ACCSCIENCE PUBLISHING

Eurasian Journal of Medicine and Oncology

ORIGINAL RESEARCH ARTICLE

An exploratory PET/MR radiomics framework for preoperative prediction of HER2 status and metastatic burden in colorectal cancer

Zengbei Yuan[1,2,3†], Kun Wang[1,2†], Wen Jiang[1,2†], Zhixian Tang[1], Xing Chen[2], Weilun Wang[2], Yihong Yang[2], Zirong Zhou[2], Jun Zhao[2], Qingqing Huang[1*], and Na Qi[1,2*]

[1]Shanghai Key Laboratory of Molecular Imaging, Shanghai University of Medicine and Health Sciences, Shanghai, China

[2]Department of Nuclear Medicine, Shanghai East Hospital, School of Medicine, Tongji University, Shanghai, China

[3]School of Medical Technology and Engineering, Fujian Medical University, Fuzhou, Fujian, China

[†]*These authors contributed equally to this work.*

***Corresponding authors:**
Qingqing Huang
(qingqinghuang80@163.com);
Na Qi
(qinaseu@163.com)

Citation: Yuan Z, Wang K, Jiang W, *et al*. An exploratory PET/MR radiomics framework for preoperative prediction of HER2 status and metastatic burden in colorectal cancer. *Eurasian J Med Oncol*.
doi: 10.36922/EJMO026270322

Received: June 30, 2026

Revised: September 12, 2026

Accepted: October 28, 2026

Published online: November 16, 2026

Copyright: © 2026 Author(s). This is an Open-Access article distributed under the terms of the Creative Commons Attribution License, permitting distribution, and reproduction in any medium, provided the original work is properly cited.

Publisher's Note: AccScience Publishing remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract

Introduction: Noninvasive assessment of human epidermal growth factor receptor 2 (HER2) status and metastatic burden in colorectal cancer remains challenging. Multimodal positron emission tomography (PET)/magnetic resonance (MR) radiomics offers an opportunity for integrated molecular and metastatic characterization.

Objective: To develop and internally validate an exploratory PET/MR radiomics framework for the preoperative prediction of HER2 status and metastatic burden (non-regional lymph nodes) in colorectal cancer (CRC) patients.

Methods: This retrospective study incorporated 61 samples with pathologically confirmed CRC who underwent preoperative ^{18}F-fluorodeoxyglucose PET/MR imaging. HER2 positivity was defined according to the refined HER2 Amplification for CRC Enhanced Stratification criteria. Metastatic burden was categorized into three groups based on preoperative imaging: non-metastatic, regional lymph node metastasis, and distant metastasis. Tumor regions were manually segmented on PET, T1- and T2-weighted imaging (T1WI, T2WI), and 2,264 radiomic features were extracted from each imaging modality. Feature selection was performed using a comprehensive multi-step pipeline. An extreme gradient boosting (XGBoost)-based model was then developed and internally validated via 5-fold cross-validation, with performance metrics reported as mean ± standard deviation.

Results: For HER2 status prediction, the multimodal combined XGBoost model demonstrated superior performance compared to single-modality models, yielding a mean AUC of 0.9214 ± 0.0779. For metastatic burden prediction, the combined model further excelled, yielding an AUC of 0.827 ± 0.136. SHapley Additive exPlanations analysis revealed imaging signatures for the two prediction tasks. HER2 status prediction was primarily driven by texture-related features, particularly those derived from the Gray-Level Run Length Matrix, Gray-Level Co-occurrence Matrix, and First-order matrices, whereas metastatic burden prediction relied mainly on multiscale texture descriptors from the first-order, Neighborhood Gray-Tone Difference Matrix, and Gray-Level Dependence Matrix.

Conclusion: PET/MR radiomics offers an exploratory approach for future risk

stratification in CRC, although independent validation is required to confirm clinical utility.

Keywords: Colorectal cancer; Machine learning; Positron emission tomography/magnetic resonance; Radiomics; SHapley Additive exPlanations

1. Introduction

Colorectal cancer (CRC) is characterized by a complex pathogenesis involving dysregulation across multiple molecular signaling pathways. Emerging evidence has highlighted the critical role of human epidermal growth factor receptor 2 (HER2) in the tumorigenesis of CRC.[1] Amplification of the *HER2* gene or overexpression of its protein product has been closely associated with increased tumor aggressiveness and poor prognosis, while also serving as a promising molecular marker for targeted therapy.[2,3] This is particularly relevant in *KRAS* wild-type CRC patients, where the assessment of HER2 status can offer significant clinical guidance for targeted interventions.[4] As a result, the evaluation of CRC based on precise molecular profiling has become a major focus in current research efforts.[5] Beyond molecular alterations, metastatic burden represents another critical determinant of clinical outcome and therapeutic decision-making in CRC. Lymph node involvement is a cornerstone of the tumor, node, metastasis (TNM) staging system and serves as one of the strongest prognostic indicators for recurrence and survival. Regional lymph nodes are those located within the lymphatic drainage area of the primary colorectal tumor. Metastases to non-regional lymph nodes are classified as distant metastases.[6] Moreover, the extent of metastatic spread, ranging from localized disease to regional lymph node metastasis (LM) and distant dissemination (involving non-regional lymph nodes), directly influences surgical planning, neoadjuvant treatment strategies, and long-term management. Accurate preoperative assessment of metastatic burden remains challenging because conventional imaging approaches often rely primarily on morphological criteria and may fail to fully capture the biological aggressiveness of tumors. Therefore, the simultaneous characterization of molecular status and metastatic burden has emerged as an important goal for precision oncology, enabling more comprehensive risk stratification and individualized treatment planning in CRC.

Positron emission tomography (PET)/magnetic resonance (MR), an advanced modality integrating metabolic and anatomical imaging, is increasingly applied in CRC diagnosis, staging, and treatment evaluation.[7] PET imaging captures metabolic activity within tumor tissues, while MR imaging provides detailed anatomical structures and superior soft tissue contrast.[8] Radiomics extracts high-throughput quantitative features (shape, texture, and intensity) from PET/MR images to reveal underlying tumor micro-heterogeneity.[9] This imaging-based analytic approach holds promise for the non-invasive evaluation of tumor biological behavior and offers a reliable foundation for formulating individualized treatment strategies.[10] These extracted radiomic features, ranging from first-order intensity statistics to complex spatial texture descriptors, collectively quantify the intricate morphological and micro-environmental complexity of tumors that is imperceptible to visual assessment.[11-14]

Researchers are committed to combining radiomics and machine learning to predict individual biological characteristics of CRC, including clinical outcomes (prognosis, treatment response[15,16]), genetic status (*KRAS* mutations and *HER2* status[17,18]), and tumor characteristics (microsatellite instability, tumor fat infiltration[19,20]). Despite these advances, most existing radiomics studies in CRC have focused on a single clinical endpoint. For instance, while several models have been proposed for HER2 prediction or TNM staging, they predominantly rely on single-modality computed tomography (CT) or MRI. Prior PET/MR radiomics work remains limited by small sample sizes and insufficient multimodal integration. Furthermore, the clinical positioning of these models, whether intended as a replacement, a triage test, or an adjunct to standard pathology, has rarely been explicitly defined. HER2 status reflects the molecular landscape and therapeutic vulnerability of tumors, whereas metastatic burden represents disease dissemination and clinical staging. We developed the integrated CRC-molecular and metastatic profiling (MMP) framework. Crucially, CRC-MMP is not intended to replace tissue-based pathological assessment, but rather to serve as a non-invasive adjunct tool to complement standard clinical workflows.

