Istituto Nazionale di Fisica Nucleare

Dipartimento Interateneo di Fisica Michelangelo Merlin

Università degli Studi di Bari Aldo Moro

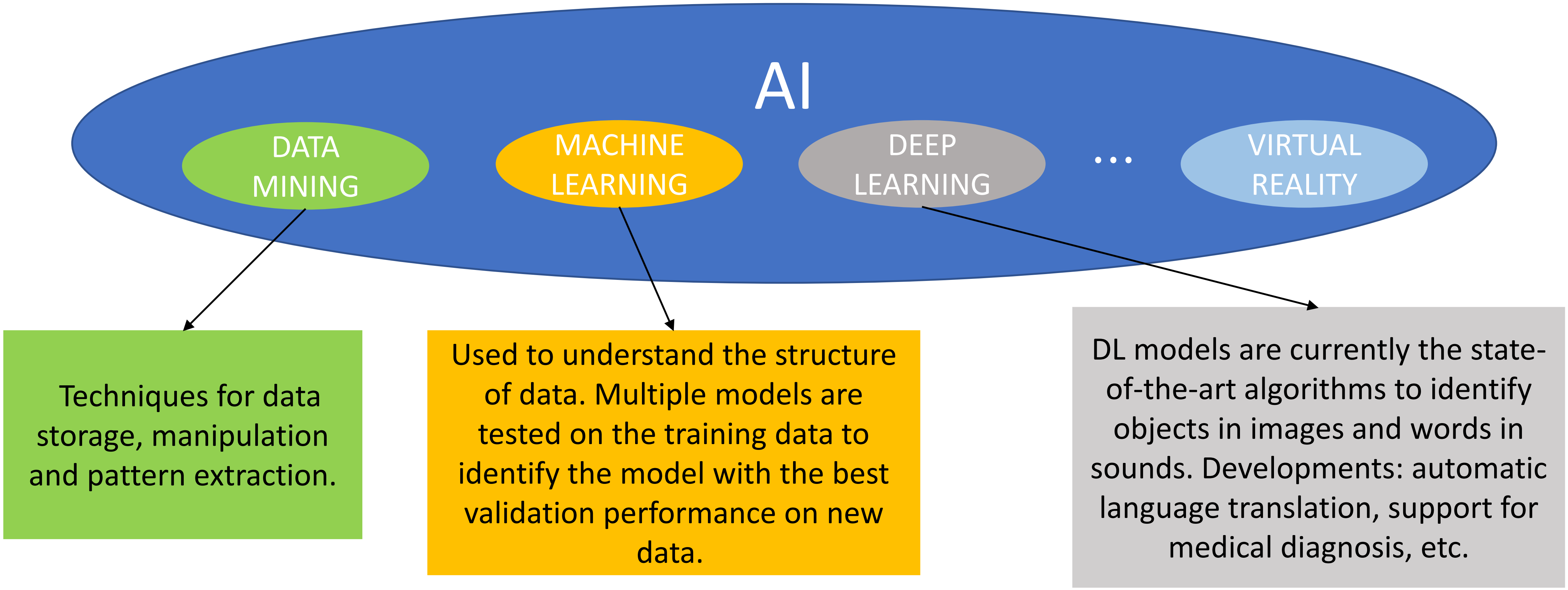
AI Applications in ReCaS Datacenter

Angela Lombardi, Nicola Amoroso, Domenico Diacono, Alfonso Monaco, Sabina Tangaro and Roberto Bellotti

