Ann Surg Oncol
https://doi.org/10.1245/s10434-025-18248-y

Annals of
SURGICAL ONCOLOGY
OFFICIAL JOURNAL OF THE SOCIETY OF SURGICAL ONCOLOGY

ORIGINAL ARTICLE – HEPATOBILIARY TUMORS

# CT-Based Radiomics Models with External Validation for Prediction of Recurrence and Disease-Specific Mortality After Radical Surgery of Colorectal Liver Metastases

**Simona Marzi, PhD[1], Antonello Vidiri, MD[2], Anna Ianiro, PhD[1], Chiara Parrino, MD[3], Sergio Ruggiero, MD[2], Claudio Trobiani, MD[2], Leonardo Teodoli, MD[2], Giulio Vallati, MD[2], Giovanni Trillò, MD[2], Maria Ciolina, MD[4], Francesca Sperati, MSc[5], Andrea Scarinci, MD[6], Matteo Virdis, MD[7], Michele Droz Dit Busset, MD[7], Tommaso Stecca, MD[8], Marco Massani, MD[8], Giovanni Morana, MD[9], and Gian Luca Grazi, MD[10]**

[1]Medical Physics Laboratory, IRCCS Regina Elena National Cancer Institute, Rome, Italy; [2]Radiology and Diagnostic Imaging Department, IRCCS Regina Elena National Cancer Institute, Rome, Italy; [3]Hepatopancreatobiliary Surgery, IRCCS, Regina Elena National Cancer Institute, Rome, Italy; [4]Radiology Unit, Department of Medical-Surgical Sciences and Translational Medicine, Sant'Andrea University Hospital, "Sapienza" University of Rome, Rome, Italy; [5]Biostatistics and Bioinformatic Unit, San Gallicano Dermatological Institute IRCCS, Rome, Italy; [6]Department of Medical Surgical Science and Translational Medicine, Sant'Andrea University Hospital, "Sapienza" University of Rome, Rome, Italy; [7]General Surgery and Liver Transplantation Unit, Fondazione IRCCS Istituto Nazionale Tumori, Milan, Italy; [8]General Surgery 1 Unit, Azienda ULSS 2 Marca Trevigiana, Treviso, Italy; [9]Department of Radiology, Ca' Foncello General Hospital, Treviso, Italy; [10]HepatoBiliaryPancreatic Surgery, AOU Careggi, Department of Experimental and Clinical Medicine (DMSC), University of Florence, Florence, Italy

## ABSTRACT

**Purpose.** To build computed tomography (CT)-based radiomics models, with independent external validation, to predict recurrence and disease-specific mortality in patients with colorectal liver metastases (CRLM) who underwent liver resection.

**Methods.** 113 patients were included in this retrospective study: the internal training cohort comprised 66 patients, while the external validation cohort comprised 47. All patients underwent a CT study before surgery. Up to five visible metastases, the whole liver volume, and the surrounding free-disease parenchymal liver were separately delineated on the portal venous phase of CT. Both radiomic features and baseline clinical parameters were considered in the models' building, using different families of machine learning (ML) algorithms.

**Results.** The Support Vector Machine and Naive Bayes ML classifiers provided the best predictive performance. A relevant role of second-order and higher-order texture features emerged from the largest lesion and the liver residual parenchyma. The prediction models for recurrence showed good accuracy, ranging from 70% to 78% and from 66% to 70% in the training and validation sets, respectively. Models for predicting disease-related mortality performed worse, with accuracies ranging from 67% to 73% and from 60% to 64% in the training and validation sets, respectively.

**Conclusions.** CT-based radiomics, alone or in combination with baseline clinical data, allowed the prediction of recurrence and disease-specific mortality of patients with CRLM, with fair to good accuracy after validation in an external cohort. Further investigations with a larger patient population for training and validation are needed to corroborate our analyses.

**Keywords** Colon rectal liver metastases · Recurrence · Radiomics · Machine learning · Modeling · Validation



First Received: 13 March 2025
Accepted: 14 August 2025

A. Vidiri, MD
e-mail: antonello.vidiri@ifo.it

Published online: 10 September 2025

Colorectal cancer (CRC) is the second most common cause of cancer-related deaths in the world; it is the third most prevalent malignancy, with incidence and mortality still increasing in recent years. During the disease course, up to 50% of patients develop liver metastasis (LM), which remains the leading cause of death in patients with CRC.[1] Early detection, more effective systemic therapy, and surgical and locoregional treatments prolong overall survival and improve the quality of life in patients with colorectal liver metastases (CRLM).[2]

Imaging plays a crucial role in the diagnosis and follow-up of patients with CRC.[3] In particular, computed tomography (CT) is widely used for preoperative clinical evaluation, providing a noninvasive assessment of the number, size, and vascular anatomy of CRLM. At the same time, magnetic resonance imaging (MRI) with hepatospecific contrast medium has the highest diagnostic accuracy, sensitivity, and specificity, providing additional information on biliary anatomy.[4] Nonetheless, these imaging techniques are unsatisfactory for defining patient stratification and outcomes. At present, prognostic stratification relies mainly on molecular profiling (*RAS* and *BRAF* mutations, Microsatellite Instability (*MSI*) status, and *HER2* amplifications) and clinical, biochemical (carcinoembryonic antigen or CEA levels) tumor, and pathological variables (tumor size, nodal status, number, and size of CRLM; growth pattern; and resection margin). Quantitative risk models have been established to increase prediction accuracy, and proteomic and microRNA biomarkers may enhance biological-driven prognostic stratification. However, genetic profiling is costly, time-consuming, and not readily available at diagnosis. At the same time, other important prognostic factors are only available after resection of CRLM, thus proving useless for patients' selection for surgical procedures and other treatment decisions. The development of quantitative imaging biomarkers directly related to clinicopathological and biological features may provide a readily available and versatile tool for an accurate and noninvasive prediction of patients' risk of progression and prognosis.[5]An increasingly important role in radiology is represented by radiomics, an emerging field in which a large amount of quantitative information can be extracted through dedicated computer software, including measurements of intensity, texture, and shape.[6] In particular, texture analysis examines pixels based on the analysis of the distribution of coarseness and irregularity of gray-level pattern intensities within a region of interest, quantifying tissue heterogeneity. The use of machine learning (ML) algorithms allows the integration of this large number of quantitative imaging features with clinical data so that texture features can be correlated with pathology data and outcomes to gain a deeper insight into tumor biology and to refine prognostic evaluation and outcome prediction.[7,8]

A particular branch of this research is represented by radiogenomics, whose task is to discover the radiomics features associated with gene expression or polymorphism related to cancer development and progression. In patients with CRLM, a radiomics-based approach, combined with the most relevant clinical and genetic factors at baseline, may significantly contribute to identifying early predictors of risk of recurrence and overall survival following surgical resection.[9] Even though several studies have already reported significant correlations between the characteristics of liver texture, derived from CT or magnetic resonance (MR)-based radiomics, and the clinical outcomes in patients with CRLM,[9–12] further research should be encouraged to define better prediction tools and promote their translation in clinical practice.

Our study aims to build three CT-based radiomics classification models, with independent external validation, to predict (1) recurrence, (2) local recurrence, and (3) disease-specific mortality in patients affected by CRLM who underwent liver resection.

## MATERIALS AND METHODS

### *Patient Population*

This study was approved by the institutional review board and conducted following the ethical statements of the Declaration of Helsinki (approval no. RS1765/22).

