Proceedings of the 10th World Congress on New Technologies (NewTech'24)
Barcelona, Spain - August 25-27, 2024
Paper No. ICBB 120
DOI: 10.11159/icbb24.120

A Fully Automatic Multi-Vendor AI-System To Segment And Predict Resistance To Treatment Of Rectal Cancer On MRI

Jovana Panic[12], Arianna Defeudis[3], Lorenzo Vassallo[4], Stefano Cirillo[5], Marco Gatti[2], Antonio Esposito[6], Serena dell'Aversana[7], Salvatore Siena[8], Angelo Vanzulli[8], Daniele Regge[3], Samanta Rosati[1], Gabriella Balestra[1], Valentina Giannini[2,3]

[1] Department of Electronics and Telecommunications, Polytechnic of Turin, Turin, Italy
jovana.panic@polito.it; samanta.rosati@polito.it; gabriella.balestra@polito.it; valentina.giannini@polito.it
[2] Department of Surgical Science, University of Turin, Turin, Italy
m.gatti@unito.it
[3] Department of Radiology, Candiolo Cancer Institute, FPO-IRCCS, Candiolo, Italy
arianna.defeudis@ircc.it; daniele.regge@ircc.it
[4] Department of Diagnostic Imaging and Radiotherapy, AOU Città della Salute e della Scienza, Turin, Italy
lvassallo@cittadellasalute.to.it
[5] Department of Radiology, A. O. Ordine Mauriziano (Ospedale Umberto I), Turin, Italy
s.cirillo@mauriziano.it
[6] School of Medicine, Vita-Salute San Raffaele University, Milan, Italy
esposito.antonio@hsr.it
[7] Department of Radiology, Santa Maria delle Grazie Hospital, ASL Napoli 2 Nord, Pozzuoli, Italy
dellaversanaserena@gmail.com
[8] Niguarda Cancer Center, Grande Ospedale Metropolitano Niguarda, Milan, Italy
salvatore.siena@ospedaleniguarda.it; angelo.vanzulli@ospedaleniguarda.it

Abstract **-** In this study, we developed and validated a fully automatic system based on pretreatment MRI to predict resistance to therapy in rectal cancer patients using a multi-center and multi-vendor database. Tumors were automatically segmented using in-house automatic U-Net segmentations and subsequently classified as responder and non-responder through a Random Forest algorithm that was fed with a subset of features selected by a customized features selection approach. Despite the strong imbalance between the two classes, the performances yielded are promising, with an area under the curve of 0.72 and a balanced accuracy of 66% on the external validation set. Even if further analyses are still required to improve the performance, our results represent a further step towards a more personalized medicine for patients with rectal cancer.

Keywords**:** radiomics, automatic segmentation, rectal cancer, multi-center database, therapy resistance

1 Introduction

Rectal cancer (RC) is the third leading cancer-related cause of death globally, among both men and women [1]. Even if there is not a single cause of RC, there are several risk factors related to both the patient's medical history and unbalanced lifestyle [2]. The current recommended diagnosis is carried out using Magnetic Resonance Imaging (MRI) for both initial and after-therapy staging, while the treatment plan is neoadjuvant chemoradiotherapy (nCRT) to decrease tumor size, followed by Total Mesorectal Excision [3][4]. However, the tumor response to nCRT remains variable, and up to 80% of patients respond partially or show a tumor progression. The latter undergo an unnecessary treatment that can also produce some serious side effects such as incontinence and sexual dysfunction [2][3]. Therefore, to improve benefits for all patients, i.e., delivering personalized treatment, it would be important to develop fully automatic non-invasive methods to early identify non-responder patients. Several studies [5-10] developed such systems, however despite the promising results, several efforts need to be made to assess the robustness and generalizability of such systems in multi-center external cohorts. Moreover, another limitation of the use of these systems is the fact that they rely on manual segmentation, which is timeconsuming and highly dependent on the reader [9]. This study aims to implement and externally validate a fully automatic AI-based system based on pretreatment MRI to segment and predict resistance to nCRT patients.

