Computing Methods for Experimental Physics and Data Analysis

Data Analysis in Medical Physics

Lecture 11: DL classifiers on medical images (CNN) and course summary

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

Today's objectives are…

- To practice with Convolutional Neural Networks (CNN) to classify images
- To summarize course content and objectives

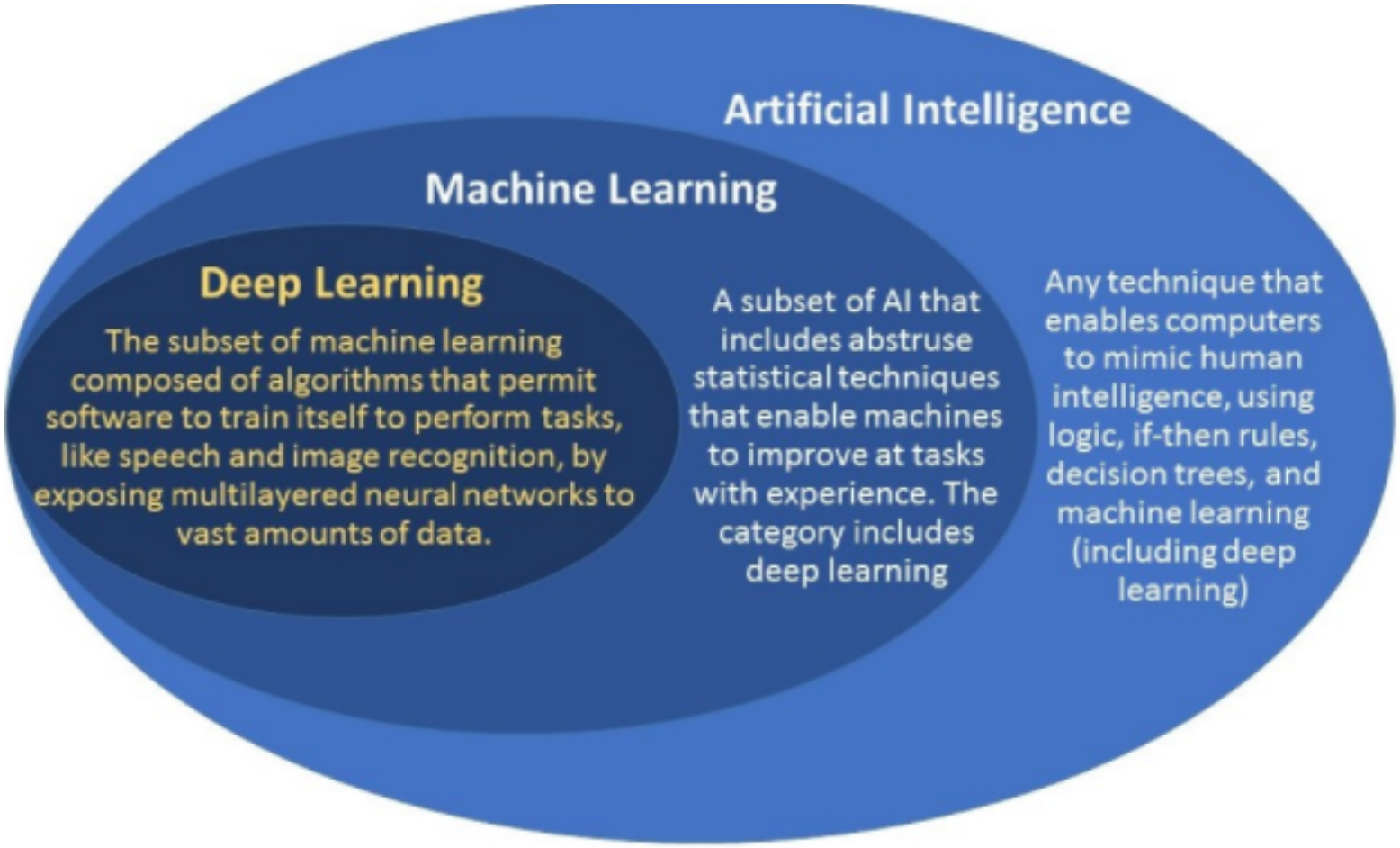
Follow and modify the code in https://github.com/retico/cmepda_medphys

- Lecture11_demo_train_CNN.mlx
- Lecture11_demo_visualizing_CNN_layers.mlx
- Lecture11_CNN.ipynb

The dataset used in the examples are:

