

Title:

Prediction of Local Control for Colorectal Liver Metastases Treated with Radiotherapy using a Radiomic Artificial Intelligence Model

Word Count: Abstract: 250, Main Text: 2996 (From Introduction to end of Discussion, includes headers)

Authors and Affiliations:

Ricky Hu^{*1}, Ishita Chen^{*2}, Jacob Peoples³, Jean-Paul Salameh¹, Mithat Gönen⁴, Paul B. Romesser², Amber L. Simpson^{3,5}, and Marsha Reyngold²

***Equal Contribution and Co-First Authorship**

¹School of Medicine, Queen's University, Kingston, ON

²Department of Radiation Oncology, Memorial Sloan Kettering Cancer Center, New York, NY

³School of Computing, Queen's University, Kingston, ON

⁴Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY

⁵Department of Biomedical and Molecular Sciences, Queen's University, Kingston, ON

Corresponding Author: Ricky Hu

Email: rhu@qmed.ca

Main Affiliation Address: 557 Goodwin Hall, Queen's University, Kingston, Ontario, K7L 2N8

Declaration of Interest:

PBR is a EMD Serono consultant and reports support for travel from Elekta and Philips healthcare and prior research funding from EMD Serono.

Author Contributions:

Ishita Chen: Conceptualization, data acquisition, data processing, data analysis, and writing.

Ricky Hu: Methodology, software development, data analysis, visualization, and writing.

Jacob Peoples: Methodology, software development, data analysis, and writing.

Jean-Paul Salameh: Data analysis and writing.

Mithat Gönen: Biostatistical support, data analysis, and writing.

Paul B. Romesser: Supervision, data acquisition, data processing, data analysis, and writing.

Amber Simpson: Supervision, conceptualization, methodology, data analysis, and writing.

Marsha Reyngold: Supervision, data acquisition, data processing, data analysis, and writing.

Abstract

Background and Purpose

Prognostic assessment of local therapies for colorectal liver metastases (CLM) is essential for guiding management in radiation oncology. Existing models using clinicopathological variables have limited accuracy. Computed tomography (CT) contains liver texture information which may be predictive of metastatic environments. We sought to build an automated system to predict progression-free survival using CT radiomics and artificial intelligence (AI) modelling.

Materials and Methods

Liver CT scans and outcomes for N=97 CLM patients treated with radiotherapy were retrospectively obtained from Memorial Sloan Kettering Cancer Center. A survival model was built by extracting 108 radiomic features from liver and tumor CT volumes for a random survival forest (RSF) to predict local progression. Accuracies were measured by concordance indices (C-index) and integrated Brier scores (IBS) with 4-fold cross-validation. This was repeated with different liver segmentations and radiotherapy clinical variables as inputs to the RSF. Predictive features were identified by perturbation importances.

Results

The AI radiomics model achieved a C-index of 0.68 (CI: 0.62 – 0.74) and IBS below 0.25. The most predictive radiomic feature was gray tone difference matrix strength (importance: 1.90 CI: 0.93 – 2.86) and most predictive treatment feature was maximum dose (importance: 3.83, CI: 1.05 – 6.62). The clinical data only model achieved a similar C-index of 0.62 (CI: 0.56 - 0.69), suggesting that predictive signals exist in radiomics and clinical data.

Conclusions

The AI model achieves good prediction accuracy for progression-free survival of CLM, providing support that radiomics or clinical data combined with machine learning may aid prognostic assessment and management.

64 **Keywords:**

65

66 Radiomics, Artificial Intelligence, Machine Learning, Computer Vision, Survival Analysis,

67

1. Introduction:

Patients with colorectal cancer develop colorectal liver metastases (CLM) in approximately 50% of cases [1] with 40% recurring within 12-months. Surgery is the standard of care for patients that present with liver-limited resectable CLM, with reported 5-year survival ranging from 28-58% [2]. Non-surgical liver-directed local therapy for CLM, such as thermal ablation, can be effective, but is invasive [3]. External beam radiation therapy (EBRT) has emerged as an alternative, non-invasive approach for localized therapy of CLM in patients who are ineligible for other treatment options. Numerous clinical factors have been shown to influence the local control of CLM associated with EBRT, including dose delivered to the lesion and the size of the lesion [4]. Prognosis of local tumor control is essential to determine appropriate treatment for CLM. The goal of this work was to build a predictive model utilizing quantitative information from radiomic features contained in computed tomography (CT) scans combined with machine learning methods to predict local control outcomes in CLM.

Several models have been published for predicting clinical outcomes in CLM patients. A common approach is utilizing multivariate Cox proportional hazard regression using clinically relevant variables and selecting high hazard ratio variables for a scoring system [5-13]. Fong et al. [6], for instance, utilized 5 clinicopathological variables (node-positive primary, interval from primary to metastases, number of hepatic tumors, if largest hepatic tumor >5cm, and carcinoembryonic antigen level > 200ng/ml) for their scoring system. Wang et al. [14] evaluated the accuracy of 9 different survival prediction scoring systems with 6 of the scoring systems resulting in a concordance index (C-index) crossing below the 0.50 threshold in its 95% confidence interval. A C-index crossing 0.50 would indicate prediction no better than random chance.

There are limitations with these scoring systems, namely that scoring systems require manual thresholding, and that other data, such as CT imaging, may provide features that can be utilized for outcome prediction. We aim to address these gaps by building a survival prediction model from two motivating ideas: the usage of CT radiomic imaging data and modelling time-dependent survival with machine learning.

CT has been standard of care for characterizing tumor response and radiomics is an emerging field that shows promise in analyzing complex details in CT scans. Radiomic features are computed textural attributes of which quantitatively characterize shape, intensity statistics, and gray-level relationships within the anatomy of interest. Specific to liver metastases, Fiz et al. [15] report 32 different studies up to June 2020 evaluating the association of radiomics to overall survival, tumor size, or response evaluation

criteria, however the studies assess for association and did not measure predictive accuracy. Ganeshan et al. [16] observed that intensity and entropy of a liver CT volume of CLM patients significantly changed after contrast injection, indicating that radiomics can capture textural changes from enhancement. Miles et al. [17] investigated radiomics in relation to CLM survival by computing intensity and uniformity features from a CT liver volume and observing that textural uniformity was significantly associated with increased survival. Creasy et al. [18] and Simpson et al. [19] also observed that increased homogeneity in liver CT volumes were associated with increased risk of hepatic recurrence.

