

Computing Methods for Experimental Physics and Data Analysis

Data Analysis in Medical Physics

Lecture 7: Extracting features from images, analysis of image features

Alessandra Retico

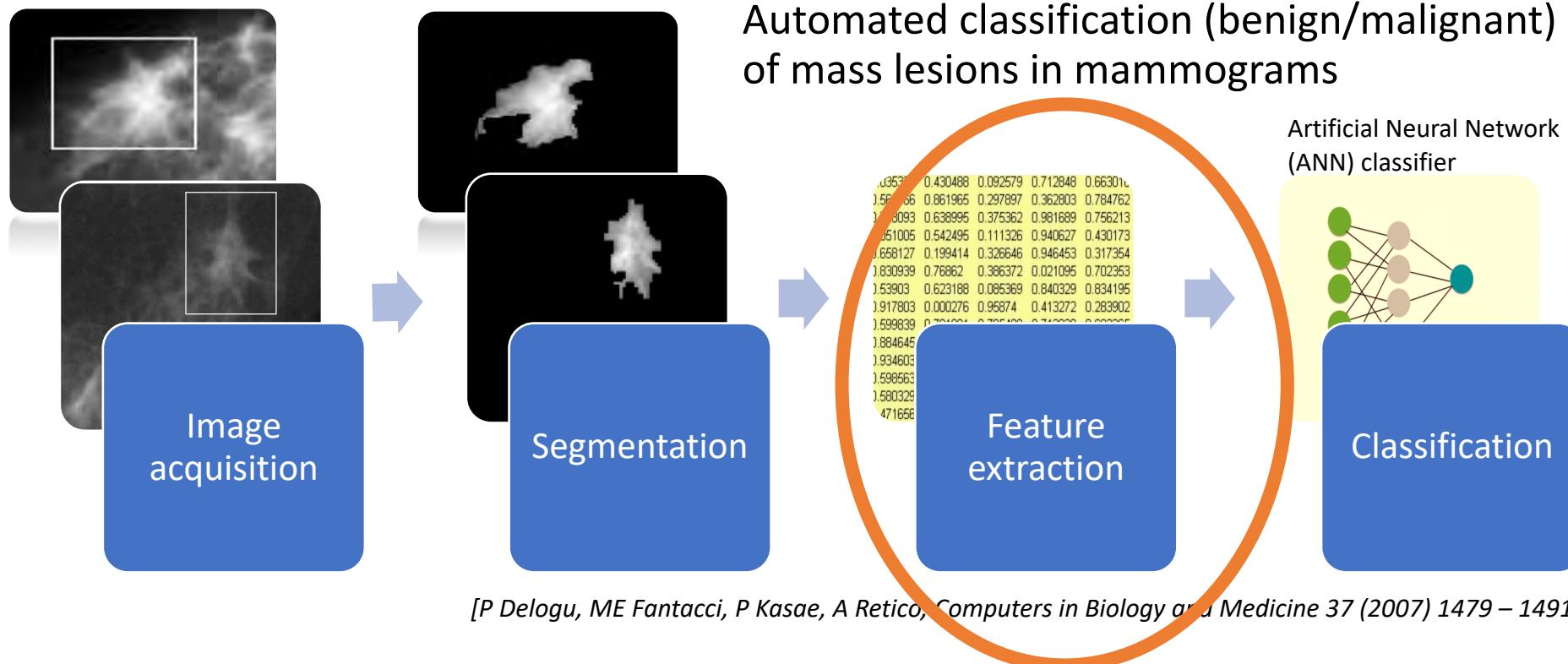
alessandra.retico@pi.infn.it

INFN - Pisa

Typical image analysis pipeline for assisted diagnosis

Example: 1) Object segmentation; 2) Hand-crafted feature extraction; 3) Machine Learning classification

2D images



See demo code: Lecture7_demo1_extract_features.m

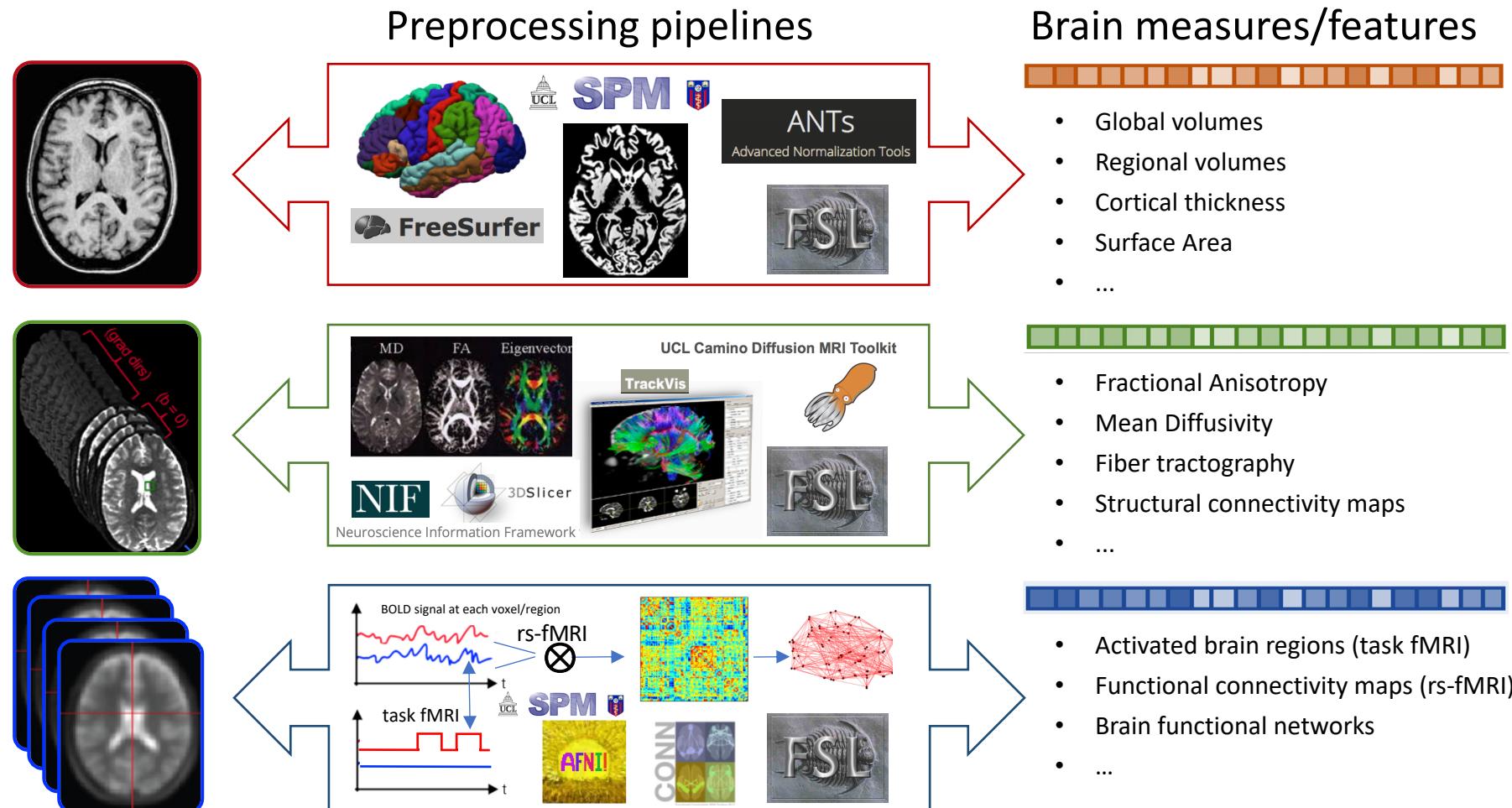
Image/ROI features

Several descriptive features of segmented ROIs are provided by the `regionprops` MATLAB function, which can return:

- Shape Measurements:
 - 'Area', 'Centroid', and 'BoundingBox'
 - 'ConvexArea', 'ConvexHull', 'ConvexImage', 'Circularity', 'EulerNumber', 'Filled Area', 'FilledImage', 'MaxFeretProperties', 'MinFeretProperties' and 'Solidity'
- Pixel Value Measurements:
 - 'MaxIntensity', 'MeanIntensity', 'MinIntensity', 'PixelValues', 'WeightedCentroid'

