
UNIVERSITAT POLITÈCNICA DE CATALUNYA
BARCELONATECH
Departament de Teoria del Senyal
i Comunicacions

barcelonaβeta
Brain Research Center

# Prevention of Alzheimer's disease: a contribution from MRI and Machine Learning

Miguel Artigues Canaves

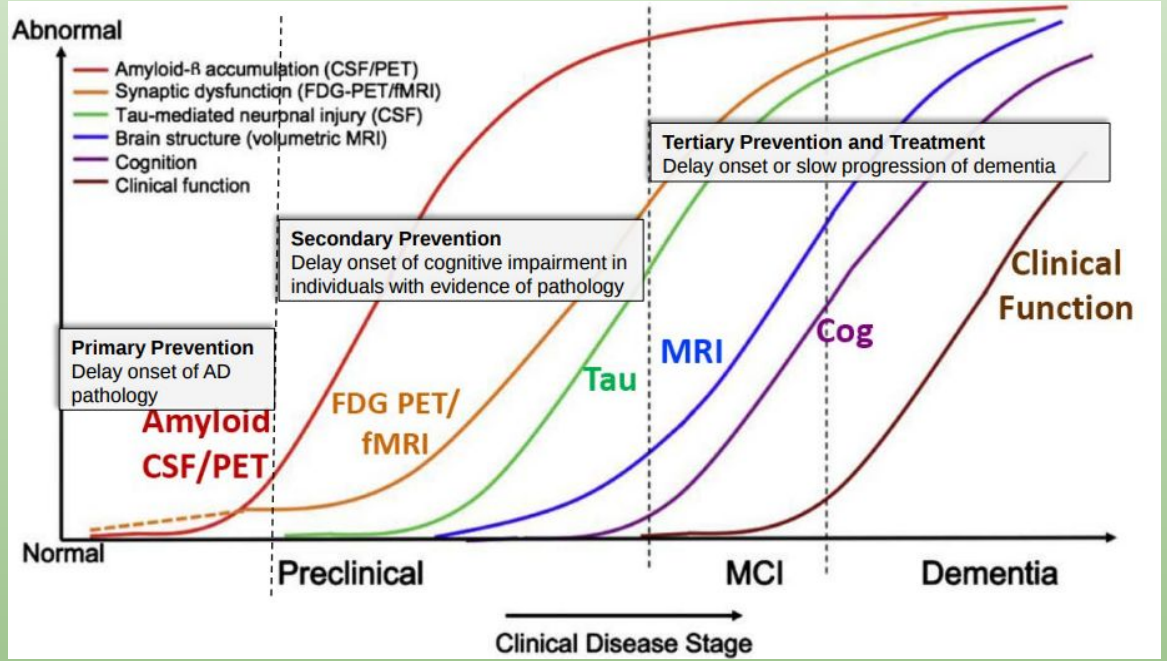
Degree thesis defense
Barcelona, 25th May 2018

# Contents

- Introduction
- Data
- Algorithm design
- Results and conclusions

# Alzheimer’s disease (AD)

- Neurodegenerative chronic and currently irreversible disorder
- One of the most common causes of dementia in elderly people
- Growing trend
- Brain alterations during the early stages of the disease



# Current situation

- Preclinical AD characterized by:
  - Cognition within normal ranges and abnormal amyloid biomarkers as measured in cerebrospinal fluid (CSF)
  - Positron emission tomography (PET)

- MRI established as one of the preferred diagnostic imaging tools
  - Cheaper
  - Minimally invasive

# Scope of the project

- Analyse T1-MRI volumes:
  - At the voxel level
  - Novelty:  longitudinal pipeline

- Objectives
  - Show how MRIs can be used for subject classification using machine learning tools
  - Identify patterns of volumetric changes specific to preclinical AD

