Laboratorio di Informatica Applicata ai Servizi Ospedalieri

Hands-on on medical data import and visualization

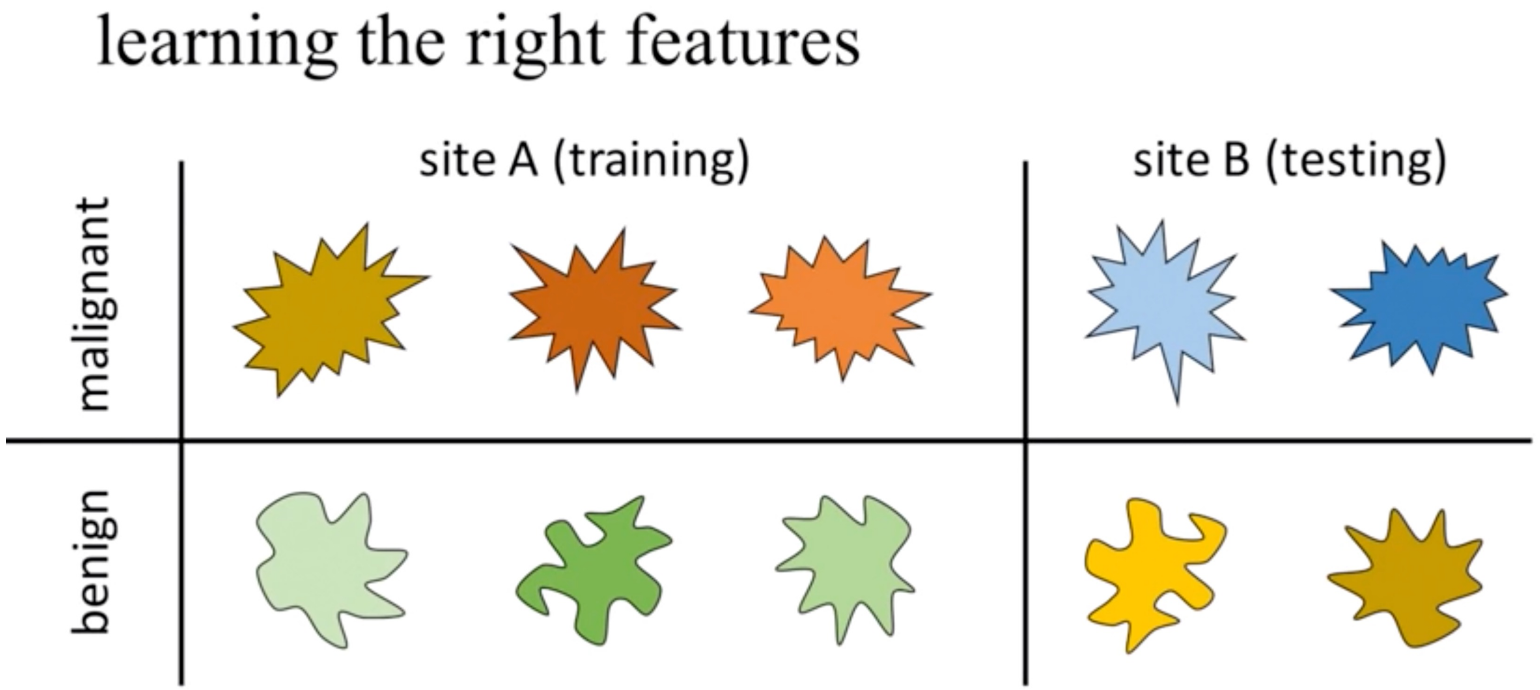
Alessandra Retico
alessandra.retico@pi.infn.it
INFN - Pisa

Todays' objectives

- Explore the use of predictive models in medical data analysis
  - Supervised classification with Machine Learning
    - Features-to-label conversion → **ML_classification.ipynb**
  - Supervised classification with CL
    - Image-to-labels conversion. → **DL_classification.ipynb**
  - Image segmentation with DL
    - Image-to-image conversion → **DL_segmentation.ipynb**

- Quantification of model performance
  - Figures of merit: Sensitivity, Specificity, AUC , Precision, Recall, F1
- Estimate of model robustness
  - Cross validation methods: k-fold cross validation, leave-one out cross validation