2 Materials and Methods

2.1 Database

We designed a multi-center retrospective study enrolling RC patients who underwent multiparametric MRI before nCRT after October 2000. This multi-center retrospective study was approved by the institutional review boards (IRBs) in each institution, with a waiver for the requirement of informed consent, as de-identified data were used. Patients were enrolled from six different institutions:

- Center 01: Candiolo Cancer Institute FPO-IRCCS (FPO) (Candiolo, Italy)
- Center 02: A. O. Ordine Mauriziano (Ospedale Umberto I) (Turin, Italy)
- Center 03: Azienda Ospedaliero-Universitaria Città della Salute e della Scienza di Torino, Ospedale Molinette (Turin, Italy)
- Center 04: Ospedale di Niguarda (Milan, Italy)
- Center 05: Presidio Ospedaliero di Pozzuoli, Ospedale Santa Maria delle Grazie (Pozzuoli, Italy)
- Center 06: Ospedale San Raffaele (OSR) (Milan, Italy)

For this study, we considered only the fast spin-echo T2 weighted (T2w) sequences on the axial plane perpendicular to the longest tumor diameter, acquired according to MRI guidelines for reporting RC staging of each center [11]. Among all centers, we identified three manufacturers: GE Medical System, Philips HealthCare, and Siemens. Table 1 shows the number of sequences for each vendor group, and the T2w parameters for each center.

We divided the sequences according to the center. The Construction set included the sequences acquired by centers 01, 02, 03, and 04, and was divided into the training set (TR) by randomly selecting 70% of patients, while the remaining 30% were included in the internal validation set (IntVAL). The centers 05 and 06 were included in the external validation (ExtVAL).

Table 1: Dataset description according to vendor and T2-weighted parameters for each center.

	Center	GE	Philips	Siemens	**Pixel resolution** (M-IQR) (mm)	**Slice thickness** (M-IQR) (mm)	**FOV** (M-IQR) (mm)
Construction Set	01	82	5	/	0.45 (0.43-0.45)	4.40 (4.00-4.40)	230 (220-230)
	02	/	24	/	0.47 (0.47-0.47)	3.50 (3.50-4.00)	240 (240-240)
	03	/	37	/	0.49 (0.49-0.49)	3.00 (3.00-3.00)	250 (205-250)
	04	2	27	/	0.55 (0.54-0.55)	5.00 (4.00-5.00)	280 (280-280)
Ext VAL	05	/	/	25	0.63 (0.63-0.63)	3.00 (3.00-3.00)	200 (200-200)
	06	/	56	/	0.72 (0.72-0.77)	4.00 (4.00-4.02)	405 (385-405)

[a]ExtVAL: External Validation; FOV: field of view, for all our sequences is a quadratic FOV, same dimension for both x and y axes; M: median value; IQR: interquartile range. NA: not available.

2.2 Reference Standard

To evaluate the performance of the segmentation network, we used masks of the tumor that were manually segmented by different radiologists, one per center, with high experience in reporting MRIs, and then revised by a centralized expert radiologist. For the development of the predictive system, the reference standard was given by the Tumor Regression Grade (TRG), evaluated by experienced pathologists, blinded to clinical information and MRI findings from the resected tumor. The TRG was assessed according to the Mandard classification [12]. These patients were classified into two different classes: responders (R+) if TRG was equal or lower than 3, and resistant (R-) if TRG was equal or greater than 4.

2.3 Data Pre-Processing

2.3.1 Sequence variability reduction

To reduce the impact of the high variability and inhomogeneity of the images on the robustness [13], first, we applied a spatial normalization method to reduce the variability in terms of anatomical structures included in the sequences and T2w characteristics. To this scope, we resampled the images, setting a pixel resolution of 0.47mm, and then resized them to have the same FOV (180x180mm2), obtaining sequences with fixed dimensions of 384x384 pixels.

2.3.2 Automatic tumor segmentation system

All tumors were automatically segmented using an internally developed and validated DL algorithm, based on a fully convolutional network, the U-Net. The network was implemented using a multi-center and multi-vendor database as well [14].

2.3.3 Feature Extraction

Radiomic features were extracted from the automatic segmentation masks using PyRadiomics, an IBSI-compliant open-source platform. The following classes were extracted from the original, wavelet, and Laplacian-filtered images: First Order Statistics; Gray Level Cooccurrence Matrix; Gray Level Dependence Matrix; Gray Level Run Length Matrix; Gray Level Size Zone Matrix; Neighboring Gray Tone Difference Matrix. The discretization was performed considering a fixed bin count of 32 bins, that was chosen after evaluation on pixel intensity ranges. To avoid interpolated isotropic voxels, the feature extraction was performed in 2.5D [15].

