ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

DIPARTIMENTO DI FISICA E ASTRONOMIA - DIFA

AIM Live Meeting

Classification of Prostate Tumours from NMR Images Texture Analysis: a Machine Learning Approach

Carlo Mengucci[1,2] Maurizio Marrale[3] Daniel Remondini[1]

[1] *Department of Physics and Astronomy, UNIBO*

[2] *Department of Agricultural and Food Sciences, UNIBO*

[3] *Department of Physics and Chemistry, UNIPA*

February 3, 2020

Dataset

A new DataFrame with 92 patients, divided into 3 classes, was extracted from the original T2 data:

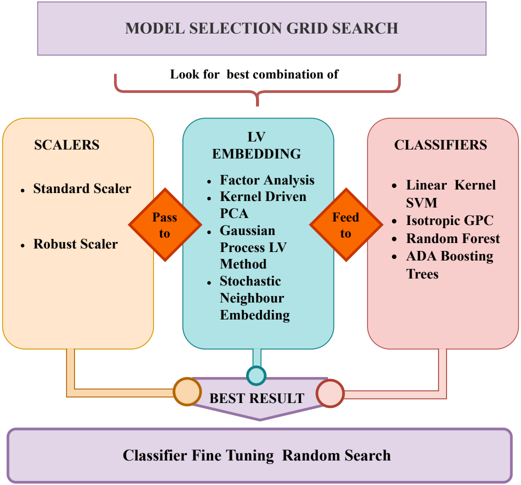
- **Control Tissues**: 46
- **Lesion Tissues**: 26
- **Lesion with Tumor Tissues**: 20

Objective: evaluate predictive power of the 290 features resulting from the intersection of features amongst patients. All features have been extracted from ROI in images by the **MazDa** software.

Challenges: Not so many images, classes have diferent size in population, many features.

Explorative Framework

What if we know nothing about the features?



Explorative Framework

Results

- **Scaler**: Robust Scaler
- **LV Embredding**: 15 Components Linear Kernel PCA
- **Classifier**: Random Forest
- **Crossvalidated Performance**: Average accuracy ~ 0.68

Single Run LOO Confusion Matrix:

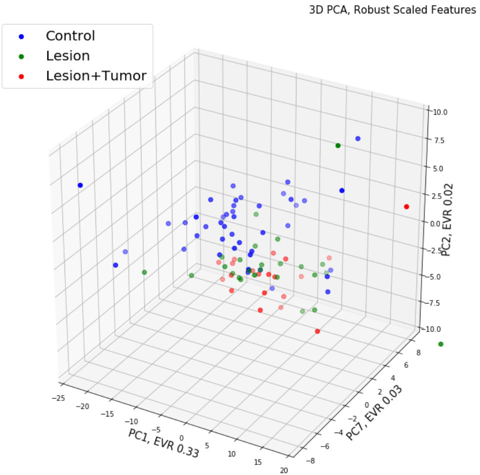
	Control	Lesion	Lesion+Tumor
Control	38	5	3
Lesion	9	5	12
Lesion+Tumor	4	4	12

Struggles to discriminate between Lesion/Lesion+Tumour

Explorative Framework

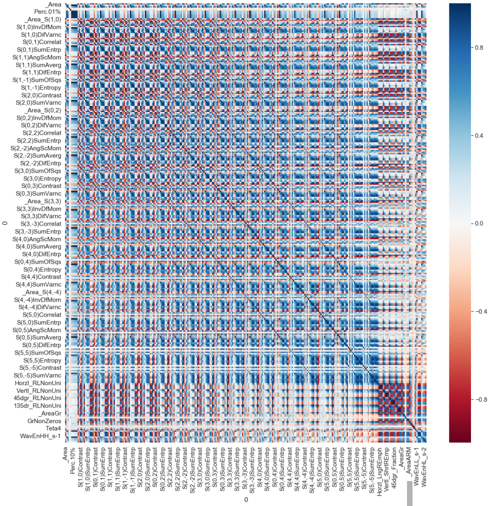
Best 3d Projection and Feature Selection with Loadings Analysis

- **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



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 expected
- Some kind of penalty model may help



Adding Knowledge About Features

Best Solution: LR with EN mixed penalty

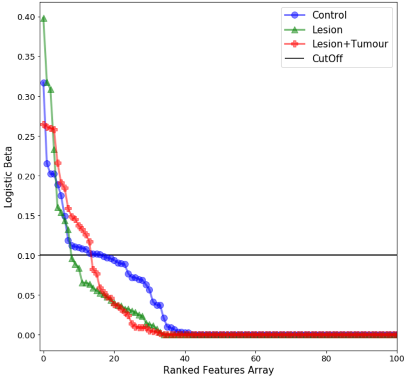
- **L1-L2 mixed Elastic Net penalization gives best balance between induced sparsity (feature selection) and retained information for predictive performances**
- **Best model for prediction and most representative features for each class are simulaneously found**
- **Uninformative features are pruned**
- **Cross Validated predictive performance:** ~ 0.72 **average accuracy!**

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

Adding Knowledge About Features

LR Results

- **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



Conclusions and Future Developments

- **Considering the difficulty of the task, this framework seems promising**
- **More images will obviously improve performance stability, especially in the Lesion gray zone**
- **A deeper knowledge of how features are extracted with MazDa will help to find a physical meaning to feature selection**
- **A more efficient random search using classifiers parameters a priori distributions is on the way**

Thanks to the research group