Measuring normality and pathology in medical imaging.

From the physical signals to diagnosis, fighting the sources of noise and variability.

A bit of philosophy…

What is Physics?

A bit of philosophy...

According to dictionaries...

The science of matter and energy and of interactions between the two, grouped in traditional fields such as *acoustics, optics, mechanics, thermodynamics,* and *electromagnetism*, as well as in modern extensions including *atomic and nuclear physics, cryogenics, solidstatephysics, particle physics, and plasma physics*.

Almost all aforementioned fields find applications in medical field.... But there is more...

A bit of philosophy…

One of the fundamental concepts in Physics is that of **measure**.

By measure, physicists generally mean the quantitative assessment of an observable but the concept entails more complex logical instruments such as theoretical models, statistics, metrics, signal and noises.

A bit of philosophy…

Physics

- Observations
- Direct / indirect
- Derived from previous experiments / better estimates of current theories

- Theory
- One or more models, depend on free parameters
- Few parameters = happy physicist
- Assumption: Complex phenomena can be described by relatively simple models

Medicine

- Observations
- Direct: Clinical practice

- Theory
- No comprehensive models
- Highly complex system
- Subsystem interactions and history not negligible

A bit of philosophy…

Physics

- Experiment
- Designed to verify key aspects of theory, prove/disprove models
- Typical paradigm: Out = signal + noise
- Reproducibility is a key factor

- Data analysis
- Designed to extract "signal" from "noise" [filters]
- Experiment characterization [noise]
- Estimate model parameters [from signal]
- Error estimation relatively simple

Medicine

- Experiment
- Clinical trials (in vitro, in vivo, ….)
- Typical paradigm: improvement / no-improvement
- Reproducibility is rarely achieved

- Data analysis
- Designed to extract "improvement probability"
- Strong a-priori assumptions
- What is "noise"?
- Error estimation generally difficult

Medical imaging: signals

Actual signals i.e. Data:

- 2D/3D matrices (X-rays, CT, MRI, PET)
- Array of real values (molecular concentration, …)

pathological 'Signals'

- Define metric
  pathology "signal" in a normalcy "background"
- no theory
  it must be deduced through group comparison
- lots of assumptions
  pathology markers are common to all individuals
  transition from normal to pathological = continuous process

Medical imaging: noises

Acquisition noises [easy peasy (almost)]

- Scanner electronic noise
- Scanner non idealities (e.g. B-field inhomogeneities,…)
- Scan quality issues (resolution, acquisition protocol, …)
- Image artifacts (subject movements during acquisition, e.g. object driven B-field distortion, …)

Physiological noises [difficult]

- Confounding variables (age, sex, education, general anamnesis,…)
- Inter-individual variability can be more significant than normalcy vs. pathology difference
- Cohort representativity

Medical imaging: noises

"Gold standard" noises [very hard]

- Group mixing (clinical assessment is not 100% accurate)
- Group purity (comorbid pathologies)
- General accuracy can change at different ages (e.g. neurodevelopment/neurodegeneration)

Data processing noises [tricky but manageable]

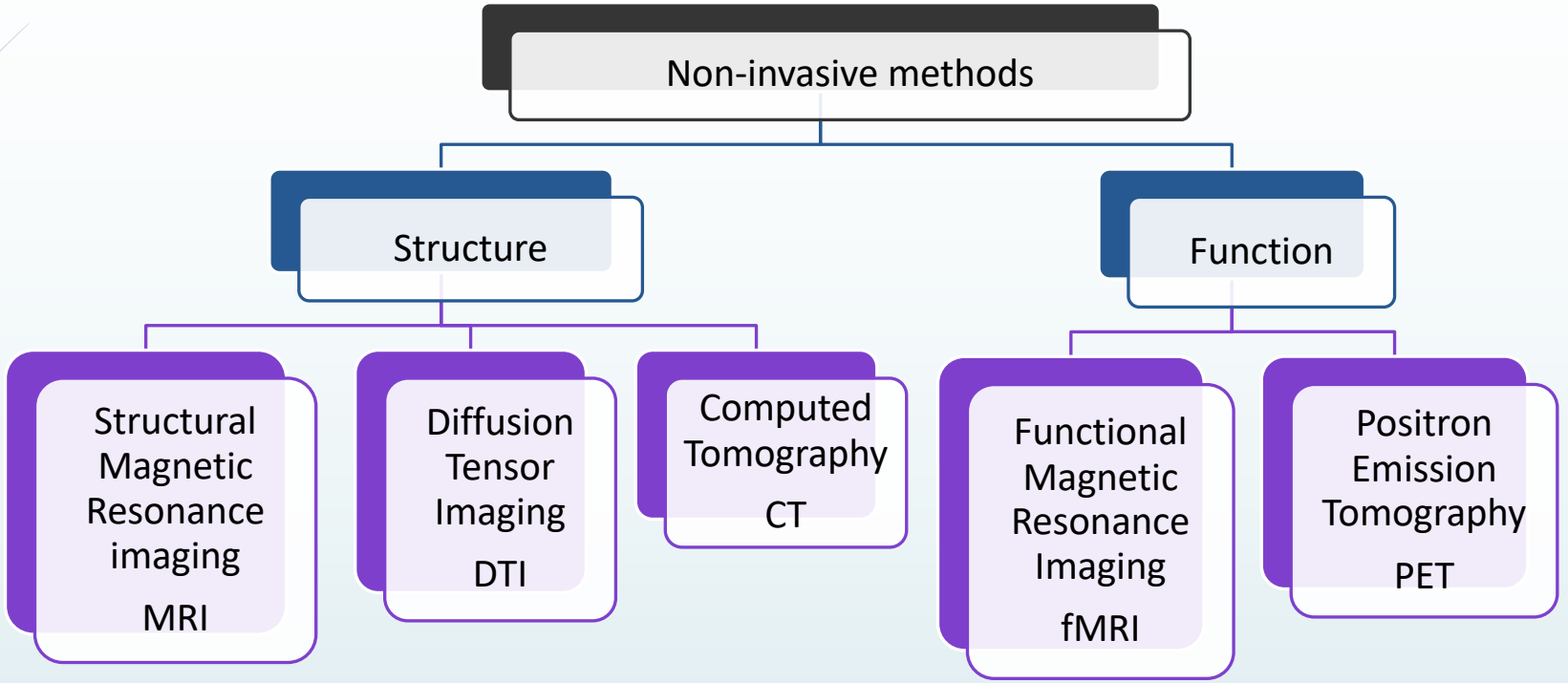
- Signal is *deduced* by comparison among cohorts: method selection is important
- Information degradation due to sub-optimal processing
- Depends on assumptions on "signal"

What do we aim at? Radiomics

Radiomics is a field of medical study that aims to extract large amount of quantitative features from medical images using data-characterisation algorithms.
These features, have the potential to uncover disease characteristics that fail to be appreciated by the naked eye.

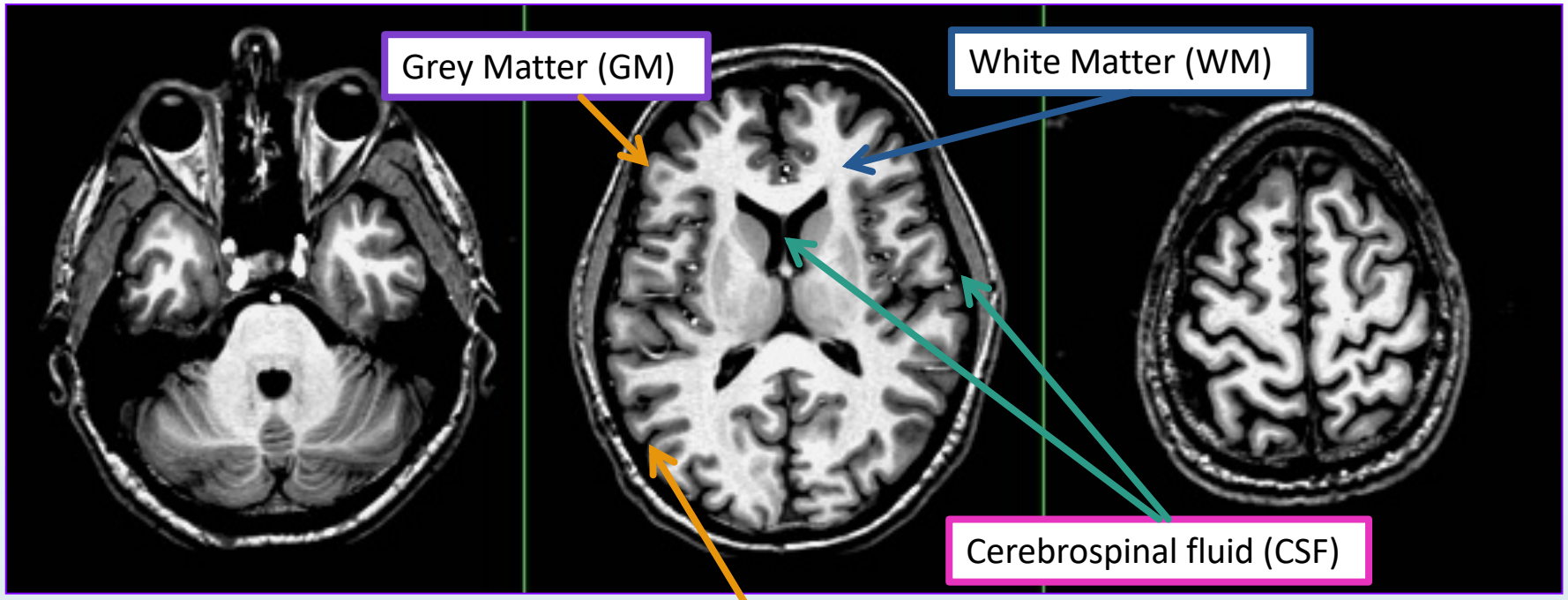
The hypothesis of radiomics is that the distinctive imaging features between disease forms may be useful for predicting prognosis and therapeutic response for various conditions, thus providing valuable information for personalised therapy.

Medical imaging of the brain



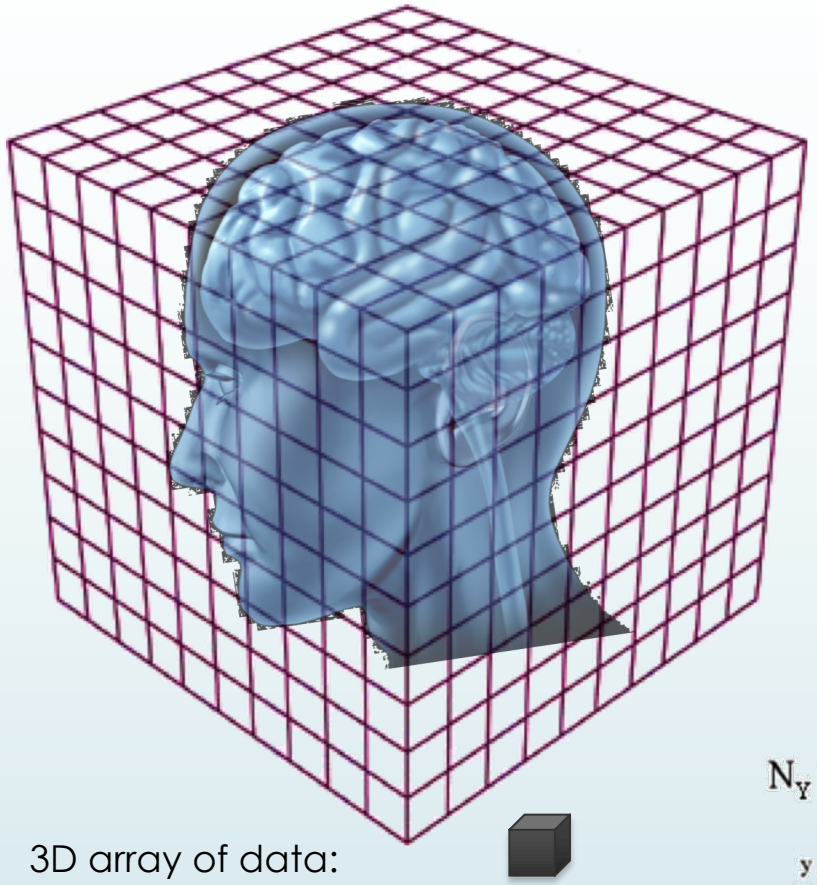
T_1-weighted images

Axial slices of a human head with spatial resolution of about 1 mm^3



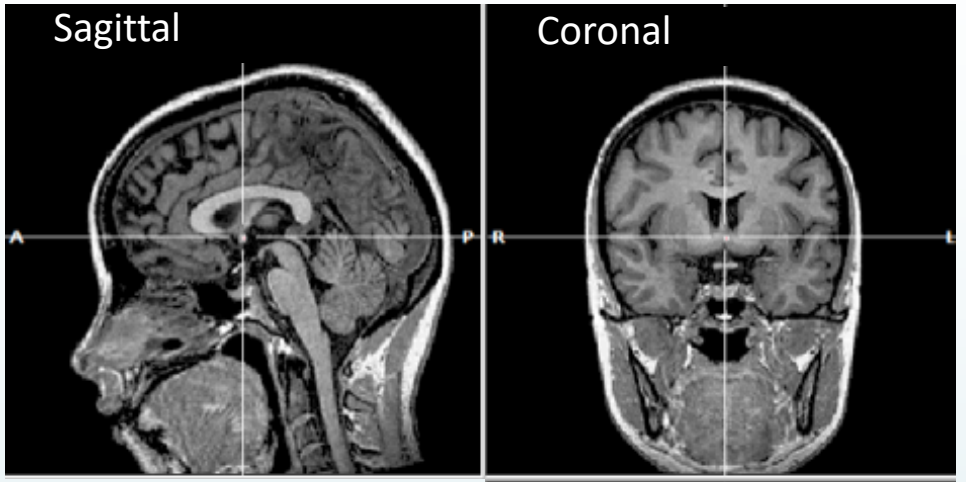
Grey Matter (GM) cortex can be localized and cortical thickness can be evaluated to investigate GM involvement in pathological conditions.