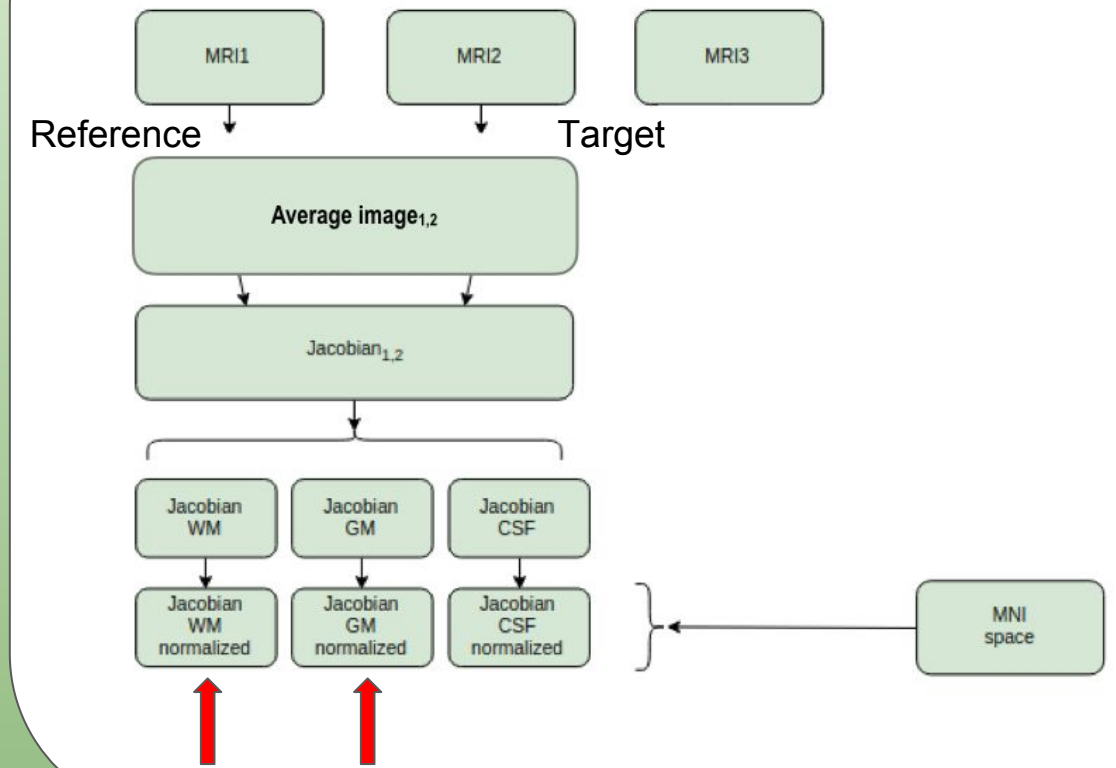
# Data

- Extracted by BBRC from a subset of the Alzheimer's Disease Neuroimaging Initiative (ADNI)
- Volunteers with cognition impairment excluded
- Atributtes:
  - Amyloid beta (**A$\beta$**)
  - Clinical diagnosis

**Table 1**
**Distribution of the number of MRI-T1 image adquisitions per subject**

| Amount of images acquired | # number of subjects |
|---|---|
| 1 | 481 |
| 2 | 391 |
| 3 | 81 |
| 4 | 39 |
| 5 | 23 |
| 6 | 7 |
| Total | 1025 subjects |

# Pre-processing



# Jacobian labeling

Threshold Amyloid beta biomarker (**Aβ**) = 192 pg/mL

NC

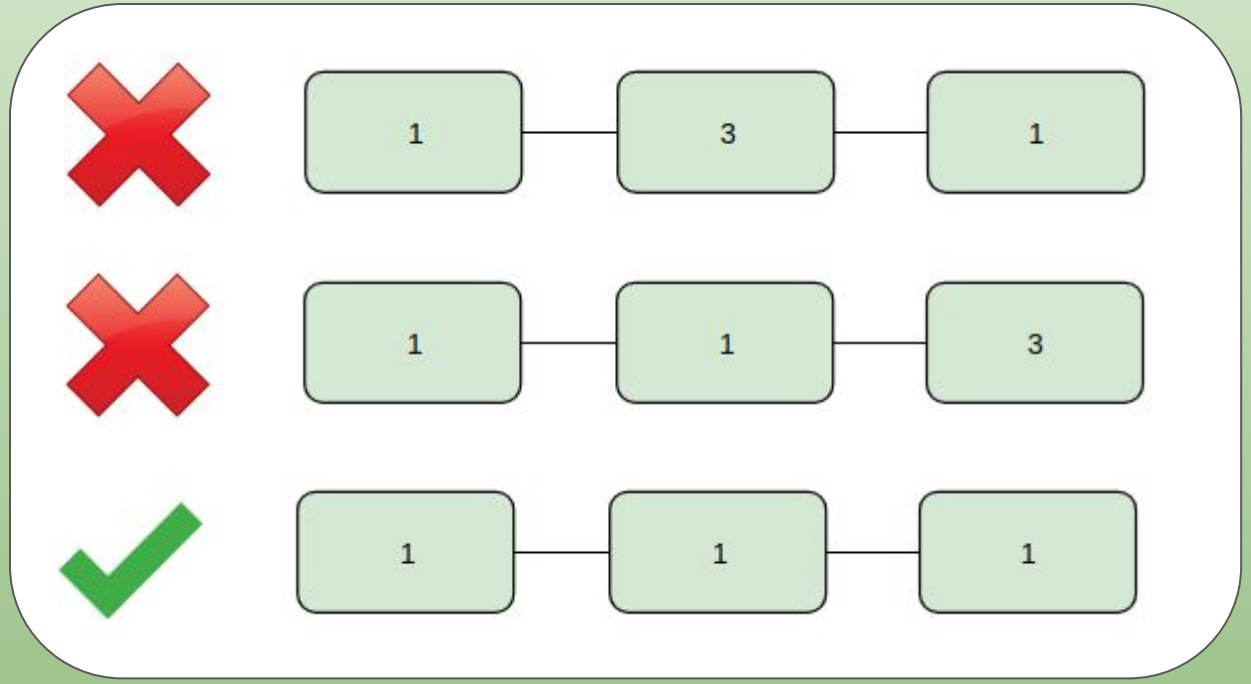
PC

**Table 2**

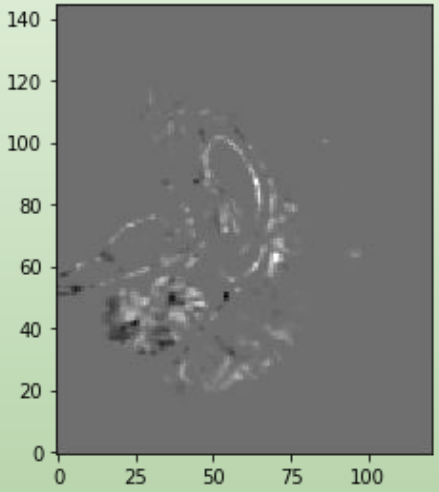
Classification rules for Jacobians based on their reference and target events

