

Radiomics in Prostate Cancer: a Machine Learning Approach

Carlo Mengucci^{1,2}, Daniel Remondini¹, Giuseppe Salvaggio⁴, Giuseppe Merennino⁴, Maurizio Marrale³

¹Department of Physics and Astronomy, UNIBO and INFN-BO

²Department of Agricultural and Food Sciences, UNIBO and INFN-BO

³Department of Biomedicine, Neurosciences and Advanced Diagnostics, UNIPA

⁴Department of Physics and Chemistry “E. Segrè”, UNIPA and INFN-CT



Next-AIM Meeting, 13th -15th February 2023

Participants UNIPA (INFN CT) and UNIBO



UNIVERSITÀ
DEGLI STUDI
DI PALERMO



Azienda Ospedaliera Universitaria
Policlinico Paolo Giaccone
di Palermo



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

Policlinico S. Orsola-Malpighi

Marrale Maurizio
Tomarchio Elio
Collura Giorgio
Bartolotta Antonio
D'Oca Maria Cristina

Giuseppe Salvaggio
Massimo Midiri
Roberto Lagalla
Cesare Gagliardo
Giuseppe Merennino

Daniel Remondini
Carlo Mengucci

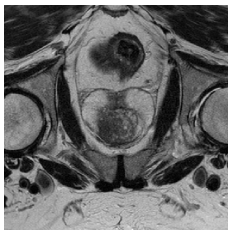
Objective: evaluate the predictive power of radiomics features
from ROIs segmented by experts

WP3 Task 3: Radiomics in prostate cancer

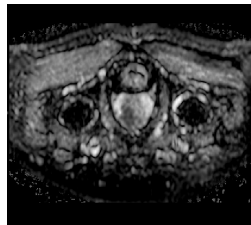
Tumor classification on MR images of prostate by ML algorithms

Prostate cancer diagnosis is performed by using T2w- and ADC MR images

T2w-MRI



ADC-MRI



Classification with PIRADS (“Prostate Imaging Reporting and Data System”) scale

American College of Radiology website. Prostate imaging reporting & data system (PI-RADS). <https://www.acr.org/media/ACR/Files/RADS/Pi-RADS/PI-RADS-V2-1.pdf?la=en>.

Boesen L, Chabanova E, Logager V, Balslev I, Thomsen HS. Apparent diffusion coefficient ratio correlates significantly with prostate cancer Gleason score at final pathology. J Magn Reson Imaging. 2015;42:446-453.

PI-RADS

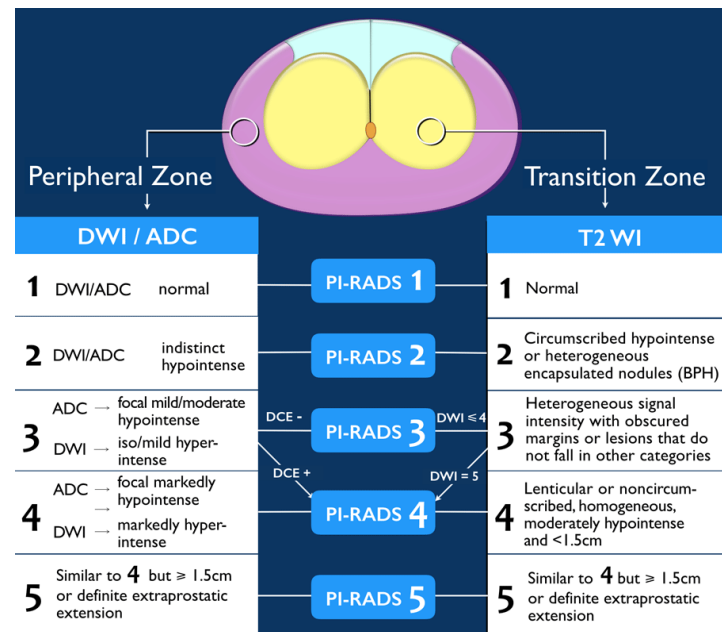
PI-RADS 1 = Very low (clinically significant cancer highly unlikely)

PI-RADS 2 = Low (clinically significant cancer unlikely)

PI-RADS 3 = Intermediate (clinically significant cancer equivocal)

PI-RADS 4 = High (clinically significant cancer likely)

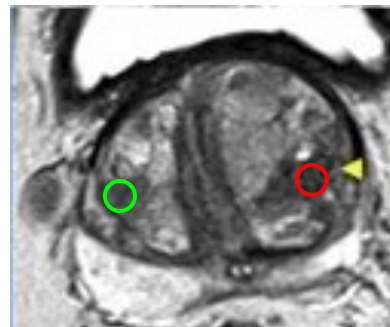
PI-RADS 5 = Very high (clinically significant cancer highly likely)



