2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Comparison of radiomics approaches to predict resistance to 1st line chemotherapy in liver metastatic colorectal cancer.

Arianna Defeudis, Lorenzo Cefaloni, Giuliana Giannetto, Giovanni Cappello, Francesco Rizzetto, Jovana Panic, Davide Barra, Giulia Nicoletti, Simone Mazzetti, Alberto Vanzulli, Daniele Regge, Valentina Giannini

***Abstract*— Colorectal cancer (CRC) has the second-highest tumor incidence and is a leading cause of death by cancer. Nearly 20% of patients with CRC will have metastases (mts) at the time of diagnosis, and more than 50% of patients with CRC develop metastases during their disease. Unfortunately, only 45% of patients after a chemotherapy will respond to treatment. The aim of this study is to develop and validate a machine learning algorithm to predict response of individual liver mts, using CT scans. Understanding which mts will respond or not will help clinicians in providing a more efficient per-lesion treatment based on patient specific response and not only following a standard treatment. A group of 92 patients was enrolled from two Italian institutions. CT scans were collected, and the portal venous phase was manually segmented by an expert radiologist. Then, 75 radiomics features were extracted both from 7x7 ROIs that moved across the image and from the whole 3D mts. Feature selection was performed using a genetic algorithm. Results are presented as a comparison of the two different approaches of features extraction and different classification algorithms. Accuracy (ACC), sensitivity (SE), specificity (SP), negative and positive predictive values (NPV and PPV) were evaluated for all lesions (*per-lesion* analysis) and patients (*per-patient* analysis) in the construction and validation sets. Best results were obtained in the *per-lesion* analysis from the 3D approach using a Support Vector Machine as classifier. We reached on the training set an ACC of 81%, while on test set, we obtained SE of 76%, SP of 67%, PPV of 69% and NPV of 75%. On the validation set a SE of 61%, SP of 60%, PPV of 57% and NPV of 64% were reached. The promising results obtained in the validation dataset should be extended to a larger cohort of patient to further validate our method.**

***Clinical Relevance*— to develop a radiomics signatures predicting single liver mts response to therapy. A personalized mts approach is important to avoid unnecessary toxicity offering more suitable treatments and a better quality of life to oncological patients.**

I. INTRODUCTION

Colorectal cancer (CRC) has the second-highest tumor incidence and is a leading cause of death by cancer. [1]. Approximately 35% of the CRC patients present with stage IV metastatic disease at the time of diagnosis and up to 70% of patients will develop liver metastases (mts) in their life [2]. It is then clear how further investigating CRC and its deriving liver metastases is of the utmost importance for a better understanding of the disease and to develop new patient-oriented treatment plans. Empirically the backbone chemotherapy standard treatment in mCRC is using FOLFOX and/or FOLFIRI. However, 55% of patients undergoing this standard first-line chemotherapy do not respond to treatment or respond for a short period of time and then progress [3-4]. Non-responders, if predicted, could benefit from alternative treatments and/or avoid toxicity. Radiomics and textural analysis have the potential to differentiate responders to non-responders and open the way to a *per-lesion* personalized approach [5] [6]. The aim of the study is to compare radiomics results using two different approaches for features extraction. A 3D approach, where features are extracted from the whole metastasis and a ROI's approach, extracting features from a 7x7 ROI crossing the metastasis. Our idea was to compare features from the whole tumor, that do not consider the metastasis heterogeneity but only a mean value, with those extracted from smaller regions inside the tumor trying to appreciate small changes in tumor heterogeneity within ROIs and to classify each metastasis according to its single behavior.

A.D., D.B., G.N., S.M., D.R., V.G. are with University of Turin, Department of Surgical Science, Torino, Italy (corresponding author email: arianna.defeudis@unito.it).

G.C., G.G., D.R. are with Candiolo Cancer Institute, FPO - IRCCS, Strada Provinciale 142, km 3.95, Candiolo (TO), Italy.

L.C., J.P. are with Polytechnic of Turin, Department of Electronics and Telecommunications, Torino, Italy.

F.R., A.V. are with Department of Radiology, ASST Grande Ospedale Metropolitano Niguarda, Milan, Italy.