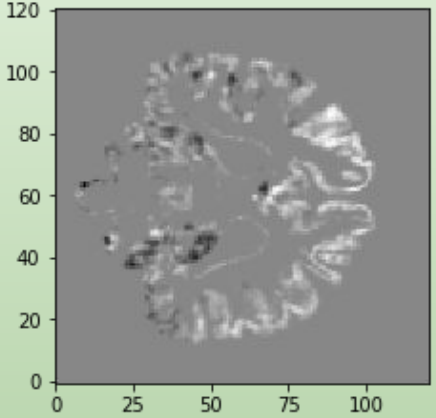
| class | Classification rules for jacobians based on their reference and target events | clinical diagnosis |
|---|---|---|
| 1 | **Aβ-/Aβ-** *(Aβneg)* | without symptoms |
| 2 | **Aβ-/Aβ+** (Aβconv) | without symptoms |
| 3 | **Aβ+/Aβ+ without symptoms***(Aβpos)* | without symptoms |
| 4 | **Aβ+**(NC)**/Aβ+**(MCI,AD) | symptoms |
| 5 | **Aβ+/Aβ+ with symptoms** *(Aβpossymp)* | symptoms |

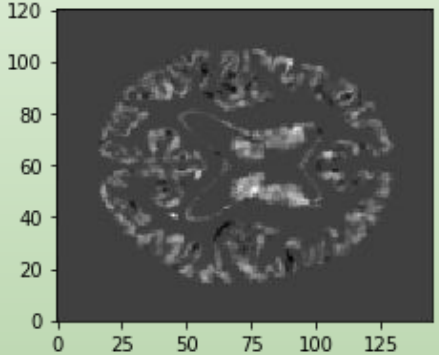
# Subject labeling



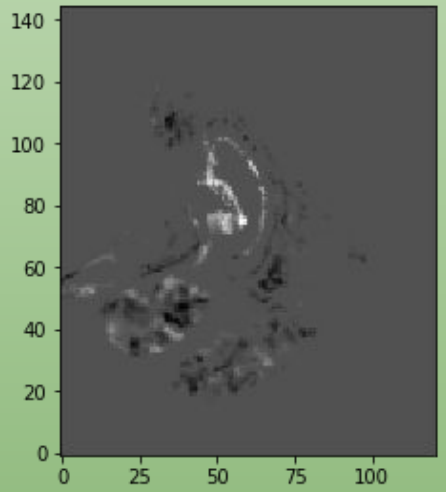
- Normal control

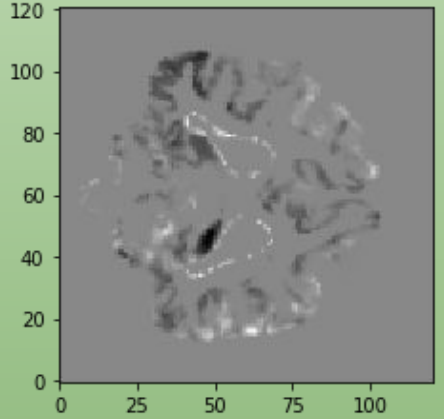


