RADIOMICS IN PROSTATE CANCER: A MACHINE LEARNING APPROACH

Carlo Mengucci[1,2], Daniel Remondini[1], Maurizio Marrale[3], Giuseppe Salvaggio[4], Giuseppe Merennino[4]

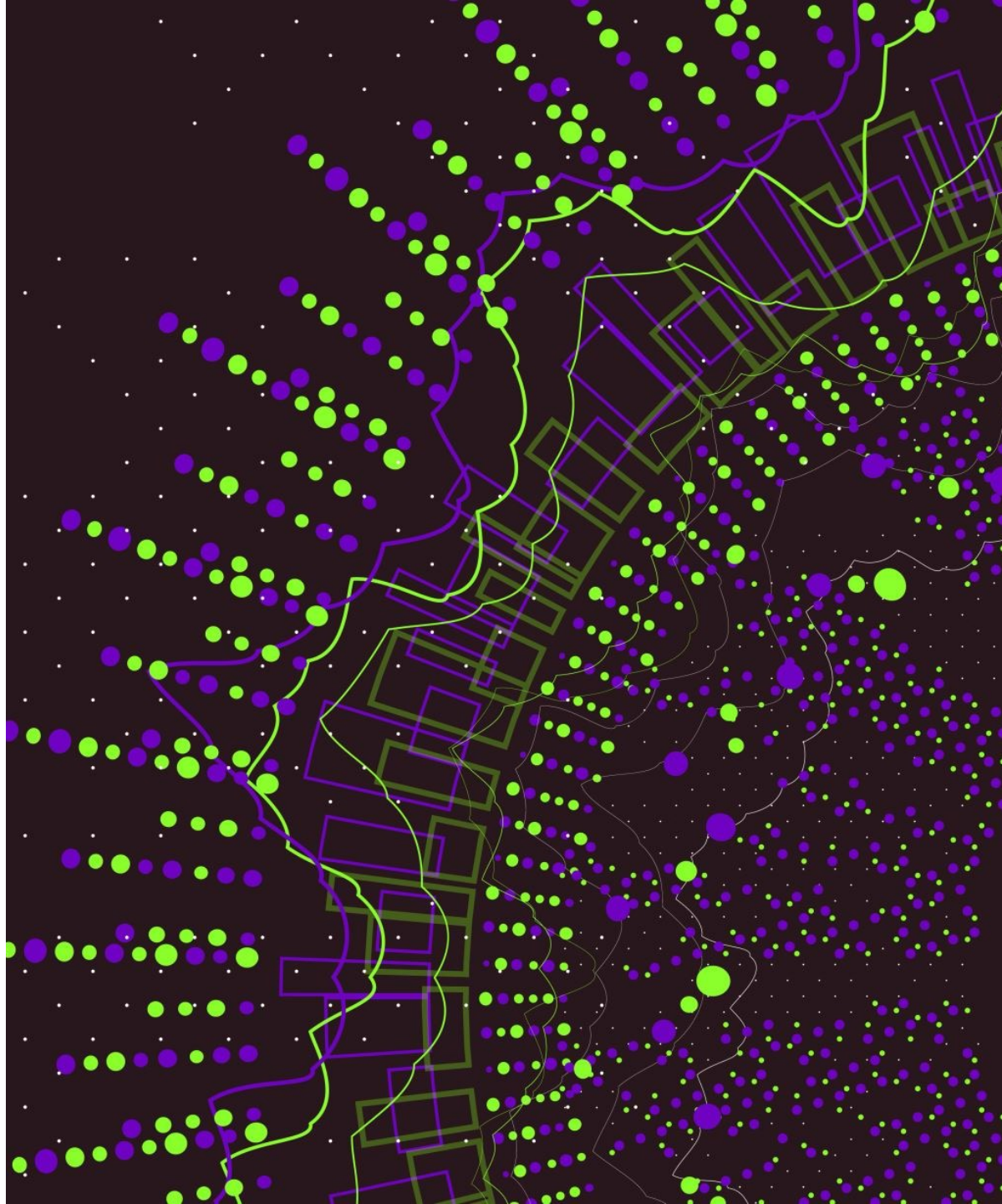
[1]Department of Physics and Astornomy, UNIBO

