CLASSIFICATION OF GOOD AND BAD RESPONDERS IN LOCALLY ADVANCED RECTAL CANCER AFTER NEOADJUVANT RADIO-CHEMOTHERAPY USING RADIOMICS SIGNATURE

Calin Gh. BUZEA [1], Camil C. MIRESTEAN [1], Maricel AGOP[2],

Viorel Puiu PAUN[3*], Dragos T. IANCU [1,4]

We present and discuss here, a study on 22 patients diagnosed with locally advanced rectal cancer proposed for neo-adjuvant radio-chemotherapy, by evaluating their CT-simulator images using radiomics signature. Programs were written in Python, first, to perform the initial stages of image processing and second, to build and save a model using a deep convolutional neural network (CNN) which in the end, proved to be able to evaluate good and bad responders from unseen CT images within a high level of prediction accuracy (over 95.0 %).

Keywords: radiomics, unsupervised learning, rectal cancer, tumor regression grading, radiotherapy.

1. Introduction

Colorectal cancer is the third leading cause of cancer-related mortality in Western countries and in approximately one-third of cases the tumor is localized in the rectum [1]. A standard for the therapeutic scheme in locally advanced rectal cancer (LARC) may be represented by surgical resection, preceded by neoadjuvant chemo-radiotherapy (CRT) or radiotherapy only (RT) [2,3]. Neoadjuvant treatment can lower the risk of local recurrence, decrease the tumor size, and facilitate subsequent successful R0 resection and sphincter – preserving surgery [4]. After CRT, pathological complete response (pCR) is achieved for approximately 15% to 30% of the patients [3,4] and in these cases a wait-and-see strategy is becoming a viable therapeutic option [5].
To improve patient management and/or stratification, it could be advantageous to determine the likelihood of pCR or near pCR before treatment also to allow clinicians to tailor therapy. Importantly, patients predicted as non-responders could benefit from alternative treatments or up-front surgery, avoiding toxicity and side effects of CRT/RT. Also, it might be important to provide physicians with accurate information using noninvasive approaches to identify complete

[1]Regional Institute of Oncology, Henry Mathias Berthelot 2-4, Iasi, Romania
[2] "Gh. Asachi" Technical University, D. Mangeron 65, Iasi, Romania
[3] University POLITEHNICA of Bucharest, Splaiul Independentei 313, Bucharest, Romania
[4] "Gr. T. Popa" University of Medicine and Pharmacy, University 16, Iasi, Romania
* Corresponding author - e-mail: viorel.paun@physics.pub.ro

responders for an alternative surgical treatment such as sphincter-saving local excision.

Previous studies have highlighted several imaging modalities for their capability to distinguish good responders from others for LARC. Examples are FDG-PET [6], T2-weighted magnetic resonance imaging (T2w-MRI) [7], dynamic contrast-enhanced MR (DCE-MRI) [8] and diffusion-weighted imaging (DWI) [9,10].

In the past 10 years, medical digital image analysis has grown dramatically as advancement of the pattern recognition tools and increase of the data collection. Furthermore, medical digital imaging could give a whole picture of the tumor shape, texture and volume, and it is also a noninvasive way to get comprehensive tumor information [11]. A relatively new field of medical investigation, known as Radiomics, is expected to be central to precision medicine due to its ability to gather detailed information describing tumour phenotypes [12].

Radiomics is a promising area of medical research that uses state-of-the-art machine learning (ML) techniques for image characteristics extraction from different types of medical imaging such as computed tomography (CT), nuclear magnetic resonance (MRI) and positron emission tomography (PET) for objective and computable characterization of tumor phenotypes. It was formally introduced by Lambin and collaborators in 2012 [13,14]. It is about extracting and studying a huge amount of radiological imaging features, using either supervised, or unsupervised learning and these data are then used to predict or decode hidden genetic and molecular traits in decision support [11,15-20]. Radiomic features include useful spatial and textural information about black and white images and correlations between pixels in images. Further, these characteristics can be modeled by computerized systems, thus supplementing as an adjuvant instrument the individual diagnosis and treatment guidance [21].