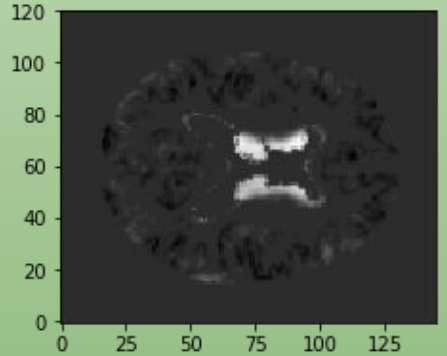




- Preclinical subject



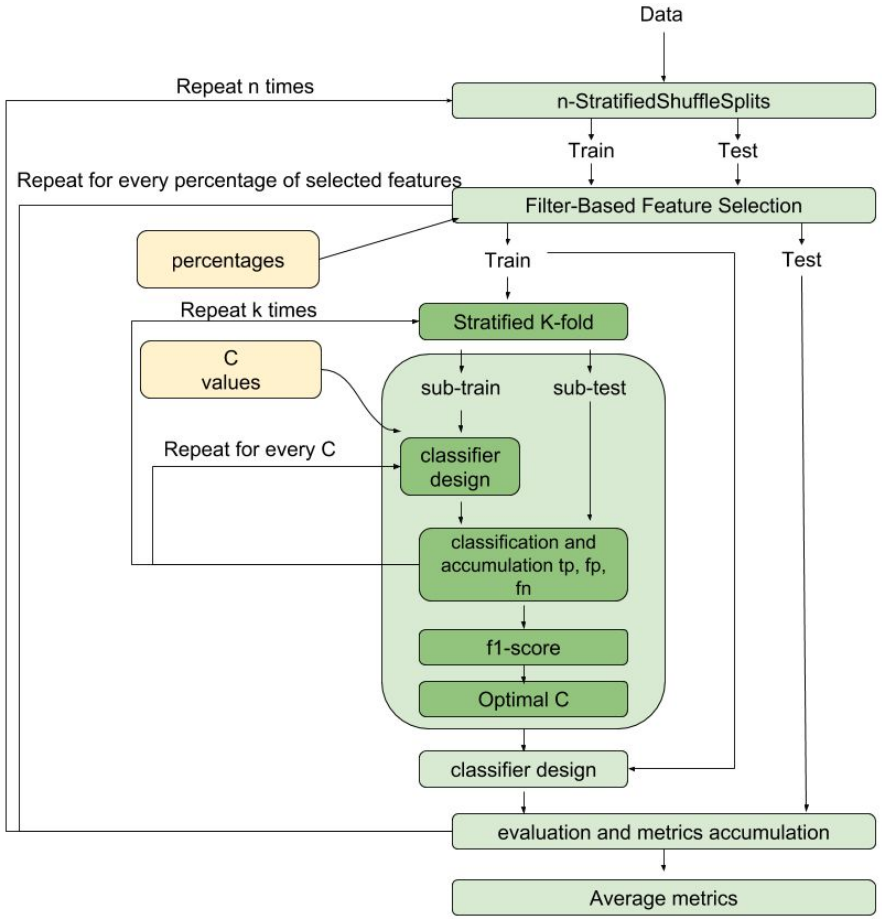



# Algorithm design

- 500000 features
- 288 samples
- 40 % controls
- 60 % preclinical

- 1000 splits
- “C” value optimization
- Logistic regression classifier

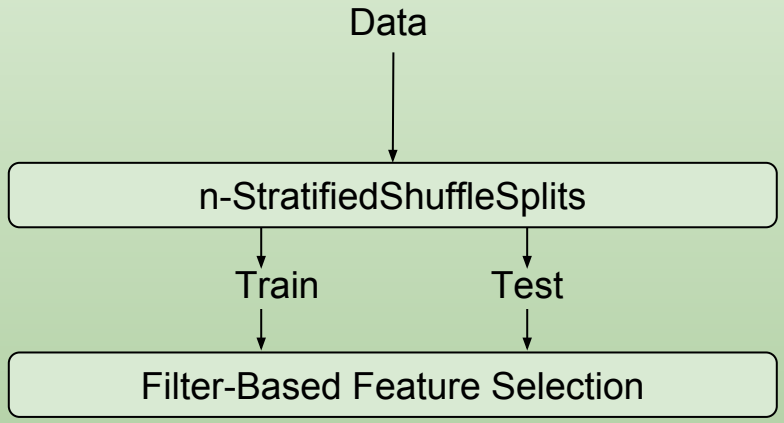


# Algorithm design

- 1000 stratified shuffle splits
- 2 feature selection strategies
  - Analysis of variance (ANOVA) F-test scores

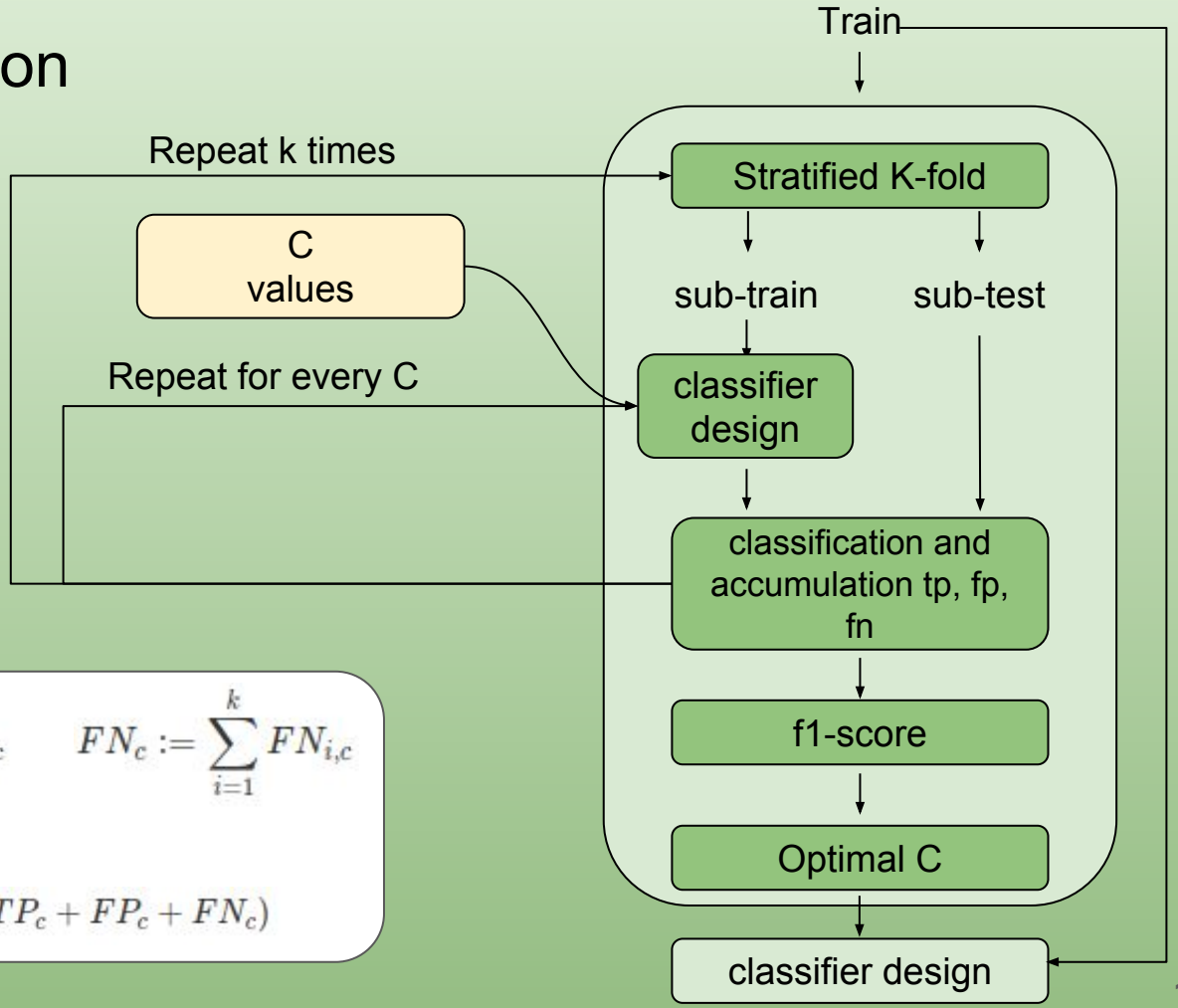
$$F := \frac{between - group \quad variability}{within - group \quad variability}$$