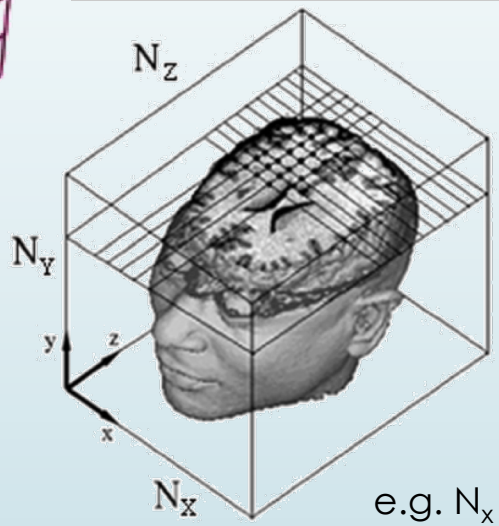
T_1-weighted images



3D array of data:

- The voxel is the "volume element"
- High resolution MRI T_1-w images :
- $v_x \times v_y \times v_z \approx 1 \times 1 \times 1$ mm^3



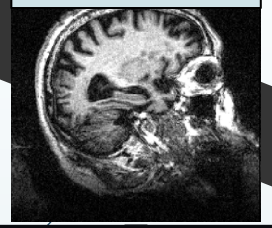


e.g. $N_x \times N_y \times N_z = 256 \times 256 \times 256$

Example: can we distinguish AD subjects from control people by their T_1-weighted images?

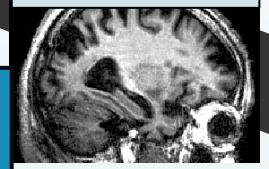
Pre-processing

RAW



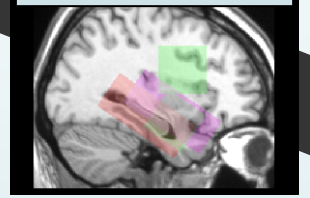
Quality assessment

NORMALIZED

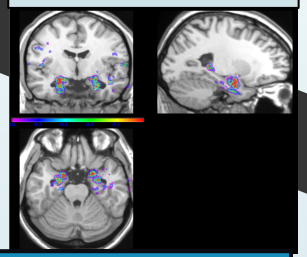


Disease-specific volumes (clinically driven)

AD-regions

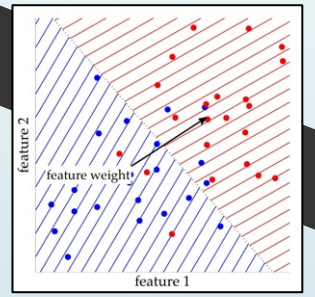


FEATURES

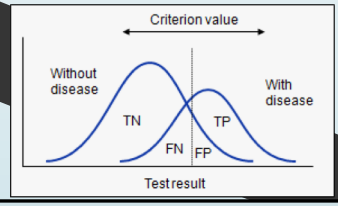


Features extraction
A ML algorithm selects those who most likely discriminate between Controls/AD

Classification
A ML classifier takes the feature subset outputs the distance between the input set and the discriminating hyper surface



A number is assigned to each subject ranging from 1 (normalcy) to -1 (AD)



CLASSIFICATION INDEX

Example: can we distinguish AD subjects from control people by their T_1-weighted images?

- Initial quality check
  - Image artifacts
  - Voxel size and aspect ratio

  Acquisition noises/variability
- Noise removal
  - Steerable pyramid de-noising

  Acquisition noises/variability
- Spatial registration
  - 3-way scalable (7 d.o.f.) + affine (12 d.o.f.)
  - Mutual information and normalized correlation metric

  Physiological noises
- Intensity normalization
  - CSF/GM/WM segmentation
  - VOI-based histogram match

  Acquisition noises/variability

Example: can we distinguish AD subjects from control people by their T_1-weighted images?

- Region (VOI) extraction
  - Template matching, rigid (6 d.o.f.) registration

Physiological noises

- Features computation
  - 4 different neighborhoods
  - Intensity & texture based filtering

- Classification
  - Random Forest (RF) important variable map
  - Support Vector Machine (SVM) classifier

"Gold standard" noises

Example: can we distinguish AD subjects from control people by their T_1-weighted images?

Physiological noises

Acquisition noises/variability

"Gold standard" noises

...But we are introducing data processing noise...
If someone else in the world would process the same data with different methods... Would he find the same results?

Example: can we distinguish AD subjects from control people by their T_1-weighted images?

- Initial quality check
  - Image artifacts
  - Voxel size and aspect ratio

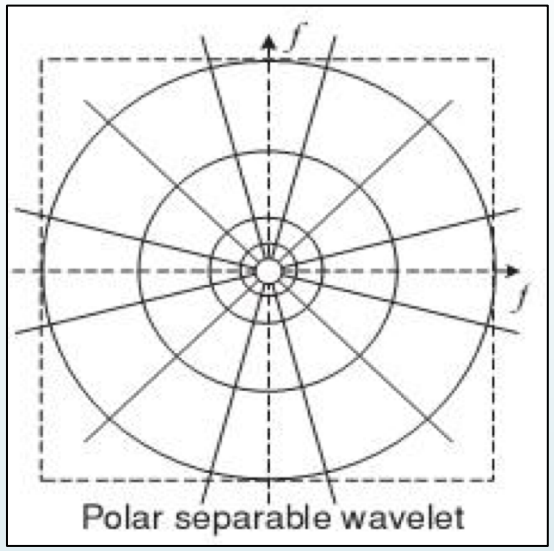
  Acquisition noises/variability
- Noise removal
  - Steerable pyramid de-noising

  Acquisition noises/variability
- Spatial registration
  - 3-way scalable (7 d.o.f.) + affine (12 d.o.f.)
  - Mutual information and normalized correlation metric

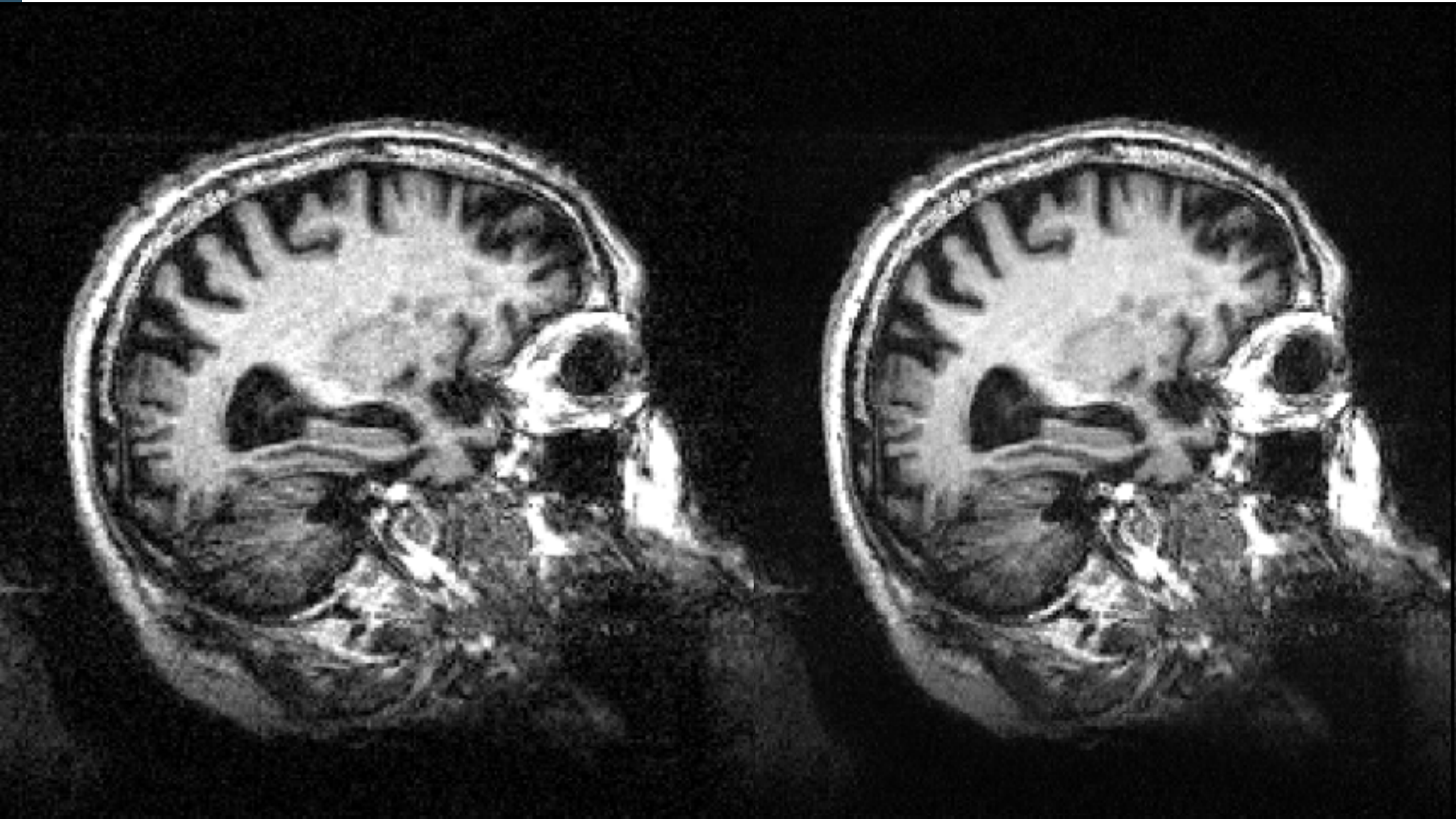
  Physiological noises
- Intensity normalization
  - CSF/GM/WM segmentation
  - VOI-based histogram match

  Acquisition noises/variability

Example: can we distinguish AD subjects from control people by their T_1-weighted images? **Denoising**



- The steerable pyramid filter performs a polar-separable decomposition in the frequency domain, thus allowing independent representation of scale and orientation



Example: can we distinguish AD subjects from control people by their T_1-weighted images?

- Initial quality check
  - Image artifacts
  - Voxel size and aspect ratio

Acquisition noises/variability

- Noise removal
  - Steerable pyramid de-noising

Acquisition noises/variability

- Spatial registration
  - 3-way scalable (7 d.o.f.) + affine (12 d.o.f.)
  - Mutual information and normalized correlation metric

Physiological noises

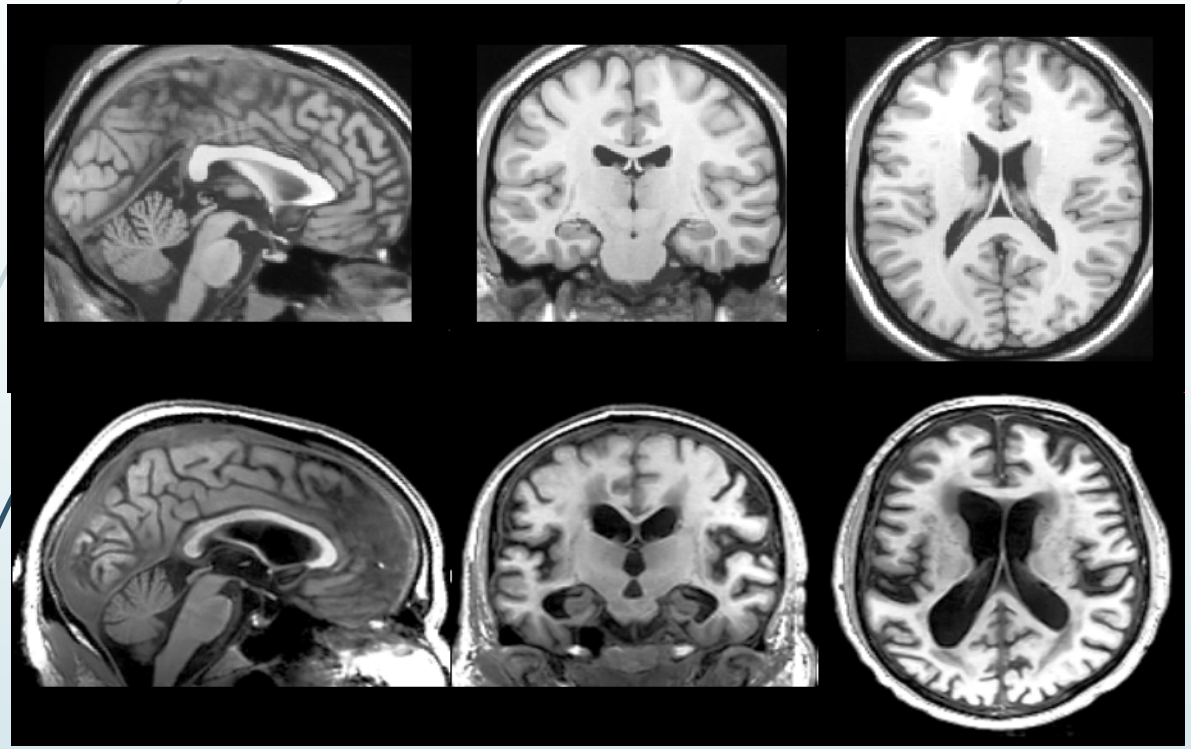
- Intensity normalization
  - CSF/GM/WM segmentation
  - VOI-based histogram match

Acquisition noises/variability

Example: can we distinguish AD subjects from control people by their T_1-weighted images?

Spatial registration

ICBM152-template



Registered T1-weighted scan

Registration = is the process of transforming different sets of data into one coordinate system with the alignement of corresponding structures

A combined 12 d.o.f. transformation is computed to minimize a given metric, mapping the MRI onto a reference image

A 12 d.o.f. transformation is a linear transformation that can manage rotations, translations, shearing and scaling.

Example: can we distinguish AD subjects from control people by their T_1-weighted images?

