Homework - 3

Submission Deadline: April 10

Prostate tumor grade classification

Background

Prostate cancer is the most common cancers found in men and the second most common among the causes of death related to cancer[1]. Accurate risk prediction of prostate cancer can avoid over-treatment and stratify patients for customized treatments.

Clinically, prostate-specific antigen (PSA) testing and random sampling of the prostate gland by transrectal ultrasound (TRUS)-guided biopsy[2] were instrumental in diagnosing clinically significant prostate cancer (csPCa). Gleason score, one of the most common prostate grading systems, can estimate possible threats posed by prostate lesions. This scoring metric allows researchers to classify prostate tumor grades.

Advancements in multi-parametric magnetic resonance imaging (MP-MRI) enable the acquisition of both anatomical and functional information about the prostate, which can increase accuracy of csPCa diagnosis[4]. However, it is essential that MRI image data are effectively analyzed to provide optimal sensitivity and specificity for clinical applications.

Radiomics, the extraction of image features, has been used to build models aiding in lesion detection and cancer prognosis[7]. Machine learning algorithms, especially classification methods, can classify extracted features and make predictions on Gleason scores[8]. Pyradiomics, a free open access Python package, can be used to extract radiomics features in both 2D and 3D[9].

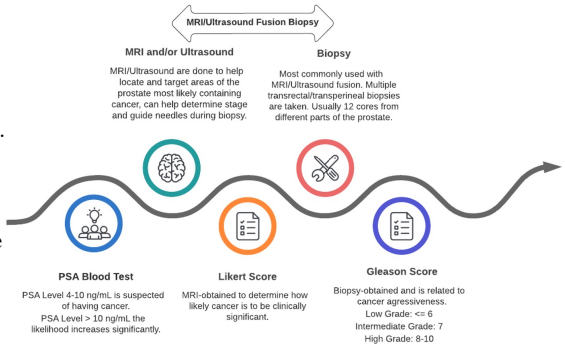
Prostate tumor grade classification

Dataset:

This dataset contains 180 prostate cancer patients who underwent TRUS biopsy and pre-operative MRI. Only t2-weighted MRI images were included and each t2-weighted MRI image corresponds to one tumor ROI, which were also included.

MRI files can be found in "MR data" folder.
Tumor ROI files can be found in "Lesion data" folder.

Clinical data are also provided. First, **PSA levels** were reported for each patient, indicating the likelihood of cancer. Second, each t2-weighed MRI was given a **Likert five-grade scale** with close correspondence to PIRADS version 2, indicating the probability of clinically significant prostate cancer. Finally, each patient underwent TRUS-biopsy and received a **Gleason score** to indicate the aggressiveness of prostate cancer.



Clinical data can be found in clinicaldata.csv

Prostate tumor grade classification

Pyradiomics Installation

1. Install via pip

Pre-built binaries are available on PyPi for installation via pip. For the python versions mentioned below, wheels are automatically generated for each release of PyRadiomics, allowing you to install pyradiomics without having to compile anything. For other python versions, a source distribution is also available, but this requires compiling the C extensions.

- Ensure that you have `python` installed on your machine, version 3.5, 3.6 or 3.7 (64-bits).
- Install PyRadiomics:

```
python -m pip install pyradiomics
```

2. Install via conda

Besides pre-built binaries for PyPi, PyRadiomics is also available on conda cloud. To install PyRadiomics on Conda, run:

```
conda install -c radiomics pyradiomics
```

Detailed Pyradiomics documentations can be found here:
https://pyradiomics.readthedocs.io/en/latest/index.html

3. Install from source

PyRadiomics can also be installed from source code. This allows for the bleeding edge version, but does require you to have a compiler set up for python, as PyRadiomics comes with C extensions for the calculation of texture matrices and some shape features.

- Ensure you have the version control system `git` installed on your machine.
- Ensure that you have `python` installed on your machine, at least version 3.5 (64-bits).
- Clone the repository:

```
git clone git://github.com/Radiomics/pyradiomics
```

- For unix like systems (MacOSX, linux):

```
cd pyradiomics
python -m pip install -r requirements.txt
python setup.py install
```

  - To use your build for interactive use and development:

```
python setup.py build_ext --inplace
```

  - If you don't have sudo/admin rights on your machine, you need to locally install numpy, nose, tqdm, PyWavelets, SimpleITK (specified in requirements.txt). In a bash shell:

```
pip install --user --upgrade pip
export PATH=$HOME/.local/bin:$PATH
pip install --user -r requirements.txt
export PYTHONPATH=$HOME/.local/lib64/python2.7/site-packages
```

Prostate tumor grade classification

In this project, your goal is to **formulate your own research question** regarding how to improve the existing diagnosis of prostate cancer using radiomics approaches. The appropriate label for prostate tumor grade classification is the **Gleason score**. After formulating your research question, you will train a ML model or a DL model to predict the tumor grade based on Gleason score using radiomics features (or "deep" features if you are using DL). You are welcome to include **Likert Scale** or **PSA** levels to your model. Tables on how to interpret Likert scale and Gleason score are provided in the following slides.

