Artificial
Intelligence in
Medicine

Harmonization of features extraction process using the ComBat tool in a multicenter study on stage III unresectable Non-Small-Cell Lung Cancer

Raffaella Cabini

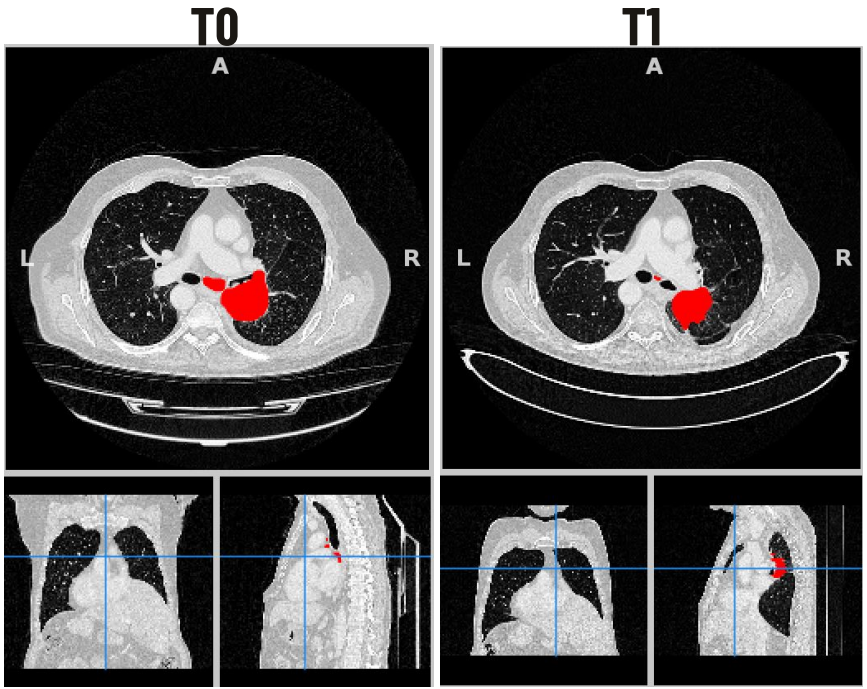
108 Congresso Nazionale Societa' Italiana di Fisica
14/09/2022

(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:

- Probability of Metastasis
- Probability of Relapses
- Response to the therapy

(Small) dataset

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

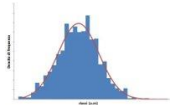
Contrast Agent
2 types

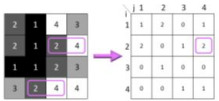
kV-peak
100 kVp, 120 kVp, 130 kVp, 140 kVp

Exposure Time
From 350 s to 1000 s

Features Extraction

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



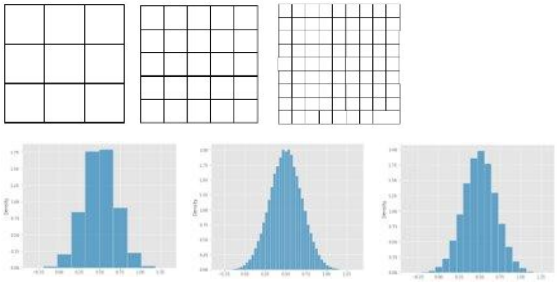


- 4 Size- and Shape-based features
- 6 First-order Statistics features
- 32 Higher order Statistics features:
  7 grey-level co-occurency matrix features, 11 grey-level run length matrix features, 3 neighboring grey-level difference matrix features, 11 grey-evel zone length matrix features

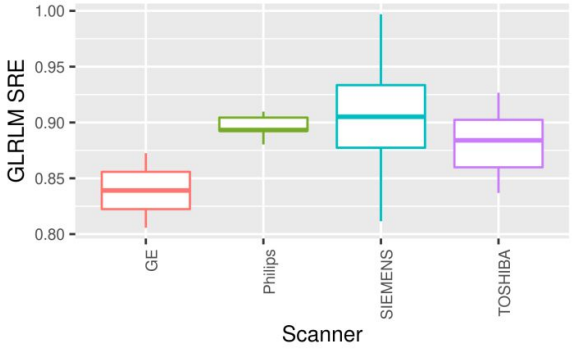
Images Parameters Harmonization

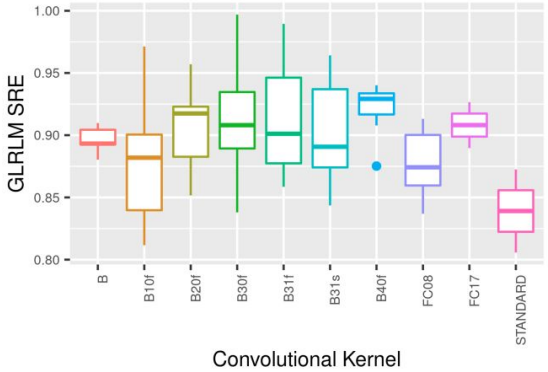
To harmonize images (before the extraction):

