

ORIGINAL ARTICLE

Open Access

Impact of inter-reader contouring variability on textural radiomics of colorectal liver metastases



Francesco Rizzetto¹, Francesca Calderoni², Cristina De Mattia², Arianna Defeudis^{3,4}, Valentina Giannini^{3,4}, Simone Mazzetti^{3,4}, Lorenzo Vassallo⁵, Silvia Ghezzi⁶, Andrea Sartore-Bianchi^{6,7}, Silvia Marsoni⁸, Salvatore Siena^{6,7}, Daniele Regge^{3,4}, Alberto Torresin^{2,9} and Angelo Vanzulli^{1,7*}

Abstract

Background: Radiomics is expected to improve the management of metastatic colorectal cancer (CRC). We aimed at evaluating the impact of liver lesion contouring as a source of variability on radiomic features (RFs).

Methods: After Ethics Committee approval, 70 liver metastases in 17 CRC patients were segmented on contrast-enhanced computed tomography scans by two residents and checked by experienced radiologists. RFs from grey level co-occurrence and run length matrices were extracted from three-dimensional (3D) regions of interest (ROIs) and the largest two-dimensional (2D) ROIs. Inter-reader variability was evaluated with Dice coefficient and Hausdorff distance, whilst its impact on RFs was assessed using mean relative change (MRC) and intraclass correlation coefficient (ICC). For the main lesion of each patient, one reader also segmented a circular ROI on the same image used for the 2D ROI.

Results: The best inter-reader contouring agreement was observed for 2D ROIs according to both Dice coefficient (median 0.85, interquartile range 0.78–0.89) and Hausdorff distance (0.21 mm, 0.14–0.31 mm). Comparing RF values, MRC ranged 0–752% for 2D and 0–1567% for 3D. For 24/32 RFs (75%), MRC was lower for 2D than for 3D. An ICC > 0.90 was observed for more RFs for 2D (53%) than for 3D (34%). Only 2/32 RFs (6%) showed a variability between 2D and circular ROIs higher than inter-reader variability.

Conclusions: A 2D contouring approach may help mitigate overall inter-reader variability, albeit stable RFs can be extracted from both 3D and 2D segmentations of CRC liver metastases.

Keywords: Colorectal neoplasms, Image processing (computer-assisted), Liver neoplasms, Radiomics, Tomography (x-ray, computed)

* Correspondence: angelo.vanzulli@unimi.it

¹Department of Radiology, ASST Grande Ospedale Metropolitano Niguarda, Piazza Ospedale Maggiore 3, 20162 Milan, Italy

⁷Department of Oncology and Hemato-Oncology, Università degli Studi di Milano, via Festa del Perdono 7, 20122 Milan, Italy

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



© The Author(s). 2020 **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Key points

- Reader contouring variability may impact on radiomic features of liver metastases from colorectal cancer (CRC).
- Stable textural features against inter-reader variability can be extracted from contrast-enhanced computed tomography images of liver metastases from CRC.
- Two-dimensional contouring seems to be less affected than three dimensional contouring by inter-reader variability.
- Two-dimensional contouring may help reduce variability of readers' lesion segmentation.

Background

In the current era of targeted therapies, the search for imaging biomarkers linking the genetic and molecular characteristics of tumours to the clinical and morpho-functional phenotype is pivotal to provide oncologic patients with more tailored treatment options [1, 2]. A special effort to achieve this goal is being made in colorectal cancer (CRC), one of the most common malignant tumours worldwide [3]. Since 20% of patients with CRC already have liver metastases at the time of diagnosis and up to 50% will develop them within the first 3 years [4], to improve the detection of molecular alterations over time and space of these lesions is crucial to optimise the patient's management [5].

In this context, great expectations were raised by radiomics, namely the quantitative analysis of medical imaging for the extraction of high-throughput data with diagnostic, prognostic and predictive value [6]. Evidence correlating the textural radiomic features (RFs) extracted from the computed tomography (CT) scans of CRC liver metastases with the clinical outcomes of the patients have accumulated in the last few years. For example, texture analysis has been used to predict the tumour grade and overall survival of patients with stage IV CRC before treatment [7, 8], the response of liver metastases to first-line chemotherapy [9, 10] and the risk of liver recurrence after hepatic resection of CRC lesions [11].

However, the extraction of the RFs is a complex process with many steps, each of them characterised by specific issues that could compromise the robustness of the results [12]. Widely studied sources of uncertainty in radiomics are the image acquisition and reconstruction settings and the preprocessing manipulations [13, 14], but inter-reader variability in lesion segmentation is also critical, especially considering that the current standard of reference is manual contouring [15, 16] and that multicentric trials, involving multiple readers, are recommended to assure adequate statistical power [6, 17]. Depending on how the regions of interest (ROIs) are encompassed in the segmentation, the subsequent

quantitative analysis can be significantly modified [18]. For liver metastases, this issue is particularly relevant: given that the tumour type and site are crucial aspects to consider, a higher inter-reader uncertainty is expected for lesions with blurred boundaries or low-contrast interface with the surrounding tissues [16, 19, 20]. Moreover, in terms of reproducibility and predictive value, controversies still exist regarding the choice of including in the segmented ROI the whole lesion or just its more representative cross-section [21–23].

At present, the role of contouring in RFs reproducibility has been addressed in several studies [24–27], but to the best of knowledge, none of them concerned hepatic CRC metastases. Therefore, the aim of this work is to assess the influence of inter-reader contouring variability on the texture analysis of CRC liver metastases, focusing on the role of three- and two-dimensional segmentation in determining RFs robustness. Since different approaches can impact on the results of radiomic studies but also on the time and resources needed for the data collection and analysis, to optimise the contouring strategy is essential.

Methods

Patients

This was an ancillary study conducted on CRC patients ($n = 31$) enrolled from 2016 to 2018 in the multicentric phase II HERACLES trial (NCT03225937), exploring the efficacy of dual human epidermal growth factor receptor 2 blockade in patients harbouring human epidermal growth factor receptor 2-amplified metastatic CRC. Clinical inclusion and exclusion criteria of the trial were previously reported [28, 29]. A further selection was performed to include only patients with liver metastases. The study was approved by the Ethics Committee, and all the patients signed written informed consent to allow the images of their diagnostic examinations to be used for scientific purposes at the time of enrolment in the study.

Reading protocol

Two residents in the radiology department, referred as R1 for hospital 1 and R2 for hospital 2 (3 and 4 years of experience, respectively) reviewed the available imaging. For each patient, only the portal venous phase of an abdominal computed tomography (CT) scan with intravenous injection of iodinated contrast agent was used. This phase is the most used for radiomics of liver metastases and provides the best visualisation of the lesions [30].

Every metastasis was individually evaluated if suitable for the analysis (*i.e.*, to provide adequate textural information) by applying the following exclusion criteria [31, 32]: (a) maximum axial diameter lesser than 10 mm; (b) tumour boundaries not surely identifiable because of artefacts or confluent lesions. Any disagreement about the

selection of specific lesions was resolved through consultation between the two readers.

Both readers measured the largest axial diameter of all the lesions with a digital calliper. The two readers contoured the whole lesion volume using a three-dimensional (3D) region of interest (ROI) and the largest and most representative area on the axial slice using a two-dimensional (2D) ROI of the metastases from the original CT images. For the segmentation task, R1 used 3D Slicer v.4.10.0 (www.slicer.org) and R2 used MIPAV (Medical Image Processing, Analysis and Visualization, <http://mipav.cit.nih.gov>), both of them allowing lesion contouring by generating polygonal meshes. All the sets of ROIs were exported as NIfTI binary labelmaps.