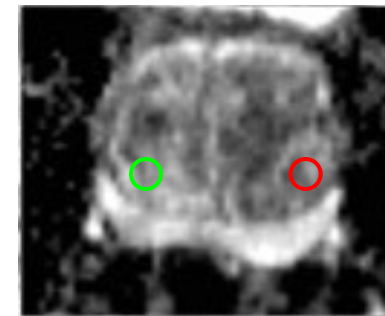
Step 1: FIRST analysis

- 46 Patients recruited at the Polyclinic University Hospital of Palermo (26 benign lesion and 20 tumoral tissues)
- T₂-weighted and ADC images
- Histological results of biopsy

T2w



ADC



Identification of a ROI in the **lesion region** and another ROI in **healthy tissue** as control

Objective: evaluate the predictive power of radiomics features from ROIs segmented by experts



Azienda Ospedaliera Universitaria
Policlinico Paolo Giaccone
di Palermo

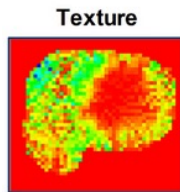
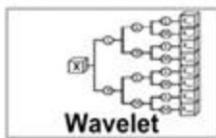
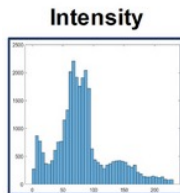


Feature extraction

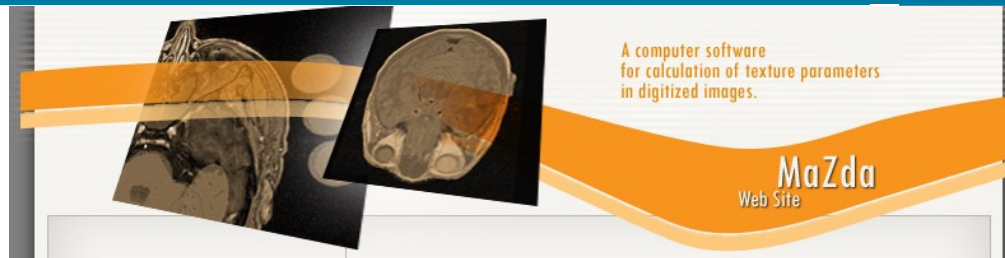
MaZda is a computer program for calculation of texture parameters (features) in digitized images.

Small ROIs with cross shape

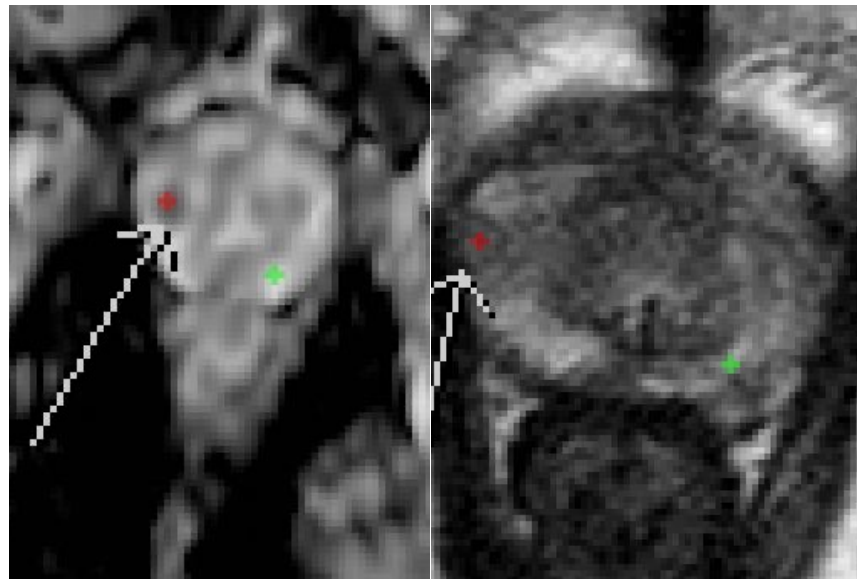
About 300 features (shape, histogram, texture, wavelet-filters)



Challenges: Not many images, unbalanced classes, possibly redundant features (about 300)



Szczypinski, P.M., Klepaczko, A., MaZda - A framework for biomedical image texture analysis and data exploration, Biomedical Texture Analysis: Fundamentals, Tools and Challenges, 2017, pp. 315-347

