A Convolutional Neural Network based system for Colorectal cancer segmentation on MRI images

A Convolutional Neural Network based system for Colorectal cancer segmentation on MRI images

Jovana Panic, Arianna Defeudis, Simone Mazzetti, Samanta Rosati, Giuliana Giannetto, Lorenzo Vassallo, Daniele Regge, Gabriella Balestra, *Member, IEEE* and Valentina Giannini

***Abstract*— The aim of the study is to present a new Convolutional Neural Network (CNN) based system for the automatic segmentation of the colorectal cancer. The algorithm implemented consists of several steps: a pre-processing to normalize and highlights the tumoral area, the classification based on CNNs, and a post-processing aimed at reducing false positive elements. The classification is performed using three CNNs: each of them classifies the same regions of interest acquired from three different MR sequences. The final segmentation mask is obtained by a majority voting. Performances were evaluated using a semi-automatic segmentation revised by an experienced radiologist as reference standard. The system obtained Dice Similarity Coefficient (DSC) of 0.60, Precision (Pr) of 0.76 and Recall (Re) of 0.55 on the testing set. After applying the leave-one-out validation, we obtained a median DSC=0.58, Pr=0.74, Re=0.54. The promising results obtained by this system, if validated on a larger dataset, could strongly improve personalized medicine.**

***Clinical Relevance*—to provide reliable segmentation system from which perform radiomics analysis aimed at predicting response to chemotherapy and personalizing treatment.**

I. Introduction

Colorectal cancer (CRC) is a malignant tumor arising from the inner wall of the colon and/or the rectum [1]. Currently, the recommended treatment for CRC is neoadjuvant chemoradiotherapy (nCRT) followed by total mesorectal excision [2]. Despite the advantages showed by nCRT, the response of the patients varies widely, ranging from completely responder (15% to 20% of the cases) to no response or tumor progression [3][4]. To avoid the main drawbacks of nCRT on non-responder patients, e.g. increase of difficulties and delays of surgical resection [5], several efforts have been made toward the implementations of method able to predict the patient response from Magnetic Resonance Imaging (MRI) before treatment [6][7]. Most of these studies, however, rely on manual or semi-automatic segmentation of CRC, which are time-consuming, and highly operator dependent. Therefore, it is of key importance to develop valid techniques for automatic detection of CRC. Inspired by the success of Deep Learning [8][9], new learning-based segmentation methods have been introduced. Jian et al. [10] presented a Fully Connected Network (FCN)-based method for the segmentation of colorectal cancer on MRI images. This study is characterized by a semi-automatic extraction of the Region of Interest (96x96 pixels) around the tumor and the use of a pretrained VGG-16 model [11]. Conversely, Trebeschi et al. [12] developed a completely automatic CNN-based process to segment a single subtype of CRC, that is adenocarcinoma, excluding mucinous cancers [13], using 21x21x3 voxels.

The aim of the study is to present a new approach based on the analysis of small size ROIs for the segmentation of the CRC.

II. Materials and Methods

A. Patients and reference standard

33 patients (22 males and 11 females) with proven locally CRC were included. Among them, 28 had adenocarcinomas and 5 mucinous carcinomas. Before nCRC, all patients underwent multiparametric (mp)MRI, consisting of T2 weighted (T2w) and Diffusion Weighted Imaging (DWI) at the Candiolo Cancer Institute, FPO-IRCCS between October 2010 and February 2016. For all T2w and DWI B1000 sequences an initial mask has been created using a k-means algorithm, and then manually revised by an expert radiologist. These semi-automatic masks have been used as reference standard. 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. Pre-processing

The pre-processing consists of three steps: the evaluation of the Apparent Diffusion Coefficient (ADC), the cropping phase and the extraction of the ROIs.

