RESEARCH ARTICLE

# Radiomics analysis of baseline computed tomography to predict oncological outcomes in patients treated for resectable colorectal cancer liver metastasis

**Emmanuel Montagnon**[1], **Milena Cerny**[1,2], **Vincent Hamilton**[3], **Thomas Derennes**[3], **André Ilinca**[3], **Mohamed El Amine Elforaici**[1,4], **Gilbert Jabbour**[3,5], **Edmond Rafie**[3], **Anni Wu**[3], **Francisco Perdigon Romero**[6], **Alexandre Cadrin-Chênevert**[7], **Samuel Kadoury**[1,4], **Simon Turcotte**[1,8], **An Tang**[1,3,9]*

1 Centre de recherche du Centre hospitalier de l'Université de Montréal (CRCHUM), Montréal, QC, Canada, 2 Department of Radiology, CISSS des Laurentides, Hôpital de Saint-Eustache, Saint-Eustache, QC, Canada, 3 Department of Radiology, Radiation Oncology and Nuclear Medicine, Université de Montréal, Montréal, QC, Canada, 4 MedICAL Laboratory, Polytechnique Montréal, Montréal, QC, Canada, 5 Division of Internal Medicine, Department of Medicine, Hôpital du Sacré-Cœur-de-Montréal, Montréal, QC, Canada, 6 Montreal AI Hub, Ericsson Canada, Montréal, QC, Canada, 7 Department of Medical Imaging, CISSS Lanaudière, Université Laval, Joliette, Québec, Canada, 8 Hepatopancreatobiliary and Liver Transplantation Division, Department of Surgery, Centre Hospitalier de l'Université de Montréal (CHUM), Montréal, QC, Canada, 9 Department of Radiology, Centre Hospitalier de l'Université de Montréal (CHUM), Montréal, QC, Canada

* an.tang@umontreal.ca

## Abstract

### Objective

The purpose of this study was to determine and compare the performance of pre-treatment clinical risk score (CRS), radiomics models based on computed (CT), and their combination for predicting time to recurrence (TTR) and disease-specific survival (DSS) in patients with colorectal cancer liver metastases.

### Methods

We retrospectively analyzed a prospectively maintained registry of 241 patients treated with systemic chemotherapy and surgery for colorectal cancer liver metastases. Radiomics features were extracted from baseline, pre-treatment, contrast-enhanced CT images. Multiple aggregation strategies were investigated for cases with multiple metastases. Radiomics signatures were derived using feature selection methods. Random survival forests (RSF) and neural network survival models (DeepSurv) based on radiomics features, alone or combined with CRS, were developed to predict TTR and DSS. Leveraging survival models predictions, classification models were trained to predict TTR within 18 months and DSS within 3 years. Classification performance was assessed with area under the receiver operating characteristic curve (AUC) on the test set.

## Results

For TTR prediction, the concordance index (95% confidence interval) was 0.57 (0.57–0.57) for CRS, 0.61 (0.60–0.61) for RSF in combination with CRS, and 0.70 (0.68–0.73) for Deep-Surv in combination with CRS. For DSS prediction, the concordance index was 0.59 (0.59–0.59) for CRS, 0.57 (0.56–0.57) for RSF in combination with CRS, and 0.60 (0.58–0.61) for DeepSurv in combination with CRS. For TTR classification, the AUC was 0.33 (0.33–0.33) for CRS, 0.77 (0.75–0.78) for radiomics signature alone, and 0.58 (0.57–0.59) for DeepSurv score alone. For DSS classification, the AUC was 0.61 (0.61–0.61) for CRS, 0.57 (0.56–0.57) for radiomics signature, and 0.75 (0.74–0.76) for DeepSurv score alone.

## Conclusion

Radiomics-based survival models outperformed CRS for TTR prediction. More accurate, noninvasive, and early prediction of patient outcome may help reduce exposure to ineffective yet toxic chemotherapy or high-risk major hepatectomies.

## Introduction

Colorectal cancer is the second cause of cancer death in Canada and worldwide [1, 2]. Most patients progress by developing colorectal cancer liver metastases (CRLM). Resistance to systemic chemotherapy is the main determinant of patient survival and clinical decision making. While consensus is growing that it is unwise to continue the practice of treating large numbers of unselected patients who must endure treatment-related morbidities knowing only a fraction will benefit, there are currently few tools to reliably predict outcome after curative-indented treatments [3].

Better prediction tools would help identify subsets of patients at lowest risk of recurrence most likely to benefit from hepatectomy alone. In contrast, CRLM patients not achieving response with standard systemic chemotherapy could be offered non-surgical liver-directed therapy such as hepatic arterial infusion chemotherapy [4, 5]. Earlier assessment of CRLM response to chemotherapy would limit surgical morbidities related to hepatotoxicity, reduce unnecessary adverse events and cost in non-responding patients, and promote earlier therapeutic change with the opportunity to improve patient outcomes. In patients with operable CRLMs, the best validated pre-operative clinical risk score, which includes five clinicopathologic factors, cannot well identify patients with particularly good recurrence free-survival, and does not take into account the impact of chemotherapy [6, 7]. Fong's clinical risk score (CRS), first proposed in [8] is commonly used in oncology for patient selection of candidates for surgical resection of CRLM and prediction of tumor recurrence.

Preoperative assessment of CRLM therapeutic response relies on visual assessment of follow-up computed tomography (CT) examinations by radiologists. However, interpretation is often limited by changes in CRLM appearance. Over the last decade, radiomics and deep learning approaches have been increasingly used in radiology [9–11] tackling various computer vision tasks such as detection [12, 13], classification, segmentation [14] but also in the context of survival analysis [15, 16]. Radiomics aim to leverage information embedded in medical images [17] by quantifying features related to textures, patterns or statistical parameters. Imaging features exhibited by liver metastases have been shown to be related to treatment response [18], overall survival [19] or disease-free survival [20]. However, there is a need to

assess the ability of radiomics to predict clinical outcomes of patients with CRLM and to determine if they can improve the prediction performance when compared to the CRS, either alone or in combination.

The purpose of this study was to determine and compare the performance of pre-treatment CRS, radiomics models based on computed (CT) examinations, and their combination for predicting time to recurrence (TTR) and disease-specific survival (DSS) in patients with CRLM. More specifically, we used random survival forest (RSF) survival models and neural networks survival models (DeepSurv) to predict TTR and DSS. We then used radiomics signatures and DeepSurv scores, alone or in combination with CRS, to predict TTR and DSS as classification tasks. Performances of various tumor aggregations on outcome predictions are also presented.

## Materials and methods

### Study design and patients

All patient provided written consent to participate to the *Centre hospitalier de l'Université de Montréal* hepatopancreatobiliary cancer biobank and prospective registry, registered by the Canadian Tumour Repository Network [21], approved by the institutional ethics board (IRB) (No. 09.237, 22-Jan-2010), by which they accepted that clinical, radiological, and histo-pathological data be collected and used for research in specific projects further approved by the IRB. The current retrospective research project was approved on 31-May-2018 (IRB No. 18.023).

Patients with upfront resectable CRLM determined by treating oncologists and surgeons, treated between 2007 and 2017, who underwent a CT-scan within 3 months before and 10 days after beginning of pre-operative chemotherapy regimen, were selected.

Flowchart of patient selection is presented in **Fig 1**.

### Clinical risk score

Based on clinical preoperative data, a clinical risk score (CRS) [22] ranging from 0 to 5 was calculated for each patient adding one point if the following features were present: node positive primary cancer, disease-free interval from primary to CRLM diagnosis of < 12 months, more than one CRLM, largest CRLM > 5 cm, and pre-operative carcinoembryonic antigen level > 200 ng/mL. The CRS was dichotomized as low risk (0-1-2) vs. high risk (3-4-5), as published [8].

