Artificial Intelligence in Medicine

Radiomics and DL-segmentation on lung tumor, the Blue Sky project (WP3.T8)

Francesca Brero
Raffaella Cabini
Ian Postuma

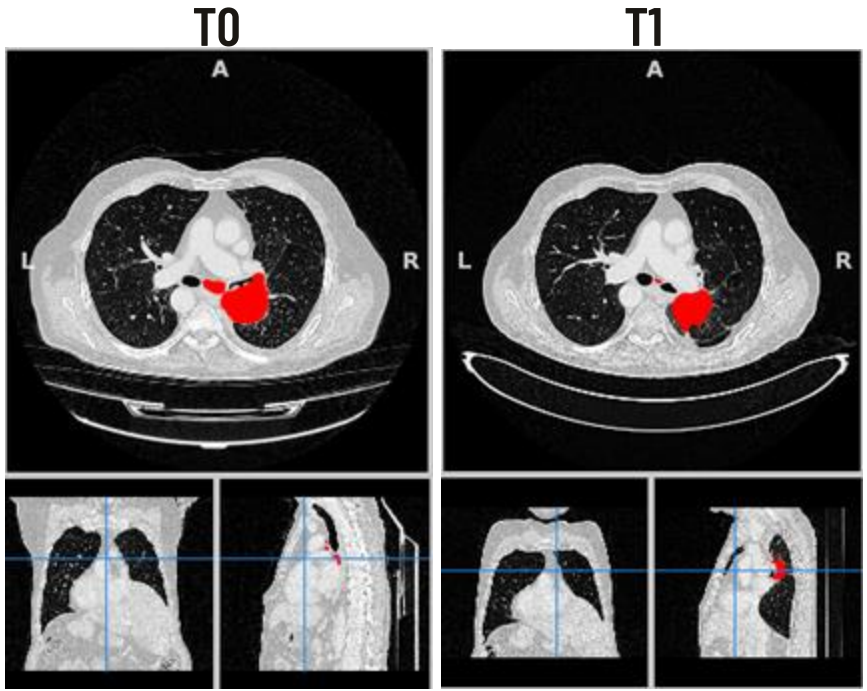
General Meeting next_AIM
13-15/02/2023

(Small) dataset

57 patients with inoperable stage III lung adenocarcinoma undergoing radio-chemo-immuno therapy.
2 CT images acquired at 2 time points:

- Baseline: at the diagnosis
- After the chemo-radio-immuno therapy

Homogeneus dataset of clinical features



Aim

Predict:

- Progression Free Survival (PFS), combined event of:
- Metastasis
- Relapses
- Death

STEPS

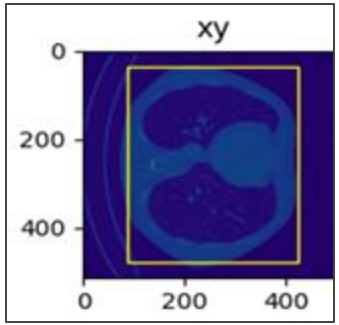
01 nnUnet for automatic segmentation of lung cancer

02 Radiomics extraction harmonization for Progression Free Survival prediction

nnUnet for automatic segmentation of lung cancer

Pipeline

Traning and Testing of two Convolutional Neural network:



Object Detection

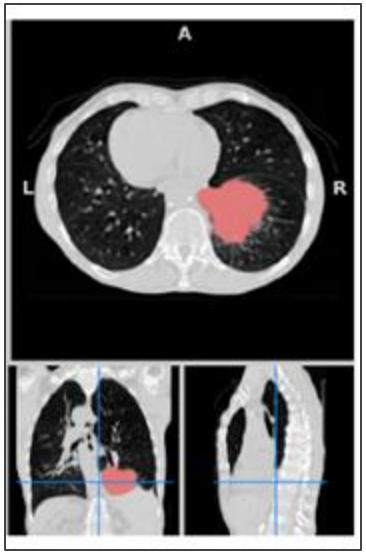
Bounding Box identification

Select the CT region that contains the lungs

Image Segmentation

Segmentation Region of Interest (ROI)

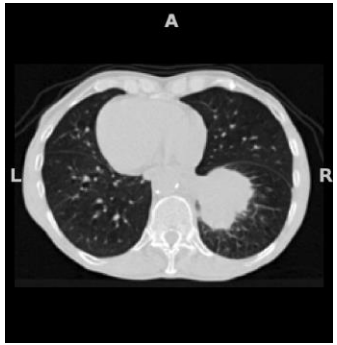
Non-Small Cells Lung Cancer



Lizzi et al. Quantification of pulmonary involvement in COVID-19 pneumonia: an upgrade of the LungQuant software for lung CT segmentation.