The first step is performed using the DWI sequences of each patient following the formula:

$$\mathrm{ADC} = -\frac{1}{b}\ln\left(\frac{S_{DWI}}{S_0}\right), \quad (1)$$

where S_0 is the DWI sequence with b-value 0, while S_{DWI} is the DWI sequence with higher b-value (in our case 1000) [14]. Second the region of interest, i.e. the tumoral area, is automatically cropped to reduce the amount of irrelevant information and to minimize the differences among patients, during the normalization [15]. To perform this step, in each slice of the DWI b1000 sequence (fig 1.a), 4 clusters are identified by the Fuzzy c-mean clustering and the one whose centroid's value is between 50th and 85th percentile of the centroids' distribution, is selected (fig 1.b). In this way most of the artefacts are removed. A 2D image is then created, by

J.P., G.G., L.V. are with Candiolo Cancer Institute, FPO -IRCCS, Strada Provinciale 142, km 3.95, Candiolo (TO), Italy. (corresponding author email: jovana.panic@ircc.it)

A.D., S.M., D.R., V.G. are with with Candiolo Cancer Institute, FPO - IRCCS, Strada Provinciale 142, km 3.95, Candiolo (TO), Italy and University of Turin, Department of Surgical Science, Torino, Italy.

G.B. is with Polytechnic of Turin, Department of Electronics and Telecommunications, Torino, Italy.

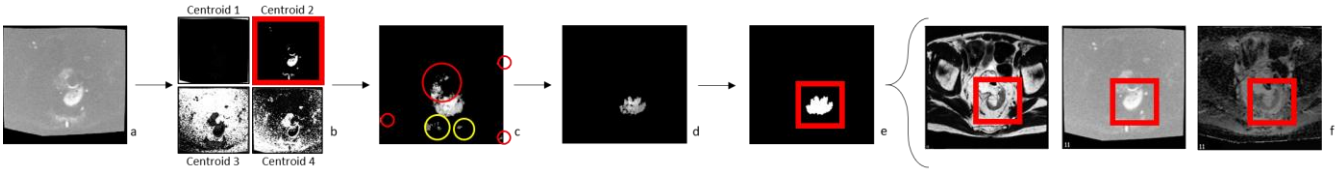

Figure 1: Cropping stage steps: from the B1000 (a), 4 clusters are identified by the Fuzzy c-mean clustering (b) and the one that satisfies the condition (centroid's value between 50th and 85th percentile) is selected for each slice (c). Then all the masks related to the considered cluster are summed in order to define the initial area of interest (c). The identified objects which are close to the borders and in the upper half of the image (red circles), and less frequently detected (yellow circles) are removed (d). Thanks to the created binary mask (e) the box crop is defined (red rectangle). The same box crop is applied on the three sequences (f).

considering all voxels selected in each slice (fig 1.c). To reduce false positive areas, all objects which are close to the borders, in the upper half of the image, and less frequently detected are removed (fig 1.d). Once the final binary mask of the tumoral area is obtained, the coordinates of the crop are evaluated (fig 1.e), and then applied on T2w, DWI, ADC sequences (fig 1.f).

Finally, all cropped images are normalized using the *min max scaling*, as follows:

$$\mathrm{img}_{\mathrm{normalized}} = \frac{\mathrm{img}-\mathrm{min}}{\mathrm{max}-\mathrm{min}}, \qquad (2)$$

where min and max are evaluated considering all the slices of the patient.

The third step consists in the extraction of ROIs that will be used by the CNN. To this scope, all cropped images are divided in regions of dimensions 3x3, 6x6, 9x9 without overlap. Considering the ground truth provided, the ROIs have been labeled in three different classes: the *tumoral ROI* (label 2), the *bright tumoral ROI* (label 1) which is characterized by the 85% of the pixels whose intensity is higher than the median intensity of the sequence, and the *dark non tumoral ROI* (label 0), which instead presents the 85% of the pixels whose intensity is lower than the median.

For a balanced dataset, all the ROIs related to the class 2 have been considered, while the ROIs of class 0 and 1 have been randomly selected to have an equal number of tumoral and non-tumoral samples.

C. Neural Networks

The CNN system proposed is characterized by three CNNs, each trained with ROI coming from a different MR sequence, i.e., T2w, DWI and ADC maps.

