

RESEARCH

Open Access



# Multi-cohort study in gastric cancer to develop CT-based radiomic models to predict pathological response to neoadjuvant immunotherapy

Ze-Ning Huang<sup>1,2†</sup>, Hao-Xiang Zhang<sup>1,2†</sup>, Yu-Qin Sun<sup>1,2,3†</sup>, Xing-Qi Zhang<sup>1,2</sup>, Yi-Fen Lin<sup>4</sup>, Cai-Ming Weng<sup>1,2</sup>, Chao-Hui Zheng<sup>1,2</sup>, Ping-Li<sup>1,2</sup>, Jia-Bin Wang<sup>1,2</sup>, Qi-Yue Chen<sup>1,2</sup>, Long-Long Cao<sup>1,2</sup>, Mi Lin<sup>1,2</sup>, Ru-Hong Tu<sup>1,2</sup>, Chang-Ming Huang<sup>1,2\*</sup>, Jian-Xian Lin<sup>1,2\*</sup> and Jian-Wei Xie<sup>1,2\*</sup> 

## Abstract

**Background** Neoadjuvant immunotherapy has been shown to improve survival in patients with gastric cancer. This study sought to develop and validate a radiomics-based machine learning (ML) model for patients with locally advanced gastric cancer (LAGC), specifically to predict whether patients will achieve a major pathological response (MPR) following neoadjuvant immunotherapy. With its predictive capabilities, this tool shows promise for enhancing clinical decision-making processes in the future.

**Methods** This study utilized a multicenter cohort design, retrospectively gathering clinical data and computed tomography (CT) images from 268 patients diagnosed with advanced gastric cancer who underwent neoadjuvant immunotherapy between January 2019 and December 2023 from two medical centers. Radiomic features were extracted from CT images, and a multi-step feature selection procedure was applied to identify the top 20 representative features. Nine ML algorithms were implemented to build prediction models, with the optimal algorithm selected for the final prediction model. The hyperparameters of the chosen model were fine-tuned using Bayesian optimization and grid search. The performance of the model was evaluated using several metrics, including the area under the curve (AUC), accuracy, and Cohen's kappa coefficient.

**Results** Three cohorts were included in this study: the development cohort (DC,  $n = 86$ ), the internal validation cohort (IVC,  $n = 59$ ), and the external validation cohort (EVC,  $n = 52$ ). Nine ML models were developed using DC

<sup>†</sup>Ze-Ning Huang, Hao-Xiang Zhang and Yu-Qin Sun contributed equally to this work and should be considered shared first authors.

\*Correspondence:  
Chang-Ming Huang  
hcmlr2002@163.com  
Jian-Xian Lin  
linjian379@163.com  
Jian-Wei Xie  
xjwhw2019@163.com

Full list of author information is available at the end of the article



© The Author(s) 2025, corrected publication 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

cases. Among these, an optimized Bayesian-LightGBM model, demonstrated robust predictive performance for MPR following neoadjuvant immunotherapy in LAGC patients across all cohorts. Specifically, within DC, the LightGBM model attained an AUC of 0.828, an overall accuracy of 0.791, a Cohen's kappa coefficient of 0.552, a sensitivity of 0.742, a specificity of 0.818, a positive predictive value (PPV) of 0.586, a negative predictive value (NPV) of 0.867, a Matthews correlation coefficient (MCC) of 0.473, and a balanced accuracy of 0.780. Comparable performance metrics were validated in both the IVC and the EVC, with AUC values of 0.777 and 0.714, and overall accuracies of 0.729 and 0.654, respectively. These results suggested good fitness and generalization of the Bayesian-LightGBM model. Shapley Additive Explanations (SHAP) analysis identified significant radiomic features contributing to the model's predictive capability. The SHAP values of the features `wavelet.LLH_gldm_SmallDependenceLowGrayLevelEmphasis`, `wavelet.HHL_grlm_RunVariance`, and `wavelet.LLH_glszm_LargeAreaHighGrayLevelEmphasis` were ranked among the top three, highlighting their significant contribution to the model's predictive performance. In contrast to existing radiomic models that exclusively focus on neoadjuvant chemotherapy, our model integrates both neoadjuvant immunotherapy and chemotherapy, thereby offering more precise predictive capabilities.

**Conclusion** The radiomics-based ML model demonstrated significant efficacy in predicting the pathological response to neoadjuvant immunotherapy in LAGC patients, thereby providing a foundation for personalized treatment strategies.

**Keywords** Radiomics, Machine learning, Neoadjuvant therapy, Immunotherapy, Prediction model

## Introduction

Gastric cancer (GC) is the fifth most common malignancy and the third leading cause of cancer-related death worldwide [1]. Surgical resection is the mainstay of treatment for gastric cancer; nevertheless, over 30% of locally advanced gastric cancer (LAGC) patients experience recurrence despite achieving complete resection and receiving adjuvant therapy [2, 3]. Therefore, neoadjuvant chemotherapy has been gradually applied in the treatment of gastric cancer and has achieved good clinical efficacy in recent years [4–6].

Immunotherapy, particularly immune checkpoint inhibitor (ICI) therapy, has markedly transformed cancer care for various malignancies [7, 8]. Currently, the application and efficacy of neoadjuvant immunotherapy combined with chemotherapy are becoming the focus of clinical studies [9, 10]. The efficacy of neoadjuvant immunotherapy is commonly assessed using the measure of pathological complete response (pCR), major pathological response (MPR), and tumor regression grading (TRG) [11–13]. Biomarkers such as tumor mutational burden (TMB) and PD-L1 expression levels have shown promising potential in predicting clinical responses [14, 15]. However, their clinical utility is limited by the complexity of their detection methods compared to blood tests and imaging techniques, as well as the spatial and temporal heterogeneity within tumors [16]. Currently, histopathological examination is still recognized as the clinical gold standard for assessing the pathological response to neoadjuvant treatment; however, it can only be performed after surgery, which may delay timely adjustments to treatment plans. Therefore, accurately predicting patient responses to neoadjuvant therapy combining PD-1

inhibitors with chemotherapy preoperatively is crucial for optimizing treatment strategies.

Computed tomography (CT) scanning is commonly used preoperatively to assess gastric cancer, though its diagnostic accuracy can vary due to differences in physician experience [17, 18]. Recently, radiomics, which generates precise image quantification information from CT images, has provided more objective data support and has been extensively applied in tumor diagnosis, treatment evaluation, and prognosis prediction [19, 20]. For instance, Gao et al. [21] analyzed contrast-enhanced CT images of gastric cancer patients, extracted and selected radiomic features, and constructed a prediction model that successfully forecasted lymph node metastasis in early-stage gastric cancer. Similarly, Sun et al. [22] utilized CT radiomics and subjective CT signs to significantly improve the accuracy and specificity of gastric cancer diagnosis. While some studies have explored radiomics-based models for predicting the efficacy of neoadjuvant chemotherapy in GC patients [23, 24], models predicting the response to neoadjuvant immunotherapy have not been verified. Neoadjuvant immunotherapy and neoadjuvant chemotherapy exhibit significant mechanistic differences, which may render the direct application of previous research findings to patients undergoing neoadjuvant immunotherapy inappropriate. To address this gap, our study constructs and validates a radiomics-based predictive model for assessing treatment efficacy in patients with gastric cancer receiving neoadjuvant immunotherapy. Moreover, these studies often rely on single-time-point radiomic features, which fail to capture the high heterogeneity of gastric cancer and do not reflect tumor progression or regression in real-time during neoadjuvant therapy. Additionally, machine learning (ML),

due to its ability to handle large datasets, discover complex patterns, and make efficient predictions, has increasingly been applied to improve the accuracy and efficiency of diagnosis, treatment, and prognosis in oncology [25, 26]. However, traditional ML models are often difficult to interpret and are considered black boxes, which poses challenges for clinical application. Shapley Additive Explanations (SHAP) is a technique that employs Shapley values to elucidate the predictions generated by machine learning models, thereby enhancing the comprehension of each feature's contribution to the model's outputs [27, 28]. In essence, a higher SHAP value signifies a more substantial influence of the corresponding feature on the prediction outcome. SHAP was used to explain the prediction model from a global perspective, addressing the shortcomings of ML models and enhancing their transparency and practicality.

In summary, this study aimed to develop and validate a ML model based on changes in radiomic features before and after neoadjuvant immunotherapy to predict the response of LAGC patients to neoadjuvant immunotherapy. Our research is predicated on the hypothesis that utilizing radiomic features obtained from pre- and post-treatment imaging, combined with machine learning techniques, can facilitate the development of an efficient and interpretable predictive model to inform neoadjuvant immunotherapy strategies for patients with locally advanced gastric cancer. This methodology not only holds the potential to enhance the precision of treatment response predictions but also aids clinicians in formulating more personalized treatment plans.