Therefore, in this study, we developed the CRC-MMP framework to integrate the prediction of molecular status and metastatic burden in CRC. We specifically prespecified three hypotheses: (i) multimodal PET/MR radiomics can noninvasively predict HER2 status in CRC; (ii) it can

effectively discriminate predefined metastatic categories; and (iii) the integration of multimodal imaging yields superior predictive performance compared to individual modalities. Following manual tumor delineation, radiomic features were independently extracted from preoperative ^{18}F-fluorodeoxyglucose (FDG) PET, T1-weighted imaging (T1WI), and T2-weighted imaging (T2WI). After feature selection using a comprehensive multi-step feature selection pipeline, ML-based radiomics models were constructed for two complementary prediction tasks: prediction of HER2 expression status and metastatic burden. Furthermore, SHapley Additive exPlanations (SHAP) were employed to characterize modality-specific feature contributions and improve model interpretability. By integrating molecular profiling and metastatic assessment within a unified PET/MR radiomics framework, CRC-MMP aims to provide a non-invasive solution for comprehensive tumor characterization, preoperative risk stratification, and precision management of CRC.

2. Materials and methods

2.1. Enrolled participants

All patients who received treatment at Shanghai East Hospital between December 2022 and March 2025 were retrospectively enrolled in this study (Table 1 and Figure 1). Due to the retrospective design, strict consecutive enrollment could not be guaranteed; however, all patients meeting the inclusion criteria within this timeframe were screened to minimize selection bias. Patients were eligible for inclusion if they fulfilled these conditions: (i) histologically proven non-mucinous adenocarcinoma; (ii) available HER2 expression testing; (iii) preoperative ^{18}F-FDG PET/MR examination performed within two weeks before pathological evaluation; and (iv) complete preoperative imaging available for metastatic burden assessment. Initially, 75 patients were screened, and 14 were excluded based on the following criteria: (i) history of other malignancies (n = 4); (ii) receipt of radiotherapy, chemotherapy, immunotherapy, or other anti-tumor treatment before PET/MR examination (n = 6); (iii) history of inflammatory bowel disease, such as ulcerative colitis (n = 1); and (iv) incomplete clinical, pathological, or imaging data (n = 3).

For the molecular prediction task, HER2 status was determined according to pathological examination and dichotomized into HER2-positive and HER2-negative groups. For the metastatic prediction task, metastatic burden was determined based on preoperative imaging findings and categorized into three groups: non-metastatic disease (LM−), regional LM (LM+), and distant LM (LM++).

All patients were required to fast for at least six hours before undergoing the ^{18}F-FDG PET/MR examination, and peripheral blood glucose levels were maintained below 11.1 mmol/L. Imaging was performed on an integrated 3.0T time-of-flight PET/MR scanner (uPMR 790, United Imaging, China) approximately 60 minutes after the intravenous injection of ^{18}F-FDG with a dose of 190 ± 35 MBq (5.13 ± 0.95 mCi). The acquisition covered the region from the skull vertex to the mid-thigh. PET acquisition time was 5 minutes per bed position for the brain and 3 minutes per bed position for the body. The MRI protocol included standard clinical sequences, T1WI and T2WI, with parameters as follows: flip angle = 10°, echo train length = 30, field of view = 549 × 384 mm^2, scanning matrix = 256 × 329, slice thickness = 2 mm, and interslice gap = 2 mm, using axial plane acquisition. PET reconstruction algorithm: ordered subset expectation maximization.

2.2. Identification of human epidermal growth factor receptor 2 status and lymph node metastasis burden

For HER2 status determination, primary tumor surgical specimens were processed according to standard histopathological protocols and independently evaluated by board-certified pathologists. HER2 protein expression was assessed using immunohistochemistry (IHC) and scored as 0, 1+, 2+, or 3+ based on staining intensity. HER2 positivity was determined according to the refined HER2 Amplification for CRC Enhanced Stratification criteria.[21] For cases with equivocal IHC results (2+), fluorescence *in situ* hybridization was performed to confirm HER2 gene amplification. HER2 testing was performed on primary tumor surgical specimens. It should be noted that in our clinical practice, HER2 testing is not routinely performed for all CRC patients but is selectively ordered, often for those with advanced disease or when targeted therapy is being considered. Consequently, our retrospective cohort, which required the availability of both high-quality preoperative PET/MR imaging and complete HER2 testing results, naturally exhibits a higher proportion of HER2-positive cases compared to the normal CRC population.

Metastatic burden was determined based on preoperative radiological findings from contrast-enhanced CT, MRI, and ^{18}F-FDG PET, as documented in imaging reports and confirmed through multidisciplinary clinical evaluation.[22] It is crucial to note that the metastatic burden evaluated in this study strictly refers to lymph node metastatic burden and does not include any distant organ metastases (e.g., liver, lung, or peritoneal metastases). Patients were categorized into three groups according to the anatomical extent of lymph node involvement: LM−, no radiological evidence of LM; LM+, regional LM

Table 1. Demographic information of enrolled participants (mean ± standard deviation)

Characteristic	Total	HER2+	HER2−	LM−	LM+	LM++
Number	61	27	34	20	26	15
Age	60.7049 ± 4.2248	58.5582 ± 15.1640	63.3704 ± 12.7223	61.8000 ± 12.7840	60.9615 ± 15.0930	58.8000 ± 15.2606
Sex						
Male	38 (62.30%)	15 (55.56%)	23 (67.65%)	16 (80.00%)	14 (53.85%)	8 (53.33%)
Female	23 (37.70%)	12 (44.44%)	11 (32.35%)	4 (20.00%)	12 (46.15%)	7 (46.67%)
Type of differentiation						
Poor differentiation	29 (47.54%)	13 (48.15%)	16 (47.06%)	8 (40.00%)	14 (53.85%)	7 (46.67%)
Well or moderately differentiated	32 (52.46%)	14 (51.85%)	18 (52.94%)	12 (60.00%)	12 (46.15%)	8 (53.33%)
CEA						
<5 ng/ml	20 (32.79%)	8 (29.63%)	12 (35.29%)	9 (45.00%)	6 (23.08%)	5 (33.33%)
≥5 ng/ml	41 (67.21%)	19 (70.37%)	22 (64.71%)	11 (55.00%)	20 (76.92%)	10 (66.67%)
CA19-9						
<30 U/ml	36 (59.02%)	19 (70.37%)	17 (50.00%)	13 (65.00%)	13 (50.00%)	10 (66.67%)
≥30 U/ml	25 (40.98%)	8 (29.63%)	17 (50.00%)	7 (35.00%)	13 (50.00%)	5 (33.33%)
SUV-max	12.2800±5.6400	12.6900±5.5500	11.9474±5.7745	11.9530±5.1447	11.8777±5.8017	13.3967±6.2133

Abbreviations: CA19-9: Carbohydrate antigen 19-9; CEA: Carcinoembryonic antigen; HER2: Human epidermal growth factor receptor 2; LM−: No metastasis; LM+: Lymph node metastasis; LM++: Distant lymph node metastasis; SUV: Standardized uptake value.