All networks implemented share the same architecture (Fig. 2), which consists of two subsequent Convolutional layers followed by Batch Normalization layer, another Convolutional layer and the Fully Connected layer. The output layer is defined by a Dense layer with *Softmax* activation function. The three Convolutional layers are defined by a 3x3 kernel and the *ReLU* activation function [16]. The optimizer is

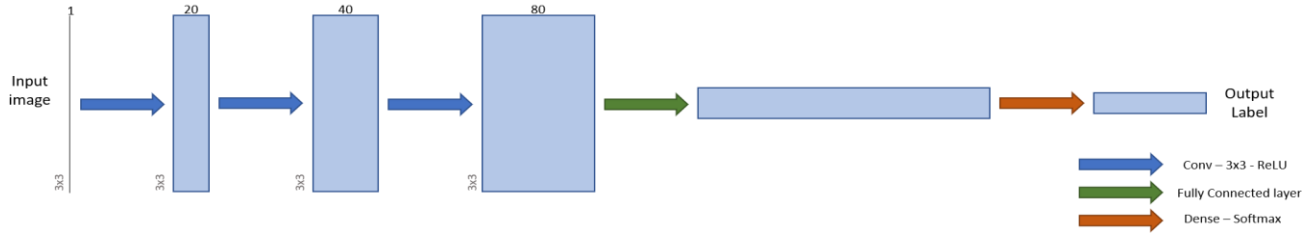

Figure 2. CNN architecture.

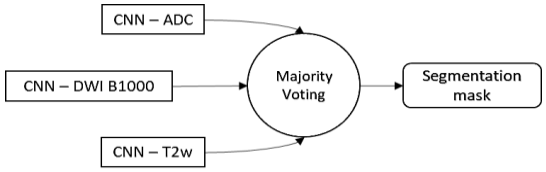

Figure 3. Application of the *majority voting* on the mask of the ADC sequence (CNN - ADC), DWI B 1000 sequence (CNN – DWI B 1000) and T2w sequence (CNN-T2w) thus obtaining the final segmentation mask.

Adam [17], with learning rate 0.001, the loss function used is the Categorical Crossentropy (3)

$$\mathrm{H_p(q)} = -\frac{1}{\mathrm{N}}\sum_{\mathrm{i=1}}^{\mathrm{N}} \mathrm{y_i}\cdot\log\left(\mathrm{p(y_i)}\right), \qquad (3)$$

where $\mathrm{y_i}$ is the label and $\mathrm{p(y_i)}$ is the predicted probability of the sample to belong to the label class. Training epochs was set to 150.

The final segmentation mask (Fig. 3) is obtained using the majority voting on the three CNNs.

In order to analyze the effects of the resolution of the ROIs on the performances, three different CNN systems have been implemented as ROIs classifiers: one for the 3x3 ROIs, one for 6x6 ROIs, and the last one for 9x9 ROIs.

For the creation of the dataset for the CNN systems, we considered three slices per patient, i.e. 99 slices for each sequence, as construction dataset, and the remaining slices as validation dataset. The latter was composed of slices with and without tumor. The construction dataset was subsequently divided into 80% for the training and the 20% for the testing set. All the networks have been implemented on Python 3.7.4 with Keras and Tensorflow libraries [18].

D. Post-processing

The post-processing phase aims to reduce the false positive elements detected by the neural network system, as it is shown in Fig. 4. First, only ROIs classified as tumor after the majority voting are considered (class 2 by at least 2 CNNs). Considering the characteristics of the CRC, all the object detected whose area is next to the border, lower than 100 pixels and spatially connected on less than three slices are removed.

