National PhD in Artificial Intelligence
Health and life sciences area
2022 PhD School

September 26-30, 2022
9:00 AM – 7:00 PM

Università Campus Bio-Medico di Roma
Via Giacomo Dina, 36 - 00128 Roma

2022 PhD school competition
Ermanno Cordelli
Università Campus Bio-Medico di Roma

Consiglio Nazionale delle Ricerche

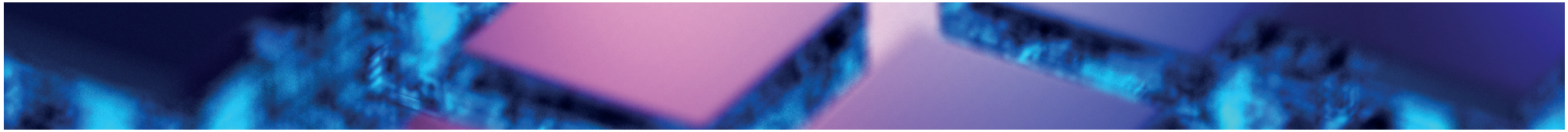
Università di Pisa

Politecnico di Torino

Università Campus Bio-Medico di Roma

Università degli Studi di Napoli Federico II

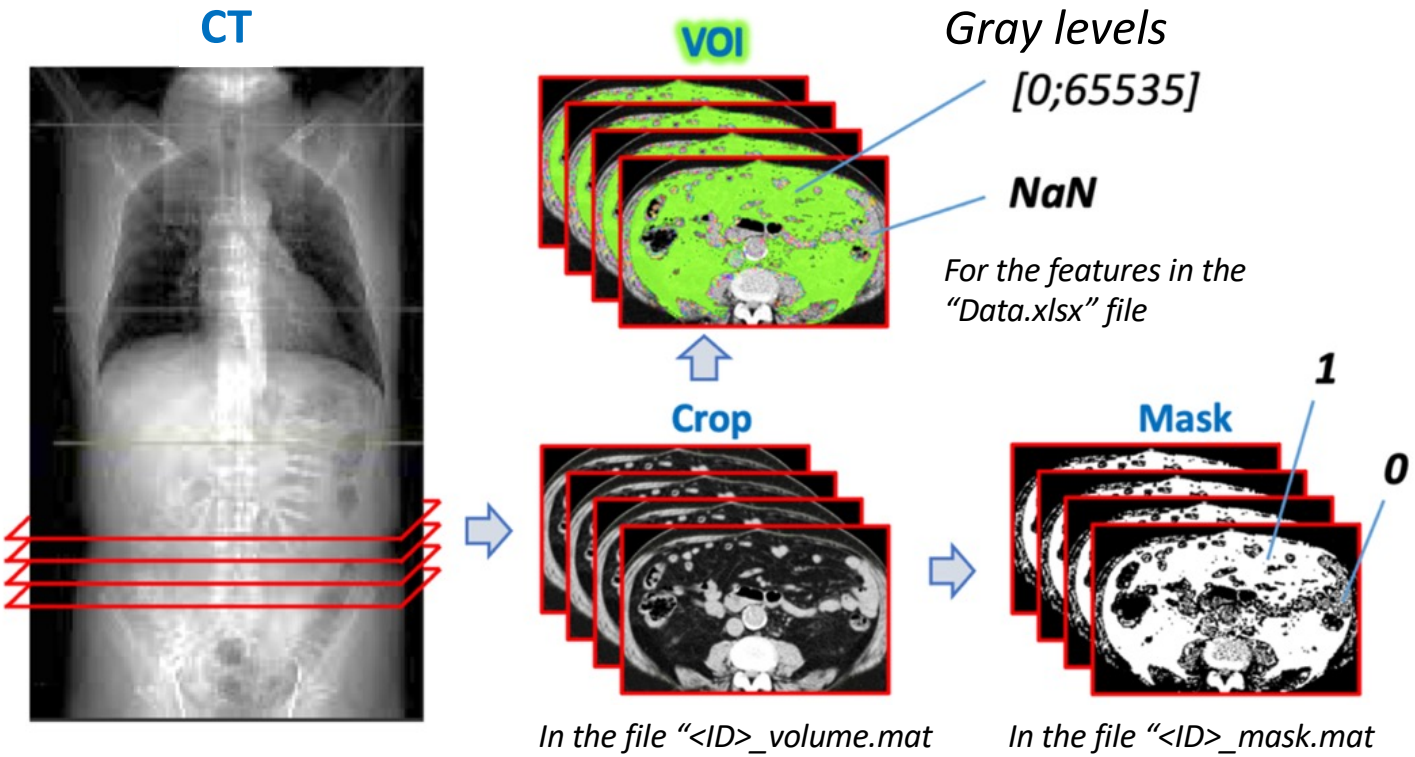
Sapienza Università di Roma



2022 PhD school competition

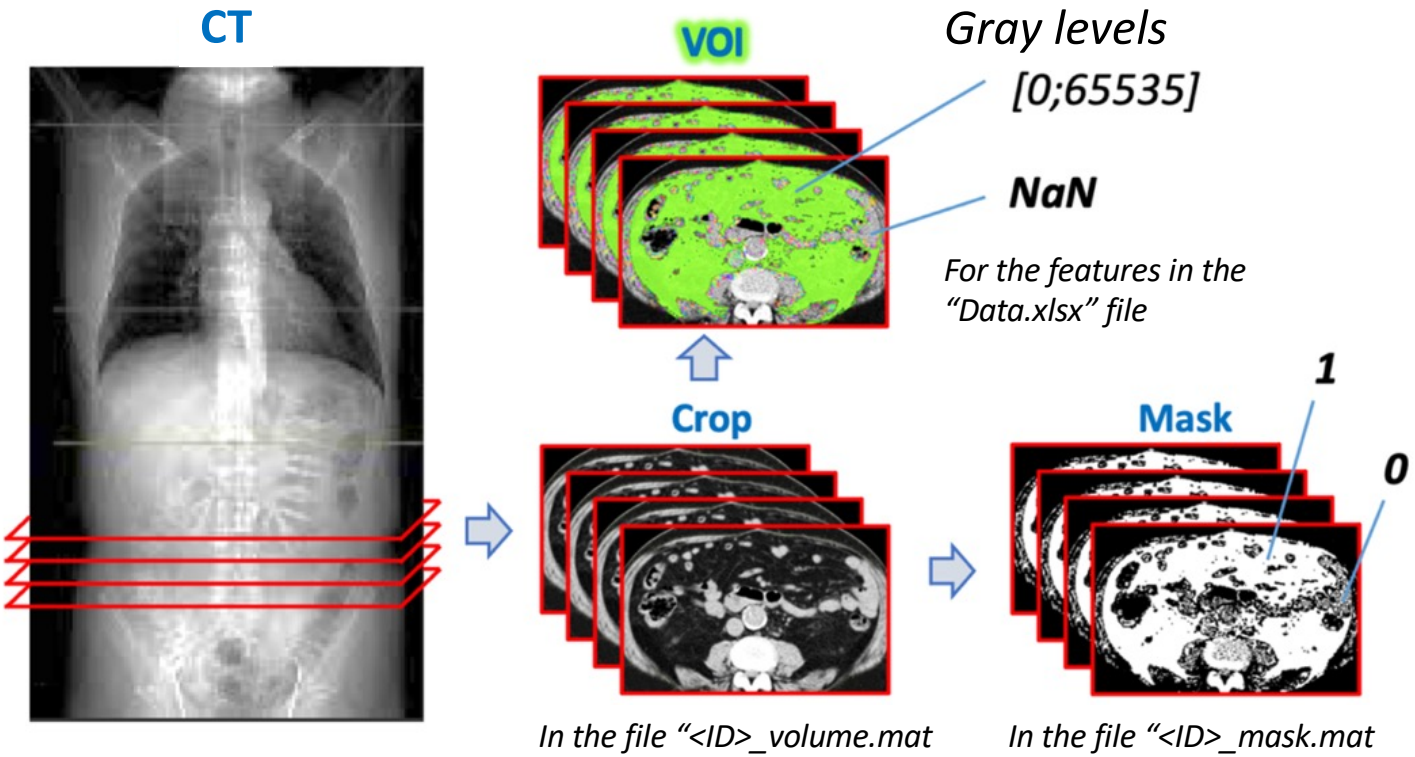
- We are in the context of radiomics
  - It is a method that extracts a large number of features from medical images using data-characterisation algorithms, and it emerged from radiology and oncology
  - These features have the potential to uncover tumoral patterns and characteristics that fail to be appreciated by the naked eye, and that may be useful for predicting prognosis and therapeutic response for various cancer types
  - It supports personalized therapy definition and deployment, being one of the most applications of AI in medical imaging , i.e., it can be applied to any medical study where a pathological process can be imaged.
- We will provide TC scans from 79 (anonymized) patients from abdomen region
- Link: sent by email to you
- All 79 patients were initially diagnosed with a malignant prostate neoplasm and underwent a radical prostatectomy.
- Subsequently, during a “control” examination, tumour recurrence was found in 45 out of the patients
- Requested tasks:
  - Develop an AI-based approach that, using CT scans, predicts the tumour recurrence (i.e. a binary task)
  - Develop an approach to explain the reasons behind the decisions taken by the AI-based approach
  - Design (and eventually implement) a software interface for the physicians (radiologists or oncologists) to use the tools