2.4 Classifier Development

2.4.1 Feature Selection

To improve the robustness of the predictive model across slightly different segmentation masks, we first identified a sub-group of robust features, using the Intraclass Correlation Coefficient (ICC) [16]. We computed the ICC between the manual and automatic segmentation masks with a volumetric overlap higher than 75%, and we selected only features showing an ICC $\geq$ 0.75, which identifies good reliability between the features [17]. Then, for each feature, we computed the Area Under the ROC Curve (AUC) on the outcome and the Spearman correlation with all the other features. If the correlation was higher than 0.90, we discarded the feature with the lowest AUC with the outcome. Finally, we sorted the selected features according to their AUC, and, starting from the one with the highest AUC, we iteratively added the others one at a time. At each iteration, we evaluated the AUC of the model on TR and IntVAL, and we selected the best number of features based on the point in which the overfitting occurs, i.e., AUC on TR increases while AUC on IntVAL starts decreasing.

2.4.2 Classification algorithm

In this study, we developed a Random Forest (RF), an algorithm that creates a set of decision trees, each trained on a different subset of the data. We set the number of trees equal to 100, and the final prediction of the random forest is obtained by taking the majority vote across all trees.

2.4.3 Statistical Analysis

For the evaluation of the performances of the automatic detection system, we used the Dice Similarity Coefficient (DSC), Precision (Pr), and Recall (Re), three metrics that evaluate overlap, over-segmentation, and under-segmentation, respectively. The Pearson correlation was applied for the evaluation of the correlation between the features. Balanced accuracy (acc), sensitivity, specificity, positive predictive value (PPV), and negative predictive value (NPV) of the predictive system were evaluated on TR, IntVAL, and ExtVAL.

To statistically compare the differences between the validation sets, we performed the chi-squared (comparison of proportions) analysis. All analyses were performed using Python 3.7, Matlab (R2023a), and MedCalc Software Ltd.

3 Results and Discussion

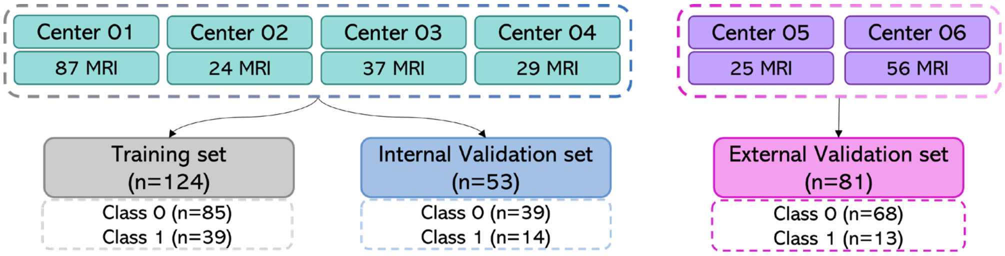

Fig. 1. Database division into Training, Internal and External Validation.

3.1 Database

273 patients were retrospectively collected, having an average age of 64 years (range 34-86). 124 patients were included in the TR, 53 in the IntVAL, and 81 in the ExtVAL (Fig. 1). All sets were highly unbalanced between the two classes. Indeed, the percentage of R- is 31%, 26%, and 16% in TR, IntVAL, and ExtVAL, respectively.

3.2 Automatic detection system

The performances in terms of median and Inter-Quartile Range (IQR) of the DL segmentation network on the enrolled sequences were DSC=0.68 (IQR: 0.57-0.78), Pr=0.55 (IQR: 0.41-0.70), and Re=0.95(IQR:0.87-0.98), as demonstrated in our previous work [14].

3.3 Automatic predictive system

665 radiomic features were initially extracted, and 184 were identified as robust according to the ICC. After the feature selection, we included in the final model 19 features: 4 Original, 6 Laplacian, and 9 Wavelet.