Bounding Box

Input



Toraco-abdominal 3D CT scans, matrix size 100x100x100

DL Model

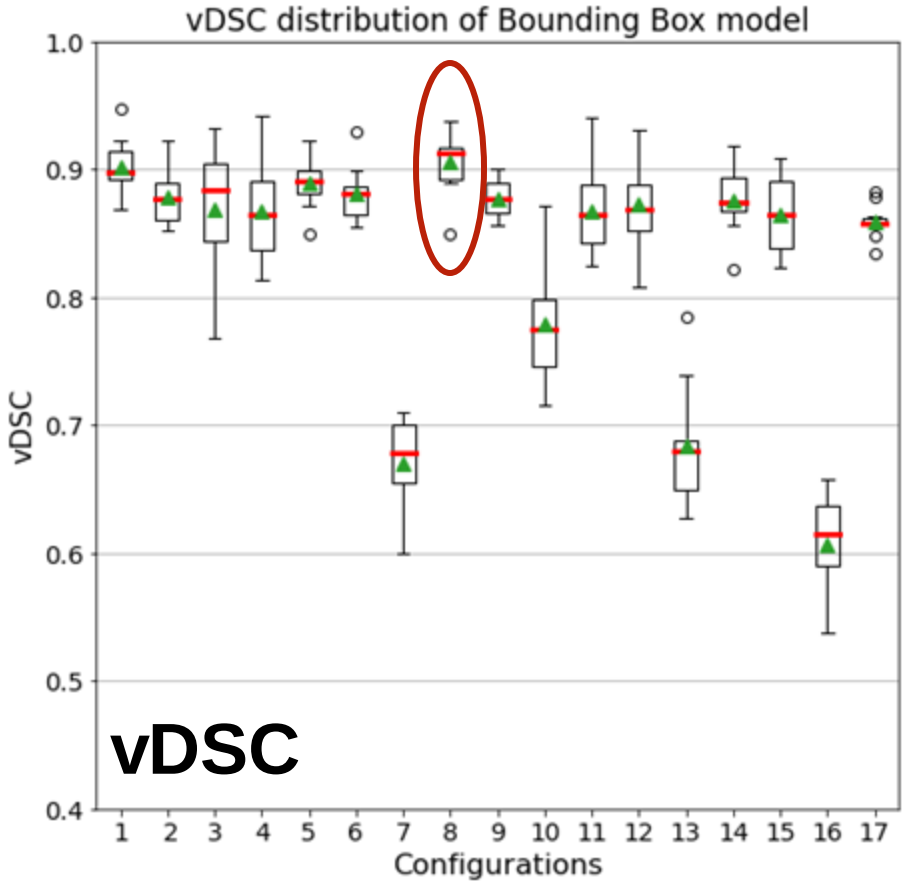
Based on AlexNet architecture

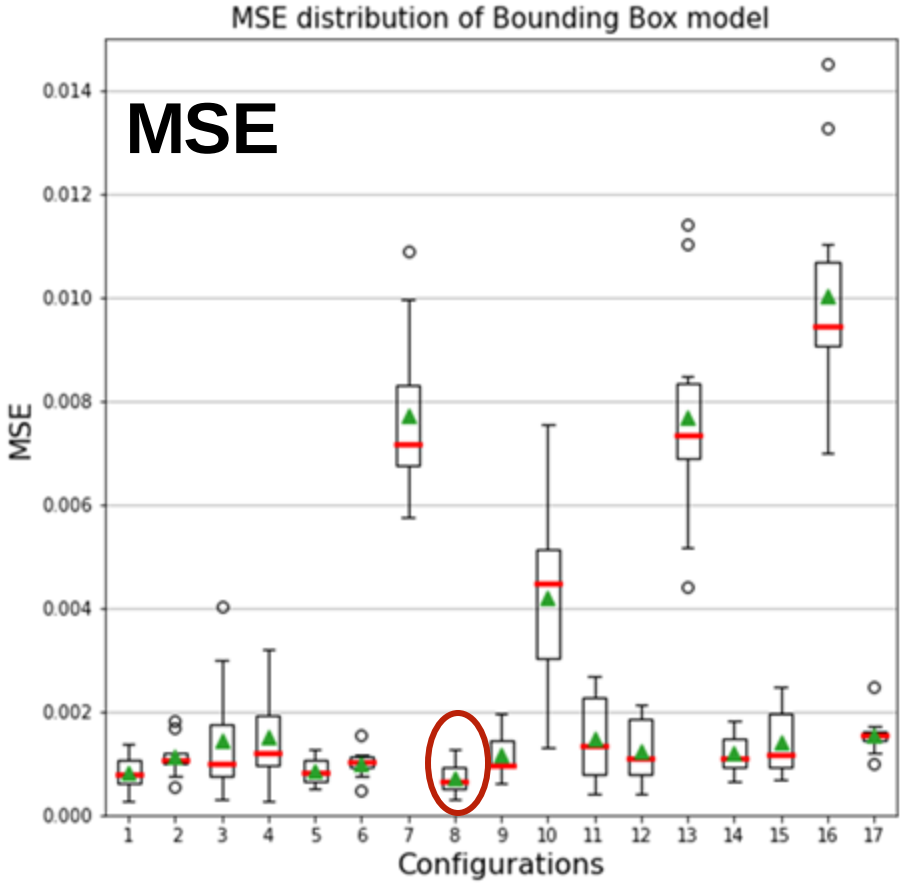
Output

Minimum and maximum coordinates of the Bounding Box

