pathologic after neoadjuvant therapy ypT: tumor stage, pathologic after neoadjuvant therapy ypTNM: Tumor-Node-Metastasis, pathologic after neoadjuvant therapy

Performance of single-modality models

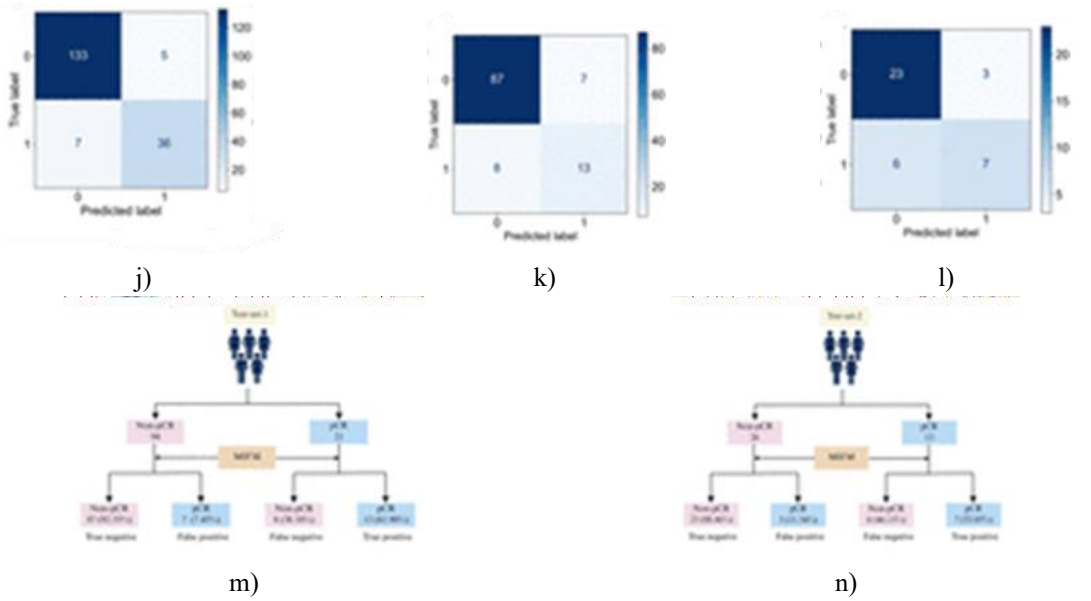
Following the feature reduction process, the unimodal signatures consisted of 14 radiomics features and 11 pathomics features. Performance comparisons across the seven candidate machine learning classifiers demonstrated that the XGBoost algorithm achieved the most stable and robust predictive performance across the training cohort as well as both validation cohorts **(Figures 2a and 2b)**. Consequently, XGBoost was chosen to construct the final radiomics-only and pathomics-only models using their respective optimized feature subsets.



a) b) c)

d) e) f)

g) h) i)



j) k) l)

m) n)

Figure 2. Overall predictive performance of the proposed models for pathological complete response (pCR). Panels a and b display radar plots summarizing the AUCs of seven machine learning classifiers trained using pathomics features alone and radiomics features alone, respectively, in the training cohort. Panel C presents a Sankey diagram illustrating how predictions from the unimodal models are redistributed after integration into the multimodal intermediate fusion model (MIFM), with reference to the true pCR status. Discriminative ability of the unimodal and multimodal approaches is illustrated using ROC curves for the training cohort (d), test cohort 1 (e), and test cohort 2 (f). Corresponding precision–recall curves are shown in panels g–i. Panels j–l report confusion matrices for the MIFM across the training and two independent test cohorts. Panels m and n depict flow charts comparing MIFM-assigned classes with observed outcomes in test cohorts 1 and 2, highlighting the distribution of correctly and incorrectly classified cases. AUC, area under the curve; AUPRC, area under the precision–recall curve; BNB, naïve Bayes, Bernoulli; GNB, naïve Bayes, Gaussian; KNN, neighbors, k-nearest; LR, regression, logistic; MIFM, model, multimodal intermediate fusion; MLFM, model, multimodal late fusion; pCR, complete response, pathological; PR, precision–recall; RF, forest, random; ROC, receiver operating characteristic; SVM, machine, support vector; XGB, Boosting, eXtreme Gradient.

Across cohorts, the pathomics-based model demonstrated an AUC of 0.88 (95% CI, 0.82–0.94) in the training cohort, which declined to 0.68 (95% CI, 0.55–0.81) and 0.67 (95% CI, 0.48–0.86) in test cohorts 1 and 2, respectively. The radiomics-based model showed slightly stronger discrimination, achieving AUCs of 0.90 (95% CI, 0.84–0.95) in training, 0.74 (95% CI, 0.62–0.85) in test cohort 1, and 0.68 (95% CI, 0.51–0.85) in test cohort 2 **(Figures 2d–2f).** In the training cohort, radiomics also outperformed pathomics in terms of classification sensitivity, accuracy, and specificity (0.84, 0.77, and 0.86 vs. 0.81, 0.72, and 0.83), with detailed metrics summarized in **Table 2**.

Table 2. Quantitative evaluation of model performance for predicting pathological complete response.