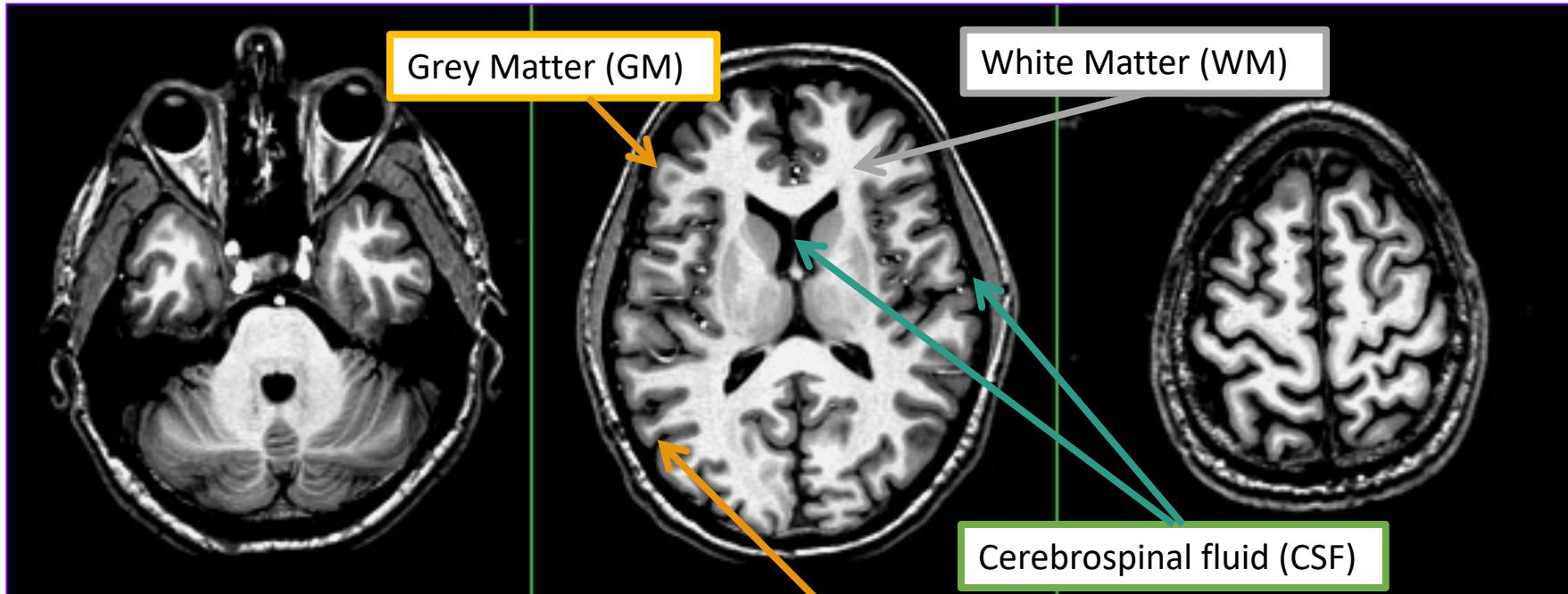
The function `regionprops3` provides analogous features for 3D ROIs

Feature extraction from brain images (few examples)



MRI T_1 -weighted brain images

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



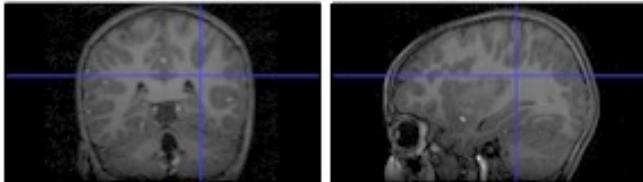
Grey Matter (GM) cortex can be followed and cortical thickness can be evaluated to investigate GM involvement in pathological conditions.

The need for tissue segmentation in MRI

- 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

 **Quantification**

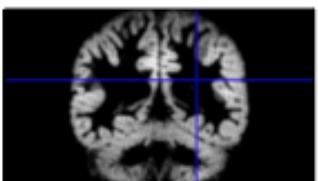
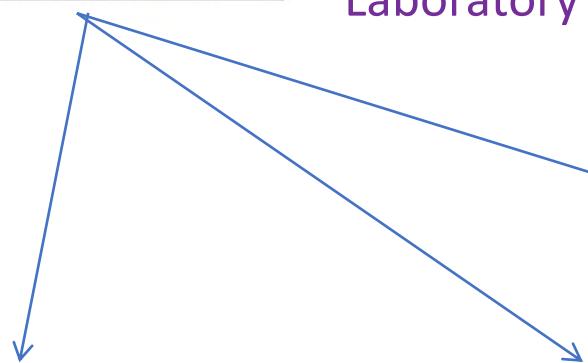
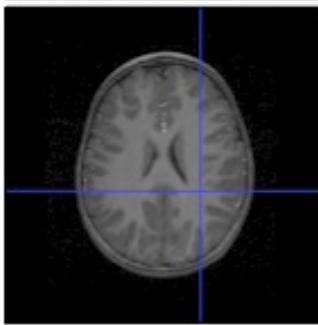
Segmentation of brain components



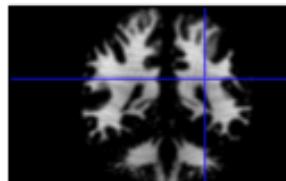
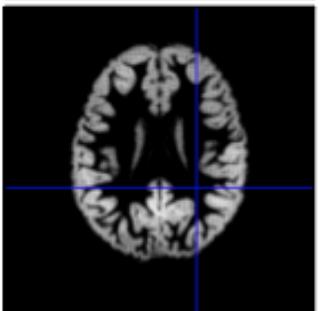
Tissue segmentation in SPM (Statistical Parametric Mapping)
Wellcome Trust Centre for Neuroimaging, Functional Imaging
Laboratory (FIL), University College of London (UCL)