### Imaging data

**CT dataset.** For each patient, imaging data consisted in contrast-enhanced computed tomography (CT) pre-treatment examinations performed either before or within then days of first round of chemotherapy. CT scans were performed with various 16- and 64-multidetector scanners at our institution. All CT examinations included at least one portal venous phase covering the upper abdomen. CT scans acquired over a ten-year range (2007–2017) were retrospectively retrieved for research purposes between January 1, 2018 and June 30, 2018. Since patients were enrolled in a biobank, the authors had access to information that could identify individual participants during or after data collection. CT scans were performed on scanners from different manufacturers (Philips, Toshiba, GE, Siemens), models, variable Mas, and reconstruction kernels. The range of acquisition parameters are provided in **S1 Table**. Overall data workflow is presented in **Fig 2**.

**Tumor segmentation.** A total of 731 liver metastases have been segmented manually on CT images by 3 medical students. The segmentation masks were then reviewed and corrected,
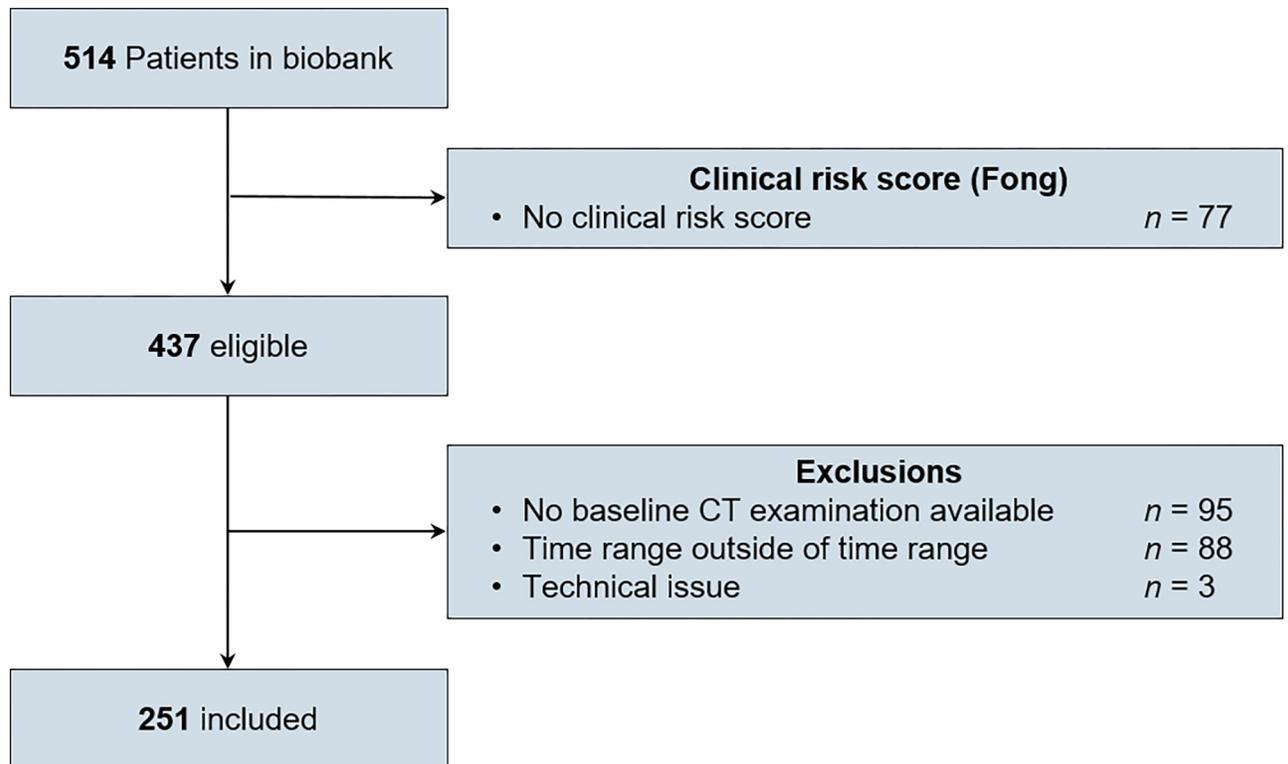
**Fig 1. Flowchart of patient selection.** Patients were selected from a prospectively maintained registry based on their imaging and treatment protocols. CT = computed tomography.

https://doi.org/10.1371/journal.pone.0307815.g001

if needed, by a fellowship-trained radiologist. These masks were used to extract the radiomics features from liver metastases on CT examinations. The segmentation process was done using a free open-source image postprocessing software (The Medical Imaging Interaction Toolkit [MITK]; Heidelberg, Germany) [23]. Binary segmentation masks were saved for each CRLM.

**Radiomics features computation.** All images and masks have been resampled to 1 x 1 x 1 mm$^3$ using the B-spline interpolation method to maintain a constant spatial sampling. Images were clipped between -100 and 200 Hounsfield units (HU) and bin width was set to 25. Masks
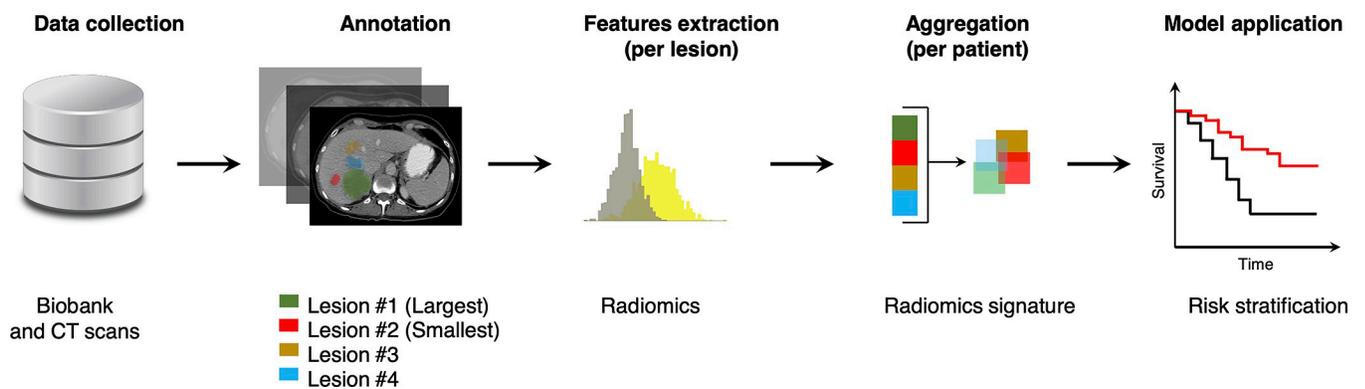


**Fig 2. Data workflow.** Patients from our prospectively maintained registry were selected based on their imaging and treatment protocols. CT exams were collected. All liver metastases were manually segmented. For each metastasis, 3-D radiomics were extracted and indexed, allowing multiple aggregation strategies. The radiomics signatures were used to stratify the risk of tumor recurrence and patient survival. CT = computed tomography.

https://doi.org/10.1371/journal.pone.0307815.g002

and images were cropped one millimeter around lesion contours. Radiomics features from 3D lesions and masks were then extracted using PyRadiomics library [24]. Extracted features included first order, shape, gray level co-occurrence matrix (GLCM), gray level run length matrix (GLRLM), gray level size zone matrix (GLSZM), neighbouring gray tone difference matrix (NGTDM), gray level dependence matrix (GLDM) applied on original image, Laplacian of Gaussian filtered images with sigma ranging from 1 to 5 mm, and wavelet transformed images.

**Radiomics features aggregation.** For each segmented metastasis, a 1-D array of 1,317 radiomics features is obtained. Thus, for a patient exhibiting $m$ metastases a 2-D array of ($m \times 1,317)$ size is obtained. Since patients may present various amount of metastases, radiomic features extracted at the lesion scale need to be aggregated in order to leverage each lesion information and to retrieve features at patient scale, aiming further prediction (*e.g.*, disease free survival, recurrence). Multiple aggregation strategies have been proposed in [25]. Investigated aggregation strategies are presented in **Fig 3**. For comparison, we also considered the smallest lesion.