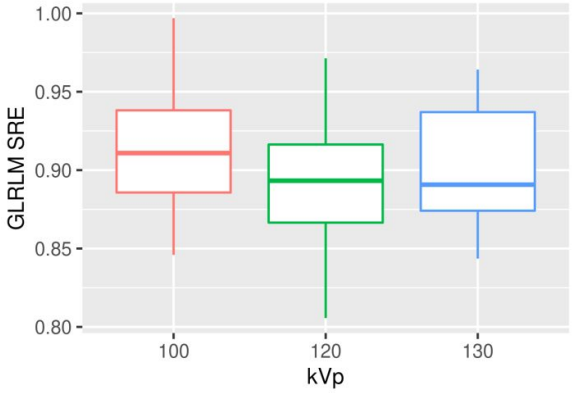
1. **Spatial Resampling:** 1mm x 1mm x 1mm
2. **Intensity Rescaling:** -1000 HU, 3000HU
3. **Intensity Discretization:** 400 bins of size 10 HU

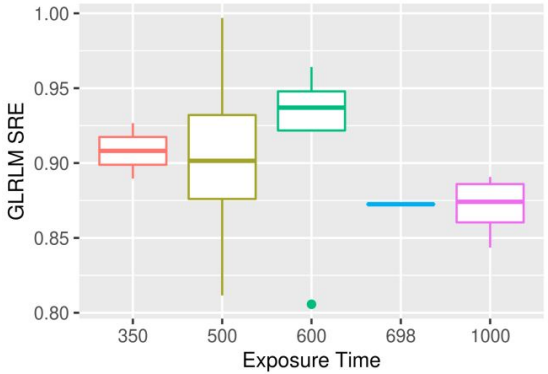


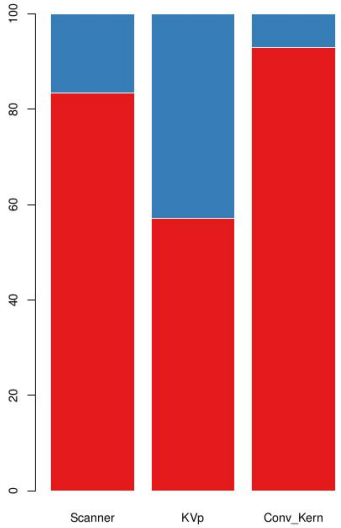
Acquisition parameters Harmonization











Percentage of features dependent on the parameter

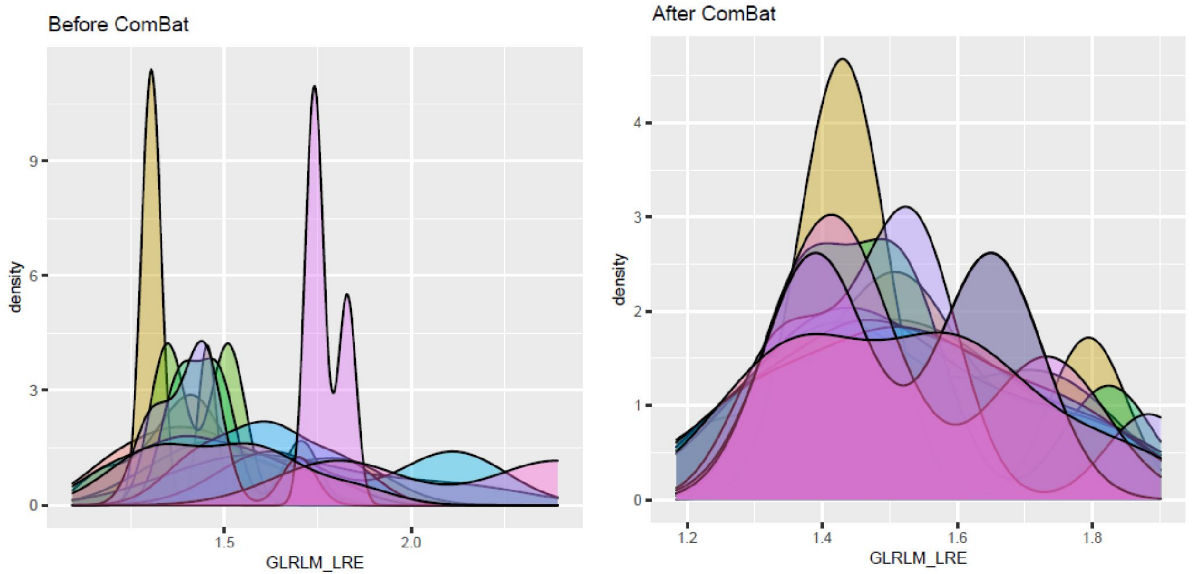
Percentage of features not dependent on the parameter