Artificial intelligence (AI) methods have shown potential in survival prediction in previous studies [20, 21]. AI, specifically machine learning, initializes models with parameters that can be optimized as more training data is available. This allows for the initialization of complex model architecture which may be more suitable to model interdependent variables than a linear models used for previous scoring systems.

We set out to evaluate, as a proof of concept, whether an automated prediction system can predict progression-free survival for liver metastasis patients treated with RT. To do this, we programmed a machine learning pipeline by implementing existing radiomic libraries to extract features from liver volumes that were manually segmented by radiation oncologists. We then implemented known machine learning models to be trained with the radiomic data and validated it with known patient recurrence outcomes.

2. Materials and Methods

2.1 Data Collection and Equipment:

This retrospective analysis was approved by the institutional review board with a waiver of informed consent at Memorial Sloan Kettering Cancer Center (MSK) (New York, NY). The MSK database was queried to obtain pre-treatment CT scans data for patients receiving radiation treatment for CLM between February 2006 to February 2019. Liver and gross tumor volumes (GTV) were segmented by radiation oncologists at MSK, as part of standard of care. This created volume subsets of liver volumes only, GTV only, and liver and GTV volume for radiomic analysis. The MSK database was also queried to obtain dosimetric treatment parameters and right-censored time-to-event data for the outcome of progression in the treated tumor (local progression).

2.2 Image Analysis

The task for the AI model was to predict the primary endpoint, defined as time from CLM radiation therapy until local tumor progression. To accomplish this, an AI survival prediction model, visualized in

Figure 1, was developed, consisting of an offline training component and a real-time prediction component. The input to the training component is a set of liver CT scans. Radiomic features are extracted from the liver and/or tumor volumes and used to train a survival model, which learns to predict a survival time interval for the patients in the training dataset. After the model is trained, new patient CT scans can be used as an input to the finalized model to compute a real-time survival prediction. The training stage contains three main components: radiomic feature extraction, feature selection, and random survival forest modelling. The AI model was programmed in Python, utilizing the PyRadiomics [22] and PySurvival libraries [23]. Concordance indices were programmed in R with the Hmisc library [24].

2.3 Radiomic Feature Extraction

In the first stage, 108 radiomic features were computed for a liver volume extracted from a CT scan. This includes computations related to shape, intensity statistics, gray-level co-occurrence matrices, gray level run-length matrices, gray level dependence matrices, and gray tone difference matrices. A full list of radiomic features is available in Supplementary Table S1. The majority of radiomic features follow the Image Biomarker Standardisation Initiative (IBSI) guidelines. Deviations from IBSI are listed in Supplementary Table S2. A set of radiomic features was computed for each lesion. Lesions were grouped together so that when they are shuffled into validation sets that no patient will have lesions both in the training and validation subset. An example of predictive radiomic features and associated outcomes is displayed in Figure 2.

2.4 Feature Selection

Retaining all 108 radiomic features would likely result in overfitting due the dimensionality of the feature space being too large for the sample size [25]. Redundant features were removed using a variance inflation factor threshold of 10 as an indicator of collinearity [26]. We then ranked remaining features using the hazard ratios predicted for each variable in a Cox proportional hazards model (CPH) [27], removing features until the ratio of features to samples was less than 1:10.

2.5 Random Survival Forest Model

To predict survival from the filtered feature set, the random survival forest (RSF) algorithm was used [28]. The algorithm creates ensemble decision tree with nodes representing features with a threshold value. The features used and the threshold values are iteratively optimized to maximize the log-rank statistic between two child nodes. The full algorithm is listed in Supplementary Equation S3. A template RSF was instantiated using the PySurvival library [23] and then hyperparameters of number of trees, maximum number of patients for a terminal node, and maximum depth were optimized with a gridsearch

algorithm. After optimization, feature importances were computed by error rates between the perturbed and unperturbed model for that feature.

2.6 Validation and Statistical Analysis

A 4 k-fold cross-validation scheme was used to provide multiple estimates of the performance of the model. The data was partitioned into 4 subsets of equal size and proportion of recurrences. The survival model was built by performing feature selection, training the RSF model, and hyperparameter optimization on 3 of the subsets and then evaluated with the remaining subset. This was repeated 4 times with a different testing subset. The concordance index (C-index), computed by Somers' Dxy rank correlation [29], and integrated Brier score (IBS) were averaged over 4 k-folds with confidence intervals computed by using the standard error of the distribution of C-indices. One limitation of this method is that the sample size may not allow larger k-fold splits for a more accurate measurement of the confidence interval [30]. All analysis was programmed with Python.

Ablation analysis was performed to investigate the performance of the model when adjustments to individual components were made. First, we defined 11 different feature sets:

1. Non-imaging and non-treatment clinical data: baseline patient variables not related to treatment information or tumor geometry from CT imaging.
2. Treatment clinical data: variables related to treatment parameters, including dosimetric variables.
3. Imaging clinical data: variables related to tumor geometry measured in CT imaging.
4. All pre-treatment clinical data: All clinical data except treatment clinical data. This represents variables that are not based on physician judgment for treatment planning.
5. All clinical data: The union feature sets 1-3.
6. Radiomics: tumor volume: radiomic features computed from the tumor volume only.
7. Radiomics: liver parenchyma: radiomic features computed from the liver parenchyma only.
8. Radiomics: liver parenchyma + tumor: radiomic features computed from the union of the tumor volume and liver parenchyma.
9. Treatment clinical data and radiomics from liver parenchyma + tumor: the union of feature sets 2. And 8.
10. Non treatment clinical data and radiomics from liver parenchyma + tumor: the union of feature sets 4. And 8.
11. All clinical data and radiomics from liver parenchyma + tumor: the union of feature sets 5. And 8.

Table 1 displays a list of categorized clinical variables.