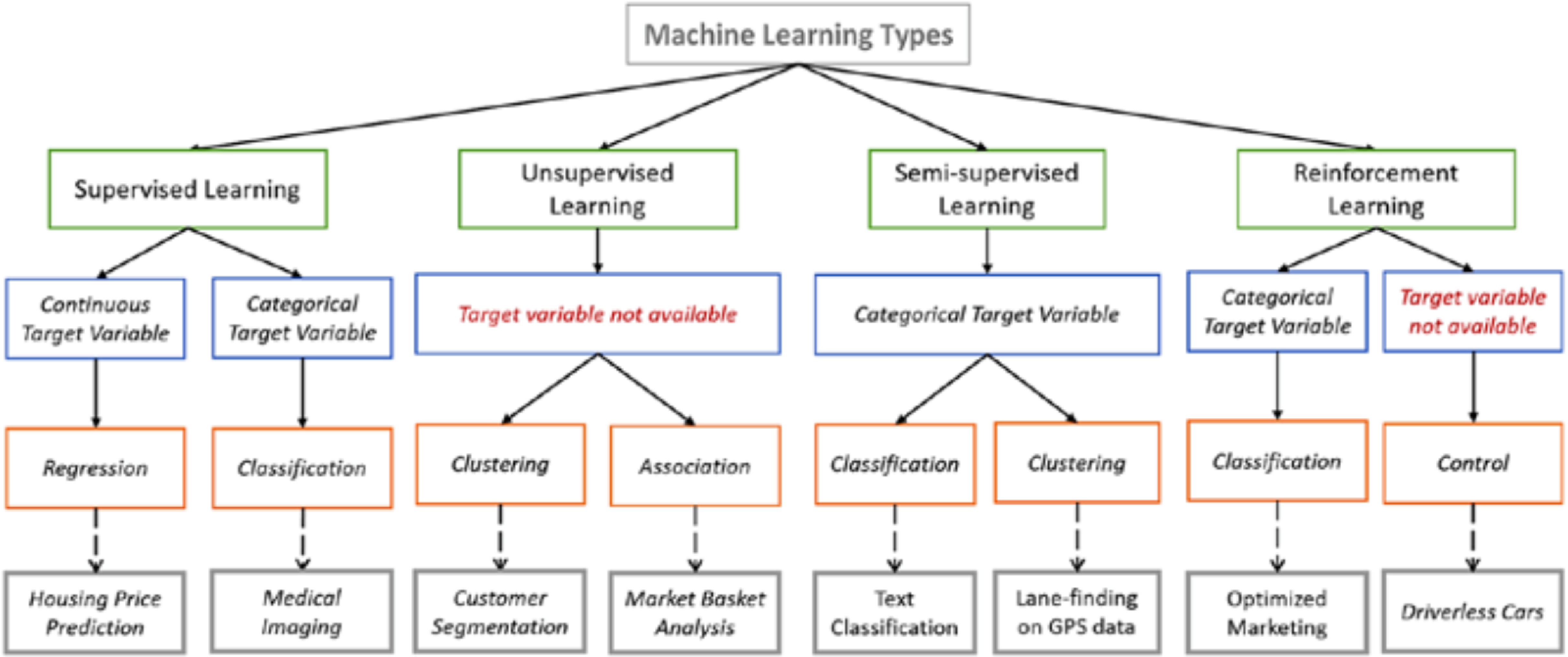
angela.lombardi@ba.infn.it

ReCaS Bari

Artificial Intelligence



Although all these methods have the common goal of extracting information, patterns and relationships that can be used to make decisions, they have different approaches and functionalities.


Machine Learning Types
Supervised Learning
Unsupervised Learning
Semi-supervised Learning
Reinforcement Learning
Continuous Target Variable
Categorical Target Variable
Target variable not available
Categorical Target Variable
Categorical Target Variable
Target variable not available
Regression
Classification
Clustering
Association
Classification
Clustering
Classification
Control
Housing Price Prediction
Medical Imaging
Customer Segmentation
Market Basket Analysis
Text Classification
Lane-finding on GPS data
Optimized Marketing
Driverless Cars

Supervised machine learning: The program is “trained” on a pre-defined set of “training examples”, which then enable its ability to reach an accurate conclusion when given new data. In other words it is the machine learning task of learning a function that maps an input to an output based on example input-output pairs.

The data entries ("istances") you feed to the algorithm for it to learn (through its "attributes", istantiated in "features", in a process called "training") include the truth info ("labels")

Classification

Attributes			Target
X	Y	Z	Class
1	2	2	A
1	2	3	B
1	4	3	B
4	2	6	A

Regression

Attributes			Target
X	Y	Z	Variable
1	2	2	1.3
1	2	3	2.1
1	4	3	4.93
4	2	6	5

Supervised ML: Classification vs Regression

A typical supervised learning task is **classification**

- predict “classes”: binary (0/1, yes/no) or multi-class (A/B/C/D)
- e.g. spam filter: trained with many example emails along with their class labels (“spam” or “not-spam”), it learns how to classify new coming emails

Another typical supervised learning task is **regression**

- predict “target numeric values” (in a continuum of values)
- e.g. a price of a house, knowing its attributes, and being given plenty of instances of other houses (both their features and price)

Note: regression algorithms exist that can be used for classification as well, and vice versa

- e.g. Logistic Regression is commonly used for classification, as it can output a value that corresponds to the probability of belonging to a given class (e.g., 20% chance of being spam)

Unsupervised machine learning: The program is given a bunch of data and must find patterns and relationships therein. It learns from test data that has not been labeled, classified or categorized.

Clustering

- Definition

Machine Learning method of grouping a set of objects in such a way that objects in the same group (called a **cluster**) are more similar (in some sense) to each other than to those in other groups (**clusters**).

ID	X	Y	Z
1	1	2	2
2	4	5	6
3	1	1	2
4	3	4	5
5	1	2	3

Clustering →

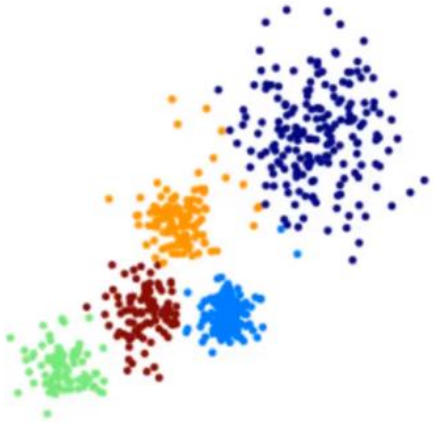
ID	X	Y	Z
1	1	2	2
5	1	2	3
3	1	1	2
4	3	4	5
2	4	5	6

K Means Clustering is an unsupervised learning algorithm that will attempt to group similar clusters together in your data.

So what does a typical clustering problem look like?