[2]Department of Agricultural and Food Sciences, UNIBO

[3]Department of Physics and Chemistry, UNIPA

[4]Department of Biomedicine, Neurosciences and Advanced Diagnostics, UNIPA

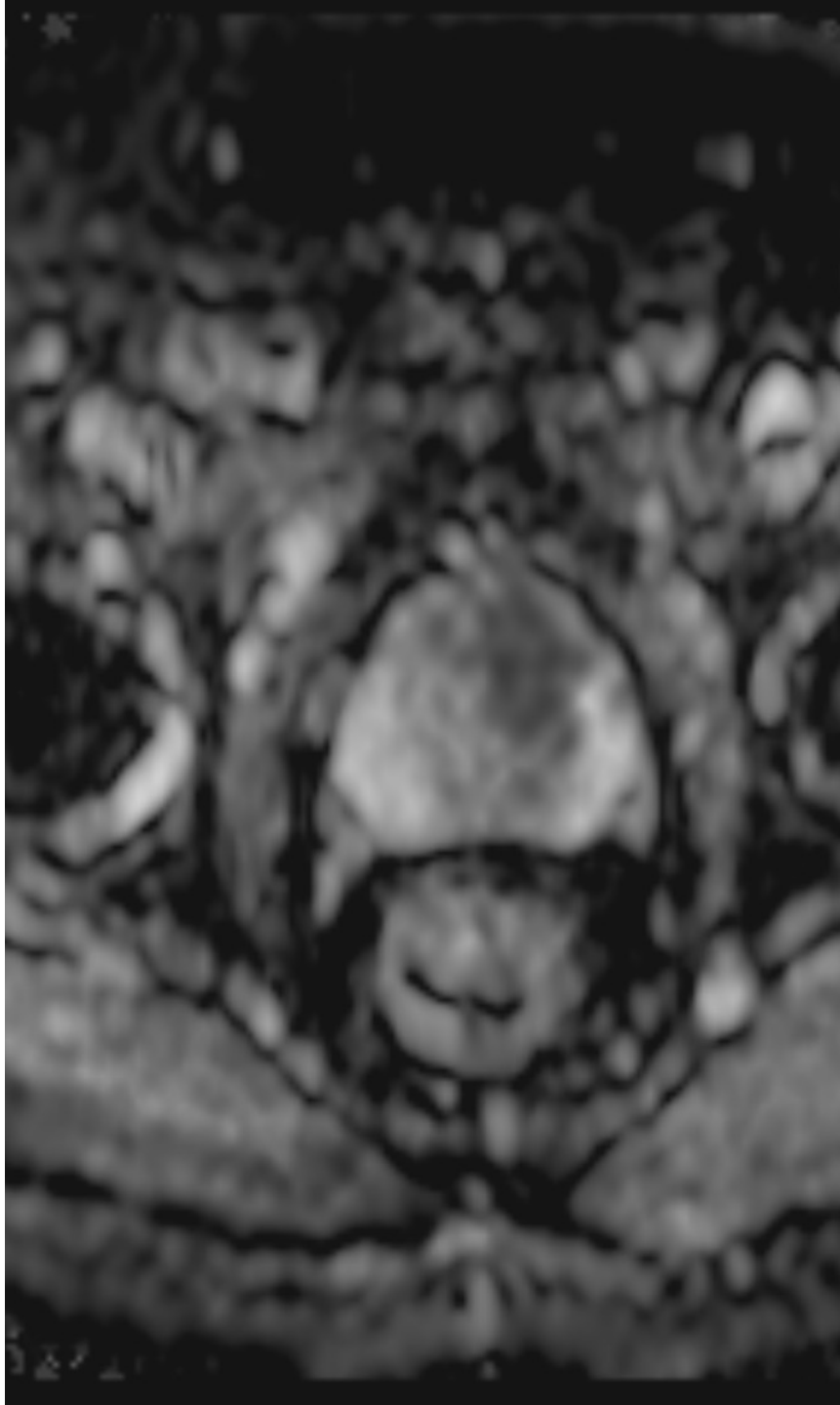
Next-AIM Kickoff Meeting, February 2022



ORIGINAL DATASET & CHALLENGES

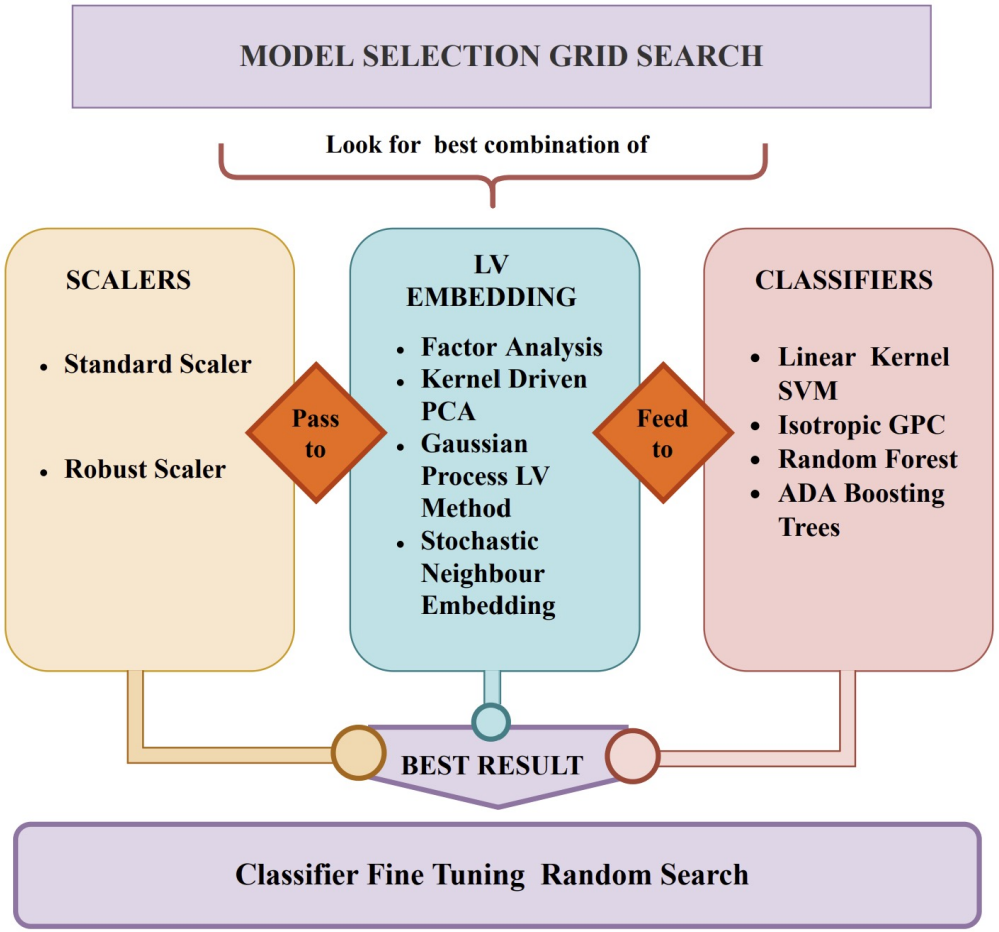
- 92 patients, from which 3 classes of samples are extracted using T2 weighted images
- 46 control tissues, 26 lesion tissues, 20 tumoral tissues with lesion
- **Objective:** evaluate the predictive power of texture analysis features, extracted with MazDa (or similar software) from ROIs segmented by experts

- **Challenges:** Not so many images, unbalanced classes, high number of possibly redundant features (circa 300)



BLIND MODEL SELECTION- LEARNING THROUGH SELF-OPTIMIZATION

Or what if we know nothing about the features?



