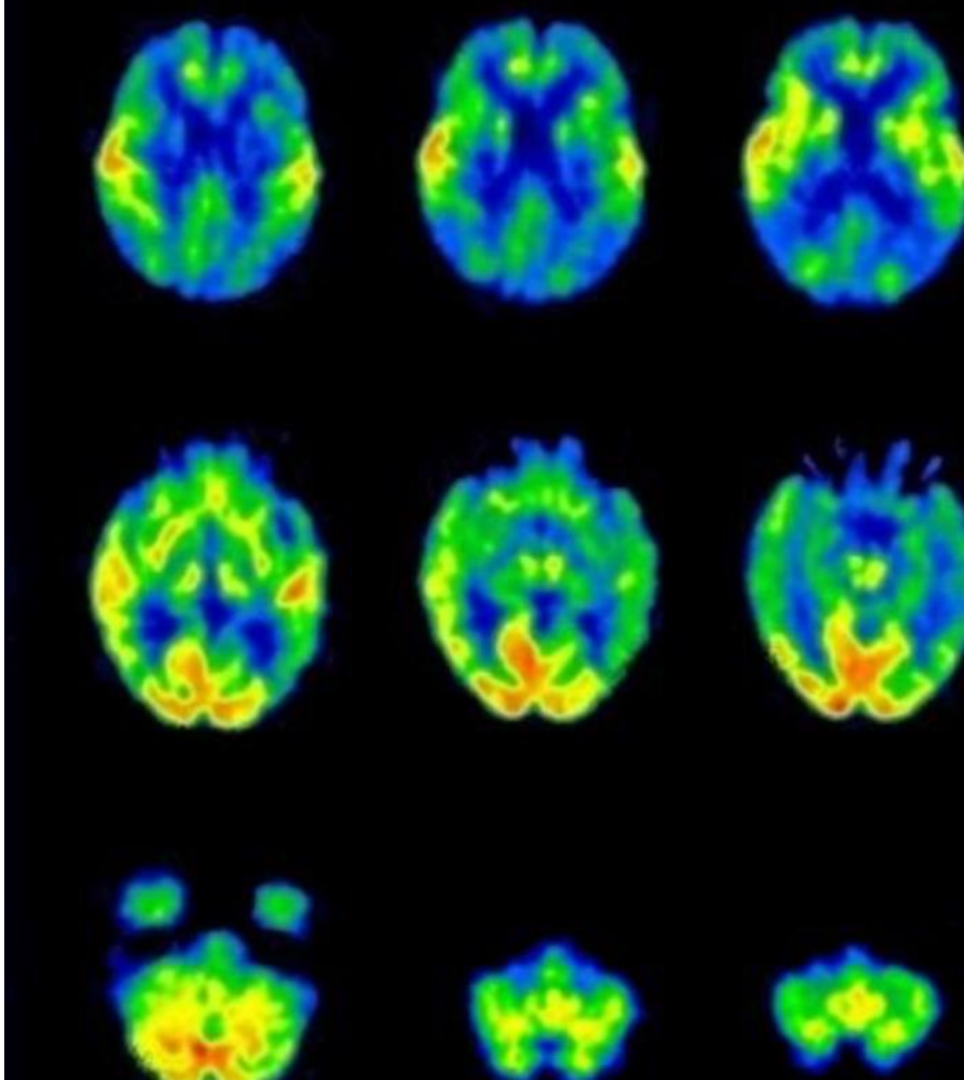
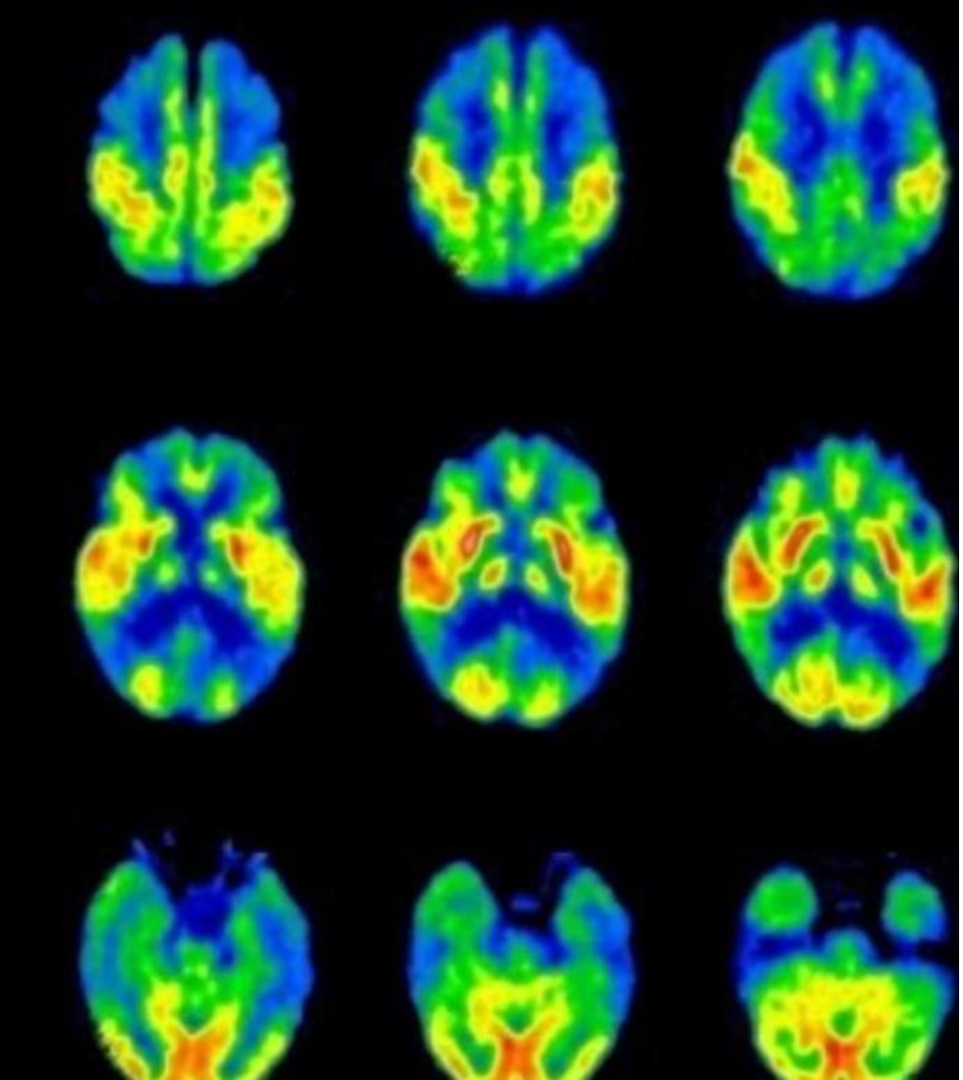
# MEDICAL IMAGING & BIG DATA PROJECT

Carnati Emanuele 781231
De Cicco Alessandro 834225
Mammana Lorenzo 807391





# Table of contents

# 1.
# Introduction

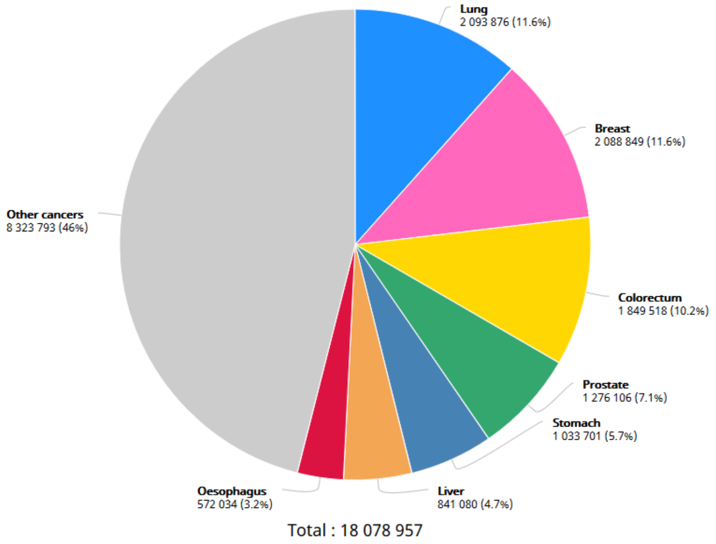



Fig. 1 Estimated number of new cases in 2018, worldwide, both sexes, all ages. GLOBOCAN 2018