Fig. 1 depicts the processes involved in the Radiomics workflow. The first step involves the acquisition of high quality and standardized imaging, for diagnostic or planning purposes. From this image, the macroscopic tumour is defined, either with an automated segmentation method or alternatively by an experienced radiologist or radiation oncologist. Quantitative imaging features are subsequently extracted from the previously defined tumour region. These features involve descriptors of intensity distribution, spatial relationships between the various intensity levels, texture heterogeneity patterns, descriptors of shape and of the relations of the tumour with the surrounding tissues. The extracted image traits are then subjected to a feature selection procedure. The most informative features are identified based on their independence from other traits, reproducibility and prominence on the data. The selected features are then analyzed for their relationship with treatment outcomes or gene expression. The ultimate goal is to provide accurate risk stratification by incorporating the imaging traits into

predictive models for treatment outcome and to evaluate their added value to commonly used predictors.

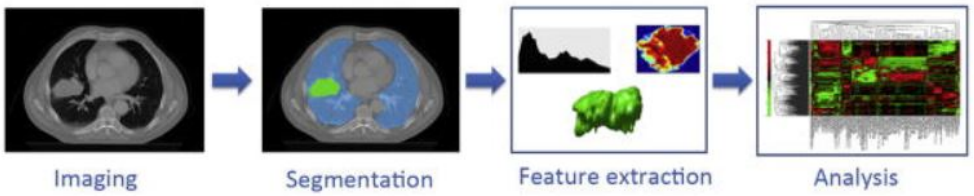


Fig. 1 - typical workflow of Radiomics

In the last few years, the literature on automated classification of images has been extensive, with applications covering different anatomical parts other than colon, such as brain, breast, prostate and lungs. Most of the proposed approaches rely on automated texture analysis, where a limited set of local descriptors are computed from patches of the original input images and then fed into a classifier. Among the most frequently used, statistical features based on grey level co-occurrence matrix (GLCM), local binary patterns (LBP), Gabor and wavelet transforms, etc. The texture descriptors, eventually encoded into a compact dictionary of visual words, are used as input of machine learning techniques such as Support Vector Machines (SVM), Random Forests or Logistic Regression classifiers [22]. In spite of the good level of accuracy obtained by some of these works, the dependence on a fixed set of handcrafted features is a major limitation to the robustness of the classical texture analysis approaches. First, because it requires a deep knowledge of the image characteristics that are best suited for classification, which is not obvious. Second, because it puts severe constraints to the generalization and transfer capabilities of the proposed classifiers, especially in presence of inter-dataset variability.

As an answer to such limitations, in the recent years the use of deep learning (DL) architectures, and more specifically Convolutional Neural Networks (CNNs), has become a major trend [23,24]. In CNNs a number of convolutional and pooling layers learns by backpropagation a set of features that are best for classification, thus avoiding the extraction of hand-crafted texture descriptors. Nonetheless, the necessity of training the networks with a huge number of independent histological samples is still an open issue, which limits the usability of the approach in the everyday clinical setting.

Radiomics has been extensively studied in oncology, with a substantial contribution from the quantitative imaging network (QIN) and the National Cancer Institute (NCI) [11]. There are studies and reports, especially for breast cancer [25], glioblastoma [18], head and neck cancers [26], lung cancer [14], esophageal cancer [27], prostate cancer [28] and rectal cancer [29]. In addition, radiomics was also used in dermatological studies [30].

All computer programs in this work were written in Python, an interpreted, object-oriented, high-level programming language with dynamic semantics. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms, and can be freely distributed.

In this paper, we evaluate a deep learning convolutional neural network (CNN) approach, which trains, validates and saves a classifier for predicting tumor regression grading of rectal tumors after neoadjuvant chemoradiotherapy. The computer program works by extracting characteristic features from the 320 CT - simulator images of a number of 22 patients diagnosed with LARC proposed for neo-adjuvant radio-chemotherapy. The images are first converted from DICOM to JPEG format and a square region of interest (ROI) is defined in each of the CT-simulator pictures, for each patient, to delimitate the image of interest from the noisy background. Note the set of images used for each patient were previously contoured by a highly trained radiation oncologist as defining the tumor. We find that the saved model, even using a simple CNN architecture, succeeds to evaluate good and bad responders (identified using Dworak tumor regression grade (TRG) system obtained after surgery of these patients) from new CT- simulator images within a high degree of classification accuracy. The results are encouraging, suggesting the wealth of imaging radiomics should be further explored to help tailoring the treatment into the era of personalized medicine.

2. Materials and Methods

Twenty-two patients with LARC treated between 13.07.2015 and 22.07.2016 with photon radiotherapy were included in the study. A 8 - 12 weeks resting interval is suggested before surgery to get a maximum effect of radio-chemotherapy.