Available material



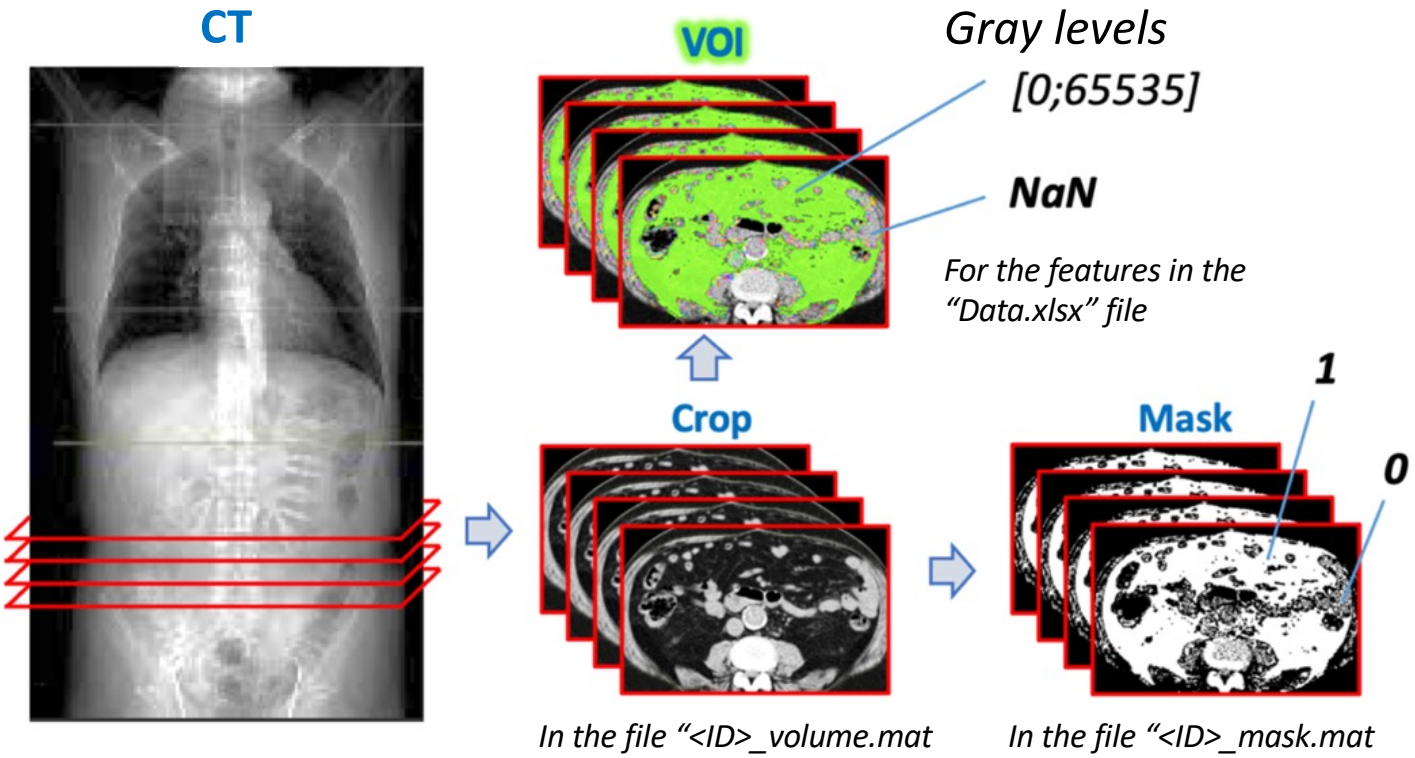
- 3D images of abdominal fat segmented from CT scans (crop of Volume Of Interest, VOI) of the patients;
- Quantitative features pre-computed for you from the VOIs:
  - 12 features describing the statistics from the first-order histogram (column from I to T)
  - 182 features computed from the GLCM 3D histograms (second order histograms), computed as 7 quantities derived from the matrices computed from 26 different directions (the name starts with GLCM);
  - 48 extracted from the first-order histogram of the TOP-LBP transformations, as 12 features from 4 different variants of the transformation;
- Qualitative features:
  - Epoca_TC: if the CT was collected before or after the radical prostatectomy (i.e., during the radiotherapy)
  - Area_grasso_periviscerale: volume of perivisceral fat;
  - Area_grasso_sottocutaneo: volume of subcutaneous fat;
  - Istologia: result of the istologic analysis of the tissue;
  - GS_alla_diagnosi: Gleason Score at the time of diagnosis;
  - TNM_alla_diagnosi: stage of the tumour at the time of diagnosis;
  - Eta_alla_RP: age of the patient at the time of radical prostatectomy