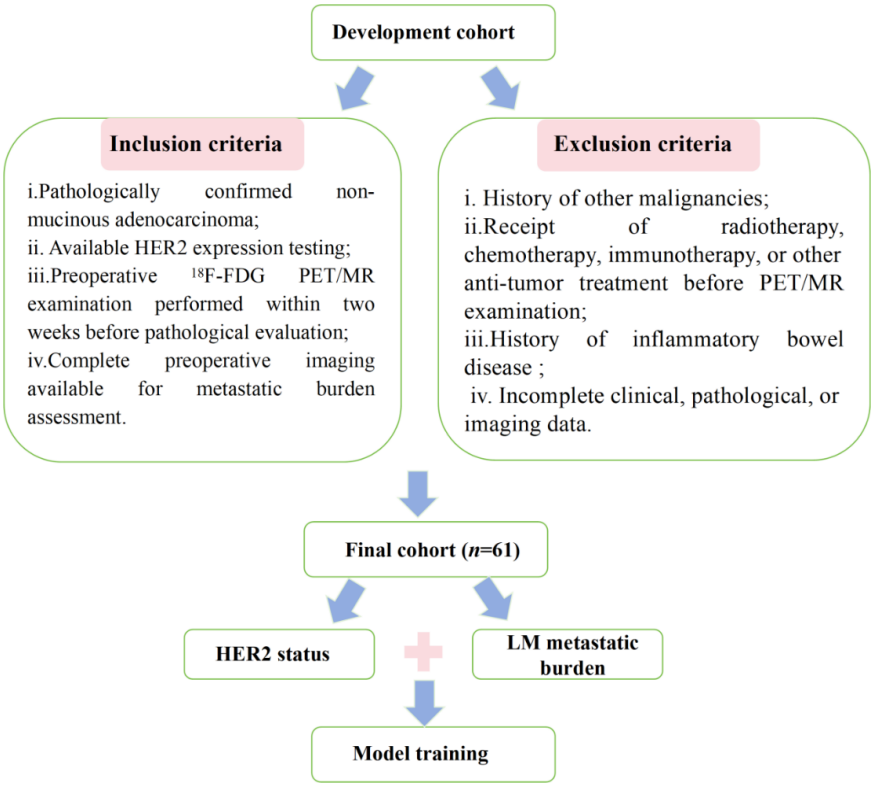


Figure 1. Schematic presentation of the research population. The flow chart demonstrates the selection criteria and patient disposition in the CRC-MMP. Abbreviations: CRC-MMP: Colorectal cancer molecular and metastatic profiling; FDG: Fluorodeoxyglucose; HER2: Human epidermal growth factor receptor 2; LM: Lymph node metastasis; MR: Magnetic resonance; PET: Positron emission tomography.

confined to locoregional draining lymph nodes according to the TNM classification; and LM++, distant LM involving non-regional lymph nodes beyond the primary lymphatic drainage area. All imaging findings were independently reviewed by experienced radiologists and confirmed by multidisciplinary consensus.

2.3. Image preprocessing and feature extraction

For each patient's primary tumor lesion, semi-automated radiomic feature extraction was performed using the LIFEx software (v25.06.1; Institut Curie, INSERM, and CNRS, France; https://lifexsoft.org).[23] A nuclear medicine physician delineated the volume of interest (VOIs) on transverse PET images based on a 40% standardized uptake value-max threshold, without access to the patients' clinical information. Crucially, the PET-derived VOIs were not blindly mapped to the MR images; rather, they served only as initial spatial references. The VOI boundaries were then manually corrected and refined slice-by-slice by the physicians directly on the high-resolution T1WI and T2WI images to ensure precise anatomical delineation, compensating for any potential motion, rectal filling, or bowel peristalsis (Figure S2). In cases with multiple VOIs, the largest three-dimensional volume was selected for analysis and independently confirmed by a second nuclear medicine physician with more than 25 years of diagnostic experience. Radiomic features were extracted using LIFEx, following the guidelines of the Imaging Biomarker Standardization Initiative. It should be clarified that while the VOI annotations are strictly required to define the spatial boundaries for feature calculation, the extracted feature types constitute a fixed, predefined set, independent of the specific VOI shapes; the VOIs solely determine the numerical values of these features. We independently extracted 2,264 quantitative features from each of the three modalities (T1WI, T2WI, and PET), yielding an initial pool of 6,792 features prior to selection. For each modality, the feature set comprised 450 first-order statistics, 14 shape features, and 1,800 texture features. The least absolute shrinkage and selection operator method was then applied separately to each modality to select the most informative features for subsequent modeling.

2.4. Feature analysis and reduction

To mitigate feature redundancy and the risk of overfitting in high-dimensional radiomics data, a multi-step feature selection pipeline was implemented:

(i) All features were standardized using *Z*-score normalization, followed by variance thresholding to remove low-information features and retain the top 500 with the highest variance.
(ii) Feature-outcome associations were evaluated using the Mann–Whitney *U* test (for sample sizes < 30) or independent-samples t-test (for sample sizes ≥ 30). Features with $p < 0.05$ were retained. If none met this threshold, the top 100 features with the lowest *p*-values were selected to ensure sufficient candidates.
(iii) Pearson correlation analysis was performed among retained features to reduce multicollinearity. For any pair of features with an absolute correlation coefficient > 0.9 ($|r| > 0.9$), one feature was removed.
(iv) Finally, the radiomics signature was constructed by retaining features with non-zero coefficients using least absolute shrinkage and selection operator logistic regression, with the optimal regularization parameter determined via 10-fold cross-validation.

It should be clarified that this entire feature selection pipeline was applied to the full dataset of each imaging modality independently prior to the 5-fold cross-validation model evaluation. The resulting modality-specific feature subsets were subsequently concatenated to construct the multimodal combined model. Through this sequential process, the high-dimensional feature set was compressed into a compact subset for downstream CRC-MMP development. All procedures were implemented in Python (Scikit-learn; v3.8, Python Software Foundation, Netherlands), and the selected features with their coefficients were exported to ensure reproducibility.

2.5. Model development and interpretability analysis

Extreme gradient boosting (XGBoost) was employed as the classification engine due to its robustness in handling high-dimensional radiomics data and capturing complex nonlinear feature interactions,[24] while effectively mitigating overfitting through built-in L1/L2 regularization. Hyperparameters were optimized via grid search coupled with 5-fold cross-validation. The specific search space was defined as follows: n_estimators ∈ {200, 300, 500}, max_depth ∈ {3, 4, 5, 6}, learning_rate ∈ {0.01, 0.05, 0.1}, reg_alpha (L1) ∈ {0, 0.1, 1}, and reg_lambda (L2) ∈ {1, 5, 10}. The scoring metric was set to F1 (or F1_macro) for the binary HER2 task, and F1_weighted for the multiclass LM task to ensure balanced evaluation across uneven class sizes. The optimal parameter combination was selected based on the highest cross-validated F1-score. To ensure strict reproducibility, a fixed random_state (42) was applied to all stochastic processes, and verbosity was set to 0. All models were implemented in Python using the XGBoost library and Scikit-learn.