For all strategies involving multiple lesions per patient, geometric features were summed while texture related features were averaged, as proposed in [25]. Thus, for each aggregation strategy, a unique radiomics dataset was obtained.
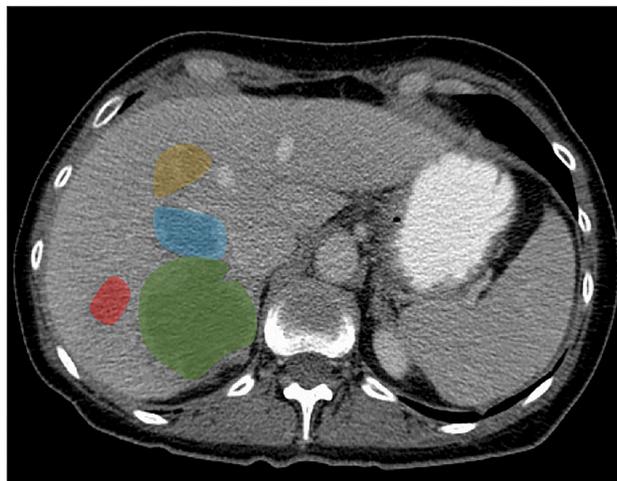
## Ground truth

Time to recurrence (TTR) was defined as the delay between date of curative surgery with complete CRLM resection and first diagnosis of tumor recurrence. Disease-specific survival (DSS) was defined as the delay between date of curative surgery with complete CRLM resection and cancer-related death. Patients who did not experience recurrence or cancer-related death were censored, with a time of observation set as the last follow up date.

## Dataset composition

For each TTR and DSS studies, a holdout test dataset representing 15% of the overall dataset was extracted, ensuring a similar ratio of events in both training and test datasets. *Radiomics signature.*—In order to reduce the number of features and to promote learning process, we first applied minimum redundancy maximum relevancy (MRMR) algorithm [26] on training datasets for both TTR and DSS studies, through its R language implementation [27], to extract the 50 first ranked features. Secondly, we used univariate Cox models to further reduce the number of features, keeping only features exhibiting a concordance index greater than 0.53 on the training dataset. Radiomics signature, which consists in a linear combination of radiomic features was established using a Cox proportional model using Lasso method as feature reduction method. Non-zero features and their associated coefficients were indexed to compute radiomics signature for each aggregation strategies. To compute the statistical distribution of radiomics features, models were trained 100 times on a random subset representing 85% of the training dataset. Such approach avoids data leaking by excluding the test dataset in the feature selection step, while allowing bootstrapping.

## Recurrence and survival prediction

**Models.** We considered two distinct survival models, namely random survival forest (RSF) [28] from the *Scikit-Surv* package [29], and neural network survival models derived from DeepSurv architecture proposed in [30]. Backbone model consisted in two consecutive blocks, each consisting in a 4-node fully connected layers, batch normalization, and a 0.1 coefficient dropout layer. Combination of radiomic features with CRS consisted in concatenating CRS to selected features in survival models training datasets.
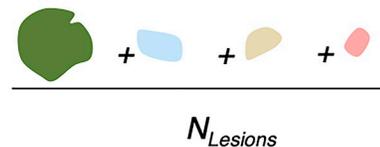
**Fig 3. Investigated aggregation strategies.** Considering multiple metastases, per lesion radiomic features are combined in six distinct schemes. Coefficients *a*, *b*,*c* and *d* indicate lesions volume ratios.

For each model, training was performed 100 times on distinct random splits of the train dataset at a 85% ratio. Metrics on holdout dataset were indexed, providing statistics for survival analysis and classification tasks. Survival models were trained using either radiomic only or in combination with CRS. Classification step consisted in applying logistic regression to survival radiomics signatures scores alone or combined with CRS (1 or 2 features). Same approach was

used for DeepSurv scores. To avoid overfitting, neural network survival models training was stopped at the lowest validation dataset loss, which consisted in the 15% remaining from the train dataset split. Adam optimizer with a 0.00001 learning rate was used for training. In order to address radiomic signatures relevance, standard classification from selected radiomic features is investigated for both TTR and DSS prediction using logistic regression and random forest classifiers.

## Statistical analysis

Concordance index, which is a common evaluation metric of survival models, relies on correlation between predicted score and observed events. Concordance index was computed for CRS alone, radiomics signature alone, and the combination of both. Area under the curve (AUC) was computed from receiver operating characteristic curves and used as classification metric. To compute the AUC, the dichotomization was based on the time interval that provided balanced groups, which was recurrence status at 18 months for TTR and survival at 3 years for DSS.

Confidence intervals were computed using bootstrapping method over two thousands iterations.

## Results

### Study database

**Table 1** summarizes characteristics of patients, tumors, clinical risk score, and chemotherapy received. The majority of patients were males (64.3%). The mean age of patients was 62.9 ± 9.5 years. The size of metastases ranged from 5 to 205 mm. The clinical risk score was distributed as follows: 7, 42, 93, 84, 14, and 1 patients had Fong's scores of 0 to 5, respectively. All patients (241/241) received chemotherapy before surgery and a majority of patients (202/241) also received chemotherapy after surgery.

### Time to recurrence prediction

Bar plots of concordance indexes obtained for TTR prediction using survival models across aggregation strategies are shown in **Fig 4**. For RSF, the highest performance was obtained with 'unweighted average' in combination with CRS which provided a concordance index of 0.61 (95% confidence interval: 0.60–0.61). For DeepSurv-44 the highest performance was obtained with 'largest lesion + lesions count' aggregation in combination with CRS which provided a concordance index of 0.70 (0.68–0.73).

### Disease-specific survival prediction

Bar plots of concordance indexes obtained for DSS prediction using survival models across aggregation strategies are shown in **Fig 5**. For RSF, the highest performance was obtained with the aggregation 'weighted average of three largest lesions' which provided a concordance index of 0.57 (0.56–0.57). For DeepSurv-44 the highest performance was obtained with the aggregation 'largest lesion only' which provided a concordance index of 0.60 (0.58–0.61).

### Time to recurrence classification

Bar plots of AUCs obtained for TTR classification using radiomic signatures across aggregation strategies are shown in **Fig 6**. For radiomics signature, the highest performance was obtained with the aggregation 'largest and total metastatic volume' providing an AUC of 0.77

**Table 1. Clinical characteristics of 241 patients.**

| Characteristics | Data |
|---|---|
| **Patients** | |
| Sex | |
| Male | 155/241 (64.3) |
| Female | 86/241 (35.7) |
| Age (years) ± SD | 62.9 ± 9.5 |
| Number of liver metastases before chemotherapy | 3.0 ± 3.0 |
| **Tumors** | |
| Metastases size (mm) | |
| Mean ± SD | 32.7 ± 22.9 |
| Range | 5.3–204.6 |
| **Clinical risk score** | |
| Fong's score | |
| 0 | 7/241 (2.9) |
| 1 | 42/241 (17.4) |
| 2 | 93/241 (38.6) |
| 3 | 84/241 (34.9) |
| 4 | 14/241 (5.8) |
| 5 | 1/241 (0.4) |
| **Chemotherapy** | |
| Received chemotherapy before surgery | 241/241 (100) |
| First chemotherapy duration (days) | 66.0 ± 50.0 |
| Folfox-based | 217/241 (90.0) |
| 5-fluorouracil-based combination | 239/241 (99.2) |
| Received chemotherapy after surgery | 202/241 (85.9) |
| First post-surgery chemotherapy duration (days) | 91.8 ± 55.6 |
| Folfox-based | 155/202 (76.7) |
| 5-fluorouracil-based combination | 187/202 (92.6) |

Numbers in parentheses are percentages. SD = standard deviation.

https://doi.org/10.1371/journal.pone.0307815.t001

(0.75–0.78). For DeepSurv-44, the highest performance was obtained with aggregation 'Largest lesion only' providing an AUC of 0.58 (0.57–0.59).

Bar plots of AUCs obtained for TTR classification using selected radiomic features across aggregation strategies are shown in **S1 Fig**. For logistic regression, the highest performance was obtained with the aggregation 'smallest' combine with CRS providing an AUC of 0.40 (0.39–0.41). For random forest, the highest performance was obtained with aggregation 'Unweighted average' combined with CRS, providing an AUC of 0.50 (0.48–0.51).