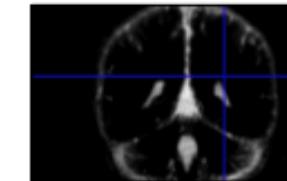
www.fil.ion.ucl.ac.uk/spm/



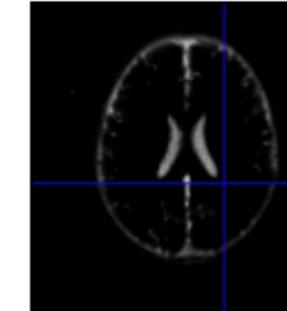
Grey matter



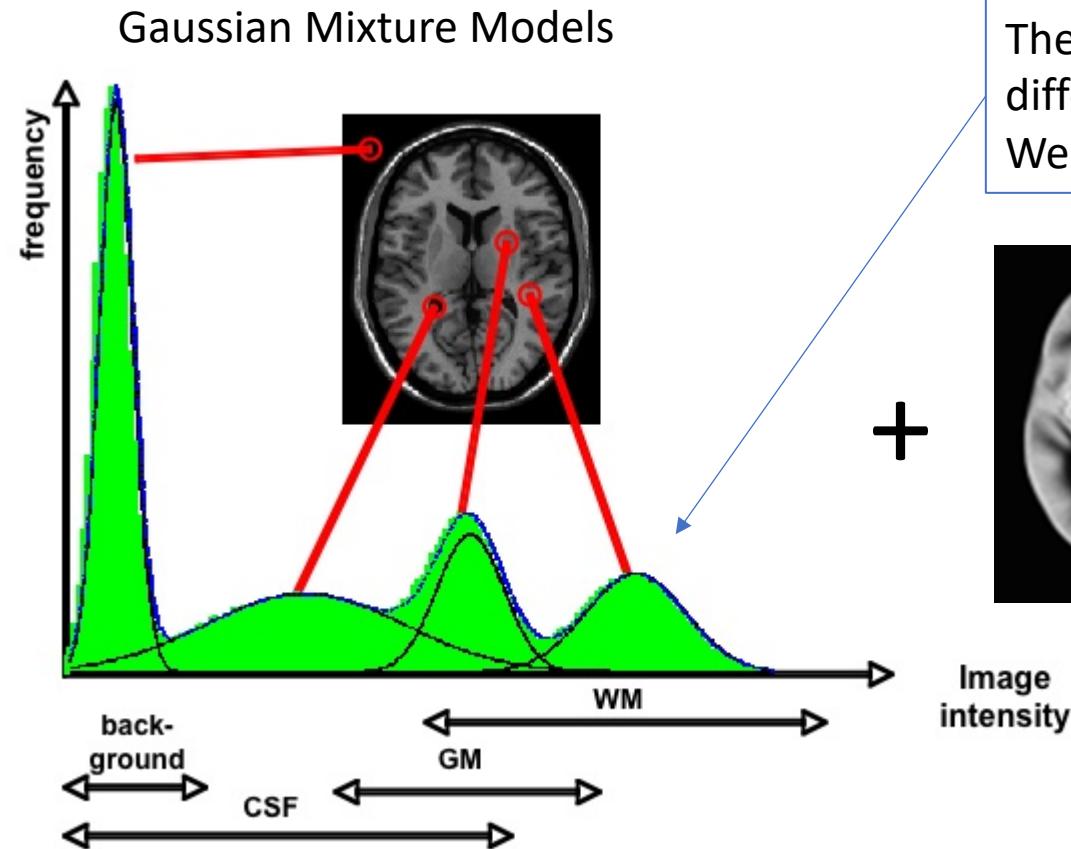
White matter



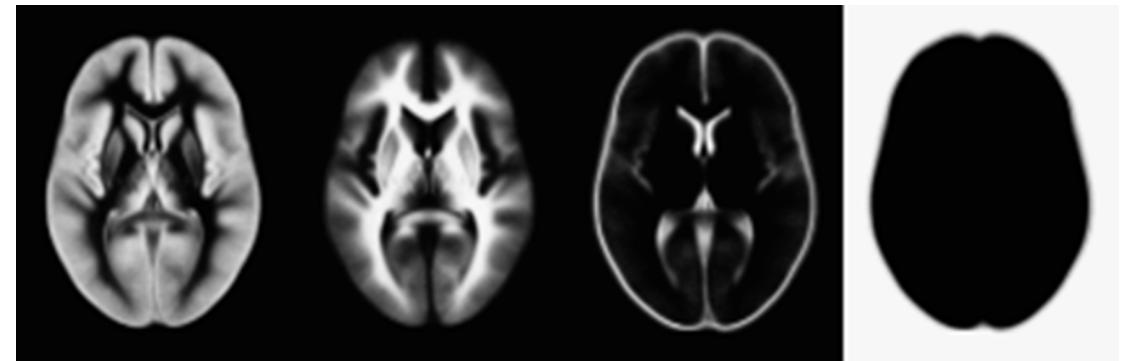
Cerebrospinal fluid



Gaussian Mixture Models (MOG)



The central values of these Gaussians are variable across different images/subjects.
We can exploit the spatial information provided by TPMs.



- Take each voxel and look at both its intensity and prior spatial knowledge, to make segmentation an easier problem to solve.

Tissue Probability Maps

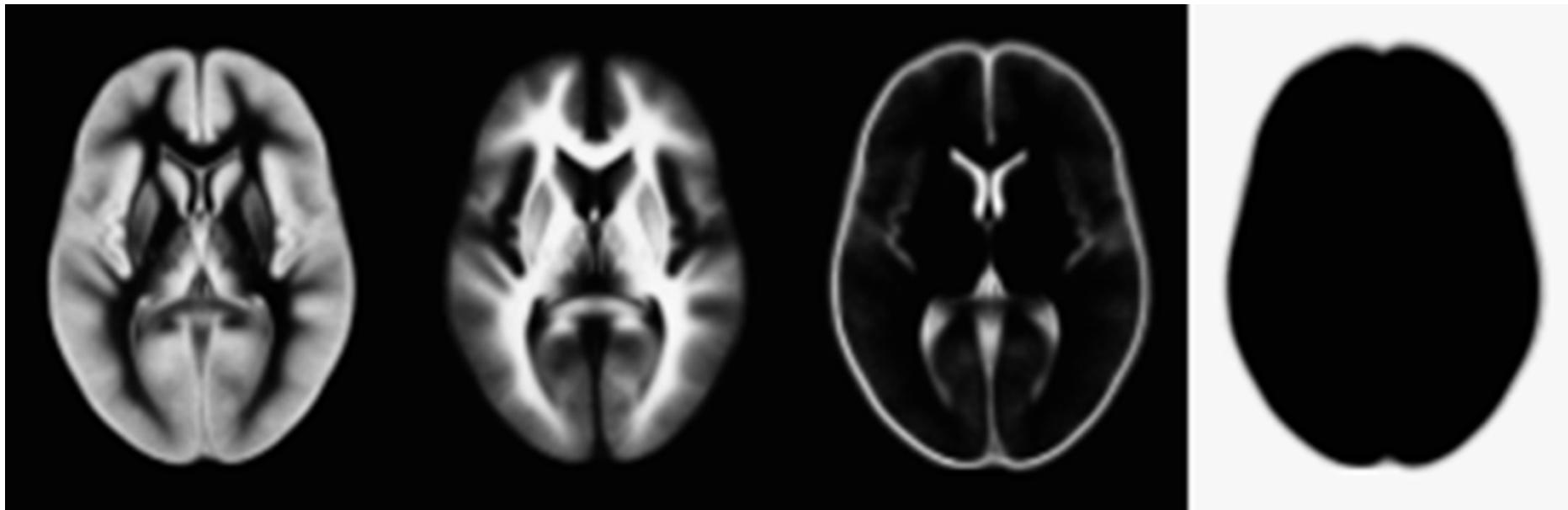
- Tissue Probability Maps (TPM): standard space structural estimates of the location of GM, WM, CSF, etc...
- Based on multiple previous segmentations

TPM,1

TPM,2

TPM,3

TPM,4



Segmentation of brain components



- 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:

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

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

$$\begin{aligned} P(y_i, c_i = k | \mu_k, \sigma_k, \gamma_k) \\ = P(y_i | c_i = k, \mu_k, \sigma_k) P(c_i = k | \gamma_k) \end{aligned}$$



Sum over K clusters (e.g. GM, WM, CSF, other components), accounting for each voxel and maximize with respect to the unknowns (μ, σ)

The freesurfer brain segmentation software



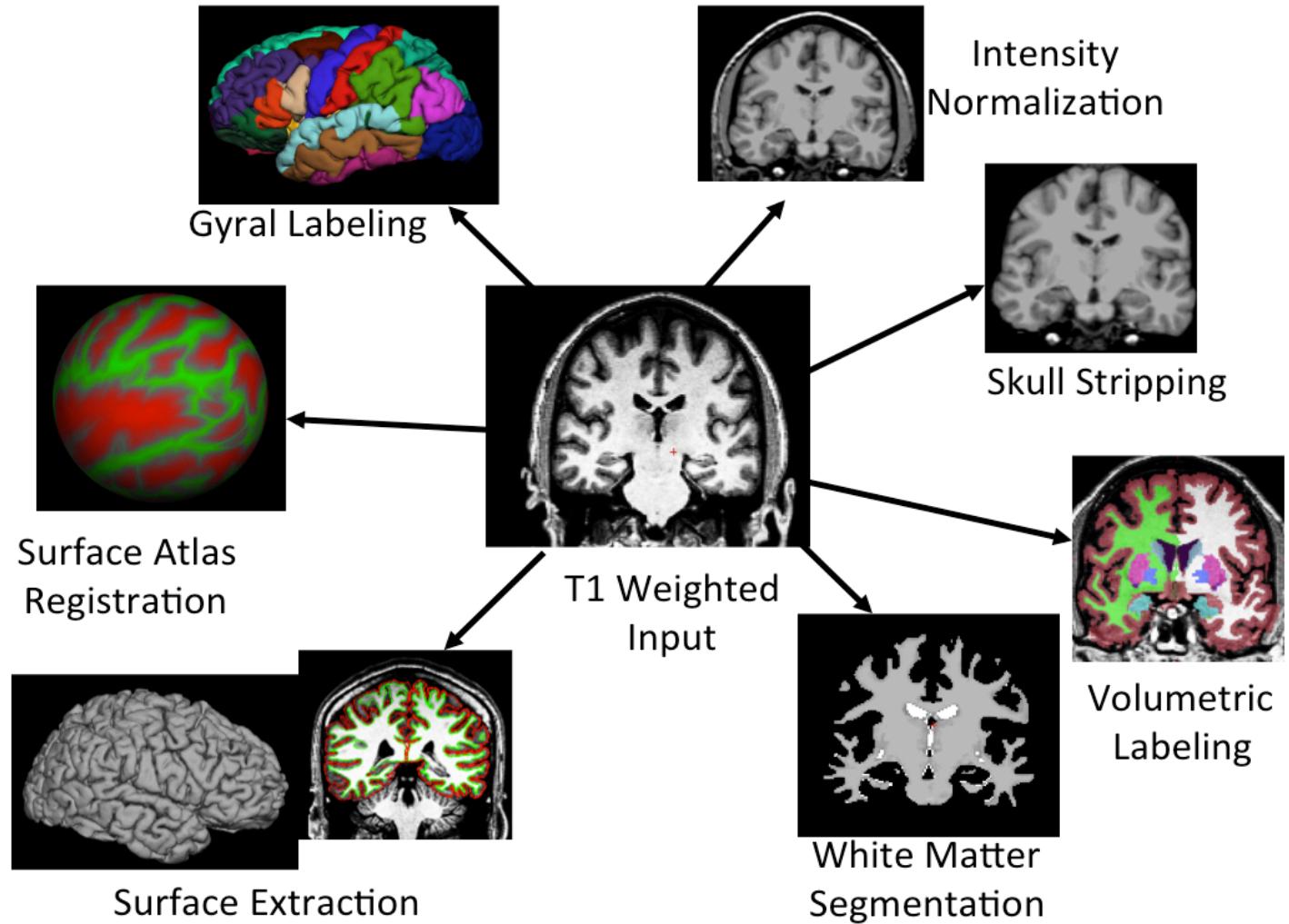
FreeSurfer v6.0
<http://freesurfer.net/>