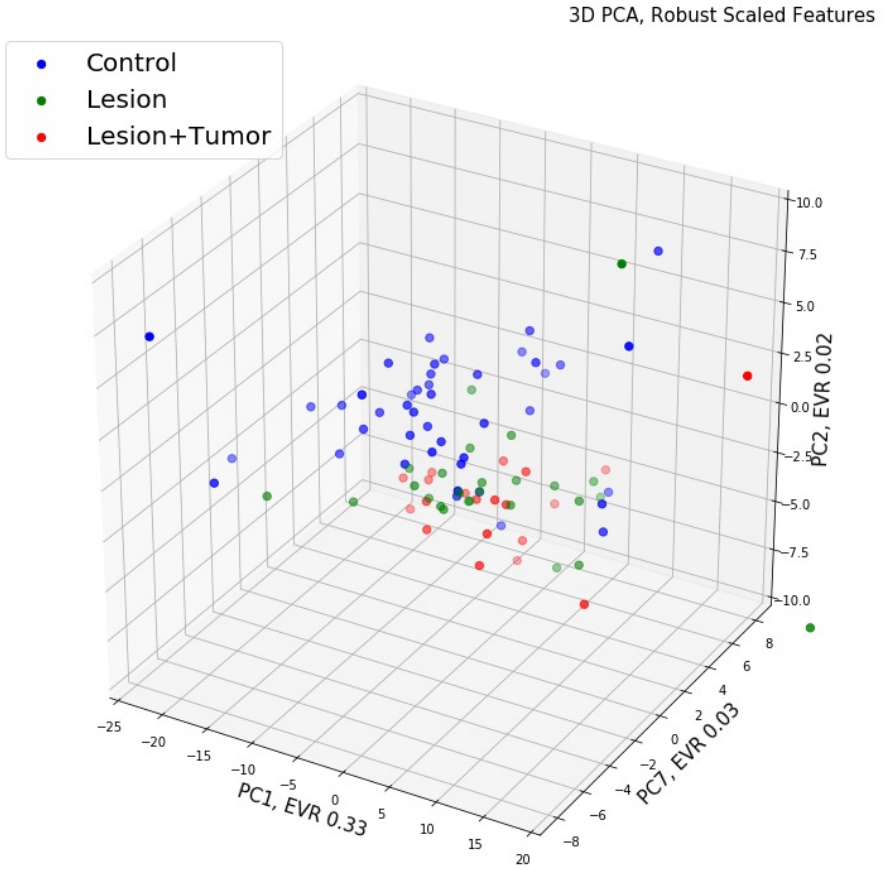
BLIND MODEL SELECTION- RESULTS

Best model:

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

Crossvalidated Performance of the 3 classes problem: average accuracy ~ 0.68

Best Self-Learned 3D Embedding of the Dataset and Loadings Analysis



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

Pros: Create and optimize models straight from features extracted with high-level software (MazDa). From technicians to classification in one click.

Cons: Struggles to discriminate between Lesion/Lesion + Tumors

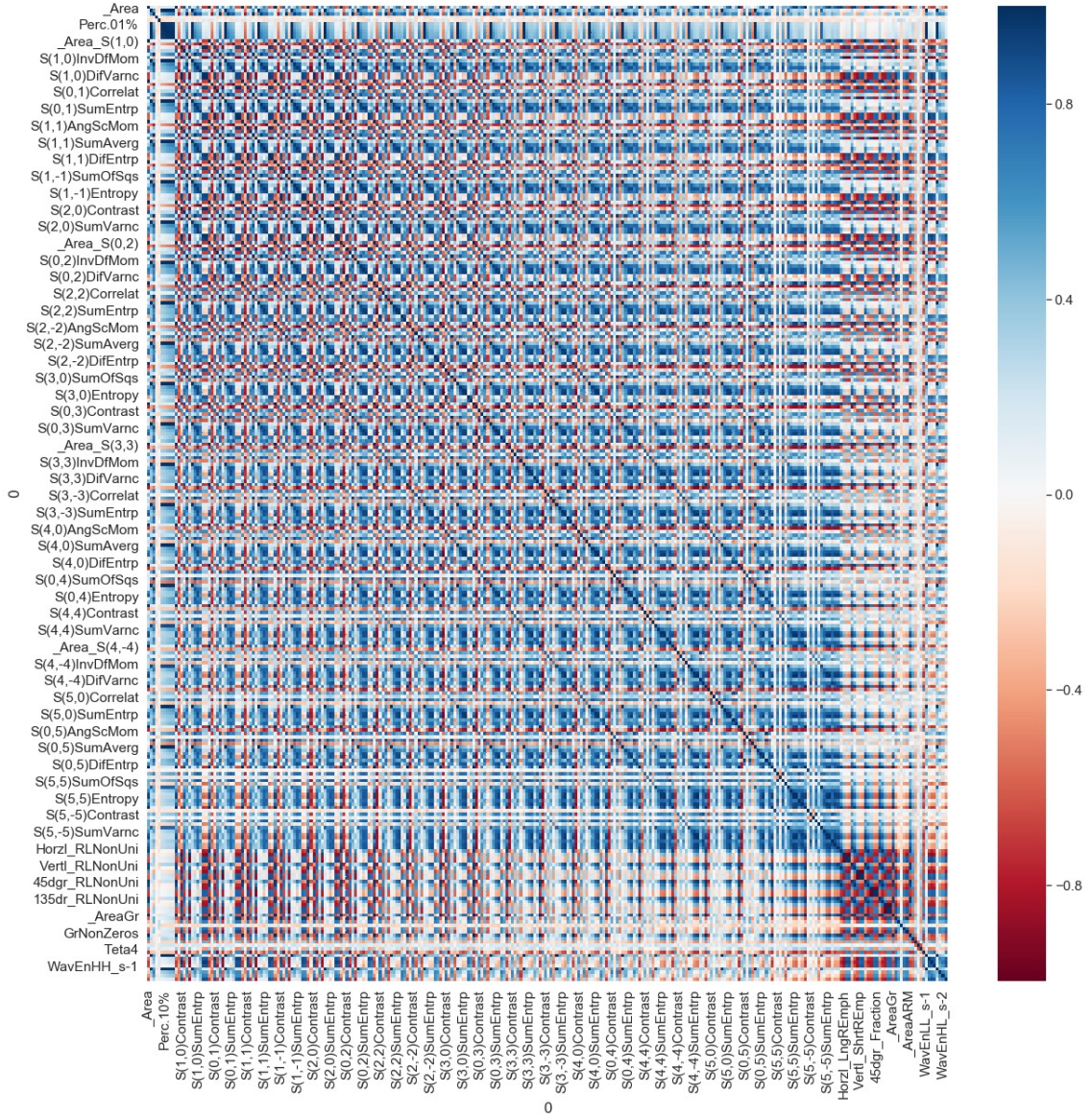
	Control	Lesion	Lesion+Tumor
Control	38	5	3
Lesion	9	5	12
Lesion+Tumor	4	4	12