## Methods

### Study design and participants

This retrospective multicenter cohort study collected data from 2019 to 2023 at Fujian Medical University Union Hospital (FMUUh) and Zhangzhou Municipal Hospital (ZZMH). The study included 268 LAGC patients who had undergone neoadjuvant immunotherapy followed by radical resection surgery. Among these, 195 cases were from FMUUh and 62 from ZZMH. All enrolled patients met the following criteria: (1) histologically confirmed gastric adenocarcinoma; (2) cT3/T4N0/+M0 and underwent neoadjuvant immunotherapy prior to surgery; (3) no history of other malignancies or distant metastases; (4) R0 resection with complete clinical and follow-up records. The exclusion criteria were: (1) other prior treatments, such as radiotherapy, targeted therapy, or chemotherapy alone; (2) remnant gastric cancer; (3) missing baseline CT (CT1) before neoadjuvant therapy or preoperative CT (CT2). Ultimately, 145 patients from FMUUh were allocated to the training set, which was further divided into a development cohort (DC,  $n=86$ ) and an internal validation cohort (IVC,

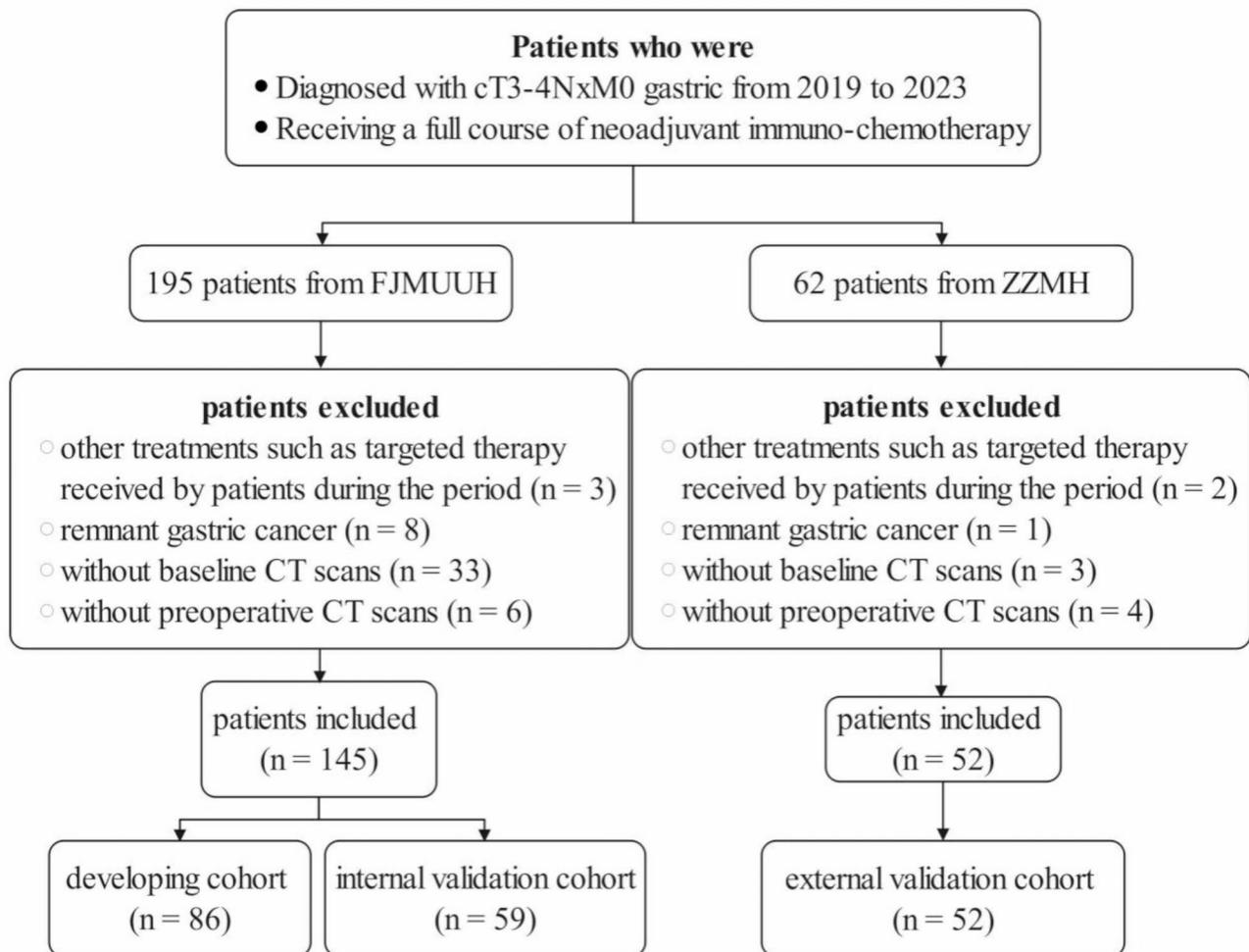
$n=59$ ) at a ratio of 6:4. Additionally, 52 patients from ZZMH were included in the external validation cohort (EVC). A patient selection flowchart is shown in Fig. 1. The study protocol was approved by the ethics committees of both hospitals and adhered to the principles of the Declaration of Helsinki.

### Definitions

The primary outcome measure was MPR, defined as the presence of <10% viable cancer cells in the primary tumor. Follow-up included physical examinations, laboratory tests (including carcinoembryonic antigen and carbohydrate antigen 19–9), and imaging studies (including chest X-ray, chest CT, abdominal ultrasound, or abdominal/pelvic CT). An annual endoscopic examination was also recommended.

### Imaging data acquisition and processing

All participants underwent abdominal contrast-enhanced CT scans at baseline (CT1) before neoadjuvant immunotherapy and within two weeks prior to surgery (preoperative CT, CT2). CT images were sourced from portal venous phase axial slices and retrieved from the Picture Archiving and Communication System. In this study, we placed a substantial emphasis on safeguarding patient privacy and anonymizing imaging data. To this end, we implemented de-identification procedures for each patient's imaging data and assigned a unique anonymous identifier to each dataset. All data were managed using secure storage and access control mechanisms to ensure both confidentiality and integrity. During data processing, we employed medical image processing software, such as 3DSlicer and Pyradiomics, to automatically remove sensitive information and conducted feature extraction and model training within a local environment. This approach minimized the risk of data transmission beyond the organizational boundaries. Furthermore, all team members signed data use agreements, committing to utilizing the data exclusively for the specified research purposes and strictly adhering to privacy protection regulations. Any breaches of the agreement would result in serious consequences. Through these measures, we have not only effectively protected patient privacy but also established a reliable data foundation for high-quality scientific research. To ensure consistency in scale during the subsequent feature extraction process, all CT images were denoised to mitigate noise interference and resampled to standardize the voxel size to  $1 \times 1 \times 1 \text{ mm}^3$ . Regions of interest (ROIs) for tumor areas were manually segmented using the maximum cross-sectional images from CT scans with 3D Slicer software (version 4.10.2, [<http://www.slicer.org>]). Semi-automatic segmentation tools, including threshold segmentation and active contour models, were also employed. This



**Fig. 1** The enrollment of patients in the FJMMU cohort and ZZMH cohort

process was conducted by a radiologist (L.Y.F.) with seven years of experience in abdominal CT imaging. The radiologist was aware of the gastric cancer diagnosis but not of other clinical details. One month later, images from 50 randomly selected patients were re-segmented by the same radiologist and another surgeon (H.Z.N.) with ten years of experience in abdominal CT imaging to evaluate inter- and intra-observer reproducibility. Radiomic features were extracted from ROIs using the open-source Pyradiomics package (version 2.12, [<https://pyradiomics.readthedocs.io/en/2.12/>]). Pyradiomics, a widely recognized open-source radiomics software tool, adheres to the guidelines set forth by the Image Biomarker Standardization Initiative (IBSI), thereby ensuring consistency and comparability in feature extraction across different studies. Pyradiomics supports a diverse array of feature extraction types, including first-order statistical features, texture features, shape features, and features derived from wavelet transformation. The diverse types of features effectively capture the heterogeneity of

tumors. Moreover, the Pyradiomics package provides flexible configuration options, enabling researchers to adjust various parameters during the feature extraction process to meet their specific requirements. A total of 832 radiomic features were extracted, comprising 18 first-order statistical features, 86 texture features, and 728 wavelet decomposition features. Throughout the feature extraction process, we adhered to the guidelines set forth by the Image Biomarker Standardisation Initiative to ensure consistency and standardization.

#### Feature standardization and selection

To assess the reliability of the extracted features, we calculated the intra- and inter-class correlation coefficients (ICCs). ICCs < 0.5 indicate poor reliability, 0.5 to 0.75 indicate moderate reliability, 0.75 to 0.9 indicate good reliability, and > 0.9 indicates excellent reliability. We randomly selected the radiomic features of 50 patients for verification. If a feature's ICC\_intra and ICC\_inter were both greater than 0.9, it was considered to have passed

the reproducibility and inter-observer consistency tests. For the features that passed these tests, deltaCT features were obtained by calculating the difference between CT2 and CT1. Furthermore, Z-score normalization was applied to the CT2 and deltaCT features to eliminate scale differences and ensure a fair comparison. The minimum Redundancy Maximum Relevance (mRMR) method was used to select the top 30 most relevant features while minimizing redundancy. The least absolute shrinkage and selection operator (LASSO) regression, combined with ten-fold cross-validation, was then applied to identify the most representative features for model construction.

### Machine learning model construction, evaluation, and validation

In selecting machine learning models to assess their capability to predict major pathological response in LAGC patients undergoing neoadjuvant immunotherapy, we considered factors such as diversity and representativeness, historical popularity and performance, complementarity, comparative analysis, and practical applicability. Consequently, we selected a range of ML models, including LightGBM, Decision Tree (DT), XGBoost, Multi-Layer Perceptron (MLP), K-Nearest Neighbors (KNN), Random Forest (RF), Elastic Net (Enet), Support Vector Machine (SVM), and Logistic Regression. Utilizing the prediction endpoint of major pathological response as a binary outcome (yes/no) alongside radiomic features as predictors, we developed predictive models employing various machine learning algorithms. Bayesian hyperparameter optimization and grid search were used to find the optimal parameters of the model. Bayesian optimization is selected for tuning the hyperparameters of machine learning models due to its efficiency, flexibility, support for mixed-type parameters, and robust black-box optimization capabilities [29]. By integrating these methodologies, our objective is to more accurately identify the optimal parameter configurations, thereby enhancing the model's performance. Training and parameter optimization were performed on the DC model, with performance re-evaluation conducted on the IVC dataset. The performance metrics included the area under curve (AUC), accuracy, Cohen's kappa coefficient, sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), Matthews correlation coefficient (MCC), and balanced accuracy. We selected the best-performing model as our final model. External validation was conducted using an independent dataset from ZZMH.

### SHAP value analysis

We calculated the SHAP values for each radiomic feature in the final model, ranking all the features based on these values. SHAP values quantify the contribution

of each feature to individual predictions. This process identified the most critical features for predicting MPR in LAGC patients undergoing neoadjuvant immunotherapy. A comprehensive understanding of the relationship between features and MPR was achieved through the analysis of mean SHAP values, as well as SHAP summary and dependence plots. The mean SHAP value analysis provides the average contribution of each feature across samples. SHAP summary plots visually display the importance and effect of features on the model's predictions across all samples. SHAP dependence plots illustrate the relationship between a single feature's SHAP value and its actual predictive value, with points representing the SHAP value of that feature for a given sample. These analyses helped identify the most relevant radiomic features and provided insight into how they influence MPR.