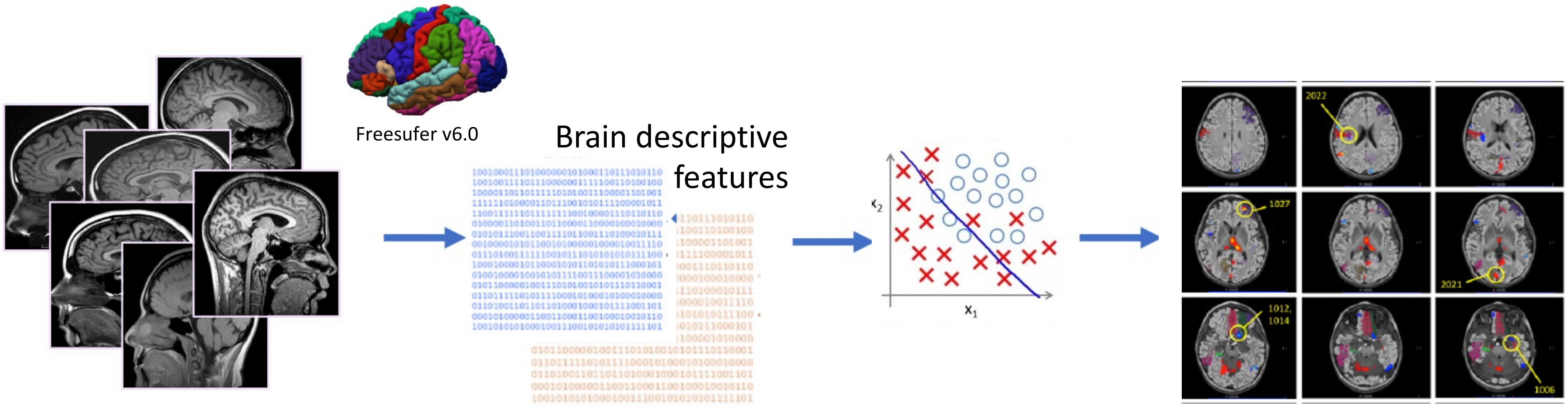
Learning the right information: confounders



- If we obtain good discrimination performance between malignant and benign masses are we sure the classifier is exploiting the right mass properties?
- A classifier trained on data from site A which learnt to distinguish masses according to color tones, will not work on data from site B.

All possible confounder variables should be accounted for in the analysis.
Classification results should be cross checked.

Analysis of MRI data: confounding parameters



- **Confounding variables** introduce bias in the classifier training phase, suggesting correlations that in fact are not there.