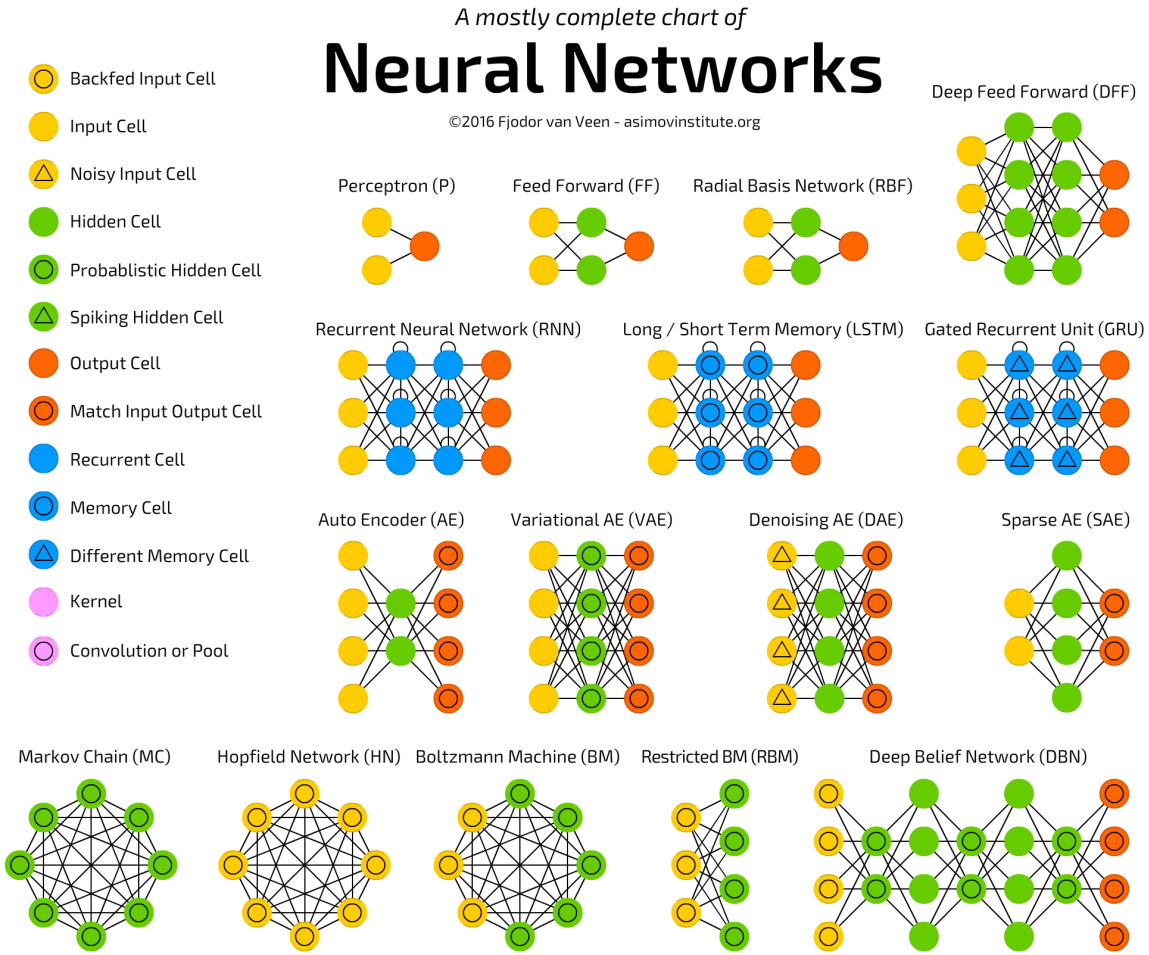
- DATASET/IMAGES/Mammography_micro
- DATASET/IMAGES/Mammography_masses