SHapley Additive exPlanations enhances model interpretability by quantifying the impact of each feature on individual predictions.[25] SHAP quantifies each feature's

marginal contribution to the model's prediction relative to a baseline (typically the dataset's mean expected value). This approach assigns meaningful values to features, offering a clear explanation of how each variable influences the final prediction compared to the baseline. Specifically, SHAP values provide a quantitative assessment of feature importance by computing the expected change in model output when a feature is included versus excluded, across all possible feature combinations. Positive SHAP values indicate that a feature contributes positively to the prediction outcome, while negative values suggest a suppressive effect.

3. Results

3.1. Multimodal framework CRC-MMP for predicting lymph node metastasis burden and HER2 status

This study aims to diagnose HER2 expression status and LM burden in CRC patients based on PET/MR radiomic features. The experimental workflow includes the following steps: (i) subject selection; (ii) data preprocessing: delineation of VOIs by experienced nuclear medicine physicians; (iii) extraction of radiomic features from T1WI, T2WI, and PET images; (iv) feature reduction using a multi-step feature selection pipeline; (v) construction of ML models based on the selected features and interpretability analysis (Figure 2).

The predictive performance of CRC-MMP was quantitatively evaluated using multiple metrics, including area under the curve (AUC), accuracy, and F1-score, and was further compared with 3 baseline models. All the results are the average values obtained from 5-fold cross-validation. In addition to performance evaluation, SHAP was employed to provide feature-level interpretability and to quantify the contribution of individual radiomic features to model predictions.

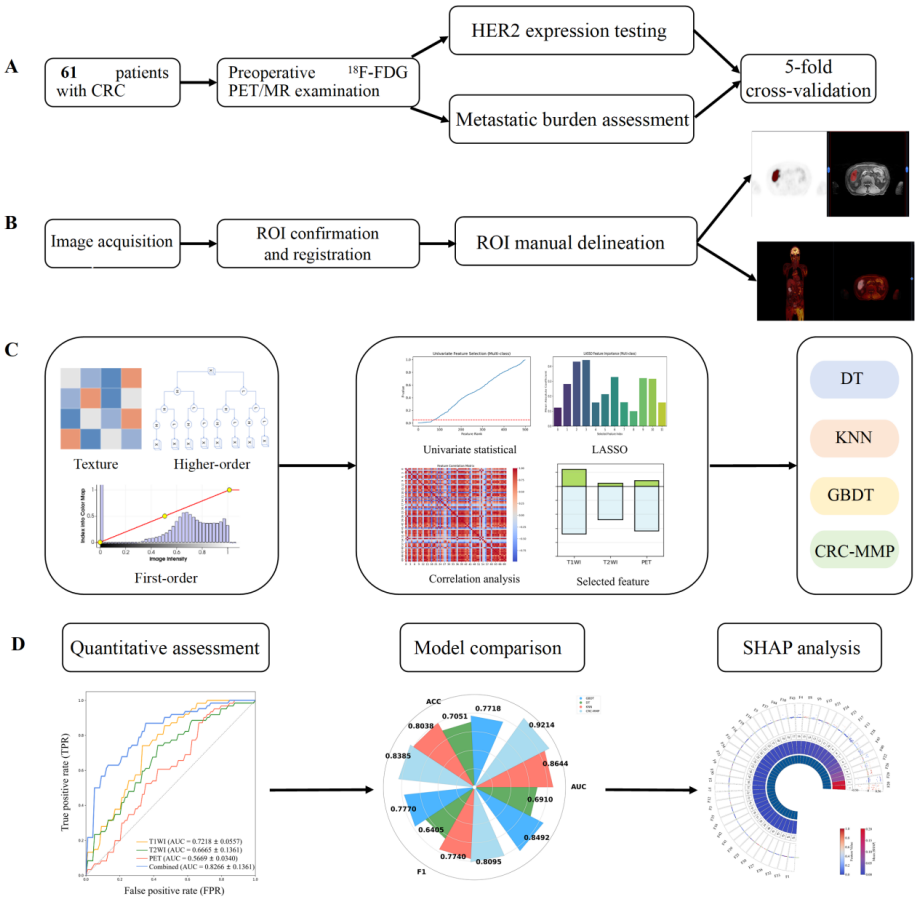


Figure 2. Overview of the multimodal and multitask CRC-MMP framework. (A) Dataset. (B) Image preprocessing and segmentation. (C) Feature extraction and model development. (D) Evaluation and interpretability analysis.
Abbreviations: CRC-MMP: Colorectal cancer molecular and metastatic profiling; DT: Decision tree; FDG: Fluorodeoxyglucose; GBDT: Gradient boosting decision tree; HER2: Human epidermal growth factor receptor 2; KNN: K-nearest neighbors; LASSO: Least absolute shrinkage and selection operator; ROI: Region of interest; SHAP: SHapley Additive exPlanations.

To mitigate feature redundancy and enhance model generalization, a dimensionality reduction strategy was applied prior to model construction. The number of retained features varied across prediction tasks. For the LM prediction task, 17 T1WI features, 12 T2WI features, and 16 PET features were ultimately selected. For the HER2 status prediction task, the final feature set comprised six T1WI features, one T2WI feature, and two PET features. Detailed information on the selected radiomic features is provided in Table S1 and Figure S1. To ensure robust uncertainty estimation, 95% confidence intervals for all performance metrics were derived via 1,000-iteration bootstrap resampling.

For the prediction of LM burden, the combined model again outperformed all single-modality models, achieving a mean AUC of 0.8266 ± 0.1361 (Table 2 and Figure 3). Among the single-modality models, the T1WI-based model demonstrated the highest discriminative ability (AUC: 0.7218 ± 0.0557), followed by the T2WI-based model (AUC: 0.6665 ± 0.1361), whereas the PET-based model yielded the lowest performance (AUC: 0.5669 ± 0.0340).

