Quantitative methods for neuroimaging data analysis

Alessandra Retico

Istituto Nazionale di Fisica Nucleare (INFN), Sezione di Pisa, IT

X Seminar on Software for Nuclear, Subnuclear and Applied Physics

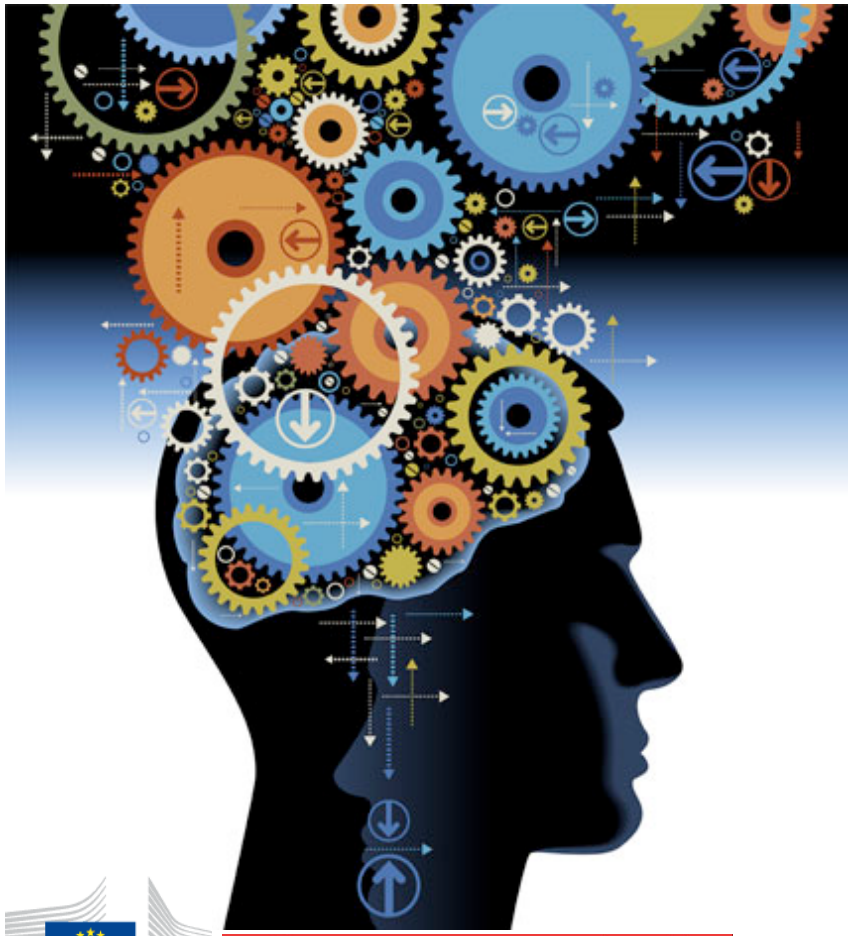
2-8 June 2013 *Porto Conte Alghero*

alessandra.retico@pi.infn.it

Outline

- Neuroimaging
  - Structural MRI of the brain
- Voxel-by-voxel analysis
  - Statistical methods (Voxel-based morphometry, VBM)
    - → Comparison between groups of subjects
  - Machine learning approaches
    - → Group comparison
    - → Single-subject classification
- Case studies:
  - Autistic Spectrum Disorders
  - Alzheimer's Disease

The human brain



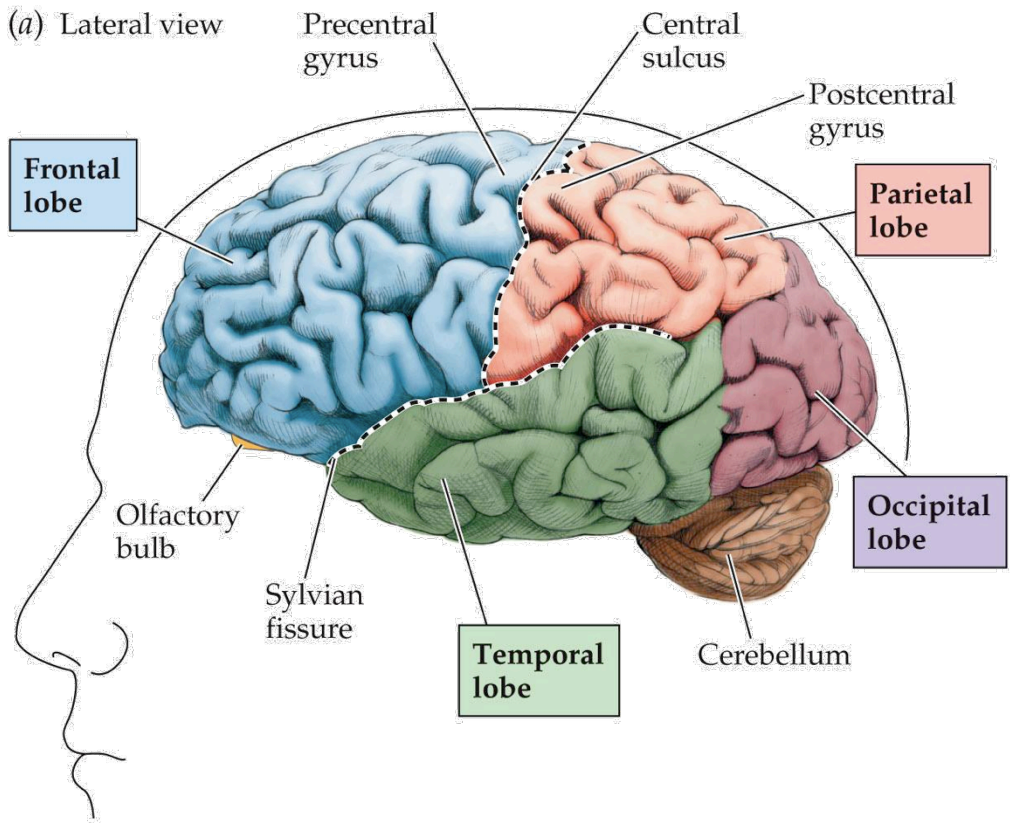
Brain numbers:

- ~ 1.3 kg
- ~ 10^{11} neurons
- ~ 10^{4} synaptic connections per neuron on average

The brain's network of neurons forms a massively parallel information processing system.

"My brain? That's my second favorite organ." Woody Allen, Sleeper, 1973

The human brain



***Biological Psychology 6e*, Figure 2.12 (Part 1)**

FRONTAL LOBE

- Concerned with reasoning, planning, parts of speech and movement (motor cortex), emotions, and problem-solving.

PARIETAL LOBE

- Concerned with perception of stimuli such as touch, pressure, temperature and pain.

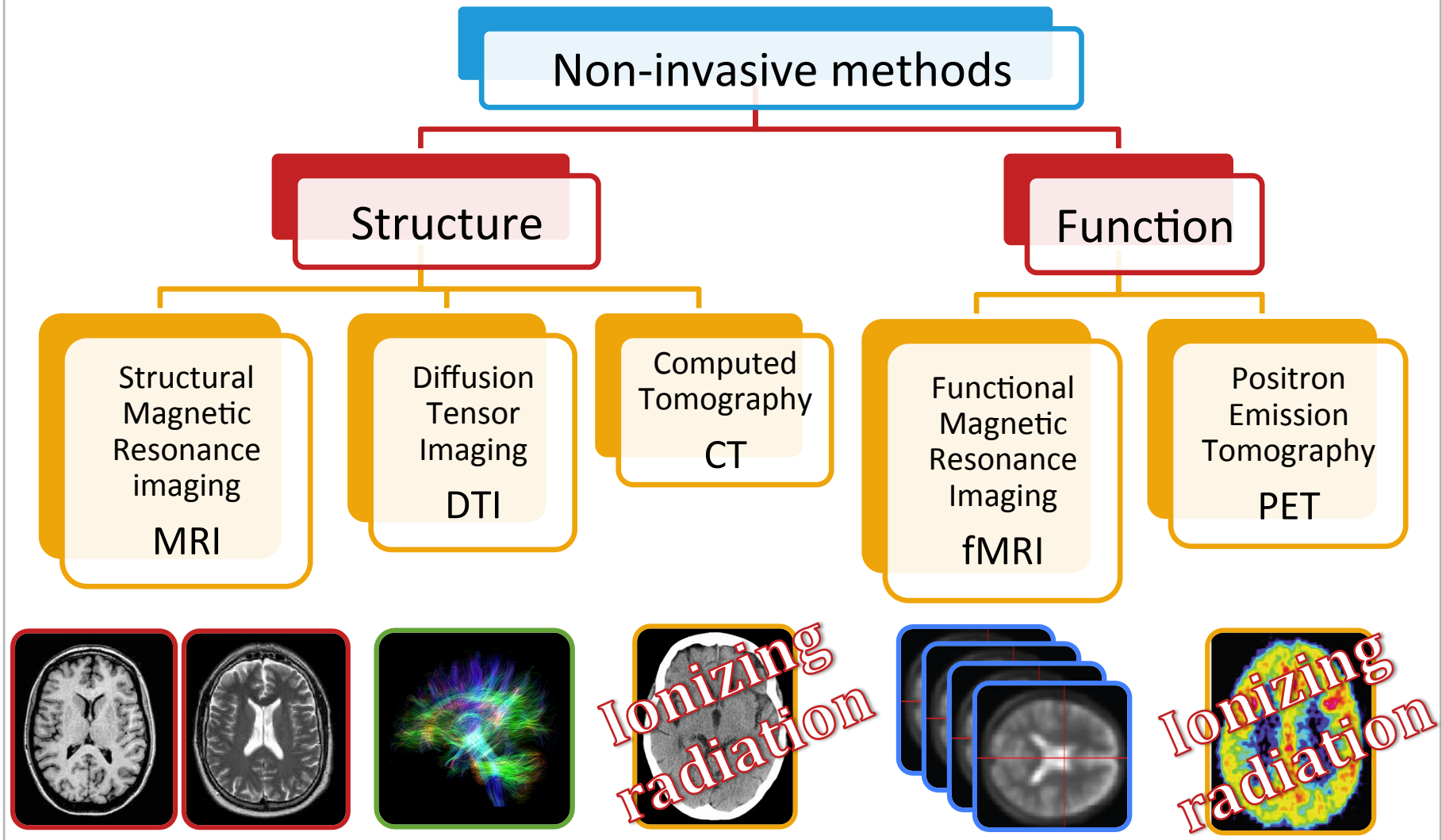
TEMPORAL LOBE

- Concerned with perception and recognition of auditory stimuli (hearing) and memory (hippocampus).

OCCIPITAL LOBE

- Concerned with many aspects of vision.

Measuring structure and function of the brain



Magnetic Resonance in Medicine

Magnetic Resonance (MR) data acquisition

1) Put subject in a static magnetic field [and leave him there]

   ⇒ Magnetizes the H nuclei in water (H_2O)

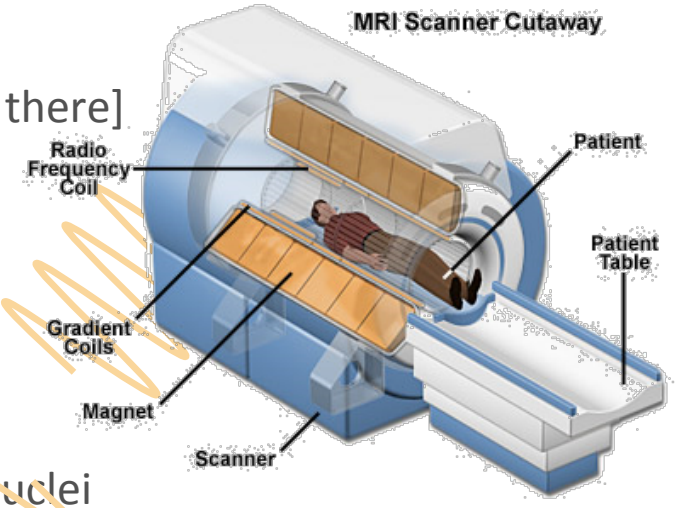
2) Transmit radio waves into subject

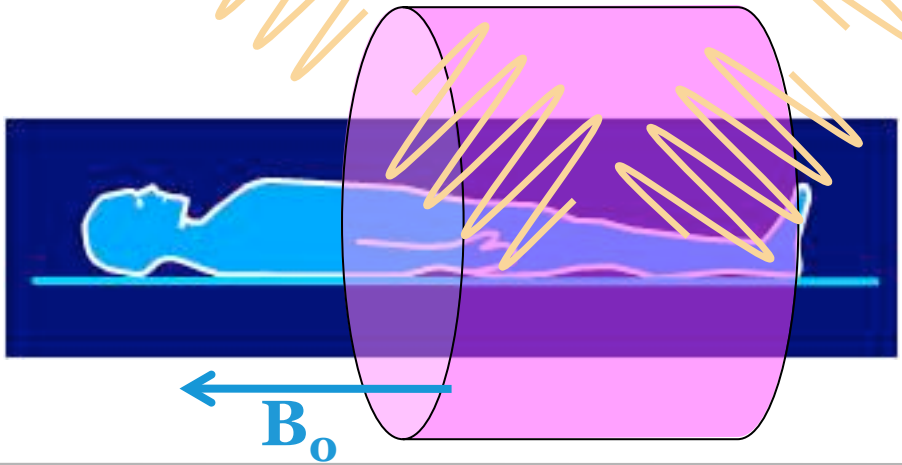
   ⇒ Perturbs the magnetization of the water

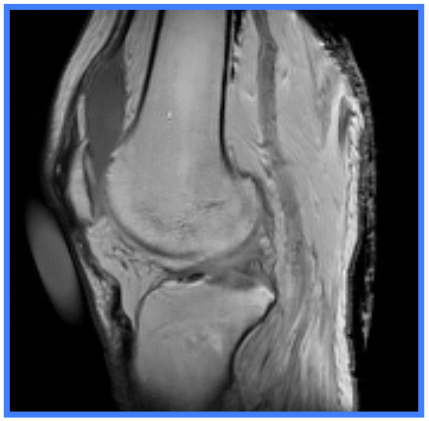
3) Turn off radio wave transmitter

4) Receive radio waves re-transmitted by subject's H nuclei

5) Process raw radio wave data to reconstruct images

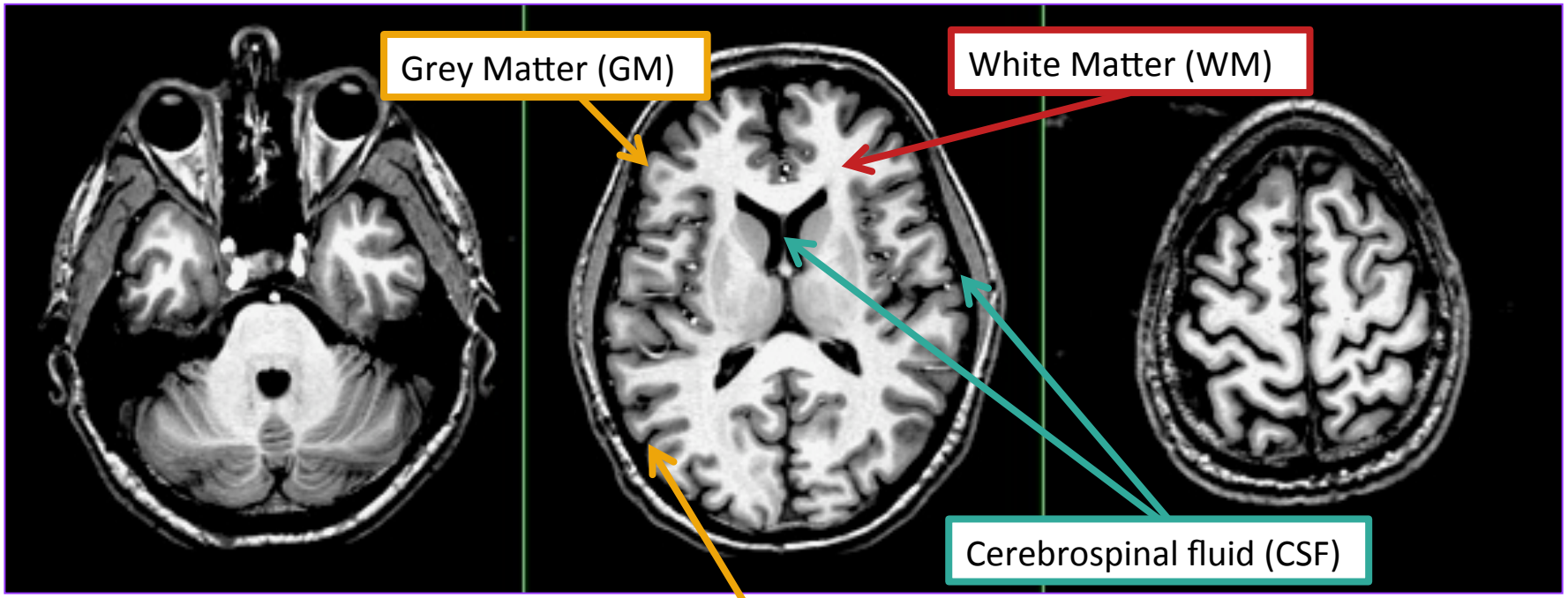






T_1-weighted images

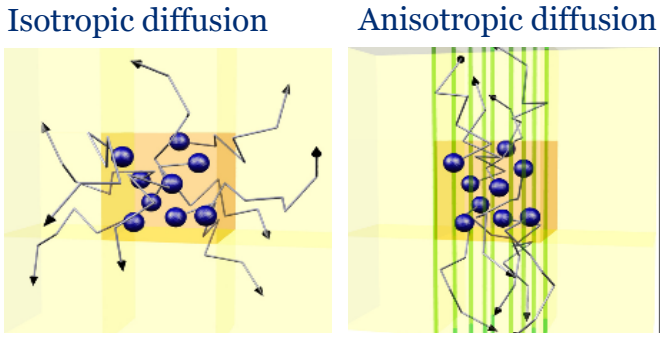
Axial slices of a human head with spatial resolution of about 1 mm^3

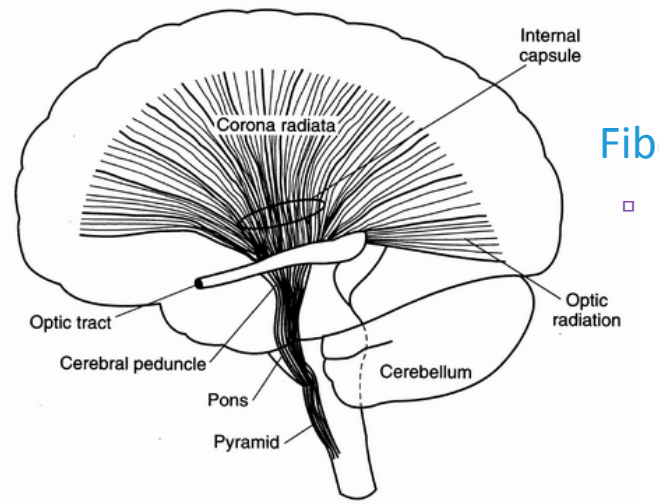


Grey Matter (GM) cortex can by followed and cortical thickness can be evaluated to investigate GM involvement in pathological conditions.