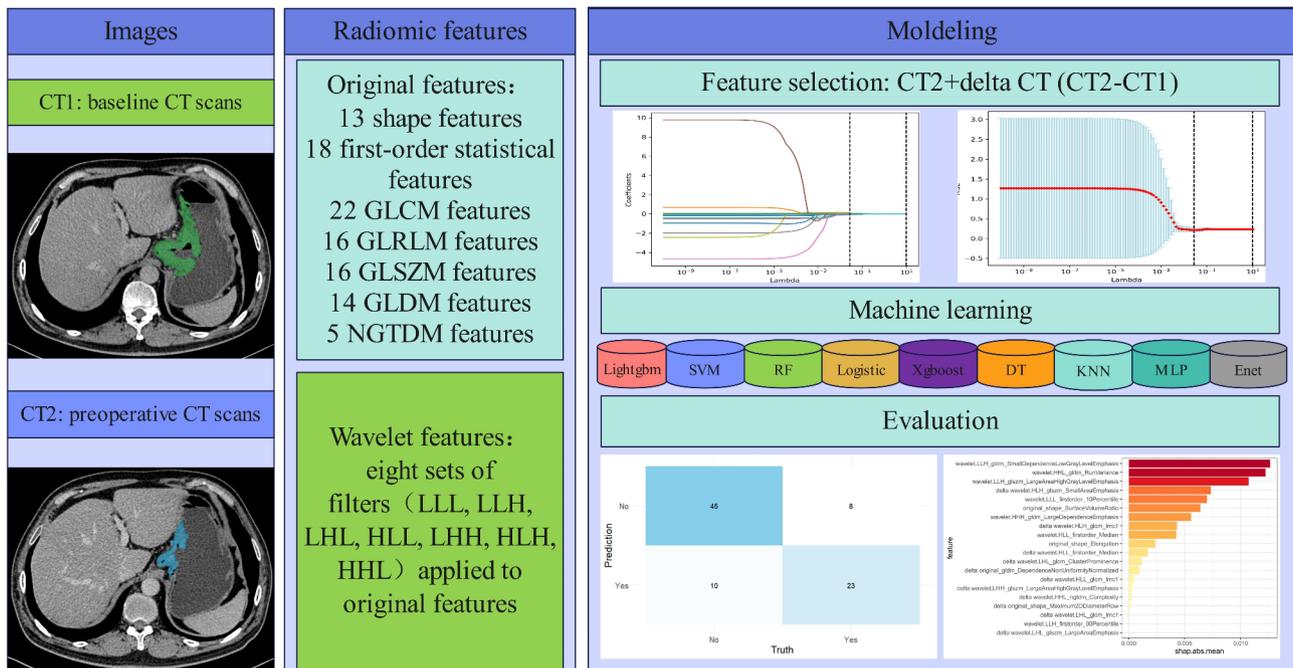
### Statistical methods

The overview of the study flow is presented in Fig. 2. Data analysis was conducted using R version 4.3.1 (<http://r-project.org>). The mRMR method and LASSO regression were used to select the most effective radiomic features. Machine learning models were constructed using the 'tidymodels' package in R, and DeLong's test was utilized to compare AUC values between DC and IVC. All statistical tests were two-sided, with p-values < 0.05 considered statistically significant.

## Results

### Baseline patient characteristics

A total of three cohorts were included in this study: a DC ( $n=86$ ), an IVC ( $n=59$ ), and an EVC ( $n=52$ ) (Table 1). To elucidate the distinctions among the cohorts, we employed circular plots to present the baseline characteristics of the three cohorts (Figure S1). The mean ages of the patients in DC, IVC, and EVC were 62.5 years, 65.0 years, and 65.0 years, respectively, with no statistically significant difference in age ( $P=0.2097$ ). The proportion of male patients in the EVC group was significantly lower (63.5%) compared to the DC (82.6%) and IVC (79.7%) groups ( $P=0.0298$ ). The proportion of tumors located in the middle third of the stomach was significantly higher in EVC (40.4%) compared to DC (12.8%) and IVC (18.6%) ( $P=0.0023$ ). In terms of the number of cycles of neoadjuvant immunotherapy received, a higher proportion of patients in the EVC group underwent 3–4 cycles of treatment (92.3%), compared to 79.1% in the DC group and 78.0% in the IVC group ( $P=0.0242$ ). Additionally, the proportion of patients at the cT3 stage in the EVC group was notably higher (23.1% versus 7.0% in the DC group and 6.8% in the IVC group,  $P=0.0064$ ). No significant differences were observed among the cohorts regarding BMI, cN, ypN, ypN, TRG classification, or the proportion



**Fig. 2** Workflow of the study. Workflow of machine learning modeling for MPR prediction in patients with locally advanced gastric cancer. CT, computed tomography; GLCM, gray level co-occurrence matrix; GLRLM, gray level run length matrix; GLSZM, gray level size zone matrix; GLDM, gray level dependence matrix; NGTDM, neighboring gray-tone difference matrix; SVM, support vector machine; RF, random forest; DT, decision tree; KNN, K nearest neighbor; MLP, multilayer perceptron; Enet, elastic net

of patients receiving adjuvant chemotherapy (all  $P > 0.05$ ). These differences suggest that the EVC group encompasses a greater number of relatively early-stage lesions, varied gender ratios, and diverse treatment protocols among its patient population. These results indicate significant differences in certain baseline characteristics among the cohorts, underscoring the importance of validating the model across diverse patient populations.

**Feature selection and model Building**

A total of 832 features were extracted from the ROIs in the baseline CT (CT1) and preoperative CT (CT2) scans. Among these, 606 features had ICCs greater than 0.9 for both intra- and inter-observer reliability (Figure S2). By calculating the differences between CT2 and CT1, we derived an additional 606 delta CT features that reflect tumor changes over time. Combining deltaCT features with CT2 features resulted in a total of 1,212 imaging features for analysis. Dimension reduction was subsequently applied. The mRMR algorithm was used to further simplify the radiomic features, retaining 30 features, which included 11 CT2 features and 19 deltaCT features for subsequent analysis. Finally, LASSO regression with cross-validation was employed to select the 20 most valuable radiomic features, which comprised 9 CT2 features and 11 deltaCT features (Figure S3 and Table S1). These 20 features were used to construct the radiomics model.

**Model performance**

The MPR was used as the predictive endpoint in this study. Nine ML models were constructed using 20 radiomic features as predictors: LightGBM, DT, XGBoost, MLP, KNN, RF, Enet, SVM, and Logistic Regression. These models were optimized using either Bayesian hyperparameter optimization or grid search. Table 2 and Figure S4 demonstrate the performance of various machine learning models under different hyperparameter optimization settings. Figure S5 shows P values derived from the DeLong test comparing AUC values between different prediction models in the development cohort. In DC, the Bayesian-optimized LightGBM model performed exceptionally well, achieving an AUC of 0.828. In the IVC, it achieved an AUC of 0.777 (DeLong test  $p = 0.518$ , which is greater than 0.05), indicating no signs of overfitting. The model demonstrated good predictive performance with an overall accuracy of 0.791, a Cohen’s kappa value of 0.552, sensitivity of 0.742, specificity of 0.818, a PPV of 0.586, an NPV of 0.867, an MCC of 0.473, and a balanced accuracy of 0.780 (Table 2). Confusion matrices displayed good consistency between predicted and actual outcomes in both DC and IVC (Fig. 3). Other models, even when optimized with Bayesian methods, performed worse in IVC (Table 2). Given its superior comprehensive performance, the Bayesian-optimized LightGBM model was selected as the optimal prediction model.

**Table 1** Baseline, surgical, and pathological information of the patients

	Developing cohort (n = 86)	Internal validation cohort (n = 59)	External validation cohort (n = 52)	P value
Age (median [IQR])	62.50 [56.0, 68.0]	65.00 [58.5, 68.5]	65.00 [59.0, 69.0]	0.2097
Sex (%)				0.0298
Male	71 (82.6)	47 (79.7)	33 (63.5)	
Female	15 (17.4)	12 (20.3)	19 (36.5)	
BMI (median [IQR])	21.56 [19.8, 23.6]	22.20 [20.8, 23.5]	22.10 [19.6, 23.2]	0.4661
Site (%)				0.0023
Upper	37 (43.0)	28 (47.5)	21 (40.4)	
Middle	11 (12.8)	11 (18.6)	21 (40.4)	
Lower	22 (25.6)	11 (18.6)	9 (17.3)	
Overlapping	16 (18.6)	9 (15.3)	1 (1.9)	
Cycle of NICT (%)				0.0242
1–2	10 (11.6)	2 (3.4)	2 (3.8)	
3–4	68 (79.1)	46 (78.0)	48 (92.3)	
≥ 5	8 (9.3)	11 (18.6)	2 (3.8)	
cT category (%)				0.0064
T3	6 (7.0)	4 (6.8)	12 (23.1)	
T4	80 (93.0)	55 (93.2)	40 (76.9)	
cN category (%)				0.9588
N0	4 (4.6)	3 (5.1)	3 (5.8)	
N+	82 (95.4)	56 (94.9)	49 (94.2)	
ypT category (%)				0.9760
T0	14 (16.3)	9 (15.3)	8 (15.4)	
T1-2	16 (18.6)	13 (22.0)	12 (23.1)	
T3-4	56 (65.1)	37 (62.7)	32 (61.5)	
ypN category (%)				0.4685
N0	41 (47.7)	29 (49.2)	20 (38.5)	
N+	45 (52.3)	30 (50.8)	32 (61.5)	
Grade (%)				0.0044
Well-differentiated	0 (0.0)	0 (0.0)	2 (3.8)	
Moderately differentiated	26 (30.2)	18 (30.5)	16 (30.8)	
Poorly differentiated	42 (48.8)	28 (47.5)	34 (65.4)	
Unknown	18 (20.9)	13 (22.0)	0 (0.0)	
Lymphovascular invasion (%)				0.0367
Absent	54 (62.8)	33 (55.9)	21 (40.4)	
Present	32 (37.2)	26 (44.1)	31 (59.6)	
Perineural invasion (%)				0.1420
Absent	48 (55.8)	29 (49.2)	20 (38.5)	
Present	38 (44.2)	30 (50.8)	32 (61.5)	
TRG (%)				0.0945
1a	14 (16.3)	9 (15.3)	8 (15.4)	
1b	17 (19.8)	12 (20.3)	11 (21.2)	
2	26 (30.2)	13 (22.0)	24 (46.2)	
3	29 (33.7)	25 (42.4)	9 (17.3)	
Adjuvant chemotherapy (%)				0.8529
No	8 (9.3)	5 (8.5)	6 (11.5)	
Yes	78 (90.7)	54 (91.5)	46 (88.5)	