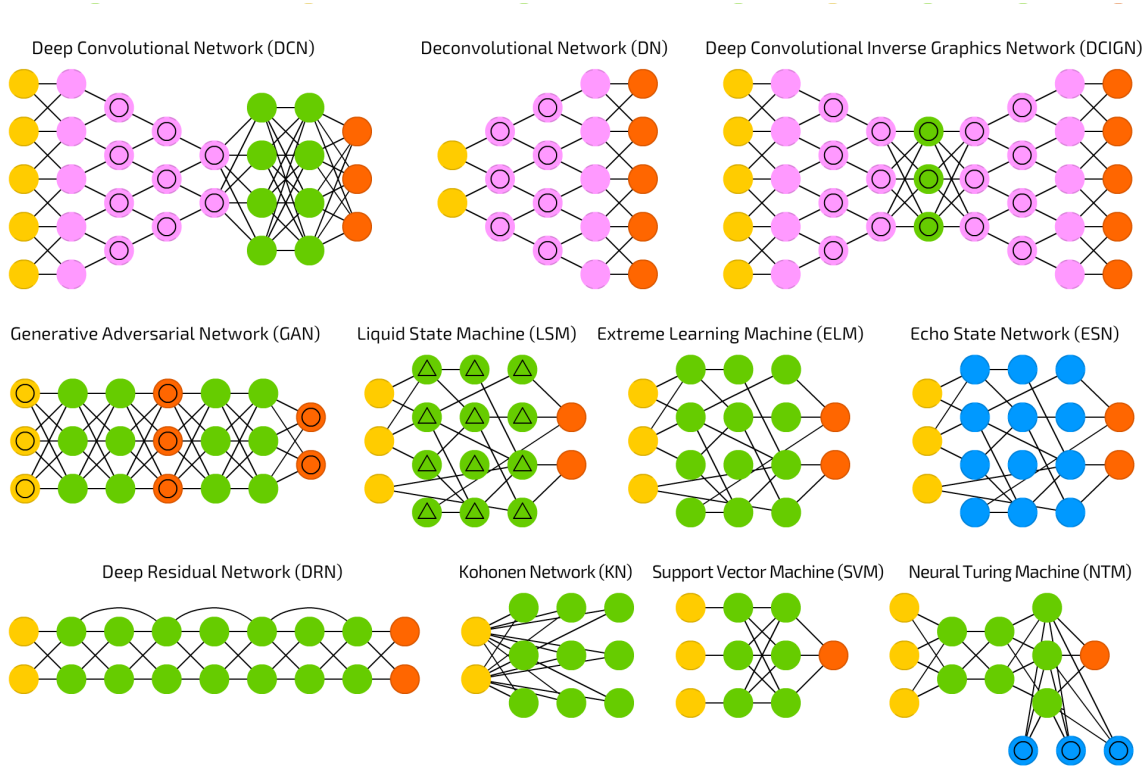
Artificial Intelligence, Machine Learning, Deep Learning