# Why medical imaging?

*Cancer is the second leading cause of death globally.*

*While medical knowledge forms the basis of health practitioners decisions, medical imaging is a vital part of confirming any diagnosis.*

*It can also inform the doctor of internal problems that a basic external examination would fail to detect.*

# The importance of radiomics

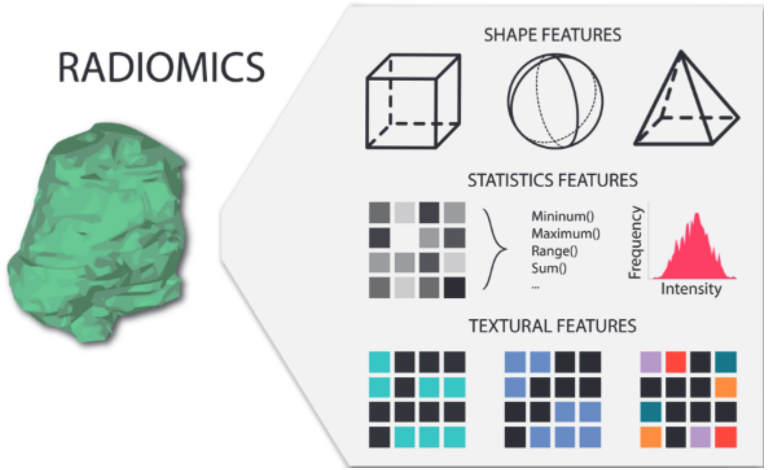



Fig. 2 Radiomics features [1]

*Radiomics is a field of medical study that aims to extract large amount of quantitative features from medical images.*

*These features have the potential to uncover disease characteristics hardly noticeable by the naked eye.*

# Machine learning applied to medical imaging

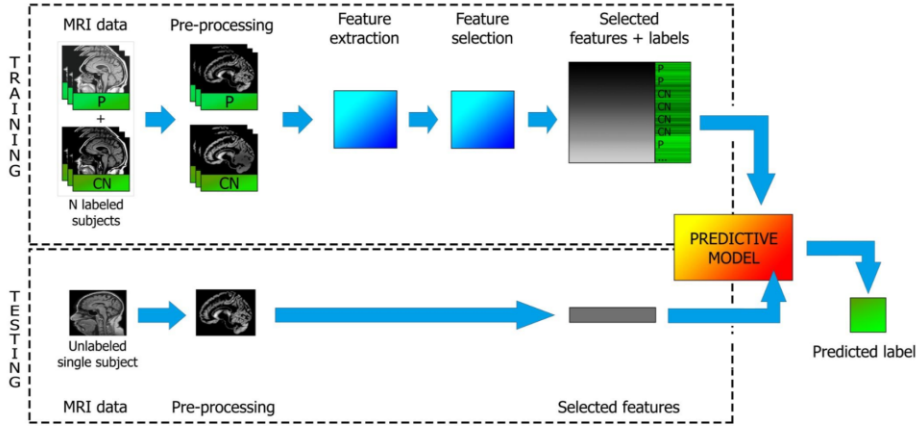


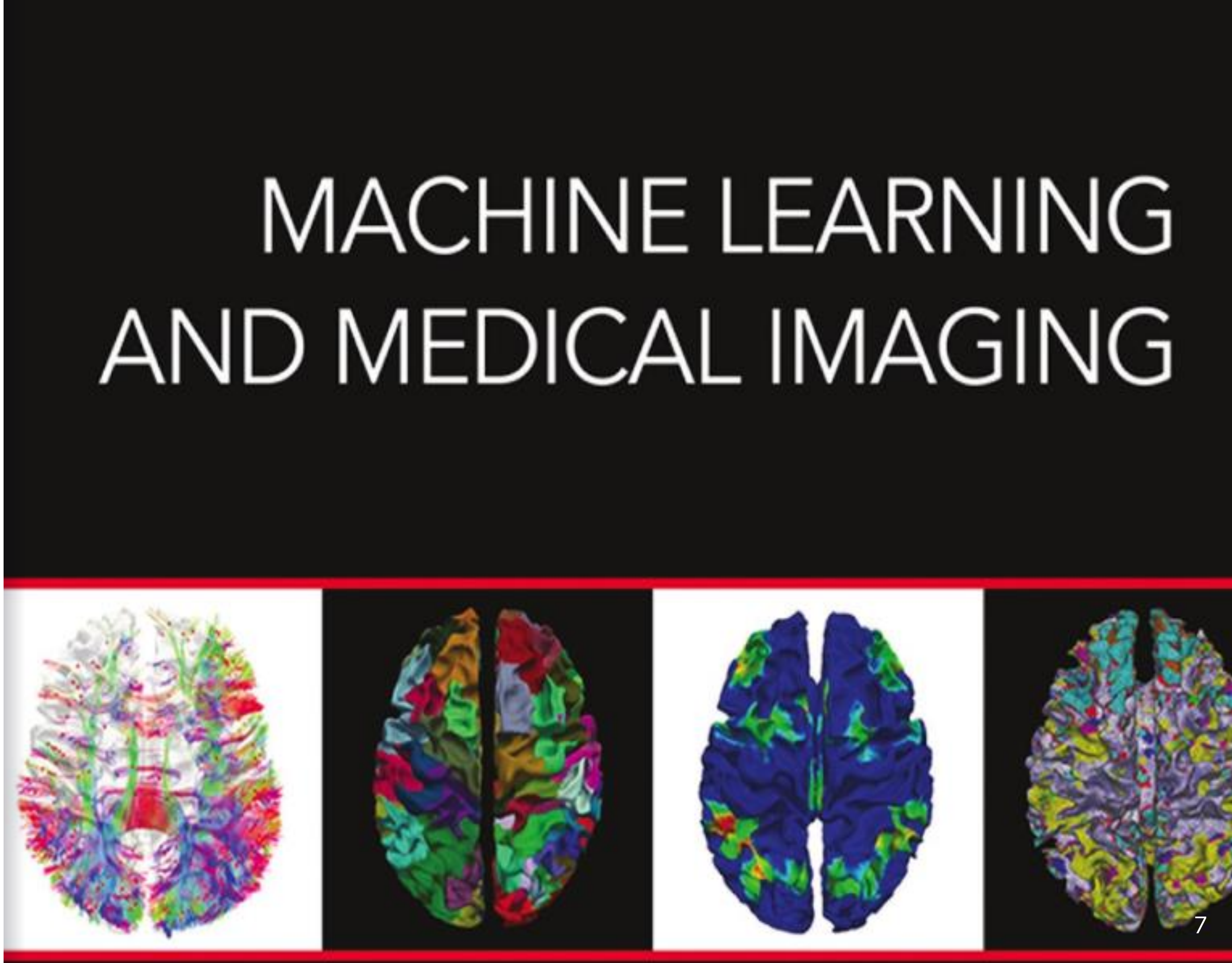

Fig. 3 Typical machine learning workflow. Christian Salvatore, Medical Imaging & Big Data course 2019, unimib

*A large number of features allow computer science experts to build machine learning models.*

*Machine learning in radiology is expected to have a substantial clinical impact with imaging examinations being routinely obtained in clinical practice, providing an opportunity to improve decision support in medical image interpretation.*