  - Logistic regression classifier weights



# Nested cross-validation

- Stratified 3-fold nested cross-validation
- 20 equispaced “C” values in range [10e-3, 10e3]
- Accumulative F1-score

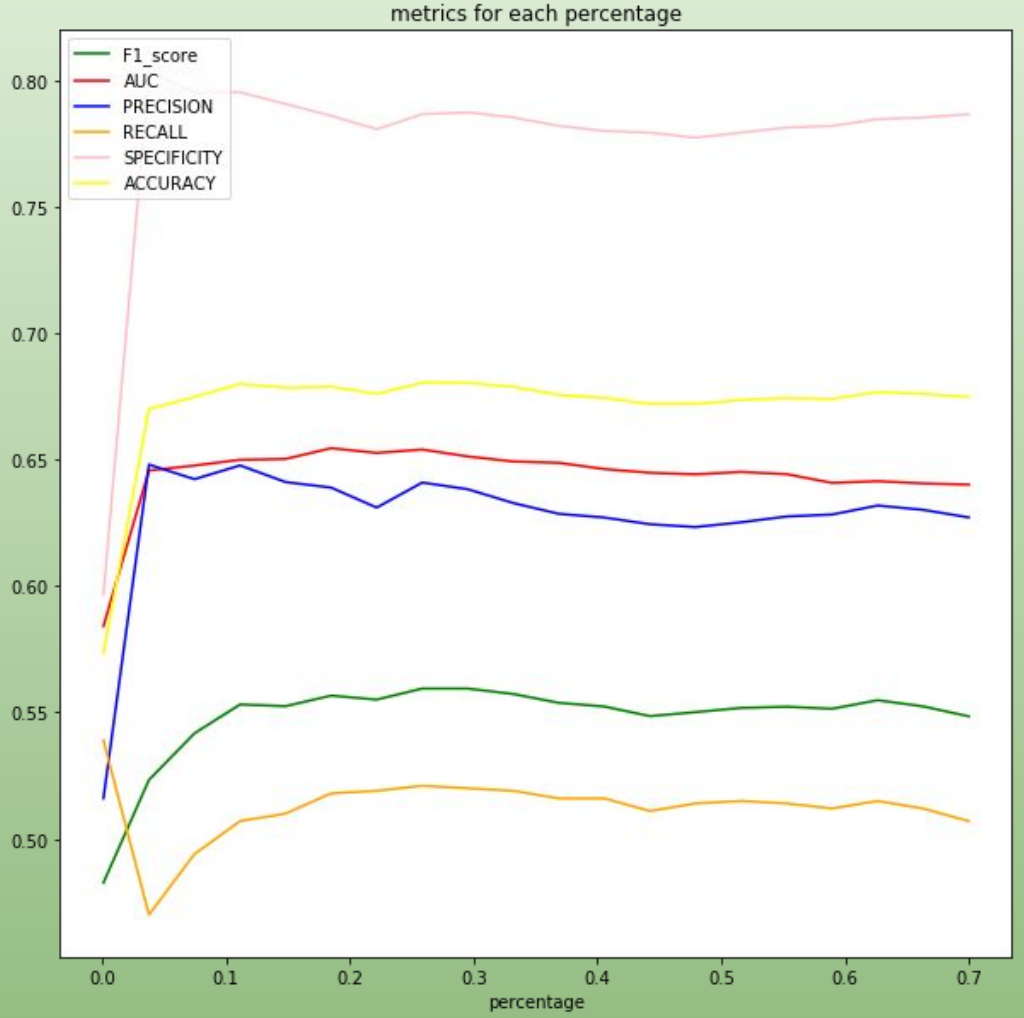



$$TP_c := \sum_{i=1}^{k} TP_{i,c} \qquad FP_c := \sum_{i=1}^{k} FP_{i,c} \qquad FN_c := \sum_{i=1}^{k} FN_{i,c}$$

