


RESEARCH ARTICLE

Open Access



# Using tumor habitat-derived radiomic analysis during pretreatment $^{18}\text{F}$ -FDG PET for predicting KRAS/NRAS/BRAF mutations in colorectal cancer

Hongyue Zhao<sup>1†</sup>, Yexin Su<sup>1†</sup>, Yan Wang<sup>1</sup>, Zhehao Lyu<sup>1</sup>, Peng Xu<sup>1</sup>, Wenchao Gu<sup>3</sup>, Lin Tian<sup>2\*</sup> and Peng Fu<sup>1\*</sup> 

## Abstract

**Background** To investigate the association between Kirsten rat sarcoma viral oncogene homolog (KRAS) / neuroblastoma rat sarcoma viral oncogene homolog (NRAS) /v-raf murine sarcoma viral oncogene homolog B (BRAF) mutations and the tumor habitat-derived radiomic features obtained during pretreatment  $^{18}\text{F}$ -fluorodeoxyglucose (FDG) positron emission tomography (PET) in patients with colorectal cancer (CRC).

**Methods** We retrospectively enrolled 62 patients with CRC who had undergone  $^{18}\text{F}$ -FDG PET/computed tomography from January 2017 to July 2022 before the initiation of therapy. The patients were randomly split into training and validation cohorts with a ratio of 6:4. The whole tumor region radiomic features, habitat-derived radiomic features, and metabolic parameters were extracted from  $^{18}\text{F}$ -FDG PET images. After reducing the feature dimension and selecting meaningful features, we constructed a hierarchical model of KRAS/NRAS/BRAF mutations by using the support vector machine. The convergence of the model was evaluated by using learning curve, and its performance was assessed based on the area under the receiver operating characteristic curve (AUC), calibration curve, and decision curve analysis. The SHapley Additive exPlanation was used to interpret the contributions of various features to predictions of the model.

**Results** The model constructed by using habitat-derived radiomic features had adequate predictive power with respect to KRAS/NRAS/BRAF mutations, with an AUC of 0.759 (95% CI: 0.585–0.909) on the training cohort and that of 0.701 (95% CI: 0.468–0.916) on the validation cohort. The model exhibited good convergence, suitable calibration, and clinical application value. The results of the SHapley Additive explanation showed that the peritumoral habitat and a high\_metabolism habitat had the greatest impact on predictions of the model. No meaningful whole tumor region radiomic features or metabolic parameters were retained during feature selection.

<sup>†</sup>Hongyue Zhao and Yexin Su contributed equally to this work and share first authorship.

\*Correspondence:

Lin Tian  
tianlin6225108@163.com  
Peng Fu  
fupeng0451@163.com

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



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, 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 changes were made. 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/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

**Conclusion** The habitat-derived radiomic features were found to be helpful in stratifying the status of KRAS/NRAS/BRAF in CRC patients. The approach proposed here has significant implications for adjuvant treatment decisions in patients with CRC, and needs to be further validated on a larger prospective cohort.

**Keywords** Colorectal cancer, KRAS/NRAS/BRAF, Habitat, Radiomic,  $^{18}\text{F}$ -FDG PET

## Introduction

Around 800,000 people across the world succumb to colorectal cancer (CRC) every year. It is the third most common type of cancer, after lung and breast cancers [1]. Epidermal growth factor receptor (EGFR) signaling is a frequent event in cancer development [2]. The gene-encoding downstream effectors of the EGFR pathway in CRC include Kirsten rat sarcoma viral oncogene homolog (KRAS), neuroblastoma rat sarcoma viral oncogene homolog (NRAS) and *v*-raf murine sarcoma viral oncogene homolog B (BRAF) [3]. The KRAS and NRAS genes are both part of the rat sarcoma viral oncogene (Ras) family of small GTPases. KRAS mutations occur in the early stage of CRC, with an incidence of 30%–50%, while NRAS has an incidence of 3% of all cases [4, 5]. The BRAF gene, which belongs to the serine/threonine kinase family, is a proto-oncogene, and mutations in it occur in about 10% of CRC patients [4, 5]. The activation of oncogenic mutations in KRAS/NRAS/BRAF can result in constitutive Ras/Raf/MEK/ERK activation, where are key drivers of cancer growth in humans and indicate poor survival [6, 7]. In other words, mutations in Ras genes (such as KRAS and NRAS) or BRAF genes can activate downstream signaling pathways without EGFR inhibition, leading to tumor proliferation [8]. Accordingly, the National Comprehensive Cancer Network recommends that the KRAS/NRAS/BRAF genotypes in metastatic CRC patients be determined [9], and has approved both panitumumab and cetuximab as anti-EGFR monoclonal antibodies for the treatment of metastatic CRC caused by KRAS/NRAS/BRAF wild-type tumors [10, 11].

Previous studies have explored the mechanisms underlying the relationship between glucose accumulation and the status of KRAS/NRAS/BRAF mutation in CRC cells [12]. However, studies on the predictive value of the metabolic parameters have yielded different conclusions [4, 13]. The accumulation of  $^{18}\text{F}$ -fluorodeoxyglucose (FDG) in cancer tissues is a complex process that is influenced by various mechanisms, such as mutated TP53, HIF-1 $\alpha$  expression, and local inflammation, that can interfere with  $^{18}\text{F}$ -FDG uptake by CRC cells [12, 14, 15]. Therefore, simple metabolic parameters struggle to identify the metabolic heterogeneity in different regions within the tumor. Moreover, the values of the

most commonly used metabolic parameters in positron emission tomography (PET) vary with different factors, including the type of PET scanner used, metabolic differences between patients, and their blood glucose levels after fasting [4]. The recent emergence of radiomics technology, which can be used to extract microscale imaging-related information from standard medical images to predict complex genomic events, provides an alternative mechanism [16]. Nevertheless, a retrospective study of 151 patients with rectal cancer showed that textural features based on  $^{18}\text{F}$ -FDG PET/computed tomography (CT) have limited value in terms of predicting KRAS/NRAS/BRAF mutations [17]. Furthermore, the use of radiomics based on magnetic resonance to assess KRAS mutations does not always yield accurate results [18, 19]. Most radiomics studies usually analyze the tumor as a whole. However, this approach assumes that the tumor is heterogeneous but well mixed, thus, ignoring local phenotypic differences within the tumor [20]. In contrast to previously proposed methods, partitioning groups of voxels with a similar tumor biology into subregions for analysis—namely, habitat imaging—allows for better quantification of heterogeneity within tumors [21, 22]. Habitat-derived radiomics analysis in particular provides a better insight into the tumor heterogeneity than traditional radiomics analysis, in terms of the prognosis and predictions of the survival of patients of esophageal cancer who have been treated with concurrent chemoradiotherapy [23]. Hence, we hypothesize that this technique can be used to predict KRAS/NRAS/BRAF mutations in CRC patients and contribute to the development of modalities of non-invasive genomic detection.

Although mutation analysis using tissues of the tumor remains the standard of care [24], overcoming the limitations of spatial and temporal heterogeneity in the tumor based on the information provided by imaging-based studies has profound implications for adjunctive personalized molecular therapy. In this pilot study, we aim to determine whether habitat-derived radiomics analysis is an effective means to distinguish between mutant genomes and wild-type genomes in CRC patients, and to compare habitat-derived radiomics analysis with whole tumor region radiomics analysis and the metabolic parameters.

## Materials and methods

### Patients

The experimental protocols were approved by the Institutional Ethics Committee of the First Affiliated Hospital of Harbin Medical University, and the requirement of written informed consent from the patients was waived owing to the retrospective nature of this study. We collected data on patients with CRC who had undergone  $^{18}\text{F}$ -FDG PET/CT imaging and surgical treatment in the hospital from January 2017 to July 2022, for a total of 245 cases. As part of the analysis, we gathered clinical information on the patients and their pathological characteristics from medical records, including age, sex, tumor location, tumor differentiation, TNM stage, carcinoembryonic antigen (CEA), and KRAS/NRAS/BRAF status. The mutation analysis of KRAS (exons 2, 3, and 4), NRAS (exons 2, 3, and 4), and BRAF (V600E) was performed by using the amplification refractory mutation system–PCR methods in formalin-fixed and paraffin-embedded tumor samples. The criteria of exclusion were as follows: (1) patients who had received antineoplastic therapy (such as surgery, radiotherapy, and chemotherapy) before PET/CT examination, (2) no PET/CT scan or incomplete clinical information, (3) the interval between PET/CT imaging and genetic testing was longer than four weeks, (4) negative  $^{18}\text{F}$ -FDG uptake or volume of interest (VOI) measuring was smaller than 64 voxels, and (5) patients with other malignant tumors. A total of 62 CRC patients were finally considered for the analysis. They consisted of 41 males and 21 females, with a mean age of 63.06 years (range, 32–83 years). The cases were then randomly included in either a training cohort ( $n=37$ ) or a validation cohort ( $n=25$ ) by using stratified sampling with a 6:4 ratio.