Considering the main lesion for each patient, in the same slice of the 2D ROI, R1 also segmented two circular ROIs (Fig. 1): the smallest inclusive of the whole metastasis and the largest one completely inside it. This additional set of segmentations was intended to assess the impact of a simplified segmentation protocol on inter-reader variability.

All segmentations were finally approved by two experienced radiologists with over 20 years of experience.

Extraction of radiomic features

The RFs were extracted from the ROIs using Imaging Biomarker Explorer (IBEX v. 1.0 β) platform, a free open-source software developed by the MD Anderson Cancer Center (Houston, USA). For this study, 32 textural RFs from the *grey level co-occurrence matrix* (GLCM) and the *grey level run length matrix* (GLRLM) were considered [33]. Also, max 3D diameter, number of voxels, and volume were extracted from the ROIs. For the calculated RFs, the consistency with the image biomarker standardisation

initiative (IBSI) standard was verified using an IBSI-validated in-house developed software [34].

The RF extraction in the study was performed on the original CT image setting in IBEX a fixed range of 480 grey levels (from -200 to 279 HU) discretised in 32 bins, with offset = 1 and symmetry = 1. All the directions allowed by the software were considered. The range was reasoned on the grey level distribution of the overall metastases, whilst the number of bins was chosen as a compromise to limit the noise contribution and the loss of texture information [35].

No resampling nor other preprocessing were applied to the CT images.

Data analysis

All analyses were performed on Microsoft® Office Excel spreadsheet, except for the calculation of the intraclass correlation coefficient (ICC) based on a single-reader two-way random-effects model, which was performed on R v.3.5.1 (“Psych” package). When required, statistical significance was established at the $p < 0.050$ level.

The inter-reader contouring agreement on both 3D and 2D ROIs was evaluated through two similarity indices: average Hausdorff distance (HD) and Dice coefficient (DC), both calculated with “SlicerRT” toolkit [36]. Hausdorff distance measures how far two subsets of a metric space are from each other, thus indicating the longest distance between the boundaries of two contours. The average HD was chosen so as to have a better representation of global contouring discrepancy [37]. Instead, the DC quantifies the spatial overlap between two contours/volumes, ranging from 0 for null overlapping and 1 for perfect overlapping [38]. The two indices emphasise different characteristics: the DC quantifies the

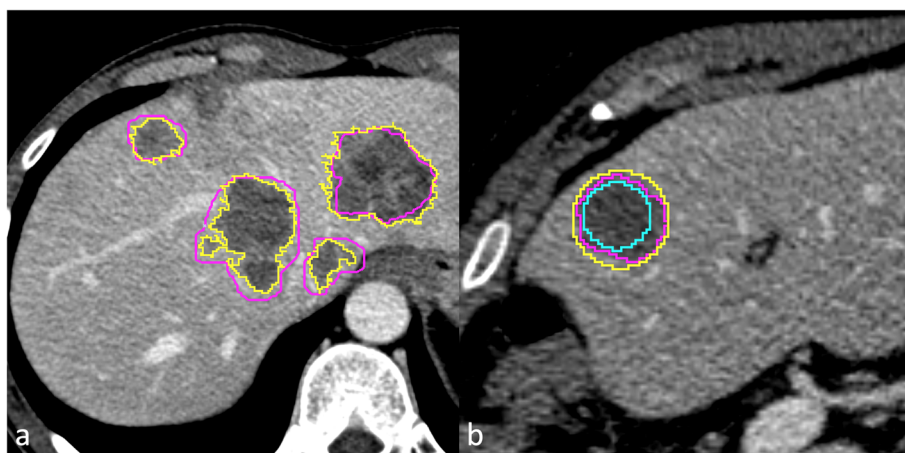


Fig. 1 For each metastasis, the whole lesion volume and the largest axial cross-section were segmented by two readers. **a** Purple line (reader 1) versus yellow line (reader 2) contouring. The largest two-dimensional (2D) region of interest (ROI) of the main lesion was confronted with two circular ROIs, one inside the metastasis and one outside it. **b** Purple line (reader 1) 2D versus yellow line (smallest circular ROI inclusive of the whole lesion) versus azure line (largest circular ROI completely inside the lesion)

discrepancy in voxel labelling, whilst the HD performs better at detecting deviations (spikes, holes, etc.) which alter the contour shape but do not substantially modify the volume [26].

For both DC and HD, calculated over all the segmentations, median value and interquartile range (IQR) were reported. The Wilcoxon signed-rank test was performed to evaluate if there was a significant difference between the indices calculated for 2D and 3D ROIs. Spearman's ρ correlation coefficient between the values of DC and HD was also calculated.

To verify if the size of the metastases could influence the inter-reader agreement, a linear regression analysis was performed to evaluate the association between the following parameters: manual largest axial diameter and max 3D diameter *versus* DC and HD; volume/area (cm^3/cm^2) *versus* DC and HD. The strength of correlation was reported following Evans's interpretation [39].

The influence of inter-reader variability on the extracted RFs was assessed considering the relative change $[(\text{RF}_{R2} - \text{RF}_{R1})/\text{RF}_{R1}]$ both on 3D and 2D ROIs for all lesions. For each RF, the mean relative change (MRC) was calculated. The inter-reader MRC in RF values was also compared with the MRC obtained from R1 2D ROIs against the circular ROIs, taking R1 values as a reference.

The ICC of the RFs between the two readers was calculated to describe how strongly the two datasets resembled each other and so to guide the selection of RFs according to robustness [40]. In accordance with the literature [41], the ICC was interpreted as follows: poor agreement for $\text{ICC} \leq 0.50$; moderate agreement for $0.50 < \text{ICC} \leq 0.75$; good agreement for $0.75 < \text{ICC} \leq 0.90$; excellent agreement for $\text{ICC} > 0.90$.

Results

Of the original 31 patients of the trial, 14 (45%) were excluded because they had no liver metastases ($n = 6$), because of the presence of imaging artefacts ($n = 3$) or because there were only lesions < 10 mm ($n = 2$) or only confluent metastases largely occupying the liver parenchyma ($n = 3$). Therefore, 17 patients from 3 different centres were finally included and, according to the selection criteria, a total of 70 lesions were considered suitable for the analysis by the readers. The demographical data of the included patients are reported in Table 1, whilst detailed information about the acquisition and reconstruction parameters of their CT scans are listed in Table 2.

The largest axial diameter of the selected lesions ranged from 10 to 80 mm, with a median value of 27 mm (IQR 17–29 mm) according to R1 and 26 mm (IQR 16–26 mm) according to R2.

Contouring variability

Moving from 3D to 2D ROIs, an increase in DC and a reduction in HD was observed. Specifically, 3D ROIs showed a median DC of 0.76 (IQR 0.71–0.82) and a median HD of 1.15 mm (IQR 0.90–1.41 mm). For 2D ROIs, the median DC was of 0.85 (IQR 0.78–0.89), and median HD was of 0.21 mm (IQR 0.14–0.31 mm). According to Wilcoxon signed-rank test, these differences were significant for both DC ($p < 0.001$) and HD ($p < 0.001$). Moreover, a very strong negative correlation was found between HD and DC for 2D ROIs ($\rho = -0.85$; $p < 0.001$), but only a weak negative correlation was found for 3D ones ($\rho = -0.38$; $p < 0.001$) (Fig. 2). An example of discrepancy between the two similarity indices is presented in Fig. 3.