### Disease-specific survival classification

Bar plots of AUCs obtained for disease-specific survival classification using radiomic signatures across aggregation strategies are shown in **Fig 7**. For radiomics signature, the highest performance was obtained with the aggregation 'largest lesion only' combined with CRS which provided an AUC of 0.57 (0.56–0.57). For DeepSurv, the highest performance was obtained with the aggregation 'largest lesion only' which provided an AUC of 0.75 (0.74–0.76).

Bar plots of AUCs obtained for disease-specific survival classification using selected radiomic signatures across aggregation strategies are shown in **S2 Fig**. For logistic regression, the

**Fig 4. Time to recurrence (TTR) prediction using survival models.** Concordance indexes obtained on holdout test dataset using survival models across aggregation strategies for TTR prediction with (A) random survival forests and (B) DeepSurv. Green dashed line indicates performance of CRS alone.

highest performance was obtained with the aggregation 'weighted average of the three largest lesions' which provided an AUC of 0.55 (0.54–0.57). For random forest, the highest performance was obtained with the aggregation 'weighted average' combined with CRS which provided an AUC of 0.63 (0.62–0.65).



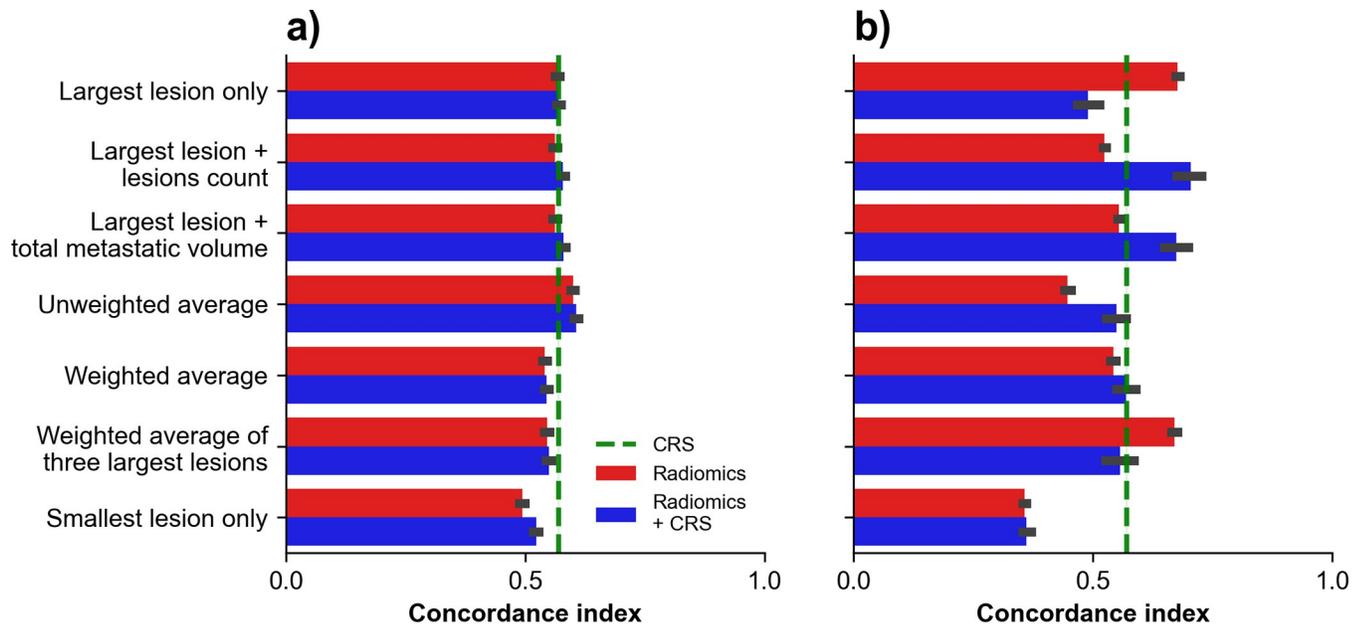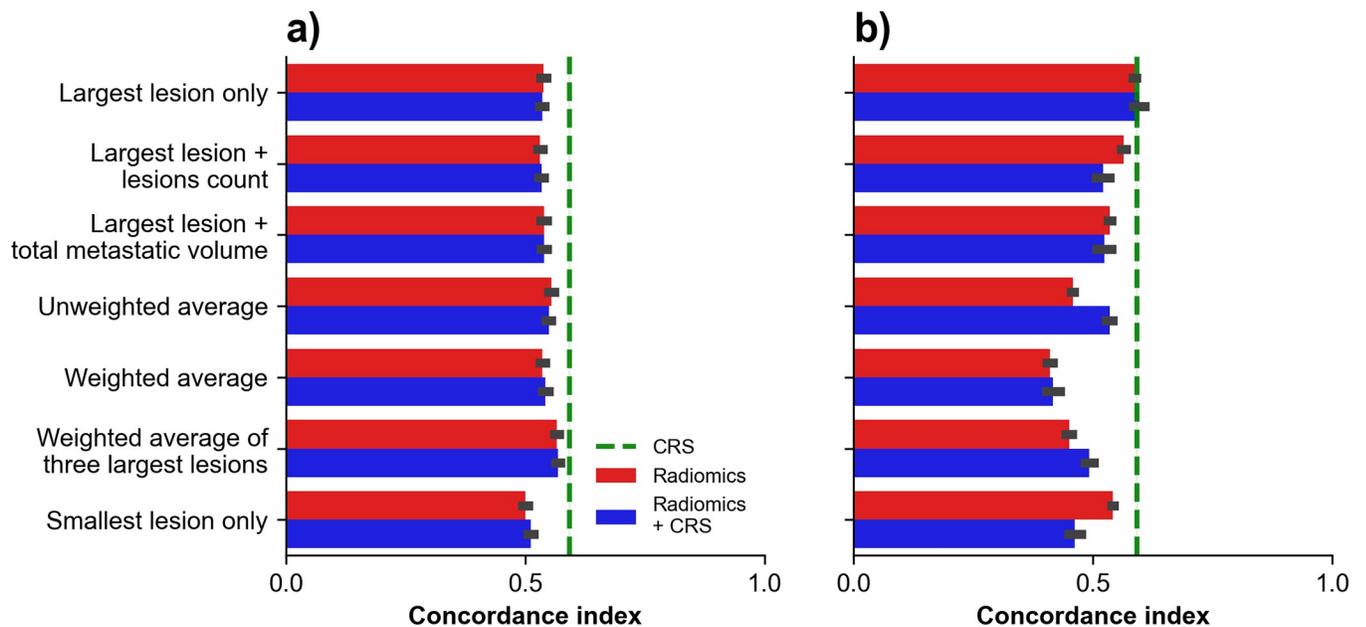**Fig 5. Disease-specific survival (DSS) prediction using survival models.** Concordance indexes obtained on holdout test dataset using survival models across aggregation strategies for DSS prediction with (A) random survival forests and (B) DeepSurv. Green dashed line indicates performances of CRS alone.

