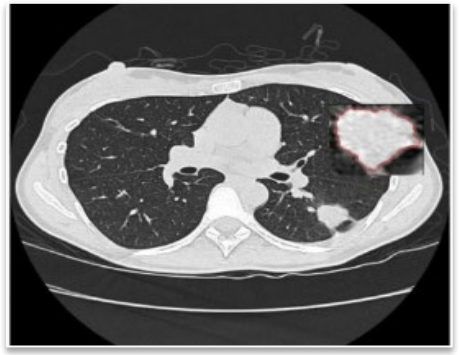
Neuroengineering A.Y. 2022/23
PW 1
Low-dose CT screening optimization through automatic Lung Cancer diagnosis using CNN

Tutor: Matteo Rossi

Supervisor: Prof. Pietro Cerveri

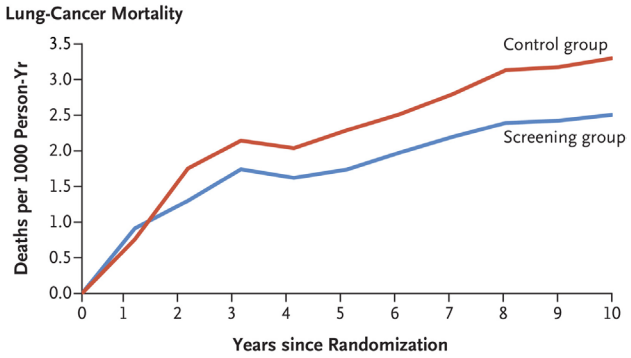
Background

X-ray diagnostic imaging for early detection of lung cancer



Gold standard: Low-dose CT (LDCT)

LDCT lung cancer screening (LDCT LCS)



Mortality reduction >20% on high risk subjects

NLST Team [2011]; Zhao [2011]; Pastorino[2015]

Clinical practice limits

- Large amount of data
- Interobserver variability
- Additional exams caused by false positives
- Economic impact in the health system

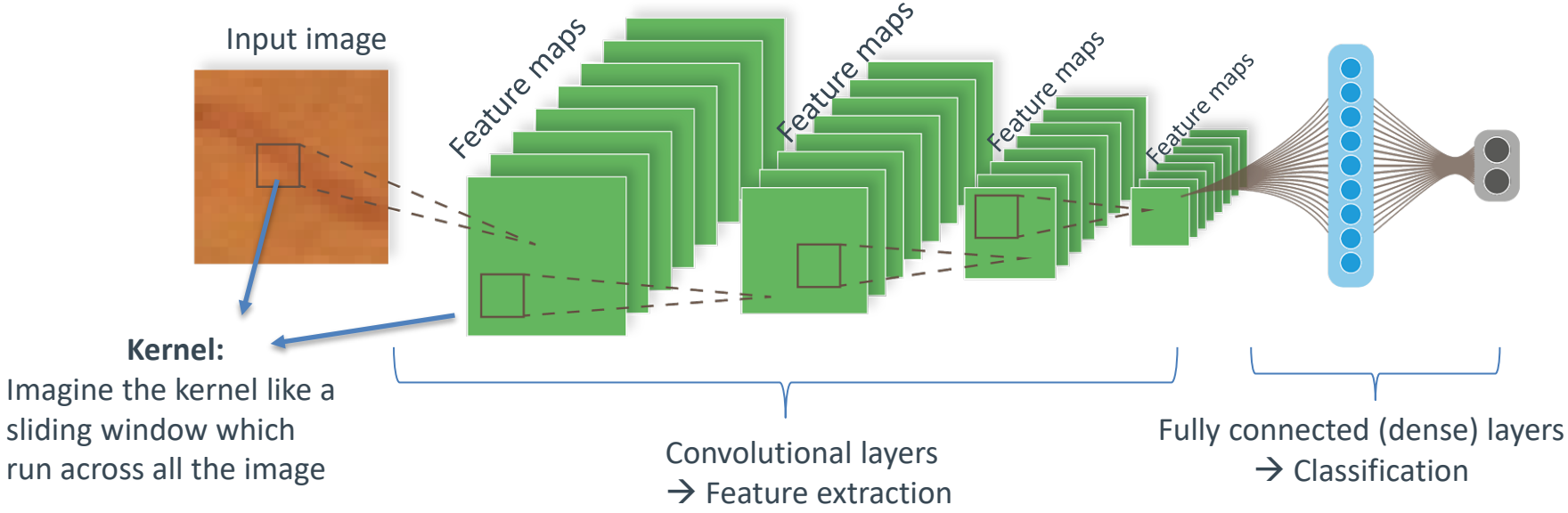
Veronesi [2020]