Abbreviations: IQR, interquartile range; BMI: body mass index; NICT: neoadjuvant immuno-chemotherapy; TRG: tumor regression grade

### External validation

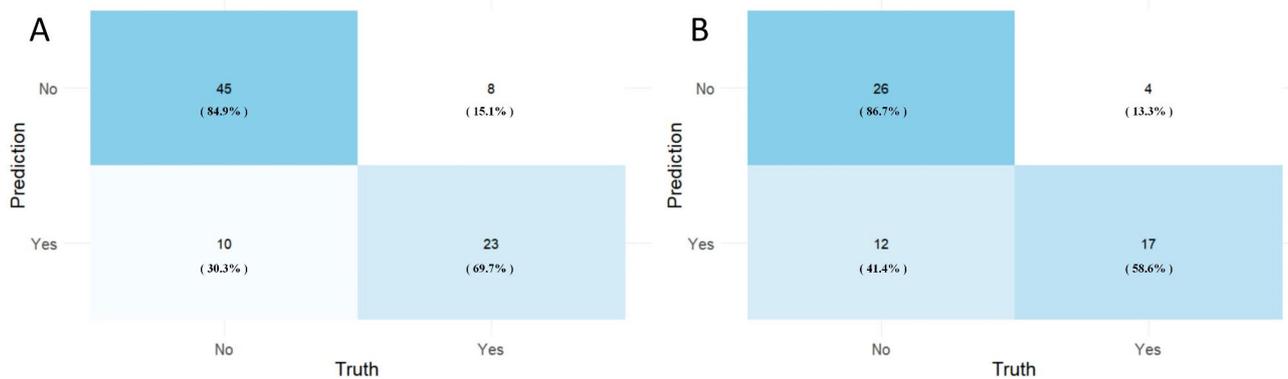
In EVC, the LightGBM model demonstrated stable predictive capability. Confusion matrices indicated good consistency between predicted and actual outcomes

in EVC (Fig. 4A). The AUC was 0.714 (Fig. 4B), with an overall accuracy of 0.654, a Cohen's kappa value of 0.300, sensitivity of 0.684, specificity of 0.636, a PPV of 0.520, an NPV of 0.778, an MCC of 0.309, and a balanced

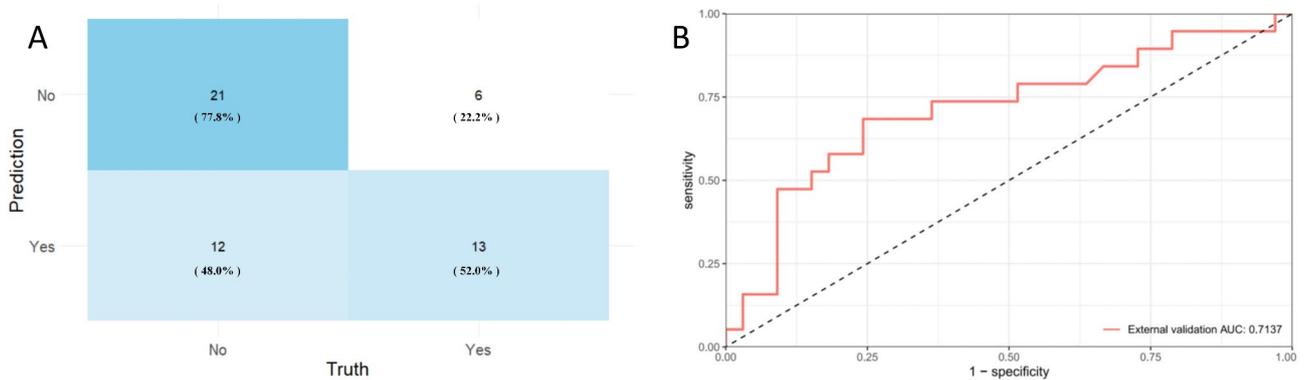
**Table 2** The radiomics machine learning model optimization, model performance in the developing cohort(DC) and the internal validation cohort(IVC)

Model	HPO	Data	AUC	p value	Accuracy	Kappa	Sensitivity	Specificity	PPV	NPV	MCC	Bal_accuracy
lightGBM	Bayes	DC	0.828		0.791	0.552	0.742	0.818	0.697	0.849	0.553	0.780
lightGBM	Bayes	IVC	0.777	0.518	0.729	0.455	0.810	0.684	0.586	0.867	0.473	0.747
lightGBM	Grid	DC	0.982		0.953	0.901	0.968	0.945	0.909	0.981	0.902	0.957
lightGBM	Grid	IVC	0.749	0.001	0.644	0.278	0.667	0.632	0.500	0.774	0.286	0.649
DT	Bayes	DC	0.903		0.884	0.748	0.839	0.909	0.839	0.909	0.748	0.874
DT	Bayes	IVC	0.686	0.014	0.712	0.378	0.619	0.763	0.591	0.784	0.378	0.691
DT	Grid	DC	0.841		0.849	0.660	0.710	0.927	0.846	0.850	0.666	0.818
DT	Grid	IVC	0.733	0.186	0.695	0.305	0.476	0.816	0.588	0.738	0.309	0.646
Xgboost	Bayes	DC	0.812		0.721	0.449	0.839	0.655	0.578	0.878	0.474	0.747
Xgboost	Bayes	IVC	0.768	0.590	0.593	0.256	0.857	0.447	0.462	0.850	0.308	0.652
Xgboost	Grid	DC	0.833		0.837	0.608	0.548	1.000	1.000	0.797	0.661	0.774
Xgboost	Grid	IVC	0.697	0.119	0.695	0.273	0.381	0.868	0.615	0.717	0.288	0.625
MLP	Bayes	DC	0.820		0.733	0.475	0.871	0.655	0.587	0.900	0.506	0.763
MLP	Bayes	IVC	0.719	0.226	0.610	0.267	0.810	0.500	0.472	0.826	0.304	0.655
MLP	Grid	DC	0.822		0.756	0.515	0.871	0.691	0.614	0.905	0.540	0.781
MLP	Grid	IVC	0.722	0.229	0.593	0.228	0.762	0.500	0.457	0.792	0.255	0.631
KNN	Bayes	DC	0.870		0.767	0.541	0.903	0.691	0.622	0.927	0.571	0.797
KNN	Bayes	IVC	0.717	0.054	0.695	0.375	0.714	0.684	0.556	0.813	0.383	0.699
KNN	Grid	DC	0.885		0.814	0.596	0.742	0.855	0.742	0.855	0.596	0.798
KNN	Grid	IVC	0.696	0.016	0.695	0.335	0.571	0.763	0.571	0.763	0.335	0.667
RF	Bayes	DC	0.916		0.872	0.725	0.839	0.891	0.812	0.907	0.725	0.865
RF	Bayes	IVC	0.741	0.020	0.712	0.378	0.619	0.763	0.591	0.784	0.378	0.691
RF	Grid	DC	0.921		0.872	0.712	0.742	0.945	0.885	0.867	0.719	0.844
RF	Grid	IVC	0.741	0.015	0.712	0.351	0.524	0.816	0.611	0.756	0.353	0.670
Enet	Bayes	DC	0.917		0.884	0.751	0.871	0.891	0.818	0.925	0.752	0.881
Enet	Bayes	IVC	0.708	0.006	0.678	0.290	0.524	0.763	0.550	0.744	0.290	0.643
Enet	Grid	DC	0.914		0.884	0.748	0.839	0.909	0.839	0.909	0.748	0.874
Enet	Grid	IVC	0.708	0.007	0.712	0.351	0.524	0.816	0.611	0.756	0.353	0.670
SVM	Bayes	DC	0.920		0.884	0.755	0.903	0.873	0.800	0.941	0.758	0.888
SVM	Bayes	IVC	0.708	0.007	0.661	0.261	0.524	0.737	0.524	0.737	0.261	0.630
SVM	Grid	DC	0.925		0.907	0.798	0.871	0.927	0.871	0.927	0.798	0.899
SVM	Grid	IVC	0.707	0.006	0.695	0.335	0.571	0.763	0.571	0.763	0.335	0.667
Logistic	NA	DC	0.937		0.942	0.869	0.839	1.000	1.000	0.917	0.877	0.919
Logistic	NA	IVC	0.714	0.004	0.661	0.210	0.381	0.816	0.533	0.705	0.216	0.598

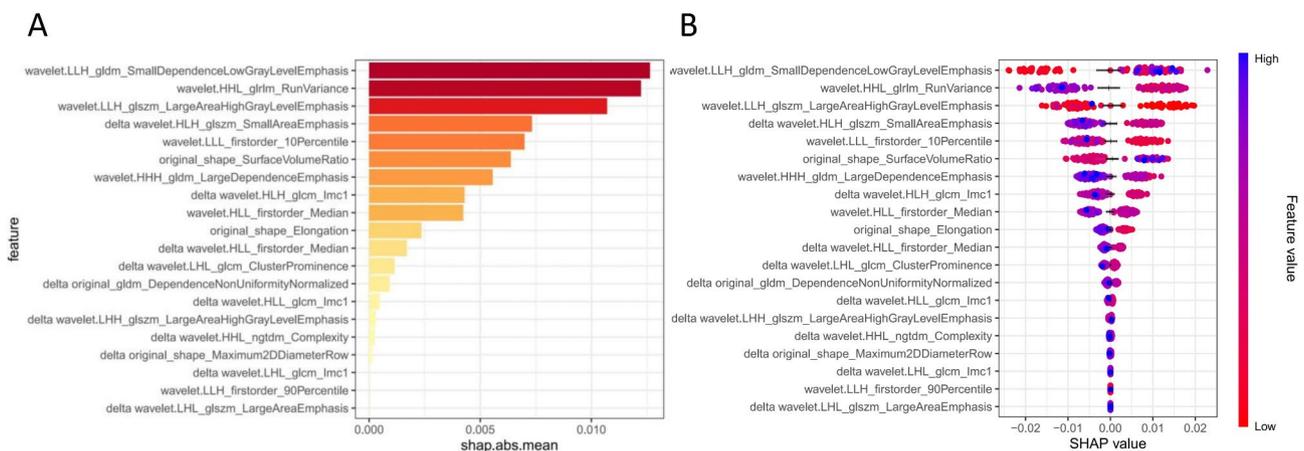
Abbreviations: SVM, support vector machine; RF, random forest; DT, decision tree; KNN, K nearest neighbor; MLP, multilayer perceptron; Enet, elastic net; HPO, hyper-parameter optimization; AUC, area under curve; PPV, positive predictive value; NPV, negative predictive value; MCC, Matthews correlation coefficient; Bal\_accuracy, balanced accuracy