Category	Variables
Imaging Clinical Data	Number of lesions at radiotherapy Other sites at radiotherapy Lesion dimension 1 Lesion dimension 2 PTV (cm ³)
Treatment Clinical Data	Biologically effective dose (Gy) Minimum dose for planning target volume (cGy) Maximum dose (cGy) Dose for 95% of target volume (% of intended prescribed dose) Systemic treatment before radiotherapy Lines of chemotherapy Hepatic arterial infusion pump before radiotherapy Reirradiation Surgery before radiotherapy Ablation before radiotherapy Yttrium-90 embolization before radiotherapy Arterial embolization before radiotherapy
Other Clinical Data	Primary tumor subsite Metastasis at diagnosis Number of liver lesions at diagnosis Other sites at diagnosis Liver location Carcinoembryonic antigen Kirsten rat sarcoma virus mutation

Table 1: The categorization of clinical variables to imaging, treatment, and other (non-imaging and non-treatment) clinical variables. The goal of this categorization was to observe if different subsets of clinical data performed better at prediction progression in the absence of other subsets.

Each feature set was used to build a RSF survival model with feature selection, without feature selection, and with a CPH model with gridsearch optimization of the regularization parameter. The goal was to evaluate the performance of radiomics compared to clinical data, whether the combination of both enhance performance, whether different radiomic volumes are more predictive, whether the lack of feature selection will result in overfitting, and whether using a CPH model is sufficient.

3. Results

The query resulted in obtaining CT imaging with intravenous contrast and chart data for N=97 patients, with 129 lesions identified. Of the 129 lesions, 55 resulted in local progression, 67 in no progression, and 7 undetermined. The baseline distribution of clinical variables is summarized in Supplementary Table S4 (grouped by lesion) and Supplementary Table S5 (grouped by patient). The mean freedom from local progression was 10.5 months. Dosage and fractions are summarized in Supplementary Table S6. Liver metastases were treated with total dose range of 24Gy-80Gy (mean: 52.6Gy) and number of fractions from 3-50 (mean: 8.6). The averaged cross-validation accuracies are summarized in Table 2. Samples of the predicted survival and IBS curves compared to ground truth are visualized in Figure 3. Feature importance computation results are summarized in Tables 4-6. The final models were uploaded to a public repository linked in Supplementary Table S7.

Input Features	Concordance Index (95% CI)	Integrated Brier Score (95% CI)
(No Feature Selection, Local Progression as Outcome)		
Other Clinical Data	0.64 [0.54, 0.75]	0.18 [0.15, 0.22]
Imaging Clinical Data	0.66 [0.61, 0.71]	0.17 [0.14, 0.20]
Treatment Clinical Data	0.69 [0.62, 0.77]	0.17 [0.14, 0.20]
All Pre-treatment Clinical Data	0.63 [0.55, 0.71]	0.22 [0.19, 0.25]
All Clinical Data	0.67 [0.58, 0.75]	0.16 [0.15, 0.18]
Radiomics: Tumor Volume	0.64 [0.52, 0.76]	0.18 [0.17, 0.18]
Radiomics: Liver Parenchyma	0.61 [0.53, 0.69]	0.21 [0.19, 0.23]
Radiomics: Liver Parenchyma + Tumor	0.66 [0.58, 0.74]	0.20 [0.17, 0.22]
Treatment Clinical Data + Radiomics from Liver Parenchyma and Tumor	0.66 [0.59, 0.73]	0.19 [0.18, 0.21]
All Pre-treatment Clinical Data + Radiomics from Liver Parenchyma and Tumor	0.66 [0.55, 0.77]	0.21 [0.17, 0.25]

All Clinical Data and Radiomics from Liver Parenchyma + Tumor	0.64 [0.60, 0.68]	0.19 [0.16, 0.22]
(With Feature Selection, Local Progression as Outcome)		
Other Clinical Data	0.66 [0.56, 0.76]	0.19 [0.16, 0.22]
Imaging Clinical Data	0.61 [0.56, 0.66]	0.17 [0.14, 0.19]
Treatment Clinical Data	0.72 [0.64, 0.79]	0.18 [0.15, 0.21]
All Pre-treatment Clinical Data	0.65 [0.58, 0.72]	0.21 [0.18, 0.24]
All Clinical Data	0.62 [0.56, 0.69]	0.19 [0.16, 0.22]
Radiomics: Tumor Volume	0.58 [0.51, 0.84]	0.19 [0.16, 0.24]
Radiomics: Liver Parenchyma	0.66 [0.60, 0.72]	0.20 [0.18, 0.22]
Radiomics: Liver Parenchyma + Tumor	0.68 [0.62, 0.74]	0.20 [0.16, 0.25]
Treatment Clinical Data + Radiomics from Liver Parenchyma and Tumor	0.73 [0.64, 0.82]	0.18 [0.15, 0.20]
All Pre-treatment Clinical Data + Radiomics from Liver Parenchyma and Tumor	0.66 [0.57, 0.75]	0.20 [0.17, 0.23]
All Clinical Data and Radiomics from Liver Parenchyma + Tumor	0.69 [0.65, 0.74]	0.23 [0.21, 0.26]
With Cox Proportional Hazards Model		
Other Clinical Data	0.53 [0.50, 0.56]	0.20 [0.18, 0.22]
Imaging Clinical Data	0.56 [0.45, 0.67]	0.25 [0.22, 0.28]
Treatment Clinical Data	0.50 [0.48, 0.52]	0.24 [0.20, 0.28]
All Pre-treatment Clinical Data	0.54 [0.48, 0.60]	0.19 [0.15, 0.23]
All Clinical Data	0.57 [0.48, 0.66]	0.21 [0.16, 0.26]
Radiomics: Tumor Volume	0.47 [0.42, 0.52]	0.22 [0.17, 0.27]
Radiomics: Liver Parenchyma	0.49 [0.42, 0.56]	0.24 [0.22, 0.26]
Radiomics: Liver Parenchyma + Tumor	0.43 [0.40, 0.46]	0.25 [0.21, 0.29]
Treatment Clinical Data + Radiomics from Liver Parenchyma and Tumor	0.53 [0.45, 0.61]	0.19 [0.15, 0.23]
All Pre-treatment Clinical Data + Radiomics from Liver Parenchyma and Tumor	0.55 [0.49, 0.61]	0.20 [0.17, 0.23]
All Clinical Data and Radiomics from Liver Parenchyma + Tumor	0.58 [0.47, 0.67]	0.22 [0.19, 0.25]

Table 2: A summary of accuracy results for each input combination to the model. The artificial intelligence model achieved good, nonrandom C-indices and feature selection decreased the variance of the cross-validation accuracies.