### PET/CT imaging

Each patient had undergone  $^{18}\text{F}$ -FDG PET/CT examination on a Gemini GXL PET/CT scanner (Philips, Amsterdam, Netherlands) that was equipped with a 16-slice CT. An HM-12 cyclotron (Sumitomo Heavy Industries Ltd., Tokyo, Japan) was used for the production of  $^{18}\text{F}$ -FDG with a high radiochemical purity ( $\geq 95\%$ ). The patients were made to fast for at least 6 h before the intravenous injection of  $^{18}\text{F}$ -FDG (3.7~7.4 MBq/kg), and their intravenous blood glucose was controlled to below 8.0 mmol/L. Images were acquired  $60 \pm 5$  min after the injection of the  $^{18}\text{F}$ -FDG tracer. The patients were scanned from the middle of the thigh to the top of the skull according to the standard institutional clinical protocols. Whole-body scanning with a low-dose CT (tube current: 50 mA, tube voltage: 120 kV, slice thickness: 5 mm) was performed without an oral or an intravenous contrast agent, followed by PET imaging (1.5 min/bed

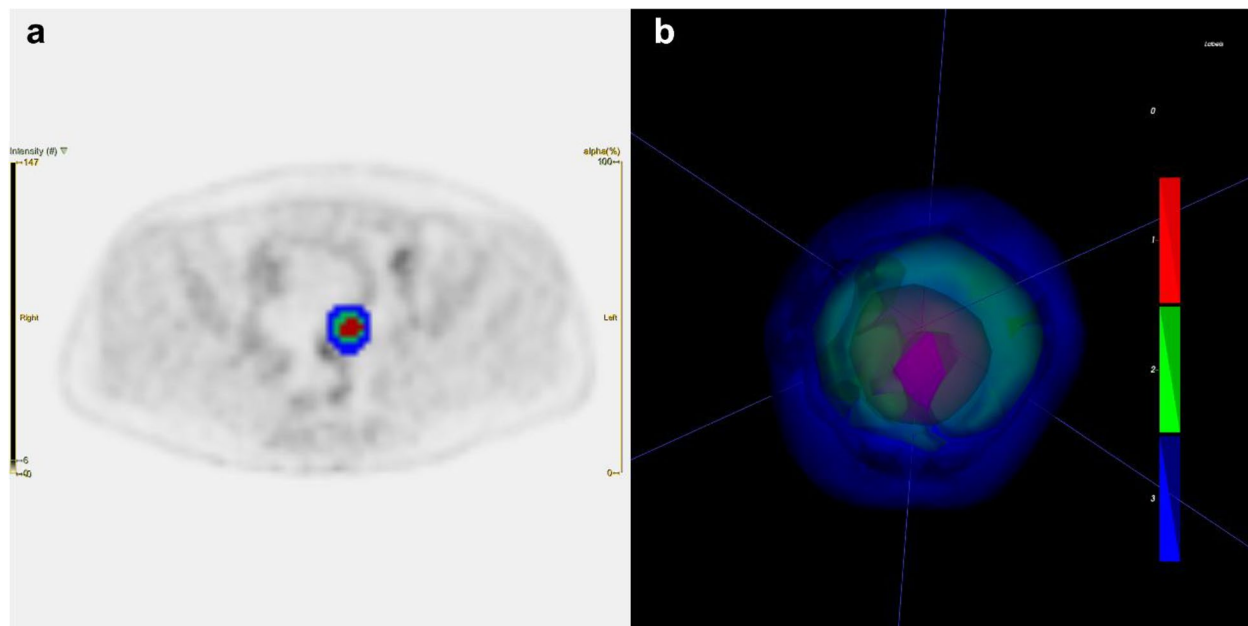
position, six to seven PET bed positions, 50% overlap between bed positions) in the same range. The data were reconstructed using the 3D-line of response RAMLA reconstruction algorithm that corrected for attenuation, attenuation decay, scatter, and dead time. The PET/CT images were fused using Syntegra V2.1 J software (Philip Medical Systems, Netherlands). Each reconstructed PET image had a matrix size of  $114 \times 114 \times 129$ , and a voxel size of  $4 \times 4 \times 4$  mm, while each reconstructed CT image had a matrix size of  $512 \times 512 \times 211$ , and a voxel size of  $1.172 \times 1.172 \times 5.000$  mm.

### Tumor segmentation and habitat generation

The acquired PET/CT images were anonymized. The PET images were normalized to injection doses of  $^{18}\text{F}$ -FDG and the body weight of the patients after decay correction to transform them into standardized uptake value (SUV) units. All images were observed in a range of threshold of SUV from 0 to 6, and the two-dimensional region of interest was manually segmented based on the boundary of the tumor lesion on each horizontal axial PET image with tumor to form a three-dimensional VOI. The VOI of the tumor was segmented by a physician experienced in nuclear medicine (Z.L.; 8 years of experience), who had been unaware of the status of mutation, by using LIFEX software (version 7.2.0; [www.lifexsoft.org](http://www.lifexsoft.org)) [25]. The Otsu threshold algorithm was implemented in Python (version 3.9.7; [www.python.org](http://www.python.org)). It can be used to count the number of pixels of each gray level in the VOI and identify an “optimal” threshold to maximize the inter-class variance [26, 27]. The corresponding VOI of the tumor in the PET images was divided into two significantly different hypermetabolic and hypometabolic regions representing different tumor habitats. We defined them as “low\_metabolism\_habitat” and “high\_metabolism\_habitat,” respectively. In addition, a ring-shaped peritumoral VOI was generated by expanding the margin of the lesion by 2 mm according to the outline of the segmented tumor, and special care was taken to avoid including other structures, especially the bladder and adjacent lymph nodes, by making additional modifications to the peritumoral VOI as needed. We explicitly considered the peritumoral tissue as a unique region of the tumor habitat to account for the relationship between mutations of genes of the tumor and the peritumoral regions. In this way, three subregions—low\_metabolism\_habitat, high\_metabolism\_habitat, and peritumoral—were obtained from the PET images, as shown in Fig. 1.

### Feature extraction

The data on each raw image were resampled to an isotropic resolution of  $1 \times 1 \times 1$  mm<sup>3</sup> by using the B-Spline interpolation method, and all VOIs defined in the raw



**Fig. 1** Schematic diagram of habitat generation. **a** Three subregions, low\_metabolism\_habitat (green area), high\_metabolism\_habitat (red area) and peritumoral (blue area) were obtained from the PET images. **b** A three-dimensional volume-rendered image of the three subregions. PET, positron emission tomography

imaging data were resampled by using the Nearest Neighbor interpolation on the data on isotropic resolution. The radiomic features were extracted by using PyRadiomics (version 3.0.1; <https://github.com/Radiomics/pyradiomics>), an open source radiomics toolkit [28]. A fixed bin width of 0.25 SUV was applied to discretize the continuous scale of SUV intensity and calculate the textural features [29, 30]. The extracted radiomic features ( $n=107$ ) were divided into seven categories: 14 shape features, 18 first-order statistics, 24 Gray-level co-occurrence matrix (glcm) features, 16 Gray-level size zone matrix (glszm) features, 16 Gray-level run length matrix (glrlm) features, 5 neighboring gray tone difference matrix (ngtdm) features, and 14 Gray-level dependence matrix (gldm) features. In addition, we applied a wavelet filter (wavelet-LHL, wavelet-LHH, wavelet-HLL, wavelet-LLH, wavelet-HLH, wavelet-HHH, wavelet-HHL, and wavelet-LLL) to the original images to obtain the corresponding derivative images in order to extract detailed high-dimensional radiomic features ( $n=744$ ) [31]. More interpretations are available from: <https://www.radiomics.io/pyradiomics.html>. A total of 851 features were thus calculated for each habitat, for a total of 2,553 ( $n=851 \times 3$ ) radiomic features across the three habitats. We repeated the above procedure to extract 851 radiomic features from the entire region of the tumor in each patient for the sake of comparison.

To assess the robustness and stability of the obtained radiomic features, the nuclear medicine physician (Z.L.) and another experienced radiologist (W.G.; 8 years of experience) independently performed repeated segmentation on 20 randomly selected cases after a one-month interval. This was done to calculate the intra-observer and inter-observer correlation coefficients (ICCs). Higher values of the ICCs reflected a higher degree of reproducibility, and features with ICCs below 0.75 were excluded from subsequent analyses.

For a comparison of predictive performance, eight metabolic parameters—minimal standardized uptake value (SUVmin), average standardized uptake value (SUVmean), maximal standardized uptake value (SUVmax), standard deviation of standardized uptake value (SUVstd), metabolic tumor volume (MTV), standardized metabolic tumor volume (sMTV), total lesion glycolysis (TLG), standardized total lesion glycolysis (sTLG) were also calculated by using voxels larger than the set threshold of 40% of SUVmax in the VOI of the tumor. The detailed formula has been provided on the relevant website ([www.lifexsoft.org](http://www.lifexsoft.org)).

#### Correlation analysis of features

Prior to feature selection and model construction, we performed the unsupervised clustering of the habitat-derived radiomic features, the whole tumor region radiomic features, and the metabolic parameters by using

the Euclidean distance. This allowed us to determine the differences between the patterns of global expression of features of the images in the presence and absence of KRAS/NRAS/BRAF mutations. Moreover, we visualized the correlation between the habitat-derived radiomic features and whole tumor region radiomic features, as well as that between the habitat-derived radiomic features and the metabolic parameters by using pairwise Pearson correlation coefficients to create correlation heatmaps. This enabled us to link different features of the images.