Dataset	Model	Area Under the Curve (95% Confidence Interval)	Overall Accuracy (95% CI)	Sensitivity (95% CI)	Specificity (95% CI)	F1 Score (95% CI)
Training Set	Unimodal Pathomics Model	0.88 (0.82–0.94)	0.81 (0.75–0.86)	0.72 (0.58–0.85)	0.83 (0.77–0.89)	0.64 (0.52–0.74)
	Unimodal Radiomics Model	0.90 (0.84–0.95)	0.84 (0.79–0.90)	0.77 (0.64–0.89)	0.86 (0.80–0.92)	0.69 (0.58–0.80)
	Multi-Input Fusion Model (MIFM)	0.97 (0.94–0.99)	0.93 (0.90–0.97)	0.84 (0.71–0.95)	0.96 (0.93–0.99)	0.86 (0.77–0.93)

	Multi-Level Fusion Model (MLFM)	0.93 (0.88–0.97)	0.89 (0.85–0.93)	0.79 (0.67–0.91)	0.92 (0.87–0.96)	0.77 (0.68–0.86)
Test Set 1	Unimodal Pathomics Model	0.68 (0.55–0.81)	0.68 (0.59–0.77)	0.52 (0.32–0.75)	0.71 (0.61–0.80)	0.37 (0.21–0.52)
	Unimodal Radiomics Model	0.74 (0.62–0.85)	0.69 (0.61–0.77)	0.62 (0.39–0.83)	0.70 (0.61–0.80)	0.42 (0.26–0.57)
	Multi-Input Fusion Model (MIFM)	0.78 (0.64–0.90)	0.87 (0.81–0.92)	0.62 (0.41–0.83)	0.93 (0.87–0.98)	0.63 (0.44–0.79)
	Multi-Level Fusion Model (MLFM)	0.77 (0.66–0.86)	0.70 (0.61–0.77)	0.57 (0.35–0.78)	0.72 (0.63–0.81)	0.41 (0.24–0.56)
Test Set 2	Unimodal Pathomics Model	0.67 (0.48–0.86)	0.69 (0.56–0.82)	0.54 (0.27–0.82)	0.77 (0.60–0.92)	0.54 (0.27–0.74)
	Unimodal Radiomics Model	0.68 (0.51–0.85)	0.59 (0.44–0.74)	0.54 (0.29–0.80)	0.62 (0.42–0.81)	0.47 (0.22–0.67)
	Multi-Input Fusion Model (MIFM)	0.76 (0.55–0.94)	0.77 (0.64–0.90)	0.54 (0.27–0.83)	0.88 (0.74–1.00)	0.61 (0.33–0.82)
	Multi-Level Fusion Model (MLFM)	0.73 (0.56–0.89)	0.64 (0.49–0.79)	0.54 (0.25–0.82)	0.69 (0.50–0.86)	0.50 (0.24–0.71)

MLFM, multimodal late fusion model; MIFM, multimodal intermediate fusion model; AUC, area under curve.

Because of the imbalance between response classes, precision–recall analysis was performed as a complementary evaluation. The pathomics-based classifier achieved AUPRC values of 0.73 in the training cohort, decreasing to 0.37 in test cohort 1 and 0.50 in test cohort 2. The radiomics-based classifier consistently showed higher PR performance, with AUPRCs of 0.81, 0.45, and 0.55 in the corresponding cohorts **(Figures 2g–2i)**.

Multimodal model performance

Using the fusion strategies defined a priori, we constructed both intermediate- and late-fusion multimodal frameworks. In all datasets, these multimodal approaches surpassed the unimodal radiomics and pathomics models across all assessed metrics (**Table 2 and Figures 2d–2i).** Direct comparison between the two fusion strategies revealed superior overall classification performance for the multimodal intermediate fusion model (MIFM), which achieved higher sensitivity, specificity, accuracy, and F1 score than the multimodal late fusion model (MLFM) **(Table 2)**. Evaluation of the MIFM confusion matrices **(Figures 2j–2n)** demonstrated a strong ability to correctly identify non-pCR cases, with 87 true negatives in test cohort 1 and 23 in test cohort 2, resulting in the highest specificity among all evaluated models **(Table 2).** The Sankey visualization illustrated how integrating modalities corrected a substantial proportion of unimodal misclassifications, with a clear net shift toward concordance with true labels **(Figure 2c).**

Exploratory survival stratification based on observed and predicted pCR

We further explored whether overall survival (OS) differed according to either pathological pCR status or pCR predicted by the MIFM using a predefined decision threshold. In the training cohort, patients who achieved a pathological pCR exhibited prolonged OS, as reflected by a clear separation of Kaplan–Meier curves **(Figure 3a).** However, this distinction was not statistically significant in either external test cohort **(Figures 3c and 3d).** When patients were stratified according to the model-predicted pCR status, a comparable trend was observed, although no cohort demonstrated statistically significant survival separation **(Figures 3b, 3d and 3f).** Consistent with these findings, univariate Cox proportional hazards analysis showed that both observed pCR and MIFM-predicted pCR were significantly associated with OS ($p < 0.005$).