TABLE I. CNN SYSTEMS PERFORMANCES ON THE TEST SET

Variables	CNN 3x3 ROIs				CNN 6x6 ROIs				CNN 9x9 ROIs			
	All mask		*Tumoral object* mask		*All* mask		*Tumoral object* mask		*All* mask		*Tumoral object* mask	
	Mean (std)	*median 25p 75p*	*Mean (std)*	*median 25p 75p*	*Mean (std)*	*median 25p 75p*	*Mean (std)*	*median 25p 75p*	*Mean (std)*	*median 25p 75p*	*Mean (std)*	*median 25p 75p*
DSC	0.55 (0.20)	0.57 0.42 0.70	0.58 (0.19)	0.60 0.47 0.73	0.53 (0.18)	0.55 0.40 0.68	0.55 (0.18)	0.57 0.43 0.68	0.45 (0.19)	0.46 0.32 0.86	0.47 (0.19)	0.48 0.34 0.60
Pr	0.70 (0.22)	0.74 0.57 0.89	0.74 (0.19)	0.76 0.63 0.90	0.70 (0.21)	0.75 0.60 0.85	0.71 (0.19)	0.76 0.61 0.85	0.67 (0.25)	0.75 0.47 0.86	0.71 (0.22)	0.76 0.58 0.87
Re	0.53 (0.24)	0.54 0.35 0.71	0.54 (0.23)	0.55 0.37 0.72	0.50 (0.22)	0.51 0.33 0.67	0.51 (0.22)	0.53 0.35 0.68	0.40 (0.20)	0.38 0.25 0.55	0.41 (0.21)	0.40 0.25 0.56

E. *Statistical analysis*

For this study the main parameters used for the analysis of the performances of the neural networks are the following:

- Dice Similarity Coefficient (DSC)

$$\text{Dice Similarity Coefficient} = \frac{2|MM \cap OM|}{|MM|+|OM|}, \quad (4)$$

where MM is the manual mask, while OM is the obtained mask.

- Precision (Pr)

$$\text{Precision} = \frac{TP}{TP+FP}, \quad (5)$$

where TP is True Positive, while FP is False Positive.

- Recall (Re)

$$\text{Recall} = \frac{TP}{TP+FN}, \quad (6)$$

where FN is False negative. All those parameters provide the information related on how much the segmentations differ, both in shape and misclassified tissues.

The parameters have been evaluated considering both the final segmentation mask and the mask containing only the *tumoral object*, which is the predicted object spatially connected with the manually segmented mask, in order to discard FP object. This operation was performed on both training and testing sets. Then, the *Leave one out* validation method was applied on the CNN system with the highest performances. This method consists on partitioning a sample of data into complementary subsets, where one is used for the training set, and the other for the testing set. In this case all slices of one patient were iteratively used as test set, while the others were used as training set.

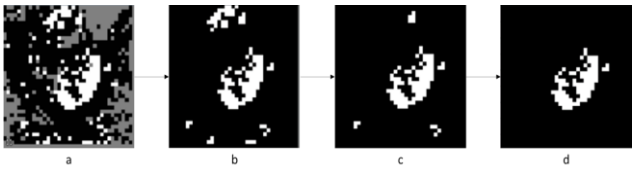

Figure 4. Post-processing phases: From the (a) mask obtained by the system (white ROI = class 2), the binary mask of the tumoral ROIs (b) is obtained; removal of the objects next to the borders (c), with area smaller than 100 pixels which are connected on less than three slices (d).

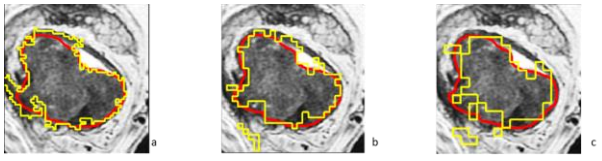

Figure 5. Example of segmentation mask obtained by CNN 3x3 (a), CNN 6x6 (b) and CNN 9x9 (c). The yellow line is the predicted mask, the red line the manual segmentation mask.

III. RESULTS

A. *CNN performances*

Testing set: As shown in table I, the best performance is obtained using the 3x3 ROIs, with DSC=0.60, Pr=0.76 and Re=0.55, when considering the mask of the *tumoral object.* The results of the best classifier related to the training set are close to the testing: in fact, the DSC is 0.64, Pr is 0.76 and Re is 0.60. Analyzing each classifier separately, it is possible to notice that the values of all parameters related to mean, median and percentiles are comparable between the entire mask and the *tumoral object* mask: considering the CNN 3x3, the median values of parameters slightly differ between the entire mask (DSC=0.57, Pr=0.74, Re=0.54) and the tumoral object (DSC=0.60, Pr=0.76, Re=0.55). Considering the training set, the performances of the CNN 3x3 considering the mask of the tumoral object are similar to the one on the testing set: DSC=0.64, Pr=0.76, Re=0.60. Fig. 5 shows the different segmentation masks (yellow line) obtained by the three CNN systems, compared with the ground truth (red line).