Competition folder

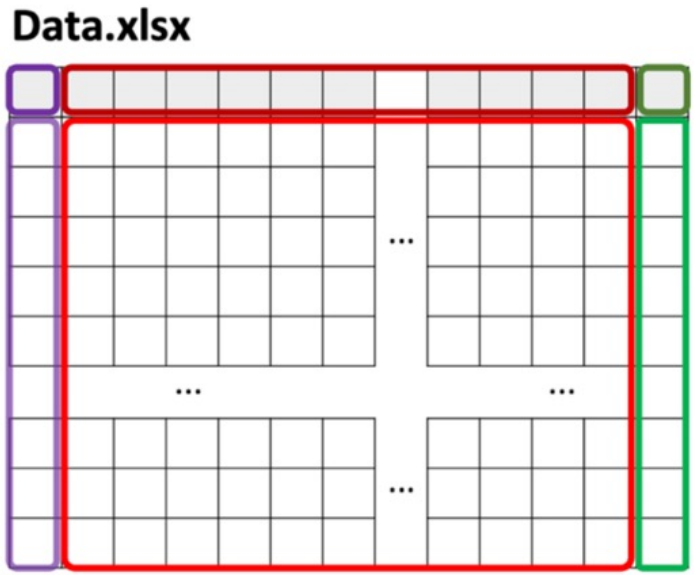


- Folder 'volumes_and_masks', in which there are 79 folders named with the ID of the corresponding patient
  - Each ID folder contains the files "<ID>_volume.mat" and "<ID>_mask.mat";
  - The file "<ID>_volume.mat" contains the crop of the VOIs extracted from the CT scans of the corresponding patient and saved as a 3D matrix of double in Matlab;
  - The file "<ID>_mask.mat" contains the crop of the VOIs extracted from the CT scans of the corresponding patient in which the voxels of the VOI were set to 1 and all others to 0 and saved as a 3D matrix of doubles in Matlab;
  - Note that a VOI contains the voxels of the tumour manually segmented by the physician;
  - The colour depth with which each scan was saved is 2^{16} shades of grey, with values in the range [0;65535].

Competition folder



- File "Data.xlsx", in which several clinical and quantitative features were extracted and listed in the table directly from the 3D volumes (VOI).
  - The table consists of 79 rows and 245 columns;
  - The first row and first column are therefore to be ignored for the purposes of computation;
- Folder "Patients_ s10fCV_folds", containing the mandatory split in training and test portion in 10-fold cross validation



Final evaluation

- The final committee is external to the organizing committee, and it includes radiologists, radiation oncologist, expert in venture capitals, industry expert, AI experts
- In the pitch you must display
  - the average performance with the standard deviation measured in terms of accuracy, sensitivity, specificity, AUC and F1 score
  - A detailed table with the same performance for each cross validation fold
  - The explanations of the decision taken on patients included in fold 10
  - Graphic rendering (and possibly the implementation) of a software for the physician to use the AI-based tools developed
- Here it is the link to inform us of the auto-defined working groups
- Length of the pitch will be communicated after group notification (roughly 7-10 minutes per group)
- Each group component must present during the pitch
- Each group must upload:
  - The presentation and its source files
  - The commented code developed
  - The explained decisions for the test folder
  - All the other material you prepared for the competition (e.g. software schema) – link for the submission
  - Link will be provided soon