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

Neural network Zoo



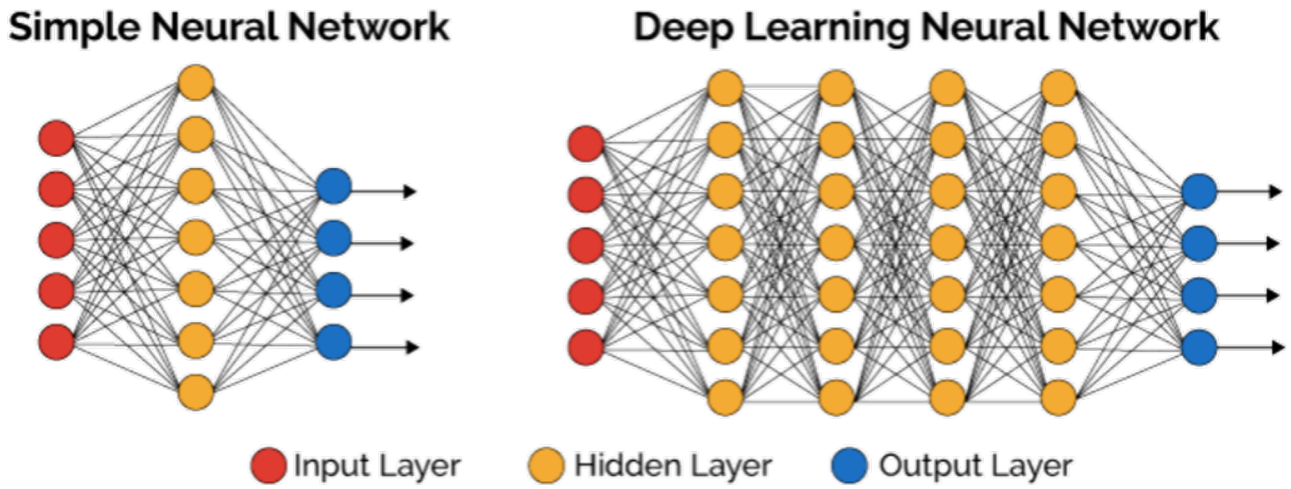


http://www.asimovinstitute.org/neural-network-zoo/

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



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

input

parameters of the network

non-linear activation function

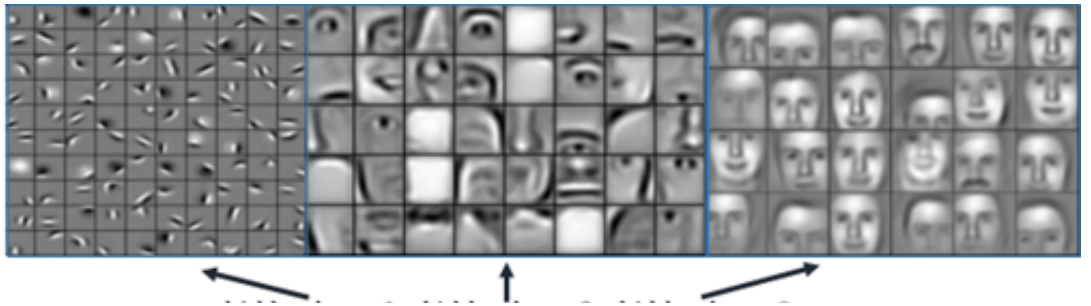
parameters of layer *L*