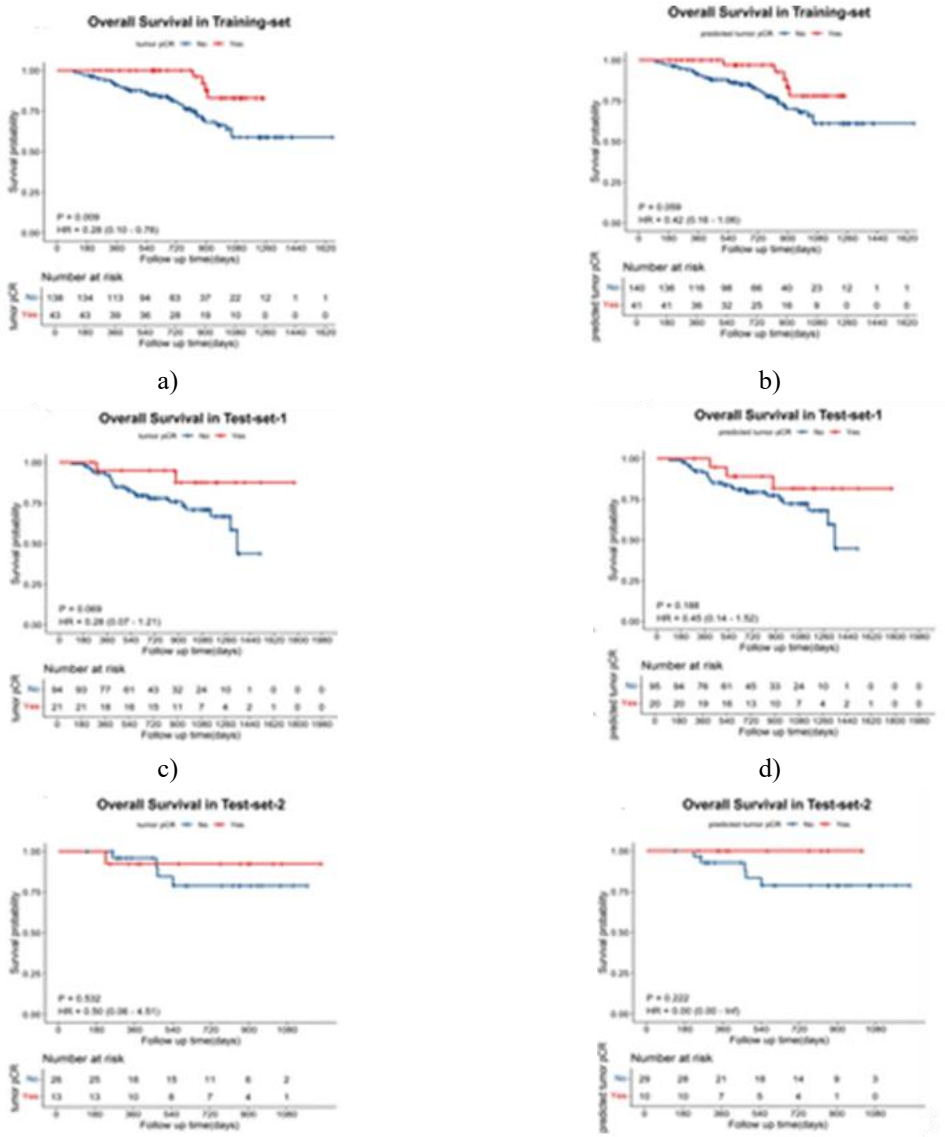
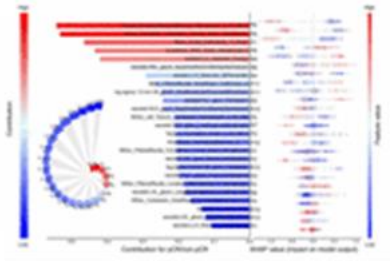


Figure 3. Evaluation of prognostic stratification capability. Kaplan–Meier survival curves for overall survival are shown according to pathological response status (pCR vs. non-pCR) in the training cohort (a), test cohort 1 (c), and test cohort 2 (e). Corresponding survival analyses based on pCR status predicted by the multimodal intermediate fusion model (MIFM) are presented for the training cohort (b), test cohort 1 (d), and test cohort 2 (f).

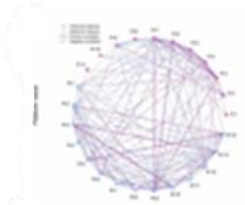
Abbreviations: KM, Kaplan–Meier; MIFM, multimodal intermediate fusion model; OS, overall survival; pCR, pathological complete response.

Model interpretability and clinical software implementation

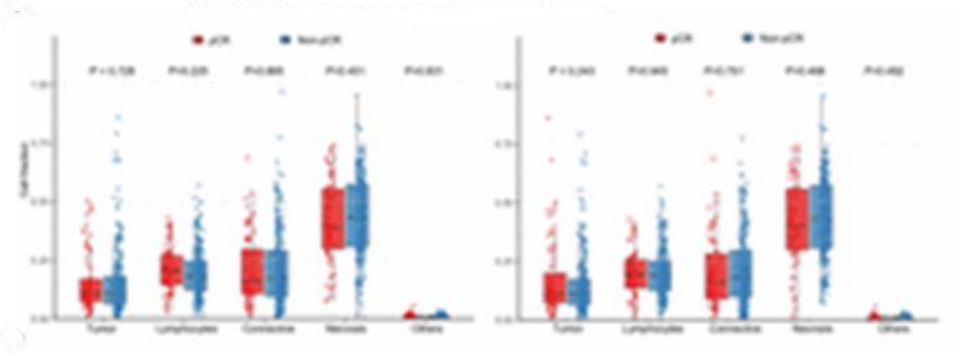
Feature attribution analysis using SHAP demonstrated how individual radiomics and pathomics variables influenced MIFM output probabilities **(Figure 4a)**. Examination of inter-feature relationships revealed weak to moderate correlations between radiomics and pathomics inputs (ranging from −0.57 to 0.69) **(Figure 4b)**, indicating that each modality contributes largely nonredundant information. This pattern supports the premise that integrating imaging- and tissue-derived features enables a more comprehensive representation of tumor characteristics and therapeutic response.