Decision support systems

Liu [2020];Benzaquen [2019]; Sahiner [2009]

Background

CNN based solutions demonstrated to be promising

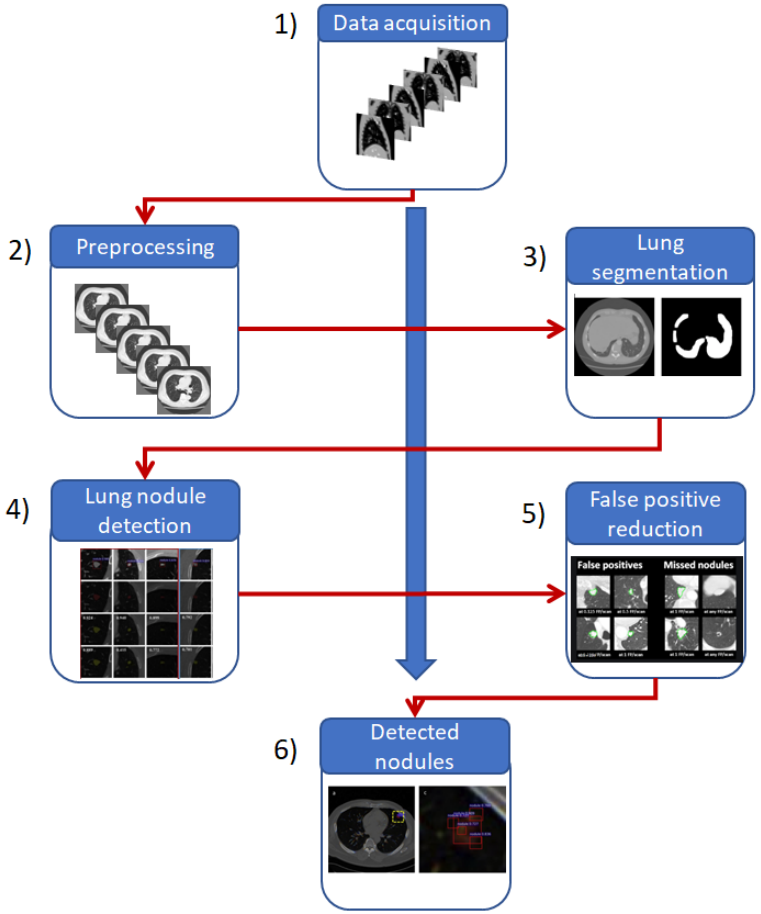


Still open challenges: **(1)** Increase stability and **(2)** Increase of interpretability

Computer assisted solutions for lung cancer prevention:

- Computer assisted Detection (CADe)
- Computer assisted Diagnosis (CADx)

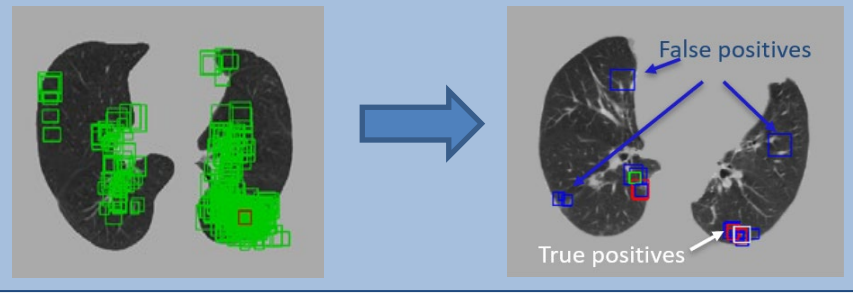
Computer Assisted Detection (CADe)