From September 2012 to January 2021, a database of patients with metastatic colorectal cancer with at least one measurable lesion (greater diameter $\geq$ 5 mm) was consecutively identified at the National Cancer Institute Regina Elena (IRE) in Rome (Italy). Owing to the study's retrospective nature, signed informed consent was waived.

The study population met the following inclusion criteria: (1) presence of histologically confirmed synchronous or metachronous CRLM; (2) no previous liver surgery; (3) curative-intent resection of CRLM; (4) availability of a primary staging CT including at least a portal venous phase (PVP); and (5) availability of clinical and radiological follow-up for at least 6 months after surgery. The exclusion criteria were: (1) primary staging CT not performed at our institution and (2) poor image quality due to motion artifacts or low resolution.

The external validation cohort was built from the surgical series of two external institutions: the Gastrointestinal Surgery and Liver Transplantation Unit of the National Cancer Institute of Milan (Italy) and the Regional Center for Hepatopancreatobiliary (HBP) Surgery of the Regional Hospital of Treviso (Italy). This cohort met the same inclusion and exclusion criteria.

*Neoadjuvant Chemotherapy and Surgical Procedures*

Each patient was evaluated by a multidisciplinary team. Several underwent neoadjuvant chemotherapy, with protocols determined by the oncologist on the basis of the tumor's molecular features, the patient's age, comorbidities, and performance status. The regimens were combinations of 5-fluorouracil/folinic acid with oxaliplatin (FOLFOX) or irinotecan (FOLFIRI) plus molecular target agents (anti-vascular endothelial growth factor (*VEGF*) or anti-fibroblast growth factor receptor (*FGFR*) monoclonal antibodies).

Subsequently, all patients underwent surgery, with the extent of resection ranging from wedge resection to major hepatectomy, according to the Brisbane classification.[13] The extent of resection was preoperatively estimated by the multidisciplinary team and precisely assessed during the procedure on the basis of the intraoperative ultrasound findings.

The timing of surgery for synchronous metastases (defined as metastases present at the time of diagnosis of the primary tumor or detected within 6 months after that) was determined on the basis of multidisciplinary consensus; the preferred strategy—when feasible—was to perform hepatic resection as early as possible, either through a "liver-first" or "simultaneous" approach.

Lesions that were no longer visible after chemotherapy (vanishing metastases) were treated with an aggressive surgical approach, guided by pretreatment imaging and routine intraoperative ultrasound. If the site was no longer identifiable, close surveillance with short-interval imaging was considered. Their eventual reappearance during follow-up was classified as local recurrence.

All resection margins were free of residual tumor (R0). After surgery and further multidisciplinary evaluation, several patients received adjuvant chemotherapy based on final pathology and patients' related features.

During follow-up, each patient underwent clinical examination and laboratory tests every 3 months during the first 3 years and every 6 months thereafter; CT scans were performed at 3 months and then every 6 months after surgery.

*Clinical Endpoints*

Recurrence was defined as a new lesion detected at the CT scans performed after liver resection; local recurrence was defined as an intrahepatic recurrence after the first liver resection. The time of recurrence/local recurrence (in months) was calculated from the date of CRLM surgery (or the date of the bowel surgery in the case of synchronous colon and liver surgery) to the date of the first recurrence/local recurrence. Early recurrence was defined as recurrence within 6 months of liver surgery.[14]

The follow-up time was calculated from the date of liver surgery to the date of the last follow-up or death. At the last follow-up, each patient was recorded as alive or alive with disease; the event of death was recorded as death due to disease or death due to other causes.

*CT Acquisition Protocol*

CT acquisition was performed using a variety of multidetector-row CT scanners from different manufacturers (Philips, Brilliance 64, Philips Medical Systems, Best, The Netherlands; Siemens Somatom Definition Flash, Siemens Healthcare, Erlangen, Germany; and GE Lightspeed VCT, GE Medical Systems, Waukesha, WI, USA). The PVP was obtained with a tube voltage of 120 kV and a tube current covering 300–500 mA with a scan delay set at 70 s. Weight-adapted contrast medium was applied intravenously at a rate of 2.0–3.0 ml/s, followed by a saline flush of 40 ml. Images were acquired with a large field of view, a slice thickness range from 2.0 to 5.0 mm, and a pitch of 0.5–0.7 mm; subsequently, a reconstruction algorithm was applied with a standard soft tissue kernel.

Similarly, in Milan and Treviso, the CT study was performed using various multidetector-row CT scanners with acquisition parameters comparable to those used in the training cohort. As detailed below, harmonization between different CT scanners was applied at the feature level using an appropriate correction method in the training and validation cohorts.[15]

*Lesion Delineation*

The volumes of interest (VOIs) used for radiomics analysis were delineated with the Pinnacle3 (Philips Healthcare Nederland B.V.) treatment planning system, version 16.2.1. The Pinnacle3 software runs on a Solaris UNIX (or UNIX-compliant) workstation. It comprises several modules for elaborating a radiation therapy plan, including the contouring module for targets and organs at risk delineation.

Up to five visible metastases, the whole liver volume, and the surrounding free-disease parenchymal liver were separately delineated on the PVP at the primary staging CT. We chose a five-lesion cutoff to balance segmentation load and information redundancy. A radiologist manually contoured the single lesions using the paintbrush tool. By contrast, the liver contour was outlined with the Pinnacle Auto-Segmentation tool, based on the application of probability-based atlases and a segmentation algorithm to the image set. The radiologist then freehand refined these contours. Vanishing metastases on preoperative CT were not included in the segmentations.

For each patient, a VOI named liver background (BKG) was obtained by cropping the contours of the whole Liver, giving an inner margin of 5 mm from the liver surface and the lesions' borders, as illustrated in Fig. 1.

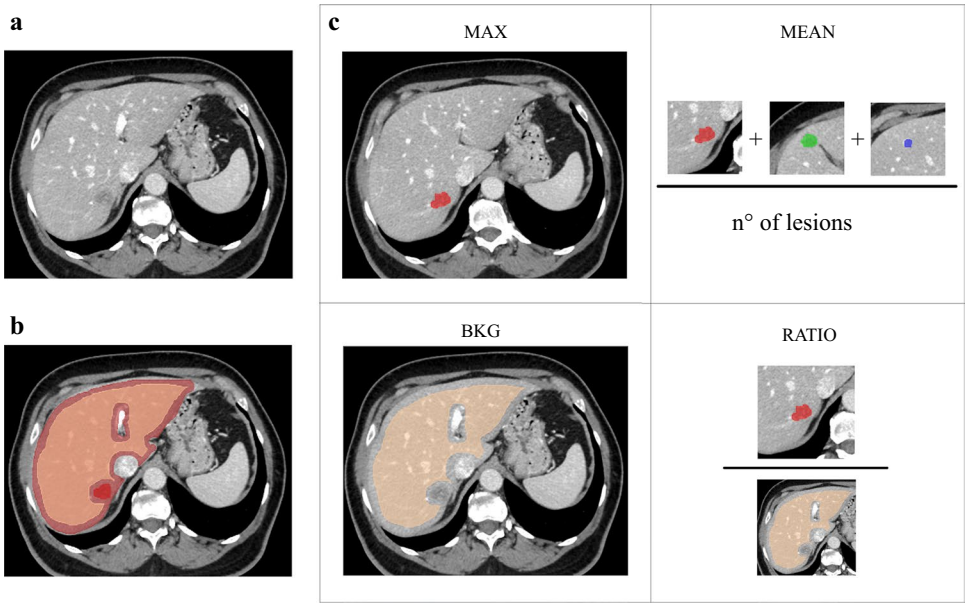


**FIG. 1** CT image of a patient with isolated CRLMs (**A**); the delineation of a lesion (red), the residual liver parenchyma background (orange) overlaid on the whole liver (brown), with the exclusion of large portal and hepatic veins (**B**); the four different sets of features derived from each patient (**C**): from the largest lesion (MAX), averaging the features of each lesion between them (MEAN), from the residual liver parenchyma background (BKG), and from the ratio between the features of the largest lesion and those of the residual liver parenchyma background (RATIO)