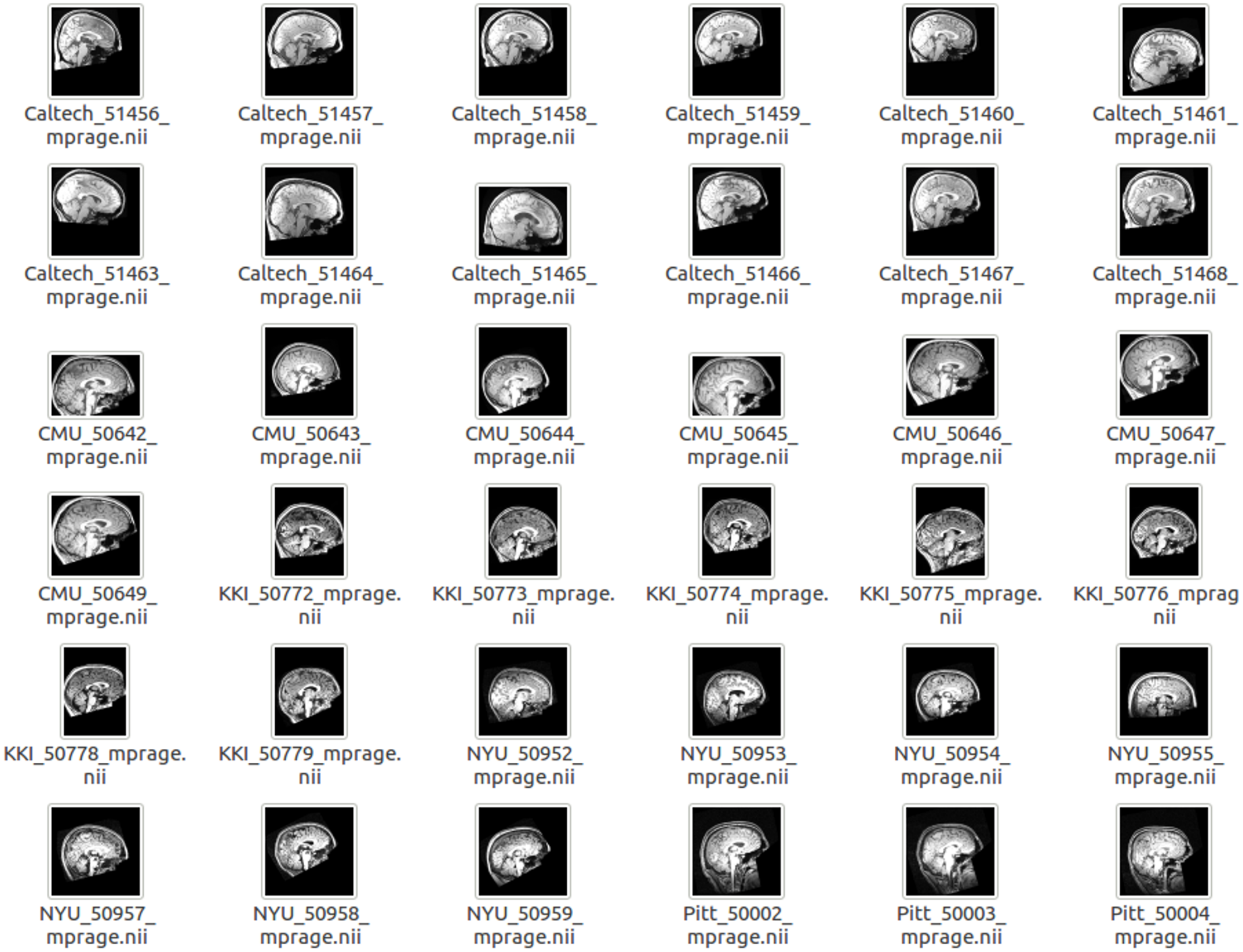
- Biases introduced by the MRI **acquisition site** strongly affect the classification results

See demo code **ML_classification.ipynb**

Multicenter MRI datasets: the ABIDE sample

Data gathered by different sites and/or acquisition systems carries local “fingerprint”, which can hide subtle information of interest.