FreeSurfer Software Suite

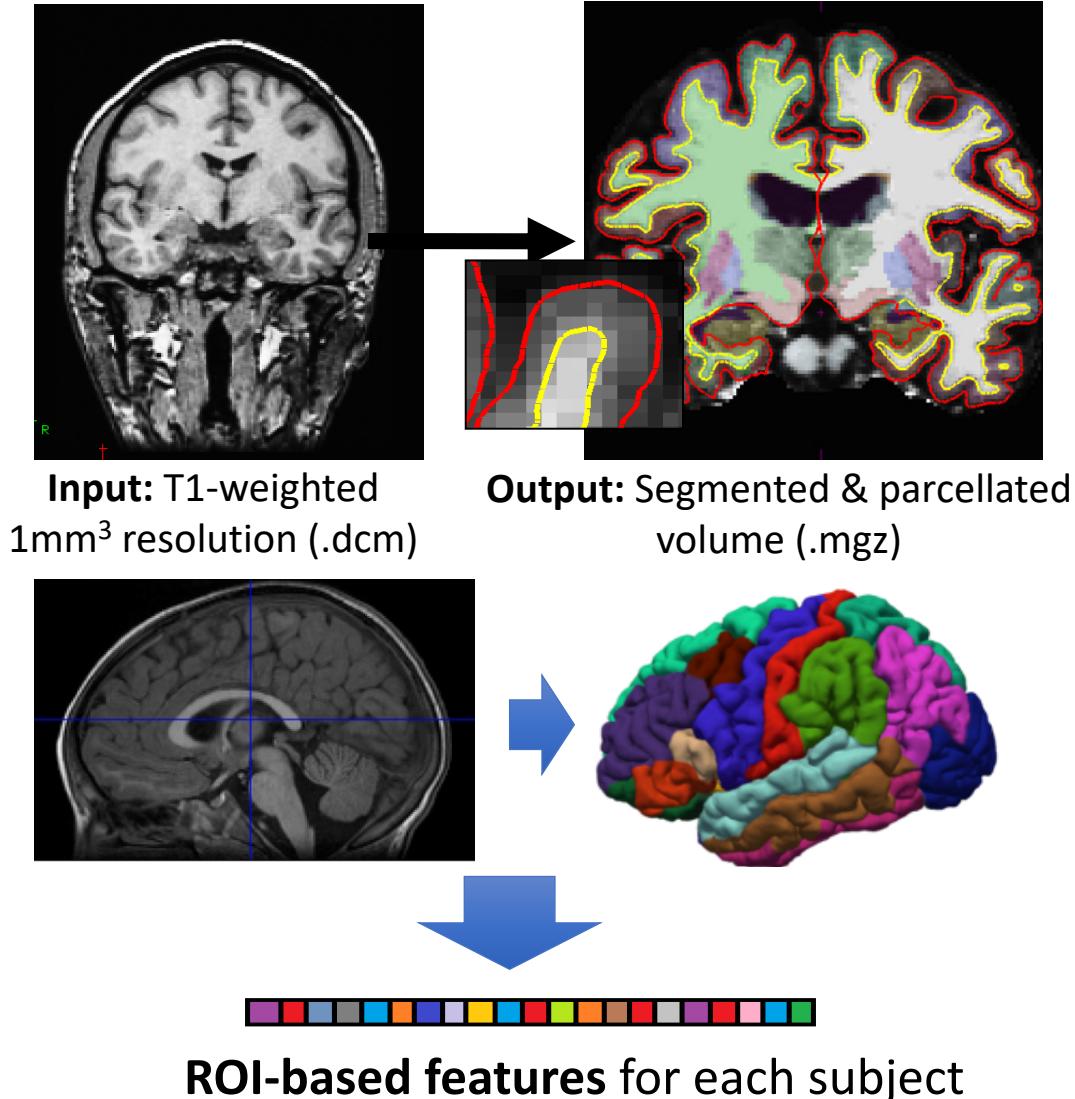
An open source software suite for processing and analyzing (human) brain MRI images.

- Skullstripping
- Image Registration
- Subcortical Segmentation
- Cortical Surface Reconstruction
- Cortical Segmentation
- Cortical Thickness Estimation
- Longitudinal Processing
- fMRI Analysis
- Tractography
- FreeView Visualization GUI
- and much more...

to run all steps in the FreeSurfer preprocessing stream: `recon-all -all`



The freesurfer brain segmentation SW



<http://surfer.nmr.mgh.harvard.edu/>

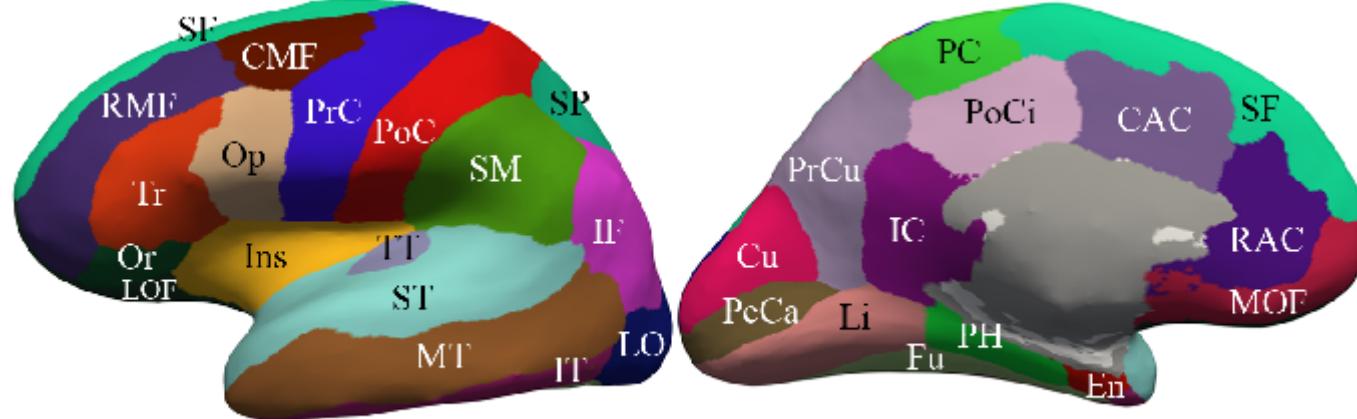
- The cerebral cortex is parcelled into a number of structures depending on the atlas we choose, e.g. **62 structures** according to Desikan-Killiany atlas.
- **Several Region-of-Interest (ROI)-based features** can be computed for each structure (**area; volume; average thickness; thickness standard deviation; mean curvature**)
- Also **Global features** can be considered, e.g.: **white surface total area** and **mean thickness of cerebral cortex** for both hemispheres.

The surface-based pipeline consists of several stages described in detail in:

- Dale, A.M., Fischl, Bruce, Sereno, M.I., Cortical Surface-Based Analysis I: Segmentation and Surface Reconstruction. *NeuroImage* 9(2):179-194. 1999.
- Fischl, B.R., Sereno, M.I., Dale, A. M. Cortical Surface-Based Analysis II: Inflation, Flattening, and Surface-Based Coordinate System. *NeuroImage* 9, 195-207. 1999.