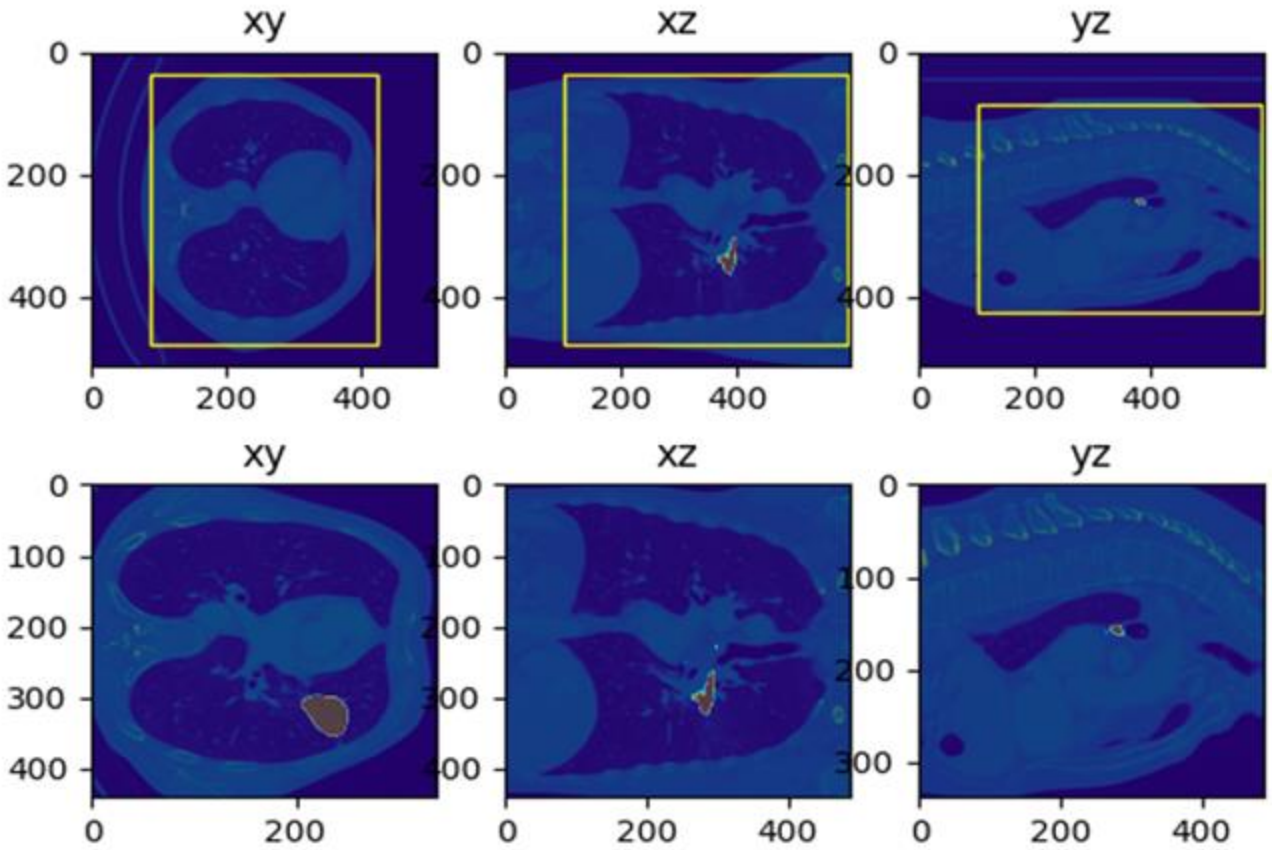
17 tested configurations (set of NN parameters) to find the best network architecture.
The test was performed on 10 CT images evaluating both MSE and Dice (vDSC) metrics.