$$F - measure_c := (2 \times TP_c)/(2 \times TP_c + FP_c + FN_c)$$

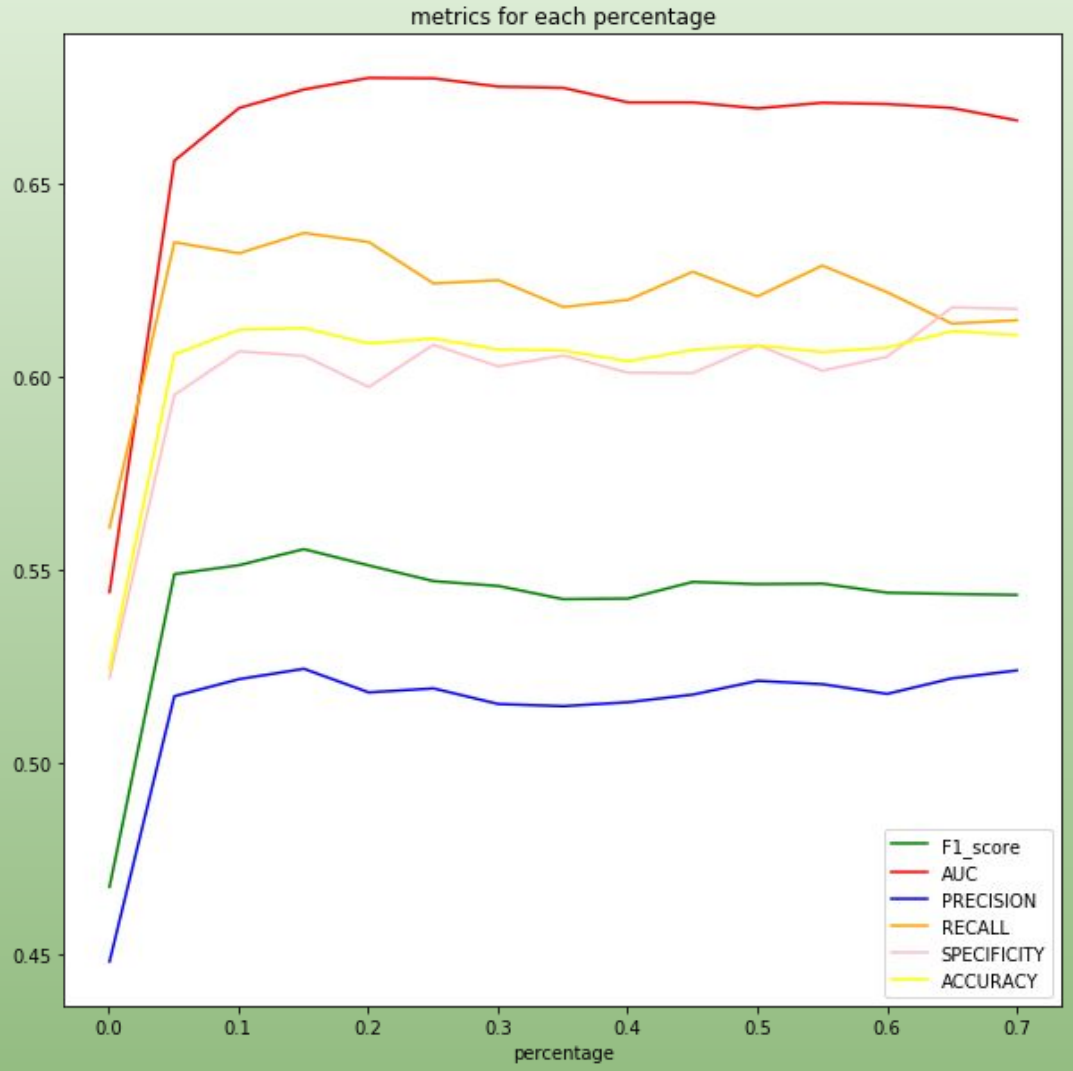
# Results

- Split per subject
- F-test based feature selection



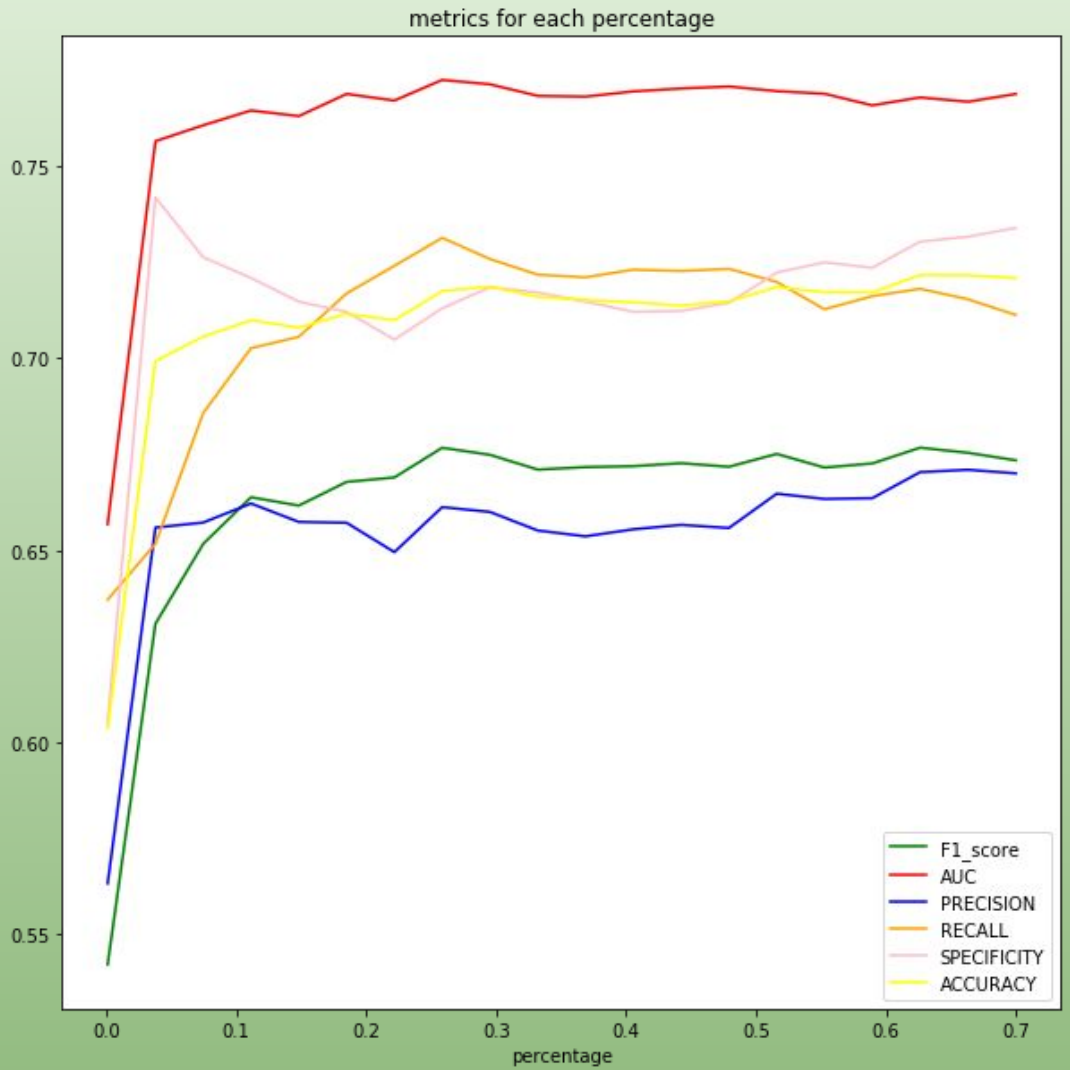