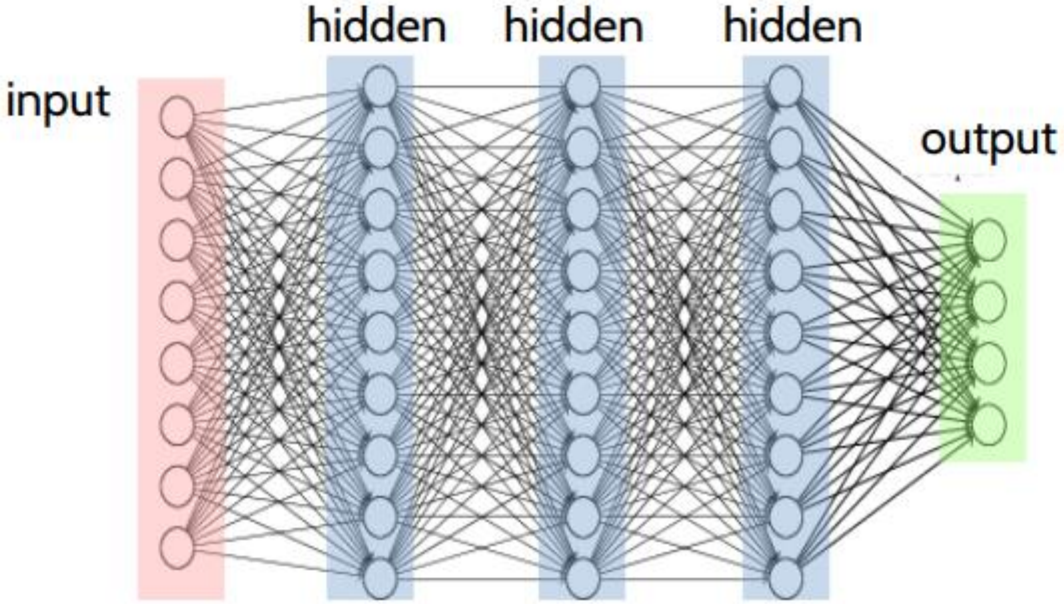
- Cluster Similar Documents
- Cluster Customers based on Features
- Market Segmentation
- Identify similar physical groups

- The overall goal is to divide data into distinct groups such that observations within each group are similar



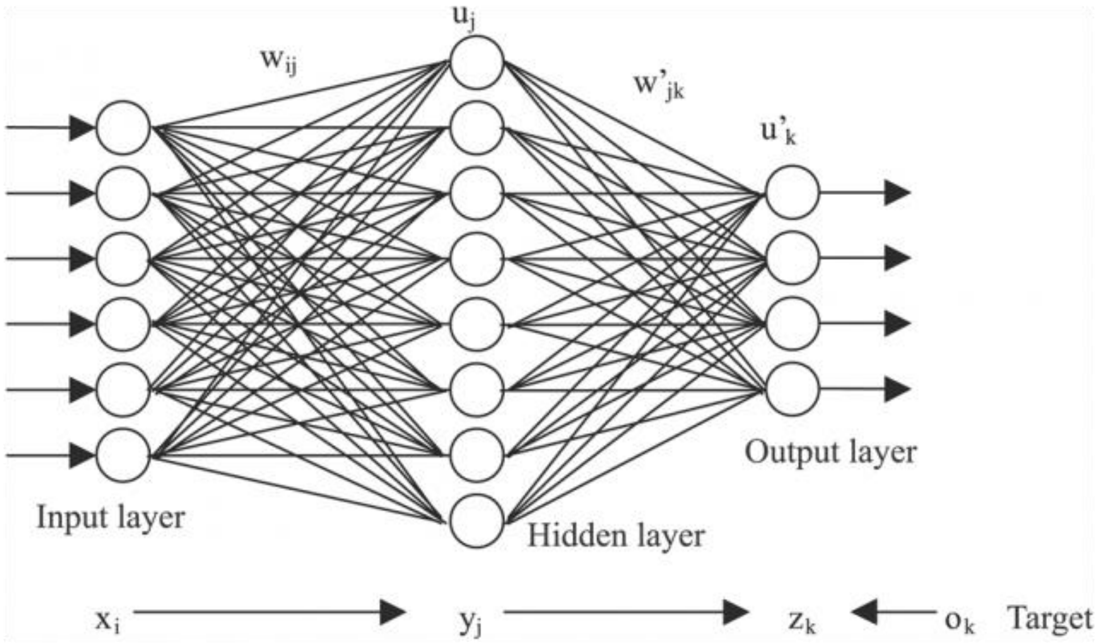
What is deep learning?

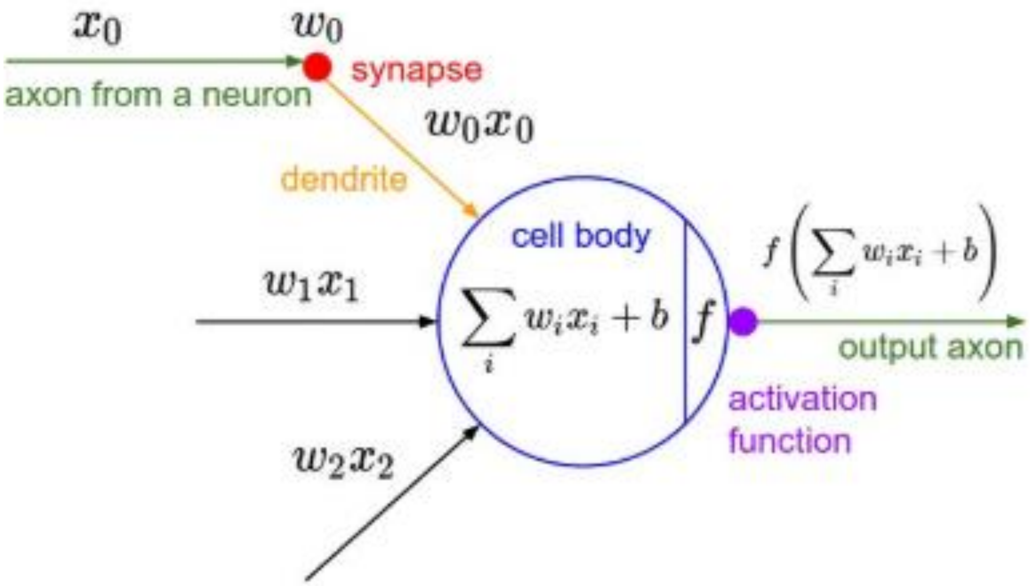
*Deep Learning means using a **neural network** with **several layers of nodes** between input and output*



Feedforward Neural Networks

Many applications of deep learning use feedforward neural network architectures, which learn to map a fixed-size input (for example an image) to a fixed-size output (for example a probability for each of several categories).