# 2.

Goal of the project & dataset

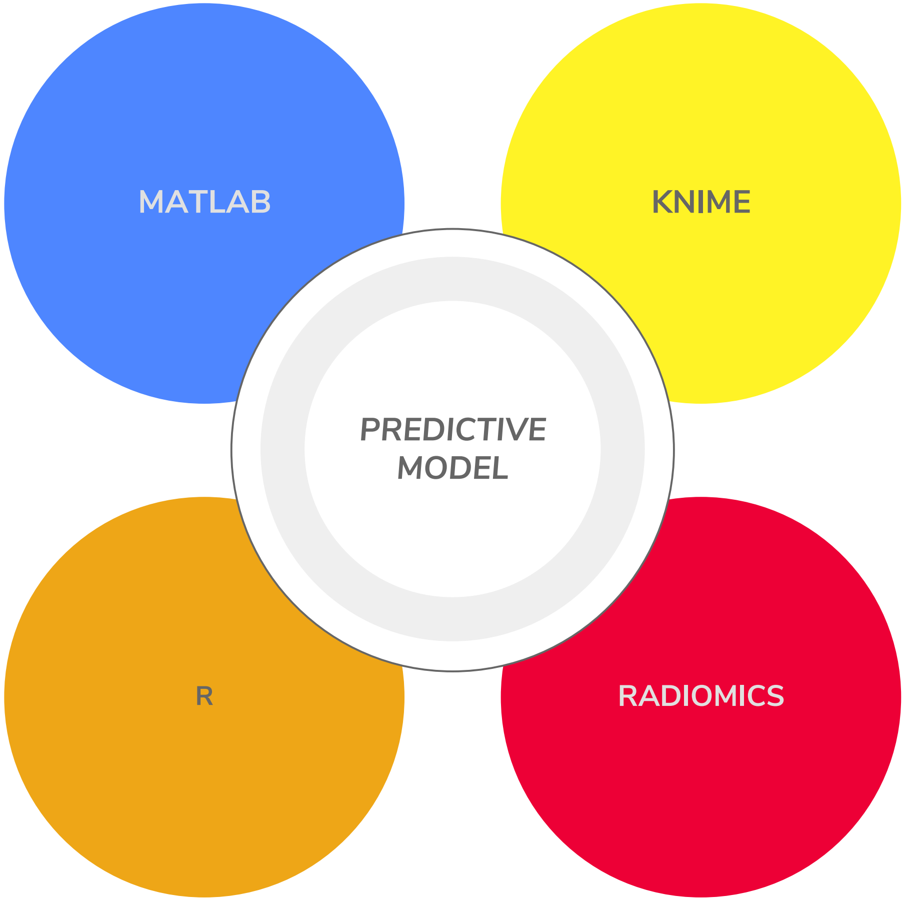


# Goal of the project

*The goal of the project is the creation of a predictive model capable to classify heterogeneous and homogeneous tumour lesions.*

*We have been working with:*

- *Matlab*
- *Knime*
- *R*



# Dataset

*44 images of tumour volumes in NIfTI format.*

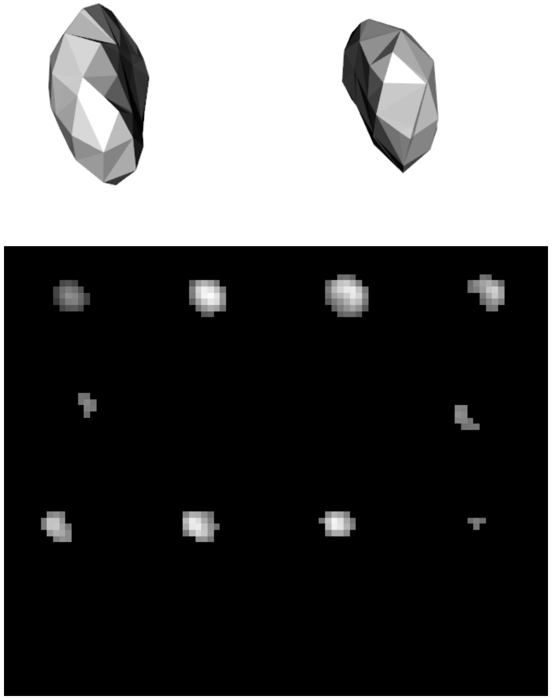
*Based on synthetic lesion constructed from real patient data.*

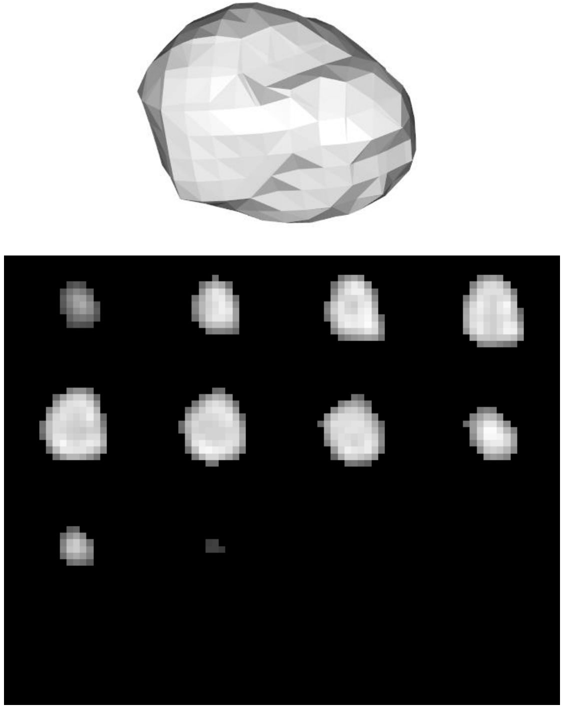
*Segmented using two different thresholds, one adaptive and one fixed [2].*