# Results

- Split per Jacobian
- F-test based feature selection



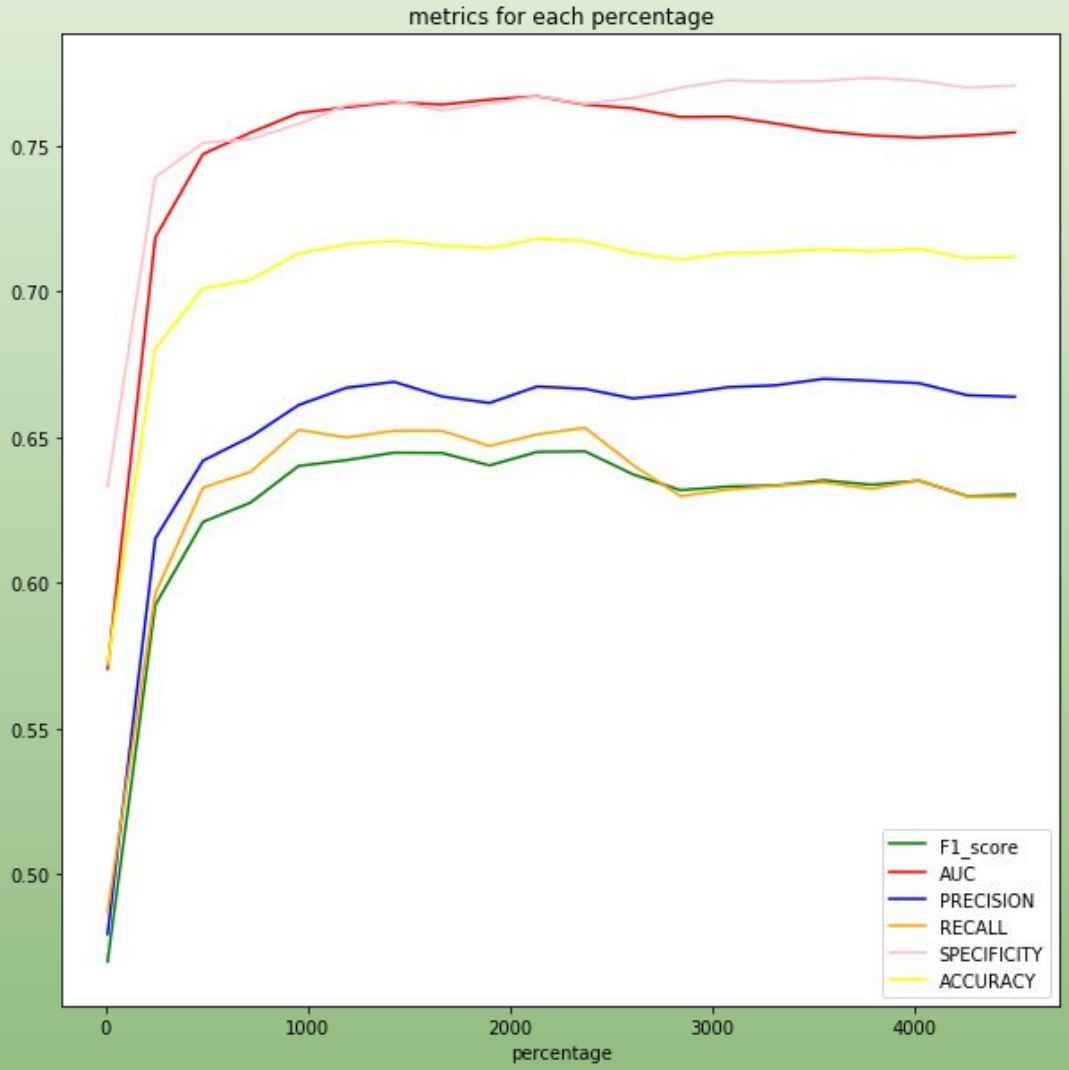
# Results

- Split per Jacobian
- F-test based feature selection
- Testing only with Jacobians with dt > 1.15 years