Treatment was planned using CT-based simulation with patient lying in supine position. Image fusion between the diagnostic MRI and the planning CT images was made. Target volumes and radiosensitive organs at risk were delineated on images obtained from CT simulation fused with diagnostic imaging with contrast agent and the treatment plan was created and delivered by 3D conformal technique.

All patients benefited from total excision of mesorectum and Dworak's degree of regression was evaluated on the pathological sample.

For each of the 22 patients we extracted from the CT simulator DICOM images only those contoured by an experienced radiation oncologist as defining the PTV-T (target volume of the primary rectal tumor) and PTV-N (target volume of lymph nodes) (Phase I) (see Table 1) where :

$$S1=(x1-x2)/Step+1; \quad S2=(x2-x3)/Step+S1 \qquad (1)$$

are the first and the last slide contoured, respectively, x4, x3, x2, x1 the coordinates of the slide on the Oy axis, along the patient, *Step* is the step of the slices taken within the CT simulator.

Commonly used tumor regression grading (TRG) for rectal cancer is the Dworak grading system (which is used here too). TRGs provides important prognostic information since complete or subtotal tumor regression has shown to be associated with better patient's outcome. The prognostic value of TRG may even exceed those of currently used staging systems (e.g., TNM staging) for tumors treated by neoadjuvant therapy.

Table 1

Number of CT simulator slides extracted for each patient, the associated Dworak index and their groups (G =good or B = bad responders)

Patient number	Step	x4	x3	x2	x1	S1	S2	Number of Slides	TRG (Dworak)	Responders (good/bad)
1	0,3	-18,9	-7,2	13,8	24,9	38	108	71	2	B
2	0,3	-19,25	-2,75	15,85	21,25	19	81	63	2	B
3	0,3	-15,2	-3,2	9,7	22,6	44	87	44	1	B
4	0,3	-11,4	-6	10,5	21	36	91	56	1	B
5	0,3	-18,1	-7,3	6,8	17,6	37	84	48	3	G
6	0,3	-16,2	-7,2	10,5	24,3	47	106	60	1	B
7	0,3	-12	-4,2	11,7	25,2	46	99	54	3	G
8	0,3	-18,9	-9	9,3	24,6	52	113	62	1	B
9	0,3	-15,1	-3,1	11	23,9	44	91	48	3	G
10	0,3	-12,3	-0,9	15	27,3	42	95	54	2	B
11	0,3	-12,25	-1,45	12,65	27,05	49	96	48	1	B
12	0,5	-16,85	-8,85	15,15	29,65	30	78	49	2	B
13	0,3	-15,75	-5,85	15,15	28,05	44	114	71	1	B
14	0,5	-14,4	-1,4	14,6	20,6	13	45	33	4	G
15	0,3	-14,1	-4,2	17,1	24,9	27	98	72	2	B
16	0,3	-15,9	-4,8	9,6	21,3	40	88	49	1	B
17	0,5	-16,35	-3,35	16,65	24,15	16	56	41	2	B
18	0,3	-21	-9,9	5,4	15,9	36	87	52	4	G
19	0,3	-12,45	-10,05	12,45	21,75	32	107	76	progression	B
20	0,3	-18	-8,7	5,4	18,6	45	92	48	2	B
21	0,3	-16,65	-4,65	13,35	23,25	34	94	61	2	B
22	0,3	-18,9	-7,2	8,7	21,9	45	98	54	3	G

Computer Program #1 (CP1) was used to convert the DICOM images selected as mentioned above, into JPEG images. Furthermore, a square region of interest (ROI) (see Fig. 2) was created for each of the CT images, for all patients, to delimitate the image of mesorectum from the noisy background. This program also generated images with the same ROIs' sizes (256 x 256 pixels) and excluded all noises from the acquired digital images. For the main Computer Program #2 (CP2) to be able to process these images, the ROIs were placed onto a black background. These images were stored as JPEG files for further image processing and complex measurements with dedicated software tools. Also, CP1 stored the features and labels into two pickle files "X_rect.pickle" and "y_rect.pickle", for the main program CP2 to be able to load and further process.

Using Dworak system, *good responders* were defined as Dworak TRG3+TRG4; *bad responders* were defined as Dworak TRG2+TRG1. The two groups (*good responders* versus *bad responders*) were defined to evaluate outcome results. CT images from the two groups were stored into two different subfolders: 160 images in GOOD, and 160 images in BAD. The program creates the DATADIR folder from the 320 images stored by us into the subfolders (BAD, GOOD).