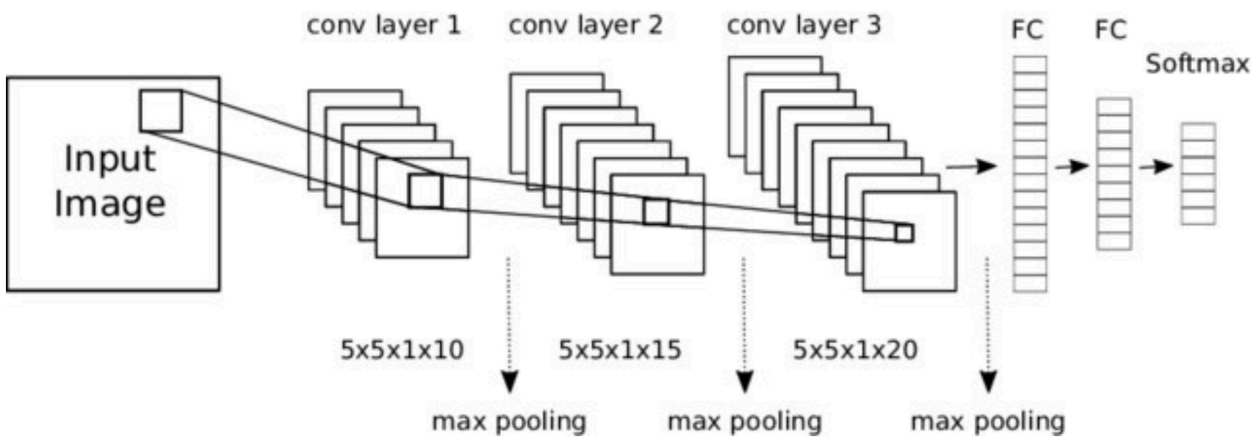
Deep Neural Networks

- 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

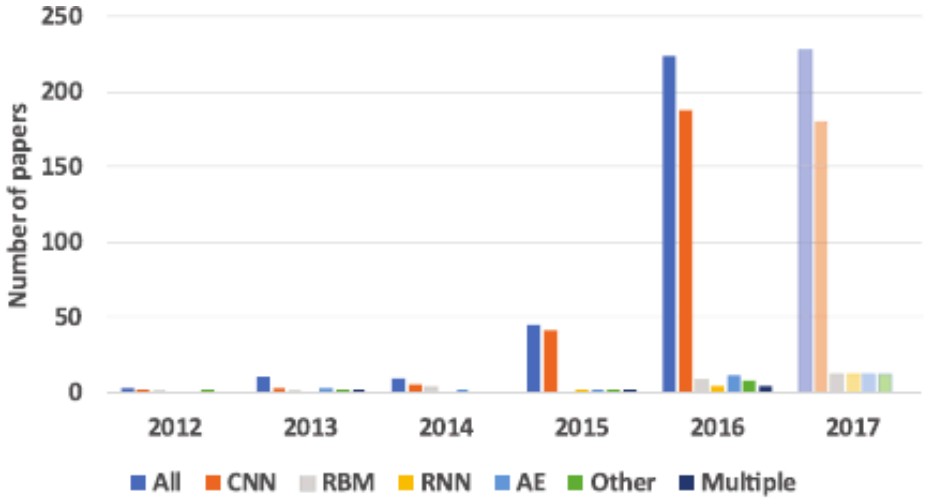


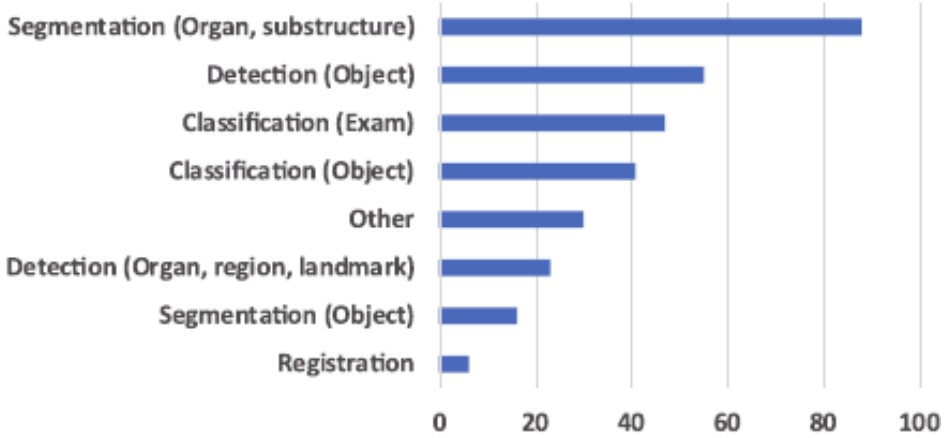


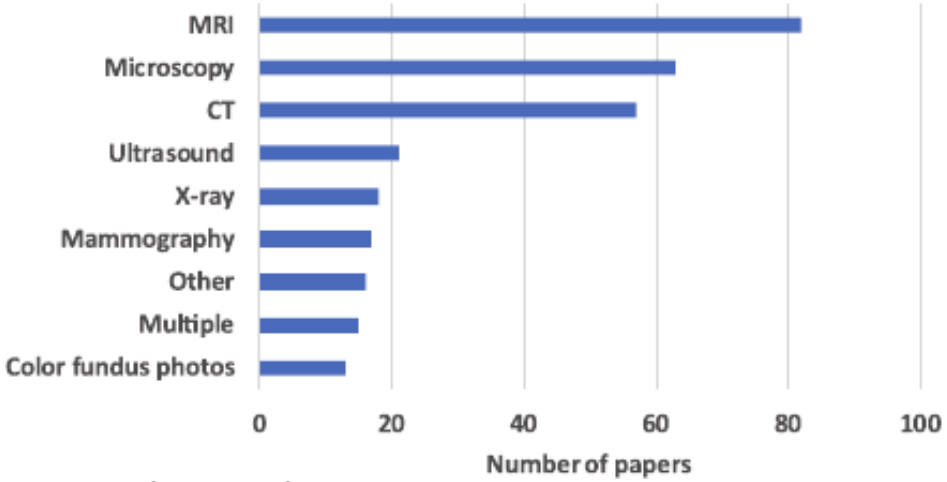
See demo code on L11_code/Lecture11_demo_train_CNN.m, Lecture11_demo_visualizing_CNN_layers.mlx

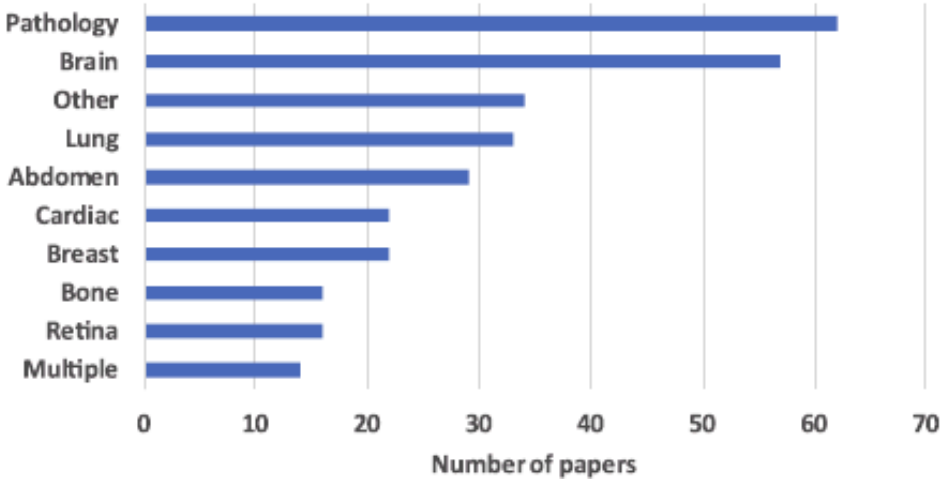
Deep Learning has become very popular in Medical Imaging