- Initial quality check
  - Image artifacts
  - Voxel size and aspect ratio

Acquisition noises/variability

- Noise removal
  - Steerable pyramid de-noising

Acquisition noises/variability

- Spatial registration
  - 3-way scalable (7 d.o.f.) + affine (12 d.o.f.)
  - Mutual information and normalized correlation metric

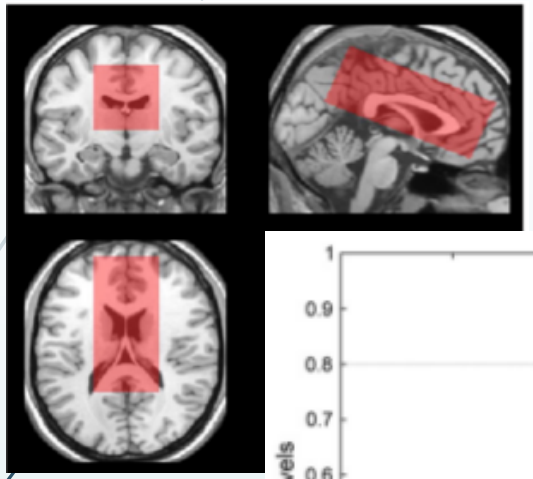
Physiological noises

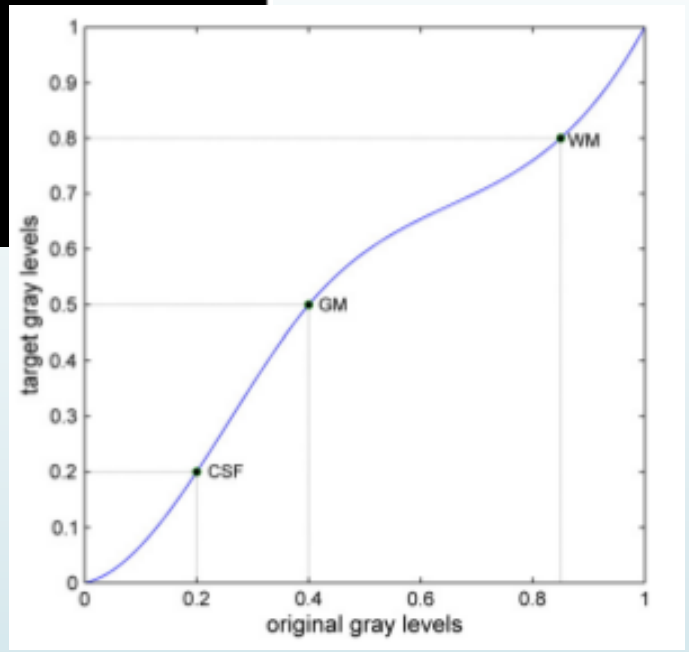
- Intensity normalization
  - CSF/GM/WM segmentation
  - VOI-based histogram match

Acquisition noises/variability

Example: can we distinguish AD subjects from control people by their T_1-weighted images?

Intensity normalization



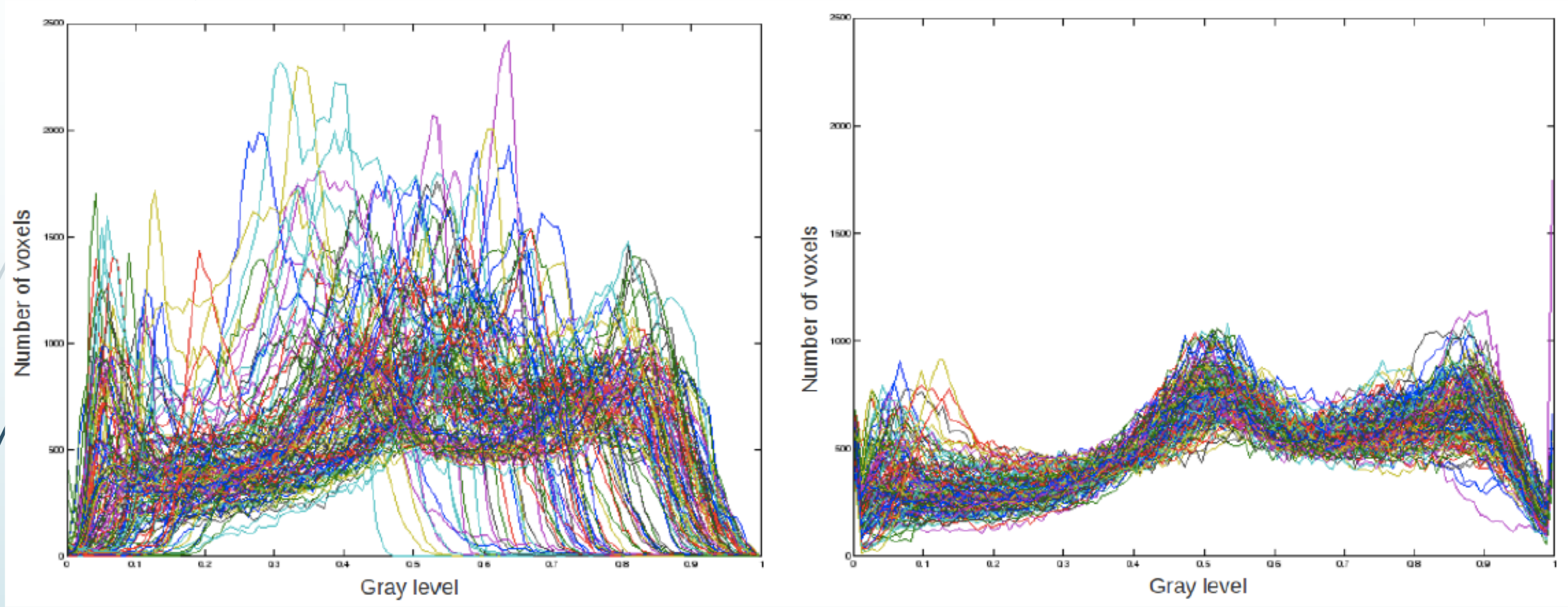


T1-weighted signal is not a quantitative measure.

We consider a reference image (template). The three clusters means of the new image are matched to the corresponding CSF/GM/WM mean levels on the reference image.
These are the fixed points through which a cubic spline is fitted. This non-linear intensity normalization pairs the three mean cluster intensities in the ROI (CSF/GM/WM) between each subject and the MNI reference image, and extends the mapping to the other gray levels by a smooth piecewise polynomial curve.

Example: can we distinguish AD subjects from control people by their T_1-weighted images?

Intensity normalization



Example: can we distinguish AD subjects from control people by their T_1-weighted images?

- Region (VOI) extraction
  - Template matching, rigid (6 d.o.f.) registration

Physiological noises

- Features computation
  - 4 different neighborhoods
  - Intensity & texture based filtering

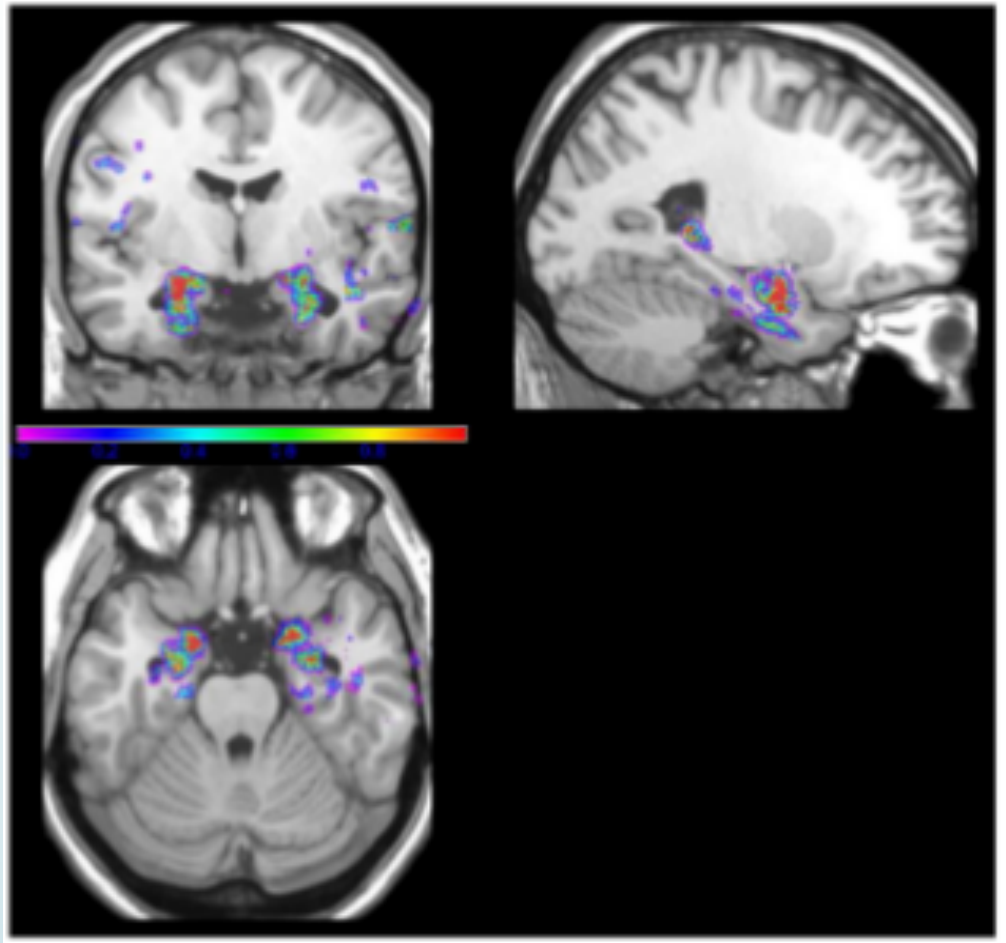
- Classification
  - Random Forest (RF) important variable map
  - Support Vector Machine (SVM) classifier

"Gold standard" noises

Example: can we distinguish AD subjects from control people by their T_1-weighted images?

Important features maps

Important features selection by a Random Forest Classifier (IFM, Important Features Map) Control and AD groups



Example: can we distinguish AD subjects from control people by their T_1-weighted images?

- Region (VOI) extraction
  - Template matching, rigid (6 d.o.f.) registration

Physiological noises

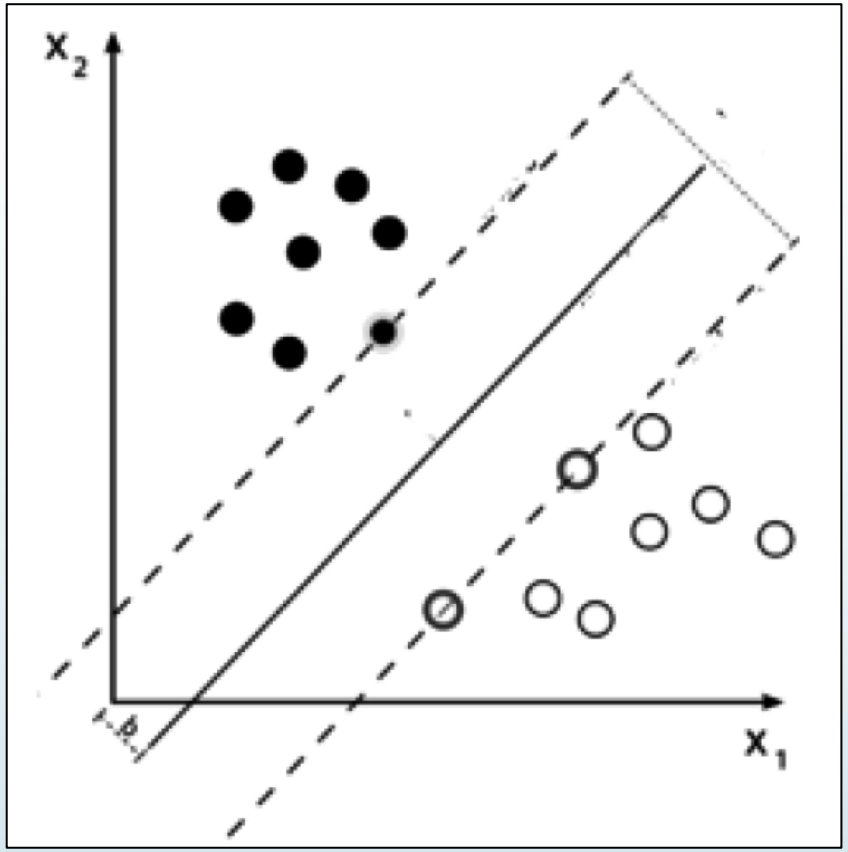
- Features computation
  - 4 different neighborhoods
  - Intensity & texture based filtering

- Classification
  - Random Forest (RF) important variable map
  - Support Vector Machine (SVM) classifier

"Gold standard" noises

Example: can we distinguish AD subjects from control people by their T_1-weighted images? **Classification**

The SVM classier training consists in the calculation of the hyperplane able to separate in the best way two N dimensional data set.



Example: can we distinguish AD subjects from control people by their T_1-weighted images?