### Feature selection and prediction model development

To begin with, each feature was z-score normalized by using the mean and standard deviation of the training cohort features. The addition of a feature-ranking procedure beforehand can help improve the final performance [32]. We thus used the entropy-based maximum relevance minimum redundancy (mRMR) technique of feature selection in the supervised stage of feature selection. Entropy-based mRMR algorithms use mutual information and conditional entropy to measure the correlation and redundancy between features. Specifically, the algorithm first calculates the mutual information between each feature and the target variable, and then calculates the conditional entropy between the selected feature and other features, to select the optimal feature that has the maximum correlation with the target variable and the minimum redundancy with other features. The top 30 radiomic features were then fed into the least absolute shrinkage and selection operator (LASSO) algorithm for more refined feature selection. The optimal coefficient of lambda shrinkage within one standard error of this value was determined using ten-fold cross-validation. By penalizing the loss function with the L1 norm of the feature coefficients, LASSO forced those coefficients corresponding to weak features to become zero, thus preventing overfitting to produce a sparse matrix [33]. The underlying relationship between the radiomic phenotypic features and the genetic status is complex, and may even be non-linear. Moreover, the sample size in this study was small. Therefore, we used the support vector machine (SVM) to build a model of the classifier. The SVM is a statistical classifier model based on the principle of structural risk minimization that has been used to solve a range of non-linear problems involving a small number of samples [34, 35]. Five-fold cross-validation and random search with 1,000 iterations were used to determine the best parameter configuration during model training. The numbers of iterations and folds of cross-validation were determined based on factors such as the trade-off of bias and variance, as well as the actual computational resources. The same workflow was used for the habitat-derived radiomic features and the whole tumor region

radiomic features. However, owing to the small number of dimensions of their features, we used univariate analysis to input features of the metabolic parameters that were significantly different into the SVM classifier, with a value of  $P$  smaller than 0.05 as the threshold.

### Model evaluation and validation

The predictive performance of the model was evaluated on the entire training cohort and validated by using the validation cohort. We applied the receiver operating characteristic (ROC) curve, calibration curve, and decision curve analysis (DCA) to evaluate the predictive accuracy and clinical feasibility of the model. We also calculated the area under the receiver operating characteristic curve (AUC), sensitivity, specificity, accuracy, recall, precision, the F1-score, and area under the precision–recall curve (AUPR) as indicators of performance. The learning curve of the model was used to determine its convergence. It is a visualization tool to identify whether the number of training samples of a given model is sufficient for its convergence by showing how the errors or accuracies of the training and validation sets change as the number of training samples increases [36]. We used the SHapley Additive explanation (SHAP) method, which is a visual tool that can calculate the contribution of each input variable to the decision of a machine learning model [37], to interpret the output of the machine learning model, and reveal the relationship between the radiomic features of the image and KRAS/NRAS/BRAF mutations.

### Statistical analysis

The Python programming language (version 3.9.7; [www.python.org](http://www.python.org)) was used for statistical analysis, visualization, and machine learning. The Student t-test was used to determine whether the continuous variables were normally distributed (the results are reported as mean  $\pm$  standard deviation), and the Mann–Whitney U test was used if they were not normally distributed (presented as the median with an interquartile range). The Chi-square test or Fisher's exact test was used for the categorical variables (presented as numbers or percentages). Two-tailed  $P$ -values of  $<0.05$  were considered to be statistically significant.

## Results

### Clinical characteristics

The patients were divided into a mutant group and a wild group according to the results of the status of KRAS/NRAS/BRAF mutations. Table 1 lists the demographics of the patients and the clinical characteristics of the tumors in both groups. No statistical differences were observed between the wild and mutant groups in terms

**Table 1** Comparison of patient demographic and clinical tumor characteristics between KRAS/NRAS/BRAF mutation group and wild type group

Characteristics	Wild type (n = 27)	Mutant type (n = 35)	P
Sex			0.644 <sup>a</sup>
Male	17(63.0)	24(68.6)	
Female	10(37.0)	11(31.4)	
Age(year)	61.11 ± 11.40	64.57 ± 10.21	0.213 <sup>b</sup>
Tumor location			0.636 <sup>a</sup>
Colon or sigmoid colon	20(74.1)	24(68.6)	
Rectum or rectosigmoid junction	7(25.9)	11(31.4)	
Differentiation			0.420 <sup>a</sup>
Well/moderate	21(77.8)	24(68.6)	
Poor	6(22.2)	11(31.4)	
TNM stage			0.044 <sup>a</sup>
I-II	17(63.0)	13(37.1)	
III-IV	10(37.0)	22(62.9)	
CEA			0.001 <sup>a</sup>
< 5.0 ng/ml	21(77.8)	12(34.3)	
≥ 5.0 ng/ml	6(22.2)	23(65.7)	

CEA carcinoembryonic antigen, KRAS Kirsten rat sarcoma viral oncogene homolog, NRAS neuroblastoma rat sarcoma viral oncogene homolog, BRAF v-raf murine sarcoma viral oncogene homolog B

<sup>a</sup> Chi-square test

<sup>b</sup> Student t-tests

of sex, age, tumor location, and differentiation ( $P > 0.05$ ). However, statistically significant differences were noted between the wild and mutant groups in terms of the TNM stage and the CEA ( $P < 0.05$ ). Table 2 lists the demographics of the patients and the clinical characteristics of the tumor in the training and validation groups. The results show that although there were differences between them in terms of sex, age, tumor location, TNM stage, and CEA, they were not statistically significant ( $P > 0.05$ ).

#### Unsupervised cluster analysis and correlation analysis

After excluding features with low reproducibility (ICCs < 0.75), the number of habitat-derived radiomic features were reduced from 2553 to 2426, and the number of whole tumor region radiomic features were reduced from 851 to 803. We then separately analyzed their clustering. The cluster map of the habitat-derived radiomic features had a clear contrast in shape with the cluster maps of the whole tumor region radiomic features and the metabolic parameters (Supplementary Information Fig. S1). This indicates that the habitat-derived radiomic features were more representative and compact. Notably, the heatmap of correlation (Supplementary Information Fig. S2) showed that habitat-driven radiomics features with potential relevance could be found for almost all whole tumor region radiomic features or

metabolic parameters (dark-blue color). However, the presence of partial habitat-derived radiomic features exhibited a low correlation with the whole tumor region radiomic features or the metabolic parameters (light-blue and red colors, respectively). This may correspond to additional deep information and be more suggestive for the status of KRAS/NRAS/BRAF mutations during model construction.

#### Feature selection

The habitat-derived radiomic features were ranked according to mRMR, and the top 30 features were then selected for further analysis. The top features included seven features from low\_metabolism\_habitat, 12 from high\_metabolism\_habitat, and 11 from the peritumoral region. Following this, we used the LASSO method to select five subregional radiomic features with non-zero coefficients under the best value of lambda, as illustrated in Fig. 2. We finally retained two features from high\_metabolism\_habitat, three from the peritumoral region, and no feature from low\_metabolism\_habitat. The rule of thumb to avoid overfitting is that the number of predictors should be kept within 1/10–1/3 of the sample size in each group of the main cohort [38]. Therefore, the number of features screened here was considered reasonable. However, when mRMR and LASSO analyses were performed on the whole tumor region radiomic features,

**Table 2** Comparison of patient demographic and clinical tumor characteristics between the training cohort and validation cohort

Characteristics	Training cohort (n = 37)	Validation cohort (n = 25)	P
Sex			0.402 <sup>a</sup>
Male	26(70.3)	15(60.0)	
Female	11(29.7)	10(40.0)	
Age(year)	64.35 ± 10.71	61.16 ± 10.85	0.257 <sup>b</sup>
Tumor location			0.473 <sup>a</sup>
Colon or sigmoid colon	25(67.6)	19(76.0)	
Rectum or rectosigmoid junction	12(32.4)	6(24.0)	
Differentiation			0.933 <sup>a</sup>
Well/moderate	27(73.0)	18(72.0)	
Poor	10(27.0)	7(28.0)	
TNM stage			0.640 <sup>a</sup>
I-II	17(45.9)	13(52.0)	
III-IV	20(54.1)	12(48.0)	
CEA			0.498 <sup>a</sup>
< 5.0 ng/ml	21(56.8)	12(48.0)	
≥ 5.0 ng/ml	16(43.2)	13(52.0)	
Mutational status			0.953 <sup>a</sup>
KRAS/NRAS/BRAF mutated	21(56.8)	14(56.0)	
Wild type	16(43.2)	11(44.0)	

CEA carcinoembryonic antigen, KRAS Kirsten rat sarcoma viral oncogene homolog, NRAS neuroblastoma rat sarcoma viral oncogene homolog, BRAF v-raf murine sarcoma viral oncogene homolog B

<sup>a</sup> Chi-square test

<sup>b</sup> Student t-tests

no feature was retained (Supplementary Information Fig. S3). We also did not find a significant difference in the metabolic parameters between the wild type group and the mutant group of KRAS/NRAS/BRAF, although all metabolic parameters exhibited a tendency to have larger values in the mutant group than in the wild type group (Supplementary Information Fig. S4). Therefore, no whole tumor region radiomic features and metabolic parameters were available for the subsequent examination of model construction.