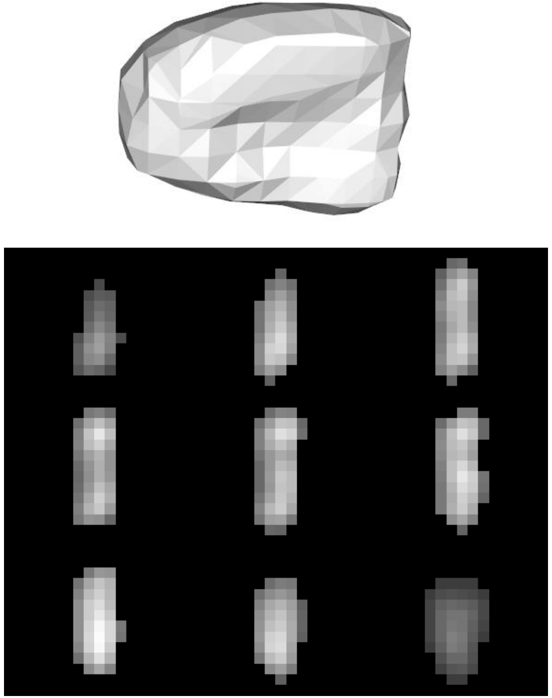
*Mostly homogeneous lesions (59%).*

*Some lesions very similar.*

# Dataset samples

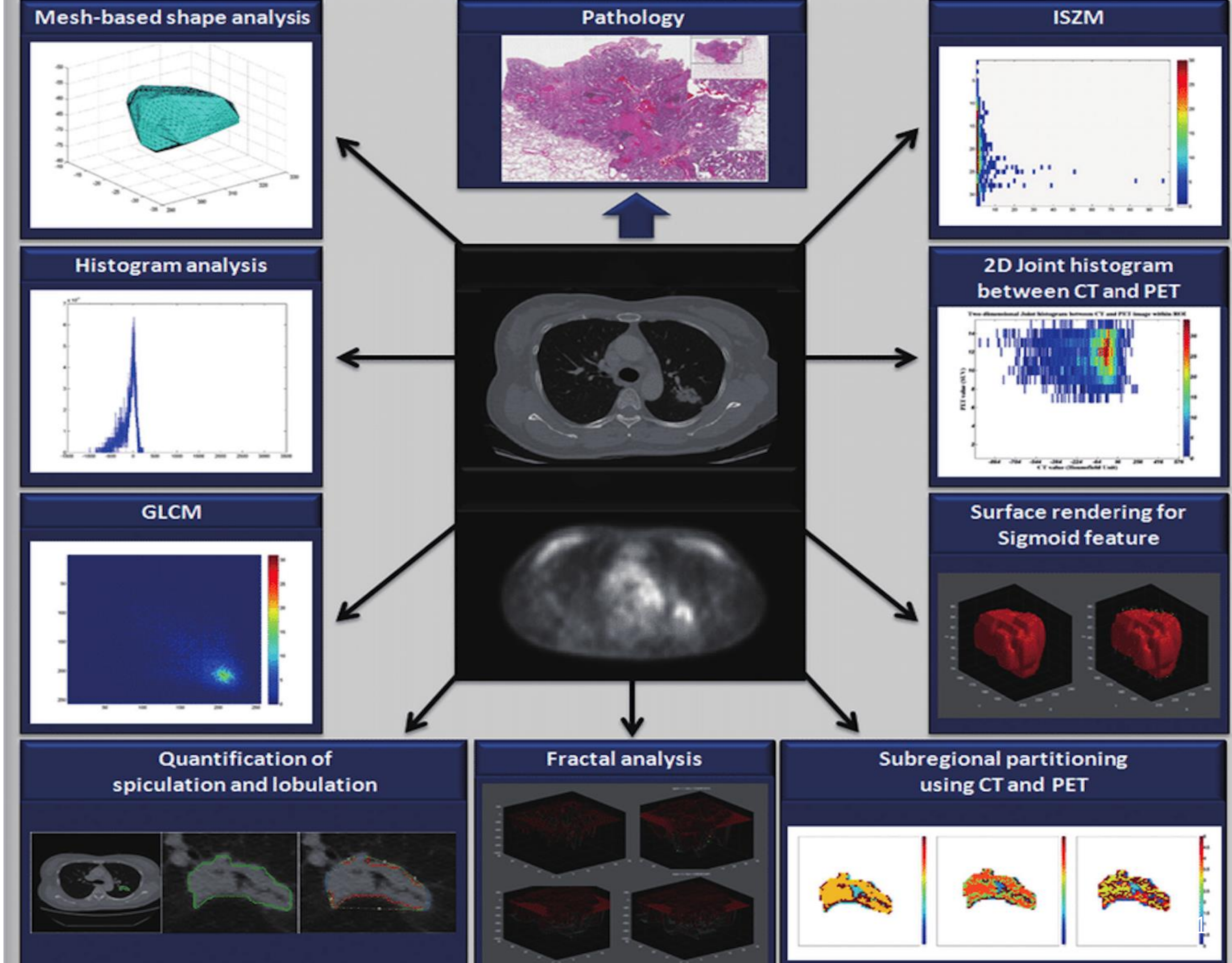






# 3.

Feature extraction



# Feature extraction with MATLAB

*For each input image we have extracted a set of 58 radiomics features mainly using the MATLAB tool developed and defined by Vallières et al. [3, 4, 5].*

*Each feature out of [0, 1] is also rescaled in this range using the formula:*

$$z_i = \frac{x_i - \min(x_i)}{\max(x_i) - \min(x_i)}$$

*Finally, each instance is labelled either:*

- *0 if homogeneous*
- *1 otherwise*

*The script* **create_dataset** *puts features and labels in one single matrix of size 44x59.*

# Features extracted

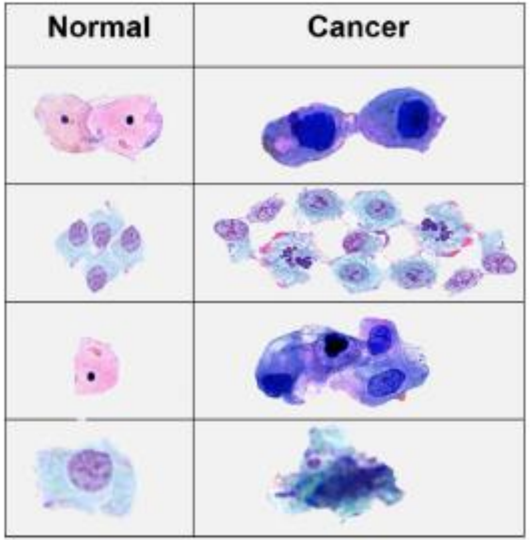
*Morphological features:* 5/59

Concern the features relating to lesions shape like surface, sphericity and metabolic target volume (MTV).

Usually, the more irregular the shape, the more aggressive is the tumour.

*Histogram-based features: 13/59*

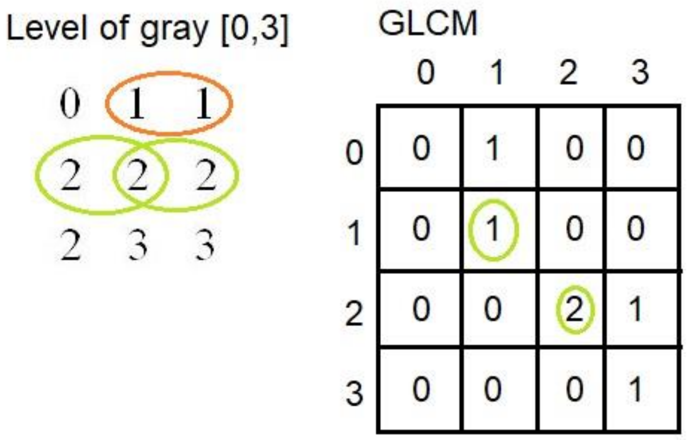
These features include various descriptive statistics indices relating voxel gray level of the lesions such as kurtosis, entropy, skewness, uniformity etc.



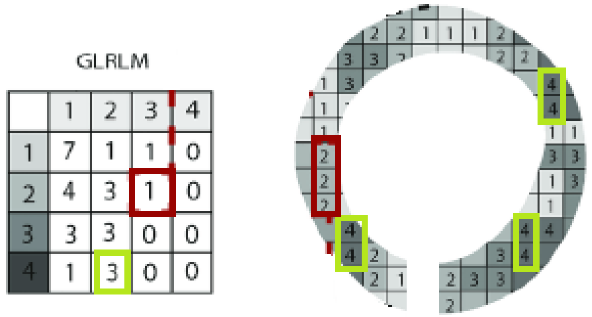

http://sphweb.bumc.bu.edu/otlt/MPH-Modules/PH/PH709_Cancer/PH709_Cancer7.html

# Features extracted

*GLCM (Gray Level Co-Occurence Matrix): 9/59*
This matrix quantifies the number of times that the combinations of level dimensions occur in two pixels. They are: energy, contrast, dissimilarity, homogeneity, correlation, autocorrelation, variance, sum average and entropy.




*GLRLM (Gray Level Run Length Matrix): 13/59*
Quantifies the length of number of consecutive pixels for each level of gray along one dimension.