This problem is similar to the management of **systematic errors**



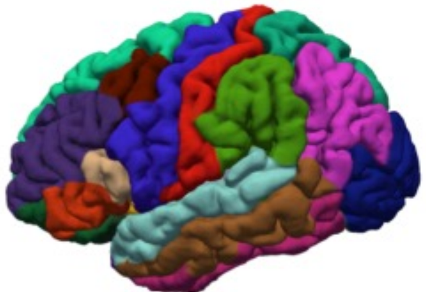
Autism Brain Imaging Data Exchange

ABIDE

NITRC NeuroImaging Tools & Resources Collaboratory

http://fcon_1000.projects.nitrc.org/indi/abide

Site dependence of sMRI data



volume and thickness of 62
brain parcels for each subject

- ABIDE healthy subjects
- $Site_i$ vs. $Site_j$ binary classification

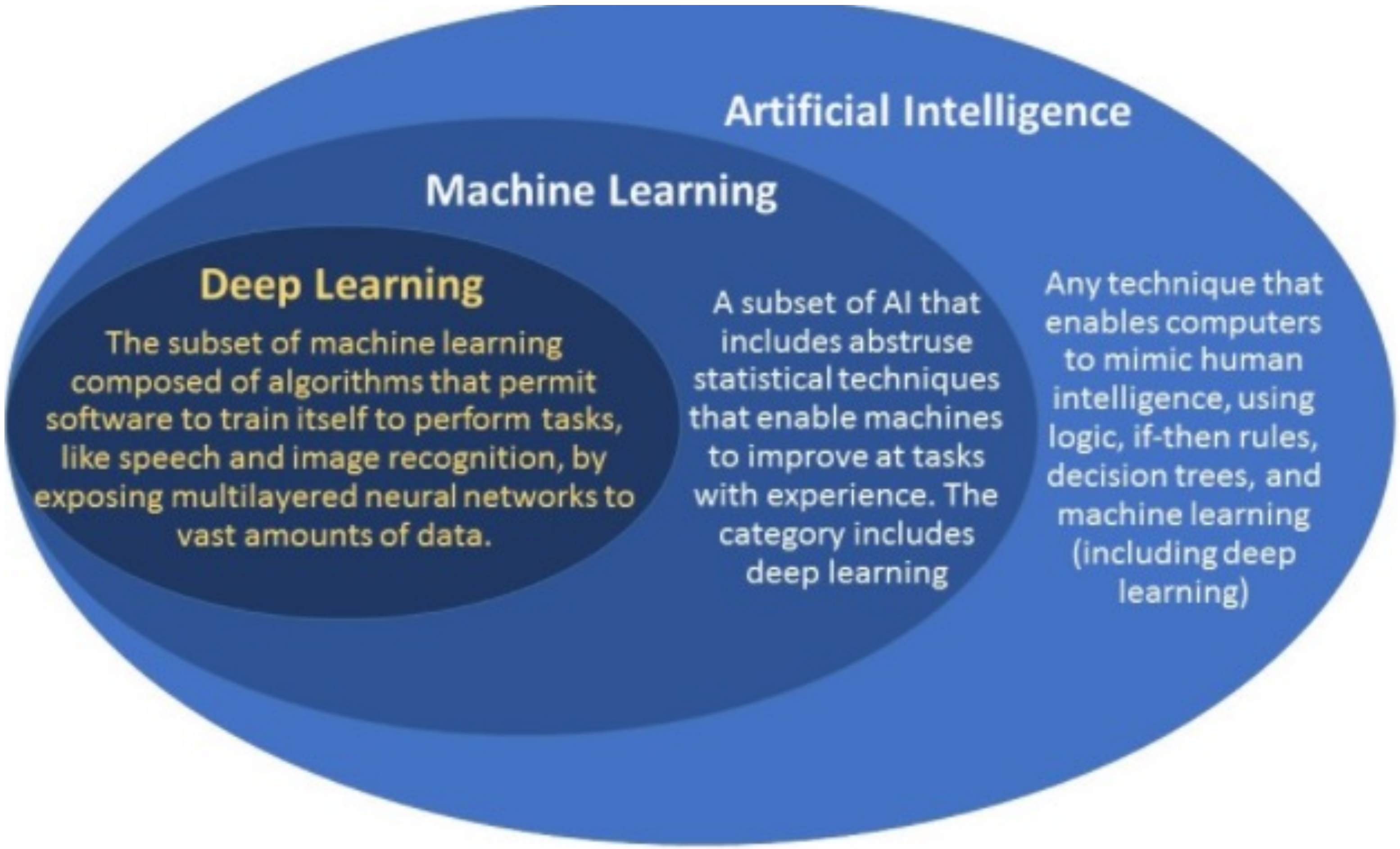
*E. Ferrari et al., "Dealing with confounders and outliers in classification medical studies: the Autism Spectrum Disorders case study", **AIIM** 108:101926, 2020*

Areas Under the ROC Curve (AUC) obtained in two-class classification

	NYU ABIDE1	NYU-1 ABIDE2	NYU-2 ABIDE2	OHSU ABIDE1	OHSU ABIDE2	USM ABIDE1	USM ABIDE2	UM-1 ABIDE1	UM-2 ABIDE1
NYU ABIDE1	-	0.78	0.89	0.99	1.00	0.99	1.00	0.99	0.98
NYU-1 ABIDE2		-	0.70	0.99	1.00	1.00	1.00	0.99	0.98
NYU-2 ABIDE2			-	1.00	0.98	0.99	0.99	1.00	1.00
OHSU ABIDE1				-	0.63	0.97	0.96	1.00	1.00
OHSU ABIDE2					-	0.99	0.96	0.98	0.98
USM ABIDE1						-	0.75	0.99	0.99
USM ABIDE2							-	0.97	0.97
UM-1 ABIDE1								-	0.96
UM-2 ABIDE1									-