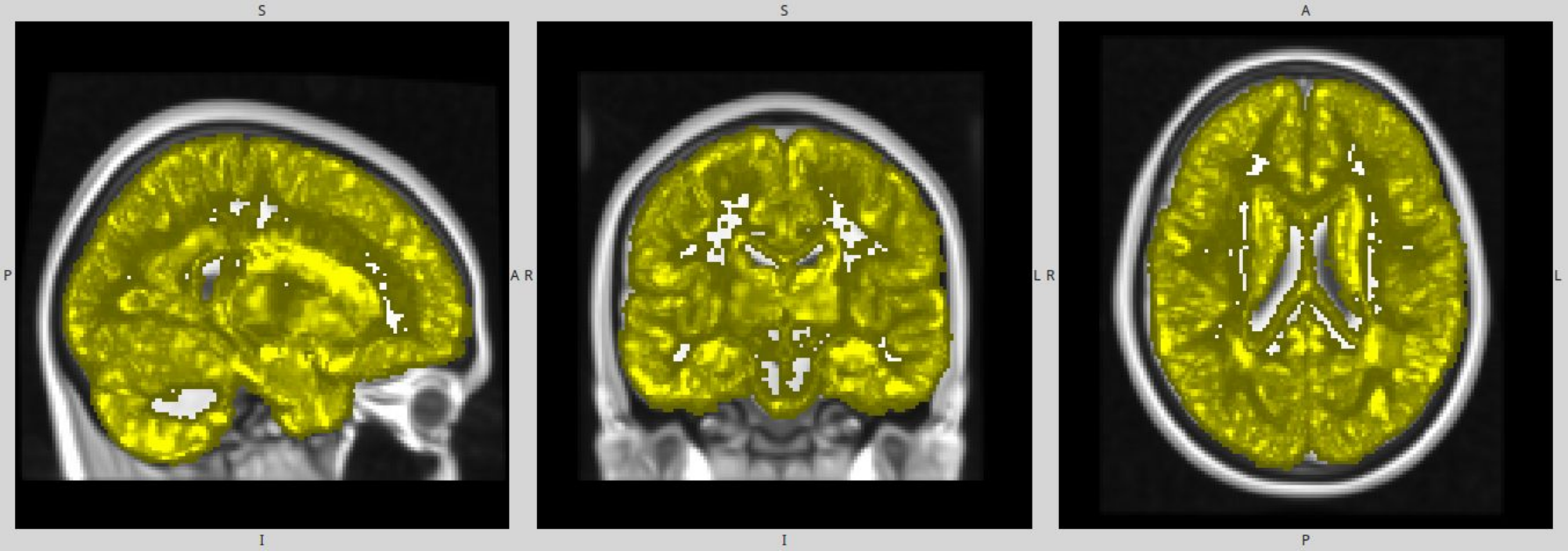


# Results

- Split per Jacobian
- LR weights based feature selection
- Testing only with Jacobians with dt > 1.15 years

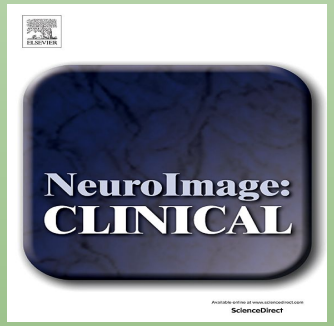


# Results: tridimensional map generated using LR weights accumulation



# Conclusions and future work

- Promissing classification performance for NC versus PC classes
- Detection of relevant brain regions
- Small dataset → Uncertainty of performance on new cohort
- Test algorithm using Jacobians from class 2 (A$\beta$conv)


- Abstract accepted and study exposed at AAIC
- Paper to be summited at Neuroimage Clinical

# F-test

$$F := \frac{between-group \quad variability}{within-group \quad variability}$$

$$between-group \quad variability := \sum_{i=1}^{K} n_i \left(\overline{Y_i} - \overline{Y}\right)^2 / (K-1)$$

$$within-group \quad variability := \sum_{i=1}^{K} \sum_{j=1}^{n_i} \left(Y_{ij} - \overline{Y_i}\right)^2 / (N-K)$$

Where:

- $n_i$ : number of samples from variable y that belong to class i
- N : Y size
- K : number of classes
- $\overline{y}$ : mean estimated value of y
- $y_i$ : mean estimated value of elements from y that belong to class i

# Classifier: logistic regression

$$\hat{y} := g(\underline{w}^T \underline{x} + w_0) \quad where \quad g(z) := \frac{1}{1 - e^{-z}}$$

$$J(\underline{w}) := \min_{w, w_0} ||\underline{w}||_1 + C \sum_{i=1}^{n} log(exp(-y_i(\underline{X_i}^T \underline{w} + w_0)) + 1)$$

$$w_i := w_i - \alpha \frac{\partial}{\partial w_i} J(\underline{w})$$