Classification

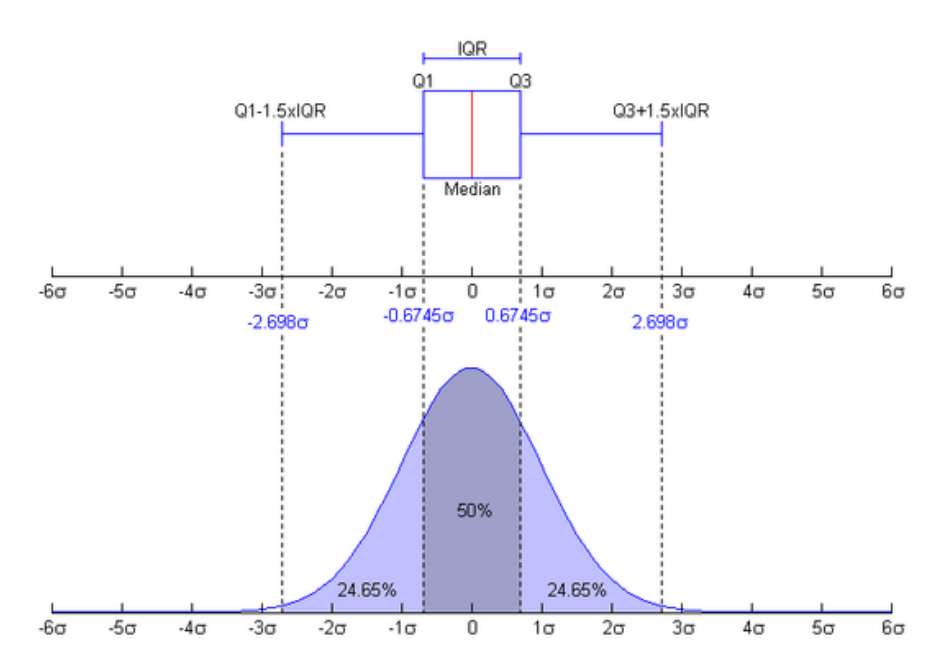
SVM classification
Support Vector Machine
(dimensionality problem
d ≈10^4 vs n ≈10^2)

We can split the problem in
1000 SVMs that classify 1/1000
of the features (d ≈ n)
And then we can average the
outcome of the 1000 classifiers

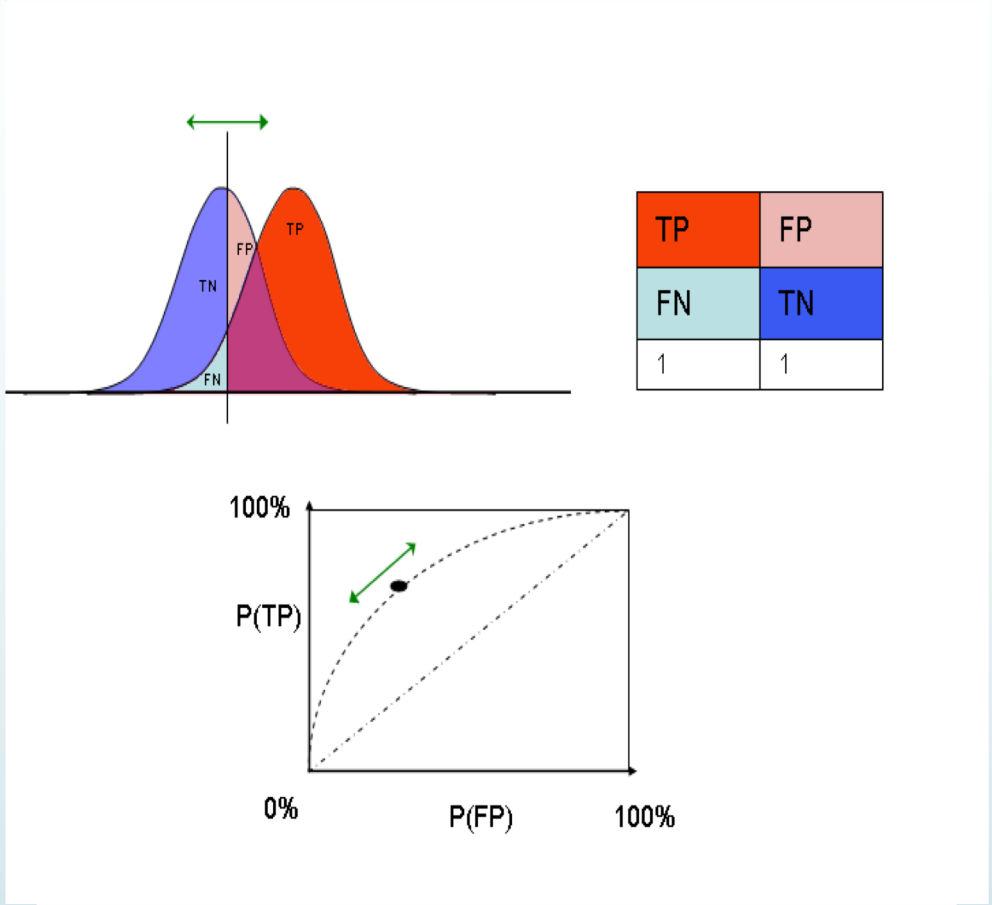
The final output is a
continuous index ranging
from -1 to 1
with -1=AD and 1=Control

Each subject is classified using this classification index
Quantitative, reproducible measure

Example: can we distinguish AD subjects from control people by their T_1-weighted images? **Classification**



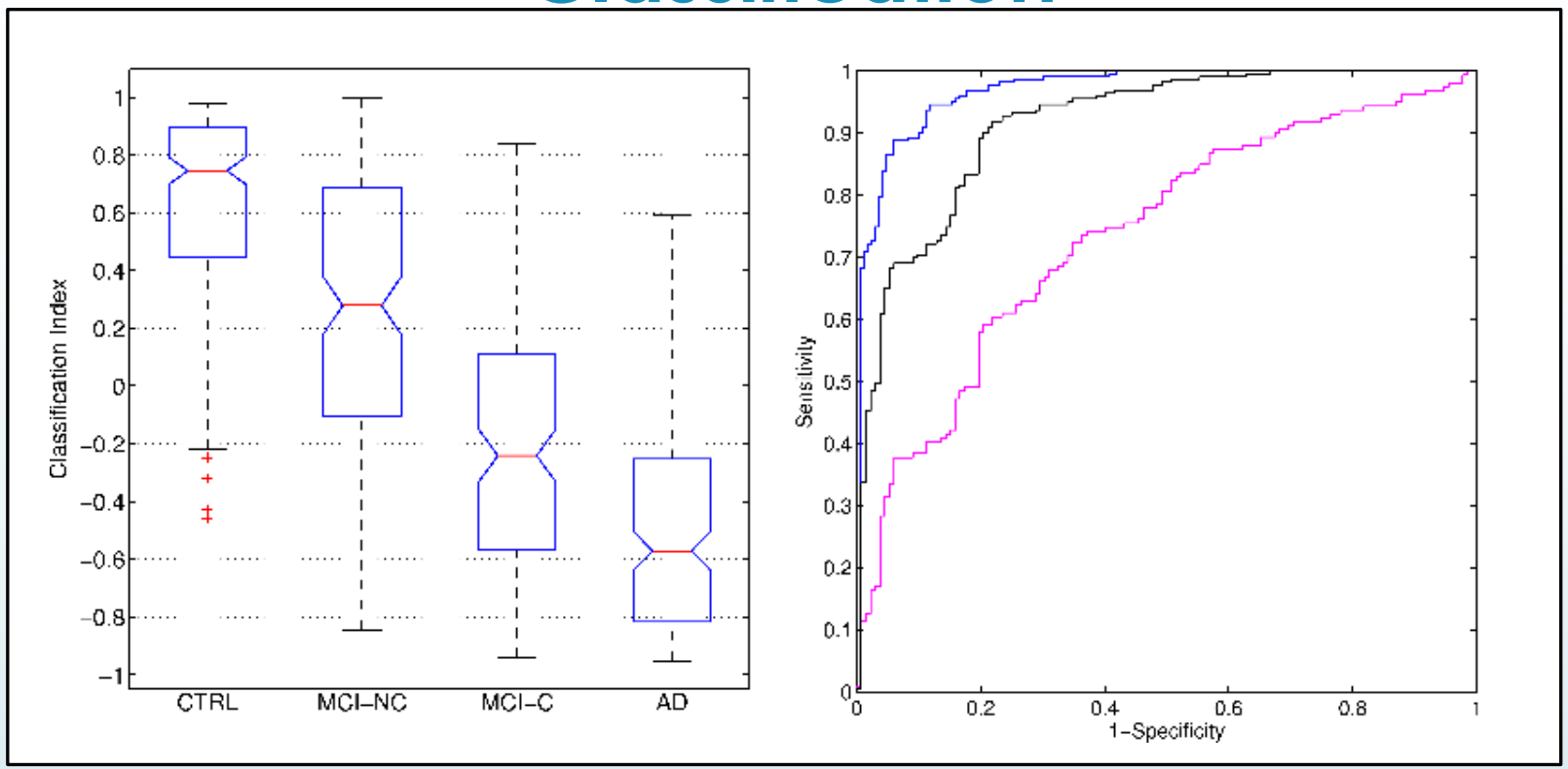
Box plot



ROC (Receiver operating characteristic) curve

Example: can we distinguish AD subjects from control people by their T_1-weighted images?

Classification



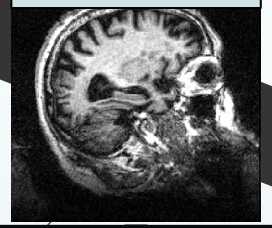
Discrimination capability between control and AD cohorts described by an area under the ROC curve (AUC) of 0.97. AUC for the ROC curve of control vs MCI converter cohorts is 0.92; AUC for the ROC curve of MCI vs MCI converter is 0.74.

Example: can we distinguish AD subjects from control people by their T_1-weighted images?

Yes

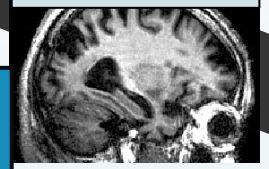
Pre-processing

RAW



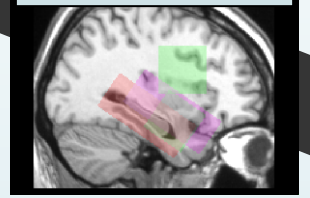
Quality assessment

NORMALIZED

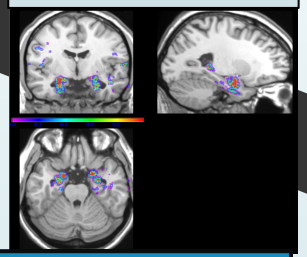


Disease-specific volumes (clinically driven)

AD-regions

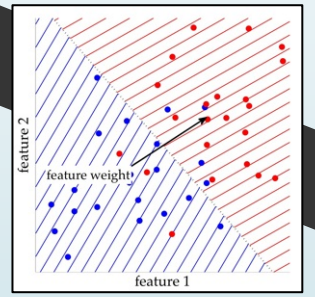


FEATURES

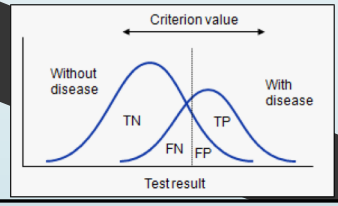


Features extraction
A ML algorithm selects those who most likely discriminate between Controls/AD

Classification
A ML classifier takes the feature subset outputs the distance between the input set and the discriminating hyper surface

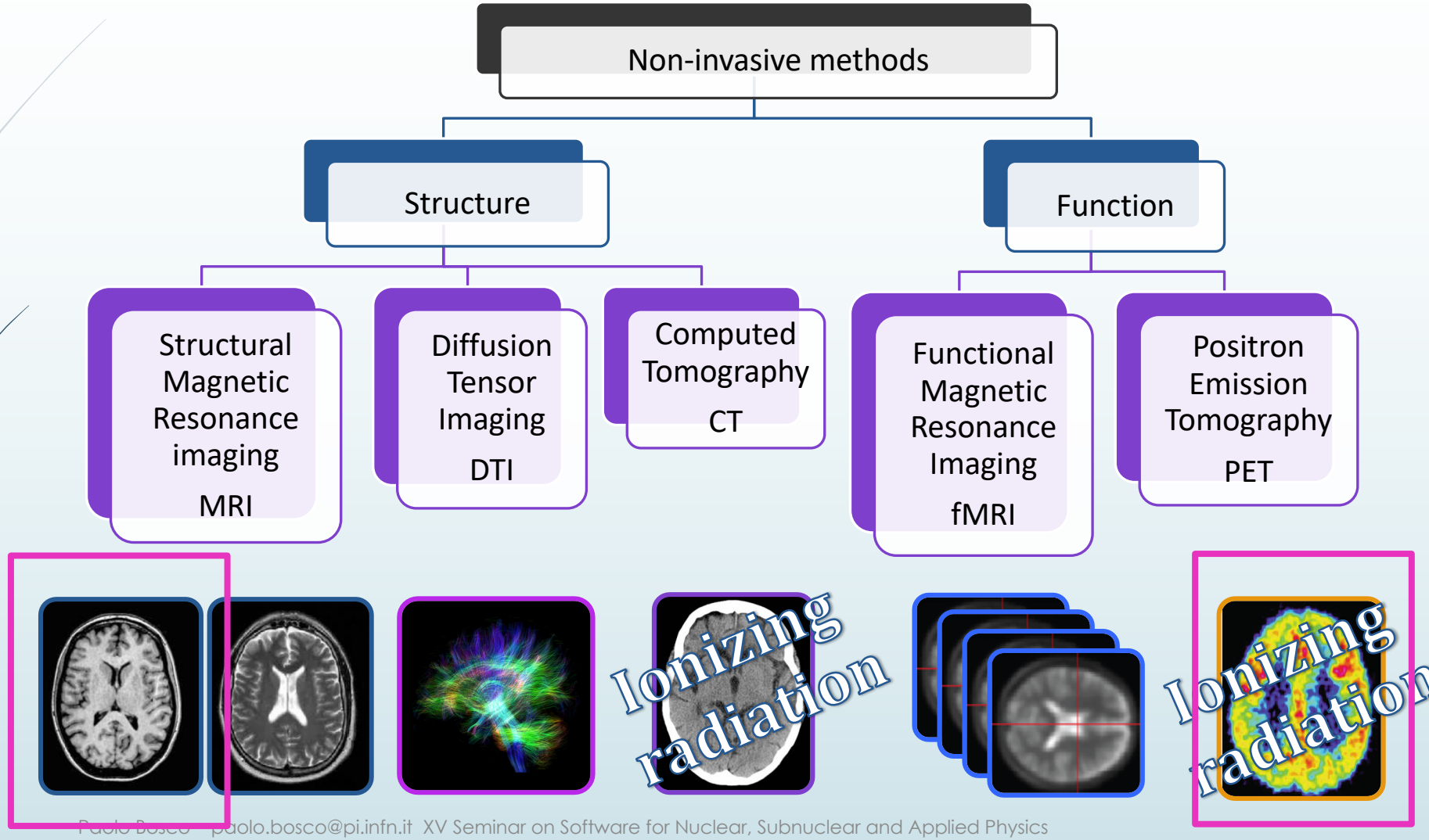


A number is assigned to each subject ranging from 1 (normalcy) to -1 (AD)