G. Litjens et al., A survey on in deep learning medical image analysis, Medical Image Analysis 42 (2017) 60–88









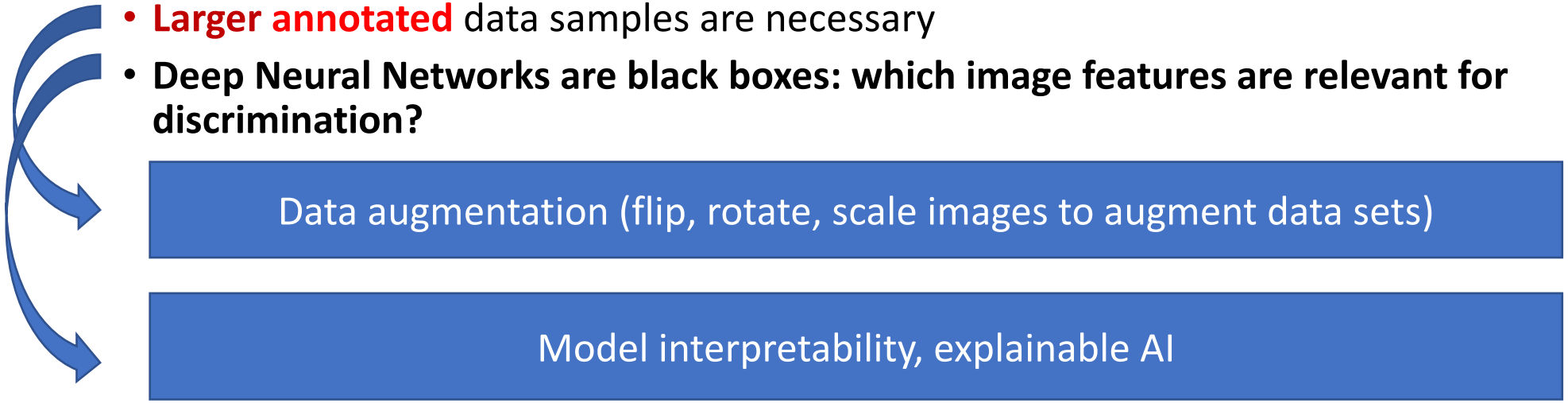
Breakdown of scientific papers by publication year, task addressed, imaging modality, and application area

DL vs. traditional ML approaches

- Deep Neural Networks can replace traditional handcrafted feature extraction

  ⟶ ***Data driven decision making***

  - **Pros:**
    - No prior selection of problem-related feature => no loss of information
  - **Cons:**
    - **Larger annotated** data samples are necessary
    - **Deep Neural Networks are black boxes: which image features are relevant for discrimination?**



Reproducibility studies

- Different analysis pipelines with the same purpose may rely both on different principles and on different algorithm implementations
- Extensive tests should be performed to assess the **reliability** of the features/measures they produce, before the output of SW packages for the analysis of medical data is used to infer results in the field of clinical research
- To estimate the precision of a SW package, a reference "gold standard" is necessary
- In most cases "gold standard" measures on large samples are not available, nevertheless, we can evaluate:
  - the robustness of a SW pipeline (**intra-method agreement**)
  - the reproducibility of the same measure across different SW pipelines (**inter-method agreement**)

- **The lack of intra-method and inter-method agreement can produce inconsistent results in medical imaging studies**

The *gold standard (ground truth)* problem

- Expert radiologists :
  - They quite often do not agree on the presence of lesions or on lesion relevance
  - Their manual segmentations will not be highly reproducible inter and intra-readers
- The most common strategy to create a gold standard is to form a panel of experts to jointly come together to a solution in doubtful cases
- Problems when performing algorithm training and validation:
  - the algorithm performance may drastically change according to the gold standard
- A possible alternative approach is to rely on **synthetic datasets!**

MATLAB and Python

- MATLAB provides a flexible, two-way integration with other programming languages, allowing you to reuse legacy code
- You can access Python functionality from MATLAB (you call Python functions and objects directly from MATLAB)
- To call Python® modules in MATLAB®, you must have a supported version of the reference implementation (CPython) installed on your system.

Install Supported Python Implementation

- MATLAB supports versions 2.7, 3.6, 3.7 and 3.8.
- MATLAB selects the version of Python based on your system path.
- To call a Python function, type py. in front of the module name and function name.
- Pass MATLAB data as arguments to Python function.
- MATLAB converts the data into types that best represent the data to the Python language.

```
>> pyenv                    # to determine which version MATLAB is using
>> py.help('len')           % To call a Python function, type py. in front of the function name.
>> L=py.os.listdir(".");
```

Python and MATLAB

Get Started with MATLAB Engine API for Python

R2019b

The MATLAB® Engine API for Python® provides a Python package named `matlab` that enables you to call MATLAB functions from Python. You install the package once, and then you can call the engine in your current or future Python sessions. For help on installing or starting the engine, refer to:

- Install MATLAB Engine API for Python
- Start and Stop MATLAB Engine for Python

The `matlab` package contains the following:

- The MATLAB Engine API for Python
- A set of MATLAB array classes in Python (see MATLAB Arrays as Python Variables)