Nearly all input dataset variations using the AI model resulted in a C-index greater than 0.50 within 95% confidence interval ranges. The highest average prediction accuracy occurred when combining both radiomics of the liver parenchyma and tumor volume with treatment data (C-index: 0.73 [0.64, 0.82]). However, this was not statistically significantly different from models utilizing only clinical data. Utilizing only radiomic data from the liver parenchyma and tumor volume resulted in a C-index of 0.68 [0.62, 0.74]. The IBS of all RSF models were below 0.25, indicating nonrandom prediction [31].

Feature (Treatment Data Only)	Importance Score (95% CI)
Maximum Dose	10.84 [6.35, 15.34]
Carcinoembryonic Antigen at Radiotherapy	2.69 [-0.43, 5.81]
Lines of Chemotherapy	2.53 [1.16, 3.9]
Pump Before Radiotherapy	-0.81 [-1.57, -0.05]
Feature (Radiomics on Liver Plus Tumor Volume Only)	
Neighborhood Gray Tone Difference Matrix Strength	3.74 [2.25, 5.22]
Neighborhood Gray Tone Difference Matrix Busyness	3.32 [2.5, 4.15]
Kurtosis	1.97 [1.58, 2.37]
Maximum 2D Diameter Slice	1.45 [0.20, 2.69]
Gray Level Size Zone Matrix Low Gray Level Emphasis	0.33 [-0.75, 1.42]
Neighborhood Gray Tone Difference Matrix Contrast	0.02 [-0.78, 0.82]
Skewness	-0.25 [-0.81, 0.31]
Gray Level Co-occurrence Matrix Cluster Shade	-0.88 [-2.78, 1.01]
Feature (Treatment Data and Radiomics on Liver Plus Tumor Volume)	
Maximum Dose	3.83 [1.05, 6.62]
Neighborhood Gray Tone Difference Matrix Strength	1.90 [0.93, 2.86]
Lines of Chemotherapy	1.36 [0.38, 2.35]
Gray Level Size Zone Matrix Low Gray Level Emphasis	1.01 [-0.37, 2.39]
KRAS Mutation	0.65 [0.10, 1.19]
Carcinoembryonic Antigen at Radiotherapy	0.48 [-1.11, 2.08]
Gray Level Size Zone Matrix Nonuniformity	0.48 [-0.32, 1.27]

Gray Level Co-occurrence Matrix Cluster Shade	0.17 [-0.98, 1.32]
Pump Before Radiotherapy	-0.08 [-1.21, 1.04]
Skewness	-0.29 [-0.73, 0.15]

Table 4: The feature importances for the random survival forest model utilizing treatment data only, radiomics data only, or the combination of both. Maximum dose was observed to be the most predictive feature, significantly with more information gain than any other treatment feature. Gray tone difference matrix computations were the most predictive when only using radiomics data. Both gray tone difference matrices and maximum dose features resulted in high predictive value in the combined model. However, the importance of maximum dose was decreased compared to when using only treatment data, indicating that the model is still able to predict survival with the remaining radiomic features.

The most predictive radiomic feature was the neighboring gray tone difference matrix (NGTDM) strength. The most predictive clinical variable was maximum dose, significantly greater than any other clinical variable.

4. Discussion

The goal of the study was to develop a method utilizing radiomics and machine learning to predict time until local progression of CLM patients. The IBS of every dataset combination was below the threshold of 0.25, indicating that the predictions by the RSF model is non-random and information-gaining [31]. This suggests that there is predictive texture within the liver parenchyma and tumor volume that may be investigated further to understand structural changes that affect prognosis. This is consistent with Simpson et al. [19], who observed that radiomic features were associated with recurrence and are potentially reflective of tissue abnormalities that create a metastatic environment.

There are several opportunities we aimed to address to improve on existing methods. First, the CPH model in theory is parameterized with lower complexity than RSF and may be unable to capture nonlinear dependencies [32]. However, from our results, this is indeterminate as the although the CPH did not significantly perform better than random chance, there was a wide confidence interval to overlap with the RSF model. When assessing performance with IBS, the CPH model was not greater than the 0.25 threshold for only the combined radiomics and clinical subsets. Recent studies modelling survival with radiomics show no significant difference between CPH and RSF models [33, 34]. Comparison of our model may require a larger sample size and to evaluate the feature selection and optimization methods other studies have used. Secondly, existing studies performing linear mapping of hazard ratios to

prediction scores and may oversimplify nonlinear dependencies between variables, particularly when relying on rounding to integer scores. Thirdly, there may be predictive information missed if only analyzing clinicopathological variables. As tumor progression results in changes in tissue, there may be observable structural changes in the liver associated with survival.

Most prior studies reported a C-index lower than 0.60 when tested on external datasets, with one model by Wang et al. achieving a C-index of 0.64 [14]. However, we did not have access to all variables used, which is required for a fairer comparison between manual scoring systems and automated RSF methods in future studies.

The radiomic model from the union of the liver parenchyma and tumor with feature selection enabled achieved a C-index (95% CI) of 0.68 (0.62 – 0.74). Utilizing tumor or liver parenchyma volumes only performed within the same confidence interval range. This suggests that both the liver parenchyma and tumor contain textural features predictive of local control. As the radiomic features are computed as a single point-data aggregate characteristic value for the volume, it is difficult to localize the exact regions of abnormal texture. Future studies that isolate patches of the liver can be conducted to localize regions with abnormal radiomic values.

Without feature selection, there was a larger variance across the cross-validation folds. This is likely due to overfitting as the number of input variables defines the dimensionality in the optimization problem for the machine learning model. The optimized solution may be too specific to the training data, resulting in lower testing accuracy.