**Fig 6. Time to recurrence (TTR) classification using radiomics signatures.** AUC obtained on holdout test dataset using radiomics signatures across aggregation strategies for TTR classification with (A) radiomics signature and (B) DeepSurv score. Green dashed line indicates performances of CRS alone.

https://doi.org/10.1371/journal.pone.0307815.g006

## Radiomics signature coefficients

Box plot depicting variability of selected features coefficients and count plot of radiomics signature coefficients selections over the splits are shown in **Fig 8** for 'Largest lesion only' in TTR study. A majority of features were selected in about half of the splits depicting their



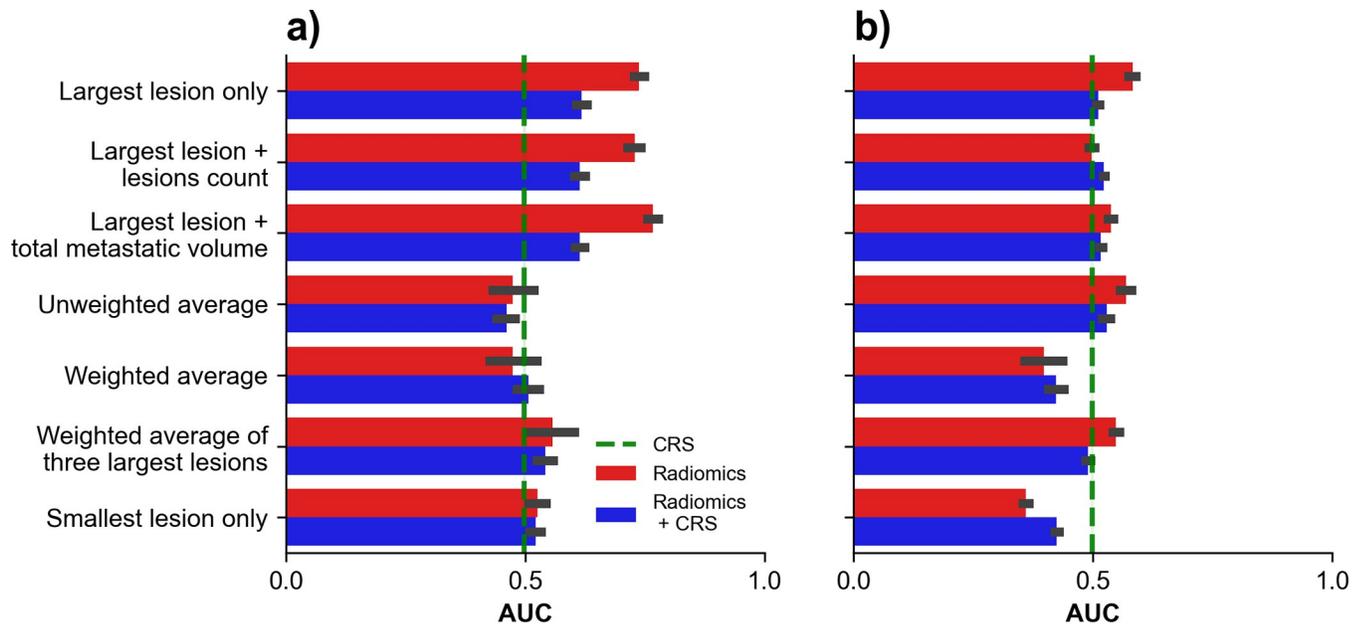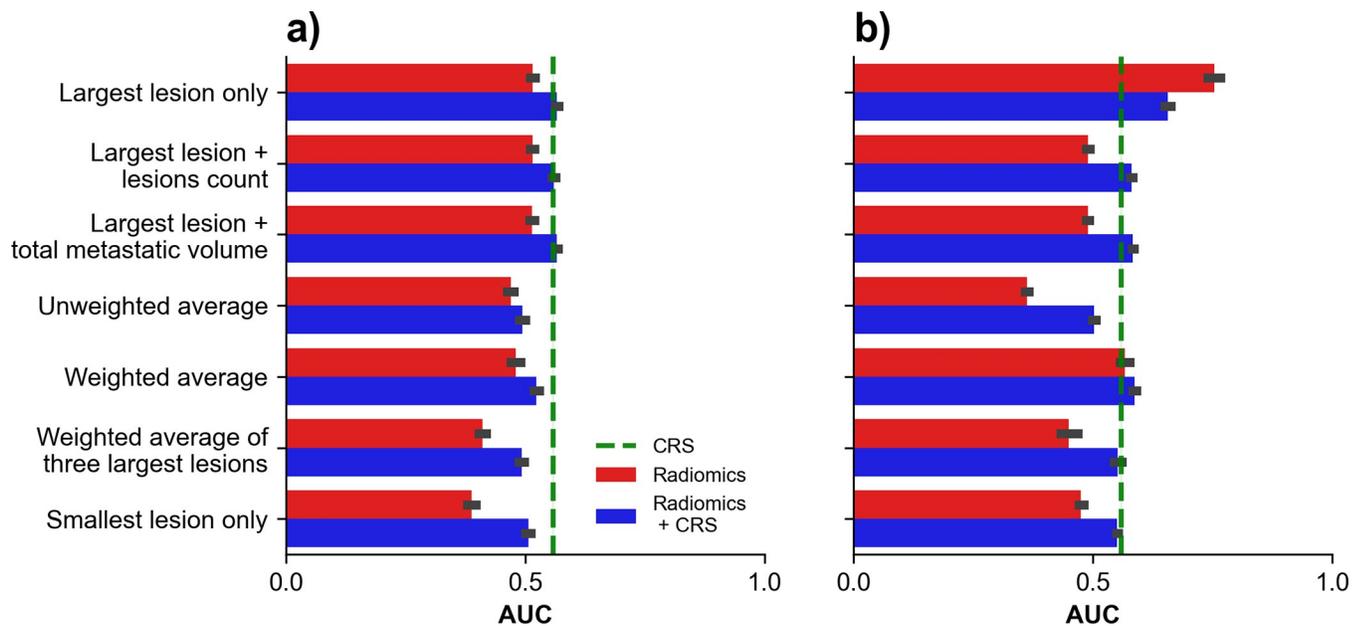**Fig 7. Disease-specific survival (DSS) classification using radiomics signatures.** AUC obtained on holdout test dataset using radiomics signatures across aggregation strategies for DSS classification with (A) radiomics signature and (B) DeepSurv score. Green dashed line indicates performances of CRS alone.
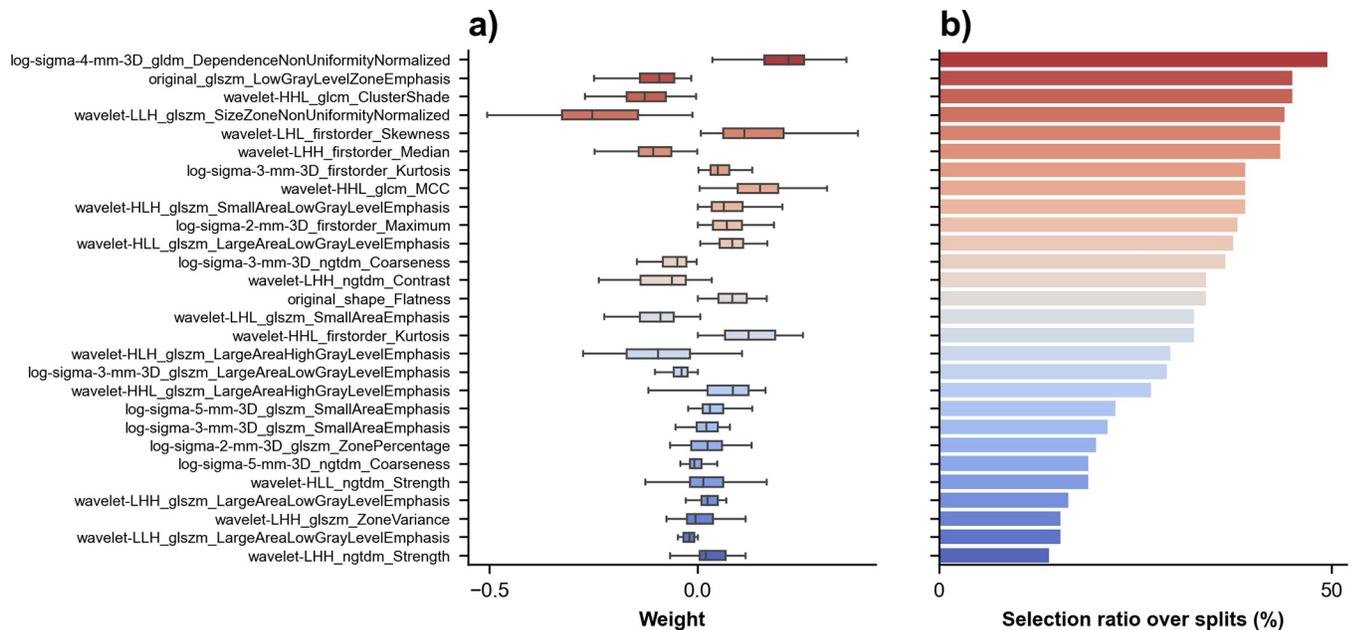
https://doi.org/10.1371/journal.pone.0307815.g007

**Fig 8. Radiomics signature coefficients for prediction of recurrence.** (A) Box plot shows selected features coefficients sorted by descending order of selection ratio over splits. (B) Count plot of radiomics signature coefficients selections over the splits for the aggregation the radiomics signature coefficients with the aggregation 'largest lesion only'.

contribution to the final radiomics signature. Similar plot for DSS study is provided in **S3 Fig**. Initial sets of selected features for 'Largest lesion only' are provided in **S2 Table**. Dataset is provided in **S1 File**.