*Statistics*

We calculated the descriptive statistics for all variables of interest: we reported continuous variables with their means and standard deviations (± SD), median and relative interval (min–max), and categorical variables with absolute and percentage frequencies. The Kolmogorov–Smirnov normality test was performed for all the continuous variables. To explore the differences between continuous variables, we performed the Student's *t*-test or Mann–Whitney test, according to the nature of the variables. We assessed the relationships between categorical variables when appropriate using the chi-squared or Fisher's exact test. The Kaplan–Meier (KM) product-limit method was used to estimate survival curves. A $p$ value $< 0.05$ was considered statistically significant. Statistical analyses were conducted using SPSS version 29.0 (SPSS Inc., Chicago, IL, USA).

*Extraction of Radiomic Features*

CT images and structure set files were imported into S-IBEX software[16] to calculate the radiomic features using the DICOM-RT format. The S-IBEX software is a standardized version of image biomarker explorer (IBEX) software,[17] adapted and validated consistently with the guidelines of the Image Biomarker Standardization Initiative (IBSI).[18]

Specifically, seven families of features were extracted from a three-dimensional (3D) analysis of CT images: intensity direct (18 features), intensity histogram (21 features), gray-level co-occurrence matrix or GLCM (25 features), gray-level run-length matrix or GLRLM (16 features), gray-level size-zone matrix or GLSZM (16 features), neighboring gray-level dependence or NGLD (15 features), and morphology (23 features), for a total of 134 features for each VOI. A detailed list of the features and the image preprocessing steps (interpolation, resegmentation, and intensity discretization) are reported in Supplementary Tables S1 and S2. Specifically, the voxel size was resampled to $1 \times 1 \times 1$ mm$^3$ (isotropic voxel) to reduce the acquisition-related feature variability.

For each patient, four different sets of features were derived: (1) from the largest lesion (MAX), (2) averaging the features of each lesion between them (MEAN), (3) from the residual liver parenchyma background (BKG), and (4) from the ratio between the features of the largest lesion and those of the residual liver parenchyma background (RATIO), as shown in Fig. 1.

The harmonization between different CT scanners was performed at the feature level using the ComBat algorithm[15] via the package neuroCombat in R Studio (RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA, USA, URL http://www.rstudio.com), as shown in Supplementary Fig. S1. The outliers, i.e., data that were more than three median absolute deviations away from the median, were removed from the value distribution of each feature to improve the reliability of radiomics.[19]

*Model Building Driven by Machine Learning Algorithms*

Three different endpoints were considered to build the classification models: (1) presence of recurrence (0/1), (2) presence of local recurrence (0/1), and (3) disease-specific mortality (0/1).

Figure 2 illustrates the model-building pipeline. For the prediction models of recurrence and local recurrence, as the patient groups in the training cohort were not equally

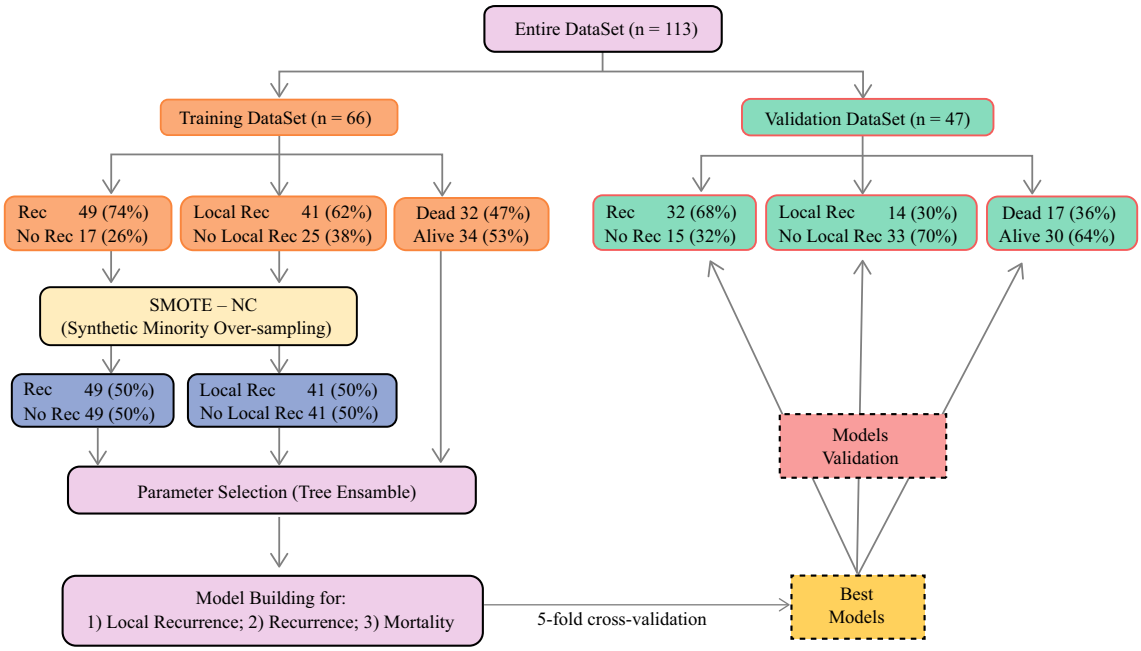


**FIG. 2** The pipeline shows the main steps in building the three prediction models for recurrence, local recurrence, and disease-related mortality

populated (49 versus 17 and 41 versus 25, respectively), an oversampling technique was used to generate synthetic data for the minority class to mitigate the data imbalance during the training,[20] by using the RSBID package in RStudio.

Both radiomic features and clinical parameters were considered in the selection process. The most significant predictors were initially selected using the Mann–Whitney test or chi-squared test with a cutoff for $p$ of 0.15, as appropriate. A second feature selection was performed using an ensemble of 100 classification trees to further reduce the number of relevant predictors.[21] Feature selection was approached with both conventional statistical methods and ML algorithms to infer the relationships between variables and optimize the prediction model's performance at the same time, as suggested by literature.[22] To explore the potential impact of this filtering strategy on model performance and to support its validity, we performed exploratory analyses comparing the performance of recurrence prediction models with and without the initial univariate filtering step (Supplementary Fig. S2).

Different families of ML algorithms, including decision tree, linear discriminant, logistic regression, naive Bayes (NB), support vector machine (SVM), $k$-nearest neighbor classifiers, and ensemble classifiers, were trained on our dataset and compared. An iterative optimization process was used to select the most appropriate hyperparameters for each algorithm for model building. A stratified fivefold cross-validation was applied to limit overfitting.

After training ML algorithms,[7] the model performance was assessed using standard metrics derived from the confusion matrix, including accuracy, sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV). The binary class labels (0/1) were obtained by applying the default decision rules. Specifically, NB predictions were based on the highest probability (i.e., a 0.5 threshold for binary classification), while SVM predictions relied on the sign of the decision function.

The confusion matrix is a tabular summary that compares the predicted labels against the true labels, displaying the number of true positives, true negatives, false positives, and false negatives.