Both GLCM and GLRLM are rotation dependent.

https://www.researchgate.net/figure/Principles-of-generating-texture-analysis-features-Principles-of-generating-the_fig2_320821651

# Features extracted

*GLSZM(Gray-Level Size Zone Matrix): 13/59*
*R*epresents the number of connected voxels that share the same gray level intensity.
These features are rotation indepedent.

| | | | |
|---|---|---|---|
| 1 | 2 | 3 | 4 |
| 1 | 3 | 4 | 4 |
| 3 | 2 | 2 | 2 |
| 4 | 1 | 4 | 1 |

| *Level* $g$ | *Size zone, s* 1 | 2 | 3 |
|---|---|---|---|
| 1 | 2 | 1 | 0 |
| 2 | 1 | 0 | 1 |
| 3 | 0 | 0 | 1 |
| 4 | 2 | 0 | 1 |

http://thibault.biz/Research/ThibaultMatrices/GLSZM/GLSZM.html

*NGTDM(Neighbouring Gray Tone Difference Matrix): 5/59*

*Quantifies the difference between a gray value and the average gray value of its neighbours.*

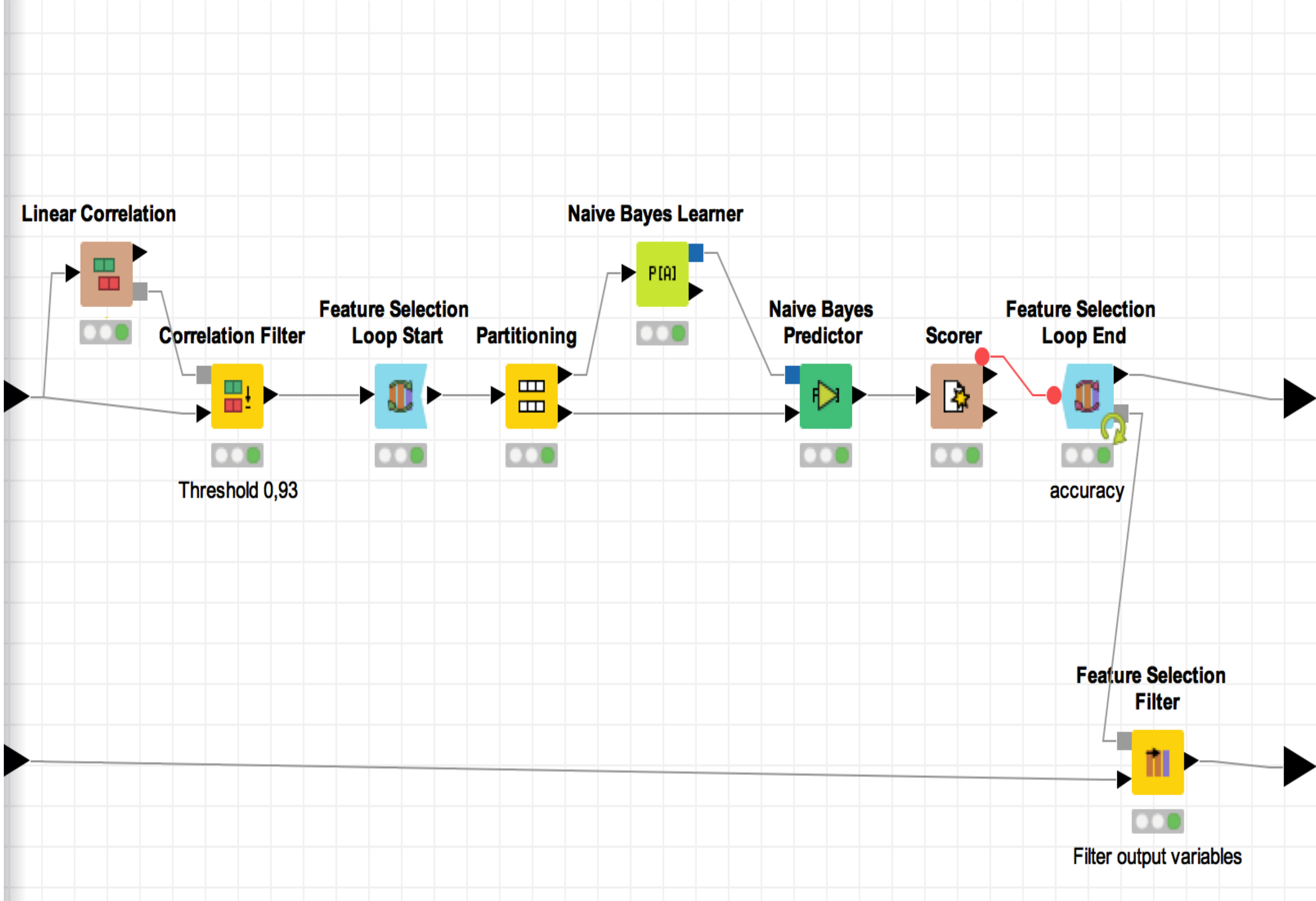
$$\mathbf{I} = \begin{bmatrix} 1 & 2 & 5 & 2 \\ 3 & 5 & 1 & 3 \\ 1 & 3 & 5 & 5 \\ 3 & 1 & 1 & 1 \end{bmatrix}$$

| $i$ | $n_i$ | $p_i$ | $s_i$ |
|---|---|---|---|
| 1 | 6 | 0.375 | 13.35 |
| 2 | 2 | 0.125 | 2.00 |
| 3 | 4 | 0.25 | 2.63 |
| 4 | 0 | 0.00 | 0.00 |
| 5 | 4 | 0.25 | 10.075 |

https://pyradiomics.readthedocs.io/en/latest/features.html

# 4.

## Feature Selection



# Feature selection

*The feature selection method is based on a naive bayesian forward selection implemented in KNIME.*

*A correlation filter is applied before performing this method. It reduces our feature set by removing highly correlated features (PCC > 0.9).*

*A smaller feature set reduces model complexity and tends to prevent overfitting.*

*After the iterations, 17 columns are chosen, beyond the label column.*

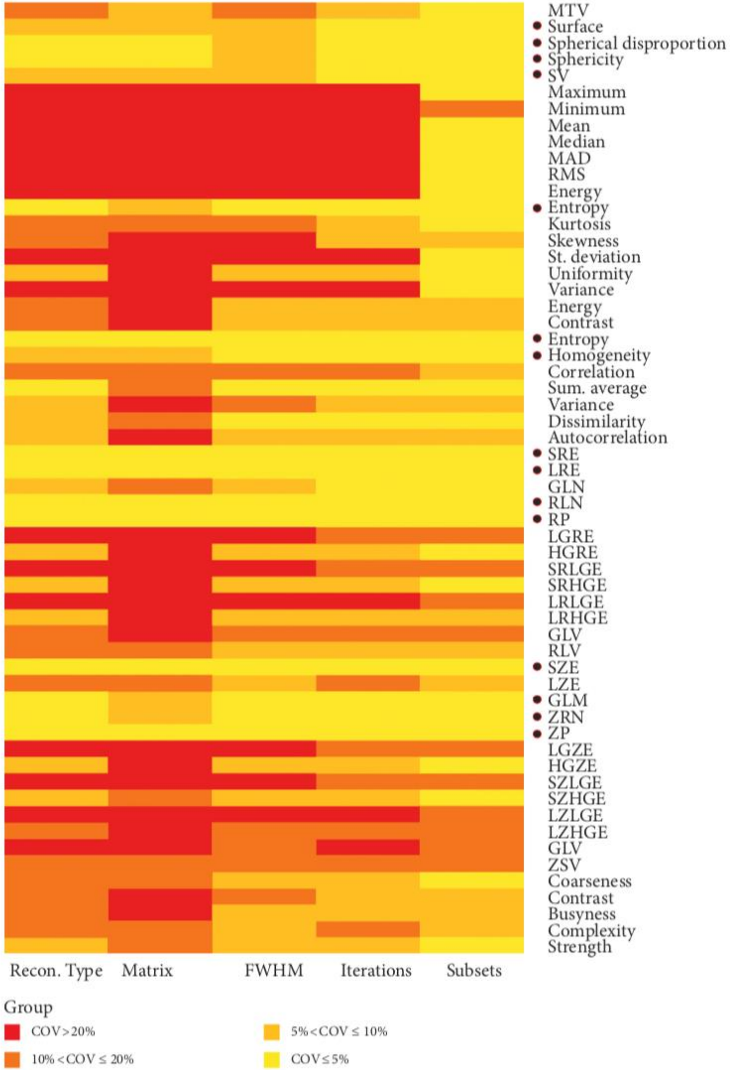
| Columns: 18 |
|---|
| mtv |
| spherical_disproportion |
| stv_ratio |
| max |
| energy |
| glcm_energy |
| glcm_homogeneity |
| glcm_correlation |
| glcm_sumaverage |
| glrlm_lgre |
| glrlm_glv |
| glrlm_rlv |
| glszm_sze |
| glszm_gln |
| glszm_szhge |
| ngtdm_coarseness |
| ngtdm_complexity |
| label |