Weak-to-moderate correlations ($-0.45 \leq \rho \leq 0.45$) were found between the two similarity indices and size parameters for both 3D and 2D ROIs (Table 3).

Impact on texture analysis

As illustrated in Fig. 4, different RFs showed to have different susceptibility to inter-reader variability. In particular, inter-reader MRC ranged from 0 to 1567% for 3D ROIs and from 0 to 752 for 2D ROIs. The inter-reader discrepancy was below 10% for more than 60% of the RFs extracted from both the sets of ROIs. For 24/32 (75%) RFs, the discrepancies were lower when calculated from 2D than 3D ROIs. Specifically, this applied to 5/11 (45%) of GLRLM RFs and to 19/21 (90%) of GLCM RFs.

The ICC gave similar results in terms of RF robustness (Fig. 5). In particular, 11/32 (34%) RFs for 3D ROIs and 17/32 (53%) RFs for 2D ROIs were found to be very robust ($\text{ICC} > 0.90$). In both cases, the ICC of inter-reader variability ranged from 0.06 to 0.99.

When combining the results from MRC and ICC analysis (Table 4), the following RFs were found most stable: long run emphasis, long run high grey level emphasis, low grey level run emphasis, run percentage, and short run emphasis= for GLRLM as well as difference entropy, dissimilarity, homogeneity 1, homogeneity 2, and inverse difference normalised for GLCM.

Comparing RF values from R1 2D ROIs with those from the circular ROIs, a lower discrepancy between R1 and R2 ROIs was observed in most cases (Table 5). In particular, taking account of the MRC, inter-reader variability was equal or preponderant for 30/32 (94%) RFs.

Discussion

In this study, the impact of inter-reader contouring variability on texture analysis of CRC liver metastases was assessed comparing the 3D and 2D ROIs of 70 lesions from 17 patients and the respectively extracted RFs.

The segmentation process of liver metastases is a challenging task due to the site and the vague boundaries of

Table 1 Demographical data and number of analysed metastases for each patient enrolled in the study

Patient	Age at CT (years)	Sex	Primary cancer site	Number of analysed metastases	Lines of treatment	Chemotherapy regimens
1	77	F	Rectum	1	2	FOLFOX + bevacizumab FOLFIRI + cetuximab
2	66	M	Colon (left)	5	2	FOLFOX FOLFIRI + cetuximab
3	62	M	Colon (left)	3	4	FOLFIRI + cetuximab Regorafenib Trifluridine/tipiracil Capecitabine
4	59	M	Colon (left)	1	2	FOLFOX FOLFIRI + cetuximab
5	56	M	Rectum	1	2	FOLFIRI + cetuximab Not available
6	40	M	Colon (left)	2	2	FOLFOX + panitumumab FOLFIRI + bevacizumab
7	61	M	Colon (left)	4	2	FOLFOX + panitumumab FOLFIRI + aflibercept
8	56	M	Colon (right)	6	2	FOLFOX + panitumumab FOLFIRI + aflibercept
9	47	M	Colon (left)	5	2	FOLFOX + cetuximab FOLFIRI + bevacizumab
10	32	M	Colon (left)	7	3	FOLFOXIRI + bevacizumab FOLFIRI + aflibercept Panitumumab
11	66	M	Colon (left)	6	2	XELIRI FOLFOX + bevacizumab
12	63	F	Colon (left)	8	2	FOLFOX + panitumumab FOLFIRI + bevacizumab
13	61	F	Colon (left)	1	2	XELOX + bevacizumab FOLFIRI + cetuximab
14	52	M	Rectum	2	2	FOLFOX FOLFOX + bevacizumab
15	41	M	Colon (left)	5	1	FOLFOX + panitumumab
16	59	M	Rectum	4	1	FOLFIRI + bevacizumab
17	60	M	Colon (left)	9	3	FOLFOX + cetuximab FOLFIRI + bevacizumab FOLFIRI

Previous drug regimens are reported in chronological order of administration. All patients had histological-confirmed adenocarcinoma of the colon/rectum with metastatic liver disease not amenable to salvage surgery. In all cases, the primary tumour was KRAS (Kirsten rat sarcoma) wild-type and HER2 (human epidermal growth factor 2) positive

F Female, M Male, FOLFIRI Leucovorin + fluorouracil + irinotecan, FOLFOX Leucovorin + fluorouracil + oxaliplatin, FOLFOXIRI Leucovorin + fluorouracil + oxaliplatin + irinotecan, XELIRI Capecitabine + irinotecan, XELOX Capecitabine + oxaliplatin

the lesions. However, we obtained satisfactory mean DC values, consistent with similar studies [19, 42, 43]. Also, as suggested by the weak correlation between the similarity indices and the lesion size, the influence of the latter on segmentation variability seemed limited.

In general, the inter-reader contouring agreement was significantly better for 2D ROIs rather than 3D ROIs. As far as the latter set, considering that HD is more sensitive to ROI shape variation than DC [26], pairs of segmentations with high values for both the similarity indices were more common. Indeed, in 3D volume segmentation, the more peripheral slices along the *z*-axis

containing the lesion suffer more for partial volume effect and the impact of all the sources of variability is greater [44, 45]. The median values of the two similarity indices and the correlation found between them for the 2D ROIs corroborated this finding.

The improvement in contouring agreement observed for 2D ROIs predictably corresponded to a reduction of inter-reader discrepancy for the majority of the RFs, although as small as the number of RFs robust to inter-reader variability was similar in the 3D and 2D sets. The robustness of these RFs was confirmed also by the ICC, so that there was correspondence between RFs with low

Table 2 Acquisition and reconstruction parameters extracted from the header DICOM of the computed tomography scans, patient by patient