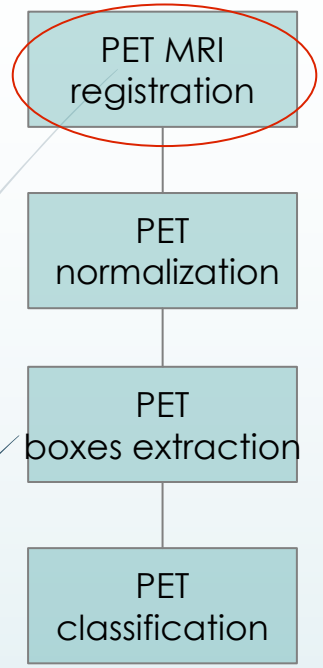


CLASSIFICATION INDEX

What about brain function? FDG-PET

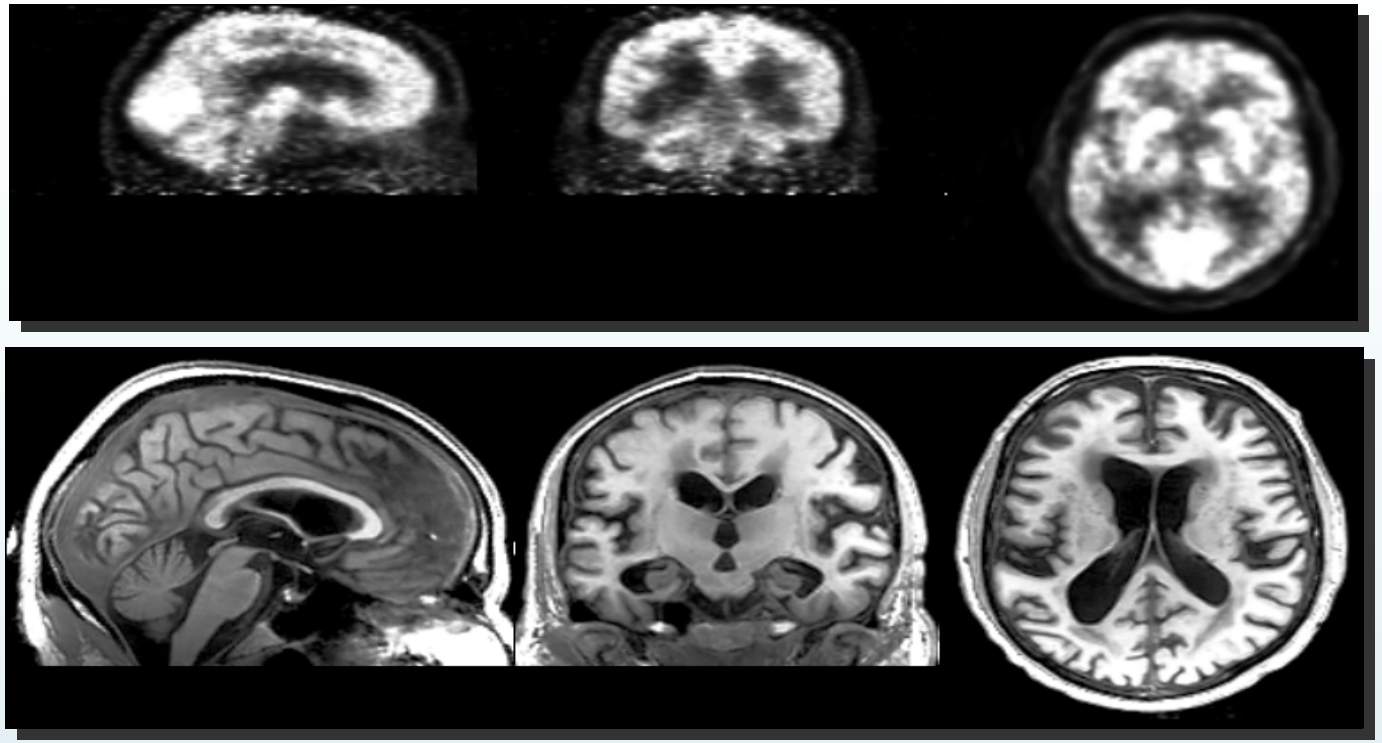


Example: can we distinguish AD subjects from control people by their FDG-PET images?



FDG-PET

MRI

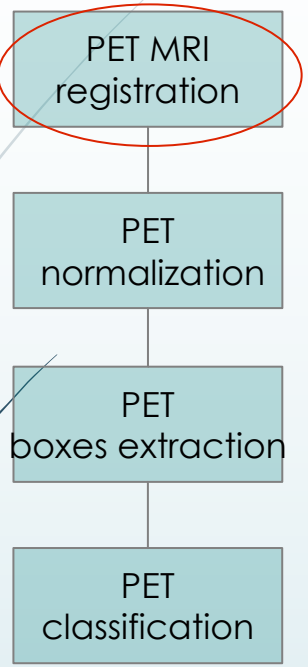


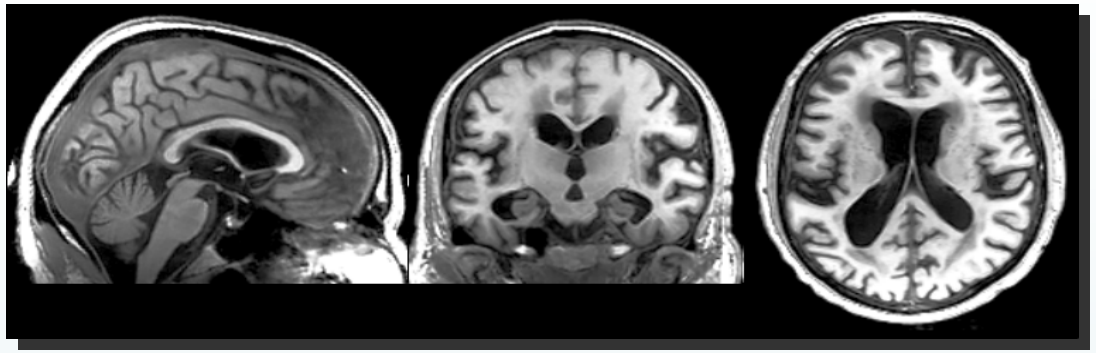
Sagittal Coronal Axial

Again... Registration = is the process of transforming different sets of data into one coordinate system with the alignement of corresponding structures

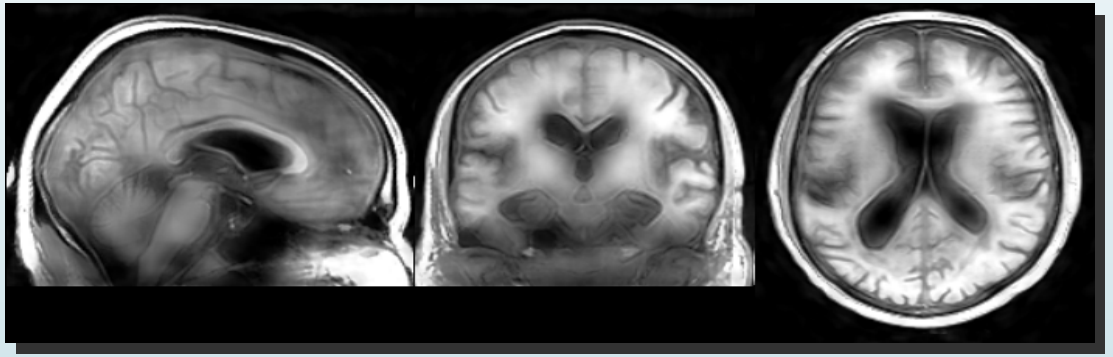
→ minimization problem on a metric

Example: can we distinguish AD subjects from control people by their FDG-PET images?





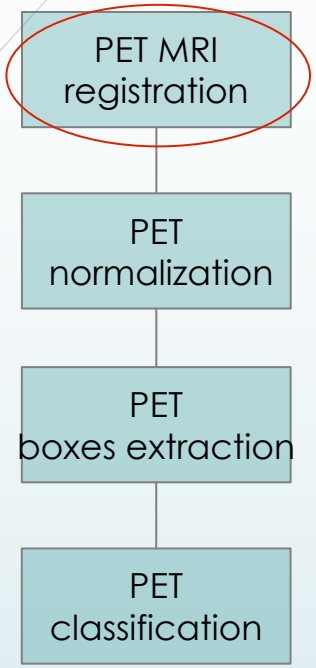
Curvelet transform
~Low pass filter
and inverted transform



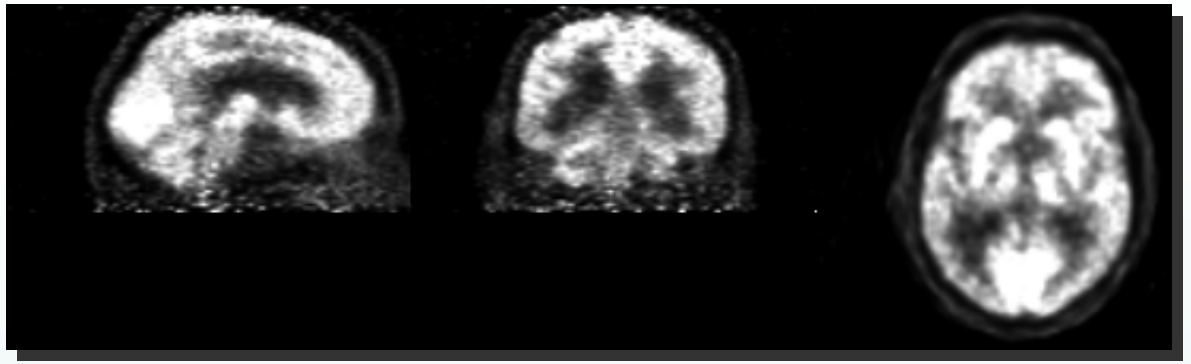
High scale (low spatial frequencies) and high directional informations preserved

Border and main CSF structures preservation

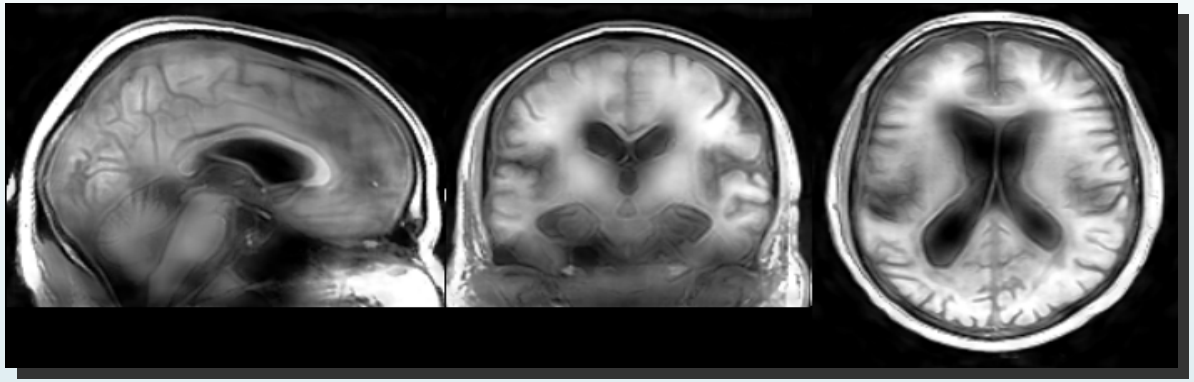
Example: can we distinguish AD subjects from control people by their FDG-PET images?



FDG-PET



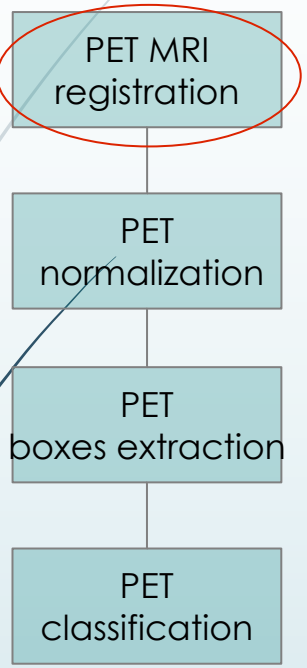
MRI after low frequency coefficients selection

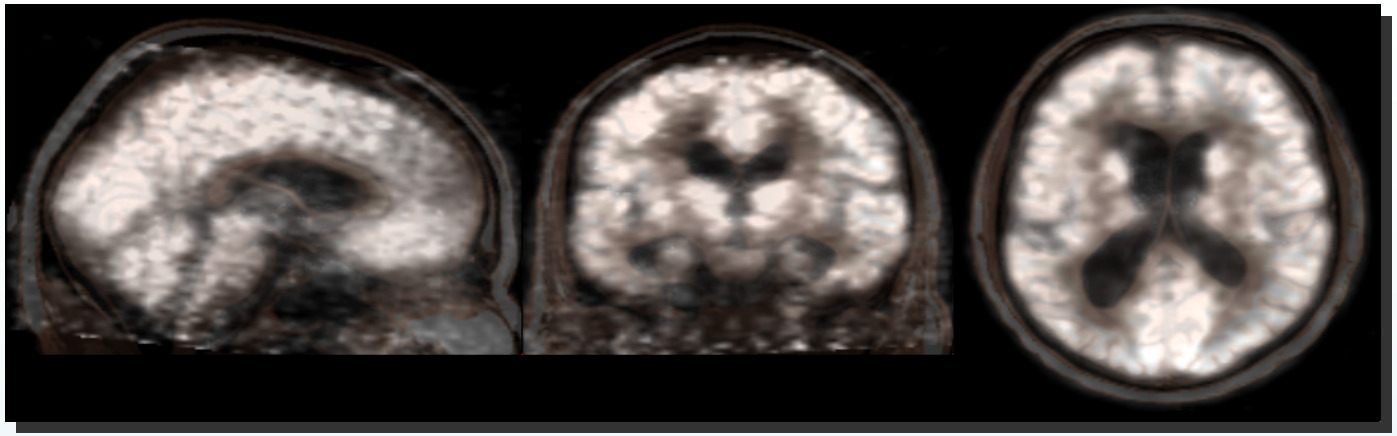


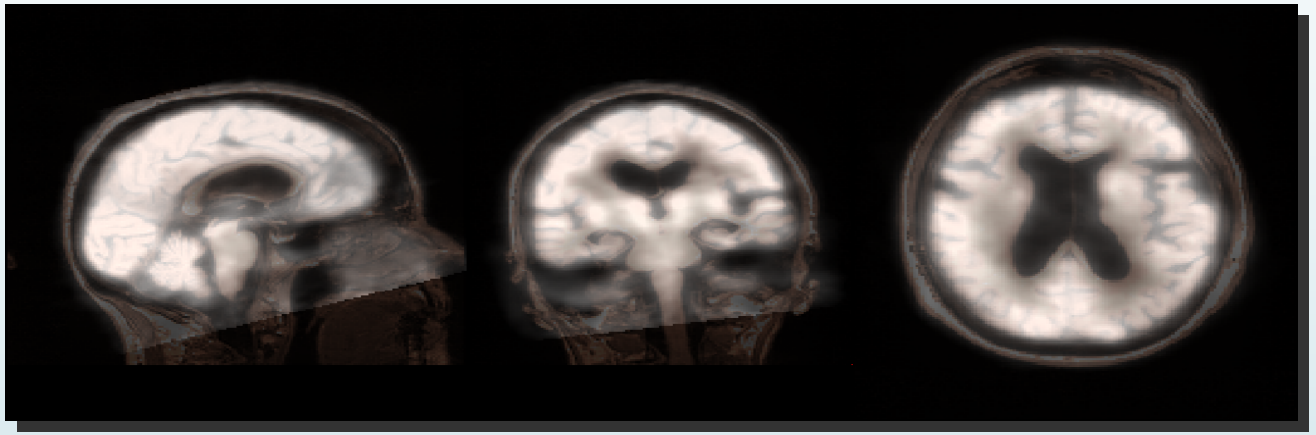
Registration by 7 d.o.f transform using mutual information as metric

$$I(X;Y) = \sum_{x,y} p(x,y) \log \frac{p(x,y)}{p(x)\,p(y)}$$

Example: can we distinguish AD subjects from control people by their FDG-PET images?