Validation: Fig. 6 shows the graph related to the trends of DSC, Pr and Re considering only the tumoral object. It is possible to notice that the values of the parameters considerably differ among some patients, such as 32, 44 and 80, while the overall trend is comparable to the results obtained during the training of the networks (mean values: DSC of 0.58, Pr of 0.74, Re of 0.54). In fig. 7 are shown the differences between the segmentations obtained by the CNN 3x3 system and the reference standard: in patient 32 (fig 7.a), the tumor presents several mucinous areas which are difficult to correctly classify since they are considerably different than adenocarcinomas. The last two cases (patient 44 - fig 7.b and patient 80 - fig. 7.c) show several misclassified areas by the system, which are due to the fact that the pixel intensities range widely among patients.

IV. DISCUSSION

In this study, a new CNN-based system is implemented for the automatic CRC segmentation, with a DSC of 0.58, precision of 0.74 and recall of 0.54 for the validation set.

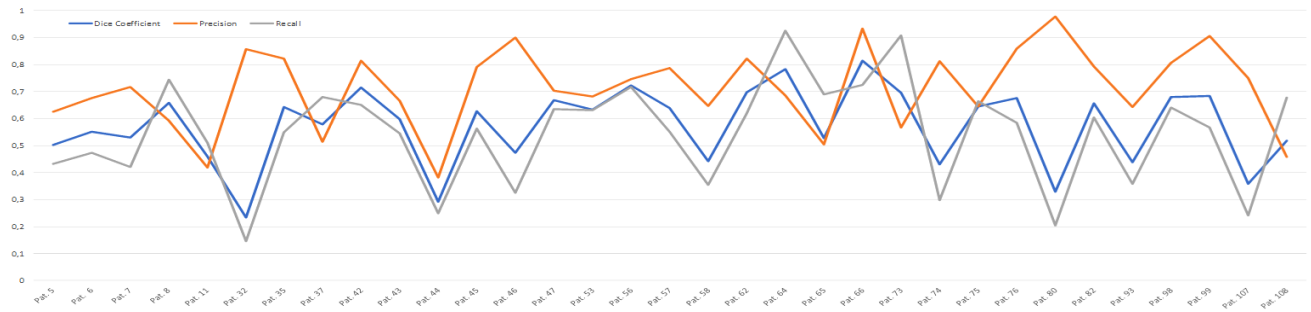

Figure 6. Trends of the dice coefficient, precision and recall for the *Leave-one-out* validation method. The values are related to the tumoral object. The values of the patients are the result of the anonymization applied to the images.

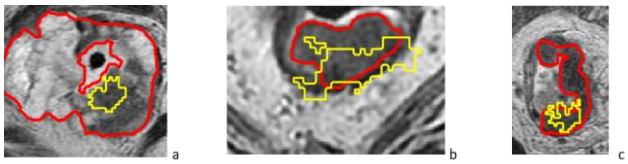

Figure 7. Segmentation masks related to patient 32 (a), 44 (b), and 80 (c), where the red line is the proven segmentation mask, the yellow line is the segmentation mask provided by the CNN 3x3 system.

These results prove that using a small dimensional ROI the classifiers are more precise, since there is less inhomogeneous information. Another important consideration comes from the comparison of the parameter values between the entire mask and the mask of the tumoral object. It is possible to assert that the post-processing stage implemented is a valid false-positive reduction method. Overall, the obtained results are promising, considering the complexity of the information provided by the mpMRI sequences, which can easily lead to wrong predictions and misunderstandings by the neural networks. Previously, Jian et al. [10] developed a method that obtained higher DSC than ours (0.83 vs 0.58), however our method is fully automatic, thus allowing more reproducible and stable results. Conversely, Trebeschi et al. [11] presented an automatic method that reached slightly higher DSC (0.68). However, they only considered the adenocarcinoma subtype of CRC. In the method here developed, we include also the mucinous cases that can considerably affects the performances, since they are characterized by bright tumoral areas, while the adenocarcinomas by dark tumoral areas.