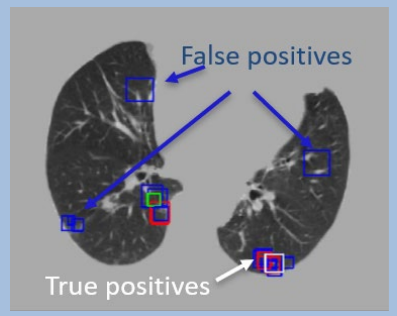
Lung nodule detection

Processing of thounsand regions of different shape and size

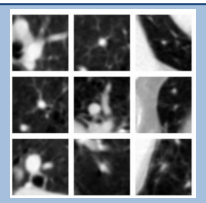
Selection of regions with highest probability of being a nodule



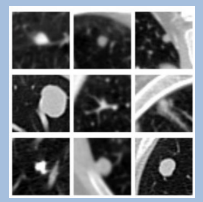
False positive reduction



FP

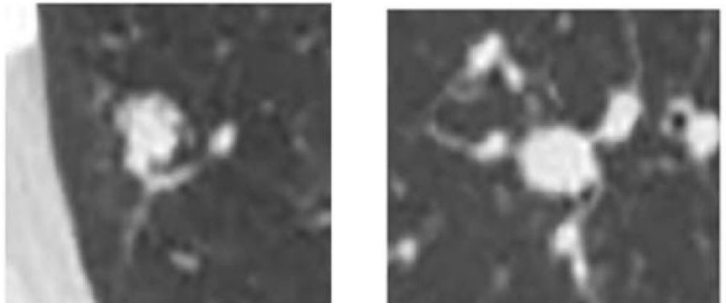


TP



Computer Assisted Diagnosis (CADx)

Malignant vs benign



Clinical practice in the management of indeterminate pulmonary nodules through LDCT

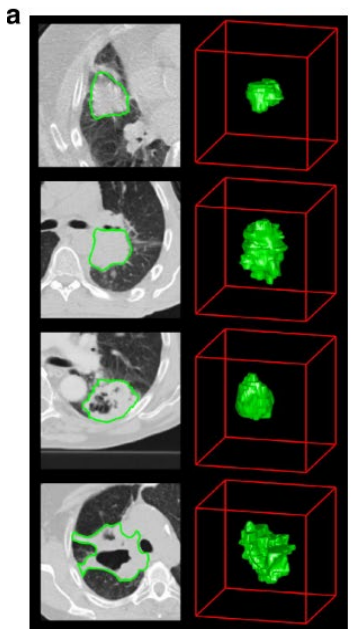
- Describe the appearance of nodules in the image
- Consider growth rate and nodule appearance changes

Artificial neural networks

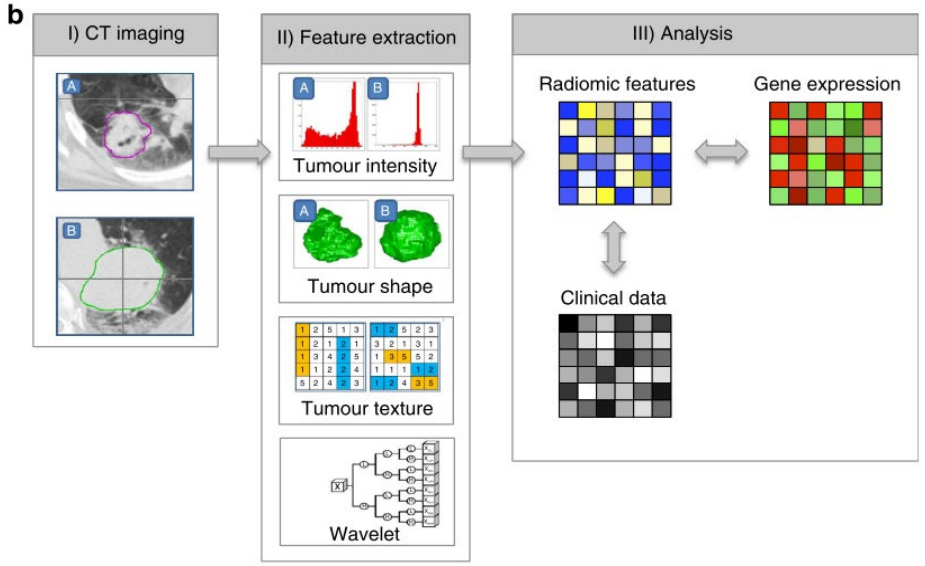
- Extract automatically features from image
- Learns how to elaborate and combine features to classify the two groups