To go from one layer to the next, a set of units compute a weighted sum of their inputs from the previous layer, and pass the result through a non linear function.

Computational Neuroscience

Recent advances in neuroimaging analysis supported by:

1. high- throughput analysis of large amounts of data. In this field standard approaches for data acquisition and analysis are required to increase global collaboration;

2. big data open sharing brain imaging repositories: Alzheimer's Disease Neuroimaging Initiative (ADNI), Autism Brain Imaging Data Exchange (ABIDE);

3. huge computational resources such as gpu hardware, grid infrastructure and network bandwidth required to support parallelization and distribution of efficient computational and communication resources.

Multiple analysis pipelines

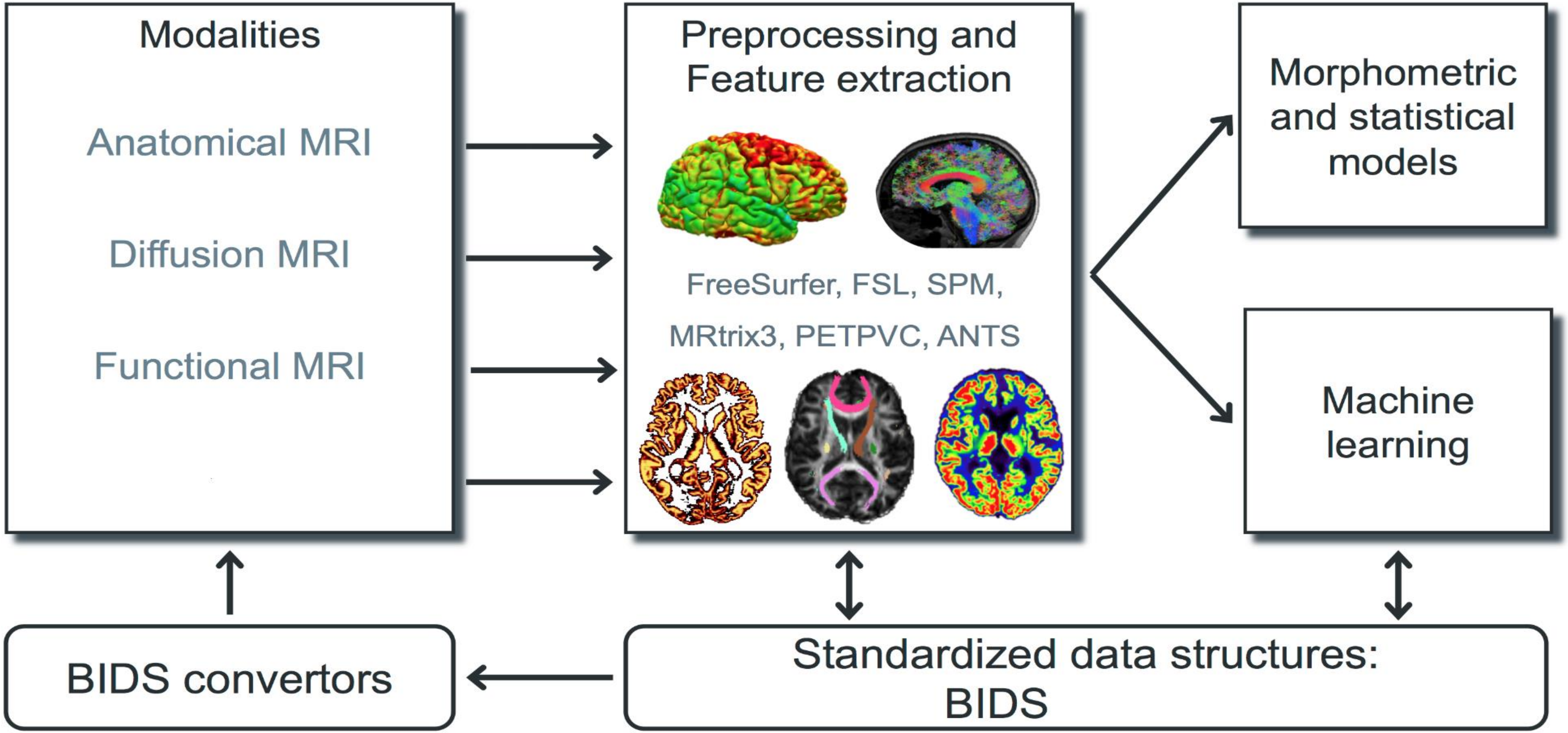
These resources facilitate the processing of large amounts of data and allow a more efficient analysis of the resulting imaging datasets.

These resources introduce a considerable complexity of management and demand of experienced users.