Table 2 shows the performances of the predictive system on TR, IntVAL, and ExtVAL. Performances on the TR were almost perfect, as we could expect from the RF algorithm, however, we obtained good results on both validation sets, meaning that the algorithm is able to generalize on an external dataset. Indeed, the median values are AUC=0.73 vs 0.71, acc=0.70 vs 0.66, and sensitivity of 0.71 vs 0.69 for IntVAL and ExtVAL, respectively. These values, were not statistically different (p-value > 0.05), as also shown in Fig.2.

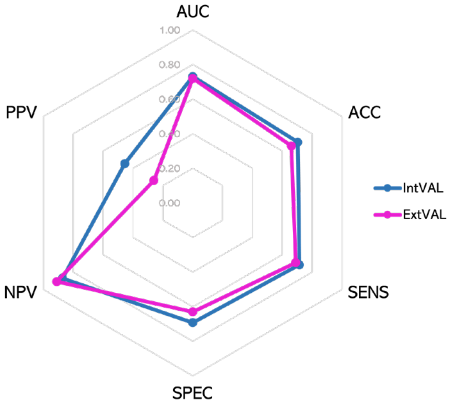

Fig. 2. Performance's radar graph on validation sets.

Table 2: Predictive system performances on the different sets

Set	AUC (95%CI)	Balanced Accuracy [rate] (95%CI)	Sensitivity [rate] (95%CI)	Specificity [rate] (95%CI)	PPV [rate] (95%CI)	NPV [rate] (95%CI)
TR	1.00 (1.00-1.00)	0.99 [123/124] (0.98-1.00)	1.00 [39/39] (1.00-1.00)	0.99 [84/85] (0.97-1.00)	0.98 [39/40] (0.93-1.00)	1.00 [84/84] (1.00-1.00)
IntVAL	0.73 (0.57-0.90)	0.70 [37/53] (0.58-0.83)	0.71 [10/14] (0.48-0.95)	0.69 [27/39] (0.55-0.84)	0.45 [10/22] (0.25-0.66)	0.87 [27/31] (0.75-0.99)
ExtVAL	0.72 (0.55-0.88)	0.66 [52/81] (0.54-0.76)	0.69 [9/13] (0.44-0.94)	0.63 [43/68] (0.52-0.75)	0.26 [9/34] (0.12-0.41)	0.91 [43/47] (0.84-0.99)

[a]Training: TR, Internal Validation: IntVAL, External Validation: ExtVAL, Area Under the Curve: AUC, Positive Predictive Value: PPV, Negative Predictive Value: NPV.

4 Discussion

In this study, we presented a fully automatic AI system based on pretreatment MRI for the prediction of resistance in RC patients, using automatic segmentation. In the ExtVAL, our multi-center system reaches an AUC of 0.71 and acc 0.66. In literature, most studies developed AI systems to evaluate the pathological complete response (TRG=1) of RC patients; to the best of our knowledge, only a few [4][18][19] have attempted the prediction of non-response to nCRT. Even if Liu et al [18] carried out a similar analysis, their results are not generalizable since they were evaluated on a very small single-center cohort (n=26). Zhou et al [4] and Zhang et al [19] developed AI systems integrating multiple information. In particular, Zhou et al. combined several multiparametric MRI sequences, while Zhang et al. integrated the Computer Tomography and clinic-pathological features, yielding very promising results on a single-center validation set: AUC=0.77 and 0.89, respectively. Even if their results should be validated on a multi-center cohort, they demonstrated that the integration of clinical data and/or multiple imaging sequences may provide more useful information to the model for the prediction aim.

Our study has some limitations. First, the dataset is quite unbalanced, especially in the external validation set (68 R+ vs 13 R-), leading to a misinterpretation of the results, especially for AUC and PPV values. At the same time, this imbalance represents the reality across RC patients that occur in hospitals. Second, even if the results achieved by our T2w-based system are in line with the literature, it could be of interest to integrate clinical data and MRI sequences, as suggested by other studies, to improve the performances.

5 Conclusion

We developed and validated on a multi-center and multi-vendor database a fully automatic AI system able to predict resistant RC patients, trying to take a further step towards personalized medicine. In the ExtVAL, our multi-center system reached an AUC of 0.71, acc 0.66, sens 0.69 and spec 0.63. Further analyses are still required to increase performance, allowing the introduction into clinical practice.