Diffusion weighted imaging

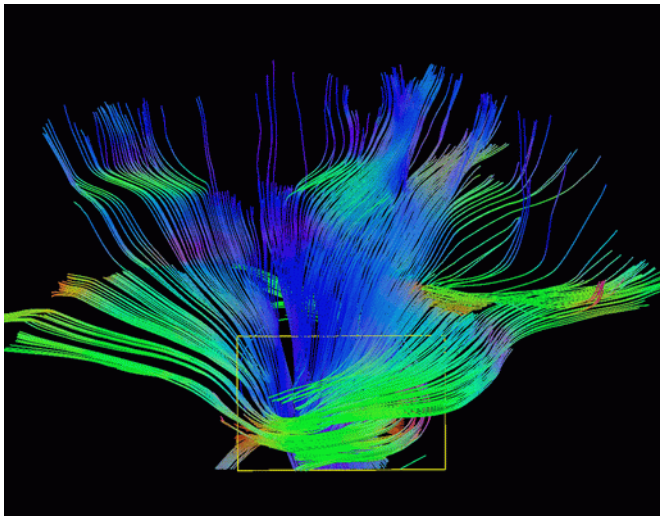
- Water molecules diffuse around during the imaging readout window
- Diffusive movement of water in brain is not isotropic
  - In white matter (WM) diffusion along axonal fiber orientation is faster
- Imaging can be made sensitive to this diffusive motion





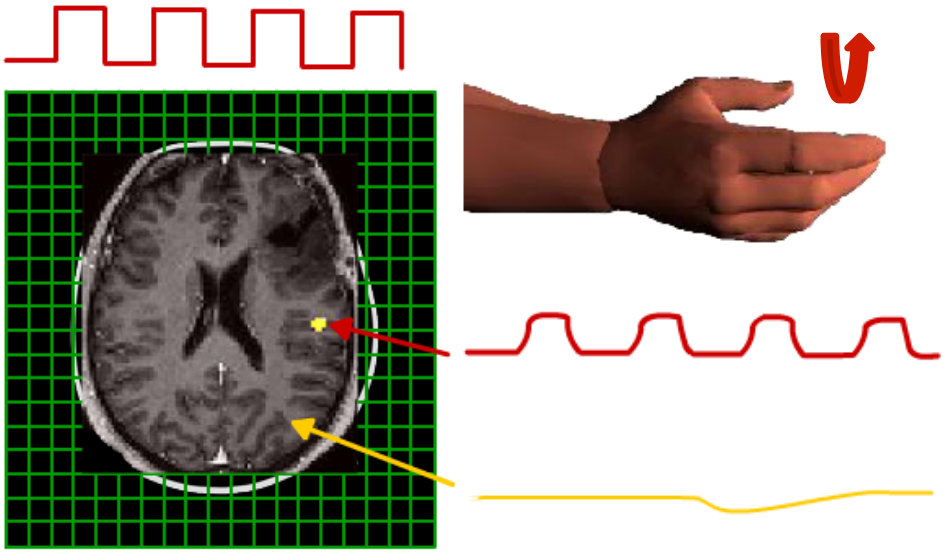
Fiber tracking:

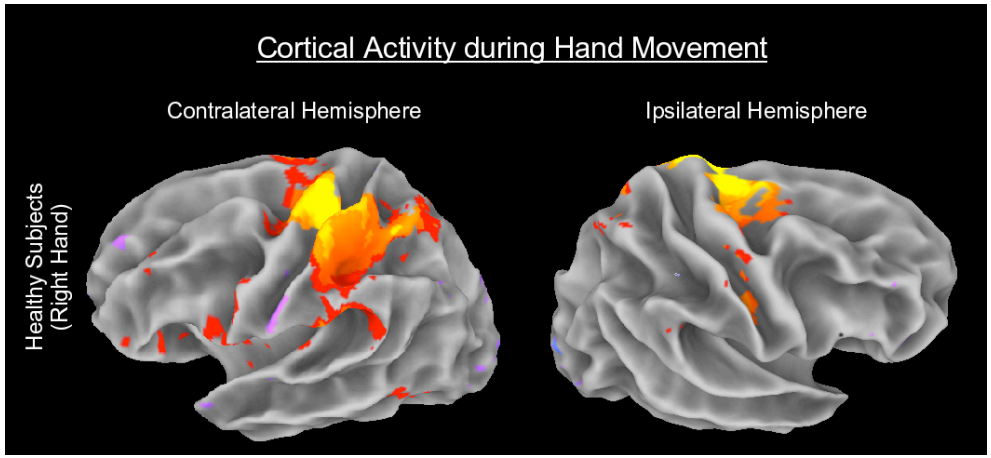
- The orientation of axonal fiber is reconstructed *in vivo*



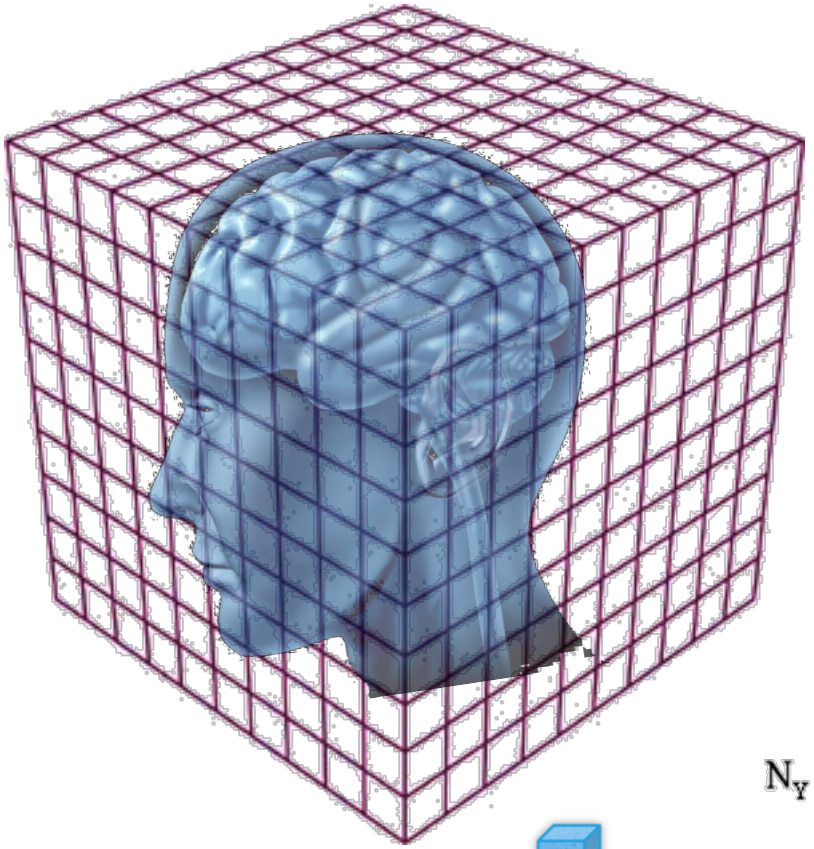
Functional MRI (fMRI)

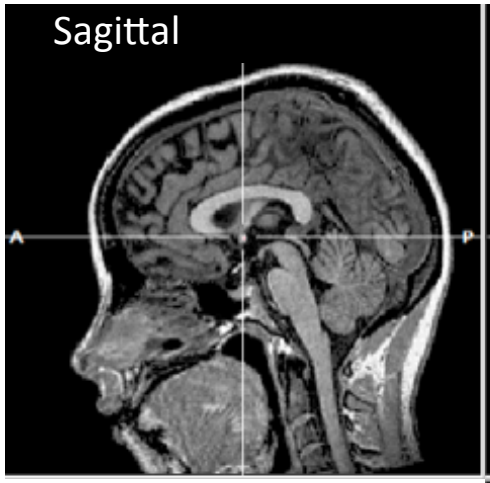
- Same basic theory and technique as structural MRI
- Uses the BOLD Response: blood-oxygen-level-dependent (BOLD) contrast
- Several images are taken over a time period
- Stimuli (visual, auditory, tactile, …) are presented during the scan
- Analysis of data time series to look for up-and-down signals that match the stimulus time series

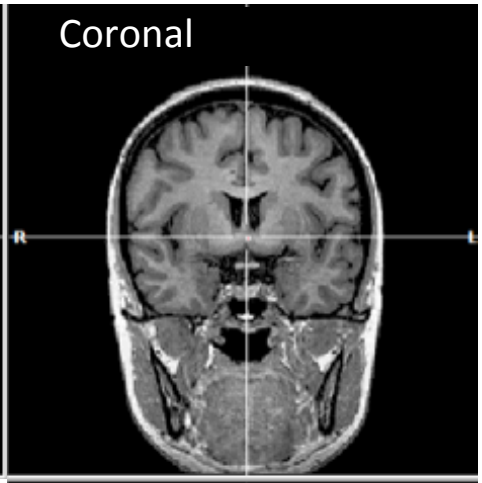


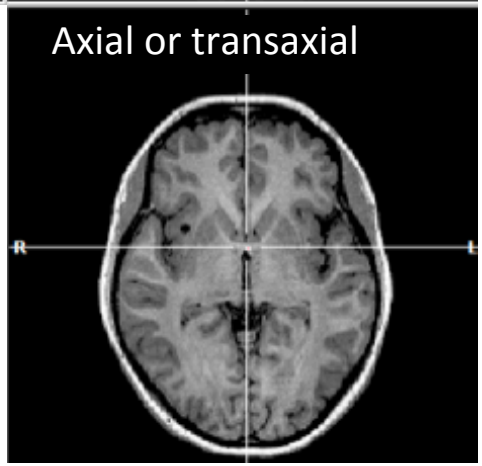


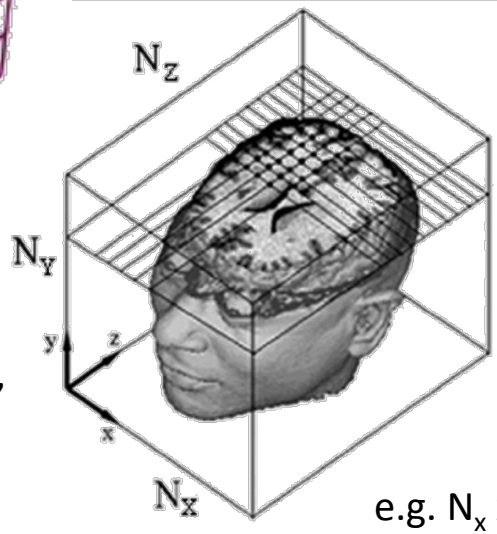
Structural MRI T_1-w images











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

Neuroimaging studies with T_1-w images

- Group comparison design:
  - To reveal systematic differences between the brain of subjects with disease and healthy controls *in vivo*; i.e. to characterize brain morphology alterations due to pathological conditions

Voxel-based morphometry (VBM)

Support Vector Machine (SVM)

  - To make prediction on the pathological condition of previously-unseen subject's data

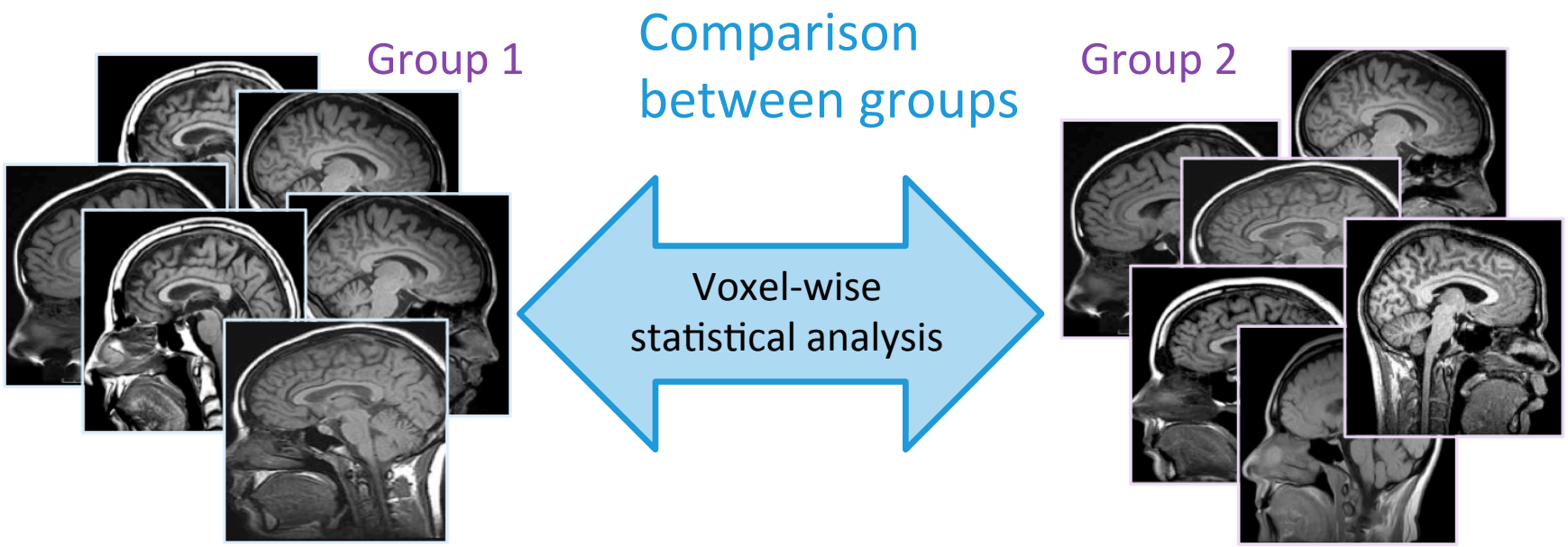
Voxel-based morphometry (VBM) analysis

[Ashburner J. and Friston K.J., Voxel-based morphometry - the methods. Neuroimage 11, 805-821 (2000)]

Wellcome Trust Centre for Neuroimaging
University College London, UK

Statistical Parametric Mapping (SPM) software package
http://www.fil.ion.ucl.ac.uk/spm/

Voxel based morphometry (VBM)



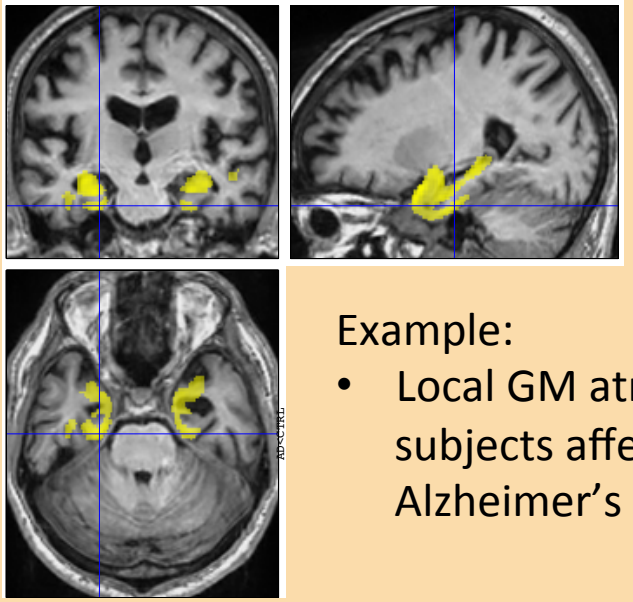
- Patients (group 1) vs. Healthy Controls (group 2):
  - Pathology specific brain alterations
- Longitudinal studies (same group after a time delay)
  - Effect of aging
  - Effect of treatments

VBM basic overview

- VBM preprocessing
  - Alignment
  - Tissue segmentation
  - "Modulation"
  - Gaussian smoothing
- Between-group voxel-wise statistical analysis

VBM output:

- a statistical parametric map (SPM) showing regions where GM (or WM) significantly differs among two groups of subjects.

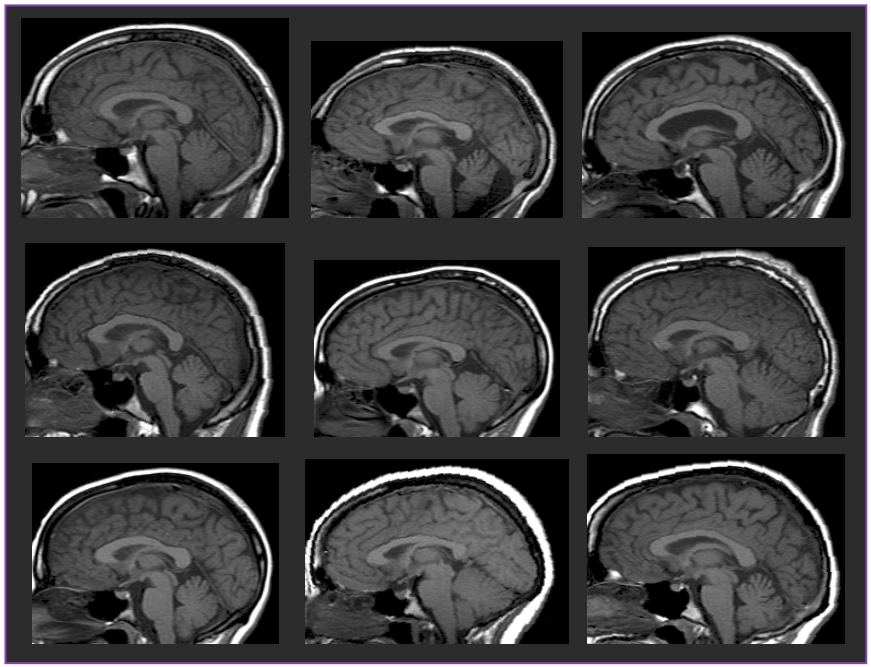


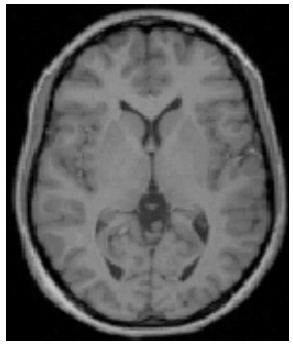
Example:

- Local GM atrophy in subjects affected by Alzheimer's Disease

Voxel-wise analysis

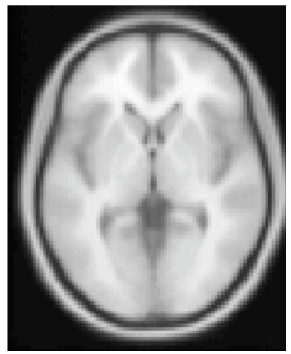
- How can we carry out a voxel-wise comparison of brains of different subjects?
- Images need to be **spatially aligned**





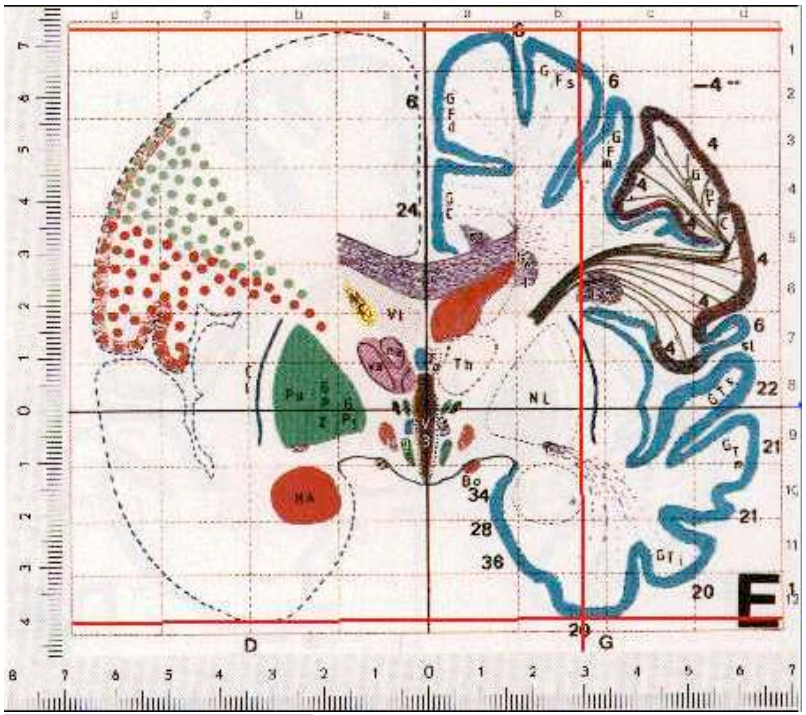
Each image is registered to a T_1 template

12 degree of freedom affine transformation, to match the size and position

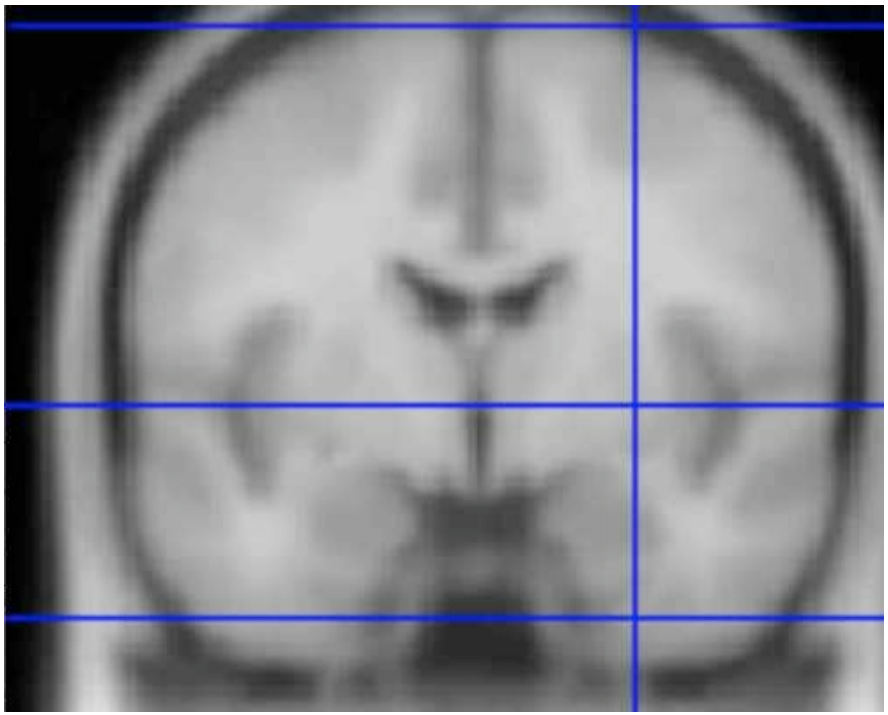


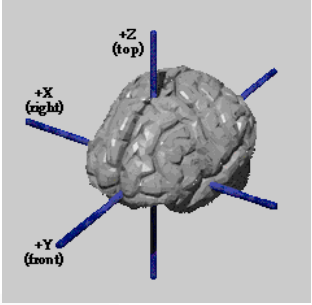
Standard reference space

The Talairach Atlas



The MNI/ICBM AVG152 Template





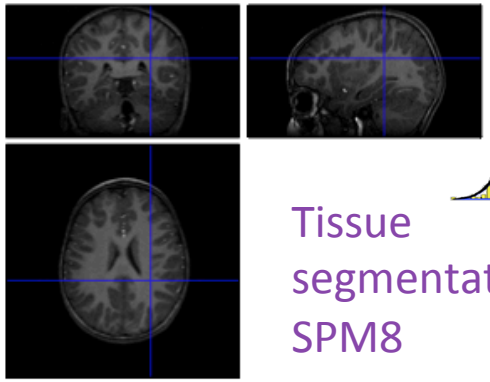
The MNI template follows the *convention* of the Talairach Atlas (Talairach and Tournoux, 1988), but does not match the *particular brain*

http://imaging.mrc-cbu.cam.ac.uk/imaging/MniTalairach

The need for tissue segmentation

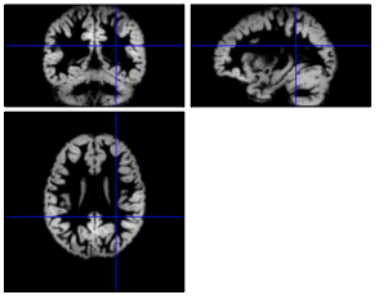
- MRI intensity is usually not quantitatively meaningful (as opposed to e.g. computed tomography images)
- Regional volumes of the three main tissue types: gray matter (GM), white matter (WM) and cerebrospinal fluid (CSF), are well-defined and potentially very interesting