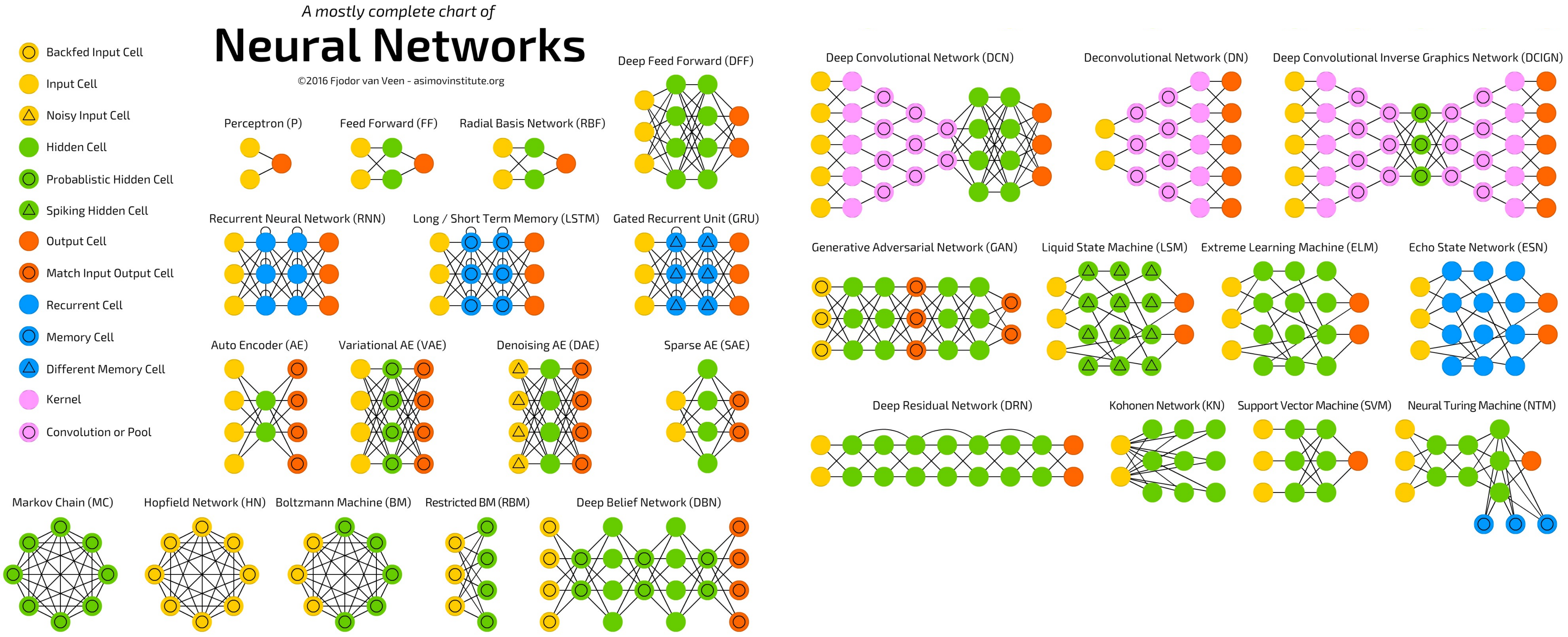
How can we eliminate/mitigate the bias due to data acquisition information?

Artificial Intelligence, Machine Learning, Deep Learning



https://www.deeplearningitalia.com/una-panoramica-introduttiva-su-deep-learning-e-machine-learning/