Fig. 2 – Example of ROI image of mesorectum

A Convolutional Neural Network (CNN) is made up of multiple locally connected trainable stages, piled one after the other, with two or more fully-connected layers as the last step. The first part of the network is devoted to learning the image representation, with successive layers learning features at a progressively increasing level of abstraction, while the last fully-connected part is devoted to classification and acts like a traditional multilayer perceptron. From a computational point of view, a CNN architecture is characterized by two main types of building blocks:

(i) Convolutional (CONV) blocks, that perform a 2D convolution operation (i.e. kernel filtering) on the input image and apply a non-linear transfer

function, such as Rectified Linear Unit (ReLU). Based on the trainable parameters of the kernels, the stage detects different types of local patterns on the input image.

(ii) Pooling (POOL) blocks that perform a non-linear down-sampling of the input (e.g. by applying a max function). This has the double effect of reducing the amount of parameters of the network to control overfitting and of making the image representation (i.e. the local pattern descriptors learnt by the network) spatially invariant.

The number of CONV and POOL blocks (i.e. the depth) of the network is directly related to the level of detail that can be achieved in the hierarchical representation of the image. Nonetheless, a higher depth also translates into a higher number of parameters, and hence on a higher computational cost.

The training paradigm chosen for the CNN is a classic backpropagation scheme: an iterative process that involves multiple passes of the whole input dataset until the model converges. At each training step, the whole dataset flows from the first to the last layer in order to compute a classification error, quantified by a loss function. Such error flows backward through the net, and at each training step the model parameters (i.e. the network weights) are tuned in the direction that minimizes the classification error on the training data.

As a trade-off between representation capabilities and computational costs, in our work we used a simple CNN model, which is represented in **Table 2**. It adopts a very simple architecture, based on piling up only 3x3 convolution and 2x2 pooling blocks. More specifically, the model consists of 2 CONV layers that can be conceptually grouped into 2 macro-blocks ending with one POOL layer each, and of a final 1-layered fully-connected (FC) stage. Nonlinearities are all based on ReLU, except for the fully-connected layer (FC) that has a sigmoid activation function. The convolution stride and the padding are fixed to 1 pixel and the maxpooling stride to 2. The net was built within Keras framework [31] and trained with a backpropagation paradigm. More specifically, we applied an ADAM gradient descent, as iterative optimization algorithm to minimize the categorical cross-entropy function between the two classes of interest (GOOD, BAD). To monitor the training and optimize the choice of hyper-parameters of the net, we used 30% of the training set as validation data i.e. from the total of 320 samples used, we train on 224 and validate on 96. This subset is completely independent from the images used for testing purposes, and was solely used to compute the validation accuracy metric upon which the training process is optimized. Based on validation, we selected a learning rate (LR) of 0.001, a momentum (M) of 0.9 and a batch size (BS) of 32 images.

The CNN was trained for 20 epochs on our colorectal cancer training dataset, which lasted 2min26sec on our computer : system manufacturer: LENOVO; model: 20354; processor: Intel(R) Core(TM) i7-4510U CPU @

2.00GHz (4 CPUs), ~2.6GHz; memory: 8192MB RAM; BIOS: InsydeH2O Version 03.73.069BCN26W; card name: NVIDIA GeForce 840M; Operating System: Windows 10.

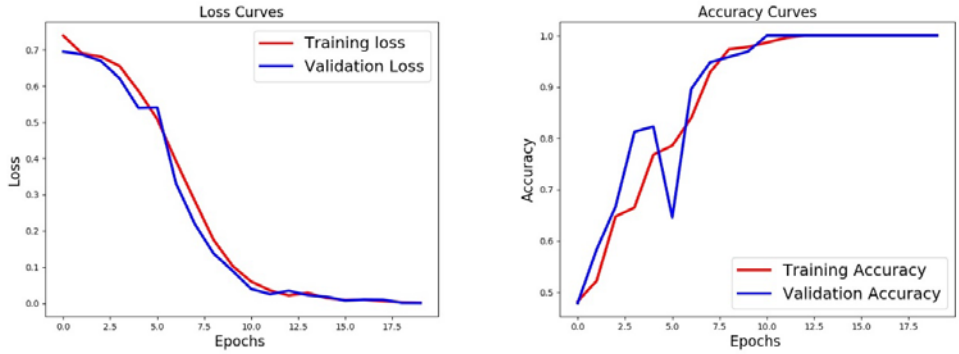
Fig. 3 shows the loss (a) and accuracy (b) curves on both the training and validation datasets. From the graphs of Fig. 3 we can derive the following observations: (i) The model seems to converge quite quickly. Indeed, while training accuracy is still increasing, the value of validation accuracy saturates within 12 epochs. (ii) The decay speed of the validation loss curve indicates that the learning rate is appropriate. (iii) The similarity of validation and training accuracies after about 10 epochs reasonably rules out overfitting.