### Model evaluation

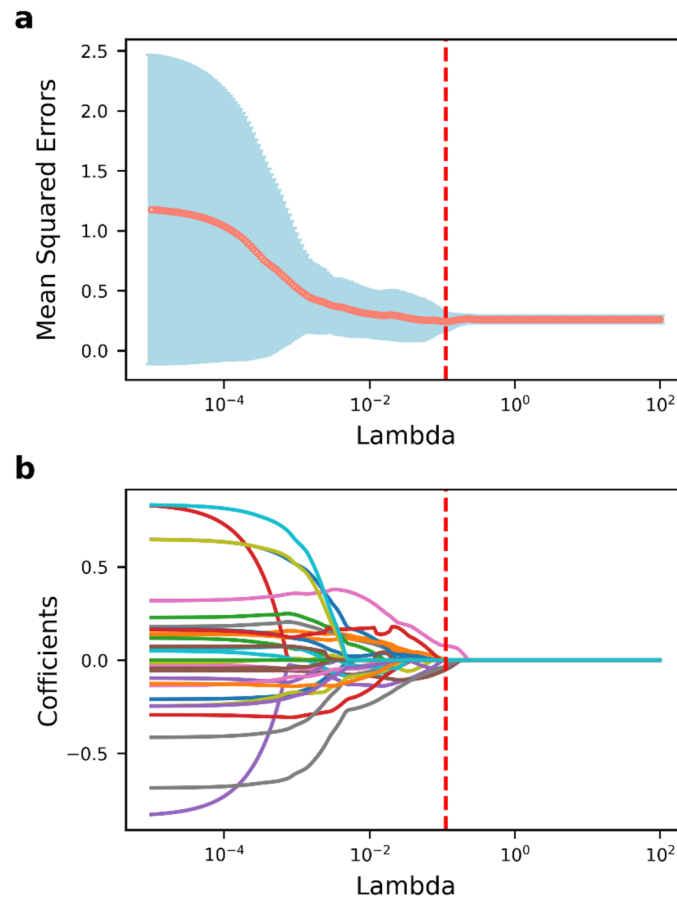
We constructed an SVM model to predict KRAS/NRAS/BRAF mutations based on the selected habitat-derived radiomic features, and determined the optimal values of the parameters as  $C=2.815$ ,  $\gamma=0.364$ , and kernel=sigmoid. The predictive performance of the model was subsequently tested on the training and validation sets. Table 3 and Fig. 3 show the performance of the model of classification and its ROC curves. The model had an AUC of 0.759 (95% CI: 0.585–0.909) on the training set, with a sensitivity of 0.810, specificity of 0.688, accuracy of 0.757, recall of 0.810, precision of 0.773, F1-score of 0.791, and AUPR of 0.808. The model in the validation group had an AUC of 0.701 (95% CI:

0.468–0.916), sensitivity of 0.786, specificity of 0.545, accuracy of 0.680, recall of 0.786, precision of 0.688, F1-score of 0.733, and AUPR of 0.682. These results reflect the acceptable predictive performance of the proposed model. In addition, its curve of calibration in Fig. 4 shows good agreement between the probability of KRAS/NRAS/BRAF mutations predicted by the model and their actual probability on both the training and validation datasets. Furthermore, the DCA plots verified the clinical utility of the SVM model based on habitat-derived radiomic features for the training group (Fig. 5a) and the validation group (Fig. 5b).

The learning curve provides a visual insight into the predictive performance of a model. The learning curve in Fig. 6 shows that values of the AUC of the proposed model on the training and validation groups tended to be stable and close to each other, and conformed to our expectations. Therefore, the proposed model converged and achieved stable performance even with a small sample size.

### Analysis of model interpretability

SHAP analysis was conducted on the validation cohort. We calculated the importance of each feature based on the SHAP values (Fig. 7a). The graph shows the results



**Fig. 2** Feature selection of habitat-derived radiomic features using the LASSO algorithm. **a** Tuning parameter lambda selection in the LASSO algorithm used ten-fold cross-validation. **b** LASSO coefficient profiles of the features. LASSO, least absolute shrinkage and selection operator

**Table 3** Performance evaluation of SVM model based on habitat-derived radiomic features in training cohort and validation cohort. SVM, support vector machine; AUC, area under the receiver operating characteristic curve; AUPR, area under the precision-recall curve

Cohort	AUC (95%CI)	Sensitivity	Specificity	Accuracy	Recall	Precision	F1-score	AUPR
Training cohort	0.759 (0.585–0.909)	0.810	0.688	0.757	0.810	0.773	0.791	0.808
Validation cohort	0.701 (0.468–0.916)	0.786	0.545	0.680	0.786	0.688	0.733	0.682

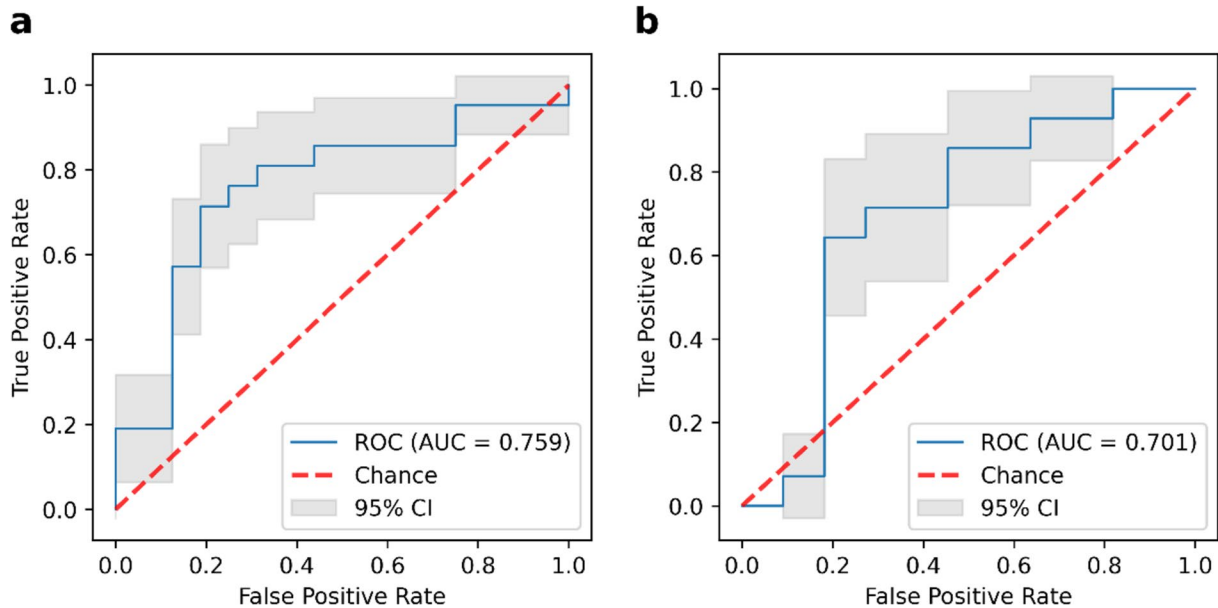
of the analysis of the importance of features in the SVM model by SHAP method. Each point in Fig. 7a represents the SHAP value of a feature in one instance. Its position on the Y-axis was determined based on the importance of the feature and its position on the X-axis was determined by its SHAP value. The colors indicate values of the features ranging from small to large to better represent the distribution of the SHAP values for each feature. We also used SHAP to account for a typical prediction of the presence and absence of KRAS/NRAS/BRAF mutations (Fig. 7b and c). The predictions were made starting with a baseline value

derived from the average of all predictions. Each band showed the effect of its feature in pushing the value of the target variable further from or closer to the baseline value, and the likelihood of increasing (positive value) or decreasing (negative value) KRAS/NRAS/BRAF mutations.

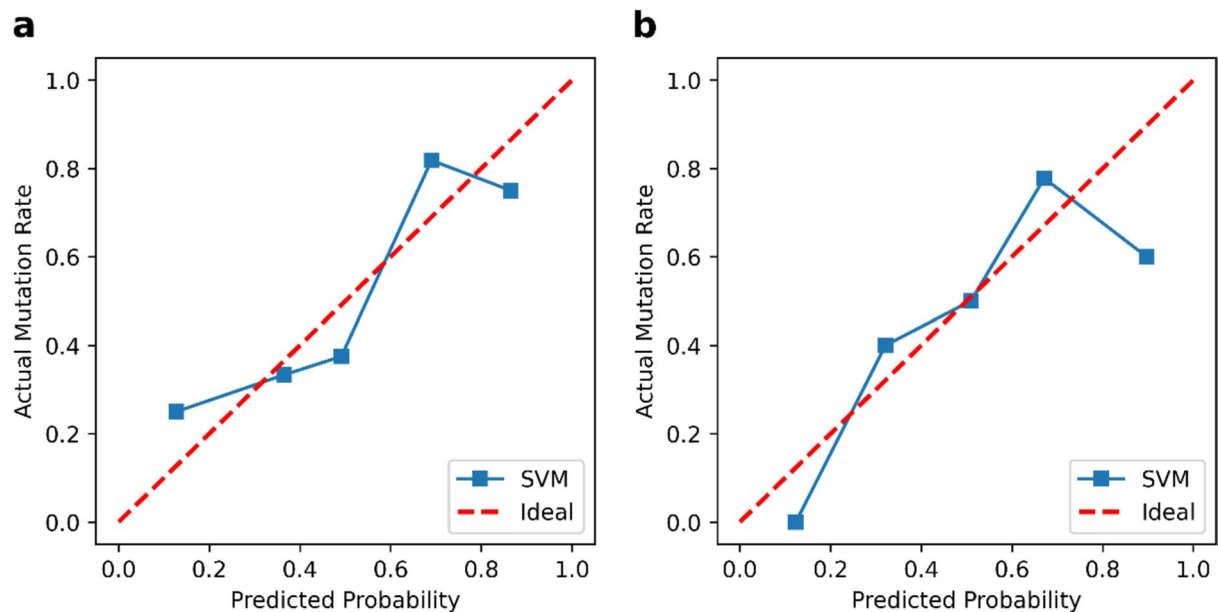
**Discussion**

With advances in precision medicine, the status of KRAS/NRAS/BRAF mutations has been applied to clinical decision-making in case of CRC [39]. Histopathology remains the gold standard for the diagnosis and