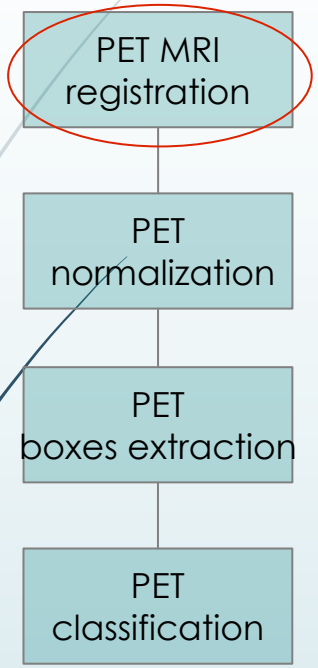


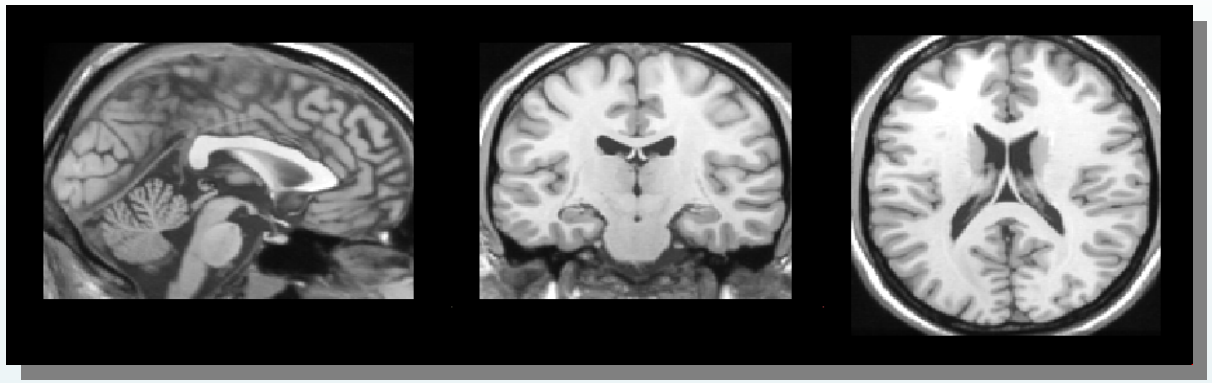


Many applications-> e.g. registration of low resolution in-beam PET to High resolution CT scans for activity evaluation during hadrontherapy.

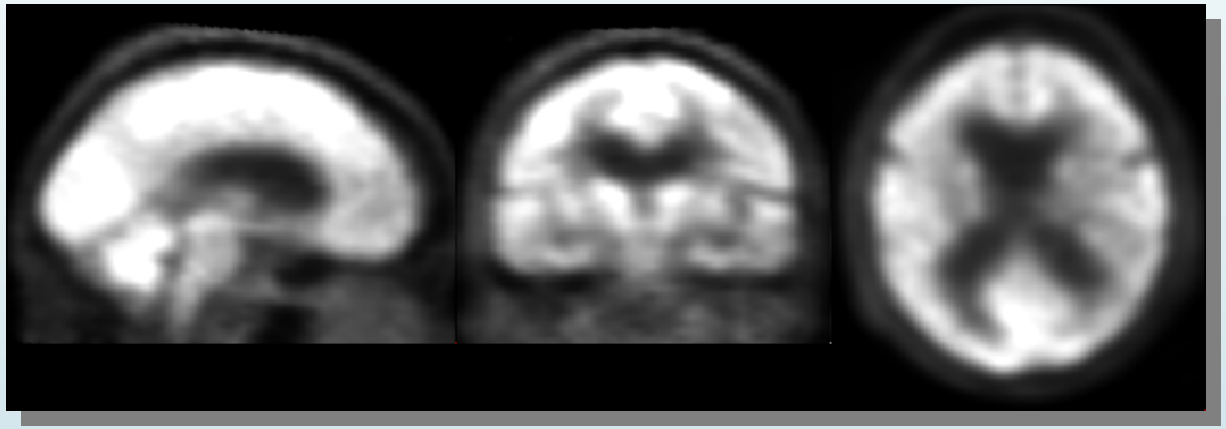
Example: can we distinguish AD subjects from control people by their FDG-PET images?

- Registration on ICBM152 template. Same transform matrix applied to PET image.



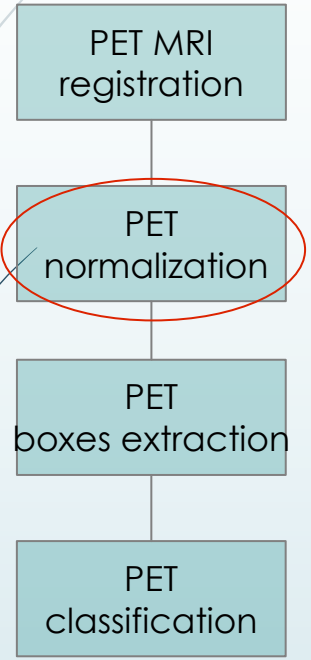


ICBM152 template

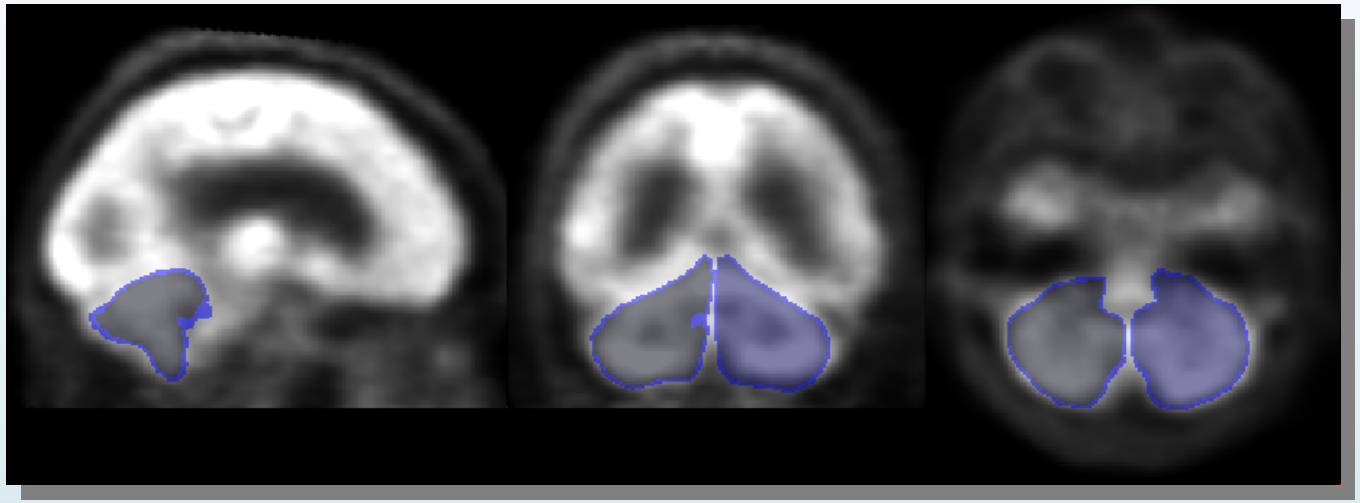


oriented PET

Example: can we distinguish AD subjects from control people by their FDG-PET images?

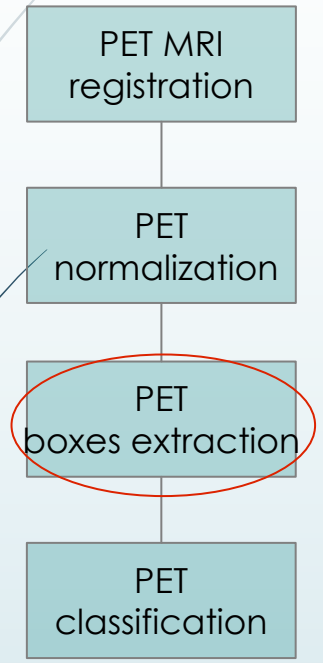


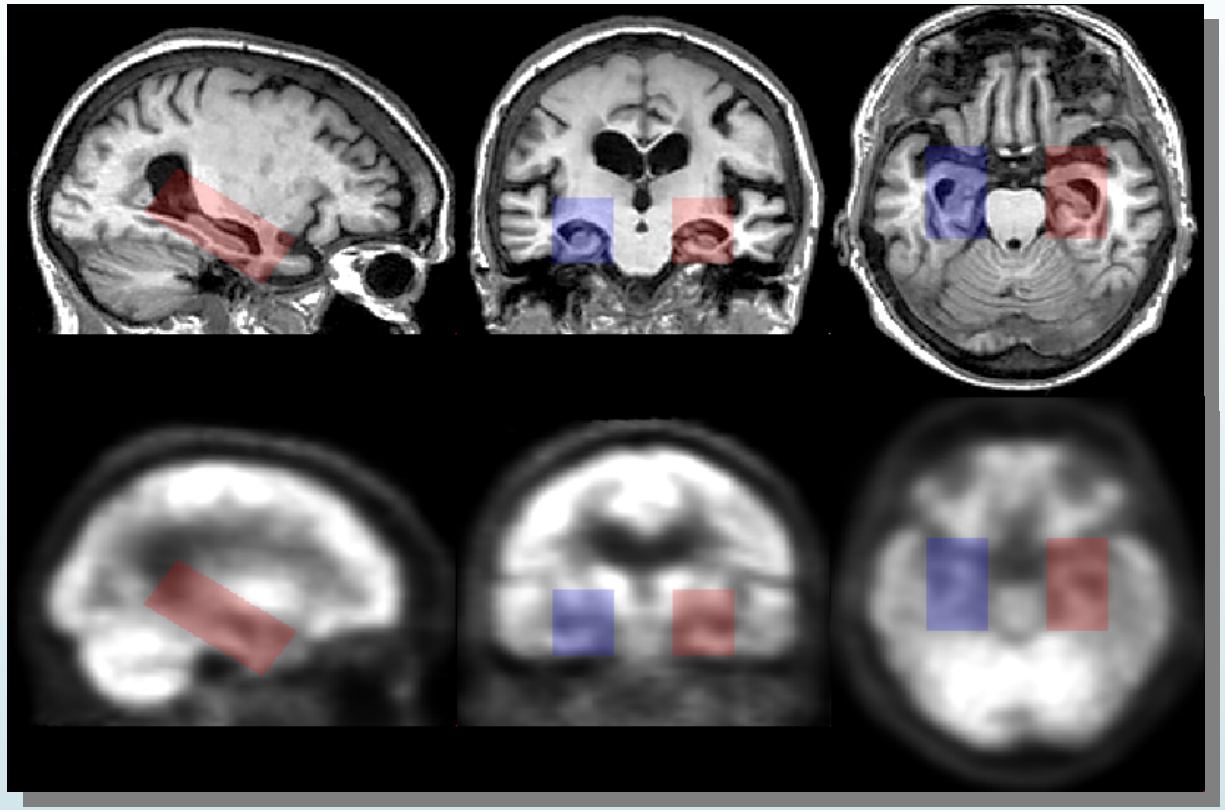
- PET signal intensity normalization
  - Cerebellum automatic segmentation
  - Global counts normalized on average cerebellum signal



Example: can we distinguish AD subjects from control people by their FDG-PET images?

- Hippocampal boxes extraction





Example: can we distinguish AD subjects from control people by their FDG-PET images?