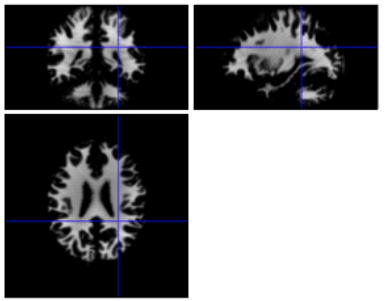
Segmentation example

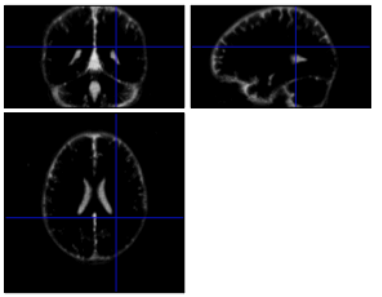


Mixture of Gaussian Model (MOG)

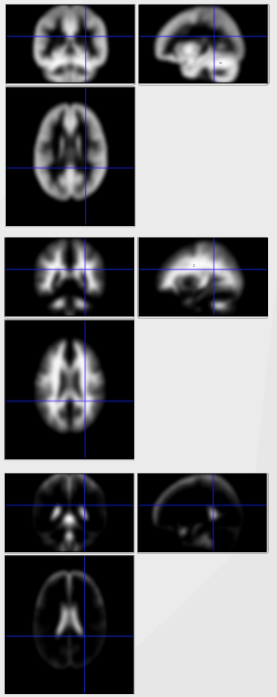
Tissue segmentation in SPM8







Tissue probability maps (TPMs) in SPM8



- In a simple MOG, the probability of obtaining a voxel with intensity y_i given that it belongs to the k^{th} cluster, i.e. Gaussian ($c_i = k$), and that the k^{th} Gaussian is parameterized by μ_k and σ_k^2 is:

$$P(y_i|c_i = k, \mu_k, \sigma_k) = \frac{1}{\sqrt{2\pi\sigma_k^2}} exp\left(-\frac{(y_i - \mu_k)^2}{2\sigma_k^2}\right)$$

- The prior probability of any voxel, irrespective of its intensity, belonging to the k^{th} is:

$$P(c_i = k|\gamma_k) = \gamma_k \qquad \sum_{k=1}^{K} \gamma_k = 1$$

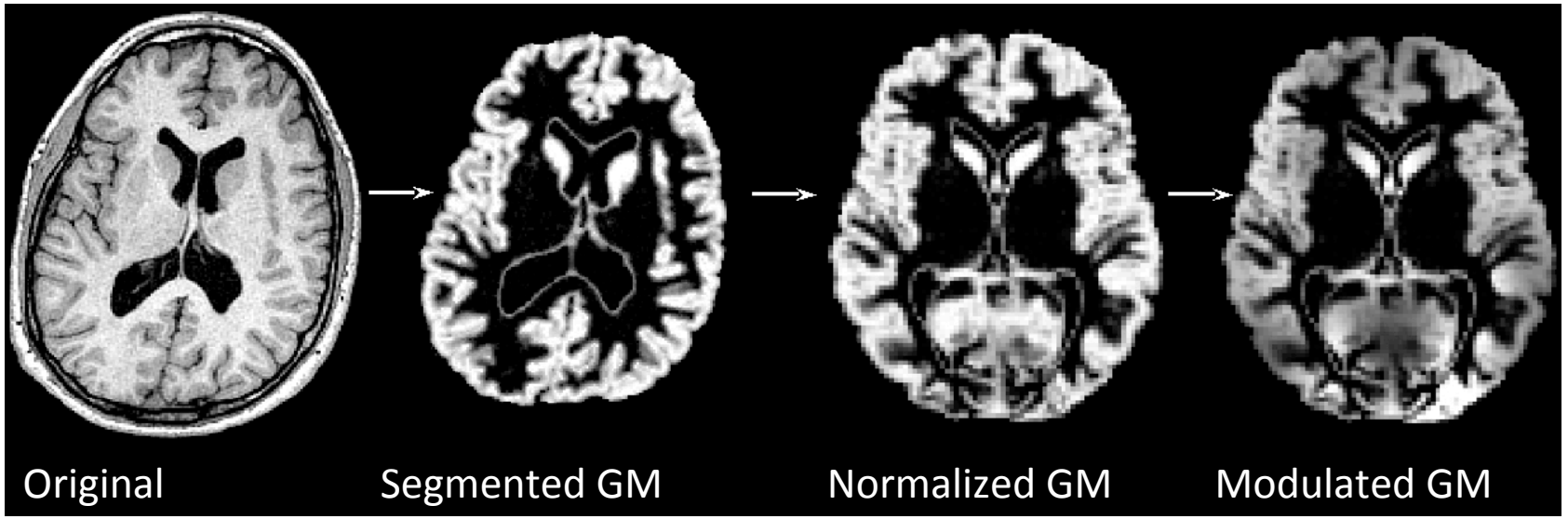
- Using Bayes rule, the joint probability of cluster k and intensity y_i is:

$$P(y_i, c_i = k|\mu_k, \sigma_k, \gamma_k)$$

$$= P(y_i|c_i = k, \mu_k, \sigma_k)P(c_i = k|\gamma_k)$$

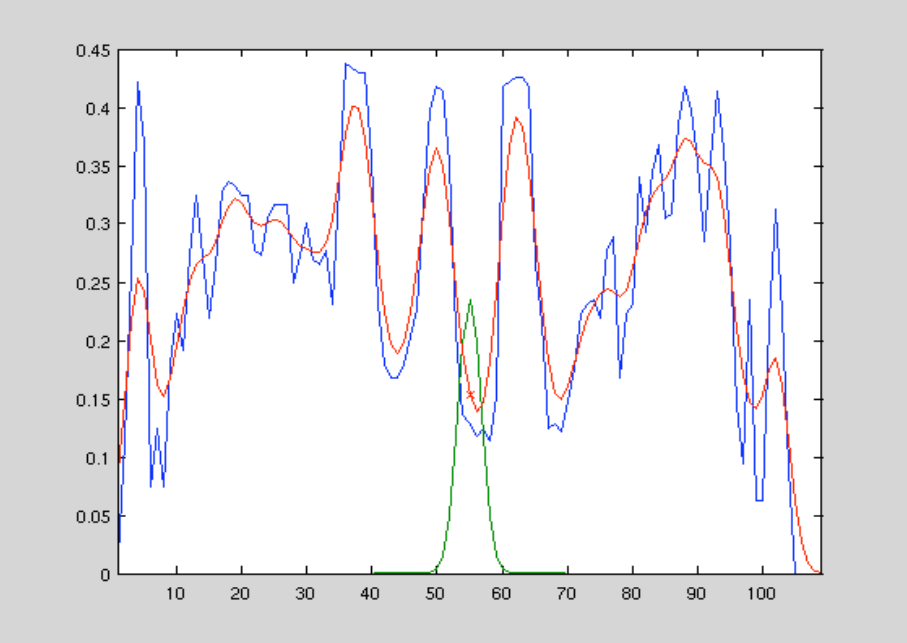
⟹ Sum over K clusters, accounting for each voxel and <u>maximize</u> with respect to the unknowns (μ, σ)

Concentration or volume differences: modulated vs. non-modulated VBM analysis



- When warping an image to match a template the information about the absolute volume of a region will be lost.
- MODULATION compensates for the effects of spatial normalization to preserve the information about absolute volumes.
- To assess volume differences across two categories of subjects MODULATED data should be analyzed.

Example of Gaussian smoothing



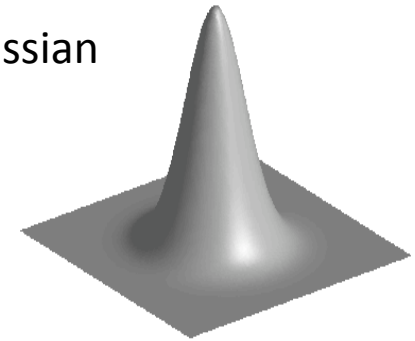
Convolution with an isotropic Gaussian kernel:

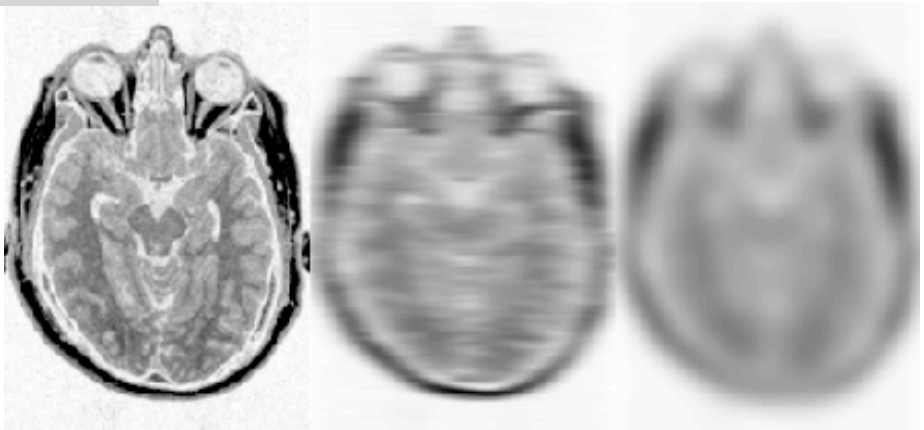
- each voxel becomes weighted average of surrounding voxels

The Gaussian kernel is **separable** we can smooth 2D data with two 1D convolutions.

It can be generalised to 3D

2D Gaussian Kernel

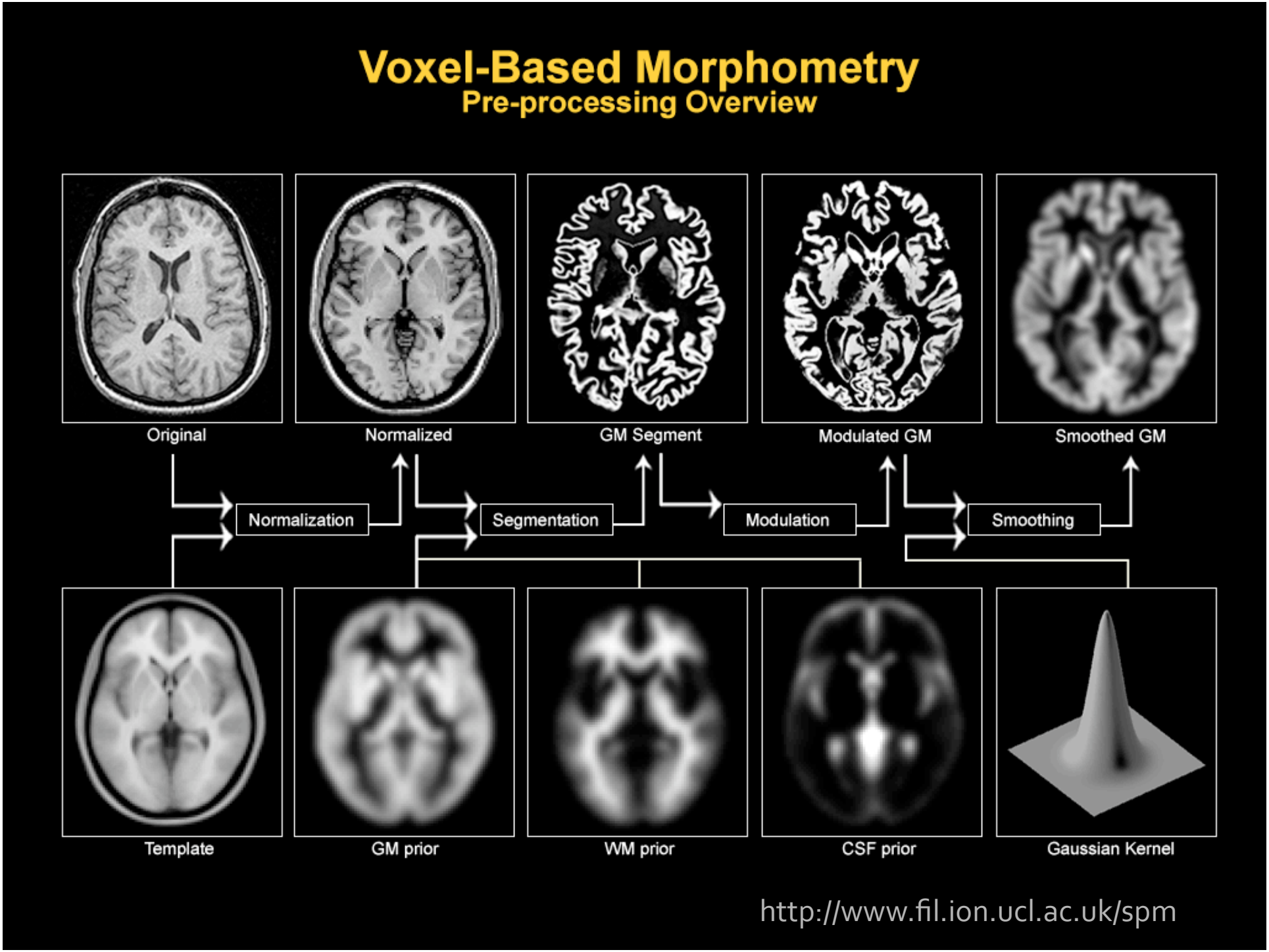




Smoothing

Why would we deliberately blur the data?

- Spatial normalization is never exact, so homologous regions can never be precisely registered
  - Smoothing suppresses noise and effects due to differences in anatomy by averaging over neighboring voxels and compensates for inaccuracies in normalization
- Smoothing renders the data more normally distributed (Central Limit theorem)
  - Required if using parametric statistics
- Filter size should match the expected effect size
  - Usually between 6 – 12 mm


Voxel-Based Morphometry
Pre-processing Overview
Original
Normalized
GM Segment
Modulated GM
Smoothed GM
Normalization
Segmentation
Modulation
Smoothing
Template
GM prior
WM prior
CSF prior
Gaussian Kernel
http://www.fil.ion.ucl.ac.uk/spm

VBM analysis of GM segments

- Voxel-wise analysis (two sample t-test)
- COVARIATES: Global measures can be modeled as confounding effects during the statistical analysis.
  - For example, when investigating differences in GM or WM volumes between populations, the total amount of GM or WM or the total intracranial volume should be included as a covariate of no interest.
- Usually, age, gender and Total Intracranial Volume (TIV), estimated as the sum of tissue segments (GM+WM+CSF), are considered as covariates.

VBM: group comparison

THE GENERAL LINEAR MODEL

- **Intensity** for each voxel (V) is a function that models the different things that account for differences between scans:

$$Y_j = x_{j1}\beta_1 + \ldots + x_{jl}\beta_l + \ldots + x_{jL}\beta_L + \epsilon_j$$

- the response variable Y_j as a linear combination of **L** explanatory variables

- In practice, the contrast of interest is usually t-test between β_1 and β_2

- e.g. "is there significantly more GM in the control than in the patients' scans?"

The Alzheimer's disease case study

VBM analysis:

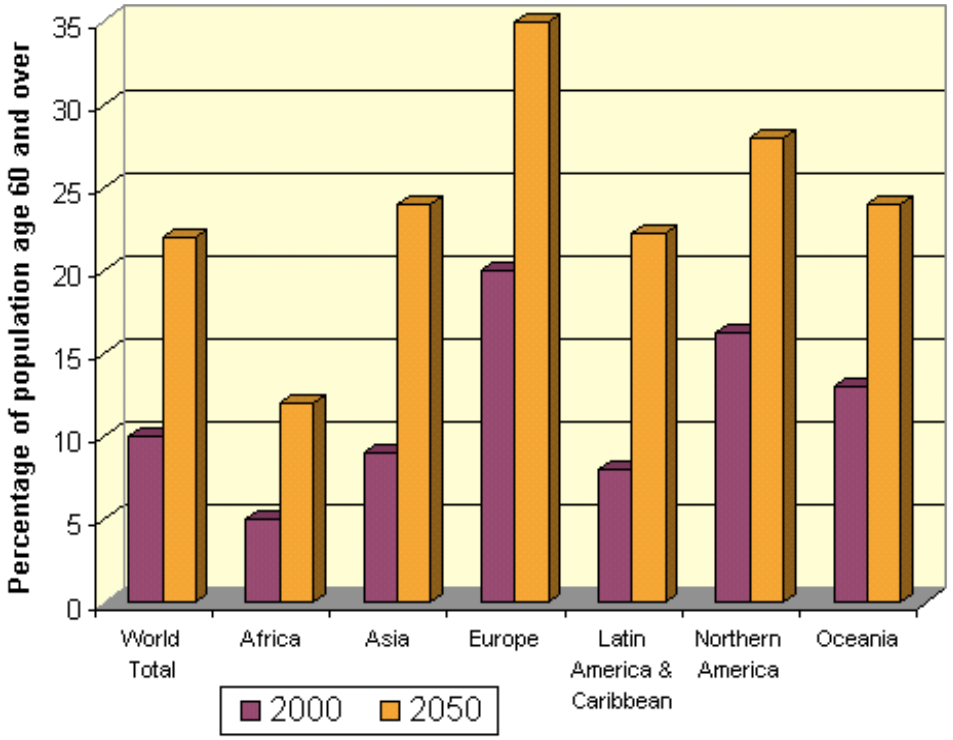
- results
- limitations

The Alzheimer's disease (AD)

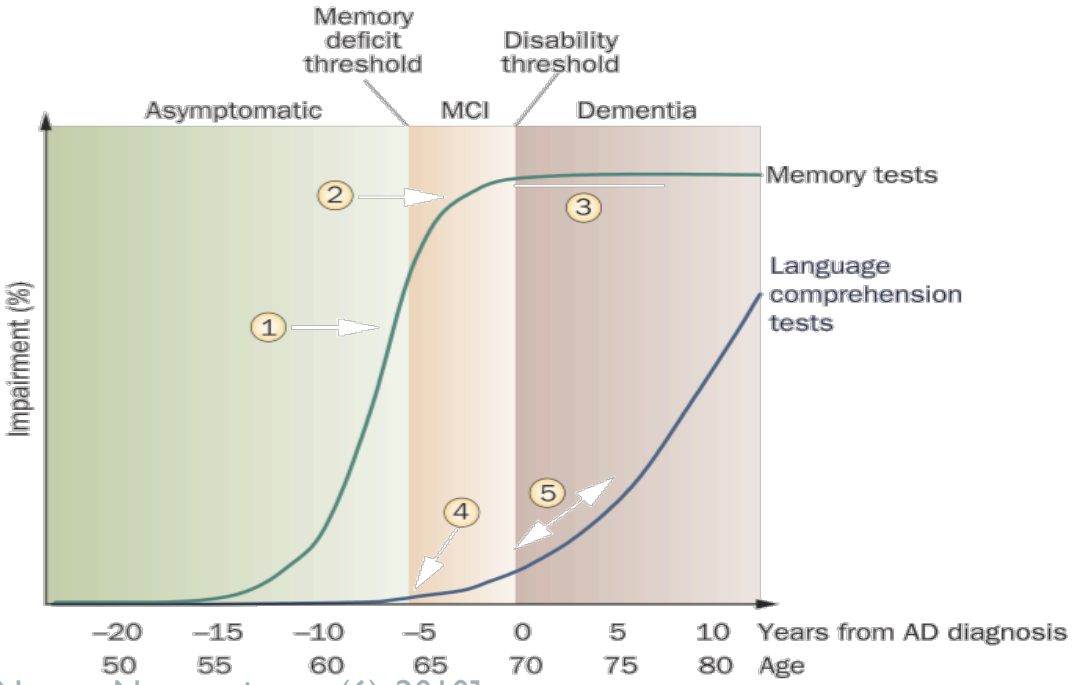
- The Alzheimer's disease (AD) is one of the most disabling and burdensome health conditions worldwide.

  [Lancet, 2005; (366); pp. 2112-17]

- Early and accurate diagnosis of AD is crucial in the perspective of future treatments.