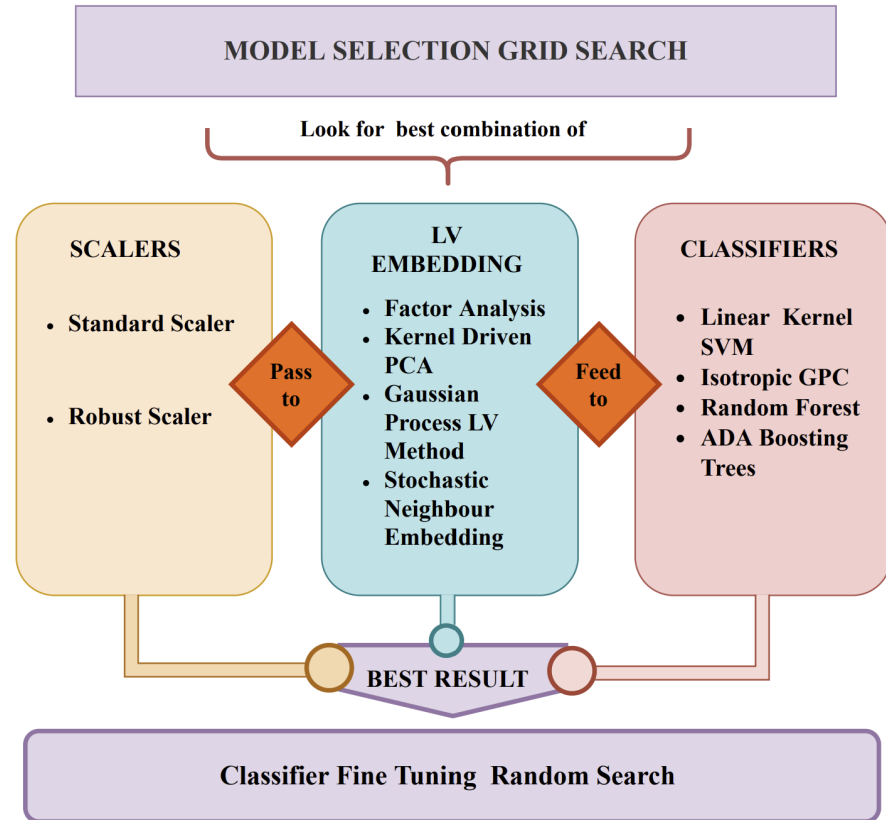


Model selection

Data scaling

Dimensionality reduction

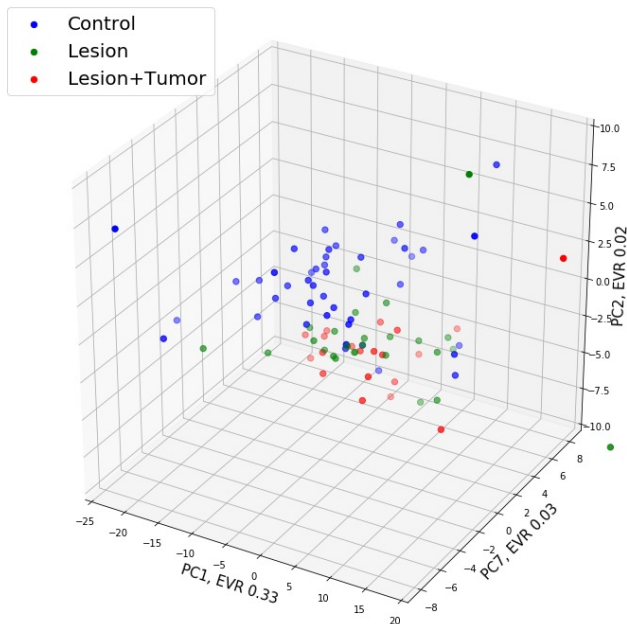
Classification



PREVIOUS RESULTS

3D Embedding of Dataset + Loadings

3D PCA, Robust Scaled Features



PC1: S(4,4)DifEntrp,
S(4,4)SumEntrp,
S(3,-3)SumVarnc,
S(4,0)AngScMom,
S(0,5)SumVarnc

PC7: S(0,5)SumAverg,
S(3,0)SumAverg,
S(0,5)AngScMom,
S(4,-4)AngScMom,
Perc. 10%

PC9: Perc. 01%,
S(5,-5)DifVarnc,
S(3,-3)AngScMom,
S(4,4)InvDfMom,
S(5,-5)AngScMom

Best model:

- Scaler: Robust Scaler
- Embedding: 15 components
linear kernel PCA
- Classifier: Random Forest