A possible way to improve the performances consists in increasing the amount of data, possibly from different centers to make the system able to generalize. Moreover, it could be possible to define new structures for the CNNs to mimic the radiologists' tumor detection process, using all the sequences provided by the MRI exam.

ACKNOWLEDGMENT

This work was funded by AIRC 5xmille Special Program Molecular Clinical Oncology - Ref. 9970, FPRC 5xmille 2013 Ministero Salute, and FPRC 5xmille 2015 Ministero della Salute (STRATEGY).

REFERENCES

[1] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2019.," *CA. Cancer J. Clin.*, vol. 69, no. 1, pp. 7–34, 2019.

[2] H. J. Schmoll *et al.*, "Esmo consensus guidelines for management of patients with colon and rectal cancer. A personalized approach to clinical decision making," *Ann. Oncol.*, vol. 23, no. 10, pp. 2479–2516, 2012.

[3] X. Yi *et al.*, "MRI-based radiomics predicts tumor response to neoadjuvant chemoradiotherapy in locally advanced rectal cancer," *Front. Oncol.*, vol. 9, no. JUN, pp. 1–10, 2019.

[4] I. J. Park *et al.*, "Neoadjuvant treatment response as an early response indicator for patients with rectal cancer," *J. Clin. Oncol.*, vol. 30, no. 15, pp. 1770–1776, 2012.

[5] Y. Li *et al.*, "A review of neoadjuvant chemoradiotherapy for locally advanced rectal cancer," *Int. J. Biol. Sci.*, vol. 12, no. 8, pp. 1022–1031, 2016.

[6] S. Rosati, C. M. Gianfreda, G. Balestra, V. Giannini, S. Mazzetti, and D. Regge, "Radiomics to Predict Response to Neoadjuvant Chemotherapy in Rectal Cancer: Influence of Simultaneous Feature Selection and Classifier Optimization," in *2018 IEEE Life Sciences Conference (LSC)*, 2018, pp. 65–68.

[7] V. Giannini *et al.*, "Multiparametric magnetic resonance imaging of the prostate with computer-aided detection: experienced observer performance study," *Eur. Radiol.*, vol. 27, no. 10, pp. 4200–4208, 2017.

[8] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *ImageNet Classif. with Deep Convolutional Neural Networks*, pp. 1097–1105, 2012.

[9] R. Yamashita, M. Nishio, R. K. G. Do, and K. Togashi, "Convolutional neural networks: an overview and application in radiology," *Insights Imaging*, vol. 9, no. 4, pp. 611–629, 2018.

[10] J. Jian *et al.*, "Fully convolutional networks (FCNs)-based segmentation method for colorectal tumors on T2-weighted magnetic resonance images," *Australas. Phys. Eng. Sci. Med.*, vol. 41, no. 2, pp. 393–401, 2018.

[11] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," pp. 1–14, 2014.

[12] H. J. W. L. Aerts *et al.*, "Deep Learning for Fully-Automated Localization and Segmentation of Rectal Cancer on Multiparametric MR," *Sci. Rep.*, vol. 7, no. 1, pp. 1–9, 2017.

[13] C. Luo, S. Cen, G. Ding, and W. Wu, "Mucinous colorectal adenocarcinoma: Clinical pathology and treatment options," *Cancer Commun.*, vol. 39, no. 1, pp. 1–13, 2019.

[14] S. Shinya, T. Sasaki, Y. Nakagawa, Z. Guiquing, F. Yamamoto, and Y. Yamashita, "The efficacy of diffusion-weighted imaging for the detection of colorectal cancer," *Hepatogastroenterology.*, 2009.

[15] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern Recognit.*, vol. 38, no. 12, pp. 2270–2285, 2005.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Adv. Neural Inf. Process. Syst.*, vol. 2, pp. 1097–1105, 2012.

[17] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," *3rd Int. Conf. Learn. Represent. ICLR 2015 - Conf. Track Proc.*, pp. 1–15, 2015.

[18] A. Agrawal *et al.*, "TensorFlow Eager: A Multi-Stage, Python-Embedded DSL for Machine Learning," 2019.