The area under the receiver operating characteristic curve (AUC) was also used to evaluate the model's performance. The bias-corrected and accelerated percentile bootstrap method, using 1000 replicates, was applied to estimate the confidence interval for the AUC. The mid-$p$-value McNemar test was applied to compare prediction accuracies between different models.

Considering the limited size and the number of events (i.e., recurrences, local recurrences, and deaths) of the training cohort, we chose the models on the basis of the best triplet of predictors in terms of accuracy among all the possible combinations of the most relevant predictors selected above.

The MATLAB Statistics and Machine Learning Toolbox was used to create the ML-based models.

*Model Performance Evaluation for Early versus Late Recurrence*

To evaluate the predictive models' discriminative performance regarding the timing of recurrence, patients with recurrence were stratified into two groups: early recurrence within 6 months of Liver surgery and late recurrence beyond 6 months. Model outputs were compared against the observed recurrence status to construct confusion matrices and compute performance metrics within each subgroup, including accuracy, sensitivity, and specificity. The same analyses were performed for the groups of patients with early and late local recurrences.

## RESULTS

An internal study population of 66 patients met the inclusion and exclusion criteria and was used as a training cohort, while the validation cohort comprised 47 patients from two external institutions. Patient and tumor characteristics of the training and validation cohorts are reported in Table 1. Although a minimum of 6 months was required for inclusion, the median follow-up in the training cohort exceeded 2 years, ensuring the adequate capture of recurrence events.

The median time to recurrence was 10.9 months (95% confidence interval (95%CI) 7.5–14.2) and 13.6 (95%CI 9.8–17.3) for the training and validation cohort, respectively; the median time to local recurrence was 13.9 months (95%CI 8.0–19.7) for the training cohort, while it was not reached in the validation cohort.

The median time to disease-specific mortality was 39.5 months (95%CI 25.2–53.7) for the training cohort, while it was not reached in the validation cohort. The corresponding KM curves are reported in Supplementary Figs. S3, S4, and S5.

As a longer follow-up time was found in the external validation cohort compared with the training cohort (36.5 vs. 23.6 months, $p = 0.018$), we decided not to propose time-to-recurrence prediction models or disease-specific survival models but only classification models for recurrence, local recurrence, and disease-related mortality.

*Performance of the Models for Recurrence and Local Recurrence*

Following exploratory analyses, we found that the hybrid feature selection approach—consisting of initial univariate filtering followed by ensemble-based selection—did not significantly compromise the final model performance for recurrence compared with the ML-only approach. Although a few solutions from the ML-only approach exhibited slightly higher accuracy, the top-performing results from both strategies substantially overlapped, as shown in Supplementary Fig. S2.

Among the different families of ML algorithms, the SVM and NB classifiers provided the best predictive performance

**TABLE 1** Patient and tumor characteristics of the training and validation cohorts

| | Training cohort $n = 66$ | Validation cohort $n = 47$ | $p$ value |
|---|---|---|---|
| Age (years) | 64.0 (34.0, 86.0) | 68.1 (45.0, 92.0) | 0.166* |
| Gender (F/M) | 23/43 | 16/31 | 0.929** |
| Fup (months) | 23.6 (0.3, 99.9) | 36.5 (1.3, 83.0) | **0.018**§ |
| Recurrences (0/1) | 17/49 | 15/32 | 0.516** |
| Local recurrences (0/1) | 25/41 | 33/14 | < **0.001**** |
| Disease-specific death (0/1) | 32/34 | 30/17 | 0.076** |
| Status (1/2/3/4)* | 19/12/34/1 | 19/11/17/0 | 0.306** |
| PNI | 48.6 (37.3, 62.8) | 48.3 (32.8, 59.0) | 0.258* |
| FONG (0/1/2/3/4) | 1/10/18/20/17 | 2/8/15/17/4 | 0.220** |
| CEA (ng/ml) | 11.4 (1.6, 175.4) | 5.2 (0.2, 252.0) | 0.137§ |
| NLR | 2.5 (0.7, 9.2) | 2.3 (0.5, 4.5) | 0.780§ |
| PLR | 124.1 (35.9, 395.8) | 150.5 (46.7, 569.2) | 0.056§ |
| Neoadj CHT (No/Yes) | 33/33 | 26/21 | 0.577** |
| Number of lesions | 2 (1, 15) | 2 (1, 14) | 0.07§ |

*Student's *t*-test, §Mann–Whitney test, **chi-squared test

Status: 1, no evidence of disease; 2, alive with disease; 3, died of disease; 4, died of other cause

Significant *p*-values in bold.

*Fup* follow-up, *PNI* prognostic nutritional index, *FONG* scoring system developed by Fong et al. [23], *CEA* carcinoembryonic antigen, *NLR* neutrophil-to-lymphocyte ratio, *PRL* platelet-to-lymphocyte ratio, *Neoadj CHT* neoadjuvant chemotherapy

for the three classification models, as illustrated in Fig. 3 and reported in more detail in Tables 2, 3, and 4. The summary statistics of all the predictors included in the models are indicated in Supplementary Tables S3, S4, and S5.

The prediction models for recurrence showed fair-to-good accuracy, ranging from 70 to 78% and from 66 to 70% in the training and validation sets, respectively. The best models included radiomics from GLRLM and GLSZM matrices and clinical factors; specifically, patients with recurrences had more lesions, with a larger proportion receiving neoadjuvant chemotherapy. The model derived from the largest lesion (MAX) provided good sensitivity and specificity in the training and validation cohorts. By contrast, when transferred to the validation cohort, the MEAN and RATIO models showed a substantial reduction in accuracy, even though no statistically significant differences were found between these models in the training and validation cohorts (Supplementary Table S6). The model derived from the BKG showed the lowest accuracies in both cohorts (70% and 66%), but it provided the highest specificity when transferred to the validation cohort (80%).

The accuracy of the models for predicting the local recurrence was 60–78% and 66–72% in the training and validation sets, respectively. They included predictors from GLSZM and NGLD matrices and shape features alone or combined with the number of lesions. The model derived from MAX had very good specificity but poor sensitivity, while those derived from MEAN, BKG, and RATIO showed comparable accuracy but higher sensitivity. In the training cohort, statistically significant differences were found between the models derived from BKG and MAX ($p = 0.035$) and BKG and RATIO ($p = 0.008$),

