Radiomics predicts response of individual HER2-amplified colorectal cancer liver metastases in patients treated with HER2-targeted therapy

Authors' names:

Valentina Giannini[1,2], Samanta Rosati[3], Arianna Defeudis[1,2], Gabriella Balestra[3], Lorenzo Vassallo[1], Giovanni Cappello[1], Simone Mazzetti[1,2], Cristina De Mattia[4], Francesco Rizzetto[5], Alberto Torresin[4,6], Andrea Sartore-Bianchi[7,8], Salvatore Siena[7,8], Angelo Vanzulli[5,7], Francesco Leone[9,10], Vittorina Zagonel[11], Silvia Marsoni[12], Daniele Regge[1,2].

Authors' affiliations:

[1]Department of Radiology, Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Italy.

[2]Department of Surgical Sciences, University of Turin, Turin, Italy.

[3]Department of Electronics and Telecommunications, Polytechnic of Turin, Turin, Italy.

[4]Department of Medical Physics, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy.

[5]Department of Radiology, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy.

[6]Department of Physics, University of Milan, Milan, Italy

[7]Department of Oncology and Hemato-Oncology, University of Milan, Milan, Italy.

[8]Niguarda Cancer Center, Grande Ospedale Metropolitano Niguarda, Milan, Italy.

[9]Medical Oncology, Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Italy.

[10]Department of Oncology, University of Turin, Turin, Italy.

[11]Medical Oncology Unit 1, Istituto Oncologico Veneto-IRCCS, Padova, Italy

[12]Precision Oncology, IFOM-The FIRC Institute of Molecular Oncology, Milan, Italy.

Corresponding author:

Valentina Giannini

Department of Radiology, Candiolo Cancer Institute, FPO-IRCCS

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1002/ijc.33271

Strada Provinciale 142 Km 3,95

10060 Candiolo, Torino, Italy

Tel: +39 (0)11 9933327

valentina.giannini@ircc.it

Keywords: radiomics, prediction of response to therapy, machine learning, CT liver metastases, genetic algorithms.

Article category: Research article (Innovative Tools and Methods)

LIST OF ABBREVIATIONS:

CRC: Colorectal Cancer

GA: Genetic Algorithms

GLCM: Grey Level Co-occurrence Matrix

GLRLM: Grey Level Run Length Matrix

lmCRC: Liver metastasis Colorectal Cancer

NPV: Negative Predictive Value

PPV: Positive Predictive Value

R+: Responders

R-: Non responders

NOVELTY AND IMPACT: We developed a radiomics signature to predict behavior of individual metastasis to targeted treatment. The algorithm is very effective in predicting responding lesions,

proving to be able to identify patients with outlier lesions, i.e. that do not respond in a general condition were most lesions respond. The model also allows identification of non-responder lesions in patients with heterogeneous response, potentially paving the way to a more aggressive diagnostic and therapeutic approach in selected patients.

Abstract

The aim of this study was to develop and validate a machine learning algorithm to predict response of individual HER2-amplified colorectal cancer liver metastases (lmCRC) undergoing dual HER2-targeted therapy. Twenty-four radiomics features were extracted following 3D manual segmentation of 141 lmCRC on pre-treatment portal CT scans of a cohort including 38 HER2-amplified patients; feature selection was then performed using genetic algorithms. lmCRC were classified as non-responders (R-), if their largest diameter increased more than 10% at a CT scan performed following 3 months of treatment, responders (R+) otherwise. Sensitivity, specificity, negative (NPV) and positive (PPV) predictive values in correctly classifying individual lesion and overall patient response were assessed on a training dataset and then validated on a second dataset using a Gaussian naïve bayesian classifier. Per-lesion sensitivity, specificity, NPV and PPV were 89%, 85%, 93%, 78% and 90%, 42%, 73%, 71% respectively in the testing and validation datasets. Per-patient sensitivity and specificity were 92% and 86%. Heterogeneous response was observed in 9 of 38 patients (24%). Five of the 9 patients were carriers of non-responder lesions correctly classified as such by our radiomics signature, including 4 of 7 harboring only one non-responder lesion. The developed method has been proven effective in predicting behavior of individual metastases to targeted treatment in a cohort of HER2 amplified patients. The model accurately detects responder lesions and identifies non-responder lesions in patients with heterogeneous response, potentially paving the way to multimodal treatment in selected patients. Further validation will be needed to confirm our findings.

1. Introduction

Colorectal cancer (CRC) consists in a group of heterogeneous diseases exhibiting substantial genetic differences evolving in time(1). Mapping of the genetic landscape allows the identification of mutations that are amenable to targeted treatments. Our group has reported the results of the treatment with trastuzumab plus lapatinib of patients with metastatic colorectal cancer with HER2 amplification or overexpression(2). This phase II trial showed clinical benefit from treatment in 70% of patients(3). Unfortunately, as we and others have shown, patients may exhibit a heterogeneous response, some metastases shrinking while others are progressing(4–6). Kessel et al. (5) have also shown that heterogeneous response, caused by the onset of new resistant tumor clones in some lesions, is a predictor of poor overall survival(7,8) . Differentiating which colorectal cancer liver metastases (lmCRC) responds and which lingers and eventually will progress in the same patient could pave the

way to truly personalized treatment. In patients where a heterogeneous response is expected, a multimodality strategy with an interventional approach on lesions that are predicted non-responders may be biologically justified and clinically resolutive. Irregularly insensitive liver lesions in an otherwise responding patient could be biopsied to detect new molecular clones or be eventually amenable to image guided ablation, with the aim of ensuring a longer period of disease control.

Response to treatment is commonly evaluated using RECIST, which measure changes in the longest axial tumor diameters after chemotherapy(9). However, medical images can provide additional information about tumor phenotype(10). This is made possible by converting digital medical images into mineable high-dimensional data, through a process called radiomics(11). Radiomics is the study of these quantitative features and their correlation with tumor phenotypes(10), and it has been shown to be useful in predicting response to medical therapy in different tumor models(12–17). In CRC correlation has been found between entropy and skewness, computed on CT scans, and tumor grade, KRAS mutational status and risk of recurrence in post-treatment future liver remnant(18,19). However, prediction of the behavior of single lmCRC under treatment has not been explored and only one published study describes a machine learning method to predict treatment response of individual liver metastases from esophagogastric cancer, achieving an AUC of 0.87(20). However, in this small series, analyses were not performed on an independent dataset.