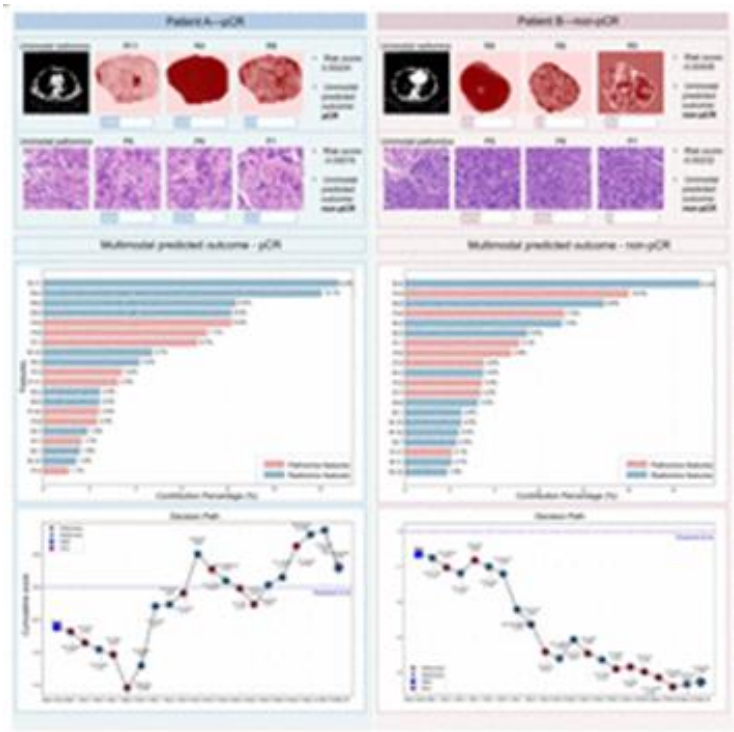
a)



b)



c)



d)

Figure 4. Model interpretability and explanatory analyses. (a) SHAP-based interpretation of the multimodal intermediate fusion model (MIFM), with a beeswarm visualization depicting feature-specific influences on individual predictions (right) and a ranked bar plot summarizing overall feature importance using mean absolute SHAP values (left); the inset illustrates relative modality contributions. (b) A Spearman correlation network illustrating relationships between retained radiomics and pathomics features, where edge thickness represents correlation strength. (c) Distribution of whole-slide image–derived cellular composition, including tumor, lymphocyte, stromal, necrotic, and other components, shown as boxplots and stratified by pathological response (left) and by MIFM-predicted response (right). (d) Representative case examples highlighting both conflicting and concordant unimodal predictions: Patient A (true pCR) demonstrates how MIFM integrates discordant radiomics and pathomics evidence, while Patient B (true non-pCR) illustrates agreement across modalities. For each case, spatial feature visualizations, rankings of the top 20 contributing variables, and

decision-path diagrams depict how modality-specific information is progressively integrated into the final output. MIFM, multimodal intermediate fusion model; pCR, pathological complete response; SHAP, SHapley Additive exPlanations; WSI, whole-slide image.

Comparative analysis of cellular composition revealed consistent trends when stratifying by either observed pathological response or model-derived predictions **(Figure 4c)**. In both comparisons, pCR groups—whether defined by histopathology or by MIFM output—were characterized by relatively increased tumor and lymphocyte proportions and reduced necrotic content compared with their respective non-pCR counterparts.

Figure 4d illustrates the interpretability workflow using two illustrative cases to clarify how predictions are generated. In the first example, disagreement between unimodal radiomics and pathomics classifiers was resolved through weighted integration within the multimodal framework. In the second example, both unimodal models reached the same conclusion. For each case, heatmaps highlight spatial patterns of influential imaging and tissue features, feature-attribution plots quantify how individual variables shifted the prediction toward or away from pCR, and decision-path visualizations outline the sequential aggregation of evidence that produced the final probability estimate. Collectively, these elements connect low-level feature signals with transparent, case-specific model reasoning.

To facilitate practical use, we developed a web-based graphical interface that operates without programming expertise. Users provide a pretreatment CT scan, the corresponding tumor region-of-interest mask, and a CellProfiler-generated CSV file, after which the system outputs an individualized pCR probability together with basic data validation results. The software prototype is intended for research purposes only.

Accurate preoperative identification of pathological complete response (pCR) after neoadjuvant chemoimmunotherapy (nCIT) remains an unmet clinical challenge in esophageal squamous cell carcinoma (ESCC). In this multicenter retrospective study spanning three academic institutions, we addressed this gap by developing and externally validating an interpretable multimodal machine learning approach that jointly leverages contrast-enhanced CT–derived radiomics and H&E-stained whole-slide image (WSI)–based pathomics. By integrating routinely acquired imaging and biopsy data, the proposed framework enables noninvasive estimation of pCR likelihood prior to surgery. Compared with single-modality models, the intermediate fusion strategy demonstrated more consistent and generalizable performance across the development cohort and two independent validation cohorts, supporting the premise that radiologic and histopathologic information provide complementary signals relevant to treatment response.