Some valid questions can be:

1). Can we implement machine learning classifiers to distinguish prostate tumor grades based on prostate MRI radiomics features?
2). Can we include clinical parameters such as PSA levels or Likert scales to improve the performance of ML models to classify prostate tumor grades?
3). Can we train a deep learning model to stratify prostate tumor grades?

Note, it is easier to classify benign vs Grade 4 or higher prostate cancer.

Prostate tumor grade classification

Likert five-grade scoring system

Likert score	
1	Clinically significant cancer is highly unlikely to be present
2	Clinically significant cancer is unlikely to be present
3	Clinically significant cancer is equivocal
4	Clinically significant cancer is likely to be present
5	Clinically significant cancer is highly likely to be present

Prostate tumor grade classification

Note: In the dataset, benign cancer have a Gleason score of 0.
Also, both grade 2 and 3 are combined in the dataset.

Risk Group*	Grade Group	Gleason Score
Low/Very Low	Grade Group 1	Gleason Score ≤ 6
Intermediate (Favorable/Unfavorable)	Grade Group 2	Gleason Score 7 (3 + 4)
	Grade Group 3	Gleason Score 7 (4 + 3)
High/Very High	Grade Group 4	Gleason Score 8
	Grade Group 5	Gleason Score 9-10

Prostate tumor grade classification

Tasks (Based on the provided Jupyter Notebook):

- Read in and visualize an MRI image (Not lesion ROI!) from one benign tumor, and one malignant tumor (GS >= 6). (Bonus 10 points)
- Formulate your own research question (20 points)
- Extract radiomic features from each patient using Pyradiomics (20 points)
- Perform feature selection using any of the following method using Sklearn:
  - ANOVA, Paired t test, LASSO, SelectK best (20 points)
- Visualize selected features in any feature space (Bonus 10 points)
- Perform classification using any of the following ML methods
  - Random forest, support vector machine, logistic regression (20 points)
- Analyze and explain the performance of your model.
  Provide a brief explanation to your research question based on your findings. (20 points)
- Implement a DL model to perform the same task (Bonus 20 points)

Please make a copy of this file to edit instead of editing the original file.

Homework submission (04/10/2022): Provide me with the links or hand in the downloaded .ipynb files.

Reference:

1. Toivonen J, Montoya Perez I, Movahedi, P, Merisaari H, Pesola M, Taimen P, et al. (2019) Radiomics and machine learning of multisequence multiparametric prostate MRI: Towards improved non-invasive prostate cancer characterization. PLoS ONE 14(7): e0217702. https://doi.org/10.1371/journal.pone.0217702.
2. Hodge KK, McNeal JE, Terris MK, Stamey TA. Random systematic versus directed ultrasound guided transrectal core biopsies of the prostate. J Urol 1989;142(1):71–74; discussion 74–75.
3. Schröder F H et al 2009 Screening and prostate-cancer mortality in a randomized European study New Engl. J. Med. 360 1320–8.
4. Hoeks C M et al 2011 Prostate cancer: multiparametric MR imaging for detection, localization, and staging Radiology 261 46–66
5. Shin T, Smyth TB, Ukimura O et al. Diagnostic accuracy of a five-point Likert scoring system for magnetic resonance imaging (MRI) evaluated according to results of MRI/ultrasonography image-fusion targeted biopsy of the prostate. BJU Int 2018; 121: 77–83
6. Brizmohun Appayya M, Adshead J, Ahmed HU et al. National implementation of multi-parametric magnetic resonance imaging for prostate cancer detection – recommendations from a UK consensus meeting. BJU Int 2018; 122: 13–25
7. Gillies RJ, Kinahan PE, Hricak H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology*. 2016;278(2):563-577. doi:10.1148/radiol.2015151169
8. Nagpal, K., Foote, D., Liu, Y. *et al.* Development and validation of a deep learning algorithm for improving Gleason scoring of prostate cancer. *npj Digit. Med.* **2,** 48 (2019). https://doi.org/10.1038/s41746-019-0112-2
9. Van Griethuysen JJM, Fedorov A, Parmar C, et al. Computational Radiomics System to Decode the Radiographic Phenotype. Cancer Res. 2017;77(21):e104-e107. doi:10.1158/0008-5472.CAN-17-0339