Crossvalidated Performance:
average accuracy ~ 0.68

	Control	Lesion	Lesion+Tumor
Control	38	5	3
Lesion	9	5	12
Lesion+Tumor	4	4	12

Adding knowledge about features

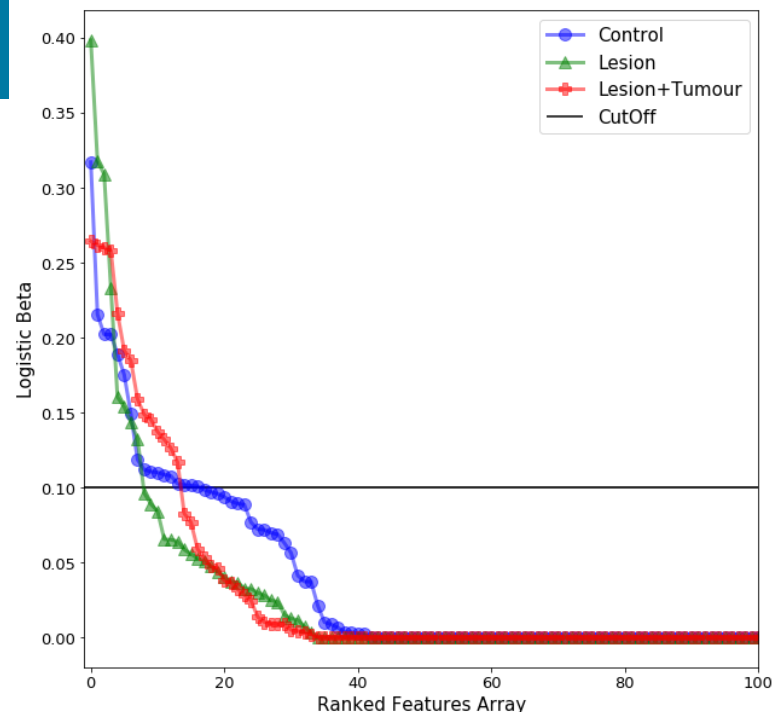
Best Solution: LR with EN mixed penalty

- L1-L2 mixed Elastic Net penalization gives best balance between induced sparsity (feature selection) and retained information for predictive performances
- Best model for prediction and most representative features for each class are simultaneously found
- Uninformative features are pruned

Controls Features: Mean, Skewness, Perc.01%,Perc.10%, S(1,0)AngScMom, S(1,0)SumAverg, S(0,1)SumAverg, S(1,1)SumAverg, S(1,-1)SumAverg, S(2,0)SumAverg, S(0,2)SumAverg, S(2,-2)SumAverg, S(3,0)SumAverg, S(3,-3)SumVarnc, S(5,0)InvDfMom, S(5,0)SumAverg, 135drLngREmph

Lesion Features: Skewness, S(4,-4)SumVarnc, S(5,0)AngScMom, S(0,5)InvDfMom, S(5,-5)Contrast, Tet4, WavEnLLs-1, WavEnLLs-2

Lesion+Tumour Features: Perc.01%, Perc.10%, S(3,-3)SumVarnc, S(4,-4)AngScMom, S(4,-4)Contrast, S(5,5)AngScMom, S(5,5)DifVarnc, S(5,-5)Contrast, S(5,-5)SumVarnc, S(5,-5)DifVarnc, GrSkewness, WavEnLLs-1, WavEnHHs-1, WavEnLLs-2



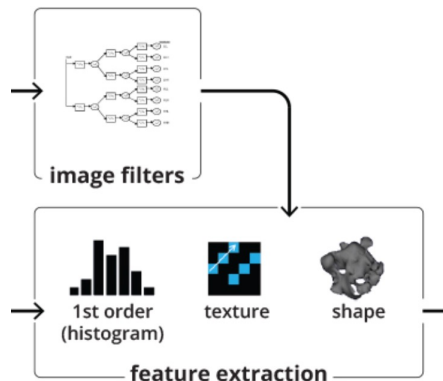
Cross Validated performance: ~ 0.72 average accuracy

	Control	Lesion	Lesion+Tumor
Control	39	4	3
Lesion	6	8	12
Lesion+Tumor	2	8	10

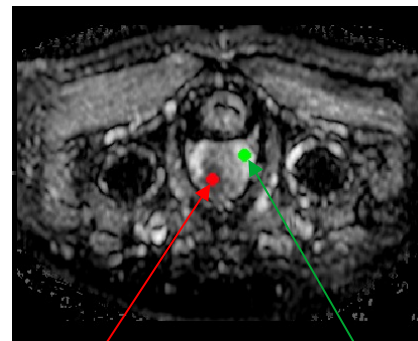
Step 2: Second analysis

Dataset

- 74 Tumors (PIRADS 4 and 5) & 74 Controls tissues (balanced classes), T2w and ADC MR data
- Circular ROI in a singular slice
- Features' extraction through pyradiomics
- ca. 800 starting features