Table 2 – used CNN architecture

Layer (type)	Output Shape	Param #
conv2d (Conv2D)	(None, 98, 98, 64)	640
activation (Activation)	(None, 98, 98, 64)	0
max_pooling2d (MaxPooling2D)	(None, 49, 49, 64)	0
conv2d_1 (Conv2D)	(None, 47, 47, 64)	36928
activation_1 (Activation)	(None, 47, 47, 64)	0
max_pooling2d_1 (MaxPooling2D)	(None, 23, 23, 64)	0
flatten (Flatten)	(None, 33856)	0
dense (Dense)	(None, 64)	2166848
activation_2 (Activation)	(None, 64)	0
dense_1 (Dense)	(None, 1)	65
activation_3 (Activation)	(None, 1)	0

Total params: 2,204,481
Trainable params: 2,204,481
Non-trainable params: 0

Train on 224 samples, validate on 96 samples



a) b)

Fig. 3 – loss and validation curves for the training of the CNN classifier.

3. Results and Discussion

In a study from 2016, in a systematic analysis of multiparametric MR imaging features in predicting pathologic response after preoperative chemoradiation therapy (CRT) for locally advanced rectal cancer (LARC), authors Ke Nie et al, enrolled forty-eight consecutive patients, and built models with improved predictive value over conventional imaging metrics. For each patient, a total of 103 imaging features were extracted and analyzed using both volume averaged and voxelized methods. Artificial neural network with 4-fold validation technique was used to select the best predictor sets to classify different response groups and the predictive performance was calculated using receiver operating characteristic (ROC) curves. The conventional volume-averaged analysis could provide an area under ROC curve (AUC) ranging from 0.54 to 0.73 in predicting pCR. Moreover, if the models were replaced by voxelized heterogeneity analysis, the prediction accuracy measured by AUC improved to 0.71–0.79 and combining all information together, the AUC could be further improved to 0.84 for pCR and 0.89 for GR prediction, respectively. [32]

Zhenyu Liu et al in 2017 enrolled 222 patients (152 in the primary cohort and 70 in the validation cohort) with clinicopathologically confirmed LARC who received chemoradiotherapy before surgery and extracted from T2- weighted and diffusion-weighted imaging, 2252 radiomic features before and after treatment. The radiomics signature comprised 30 selected features and showed good discrimination performance in both the primary and validation cohorts. The individualized radiomics model, which incorporated the radiomics signature and tumor length, also showed good discrimination, with an area under the receiver operating characteristic curve of 0.9756 (95% confidence interval, 0.9185–0.9711) in the validation cohort, and good calibration. Using pre- and

posttreatment MRI data, they developed a radiomics model with excellent performance for individualized, noninvasive prediction of pCR. [33]

In a 2018 study on ninety-five patients, 49 males (52%) and 46 females (48%), with twenty-two patients (23%) having pathologic complete response after chemoradiation, Jean-Emmanuel Bibault et al extracted one thousand six hundred eighty-three radiomics features for the tumor volume from the treatment planning CT Scan. A Deep Neural Network (DNN) was created to predict complete response, as a methodological proof-of-principle. The results were compared to a baseline Linear Regression model using only the TNM stage as a predictor and a second model created with Support Vector Machine on the same features used in the DNN. The DNN predicted complete response with an 80% accuracy, which was better than the Linear Regression model (69.5%) and the SVM model (71.58%). This model correctly predicted complete response after neo-adjuvant rectal chemoradiotherapy in 80% of the patients of this multicenter cohort. [34]

In this paper, we build a computer program in Python using the Keras framework based on a CNN simple architecture to develop a classifier for tumor regression grading of rectal tumors after neoadjuvant chemoradiotherapy. Using features extracted from a set of CT for radiotherapy planning images, used for training and validation, this model succeeded to evaluate good and bad responders from new CT for radiotherapy planning images with a loss function val_loss: 0.1537 and overall accuracy val_acc: 0.9688. Notice also, the results we found are in the range of those presented in the literature 69.5 - 96%.