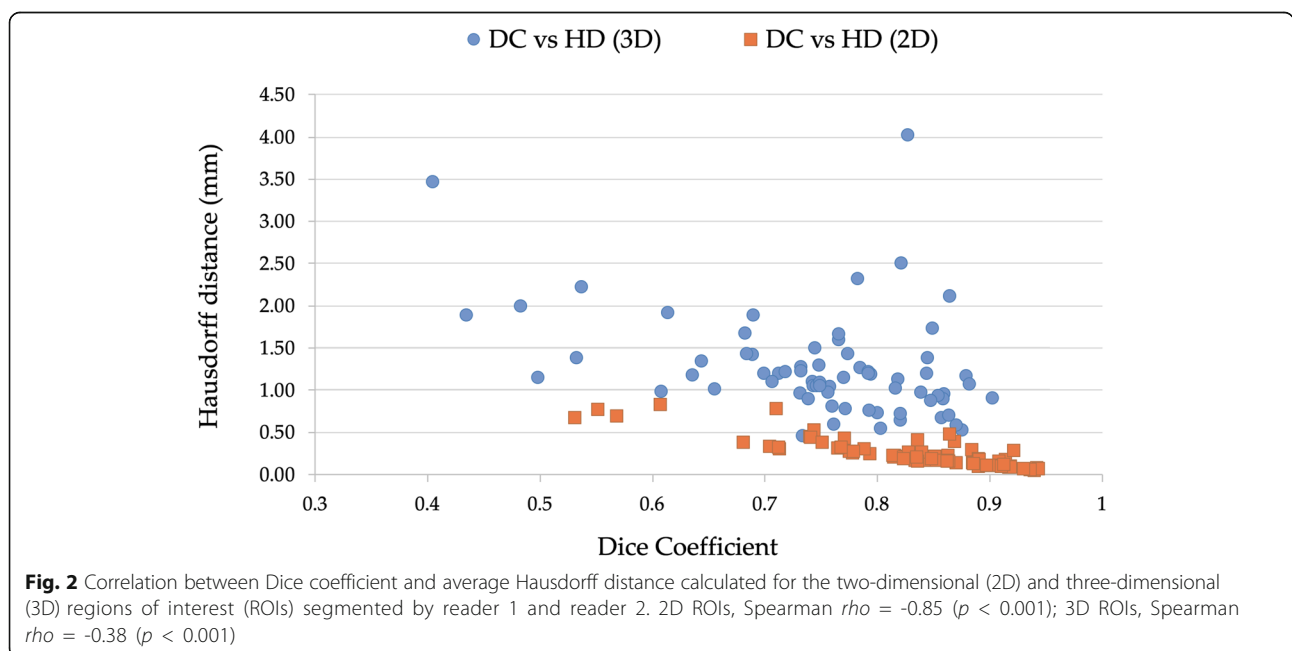
Patient	Manufacturer	Model	Slice thickness (mm)	Increment (mm)	Pixel size (mm ²)	kVp	Kernel
1	Siemens	Sensation 64	3	3	0.8242 × 0.8242	120	B30f
2	Siemens	Somatom Definition	3	3	0.7031 × 0.7031	120	B30f
3	Philips	Brilliance 64	3	3	0.8730 × 0.8730	100	B
4	Siemens	Sensation 64	3	3	0.7812 × 0.7812	120	B30f
5	Siemens	Sensation 64	3	3	0.8750 × 0.8750	120	B30f
6	Toshiba	Aquilion	3	3	0.7210 × 0.7210	120	FC13
7	Siemens	Sensation 64	3	3	0.7852 × 0.7852	120	B30f
8	Siemens	Sensation 64	3	3	0.8047 × 0.8047	120	B30f
9	Siemens	Somatom Definition	3	3	0.7773 × 0.7773	100	B30f
10	Hitachi	Eclos	2.5	2.5	0.7100 × 0.7100	120	32
11	Hitachi	Eclos	2.5	2.5	0.8410 × 0.8410	120	32
12	Siemens	Somatom Definition	3	2	0.6875 × 0.6875	100	I30f/3
13	GE	Optima CT520 Series	2.5	2.5	0.8477 × 0.8477	120	Standard
14	GE	LightSpeed Pro 32	0.625	0.625	0.8926 × 0.8926	120	Standard
15	Siemens	Somatom Definition	3	2	0.6328 × 0.6328	100	I30f/3
16	Siemens	Somatom Definition	3	2	0.7969 × 0.7969	100	I30f/3
17	Siemens	Somatom Definition	3	2.5	0.7344 × 0.7344	120	I30f/3

All images had a matrix size of 512 × 512 and were acquired 70–80 s after contrast injection with an automatic exposure control system

inter-reader variability and RFs with a good or excellent ICC.

Analysing the RFs with the greatest instability, it is reasonable to believe that mathematical issues, like the high exponents (e.g., power 3 or 4) in the formula of the

“cluster” features, contribute to amplify the differences in the ROIs. On the other hand, the RFs most influenced by contouring variability may also be the most sensitive ones to texture variation, i.e., those with the best capability to capture the information within the CT images of CRC



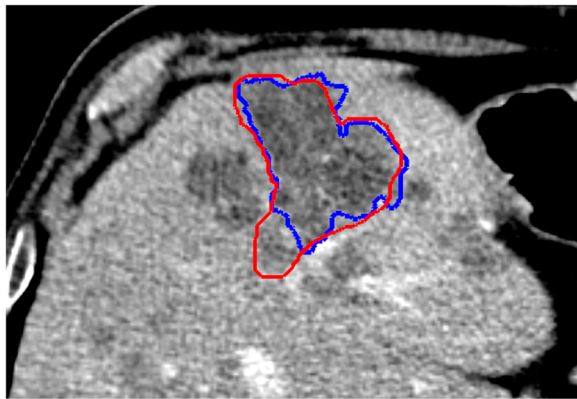


Fig. 3 Example of discrepancy between similarity indices (patient number 7): Dice coefficient was 0.86 (median two-dimensional, 0.85), whilst average Hausdorff distance was 0.48 mm (median two-dimensional, 0.21 mm). The regions of interest (blue and red lines) were approximately overlapping, but the readers differently interpreted the nature of a hypodense area adjacent to the metastasis

liver metastases, and thus conceivably, the RFs with the best potential predictive value. For example, Simpson et al. [11] found that “contrast, correlation and homogeneity” were associated with hepatic disease-free survival in patients with CRC liver metastases. In the current analysis, the first two RFs showed a mild-to-high inter-reader variability, which is consistent with a greater sensitivity to texture variation.

These aspects must be considered when choosing the RFs to create radiomics predictive models since the “noise” related to inter-reader variability could eclipse meaningful information in the texture of CRC liver metastases, but the selection of only very robust RFs may be inadequate to detect differences in the image texture as well.

The ideal solution to eliminate the interference of inter-reader variability would be to dispose of semiautomatic or, preferably, automatic methods for the segmentation of liver metastases [15, 26]. However, the tools currently available are not yet reliable enough, as shown by testing 24 valid state-of-the-art liver tumour segmentation algorithms [43], so that operator input remains indispensable [46].

Interestingly, as shown by the comparison between standard ROIs and circular ROIs, when one of the readers drew simple geometric ROIs, less tailored on the lesion boundaries, the discrepancy in RFs values were lower or comparable to that relative to the other reader. This suggests that in the multicentric setting inter-reader variability may be handled in two ways: involving a large number of readers, so as to allow the selection of robust RFs according to individual reproducibility (e.g., including RFs with ICC > 0.90 in final models) [47]; or with a “centralised” approach based on few readers to minimise variability. In the second case, a simplified segmentation protocol to accelerate the contouring task could be followed, as it would introduce a variability at most equivalent to that determined by multiple readers.

However, such analysis was limited to the 2D ROIs due to the complexity of applying it to the 3D ones, so it should be verified with larger samples. A viable compromise between assessing the lesion in its entirety and limiting the inter-reader disagreement could be to exclude from the segmentation the most peripheral slices along the Z-axis of the metastasis. Alternatively, clinical radiomic-based models could mix RFs extracted from 3D and 2D ROIs on the basis of their dependency on inter-reader variability, provided that the selection and extraction of the 2D ROIs may require additional work unless implementing automatic processes.