The multimodal combined model achieved the optimal performance for HER2 status prediction, yielding a mean AUC of 0.9214 ± 0.0779. Among the single-modality

Table 2. Diagnosis results for different modalities (mean ± standard deviation [95% confidence interval])

Task	Modality	AUC	ACC	F1
LM	PET	0.5669 ± 0.0340 [0.489, 0.673]	0.3782 ± 0.0990 [0.246, 0.508]	0.3510 ± 0.1073 [0.248, 0.513]
	T1WI	0.7218 ± 0.0557 [0.631, 0.815]	0.4744 ± 0.0998 [0.344, 0.591]	0.4544 ± 0.0859 [0.342, 0.600]
	T2WI	0.6665 ± 0.1361 [0.544, 0.769]	0.4295 ± 0.1168 [0.312, 0.557]	0.4039 ± 0.1550 [0.315, 0.555]
	Combined	0.8266 ± 0.1361 [0.728, 0.896]	0.6269 ± 0.1631 [0.508, 0.738]	0.6071 ± 0.1850 [0.499, 0.739]
HER2	PET	0.8599 ± 0.1111 [0.714, 0.925]	0.7385 ± 0.0643 [0.623, 0.853]	0.6710 ± 0.1297 [0.533, 0.828]
	T1WI	0.7676 ± 0.0806 [0.633, 0.873]	0.6897 ± 0.1768 [0.574, 0.787]	0.6280 ± 0.2435 [0.465, 0.778]
	T2WI	0.5800 ± 0.1834 [0.396, 0.698]	0.5397 ± 0.1163 [0.361, 0.623]	0.4550 ± 0.2126 [0.204, 0.542]
	Combined	0.9214 ± 0.0779 [0.823, 0.975]	0.8385 ± 0.1090 [0.738, 0.918]	0.8095 ± 0.1151 [0.679, 0.909]

Abbreviations: ACC: Accuracy; AUC: Area under the curve; HER2: Human epidermal growth factor receptor 2; LM: Lymph node metastasis; PET: Positron emission tomography; T1WI: T1-weighted imaging; T2WI: T2-weighted imaging.

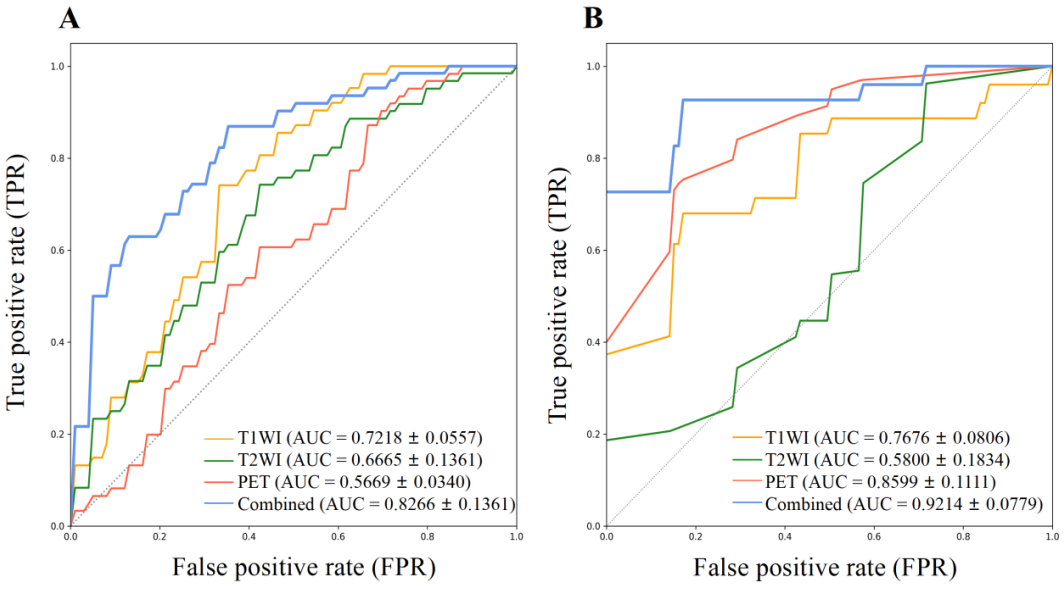


Figure 3. The receiver operating characteristic curves for different modalities across (A) LM burden and (B) HER2 status prediction
Abbreviations: AUC: Area under the curve; HER2: Human epidermal growth factor receptor 2; LM: Lymph node metastasis; PET: Positron emission tomography; T1WI: T1-weighted imaging; T2WI: T2-weighted imaging.

models, the PET-based model demonstrated the strongest discriminative capability, with an AUC of 0.8599 ± 0.1111. The T1WI-based model also showed moderate predictive value (AUC: 0.7676 ± 0.0806). In contrast, the T2WI-based model exhibited the lowest performance among the three single modalities (AUC: 0.5800 ± 0.1834).

3.2. Comparative evaluation with other baseline models

To contextualize the performance of the proposed CRC-MMP framework, its predictive metrics were compared against three conventional machine learning baseline models: Gradient Boosting Decision Tree (GBDT), K-nearest neighbors (KNN), and decision tree (DT).[26-29]

The CRC-MMP again exhibited the most favorable overall classification performance for the LM burden prediction, achieving a macro-average AUC of 0.8266 (Table 3 and Figure 4). It slightly surpassed the GBDT model (AUC = 0.8176, ΔAUC = 0.0090), which was the strongest competitor in this multiclass task. More importantly, CRC-MMP showed substantial AUC improvements over KNN (AUC = 0.7507) and DT (AUC = 0.6809). Although GBDT achieved marginally higher accuracy (0.6564 vs. 0.6269) and F1-score (0.6387 vs. 0.6071) under the default Argmax decision rule, the superior AUC of CRC-MMP indicated its better overall probability estimation and global discriminative ability across the three distinct LM classes.

In predicting HER2 status, the CRC-MMP model achieved the highest discriminative performance (AUC = 0.9214), surpassing the best-performing baseline, KNN

Table 3. Performance comparison of different baseline models

Task	Model	AUC	ACC	F1
LM	GBDT	0.8176	0.6564	0.6387
	DT	0.6809	0.5782	0.5605
	KNN	0.7507	0.5603	0.5002
	CRC-MMP	0.8266	0.6269	0.6071
HER2	GBDT	0.8492	0.7718	0.7770
	DT	0.6910	0.7051	0.6405
	KNN	0.8644	0.8038	0.7740
	CRC-MMP	0.9214	0.8385	0.8095

Abbreviations: CRC-MMP: Colorectal cancer molecular and metastatic profiling; DT: Decision tree; GBDT: Gradient boosting decision tree; HER2: Human epidermal growth factor receptor 2; KNN: K-nearest neighbors; LM: Lymph node metastasis.

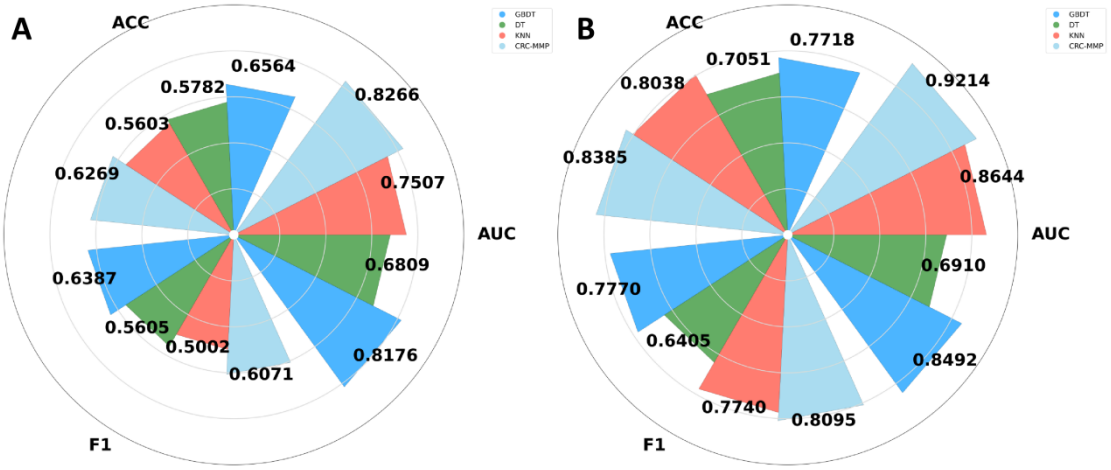


Figure 4. Radar chart comparing AUC, ACC, and F1-score metrics for different baseline models in (A) LM burden and (B) HER2 status prediction
Abbreviations: ACC: Accuracy; AUC: Area under the curve; CRC-MMP: Colorectal cancer molecular and metastatic profiling; DT: Decision tree; GBDT: Gradient boosting decision tree; HER2: Human epidermal growth factor receptor 2; KNN: K-nearest neighbors; LM: Lymph node metastasis.