In Machine Learning, performance measurement is an essential task. So when it comes to a classification problem, we can count on an AUC - ROC Curve. When we need to check or visualize the performance of the multi - class classification problem, we use AUC (Area Under The Curve) ROC (Receiver Operating Characteristics) curve. It is one of the most important evaluation metrics for checking any classification model's performance. It is also written as AUROC (Area Under the Receiver Operating Characteristics). Generally speaking, ROC describes the discriminative power of a classifier independent of class distribution and unequal prediction error costs (false positive and false negative cost). ROC is a probability curve and AUC represents degree or measure of separability. It tells us how much the model is capable of distinguishing between classes. Higher the AUC, better the model is at predicting 0s as 0s and 1s as 1s. By analogy, higher the AUC, better the model is at distinguishing between patients with disease and no disease. The ROC curve is plotted with true positive rate (TPR) against the false positive rate (FPR) where TPR is on y-axis and FPR is on the x-axis. The question it answers is the following: "When it is actually the negative result, how often does it predict incorrectly?"

We give in Figure 4 the AUROC for our problem. With a value of 0.998 for the AUC we are sure the model distinguishes very well between patients with GOOD and BAD responders.

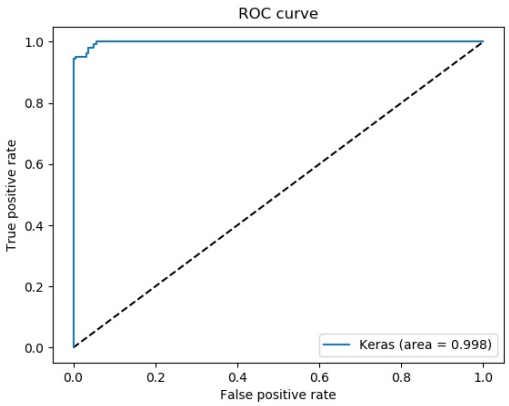
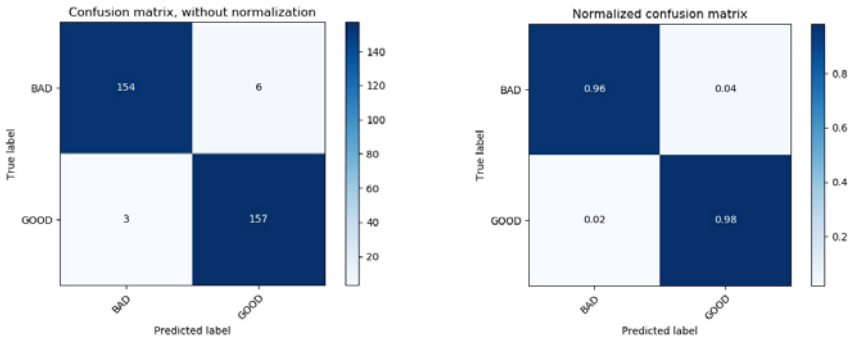


Fig. 4 - Receiver Operating Characteristics curve for the trained classifier

A **confusion matrix** (in unsupervised learning usually called **matching matrix**) is a table that is often used to **describe the performance of a classification model** (or "classifier") on a set of test data for which the true values are known. The confusion matrix itself is relatively simple to understand, but the related terminology can be confusing. In figure 5 we print the confusion matrix for our model.



a) b)

Fig. 5 – confusion matrix without normalization a) and normalized b) for the model classifier trained in this study.

As can be seen from the results in Fig. 5 the expected values are represented by rows and predicted values by columns. The diagonal represents the elements where the predicted values where equal to the expected values, and the off-diagonal values represent the elements where the classifier got the prediction wrong. The higher the proportion of values on the diagonal of the matrix in relation to values off of the diagonal, the better the classifier is. High number of

values off the diagonal indicate problem areas. In the model developed in this paper the classifier literally confused BAD with GOOD responders in 6 out of 160 instances, and GOOD with BAD responders in 3 out of 160 instances. Out of 320 images, the classifier predicted correctly 311 of them.