II. MATERIALS & METHODS

A. Patients and reference standard

We retrospectively evaluated 92 patients with a newly diagnosed stage IV CRC treated with a standard first-line chemotherapy and having at least one measurable secondary liver lesion as defined by the RECIST Criteria (greater diameter ≥ 10 mm). 31 patients were enrolled in a clinical trial at the Candiolo Cancer Institute, FPO-IRCCS (Center A) and 61 at the Niguarda Cancer Center (Center B). All patients underwent a CT examination with contrast agent injection within 2 weeks from the beginning of the first line treatment (baseline CT) and after 3 months of therapy. A resident radiologist, with 5 years of experience in reading CT exams, manually segmented all liver mts with a diameter ≥ 10 mm using an open-source software (ITK-snap) on the portal phase of the baseline CT exam. For each patient, a maximum number of 10 mts were selected (excluding confluent/subdiaphragmatic mts, or those containing large vessels). For each segmented mts, the radiologist measured the longest diameter at baseline and after 3 months of

978-1-7281-1178-0/21/$31.00 ©2021 IEEE

therapy (time point -TP- 1). Mts were accordingly classified as non-responder (R-) and responder (R+), as following:

- if the baseline diameter was > 20 mm (large lesions), a decrease in diameter between baseline and TP1 greater than 30% was classified as R+, otherwise R-.
- if the baseline diameter was <= 20 mm (small lesions), a decrease in diameter between baseline and TP1 greater than 4 mm was classified as R+, otherwise R-. Stable disease was considered R- because, given that all patients were at their first treatment, a more prominent answer was expected.

The study was approved by the local Ethics Committee, in accordance with the Helsinki Declaration; signed informed consent to use and analyze imaging data was obtained from all participants before entering the study.

B. *Features extraction (FE)*

Radiomics features (RF) were extracted on the baseline CT image from both a 7x7 ROI, that moved across the image by step of 2 pixels (*ROI's approach*) and from the whole tumor volume (*3D approach*). To guarantee reproducibility, an in-house software, *TexTO*, compliant with the Image Biomarker Standardization Initiative (IBSI) was used, implemented in *C++* and *ITK libraries* [7].

A total of 75 features were computed: 17 of the intensity based statistical features (STAT), 17 from the Intensity Histogram (IH), 25 from the Grey Level Co-occurrence Matrix (GLCM), 16 from Grey Level Run Length Matrix (GLRLM). To extract texture parameters, distance equal to 1 was used to evaluate the closest neighboring voxels and number of bins equal to 64 was used and intensity histogram rescaled between the 1st and the 99th percentile of each region. ROIs were classified as R+ or R- based on the classification of the lesion to whom they belonged. Patients were divided into a construction dataset and a validation set.

C. *Dataset construction: ROI's approach*

Patients were divided based on the number of lesions belonging to each class in a construction and validation datasets. The validation set was left out from preliminary outlier analysis. Outlier ROIs were detected if at least one of each feature's value deviated more than three times the standard deviation from the mean of all the values of that feature, on all ROIs. Outlier ROIs were removed from the training and included in the test set. A dendrogram clustering was performed to build a strong training set. For each lesion, a dendrogram was constructed and a percentage of ROIs belonging to each cluster was randomly extracted. 5 different training sets were created:

- TRS1 was built by extracting, from each cluster, 30% of ROIs for small lesions and 40% for big ones
- TRS2 was built by extracting, from each cluster, 40% of ROIs for small lesions and 50% for big ones
- TRS3 was created as TRS1, though removing ROIs belonging to the metastases that were found to be outliers
- TRS4 was built as TRS1 but considering only ROIs belonging to the slice with the maximum lesion diameter (widest).
- TRS5 was built as TRS1 considering the ROIs belonging to the widest slice of lesion plus the one above and below.

For each group, the remaining ROIs were used as test sets.

D. *Dataset construction: 3D approach*

For the *3D approach,* class imbalance did not appear to be a relevant issue since the construction/validation set partition resulted in a balanced per-lesion division, nevertheless stratified sampling for each class was performed.

Two different sets were created:

- TRS1_3D by randomly extracting 70% of lesions of both classes, R+ and R -.
- TRS2_3D by randomly extracting 80% of lesions of class 0 and 60% of lesions of class 1.

For each partition, the remaining lesions were used as test sets. In conclusion, 5 training sets were created from dataset 1 (*ROI's approach*) and 2 were created from dataset 2 (*3D*).