**Fig. 3** Performance evaluation of the optimal machine learning model. The confusion matrix of lightGBM in the (A) developing cohort and (B) internal validation cohort



**Fig. 4** Performance evaluation of the lightGBM model in the external validation cohort (EVC). (A) The confusion matrix in the EVC. (B) Area under the curve of the lightGBM model in the EVC



**Fig. 5** Feature importance SHAP summary chart and bar chart. (A) The bars represent the importance of the variables and their overall contribution to the model predictions. (B) The dot plot represents the direction of contribution of each value of each variable, with blue representing larger values and red representing lower values of each variable

accuracy of 0.660 (Table S2), confirming the model's good generalizability.

**SHAP value analysis**

Mean SHAP value analysis revealed the importance of each feature in contributing to the model's

predictions (Fig. 5A). Larger mean SHAP values indicate a greater impact on the model's output. SHAP summary plots further revealed the impact direction of different feature values on the model's predictions (Fig. 5B). When SHAP values exceed zero, the model's prediction tends toward the 'MPR' category. The nine

features with the highest SHAP values were: wavelet.LLH\_gldm\_SmallDependenceLowGrayLevelEmphasis; wavelet.HHL\_glrmlm\_RunVariance; wavelet.LLH\_glszm\_LargeAreaHighGrayLevelEmphasis; deltawavelet.HLH\_glszm\_SmallAreaEmphasis; wavelet.LLL\_firstorder\_10Percentile; original\_shape\_SurfaceVolumeRatio; wavelet.HHH\_gldm\_LargeDependenceEmphasis; deltawavelet.LHL\_glcm\_Imc1; wavelet.HLL\_firstorder\_Median. Notably, deltawavelet.HLH\_glszm\_SmallAreaEmphasis and deltawavelet.LHL\_glcm\_Imc1 were derived from deltaCT features, indicating the importance of dynamic data in the model's predictions. SHAP dependence plots demonstrated how individual features affected model output, with each point representing a patient sample. The x-axis represents the variable size, while the y-axis indicates the SHAP value (Figure S6). Through the analysis of SHAP dependence plots, we identified that the values of specific key features are significantly associated with the probability of a MPR. For example, elevated values of the feature wavelet.LLH\_gldm\_SmallDependenceLowGrayLevelEmphasis are generally associated with more positive SHAP values, indicating an increased probability of achieving MPR in patients. In contrast, higher values of the wavelet HHL\_glrmlm\_RunVariance feature tend to correspond with more negative SHAP values, suggesting a reduced likelihood of achieving MPR. These results suggest that by quantifying dynamic changes in textural radiological attributes between baseline and post-treatment CT scans (i.e., deltaCT features), we can assess tumor response to therapy more precisely. These features played a crucial role in the model's predictions, further highlighting the significance and advancement of dynamic data within the model. Through SHAP value analysis, we can intuitively see how specific features influence the model's predictions, aiding clinicians in understanding and interpreting the logic behind these predictions, thus enhancing the credibility and practicality of the model in clinical applications.

## Discussion

In this study, we successfully developed and validated a ML model based on the radiomic features aimed at predicting the efficacy of neoadjuvant immunotherapy in LAGC patients. The results demonstrate that the constructed LightGBM model exhibited significant predictive performance in DC, IVC, and an independent EVC. The effectiveness of this model supports personalized treatment planning for patients with gastric cancer, demonstrating substantial clinical applicability.

Currently, histopathological examination remains the gold standard for evaluating the efficacy of neoadjuvant therapy. However, its limitation lies in the necessity for post-surgical evaluation, which could potentially delay timely adjustments to treatment plans. In our study, we

chose MPR as the predictive endpoint. Achieving MPR is associated with better clinical outcomes, including higher survival rates and lower recurrence risks [30]. The significance of our study lies in providing key preoperative information through a non-invasive CT-based prediction model, enabling clinicians to adjust treatment strategies earlier. In the context of predicting a negative MPR to neoadjuvant immunotherapy—which suggests that the patient may not respond favorably to the current treatment regimen—it is advisable to consider adjustments to the treatment plan. This approach may involve the combination of targeted pharmacological agents or the transition to alternative immunotherapy options. The primary objective is to formulate a personalized treatment strategy that enhances therapeutic efficacy and improves patient quality of life. Conversely, if the predictive model indicates a favorable response to neoadjuvant therapy, such as predicting a MPR, surgical intervention can be pursued without unnecessary delay, thereby minimizing the physical, psychological, and financial burdens on the patient. These scenarios underscore the direct utility of predictive models in optimizing treatment strategies and improving patient outcomes, offering substantial clinical guidance.

Previous studies have shown that a prediction model based on radiomic and deep learning features from CT images taken before neoadjuvant chemotherapy, performed well in predicting TRG 0–1, achieving an AUC of 0.848 in the training set [31]. Shen et al. predicted patients' three-year overall survival using deltaCT features, achieving an AUC of 0.827 [32]. In Wang et al.'s study, a model using pre-treatment and pre-surgery radiomic features performed excellently in predicting three-year survival, with an AUC of 0.769 [33]. However, these models primarily focused on patients undergoing neoadjuvant chemotherapy and did not differentiate between those receiving neoadjuvant immunotherapy combined with chemotherapy. It is recognized that the tumor microenvironment (TME) plays a crucial role in influencing treatment outcomes. Different TMEs not only affect treatment efficacy but also undergo dynamic changes and adaptations during therapy, involving complex interactions. For example, chemotherapeutic agents such as fluorouracil, oxaliplatin, and pemetrexed can enhance the activation of anti-tumor immune cell subsets, including dendritic cells and tumor-infiltrating lymphocytes [34–36]. Additionally, monocytes and macrophages, which can constitute up to 50% of tumor mass in certain cases, play a pivotal role in the tumor microenvironment [37]. Macrophages are not only integral to cancer-immune interactions but also have potential applications as diagnostic imaging tools and as vehicles for targeted chemotherapeutic delivery to cancer cells [38, 39]. Conversely, immune checkpoint blockade (ICB)

therapies primarily act on T cells to restore their anti-tumor activity and function [40]. ICB also has significant regulatory effects on other immune cells within the TME [41, 42]. Studies have shown that radiomic features are closely related to the composition and characteristics of the TME [43]. Therefore, the mechanisms by which chemotherapy and immunotherapy affect the TME differ, leading to variations in radiomic feature representation. Thus, radiomic models that were previously used to predict the efficacy of neoadjuvant chemotherapy in gastric cancer cannot be directly applied to patients receiving neoadjuvant immunotherapy in combination with chemotherapy. Constructing a specific model to predict the response of patients undergoing neoadjuvant immunotherapy is essential.

CT radiomics offers a non-invasive method to evaluate tumor characteristics without the need for direct tissue sampling, thereby avoiding the risks and discomfort associated with biopsies. Furthermore, CT radiomics extracts multidimensional features from tumors, comprehensively reflecting tumor heterogeneity, including morphological, textural, and density information [44, 45]. Additionally, CT imaging can be performed pre-treatment, during treatment, and post-treatment, providing continuous, dynamic information about tumor changes and allowing the model to capture the overall features of the tumor longitudinally. In this study, we applied radiomics approaches to extract numerous radiomic features from CT images and used various feature selection techniques, including the mRMR algorithm and LASSO regression, to identify the most representative features. This feature selection process effectively reduced data dimensionality, removed redundant features, and retained variables most relevant to treatment response prediction. Through Bayesian optimization and grid search, we selected the best-performing LightGBM model from among several ML models. Furthermore, we conducted SHAP analysis to identify key decision factors in the model, thereby enhancing its transparency and interpretability. These steps not only increased the model's credibility but also laid a solid data foundation for its application in clinical practice. Previous radiomic investigations into neoadjuvant treatment for gastric cancer have predominantly utilized traditional radiomics methodologies [46–48], which generally concentrate on single-time-point analyses and fail to adequately account for the temporal evolution of radiomic features [44]. This limitation impedes the effective capture of dynamic tumor changes during treatment, potentially leading to the exclusion of critical information. Delta radiomics, a branch of radiomics, differs from traditional radiomics, which focuses on imaging data at a single-time-point, by acquiring imaging data repeatedly and extracting quantitative features to reveal changes over time. Studies have shown that this method

is more sensitive and specific than traditional methods, capable of detecting changes in treatment effects earlier [49, 50]. Our study utilized deltaCT radiomic features of patients with gastric cancer before and after neoadjuvant immunotherapy, reflecting post-treatment imaging attributes and quantifying dynamic changes in textural radiological properties between baseline and post-treatment CT scans, allowing for dynamic monitoring of changes during treatment and demonstrating advanced precision in assessing treatment effects. Our multi-time-point radiomics approach utilizes the planned CT scans that patients undergo during their treatment, thereby avoiding additional exposure to ionizing radiation. This method not only mitigates the risks associated with additional radiation but also enhances clinical decision-making by providing more dynamic information, all without incurring additional costs. Our strategy increases the informational value of radiological assessments without compromising patient safety or budgetary constraints, offering a pragmatic solution for improving personalized therapeutic strategies.