In what follows TP stands for true positive, TN for true negative, FP for false positive and FN for false negative. From the matrix above if we calculate the "accuracy" defined as the ratio (TP+TN)/(TP+TN+FP+FN), we obtain again that 97% of the predicted outputs were correctly classified. Precision calculated as the ratio TP/(TP+FP) is 0.963 and it answers the question: "When it predicts the positive result, how often is it correct?" and it is usually used when the goal is to *limit the number of FP*. Recall, or TPR defined as the ratio TP/(TP+FN) is 0.981 and it answers the question: "When it is actually the positive result, how often does it predict correctly? Recall, usually is used when the goal is to *limit the number of FN*.

4. Conclusions

According to the results of this study, in practice, we can ultimately feed diagnostic, anatomic and metabolic (CT, RMN, PET CT) and/or CT simulator images of patients with rectal tumors to a specially built computer program/algorithm and the resulted classifier could predict, with a quite high precision rate, the tumor regression grading of these patients, after neoadjuvant chemoradiotherapy.

Basically, the possibility of predicting a response to the treatment based on images correlated to clinical, biological and pathological factors (e.g., tumor stage, histology, lymphovascular invasion) can lead to a stratification of patients in order to intensify or lessen the neoadjuvant therapy. Moreover, the system can help medical experts in completing a superior assessment of medical pictures, by indicating out patterns related to the disease.

REFERENCES

[1] *Siegel R, Miller KD, Jemal A.* Cancer statistics, 2017. CA Cancer J Clin. 2017;67:7–30.

[2] NCCN Clinical Practice Guidelines in Oncology: Rectal Cancer, Version 1.2016. (2016). NCCN.org. Accessed 11 Jan 2019.

[3] *Sauer R, Becker H, Hohenberger W, et al.* Preoperative versus postoperative chemoradiotherapy for rectal cancer. N Engl J Med. 2004;351(17):1731–40.

[4] *Li Y, Wang J, Ma X, Tan L, Yan Y, Xue C, et al.* A review of neoadjuvant chemoradiotherapy for locally advanced rectal cancer. Int J Biol Sci. 2016;12(8):1022–31.

[5] *Maas M, Nelemans PJ, Valentini V, et al.* Long-term outcome in patients with a pathological complete response after chemoradiation for rectal cancer: a pooled analysis of individual patient data. Lancet Oncol. 2010;11:835–44.

[6] *Janssen MH, Ollers MC, Riedl RG, van den Bogaard J, Buijsen J, van Stiphout RG, et al.* Accurate prediction of pathological rectal tumor response after two weeks of preoperative

radiochemotherapy using (18)F-fluorodeoxyglucose-positron emission tomography-computed tomography imaging. Int J Radiat Oncol Biol Phys 2010;77:392–9.

[7] *Engelen SM, Beets-Tan RG, Lahaye MJ, Lammering G, Jansen RL, van Dam RM, et al.* MRI after chemoradiotherapy of rectal cancer: a useful tool to select patients for local excision. Dis Colon Rectum 2010;53:979–86.

[8] *Intven M, Reerink O, Philippens ME*. Dynamic contrast enhanced MR imaging for rectal cancer response assessment after neo-adjuvant chemoradiation. J Magn Reson Imaging 2015;41:1646–53.

[9] *Lambregts DM, Vandecaveye V, Barbaro B, Bakers FC, Lambrecht M, Maas M, et al*. Diffusion weighted MRI for selection of complete responders after chemoradiation for locally advanced rectal cancer: a multicenter study. Ann Surg Oncol 2011;18:2224–31.

[10] *Intven M, Reerink O, Philippens ME*. Diffusion-weighted MRI in locally advanced rectal cancer: pathological response prediction after neoadjuvant radiochemotherapy. Strahlenther Onkol 2012;189:117–22.

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

[12] *Chen B, Zhang R, Gan Y, et al*. Development and clinical application of radiomics in lung cancer. Radiat Oncol 2017;12:154.

[13] *Lambin P, Rios-Velazquez E, Leijenaar R, et al.,* 2012. Radiomics: extracting more information from medical images using advanced feature analysis. *Eur J Cancer*, 48(4):441-446.

[14] *Scrivener M, de Jong EEC, van Timmeren JE, et al*., 2016. Radiomics applied to lung cancer: a review. *Transl Cancer Res*, 5(4):398-409.