E. *Feature selection*

For feature selection a Genetic Algorithm (GA) was used [8]. Each GA solution was used to train a machine learning model using the training set, afterwards used to obtain a first prediction on the test set. The goodness of each solution explored by the algorithm was evaluated by 2 different fitnesses (F):

$$F1 = 1 - \frac{SE+SP}{2} \quad (1)$$

$$F2 = 1 - \frac{SE+SP}{2} + (0.3 * NPV * SP) \quad (2)$$

Where SE and SP represent sensitivity and specificity, respectively, and NPV negative predictive value, of the trained model constructed using the current features subset. It was decided to prefer SE and SP and not the overall accuracy, with reference F1, to maximize per class classification. On the other hand, a penalty term was added in F2 to account for class R-, we decide to favor specificity, since the aim is mainly to avoid treating lesions that do not respond to therapy. The algorithm started with an initial population of 600 randomly generated solutions. To extract 80% of the solutions, a roulette wheel selection was carried out, favoring those that minimized their fitness value and applying a 4-point crossover operator with probability equal to 0.8 and bit mutation probability equal to 0.3. During each GA iteration the best solution was stored, until either 2500 iterations were reached or no changed occurred for 50 consecutive iterations. To consider the random component of GA, the algorithm was run 5 times starting from the same initial population. For each repetition, the best solution was saved, the one that minimized the fitness value.

F. *Predictive Models*

Three different predictive models were chosen to carry out classification.

- Gaussian Naive Bayes classifier (NB) usually used if the dimensionality of the input set in high.
- Multilayer Perceptron (NN) tested with different structures, from one hidden layer up to 4 hidden layers. Additional 4 bits were added in the GA solution, allowing thus to include 16 possible structural combinations in the minimization of the fitness function.
- Support Vector Machine (SVM) for which some parameter tuning was performed. An additional 5 bits were

encoded in the solution, 3 of which used to test 8 different box constrains and 2 to test 4 possible kernel functions. For the *ROI's approach*, due to the extensive computational times required and the amount of data to be processed, SVM was discarded, and not inserted in the study.

G. Statistical Analysis

After having trained the model and evaluated its performances on the test set in the GA, the solution that minimized the fitness function was selected as the best one and evaluated on the validation dataset. The latter was never seen by the classifier during the training phase. Since the goal of the study is to correctly classify mts, particular attention was given to *per-lesion analysis*. For each lesion, the percentage of ROI classified as R+ was computed and the corresponding ROC curve was constructed. Each lesion having a percentage of R+ ROIs higher than the value represented by the Youden Index (YI) derived from the ROC curve of the construction set was considered as R+, otherwise R-. The YI evaluated through the construction set was used also on the validation set. *Per-Roi, per-lesion* and *per-patient* performances were evaluated with SE, SP, NPV, positive predictive value (PPV), and overall accuracy (ACC), both on construction and validation datasets. Statistics analysis was performed with MATLAB 2019 ®.

III. RESULTS

A. Patients

92 patients were included in the analysis. 27 of them had all R+ lesions, 30 all R- lesions and 35 mixed (M) response (both R+ and R). Patients were divided based on the number of lesions belonging to each class in the construction and validation sets. The construction set was composed of 54 patients (15 R+, 17 R- and 22 M) and 259 mts (127 R+ and 132 R-). The remaining 38 patients belonged to the validation set (12 R+, 13 R-, 13 M) and 127 mts (59 R+ 68 R-).

B. ROI's approach

Using features extracted from each ROIs, a total of 20 GAs were run, combining 5 different training sets, 2 fitness functions, and 2 machine learning models. The NN trained with TRS2 and F1, showed the highest NPV values across training, test and validation set. The *per-ROI* analysis yielded 49% in SP and 65% in NPV, 77% in SE and 65% in PPV, while ACC was 65%, in the training set. Performances were quite similar on the test set, with 38% in SP and 63% in NPV, 79% in SE and 59% in PPV, and ACC of 60%. *Per-ROI* validation set analysis reported 43% in SP and 69% in NPV, 64% in SE and 36% in PPV and ACC equal to 50%. The corresponding ROC curve was constructed, and the best cut-off point to maximize SE and SP was calculated on the construction set. The deriving Youden Index used was equal to 0.8. This means that for a lesion to be R+ it must have at least 80% of its ROIs classified as R+. Using this cut-off on both the construction and the validation sets, we obtained the per-lesion results showed in Table 1, *ROI's approach*. Our algorithm reached an ACC of 63% (164/259) and an Area Under the Curve (AUC) value of 0.68 in the construction set and ACC of 48% (61/127) and AUC of 0.56 in the validation set. For *per-patient* analysis only, patients with the same number of R+ and R- lesions were not included, 3/22 in the construction set were removed and 2/13 in the validation set. Performances showed an ACC of 60% (29/48) on the construction set, and 61% (16/26) on the validation. Additional results are shown in Table 1.