Cortical parcellation

- Cortical parcellation according to the Desikan-Killiany atlas. Both hemispheres contain the same regions.



Region (gyrus)	Abb.
Caudal anterior cingulate	CAC
Caudal middle frontal	CMF
Cuneus	Cu
Entorhinal	En
Fusiform	Fu
Inferior parietal	IP
Inferior temporal	IT
Isthmus cingulate	IC
Lateral occipital	LO
Lateral orbitofrontal	LOF
Lingual	Li

Region (gyrus)	Abb.
Medial orbitofrontal	MOF
Middle temporal	MT
Parahippocampal	PH
Paracentral	PC
Pars opercularis	Op
Pars orbitalis	Or
Pars triangularis	Tr
Pericalcarine	PeCa
Postcentral	PoC
Posterior cingulate	PoCi

Region (gyrus)	Abb.
Precentral	PrC
Precuneus	PrCu
Rostral anterior cingulate	RAC
Rostral middle frontal	RMF
Superior frontal	SF
Superior parietal	SP
Superior temporal	ST
Supramarginal	SM
Transverse temporal	TT
Insula	Ins

R. S. Desikan, F. Sgonne, B. Fischl, B. T. Quinn, B. C. Dickerson, D. Blacker, R. L. Buckner, A. M. Dale, R. P. Maguire, B. T. Hyman, "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest", *Neuroimage*, vol. 31, no. 3, pp. 968-980, 2006.

Subcortical volume segmentation

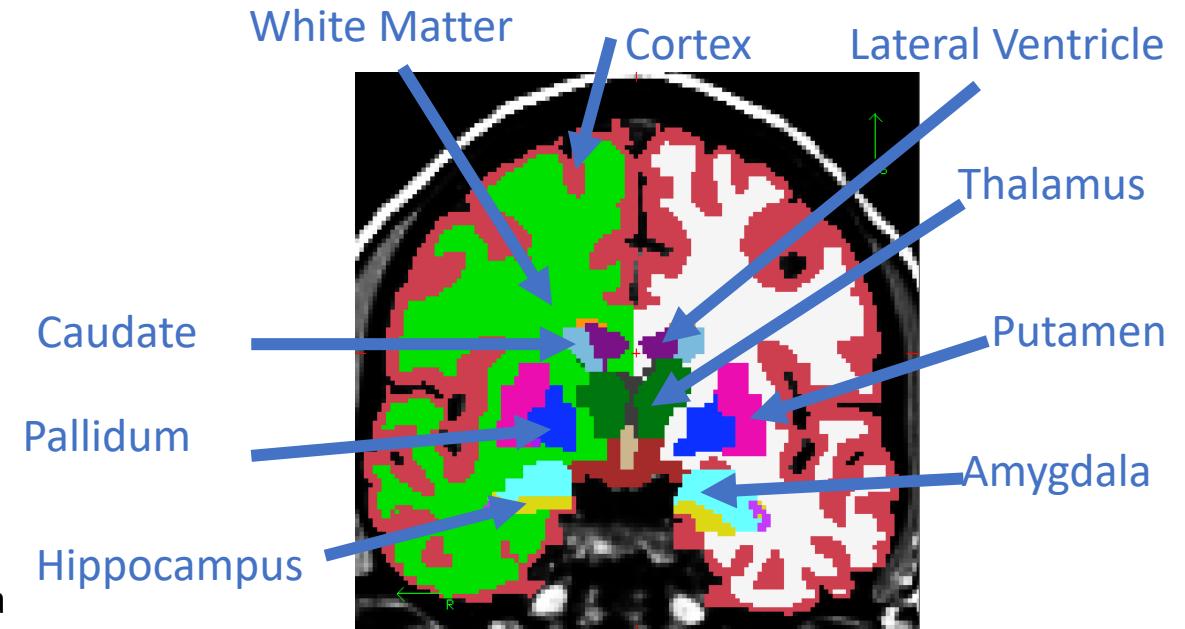
The volume-based stream to label subcortical tissue classes consists of five stages:

- affine registration with MNI305
- initial volumetric labeling*
- correction of variation in intensity due to inhomogeneities in the B1 magnetic field
- high dimensional nonlinear volumetric alignment to the MNI305 atlas
- final labeling

* The atlas used for labeling is built from a training set, i.e., a set of subjects whose brains (surfaces or volumes) have been labeled by hand.

The volume-based stream is designed to label subcortical tissue classes is described in detail in:

- Fischl B, Salat D, Busa E, Albert M, Dieterich M, Haselgrove C, van der Kouwe A, Killiany R, Kennedy D, Klaveness S, Montillo A, Makris N, Rosen B, Dale A. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* Jan 31;33(3):341-55. 2002.
- Fischl B, van der Kouwe A, Destrieux C, Halgren E, Segonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D, Caviness V, Makris N, Rosen B, Dale AM. Automatically parcellating the human cerebral cortex. *Cereb Cortex*. 2004 Jan;14(1):11-22.