The main objective of this study was to develop and validate retrospectively a machine learning algorithm to predict response of individual lmCRC in patients with HER2 amplification undergoing dual target therapy.

2. Material and methods

2.1. Study design and patients

The study population was composed of patients enrolled in the HERACLES Trial (NCT03225937). All included patients had histologically confirmed RAS wild-type and HER2 positive adenocarcinoma of the colon or rectum with metastatic disease no longer responding to standard chemotherapy nor anti-EGFR targeted therapy. Patients enrolled in the trial, were treated with anti-HER2 therapy either received lapatinib + trastuzumab or pertuzumab + trastuzumab-emantansine.

Inclusion criteria, other than those listed in the HERACLES study(2), were the availability of a: 1) baseline portal-phase contrast-enhanced CT scan with at least one liver metastasis of at least 10 mm; 2) second contrast-enhanced CT performed after 3 months of antiHER2 therapy. Patients

included in the study were randomly assigned to one of two groups: the first for model construction (training set) and the second for validation of the model (validation set). The proportion of patients in each of the following subgroups, i.e., all R+ lmCRC, all R- lmCRC, and mixed response, was similar in both the training and the validation dataset.

2.2. CT scans

Exams were acquired using different CT scanners (Philips, Siemens, Hitachi, Toshiba, General Electrics) and protocols. Across the database, median slice thickness was 3 mm (range 0.625-3 mm), median slice increment 3 mm (range 0.625-3 mm), median pixel size 0.789x0.789 mm (range 0.611x0.611 – 0.977x0.977 mm), median kVp 120 (range 80-120) and matrix size was 512x512. An experienced radiologist (>20 years in reporting CT examinations) measured the largest diameter of each liver metastases at baseline and time point 1 (TP1), after 3 months of antiHER2 treatment. Lesions were classified as: responders (R+) if the largest diameter decreased more than 10% from baseline to TP1, or if they remained stable (± 10%); non-responders (R-) if the largest diameter increased more than 10% from baseline to TP1. The cut-off of 10% was chosen considering for intra-observer variability. Previous work has demonstrated that an increases/decreases higher than 10% can be considered as true, rather than measurement variation, with a 95% confidence(21). Reference standard was represented by the above reported dichotomic variable.

2.3. Features extraction

All liver metastases with a diameter 10 mm or more were manually contoured using Mipav software (https://mipav.cit.nih.gov). A resident radiologist used painting tools to segment the area of the tumor on each slice on portal phase pre-treatment CT and saved the 3D segmentation as a binary mask (1 lesion, 0 background). All segmentation masks were then reviewed and if necessary modified by an experienced radiologist (>20 years of experience in reporting CT scans).

The following 24 radiomics features were extracted from all voxels belonging to the 3D mask of the baseline portal phase CT: (a) volume of the tumor mask, (b) 4 first-order parameters, i.e., mean intensity, standard deviation, skewness and kurtosis; (c) 19 second-order texture parameters, 15 derived from the Grey-Level Co-Occurrence matrix (GLCM) (22) and 4 derived from the Grey-Level Run Length Matrix (GLRLM) (23). To extract the texture parameters, we used distance=1 voxel in order to evaluate the closest neighboring voxels, number of bins=64, intensities histogram of each ROI rescaled between the 1st and the 99th percentile of the ROI (14). The range of grey levels was

symmetrically enlarged to obtain an integer number of grey levels per bin. The GLCM and GLRLM matrices were generated for each of the 13 unique directions of a 3D image, and then averaged to make the method rotationally invariant to the distribution of texture. Texture features were computed using using an in-house framework based on C++ and IT libraries that was compliant to the Image Biomarker Standardization Initiative (IBSI) (24).

2.4. Feature selection and radiomics model development

To perform feature selection we used genetic algorithms (GAs), i.e. heuristic algorithms belonging to the computational intelligence field (25,26) . Each GA solution was coded as a binary 24 bits vector, one bit representing a feature: a "0" in a given position identified a feature not selected whereas a "1" labeled a feature included in the final subset. Each solution of the GA was evaluated by a fitness function that measured the ability of the corresponding feature subset to obtain a Gaussian naïve Bayesian classifier able to classify metastases of the training set. Since the extracted features presented very different ranges, the min-max scaling was applied to the training set to normalize each feature in the range [0, 1]. Then, only the features selected by the current solution (corresponding to bits equal to "1") were kept and used to build the classifier. The leave-one-out cross-validation (LOOCV) was applied to evaluate the generalization capability of the classifier and to avoid overfitting. In each iteration of the LOOCV all metastases of the same patient were excluded from the training set: a Gaussian naïve Bayesian classifier was constructed using the remaining metastases and the left-out metastases were used as test set. This procedure was repeated for all patients in the training set and the probability to be a R+ lesion was calculated for all lesions. To assign a class (R+ or R-) to the lesions, we applied the cut-off that optimized sensitivity and specificity, i.e., the Youden Index. Finally, we computed the fitness value of the current solution as:

$$fitness = 1 - \frac{SE+SP}{2} + 0.3 * (1 - npv * SP) \qquad (1)$$

where sensitivity (SE), specificity (SP) and negative predictive value (NPV) represent the SE, SP and NPV of the Gaussian naïve Bayesian classifier constructed using the current feature subset. Lower fitness values corresponded to better feature subsets. We decided to implement eq. (1) in order to use the average between SE and SP instead of the total accuracy to avoid bias due to the different number of R+ and R- metastases in the training set. A penalty term was added (the last part of eq. 1) to subsets producing inaccurate recognition of R- lmCRC.