These methods are worthy of future investigation, considering that the main limitation of our study is not being able to assess how the improvement of RFs stability against contouring variability impact on the predictive performance due to cohort size. Indeed, only few patients were assessed, but each metastasis was singularly considered, so that the number of lesions analysed was consistent with similar works. Another limitation is that the impact of the acquisition/reconstruction settings of CT scans was not considered. The heterogeneity of scanning equipment and protocols, due to the time span and referral of patients from different institutions, could have reduced the congruency of the segmentation, but this rather strengthens the results about the textural features found to be stable. Also, two different contouring softwares were used, although eventual differences hence derived can be considered incorporable in the concept

Table 3 Correlation results (Spearman’s ρ coefficients) between similarity indices (Dice coefficient and average Hausdorff distance) and size parameters of the segmented metastases for both 2D and 3D ROIs

Correlation	2D ROI		3D ROI	
	DC	HD	DC	HD
ROI manual axial diameter	0.42 ($p < 0.001$)	-0.04 ($p = 0.760$)	0.45 ($p < 0.001$)	0.45 ($p < 0.001$)
ROI maximum 3D diameter	0.37 ($p < 0.001$)	-0.12 ($p = 0.382$)	0.42 ($p < 0.001$)	0.41 ($p = 0.002$)
ROI volume/area	0.30 ($p < 0.001$)	-0.17 ($p = 0.209$)	0.36 ($p < 0.001$)	0.30 ($p = 0.024$)

2D Two dimensional, 3D Three dimensional, DC Dice coefficient, HD Hausdorff distance (average), ROI Region of interest

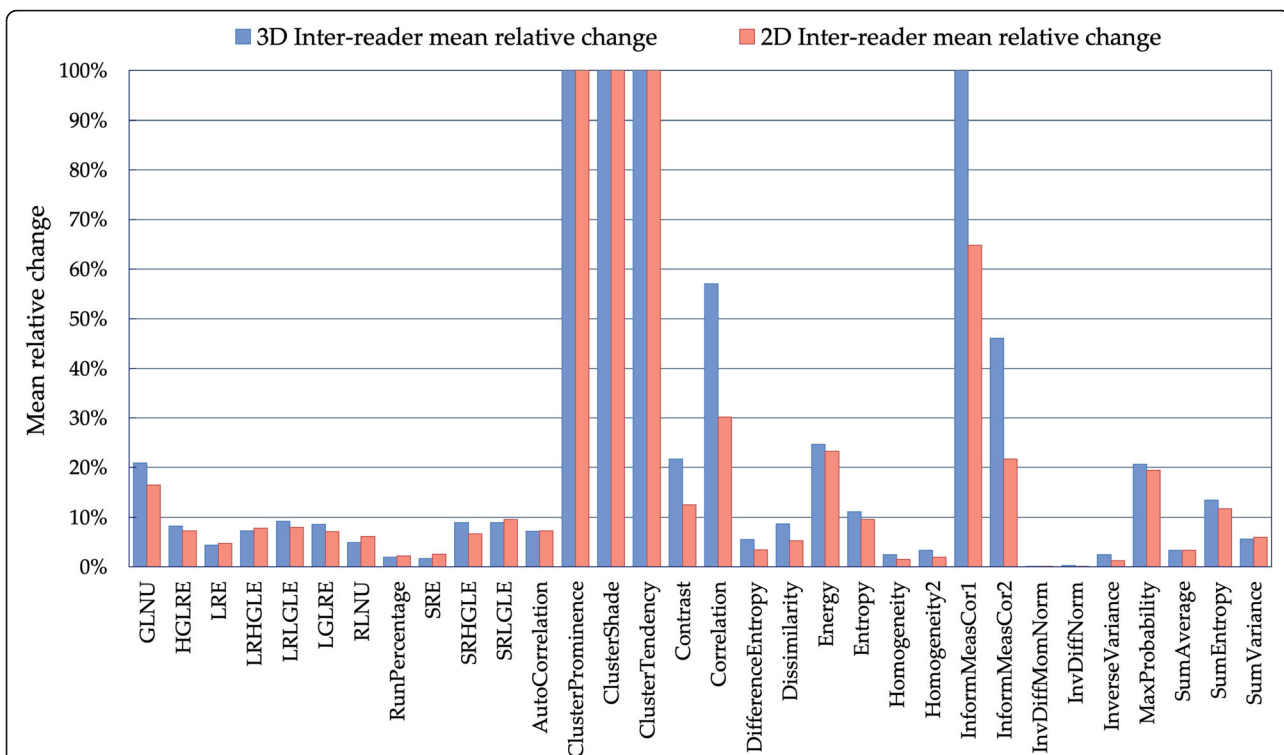


Fig. 4 Means of relative changes between the RFs extracted from each lesion ($n = 70$) contoured by the two readers. The results from two-dimensional and three-dimensional segmentations were compared. Out of scale values have been truncated. The cluster features showed the greatest instability between readers. *GLNU* Grey level non-uniformity, *HGLRE* High grey level run emphasis, *LGLRE* Low grey level run emphasis, *LRE* Long run emphasis, *LRHGLE* Long run high grey level emphasis, *LRLGLE* Long run low grey level emphasis, *RLNU* Run length non-uniformity, *SRE* Short run emphasis, *SRHGLE* Short run high grey level emphasis, *SRLGLE* Short run low grey level emphasis

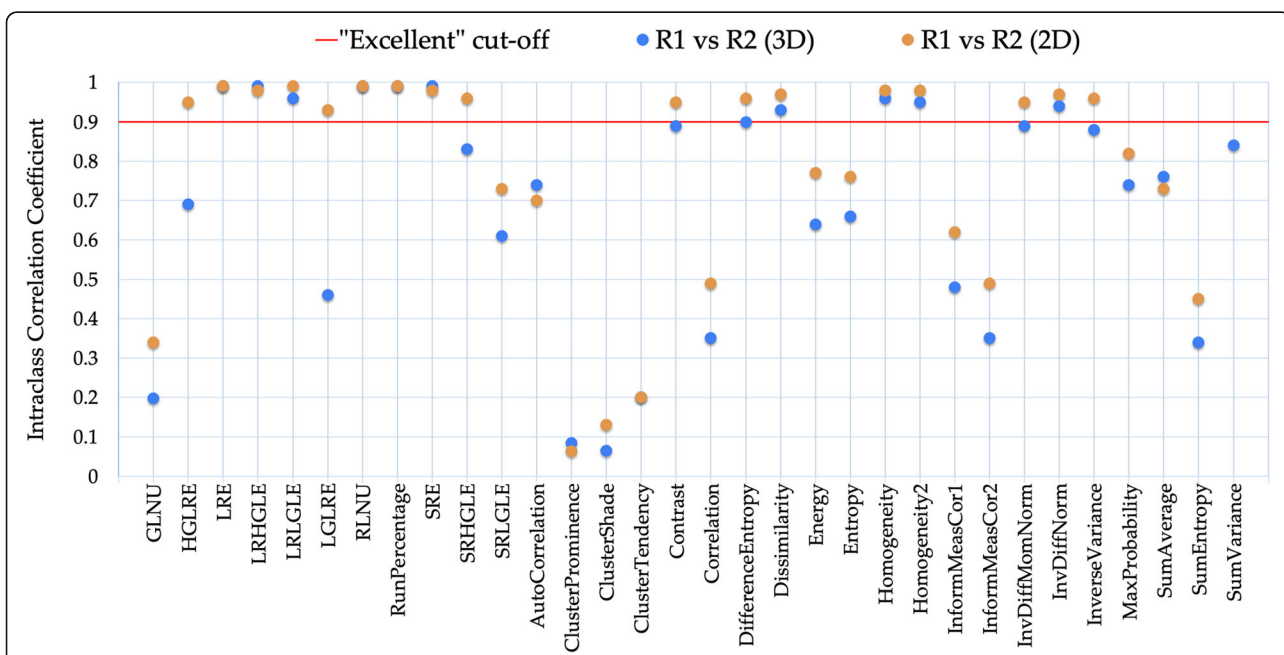


Fig. 5 For all radiomic features, the intraclass correlation coefficients (ICC) of inter-reader variability are plotted and compared between three-dimensional and two-dimensional segmentations. "Excellent" ICC cutoff is shown as a red line. *R1* Reader 1, *R2* Reader 2, *GLNU* Grey level non-uniformity, *HGLRE* High grey level run emphasis, *LGLRE* Low grey level run emphasis, *LRE* Long run emphasis, *LRHGLE* Long run high grey level emphasis, *LRLGLE* Long run low grey level emphasis, *RLNU* Run length non-uniformity, *SRE* Short run emphasis, *SRHGLE* Short run high grey level emphasis, *SRLGLE* Short run low grey level emphasis

