

ALMA MATER STUDIORUM · UNIVERSITY OF BOLOGNA

School of Science
Department of Physics and Astronomy
Master Degree in Physics

Implementation of an automated pipeline for
predicting the response to neo-adjuvant
chemo-radiotherapy of colorectal cancer

Supervisor:
Prof. Gastone Castellani

Submitted by:
Giuseppe Filitto

Co-supervisor:
Dr. Nico Curti

Academic Year 2020/2021

Abstract

Colorectal cancer is a malignant neoplasm of the large intestine resulting from the uncontrolled proliferation of the cells making up the colorectal tract. Colorectal cancer is the second malignant tumor per number of deaths after the lung cancer and the third for number of new cases after the breast and lung cancer. Risk factors for this kind of cancer include colon polyps, long-standing ulcerative colitis, diabetes of type II and also genetic history (HNPCC or Lynch syndrome). In order to get information about diagnosis, therapeutic effect evaluation on colorectal cancer, radiomic analysis can be performed on radiological images through the application of dedicated radiomic algorithms based on segmentation and features extraction. By segmentation we mean the determination of the regions of interest (ROI) in images that are going to be analyzed. In clinical routines, it is carried with manual or semi-manual techniques by radiologists, but this process is time-consuming, highly operator-dependent and subject to operator expertise. By Radiomic features we meant the features coming from radiographic medical images, which can potentially uncover disease characteristics that fail to be appreciated by the naked eye. The aim of this project is to implement an automated pipeline based on automatic segmentation of T2 weighted Magnetic Resonance (MR) images exploiting Convolutional Neural Networks in order to predict the response to neo-adjuvant chemo-radiotherapy of colorectal cancer using radiomics features.

... To my family and Nicole

Contents

1	Introduction	4
2	Segmentation	7
2.1	Medical Images	7
2.2	Methods review	7
2.2.1	Thresholding	7
2.2.2	Region growing	7
2.2.3	Clustering	7
2.2.4	U-Net	7
3	Radiomics	8
3.1	First Order Features	8
3.2	Second Order Features	8
4	Pipeline	9
4.1	Description	9
4.2	Implementation	9
4.2.1	Pre-processing	9
4.2.2	Training	9
4.2.3	Segmentation outcomes	9
4.2.4	Features Extraction	9
4.2.5	Predictions	9
5	Results	10
5.1	Dataset	10
5.2	Segmentation	10
5.2.1	Training	10
5.2.2	Segmentation outcomes	10

5.3	Features Extraction	10
5.4	Predictions	10
6	Conclusions	11

Introduction

Colorectal cancer is a malignant neoplasm of the large intestine resulting from the uncontrolled proliferation of the cells making up the colorectal tract. Colorectal cancer is the second malignant tumor per number of deaths after the lung cancer and the third for number of new cases after the breast and lung cancer[1]. Among the risk factors for this kind of cancer non hereditary could range from colon polyps to long-standing ulcerative colitis, from Crohn's disease to old age. Also genetic history (HNPCC or Lynch syndrome) and nutritional factors as diabetes II can increase the probability of develop cancer [2]. Preventive measures for colorectal cancer include physical activity, reducing the consumption of processed meat and alcohol, and avoiding smoking[3].

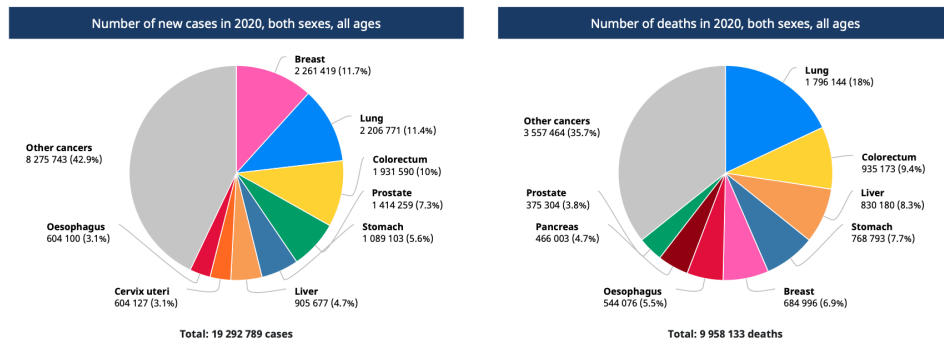


Figure 1.1: World's cancer cases and deaths. From [1]

Screening and diagnosis methods for colorectal cancer can be based on different techniques such as colonoscopy, MRI and CT scans [4]. The gold standard in medical routines is colonoscopy which is an invasive technique. However, non invasive/destructive methods like Magnetic Resonance and Computed Tomography are also use for tumor staging. In particular MRI, thanks to the high spatial resolution is used for pre-operative predictions and for the evaluation of the neoadjuvant therapy in Colorectal cancer [2].