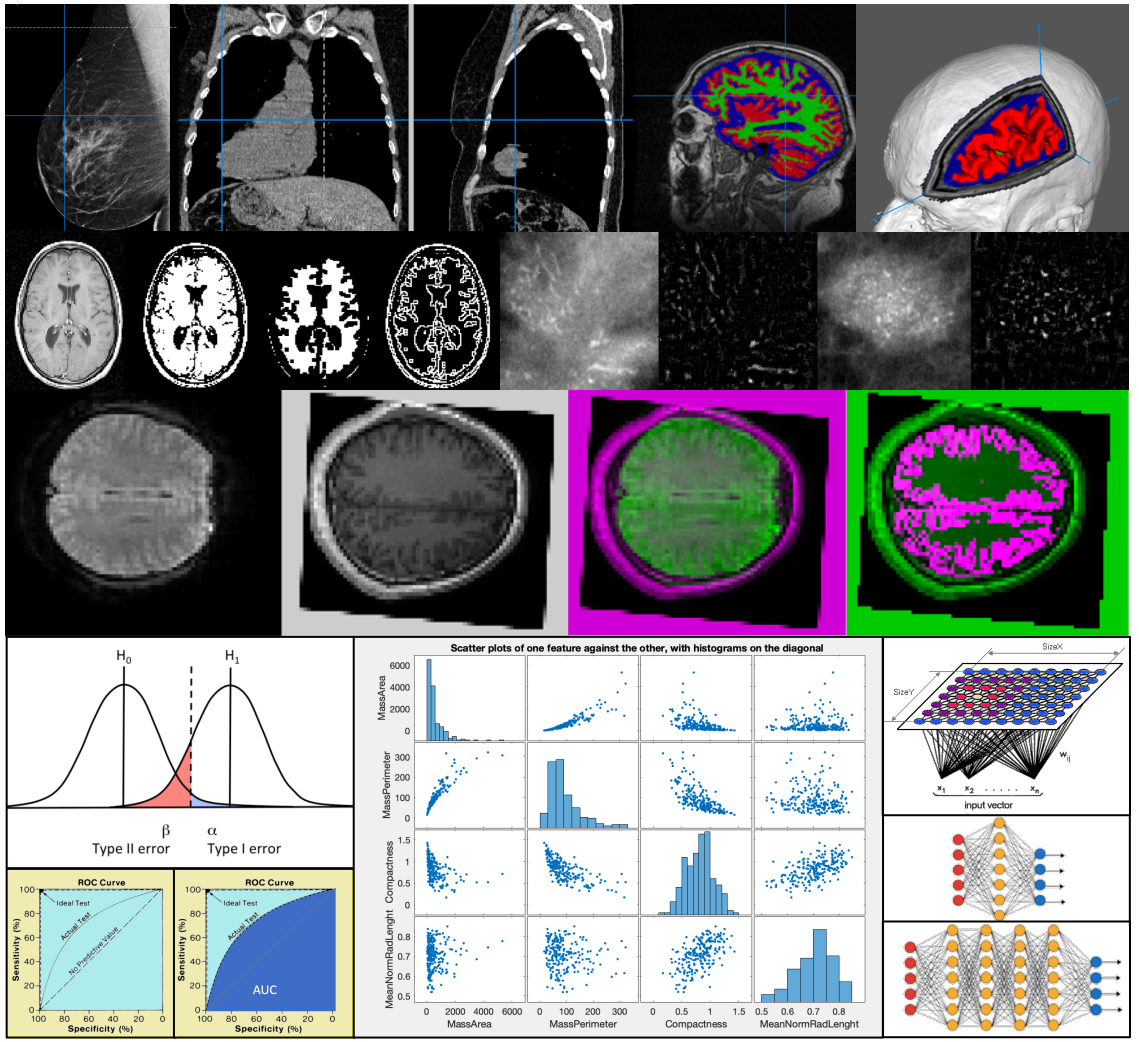
The engine provides functions to call MATLAB, and the array classes provide functions to create MATLAB arrays as Python objects. You can create an engine and call MATLAB functions with `matlab.engine`. You can create MATLAB arrays in Python by calling constructors of an array type (for example, `matlab.double` to create an array of doubles). MATLAB arrays can be input arguments to MATLAB functions called with the engine.

```
>>>  import matlab.engine
>>> eng = matlab.engine.start_matlab()
```

