Machine learning applications for Radiomics

Societal impact and valorizations

In this thesis, state of the art computational approaches of mining non-invasive medical imaging data to derive predictive and prognostic inferences in clinical oncology are thoroughly described. Although the cohorts employed in this thesis were focused on specific oncological diseases – lung cancer and head and neck cancer, the described computational approaches harbor broad general applicability. In this section, we discuss the foreseeable societal impact of the presented research along with the policies that could be followed for the enhancement of personalized medicine.

We observe an exponential increase in data and information in this recent "big data" era. Healthcare is no different to this trend. The healthcare institutions of the developed countries archive immense amount of patient data and these practices are being adopted by the highly populated developing countries as well. This increasing amount of patient specific data could transform the routine patient care by tailoring and customizing treatments to individuals. More powerful predictive/prognostic tools are becoming available, enhancing the routine clinical care. This is aligned with the aimed personalized care – bringing the right treatment to the right patient. The use of predictive/prognostic models is increasing, with many online tools already available:

- http://www.predict.nhs.uk/ is an online tool used to help deciding the ideal course of treatment following breast cancer surgery, based on one's cancer histopathology;
- http://www.cancer.gov/bcrisktool/ allows estimating woman's risk of developing invasive breast cancer;
- https://www.mskcc.org/nomograms a long list of nomograms developed for a great variety of solid tumors, and based on hundreds or even thousands of patients, can be used to predict cancer outcomes or assess disease risk;
- http://www.predictcancer.org/ a platform making available published models developed at Maastro for lung, rectum, head and neck, and endometrium cancer

Furthermore, several collaborative consortiums for sharing data have been initiated:

- http://www.cancerimagingarchive.net/ is an online repository consisting imaging and clinical data related to different cancer types. It provides a platform to different healthcare institutions of the world to share data and collaborate research.

- https://www.cancerdata.org/ is an initiative by Maastro clinic for sharing cancer data for research.

We believe that great knowledge gain could be obtained by data sharing. It allows having an access of external cohorts for validation, which is essential for the validity of research methodologies and proposed hypotheses of data driven research like Radiomics. Moreover, data sharing helps in reproducing the successful predictive/prognostic models, which latter could also be improved with additional data or by using better computational approaches.

As far as the research presented in this thesis is concerned, majority of the chapters employed one or more publicly available data. Furthermore, majority of the applied computational approaches and analysis tools are open source.

The first part of the thesis focused on evaluating different publicly available semi-automatic segmentation methods keeping in mind their applicability for raidomics. Two different semi-automatic segmentation approaches available in a free and publicly available 3D-Slicer software tool, were compared to the manual segmentations in terms of their efficiency and accuracy. 3D-Slicer is an open source software platform for medical image informatics, image processing, and three-dimensional visualization. Built over two decades through support from the National Institutes of Health and a worldwide developer community, Slicer brings free, powerful cross-platform processing tools to physicians, researchers, and the general public (https://www.slicer.org). Moreover, all the cohorts used in the first part of the thesis are available at TCIA (http://www.cancerimagingarchive.net) or at cancerdata.org.

The second part of the thesis covered the radiomics based predictive/prognostic analyses, where different machine learning approaches of feature selection and modelling were investigated for the radiomic cohorts of lung and/or head & neck cancer patients. The work presented in chapter 5, was one of the very first investigations investigating more than 400 quantitative imaging features in over a 1000 patients with lung and head and neck cancer, with independent validation datasets. We have shown for the first time the true independent validated impact of radiomics in both lung and head and neck cancer, showing the translational capability of our findings. Additionally, with the publication of the study presented in Chapter 5, we shared imaging datasets to the cancer imaging archive (TCIA) of 511 NSCLC patients of which 89 with gene-expression data. This again, will facilitate reproducibility of research and allow the investigation of other methodologies and hypotheses in the same data. This chapter received attention from the media, including a press release from the Maastricht University Medical Centre[1] and an internal press release in the Dana-Farber Cancer Institute in Boston, USA. This study was also featured in the outlook issue of the prestigious scientific magazine Nature, in a special issue on the outlook of lung cancer[2].

The studies reported in chapters 6-9, were also one of those very first investigations investigating the impact of the advanced machine-learning approaches for radiomic analyses with large independent validation cohorts and two different cancer types.

Predictive and prognostic models with high accuracy, reliability, and efficiency are vital factors driving the success of radiomics. Therefore, it is essential to compare different machine learning models for radiomics based clinical biomarkers. Moreover, we used open source analysis libraries and published parameter settings for the implementation of different machine learning methods enabling the unbiased evaluation of these methods. These articles have received good acceptance from the scientific community resulting in a decent number of citations within a short timespan.

Overall, in order to achieve significantly higher and positive societal impact, collaborative steps in terms of sharing data and knowledge are required. We are actively involved in these initiatives by participating in different data and tools sharing consortium and collaborating with different institutions across the glob. Our work has contributed towards the release of publicly available radiomics toolboxes[3]. By involving more institutions in sharing data and knowledge, we can evaluate the validity of radiomics based predictive/prognostic models, which could enhance the personalized medicine and hence the public health.

(1) http://maastrichtumchoofdsite.createsend1.com/t/ViewEmail/t/CADC497D2FEB9C5D

(2) http://www.nature.com/nature/journal/v513/n7517_supp/full/513S4a.html

(3) https://www.radiomics.io/code.html