Mild Cognitive Impairment (MCI)



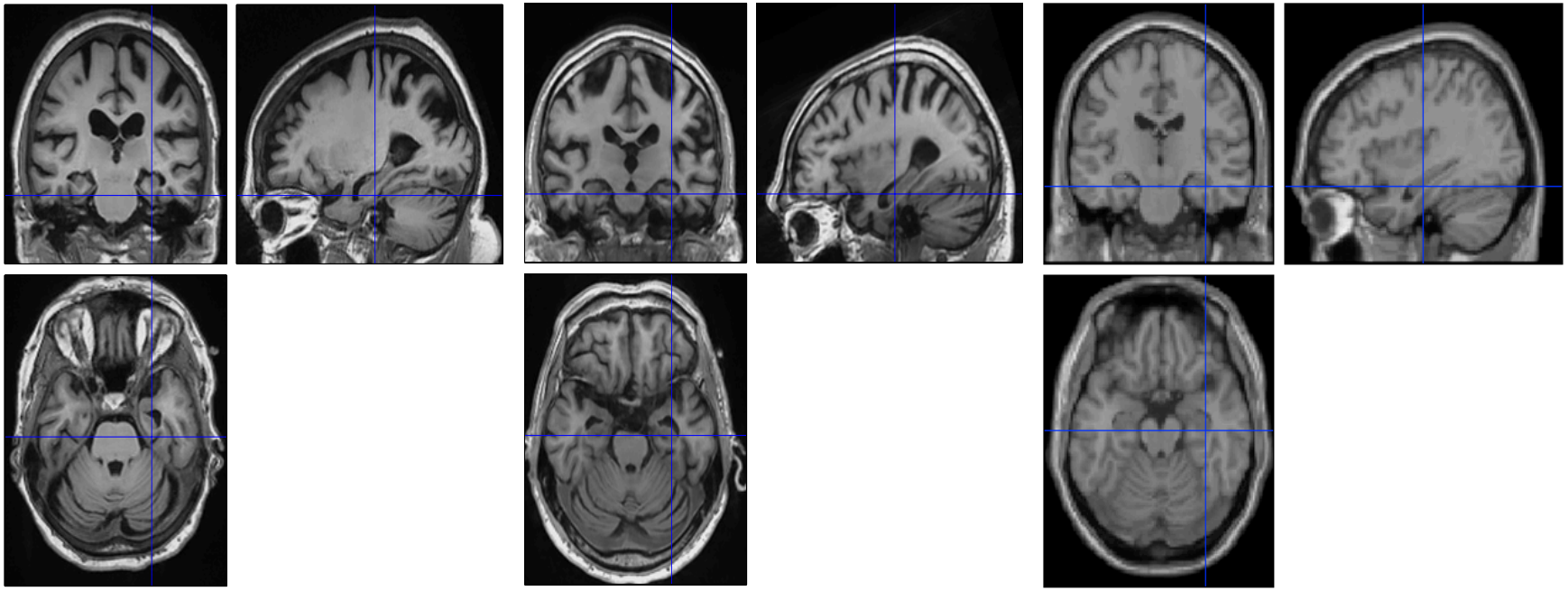
[Nature Neuroscience, (6), 2010]

So far, Medial Temporal Lobe (MTL) atrophy is one of the key biomarkers to detect early neurodegenerative changes in the course of Alzheimer's disease.

- Mild Cognitive Impairment (MCI):
  a transitional stage between normal condition and dementia.
- Part of MCI subjects will develop dementia… but another substantial part will not!
- Subjects in the MCI condition have a yearly rate of 10-15% of AD conversion

  [Petersen RC et al, Arch Neurol 66, 1447-55 (2009)].

Example of MTL atrophy visibility on MRI structural data



AD

CTRL

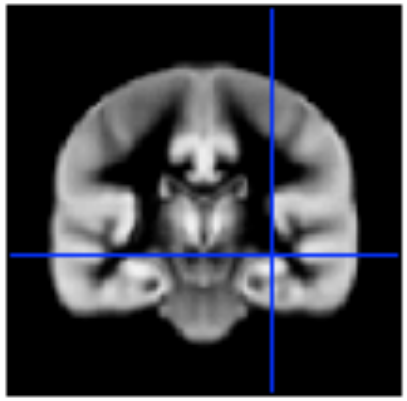
Healthy younger subject

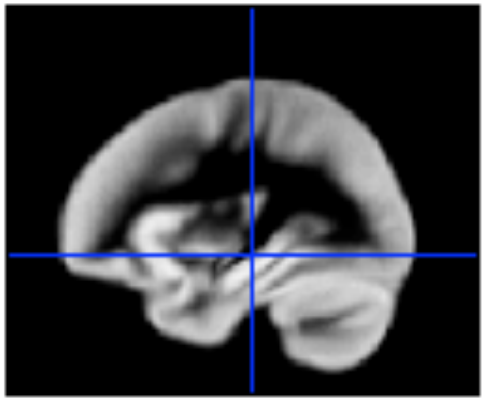
DARTEL

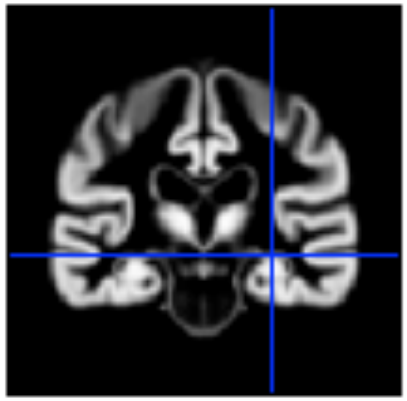
- When analyzing older or younger subjects it's better to build a population-based template.
- The ICBM452 atlas supplied with SPM is an average of T_1-weighted MRIs of normal young adult brains.
- DARTEL is a tool for achieving more accurate inter-subject registration of brain images.
  - *Diffeomorphic Anatomical Registration Through Exponentiated Lie Algebra*

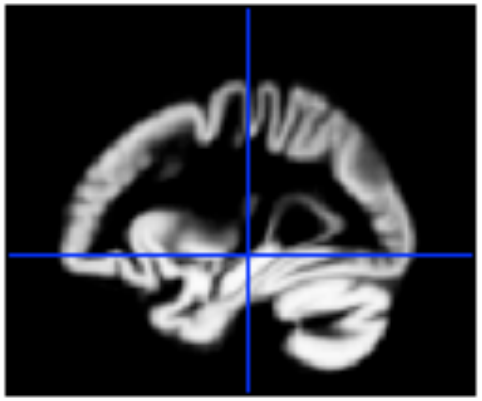
[J. Ashburner. A Fast Diffeomorphic Image Registration Algorithm. NeuroImage, 38(1):95-113, 2007]

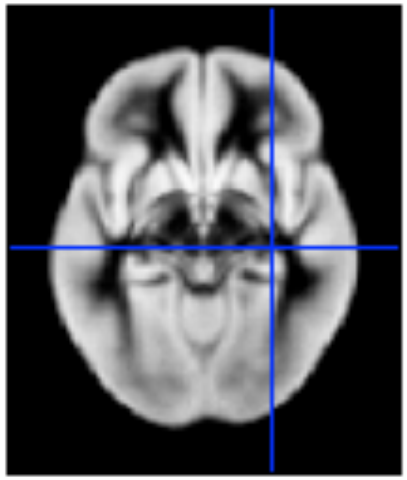
SPM GM tissue probability map vs. DARTEL-generated GM template



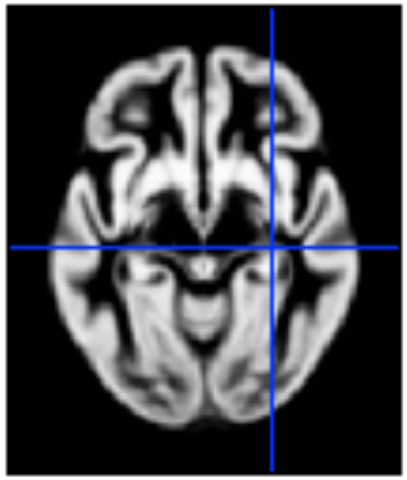








SPM GM tpm



DARTEL generated GM tpm

Dataset used in this study

- The Alzheimer's Disease Neuroimaging Initiative (ADNI)

The mission of the ADNI is to define the progression of Alzheimer's Disease. A major goal of the ADNI study has been to collect and validate data such as MRI and PET images, cerebral spinal fluid, and blood biomarkers as predictors of the disease. Data from Alzheimer's Disease patients, Mild Cognitive Impairment subjects, and elderly controls are available to download.

- Dataset used in this study:

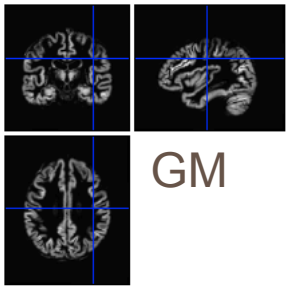
Cohort	Sample size	Age	Gender (M/F)	MMSE
CTRL	189	76.6±5.1	95/94	29.1±0.9
MCI-NC	166	75.7±7.3	106/60	27.2±2.4
MCI-C	136	75.1±7.1	80/56	25.2±2.7
AD	144	75.5±7.5	78/66	22.3±3.3
TOT	635			

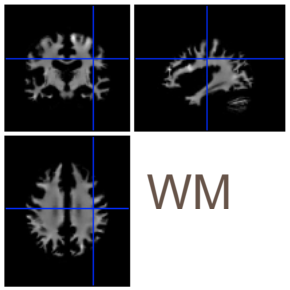
- Train & test: 144 AD + 189 age-matched Controls (CTRL)
- Trial on 302 MCI subjects: 136 converters (MCI-C) + 166 non converters (MCI-NC)

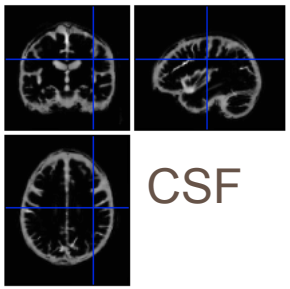
- This dataset was selected with the requirement of having at least 2-year information in addition to the baseline scan:
  - the AD/CTRL subjects were confirmed to be healthy controls/AD at the follow-up assessment;
  - MCI-C subjects converted to AD in a time-frame of 2 years from the baseline scan.

Voxel-Based Morphometry (VBM) analysis with SPM8

The VBM whole-brain analysis provides statistical parametric maps (SPM) of regions where GM significantly differs among groups of subjects

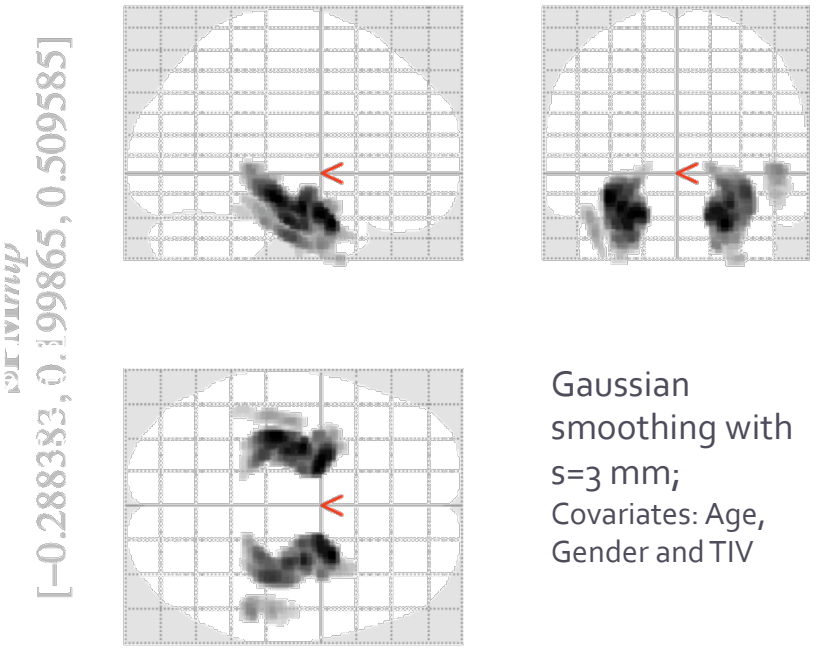

GM


WM


CSF

- VBM preprocessing:
  - SPM tissue segmentation of GM, WM and CSF
  - DARTEL registration and study specific template
  - Alignment to MNI space
- VBM analysis (voxel-wise t-test) to identify the regions where AD and CTRL differ (Gaussian smoothing with s=3mm; i>0.1; covariates: Age, Gender and TIV)
  - The amount of GM in AD is significantly lower than in CTRL, especially in the Temporal and Limbic Lobes

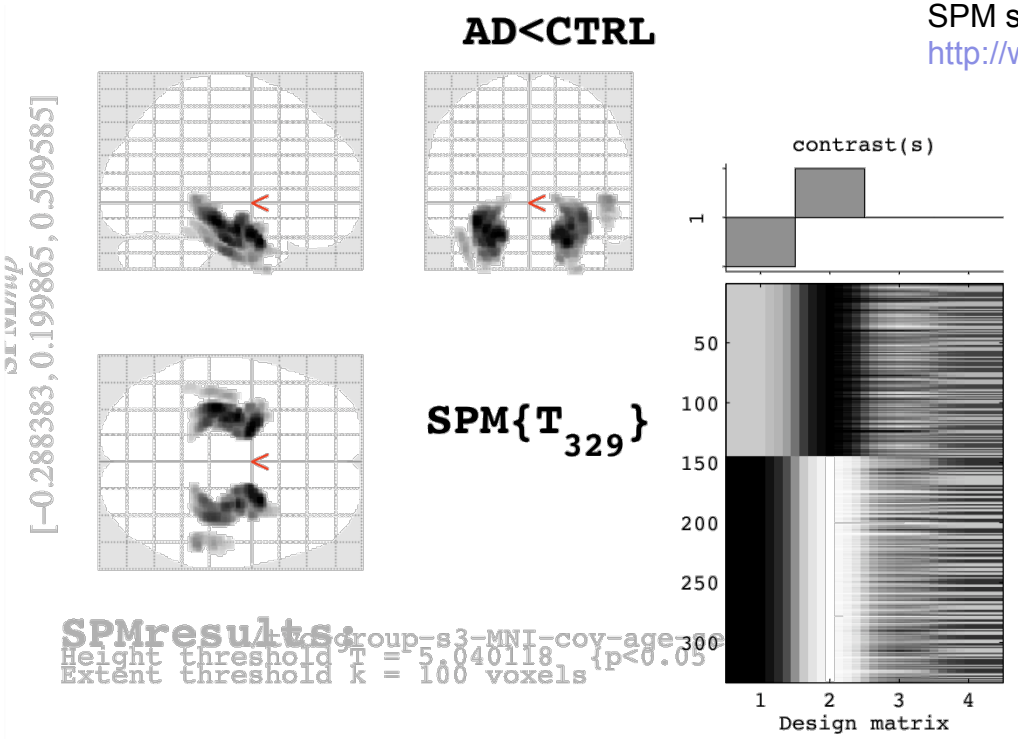
AD<CTRL



Gaussian smoothing with s=3 mm;
Covariates: Age, Gender and TIV

VBM results: Structural abnormalities in AD subjects vs. Controls

SPM software package
http://www.fil.ion.ucl.ac.uk/spm/

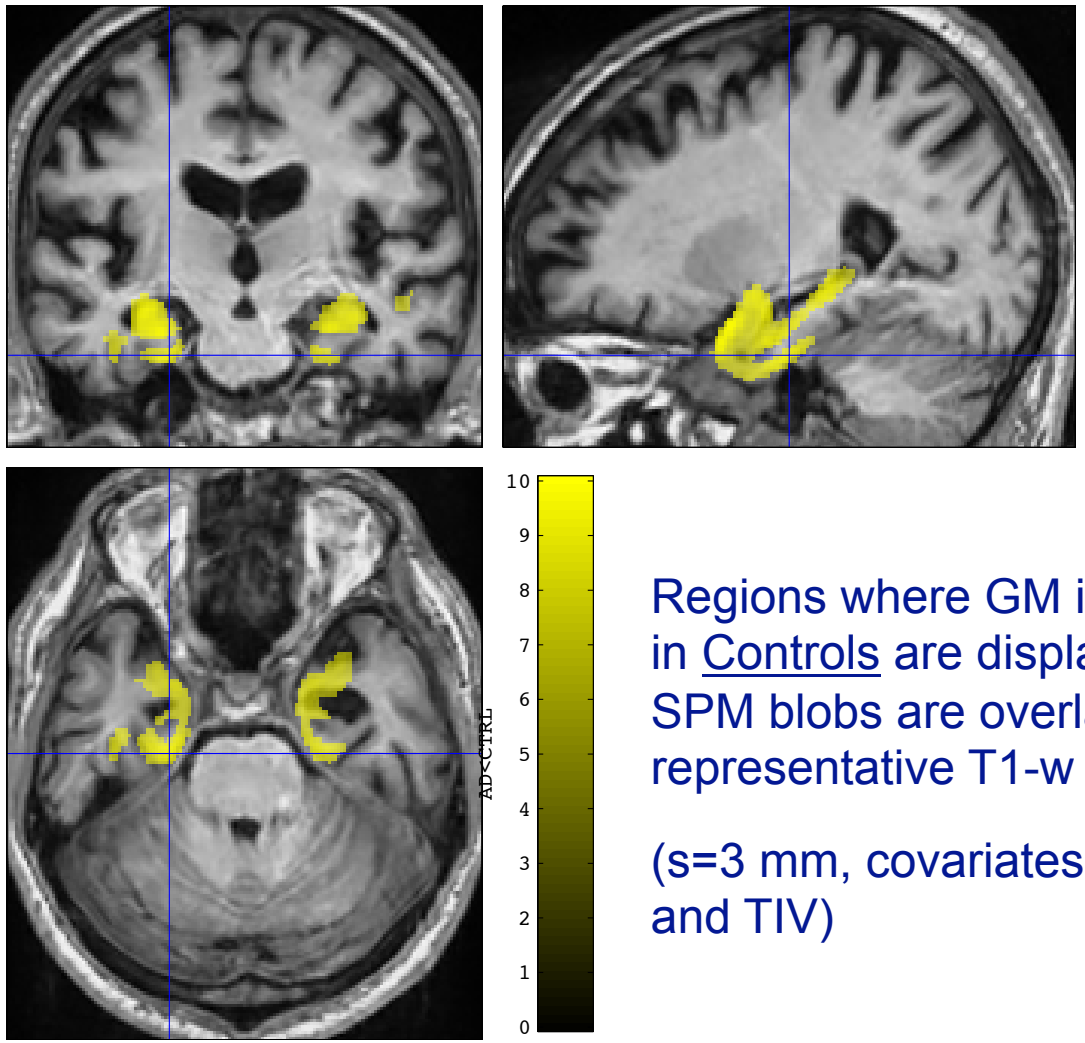


Statistics: *p-values adjusted for search volume*

set-level		cluster-level				peak-level					mm	mm	mm
p	c	$p_{FWE-corr}$	$q_{FDR-corr}$	k_E	p_{uncorr}	$p_{FWE-corr}$	$q_{FDR-corr}$	T	($Z_=$)	p_{uncorr}			
0.0004		**0.000**	**0.000**	**3835**	**0.000**	**0.000**	**0.000**	**10.02**	**Inf**	**0.000**	**-35**	**-24**	**-12**
						0.000	0.000	9.87	Inf	0.000	-21	-0	-20
						0.000	0.000	9.76	Inf	0.000	-29	-9	-17
		0.000	**0.000**	**3690**	**0.000**	**0.000**	**0.000**	**10.02**	**Inf**	**0.000**	**25**	**6**	**-25**
						0.000	0.000	9.47	Inf	0.000	21	0	-18
						0.000	0.000	9.29	Inf	0.000	27	-9	-15
		0.000	**0.000**	**426**	**0.000**	**0.000**	**0.000**	**6.96**	**6.72**	**0.000**	**52**	**-35**	**1**
						0.000	0.002	6.47	6.27	0.000	52	-25	-4
		0.000	**0.000**	**332**	**0.000**	**0.000**	**0.002**	**6.50**	**6.30**	**0.000**	**-42**	**-15**	**-30**
						0.000	0.003	6.40	6.21	0.000	-45	-27	-21
						0.006	0.140	5.53	5.41	0.000	-48	-39	-20