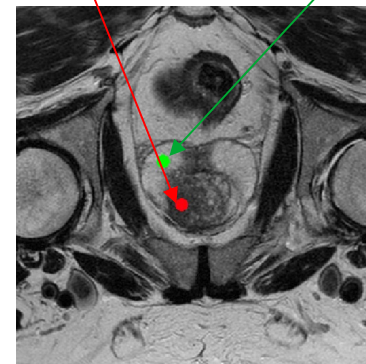


ROI identification



Lesion

Health tissue

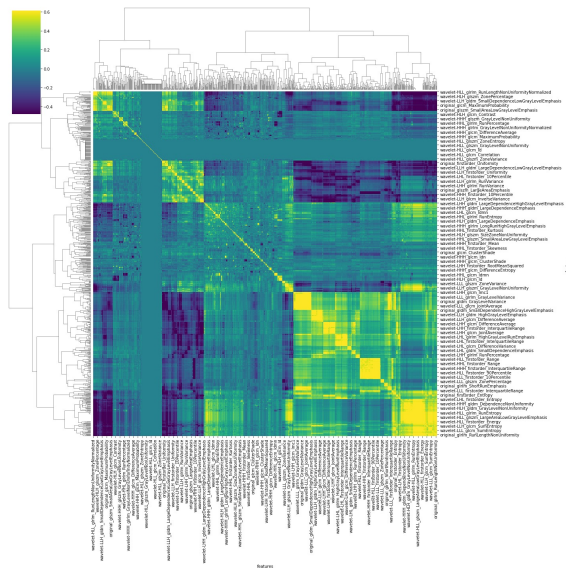


van Griethuysen, J. J. M., Fedorov, A., Parmar, C., Hosny, A., Aucoin, N., Narayan, V., Beets-Tan, R. G. H., Fillon-Robin, J. C., Pieper, S., Aerts, H. J. W. L. (2017). Computational Radiomics System to Decode the Radiographic Phenotype. *Cancer Research*, 77(21), e104–e107.

MACHINE LEARNING METHODS

PCA+fine tune random forest (f1, accuracy 0.76 - 0.8)

BORUTA feature selection + fine tuned random forest (f1, accuracy 0.86 - 0.9)



Evolution of the first pipeline with a more robust feature selection that does not depend on the various user-defined thresholds

BORUTA Robust Features

```
original_firstorder_10Percentile
original_firstorder_90Percentile
original_firstorder_Maximum
original_firstorder_Mean
original_firstorder_Median
original_firstorder_Minimum
original_firstorder_RootMeanSquare
d
wavelet-
HLL_firstorder_10Percentile
wavelet-
HLL_firstorder_90Percentile
wavelet-
HLL_firstorder_InterquartileRange
wavelet-HLL_firstorder_Maximum
wavelet-
HLL_firstorder_MeanAbsoluteDeviati
on
wavelet-HLL_firstorder_Mean
wavelet-HLL_firstorder_Median
wavelet-
HLL_firstorder_RobustMeanAbsoluteD
eviation
wavelet-HLL_firstorder_Variance
wavelet-
HLH_firstorder_MeanAbsoluteDeviati
on
wavelet-
HLH_firstorder_RobustMeanAbsoluteD
eviation
```

```
wavelet-
HLH_firstorder_Variance
wavelet-
HHL_firstorder_10Percentile
wavelet-
HHL_firstorder_90Percentile
wavelet-
HHL_firstorder_MeanAbsoluteDe
viation
wavelet-
HHL_firstorder_Variance
wavelet-
HHH_firstorder_Maximum
wavelet-
HHH_firstorder_MeanAbsoluteDe
viation
wavelet-HHH_firstorder_Range
wavelet-
HHH_firstorder_RobustMeanAbso
luteDeviation
wavelet-
HHH_firstorder_Variance
wavelet-
LLL_firstorder_10Percentile
wavelet-
LLL_firstorder_90Percentile
wavelet-
LLL_firstorder_Maximum
wavelet-LLL_firstorder_Mean
wavelet-LLL_firstorder_Median
wavelet-
LLL_firstorder_Minimum
wavelet-
LLL_firstorder_RootMeanSquare
d
```

Future analyses

- Analyses using the 2nd protocol on T2w-MRI data from PIRADS3 data (at least 40 patients)
- Classification by using all PIRADS (3, 4 and 5) data
- Analyses on ADC data and combination with T2 data
- Analyses with further classification algorithms