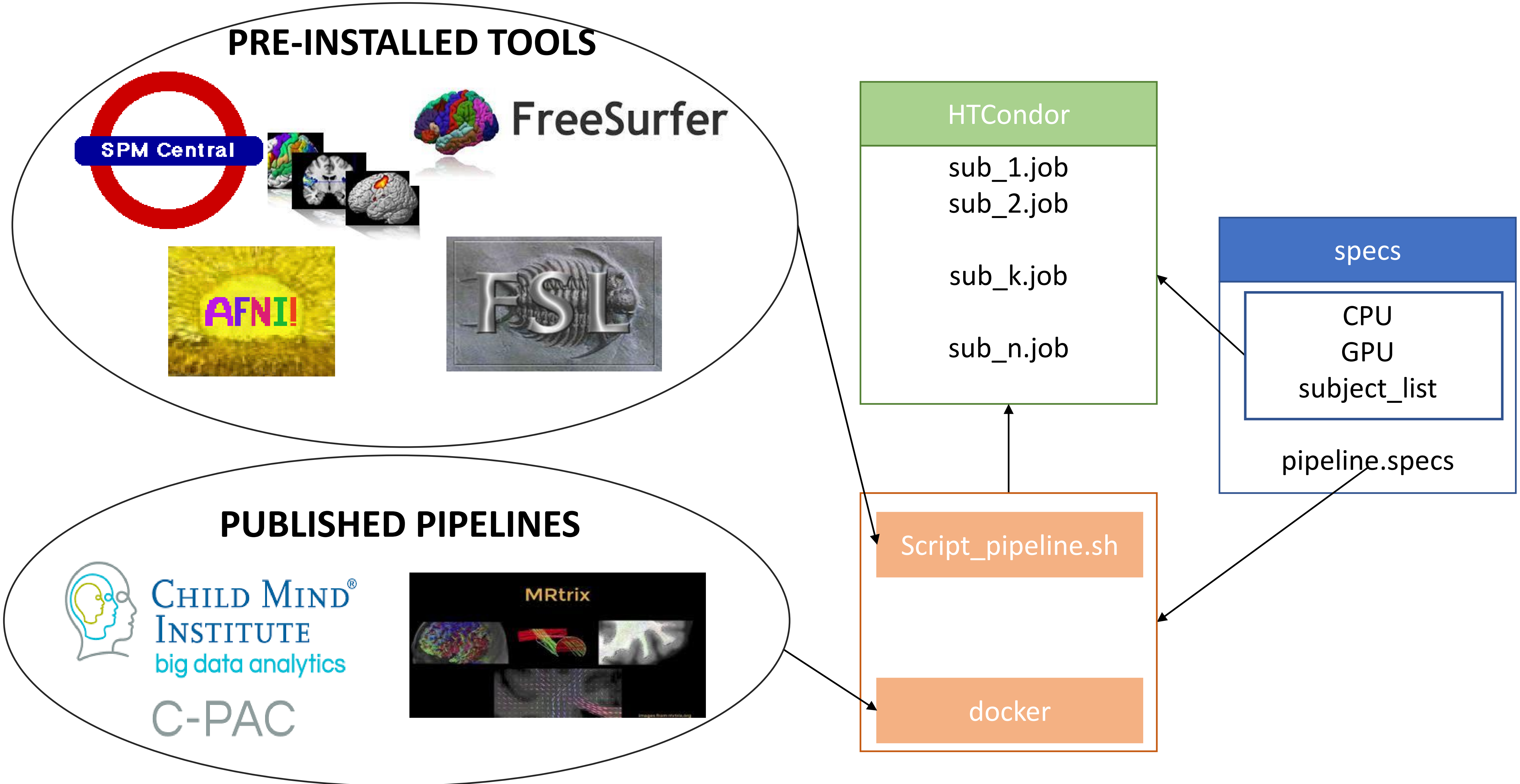
Neuroscience pipelines that:

a) formalize the main steps to be taken to achieve high-quality results.
b) multiple analysis pipelines can produce signicantly different results, raising questions about the replicability and reliability of scientific findings.
c) explore the impact of processing decisions with a different set of preprocessing options.