MNI coordinates

VBM results: Structural abnormalities in AD subjects vs. Controls

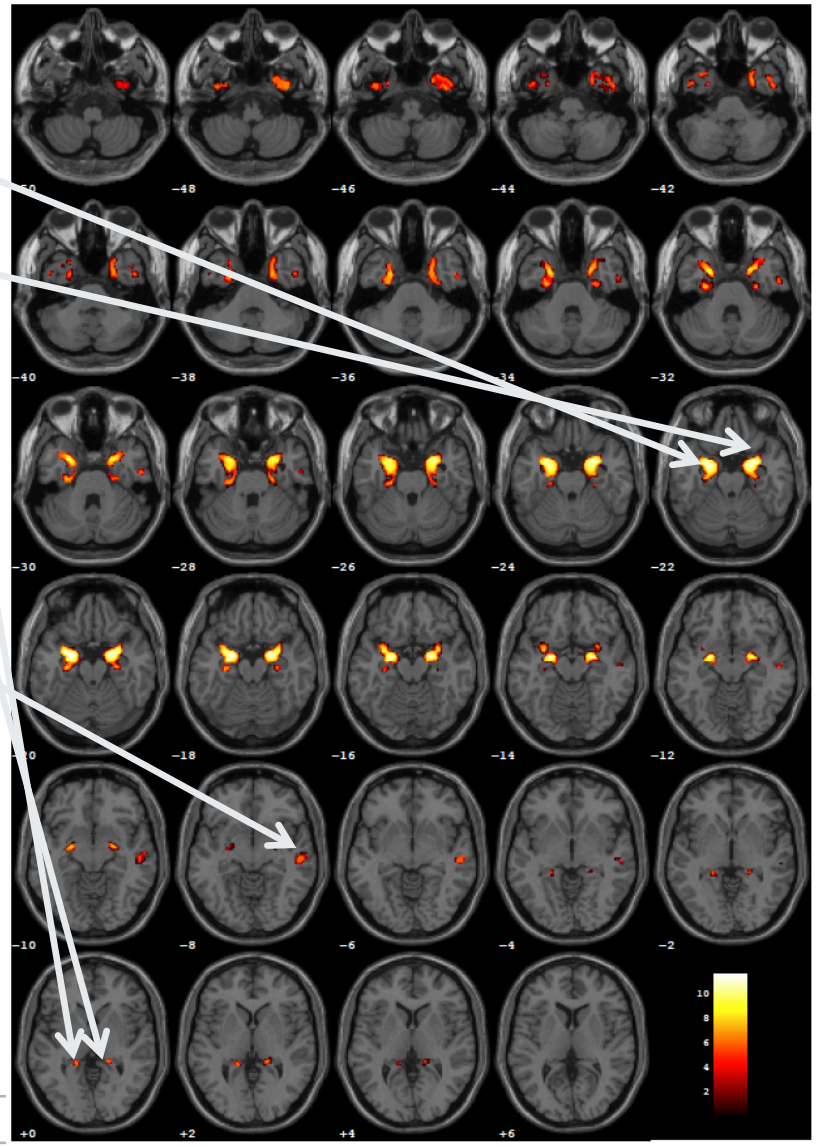


Regions where GM in AD < than GM in Controls are displayed: significant SPM blobs are overlaid to a representative T1-w image

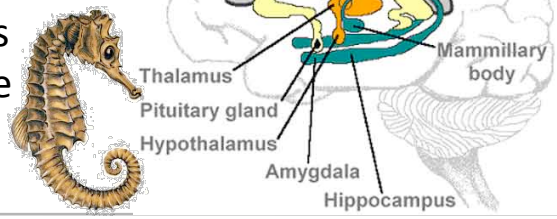
(s=3 mm, covariates: age, gender and TIV)

Voxel-Based Morphometry (VBM) analysis

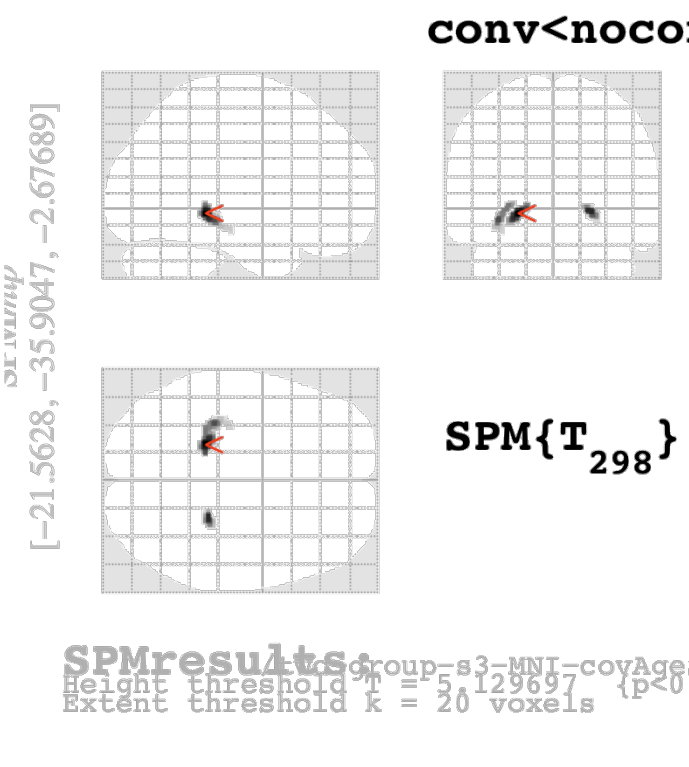
# voxels	MNI coordinates			Talairach coordinates			ROIs		
2646	-22	-6	-21	-21	-5	-16	LC	Limbic Lobe	Parahippocampal Gyrus (Amygdala)
2892	21	-4	-18	19	-4	-12	RC	Limbic Lobe	Parahippocampal Gyrus (Amygdala)
62	-16	-39	1	-16	-38	1	LC	Limbic Lobe	Parahippocampal Gyrus
57	18	-36	1	16	-35	2	RC	Limbic Lobe	Parahippocampal Gyrus
158	56	-19	-8	51	-19	-4	RC	Temporal Lobe	Superior Temporal Gyrus

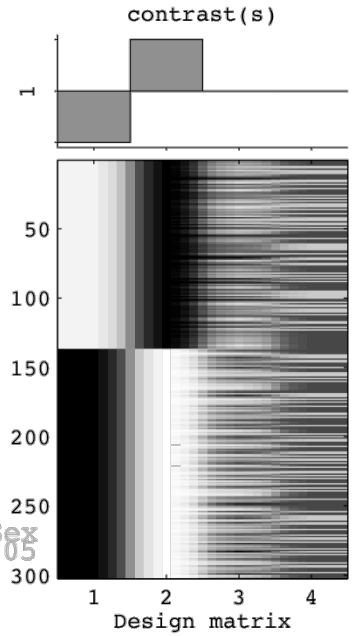


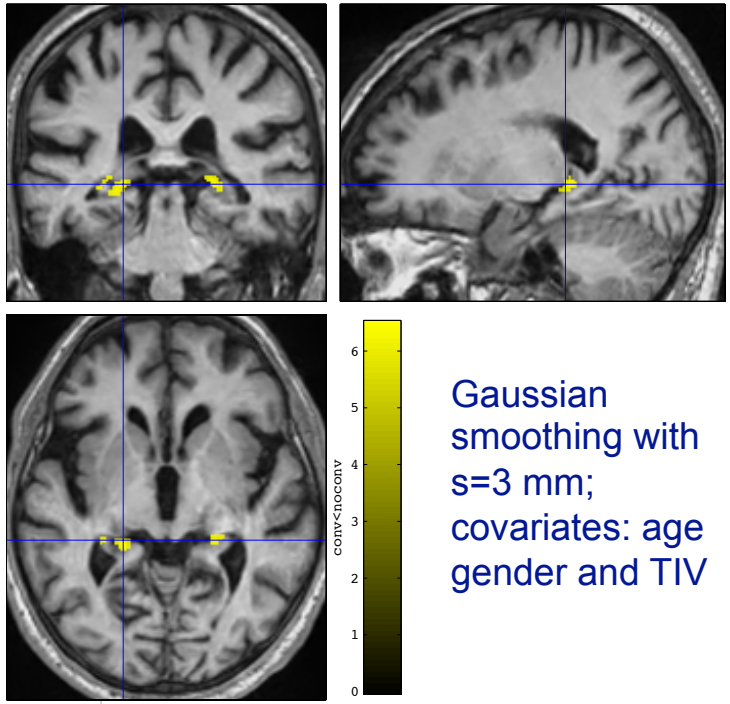
In AD, the hippocampus is one of the first regions of the brain to suffer damage; memory loss and disorientation are included among the early AD symptoms.



VBM results: Structural abnormalities in MCI converters vs. non converters







Gaussian smoothing with s=3 mm; covariates: age gender and TIV

Statistics: ***p-values adjusted for search volume***

set-level		cluster-level				peak-level					mm	mm	mm
p	c	$p_{FWE-corr}$	$q_{FDR-corr}$	k_E	p_{uncorr}	$p_{FWE-corr}$	$q_{FDR-corr}$	T	($Z_=$)	p_{uncorr}			
0.0003		0.000	0.000	96	0.000	0.000	0.029	6.49	6.27	0.000	−22	−36	−3
		0.000	0.000	47	0.000	0.000	0.040	6.22	6.03	0.000	22	−34	−1
		0.000	0.000	64	0.000	0.001	0.086	5.97	5.79	0.000	−34	−30	−7
						0.003	0.198	5.73	5.58	0.000	−28	−36	2
						0.028	0.634	5.26	5.14	0.000	−34	−21	−12

Regions where GM in MCI converters < than GM in MCI non converters are displayed: significant SPM blobs are overlaid to a representative T1-w image

Limitation of the VBM approach

- VBM allows localizing brain regions where GM (WM) significantly differs among groups of subjects.

- VBM limitations:
  - VBM allows for comparisons between groups of subjects, not for single subject analysis.
  - The inference on new cases is not possible.

Support Vector Machine (SVM) analysis

[Vapnik, V.N., The Nature of Statistical Learning Theory. New York: Springer (1995)]

Why using machine-learning techniques?

- Some tasks cannot be defined well, except by examples.
- Relationships and correlations can be hidden within large amounts of data.
  - Machine Learning/Data Mining may be able to find these relationships.
- The amount of knowledge available about certain tasks might be too large for explicit encoding by humans (e.g., medical diagnostic).
  - Learning from examples

Vocabulary:

- Training set: a subset of items, with known classification
- Training: allowing the algorithm to set weights and "learn" how to classify your data, using the training set
- This produces classifier, i.e. a quantitative mapping from features to class.
- Classification: using these training-set weights to classify data of previously unknown category

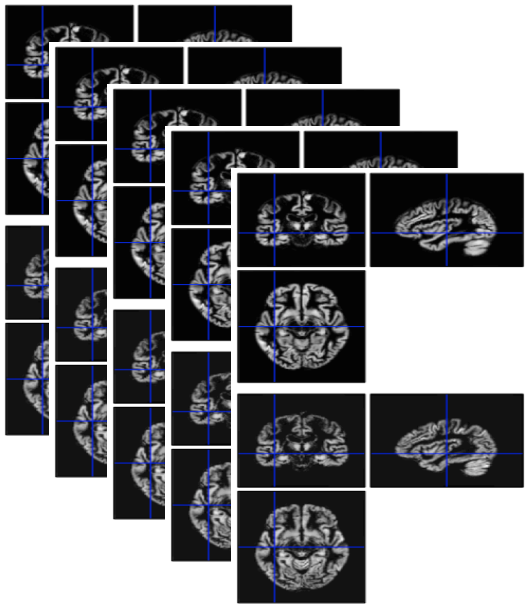
Pattern classification in MRI data of Alzheimer's disease

- Automatic classification of subjects into the Alzheimer's Disease (AD) and the Healthy Control (HC) categories

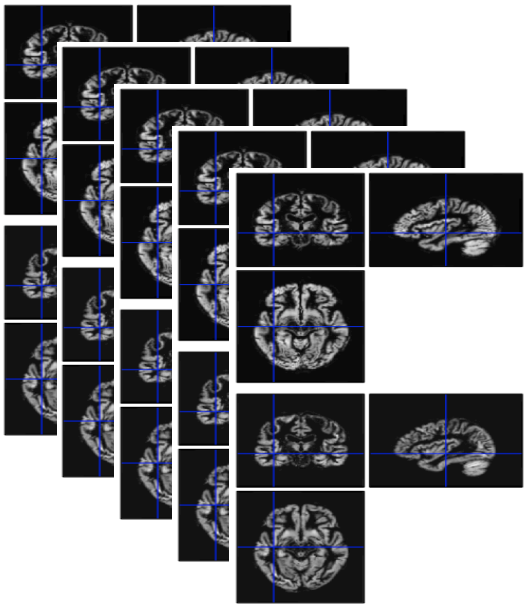
- The predictive value of structural MRI can be investigated to:
  - classify newly acquired subjects;
  - infer on the outcome of Mild Cognitive Impairment (MCI) subjects.

Support Vector Machine (SVM) classification

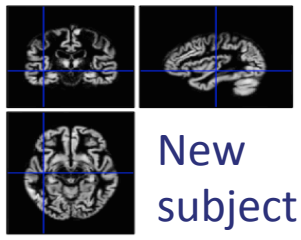
AD subjects



Healthy Controls



Training on known cases



New subject

Trained SVM

SVM OUTPUT

Prediction on MCI outcome

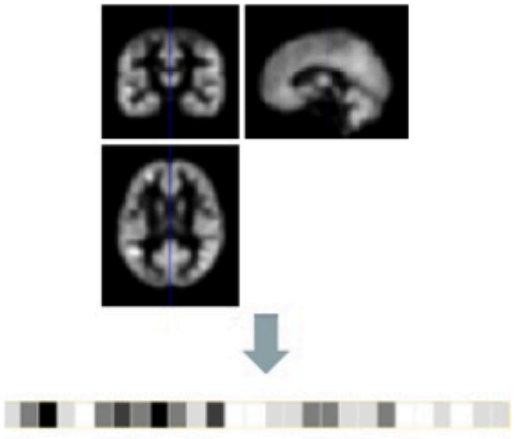
Pattern classification

- Support Vector Machines (SVM)
  - Linear kernel SVM
- Classifier performance evaluation
  - Train, test and validation sets
  - Cross validation procedure
  - Figures of merit
    - Sensitivity, Specificity and Receiver Operative Characteristic curve (ROC)
    - Area under the ROC curve (AUC)

Support Vector Machine (SVM) classification

- In the machine learning framework, each image $\mathbf{x}_i \in \mathbf{R}^n$ is considered as a point in a n-dimensional space (n is the number of voxels in the image).

VBM preprocessed data: segmented, normalized, modulated and smoothed GM for each subject



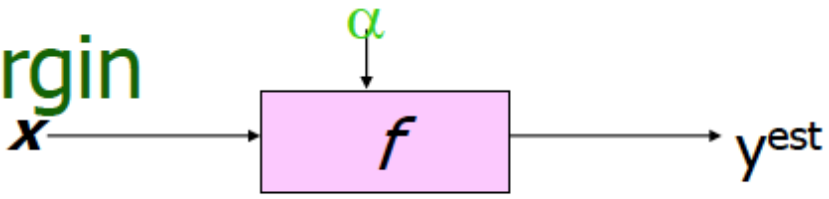
Binary classification

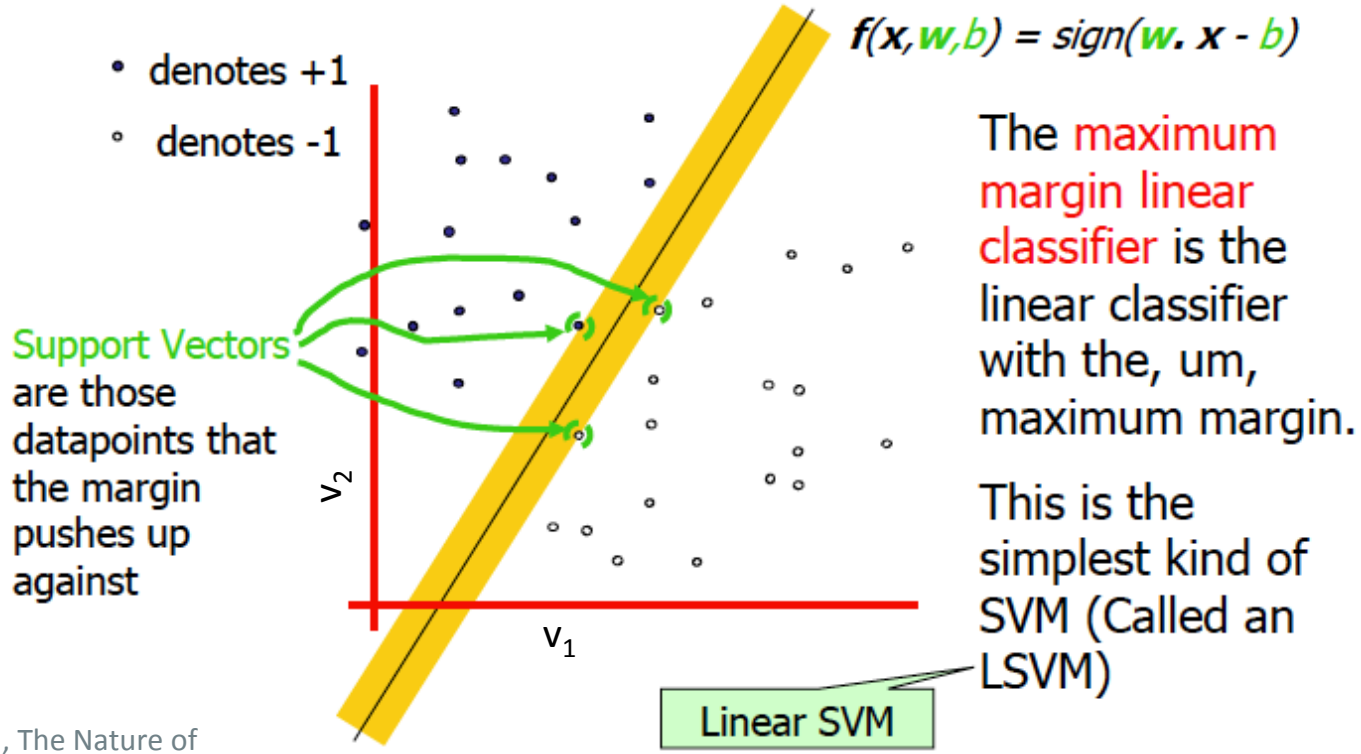
- In a two-class classification (e.g. patients vs. controls) the i-th image can be labelled with y_i:

$$y_i \in \{-1, 1\} \qquad \text{where } i = 1, \ldots, l.$$

Support Vector Machines (SVM)

Maximum Margin





[Vapnik VN, The Nature of Statistical Learning Theory. New York: Springer (1995)]

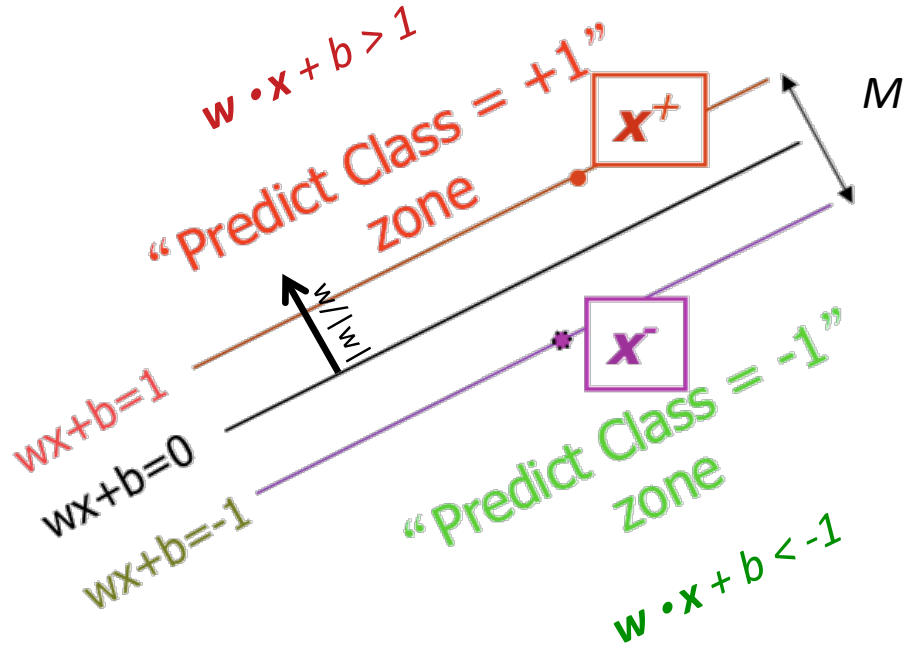
Slides from Andrew Moore's SVM tutorial http://www.cs.cmu.edu/~awm/tutorials

Support Vector Machines (SVM)

Computing the margin width

What we know:

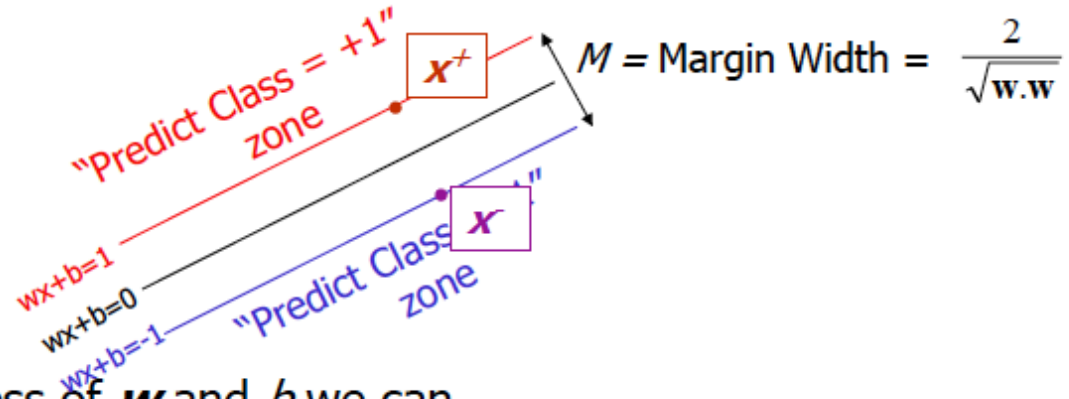
- $\boldsymbol{w} \bullet \boldsymbol{x}^{+} + b = +1$
- $\boldsymbol{w} \bullet \boldsymbol{x}^{-} + b = -1$
- $\boldsymbol{w} \bullet (\boldsymbol{x}^{+} - \boldsymbol{x}^{-}) = 2$



- $M = \boldsymbol{w}/|\boldsymbol{w}| \bullet (\boldsymbol{x}^{+} - \boldsymbol{x}^{-}) = 2/|w|$

Support Vector Machines (SVM)

Learning the Maximum Margin Classifier



Given a guess of ***w*** and *b* we can

- Compute whether all data points in the correct half-planes
- Compute the width of the margin

So now we just need to write a program to search the space of **w**'s and *b*'s to find the widest margin that matches all the datapoints. *How?*

Gradient descent? Simulated Annealing? Matrix Inversion? EM? Newton's Method?