C. 3D approach

Using features extracted from the whole volume, a further investigation was conducted. A total of 12 GA features selection algorithms were run, combining 2 different training sets, 2 fitness functions and 3 machine learning models. The gaussian SVM trained with TRS1_3D and F2 showed the high performance, especially in the validation dataset. Training set showed an ACC of 82% (148/180), 82% in SP, 81% SE, 82% in NPV and 81% in PPV. Test set performance showed an ACC equal to 72% (57/79). Promising results were observed also on the validation set, where the ACC was 61% (77/127). Other results are shown on Table 1, *3D approach*. *Per-patient* analysis was also carried out using the best SVM model. Patients with same number of R+ and R- lesions were also excluded, 6% (3/54) and 5% (2/38) of the construction and validation sets. The best model yielded an ACC of 73% (31/42), but only 32% (10/31) on the validation. In Figure 1 accuracy trend is shown: it represents the range between min-max values of ACC we obtained after tuning each classifier. 3D approach reaches always higher ACC values, around 20% more in *per-lesion* analysis, with the SVM ranging 76-90% in training and 55-62% in validation set. While ROI approach has a stabler behavior between classifiers and type of analysis, with NN in *per-patient* ranging between 45 to 60.

	Construction set				**Validation set**			
Per-lesion	***Sensitivity***	***Specificity***	***NPV***	***PPV***	***Sensitivity***	***Specificity***	***NPV***	***PPV***
ROI's approach	61 (53-69)	65 (53-73)	63 (55-71)	62 (54-70)	40 (42-66)	54 (33-53)	51 (39-63)	43 (30-56)
3D approach	76 (63-89)	67 (52-82)	75 (62-88)	69 (54-84)	61 (49-73)	60 (48-72)	64 (53-75)	57 (44-70)
Per-patient								
ROI's approach	66 (47-85)	54 (34-74)	61 (41-81)	59 (39-79)	58 (39-89)	64 (39-89)	64 (39-89)	61 (30-86)
3D approach	80 (62-98)	68 (49-87)	78 (61-95)	69 (49-89)	41 (18-64)	21 (15-42)	23 (10-45)	38 (15-61)

Table 1: BEST RESULTS OF THE PER-LESION and PER-PATIENT ANALYSIS. VALUES ARE EXPRESSED IN PERCENTAGE (95% C.I.).

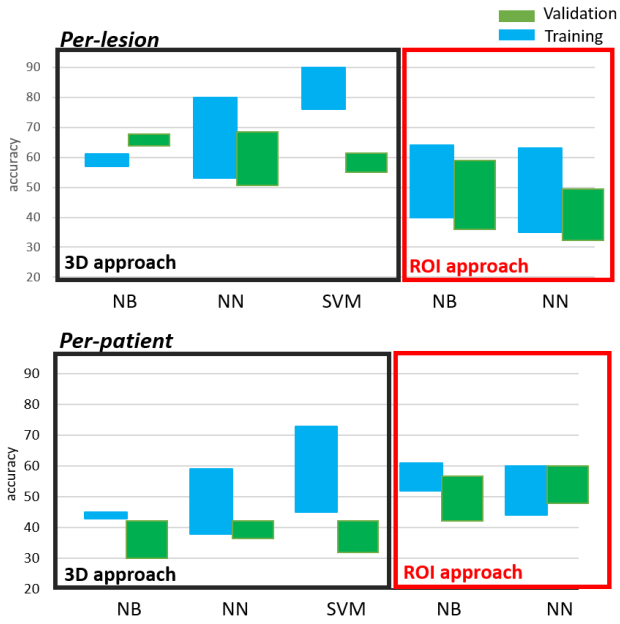

Figure 1: Accuracy trend after tuning of each classifier. Values are in %.