[15] *Cook GJR, Siddique M, Taylor BP, et al*., 2014. Radiomics in PET: principles and applications. *Clin Transl Imaging*, 2(3):269-276.

[16*] Kumar V, Gu Y, Basu S, et al*., 2012. Radiomics: the process and the challenges. *Magn Reson Imaging*, 30(9):1234-1248.

[17] *Court LE, Fave X, Mackin D, et al*., 2016. Computational resources for radiomics. *Transl Cancer Res*, 5(4): 340-348.

[18] *Narang S, Lehrer M, Yang D, et al.,* 2016. Radiomics in glioblastoma: current status, challenges and potential opportunities. *Transl Cancer Res*, 5(4):383-397.

[19] *Yip SSF*, Aerts HJWL, 2016. Applications and limitations of radiomics. *Phys Med Biol*, 61(13):R150-R166.

[20] *Sala E, Mema E, Himoto Y, et al.,* 2017. Unravelling tumour heterogeneity using next-generation imaging: radiomics, radiogenomics, and habitat imaging. *Clin Radiol*, 72(1):3-10.

[21] *Parekh V, Jacobs MA*, 2016. Radiomics: a new application from established techniques. *Expert Rev Precis Med Drug Dev*, 1(2):207-226.

[22] *Di Cataldo, S. and Ficarra, E.* (2017). Mining textural knowledge in biological images: Applications, methods and trends. Computational and Structural Biotechnology Journal, 15:56 – 67.

[23] *Janowczyk, A. and Madabhushi, A.* (2016). Deep learning for digital pathology image analysis: A comprehensive tutorial with selected use cases. Journal of Pathology Informatics, 7(1):29.

[24] *Korbar, B., Olofson, A. M., Miraflor, A. P., Nicka, C. M., Suriawinata, M. A., Torresani, L., Suriawinata, A. A., and Hassanpour, S.* (2017). Deep learning for classification of colorectal polyps on whole-slide images. Journal of Pathology Informatics, 8:30.

[25] *Wu WM, Parmar C, Grossmann P, et al.,* 2016. Exploratory study to identify radiomics classifiers for lung cancer histology. *Front Oncol*, 6:71.

[26] *Wong AJ, Kanwar A, Mohamed AS, et al.,* 2016. Radiomics in head and neck cancer: from exploration to application. *Transl Cancer Res*, 5(4):371-382.
[27] *Van Rossum PSN, Xu C, Fried DV, et al.*, 2016. The emerging field of radiomics in esophageal cancer: current evidence and future potential. *Transl Cancer Res*, 5(4):410-423.
[28] *Stoyanova R, Takhar M, Tschudi Y, et al.*, 2016. Prostate cancer radiomics and the promise of radiogenomics. *Transl Cancer Res*, 5(4):432-447.
[29] *Dinapoli N, Casa C, Barbaro B, et al.*, 2016. Radiomics for rectal cancer. *Transl Cancer Res*, 5(4):424-431.
[30] *Cho DS, Clausi DA, Wong A,* 2015. Dermal radiomics for melanoma screening. *Vision Lett*, 1(1):23.
[31] *Chollet, F. et al.* (2015). Keras. https://github.com/fchollet/keras.
[32] *Ke Nie, Liming Shi, Qin Chen, Xi Hu, Salma K. Jabbour, Ning Yue, Tianye Niu and Xiaonan Sun,* Rectal Cancer: Assessment of Neoadjuvant Chemoradiation Outcome based on Radiomics of Multiparametric MRI, Clinical Cancer Res; 22(21) November 1, (2016).
[33] *Zhenyu Liu, Xiao-Yan Zhang,Yan-Jie Shi, Lin Wang, Hai-Tao Zhu, Zhenchao Tang, Shuo Wang, Xiao-Ting Li, Jie Tian and Ying-Shi Sun,* Radiomics Analysis for Evaluation of Pathological Complete Response to Neoadjuvant Chemoradiotherapy in Locally Advanced Rectal Cancer, Clinical Cancer Res; 23(23) December 1, 2017
[34] *Jean-Emmanuel Bibault, Philippe Giraud, Martin Housset, Catherine Durdux, Julien Taieb, Anne Berger, Romain Coriat, Stanislas Chaussade, Bertrand Dousset, Bernard Nordlinger and Anita Burgun,* Deep Learning and Radiomics predict complete response after neo-adjuvant chemoradiation for locally advanced rectal cancer, Nature - Scientific Reports (2018) 8: 12611