Moreover, the ML model adopted in this study demonstrates significant advancements. By employing Bayesian optimization techniques to optimize the hyperparameters of the LightGBM model, we achieved high accuracy and stability in predicting pathological responses. Compared to traditional statistical methods, ML models can handle more complex nonlinear relationships and effectively reduce redundant features, thereby enhancing the model's generalizability. Thus, this study not only innovates in the field of radiomics but also demonstrates significant potential for the application of ML models, providing robust support for clinical decision-making. It is imperative to highlight that, in the process of selecting the optimal model, we initially ensured the absence of statistically significant differences in the AUC between the training and validation datasets to mitigate the risk of overfitting. Subsequently, we evaluated a comprehensive array of performance metrics for both datasets, including AUC, accuracy, sensitivity, specificity, Cohen's kappa coefficient, and Matthews correlation coefficient. The final model selected was the one that exhibited the most consistent performance across all metrics following a thorough evaluation. Although the LightGBM model demonstrated suboptimal performance on certain metrics relative to other models, the selection process prioritized stability over merely selecting the model with the highest performance in a single metric.

In the development cohort, the models achieved a sensitivity of 0.742 and a specificity of 0.818, reflecting a high degree of accuracy in distinguishing patients with favorable versus unfavorable treatment responses. Such high sensitivity and specificity are vital for clinical decision-making, as they enable clinicians to accurately evaluate

patient responses and tailor treatment plans accordingly. Furthermore, the positive predictive value was 0.697, and the negative predictive value was 0.849, underscoring the model's reliability in predicting both MPR and non-MPR patients. The Cohen's kappa coefficient was 0.552 in the development cohort and 0.455 in the internal validation cohort, suggesting moderate to good consistency in MPR prediction. Overall, our model demonstrated robust predictive performance across both the development and internal validation cohorts, and exhibited significant generalization capability in the external validation cohort. The findings of this study indicate that our model, which leverages radiomic features, is capable of accurately predicting the responses of patients with locally advanced gastric cancer to neoadjuvant immunotherapy. This capability provides clinicians with a valuable tool for optimizing treatment strategies. Nevertheless, despite the model's demonstrated predictive capacity, further research is necessary to validate its long-term efficacy and to explore the translation of these findings into practical applications for routine clinical practice. Given the increasing volume of data and advancements in technology, continuous optimization of the model will be crucial for enhancing predictive accuracy.

In the external validation cohort, we observed that metrics such as sensitivity and specificity were somewhat lower compared to those in the development cohort and internal validation cohort. Nonetheless, the model exhibited satisfactory overall performance, with an AUC of 0.714, demonstrating robust predictive capability. The decline in performance metrics may be attributed to several factors. Primarily, the relatively small sample size in the external validation cohort could result in greater variability in model performance and insufficient representation of certain features due to limited data, potentially affecting predictive accuracy. Furthermore, the data from the external validation cohort were sourced from various medical centers, resulting in variations in patient characteristics, treatment protocols, and imaging equipment when compared to the development cohort and internal validation cohort datasets. These discrepancies further influence the predictive efficacy of the model. Despite the application of the efficient LightGBM algorithm, the optimization of hyperparameters through Bayesian optimization, and the reduction of model complexity via feature selection with LASSO regression, the model remains susceptible to overfitting. This vulnerability is primarily attributed to the small and highly heterogeneous sample size of the external validation cohort, which poses challenges in generalizing the model to new datasets. In summary, the observed decline in the machine learning model's performance within the external validation cohort can be attributed to multiple contributing factors. This indicates that, although the model has undergone

extensive training and optimization, there is still potential for enhancement, particularly when applied to diverse real-world data.

Through SHAP analysis, we identified key radiomic features, some of which are associated with tumor biology. For example, the features `SmallDependenceLowGrayLevelEmphasis` and `DependenceNonUniformityNormalized` show a strong association with CD8-positive cell infiltration [51], which plays a crucial role in the tumor immune response. Additionally, the `LargeAreaHighGrayLevelEmphasis` feature is correlated with tertiary lymphoid structures, particularly mature tertiary lymphoid tissues, which are significant components of the tumor microenvironment [52]. The `Elongation` and `firstorder_90 Percentile` features are linked to a T-cell inflamed tumor microenvironment [53], often associated with improved responses to immunotherapy. Furthermore, the `firstorder_Median` feature can identify immune phenotypes [54]. These analyses suggest that radiomic features not only capture the macroscopic and microscopic structural characteristics of tumors but also offer a non-invasive means of understanding tumor biological behavior. Importantly, while some radiomic features have demonstrated potential biological significance in other studies, the biological context underlying most features remains largely unclear. In future research endeavors, we intend to incorporate foundational research methodologies to further explore the biological mechanisms underlying these features.

In this study, MPR was selected as the predictive endpoint instead of pCR due to several considerations. Firstly, evidence suggests that patients achieving MPR generally demonstrate higher survival rates and lower recurrence risks compared to those who do not achieve MPR [30]. Secondly, in clinical practice, pCR is an exceedingly stringent criterion, with only a small proportion of patients meeting this standard [55]. Conversely, MPR serves as a more attainable yet highly relevant metric, more frequently observed in real-world treatment settings, thereby enhancing its suitability for guiding clinical decisions. For patients who do not achieve pCR but exhibit significant pathological responses, MPR provides a more realistic and achievable target, enabling clinicians to adjust or optimize treatment strategies to improve therapeutic outcomes. In summary, MPR is not only a robust indicator closely associated with long-term patient prognosis but also offers a more flexible and practical alternative compared to pCR. This study assists clinicians in developing more personalized treatment strategies tailored to individual patient circumstances, thereby improving overall therapeutic efficacy.

This study aims to develop a predictive model specifically tailored to evaluate the response to neoadjuvant immunotherapy in patients with gastric cancer. The

biological characteristics of tumors and their radiomic features in patients who have previously undergone targeted therapy, radiotherapy, or only chemotherapy may have been modified by these prior treatments. For example, radiotherapy can induce local tissue fibrosis [56], while targeted therapies may alter the metabolic and angiogenic properties of tumors [57]. Such alterations could affect the extraction and analysis of radiomic features, thereby complicating the accurate assessment of responses to neoadjuvant immunotherapy. By excluding patients who have received prior treatments or only chemotherapy, we aim to more accurately assess the efficacy of neoadjuvant immunotherapy without confounding factors. The predictive model developed in this study offers robust support for clinical decision-making. Nevertheless, extrapolating this model to other cancer types may present challenges due to the distinct biological characteristics and radiological manifestations inherent to different cancers. For example, gastric, colorectal, and pancreatic cancers exhibit significant variations in their tumor microenvironments (TMEs) and radiomic features [58–60], which may influence the model's applicability. Consequently, directly applying the gastric cancer model to other cancer types may not adequately capture these distinctions, potentially affecting the model's predictive performance. Future research should incorporate more fundamental metrics and multimodal data to enhance the model's generalization capability across various cancers, thereby improving its clinical applicability.

Current radiomic models predominantly depend on features derived from single-time-point data, which inadequately capture the dynamic changes occurring within tumors. Furthermore, these models often exhibit limitations in feature selection and model optimization [46–48], frequently relying on basic statistical or texture features while overlooking the potential of more complex features, such as those derived from wavelet decomposition [61]. Additionally, these models are often criticized as 'black boxes' due to their lack of interpretability in predictive outcomes and their limited generalization capabilities. The strength of our research lies in addressing these limitations by incorporating dynamic data, implementing rigorous feature selection and model optimization, enhancing model interpretability, and utilizing multi-center datasets. These advancements not only improve the predictive performance of the models but also provide more reliable and practical tools for clinical application. The strengths of this study include the use of data from a multicenter cohort, which increases the model's generalizability. Strict radiomic feature selection and model optimization processes ensured the scientific rigor and reliability of the model, contributing to its robustness in clinical applications. Additionally, the use of SHAP analysis enhanced the model's interpretability,

allowing clinicians to better understand the results and effectively apply the prediction model.

Despite the advances presented in this study, our approaches had some limitations that should be acknowledged. Although the model's generalization capability and the robustness of external validation are enhanced through a multi-center design, variations in CT imaging acquisition protocols across different medical institutions may influence the extraction of radiomic features and, consequently, affect model performance [62]. Different centers may employ diverse scan parameter settings, such as tube voltage, tube current, slice thickness, and reconstruction algorithms, which can affect the consistency and reproducibility of radiomic features, thereby influencing model performance [63]. Furthermore, the models and technical specifications of CT equipment utilized by various centers directly impact image quality and resolution [64]. Completely eliminating biases arising from differences in CT acquisition protocols remains a significant challenge. Future research should focus on addressing this issue by enhancing data collection methods or developing more robust feature extraction techniques, ensuring that radiomic models maintain stable and reliable performance across diverse clinical settings.

Despite the evaluation of model performance through multi-cohort validation, it is crucial to remain vigilant regarding potential overfitting issues. The limited sample size in the external validation dataset may affect the stability of the validation results, potentially impacting the accurate assessment of the model's generalization capability. During the feature selection phase, we employed techniques such as minimum Redundancy Maximum Relevance and Least Absolute Shrinkage and Selection Operator regression to identify the 20 most representative features from a plethora of radiomic features. However, the limited sample size may have led to an excessive number of selected features, thereby heightening the risk of overfitting. Additionally, while Bayesian optimization and grid search enhanced the model's performance on the training dataset during model optimization, they may have also intensified the risk of overfitting. Despite validating the model's performance using both the internal validation cohort and the external validation cohort, the model's robustness remains potentially compromised by the small sample sizes of these cohorts.

By employing a radiomic approach, we incorporated a comprehensive set of characteristics to thoroughly characterize tumor attributes, including shape metrics, first-order statistical measures, and texture descriptors, among others. Through feature selection and machine learning techniques, we meticulously identified the most representative radiomic variables for predicting major pathological response. These features effectively capture tumor heterogeneity and biological behavior, thereby

significantly enhancing the model's predictive performance. Nevertheless, the Pyradiomics package has certain limitations that prevented us from directly extracting information regarding tumor location. Consequently, future research will aim to further optimize the model to encompass a wider array of tumor characteristics. In addition, we intend to investigate a broader spectrum of biomarkers to elucidate the mechanisms underlying tumor response to treatment from multiple perspectives, thereby offering more precise criteria for clinical treatment decisions.