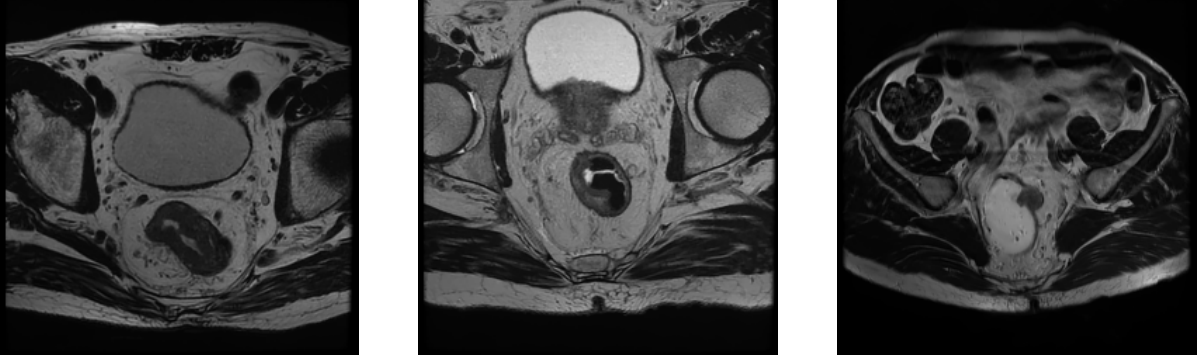


Figure 1.2: T2 weighted axial colorectal cancer MR images. From original dataset

In order to get information about diagnosis, therapeutic effect evaluation on colorectal cancer, radiomic analysis can be performed on radiological images through the application of dedicated radiomic algorithms based on segmentation and features extraction. By segmentation we mean the determination of the regions of interest (ROI) in images that are going to be analyzed. In clinical routines, it is carried with manual or semi-manual techniques by radiologists, but this process is time-consuming, highly operator-dependent and subject to operator expertise [4, 2, 5]. Lately, several deep learning techniques in the medical field have been developed exploiting Convolutional Neural Networks (CNN) for medical images segmentation [5, 4, 2].

As regard the colorectal cancer segmentation on T2-weighted magnetic resonance, one of the most important study of Trebeschi et al. [5], shows how the deep learning approach can perform accurate localization and segmentation of rectal cancer in MR imaging in the majority of patients despite the difficulty in segmenting a quite important Field-of-View (FOV) as in colorectal MR images. In fact, it has been shown how the network performances can vary training the same model with different image annotations or 'labels', taken from different experts, resulting in different Dice-Coefficient (DSC) scores, DSC=0.68 for "reader 1" and DSC=0.70 for "reader 2" respectively. Moreover, not only the labels but also the kind of image, that is to say the kind of tumor can affect the performances of the Networks as resulted in the study of Jovana Panic et al.[6], where including cases of mucinous characterized by bright tumoral areas compared to adenocarcinoma, the Dice-Coefficient results in a lower value, in particular DSC=0.58. An overall study about the comparison of different models, in particular different U-Net architectures, for the segmentation of colorectal cancer on T2-weighted magnetic resonance has been carried by Yi-Jie Huang et al.[7], resulting in a Dice-Coefficient score range between 0.52 and 0.75.

As regard Radiomic features we meant the features coming from radiographic medical images, which can potentially uncover disease characteristics that fail to be appreciated by the naked eye[8].

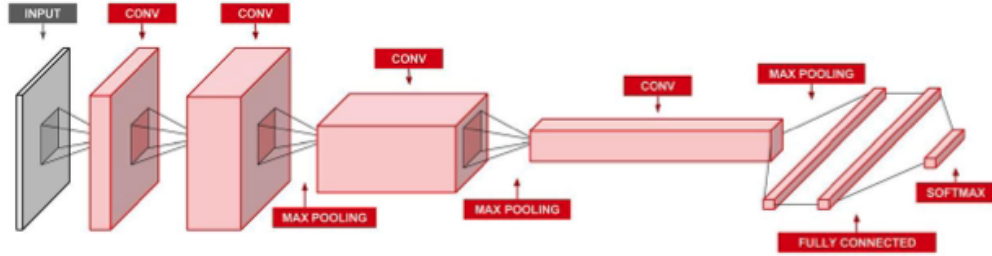


Figure 1.3: Example of CNN. From [5]

Radiomics features can be divided into different classes: shape-size based-features, Haralick's features (Energy, Entropy, Contrast etc...), image voxel descriptors features called GLCM(gray-level co-occurrence matrix) and so on. In order to get useful information about the tumor, after segmentation, different approaches have been exploited in literature. For example, in a study of Soomoro et al. [9], on T2-weighted magnetic resonance images of colorectal cancer, Haralick's features have been exploited to study the cancer evolution after a segmentation and classification of tumors depending on the stage malignancy. Another approach consists in clustering all the extracted features after segmentation, in order to get the most relevant ones, as carried in a study of Parmar et al.[10].