Table 4 Mean relative changes and intraclass correlation coefficients are reported for all the textural features and both the 2D and 3D ROI sets

		2D ROI		3D ROI		
		Relative change	ICC	Relative change	ICC	
GLRLM	GLNU	16%	0.34	21%	0.20	
	HGLRE	7%	0.95	8%	0.69	
	LRE	5%	0.99	4%	0.99	
	LRHGLE	8%	0.98	7%	0.99	
	LRLGLE	8%	0.99	9%	0.96	
	LGLRE	7%	0.93	9%	0.46	
	RLNU	6%	0.99	5%	0.99	
	Run percentage	2%	0.99	2%	0.99	
	SRE	3%	0.98	2%	0.99	
	SRHGLE	7%	0.96	9%	0.83	
	SRLGLE	10%	0.73	9%	0.61	
	GLCM	AutoCorrelation	7%	0.70	7%	0.74
		Cluster prominence	418%	0.06	641%	0.08
ClusterShade		752%	0.13	1567%	0.06	
Cluster tendency		100%	0.20	116%	0.20	
Contrast		13%	0.95	22%	0.89	
Correlation		30%	0.49	57%	0.35	
Difference entropy		3%	0.96	6%	0.90	
Dissimilarity		5%	0.97	9%	0.93	
Energy		23%	0.77	25%	0.64	
Entropy		10%	0.76	11%	0.66	
Homogeneity		1%	0.98	2%	0.96	
Homogeneity2		2%	0.98	3%	0.95	
InformationMeasureCorrel1		65%	0.62	136%	0.48	
InformationMeasureCorrel2		22%	0.49	46%	0.35	
InverseDifferMomentNormal		0%	0.95	0%	0.89	
InverseDifferNormal		0%	0.97	0%	0.94	
InverseVariance		1%	0.96	2%	0.88	
MaxProbability		19%	0.82	21%	0.74	
SumAverage		3%	0.73	3%	0.76	
SumEntropy		12%	0.45	14%	0.34	
SumVariance	6%	0.80	6%	0.84		

Bold text is used for features found robust against inter-reader variability (ICC > 0.90 and mean relative change < 10%)

2D Two dimensional, 3D Three dimensional, GLCM Grey level co-occurrence matrix, GLNU Grey level non-uniformity, GLRLM Grey level run length matrix, HGLRE High grey level run emphasis, ICC Intraclass correlation coefficient, LGLRE Low grey level run emphasis, LRE Long run emphasis, LRHGLE Long run high grey level emphasis, LRLGLE Long run low grey level emphasis, RLNU Run length non-uniformity, SRE Short run emphasis, SRHGLE Short run high grey level emphasis, SRLGLE Short run low grey level emphasis.

of inter-reader variability itself and, in general, it better replicated a likely situation in multicentric settings. Finally, the study focused only on the second-order features.

In conclusion, the current study highlighted the possibility to extract textural RFs robust against contouring variability from CRC liver metastases. This is essential to

translate radiomics into clinical practice since the creation of large labelled imaging datasets will necessarily require the involvement of multiple readers. For the most stable RFs, both 3D and 2D segmentations were reliable, but a 2D approach, which is more pragmatic and less time-consuming, could mitigate inter-reader contouring variability. This may expand the choice of RFs

Table 5 Comparison of the radiomic features obtained using manual 2D ROIs by R1 versus R2 and, for R1, using manual versus circular 2D ROIs

GLRLM	R1 vs R2	R1 versus extROIs	R1 versus intROIs	GLCM	R1 versus R2	R1 versus extROIs	R1 versus intROIs
GLNU	17%	-9%	+9%	AutoCorrelation	7%	+3%	-4%
HGLRE	7%	+3%	-5%	Cluster Prominence	503%	+169%	-29%
LRE	3%	0%	+2%	ClusterShade	948%	+192%	-142%
LRHGLE	6%	+3%	-2%	ClusterTendency	86%	+53%	-16%
LRLGLE	6%	-2%	+6%	Contrast	15%	+3%	-3%
LGLRE	5%	-2%	+5%	Correlation	35%	+20%	-15%
RLNU	2%	-1%	-3%	DiffEntropy	4%	+1%	-1%
Run Percentage	1%	0%	-1%	Dissimilarity	6%	+1%	-2%
SRE	1%	-1%	-1%	Energy	23%	-13%	+10%
SRHGLE	6%	+2%	-7%	Entropy	9%	+5%	-3%
SRLGLE	5%	-2%	+5%	Homogeneity	1%	0%	+1%
				Homogeneity2	2%	0%	+1%
				InfMeasCor1	75%	+45%	-11%
				InfMeasCor2	25%	+16%	-6%
				InvDiff MomNorm	0%	0%	0%
				InvDiffNorm	0%	0%	0%
				InverseVariance	1%	0%	-1%
				MaxProbability	19%	-10%	+6%
				SumAverage	3%	+1%	-2%
				SumEntropy	11%	+7%	-4%
				SumVariance	6%	+2%	-4%

For the main lesion of each patient, the RFs from R1 2D ROIs were compared to R2 2D ROIs and to the circular 2D ROIs (extROIs and intROIs). Mean relative discrepancy taking R1 values as reference is reported. Inter-reader variability was preponderant for nearly all RFs. As expected, RFs from extROIs and intROIs had a divergent behaviour in respect of R1 2D ROIs, whose characteristics were intermediate. The colour code refers to the absolute value of discrepancy

2D Two-dimensional, extROI Smallest circular segmentation including the whole lesion, GLCM Grey level co-occurrence matrix, GLNU Grey level non-uniformity, GLRLM Grey level run length matrix, HGLRE High grey level run emphasis, intROI Largest circular segmentation completely inside the lesion, LGLRE Low grey level run emphasis, LRE Long run emphasis, LRHGLE Long run high grey level emphasis, LRLGLE Long run low grey level emphasis, R1 Reader 1, R2 Reader 2, RFs Radiomic features, RLNU Run length non-uniformity, ROIs Region of interest, SRE Short run emphasis, SRHGLE Short run high grey level emphasis, SRLGLE Short run low grey level emphasis

suitable for building clinical models, but further studies evaluating the relationship between segmentation strategy and outcome predictivity are warranted, so as to optimise the extraction of meaningful information from the CT texture of CRC liver metastases.