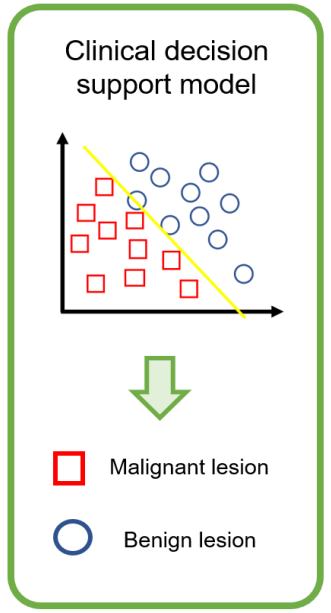
View from the clinician as black box – low interpretability

Radiomics workflow



[Aerts 2014]





[Peikert 2018; Tu 2018; Choi 2018; Liu 2017]

- **Quantitative interpretation of medical images**
- **More clinical information with respect the reading procedure currently used**
- **Limitation of invasive exams to confirm the hypothesis given by radiologist**

Focus of the project

Investigation of CNN based solution for:

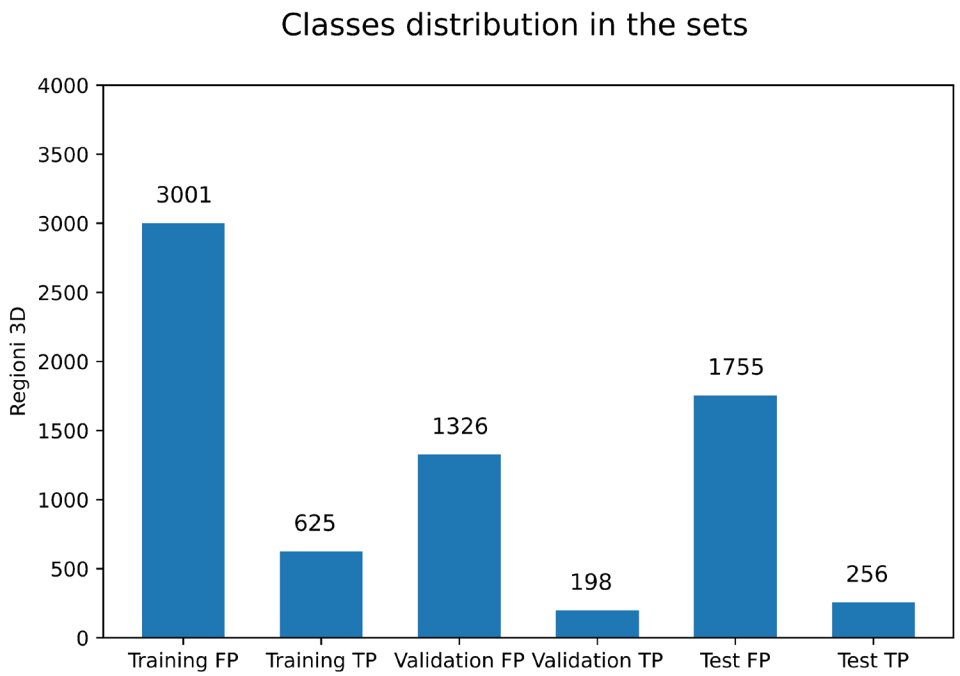
- False Positive reduction
- Malignancy identification