Multidimensional neuroimaging processing

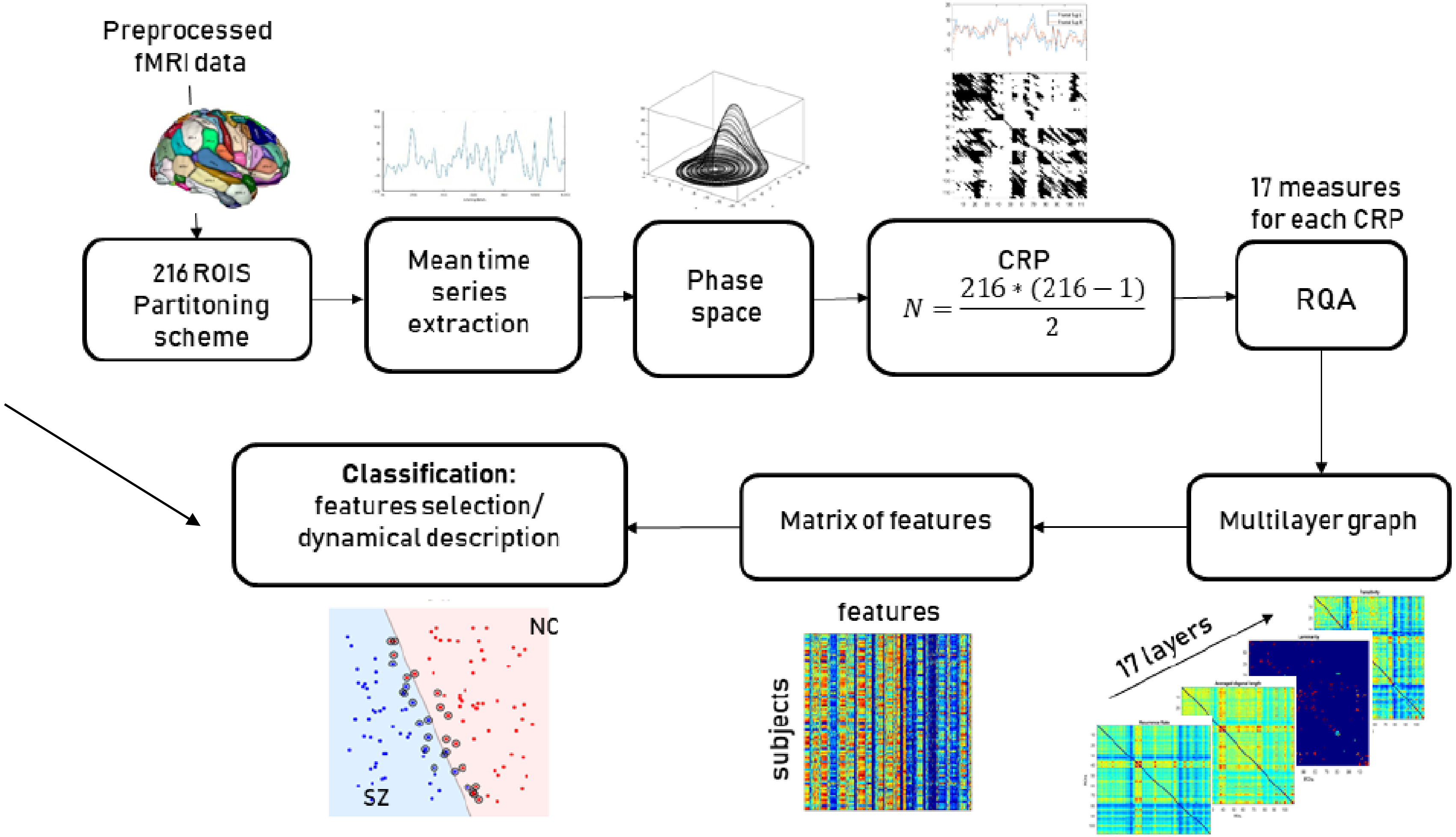


Multidimensional neuroimaging processing



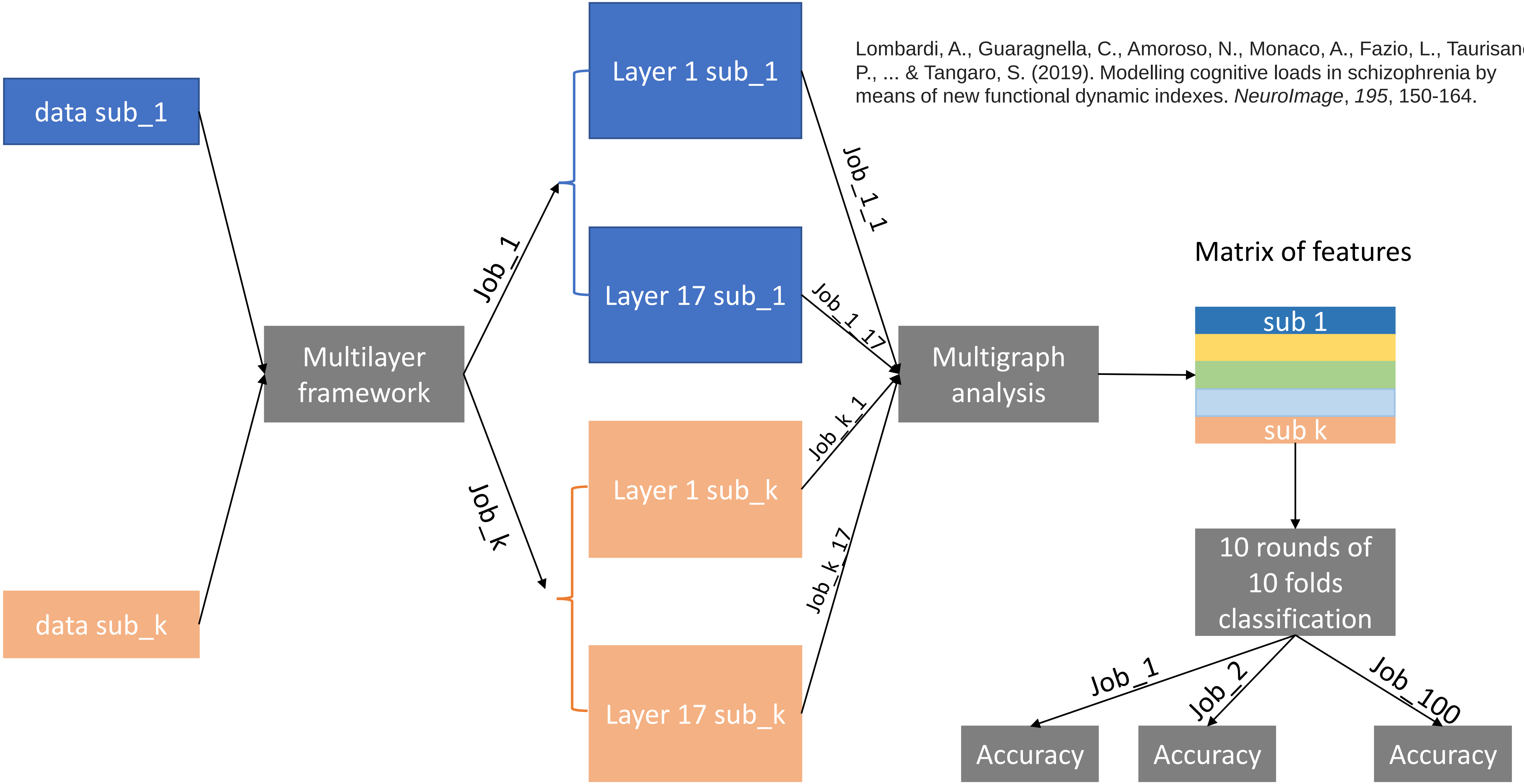
Example: fMRI for Schizophrenia classification

fMRI data of 49 participants and 42 patients with schizophrenia performing 2-back and 0-back experiments were used to dense task-evoked networks.