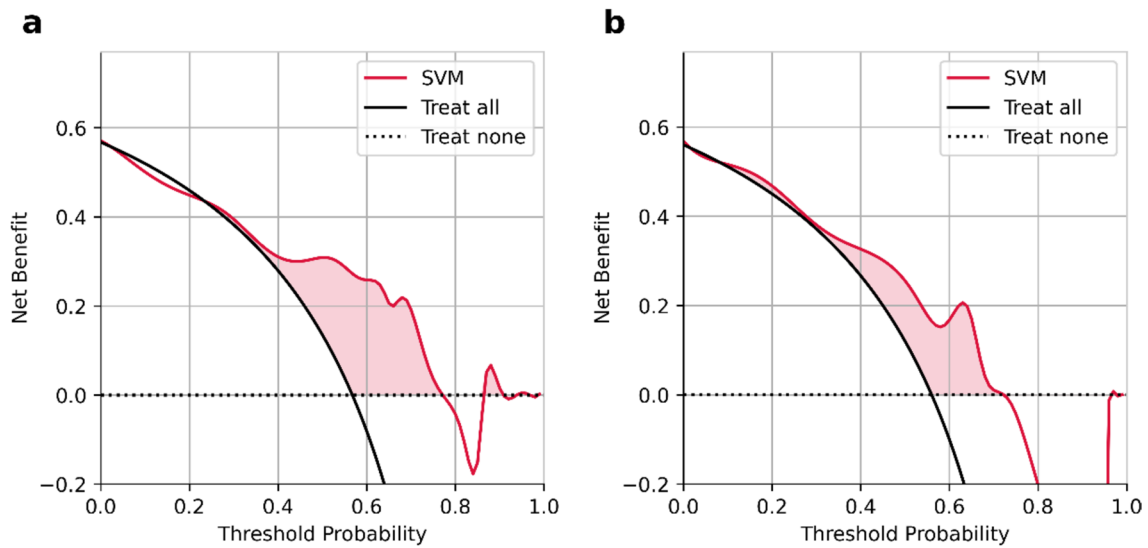
**Fig. 3** The ROC curves of the SVM model based on habitat-derived radiomic features in the training group (a) and the validation group (b). ROC, receiver operating characteristic; SVM, support vector machine; AUC, area under the receiver operating characteristic curve



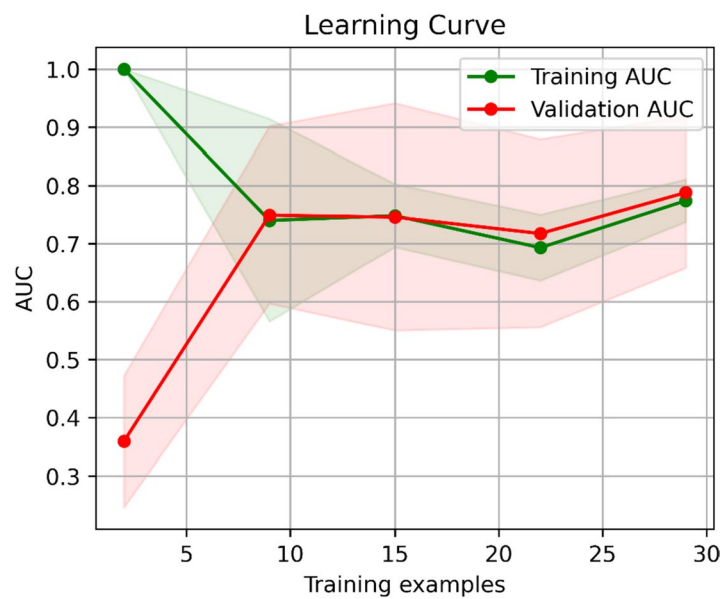
**Fig. 4** The calibration curves of the SVM model based on habitat-derived radiomic features in the training group (a) and the validation group (b). SVM, support vector machine

classification of tumors. However, there is increasing recognition of the limitations of tissue-based genetic testing, including intra-tumor heterogeneity, clonal evolution, and poor DNA quality, especially in biopsy samples [40]. This can lead to a suboptimal profile of the genetic signature of the tumor. Therefore, we sought to determine the most effective way to image phenotypes

by using habitat-derived quantitative radiomic measurements to distinguish between CRC mutants and wild type genomes in this pilot study. The AUC of our SVM model was 0.759 on the training group and 0.701 on the validation group. It exhibited sound calibration and clinical utility, and achieved good convergence even with a small sample size. This suggests that it can complement



**Fig. 5** The DCA of the SVM model based on habitat-derived radiomic features for the training (a), and validation (b) groups. SVM, support vector machine; DCA, decision curve analysis

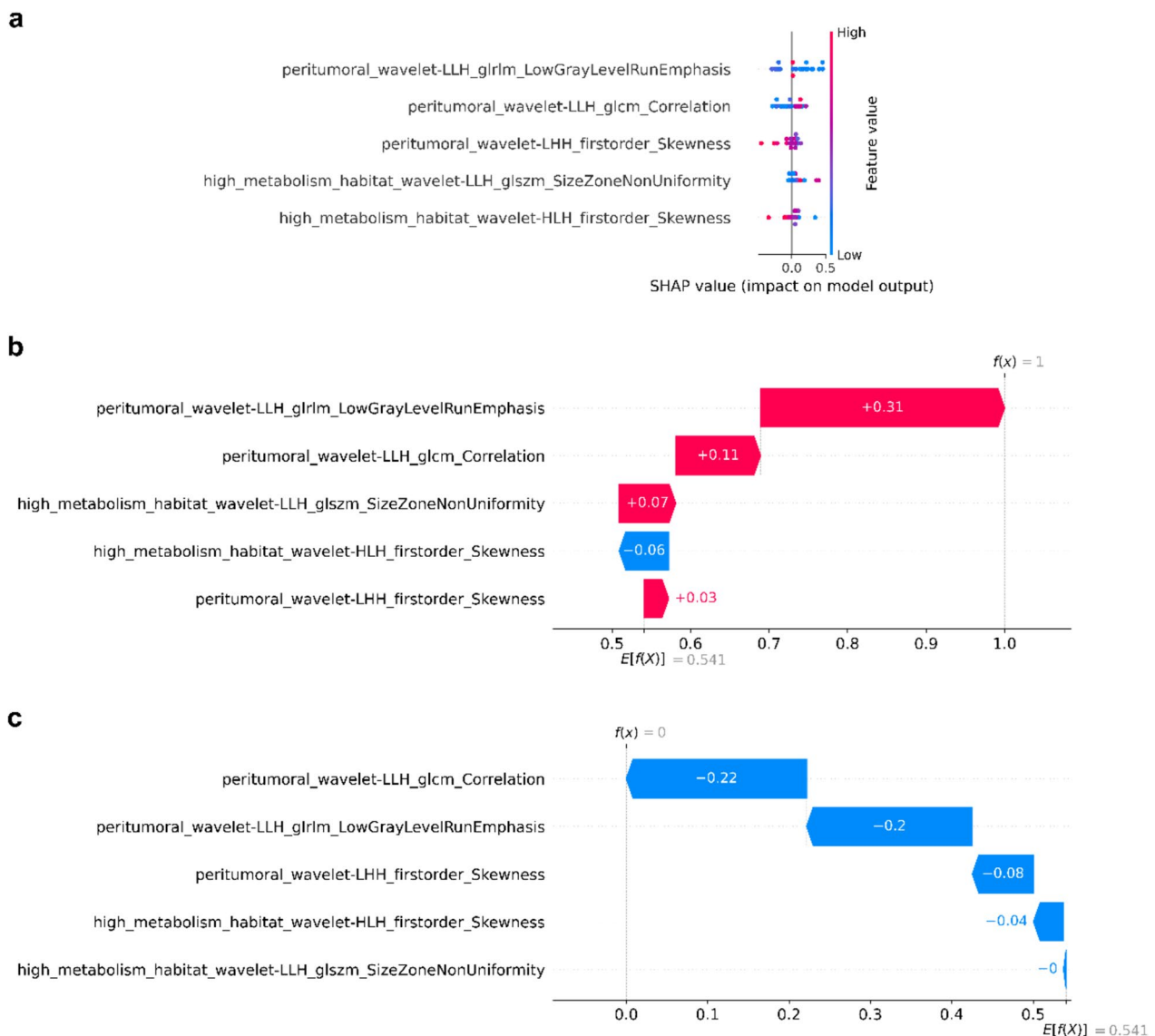


**Fig. 6** Learning curve of the SVM model based on habitat-derived radiomic features. SVM, support vector machine; AUC, area under the receiver operating characteristic curve

the status of mutation of genes to enhance the optimal treatment strategy for patients with CRC.