The aim of this project is to implement an automated pipeline based on segmentation and features extraction. As regards the former CNN such as U-Net has been trained from original T2-weighted magnetic resonance images of colorectal cancer dataset, coming from Sant'Orsola Hospital in Bologna, in order to obtain a model to segment colorectal tumor. As regards the latter, all the possible features have been extracted for all the patients' examinations and after a reduction analysis the most important ones have been used, integrating clinical data, to predict the response of the patient to the neo-adjuvant chemo-radiotherapy on colorectal cancer. (provisorio)

Chapter 2

Segmentation

2.1 Medical Images

DICOM

ROI

2.2 Methods review

2.2.1 Thresholding

2.2.2 Region growing

2.2.3 Clustering

2.2.4 U-Net

Chapter 3

Radiomics

3.1 First Order Features

3.2 Second Order Features

Chapter 4

Pipeline

4.1 Description

4.2 Implementation

4.2.1 Pre-processing

4.2.2 Training

4.2.3 Segmentation outcomes

4.2.4 Features Extraction

4.2.5 Predictions

Results

5.1 Dataset

5.2 Segmentation

5.2.1 Training

5.2.2 Segmentation outcomes

5.3 Features Extraction

5.4 Predictions

Chapter 6

Conclusions

Bibliography

- [1] International Agency for Research on Cancer. Rectum-fact-sheet, 2020. URL: <https://gco.iarc.fr/today/data/factsheets/cancers/9-Rectum-fact-sheet.pdf>.
- [2] Jia Cheng Yuan. *Analisi radiomica per predire la risposta alla chemio-radioterapia neoadiuvante nel carcinoma del colon retto*. Tesi di laurea, Alma Mater Studiorum Università di Bologna, 2019/2020.
- [3] Jemal A Siegel RL, Miller KD. Cancer statistics, 2019. *CA Cancer J Clin*, 69(1)(7-34), 2019.
- [4] Jovana Panic. *Colorectal cancer segmentation on MRI images using Convolutional Neural Networks*. Tesi di laurea magistrale, Politecnico di torino, 2018/2019.
- [5] Trebeschi et al. Deep learning for fully-automated localization and segmentation of rectal cancer on multiparametric mr. *Scientific Reports*, 7(1):5301, 2017. doi: 10.1038/s41598-017-05728-9.
- [6] Jovana Panic at al. A convolutional neural network based system for colorectal cancer segmentation on mri images. *Annu Int Conf IEEE Eng Med Biol Soc*, 2020. doi:10.1109/EMBC44109.2020.9175804.
- [7] Yi-Jie Huang, Qi Dou, Zi-Xian Wang, Li-Zhi Liu, Ying Jin, Chao-Feng Li, Lisheng Wang, Hao Chen, and Rui-Hua Xu. 3-d roi-aware u-net for accurate and efficient colorectal tumor segmentation. *IEEE Transactions on Cybernetics*, page 1–12, 2020. URL: <http://dx.doi.org/10.1109/TCYB.2020.2980145>, doi:10.1109/tcyb.2020.2980145.
- [8] Radiomics. URL: <https://en.wikipedia.org/wiki/Radiomics>.

-
- [9] Mumtaz Hussain Soomro, Gaetano Giunta, Andrea Laghi, Damiano Caruso, Maria Ciolina, Cristiano De Marchis, Silvia Conforto, and Maurizio Schmid. Haralick's texture analysis applied to colorectal t2-weighted mri: A preliminary study of significance for cancer evolution. In *2017 13th IASTED International Conference on Biomedical Engineering (BioMed)*, pages 16–19, 2017. doi:10.2316/P.2017.852-019.
- [10] Chintan Parmar, Ralph T. H. Leijenaar, Patrick Grossmann, Emmanuel Rios Velazquez, Johan Bussink, Derek Rietveld, Michelle M. Rietbergen, Benjamin Haibe-Kains, Philippe Lambin, and Hugo J. W. L. Aerts. Radiomic feature clusters and prognostic signatures specific for lung and head & neck cancer. *Scientific Reports*, 5(1):11044, 2015. doi:10.1038/srep11044.