(AUC = 0.8644), by an absolute improvement of 0.0570. Furthermore, it significantly outperformed GBDT (AUC = 0.8492) and DT (AUC = 0.6910). Consistent with its AUC advantage, CRC-MMP also secured the optimal accuracy (0.8385) and F1-score (0.8095), demonstrating its robust and comprehensive classification capability for molecular status characterization.

3.3. Model explanation with SHapley Additive exPlanations

The SHAP analysis was performed on the final combined XGBoost model trained on the entire dataset to provide a global explanation of feature importance. To mitigate the impact of feature collinearity on SHAP attribution, highly correlated features ($|r| > 0.9$) had been rigorously removed during the prior feature selection phase. SHAP analysis revealed that the HER2 prediction model was predominantly driven by texture-related radiomic features (Figure 5). The highest-ranking features included F1 (T1-Gray-Level Run Length Matrix [GLRLM]-wavelet), F2 (T1-First-order-wavelet), F4 (T1-Gray-Level Co-occurrence Matrix [GLCM]-log), and F7 (T1-GLCM-wavelet), primarily derived from the first-order, GLCM, and GLRLM matrices. These findings indicate that intratumoral texture heterogeneity, particularly multiscale texture patterns extracted from filtered images, played a major role in distinguishing HER2 expression status.

For LM burden prediction, SHAP analysis identified F20 (T2-First-order-log), F22 (T2-Neighborhood Gray-Tone Difference Matrix [NGTDM]-log), F26 (T2-First-order-Kur), and F29 (T2-Gray-Level Dependence Matrix [GLDM]-SD) as the most influential features (Figure 5). These features were mainly derived from the First-order, GLDM, and NGTDM texture matrices, suggesting that spatial heterogeneity and local structural complexity captured by multiscale texture descriptors contributed substantially to differentiating metastatic burden across the three clinical categories.

4. Discussion

This study developed and internally validated a multimodal predictive framework, termed CRC-MMP, based on 18F-FDG PET/MR radiomics, for the simultaneous preoperative prediction of HER2 status and LM burden in CRC from a single imaging examination. By integrating radiomic features extracted from PET/MR and incorporating machine learning algorithms, CRC-MMP demonstrated favorable predictive performance in both HER2 status classification and LM burden assessment. Furthermore, SHAP-based explainability analysis identified the key radiomic features driving both predictive tasks. The results indicate that CRC-MMP enables the concurrent provision of tumor molecular characteristics and LM metastatic burden information from a single ^{18}F-FDG PET/MR examination, thereby offering supplementary evidence to support clinical TNM staging evaluation and individualized therapeutic decision-making.

Integrated PET/MR imaging combines the superior soft-tissue resolution of MRI with the high sensitivity of PET for tumor metabolic activity, thereby enabling a more

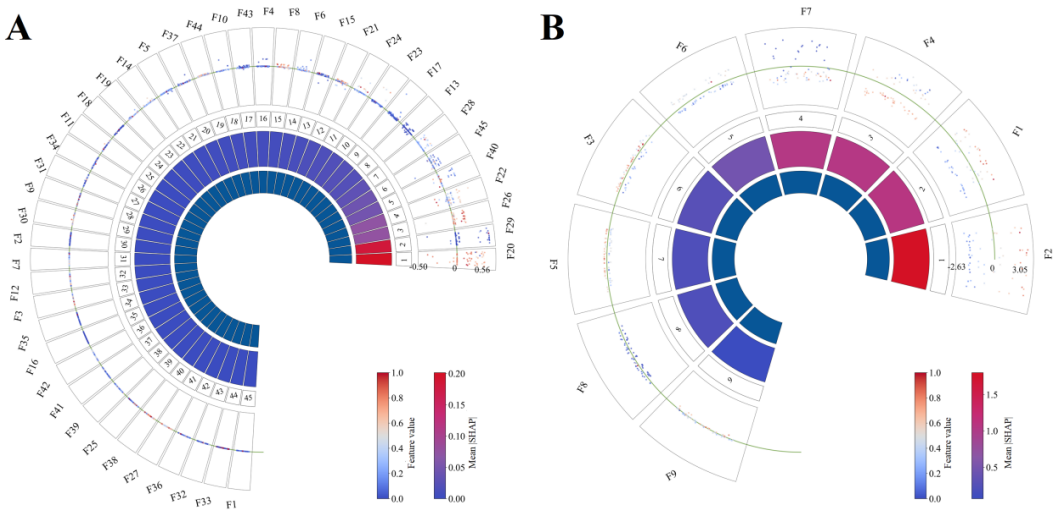


Figure 5. SHapley Additive exPlanations (SHAP) analysis results for (A) lymph node metastasis burden and (B) human epidermal growth factor receptor 2 status prediction

comprehensive characterization of tumor biological features from both anatomical and functional perspectives. In this study, for both HER2 status prediction and LM burden assessment, the multimodal joint model significantly outperformed any unimodal model (PET, T1WI, and T2WI-only), suggesting that different modalities provide highly complementary information. This finding aligns with previous multimodal radiomics studies, further underscoring the value of PET/MR-derived multimodal data fusion in complex clinical prediction tasks.

Although both HER2 status and LM burden were derived from the same PET/MR imaging, the key radiomic features underlying the two tasks exhibited notable discrepancies. SHAP analysis revealed that HER2 status prediction relied predominantly on first-order, GLCM, and GLRLM features, which primarily reflect tumor gray-level distribution and local textural heterogeneity. In contrast, LM burden prediction was more dependent on multiscale texture features, including first-order, GLDM, and NGTDM, which more comprehensively characterize the complex spatial organizational patterns and textural architectures within the tumor. This divergence suggests that the two tasks capture distinct imaging phenotypes within the tumor. Notably, the primary-tumor texture features predicting LM burden do not merely describe local anatomy; rather, they capture the intrinsic aggressive behaviors (e.g., spatial heterogeneity and microenvironmental remodeling) that drive metastatic dissemination, serving as a biological footprint of metastatic potential. However, it is crucial to emphasize that SHAP analysis provides mathematical interpretability rather than direct biological proof. The identified radiomic patterns represent statistical associations with HER2 expression and metastatic burden, rather than direct evidence of specific biological mechanisms. While these texture features may conceptually correlate with underlying cellular proliferative activity, microenvironmental alterations, or invasive potential, their exact biological substrates require further validation through multi-omics profiling and detailed histopathological correlation.