Utilizing only clinical data did not result in a statistically significant decrease in accuracy than with radiomics alone. In the combined mode, the two most predictive features were similarly maximum dose with a feature importance score (95% CI) of 3.83 (1.05 – 6.62) and NGTDM strength with a feature importance score (95% CI) of 1.90 (0.93 – 2.86). Moreover, the feature importance (95% CI) of maximum dose decreased from 10.84 (6.35, 15.34) in the treatment data only model to 3.83 (1.05, 6.62) in the combined model, indicating that the radiomic features contribute to prediction even when treatment data is available. There are variables similar to maximum dose, such as dose covering 95% of the planning target volume, that were removed by the feature selection algorithm due to collinearity. It should be noted that dosage is increased for tumors that may have shown radioresistance, hence some expert prior knowledge is required for this variable whereas the radiomic features are dependent only on the image.

Despite the observed accuracy of the model, further validation with a diverse patient population from different centers for instance is required to evaluate generalizability. With more samples, 5 k-fold or 10 k-fold cross validation may be conducted to have more testing sets so that accuracy evaluation has more statistical power with less aggressive feature selection [35]. As the samples are limited to patients treated with primary or adjuvant RT, future studies may include patients before and after radiotherapy, as texture in CT scans may change after treatment. Another exclusion is of patients who are deceased, and we were unable to evaluate the effect of death on the recurrence prediction model, which may require reparameterization with competing risks. With sufficient data, deep learning with convolutional neural networks is another potential method to predict survival and has been used previously to predict response to chemotherapy for CLM [36]. However, deep learning models are more difficult to interpret than radiomic features due to the multiple layers of matrix convolutions.

It has been a reported challenge of radiomics that there is no standardized cutoff or clinical interpretation of features [15]. For instance, high skewness indicates that the intensities are not symmetrically distributed. This may indicate inhomogeneous interactions of electromagnetic radiation with the liver tissue, but the cause of the inhomogeneity is not well understood. Hence, the features determined to be predictive require future studies to understand how they are related to pathophysiology of metastases.

In this work, we have developed a time-dependent tumor progression prediction model for CRM treated with primary or adjuvant RT utilizing radiomic features from CT scans and an AI random survival forest. The model was able to achieve good C-indices utilizing radiomic features from the liver parenchyma and tumor volume or with treatment data. As a proof of concept, this study provides support that radiomic AI methods may be further developed to aid in prognostic decision making in radiation oncology. Radiomic features determined to be predictive may be investigated in the future to understand structural changes reflected in radiomic observations in the CT scan to provide new directions for clinician analysis of liver texture.

335 **Acknowledgments:**

336 We acknowledge funding supported by the National Institutes of Health/National Cancer Institute Support
337 Grant P30 CA008748, the National Institutes of Health/National Cancer Institute early career
338 development award K08 CA255574 and National Cancer Institute R01 CA233888. The content is solely
339 the responsibility of the authors and does not necessarily represent the official views of the National
340 Institutes of Health.

References:

- [1] Abdalla EK, Vauthey JN, Ellis LM, Ellis V, Pollock R, Broglio K, et al. Recurrence and outcomes following hepatic resection, radiofrequency ablation, and combined resection/ablation for colorectal liver metastases. *Ann Surg.* 2004 ;239(6):818-27. <https://doi.org/10.1097/01.sla.0000128305.90650.71>.
- [2] Leung U, Gönen M, Allen PJ, Kingham TP, DeMatteo RP, Jarnagin WR, D'Angelica MI. Colorectal Cancer Liver Metastases and Concurrent Extrahepatic Disease Treated With Resection. *Ann Surg.* 2017;265:158–65. <https://doi.org/10.1097/SLA.0000000000001624>.
- [3] Ruers T, Van Coevorden F, Punt CJ, Pierie JE, Borel-Rinkes I, Ledermann JA, et al. Local Treatment of Unresectable Colorectal Liver Metastases: Results of a Randomized Phase II Trial. *J Natl Cancer Inst.* 2017;109(9):dix015. <https://doi.org/10.1093/jnci/dix015>.
- [4] Mahadevan A, Blanck O, Lanciano R, Peddada A, Sundararaman S, D'Ambrosio D, Sharma S, et al. Stereotactic Body Radiotherapy (SBRT) for liver metastasis - clinical outcomes from the international multi-institutional RSSearch® Patient Registry. *Radiat Oncol.* 2018;13(1):26. <https://doi.org/10.1186/s13014-018-0969-2>.
- [5] Nordlinger B, Guiguet M, Vaillant JC, Balladur P, Boudjema K, Bachellier P, Jaeck D. Surgical resection of colorectal carcinoma metastases to the liver. A prognostic scoring system to improve case selection, based on 1568 patients. *Association Française de Chirurgie. Cancer.* 1996;77(7):1254-62.
- [6] 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):309-21. <https://doi.org/10.1097/00000658-199909000-00004>.
- [7] Iwatsuki S, Dvorchik I, Madariaga JR, Marsh JW, Dodson F, Bonham AC, et al. Hepatic resection for metastatic colorectal adenocarcinoma: a proposal of a prognostic scoring system. *J Am Coll Surg.* 1999;189(3):291-9. [https://doi.org/10.1016/s1072-7515\(99\)00089-7](https://doi.org/10.1016/s1072-7515(99)00089-7).
- [8] Konopke R, Kersting S, Distler M, Dietrich J, Gastmeier J, Heller A, et al. Prognostic factors and evaluation of a clinical score for predicting survival after resection of colorectal liver metastases. *Liver Int.* 2009;29(1):89-102. <https://doi.org/10.1111/j.1478-3231.2008.01845.x>.