**S3 Table** indexes selected features for DeepSurv-44 training per aggregation type.

**Fig 9** shows representative examples of patients with shortest and longest survival times.

## Discussion

In this retrospective study derived from a prospective registry, we evaluated the use of radiomics features and signatures on pre-treatment, baseline CT images, to predict recurrence and survival after complete resection of CRLM. We initially identified radiomics features providing the highest prediction performance using dedicated feature selection methods. We compared different feature aggregation strategies to account for the presence of multiple CRLM in patients.

Demircioğlu A et al. [31] has shown that performing feature selection before cross-validation may induce data leakage and bias results. In this study, radiomics features were selected on the training dataset only, thus avoiding any data leakage from the holdout test dataset. We considered linear combinations of radiomics features at the patient level (aggregation of multiple radiomics features across lesions) at each training iteration, over the 100 training dataset splits. We then compared survival models based on CRS, radiomics signatures, and their combinations. A multivariate analysis including neoadjuvant chemotherapy characteristics and histological results has not been performed due to the wide range of treatment strategies. We focused on imaging and clinical data available at baseline, aiming for early prediction of oncological outcomes.

Because patients often have multiple CRLM, we had to investigate various aggregation strategies. 'Largest lesion only' aggregation strategy and its derivatives ranked first in both TTR and
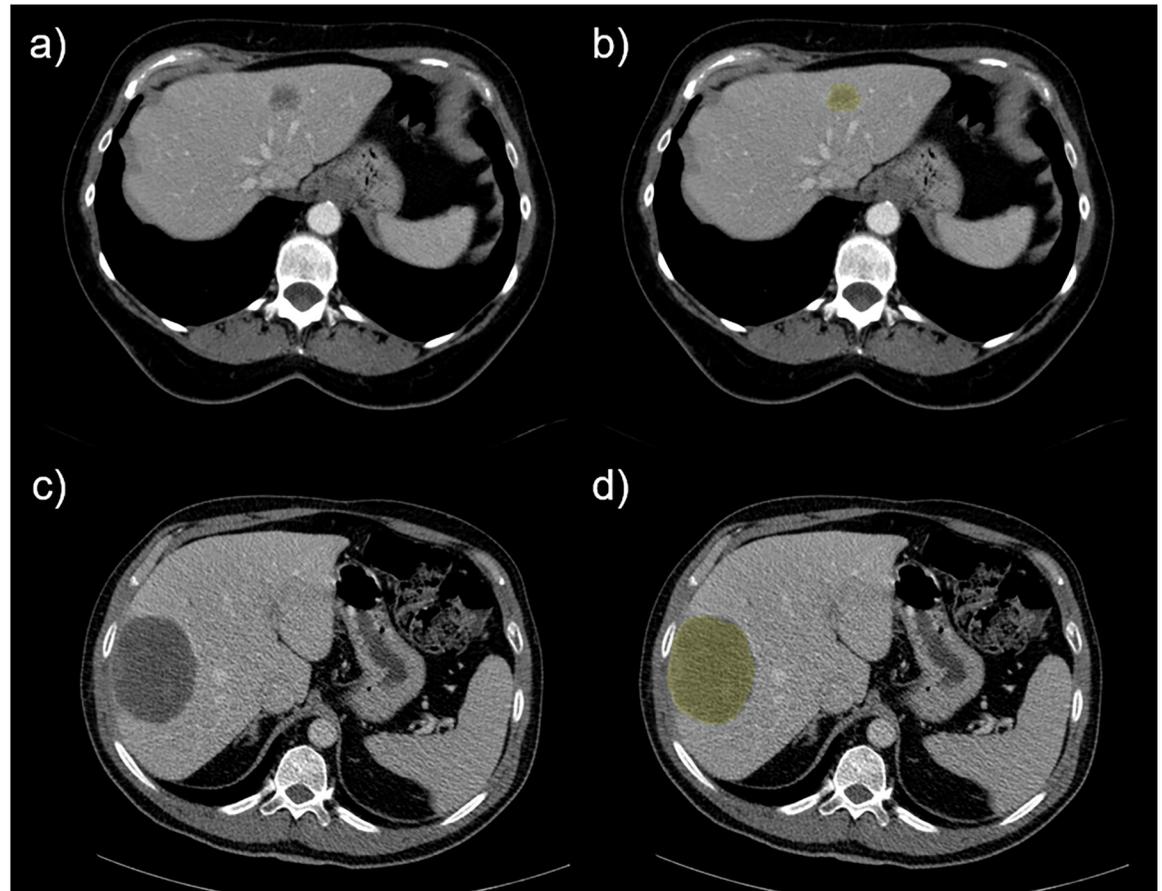
**Fig 9. Liver metastases examples with their associated segmentations.** First row exhibits (A) initial CT image (left) and (B) a segmented metastasis (yellow area) of the patient with the shortest survival in our dataset. Second row exhibits (C) initial CT image (left) and (D) a segmented metastasis (yellow) of the patient with the longest survival in our dataset.

https://doi.org/10.1371/journal.pone.0307815.g009

DSS classification tasks, regardless of the predictor used (radiomics signature or DeepSurv score). This finding is in agreement with the common practice of using the largest tumor among index lesions followed over time for the assessment of treatment response and also with the concept of tumor burden, a well known prognostic factor [32]. Conversely, 'Smallest lesion only' aggregation strategy ranked last in most scenarios, further reinforcing the finding that larger observations have a higher predictive value. Average-based aggregation strategies led to poor-to-moderate performance in almost all cases, except processed by neural networks in some cases. Averaging may induce a loss of characteristic imaging information, especially regarding texture-based features, as large metastases exhibit different textures compared to small ones.

Some recent studies have examined radiomics features to predict treatment response. Lubner et al. has performed CT texture analysis of untreated CRLM and found that different radiomics features predicted pathological response to preoperative chemotherapy and clinical outcomes [33]. Rao et al. previously examined radiomics features such as relative differences in CT texture occurring after treatment of CRLM and found that $\Delta$uniformity and $\Delta$entropy were better for predicting response to chemotherapy than changes in lesion size or volume [18]. Froelich et al. found that increased attenuation of untreated CRLM constituted a prognostic factor of prolonged overall survival [34]. Considering features from 'Largest lesion only'

survival radiomics signature presented in **Fig 8**, one can note the importance of wavelet-filtered image radiomics and the recurrence of features derived from gray level emphasis. Such trend is aligned with observations from Mostafavi et al. [35] regarding treatment response. Furthermore, like Ravanelli et al. [36] and Simpson et al. [37] who highlighted the link between density and correlation with overall survival, we observe three density related features and two correlation related features in our DSS radiomic signature shown in **S3 Fig**.

Including the CRS in radiomics-based models lead to improvements in prediction performance in only a few specific configurations, such as TTR prediction. However, CRS alone provided good concordance indexes despite its apparent simplicity. In this context, CRS or other prognostic system [38] may be used in future studies as a basis for comparison or in a multi-omics context.

The proposed neural network survival models outperformed RSF in our dataset. Since RSF are prone to overfitting, fine-tuned architectures may be required to increase performances, such as reducing the number of classifiers, or maximum depth. This would be consistent with the light neural network architecture used in study, which showed better generalization on test dataset in most of investigated configurations.

In the classification task of TTR and DSS, both logistic regression and random survival forest applied to selected radiomics features provided lower AUCs than radiomics signatures scores, combined or not with CRS, thus confirming the relevance of the proposed approach.