A defining feature of the present work is its emphasis on clinical feasibility and transparency. All input data are part of standard diagnostic workflows, eliminating the need for additional assays, specialized imaging, or increased cost. Moreover, model interpretability was embedded as a design constraint rather than a post hoc consideration. Both radiomics and pathomics features were explicitly defined using mathematical and morphologic descriptors, and model outputs were accompanied by explanations at multiple levels, including global feature importance, spatial feature localization, and case-specific decision pathways. To facilitate exploratory clinical use, we further implemented a browser-based graphical interface that requires no programming expertise and returns individualized pCR probabilities. Together, these elements provide a pragmatic and interpretable foundation for future clinical decision support in nCIT-treated ESCC.

The clinical motivation for this work arises from the evolving neoadjuvant treatment landscape in esophageal cancer. The CROSS trial firmly established neoadjuvant chemoradiotherapy (nCRT) as superior to surgery alone in locally advanced ESCC [31]. However, despite improved local control, distant metastasis remains the predominant mode of failure after nCRT, occurring far more frequently than local recurrence (22.0% vs 5.9%). This observation has prompted ongoing efforts to intensify systemic therapy. Subsequent studies demonstrated that more intensive chemotherapy regimens can further improve overall survival (OS) and local disease control compared with nCRT [32, 33], raising interest in treatment de-escalation strategies such as omission of esophagectomy in patients achieving pCR [34]. Unfortunately, reliable noninvasive tools to identify pCR are lacking.

Concurrently, immunotherapy has reshaped systemic treatment paradigms for esophageal cancer. Several studies have reported encouraging outcomes with combined chemotherapy and immunotherapy in the first-line treatment of advanced disease [35, 36], suggesting potential applicability in the neoadjuvant setting. Direct comparisons between nCIT and nCRT for locally advanced ESCC are ongoing. In a prospective multicenter study across eight

high-volume centers, Guo *et al* reported superior 2-year OS and disease-free survival with nCIT compared with nCRT, while pCR rates were similar and major pathological response favored nCRT [6]. Although the optimal neoadjuvant regimen remains unsettled, these findings collectively point to a growing role for chemoimmunotherapy and reinforce the need for predictive tools tailored to this treatment context.

From a clinical management perspective, the inability to determine pCR status before surgery limits individualized treatment strategies. Patients who ultimately achieve pCR may be candidates for nonoperative management, potentially avoiding the morbidity of esophagectomy and preserving quality of life [37–39]. In contrast, patients unlikely to achieve pCR benefit from timely surgical resection to eradicate residual disease. Accurate preoperative stratification is therefore essential for balancing the risks of overtreatment and undertreatment. Our multimodal framework directly addresses this clinical need by providing a noninvasive estimate of pCR probability, thereby informing decisions between surveillance and prompt surgery.

Multimodal data integration has shown promise for response prediction in multiple oncologic settings. Mao *et al* combined MRI, WSIs, and clinical variables to predict pCR after neoadjuvant chemotherapy in breast cancer [40], though their reliance on deep learning–derived features limited semantic interpretability. In ESCC, Qi *et al* demonstrated the feasibility of combining CT imaging with WSIs to predict pCR after nCIT [41], but their study involved a relatively small paired dataset and pathomics features extracted via deep learning. In contrast, the present study represents one of the largest multimodal investigations of pCR prediction after nCIT in ESCC to date, incorporating paired pretreatment CT scans and biopsy WSIs from three centers with external validation. By relying on handcrafted, explicitly defined radiomics and pathomics features, our approach enhances interpretability while maintaining competitive predictive performance.

Notably, the multimodal intermediate fusion model (MIFM) consistently achieved higher specificity than unimodal models across all cohorts **(Table 2).** This operating characteristic is clinically meaningful, as it reduces the likelihood of falsely classifying non-pCR patients as pCR, thereby minimizing inappropriate adoption of watch-and-wait strategies. Patients predicted as pCR could instead be considered for cautious surveillance protocols with confirmatory assessments, while those predicted as non-pCR could proceed directly to surgery. Determining safe operating thresholds and workflows for such strategies will require prospective evaluation.

To address the challenges posed by high-dimensional feature spaces and limited cohort sizes [42, 43], we implemented a stringent feature selection pipeline confined to the training cohort. From an initial pool of 1,094 radiomics and 4,892 pathomics features, two complementary selection methods—LASSO regression and SVM recursive feature elimination—were applied, and only overlapping features were retained. Multimodal integration was performed using both intermediate and late fusion approaches, combining 14 radiomics and 11 pathomics features. Although late fusion has been reported to offer robustness in some contexts [44], our findings indicate that intermediate fusion—by explicitly modeling interactions between modalities—can yield superior specificity in this setting **(Table 2 and Figures 2d–2i).** Preserving modality-specific information while exploiting cross-modality complementarity may therefore enhance both discrimination and robustness.

Interpretability analyses further linked model behavior to biologically plausible imaging and histologic characteristics. SHAP analysis highlighted radiomics features capturing textural heterogeneity and intensity distribution, such as wavelet-based size zone nonuniformity and high-percentile intensity metrics, which may reflect tumor heterogeneity, vascularity, or necrosis. Pathomics features describing nuclear texture and morphology, including measures of hematoxylin staining uniformity and nuclear asymmetry, were also influential. These features align with known associations between nuclear irregularity, proliferative activity, and malignant potential [45], providing a conceptual bridge between quantitative image features and tumor biology.