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

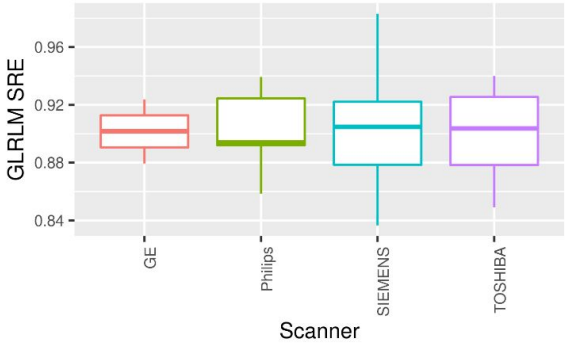
Acquisition parameters Harmonization

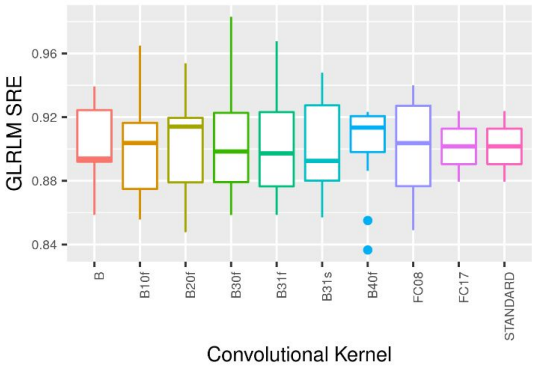
ComBat: realigns features as a function of median and variance.
Batch: numeric value associated with each combination of parameters.

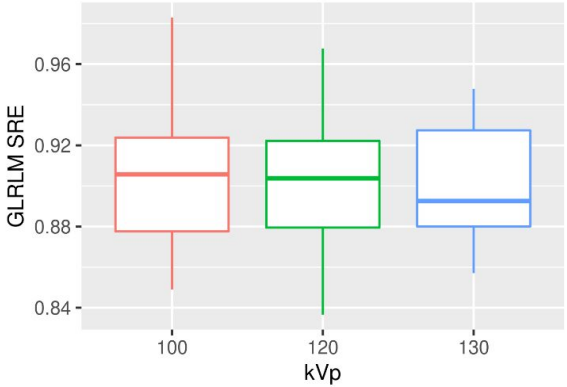
kV peak - convolution kernel combinations define **the harmonization batches**

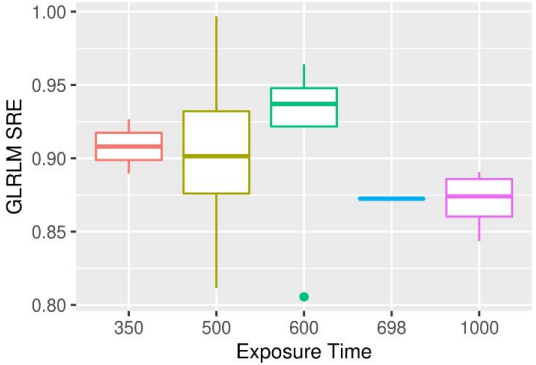


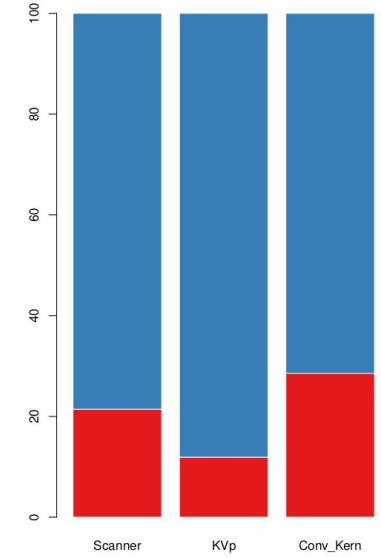
Acquisition parameters Harmonization











Percentage of features dependent on the parameter

Percentage of features not dependent on the parameter

Predictive Models

Classification task: predict which patients will develop metastases.