vDSC and MSE distributions for the Bounding Box model





Bounding Box results



Lung cancer segmentation

Input

Images generated from Bounding Box model

DL Model

Based on nnU-Net

Output

Segmentation masks of lung volume and lung cancer

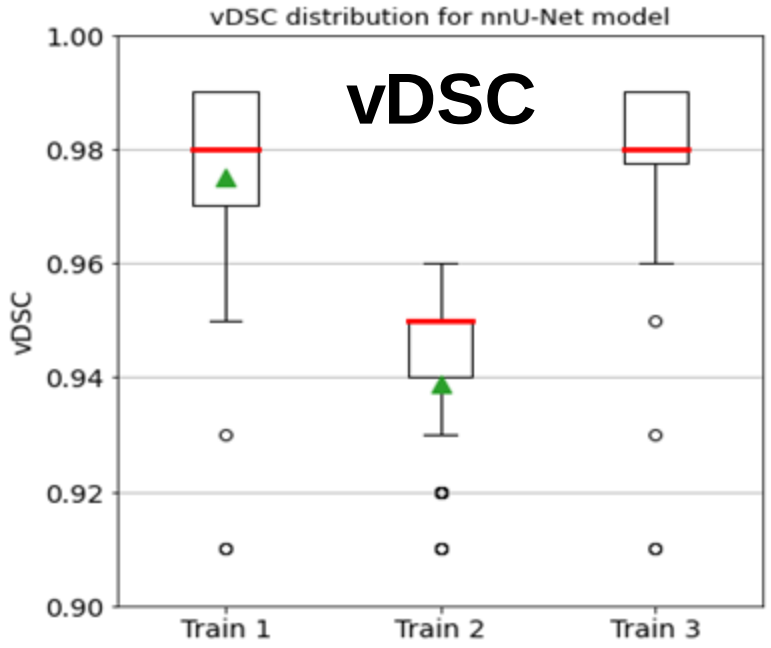
NnUnet training

Training and test: different configurations using both **private images (171 CT)** (IRCCS, Bluesky project of Policlinico S. Matteo, Pavia) and images **(511 CT)** of a **public dataset** (Lung1).