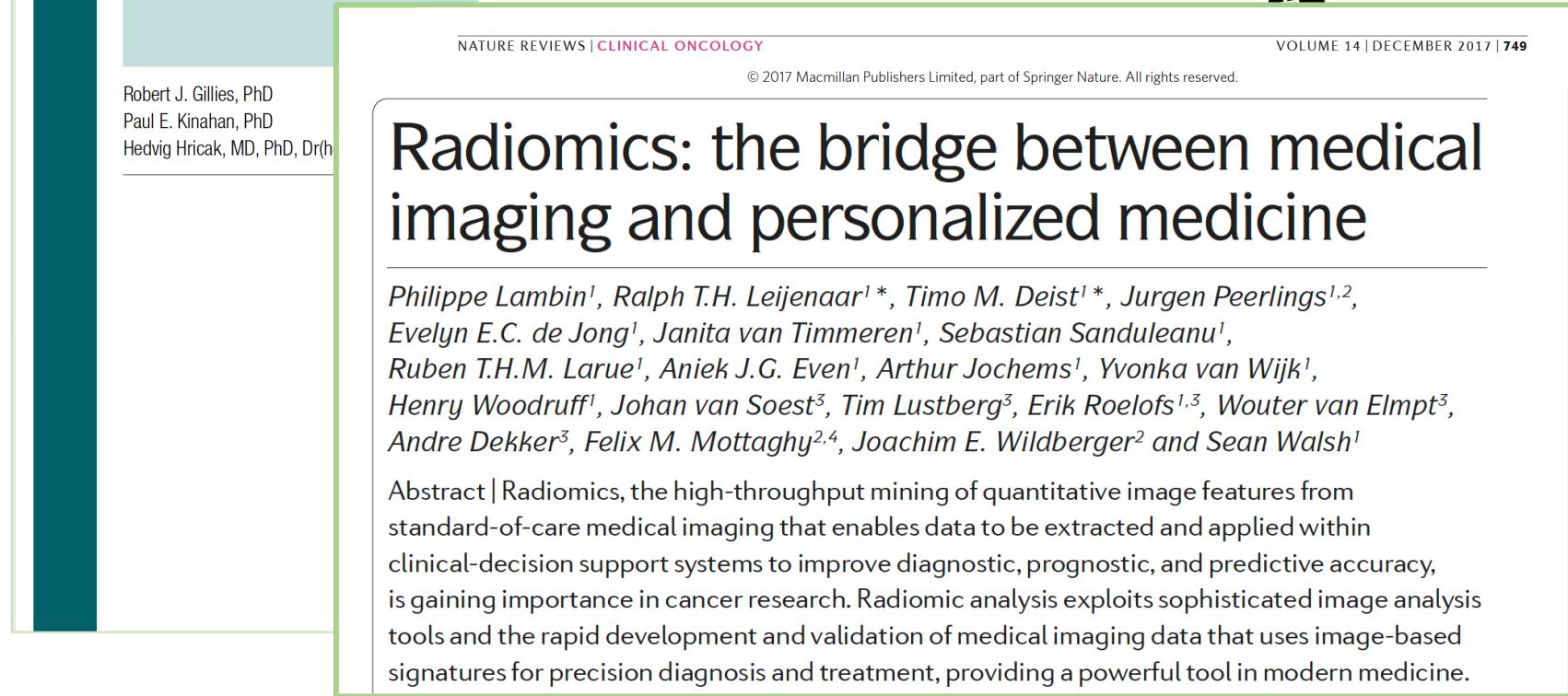


Radiomics

- 2016



- 2017



Radiomics

Radiology: Volume 278: Number 2—February 2016 • radiology.rsna.org

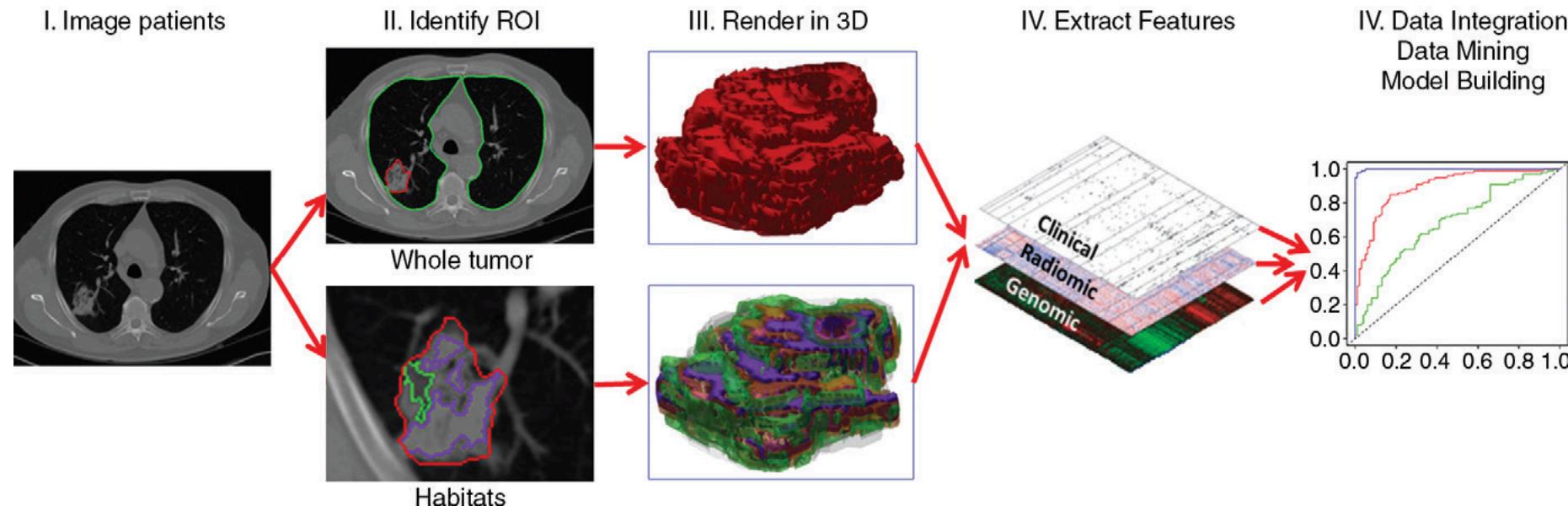


Figure 1: Flowchart shows the process of radiomics and the use of radiomics in decision support. Patient work-up requires information from disparate sources to be combined into a coherent model to describe where the lesion is, what it is, and what it is doing. Radiomics begins with acquisition of high-quality images. From these images, a region of interest (*ROI*) that contains either the whole tumor or subregions (ie, habitats) within the tumor can be identified. These are segmented with operator edits and are eventually rendered in three dimensions (*3D*). Quantitative features are extracted from these rendered volumes to generate a report, which is placed in a database along with other data, such as clinical and genomic data. These data are then mined to develop diagnostic, predictive, or prognostic models for outcomes of interest.

What's new ?

- Radiological features -> quantitative measures to be reported in image annotations
- Integration with clinical and genomic data

PyRadiomics: open-source python package for the extraction of Radiomics features from medical images

- <https://pyradiomics.readthedocs.io/en/latest/>

- Start the python interactive session:

- `python`

- Import the necessary classes:

```
from radiomics import firstorder, glcm, imageoperations, shape, glrlm, glszm, getTestCase
import SimpleITK as sitk
import six
import sys, os
```



- Set up a data directory variable:

```
dataDir = '/path/to/pyradiomics/data'
```

- You will find sample data files brain1_image.nrrd and brain1_label.nrrd in that directory.
- Use SimpleITK to read a the brain image and mask:

```
imageName, maskName = getTestCase('brain1', dataDir)
image = sitk.ReadImage(imageName)
mask = sitk.ReadImage(maskName)
```

- Calculate the first order features:

```
firstOrderFeatures = firstorder.RadiomicsFirstOrder(image,mask)
firstOrderFeatures.enableAllFeatures() # On the feature class level, all features are disabled by default
firstOrderFeatures.calculateFeatures()
for (key,val) in six.iteritems(firstOrderFeatures.featureValues):
    print("\t%s: %s" % (key, val))
```

Radiomic Features

This section contains the definitions of the various features that can be extracted using PyRadiomics. They are subdivided into the following classes:

- `First Order Statistics` (19 features)
- `Shape-based (3D)` (16 features)
- `Shape-based (2D)` (10 features)
- `Gray Level Cooccurrence Matrix` (24 features)
- `Gray Level Run Length Matrix` (16 features)
- `Gray Level Size Zone Matrix` (16 features)
- `Neighbouring Gray Tone Difference Matrix` (5 features)
- `Gray Level Dependence Matrix` (14 features)

All feature classes, with the exception of shape can be calculated on either the original image and/or a derived image, obtained by applying one of several filters. The shape descriptors are independent of gray value, and are extracted from the label mask. If enabled, they are calculated separately of enabled input image types, and listed in the result as if calculated on the original image.

Most features defined below are in compliance with feature definitions as described by the Imaging Biomarker Standardization Initiative (IBSI), which are available in a separate document by Zwanenburg et al. (2016) [1]. Where features differ, a note has been added specifying the difference.

See demo code:

- `Lecture7_demo1_extract_features_pyradiomics.ipynb`

Analysis of image features

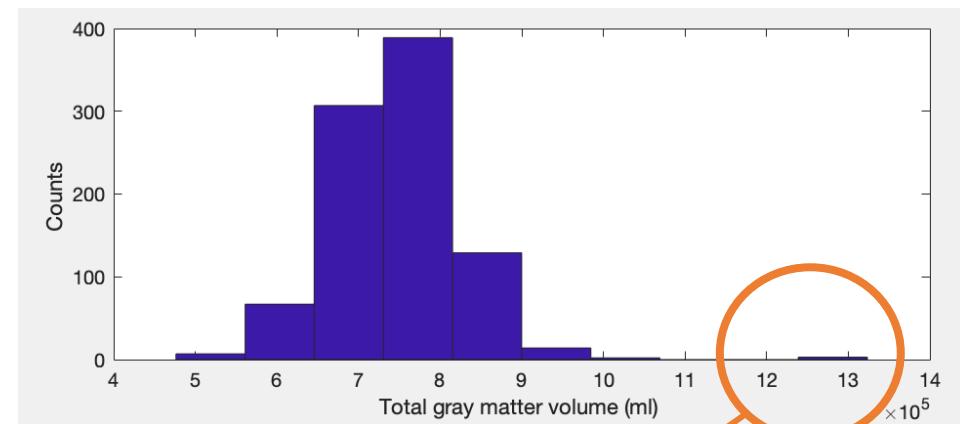
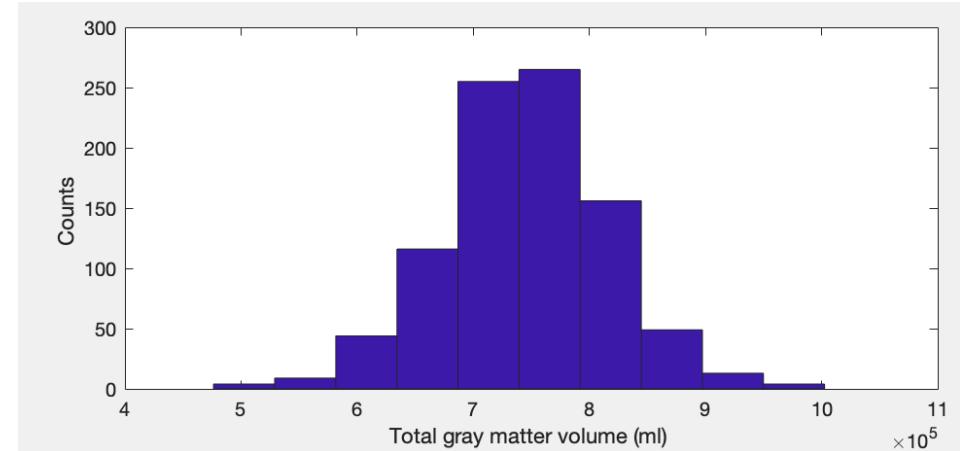
- Showing distributions and computing descriptive statistics
- Within group and between-group analysis:
 - Study of correlations between descriptive features and clinical variables
 - Comparison between two groups of subjects
 - Multiple testing