Our algorithm started with an initial population of 500 randomly generated solutions. A roulette wheel selection (27) was then applied to select the 90% of solutions to be used as parents of the next generation: the probability of each solution to be selected was inversely proportional to its fitness value. Starting from the parent solutions, a set of newborns' solutions was generated applying a 4-point crossover operator with probability equal to 0.9, and the mutation operator in which bits of the solutions were complemented with a probability equal to 0.3. Finally, parents' and newborns' solutions were pooled together, and 500 solutions were randomly extracted and used to restart the algorithms. During the GA evolution, the best current solution, i.e. the one with the lowest fitness value in the actual population, was stored. This loop was iterated until either 500 iterations were reached, or no change of the best current solution occurred for 100 consecutive iterations.

To consider the random component of GA, the entire algorithm was repeated 10 times starting from the same initial population of random solutions. Thus, 10 feature subsets resulted from our algorithm: the best subset of features was identified as that with the lowest fitness value and the lowest number of selected features.

Once the best subset of features was selected, a Gaussian naïve Bayesian classifier was created on the training set and subsequently validated on the validation dataset. Since this classifier returns, for each metastasis, a score representing the probability to be a R+ lesion, we chose as best cut-off the one optimizing sensitivity and specificity, i.e., the Youden Index. The cut-off was chosen on the training set and applied to the validation dataset.

2.5. Statistical analysis

The primary endpoint of the study was to assess if radiomics features could predict response to treatment on a per-lesion basis. In this analysis sensitivity (SE), specificity (SP), NPV and positive predictive value (PPV) were computed on both the training and the validation sets. Sensitivity was defined as the ratio between the number of correctly classified R+ metastases over the total number of R+ metastases, specificity as the ratio between the number of correctly classified R- metastases over the total number of R- metastases, PPV as the ratio between the number of correctly classified R+ lmCRC over the total number of lmCRC classified as R+, while NPV as the ratio between the number of correctly classified R- lmCRC over the total number of lmCRC classified as R-. Differences in diameters between false positive, false negative and correctly classified metastases were evaluated using the Wilcoxon test. A p-value <0.05 was considered statistically significant.

In order to assess if radiomics features could predict overall response to treatment, we performed a per-patient analysis in which a patient was defined either R+ or R- if the majority of his metastasis were classified as R+ or R-, respectively. Patients having an even number of R+ and R- lmCRC, either in the reference standard or after the classification, were discarded in this analysis. Due to the low number of patients, per patient SE, SP, PPV and NPV were computed combining all patients from the training and validation dataset.

3. Results

List of extracted features and results of the feature selection process are shown in supplementary Table and Figure 1. The best feature subset, i.e. the one having the lowest fitness and the lowest number of selected features, was number 7. In total 12 features were selected: 2 first-order statistics, 9 from the GLCM matrix and 1 from the GLRLM matrix.

3.1. Results of per-lesion analysis

The study flow-chart is presented in Figure 2. A total of 141 lmCRC were evaluated in 38 patients (32 males, 6 females; median age 59±12 years) including: 89 with a diameter < than 30mm, 27 between 30 and 50 mm, and 25 > than 50 mm. One-hundred-eight lmCRC from 28 patients were included in the training dataset, 33 lmCRC from 10 patients in the validation dataset. In the training set 75 of 108 lesions (69%) were classified as R+, the remaining 33 (31%) as R-; in the validation dataset 21 of 33 liver metastases (64%) were classified as R+, the remaining 12 (36%) as R-. Mean lesion diameter on baseline CT scans was 33 mm (SD: ± 22mm) and no differences in diameter were observed between training and validation datasets. R+ lmCRC were larger than R- lmCRC (mean diameter of R+ lesions 37±24 mm versus 27±17 mm of R- lesions; p=0.004).

The final Bayesian classifier retained 12 of the 24 examined radiomics features. Figure 1 shows the absolute value of weights of each parameter in the linear combination of the Bayesian Classifier normalized by the sum of the total weights.

Per-lesion performance in predicting response to treatment of our algorithm is shown in Table 1. After applying feature selection to the training set, we obtained a sensitivity and specificity of 89% and 85%, respectively. When the same model and cut-off (0.465) were applied to the validation set, the sensitivity and specificity in detecting R+ lesions were 90% and 42% respectively. Classification error was independent on lesion diameters (p>0.16).

Figure 3 A shows a case of a patient with mixed response to therapy. Figure 4 shows waterfall plots of all the liver metastases in the training and validation sets. Four of 7 metastases in the validation dataset having a high radiomics score despite being R-, belonged to the same patient (Figure 3B).

3.2. Per-patient results

Table 2 shows findings of per-patient analysis. Among the 38 examined patients, 8 had all R- lmCRC (21%), 21 had all R+ lmCRC (55%), and 9 showed heterogeneous response (24%). All lmCRC were correctly classified by the radiomics algorithm in 24 patients while in 2 patients (n.10 and n.48, table 2) all lmCRC were misclassified. In the remaining 12 patients lesions were partially misclassified including: 7 were the majority of lmCRC were classified correctly, 4 where an equal number of lmCRC were correctly and incorrectly classified and one patient where the majority of lmCRC were misclassified (n. 33 Table 2).

Five patients had an equal number of R+ and R- lmCRC and were therefore excluded from per-patient analysis. Considering the remaining 33 patients, of whom 7 R- and 26 R+, overall per-patient sensitivity was 92% (24/26;95%CI:75-99%), specificity 86% (6/7; 95%CI:42-100%), PPV 96% (24/25; 95%CI:80-99%), and an NPV 75% (6/8; 95%CI:43-92.5%). According to our findings, 2 of the 33 patients (6%) misclassified as R- would not have undergone a beneficial treatment. The first patient (n. 33 Table 2) had 3 R+ lmCRC that were classified by the radiomics algorithm as 1 R+ and 2 R-; the second patient (n. 2, Table 2) had one stable lmCRC that was classified as R-.