ADDING KNOWLEDGE ABOUT FEATURES

Features derived from ROIs are:

- Histogram-based,
- Gradient-based,
- Run length matrix-based,
- Co-occurence matrix-based,
- Auto regressive model-based,
- Wavelet parameter-based

Collinearity is to be expected: some kind of penalized model may help



ADDING KNOWLEDGE ABOUT FEATURES

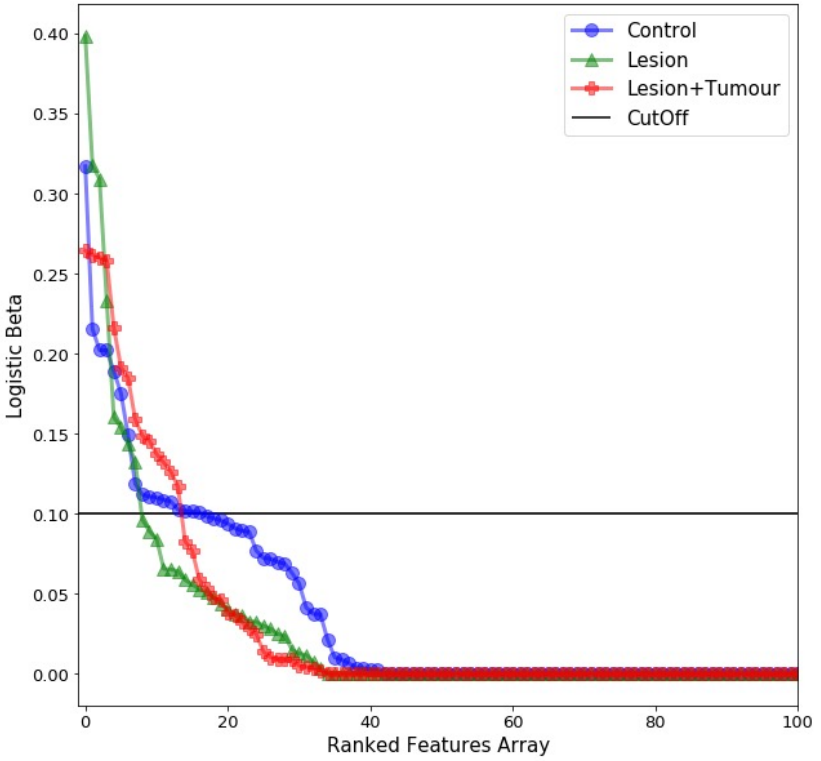
Best Solution: LR with EN mixed penalty

- 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, Teta4, 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



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

Cross Validated predictive performance: ~ 0.72 average accuracy!

NEXT AIM: WHAT'S NEXT?

- Dataset expansion
- New Segmentation
- Data Augmentation
- New low-level feature extraction (PyRadiomics)
- ADC Weighted Image Integration
- Pirads 4 & 5 model Fine Tuning (optimize binary class problem to enhance «lesion gray zone» classification)