Abbreviations

CRC: Colorectal cancer; DC: Dice coefficient; GLCM: Grey level co-occurrence matrix; GLRLM: Grey level run length matrix; HD: Hausdorff distance; IBEX: Imaging Biomarker Explorer; IBSI: Image Biomarker Standardisation Initiative; ICC: Intraclass correlation coefficient; IQR: Interquartile range; MRC: Mean relative change; RFs: Radiomic features; ROI: Region of interest

Authors' contributions

FR: investigation, data curation, formal analysis, writing—original draft preparation; FC and SM: data curation, writing—review and editing; CDM, AD, and VG: methodology, software, writing—review and editing; LV: investigation, data curation; ASB and SS: clinical project administration, patient selection, and care; SM, SS, DR, AT, and AV: conceptualisation, resources, supervision. The authors read and approved the final manuscript.

Funding

This work was supported by Fondazione Oncologia Niguarda Onlus and Fondazione AIRC under 5 per Mille 2018 – ID project 21091 (GLs: Bardelli A., Marsoni S., Regge D., Siena S.; HERACLES clinical trial GLs: Siena S., Marsoni S.).

Availability of data and materials

The datasets generated and/or analysed during the current study are not publicly available because of the terms of the research participant consent but are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The study was approved by the Institution Review Board (Ethics Committee of Milan Area C; decision number: 181_05/2012), and all the patients signed written informed consent to allow the images of their diagnostic examinations to be used for scientific purposes at the time of enrolment in the study.

Consent for publication

Not applicable

Competing interests

The authors declare that they have no competing interests

Author details

¹Department of Radiology, ASST Grande Ospedale Metropolitano Niguarda, Piazza Ospedale Maggiore 3, 20162 Milan, Italy. ²Department of Medical Physics, ASST Grande Ospedale Metropolitano Niguarda, Piazza Ospedale Maggiore 3, 20162 Milan, Italy. ³Department of Surgical Sciences, University of Turin, via Verdi 8, 10124 Turin, Italy. ⁴Radiology Unit, Candiolo Cancer Institute, FPO-IRCCS, Strada Provinciale 142 km 3.95, 10060 Candiolo, Turin, Italy. ⁵Radiology Unit, SS Annunziata Hospital ASL CN1 Cuneo, via Ospedali 14, 12038 Cuneo, Savigliano, Italy. ⁶Niguarda Cancer Center, Grande Ospedale Metropolitano Niguarda, Piazza Ospedale Maggiore 3, 20162 Milan, Italy. ⁷Department of Oncology and Hemato-Oncology, Università degli Studi di Milano, via Festa del Perdono 7, 20122 Milan, Italy. ⁸Precision Oncology, IFOM – The FIRCC Institute of Molecular Oncology, via Adamello 16, 20139 Milan, Italy. ⁹Department of Physics, Università degli Studi di Milano, via Giovanni Celoria 16, 20133 Milan, Italy.

Received: 22 June 2020 Accepted: 13 October 2020

Published online: 10 November 2020

References

- García-Figueiras R, Baleato-González S, Padhani AR et al (2019) How clinical imaging can assess cancer biology. *Insights Imaging* 10:28 <https://doi.org/10.1186/s13244-019-0703-0>
- Hassani C, Varghese BA, Nieva J, Duddalwar V (2019) Radiomics in pulmonary lesion imaging. *AJR Am J Roentgenol* 212:497–504 <https://doi.org/10.2214/AJR.18.20623>
- Ferlay J, Ervik M, Lam F, et al (2019) Global Cancer Observatory: cancer today. <http://gco.iarc.fr/today/fact-sheets-cancers>.
- Vera R, González-Flores E, Rubio C et al (2020) Multidisciplinary management of liver metastases in patients with colorectal cancer: a consensus of SEOM, AEC, SEOR, SERVEI, and SEMNIM. *Clin Transl Oncol* 22: 647–662 <https://doi.org/10.1007/s12094-019-02182-z>
- Van Cutsem E, Cervantes A, Adam R et al (2016) ESMO consensus guidelines for the management of patients with metastatic colorectal cancer. *Ann Oncol* 27:1386–1422 <https://doi.org/10.1093/annonc/mdw235>
- Gillies RJ, Kinahan PE, Hricak H (2016) Radiomics: images are more than pictures, they are data. *Radiology* 278:563–577 <https://doi.org/10.1148/radiol.2015151169>
- Lubner MG, Stabo N, Lubner SJ et al (2015) CT textural analysis of hepatic metastatic colorectal cancer: pre-treatment tumor heterogeneity correlates with pathology and clinical outcomes. *Abdom Imaging* 40:2331–2337 <https://doi.org/10.1007/s00261-015-0438-4>
- Beckers RCJ, Trebeschi S, Maas M et al (2018) CT texture analysis in colorectal liver metastases and the surrounding liver parenchyma and its potential as an imaging biomarker of disease aggressiveness, response and survival. *Eur J Radiol* 102:15–21 <https://doi.org/10.1016/j.ejrad.2018.02.031>
- Dohan A, Gallix B, Guiu B et al (2020) Early evaluation using a radiomic signature of unresectable hepatic metastases to predict outcome in patients with colorectal cancer treated with FOLFIRI and bevacizumab. *Gut* 69:531–539 <https://doi.org/10.1136/gutjnl-2018-316407>
- Rao S-X, Lambregts DMJ, Schnerr RS et al (2016) CT texture analysis in colorectal liver metastases: a better way than size and volume measurements to assess response to chemotherapy? *United Eur Gastroenterol J* 4:257–263 <https://doi.org/10.1177/2050640615601603>
- Simpson AL, Doussot A, Creasy JM et al (2017) Computed tomography image texture: a noninvasive prognostic marker of hepatic recurrence after hepatectomy for metastatic colorectal cancer. *Ann Surg Oncol* 24:2482–2490 <https://doi.org/10.1245/s10434-017-5896-1>
- Lambin P, Leijenaar RTH, Deist TM et al (2017) Radiomics: the bridge between medical imaging and personalized medicine. *Nat Rev Clin Oncol* 14:749–762 <https://doi.org/10.1038/nrclinonc.2017.141>
- Fave X, Zhang L, Yang J et al (2016) Impact of image preprocessing on the volume dependence and prognostic potential of radiomics features in non-small cell lung cancer. *Transl Cancer Res* 5:349–363 <https://doi.org/10.21037/tcr.2016.07.11>
- Traverso A, Wee L, Dekker A, Gillies R (2018) Repeatability and reproducibility of radiomic features: a systematic review. *Int J Radiat Oncol Biol Phys* 102:1143–1158 <https://doi.org/10.1016/j.ijrobp.2018.05.053>
- Avanzo M, Stancanella J, El Naqa I (2017) Beyond imaging: the promise of radiomics. *Phys Med* 38:122–139 <https://doi.org/10.1016/j.jemp.2017.05.071>
- Liu Z, Wang S, Dong D et al (2019) The applications of radiomics in precision diagnosis and treatment of oncology: opportunities and challenges. *Theranostics* 9:1303–1322 <https://doi.org/10.7150/thno.30309>
- Morin O, Vallières M, Jochems A et al (2018) A deep look into the future of quantitative imaging in oncology: a statement of working principles and proposal for change. *Int J Radiat Oncol Biol Phys* 102:1074–1082 <https://doi.org/10.1016/j.ijrobp.2018.08.032>
- Kocak B, Durmaz ES, Kaya OK, Ates E, Kilickesmez O (2019) Reliability of single-slice-based 2D CT texture analysis of renal masses: influence of intra- and interobserver manual segmentation variability on radiomic feature reproducibility. *AJR Am J Roentgenol* 213:377–383 <https://doi.org/10.2214/AJR.19.21212>
- Pavic M, Bogowicz M, Würms X et al (2018) Influence of inter-observer delineation variability on radiomics stability in different tumor sites. *Acta Oncol* 57:1070–1074 <https://doi.org/10.1080/0284186X.2018.1445283>
- Zhao B, Tan Y, Bell DJ et al (2013) Exploring intra- and inter-reader variability in uni-dimensional, bi-dimensional, and volumetric measurements of solid tumors on CT scans reconstructed at different slice intervals. *Eur J Radiol*. <https://doi.org/10.1016/j.ejrad.2013.02.018>
- Shen C, Liu Z, Guan M et al (2017) 2D and 3D CT radiomics features prognostic performance comparison in non-small cell lung cancer. *Transl Oncol* 10:886–894 <https://doi.org/10.1016/j.tranon.2017.08.007>
- Zhao B, Tan Y, Tsai WY et al (2016) Reproducibility of radiomics for deciphering tumor phenotype with imaging. *Sci Rep* 6:23428 <https://doi.org/10.1038/srep23428>
- Ng F, Kozarski R, Ganeshan B, Goh V (2013) Assessment of tumor heterogeneity by CT texture analysis: can the largest cross-sectional area be