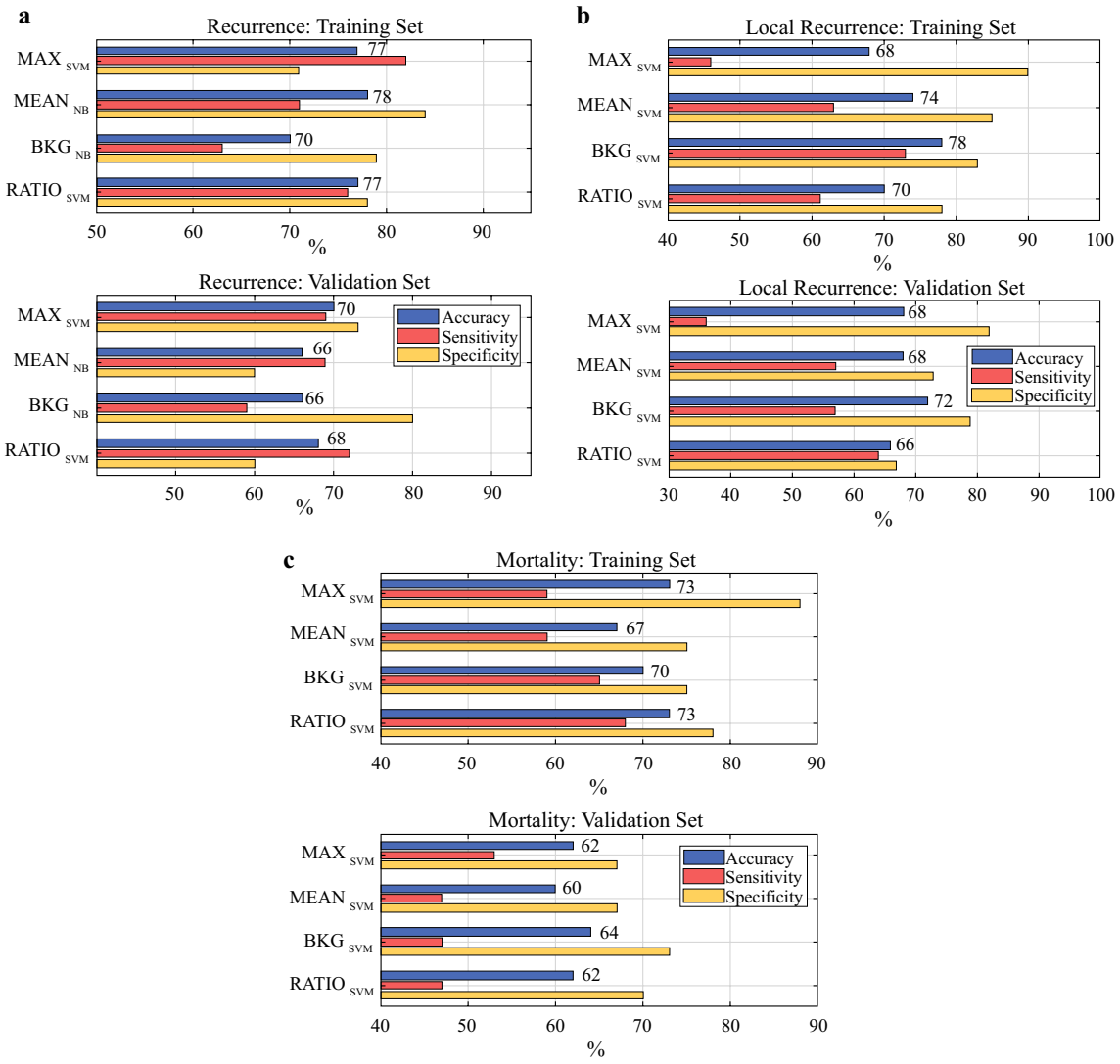


**FIG. 3** Performance of the best models for predicting recurrence (**A**), local recurrence (**B**), and disease-specific mortality (**C**) in the training and validation sets based on the four different feature sets

**TABLE 2** Performance of prediction models for recurrence on the training and validation cohort (in bold)

| Selected features | AUC | Accuracy (%) | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) |
|---|---|---|---|---|---|---|
| $MAX_{SVM}$ | 0.86 | 77 | 82 | 71 | 74 | 80 |
| (1) GLRLM run entropy | [0.77, 0.92] | [67,85] | [68,91] | [57, 83] | [64, 82] | [68, 88] |
| (2) GLSZM small-zone low-gray-level emphasis | **0.71** | **70** | **69** | **73** | **85** | **52** |
| (3) Number of lesions | **[0.55, 0.83]** | **[55,83]** | **[50, 84]** | **[45, 92]** | **[70, 93]** | **[38, 67]** |
| $MEAN_{NB}$ | 0.74 | 78 | 71 | 84 | 81 | 75 |
| (1) Intensity direct energy | [0.63, 0.84] | [64,85] | [57, 83] | [70, 93] | [65, 89] | [65, 82] |
| (2) Neoadjuvant chemotherapy | **0.64** | **66** | **69** | **60** | **79** | **47** |
| (3) Intensity direct skewness | **[0.48, 0.78]** | **[51,79]** | **[50, 84]** | **[32, 84]** | **[65, 88]** | **[32, 64]** |
| $BKG_{NB}$ | 0.78 | 70 | 63 | 78 | 74 | 68 |
| (1) Neoadjuvant Chemotherapy | [0.69, 0.87] | [60,79] | [48, 77] | [63, 88] | [62, 82] | [59, 76] |
| (2) Number of lesions | **0.70** | **66** | **59** | **80** | **86** | **48** |
| (3) GLSZM size zone entropy | **[0.52, 0.81]** | **[51,79]** | **[41, 76]** | **[52, 96]** | **[69, 95]** | **[36, 60]** |
| $RATIO_{SVM}$ | 0.87 | 77 | 76 | 78 | 77 | 76 |
| (1) GLRLM run entropy | [0.79, 0.92] | [67, 85] | [61, 857] | [63, 88] | [66, 85] | [65, 84] |
| (2) Intensity direct max | **0.66** | **68** | **72** | **60** | **79** | **50** |
| (3) Neoadjuvant chemotherapy | **[0.51, 0.80]** | **[53 81]** | **[53, 86]** | **[32, 84]** | **[67, 88]** | **[33, 67]** |

Squared brackets indicate the 95% confidence interval

*GLRLM* gray-level run-length matrix, *GLSZM* gray-level size-zone matrix, *SVM* support vector machine algorithm, *NB* naive Bayes algorithm, *AUC* area under the receiver operating characteristic curve, *PPV* positive predictive power, *NPV* negative predictive power

as also confirmed in the validation cohorts (Supplementary Table S7). This finding indicated a superiority of the BKG-based model compared with the others in predicting local recurrence, which was obtained by including only radiomic features.

### *Performance of the Models for Disease-Specific Mortality*

Models for predicting disease-specific mortality performed worse, with accuracies ranging from 67 to 73% and from 60 to 64% in the training and validation sets, respectively. Similar to previous models, sensitivity values were lower than the specificity ones; radiomic features from intensity direct, intensity histogram, and shape families were included, together with the maximum surgical diameter. No statistically significant differences were found between these models in the training and validation cohorts (Supplementary Table S8).

### *Model Performance for Early Versus Late Recurrence*

In the training set, 22 out of 49 patients (45%) experienced an early recurrence, while 27 (55%) had a late recurrence. In the validation set, 13 out of 32 patients (41%) had an early recurrence, and 19 (59%) had a late recurrence. Similarly, for local recurrence, 19 out of 41 patients (46%) in the training set had an early event, while 22 (54%) had a late local recurrence. In the validation set, early and late local recurrences were equally distributed, with 7 out of 14 patients (50%) in each group.

Confusion matrices illustrating the performance of the proposed classification models in these groups of patients are depicted in Supplementary Figs. S6 and S7, with the other performance metrics reported in Supplementary Tables S9 and S10 and Supplementary Figs. S8 and S9.