Slides from Andrew Moore's SVM tutorial http://www.cs.cmu.edu/~awm/tutorials

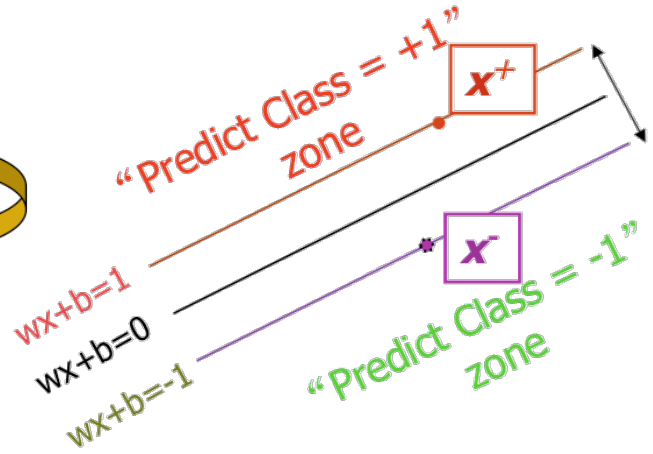
Support Vector Machines (SVM)

- Goal: **1) Correctly classify all training data**

$$wx_i + b \geq 1 \quad \text{if } y_i = +1$$
$$wx_i + b \leq -1 \quad \text{if } y_i = -1$$
$$y_i(wx_i + b) \geq 1 \quad \text{for all i}$$

2) Maximize the Margin $M = \frac{2}{|w|}$

same as minimize $\frac{1}{2} w^t w$



- **We can formulate a Quadratic Optimization Problem and solve for w and b**

Find $\boldsymbol{w}$ and b such that
$\boldsymbol{\Phi(w) = \frac{1}{2}\, w^t w}$ is minimized;
and for all $\{(\boldsymbol{x}_i, y_i)\}$: $y_i\,(\boldsymbol{w}^T \boldsymbol{x}_i + b) \geq 1$

- Quadratic optimization problems are a well-known class of mathematical programming problems, and many (rather intricate) algorithms exist for solving them.
- The solution involves constructing a *dual problem* where a *Lagrange multiplier* α_i is associated with every constraint in the primary problem.

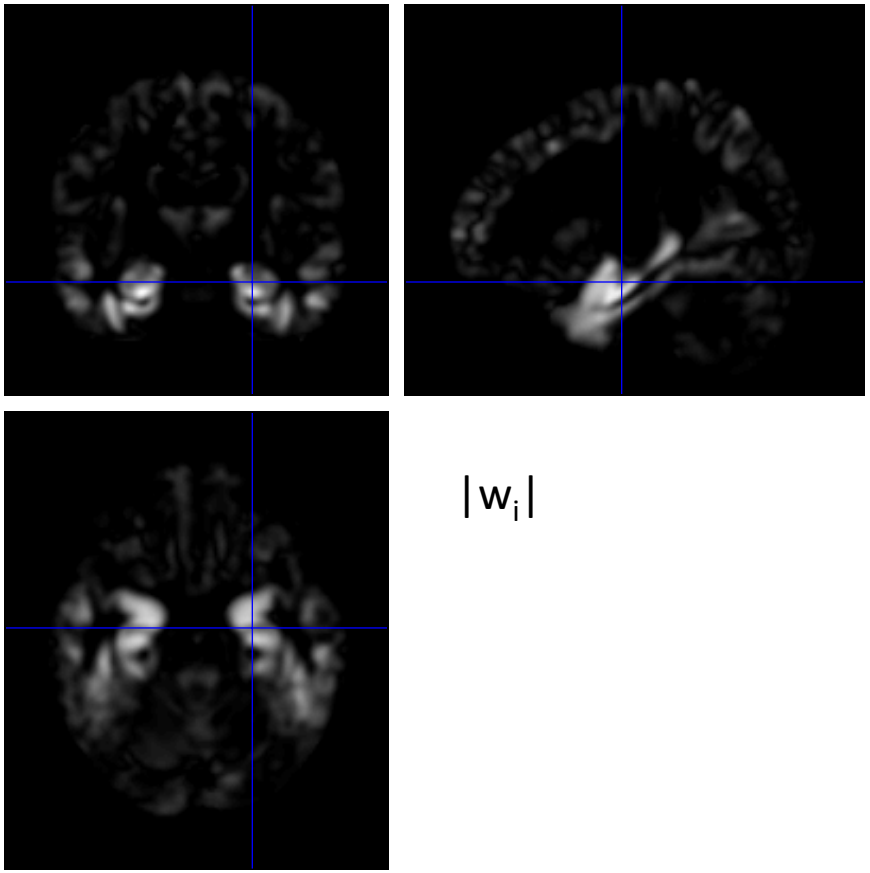
Discrimination maps

- During the SVM training the separating hyperplane is identified so that

$$\underline{w} \bullet \underline{x} + b = 0$$

where $\underline{x}$ is a data pattern, $\underline{w}$ is the weight vector and b is an offset

- $\underline{w}$ can be used to generate a map of the most discriminating voxels/regions

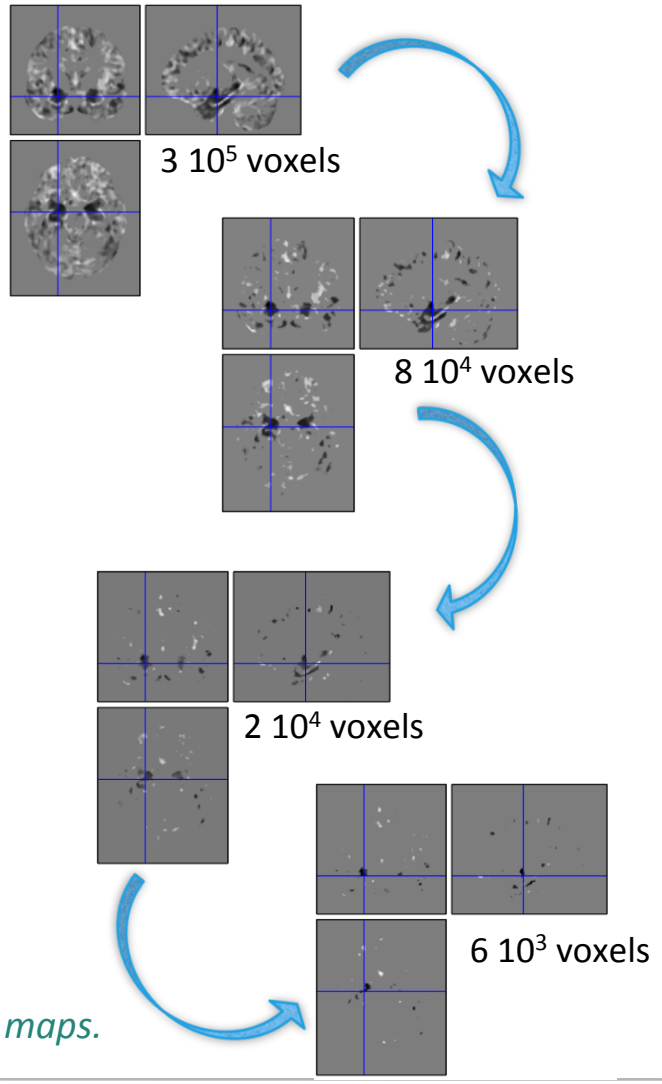


$|w_i|$

Linear-kernel SVMs allow direct extraction of the weight vector as an image.

SVM Recursive Feature Elimination (SVM-RFE)

- Features/voxels are iteratively removed from the data set with the aim of retaining only those voxels that carry discriminative information [Guyon et al, Mach Learn (2002)]
  - The weight magnitude $|w_i|$ is used as ranking criterion:
    - features/voxels with low $|w_i|$ are removed from the dataset,
    - features/voxels with high $|w_i|$ are retained.
  - A new SVM is trained at each iteration.
  - The SVM classification performance (AUC) is evaluated at each iteration (i.e. as a function of the number of retained GM voxels)



The SVM-RFE discrimination maps can be compared to the VBM maps.

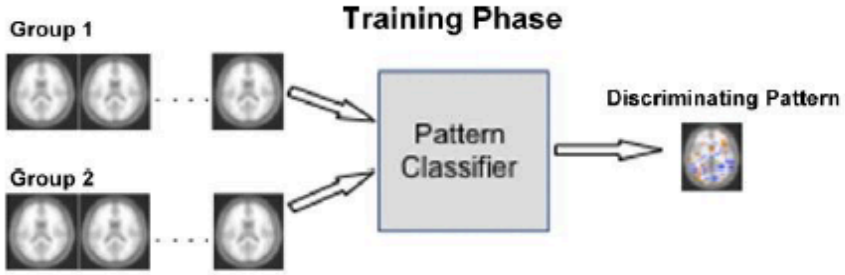
SVM training procedure

The machine learning method consists of two phases:

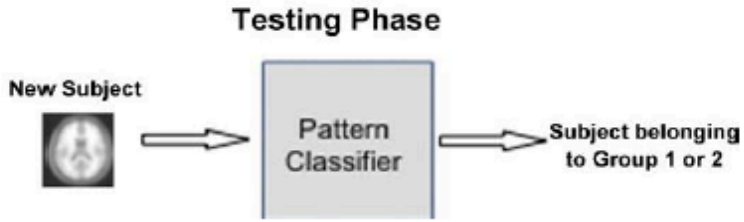
- training phase;
- testing phase.

Hence, the data set is divided in two groups: the training set and the test set.

1. During the training phase, the algorithm estimates a hyperplane that separates the examples contained in the training set according to their labels.



2. The decision function that has been learned from the training data then can be used during the testing phase to predict the class of a new test example.

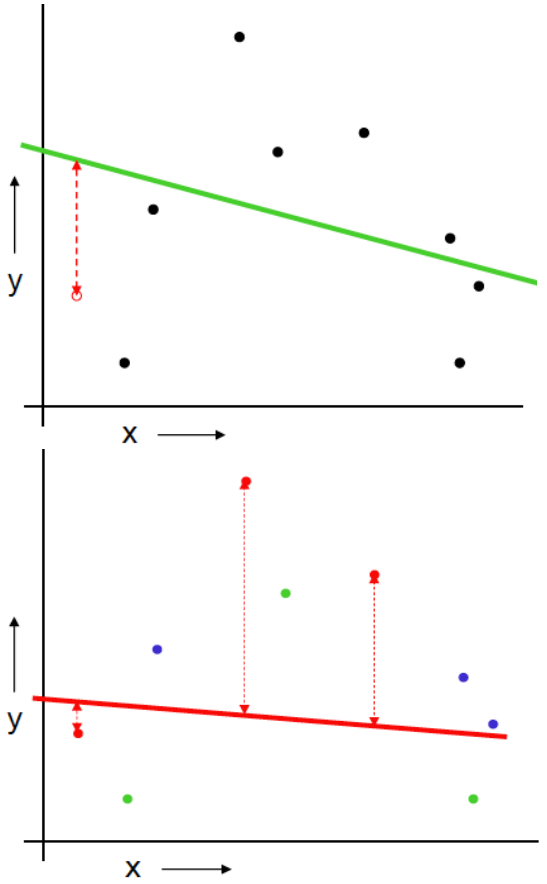


Cross validation procedure

- Partitioning sample into subsets for training and testing the classification algorithms.
- Analysis is initially performed on a single subset, while the other subset(s) are retained for subsequent use in validating the initial analysis.

Training set: the initial subset of data

Validation (testing) set: remaining subset



Leave-one-out cross-validation (LOOCV):

- Uses a single observation from original sample as validation data, and the remaining observations as training data.
- This is repeated such that each observation in the sample is used once as the validation data.

K-fold cross-validation:

- Original sample partitioned into K subsamples. Of K subsamples, one is retained as validation data while remaining (K − 1) subsamples are used as training data.
- Cross-validation process is repeated K times (the folds), with each of the K subsamples used exactly once as the validation data.
- The K results from the folds then can be averaged (or otherwise combined) to produce a single estimation.

Figures of merit

True Positive
The actual value is positive
and it is classified as positive

False Negative (Type II Error)
The actual value is positive
but it is classified as negative

True Negative
The actual value is negative
and it is classified as negative

False Positive (Type I Error)
The actual value is negative
but it is classified as positive

Confusion Matrix:

Hypothesized class \ True class	p	n
Y	True Positives	False Positives
N	False Negatives	True Negatives

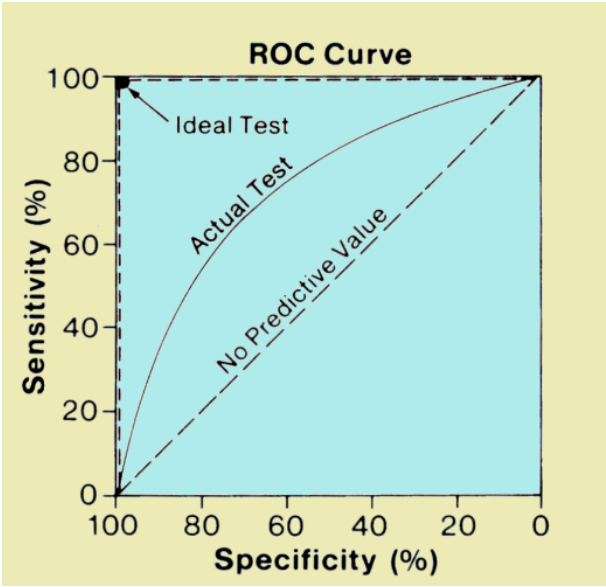
(Class assigned by the decisional system)

- True Positive Rate (TPR)
  - Positives correctly classified / Total positives

  = Sensitivity

- False Positive Rate (FPR)
  - Negatives incorrectly classified / Total negatives

  = 1 - Specificity



Area under the ROC curve (AUC)

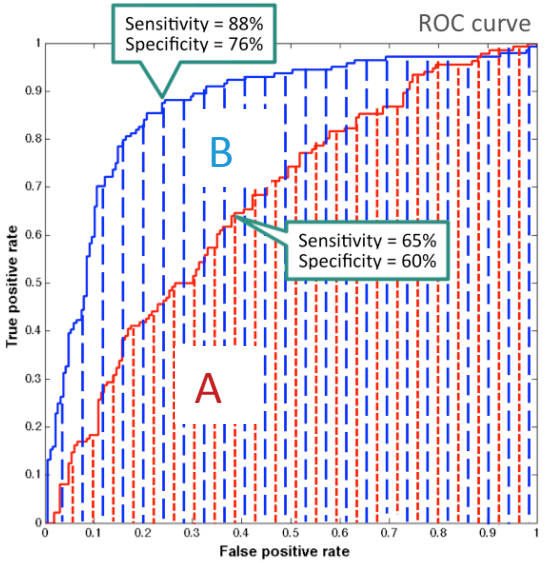
- To compare classifiers we may want to reduce the ROC performance to a single scalar value representing expected performance

  → Calculate the AUC

- Since the AUC is a portion of the area of the unit square, its value will always be between 0 and 1

- However, because random guessing produces the diagonal line between (0, 0) and (1, 1), which has an area of 0.5, *no realistic classifier should have an AUC less than 0.5*

- An ideal classifier has an area of 1



Comparing two ROC curves:

- The graph represents the areas under two ROC curves, A and B. Classifier B has greater area and therefore better average performance

- **Important statistical property:** AUC is equivalent to the probability that the classifier will rank a randomly chosen positive instance higher than a randomly chosen negative instance

[Hanley & McNeil, The meaning and use of the area under the Receiver Operating Characteristic (ROC) curve, *Radiology* 1982 143 29-36]

The Autistic Spectrum Disorder (ASD) case study

Pattern recognition approach with SVM

Autism Spectrum Disorders (ASD)

- ASD subjects have impairments in socialization, communication, and restricted interests and repetitive behaviors
- ASDs are a heterogeneous group of neurodevelopmental pathologies with a strong genetic basis and early altered neuroanatomical correlates whose diagnosis is still based on behavioral symptoms.

- The incidence of ASD is 1 in 88 children in USA

  [Centers for Disease Control and Prevention, 2012]

- ASD children may benefit from early intervention at a young age, when brain plasticity is maximal and environmental variables may have major effects on neurodevelopment

  [Dawson *et al.*, Randomized, controlled trial of an intervention for toddlers with autism: The Early Start Denver Model. *Pediatrics*, 125:17-23, 2010]

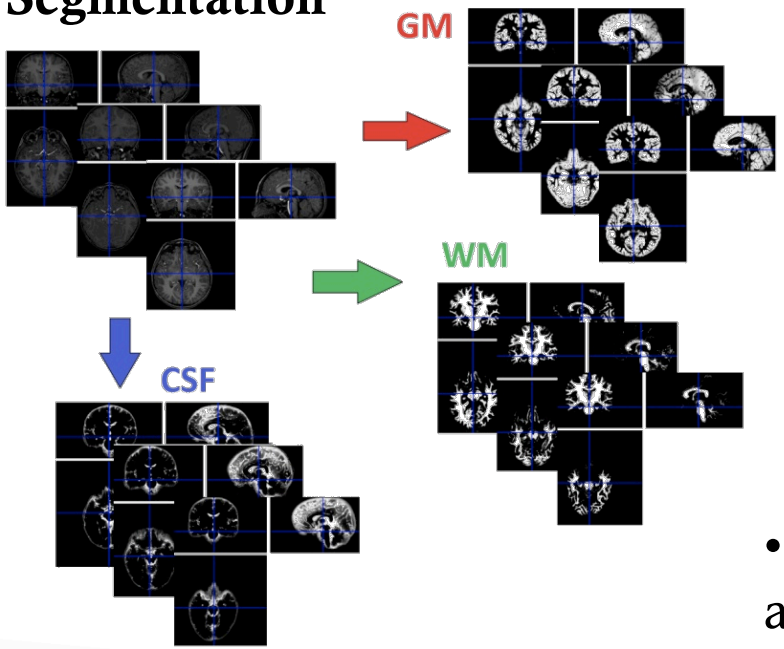
VBM-DARTEL preprocessing

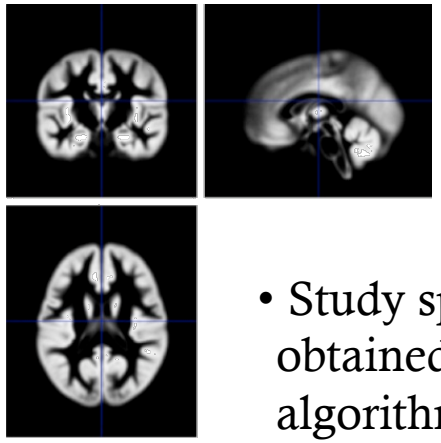
	Subject group, mean ± std [range]	
	ASD (n=76)	Ctrl (n=76)
Age (months)	53 ± 17 [25 - 88]	53 ± 18 [22 - 89]
NVIQ	71 ± 22 [30 - 113]	73 ± 23 [35 - 112]

A data set composed by 152 T_1-w MRI acquired at the IRCCS Stella Maris Foundation (Pisa, IT) has been preprocessed according to the VBM-DARTEL procedure using SPM8

(NVIQ: *nonverbal* intelligence quotient)

- **Segmentation**





- Study specific template obtained with DARTEL algorithm

- Spatial normalization to MNI space and smoothing with s= 8mm

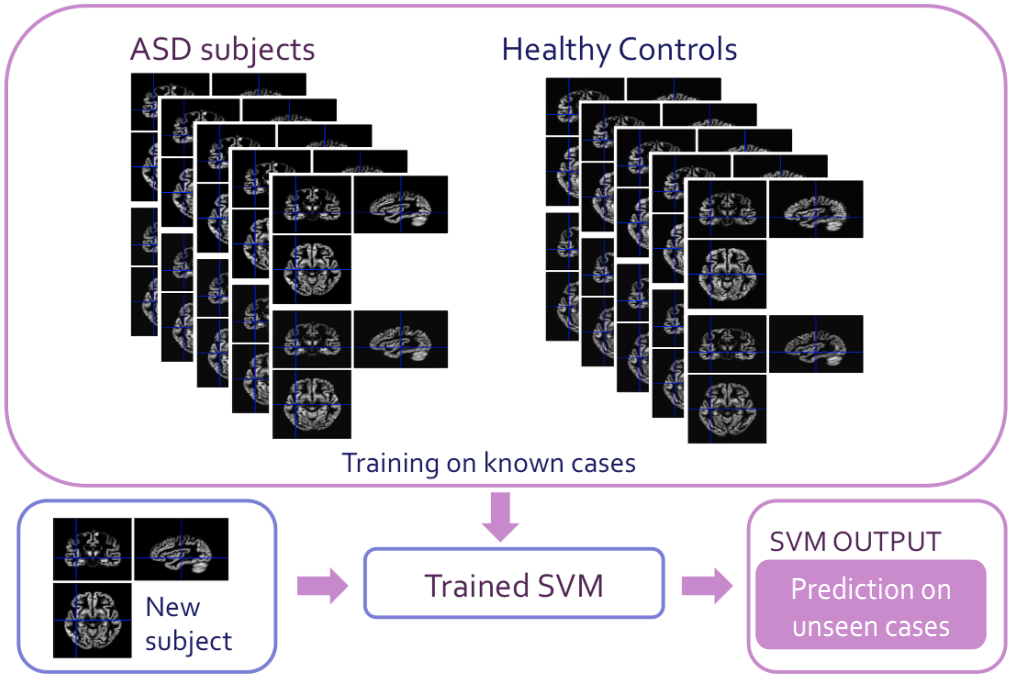
Sample characteristics and global volumes

- The brain regional absolute volumes (GM, WM, CSF and TIV) were compared between the two groups (ASD subjects and controls)

Variable	Subject group, mean ± SD [range]		t test	
	ASD (n=76)	Ctrl (n=76)	t	p value
Age (months)	53 ± 17 [25 - 88]	53 ± 18 [22 - 89]	0.0002	0.99
NVIQ	71 ± 22 [30 - 113]	73 ± 23 [35 - 112]	0.133	0.72
GM (ml)	662 ±67	629 ± 80	7.39	0.007
WM (ml)	424 ± 47	400 ± 55	8.84	0.003
CSF (ml)	225 ± 25	218 ± 34	2.41	0.123
TIV (ml)	1311 ± 134	1247 ± 162	7.18	0.008

- no meaningful between-group differences on age and NVIQ
- significant differences on GM, WM and TIV ($p<0.05$)

SVM classification of GM segments



SVM training:

- 150 training examples
- ~ 5 x 10^5 entries for each example (GM ~ 5 x 10^5 voxels)

overtraining condition

Leave-pair-out cross-validation (LPO-CV): a pair of matched subjects (one ASD patient and a matched control) in turn is left out and the classier is trained on the remaining instances.