See demo code:

- Lecture7_demo2_exploring_features mlx
- Lecture7_demo3_analyzing_features mlx

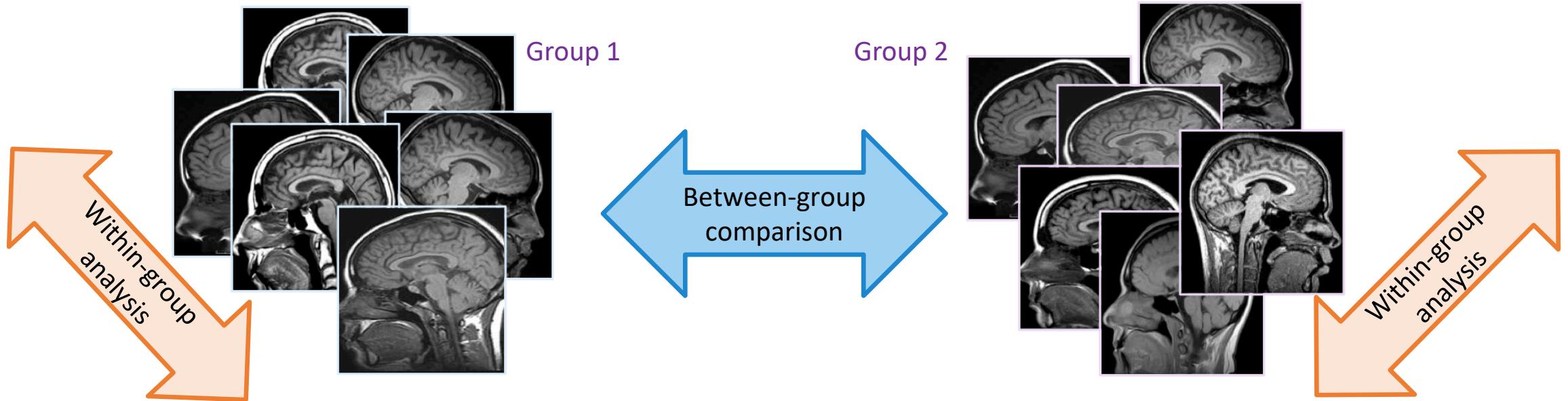
Feature distributions

- The distributions of image descriptive features should be visualized and explored, e.g. by computing:
 - Mean, median, standard deviation, ranges for each group
- Outliers should be searched for
 - Non-necessarily they have to be excluded
 - They can provide evidence that something has gone wrong in preprocessing/feature generation steps



Outliers of the distribution + unrealistic values
→ these subjects have to be excluded from the analysis

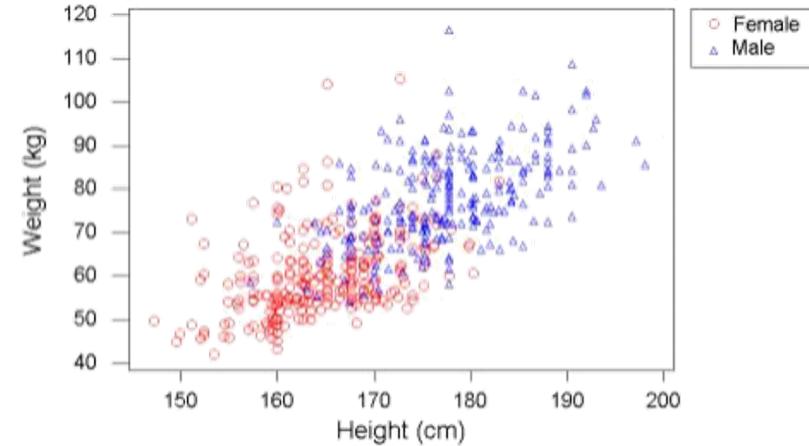
Statistical analysis of image features



- Within-group analysis: Study of the correlations between image features and other demographic/clinical characteristics of subjects
- Between-group analysis: Identification of the statistically significant different features between the two groups (case-control, longitudinal, etc.)
- Statistical analysis of descriptive features (voxel-wise/regional features/connectivity maps, etc.), both within-group and between-group require to carry out hundreds/thousands of statistical tests (e.g. Pearson correlation, two-sample t-test) on non-independent variables (hundreds/thousands of features): In multiple testing, the p value should be corrected for multiple comparisons

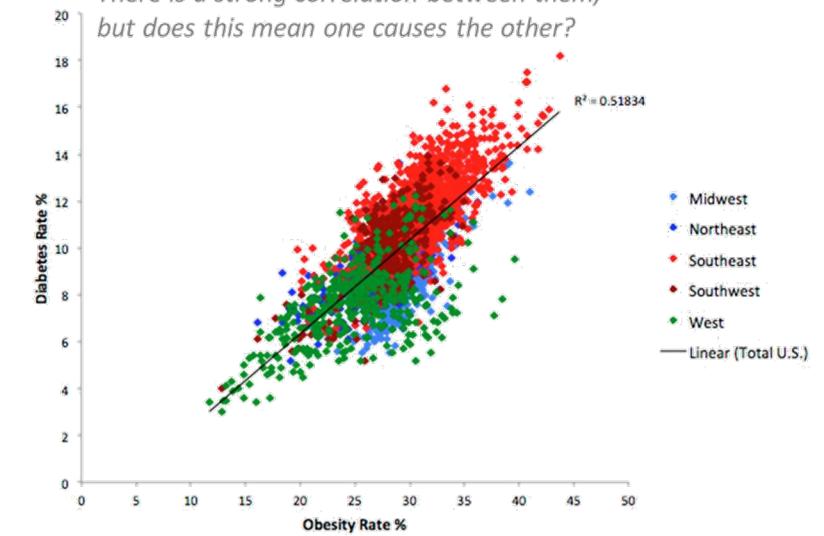
Study of the correlations between variables

- In statistics, dependence or association is any statistical relationship, whether causal or not, between two random variables.
- Two or more variables are related if, in a sample of observations, the values of those variables are distributed in a consistent manner.
- The presence of a correlation is not sufficient to infer the presence of a causal relationship. Data from correlational research can only be "interpreted" in causal terms based on some theories that we have, but correlational data cannot conclusively prove causality.



Diabetes and obesity are '**risk factors**' of each other.

*There is a strong correlation between them,
but does this mean one causes the other?*



<http://diabetes-obesity.findthedata.org/b/240/Correlations-between-diabetes-obesity-and-physical-activity>