In the training set, the accuracy for early recurrence ranged from 67 to 74%, and from 57 to 70% for late recurrence, depending on the four different feature sets. In the

**TABLE 3** Performance of prediction models for local recurrence on the training and validation cohort (in bold)

| Selected features | AUC | Accuracy (%) | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) |
|---|---|---|---|---|---|---|
| $MAX_{SVM}$<br>(1) GLSZM small-zone low-gray-level emphasis<br>(2) GLSZM large-zone high-gray-level emphasis<br>(3) Number of lesions | 0.67<br>[0.54, 0.78]<br>**0.59**<br>**[0.46, 0.75]** | 68<br>[57, 78]<br>**68**<br>**[53, 81]** | 46<br>[31, 63]<br>**36**<br>**[13, 65]** | 90<br>[77, 97]<br>**82**<br>**[65, 93]** | 83<br>[64, 93]<br>**45**<br>**[23,70]** | 63<br>[55, 69]<br>**75**<br>**[66,82]** |
| $MEAN_{SVM}$<br>(1) Center of mass shift<br>(2) NGLD normalized dependence count nonuniformity<br>(3) Number of lesions | 0.79<br>[0.66, 0.88]<br>**0.65**<br>**[0.49, 0.79]** | 74<br>[64, 83]<br>**68**<br>**[53, 81]** | 63<br>[47, 78]<br>**57**<br>**[29,82]** | 85<br>[71, 94]<br>**73**<br>**[54, 87]** | 81<br>[67, 90]<br>**47**<br>**[30, 65]** | 70<br>[60, 78]<br>**80**<br>**[68, 88]** |
| $BKG_{SVM}$<br>(1) NGLD high-dependence high-gray-level emphasis<br>(2) NGLD dependence count nonuniformity<br>(3) Elongation | 0.79<br>[0.68, 0.88]<br>**0.73**<br>**[0.59,0.87]** | 78<br>[68, 86]<br>**72**<br>**[57, 84]** | 73<br>[57, 86]<br>**57**<br>**[29, 82]** | 83<br>[68, 93]<br>**79**<br>**[61, 91]** | 81<br>[68, 90]<br>**53**<br>**[34, 72]** | 76<br>[65, 84]<br>**81**<br>**[70, 89]** |
| $RATIO_{SVM}$<br>(1) GLRLM run entropy<br>(2) NGLD normalized dependence count nonuniformity<br>(3) GLCM information correlation | 0.80<br>[0.68, 0.87]<br>**0.66**<br>**[0.49,0.81]** | 72<br>[61, 81]<br>**66**<br>**[51, 79]** | 63<br>[47, 78]<br>**64**<br>**[35, 87]** | 80<br>[65, 91]<br>**67**<br>**[48, 82]** | 76<br>[63, 86]<br>**45**<br>**[31, 60]** | 69<br>[59, 77]<br>**81**<br>**[68, 90]** |

Squared brackets indicate the 95% confidence interval

*GLCM* gray-level co-occurrence matrix, *GLRLM* gray-level run-length matrix, *GLSZM* gray-level size-zone matrix, *NGLD* neighborhood gray-level dependence, *SVM* support vector machine algorithm, *AUC* area under the receiver operating characteristic curve, *PPV* positive predictive power, *NPV* negative predictive power

validation set, accuracy ranged from 64 to 71% for early recurrence and from 65 to 71% for late recurrence.

Similarly, in the training set, the accuracy for early local recurrence ranged from 68 to 77% and from 66 to 77% for late local recurrence. In the validation set, it ranged from 65 to 78% for early local recurrence and from 68 to 75% for late local recurrence. Notably, the sensitivity of the local recurrence models decreased when applied to the subgroup of patients with late local recurrence, except for the model based on the RATIO features.

## DISCUSSION

In this study, we proposed CT-based radiomic models, including clinical data at baseline, to noninvasively predict recurrence, local recurrence, and disease-specific mortality

**TABLE 4** Performance of prediction models for disease-specific mortality on the training and validation cohort (in bold)

| Selected features | AUC | Accuracy (%) | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) |
|---|---|---|---|---|---|---|
| $MAX_{SVM}$ (1) GLCM sum variance (2) Major axis length (3) Maximum surgical diameter | 0.73 [0.59, 0.85] | 73 [60, 83] | 59 [41, 75] | 88 [71, 96] | 83 [66, 93] | 67 [57, 75] |
| | **0.60 [0.45, 0.74]** | **62 [46, 75]** | **53 [28, 77]** | **67 [47, 83]** | **47 [31, 64]** | **71 [59, 81]** |
| $MEAN_{NB}$ (1) Center of mass shift (2) Intensity histogram maximum (3) Maximum surgical diameter | 0.67 [0.52, 0.78] | 67 [54, 78] | 59 [41, 75] | 75 [57, 89] | 71 [56, 83] | 63 [52, 73] |
| | **0.57 [0.43, 0.72]** | **60 [44, 74]** | **47 [23, 72]** | **67 [47, 83]** | **44 [28, 62]** | **69 [57, 79]** |
| $BKG_{NB}$ (1) Intensity histogram mean absolute deviation (2) Center of mass shift (3) Intensity direct energy | 0.73 [0.58, 0.84] | 70 [57, 80] | 65 [46, 80] | 75 [57, 89] | 73 [59, 84] | 67 [55, 77] |
| | **0.60 [0.45, 0.74]** | **64 [49, 77]** | **47 [23, 72]** | **73 [54, 88]** | **50 [31, 69]** | **71 [60, 80]** |
| $RATIO_{SVM}$ (1) Intensity direct energy (2) Asphericity (3) Maximum surgical diameter | 0.76 [0.64, 0.87] | 73 [60, 83] | 68 [50, 83] | 78 [60, 91] | 77 [62, 87] | 69 [57, 79] |
| | **0.59 [0.44, 0.73]** | **62 [46, 75]** | **47 [23, 72]** | **70 [51, 85]** | **47 [30, 65]** | **70 [58, 79]** |

Squared brackets indicate the 95% confidence interval

*GLCM* gray-level co-occurrence matrix, *SVM* support vector machine algorithm, *NB* naive bayes algorithm, *AUC* area under the receiver operating characteristic curve, *PPV* positive predictive power, *NPV* negative predictive power

after hepatic resection for colorectal cancer liver metastases. Early identification of high-risk patients could enhance postoperative surveillance and allow more tailored postsurgical therapies, improving overall patient outcomes.

The performances of these classification models were tested on an independent external validation cohort to measure the transportability and generalization of our findings, as strongly suggested by international guidelines.[6] The variability in CT acquisition protocols, methods of image analysis, or patient management among different institutions may impair the study validation, as recently reported by Van der Reijd et al.;[24] consequently, the reproducibility of the radiomics models' performance should be assessed in different settings.[25]

As advised by previous investigators,[26–30] the radiomic analysis was not limited to the largest CRLM, but it was extended to multiple lesions (up to five by averaging the features of each lesion) and to the residual liver parenchyma background, also exploring the potential of the ratio between the features of the largest lesion and those of the surrounding liver parenchyma.[31]

Structural changes within the liver parenchyma due to angiogenesis and extracellular matrix remodeling have been described in patients with CRLM. These changes may capture biological characteristics of the evolution of metastatic disease that cannot be detected on the basis of the analysis of liver lesions or the surrounding liver separately.[29–31] The parenchymal texture analysis might suggest the presence of occult metastases or other underlying parenchymal changes, which can be associated with a worse prognosis in patients with CRLM.

The study revealed that models for predicting recurrence and local recurrence based on liver residual parenchyma (BKG) were superior in specificity when transferred to

the validation cohort than the other models. Moreover, the BKG-based model better predicted local recurrence than the others, with the liver residual parenchyma showing an increased elongation and a higher value of the NGLD high-dependence high-gray-level emphasis in patients experiencing a relapse. This suggests that the residual liver parenchyma had larger and more uniform high-gray-level regions in these patients.