4. Discussion

In this study heterogeneous radiological response to a targeted therapy was observed in 9 of 38 (24%) patients, more than twice the rate reported in the literature for response to chemotherapy (5,6), and definitely in contrast to the homogeneous pattern of response recently reported in mismatch repair deficiency CRC patients treated with immunotherapy (28). Lesion specific response has been previously documented in patients treated for lmCRC with targeted agents(7,29,30). Russo et al.(7) suggest that molecular heterogeneity may trigger unique patterns of response in different metastatic deposits. In the HERACLES-A trial (4), we have observed distinct patterns of radiologic and genomic evolution and have considered that lesion specific patterns of response might not be accurately evaluated by the RECIST criteria. Nevertheless, a recent pooled database analysis on a large data set of patients with solid tumors confirmed RECIST good performance in assessing response also to

targeted agents. Data warehouse analyses, however, are tremendously influenced by the choice of target lesions (31), lacks granularity in general (32) and in the specific focused for CRC predominantly on antiangiogenic treatment which is used in genetically unselected patients.

Moreover, at the present no imaging criteria are available to predict response before the start of the therapy and radiologists cannot gain sufficient visual information on the baseline examination to subjectively predict which lesions will respond to the therapy. In this study, we presented an innovative radiomics tool to predict response to treatment on a per-lesion basis in patients with HER2 amplified tumors treated with antiHER2 agents.

We assessed the radiomics score of individual metastases obtaining a per-lesion sensitivity and specificity in the validation set respectively of 90% and 42% at the best cut-off value. The algorithm was very effective in predicting responding lesions (R+). Indeed, only 10% of these (2/21), were misclassified as non-responding. Conversely, the radiomics model was not accurate in predicting which lesions would not have benefited from treatment (R-). Indeed, of the 12 lesions classified by the algorithm as non-responders, 7 were responding lesions. However, if the scope of radiomics is to identify patients with outlier lesions, i.e. that do not respond in a general condition where most lesions respond, then results of this study should be well received. Indeed, in the group of patients with heterogeneous response 5 of 9 were carriers of non-responding lesions that were correctly classified as such by our radiomics algorithm, including 4 of 7 with only one non-responding lesion. In this group of patients correctly identified by our method, biopsy could have revealed a different genetic makeup prompting the use of a different target molecule or, in absence of extrahepatic lesions, suggested the local ablation of outlier metastases.

The two most important radiomics features in our model are difference variance and homogeneity, weighing cumulatively over 50% in the Bayesian classifier. The former measures the dispersion (with respect to the mean) of the gray level difference distribution of the image while the latter measures the smoothness of the gray level distribution of the image, and both are somehow related to the heterogeneity of the region of interest. This finding could be translated in clinical practice. Indeed, if our observation is confirmed the identification of a more heterogeneous pattern within a lmCRC could be predictive of R+ (Figure 3A).

The predictive value of radiomics models has been previously explored(20,33). In particular Klaassen et al. have shown that a CT radiomics approach using Random Forest models was able to discriminate response of individual liver metastasis in esophagogastric cancer patients, being tumor heterogeneity the most predictive marker (20). In addition, Ahn et al.(33) identified radiomics

markers of increased heterogeneity as predictors of poor patient prognosis and opposite to this study they postulated that greater homogeneity was associated with a favorable clinical response. We therefore hypothesize lmCRC have different textures characteristics at diagnosis and after several lines of treatment and each condition could therefore benefit from a different radiomics signature. Ahn et al.(33) considered the largest lesion as representative of overall tumor behavior; they therefore did not explore behavior of individual liver metastasis.

There are limitations to this study. First, it is known that radiomics features are affected by CT equipment and protocols. However, Buch et al.(34) have demonstrated that some texture features are independent of CT parameters. In this retrospective multicenter study, we were able to overcome equipment and protocol variability by developing a feature selection algorithm that includes a cross-validation procedure, in which one patient was iteratively excluded from the training set and all his lmCRC used for testing. Therefore, only stable features among the different protocols were included in the final solution. Second, in this study specificity of the validation dataset was lower than that observed in the training dataset. This could depend on the size of the training and validation datasets and on the low number of R- lesions available overall, i.e. 33% of the total. Furthermore, we observed that in the validation dataset 4 of the 7 R+ lesions that were classified as R- had a very unique target pattern which could have affected results (Figure 3B). Of note, this study was performed on a group of 141 HER2-amplified lmCRC from a unique cohort of patients representing only 2% of all CRCs and the sample size might have been insufficient to answer our clinical questions. However, it must be noted that when applying machine learning it is not possible to know “a priori” the sample size that will be necessary to achieve adequate performance levels. Sample size may depend on several factors, including lesions characteristics and the addressed clinical question. While it is very hard to predict the minimum number of cases, it is usually recommended empirically that at least 100 cases are available (35). In order to assess the robustness of our radiomics model, we are planning to validate it on an external dataset of HER2 amplified patients. Finally, we are aware that this heavily pretreated cohort of patients might affect the predictive value of the radiomics model we developed. Further studies on first line patients should be performed, considering different molecular signatures of lmCRC as well, including non HER2 amplified patients.

5. Conclusions

In this study we have developed a radiomics signature to predict behavior of individual metastasis to targeted treatment in a cohort of HER2 amplified CRC patients with a high rate of

heterogeneous response. The model shows promising results in predicting responder lesions on the baseline CT and in the identification of non-responder lesions in patients with heterogeneous response, potentially paving the way to a more aggressive diagnostic and therapeutic approach in selected patients. Further validation will be needed to confirm our findings.

Funding: The research leading to these results has received funding from FONDAZIONE AIRC under 5 per Mille 2018 - ID. 21091 program – P.I. Bardelli Alberto, G.L. Regge Daniele.

Conflicts of interest: The authors declare no conflict of interest.

Data Availability Statement: The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Ethics statement: the study was performed in accordance with the principles of the Declaration of Helsinki and the International Conference on Harmonization and Good Clinical Practice guidelines. The institutional review boards of the participating centers approved the study and all patients provided written informed consent (registration number on clinicaltrials.gov: NCT03225937).

References

1. Blank A, Roberts DE, Dawson H, Zlobec I, Lugli A. Tumor heterogeneity in primary colorectal cancer and corresponding metastases. Does the apple fall far from the tree? Front Med [Internet]. 2018;5. Available from: https://www.frontiersin.org/article/10.3389/fmed.2018.00234/full

2. Meric-Bernstam F, Hurwitz H, Raghav KPS, McWilliams RR, Fakih M, VanderWalde A, et al. Pertuzumab plus trastuzumab for HER2-amplified metastatic colorectal cancer (MyPathway): an updated report from a multicentre, open-label, phase 2a, multiple basket study. Lancet Oncol. 2019;20:518–30.