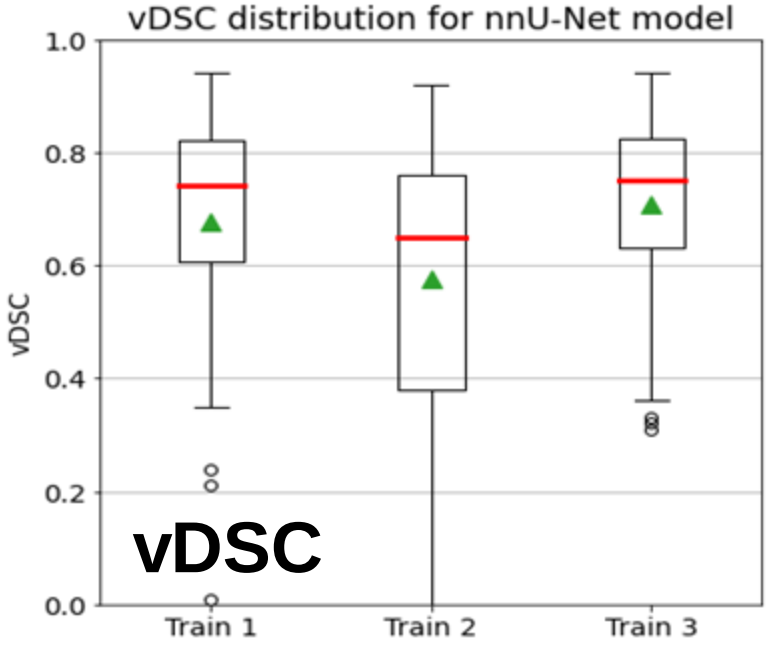
	Training set	Validation set	Test set
Train 1	Public and private images **415**	Public and private images **103**	Public and private images **59**
Train 2	Public images **325**	Public images **81**	Private images **171**
Train 3	Public and private images **409** (without corrupted data)	Public and private images **102** (without corrupted data)	Public and private images **59**