UNIVERSITA' DEGLI STUDI DI MILANO BICOCCA

*We have then evaluated the stability of the selected features with literature [2] in order to analyze how much our model could work with unseen data from different experimental conditions.*

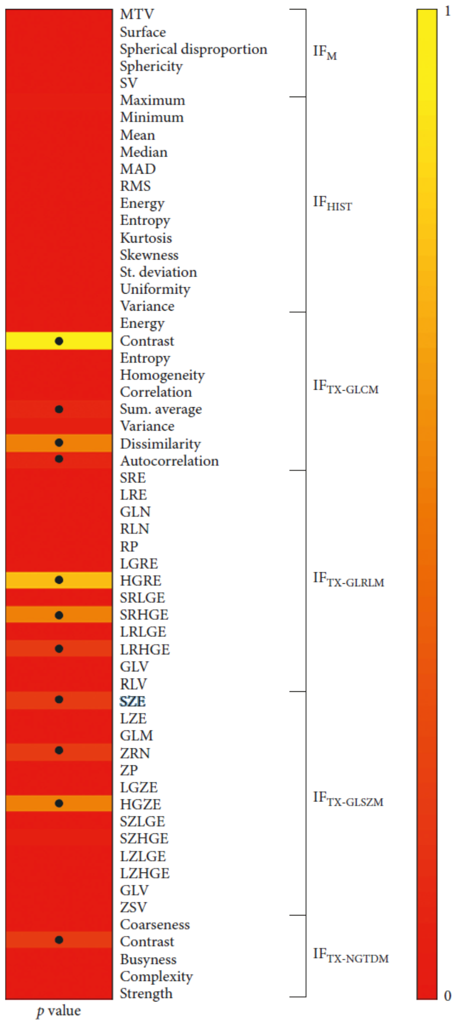
*W.r.t the chosen reconstruction method of the lesions our model contains 5 of the 15 features indicated to be stable (COV ≤ 10 %):*

- *Spherical disproportion*
- *SV*
- *GLCM_Homogeneity*
- *GLSZM_SZE*
- *GLSZM_GLM*



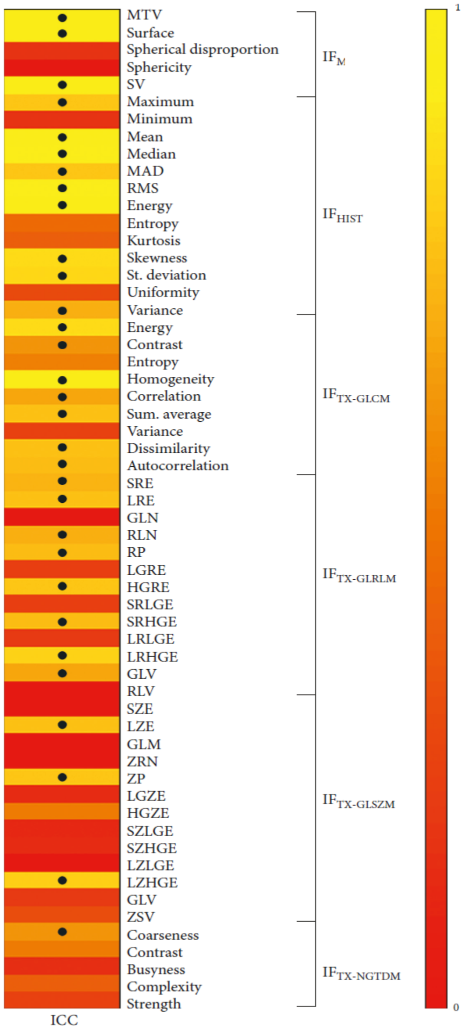
With respect to the chosen segmentation method of the lesions our model contains just 2 of the 11 features indicated to be stable (p value ≥ 0.05):

- GLCM_Sumaverage
- GLSZM_SZE



*Finally 10 out of 31 reliable features extracted using a test-retest experiment (ICC ≥ 0.6) have been selected.*

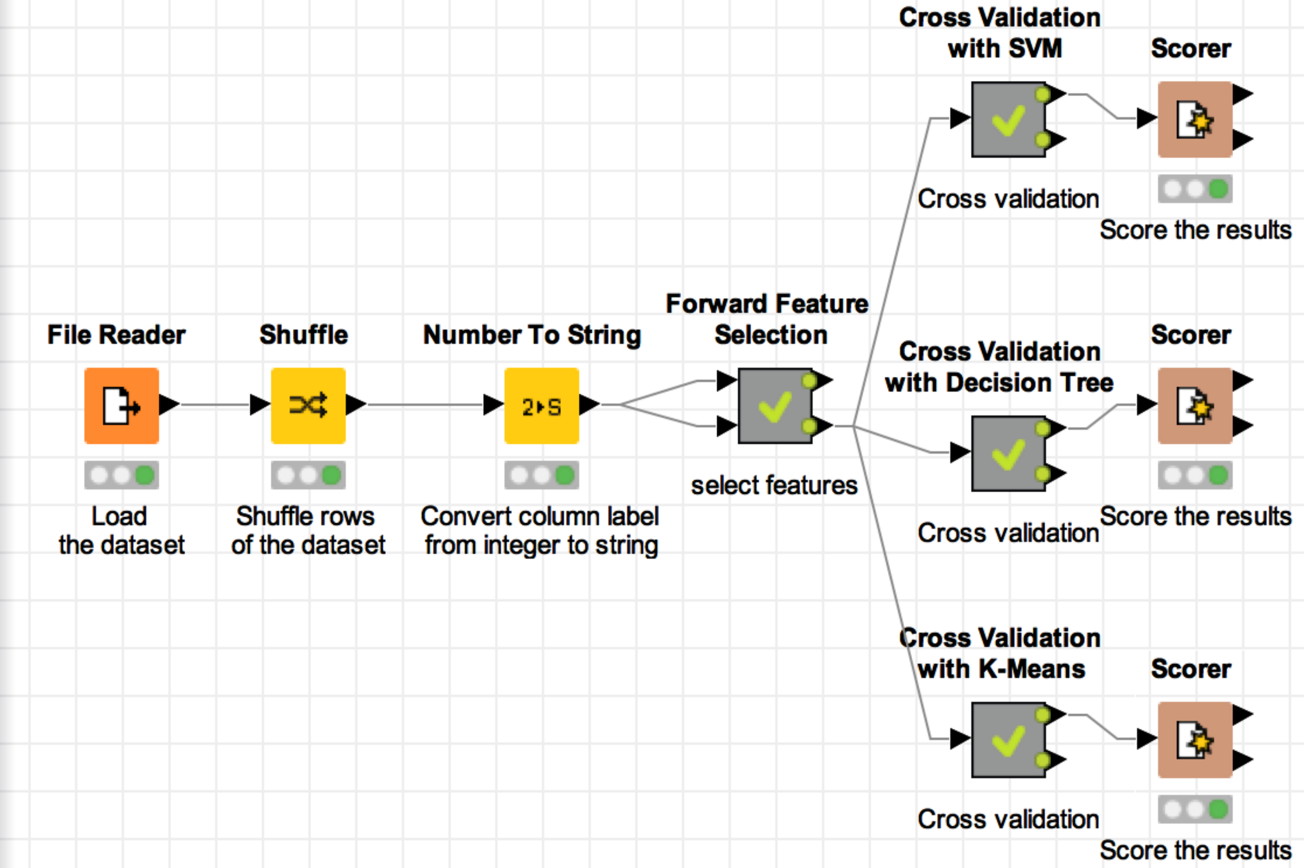
*The features selected are mainly good because they can be considered reliable over time, but the final model could have problem when new data come segmented or reconstructed in a different way of the original training data.*