3. Sartore-Bianchi A, Trusolino L, Martino C, Bencardino K, Lonardi S, Bergamo F, et al. Dual-targeted therapy with trastuzumab and lapatinib in treatment-refractory, KRAS codon 12/13 wild-type, HER2-positive metastatic colorectal cancer (HERACLES): a proof-of-concept, multicentre, open-label, phase 2 trial. Lancet Oncol [Internet]. 2016;17:738–46. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1470204516001509

4. Siravegna G, Lazzari L, Crisafulli G, Sartore-Bianchi A, Mussolin B, Cassingena A, et al. Radiologic and Genomic Evolution of Individual Metastases during HER2 Blockade in Colorectal Cancer. Cancer Cell [Internet]. 2018;34:148-162.e7. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1535610818302617

5. Van Kessel CS, Samim M, Koopman M, Van Den Bosch MAAJ, Borel Rinkes IHM, Punt CJA, et al. Radiological heterogeneity in response to chemotherapy is associated with poor survival in patients with colorectal liver metastases. Eur J Cancer [Internet]. 2013;49:2486–93. Available from: https://linkinghub.elsevier.com/retrieve/pii/S095980491300261X

6. Brunsell TH, Cengija V, Sveen A, Bjørnbeth BA, Røsok BI, Brudvik KW, et al. Heterogeneous radiological response to neoadjuvant therapy is associated with poor prognosis after resection of colorectal liver metastases. Eur J Surg Oncol. 2019;45:2340–6.

7. Russo M, Siravegna G, Blaszkowsky LS, Corti G, Crisafulli G, Ahronian LG, et al. Tumor heterogeneity and Lesion-Specific response to targeted therapy in colorectal cancer. Cancer Discov. 2016;6:147–53.

8. Piotrowska Z, Niederst MJ, Karlovich CA, Wakelee HA, Neal JW, Mino-Kenudson M, et al. Heterogeneity underlies the emergence of EGFRT790wild-type clones following treatment of T790M-positive cancers with a third-generation EGFR inhibitor. Cancer Discov. 2015;5:713–23.

9. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, Rubinstein L, et al. New guidelines to evaluate the response to treatment in solid tumors. J Natl Cancer Inst. 2000;92:205–16.

10. Aerts HJWL, Grossmann P, Tan Y, Oxnard GG, Rizvi N, Schwartz LH, et al. Defining a Radiomic Response Phenotype: A Pilot Study using targeted therapy in NSCLC. Sci Rep [Internet]. 2016;6:33860. Available from: http://www.nature.com/articles/srep33860

11. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images are more than pictures, they are data. Radiology. 2016;278:563–77.

12. Giannini V, Mazzetti S, Marmo A, Montemurro F, Regge D, Martincich L. A computer-aided diagnosis (CAD) scheme for pretreatment prediction of pathological response to neoadjuvant therapy using dynamic contrast-enhanced MRI texture features. Br J Radiol [Internet]. 2017;90:20170269. Available from: http://www.birpublications.org/doi/10.1259/bjr.20170269

13. Giannini V, Rosati S, Castagneri C, Martincich L, Regge D, Balestra G. Radiomics for pretreatment prediction of pathological response to neoadjuvant therapy using magnetic resonance imaging: Influence of feature selection. Proc - Int Symp Biomed Imaging. 2018. page 285–8.

14. Giannini V, Mazzetti S, Bertotto I, Chiarenza C, Cauda S, Delmastro E, et al. Predicting locally advanced rectal cancer response to neoadjuvant therapy with 18 F-FDG PET and MRI radiomics features. Eur J Nucl Med Mol Imaging [Internet]. 2019;46:878–88. Available from: http://link.springer.com/10.1007/s00259-018-4250-6

15. Trebeschi S, Drago SG, Birkbak NJ, Kurilova I, Călin AM, Delli Pizzi A, et al. Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers. Ann Oncol. 2019;30:998–1004.

16. Cain EH, Saha A, Harowicz MR, Marks JR, Marcom PK, Mazurowski MA. Multivariate machine learning models for prediction of pathologic response to neoadjuvant therapy in breast cancer using MRI features: a study using an independent validation set. Breast Cancer Res Treat [Internet]. 2019;173:455–63. Available from: http://link.springer.com/10.1007/s10549-018-4990-9

17. Bera K, Velcheti V, Madabhushi A. Novel Quantitative Imaging for Predicting Response to Therapy: Techniques and Clinical Applications. Am Soc Clin Oncol Educ B [Internet]. 2018;1008–18. Available from: http://ascopubs.org/doi/10.1200/EDBK_199747

18. 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 [Internet]. 2015;40:2331–7. Available from: http://link.springer.com/10.1007/s00261-015-0438-4

19. Simpson AL, Doussot A, Creasy JM, Adams LB, Allen PJ, DeMatteo RP, et al. Computed Tomography Image Texture: A Noninvasive Prognostic Marker of Hepatic Recurrence After Hepatectomy for

Metastatic Colorectal Cancer. Ann Surg Oncol. 2017;24:2482–90.

20. Klaassen R, Larue RTHM, Mearadji B, van der Woude SO, Stoker J, Lambin P, et al. Feasibility of CT radiomics to predict treatment response of individual liver metastases in esophagogastric cancer patients. Tian J, editor. PLoS One [Internet]. 2018;13:e0207362. Available from: http://dx.plos.org/10.1371/journal.pone.0207362

21. McErlean A, Panicek DM, Zabor EC, Moskowitz CS, Bitar R, Motzer RJ, et al. Intra- and interobserver variability in CT measurements in oncology. Radiology. 2013;269:451–9.

22. Haralick RM, Dinstein I, Shanmugam K. Textural Features for Image Classification. IEEE Trans Syst Man Cybern. 1973;SMC-3:610–21.