To be noted, in patients with a higher risk of recurrence/local recurrence and disease-specific mortality, the largest lesion showed higher heterogeneity and greater variability in the intensity patterns, as consistently indicated by increased values of GLRLM run entropy, GLSZM large-zone high-gray-level emphasis, and GLCM sum variance. Several studies suggested that a high value of heterogeneity and entropy correlated with tumor aggressiveness.[9,11,32–35] In particular, the heterogeneity of the tumor is a significant challenge in the treatment of metastatic colorectal cancer, and it is associated with a low response rate to treatment and, eventually, a worse overall survival after surgery.[28] Chun et al.[36] consistently showed that a transition from a heterogeneous tumor appearance to a homogeneous tumor appearance at preoperative CT scan following neoadjuvant chemotherapy was associated with pathologic response and improved survival.

According to earlier investigations, our models confirmed the predictive role of the number of lesions and their dimension, evaluated in terms of major axis length or maximum surgical diameter.[2,5,10] At the same time, neoadjuvant chemotherapy emerged as one of the top three predictors in our recurrence models; this can be explained by considering that neoadjuvant chemotherapy is more commonly administered to patients with high-risk disease at presentation—i.e., multiple ($\geq 3$) or bilobar lesions, borderline future-liver-remnant, or synchronous primary-tumor surgery.[37] As such, the higher recurrence in this subgroup may reflect that these patients were already at higher risk, probably because of a more aggressive tumor biology which was not entirely mitigated by chemotherapy; this reinforces the potential value of radiomics in pretreatment risk assessment, possibly identifying nonresponders who may benefit from alternative or intensified treatment strategies.

Notably, models for predicting disease-related mortality performed worse than the others, with fair accuracies (60–64%), poor sensitivities, and fair-to-good specificities in the validation set. This can be explained by considering that patient survival can be affected by several factors that cannot be adequately represented by radiomics features or baseline clinical data, consistent with other investigators.[10]

According to other investigators, pretreatment *CEA* levels, *FONG* score, prognostic nutritional index (*PNI*), and platelet-to-lymphocyte ratio (*PRL*) were considered in the analyses but not included in the final models after the feature selection process.[2,5,38]

Despite being trained for binary classification tasks (i.e., recurrence and local recurrence), the models demonstrated comparable performance across early and late recurrence subgroups. They showed only a slight reduction in accuracy in the validation set and a decrease in sensitivity for the local recurrence models when applied to the subset of patients with late events, except for the model based on the RATIO features. Clinically, predicting early recurrence is particularly valuable, as early recurrence is often associated with more aggressive tumor biology, worse prognosis, and reduced therapeutic options.[14,38] Accurate identification of patients at high risk for early recurrence may support intensified surveillance strategies, guide perioperative treatment planning, and facilitate timely enrollment into clinical trials designed for high-risk populations. It should be noted, however, that the limited size of our dataset may have impaired the robustness of the predictive performance in stratified subgroups, which would have required the training of risk prediction models on larger patient cohorts.

The main result of our study was building naive predictive models from preoperative CT based on a few CRLM radiomic features. These models demonstrated fair-to-good accuracies for each classification task and outperformed the clinical models alone.

Our study has several limitations: the retrospective nature of the research and the small sample size of the patient population can have introduced biases and confounding factors. Several patients undergoing radical surgery for colorectal liver metastasis in our institution were excluded from the study because their first staging CT was performed in other hospitals, and this strongly limited the sample size of the training cohort. The significant difference in follow-up duration between the internal and external cohorts prevented us from proposing time-to-recurrence prediction models or disease-specific survival models, thus limiting our investigation to classification models. Moreover, the observed differences in local recurrence and disease-specific mortality between cohorts may also be attributed to additional factors, including potential selection bias, given that tertiary referral centers may treat more advanced cases, and center-specific management variability. Additionally, we could not include genomic alterations in our analyses due to the lack of data in both training and validation cohorts. Information about patients' mutational status could have improved our models' predictive ability, given the role of *KRAS/NRAS/BRAF* status as independent prognostic factors for early recurrence and overall survival.[2,38] Our exploratory analyses on feature selection strategies suggested that applying an initial filtering step followed by ensemble selection may facilitate the development of interpretable models at a low computational cost. However, a more comprehensive assessment of the model's performance, potentially involving alternative ML-based feature selection strategies, would require a

dedicated methodological investigation to be considered in future work. Lastly, further efforts are needed to standardize and automate the radiomic feature extraction process so that the ML-driven models can be widely implemented and integrated into the clinical workflow.

## CONCLUSIONS

CT-based radiomics, alone or in combination with baseline clinical data, allowed the prediction of overall and local recurrence and disease-specific mortality of patients with CRLM, with fair-to-good accuracy after validation in an external cohort.

A relevant role of second-order and higher-order texture features emerged from both the largest lesion and the liver residual parenchyma; these feature families allowed the quantification of image heterogeneous patterns, providing insights into irregular tissue structures and microvasculature and contributing to the early identification of high-risk patients. Further investigations involving larger patient cohorts with longer follow-up durations for training and validation are warranted to corroborate our findings.

**SUPPLEMENTARY INFORMATION** The online version contains supplementary material available at https://doi.org/10.1245/s10434-025-18248-y.

**DISCLOSURE** None of the authors have a potential conflict of interest, including any financial, personal, or other relationships with other people or organizations that could inappropriately influence this work. This work was partially supported by the Italian Ministry of Research under the complementary actions to the National Recovery and Resilience Plan (NRRP) "D34Health – Digital Driven Diagnostics, Prognostics, and Therapeutics for Sustainable Health Care" Grant (grant no. PNC0000001).