IV. DISCUSSION

In our study, we compare the feasibility of developing a radiomics model able to predict response of single liver mts in patients with CRC using a ROI and a 3D FE approach. Our aim is to analyze whether a more in-deep analysis of FE could bring benefit and detail of the heterogeneity within tumor and a more accurate radiomics classification. The most model with best performance we obtained was the 3D approach using a GA for FS and a SVM algorithm for classification, reaching ACC in the training, testing and validation set of 61%, 72%, 82%, respectively. While, in the ROI's approach we reached an accuracy of 63% on the construction set, but only equal to 48% on the validation cohort. As shown in Figure 1, we explored several possible tunings of classifiers (fitness and training set partitions), obtaining results ranging between similar values. In the literature, the predictive values of RF in metastatic CRC have been previously analyzed, but most of these studies performed only a *per-patients* analysis or a not fully complete *per-lesion* analysis. This means that they investigated only the largest hepatic mts; conversely, we assessed both per-lesion and *per-patient analysis* and we evaluated a large number of mts per patient [9][10][11]. The few studies that compared differences between R+ and R- in each single mts used a single dataset for model implementation, while we validated our results on an independent cohort [12][13]. As far as we know, this is the first work evaluating a ROI's approach to predict response of single liver mts on first-line chemotherapy and it could be considered a preliminary study. According to our findings, mts heterogeneity is not well represented by 7x7 ROIs, probably because they are too small to capture such tumor characteristics. Our innovation is dual: we compared different techniques of features extraction, exploring the ROI's approach and we trained and validated machine learning models to predict response to treatment of both single mts and patients. Our work has some limitations. First, the *per-patient* performance, especially for the 3D approach are quite low. However, our scope was to identify patients with outlier lesions, that did not respond in a general condition where most lesions responded to therapy. In addition, the total number of patients is small and should be increased to better generalize these preliminary results. Second, we did not perform an inter-reader analysis of manual segmentations, but we are already working on the assessment of variability between manual and automatic segmentations, building a deep learning model for a better understanding of how FS is affected by segmentation. Third, further studies are needed to better explore the ROI's approach, increasing the size of the ROI.

In conclusion, we compared different radiomic approaches to classify single liver mts, predicting response to therapy. The results are promising and could pave the way in future radiomics studies using the new ROI-based approach.

ACKNOWLEDGMENT

This work was funded by FONDAZIONE AIRC under 5 per Mille 2018-ID.21091 program –P.I. Bardelli Alberto, G.L.Regge Daniele.

REFERENCES

[1] R. Vera et al, "Multidisciplinary management of liver metastases in patients with colorectal cancer: a consensus of SEOM, AEC, SEOR, SERVEI, and SEMNIM", Clin Transl Oncol 22, 2020.

[2] S. Trebeschi et al., "Predicting response to cancer immunotherapy using noninvasive radiomic biomarkers," Ann. Oncol., 2019.

[3] Aparicio et al. "Metastatic Colorectal Cancer. First Line Therapy for Unresectable Disease", Journal of Clinical Medicine, 2020.

[4] Vogel et al. "First-line molecular therapies in the treatment of metastatic colorectal cancer – a literature-based review of phases II and III trials", Innov Surg Sci 2018.

[5] Giannini et al. "Radiomics predicts Response of individual HER2-amplified Colorectal Cancer Liver Metastases in Patients Treated with HER2- targeted Therapy", IJC, 2020.

[6] Giannini et al. "An innovative radiomics approach to predict response to chemotherapy of liver metastases based on CT images" in 2020 IEEE Engineering in Medicine and Biology Society (EMBC), 2020.

[7] A. Zwanenburg et al., "The Image Biomarker Standardization Initiative: Standardized Quantitative Radiomics for High-Throughput Image-based Phenotyping," Radiology, 2020.

[8] S. Rosati et al., "Radiomics to predict response to neoadjuvant chemotherapy in rectal cancer: Influence of simultaneous feature selection and classifier optimization," in 2018 IEEE LSC.

[9] A. Dohan et al., "Early evaluation using a radiomic signature of unresectable hepatic metastases to predict outcome in patients with colorectal cancer treated with FOLFIRI and bevacizumab", British Society of Gastroenterology, BMJ, 2020.

[10] S. J. Ahn et al., "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.

[11] R. Nakanishi et al.,"Radiomics Texture Analysis for the Identification of Colorectal Liver Metastases Sensitive to First-Line Oxaliplatin-Based Chemotherapy", Ann Surg Oncol, 2021.

[12] R. C. J. Beckers et al., "CT texture analysis in colorectal liver metastases and the surrounding liver parenchyma and its potential as an imaging biomarker of disease aggressiveness, response and survival," Eur. J. Radiol., 2018.

[13] A. Saini et al., "Radiogenomics and radiomics in liver cancers," Diagnostic. 2019.