Although various measures were taken to minimize overfitting during model construction, the generalizability of the model to a broader population still requires further validation due to limitations in sample size. Future research should expand the sample size and validate the model using data from multicenter, large-scale clinical trials. Moreover, while radiomics can provide a wealth of information on tumor heterogeneity, issues of consistency across different scanning devices, protocols, and data processing workflows need to be addressed. Integrating other biomarkers, such as genomic data, tumor mutation burden, and TME characteristics, with radiomics for multimodal data fusion may further enhance the predictive accuracy and clinical value of the model.

## Conclusion

In conclusion, this study demonstrates the potential of CT-based radiomic models in predicting the response of LAGC patients to neoadjuvant immunotherapy, providing new insights into advancing precision medicine in gastric cancer treatment.

## Abbreviations

MPR	Major pathological response
LAGC	Locally advanced gastric cancer
CT	Computed tomography
AUC	Area under the curve
DC	Development cohort
IVC	Internal validation cohort
EVC	External validation cohort
PPV	Positive predictive value
NPV	Negative predictive value
MCC	Matthews correlation coefficient
SHAP	Shapley Additive Explanations
GC	Gastric cancer
ICI	Immune checkpoint inhibitor
pCR	Pathological complete response
TRG	Tumor regression grade
TMB	Tumor mutational burden
ML	Machine learning
ROI	Regions of interest
ICC	Interclass correlation coefficient
mRMR	Minimum Redundancy Maximum Relevance
LASSO	Least absolute shrinkage and selection operator
ROC	Receiver operating characteristic
DT	Decision Tree
MLP	Multi-layer Perceptron
KNN	K-Nearest Neighbors
RF	Random Forest

Enet	Elastic Net
SVM	Support Vector Machine

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12967-025-06363-z>.

Supplementary Material 1: Table S1. The results of CT radiomic features selection. Table S2. The radiomics machine learning model performance in the external validation cohort. Figure S1. Baseline, surgical, and pathological information of the patients. Figure S2. Histogram of the frequency distribution generated in feature screening (based on ICC values). Figure S3. Predictor variable selection based on the least absolute shrinkage and selection operator regression method. (A) Optimal parameter (lambda) selection in the least absolute shrinkage and selection operator model; (B) Least absolute shrinkage and selection operator coefficient profiles of the candidate features. Figure S4. Radar charts for the comparison of the performance among different models. (A) ML models constructed based on Bayesian hyperparameter optimization in the development cohort; (B) ML models constructed based on Bayesian hyperparameter optimization in the internal validation cohort; (C) ML models constructed based on Grid search hyperparameter optimization in the development cohort; (D) ML models constructed based on Grid search hyperparameter optimization in the internal validation cohort. Figure S5. P values derived from the DeLong test comparing AUC values between different prediction models in the development cohort. Figure S6. SHAP dependence plot. Each dependence plot shows how a single feature affects the output of the prediction model, and each dot represents a single patient. SHAP values are represented by the y-axis, and actual values are represented by the x-axis. The SHAP values for specific features exceeding zero push the decision towards the "Yes" class.

## Acknowledgements

We thank who have devoted a lot to this study, including nurses, further-study doctors, statisticians, reviewers and editors.

## Author contributions

Jian-Wei Xie had full access to all the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis. Ze-Ning Huang, Hao-Xiang Zhang and Yu-Qin Sun contributed equally to this work and should be considered as shared first authors. Chang-Ming Huang, Jian-Xian Lin and Jian-Wei Xie contributed equally as shared last authors. Concept and design: Ze-Ning Huang, Hao-Xiang Zhang, Jian-Xian Lin, and Jian-Wei Xie. Acquisition, analysis, or interpretation of data: Ze-Ning Huang, Hao-Xiang Zhang, Yu-Qin Sun, Xing-Qi Zhang, Cai-Ming Weng, Ping-Li, Jia-Bin Wang, Qi-Yue Chen, and Long-Long Cao. Drafting of the manuscript: Ze-Ning Huang, Hao-Xiang Zhang, Yi-Fen Lin, and Yu-Qin Sun. Statistical analysis: Ze-Ning Huang, Hao-Xiang Zhang, Yi-Fen Lin, and Yu-Qin Sun. Administrative, technical, or material support: Yu-Qin Sun, Xing-Qi Zhang, Mi Lin, Ru-Hong Tu, Chang-Ming Huang, Chao-Hui Zheng, Jian-Xian Lin, and Jian-Wei Xie. Supervision: Chang-Ming Huang, Jian-Xian Lin, and Jian-Wei Xie. All authors read and approved the final manuscript.

## Funding

This study was supported by the Fujian Province Medical "Creating high-level hospitals, high-level medical centers and key specialty projects" (MWY2 [2021] No.76) and the Natural Science Foundation of Fujian Province grants (2023J01674).

## Data availability

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

## Declaration

### Ethics approval and consent to participate

This study obtained approval from the Independent Ethics Committee of the Fujian Medical University Union Hospital and Zhangzhou Affiliated Hospital of Fujian Medical University.

### Consent for publication

Not applicable.

### Competing interests

The authors declare that they have no competing interests.

### Author details

<sup>1</sup>Department of Gastric Surgery, Fujian Medical University Union Hospital, No. 29 Xin-quan Road, Fuzhou, Fujian Province 350001, China

<sup>2</sup>Key Laboratory of Ministry of Education of Gastrointestinal Cancer, Fujian Medical University, Fuzhou, China

<sup>3</sup>Department of Gastrointestinal Surgery, Zhangzhou Affiliated Hospital of Fujian Medical University, Zhangzhou, China

<sup>4</sup>Department of Imaging, Zhangzhou Affiliated Hospital of Fujian Medical University, Zhangzhou, China

Received: 1 November 2024 / Accepted: 8 March 2025

Published online: 24 March 2025

## References

- Hyuna S, Jacques, Ferlay, Rebecca L, Siegel, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 Countries. *CA Cancer J Clin.* 2021;71:0.
- Noh Sung Hoon, Park Sook Ryun, Yang Han-Kwang. Adjuvant capecitabine plus oxaliplatin for gastric cancer after D2 gastrectomy (CLASSIC): 5-year follow-up of an open-label, randomised phase 3 trial. *Lancet Oncol.* 2014;15:1389–96.
- Sasako Mitsuru, Sakuramoto Shinichi, Katai Hitoshi. Five-year outcomes of a randomized phase III trial comparing adjuvant chemotherapy with S-1 versus surgery alone in stage II or III gastric cancer. *J Clin Oncol.* 2011;29:4387–93.
- David C, William H, Allum, Sally P, Stenning, et al. Perioperative chemotherapy versus surgery alone for resectable gastroesophageal cancer. *N Engl J Med.* 2006;355:0.
- Marc Y, Valérie BJ-P. Perioperative chemotherapy compared with surgery alone for resectable gastroesophageal adenocarcinoma: an FNCLCC and FFOCD multicenter phase III trial. *J Clin Oncol.* 2011;29:0.
- Salah-Eddin, Al-Batran, Ralf D, Hofheinz, Claudia P et al. Histopathological regression after neoadjuvant docetaxel, oxaliplatin, fluorouracil, and leucovorin versus epirubicin, cisplatin, and fluorouracil or capecitabine in patients with resectable gastric or gastro-oesophageal junction adenocarcinoma (FLOT4-AIO): results from the phase 2 part of a multicentre, open-label, randomised phase 2/3 trial. *Lancet Oncol.* 17: 0.
- Livingstone E, Zimmer L, Hassel JC, et al. Adjuvant nivolumab plus ipilimumab or nivolumab alone versus placebo in patients with resected stage IV melanoma with no evidence of disease (IMMUNED): final results of a randomised, double-blind, phase 2 trial. *Lancet.* 2022, 400: 1117–1129.
- Gandhi Leena, Rodríguez-Abreu Delvys, Gadgeel Shirish, Engl. Pembrolizumab plus chemotherapy in metastatic Non-Small-Cell lung Cancer. *N J Med.* 2018;378:2078–92.
- Janjigian Yelena Y, Shitara Kohei, Moehler Markus. 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. *Lancet.* 2021;398: 27–40.
- Kang Yoon-Koo, Chen Li-Tzong, Ryu Min-Hee et al. Nivolumab plus chemotherapy versus placebo plus chemotherapy in patients with HER2-negative, untreated, unresectable advanced or recurrent gastric or gastro-oesophageal junction cancer (ATTRACTION-4): a randomised, multicentre, double-blind, placebo-controlled, phase 3 trial. *Lancet Oncol.* 2022, 23: 234–47.
- Hua-Long LJ-X et al. Neoadjuvant camrelizumab and apatinib combined with chemotherapy versus chemotherapy alone for locally advanced gastric cancer: a multicenter randomized phase 2 trial. *Nat Commun.* 2024, 15: 41.
- Alessandra R, Federica P, Michele P et al. Tremellumab and durvalumab combination for the Non-Operative management (NOM) of microsatellite instability (MSI)-High resectable gastric or gastroesophageal junction cancer: the multicentre, Single-Arm, Multi-Cohort, phase II INFINITY Study. *Cancers (Basel).* 2021, 13: 0.
- Song L, Wenbin Y, Fei X, et al. Neoadjuvant therapy with immune checkpoint Blockade, antiangiogenesis, and chemotherapy for locally advanced gastric cancer. *Nat Commun.* 2023;14:0.
- Di Federico A, Alden SL, Smithy JW, et al. Inpatient variation in PD-L1 expression and tumor mutational burden and the impact on outcomes to immune checkpoint inhibitor therapy in patients with non-small-cell lung cancer. *Ann Oncol.* 2024;35:0.
- Minsuk K, Minae, An, Samuel J, Klempner et al. Determinants of response and intrinsic resistance to PD-1 Blockade in microsatellite Instability-High gastric Cancer. *Cancer discov.* 2021, 11: 0.
- Qingzhu J, Aoyun W, Yixiao Y, et al. Heterogeneity of the tumor immune microenvironment and its clinical relevance. *Exp Hematol Oncol.* 2022;11:0.
- Seevaratnam Rajini, Cardoso Roberta, McGregor, Caitlin et al. How useful is preoperative imaging for tumor, node, metastasis (TNM) staging of gastric cancer? A meta-analysis. *Gastric Cancer.* 2012, null: S3-18.
- Eisenhauer EA, Therasse P, Bogaerts J, et al. New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). *Eur J Cancer.* 2009;45:228–47.
- Weiqiu J, Yu T, Wendi X, et al. Non-linear modifications enhance prediction of pathological response to pre-operative PD-1 Blockade in lung cancer: A longitudinal hybrid radiological model. *Pharmacol Res.* 2023;198:0.
- Yong C, Jinling J, Chao Y, et al. Prediction of tumor regression grade in far-advanced gastric cancer after preoperative immuno-chemotherapy using dual-energy CT-derived extracellular volume fraction. *Eur Radiol.* 2024;0:0.
- Xujie G, Tingting M, Jingli C, et al. A CT-based radiomics model for prediction of lymph node metastasis in early stage. *Gastric Cancer. Acad Radiol.* 2020;28:0.
- Zong-Qiong SS-D, Jie H. Radiomics study for differentiating gastric cancer from gastric stromal tumor based on contrast-enhanced CT images. *J J Xray Sci Technol.* 2019;27:0.
- Yonghe C, Kaikai W, Dan L, et al. A machine learning model for predicting a major response to neoadjuvant chemotherapy in. *Adv Gastric Cancer. Front Oncol.* 2021;11:0.
- Jiayi Z, Yanfen C, Kaikai W, et al. Deep learning predicts resistance to neoadjuvant chemotherapy for locally advanced gastric cancer: a multicenter study. *Gastric Cancer.* 2022;25:0.
- Chen Y, Yizi, et al. Metabolomic ML predictor for diagnosis and prognosis of gastric cancer. *Nat Commun.* 2024;15:1657.
- Pera Manuel, Gibert Joan, Gimeno Marta. Machine learning risk prediction model of 90-day mortality after gastrectomy for Cancer. *Ann Surg.* 2022;276:776–83.
- Hu J, Xu J, Li M, et al. Identification and validation of an explainable prediction model of acute kidney injury with prognostic implications in critically ill children: a prospective multicenter cohort study. *EClinicalMedicine.* 2024, 68: 102409.
- Wei-Ju TS-FYC-T et al. Development and validation of an insulin resistance model for a population without diabetes mellitus and its clinical implication: a prospective cohort study. *EClinicalMedicine.* 2023, 58: 101934.
- Jing Z, Kaixing F, Dawei W, et al. Refining hydrogel-based sorbent design for efficient toxic metal removal using machine learning-Bayesian optimization. *J Hazard Mater.* 2024;479:0.
- Zhichao J, Yibin X, Wen Z, et al. Perioperative chemotherapy with docetaxel plus oxaliplatin and S-1 (DOS) versus oxaliplatin plus S-1 (SOX) for the treatment of locally advanced gastric or gastro-oesophageal junction adenocarcinoma (MATCH): an open-label, randomized, phase 2 clinical trial. *Gastric Cancer.* 2024;27:0.
- Yanfen C, Jiayi Z, Zhenhui, Li et al. A CT-based deep learning radiomics nomogram for predicting the response to neoadjuvant chemotherapy in patients with locally advanced gastric cancer: A multicenter cohort study. *EClinicalMedicine.* 2022, 46: 0.
- Li-Li SH-L, Fang-Hui Z. Delta computed tomography radiomic features-based nomogram predicts long-term efficacy after neoadjuvant chemotherapy in advanced gastric cancer. *Radiol Med.* 2023;128:0.
- Yinkui W, Lei T, Xiangji Y, et al. Pre- and Post-treatment Double-Sequential-Point dynamic radiomic model in the response prediction of gastric cancer to neoadjuvant chemotherapy: 3-Year survival Analysis. *Ann Surg Oncol.* 2023;31:0.