Case-level explanations illustrated how multimodal integration resolves conflicting evidence. In a representative pCR case, discordant predictions from unimodal radiomics and pathomics models were reconciled by the MIFM through weighted integration of modality-specific features, resulting in a correct final classification. In contrast, a non-pCR case with concordant unimodal predictions demonstrated balanced contributions from both modalities. Decision-path visualizations explicitly traced how feature thresholds accumulated to cross—or remain below—the decision boundary, offering transparency into the model's reasoning process. These examples underscore the synergistic relationship between macroscopic radiologic patterns and microscopic tissue characteristics.

We also explored associations between tumor microenvironment (TME) composition and response status. The TME, comprising malignant cells, stromal components, vasculature, and immune infiltrates, plays a central role in tumor progression and response to immunotherapy [46–48]. Using H&E-stained WSIs, we quantified cell-type fractions and compared distributions according to observed and model-predicted pCR status. In both analyses,

pCR groups tended to show higher tumor and lymphocyte fractions and lower necrotic content, consistent with prior reports linking lymphocytic infiltration and preserved tumor architecture to treatment sensitivity [49, 50]. Although these trends did not reach statistical significance—likely due to limited sample size, biopsy heterogeneity, and segmentation variability—they are biologically plausible and warrant further investigation in larger prospective cohorts.

Several limitations should be acknowledged. The retrospective design and modest size of the external validation cohorts introduce potential bias and limit generalizability. Established immunotherapy biomarkers such as tumor mutational burden and PD-L1 expression were not incorporated because they were not uniformly available and would increase cost; future integration of such markers may improve performance. Manual segmentation and quality control of CT images and WSIs, despite standardized protocols, introduce subjectivity, highlighting the need for automated pipelines. Additionally, stain normalization and color augmentation were not applied and should be explored in future studies. Survival analyses were exploratory, as statistically significant Kaplan–Meier separation was observed only for observed pCR in the training cohort, likely reflecting limited event numbers and the multifactorial determinants of long-term outcomes. Finally, while this study emphasizes interpretability using handcrafted features, future work may benefit from interpretable deep learning approaches and validation of generated biological hypotheses using genomic data.

Conclusion

In summary, we present an interpretable multimodal machine learning framework that integrates contrast-enhanced CT radiomics with H&E-stained WSI pathomics to predict pCR before surgery in ESCC patients undergoing nCIT. By combining clinical practicality, external validation, and transparent model reasoning, this approach demonstrates potential to support individualized treatment decisions between surveillance and timely surgery. Prospective, large-scale studies will be essential to confirm clinical utility and enable translation into routine practice.

Acknowledgments: Figure 1 was created in BioRender. Zhang, Z (2025) https:// BioRender.com/ijbkjms.

Conflict of Interest: None

Financial Support: None

Ethics Statement: None

References

1. Bray F, Laversanne M, Sung H, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA A Cancer J Clinicians. 2024;74(1):229–63. doi:10.3322/caac.21834
2. Yang H, Liu H, Chen Y, et al. Neoadjuvant Chemoradiotherapy Followed by Surgery Versus Surgery Alone for Locally Advanced Squamous Cell Carcinoma of the Esophagus (NEOCRTEC5010): A Phase III Multicenter, Randomized, Open-Label Clinical Trial. J Clin Oncol. 2018;36(29):2796–803. doi:10.1200/JCO.2018.79.1483
3. Ajani JA, D'Amico TA, Bentrem DJ, et al. Esophageal and Esophagogastric Junction Cancers, Version 2.2023, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 2023;21(4):393–422. doi:10.6004/jnccn.2023.0019
4. Liu J, Yang Y, Liu Z, et al. Multicenter, single-arm, phase II trial of camrelizumab and chemotherapy as neoadjuvant treatment for locally advanced esophageal squamous cell carcinoma. J Immunother Cancer. 2022;10(11). doi:10.1136/jitc-2021-004291
5. Yan X, Duan H, Ni Y, et al. Tislelizumab combined with chemotherapy as neoadjuvant therapy for surgically resectable esophageal cancer: A prospective, single-arm, phase II study (TD-NICE). Int J Surg. 2022;103:106680. doi:10.1016/j.ijsu.2022.106680
6. Guo X, Chen C, Zhao J, et al. Neoadjuvant Chemoradiotherapy vs Chemoimmunotherapy for Esophageal Squamous Cell Carcinoma. JAMA Surg. 2025;160(6):565–74. doi:10.1001/jamasurg.2025.0220