## REFERENCES

1. Siegel RL, Wagle NS, Cercek A, Smith RA, Jemal A. Colorectal cancer statistics, 2023. *CA Cancer J Clin*. 2023;73(3):233–54. https://doi.org/10.3322/caac.21772.
2. Buisman FE, Giardiello D, Kemeny NE, et al. Predicting 10-year survival after resection of colorectal liver metastases; an international study including biomarkers and perioperative treatment. *Eur J Cancer*. 2022;168:25–33. https://doi.org/10.1016/j.ejca.2022.01.012.
3. Caruso D, Polici M, Bellini D, Laghi A. ESR essentials: imaging in colorectal cancer—practice recommendations by ESGAR. *Eur Radiol*. 2024;34(9):5903–10. https://doi.org/10.1007/s00330-024-10645-3.
4. Renzulli M, Clemente A, Ierardi AM, et al. Imaging of colorectal liver metastases: new developments and pending issues. *Cancers*. 2020;12(1):151. https://doi.org/10.3390/cancers12010151.
5. Vadisetti SN, Kazi M, Patkar S, et al. Patterns and predictors of recurrence after curative resection of colorectal liver metastasis (CRLM). *J Gastrointest Cancer*. 2024;55(4):1559–68. https://doi.org/10.1007/s12029-024-01105-8.
6. Lambin P, Leijenaar RTH, Deist TM, et al. Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol*. 2017;14(12):749–62. https://doi.org/10.1038/nrclinonc.2017.141.
7. Erickson BJ, Korfiatis P, Akkus Z, Kline TL. Machine learning for medical imaging. *RadioGraphics*. 2017;37(2):505–15. https://doi.org/10.1148/rg.2017160130.
8. Inchingolo R, Maino C, Cannella R, et al. Radiomics in colorectal cancer patients. *World J Gastroenterol*. 2023;29(19):2888–904. https://doi.org/10.3748/wjg.v29.i19.2888.
9. Fiz F, Viganò L, Gennaro N, et al. Radiomics of liver metastases: a systematic review. *Cancers*. 2020;12(10):2881. https://doi.org/10.3390/cancers12102881.
10. Luo X, Deng H, Xie F, et al. Prognostication of colorectal cancer liver metastasis by CE-based radiomics and machine learning. *Transl Oncol*. 2024;47:101997. https://doi.org/10.1016/j.tranon.2024.101997.
11. Granata V, Fusco R, Barretta ML, et al. Radiomics in hepatic metastasis by colorectal cancer. *Infect Agents Cancer*. 2021;16(1):39. https://doi.org/10.1186/s13027-021-00379-y.
12. Lubner MG, Stabo N, Lubner SJ, et al. CT textural analysis of hepatic metastatic colorectal cancer: pre-treatment tumor heterogeneity correlates with pathology and clinical outcomes. *Abdom Imaging*. 2015;40(7):2331–7. https://doi.org/10.1007/s00261-015-0438-4.
13. Pang YY, Strasberg SM. The Brisbane 2000 terminology of liver anatomy and resections. *HPB* 2000; 2:333–39. *HPB*. 2002;4(2):99–100. https://doi.org/10.1080/136518202760378489
14. Tian Y, Wang Y, Wen N, Wang S, Li B, Liu G. Prognostic factors associated with early recurrence following liver resection for colorectal liver metastases: a systematic review and meta-analysis. *BMC Cancer*. 2024;24(1):426. https://doi.org/10.1186/s12885-024-12162-4.
15. Fortin JP, Parker D, Tunç B, et al. Harmonization of multi-site diffusion tensor imaging data. *NeuroImage*. 2017;161:149–70. https://doi.org/10.1016/j.neuroimage.2017.08.047.
16. Bettinelli A, Branchini M, De Monte F, Scaggion A, Paiusco M. Technical note: an IBEX adaption toward image biomarker standardization. *Med Phys*. 2020;47(3):1167–73. https://doi.org/10.1002/mp.13956.
17. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE. ibex: an open infrastructure software platform to facilitate collaborative work in radiomics. *Med Phys*. 2015;42(3):1341–53. https://doi.org/10.1118/1.4908210.
18. Whybra P, Zwanenburg A, Andrearczyk V, et al. The image biomarker standardization initiative: standardized convolutional filters for reproducible radiomics and enhanced clinical insights. *Radiology*. 2024;310(2):e231319. https://doi.org/10.1148/radiol.231319.
19. Park BW, Kim JK, Heo C, Park KJ. Reliability of CT radiomic features reflecting tumour heterogeneity according to image quality and image processing parameters. *Sci Rep*. 2020;10(1):3852. https://doi.org/10.1038/s41598-020-60868-9.
20. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: synthetic minority over-sampling technique. *Jair*. 2002;16:321–57. https://doi.org/10.1613/jair.953.
21. Wu X, Kumar V, Ross Quinlan J, et al. Top 10 algorithms in data mining. *Knowl Inf Syst*. 2008;14(1):1–37. https://doi.org/10.1007/s10115-007-0114-2.
22. Rajula HSR, Verlato G, Manchia M, Antonucci N, Fanos V. Comparison of conventional statistical methods with machine learning in medicine: diagnosis, drug development, and treatment. *Medicina*. 2020;56(9):455. https://doi.org/10.3390/medicina56090455.

23. Fong Y, Fortner J, Sun RL, Brennan MF, Blumgart LH. Clinical score for predicting recurrence after hepatic resection for metastatic colorectal cancer: analysis of 1001 consecutive cases. *Ann Surg*. 1999;230(3):309. https://doi.org/10.1097/00000658-199909000-00004.
24. Van Der Reijd DJ, Guerendel C, Staal FCR, et al. Independent validation of CT radiomics models in colorectal liver metastases: predicting local tumour progression after ablation. *Eur Radiol*. 2023;34(6):3635–43. https://doi.org/10.1007/s00330-023-10417-5.
25. Huang EP, O'Connor JPB, McShane LM, et al. Criteria for the translation of radiomics into clinically useful tests. *Nat Rev Clin Oncol*. 2023;20(2):69–82. https://doi.org/10.1038/s41571-022-00707-0.
26. Miles KA, Ganeshan B, Griffiths MR, Young RCD, Chatwin CR. Colorectal cancer: texture analysis of portal phase hepatic CT images as a potential marker of survival. *Radiology*. 2009;250(2):444–52. https://doi.org/10.1148/radiol.2502071879.
27. Simpson AL, Doussot A, Creasy JM, et al. Computed tomography image texture: a noninvasive prognostic marker of hepatic recurrence after hepatectomy for metastatic colorectal cancer. *Ann Surg Oncol*. 2017;24(9):2482–90. https://doi.org/10.1245/s10434-017-5896-1.
28. Montagnon E, Cerny M, Hamilton V, et al. Radiomics analysis of baseline computed tomography to predict oncological outcomes in patients treated for resectable colorectal cancer liver metastasis. *PLoS ONE*. 2024;19(9):e0307815. https://doi.org/10.1371/journal.pone.0307815.
29. Tabari A, D'Amore B, Noh J, Gee MS, Daye D. Quantitative peritumoral magnetic resonance imaging fingerprinting improves machine learning-based prediction of overall survival in colorectal cancer. *Explor Target Anti-tumor Ther*. 2024;5(1):74–84. https://doi.org/10.37349/etat.2024.00205.
30. Fiz F, Ragaini EM, Sirchia S, et al. Radiomic gradient in peritumoural tissue of liver metastases: a biomarker for clinical practice? Analysing density, entropy, and uniformity variations with distance from the tumour. *Diagnostics*. 2024;14(14):1552. https://doi.org/10.3390/diagnostics14141552.
31. Beckers RCJ, Trebeschi S, Maas M, et al. CT texture analysis in colorectal liver metastases and the surrounding liver parenchyma and its potential as an imaging biomarker of disease aggressiveness, response and survival. *Eur J Radiol*. 2018;102:15–21. https://doi.org/10.1016/j.ejrad.2018.02.031.
32. Wang Y, Ma LY, Yin XP, Gao BL. Radiomics and radiogenomics in evaluation of colorectal cancer liver metastasis. *Front Oncol*. 2022;11:689509. https://doi.org/10.3389/fonc.2021.689509.
33. Devoto L, Ganeshan B, Keller D, et al. Using texture analysis in the development of a potential radiomic signature for early identification of hepatic metastasis in colorectal cancer. *Eur J Radiol Open*. 2022;9:100415. https://doi.org/10.1016/j.ejro.2022.100415.
34. Li Y, Eresen A, Shangguan J, et al. Establishment of a new non-invasive imaging prediction model for liver metastasis in colon cancer. *Am J Cancer Res*. 2019;9(11):2482–92.
35. Han Y, Chai F, Wei J, et al. Identification of predominant histopathological growth patterns of colorectal liver metastasis by multi-habitat and multi-sequence based radiomics analysis. *Front Oncol*. 2020;10:1363. https://doi.org/10.3389/fonc.2020.01363.
36. Chun YS. Association of computed tomography morphologic criteria with pathologic response and survival in patients treated with bevacizumab for colorectal liver metastases. *JAMA*. 2009;302(21):2338. https://doi.org/10.1001/jama.2009.1755.
37. Carrion-Alvarez L, Primavesi F, Søreide K, et al. Liver metastases from colorectal cancer: a joint ESSO–EAHPBA–UEMS core curriculum collaboration. *Eur J Surg Oncol*. 2025;51(6):109728. https://doi.org/10.1016/j.ejso.2025.109728.
38. Fu S, Chen D, Zhang Y, et al. A CT-based radiomics tumor quality and quantity model to predict early recurrence after radical surgery for colorectal liver metastases. *Clin Transl Oncol*. 2024. https://doi.org/10.1007/s12094-024-03645-8.