34. Aleksandar O, Casey A, Mikko T, et al. Systematic Elucidation and Pharmacological targeting of tumor-infiltrating regulatory T cell master regulators. *Cancer Cell*. 2023;41:0.
35. David A, Schaer, Sandaruwan G, Nelusha A et al. The folate pathway inhibitor pemetrexed pleiotropically enhances effects of cancer Immunotherapy.[J]. *Clin cancer res*, 25: 0.
36. Darya A, Malika, Trad, Neale T, Hanke, et al. Doxorubicin eliminates myeloid-derived suppressor cells and enhances the efficacy of adoptive T-cell transfer in breast cancer. *Cancer Res*. 2013;74:0.
37. Christopher B, Rodell, Sean P, Arlauckas, Michael F, Cuccarese, et al. TLR7/8-agonist-loaded nanoparticles promote the polarization of tumour-associated macrophages to enhance cancer immunotherapy. *Nat Biomed Eng*. 2019;2:0.
38. Riddha D, Joseph, Hardie, Bishnu P, Joshi, et al. Macrophage-Encapsulated Bioorthogonal Nanozymes Target Cancer Cells. *JACS Au*. 2022;2:0.
39. Bishnu P, Joshi, Joseph HMA, Mingroni, et al. Surface-Modified macrophages facilitate tracking of breast Cancer-Immune interactions. *ACS Chem Biol*. 2018;13:0.
40. Giacomo O, Catherine J. Wu, Dynamics and specificities of T cells in cancer immunotherapy. *Nat Rev Cancer*. 2023;23:0.
41. Rui-Qi WX-M, Dong-Ping L. Chen Immune checkpoint therapy-elicited sialylation of IgG antibodies impairs antitumorigenic type I interferon responses in hepatocellular carcinoma. *Immunity*, 2022, 56: 0.
42. Junjie H, Lele Z, Haoran X et al. Tumor microenvironment remodeling after neoadjuvant immunotherapy in non-small cell lung cancer revealed by single-cell RNA sequencing. *Genome med*, 15: 0.
43. Guan-Hua S, Yi X, Lin J, et al. Radiomic features for assessing tumor-infiltrating lymphocytes correlate with molecular traits of triple-negative breast cancer. *J Transl Med*. 2022;20:0.
44. Philippe L, Emmanuel R-V, Ralph L, et al. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*. 2012;48:0.
45. Janita E, van Davide T, Stephanie C. Radiomics in medical imaging-how-to. *Guide Crit Reflect. Insights Imaging*. 2020;11:0.
46. Zongqiong S, Xiaofang C, Yuxi G, et al. An application study of low-dose computed tomography perfusion imaging for evaluation of the efficacy of neoadjuvant chemotherapy for advanced gastric adenocarcinoma. *Gastric Cancer*. 2017;21:0.
47. Jia F, Lei TZ-Y, Li, et al. Diffusion kurtosis imaging in the prediction of poor responses of locally advanced gastric cancer to neoadjuvant chemotherapy. *Eur J Radiol*. 2020;128:0.
48. Xiaoyuan G, Yang Z, Fei Y, et al. Locally advanced gastric cancer: total iodine uptake to predict the response of primary lesion to neoadjuvant chemotherapy. *J Cancer Res Clin Oncol*. 2018;144:0.
49. Mohammadhadi K, Prateek P, Amit G, et al. Changes in CT radiomic features associated with lymphocyte distribution predict overall survival and response to immunotherapy in Non-Small. *Cell Lung Cancer*. *Cancer Immunol Res*. 2019;8:0.
50. Nikita S, Leonardo R, Luis A, et al. Time series radiomics for the prediction of prostate cancer progression in patients on active surveillance. *Eur Radiol*. 2023;33:0.
51. Colin Y, Wang. Daniel Thomas, Ginat, Preliminary computed tomography radiomics model for predicting pretreatment CD8+T-Cell infiltration status for primary head and neck squamous cell Carcinoma. *J Comput Assist Tomogr*. 2021;45:0.
52. Qing W, Yushuai Y, Chenxi, Wang, et al. Heterogeneity of tertiary lymphoid structures predicts the response to neoadjuvant therapy and immune micro-environment characteristics in triple-negative breast cancer. *Br J Cancer*. 2024;132:0.
53. Kinga B, Ramon A, Olivia P, et al. Radiomics signature for dynamic monitoring of tumor inflamed microenvironment and immunotherapy response prediction. *J Immunother Cancer*. 2025;13:0.
54. Jun Z, Zhenru W, Jian Z, et al. Intrahepatic cholangiocarcinoma: MRI texture signature as predictive biomarkers of immunophenotyping and survival. *Eur Radiol*. 2020;31:0.
55. Yuzhou Z, Danyang, Li, Jing Z, et al. Comprehensive multi-omics analysis of resectable locally advanced gastric cancer: assessing response to neoadjuvant camrelizumab and chemotherapy in a single-center, open-label, single-arm phase II trial. *Clin Transl Med*. 2024;14:0.
56. June Y Choi, Myung Jun, Kim, Young Joo, Lee et al. Prevention of radiotherapy-induced pro-tumorigenic microenvironment by SFK inhibitors. *Theranostics*. 2025;15:0.
57. Alexandra T. Ari, Ristimäki, Targeted therapy in gastric cancer.[J]. *APMIS*, 2015, 123: 0.
58. Arefeh ZT. Sogand, vahidi, unraveling the interplay of CD8+T cells and MicroRNA signaling in cancer: implications for immune dysfunction and therapeutic approaches. *J Transl Med*. 2024;22:0.
59. Tim R, de Back, Sander R, van Hooft, Dirkje W, Sommeijer, et al. Transcriptomic Subtyping Gastrointestinal Malignancies. *Trends Cancer*. 2024;10:0.
60. Xun Y, Caixia S, Fei X, et al. Radiomic signature-based nomogram to predict disease-free survival in stage II and III colon cancer[. *J Eur J Radiol*. 2020;131:0.
61. Sun Roger, Limkin Elaine Johanna, Vakalopoulou Maria. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol*. 2018;19:1180–91.
62. Philippe L, Ralph TH, Leijenaar, Timo M, Deist, et al. Radiomics: the Bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14:0.
63. Aerts HJWL, Velazquez ER, Leijenaar RTH et al. Corrigendum: decoding tumour phenotype by noninvasive imaging using a quantitative radiomics approach. *Nat Commun*, 2014, 5.
64. Midya A, Chakraborty J, Gnen M, et al. Influence of CT acquisition and reconstruction parameters on radiomic feature reproducibility. *J Med Imaging*. 2018;5(1):011020.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.