Neural network Zoo

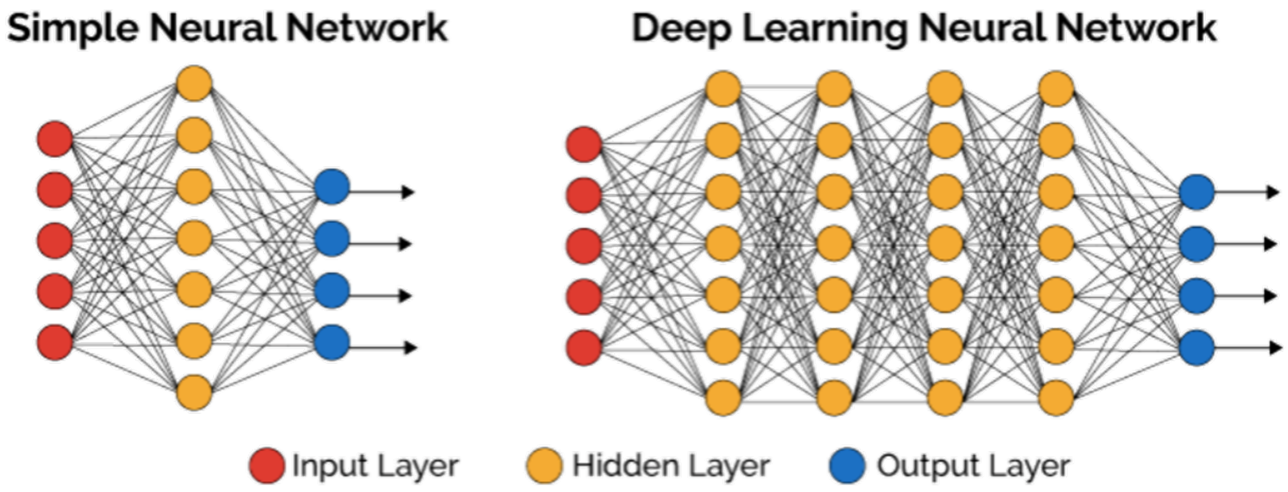


http://www.asimovinstitute.org/neural-network-zoo/

Deep Neural Networks

Deep Learning (DL) means using a neural network with several layers of nodes between input and output

DL models are a family of parametric models which learn non-linear hierarchical representations



$$a_L(\mathbf{x};\mathbf{\Theta}) = h_L(h_{L-1}(...(h_1(\mathbf{x},\boldsymbol{\theta}_1),\boldsymbol{\theta}_{L-1}),\boldsymbol{\theta}_L)$$

input

parameters of the network

non-linear activation function

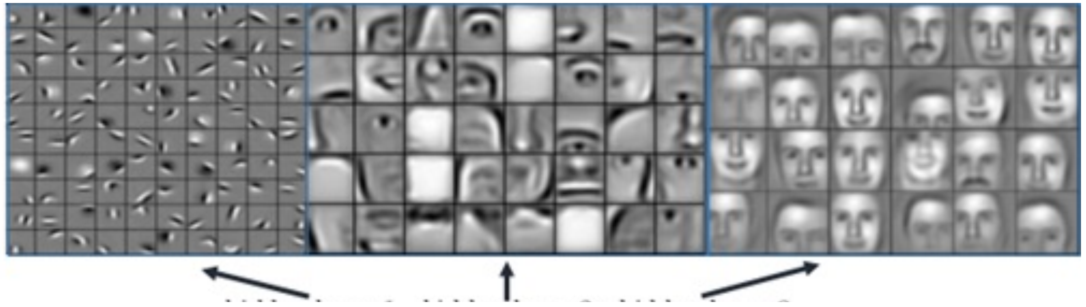
parameters of layer *L*

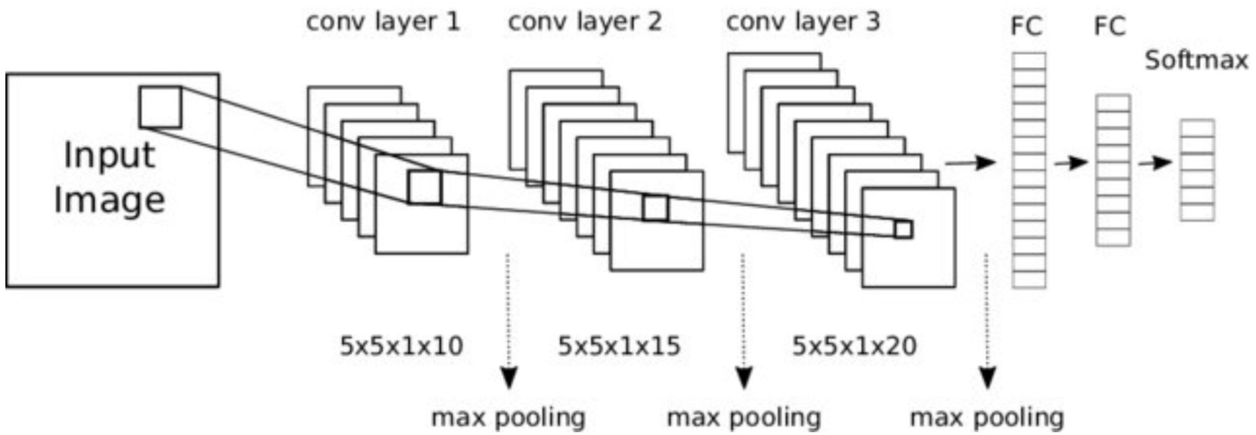
Deep Neural Networks

- Deep neural networks are generally better than other ML methods on images
- The series of layers between input and output compute relevant features automatically in a series of stages, just as our brains seem to.