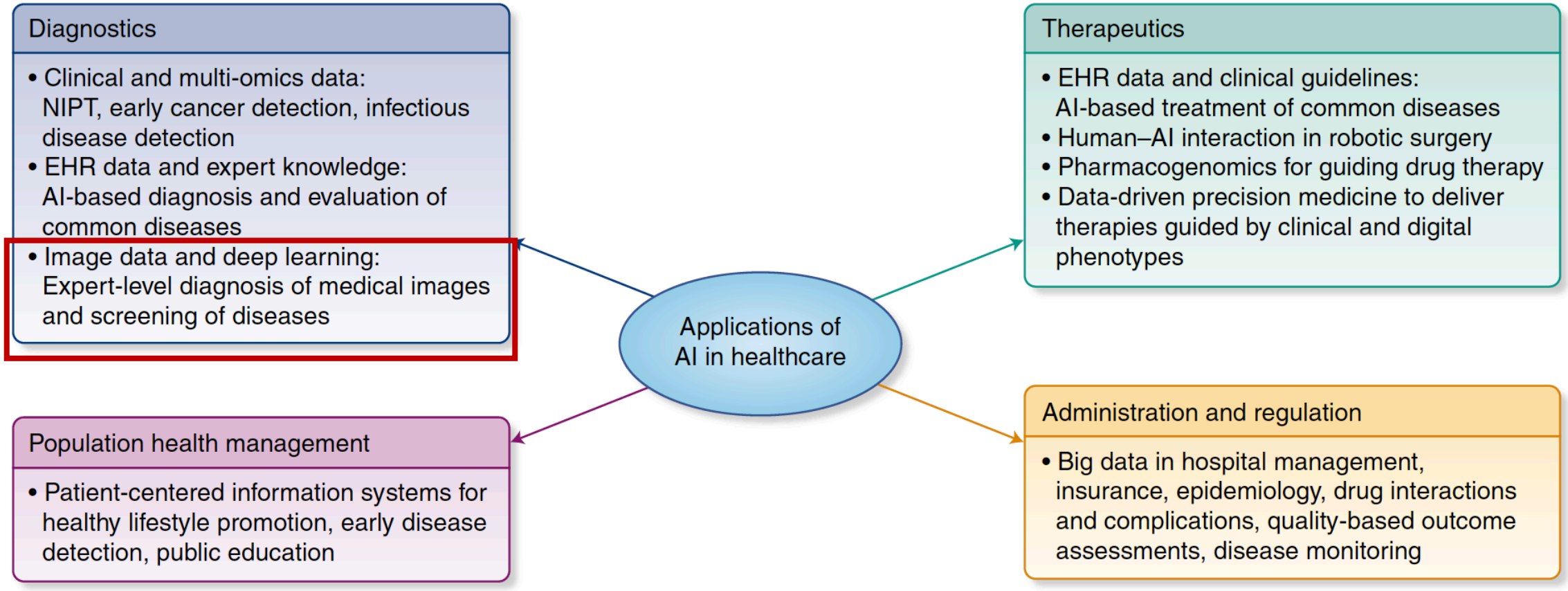
Course sum up

Medical data processing, feature extraction, feature/image classification:

- ✓ Handling standard-format medical data (DICOM), data anonymization, visualization
- ✓ Image filtering
- ✓ Deriving features form images, image segmentation
- ✓ Data quality control, outlier removal, dimensionality reduction
- ✓ Data analysis and classification
- ✓ Performance evaluation: figures of merit, cross-validation, permutation test
- ✓ Machine-learning and deep-learning tools for segmentation and classification



Artificial Intelligence applications in healthcare



Legend: HER, Electronic Health Records; NIPT, noninvasive prenatal test

*J He et al., The practical implementation of artificial intelligence technologies in medicine, Nature Medicine **25**, 30–36 (2019)*

Concluding remarks

- Medical diagnostic imaging daily produces to an incredible amount of digital information
  - ➢ not fully exploited neither for diagnosis nor for research!

- Clinicians need to be supported by reliable, effective and easy-to-use tools for diagnosing and monitoring a wide range of disease conditions

- Large Consortia are sharing multimodal and multicenter data in different medical fields

- The Medical Imaging community still lacks:
  - automated data quality pipelines
  - data harmonization strategies both for longitudinal and multicenter studies
  - new computational approaches to process and to mine multimodal data (imaging, genetics, clinical, demographic, etc.)
  - validated expert systems to support diagnosis and follow up of patients

References, sources and useful links

- References-Books
  - Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel: Backpropagation Applied to Handwritten Zip Code Recognition, Neural Computation, 1(4):541-551, Winter 1989
  - S. Haykin, Neural Networks: A Comprehensive Foundation, Prentice-Hall
  - I. Goodfellow, Y. Bengio, A. Courville. Deep Learning, The MIT Press
  - A. Géron, Hands-On Machine Learning with Scikit-Learn and TensorFlow, O'Reilly
  - Python Data Science Handbook by Jake VanderPlas https://colab.research.google.com/github/jakevdp/PythonDataScienceHandbook/blob/master/notebooks/Index.ipynb
- Sources
  - https://www.deeplearningitalia.com/una-panoramica-introduttiva-su-deep-learning-e-machine-learning/
  - http://www.asimovinstitute.org/neural-network-zoo/
  - https://towardsdatascience.com
  - https://it.mathworks.com/help/matlab/matlab-engine-for-python.html