- [9] Nagashima I, Takada T, Matsuda K, Adachi M, Nagawa H, Muto T, et al. A new scoring system to classify patients with colorectal liver metastases: proposal of criteria to select candidates for hepatic resection. *J Hepatobiliary Pancreat Surg.* 2004;11(2):79-83. <https://doi.org/10.1007/s00534-002-0778-7>.
- [10] Imai K, Allard MA, Castro Benitez C, Vibert E, Sa Cunha A, Cherqui D, et al. Nomogram for prediction of prognosis in patients with initially unresectable colorectal liver metastases. *Br J Surg.* 2016;103(5):590-9. <https://doi.org/10.1002/bjs.10073>.
- [11] Sasaki K, Morioka D, Conci S, Margonis GA, Sawada Y, Ruzzenente A, et al. The Tumor Burden Score: A New "Metro-ticket" Prognostic Tool For Colorectal Liver Metastases Based on Tumor Size and Number of Tumors. *Ann Surg.* 2018;267(1):132-41. <https://doi.org/10.1097/SLA.0000000000002064>.
- [12] Rees M, Tekkis PP, Welsh FK, O'Rourke T, John TG. Evaluation of long-term survival after hepatic resection for metastatic colorectal cancer: a multifactorial model of 929 patients. *Ann Surg.* 2008;247(1):125-35. <https://doi.org/10.1097/SLA.0b013e31815aa2c2>.
- [13] Brudvik KW, Jones RP, Giuliante F, Shindoh J, Passot G, Chung MH, et al. RAS Mutation Clinical Risk Score to Predict Survival After Resection of Colorectal Liver Metastases. *Ann Surg.* 2019;269(1):120-6. <https://doi.org/10.1097/SLA.0000000000002319>.
- [14] Wang K, Liu W, Yan XL, Li J, Xing BC. Long-term postoperative survival prediction in patients with colorectal liver metastasis. *Oncotarget.* 2017;8(45):79927-34. <https://doi.org/10.18632/oncotarget.20322>.
- [15] Fiz F, Viganò L, Gennaro N, Costa G, La Bella L, Boichuk A, et al. Radiomics of Liver Metastases: A Systematic Review. *Cancers.* 2020;12(10):2881. <https://doi.org/10.3390/cancers12102881>.
- [16] Ganeshan B, Burnand K, Young R, Chatwin C, Miles K. Dynamic contrast-enhanced texture analysis of the liver: initial assessment in colorectal cancer. *Invest Radiol.* 2011;46(3):160-8. <https://doi.org/10.1097/RLI.0b013e3181f8e8a2>.

- [17] Miles KA, Ganeshan B, Griffiths MR, Young RC, 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>.
- [18] Creasy JM, Cunanan KM, Chakraborty J, McAuliffe JC, Chou J, Gonen M, et al. Differences in Liver Parenchyma are Measurable with CT Radiomics at Initial Colon Resection in Patients that Develop Hepatic Metastases from Stage II/III Colon Cancer. *Ann Surg Oncol*. 2021;28(4):1982-9. <https://doi.org/10.1245/s10434-020-09134-w>.
- [19] Simpson AL, Doussot A, Creasy JM, Adams LB, Allen PJ, DeMatteo RP, 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>.
- [20] Kim DW, Lee S, Kwon S, Nam W, Cha IH, Kim HJ. Deep learning-based survival prediction of oral cancer patients. *Sci Rep*. 2019;9(1):6994. <https://doi.org/10.1038/s41598-019-43372-7>.
- [21] Wang W, Liu W. Integration of gene interaction information into a reweighted random survival forest approach for accurate survival prediction and survival biomarker discovery. *Sci Rep*. 2018;8(1):13202. <https://doi.org/10.1038/s41598-018-31497-0>.
- [22] 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. <https://doi.org/10.1158/0008-5472.CAN-17-0339>
- [23] Fotso, S. PySurvival: Open source package for survival analysis modeling, <https://square.github.io/pysurvival/>; 2019 [accessed 15 April 2022].
- [24] Harrell FE, Dupont C. Hmisc: Harrell miscellaneous R package version 4.6-0, <https://cran.r-project.org/web/packages/Hmisc/index.html>; 2021 [accessed 15 April 2022].
- [25] Liu R, Gillies DF. Overfitting in linear feature extraction for classification of high-dimensional image data. *Pattern Recognit*. 2016;53(C):73–86. <https://doi.org/10.1016/j.patcog.2015.11.015>.

- [26] Salmeron R, Garcia CB, Garcia J. Variance Inflation Factor and Condition Number in multiple linear regression. *J Stat Comput Simul.* 2018;88(12):2365-84. <https://doi.org/10.1080/00949655.2018.1463376>.
- [26] Bourgon R, Gentleman R, Huber W. Independent filtering increases detection power for high-throughput experiments. *Proceedings of the National Academy of Sciences of the United States of America.* 2010;107(21), 9546–51. <https://doi.org/10.1073/pnas.0914005107>.
- [28] Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. *Ann Appl Stat.* 2008;2:841–60. <https://doi.org/10.1214/08-AOAS169>.
- [29] Newson, R. Confidence Intervals for Rank Statistics: Somers' D and Extensions. *Stata J.* 2006;6(3):309-34. <https://doi.org/10.1177/1536867X0600600302>.
- [30] Rodriguez JD, Perez A, Lozano JA. Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE PAMI.* 2009;32(3):569-75. <https://doi.org/10.1109/TPAMI.2009.187>.
- [31] Steyerberg EW, Vickers AJ, Cook NR, Gerds T, Gonen M, Obuchowski N, et al. Assessing the performance of prediction models: a framework for traditional and novel measures. *Epidemiology.* 2010;21(1):128-38. <https://doi.org/10.1097/EDE.0b013e3181c30fb2>.
- [32] Lin DY, Wei LJ. The robust inference for the cox proportional hazards model. *J Am Stat Assoc.* 1989;84:1074–8. <https://doi.org/10.2307/2290085>.
- [33] Leger S, Zwanenburg A, Pilz K, Lohaus F, Linge A, Zöphel K, et al. A comparative study of machine learning methods for time-to-event survival data for radiomics risk modelling. *Sci Rep.* 2017;7(1):13206. <https://doi.org/10.1038/s41598-017-13448-3>.
- [34] Chang E, Joel MZ, Chang HY, Du J, Khanna O, Omuro A, et al. Comparison of radiomic feature aggregation methods for patients with multiple tumors. *Sci Rep.* 2021;11(1):9758. <https://doi.org/10.1038/s41598-021-89114-6>.
- [35] Menze BH, Kelm BM, Masuch R, Himmelreich U, Bachert P, Petrich W, et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data. *BMC Bioinform.* 2009;10:213. <https://doi.org/10.1186/1471-2105-10-213>.

475

476 [36] Wei J, Cheng J, Gu D, Chai F, Hong N, Wang Y, et al. Deep learning-based radiomics predicts
477 response to chemotherapy in colorectal liver metastases. *Med Phys.* 2021;48(1):513-22.
478 <https://doi.org/10.1002/mp.14563>.