Acknowledgment

This research was funded by FPRC 5xmille 2015 Ministero della Salute (STRATEGY) and Fondazione AIRC under 5 per Mille 2018—ID. 21091 program—P.I.Bardelli Alberto, G.L. Regge Daniele.

References

[1] R. L. Siegel, K. D. Miller, H. E. Fuchs, and A. Jemal, 'Cancer statistics, 2022', CA Cancer J Clin, vol. 72, no. 1, pp. 7–33, 2022, doi: 10.3322/caac.21708.

[2] G. Argilés, J. Tabernero, R. Labianca, D. Hochhauser, R. Salazar, T. Iveson, P. Laurent-Puig, P. Quirke, T. Yoshino, J. Taieb, E. Martinelli, D. Arnold, 'Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†', Annals of Oncology, vol. 31, no. 10, pp. 1291–1305, 2020, doi: 10.1016/j.annonc.2020.06.022.

[3] A. B. Benson, Venook, A. P., Al-Hawary, M. M., Azad, N., Chen, Y. J., Ciombor, K. K., Cohen, S., Cooper, H. S., Deming, D., Garrido-Laguna, I., Grem, J. L., Gunn, A., Hecht, J. R., Hoffe, S., Hubbard, J., Hunt, S., Jeck, W., Johung, K. L., Kirilcuk, N., Krishnamurthi, S., … Gurski, L., 'Rectal Cancer, Version 2.2022', JNCCN Journal of the National Comprehensive Cancer Network, vol. 20, no. 10, pp. 1139–1167, 2022, doi: 10.6004/jnccn.2022.0051.

[4] X. Zhou, Y. Yi, Z. Liu, W. Cao, B. Lai, K. Sun, L. Li, Z. Zhou, Y. Feng, J. Tian, 'Radiomics-Based Pretherapeutic Prediction of Non-response to Neoadjuvant Therapy in Locally Advanced Rectal Cancer', Ann Surg Oncol, 2019, doi: 10.1245/s10434-019-07300-3.

[5] L. Shao, Z. Liu, L. Feng, X. Lou, Z. Li, X. Zhang, X. Wan, X. Zhou, K. Sun, D. Zhang, L. Wu, G. Yang, Y. Sun, R. Xu, X. Fan, J. Tian, 'Multiparametric MRI and Whole Slide Image-Based Pretreatment Prediction of Pathological Response to Neoadjuvant Chemoradiotherapy in Rectal Cancer: A Multicenter Radiopathomic Study', Ann Surg Oncol, vol. 27, no. 11, pp. 4296–4306, 2020, doi: 10.1245/s10434-020-08659-4.

[6] Z. Guo Chao, X. Yan Yan, W. Ying Chao, C. Nuo, L. Rui, and W. Xin, 'Value of Pretreatment Inflammation-nutrition Score to Predict Non-response to Neoadjuvant Chemotherapy in Locally Advanced Rectal Cancer *', Biomed Environ Sci, vol. 36, no. 10, pp. 940–948, 2023, doi: 10.3967/bes2023.121.

[7] M. Liu, H. Lv, L. H. Liu, Z. H. Yang, E. H. Jin, and Z. C. Wang, 'Locally advanced rectal cancer: predicting non-responders to neoadjuvant chemoradiotherapy using apparent diffusion coefficient textures', Int J Colorectal Dis, vol. 32, no. 7, pp. 1009–1012, Jul. 2017, doi: 10.1007/s00384-017-2835-3.

[8] J. Song, Y. Yin, H. Wang, Z. Chang, Z. Liu, and L. Cui, 'A review of original articles published in the emerging field of radiomics', European Journal of Radiology, vol. 127. Elsevier Ireland Ltd, Jun. 01, 2020. doi: 10.1016/j.ejrad.2020.108991.

[9] A. Defeudis, S. Mazzetti, J. Panic, M. Micilotta, L. Vassallo, G. Giannetto, M. Gatti, R. Faletti, S. Cirillo, D. Regge, V. Giannini, 'MRI-based radiomics to predict response in locally advanced rectal cancer: comparison of manual and automatic segmentation on external validation in a multicentre study', Eur Radiol Exp, vol. 6, no. 1, Dec. 2022, doi: 10.1186/s41747-022-00272-2.