To our knowledge, this is the first study comparing combination of radiomics signatures and clinical risk score across multiple aggregation strategies through RSF and survival network models and validated on a holdout test dataset to provide insights on clinical outcomes in patients with CRLM.

## Limitations

External and prospective validation on independent datasets should be pursued for validation of our findings. Technical parameters such as slice thickness, reconstruction kernel, and scanner manufacturer are known to be sources of variability affecting the statistical distribution of radiomic features [39]. Radiomics reproducibility and robustness remains a current active topic in machine learning applied to medicine [40–42]. Since the objective of this work was to predict outcomes on baseline CT, we did not evaluate use the Response Evaluation Criteria In Solid Tumors (RECIST), which is also not used to guide surgical decision in patients with resectable CRLM in practice.

## Conclusions

In conclusion, radiomics features using baseline CT provided a good accuracy for predicting recurrence and disease-specific survival in patients with CRLM undergoing chemotherapy and complete resection. These results await validation of these imaging-based prognostic biomarkers in larger cohort to guide the design of future prospective studies.

## Supporting information

**S1 Fig. Time to recurrence (TTR) classification using selected radiomic features.** AUC obtained on holdout test dataset using radiomics features across aggregation strategies for TTR classification using (A) logistic regression and (B) random forest. Gray dashed line indicates randomness.
(TIFF)

**S2 Fig. Disease-specific survival (DSS) classification using selected radiomic features.** AUC obtained on holdout test dataset using radiomics features across aggregation strategies for DSS classification using (A) logistic regression and (B) random forest. Gray dashed line indicates randomness.
(TIFF)

**S3 Fig. Radiomics signature coefficients for prediction of survival.** (A) Box plot shows selected features coefficients sorted by descending order of selection ratio over splits. (B) Count plot of radiomics signature coefficients selections over the splits for the aggregation the radiomics signature coefficients with the aggregation 'largest lesion only'.
(TIFF)

**S1 Table. Scanners and imaging parameters characteristics.**
(DOCX)

**S2 Table. Radiomic features selected prediction of recurrence and survival for 'Largest lesion only' aggregation.** Radiomic features selected using 'Largest lesion only' aggregation for recurrence and survival predictions, and the intersection of the two features sets.
(DOCX)

**S3 Table. Selected radiomic features for DeepSurv-44 per aggregation type.**
(DOCX)

**S1 File. Radiomics datasets per aggregation type.**
(ZIP)

**S4 CRLM Dataset.**
(ZIP)

## Acknowledgments

## Author Contributions

**Conceptualization:** An Tang.

**Data curation:** Emmanuel Montagnon, Milena Cerny, Vincent Hamilton, Thomas Derennes, André Ilinca.

**Funding acquisition:** An Tang.

**Investigation:** Emmanuel Montagnon, Milena Cerny, Mohamed El Amine Elforaici, Gilbert Jabbour, Edmond Rafie, Anni Wu, Francisco Perdigon Romero, Alexandre Cadrin-Chênevert.

**Methodology:** Vincent Hamilton, Thomas Derennes, André Ilinca, Gilbert Jabbour, Edmond Rafie, Anni Wu, Francisco Perdigon Romero, Samuel Kadoury, Simon Turcotte.

**Project administration:** An Tang.

**Resources:** Simon Turcotte.

**Software:** Francisco Perdigon Romero.

**Supervision:** Samuel Kadoury, Simon Turcotte, An Tang.

**Validation:** Alexandre Cadrin-Chênevert, Samuel Kadoury.

**Visualization:** Mohamed El Amine Elforaici, Francisco Perdigon Romero.

**Writing – original draft:** Emmanuel Montagnon, Milena Cerny, Mohamed El Amine Elforaici, Alexandre Cadrin-Chênevert, Samuel Kadoury, Simon Turcotte, An Tang.

**Writing – review & editing:** Emmanuel Montagnon, Milena Cerny, Gilbert Jabbour, Edmond Rafie, Anni Wu, Francisco Perdigon Romero, Alexandre Cadrin-Chênevert, Samuel Kadoury, Simon Turcotte, An Tang.

# References

1. Brenner DR, Weir HK, Demers AA, Ellison LF, Louzado C, Shaw A, et al. Projected estimates of cancer in Canada in 2020. CMAJ. 2020 Mar 2; 192(9):E199–205. https://doi.org/10.1503/cmaj.191292 PMID: 32122974

2. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2018 Nov; 68(6):394–424. https://doi.org/10.3322/caac.21492 PMID: 30207593

3. Vecchio FM, Valentini V, Minsky BD, Padula GDA, Venkatraman ES, Balducci M, et al. The relationship of pathologic tumor regression grade (TRG) and outcomes after preoperative therapy in rectal cancer. Int J Radiat Oncol Biol Phys. 2005 Jul 1; 62(3):752–60. https://doi.org/10.1016/j.ijrobp.2004.11.017 PMID: 15936556

4. Mocellin S, Pasquali S, Nitti D. Fluoropyrimidine-HAI (hepatic arterial infusion) versus systemic chemotherapy (SCT) for unresectable liver metastases from colorectal cancer. Cochrane Database Syst Rev. 2009 Jul 8;(3):CD007823. https://doi.org/10.1002/14651858.CD007823.pub2 PMID: 19588444

5. Kusano M, Honda M, Okabayashi K, Akimaru K, Kino S, Tsuji Y, et al. Randomized controlled Phase III study comparing hepatic arterial infusion with systemic chemotherapy after curative resection for liver metastasis of colorectal carcinoma: JFMC 29–0003. J Cancer Res Ther. 2017 Mar; 13(1):84–90. https://doi.org/10.4103/0973-1482.184524 PMID: 28508838

6. Fong Y, Sun RL, Jarnagin W, Blumgart LH. An analysis of 412 cases of hepatocellular carcinoma at a Western center. Ann Surg. 1999 Jun; 229(6):790–9; discussion 799–800. https://doi.org/10.1097/00000658-199906000-00005 PMID: 10363892

7. Tomlinson JS, Jarnagin WR, DeMatteo RP, Fong Y, Kornprat P, Gonen M, et al. Actual 10-year survival after resection of colorectal liver metastases defines cure. J Clin Oncol Off J Am Soc Clin Oncol. 2007 Oct 10; 25(29):4575–80. https://doi.org/10.1200/JCO.2007.11.0833 PMID: 17925551

8. Fong Y, Fortner J, Sun RL, Brennan MF, Blumgart LH. Clinical Score for Predicting Recurrence After Hepatic Resection for Metastatic Colorectal Cancer. Ann Surg. 1999 Sep; 230(3):309.

9. Zhou SK, Greenspan H, Davatzikos C, Duncan JS, van Ginneken B, Madabhushi A, et al. A review of deep learning in medical imaging: Image traits, technology trends, case studies with progress highlights, and future promises. ArXiv200809104 Cs Eess [Internet]. 2020 Aug 2 [cited 2021 Feb 10]; Available from: http://arxiv.org/abs/2008.09104

10. Choy G, Khalilzadeh O, Michalski M, Do S, Samir AE, Pianykh OS, et al. Current Applications and Future Impact of Machine Learning in Radiology. Radiology. 2018 Jun 26; 288(2):318–28. https://doi.org/10.1148/radiol.2018171820 PMID: 29944078

11. Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. Med Image Anal. 2017 Dec 1; 42:60–88. https://doi.org/10.1016/j.media.2017.07.005 PMID: 28778026

12. Girshick R, Donahue J, Darrell T, Malik J. Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014. p. 580–7.

13. Ren S, He K, Girshick R, Sun J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In: Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R, editors. Advances in Neural Information Processing Systems 28. Curran Associates, Inc.; 2015. p. 91–9.