Data from basic studies suggest that mutated KRAS may lead to increased <sup>18</sup>F-FDG accumulation through the upregulation of GLUT1 under normoxic conditions, and that HIF-1 $\alpha$  increases <sup>18</sup>F-FDG accumulation in hypoxic lesions [12]. Therefore, it is theoretically feasible to characterize the phenotype of the mutants of tumors by using

metabolic parameters. However, like previous studies [13], the results of ours showed that although the metabolic parameters of KRAS/NRAS/BRAF mutants tended to have larger values than in wild type genomes, none of these was statistical significance. Several parameters based on SUV (SUVmin, SUVmean, SUVmax, SUVstd) reflect the metabolic activity of a certain part or a certain point of the tumor tissue, but cannot reflect the overall



**Fig. 7** a SHAP summary graph and distribution of SHAP values for trained SVM model based on habitat-derived radiomic features. Characteristic SHAP value influence diagram of individual KRAS/NRAS/BRAF mutant (b) and wild type (c) samples. SHAP, SHapley Additive explanation; SVM, support vector machine; KRAS, Kirsten rat sarcoma viral oncogene homolog; NRAS, neuroblastoma rat sarcoma viral oncogene homolog; BRAF, v-raf murine sarcoma viral oncogene homolog B; glrIm, gray-level run length matrix; glcm, gray-level co-occurrence matrix; glszm, gray-level size zone matrix

metabolism of the tumor. The derived parameters (MTV, sMTV, TLG, sTLG) make up for this deficiency. For example, MTV and TLG can reflect the metabolic activity and the tumor burden. However, these parameters cannot distinguish the heterogeneity of different regions within the tumor [41]. We also observed that all whole tumor region radiomic features were removed during feature selection. Some previous studies have reported the same results, whereby whole tumor region textural features obtained by using  $^{18}\text{F}$ -FDG PET/CT are not

associated with the presence of Ras mutations in patients of rectal cancer [17]. Fortunately, our SVM model based on habitat-derived radiomic features yielded good performance in predicting the presence of KRAS/NRAS/BRAF mutations, where this further demonstrates the value of techniques of habitat analysis in characterizing the complex microenvironment of tumors. It is important to note that different regions within the tumor may have different biological characteristics. Habitat-based methods deliver better performance, possibly due to the

clustering of similar voxels through habitat quantification, which allows for the delineation of metabolic heterogeneity driven by genetic heterogeneity through a more fine-grained characterization of the microenvironment of the tumor [42]. The advantage of this promising approach to describing tumor heterogeneity has been demonstrated in several previous studies. For example, in their comparison of the performance of  $^{18}\text{F}$ -FDG PET/CT whole tumor region radiomics with habitat-derived radiomics in determining Ki-67 expression in case of high-grade serous ovarian cancer, Wang et al. found that the habitat-based method was significantly better than the whole tumor region radiomics analysis (DeLong test,  $P < 0.05$ ) [26]. In another study, Xie et al. used habitat-derived radiomic features in CT images to construct a model to assess the prognosis of esophageal squamous cell carcinoma treated by definitive concurrent chemoradiotherapy. The concordance index of the model was 0.729 on the training cohort and 0.705 on the validation cohort. This is superior to that of the whole tumor region radiomics analysis, although it was not statistically significant [23].

We also explored reasons for the poor performance of the metabolic parameters and the whole tumor region radiomics in comparison with that of habitat-derived radiomics by using unsupervised clustering analysis. The results showed that the unsupervised clustering of habitat-derived radiomics led to a better match with the semantic labels (mutant/wild type genomes). Moreover, the analysis of correlations between features showed that almost all whole tumor region radiomics features and metabolic parameters could find the corresponding habitat-derived radiomic features with correlation, and some habitat-derived radiomic features had a low correlation with the whole tumor region radiomic features and the metabolic parameters. Considering that our approach based on habitat-derived radiomics delivered better performance, it is reasonable to conclude that the relevant features not only cover almost all information on the metabolic parameters and the whole tumor region radiomic features, but also contain key discriminating information that is not found in these two sets of features of images. This also nicely explains the higher classification-related performance of habitat-derived radiomics.

We used SHAP analysis to interpret the “black-box model” at both the global and the local levels [43]. The SHAP method enhances the interpretability of a machine learning model, and can estimate the positive and negative contributions of each feature to its prediction [44]. The SHAP results showed that the ranking of the features according to importance was peritumoral\_wavelet-LLH\_glrIm\_LowGrayLevelRunEmphasis, peritumoral\_wavelet-LLH\_glcm\_Correlation, peritu-

moral\_wavelet-LHH\_firstorder\_Skewness, high\_metabolism\_wavelet-LLH\_glszm\_SizeZoneNonUniformity, and high\_metabolism\_wavelet-HLH\_firstorder\_Skewness. The radiomic features obtained from the peritumoral regions were the three most important features, followed by those from the hypermetabolic regions, while none of the features from the hypometabolic regions was included in the final model. Tumor invasion is a complex biological process at the microscopic level, in which tumor cells tend to separate from the primary tumor and invade the surrounding tissues [45]. Tumors with gene mutations in particular are likely to invade and exhibit metastasized phenotypes [46]. Therefore, it may be beneficial to capture information on KRAS/NRAS/BRAF mutations from a macroscopic perspective by using radiomics in the peritumoral region. Multiple circular peritumoral regions were not observed in this study. According to a recent report, the closer the peritumoral area is to the intra-tumoral area, the greater is the amount of information that it contains [47]. The excessive expansion of information on the surrounding edges may introduce noise to impair the capability of classification of the model [48]. At the same time, the high\_metabolism features also helped facilitate the stratification of the mutations. Similar to the work in this study, Mu et al. selected the main habitat-related features for a model of the prognostic assessment of locally advanced cervical cancer based on  $^{18}\text{F}$ -FDG PET/CT images from the hypermetabolic regions [27]. Moreover, Wu et al. showed that the volume of most metabolically active and metabolically heterogeneous tumor solid components can be used as a predictor of overall survival and out-of-field progression in patients of lung cancer [49]. Based on this result, we hypothesize that the hypermetabolic habitats of tumors in PET images, which generally reflect the metabolism of active tumor cells, may harbor more information about mutations in tumors than hypometabolic habitat. It is evident from the results, that using 40% of SUVmax as the threshold to generate hypermetabolic VOI is a more convenient method to extract radiomics features from hypermetabolic tumor regions. Therefore, to further validate the effectiveness of the habitat-derived radiomic analysis, we generated a VOI with the SUVmax of 40% as the threshold based on the VOI of the whole tumor region obtained in this study, and carried out the same feature screening and modeling process. However, the model was severely overfitted during training (not shown in the results). Tumor heterogeneity refers not only to the heterogeneity within a tumor, but also to the heterogeneity between tumors. Therefore, using a fixed threshold to evaluate the degree of metabolism of all tumors interferes with

the quantification of tumor heterogeneity based on radiomics features. Moreover, all the features considered here were wavelet features, which indicates the importance of wavelet decomposition for radiomics analysis. The wavelet features characterize heterogeneity at multiple spatial scales, and may be an important radiomic feature to quantify KRAS/NRAS/BRAF mutations [32].

Our analysis of clinicopathological factors showed that patients with high CEA levels and high TNM stages had a higher probability of KRAS/NRAS/BRAF mutations. KRAS/NRAS/BRAF mutations are associated with poor prognosis of patients with CRC [50, 51], while TNM staging prediction systems and markers such as CEA play an important role in monitoring the prognosis of gastrointestinal cancer [52, 53]. Therefore, there are potential associations between KRAS/NRAS/BRAF mutations and CEA level as well as TNM stage.

Deep learning has exhibited significant value in oncology [54, 55]. In particular, Wu et al. have reported that adding deep learning features to manual radiomic features can improve the ability to predict the KRAS mutation in case of CRC [56]. However, because of the small sample size used in this study, and given that deep learning is a data-intensive technique, we did not use it here. Its appropriate use in this domain is another important area that warrants further research.

This study has certain limitations. First, this was a preliminary exploration of a small sample collected from a single hospital. Our model delivered good predictive performance, and the results of its learning curve showed that it achieved convergence even with a small number of samples. However, using a larger sample size and providing external validation will further verify its clinical significance. Second, even though it is the mainstream method in radiomics research, manual segmentation may introduce uncertainty due to human participation. Automatic segmentation can further optimize research on the radiomics-based prediction of mutations in the CRC gene. Finally, the probability of KRAS/NRAS/BRAF mutations obtained in this study was higher than that reported in the literature, and the impact of this outcome on our model needs to be further evaluated.

## Conclusion

In this study, we showed that habitat-derived radiomic features are superior to conventional metabolic parameters and whole tumor region radiomic features in predicting KRAS/NRAS/BRAF mutations in CRC patients. In addition, the SVM model based on habitat-derived radiomic features has the potential to assist in decision-making regarding the treatment of CRC patients in clinical practice, but its performance needs to be validated on a larger prospective cohort than was considered here.