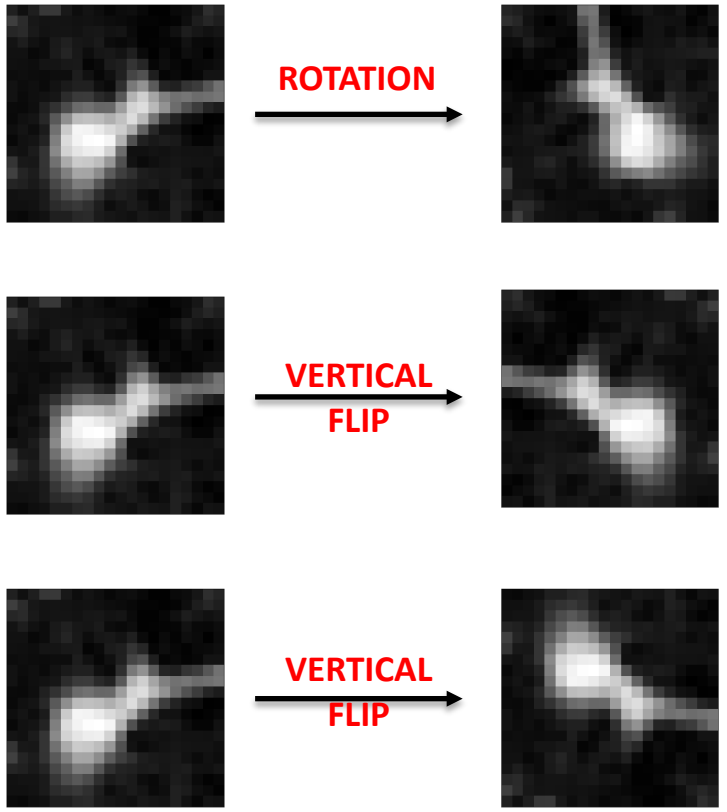
→ Binary classification problems

→ Unbalance datasets

How to deal with unbalance dataset?