Segmentation results

Lung Segmentation

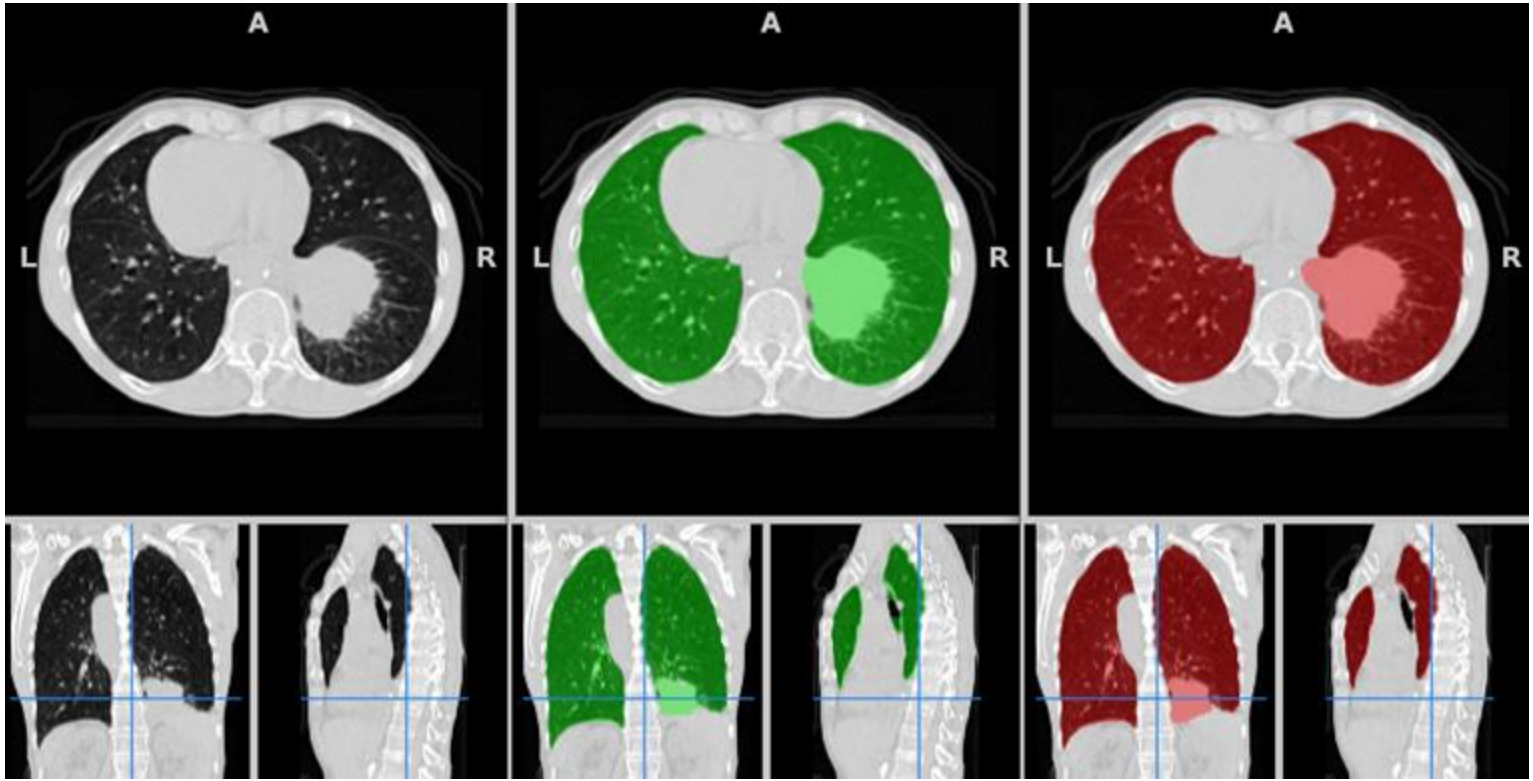


Tumor Segmentation

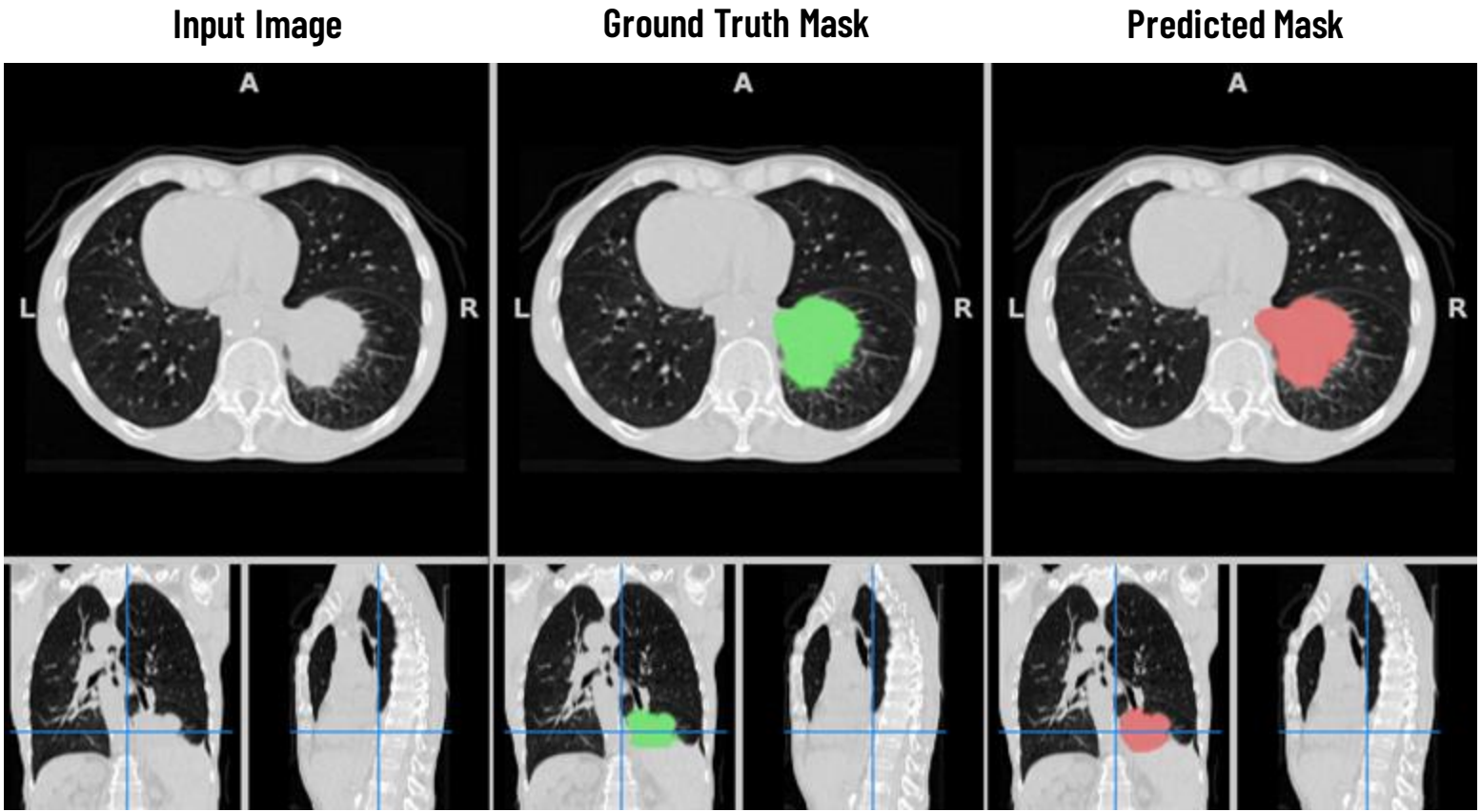


Lung results

Input Image | Ground Truth Mask | Predicted Mask



Lung cancer results



Radiomics extraction harmonization for Progression Free Survival prediction

(Small) dataset

57 Images from 10 different centers:

Image Parameters:

Pixel
From (0.62 x 0.62) mm^2 to (0.98 x 0.98) mm^2

Slice thickness
From 0.3 mm to 3.0 mm

Reconstruction Parameter:

Convolutional Kernel
11 types

Acquisition Parameters:

Scanner
4 different vendors

Current
From 56 mA to 581 mA

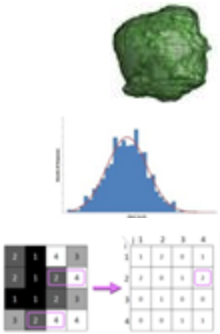
Contrast Agent
2 types

kV-peak
100 kVp, 120 kVp, 130 kVp, 140 kVp

Exposure Time
From 350 s to 1000 s

Features Extraction

42 features computed by Lifex (IBSI standard compliant), including six categories of features:

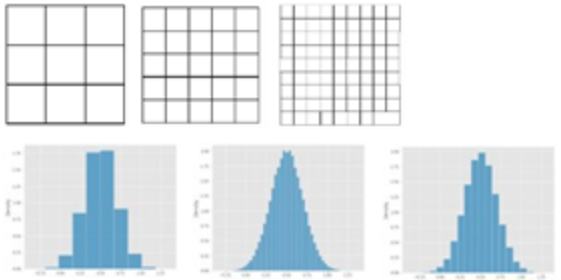


- 4 Size- and Shape-based features
- 6 First-order Statistics features
- 32 Higher order Statistics features:
  7 grey-level co-occurency matrix features, 11 grey-level run length matrix features, 3 neighboring grey-level difference matrix features, 11 grey-evel zone length matrix features

Images Parameters Harmonization

To harmonize images (before the extraction):

1. **Spatial Resampling:** 1mm x 1mm x 1mm
2. **Intensity Rescaling:** -1000 HU, 3000HU
3. **Intensity Discretization:** 400 bins of size 10 HU