## Abbreviations

AUC	Area under the receiver operating characteristic curve
AUPR	Area under the precision–recall curve
BRAF	V-raf murine sarcoma viral oncogene homolog B
CRC	Colorectal cancer
CT	Computed tomography
CEA	Carcinoembryonic antigen
DCA	Decision curve analysis
EGFR	Epidermal growth factor receptor
FDG	Fluorodeoxyglucose
glcm	Gray-level co-occurrence matrix
glzsm	Gray-level size zone matrix
glrlm	Gray-level run length matrix
gldm	Gray-level dependence matrix
ICCs	Inter-observer and intra-observer correlation coefficients
KRAS	Kirsten rat sarcoma viral oncogene homolog
LASSO	Least absolute shrinkage and selection operator
MTV	Metabolic tumor volume
mRMR	Maximum relevance minimum redundancy
NRAS	Neuroblastoma rat sarcoma viral oncogene homolog
ngtdm	Neighboring gray tone difference matrix
PET	Positron emission tomography
Ras	Rat sarcoma viral oncogene
ROC	Receiver operating characteristic
SUV	Standardized uptake value
SUVmin	Minimal standardized uptake value
SUVmean	Average standardized uptake value
SUVmax	Maximal standardized uptake value
SUVstd	Standard deviation of standardized uptake value
sMTV	Standardized metabolic tumor volume
sTLG	Standardized total lesion glycolysis
SVM	Support vector machine
SHAP	SHapley Additive explanation
TLG	Total lesion glycolysis
VOI	Volume of interest

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s40644-024-00670-2>.

**Additional file 1: Fig. S1.** Cluster heatmaps of habitat-derived radiomic features (a), whole-tumor radiomic features (b), and metabolic parameters (c). The X-axis represents 62 patients, and the Y-axis represents image features. Patients in the same cluster shared similar image features in Euclidean space. The semantic labels (mutant and wild types) for each patient are shown on the red and blue bars above. **Fig. S2.** (a) Heatmap of the correlation between habitat-derived radiomic features and whole-tumor radiomic features. (b) Heatmap of the correlation between habitat-derived radiomic features and metabolic parameters. **Fig. S3.** Feature selection of whole tumor region radiomics features using the LASSO algorithm. (a) Tuning parameter lambda selection in the LASSO algorithm used 10-fold cross-validation. (b) LASSO coefficient profiles of the features. LASSO, least absolute shrinkage and selection operator. **Fig. S4.** Violin plot of the distribution of the metabolic parameters. SUVmin, minimal standardized uptake value; SUVmean, average standardized uptake value; SUVmax, maximal standardized uptake value; SUVstd, standard deviation of standardized uptake value; MTV, metabolic tumor volume; sMTV, standardized metabolic tumor volume; TLG, total lesion glycolysis, sTLG, standardized total lesion glycolysis.

## Acknowledgements

The authors thank all their coworkers involved in the study for their support and assistance.

## Authors' contributions

All authors contributed to the study conception and design. Hongyue Zhao, Yexin Su provided the overall design of the experiments and contributed equally to the experiments. Yan Wang and Zhehao Lyu were responsible for

the collection of clinical and imaging data. Zhehao Lyu and Wenchao Gu completed tumor segmentation. Hongyue Zhao, Yexin Su, Yan Wang and Peng Xu performed the radiomics analysis and model building. Hongyue Zhao, Yexin Su, Lin Tian and Peng Fu wrote and edited the manuscript. Peng Fu and Lin Tian supervised the study. All authors read and approved the final manuscript.

### Funding

This work was supported by funds from the National Natural Science Foundation of China (82371994), 2022 Innovation Fund for Heilongjiang Provincial Institutions of Higher Learning (2022-KYYWF-0273), Scientific Research and Innovation Fund of the First Affiliated Hospital of Harbin Medical University (No. 2021M32) and Scientific Research and Innovation Fund of the First Affiliated Hospital of Harbin Medical University (No. 2021M16).

### Availability of data and materials

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

### Declarations

#### Ethics approval and consent to participate

This retrospective study was approved by the medical ethics committee of our institution. This work was performed in adherence to the tenets of the Declaration of Helsinki.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Department of Nuclear Medicine, The First Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang, China. <sup>2</sup>Department of Pathology, The First Affiliated Hospital of Harbin Medical University, Harbin, Heilongjiang, China. <sup>3</sup>Department of Diagnostic and Interventional Radiology, University of Tsukuba, Ibaraki, Japan.

Received: 14 August 2023 Accepted: 29 January 2024

Published online: 12 February 2024

### References

- Qi J, Sun H, Zhang Y, Wang Z, Xun Z, Li Z, et al. Single-cell and spatial analysis reveal interaction of FAP+ fibroblasts and SPP1+ macrophages in colorectal cancer. *Nat Commun*. 2022;13(1):1742.
- Wang TH, Wu CC, Huang KY, Chuang WY, Hsueh C, Li HJ, et al. Profiling of subcellular EGFR interactome reveals hnRNP A3 modulates nuclear EGFR localization. *Oncogenesis*. 2020;9(4):40.
- Denis JA, Patroni A, Guillermin E, Pépin D, Benali-Furet N, Wechsler J, et al. Droplet digital PCR of circulating tumor cells from colorectal cancer patients can predict KRAS mutations before surgery. *Mol Oncol*. 2016;10(8):1221–31.
- He P, Zou Y, Qiu J, Yang T, Peng L, Zhang X. Pretreatment <sup>18</sup>F-FDG PET/CT imaging predicts the KRAS/NRAS/BRAF gene mutational status in colorectal cancer. *J Oncol*. 2021;2021:6687291.
- Johnson H, El-Schich Z, Ali A, Zhang X, Simoulis A, Wingren AG, et al. Gene-mutation-based algorithm for prediction of treatment response in colorectal cancer patients. *Cancers (Basel)*. 2022;14(8):2045.
- Nicolazzo C, Belardinelli F, Vestri A, Magri V, De Renzi G, De Meo M, et al. RAS mutation conversion in bevacizumab-treated metastatic colorectal cancer patients: a liquid biopsy based study. *Cancers (Basel)*. 2022;14(3):802.
- Madej E, Brożyna AA, Adamczyk A, Wronski N, Harazin-Lechowska A, Muzyk A, et al. Vemurafenib and dabrafenib downregulates RPK4 level. *Cancers (Basel)*. 2023;15(3):918.
- Liu J, Zeng W, Huang C, Wang J, Xu L, Ma D. Upregulation of c-mesenchymal epithelial transition expression and RAS mutations are associated with late lung metastasis and poor prognosis in colorectal carcinoma. *Exp Ther Med*. 2018;15(5):4229–42.
- Benson AB, Venook AP, Al-Hawary MM, Arain MA, Chen YJ, Ciombor KK, et al. Colon cancer, version 2.2021, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. 2021;19(3):329–59.
- Montagut C, Argilés G, Ciardiello F, Poulsen TT, Dienstmann R, Kragh M, et al. Efficacy of Sym004 in patients with metastatic colorectal cancer with acquired resistance to Anti-EGFR therapy and molecularly selected by circulating tumor DNA analyses: a phase 2 randomized clinical trial. *JAMA Oncol*. 2018;4(4):e175245.
- Morano F, Corallo S, Lonardi S, Raimondi A, Cremolini C, Rimassa L, et al. Negative hyperselection of patients with RAS and BRAF wild-type metastatic colorectal cancer who received panitumumab-based maintenance therapy. *J Clin Oncol*. 2019;37(33):3099–110.
- Iwamoto M, Kawada K, Nakamoto Y, Itatani Y, Inamoto S, Toda K, et al. Regulation of 18F-FDG accumulation in colorectal cancer cells with mutated KRAS. *J Nucl Med*. 2014;55(12):2038–44.
- Krikelis D, Skoura E, Kotoula V, Rondogianni P, Pianou N, Samartzis A, et al. Lack of association between KRAS mutations and 18F-FDG PET/CT in Caucasian metastatic colorectal cancer patients. *Anticancer Res*. 2014;34(5):2571–9.
- Jiang P, Du W, Wang X, Mancuso A, Gao X, Wu M, et al. p53 regulates biosynthesis through direct inactivation of glucose-6-phosphate dehydrogenase. *Nat Cell Biol*. 2011;13(3):310–6.
- Taguchi N, Oda S, Yokota Y, Yamamura S, Imuta M, Tsuchigame T, et al. CT texture analysis for the prediction of KRAS mutation status in colorectal cancer via a machine learning approach. *Eur J Radiol*. 2019;118:38–43.
- Zinn PO, Singh SK, Kotrotsou A, Abrol S, Thomas G, Mosley J, et al. Distinct radiomic phenotypes define glioblastoma TP53-PTEN-EGFR mutational landscape. *Neurosurg*. 2017;64(CN\_suppl\_1):203–10.
- Lovinfosse P, Koopmansch B, Lambert F, Jodogne S, Kustermans G, Hatt M, et al. (18)F-FDG PET/CT imaging in rectal cancer: relationship with the RAS mutational status. *Br J Radiol*. 2016;89(1063):20160212.
- Cui Y, Liu H, Ren J, Du X, Xin L, Li D, et al. Development and validation of a MRI-based radiomics signature for prediction of KRAS mutation in rectal cancer. *Eur Radiol*. 2020;30(4):1948–58.
- Meng X, Xia W, Xie P, Zhang R, Li W, Wang M, et al. Preoperative radiomic signature based on multiparametric magnetic resonance imaging for noninvasive evaluation of biological characteristics in rectal cancer. *Eur Radiol*. 2019;29(6):3200–9.
- Gatenby RA, Grove O, Gillies RJ. Quantitative imaging in cancer evolution and ecology. *Radiology*. 2013;269(1):8–15.
- Kim M, Park JE, Kim HS, Kim N, Park SY, Kim YH, et al. Spatiotemporal habitats from multiparametric physiologic MRI distinguish tumor progression from treatment-related change in post-treatment glioblastoma. *Eur Radiol*. 2021;31(8):6374–83.
- Wu J, Cao G, Sun X, Lee J, Rubin DL, Napel S, et al. Intratumoral spatial heterogeneity at perfusion MR imaging predicts recurrence-free survival in locally advanced breast cancer treated with neoadjuvant chemotherapy. *Radiology*. 2018;288(1):26–35.
- Xie C, Yang P, Zhang X, Xu L, Wang X, Li X, et al. Sub-region based radiomics analysis for survival prediction in oesophageal tumours treated by definitive concurrent chemoradiotherapy. *EBioMedicine*. 2019;44:289–97.
- Chen SW, Shen WC, Chen WT, Hsieh TC, Yen KY, Chang JG, et al. Metabolic imaging phenotype using radiomics of [<sup>18</sup>F]FDG PET/CT associated with genetic alterations of colorectal cancer. *Mol Imaging Biol*. 2019;21(1):183–90.
- Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, 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):4786–9.
- Wang X, Xu C, Grzegorzec M, Sun H. Habitat radiomics analysis of pet/ct imaging in high-grade serous ovarian cancer: application to Ki-67 status and progression-free survival. *Front Physiol*. 2022;13:948767.
- Mu W, Liang Y, Hall LO, Tan Y, Balagurunathan Y, Wenham R, et al. <sup>18</sup>F-FDG PET/CT habitat radiomics predicts outcome of patients with cervical cancer treated with chemoradiotherapy. *Radiol Artif Intell*. 2020;2(6):e190218.
- van Griethuysen JJM, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational radiomics system to decode the radiographic phenotype. *Cancer Res*. 2017;77(21):e104–7.