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;
- **OUTPUT:** incidence of metastases or not.

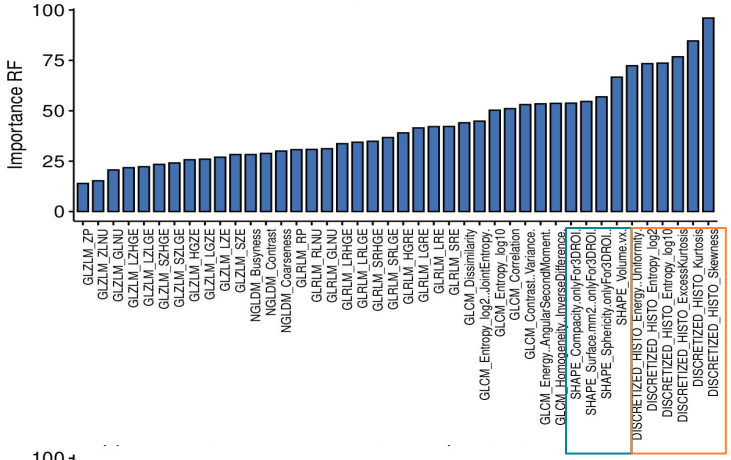
Leave-One-Out cross validation

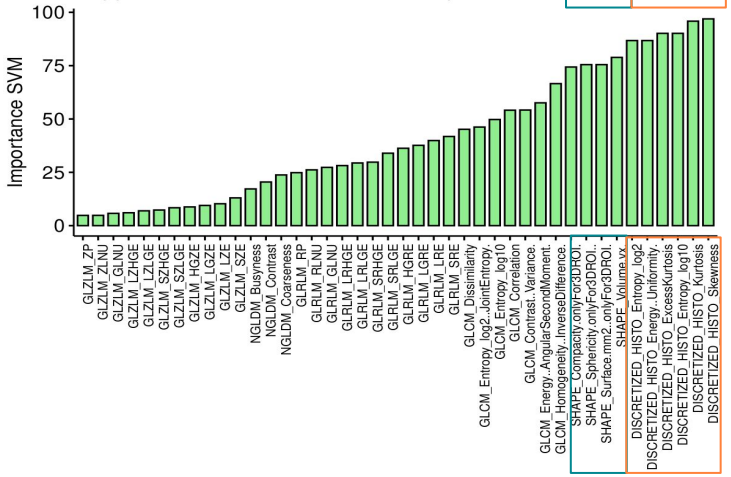
Three classification models are compared:

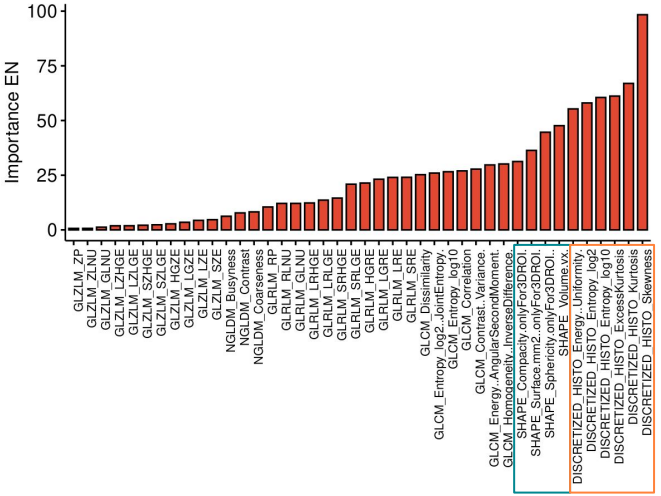
- **Elastic Net**;
- **Random forest**;
- **Support Vector Machine.**

Model	**Accuracy**	**AUC**
ElasticNet	0.67	0.58
Random Forest	0.68	0.53
Support Vector Machine	0.75	0.65

Variable importances







The most important features for the classification task are **First Order Statistics** features and **Size- and Shape-based** features

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!