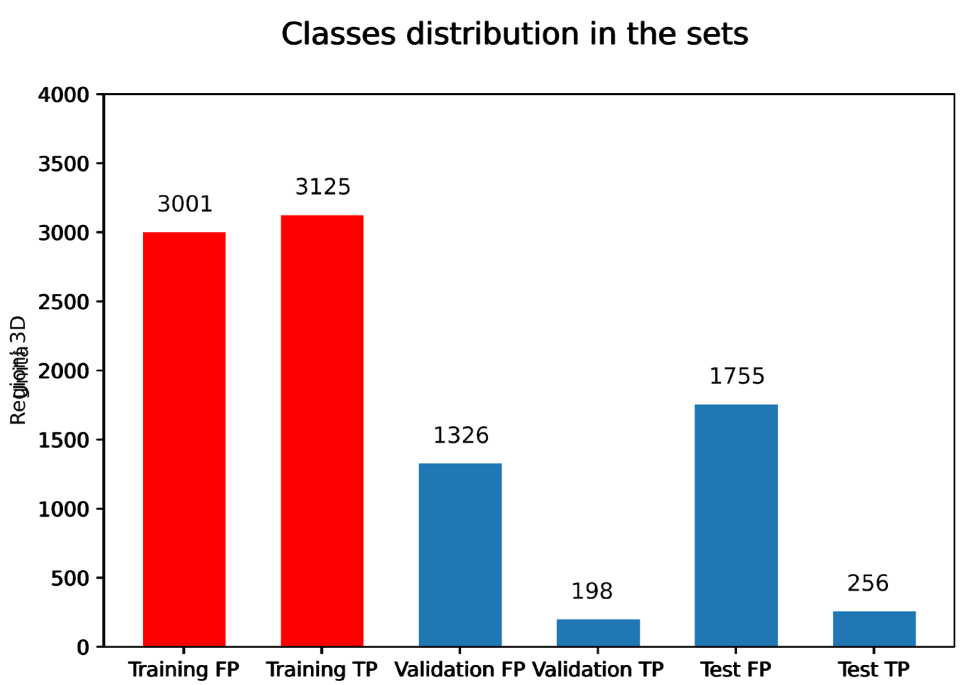
(1) Data augmentation



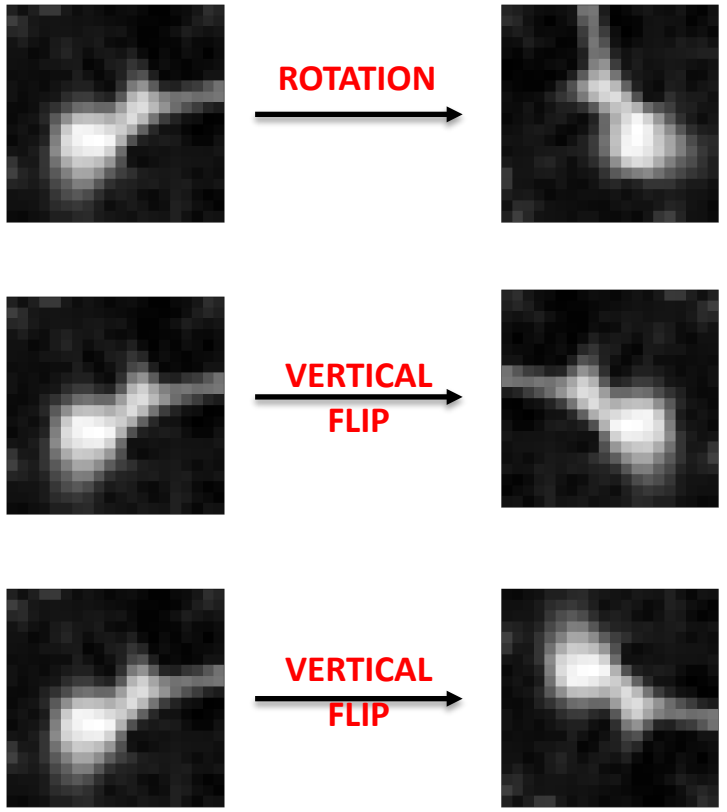


How to deal with unbalance dataset?

(1) Data augmentation



Minority class duplicate four times



(2) Weighted loss function

Dataset – False Positive Reduction

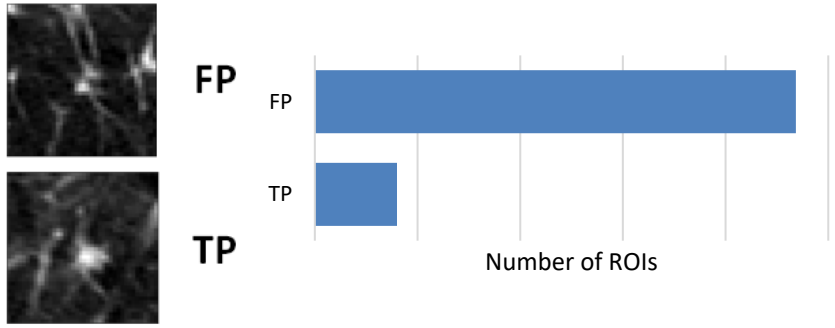
IEO Istituto Europeo di Oncologia

COSMOS [**C**ontinuous **O**bservation of **SMO**king **S**ubjects]

Longitudinal study: 5000 subjects monitored for 10 consecutive years (at least 1 CT scan/year for each patient)

Subset for False Positive Reduction

- **3D regions predicted as potential pulmonary nodules by a detection algorithm (Faster-RCNN)**
- **Higly imbalance dataset**



We will exploit all the strategies to compensate the lower number of TP

Dataset – Malignancy Prediction

IEO Istituto Europeo di Oncologia

COSMOS [**C**ontinuous **O**bservation of **SMO**king **S**ubjects]

Longitudinal study: 5000 subjects monitored for 10 consecutive years (at least 1 CT scan/year for each patient)