23. Thibault G, Angulo J, Meyer F. Advanced statistical matrices for texture characterization: Application to cell classification. IEEE Trans Biomed Eng. 2014;61:630–7.

24. Lambin P. Radiomics Digital Phantom. CancerData [Internet]. 2016;41:366–73. Available from: https://www.cancerdata.org/resource/doi:10.17195/candat.2016.08.1

25. De Leonardis G, Rosati S, Balestra G, Agostini V, Panero E, Gastaldi L, et al. Human Activity Recognition by Wearable Sensors : Comparison of different classifiers for real-time applications. MeMeA 2018 - 2018 IEEE Int Symp Med Meas Appl Proc [Internet]. IEEE; 2018. page 1–6. Available from: https://ieeexplore.ieee.org/document/8438750/

26. Rosati S, Gianfreda CM, Balestra G, Giannini V, Mazzetti S, Regge D. Radiomics to predict response to neoadjuvant chemotherapy in rectal cancer: Influence of simultaneous feature selection and classifier optimization. 2018 IEEE Life Sci Conf LSC 2018. 2018. page 65–8.

27. Pedrycz W, Sillitti A, Succi G. Computational intelligence: An introduction. Stud Comput Intell [Internet]. 2016. page 13–31. Available from: http://link.springer.com/10.1007/978-3-319-25964-2_2

28. Osorio JC, Arbour KC, Le DT, Durham JN, Plodkowski AJ, Halpenny DF, et al. Lesion-level response dynamics to programmed cell death protein (PD-1) blockade. J Clin Oncol. 2019;37:3546–55.

29. Oddo D, Siravegna G, Gloghini A, Vernieri C, Mussolin B, Morano F, et al. Emergence of MET hyper-amplification at progression to MET and BRAF inhibition in colorectal cancer. Br J Cancer. 2017;117:347–52.

30. Pietrantonio F, Oddo D, Gloghini A, Valtorta E, Berenato R, Barault L, et al. MET-driven resistance to dual EGFR and BRAF blockade may be overcome by switching from EGFR to MET inhibition in

BRAF-mutated colorectal cancer. Cancer Discov. 2016;6:963–71.

31. Kuhl CK, Alparslan Y, Schmoee J, Sequeira B, Keulers A, Brümmendorf TH, et al. Validity of RECIST version 1.1 for response assessment in metastatic cancer: A prospective, multireader study. Radiology. 2019;290:349–56.

32. Kuhl CK. RECIST needs revision: A wake-up call for radiologists. Radiology. 2019. page 110–1.

33. Ahn SJ, Kim JH, Park SJ, Han JK. Prediction of the therapeutic response after FOLFOX and FOLFIRI treatment for patients with liver metastasis from colorectal cancer using computerized CT texture analysis. Eur J Radiol. 2016;85:1867–74.

34. Buch K, Li B, Qureshi MM, Kuno H, Anderson SW, Sakai O. Quantitative assessment of variation in CT parameters on texture features: Pilot study using a nonanatomic phantom. Am J Neuroradiol. 2017;38:981–5.

35. Masutani Y, Nemoto M, Nomura Y, Hayashi N. Clinical Machine Learning in Action: CAD System Design, Development, Tuning, and Long-Term Experience. In: Suzuki K, editor. Mach Learn Comput Diagnosis Med Imaging Intell Anal [Internet]. Hershey, PA, USA: IGI Global; 2012. page 159–76. Available from: http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-4666-0059-1.ch008

Table 1: Per-lesions results of the training and validation dataset, before and after applying feature selection.

	TRAINING				VALIDATION			
	Sensitivity (95%CI)	**Specificity (95%CI)**	**PPV (95%CI)**	**NPV (95%CI)**	**Sensitivity (95%CI)**	**Specificity (95%CI)**	**PPV (95%CI)**	**NPV (95%CI)**
All features	91% (82-96%) [68/75]	91% (76-98%) [30/33]	96% (88-99%) [68/71]	81% (68-90%) [30/37]	76% (53-92%) [16/21]	42% (15-72%) [5/12]	70% (57-80%) [16/23]	50% (27-73%) [5/10]
Features selection	89% (80-95%) [67/75]	85% (68-95%) [28/33]	93% (70-99%) [67/72]	78% (64-87%) [28/36]	90% (70-99%) [19/21]	42% (15-72%) [5/12]	73% (62-82%) [19/26]	71% (36-92%) [5/7]

Numbers in brackets represent absolute values

Table 2: Total number of responder and non-responder metastases for each patients and results of the Bayesian classifier.

	Patient #	Total #mts	#R+ mts	#R- mts	Mixed response	#correctly classified R+	#correctly classified R-
TRAINING	1	2	2	0		2/2	-
	2	1	1	0		0/1	-
	5	2	0	2		-	2/2
	6	5	5	0		3/5	-
	10	1	0	1		-	1/1
	17	2	1	1	Yes	1/1	1/1
	19	6	6	0		5/6	-
	20	2	2	0		2/2	-
	22	1	1	0		1/1	-
	25	3	3	0		3/3	-
	26	5	4	1	Yes	3/4	1/1
	28	7	6	1	Yes	6/6	1/1
	29	1	1	0		1/1	-
	31	5	4	1	Yes	4/4	0/1
	32	9	6	3	Yes	6/6	2/3
	37	8	0	8		-	8/8
	38	7	7	0		7/7	-
	41	4	4	0		4/4	-
	45	4	3	1	Yes	3/3	0/1
	46	1	1	0		1/1	-
	48	2	0	2		-	0/2
	49	3	3	0		3/3	-
	50	1	0	1		-	1/1
	52	2	1	1	Yes	0/1	1/1
	54	5	5	0		5/5	-
	55	1	1	0		1/1	-
	62	8	8	0		6/8	-
	66	10	0	10		-	10/10
VALIDATION	4	1	1	0		1/1	-
	12	2	1	1	Yes	1/1	0/1
	18	6	6	0		6/6	-
	23	4	4	0		4/4	-
	27	4	2	2	Yes	2/2	0/2
	30	1	0	1		-	1/1
	33	3	3	0		1/3	-
	36	1	1	0		1/1	-
	57	3	3	0		3/3	-
	63	8	0	8		-	4/8