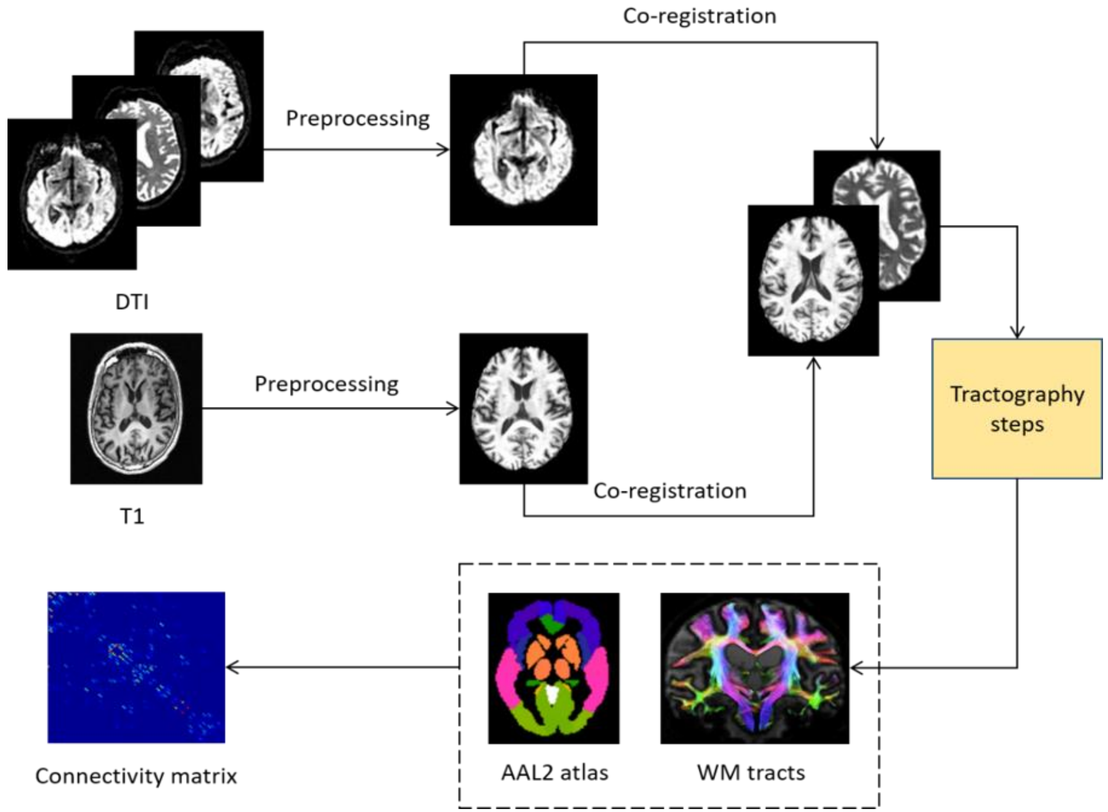
Example:

Lombardi, A., Guaragnella, C., Amoroso, N., Monaco, A., Fazio, L., Taurisano, P., ... & Tangaro, S. (2019). Modelling cognitive loads in schizophrenia by means of new functional dynamic indexes. *NeuroImage*, *195*, 150-164.



Example: DWI images for Alzheimer's disease

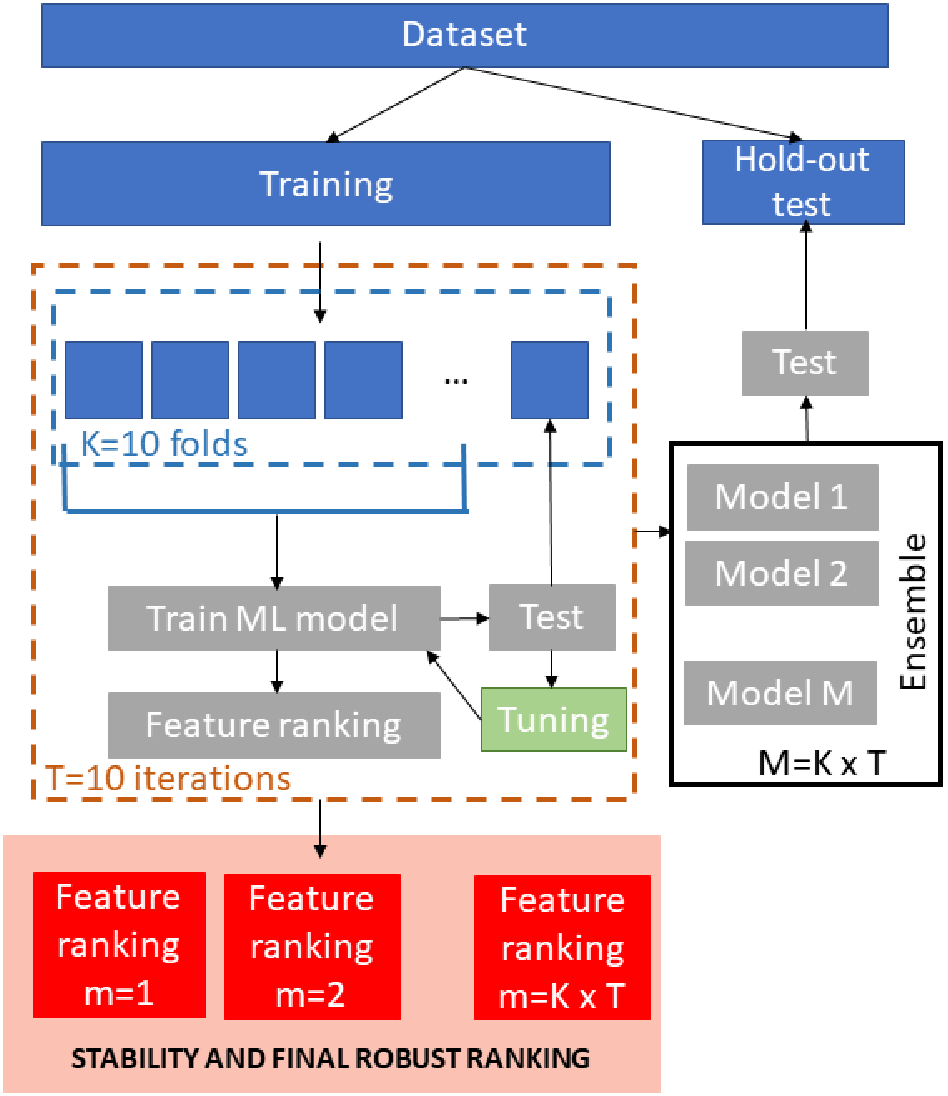
We analyzed DWI images of 122 subjects, both male and female. In accordance with diagnosis, the subjects were grouped into 52 HCs, 40 AD patients and 30 MCI converter subjects. We employed the diffusion toolkit FDT of FSL to define a processing pipeline.