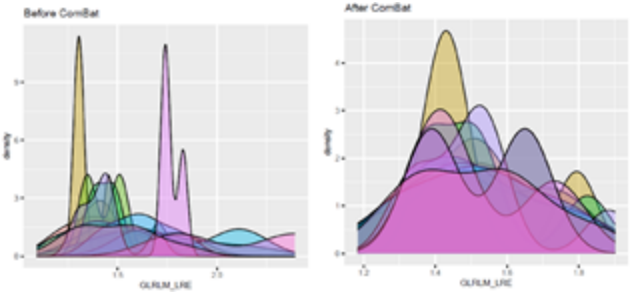


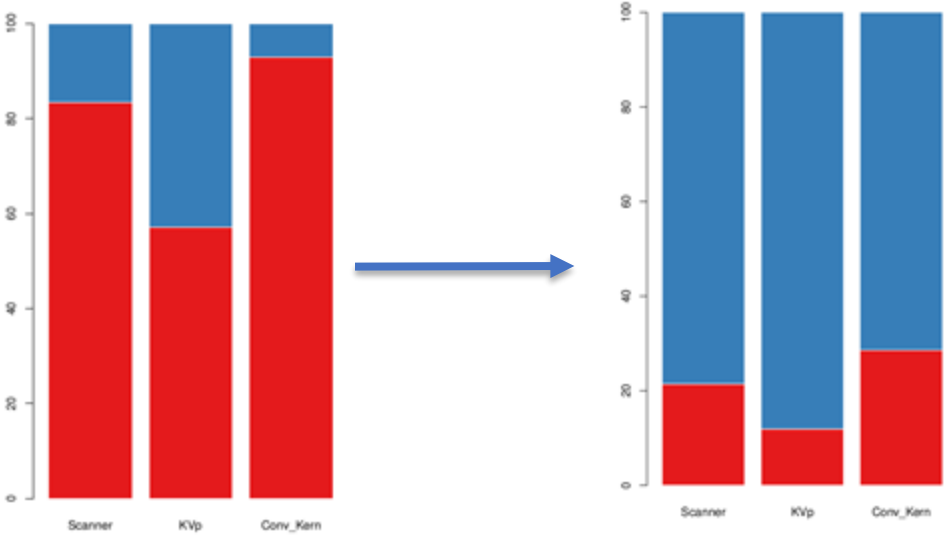
Acquisition parameters Harmonization

Scanner vendor, kVp and Convolutional Kernel have a statistically significant impact on features distributions

ComBat: realigns features as a function of median and variance.
Batch: numeric value associated with each combination of parameters.

kV peak - convolution kernel combinations define **the harmonization batches**





Percentage of features dependent on the parameter

Percentage of features not dependent on the parameter

Predictive Models

Classification task: predict Progression Free Survival.

Objective: find a mathematical model that, based on the data provided in the training phase, learns to automatically classify the patient's prognosis:

- **INPUT:** radiomic features and covariates (5 clinical features, e.g. date of administration of immunotherapy, Sequential or concurrent RT ...)
- **OUTPUT:** Progression Free or not (event of Metastasis, Relapses, Death).

Leave-One-Out cross validation

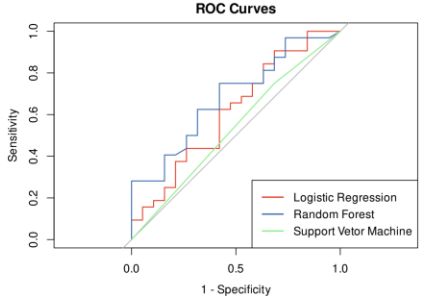
Three classification models are compared:

- Elastic Net;
- Random forest;
- Support Vector Machine.

Predictive Models

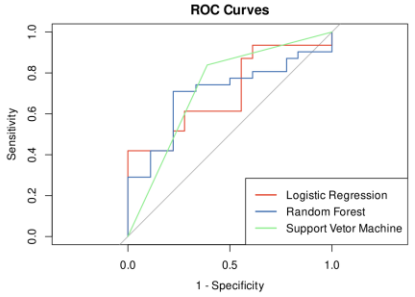
Radiomic Features

	Accuracy	AUC
ElasticNet	0.59	0.61
RandomForest	0.65	0.68
SupportVectorMachine	0.63	0.50



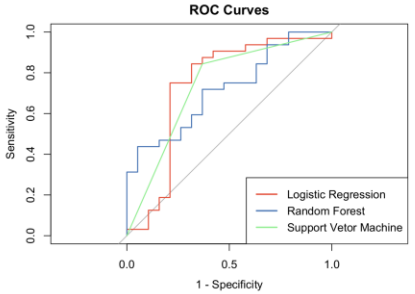
Clinical Features

	Accuracy	AUC
ElasticNet	0.69	0.70
RandomForest	0.63	0.70
SupportVectorMachine	0.76	0.72



Radiomics Features + Clinical Features

	Accuracy	AUC
ElasticNet	0.78	0.74
RandomForest	0.61	0.72
SupportVectorMachine	0.76	0.74



Conclusions

1. Importance of the development of robust analysis pipeline for small-datasets
2. Images and features harmonization steps are necessary with small-datasets
3. Possibility of applying ML for prognosis prediction
4. Automatic segmentation to prevent human segmentation variability

Thank you!

Leonardo Barzaghi, Elena Ballante, Giulia Colelli, Ian Postuma,
Sara Panebianco, Agnese Robustelli Test, Ilaria Villa