# 5.

## Classification models

- *SVM*
- *Decision Tree*
- *k-Means Clustering*



# 4-fold cross validation

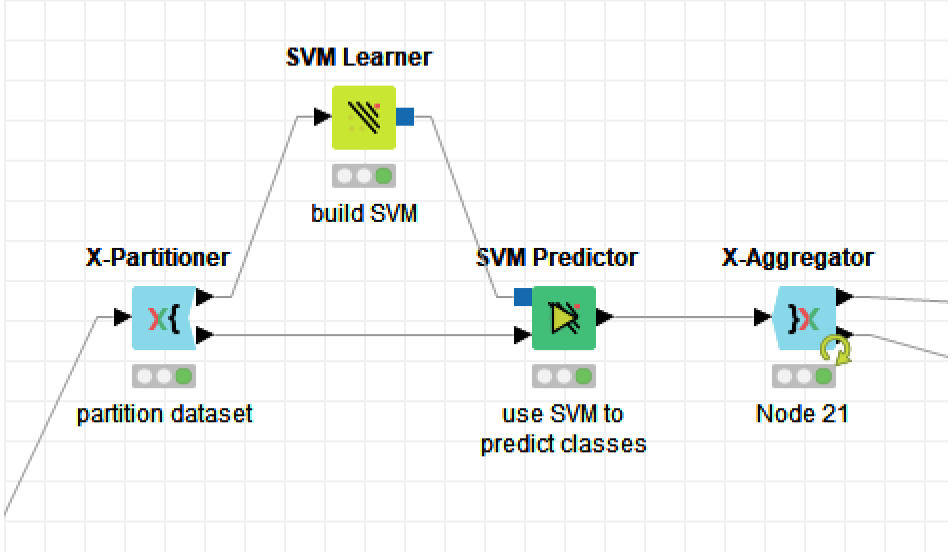

*All the 3 classification models implemented are based on stratified 4-fold cross validation.*
*The dataset is splitted into 4 subsets and iteratively 3 subsets are selected as training set while the remaining subset is used for the evaluation of the trained model.*

*The choice of 4 folds is based on the size of the dataset. With more folds, such a small dataset would have lead to insignificant test sets and thus uncorrected evaluations.*

*Advantages of cross validation:*

- *Each sample of the original dataset is used once for validation.*
- *All samples being used for both training and testing phases.*

# SVM – feature selection




*RBF-kernel with $\gamma = 0.5$ and $C = 1$*

| label \ Prediction (label) | 0 | 1 |
|---|---|---|
| 0 | 25 | 1 |
| 1 | 8 | 10 |

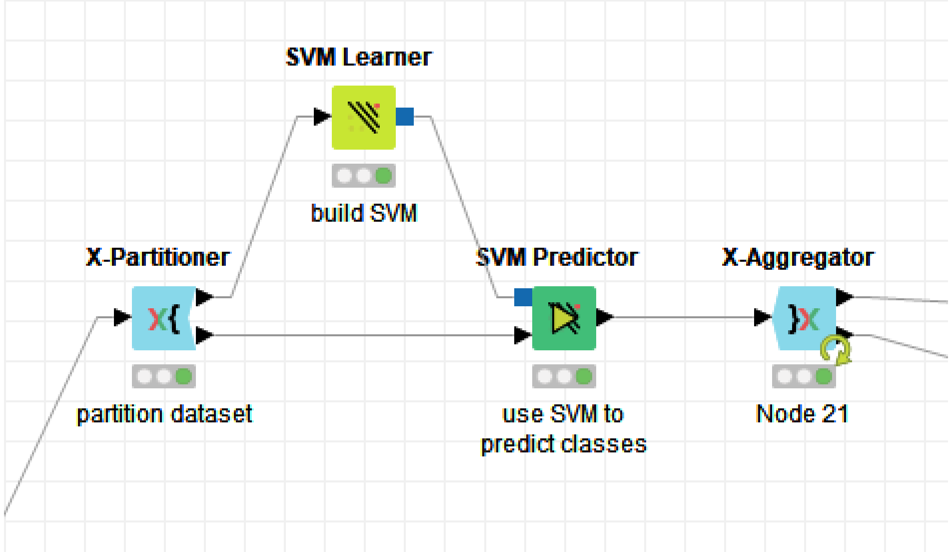
Correct classified: 35 Wrong classified: 9

Accuracy: 79,545 % Error: 20,455 %

Cohen's kappa (κ) 0,55

| Row ID | TruePositives | FalsePositives | TrueNegatives | FalseNegatives | Recall | Precision | Sensitivity | Specifity | F-measure | Accuracy | Cohen's kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 25 | 8 | 10 | 1 | 0.962 | 0.758 | 0.962 | 0.556 | 0.847 | ? | ? |
| 1 | 10 | 1 | 25 | 8 | 0.556 | 0.909 | 0.556 | 0.962 | 0.69 | ? | ? |
| Overall | ? | ? | ? | ? | ? | ? | ? | ? | ? | 0.795 | 0.55 |

# SVM – all features




*RBF-kernel with $\gamma = 0.5$ and $C = 1$*

| label \ Prediction (label) | 0 | 1 |
|---|---|---|
| 0 | 26 | 0 |
| 1 | 13 | 5 |

Correct classified: 31    Wrong classified: 13

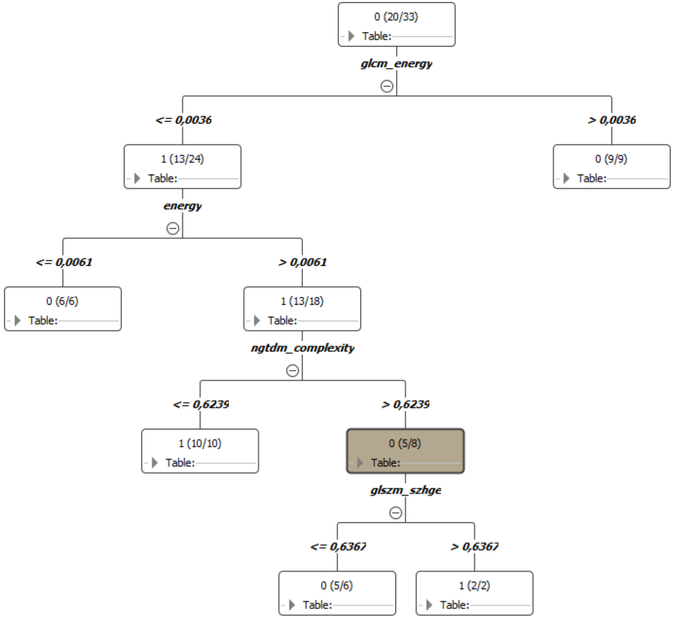
Accuracy: 70,455 %    Error: 29,545 %

Cohen's kappa (κ)

| Row ID | TruePositives | FalsePositives | TrueNegatives | FalseNegatives | Recall | Precision | Sensitivity | Specifity | F-measure | Accuracy | Cohen's kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 26 | 13 | 5 | 0 | 1 | 0.667 | 1 | 0.278 | 0.8 | ? | ? |
| 1 | 5 | 0 | 26 | 13 | 0.278 | 1 | 0.278 | 1 | 0.435 | ? | ? |
| Overall | ? | ? | ? | ? | ? | ? | ? | ? | ? | 0.705 | 0.312 |

UNIVERSITA' DEGLI STUDI DI MILANO BICOCCA

# Decision Tree – feature selection




| label \ Prediction (label) | 0 | 1 |
|---|---|---|
| 0 | 18 | 8 |
| 1 | 6 | 12 |