These findings highlight the promising role of PET/MR radiomics as a non-invasive adjunct for comprehensive tumor profiling, paving the way for more personalized management in CRC. It is crucial to emphasize that radiomics cannot replace tissue-based HER2 testing, as targeted therapy requires analytically validated pathological assessment. CRC-MMP enables the preoperative concurrent acquisition of HER2 molecular status and LM burden information, thereby supporting integrated precision diagnostics and therapeutics for CRC patients. Accurate identification of HER2-positive patients facilitates the selection of potential candidates who may benefit from HER2-targeted therapies, thereby aiding in the formulation of individualized first-line treatment strategies. Meanwhile, accurate preoperative assessment of LM burden contributes to patient risk stratification, surgical planning optimization, and evaluation of radiotherapy indications. Compared with conventional evaluation workflows that rely on histopathological testing or multiple imaging examinations, PET/MR radiomics allows the simultaneous acquisition of tumor molecular characteristics and metastatic information through a single non-invasive examination. While these findings serve as an exploratory proof-of-concept, they highlight the potential of this approach as a non-invasive adjunct tool for the comprehensive assessment of tumor biological behavior, pending further rigorous clinical validation.

Despite these promising results, several limitations should be acknowledged. First, the retrospective, single-center design and modest sample size—particularly within the metastatic subgroups—may limit the broader generalizability of our findings. The cohort was inherently enriched for patients who underwent preoperative PET/MR and had available HER2 testing. Because HER2 testing is often selectively ordered for advanced or refractory cases in clinical practice, this introduces spectrum and selection bias, resulting in a higher HER2-positive rate than expected in an unselected CRC population. Therefore, the generalizability of the model to a broader, unselected CRC population requires further verification through large-scale, prospective, multicenter studies. Second, the feature selection pipeline was applied to the entire dataset prior to cross-validation. This constitutes a form of data leakage, as information from the test folds inevitably influenced the feature selection process, potentially introducing optimistic bias. Given the computational instability of strict nested cross-validation in our highly imbalanced, small-sample cohort, this methodological limitation must be explicitly acknowledged. Furthermore, the lack of an independent multicenter external validation cohort necessitates future large-scale studies to verify the model's stability and generalizability. Third, the retrospective nature of this cohort precluded the systematic collection of several key variables, including precise tumor location, comprehensive molecular profiling (e.g., *KRAS*/*BRAF*/microsatellite instability), detailed HER2 IHC scores, and macroscopic PET parameters (e.g., metabolic tumor volume, total lesion glycolysis), thereby limiting the construction of a clinical-radiomics nomogram. Furthermore, the reference standard for metastatic burden relied predominantly on preoperative imaging rather than strict pathological confirmation. Although multidisciplinary consensus was applied, this introduces potential incorporation bias. Future

prospective studies with systematic multiparametric data collection and strict pathological reference standards are warranted to address these limitations. Finally, although SHAP analysis enhanced the interpretability of the model and revealed the radiomic features that contributed most significantly to prediction, the underlying biological mechanisms linking these imaging phenotypes to HER2 expression and tumor metastasis have not been fully elucidated. To further enhance the model's generalizability, robustness, and clinical utility, future studies should incorporate large-scale, multicenter cohorts and integrate multi-omics data with clinical variables. Although radiomics captures intratumoral imaging heterogeneity, cancer prognosis and metastatic behavior may also be influenced by systemic biological factors that are not directly represented in the primary-tumor image.[30] Although CRC was not included in that study and causality cannot be inferred, these observations illustrate the potential value of combining quantitative imaging with circulating metabolic, nutritional, inflammatory, and molecular biomarkers. Future CRC-MMP development could therefore assess whether the integration of clinical and serum-marker data improves prediction beyond PET/MR radiomics alone, while recognizing that such systemic measurements remain observational and are not yet ready for routine CRC clinical decision-making.

5. Conclusion

In conclusion, the CRC-MMP framework enables simultaneous characterization of complementary dimensions of CRC, encompassing both molecular status and metastatic burden, through the integration of multimodal 18F-FDG PET/MR radiomic features. Furthermore, the SHAP-based explainability analysis identified the key radiomic features driving model predictions, substantially enhancing the transparency of the model. Collectively, these findings indicate that the combination of multimodal PET/MR radiomic data with machine learning techniques, allowing comprehensive tumor phenotyping via a single non-invasive imaging examination, provides valuable assessment information for preoperative risk stratification and first-line clinical decision-making in CRC.

Acknowledgments

None.

Funding

This work was supported by the construction project of the Shanghai Key Laboratory of Molecular Imaging (18DZ2260400).

Conflict of interest

The authors declare no conflict of interest.

Author contributions

Conceptualization: Zengbei Yuan
Data curation: Zengbei Yuan, Kun Wang, Wen Jiang, Yihong Yang, Zirong Zhou, Jun Zhao
Formal analysis: Zhixian Tang, Xing Chen, Weilun Wang
Funding acquisition: Qingqing Huang, Na Qi
Investigation: Zengbei Yuan, Kun Wang, Wen Jiang
Methodology: Qingqing Huang, Na Qi
Supervision: Qingqing Huang, Na Qi
Validation: Zhixian Tang, Xing Chen, Weilun Wang
Visualization: Yihong Yang, Zirong Zhou, Jun Zhao
Writing–original draft: Zengbei Yuan, Kun Wang, Wen Jiang
Writing–review & editing: Qingqing Huang, Na Qi

Ethics approval and consent to participate

This study was approved by the Institutional Review Board of Shanghai East Hospital (IRB number: No.2024-LYS-098, Approval Date: Aug 2024) and performed in accordance with the principles of the 1964 Declaration of Helsinki and its later amendments. The written informed consent was waived by the Institutional Ethics Review Board due to the retrospective nature of the study.

Consent for publication

Not applicable.