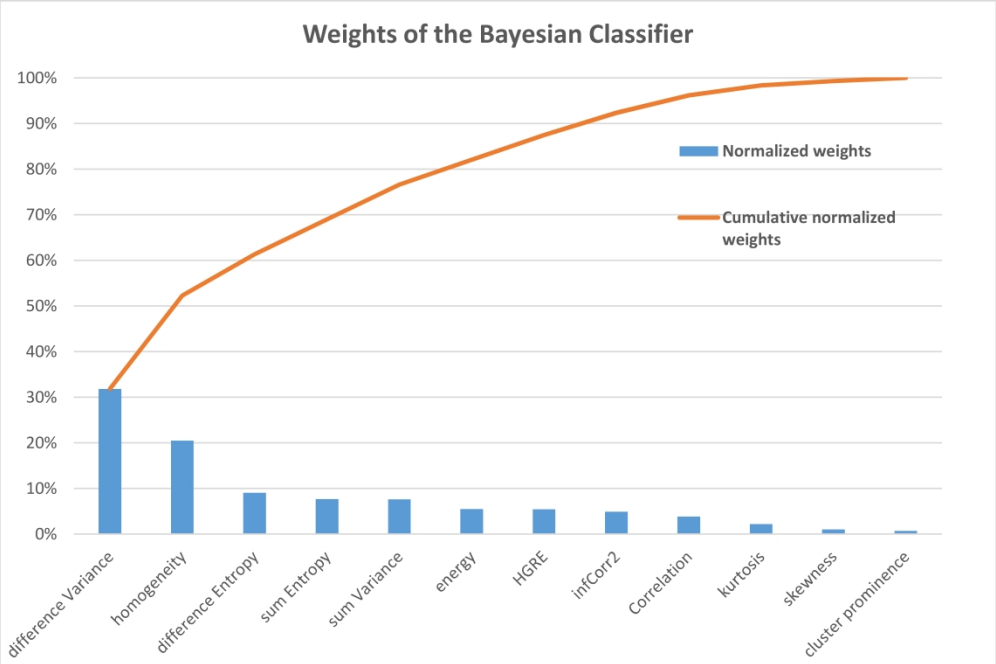


Figure 1: Absolute value of weights obtained by each selected parameter in the final Bayesian Classifier normalized by the sum of the weights.

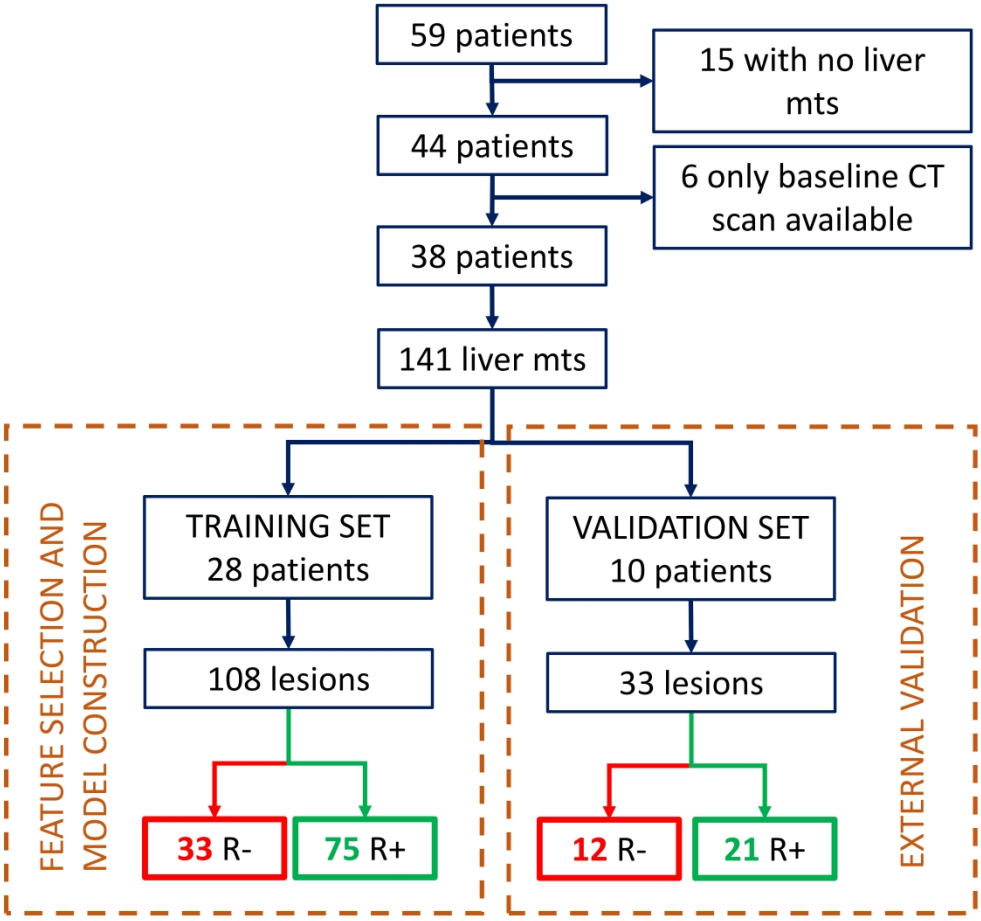
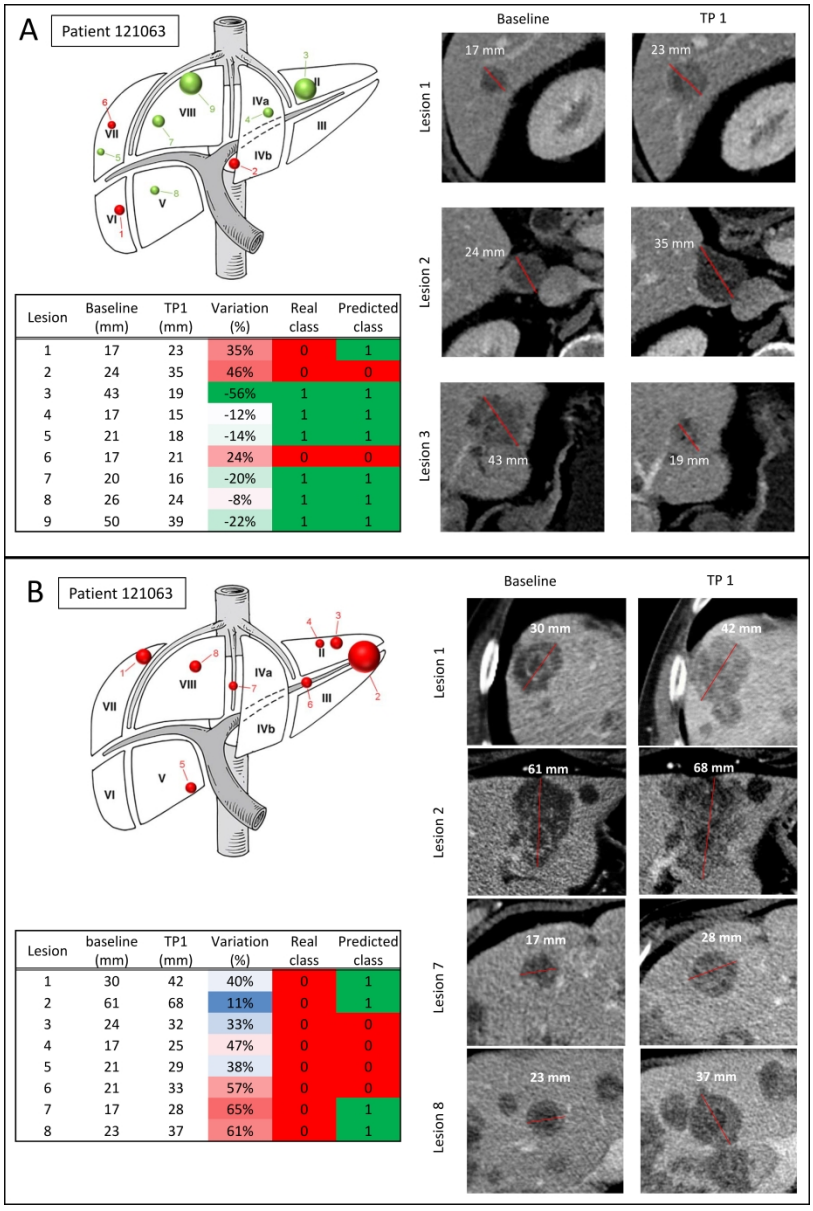