479

480

481 **Supplementary Material:**

482

483 **Supplementary Table S1: List of Radiomic Features**

Feature Category	Features
First Order Statistics	Energy Total Energy Entropy Minimum 10th Percentile 90th Percentile Maximum Mean Median Interquartile Range Range Mean Absolute Deviation Robust Mean Absolute Deviation Root Mean Squared Standard Deviation Skewness Kurtosis Variance Uniformity
3D Shape	Mesh Volume Voxel Volume Surface Area Surface Area to Volume Ratio Sphericity Compactness Spherical Disproportion Maximum 3D Diameter Maximum 2D Diameter (Axial) Maximum 2D Diameter (Coronal) Maximum 2D Diameter (Sagittal)

	Major Axis Length Minor Axis Length Least Axis Length Elongation Flatness
Gray level Co-occurrence Matrix	Autocorrelation Joint Average Cluster Prominence Cluster Shade Cluster Tendency Contrast Correlation Difference Average Difference Entropy Difference Variance Difference Average Joint Energy Joint Entropy Informational Correlation Inverse Difference Moment Inverse Difference Moment Normalized Inverse Difference Inverse Difference Normalized Inverse Variance Maximum Probability Sum Average Sum Entropy Sum of Squares
Gray Level Size Zone Matrix	Small Area Emphasis Large Area Emphasis Gray Level Non-Uniformity Gray Level Non-Uniformity Normalized Size-Zone Non-Uniformity Size-Zone Non-Uniformity Normalized

	Zone Percentage Gray Level Variance Zone Variance Zone Entropy Low Gray Level Zone Emphasis High Gray Level Zone Emphasis Small Area Low Gray Level Emphasis Small Area High Gray Level Emphasis Large Area Low Gray Level Emphasis Large Area High Gray Level Emphasis
Gray Level Run Length Matrix	Short Run Emphasis Long Run Emphasis Gray Level Non-Uniformity Gray Level Non-Uniformity Normalized Run Length Non-Uniformity Run Length Non-Uniformity Normalized Run Percentage Gray Level Variance Run Variance Run Entropy Low Gray Level Run Emphasis High Gray Level Run Emphasis Short Run Low Gray Level Emphasis Short Run High Gray Level Emphasis Long Run Low Gray Level Emphasis Long Run High Gray Level Emphasis
Gray Level Dependence Matrix	Small Dependence Emphasis Large Dependence Emphasis Gray Level Non-Uniformity Dependence Non-Uniformity Dependence Non-Uniformity Normalized Gray Level Variance Dependence Variance Dependence Entropy

	Low Gray Level Emphasis High Gray Level Emphasis Small Dependence Low Gray Level Emphasis Small Dependence High Gray Level Emphasis Large Dependence Low Gray Level Emphasis Large Dependence High Gray Level Emphasis
Neighboring Gray Tone Difference Matrix	Coarseness Contrast Busyness Complexity Strength

Supplementary Table S1: A list of radiomic features extracted from a liver volume. The features include computations related to the statistics, shape, and gray-level relationships of the image.

488 **Supplementary Table S2: Radiomic Deviations from IBSI Standards**

Computation	PyRadiomics Implementation	IBSI Guidelines
Binning	Discretizes gray values with fixed bins with edges equally spaced from 0.	Discretizes using fixed bin width equally spaced from minimum of resegmentation range
Resampling	Aligns to the corner of the original voxel	Aligns to the center of the image
Gray value rounding	Does not resample to similar resolution if original intensity is lower precision	Resamples to similar resolution if original intensity is lower precision
Mask resampling	Resamples to nearest neighbor	Allows selection of different interpolators for resampling

489
490 Supplementary Table S2: A list of deviations from the feature extraction guidelines by the Image
491 Biomarker Standardisation Initiative (IBSI).
492

Supplementary Equation S3: Random Survival Forest Algorithm

To build a survival tree that predicts survival from an input vector of radiomic features, the following steps are taken:

1. Select N samples from the dataset.
2. For each sample $i = 1, 2, \dots, N$, initialize a binary decision tree with max depth D .
3. At each node, iterate through set of features $X = \{x_1, x_2, \dots, x_N\}$ and its range of feature values $S = \{S_{min}, S_{max}\}$ to select feature x_i and a threshold split value s_i such that:

$$L(x_i, s_i) \geq L(x, s) \forall x \in X, s \in S$$

Where $L(x, s)$ is the log rank test such that

$$L(x, c) = \frac{\sum_{i=1}^N E_{i,1} - Y_{i,1} \frac{E_i}{Y_i}}{\sqrt{\sum_{i=1}^N \left(\frac{Y_{i,1}}{Y_i} \right) \left(1 - \frac{Y_{i,1}}{Y_i} \right) \left(\frac{Y_i - E_i}{Y_i - 1} \right) E_i}}$$

Where at time t_i , E_i is the number of events at time t_i , $E_{i,j}$ is the number of events at a daughter node j , Y_i is the number of patients with an events or at risk at time t_i , and $Y_{i,j}$ is the number of patients with an event or at risk at a daughter node j

4. Continue to grow children nodes unless the children node has no more than M surviving samples, where M is a user-defined hyperparameter
5. Calculate the cumulative hazard function for the decision tree with the Nelson-Aalen estimator:

$$H_q(t) = \sum_{t_{p,q} \leq t} \frac{E_{p,q}}{Y_{p,q}}$$

Where p is a patient in the set of M patients in set $P = \{p_1, p_2, \dots, p_M\}$, q is a node in the set of N nodes in set $Q = \{q_1, q_2, \dots, q_N\}$, $E_{p,q}$ is the number of events at time $t_{p,q}$, and $Y_{p,q}$ is the number of patients with an event or at risk at time $t_{p,q}$.

6. Repeat steps 1-5 K times to create K separately initialized trees, where K is a user-defined hyperparameter.
7. Average the cumulative hazard function over all trees to compute the ensembled cumulative hazard.