Availability of data

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

References

1. Chen N, He L, Zou Q, Deng H. HER2 targeted therapy in colorectal Cancer: Current landscape and future directions. *Biochem Pharmacol.* 2024;223:116101.

   doi: 10.1016/j.bcp.2024.116101

2. Zhang R, Su C, Jia Y, Xing M, Jin S, Zong H. Molecular mechanisms of HER2-targeted therapy and strategies to overcome the drug resistance in colorectal cancer. *Biomed Pharmacother.* 2024;179:117363.

   doi: 10.1016/j.biopha.2024.117363

3. Takegawa N, Yonesaka K. HER2 as an Emerging Oncotarget for Colorectal Cancer Treatment After Failure of Anti-Epidermal Growth Factor Receptor Therapy. *Clin Colorectal Cancer.* 2017;16:247-251.

   doi: 10.1016/j.clcc.2017.03.001

4. Guarini C, Grassi T, Pezzicoli G, Porta C. Beyond RAS and BRAF: HER2, a New Actionable Oncotarget in Advanced Colorectal Cancer. *Int J Mol Sci.* 2021;22(13):6813.

doi: 10.3390/ijms22136813

5. Yoshikawa A, Nakamura Y. Molecular Basis of HER2-Targeted Therapy for HER2-Positive Colorectal Cancer. *Cancers.* 2022;15(1):183.

doi: 10.3390/cancers15010183

6. Jin M, Frankel WL. Lymph Node Metastasis in Colorectal Cancer. *Surg Oncol Clin N Am.* 2018;27(2):401-412.

doi: 10.1016/j.soc.2017.11.011

7. Lee DH, Lee JM. Whole-body PET/MRI for colorectal cancer staging: Is it the way forward? *J Magn Reson Imaging.* 2017;45(1):21-35.

doi: 10.1002/jmri.25337

8. Fraum TJ, Owen JW, Fowler KJ. Beyond Histologic Staging: Emerging Imaging Strategies in Colorectal Cancer with Special Focus on Magnetic Resonance Imaging. *Clin Colon Rectal Surg.* 2016;29(3):205-215.

doi: 10.1055/s-0036-1584288

9. Shen G, Hu S, Liu B, Kuang A. Diagnostic Performance of Whole-Body PET/MRI for Detecting Malignancies in Cancer Patients: A Meta-Analysis. *PLoS ONE.* 2016;11(4):e0154497.

doi: 10.1371/journal.pone.0154497

10. Guglielmo P, Marturano F, Bettinelli A, Gregianin M, Paiusco M, Evangelista L. Additional Value of PET Radiomic Features for the Initial Staging of Prostate Cancer: A Systematic Review from the Literature. *Cancers.* 2021;13(23):6026.

doi: 10.3390/cancers13236026

11. Vuong D, Tanadini-Lang S, Huellner MW, *et al.* Interchangeability of radiomic features between [18F]-FDG PET/CT and [18F]-FDG PET/MR. *Med Phys.* 2019;46(4):1677-1685.

doi: 10.1002/mp.13422

12. Palumbo B, Bianconi F, Palumbo I, *et al.* Value of Shape and Texture Features from 18F-FDG PET/CT to Discriminate between Benign and Malignant Solitary Pulmonary Nodules: An Experimental Evaluation. *Diagnostics.* 2020;10(9):696.

doi: 10.3390/diagnostics10090696

13. Zhao Q, Shi C-Z, Luo L-P. Role of the texture features of images in the diagnosis of solitary pulmonary nodules in different sizes. *Chin J Cancer Res.* 2014;26:451-458.

doi: 10.3978/j.issn.1000-9604.2014.08.07

14. Vallières M, Freeman CR, Skamene SR, El Naqa I. A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. *Phys Med Biol.* 2015;60(4):5471-5496.

doi: 10.1088/0031-9155/60/14/5471

15. Mirshahvalad SA, Hinzpeter R, Kohan A, *et al.* Diagnostic performance of [18F]-FDG PET/MR in evaluating colorectal cancer: a systematic review and meta-analysis. *Eur J Nucl Med Mol Imaging.* 2022;49(12):4205-4217.

doi: 10.1007/s00259-022-05871-0

16. Feng L, Liu Z, Li C, *et al.* Development and validation of a radiopathomics model to predict pathological complete response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer: a multicentre observational study. *Lancet Digit Health.* 2022;4(1):e8-e17.

doi: 10.1016/S2589-7500(21)00215-6

17. Xue T, Peng H, Chen Q, Li M, Duan S, Feng F. Preoperative prediction of KRAS mutation status in colorectal cancer using a CT-based radiomics nomogram. *Br J Radiol.* 2022;95(1134):20211014.

doi: 10.1259/bjr.20211014

18. Xue T, Peng H, Chen Q, Li M, Duan S, Feng F. Preoperative Prediction of BRAF Mutation Status in Colorectal Cancer Using a Clinical-radiomics Model. *Acad Radiol.* 2022;29(9):1298-1307.

doi: 10.1016/j.acra.2021.12.016

19. Golia Pernicka JS, Gagniere J, Chakraborty J, *et al.* Radiomics-based prediction of microsatellite instability in colorectal cancer at initial computed tomography evaluation. *Abdom Radiol (NY).* 2019;44(11):3755-3763.

doi: 10.1007/s00261-019-02117-w

20. Wang H, Chen X, Ding J, *et al.* Novel multiparametric MRI-based radiomics in preoperative prediction of perirectal fat invasion in rectal cancer. *Abdom Radiol (NY).* 2023;48(2):471-485.

doi: 10.1007/s00261-022-03759-z

21. Valtorta E, Martino C, Sartore-Bianchi A, *et al.* Assessment of a HER2 scoring system for colorectal cancer: results from a validation study. *Mod Pathol.* 2015;28(11):1481-1491.

doi: 10.1038/modpathol.2015.98

22. Oliva MR, Saini S. Liver cancer imaging: role of CT, MRI, US and PET. *Cancer Imaging.* 2004;4(suppl A):S42-S46.

doi: 10.1102/1470-7330.2004.0011

23. Nioche C, Orlhac F, Boughdad S, *et al.* LIFEx: A Freeware for Radiomic Feature Calculation in Multimodality Imaging to Accelerate Advances in the Characterization of Tumor Heterogeneity. *Cancer Res.* 2018;78(16):4786-4789.

doi: 10.1158/0008-5472.CAN-18-0125

24. Zhang Y, Zhou X, Li P, Li C, Ke L. XGBoost-based model for predicting five-year survival in gastric cancer using clinical indicators. *Sci Rep.* 2026;16(1).

doi: 10.1038/s41598-026-50043-x

25. Li T, Qi W, Mao X, *et al.* Prediction of lumbar disc degeneration based on interpretable machine learning models: retrospective cohort study. *Spine J.* 2025;25(10):2335-2348.

doi: 10.1016/j.spinee.2025.04.004

26. Seto H, Oyama A, Kitora S, *et al.* Gradient boosting decision tree becomes more reliable than logistic regression in predicting probability for diabetes with big data. *Sci Rep.* 2022;12(1):15889.

doi: 10.1038/s41598-022-20149-z

27. Srisuradetchai P, Suksrikran K. Random kernel k-nearest neighbors regression. *Front Big Data.* 2024;7:1402384.

doi: 10.3389/fdata.2024.1402384

28. Narayanan H, Sokolov M, Butté A, Morbidelli M. Decision Tree-PLS (DT-PLS) algorithm for the development of process: Specific local prediction models. *Biotechnol Prog.* 2019;35(4).

doi: 10.1002/btpr.2818

29. Luo X, Wen X, Zhou M, Abusorrah A, Huang L. Decision-Tree-Initialized Dendritic Neuron Model for Fast and Accurate Data Classification. *IEEE Trans Neural Netw Learn Syst.* 2022;33(9):4173-4183.

doi: 10.1109/TNNLS.2021.3055991

30. Lubiński J, Lener MR, Marciniak W, *et al.* Serum Essential Elements and Survival after Cancer Diagnosis. *Nutrients.* 2023;15(11):2611.

doi: 10.3390/nu15112611