29. Pfaehler E, Mesotten L, Zhovannik I, Pieplenbosch S, Thomeer M, Vanhove K, et al. Plausibility and redundancy analysis to select FDG-PET textural features in non-small cell lung cancer. *Med Phys*. 2021;48(3):1226–38.
30. Somasundaram A, García DV, Pfaehler E, Jauw YWS, Zijlstra JM, van Dongen GAMS, et al. Noise sensitivity of <sup>89</sup>Zr-Immuno-PET radiomics based on count-reduced clinical images. *EJNMMI Phys*. 2022;9(1):16.
31. Wang J, Xiong X, Ye J, Yang Y, He J, Liu J, et al. A radiomics nomogram for classifying hematoma entities in acute spontaneous intracerebral hemorrhage on non-contrast-enhanced computed tomography. *Front Neurosci*. 2022;16:837041.
32. Wang X, Zhao X, Li Q, Xia W, Peng Z, Zhang R, et al. Can peritumoral radiomics increase the efficiency of the prediction for lymph node metastasis in clinical stage T1 lung adenocarcinoma on CT? *Eur Radiol*. 2019;29(11):6049–58.
33. Uttam S, Stern AM, Sevinsky CJ, Furman S, Pullara F, Spagnolo D, et al. Spatial domain analysis predicts risk of colorectal cancer recurrence and infers associated tumor microenvironment networks. *Nat Commun*. 2020;11(1):3515.
34. Cheng CW, Zhou Y, Pan WH, Dey S, Wu CY, Hsu WL, et al. Hierarchical and programmable one-pot synthesis of oligosaccharides. *Nat Commun*. 2018;9(1):5202.
35. Zhao H, Su Y, Wang M, Lyu Z, Xu P, Jiao Y, et al. The machine learning model for distinguishing pathological subtypes of non-small cell lung cancer. *Front Oncol*. 2022;12:875761.
36. Chai KE, Anthony S, Coiera E, Magrabi F. Using statistical text classification to identify health information technology incidents. *J Am Med Inform Assoc*. 2013;20(5):980–5.
37. Choi HJ, Kim I, Lee HJ, Oh HJ, Ahn MK, Baek WJ, et al. Clinical characteristics of neonatal cholestasis in a tertiary hospital and the development of a novel prediction model for mortality. *EBioMedicine*. 2022;77:103890.
38. Yang L, Dong D, Fang M, Zhu Y, Zang Y, Liu Z, et al. Can CT-based radiomics signature predict KRAS/NRAS/BRAF mutations in colorectal cancer? *Eur Radiol*. 2018;28(5):2058–67.
39. Srivatsa S, Paul MC, Cardone C, Holcman M, Amberg N, Pathria P, et al. EGFR in tumor-associated myeloid cells promotes development of colorectal cancer in mice and associates with outcomes of patients. *Gastroenterology*. 2017;153(1):178–90.e10.
40. Sciafani F, Chau I, Cunningham D, Hahne JC, Vlachogiannis G, Eltahir Z, et al. KRAS and BRAF mutations in circulating tumour DNA from locally advanced rectal cancer. *Sci Rep*. 2018;8(1):1445.
41. Liu X, Wang SC, Ni M, Xie Q, Zhang YF, Lv WF, et al. Correlation between <sup>18</sup>F-FDG PET/CT intra-tumor metabolic heterogeneity parameters and KRAS mutation in colorectal cancer. *Abdom Radiol (NY)*. 2022;47(4):1255–64.
42. Cho HH, Kim H, Nam SY, Lee JE, Han BK, Ko EY, et al. Measurement of perfusion heterogeneity within tumor habitats on magnetic resonance imaging and its association with prognosis in breast cancer patients. *Cancers (Basel)*. 2022;14(8):1858.
43. Gou W, Ling CW, He Y, Jiang Z, Fu Y, Xu F, et al. Interpretable machine learning framework reveals robust gut microbiome features associated with type 2 diabetes. *Diabetes Care*. 2021;44(2):358–66.
44. Liu CL, Tain YL, Lin YC, Hsu CN. Prediction and clinically important factors of acute kidney injury non-recovery. *Front Med (Lausanne)*. 2022;8:789874.
45. Cui L, Yu T, Kan Y, Dong Y, Luo Y, Jiang X. Multi-parametric MRI-based peritumoral radiomics on prediction of lymph-vascular space invasion in early-stage cervical cancer. *Diagn Interv Radiol*. 2022;28(4):312–21.
46. Zhang G, Chen L, Liu A, Pan X, Shu J, Han Y, et al. Comparable performance of deep learning-based to manual-based tumor segmentation in KRAS/NRAS/BRAF mutation prediction with MR-based radiomics in rectal cancer. *Front Oncol*. 2021;11:696706.
47. Tang X, Huang H, Du P, Wang L, Yin H, Xu X. Intratumoral and peritumoral CT-based radiomics strategy reveals distinct subtypes of non-small-cell lung cancer. *J Cancer Res Clin Oncol*. 2022;148(9):2247–60.
48. He K, Liu X, Li M, Li X, Yang H, Zhang H. Noninvasive KRAS mutation estimation in colorectal cancer using a deep learning method based on CT imaging. *BMC Med Imaging*. 2020;20(1):59.
49. Wu J, Gensheimer MF, Dong X, Rubin DL, Napel S, Diehn M, et al. Robust intratumor partitioning to identify high-risk subregions in lung cancer: a pilot study. *Int J Radiat Oncol Biol Phys*. 2016;95(5):1504–12.
50. Smeijer J, Sveen A, Merok MA, Danielsen SA, Eilertsen IA, Guren MG, et al. CMS-dependent prognostic impact of KRAS and BRAFV600E mutations in primary colorectal cancer. *Ann Oncol*. 2018;29(5):1227–34.
51. Yanai Y, Hayashi T, Akazawa Y, Yatagai N, Tsuyama S, Yao T, et al. Clinicopathological and mutational differences between tumors with multiple metastases and single lung metastasis in colorectal cancer. *Oncol Lett*. 2020;20(1):541–50.
52. Peng Q, Zhao P, Shen Y, Cheng M, Wu Y, Zhu Y. Prognostic implication and functional exploration for microRNA-20a as a molecular biomarker of gastrointestinal cancer. *BMC Cancer*. 2020;20(1):420.
53. Thirunavukarasu P, Talati C, Munjal S, Attwood K, Edge SB, Francescotti V. Effect of incorporation of pretreatment serum carcinoembryonic antigen levels into AJCC staging for colon cancer on 5-year survival. *JAMA Surg*. 2015;150(8):747–55.
54. Zhao H, Su Y, Lyu Z, Tian L, Xu P, Lin L, et al. Non-invasively discriminating the pathological subtypes of non-small cell lung cancer with pretreatment <sup>18</sup>F-FDG PET/CT using deep learning. *Acad Radiol*. 2023;S1076–6332(23):00167–8.
55. Wang S, Yu H, Gan Y, Wu Z, Li E, Li X, et al. Mining whole-lung information by artificial intelligence for predicting EGFR genotype and targeted therapy response in lung cancer: a multicohort study. *Lancet Digit Health*. 2022;4(5):e309–19.
56. Wu X, Li Y, Chen X, Huang Y, He L, Zhao K, et al. Deep learning features improve the performance of a radiomics signature for predicting KRAS status in patients with colorectal cancer. *Acad Radiol*. 2020;27(11):e254–62.

## Publisher's Note

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