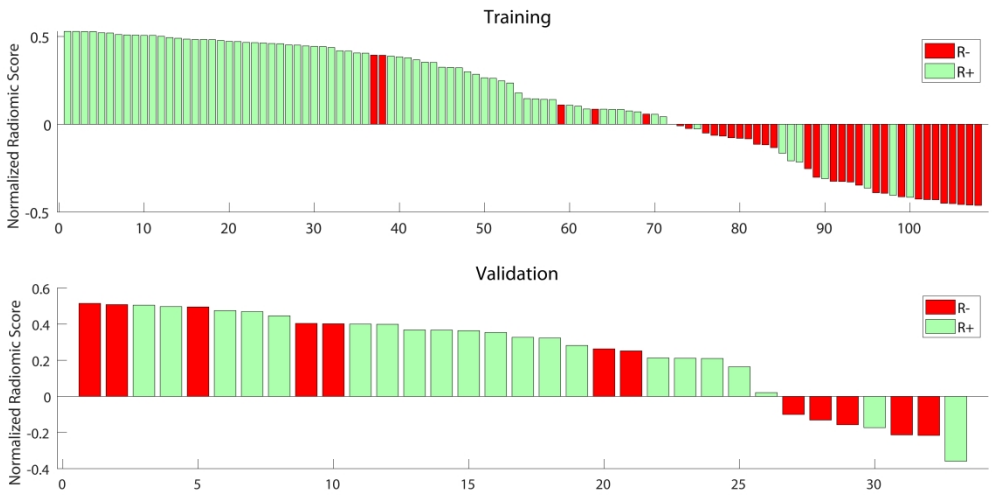
Figure 2: flowchart of the study



Lesion	Baseline (mm)	TP1 (mm)	Variation (%)	Real class	Predicted class
1	17	23	35%	0	1
2	24	35	46%	0	0
3	43	19	-56%	1	1
4	17	15	-12%	1	1
5	21	18	-14%	1	1
6	17	21	24%	0	0
7	20	16	-20%	1	1
8	26	24	-8%	1	1
9	50	39	-22%	1	1

Lesion	baseline (mm)	TP1 (mm)	Variation (%)	Real class	Predicted class
1	30	42	40%	0	1
2	61	68	11%	0	1
3	24	32	33%	0	0
4	17	25	47%	0	0
5	21	29	38%	0	0
6	21	33	57%	0	0
7	17	28	65%	0	1
8	23	37	61%	0	1

Figure 3: A) Example of a patient with 9 liver metastases and a heterogeneous response. Only liver lesion1 was misclassified by the radiomics model. The table lists the patient's liver metastases, size at baseline and at time point 1 and percentage size variations (increase in size is represented in red, decrease in green). The last 2 columns show respectively the response of each lesion based on size variations and the classification as predicted by the classifiers. TP1=time point 1; mm=millimeters. B) Example of a patient with 8 liver metastases including 4 (lesion 1,2,7,8) that were misclassified by the radiomics model, with an uncharacteristic target-like appearance. The table lists the patient's liver metastases, size at baseline and at time point 1 and percentage size variations (increase in size is represented in red, decrease in green). The last 2 columns show respectively the response of each lesion based on size variations and the classification as predicted by the classifiers. TP1=time point 1; mm=millimeters.



Caption : Figure 4: Waterfall plot of all lesions. The green marks indicate the responder lesions, while the red marks represent the non-responder lesions. The y-axis represents the radiomic score produced by the naïve Bayesian Classifier normalized with the cut-off computed on the training set using the Youden Index.