[10] A. Defeudis, C. De Mattia, F. Rizzetto, F. Calderoni, S. Mazzetti, A. Torresin, A. Vanzulli, D. Regge, V. Giannini, 'Standardization of CT radiomics features for multi-center analysis: Impact of software settings and parameters', Phys Med Biol, vol. 65, no. 19, 2020, doi: 10.1088/1361-6560/ab9f61.

[11] R. G. H. Beets-Tan, Lambregts, D. M. J., Maas, M., Bipat, S., Barbaro, B., Curvo-Semedo, L., Fenlon, H. M., Gollub, M. J., Gourtsoyianni, S., Halligan, S., Hoeffel, C., Kim, S. H., Laghi, A., Maier, A., Rafaelsen, S. R., Stoker, J., Taylor, S. A., Torkzad, M. R., & Blomqvist, L. 'Magnetic resonance imaging for clinical management of rectal

cancer: Updated recommendations from the 2016 European Society of Gastrointestinal and Abdominal Radiology (ESGAR) consensus meeting', Eur Radiol, vol. 28, no. 4, pp. 1465–1475, 2018, doi: 10.1007/s00330-017-5026-2.

[12] A. M. Mandard, F. Dalibard, J. Mandard, J. Marnay, M. Henry-Amar, J. Petiot, A. Roussel, J. Jacob,P. Segol, G. Samama, 'Pathologic assessment of tumor regression after preoperative chemoradiotherapy of esophageal carcinoma. Clinicopathologic correlations', Cancer, vol. 73, no. 11, pp. 2680–2686, 1994, doi: 10.1002/1097-0142(19940601)73:11<2680::aid-cncr2820731105>3.0.co;2-c.

[13] J. Panic, A. Defeudis, G. Balestra, V. Giannini, and S. Rosati, 'Normalization Strategies in Multi-Center Radiomics Abdominal MRI: Systematic Review and Meta-Analyses', IEEE Open J Eng Med Biol, vol. 4, pp. 67–76, 2023, doi: 10.1109/OJEMB.2023.3271455.

[14] J. Panic, A. Defeudis, S. Mazzetti, S. Rosati, G. Giannetto, M. Micilotta, L. Vassallo, M. Gatti, D. Regge, G. Balestra, V. Giannini, 'A fully automatic deep learning algorithm to segment rectal Cancer on MR images: a multi-center study', in Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS, Institute of Electrical and Electronics Engineers Inc., 2022, pp. 5066–5069. doi: 10.1109/EMBC48229.2022.9871326.

[15] L. S. Zwanenburg A, Leger S, Vallières M, 'Image biomarke standardization initiative', arXiv preprint arXiv:1612.07003, 2016, doi: 10.17195/candat.2016.08.1.

[16] A. Zwanenburg, S. Leger, L. Agolli, K. Pilz, E. Troost, C. Richter, S. Löck, 'Assessing robustness of radiomic features by image perturbation', Sci Rep, vol. 9, no. 1, Dec. 2019, doi: 10.1038/s41598-018-36938-4.

[17] T. K. Koo and M. Y. Li, 'A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research', J Chiropr Med, vol. 15, no. 2, pp. 155–163, 2016, doi: 10.1016/j.jcm.2016.02.012.

[18] M. Liu, H. Lv, L. H. Liu, Z. H. Yang, E. H. Jin, and Z. C. Wang, 'Locally advanced rectal cancer: predicting non-responders to neoadjuvant chemoradiotherapy using apparent diffusion coefficient textures', Int J Colorectal Dis, vol. 32, no. 7, pp. 1009–1012, Jul. 2017, doi: 10.1007/s00384-017-2835-3.

[19] Z. Zhang, X. Yi, Q, Pei, Y. Fu, B. Li, H. Liu, Z. Han, C. Chen, P. Pang, H. Lin, G. Gong, H. Yin, H. Zai, B. Chen, 'CT radiomics identifying non-responders to neoadjuvant chemoradiotherapy among patients with locally advanced rectal cancer', Cancer Med, vol. 12, no. 3, pp. 2463–2473, Feb. 2023, doi: 10.1002/cam4.5086.