After the preprocessing and co-registration steps, instead a structural connectome generation pipeline was performed. We run the docker images of the MRtrix pipeline by specifying the parameters in an input file for each subject.

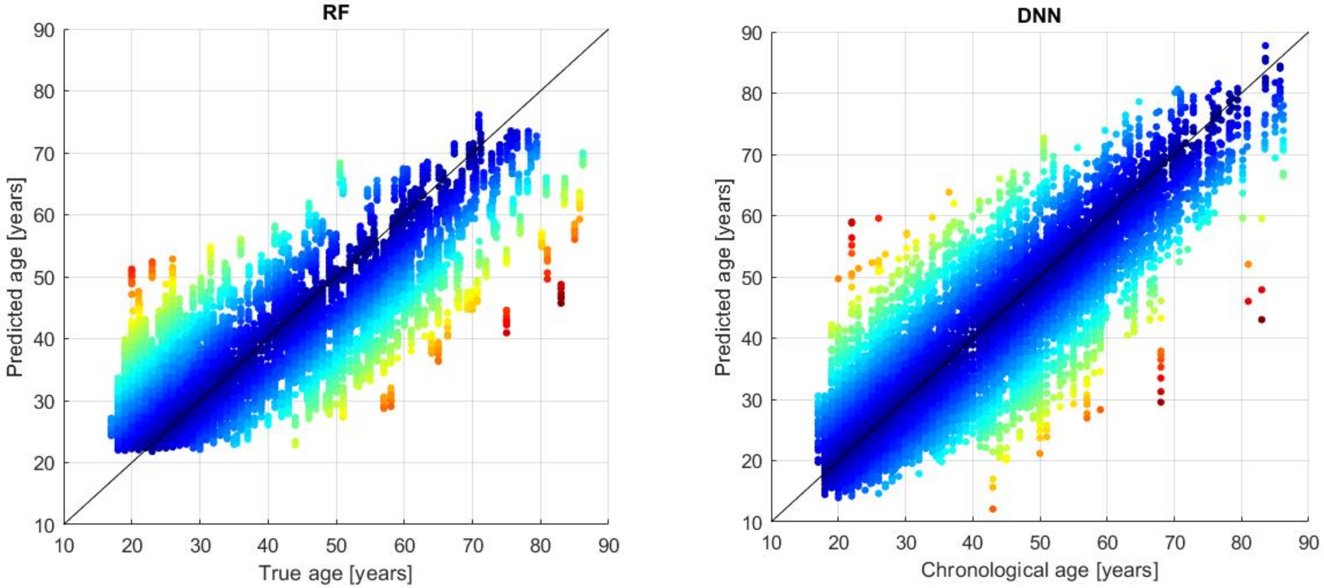
Lella, E., Lombardi, A., Amoroso, N., Diacono, D., Maggipinto, T., Monaco, A., ... & Tangaro, S. (2020). Machine learning and dwi brain communicability networks for alzheimer's disease detection. *Applied Sciences*, *10*(3), 934.

Example: Brain age prediction – PAC CHALLENGE



Lombardi, A., Monaco, A., Donvito, G., Amoroso, N., Bellotti, R., & Tangaro, S. (2020). Brain Age Prediction With Morphological Features Using Deep Neural Networks: Results From Predictive Analytic Competition 2019. *Frontiers in Psychiatry*, *11*, 1613.

Development of a predictive model based on deep neural networks (DNN) proposed during the **Predictive Analytic Competition 2019 for brain age prediction of 2638 healthy individuals.** We used FreeSurfer software to extract some morphological descriptors from the raw MRI scans of the subjects collected from 17 sites. We compared the proposed DNN architecture with other ML algorithms commonly used in the literature (RF, SVR, Lasso).



Conclusions

- ReCaS environment offers the most important features for **manage big datasets, process them, store results** in efficient manner and make all the pipeline steps available for reproducible data analysis.
- The neuroimaging processing environment is optimized for large datasets since it has been designed to reliably preprocess and **analyize data for hundreds of subjects in a single run,** on a distributed environment using HTCondor.
- **Run multiple machine learning algorithms for a multivariate experimental comparison of the predictive power of neuroimaging-based features.**

Thank you
for your attention

http://medphys.ba.infn.it/