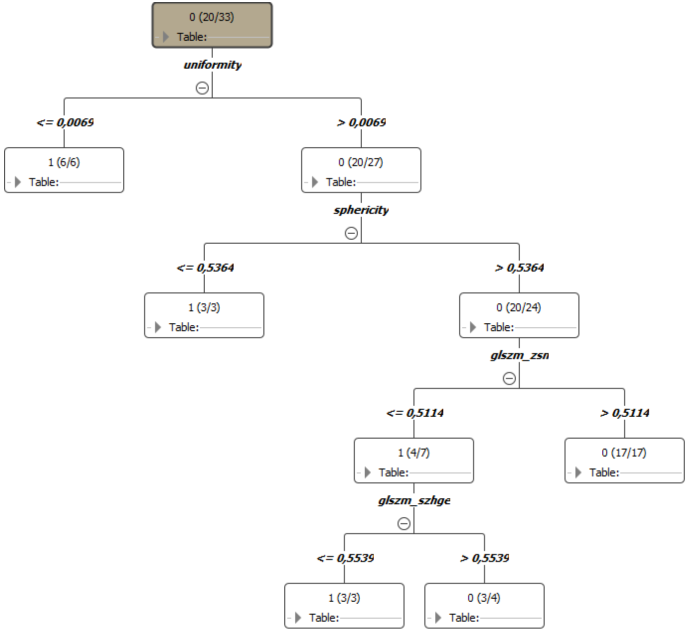
Correct classified: 30 Wrong classified: 14

Accuracy: 68,182 % Error: 31,818 %

Cohen's kappa (κ) 0,353

| Row ID | TruePositives | FalsePositives | TrueNegatives | FalseNegatives | Recall | Precision | Sensitivity | Specifity | F-measure | Accuracy | Cohen's kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 18 | 6 | 12 | 8 | 0.692 | 0.75 | 0.692 | 0.667 | 0.72 | ? | ? |
| 1 | 12 | 8 | 18 | 6 | 0.667 | 0.6 | 0.667 | 0.692 | 0.632 | ? | ? |
| Overall | ? | ? | ? | ? | ? | ? | ? | ? | ? | 0.682 | 0.353 |

# Decision Tree – all features




| label \ Prediction (label) | 0 | 1 |
|---|---|---|
| 0 | 14 | 12 |
| 1 | 8 | 10 |

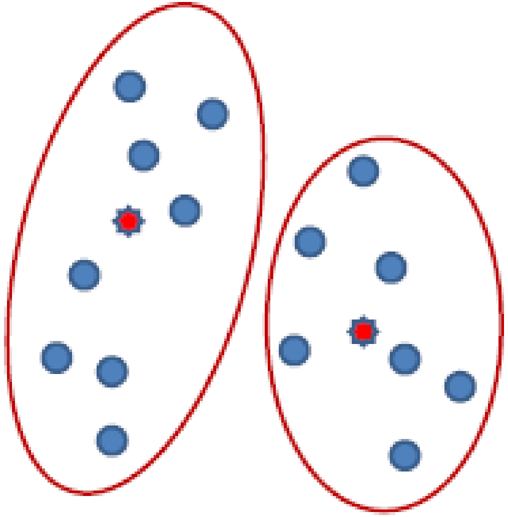
Correct classified: 24 Wrong classified: 20

Accuracy: 54,545 % Error: 45,455 %

Cohen's kappa (κ)

| Row ID | TruePositives | FalsePositives | TrueNegatives | FalseNegatives | Recall | Precision | Sensitivity | Specifity | F-measure | Accuracy | Cohen's kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 14 | 8 | 10 | 12 | 0.538 | 0.636 | 0.538 | 0.556 | 0.583 | ? | ? |
| 1 | 10 | 12 | 14 | 8 | 0.556 | 0.455 | 0.556 | 0.538 | 0.5 | ? | ? |
| Overall | ? | ? | ? | ? | ? | ? | ? | ? | ? | 0.545 | 0.091 |

# K-means clustering – feature selection



*2 cluster, 99 iterations*

| label \ prediction(label) | 0 | 1 |
|---|---|---|
| 0 | 18 | 8 |
| 1 | 9 | 9 |

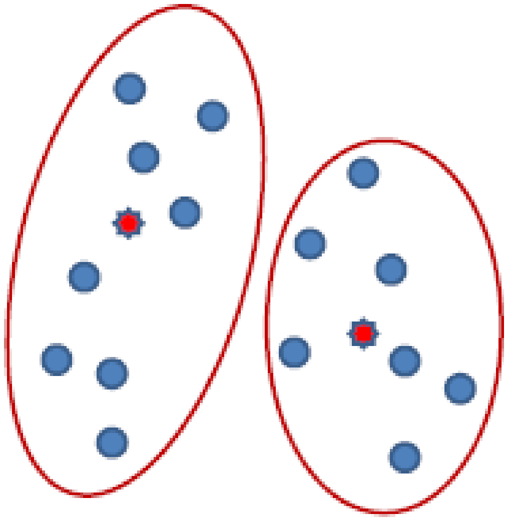
Correct classified: 27 Wrong classified: 17

Accuracy: 61,364 % Error: 38,636 %

Cohen's kappa (κ)

| Row ID | TruePositives | FalsePositives | TrueNegatives | FalseNegatives | Recall | Precision | Sensitivity | Specifity | F-measure | Accuracy | Cohen's kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 18 | 9 | 9 | 8 | 0.692 | 0.667 | 0.692 | 0.5 | 0.679 | ? | ? |
| 1 | 9 | 8 | 18 | 9 | 0.5 | 0.529 | 0.5 | 0.692 | 0.514 | ? | ? |
| Overall | ? | ? | ? | ? | ? | ? | ? | ? | ? | 0.614 | 0.194 |

# K-means clustering – all features



*2 cluster, 99 iterations*

| label \ prediction(label) | 0 | 1 |
|---|---|---|
| 0 | 21 | 5 |
| 1 | 12 | 6 |

| | |
|---|---|
| Correct classified: 27 | Wrong classified: 17 |
| Accuracy: 61,364 % | Error: 38,636 % |
| Cohen's kappa (κ) | |

| Row ID | TruePositives | FalsePositives | TrueNegatives | FalseNegatives | Recall | Precision | Sensitivity | Specifity | F-measure | Accuracy | Cohen's kappa |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 21 | 12 | 6 | 5 | 0.808 | 0.636 | 0.808 | 0.333 | 0.712 | ? | ? |
| 1 | 6 | 5 | 21 | 12 | 0.333 | 0.545 | 0.333 | 0.808 | 0.414 | ? | ? |
| Overall | ? | ? | ? | ? | ? | ? | ? | ? | ? | 0.614 | 0.15 |

# Results analysis

*The best model in term of accuracy is the SVM which achieves an accuracy of 79.5 %. Despite the high accuracy, the model is not great because fails to classify positive examples (label 1), this is clearly visible since the sensitivity is very low (0.556).*

*The decision tree performs a little better on the positive class but fails a lot more on the negative class.*

*Finally the k-means clustering shows very low performances on both positive and negative classes.*

*All the models except k-means have beneficiated of feature selection with an accuracy increase of circa 10 %.*

# THANK YOU FOR YOUR ATTENTION

# *Bibliography*


*[1] Coroller, T. P. (2017). Combining data science and medical imaging: Advancing cancer precision medicine with radiomics. Maastricht: Datawyse / Universitaire Pers Maastricht*

*[2] Francesca Gallivanone, Matteo Interlenghi, Daniela D'Ambrosio, Giuseppe Trifirò, and Isabella Castiglioni, "Parameters Influencing PET Imaging Features: A Phantom Study with Irregular and Heterogeneous Synthetic Lesions," Contrast Media & Molecular Imaging, vol. 2018*

*[3] Vallières, M. et al. (2015). A radiomics model from joint FDG-PET and MRI texture features for the prediction of lung metastases in soft-tissue sarcomas of the extremities. Physics in Medicine and Biology, 60(14), 5471-5496. doi:10.1088/0031-9155/60/14/5471*

*[4] Zhou, H., Vallières, M., Bai, H.X. et al. (2017). MRI features predict survival and molecular markers in diffuse lower-grade gliomas. Neuro-Oncology, 19(6), 862-870. doi:10.1093/neuonc/now256*

*[5] Vallière, M. et al. (2017). Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Scientific Reports, 7:10117. doi:10.1038/s41598-017-10371-5*