7. Wang H, Jiang Z, Wang Q, et al. Pathological response and prognostic factors of neoadjuvant PD-1 blockade combined with chemotherapy in resectable oesophageal squamous cell carcinoma. Eur J Cancer. 2023;186:196–210. doi:10.1016/j.ejca.2023.03.008
8. Wang H, Tang H, Fang Y, et al. Morbidity and Mortality of Patients Who Underwent Minimally Invasive Esophagectomy After Neoadjuvant Chemoradiotherapy vs Neoadjuvant Chemotherapy for Locally Advanced Esophageal Squamous Cell Carcinoma. JAMA Surg. 2021;156(5):444. doi:10.1001/jamasurg.2021.0133
9. Luchini C, Bibeau F, Ligtenberg MJL, et al. ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. Ann Oncol. 2019;30(8):1232–43. doi:10.1093/annonc/mdz116
10. Raimondi A, Lonardi S, Murgioni S, et al. Tremelimumab and durvalumab as neoadjuvant or non-operative management strategy of patients with microsatellite instability-high resectable gastric or gastroesophageal junction adenocarcinoma: the INFINITY study by GONO. Ann Oncol. 2025;36(1):285–96. doi:10.1016/j.annonc.2024.11.016
11. Wang C, Ju C, Du D, et al. CircNF1 modulates the progression and immune evasion of esophageal squamous cell carcinoma through dual regulation of PD-L1. Cell Mol Biol Lett. 2025;30(1):1–11. doi:10.1186/s11658-025-00712-y
12. Yang F, Dan M, Shi J, et al. Efficacy and safety of PD-1 inhibitors as second-line treatment for advanced squamous esophageal cancer: a systematic review and network meta-analysis with a focus on PD-L1 expression levels. Front Immunol. 2025;15. doi:10.3389/fimmu.2024.1510145
13. Anagnostou V, Bardelli A, Chan TA, et al. The status of tumor mutational burden and immunotherapy. Nat Cancer. 2022;3(8):652–6. doi:10.1038/s43018-022-00382-1
14. Zhou KI, Peterson B, Serritella A, et al. Spatial and Temporal Heterogeneity of PD-L1 Expression and Tumor Mutational Burden in Gastroesophageal Adenocarcinoma at Baseline Diagnosis and after Chemotherapy. Clin Cancer Res. 2020;26(24):6453–63. doi:10.1158/1078-0432.CCR-20-2085
15. Niknafs N, Najjar M, Dennehy C, et al. Of Context, Quality, and Complexity: Fine-Combing Tumor Mutational Burden in Immunotherapy-Treated Cancers. Clin Cancer Res. 2025;31(8):2850–63. doi:10.1158/1078-0432.CCR-23-0824
16. Boehm KM, Khosravi P, Vanguri R, et al. Harnessing multimodal data integration to advance precision oncology. Nat Rev Cancer. 2022;22(2):114–26. doi:10.1038/s41568-021-00408-3
17. Lin H, Hua J, Gong Z, et al. Multimodal radiopathological integration for prognosis and prediction of adjuvant chemotherapy benefit in resectable lung adenocarcinoma: A multicentre study. Cancer Lett. 2025;616:217557. doi:10.1016/j.canlet.2025.217557
18. Qi Y-J, Su G-H, You C, et al. Radiomics in breast cancer: Current advances and future directions. Cell Rep Med. 2024;5(7):101719. doi:10.1016/j.xcrm.2024.101719
19. Li B, Qin W, Yang L, et al. From pixels to patient care: deep learning-enabled pathomics signature offers precise outcome predictions for immunotherapy in esophageal squamous cell cancer. J Transl Med. 2024;22(1):499. doi:10.1186/s12967-024-04997-z
20. Zhang Z, Luo T, Yan M, et al. Voxel-level radiomics and deep learning for predicting pathologic complete response in esophageal squamous cell carcinoma after neoadjuvant immunotherapy and chemotherapy. J Immunother Cancer. 2025;13(1). doi:10.1136/jitc-2024-011149
21. Kumar N, Verma R, Chen C, et al. Computer-extracted features of nuclear morphology in hematoxylin and eosin images distinguish stage II and IV colon tumors. J Pathol. 2022;257(1):17–28. doi:10.1002/path.5864
22. Wang Y, Pan X, Lin H, et al. Multi-scale pathology image texture signature is a prognostic factor for resectable lung adenocarcinoma: a multi-center, retrospective study. J Transl Med. 2022;20(1):595. doi:10.1186/s12967-022-03777-x
23. van der Velden BHM, Kuijf HJ, Gilhuijs KGA, et al. Explainable artificial intelligence (XAI) in deep learning-based medical image analysis. Med Image Anal. 2022;79:102470. doi:10.1016/j.media.2022.102470
24. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. Nat Rev Clin Oncol. 2017;14(12):749–62. doi:10.1038/nrclinonc.2017.141