"Feature selection" is necessary to reduce overfitting

- Recursive Feature Elimination (RFE) technique

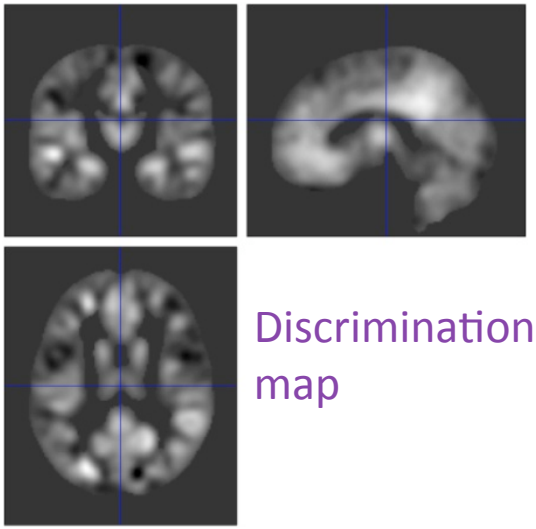
SVM-RFE

SVM-RFE aims to retain in the data set only those features/voxels that carry high discriminant information:

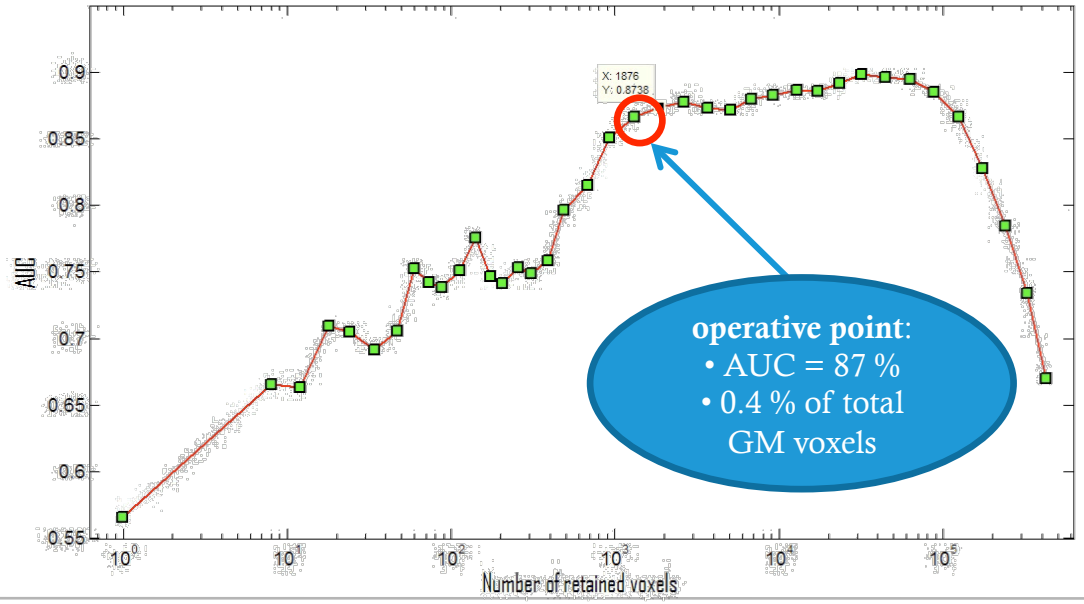
- the weight magnitude $|w_i|$ is used as ranking criterion (i.e. features/voxels with $|w_i|$ under a threshold $|T_j|$ are iteratively removed from the dataset):

$$T_j = min|w_i| + j(max|w_i| - min|w_i|)/N, \qquad \text{with} \quad j = 0, \ldots, N,$$

- a new SVM is trained at each iteration

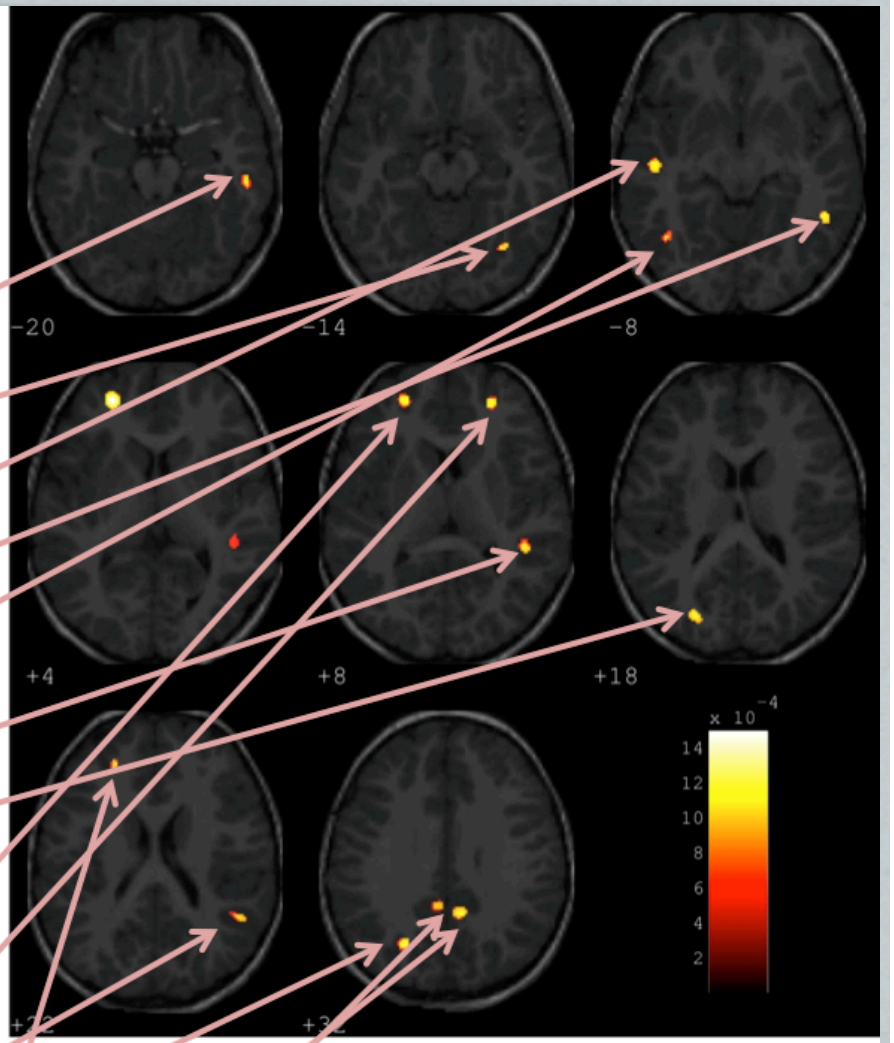


Discrimination map



Most discriminant GM areas (ASD>CTRL)

# voxels	Talairach coordinates	Brain Area			
40	(51, -27, -15)	RC	Temporal lobe	Inferior temporal gyrus	BA 20
29	(31, -65, -13)	RC	Posterior lobe	Cerebellum	
48	(-41, -59, -7)	LC	Temporal lobe	Fusiform gyrus	BA 37
41	(48, -49, -7)	RC	Temporal lobe	Sub-gyral	BA 37
78	(-49, -19, -6)	LC	Temporal lobe	Superior temporal gyrus	BA 22
77	(44, -38, -8)	RC	Temporal lobe	Superior temporal gyrus	BA 41
56	(-26, -79, 12)	LC	Occipital lobe	Middle occipital gyrus	BA 19
253	(-25, 47, 12)	LC	Frontal lobe	Superior frontal gyrus	BA 10
79	(25, 44, 18)	RC	Frontal lobe	Superior frontal gyrus	BA 10
74	(45, -53, 22)	RC	Temporal lobe	Superior temporal gyrus	BA 39
87	(-26, -68, 25)	LC	Occipital lobe	Precuneus	BA 31
25	(-25, 36, 28)	LC	Frontal lobe	Middle frontal gyrus	BA 9
116	(6, -50, 29)	RC	Parietal lobe	Precuneus	BA 31
80	(-7, -47, 30)	LC	Parietal lobe	Precuneus	BA 31

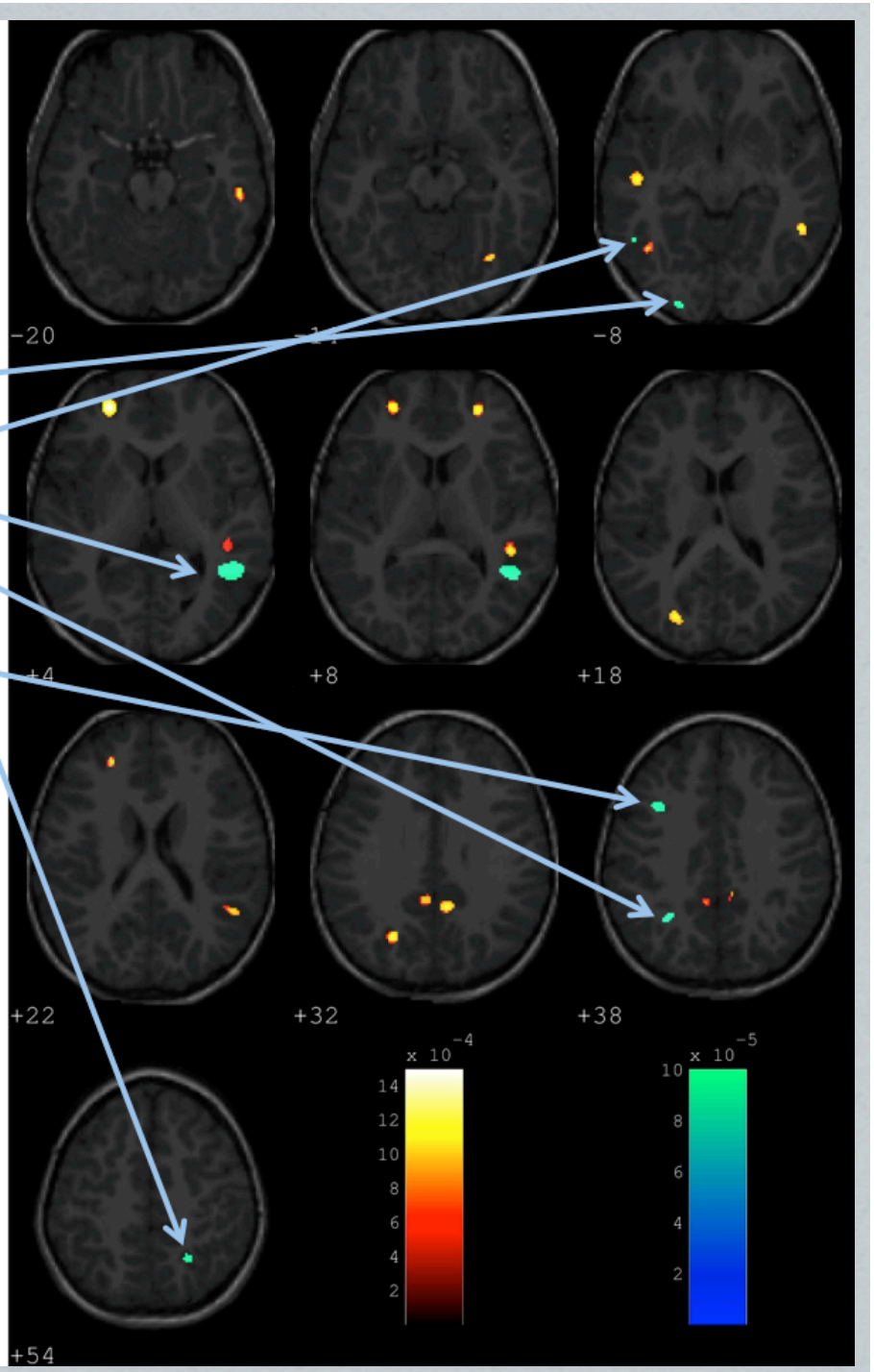


AUC: 87 %
Sensitivity: 82 %
Specificity: 80 %
Retained voxels: 0.4 %

Most discriminant GM areas (ASD<CTRL)

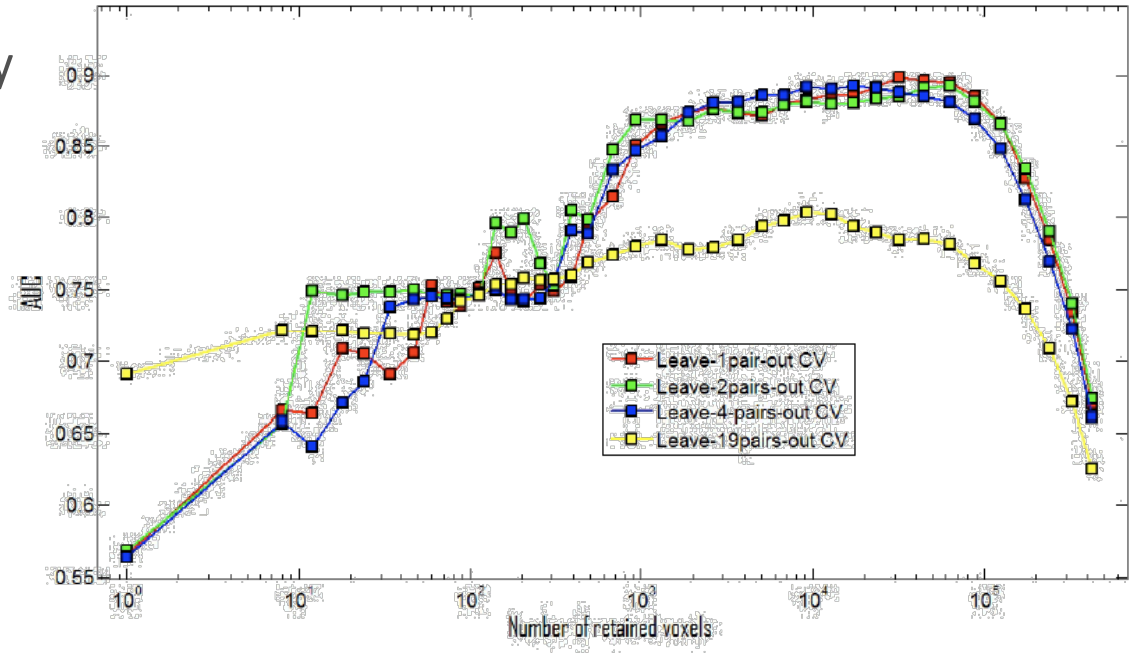
# voxels	Talairach coordinates	Brain Area			
49	(-24, -94, -10)	LC	Occipital lobe	Fusiform gyrus	BA 18
33	(-50, -54, -6)	LC	Temporal lobe	Inferior temporal gyrus	BA 37
340	(45, -51, 4)	RC	Temporal lobe	Inferior temporal gyrus	BA 37
65	(-31, -57, 33)	LC	Parietal lobe	Angular gyrus	BA 39
171	(-37, 8, 42)	LC	Frontal lobe	Middle frontal gyrus	BA 6
57	(20, -59, 48)	RC	Parietal lobe	Precuneus	BA 7

AUC: 87 %
Sensitivity: 82 %
Specificity: 80 %
Retained voxels: 0.4 %



Impact of the training set on the SVM analysis

To evaluate the reliability and the generalization ability of the SVM-based analysis we can estimate the dependency on the training sample size



CV-procedure	training set (n. of subjects)	validation set (n. of subjects)	n. of folds (CV iterations)
leave-1pair-out	150	2	76
leave-2pairs-out	148	4	38
leave-4pairs-out	144	8	19
leave-19pairs-out	114	38	4

Advantages of SVM-RFE analysis

- In contrast to VBM, SVM is a multivariate approach thus automatically takes into account interregional correlations.
  - Identification of patterns of brain alterations

- In addition SVM offers a predictive value
  - Once the decision function is learned from the training data (AD vs. CTRL, ASD vs. CTRL) it can be used to predict the class membership of new test examples (MCI prediction of conversion, ASD predictive diagnosis).

Region of Interest (ROI)-based analysis vs. whole-brain approaches

Application to the Alzheimer's disease data set

SVM-RFE analysis of AD-MCI and Control data

- Training set (20-fold cross validation)
  - AD and CTRL data
- Validation set
  - MCI converters and non converters

Cohort	Sample size
CTRL	189
MCI-NC	166
MCI-C	136
AD	144
TOT	635

- Aim of the analysis:
  - To setup a decision-making system to predict the outcome of MCI patients.