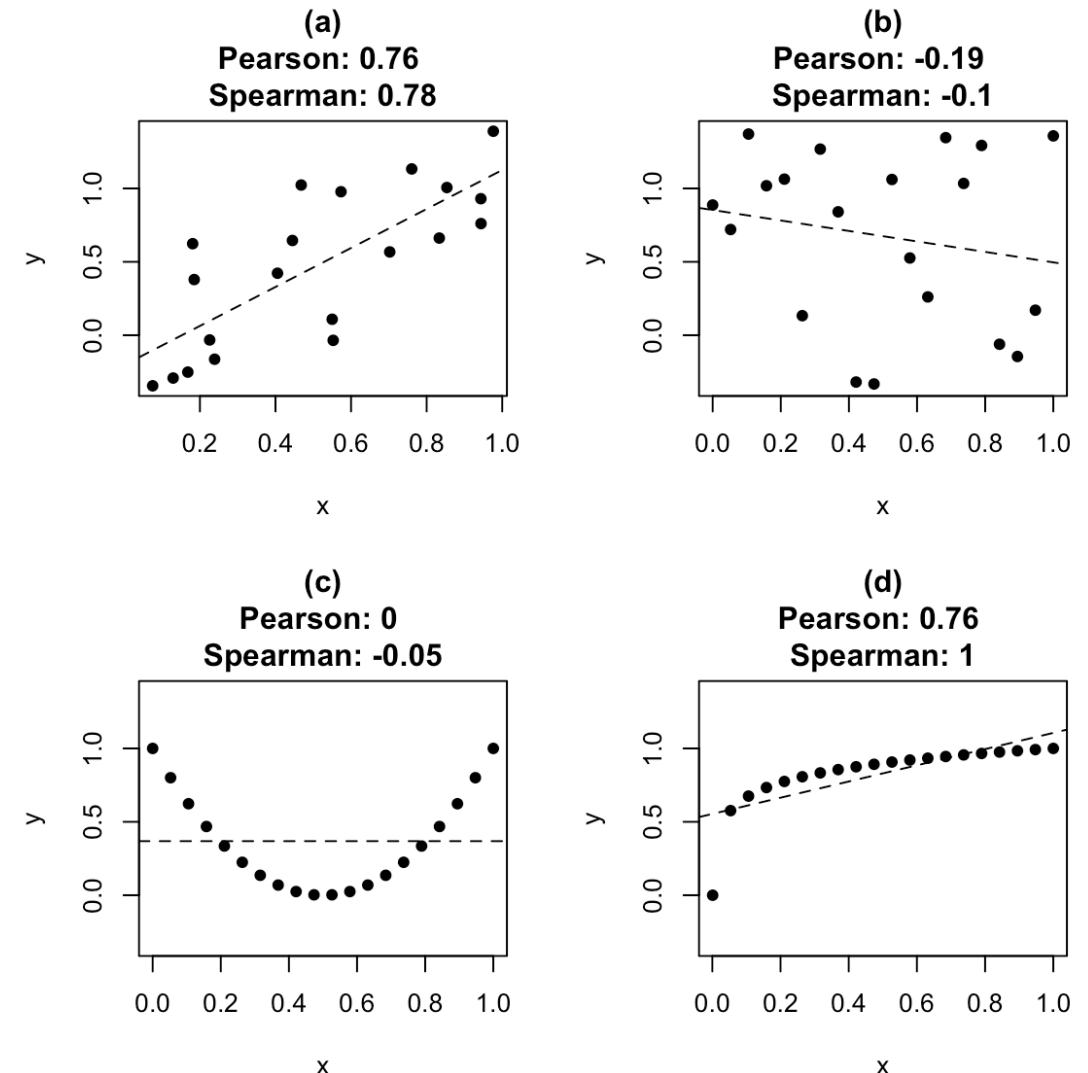
Study of the correlations between variables

- Pearson's correlation coefficient: it is a measure of the linear correlation between two variables X and Y

$$\rho_{X,Y} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y} = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_j (x_j - \bar{x})^2 \sum_k (y_k - \bar{y})^2}}$$

The covariance is a measure of the joint variability of two random variables

- Spearman's rank correlation coefficient: it assesses how well the relationship between two variables can be described using a monotonic function. It is defined as the Pearson correlation coefficient between the rank variables.



Comparing two groups for statistical differences

To test if two population means are equal the **two-sample Student's t-test** can be performed

- It is based on Student's t-statistic, which assumes that variable is normally distributed and mean is known and population variance is calculated from the sample.
- In the t-test, the null hypothesis takes the form of

$$H_0: \mu(x) = \mu(y)$$

against alternative hypothesis

$$H_1: \mu(x) \neq \mu(y),$$

wherein $\mu(x)$ and $\mu(y)$ represents the population means.

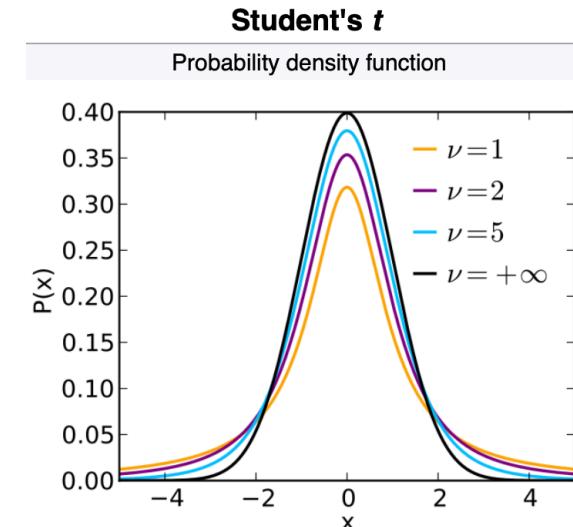
The degree of freedom of t-test is $n_1 + n_2 - 2$

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

t-test test statistics

$$s^2 = \frac{\sum_{i=1}^{n_1} (x_i - \bar{x}_1)^2 + \sum_{j=1}^{n_2} (x_j - \bar{x}_2)^2}{n_1 + n_2 - 2}$$

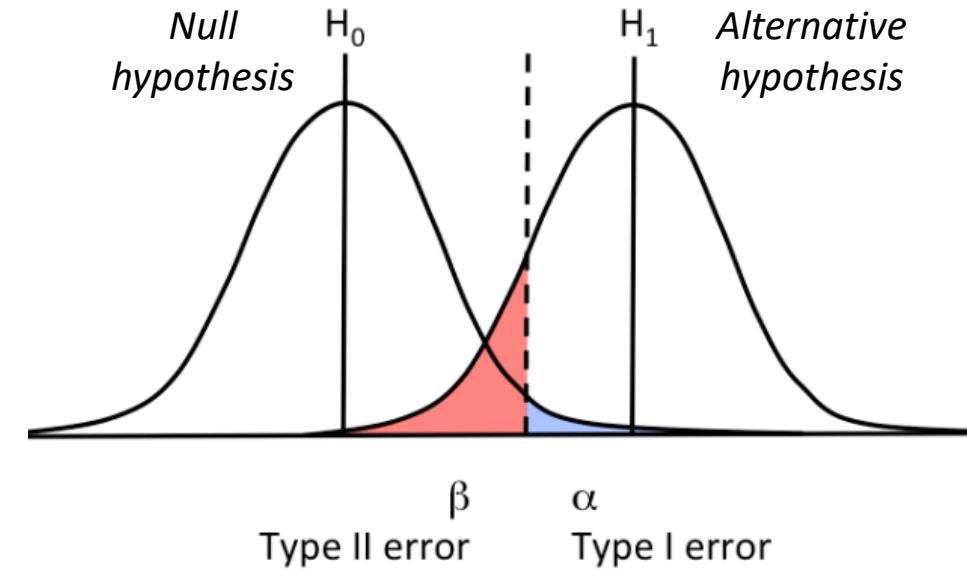
s² is the pooled sample variance



Once the t value and degrees of freedom are determined, a p-value can be found using a table of values from Student's t-distribution. If the p-value is below the threshold chosen for statistical significance, the null hypothesis is rejected in favor of the alternative hypothesis.

Statistical test significance: p value

- The statistical significance of a result is the probability that the observed relationship (e.g. between variables) or a difference (e.g. between means) in a sample occurred by pure chance, and that in the real population, no such relationship or differences exist.
- The p-value represents an index of the reliability of a result.
- Typically, results that yield $p=0.05$ are considered borderline statistically significant; this level of significance still involves a pretty high probability of Type I error (5%).



		Reality	
		Positive	Negative
Study Finding	Positive	True Positive (Power) ($1-\beta$)	False Positive Type I Error (α)
	Negative	False Negative Type II Error (β)	True Negative

Type I error:
incorrect rejection of
the null hypothesis

Type II error: non-
rejection of a false
null hypothesis

Multiple testing: corrected p-values

- The more analyses we perform on a data set, the more results will meet "by chance" the conventional significance level ($p \leq 0.05$):
 - if we carry out 100 statistical tests, by chance 5 will come out to be significant ($p < 0.05$) by chance.
- When we carry out many comparisons, we have to include some "correction" or adjustment for the total number of comparisons:
 - **Bonferroni correction** (very stringent): the conventional cutoff of significant level ($p=0.05$) is rescaled by the number of statistical tests, as $p_{\text{cutoff}} = 0.05/N_{\text{tests}}$
 - **False Discovery Rate (FDR)** approach is less stringent than Bonferroni: it is designed to control the expected proportion of "discoveries" (rejected null hypotheses) that are false (incorrect rejections).

References and sources

- Books
 - <http://www.statsoft.com/Textbook>
 - <http://www.statsoft.com/Textbook/General-Linear-Models>
- Sources
 - <https://www.statisticshowto.datasciencecentral.com>
 - <https://it.mathworks.com/help/bioinfo/ref/mafdr.html>
- See also
 - www.fil.ion.ucl.ac.uk/spm/
 - <http://freesurfer.net>
 - <https://pyradiomics.readthedocs.io>