- used as an alternative to whole tumor analysis? *Eur J Radiol* 82:342–348 <https://doi.org/10.1016/j.ejrad.2012.10.023>
24. Qiu Q, Duan J, Duan Z et al (2019) Reproducibility and non-redundancy of radiomic features extracted from arterial phase CT scans in hepatocellular carcinoma patients: impact of tumor segmentation variability. *Quant Imaging Med Surg* 9:453–464 <https://doi.org/10.21037/qims.2019.03.02>
 25. Yamashita R, Perrin T, Chakraborty J et al (2020) Radiomic feature reproducibility in contrast-enhanced CT of the pancreas is affected by variabilities in scan parameters and manual segmentation. *Eur Radiol* 30: 195–205 <https://doi.org/10.1007/s00330-019-06381-8>
 26. Owens CA, Peterson CB, Tang C et al (2018) Lung tumor segmentation methods: impact on the uncertainty of radiomics features for non-small cell lung cancer. *PLoS One* 13:e0205003 <https://doi.org/10.1371/journal.pone.0205003>
 27. Liu R, Elhalawani H, Radwan Mohamed AS et al (2020) Stability analysis of CT radiomic features with respect to segmentation variation in oropharyngeal cancer. *Clin Transl Radiat Oncol* 21:11–18 <https://doi.org/10.1016/j.ctro.2019.11.005>
 28. (2017) Evaluation of trastuzumab in combination with lapatinib or pertuzumab in combination with trastuzumab-emtansine to treat patients with HER2-positive metastatic colorectal cancer (HERACLES). <https://clinicaltrials.gov/ct2/show/record/NCT03225937>
 29. Sartore-Bianchi A, Martino C, Lonardi S et al (2019) LBA35Phase II study of pertuzumab and trastuzumab-emtansine (T-DM1) in patients with HER2-positive metastatic colorectal cancer: the HERACLES-B (HER2 amplification for colo-rectal cancer enhanced stratification, cohort B) trial. *Ann Oncol*. <https://doi.org/10.1093/annonc/mdz394.024>
 30. Lincke T, Zech CJ (2017) Liver metastases: detection and staging. *Eur J Radiol* 97:76–82 <https://doi.org/10.1016/j.ejrad.2017.10.016>
 31. Court LE, Fave X, Mackin D, Lee J, Yang J, Zhang L (2016) Computational resources for radiomics. *Transl Cancer Res* 5:340–348 <https://doi.org/10.21037/tcr.2016.06.17>
 32. Perrin T, Midya A, Yamashita R et al (2018) Short-term reproducibility of radiomic features in liver parenchyma and liver malignancies on contrast-enhanced CT imaging. *Abdom Radiol (NY)* 43:3271–3278 <https://doi.org/10.1007/s00261-018-1600-6>
 33. Zhang L, Fried DV, Fave XJ, Hunter LA, Yang J, Court LE (2015) Ibox: an open infrastructure software platform to facilitate collaborative work in radiomics. *Med Phys*. <https://doi.org/10.1118/1.4908210>
 34. Defeudis A, De Mattia C, Rizzetto F et al (2020) Standardization of CT radiomics features for multi-center analysis: impact of software settings and parameters. *Phys Med Biol* 65:195012 <https://doi.org/10.1088/1361-6560/ab9f61>
 35. van Rossum PSN, Xu C, Fried DV, Goense L, Court LE, Lin SH (2016) The emerging field of radiomics in esophageal cancer: current evidence and future potential. *Transl Cancer Res* 5:410–423 <https://doi.org/10.21037/tcr.2016.06.19>
 36. Pinter C, Lasso A, Wang A, Jaffray D, Fichtinger G (2012) SlicerRT: radiation therapy research toolkit for 3D Slicer. *Med Phys* 39:6332–6338 <https://doi.org/10.1118/1.4754659>
 37. Bogoya J, Vargas A, Cuate O, Schütze O (2018) A (p,q)-averaged Hausdorff distance for arbitrary measurable sets. *Math Comput Appl* 23:51. <https://doi.org/10.3390/mca23030051>
 38. Dice LR (1945) Measures of the amount of ecologic association between species. *Ecology* 26:297–302 <https://doi.org/10.2307/1932409>
 39. Evans JD (1996) *Straightforward statistics for the behavioral sciences*. Thomson Brooks/Cole Publishing Co, Belmont, CA, US
 40. Zwanenburg A, Leger S, Agolli L et al (2019) Assessing robustness of radiomic features by image perturbation. *Sci Rep* 9:614 <https://doi.org/10.1038/s41598-018-36938-4>
 41. Koo TK, Li MY (2016) A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *J Chiropr Med* 15:155–163 <https://doi.org/10.1016/j.jcm.2016.02.012>
 42. Moghbel M, Mashohor S, Mahmud R, Saripan MI Bin (2016) Automatic liver tumor segmentation on computed tomography for patient treatment planning and monitoring. *EXCLI J* 15:406–423. <https://doi.org/10.17179/excli2016-402>
 43. Bilic P, Christ PF, Vorontsov E, et al (2019) The liver tumor segmentation benchmark (LITS). 1–43 arXiv:1901.04056. Available via <https://arxiv.org/abs/1901.04056>
 44. Bellon MR, Siddiqui MS, Ryu S, Chetty IJ (2014) The effect of longitudinal CT resolution and pixel size (FOV) on target delineation and treatment planning in stereotactic radiosurgery. *J Radiosurg SBRT* 3:149–163
 45. Moskowitz CS, Hricak H, Zabor EC et al (2013) Intra- and interobserver variability in CT measurements in oncology. *Radiology* 269:451–459 <https://doi.org/10.1148/radiol.13122665>
 46. Kumar V, Gu Y, Basu S et al (2012) Radiomics: the process and the challenges. *Magn Reson Imaging* 30:1234–1248 <https://doi.org/10.1016/j.mri.2012.06.010>
 47. Zwanenburg A, Leger S, Agolli L et al (2019) Assessing robustness of radiomic features by image perturbation. *Sci Rep* 9:1–10 <https://doi.org/10.1038/s41598-018-36938-4>

Publisher's Note

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

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)