520 **Supplementary Table S4: Baseline variable distributions by lesion**

Characteristics	All lesions (n=129)
Number of patients/lesions	97/129
Sex, n (%)	
Male	83 (64.3)
Female	46 (36.7)
Metastasis at time of diagnosis, n (%)	
M0	40 (30.0)
M1	89 (70.0)
Colorectal Histology, n (%)	
Colon	104 (80.6)
Rectal	20 (15.6)
Undetermined	5 (3.8)
Number of liver lesions at diagnosis, n (%)	
0	5 (3.9)
1	25 (19.4)
2	9 (7.0)
3-5	32 (24.8)
> 5	53 (41.0)
Undetermined	5 (3.9)
Other sites at diagnosis, n (%)	
None	101 (78.3)
Lung	12 (9.3)
Non-regional LN	3 (2.3)
Lung and non-regional LN	4 (3.1)
Other	7 (5.4)
Undetermined	2 (1.6)
RT to other sites, n (%)	
No	75 (58.1)
Before liver RT	28 (21.7)
After Liver RT	21 (16.3)
Before and after liver RT	2 (1.6)
Undetermined	3 (2.3)
RT fraction delivered, Median (IQR)	6 (5-15)
RT dose delivered, Median (IQR)	4500 (3000 - 6000)
Dose Painting - Yes, n (%)	55 (42.6)
Intended Dose Median (IQR)	6000 (4000 - 6750)
Mean RT length \pm SD (Days)	11.6 (8.5)
PTV volume (cm3), Median (IQR)	94.4 (39.2 - 174.4)

Mean D95 \pm SD (% of intended dose)	97.7 (11.0)
Reirradiation - Yes, n (%)	8 (6.2)
Surgery before RT, n (%)	91 (70.5)
Systemic before RT, n (%)	126 (97.7)
Pump before RT, n (%)	81 (62.8)
Lines of Chemo, Median (IQR)	3 (2 - 4)
RFA before RT, n (%)	45 (34.9)
RFA to RT lesions - Yes, n (%)	13 (10.1)
Y90 before RT - Yes, n (%)	10 (7.8)
Embolization before RT, n (%)	12 (9.3)
CEA at diagnosis, Median (IQR)	15.7 (3.38 - 176.9)
CEA at RT, Median (IQR)	18.7 (4.8 - 127.2)
Number of liver lesions at RT, n (%)	
1	57 (44.2)
2	43 (33.3)
3	12 (9.4)
≥ 4	16 (12.4)
Undetermined	1 (0.7)
Other sites at RT, n (%)	
None	52 (40.3)
Lung	27 (21.0)
Non-regional LN	10 (7.8)
Lung and non-regional LN	25 (19.3)
Other	15 (11.6)
Mean lesion 1 dimension 2 \pm SD	35.2 (22.3)
Mean lesion 1 dimension 1 \pm SD	26.0 (17.9)
Freedom from local progression (FFLP), n (%)	
Progression	55 (42.6)
No progression	67 (52.0)
Undetermined	7 (5.4)
Mean FFLP (months) \pm SD	10.5 (0.4)
Any hepatic progression (AHP), n (%)	
Progression	99 (76.8)
No progression	25 (19.4)
Undetermined	5 (3.8)
Mean time to AHP (months) \pm SD	7.3 (7.1)

Abbreviations: LN = lymph node, RT = radiotherapy, PTV = planning target volume, CEA = carcinoembryonic antigen, HAIP = hepatic arterial infusion pump, TARE = transarterial radioembolization.

Supplementary Table S4: A table of baseline clinical variables recorded as part of standard of care, with averages computed from the set of variables per lesion. The clinical variables will be utilized alongside computational radiomic features from computed tomography scans as input data to a machine learning model to predict local progression.

532 **Supplementary Table S5: Baseline variable distributions by patient**

533

Characteristics	All Patients (n=97)
Sex, n (%)	
Male	63 (64.9)
Female	34 (35.1)
Metastasis at time of diagnosis, n (%)	
M0	31 (32)
M1	65 (67)
Other sites at diagnosis, n (%)	
None	74 (76.3)
Lung	9 (9.3)
Non-regional LN	3 (3.1)
Lung and non-regional LN	4 (4.1)
Other	5 (5.2)
Undetermined	2 (2.1)
RT to other sites, n (%)	
No	57 (58.8)
Before liver RT	23 (23.7)
After Liver RT	13 (13.4)
Before and after liver RT	2 (2.1)
Undetermined	1 (1)
Number of liver lesions at RT, n (%)	
1	56 (57.7)
2	25 (25.8)
3	6 (6.2)
≥ 4	9 (9.3)
Undetermined	1 (1)
Other sites at RT, n (%)	
None	37 (38.1)
Lung	21 (21.6)
Non-regional LN	8 (8.2)
Lung and non-regional LN	17 (17.5)
Other	14 (14.4)
Freedom from local progression (FFLP), n (%)	
Progression	50 (51.5)
No progression	40 (41.2)
Undetermined	7 (7.2)
Mean time to FFLP (months) ± SD	10.5 (8.8)

Any hepatic progression (AHP), n (%)	
Progression	76 (78.4)
No progression	16 (16.5)
Undetermined	5 (5.2)
Mean time to AHP (months) \pm SD	7.4 (6.9)

534

535 Abbreviations: LN = lymph node, RT = radiotherapy

536

537 Supplementary Table S5: A table of baseline clinical variables recorded as part of standard of care, with

538 averages computed from the set of variables per patient. Lesion-specific variables were excluded.

539

540 **Supplementary Table S6: Dosages and number of fractions to liver metastases**

541

Total Dose (Gy)	Fractions	BED ₁₀ (Gy)	Patient Count
24	1	82	9
24	3	43	1
27	3	51	1
30	3	60	1
30	5	48	8
30	10	39	1
35	3	60	2
36	6	68	1
38	15	48	1
40	5	72	4
45	3	113	5
45	5	86	1
50	5	100	12
50	10	75	1
60	3	180	2
60	5	132	4
60	6	120	2
60	10	96	5
60	15	84	1
67.5	15	98	15
70	10	119	9
75	3	263	1

75	5	188	2
75	15	113	1
75	25	98	5
75	50	86	1
80	10	144	1

542
543 Supplementary Table S6: A list of doses to liver metastases, fractions, biologically effective dose (BED),
544 and number of patients treated with the combination.
545

546 **Supplementary Material S7: Link to survival models**

547

548 The survival models from the different feature sets are uploaded to an open repository at:

549 https://github.com/ricky-hu/local_control_radiomics_survival_model

550

551