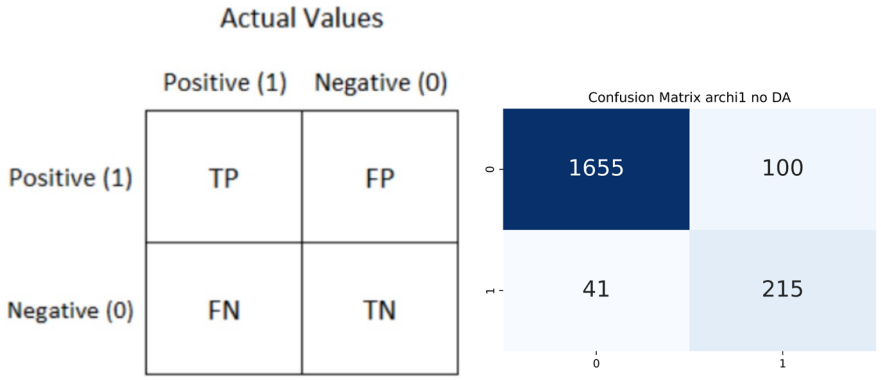
Subset Malignancy Prediction

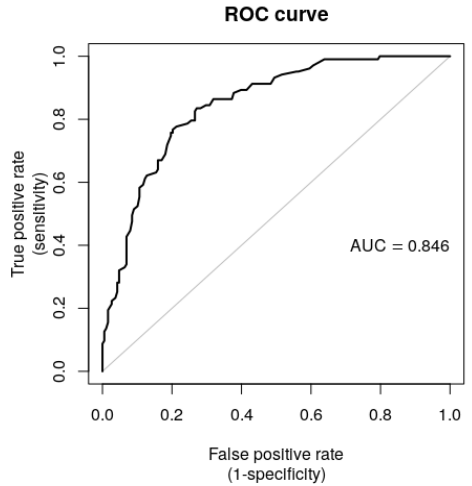
- **2D regions centered into a pulmonary nodule**
- **Each region is associated to a target value (1 in case of maligancy and 0 otherwise)**
- **Each region is associated to a set of Radiomic features**

- Investigation of strategies to compensate the lower number of positive samples
- higher importance will be given to exploit the integration of deep learning features with Radiomics features

Metrics used to evaluate a classifier performance

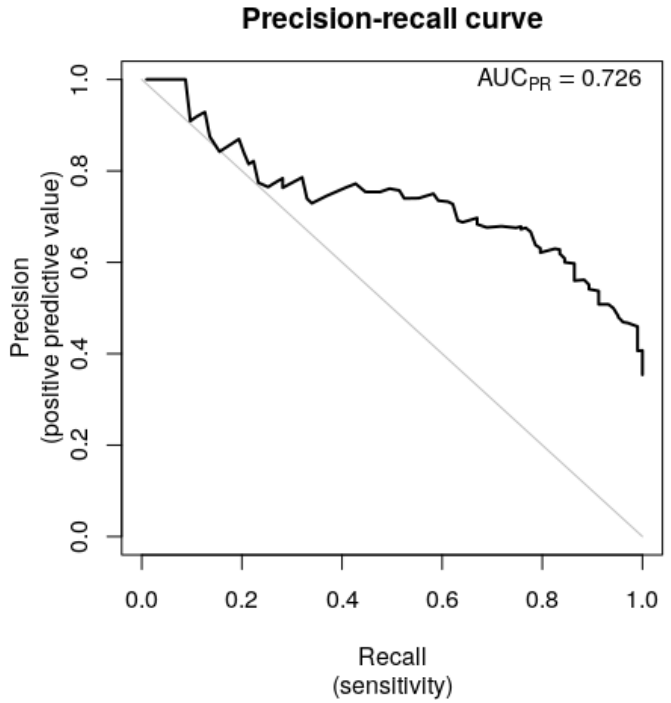
- Classification Accuracy (ACC) : ratio between the number of correctly classified points to the total number of points.
- Confusion Matrix is a summary of predicted results
- Area Under the Curve (AUC): area under the **Receiver Operating Characteristics (**ROC **)** curve is a performance measurement for the classification problems at various threshold settings





Metrics used to evaluate a classifier performance

- **Precision (Positive Predictive Value)** is the rate between the correctly classified positive instances over the total number of instances predicted as positives.
- **Recall** (sensitivity) is the fraction of the correctly classified positive instances from the total positive instances (TP/TP+FN)
- **F1-score**: harmonic mean between Precision and Recall
- Area Under Precision-Recall Curve (AUPRC): **shows the tradeoff between precision and recall for different probability thresholds**.



Project Material

Google Drive Folder:

https://drive.google.com/drive/folders/16Kze28P-pPSDykujMwl6JzdCbwnTx2kV?usp=share_link