14. Cai L, Gao J, Zhao D. A review of the application of deep learning in medical image classification and segmentation. Ann Transl Med. 2020 Jun; 8(11):713. https://doi.org/10.21037/atm.2020.02.44 PMID: 32617333

15. Reddy CK, Li Y. Machine Learning for Survival Analysis. Stat Methods.: 107.

16. Zou ZM, Chang DH, Liu H, Xiao YD. Current updates in machine learning in the prediction of therapeutic outcome of hepatocellular carcinoma: what should we know? Insights Imaging. 2021 Mar 6; 12(1):31. https://doi.org/10.1186/s13244-021-00977-9 PMID: 33675433

17. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. Radiology. 2016 Feb; 278(2):563–77. https://doi.org/10.1148/radiol.2015151169 PMID: 26579733

18. Rao SX, Lambregts DM, Schnerr RS, Beckers RC, Maas M, Albarello F, et al. CT texture analysis in colorectal liver metastases: A better way than size and volume measurements to assess response to chemotherapy? United Eur Gastroenterol J. 2016 Apr; 4(2):257–63. https://doi.org/10.1177/2050640615601603 PMID: 27087955

19. Dercle L, Lu L, Schwartz LH, Qian M, Tejpar S, Eggleton P, et al. Radiomics Response Signature for Identification of Metastatic Colorectal Cancer Sensitive to Therapies Targeting EGFR Pathway. JNCI J Natl Cancer Inst. 2020 Sep 1; 112(9):902–12. https://doi.org/10.1093/jnci/djaa017 PMID: 32016387

20. Fiz F, Viganò L, Gennaro N, Costa G, La Bella L, Boichuk A, et al. Radiomics of Liver Metastases: A Systematic Review. Cancers. 2020 Oct; 12(10):2881. https://doi.org/10.3390/cancers12102881 PMID: 33036490

21. Canadian Tissue Repository Network [Internet]. CTRNet. [cited 2022 Feb 23]. Available from: https://www.ctrnet.ca/en/resources/biobanking-software/

22. Chen YY, Perera DS, Yan TD, Schmidt LM, Morris DL. Applying Fong's CRS Liver Score in Patients with Colorectal Liver Metastases Treated by Cryotherapy. Asian J Surg. 2006 Oct 1; 29(4):238–41. https://doi.org/10.1016/S1015-9584(09)60095-6 PMID: 17098655

23. The Medical Imaging Interaction Toolkit (MITK)—mitk.org [Internet]. [cited 2021 Aug 24]. Available from: https://www.mitk.org/wiki/The_Medical_Imaging_Interaction_Toolkit_(MITK)

24. van Griethuysen JJ, Fedorov A, Parmar C, Hosny A, Aucoin N, Narayan V, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res. 2017 Nov 1; 77(21):e104–7. https://doi.org/10.1158/0008-5472.CAN-17-0339 PMID: 29092951

25. 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 May 7; 11(1):9758. https://doi.org/10.1038/s41598-021-89114-6 PMID: 33963236

26. Ding C, Peng H. Minimum redundancy feature selection from microarray gene expression data. J Bioinform Comput Biol. 2005 Apr; 3(2):185–205. https://doi.org/10.1142/s0219720005001004 PMID: 15852500

27. De Jay N, Papillon-Cavanagh S, Olsen C, El-Hachem N, Bontempi G, Haibe-Kains B. mRMRe: an R package for parallelized mRMR ensemble feature selection. Bioinformatics. 2013 Sep 15; 29 (18):2365–8. https://doi.org/10.1093/bioinformatics/btt383 PMID: 23825369

28. Ishwaran H, Kogalur UB, Blackstone EH, Lauer MS. Random survival forests. Ann Appl Stat. 2008; 2 (3):841–60.

29. Pölsterl S. scikit-survival: A Library for Time-to-Event Analysis Built on Top of scikit-learn. J Mach Learn Res. 2020;

30. Katzman J, Shaham U, Bates J, Cloninger A, Jiang T, Kluger Y. DeepSurv: Personalized Treatment Recommender System Using A Cox Proportional Hazards Deep Neural Network. BMC Med Res Methodol. 2018 Dec; 18(1):24. https://doi.org/10.1186/s12874-018-0482-1 PMID: 29482517

31. Demircioğlu A. Measuring the bias of incorrect application of feature selection when using cross-validation in radiomics. Insights Imaging. 2021 Nov 24; 12(1):172. https://doi.org/10.1186/s13244-021-01115-1 PMID: 34817740

32. Miszczuk M, Chapiro J, Do Minh D, van Breugel JMM, Smolka S, Rexha I, et al. Analysis of Tumor Burden as a Biomarker for Patient Survival with Neuroendocrine Tumor Liver Metastases Undergoing Intra-Arterial Therapies: A Single-Center Retrospective Analysis. Cardiovasc Intervent Radiol. 2022 Oct; 45(10):1494–502. https://doi.org/10.1007/s00270-022-03209-9 PMID: 35941241

33. Lubner MG, Stabo N, Lubner SJ, del Rio AM, Song C, Halberg RB, et al. CT textural analysis of hepatic metastatic colorectal cancer: pre-treatment tumor heterogeneity correlates with pathology and clinical outcomes. Abdom Imaging. 2015 Oct; 40(7):2331–7. https://doi.org/10.1007/s00261-015-0438-4 PMID: 25968046

34. Froelich MF, Heinemann V, Sommer WH, Holch JW, Schoeppe F, Hesse N, et al. CT attenuation of liver metastases before targeted therapy is a prognostic factor of overall survival in colorectal cancer patients. Results from the randomised, open-label FIRE-3/AIO KRK0306 trial. Eur Radiol. 2018 Dec; 28 (12):5284–92. https://doi.org/10.1007/s00330-018-5454-7 PMID: 29882070

35. Mostafavi L, Homayounieh F, Lades F, Primak A, Muse V, Harris GJ, et al. Correlation of Radiomics with Treatment Response in Liver Metastases. Acad Radiol. 2023 Dec 11;S1076-6332(23)00625-6. https://doi.org/10.1016/j.acra.2023.11.007 PMID: 38087718

36. Ravanelli M, Agazzi GM, Tononcelli E, Roca E, Cabassa P, Baiocchi G, et al. Texture features of colorectal liver metastases on pretreatment contrast-enhanced CT may predict response and prognosis in patients treated with bevacizumab-containing chemotherapy: a pilot study including comparison with standard chemotherapy. Radiol Med (Torino). 2019 Sep 1; 124(9):877–86. https://doi.org/10.1007/s11547-019-01046-4 PMID: 31172448

37. 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 Sep; 24(9):2482–90. https://doi.org/10.1245/s10434-017-5896-1 PMID: 28560599

38. Liu W, Zhang W, Xu Y, Li YH, Xing BC. A Prognostic Scoring System to Predict Survival Outcome of Resectable Colorectal Liver Metastases in this Modern Era. Ann Surg Oncol. 2021 Nov; 28(12):7709–18. https://doi.org/10.1245/s10434-021-10143-6 PMID: 34023948

39. Meyer M, Ronald J, Vernuccio F, Nelson RC, Ramirez-Giraldo JC, Solomon J, et al. Reproducibility of CT Radiomic Features within the Same Patient: Influence of Radiation Dose and CT Reconstruction Settings. Radiology. 2019 Oct 1; 293(3):583–91. https://doi.org/10.1148/radiol.2019190928 PMID: 31573400

40. Ger RB, Zhou S, Chi PCM, Lee HJ, Layman RR, Jones AK, et al. Comprehensive Investigation on Controlling for CT Imaging Variabilities in Radiomics Studies. Sci Rep. 2018 Aug 29; 8(1):13047. https://doi.org/10.1038/s41598-018-31509-z PMID: 30158540

41. Mackin D, Fave X, Zhang L, Fried D, Yang J, Taylor B, et al. Measuring Computed Tomography Scanner Variability of Radiomics Features. Invest Radiol. 2015 Nov; 50(11):757–65. https://doi.org/10.1097/RLI.0000000000000180 PMID: 26115366

42. Orlhac F, Frouin F, Nioche C, Ayache N, Buvat I. Validation of A Method to Compensate Multicenter Effects Affecting CT Radiomics. Radiology. 2019 Jan 29; 291(1):53–9. https://doi.org/10.1148/radiol.2019182023 PMID: 30694160