25. Shi Z, Zhang Z, Liu Z, et al. Methodological quality of machine learning-based quantitative imaging analysis studies in esophageal cancer: a systematic review of clinical outcome prediction after concurrent chemoradiotherapy. Eur J Nucl Med Mol Imaging. 2022;49(8):2462–81. doi:10.1007/s00259-021-05658-9
26. Shi C, Jordan B. Protocol for the examination of specimens from patients with carcinoma of the esophagus. College of American Pathologists Cancer Protocols. 2017.
27. Fedorov A, Beichel R, Kalpathy-Cramer J, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. Magn Reson Imaging. 2012;30(10):1323–41. doi:10.1016/j.mri.2012.05.001
28. van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res. 2017;77(14 Suppl):e104–7. doi:10.1158/0008-5472.CAN-17-0339
29. Stirling DR, Carpenter AE, Cimini BA, et al. CellProfiler Analyst 3.0: accessible data exploration and machine learning for image analysis. Bioinformatics. 2021;37(23):3992–4. doi:10.1093/bioinformatics/btab634
30. Graham S, Vu QD, Raza SEA, et al. Hover-Net: Simultaneous segmentation and classification of nuclei in multi-tissue histology images. Med Image Anal. 2019;58:101563. doi:10.1016/j.media.2019.101563
31. van den Eyck BM, van Lanschot JJB, Hulshof MCCM, et al. Ten-Year Outcome of Neoadjuvant Chemoradiotherapy Plus Surgery for Esophageal Cancer: The Randomized Controlled CROSS Trial. JCO. 2021;39(17):1995–2004. doi:10.1200/JCO.20.03614
32. Hoeppner J, Brunner T, Schmoor C, et al. Perioperative Chemotherapy or Preoperative Chemoradiotherapy in Esophageal Cancer. N Engl J Med. 2025;392(4):323–35. doi:10.1056/NEJMoa2409408
33. Hoeppner J, Schmoor C, Brunner T, et al. Recurrence Patterns of Esophageal Adenocarcinoma in the Phase III ESOPEC Trial Comparing Perioperative Chemotherapy With Preoperative Chemoradiotherapy. J Clin Oncol. 2025;43(24):3451–6. doi:10.1200/JCO-25-00948
34. van der Wilk BJ, Eyck BM, Wijnhoven BPL, et al. Neoadjuvant chemoradiotherapy followed by active surveillance versus standard surgery for oesophageal cancer (SANO trial): a multicentre, stepped-wedge, cluster-randomised, non-inferiority, phase 3 trial. Lancet Oncol. 2025;26(5):425–36. doi:10.1016/S1470-2045(25)00027-0
35. Sun J-M, Shen L, Shah MA, et al. Pembrolizumab plus chemotherapy versus chemotherapy alone for first-line treatment of advanced oesophageal cancer (KEYNOTE-590): a randomised, placebo-controlled, phase 3 study. Lancet. 2021;398(10304):759–71. doi:10.1016/S0140-6736(21)01234-4
36. Janjigian YY, Shitara K, Moehler M, et al. First-line nivolumab plus chemotherapy versus chemotherapy alone for advanced gastric, gastro-oesophageal junction, and oesophageal adenocarcinoma (CheckMate 649): a randomised, open-label, phase 3 trial. The Lancet. 2021;397(10071):27–40. doi:10.1016/S0140-6736(21)00797-2
37. Kubo Y, Nozaki R, Igaue S, et al. Neoadjuvant Chemotherapy Improves Feasibility of Larynx Preservation and Prognosis in Resectable Locally Advanced Cervical Esophageal Cancer. Ann Surg Oncol. 2024;31(10):5083–91. doi:10.1245/s10434-024-15432-4
38. Song X-Y, Lin L, Yang Y, et al. Radiotherapy as an organ-preserving alternative to surgery in patients with locally advanced esophageal squamous cell carcinoma achieving major pathologic response after induction immunochemotherapy. Int J Cancer. 2025;157(7):1680–93. doi:10.1002/ijc.35515
39. Matsuda S, Kawakubo H, Irino T, et al. Role sharing between minimally invasive oesophagectomy and organ preservation approach for surgically resectable advanced oesophageal cancer. Jpn J Clin Oncol. 2022;52(2):108–13. doi:10.1093/jjco/hyab192
40. Mao N, Dai Y, Zhou H, et al. A multimodal and fully automated system for prediction of pathological complete response to neoadjuvant chemotherapy in breast cancer. Sci Adv. 2025;11(3). doi:10.1126/sciadv.adr1576
41. Qi Y, Hu Y, Lin C, et al. A preoperative predictive model based on multi-modal features to predict pathological complete response after neoadjuvant chemoimmunotherapy in esophageal cancer patients. Front Immunol. 2025;16. doi:10.3389/fimmu.2025.1530279
42. Mao Y, Hou X, Fu S, et al. Transcriptomic and machine learning analyses identify hub genes of metabolism and host immune response that are associated with the progression of breast capsular contracture. Genes Dis. 2024;11(4):101087. doi:10.1016/j.gendis.2023.101087

43. Li Y, Yu J, Li R, et al. New insights into the role of mitochondrial metabolic dysregulation and immune infiltration in septic cardiomyopathy by integrated bioinformatics analysis and experimental validation. Cell Mol Biol Lett. 2024;29(1):1–17. doi:10.1186/s11658-024-00536-2
44. Captier N, Lerousseau M, Orlhac F, et al. Integration of clinical, pathological, radiological, and transcriptomic data improves prediction for first-line immunotherapy outcome in metastatic non-small cell lung cancer. Nat Commun. 2025;16(1):1–14. doi:10.1038/s41467-025-55847-5
45. Conner S, Guarin JR, Le TT, et al. Cell morphology best predicts tumorigenicity and metastasis in vivo across multiple TNBC cell lines of different metastatic potential. bioRxiv. 2023. doi:10.1101/2023.06.14.544969
46. Zheng S, Wang W, Shen L, et al. Tumor battlefield within inflamed, excluded or desert immune phenotypes: the mechanisms and strategies. Exp Hematol Oncol. 2024;13(1):129. doi:10.1186/s40164-024-00543-1
47. Khosravi G-R, Mostafavi S, Bastan S, et al. Immunologic tumor microenvironment modulators for turning cold tumors hot. Cancer Commun (Lond). 2024;44(1):521–53. doi:10.1002/cac2.12539
48. Jin MZ, Jin WL. The updated landscape of tumor microenvironment and drug repurposing. Sig Transduct Target Ther. 2020;5(1):166. doi:10.1038/s41392-020-00280-x
49. Hwang HW, Jung H, Hyeon J, et al. A nomogram to predict pathologic complete response (pCR) and the value of tumor-infiltrating lymphocytes (TILs) for prediction of response to neoadjuvant chemotherapy (NAC) in breast cancer patients. Breast Cancer Res Treat. 2019;173(2):255–66. doi:10.1007/s10549-018-4981-x
50. Sang S, Sun Z, Zheng W, et al. TME-guided deep learning predicts chemotherapy and immunotherapy response in gastric cancer with attention-enhanced residual Swin Transformer. Cell Rep Med. 2025;6(3):102242. doi:10.1016/j.xcrm.2025.102242