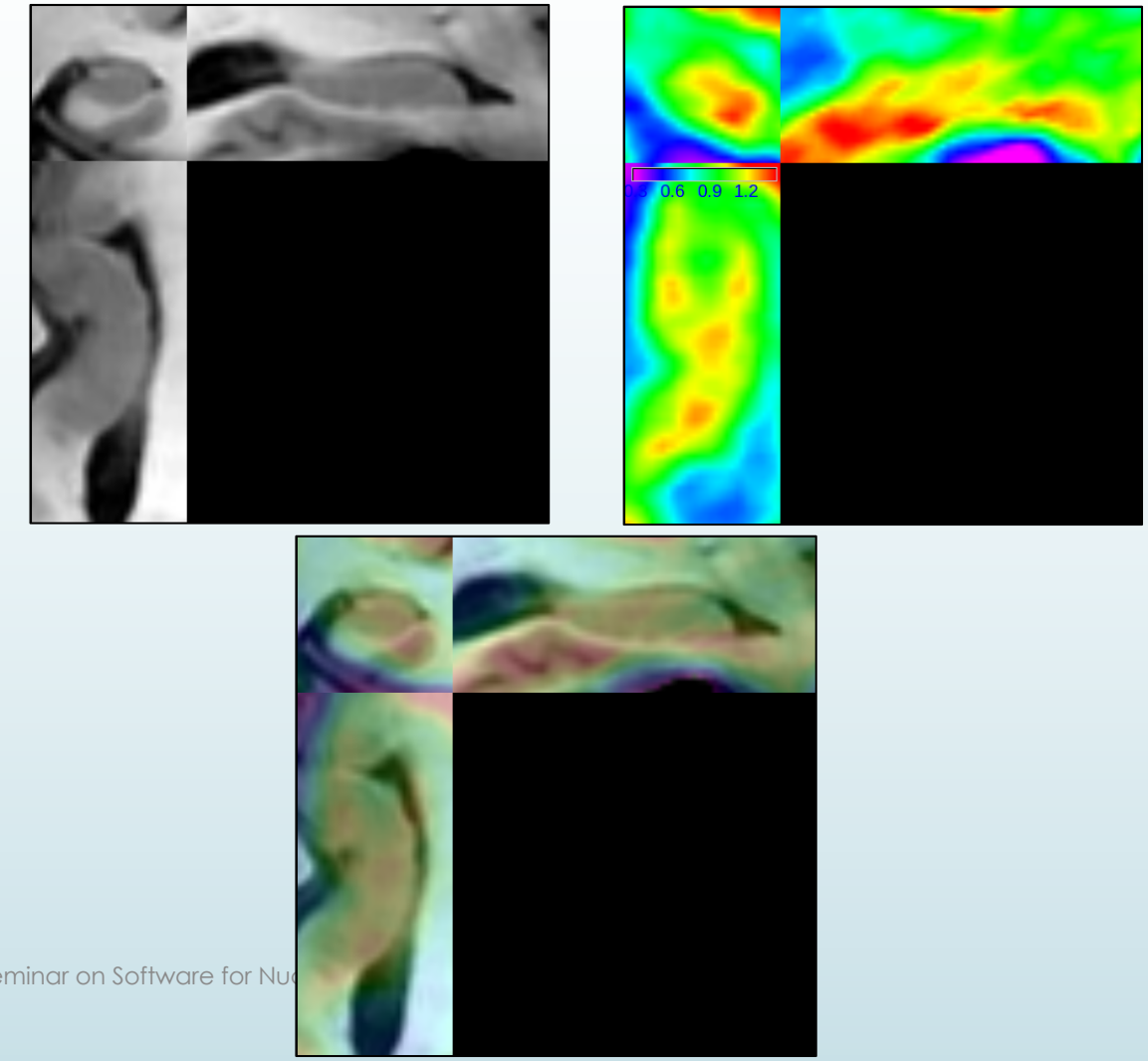
PET MRI registration

PET normalization

PET boxes extraction

PET classification

Hippocampal ROIs FDG-PET intensities (false color) overlaid to MRI signal



Example: can we distinguish AD subjects from control people by their FDG-PET images?

Hippocampal ROIs FDG-PET intensities (false color) overlaid to MRI signal

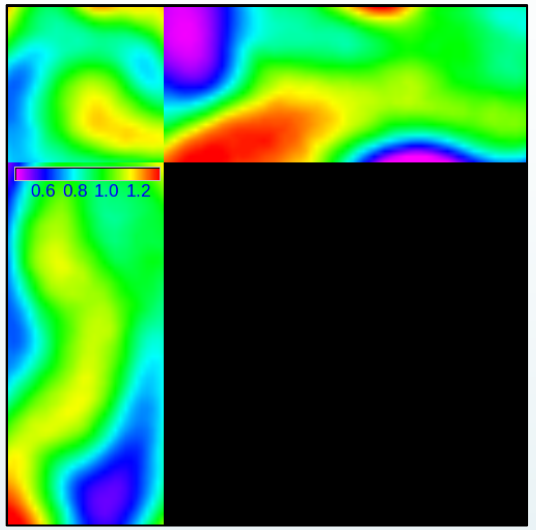
PET MRI registration

PET normalization

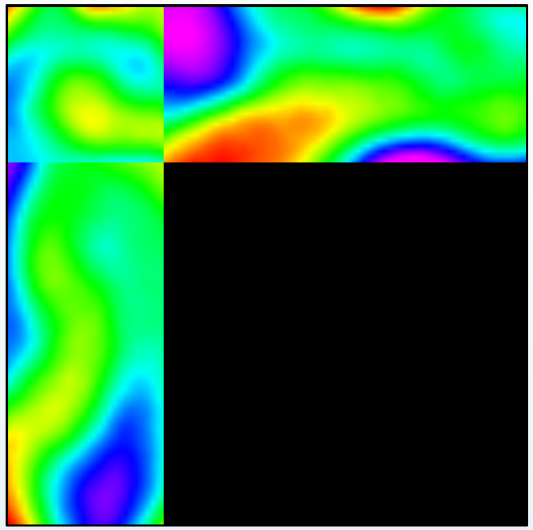
PET boxes extraction

PET classification

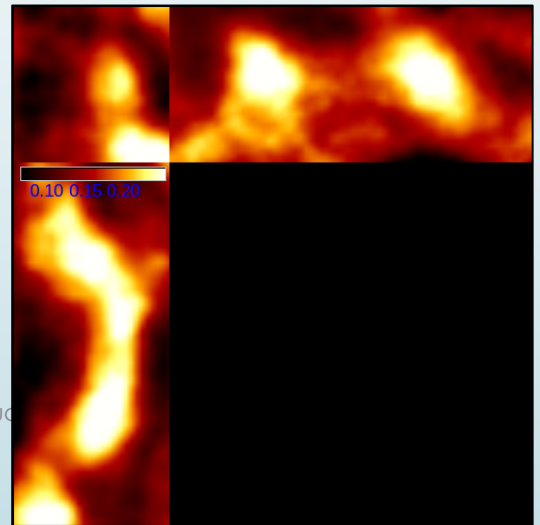
Median CTRL



–



Median AD



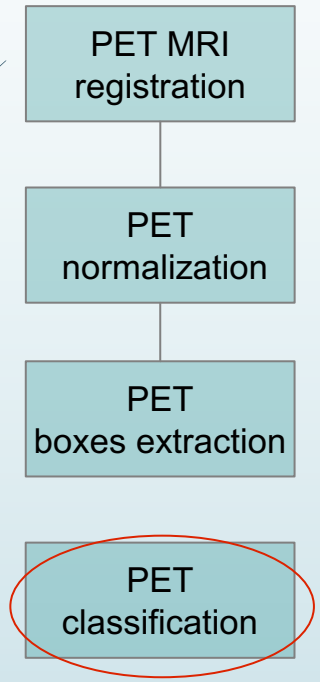
Delta

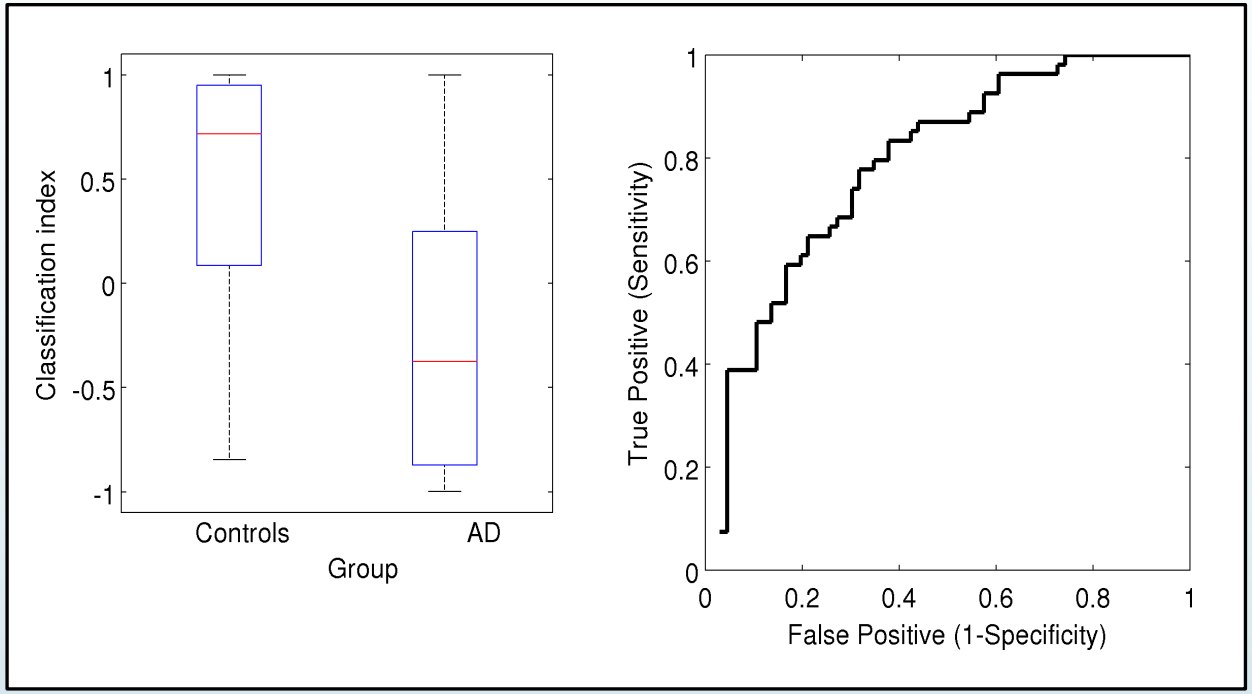
Example: can we distinguish AD subjects from control people by their FDG-PET images?

Yes, quite well

Functional index distribution calculated on hippocampal boxes

Controls/AD separation. 26 Controls (age 76 ± 4) 29 AD (age 75 ± 6)





AUC=0.80

What about data processing noise? An example from an Autism Spectrum Disorder study

Let's choose a quite simple region of interest potentially involved in ASD: the brainstem

- Deficit in social communication abilities and the presence of restricted, repetitive behaviours represent the core features of autism spectrum disorder (ASD).

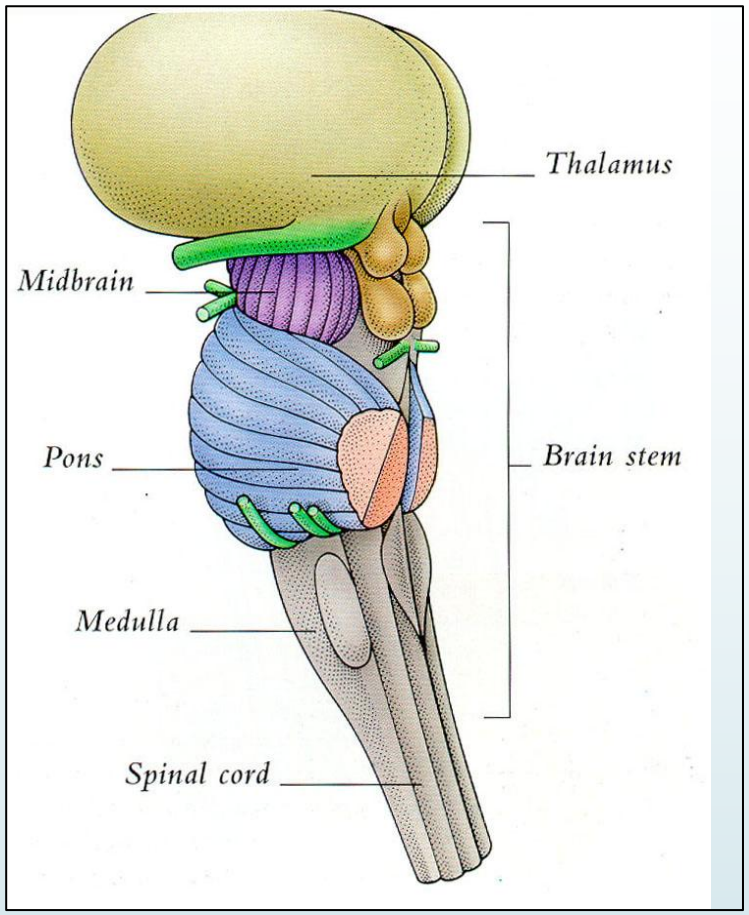
[DSM_5]

- In addition, sensorimotor abnormalities have been consistently reported in ASD individuals as an early impairment that may precede the development of defining characteristics

[Teitelbaum, *Proc Natl Acad Sci USA* 1998, Ozonoff, *Autism Res* 2008, Hilton, *Res Autism Spectr Disord*, 2007]

What about data processing noise? An example from an Autism Spectrum Disorder study

- Motor abilities depend on multiple interacting pathways including cortico-cortical, cortical-subcortical, and cortico-cerebellar connections that reach spinal motor neurons through the brainstem.
- From the anatomical point of view, the brainstem consists of midbrain (or mesencephalon), pons and medulla oblongata along the rostro-caudal axis.
- It's involved in several basic functions, including regulation of heart rate, breathing, alertness, sleeping, and eating It also plays a pivotal role in sensory information processing, in eliciting goal-oriented behaviour, in the regulation of social attention, and in the modulation of emotions.