Deep neural networks learn in multiple levels of representation, corresponding to different levels of abstraction

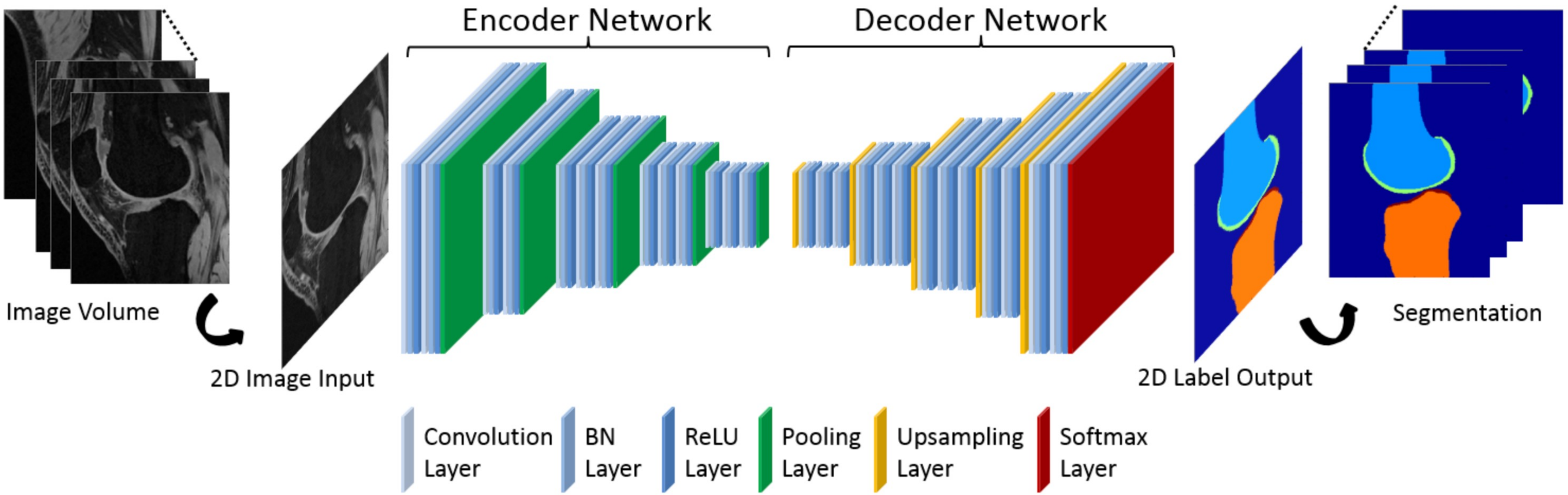
https://towardsdatascience.com



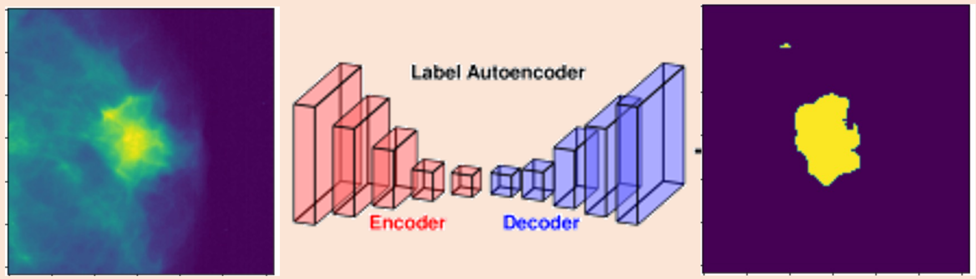


See demo code **DL_classification.ipynb**

Convolutional Auto-Encoders for image segmentation



Liu et al, Deep Convolutional Auto-Encoder and 3D Deformable Approach for Tissue Segmentation in Magnetic Resonance Imaging, Proc. Intl. Soc. Mag. Reson. Med. 25, 2017



See demo code **DL_segmentation.ipynb**