Scheme of the analysis

- Region of interest (ROI)-based methods:
  - require prior knowledge of interesting ROIs

- Whole-brain methods:
  - NO prior knowledge is needed
  - Data-driven approaches

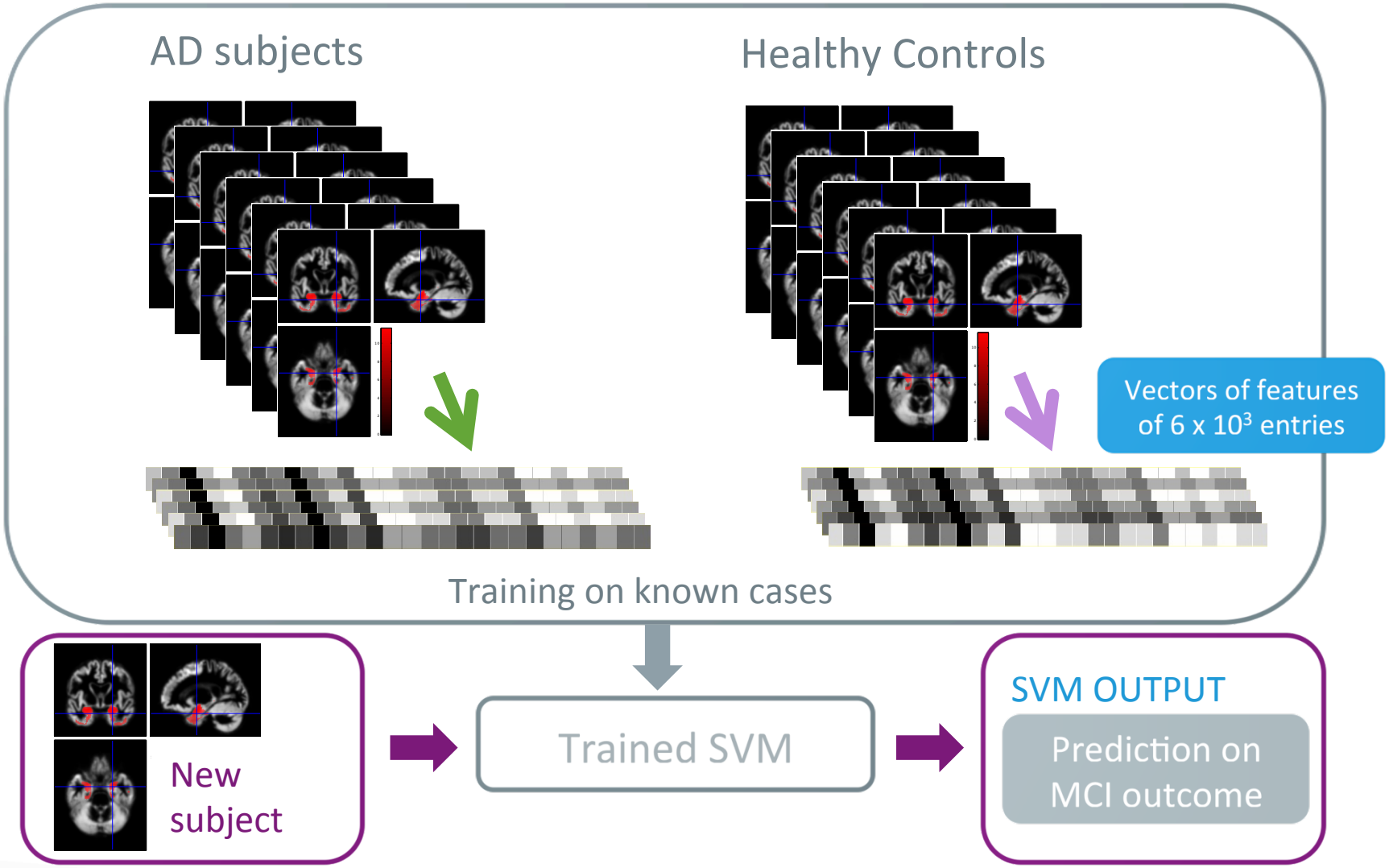
Hybrid ROI-based method:

- the ROI is the region where the between-group difference is significant according to VBM analysis (voxel-wise two-sample t-test)
- A SVM classifier is implemented to make single-subject inference

Whole-brain analysis:

- The whole GM is considered
- A SVM classifier is implemented to make single-subject inference
- The Recursive-Feature-Elimination allows for:
  - Feature selection (dimensionality reduction)
  - Localization of the "relevant information" (automated selection of ROIs)

Support Vector Machine (SVM) classification



Hybrid ROI-based method: results

SVM classification of VBM ROIs:

- Linear-kernel SVM (c parameter optimized on training data)

Train and test set:

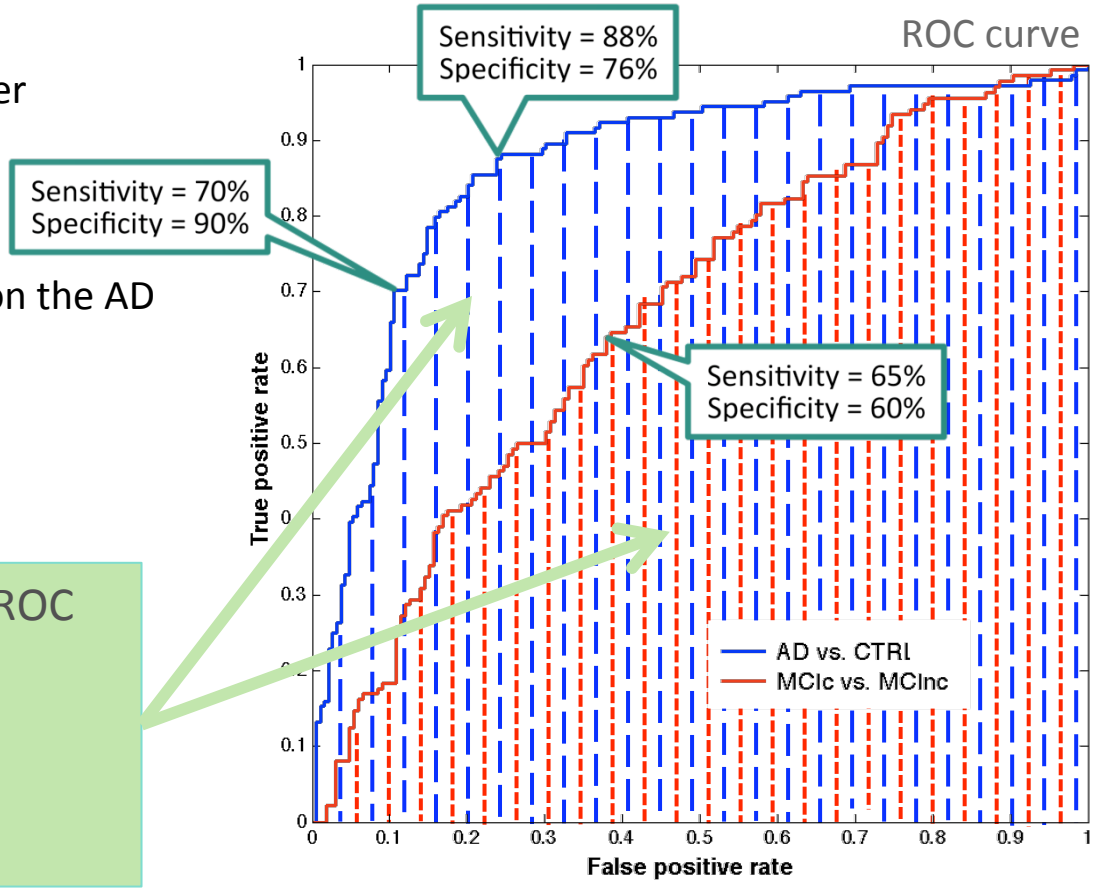
- k-fold cross validation (k=20) on the AD +CTRL data (333 subjects)

Validation

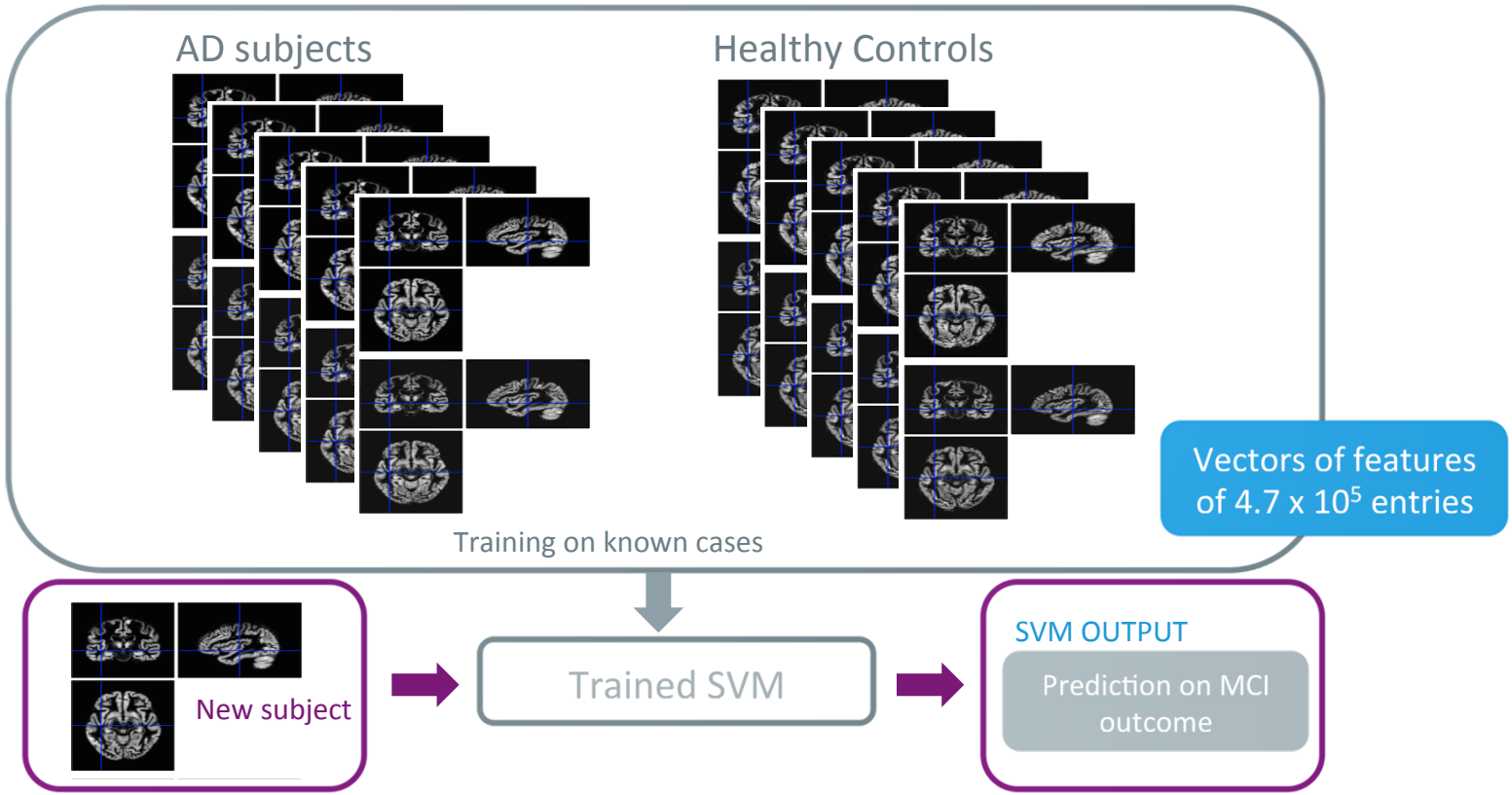
- MCI data (302 subjects)

We use the Area Under the ROC curve as a figure of merit to compare different methods:

- $AUC_{AD\text{-}CTRL} = 0.87$
- $AUC_{MCIc\text{-}MCInc} = 0.67$



Whole-GM SVM-RFE classification



Is it worth performing feature selection?

[R. Cuingnet et al, ADNI, NeuroImage 56 (2011) 766-781]

Does feature selection improve classification accuracy?

[C. Chu et al, ADNI, NeuroImage 60 (2012) 59-70]

Local vs. global approach

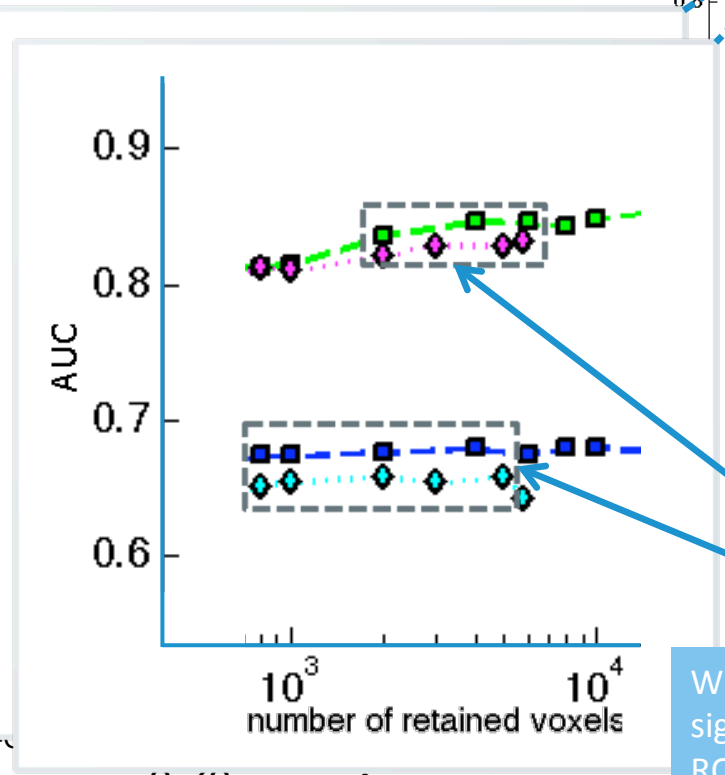
SVM-RFE has been applied to both the *Hybrid* ROI-based method and the whole-GM classification method.

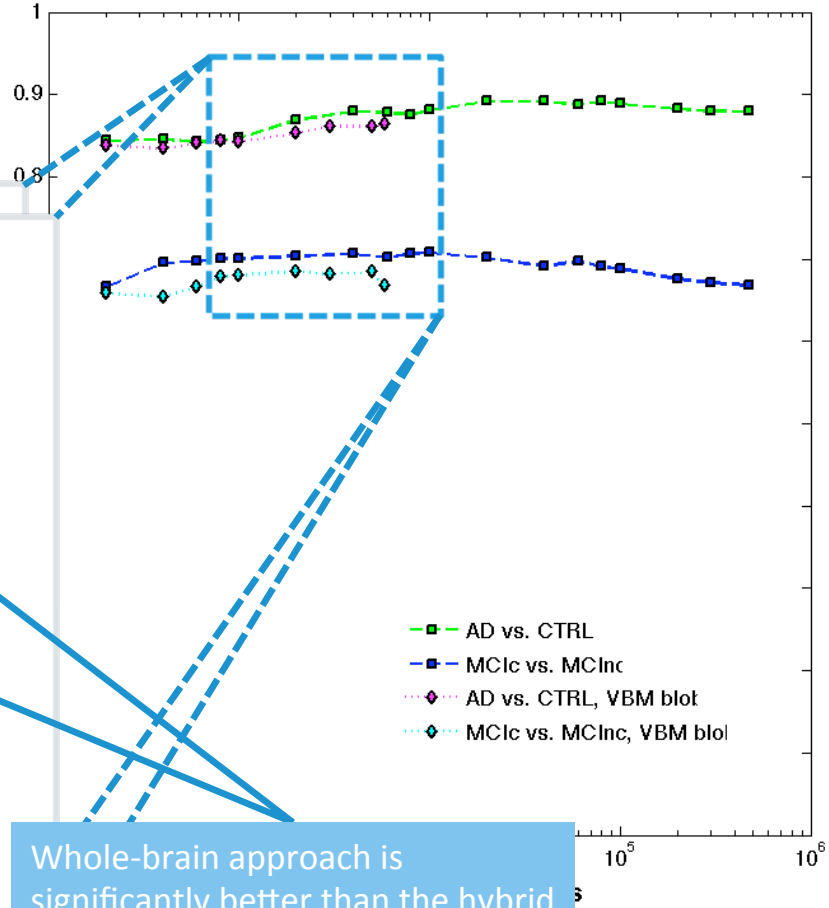
Hybrid RO

- AUC_{AD}
- AUC_{MC}

Whole GM

- AUC_{AD}
- AUC_{MC}
- AUC_{AD}
- AUC_{MC}
- AUC_{AD-C}
- $AUC_{MCIc-MCInc} = 0.70$ (6x10^3 voxels ≈ VBM blob si

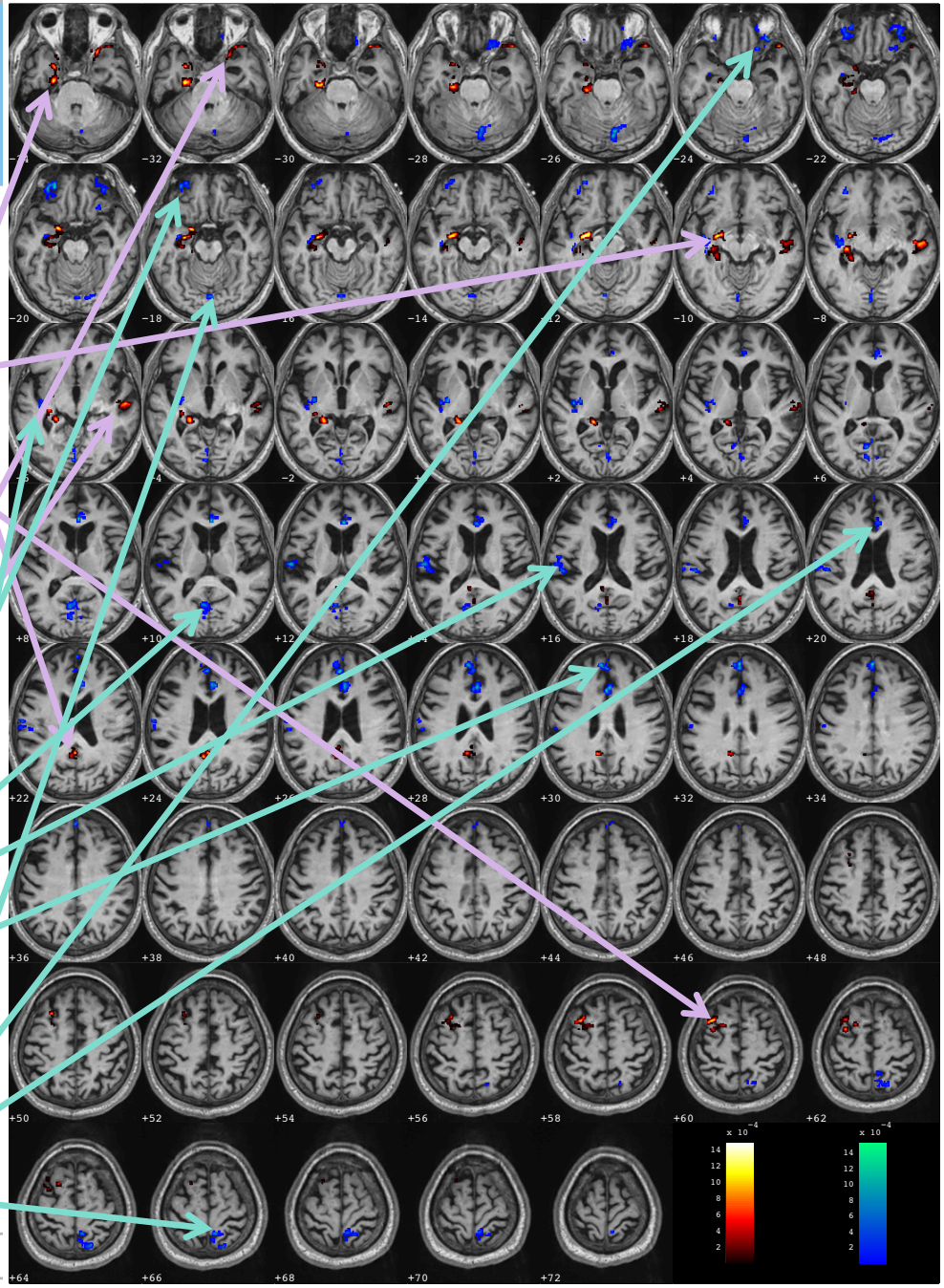




Whole-brain approach is significantly better than the hybrid ROI-based one ($p < 0.05$)

Most discriminant GM areas

# voxels	MNI coordinates			Talairach coordinates			ROIs		
446	-24	-9	-32	-23	-7	-26	LC	Limbic Lobe	Parahippocampal Gyrus
762	-25	-20	-11	-24	-19	-8	LC	Temporal Lobe	Sub-Gyral
267	-5	-53	23	-6	-53	20	LC	Limbic Lobe	Posterior Cingulate
303	-28	7	59	-28	0	57	LC	Frontal Lobe	Middle Frontal Gyrus
205	32	24	-32	29	23	-22	RC	Frontal Lobe	Inferior Frontal Gyrus
344	56	-20	-6	51	-20	-2	RC	Temporal Lobe	Superior Temporal Gyrus
240	-29	47	-17	-28	44	-7	LC	Frontal Lobe	Middle Frontal Gyrus
277	-37	-16	-5	-35	-16	-2	LC	Sub-lobar	Extra-Nuclear
221	-6	-70	9	-7	-68	6	LC	Occipital Lobe	Cuneus
312	-58	-20	18	-55	-22	18	LC	Parietal Lobe	Postcentral Gyrus
269	-3	50	30	-4	42	35	LC	Frontal Lobe	Medial Frontal Gyrus
351	5	-78	-17	4	-73	-18	RC	Posterior Lobe	Declive
235	25	35	-24	22	33	-14	RC	Frontal Lobe	Middle Frontal Gyrus
442	4	29	20	3	23	25	RC	Limbic Lobe	Anterior Cingulate
238	10	-59	65	7	-62	57	RC	Parietal Lobe	Precuneus



A. Retico - INFN Pisa

Conclusions

- Machine learning approaches are gaining popularity in the Neuroimaging community
- In addition to group characterization, they allow categorization of individual's previously unseen data: predictive diagnosis.
- Interesting studies have already been carried out on neurological and psychiatric disorders (Alzheimer's disease, Parkinson's disease, autistic spectrum disorders, schizophrenia, bipolar disorder,…)

  ….. and many more are in progress!

Acknowledgments

Thank you for your kind attention!

Data shown in this seminar have been collected by:

- the Alzheimer's Disease Neuroimaging Initiative (ADNI) (National Institutes of Health Grant U01 AG024904)
- the IRCCS Stella Maris Foundation

The analyses have been carried out within:

- the Italian MAGIC-5 research project funded by *Istituto Nazionale di Fisica Nucleare* (INFN, CSN5)
- the research collaboration between IRCCS Stella Maris and INFN (Ministry of Health and Tuscany Government grant GR-2010-2317873)