What about data processing noise? An example from an Autism Spectrum Disorder study

Variable	Subject group, mean ± std [range]			
Age (months)	ASD (n=76)		Controls (n=76)	
	53 ± 17 [25 - 88]		53 ± 18 [22 - 89]	
NVIQ	71 ± 22 [30 - 113]		73 ± 23 [35 - 112]	
Age (months)	Males (n=38)	Females (n=38)	Males (n=38)	Females (n=38)
	53 ± 16 [27 - 87]	53 ± 18 [25 - 88]	53 ± 17 [24 - 88]	53 ± 19 [22 - 89]
NVIQ	71 ± 21 [39 - 113]	71 ± 22 [30 - 103]	74 ± 23 [43 - 112]	71 ± 24 [35 - 100]

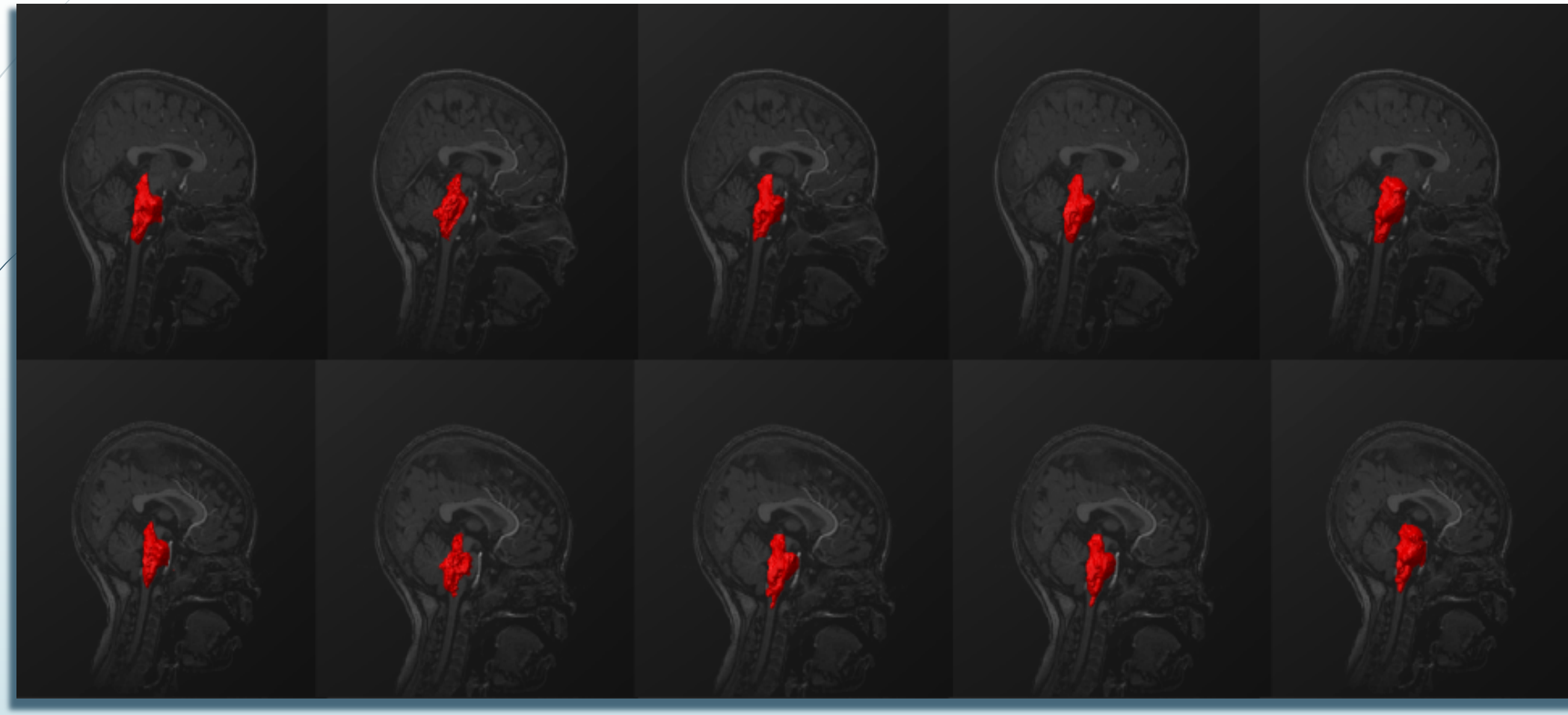
ASD children met the criteria for a diagnosis in the autism spectrum according to DSM-5, and underwent a MRI scan (T1-weighted series (FSPGR), 1.5 T GE scanner).

What about data processing noise?
An example from an Autism Spectrum Disorder study

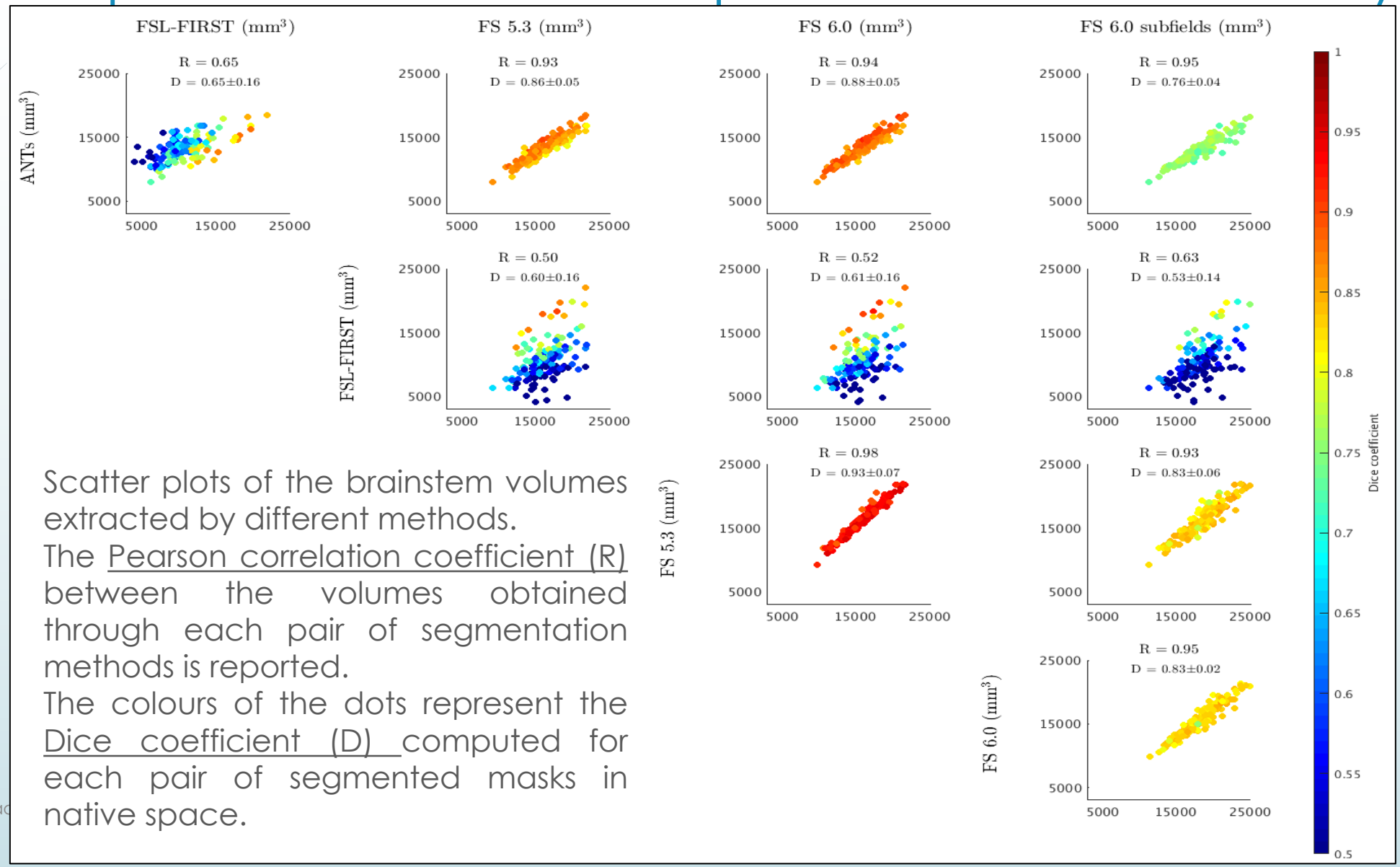
Brainstem segmentation methods

- FIRST (FSL version 5.0.8)
- FreeSurfer (version 5.3, 6.0 and 6.0 with brainstem substructure extraction)
- ANTs (Advanced Normalization Tools)

What about data processing noise?
An example from an Autism Spectrum Disorder study

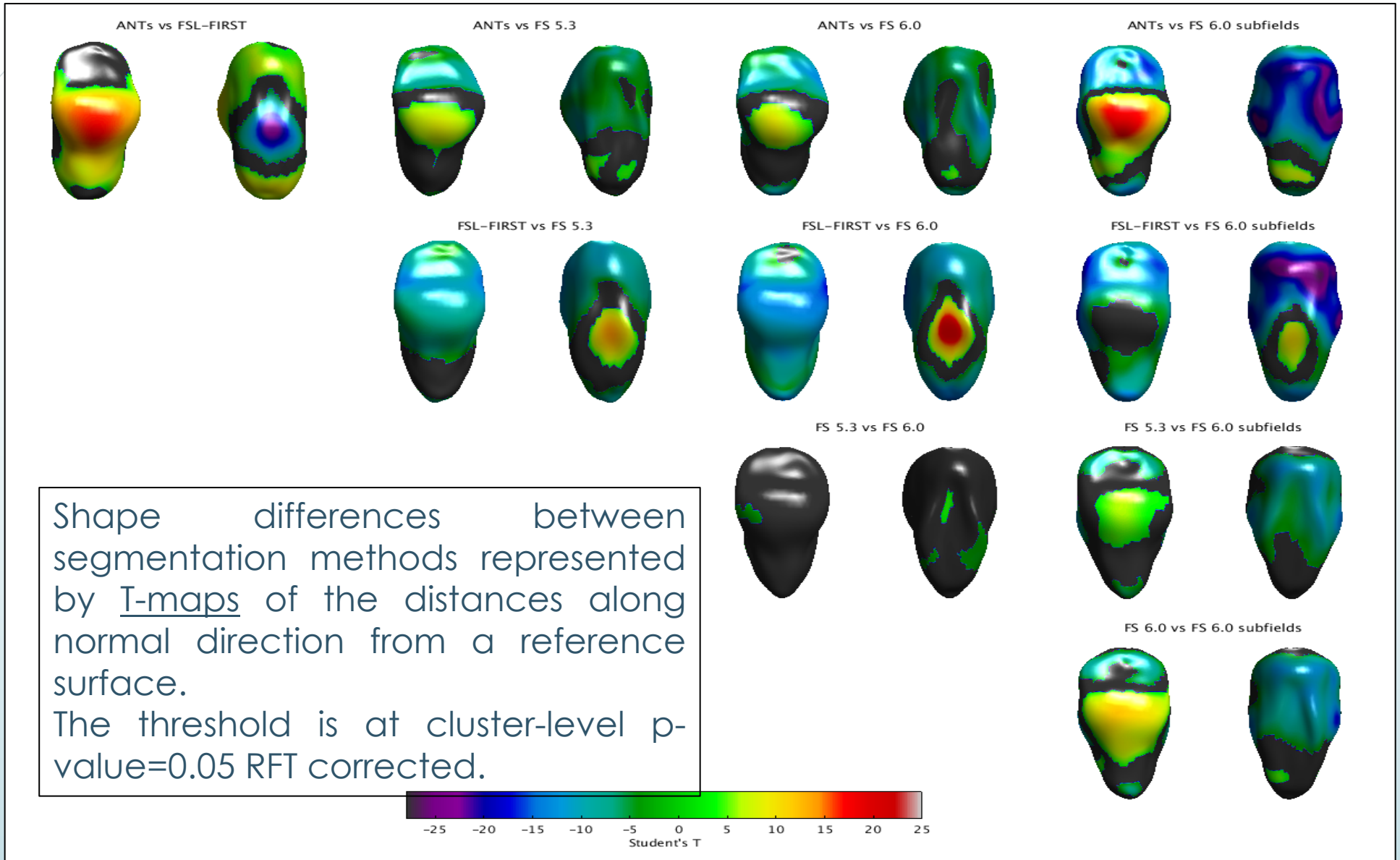


What about data processing noise?
An example from an Autism Spectrum Disorder study



Scatter plots of the brainstem volumes extracted by different methods.
The Pearson correlation coefficient (R) between the volumes obtained through each pair of segmentation methods is reported.
The colours of the dots represent the Dice coefficient (D) computed for each pair of segmented masks in native space.

What about data processing noise?
An example from an Autism Spectrum Disorder study



Shape differences between segmentation methods represented by T-maps of the distances along normal direction from a reference surface.
The threshold is at cluster-level p-value=0.05 RFT corrected.

What about data processing noise?
An example from an Autism Spectrum Disorder study

Brainstem volume analysis (ANOVA)

Variable		MF			M			F		
		Mean (SE)		ANOVA	Mean (SE)		ANOVA	Mean (SE)		ANOVA
		ASD	CTRL	p	ASD	CTRL	P	ASD	CTRL	p
BS volume (ml)	ANTs	13.53 (0.14)	13.03 (0.14)	0.016*	14.29 (0.22)	13.62 (0.22)	0.043*	12.77 (0.18)	12.45 (0.18)	0.230
	FSL -FIRST	10.60 (0.32)	10.73 (0.32)	0.776	11.92 (0.48)	11.62 (0.48)	0.668	9.40 (0.43)	9.87 (0.42)	0.452
	FS 5.3	16.62 (0.21)	15.90 (0.20)	0.016*	17.58 (0.32)	16.43 (0.31)	0.013*	15.63 (0.26)	15.43 (0.26)	0.597
	FS 6.0	16.21 (0.19)	15,56 (0.19)	0.020*	17.01 (0.29)	16.01 (0.29)	0.018*	15.33 (0.25)	15.15 (0.25)	0.647
	FS 6.0 substructures	18.81 (0.20)	18.12 (0.20)	0.021*	19.79 (0.31)	18.84 (0.31)	0.040*	17.73 (0.25)	17.46 (0.25)	0.466

What about data processing noise? An example from an Autism Spectrum Disorder study

Lesson to be learned

Even in a monocentric study (no different scanners), we can obtain controversial results due to a different image processing

Differences can be caused by

- varying definitions of brainstem structure,
- the use of different templates
- other varying effects on methods
- imaging artifacts and acquisition settings which may have different impact on different methods.

Let's recap a bit

- Medical images are no more just pictures observed by the physicians.
- They can be treated as signals and, as such, they can be handled with many of the techniques developed in Physics for signals analysis.
- Through the so called Radiomics, we can extract large amount of quantitative features: anatomical features (volumes, shapes etc), functional features etc.
- Many sources of variability can hinder this possibility (Acquisition noises/variability, Physiological noises/variability, Gold standard noise)
- We can handle many of these sources of noises BUT we must keep in mind not to introduce other sources of variability (data processing noise)

Future directions

- Data mining techniques are in a bursting evolution through the so called DEEP LEARNING.
- The clinical input will be less and less relevant since these techniques can alone identify which are the important region of interest and features.
- BUT they require HUGE amount of data to be trained

↓

Data storage, data sharing (Privacy handling, data non-homogeneity, SOPs to be developed), computing power (GPUs) etc etc...

For neuroimaging... Human Brain Project

Thank you for your attention!

I acknowledge all collaborators from INFN and Stella Maris and all people involved in the ARIANNA project for their precious contribution to the research presented in this talk (Many of the thoughts here reported have been originated by nice chats with Dott. Andrea Chincarini and Prof. Alessandra Retico).

ARIANNA was supported by the Italian Ministry of Health and Tuscany Region (grant GR-2010-2317873), by INFN-CNS5 (nextMR project) and by the Tuscany Region PAR-FAS 2007-2013 FAS Salute (ARIANNA project).

IRCCS FONDAZIONE STELLA MARIS

INFN Istituto Nazionale di Fisica Nucleare

Regione Toscana

FAS Fondo Aree Sottoutilizzate 2007-2013

REPUBBLICA ITALIANA