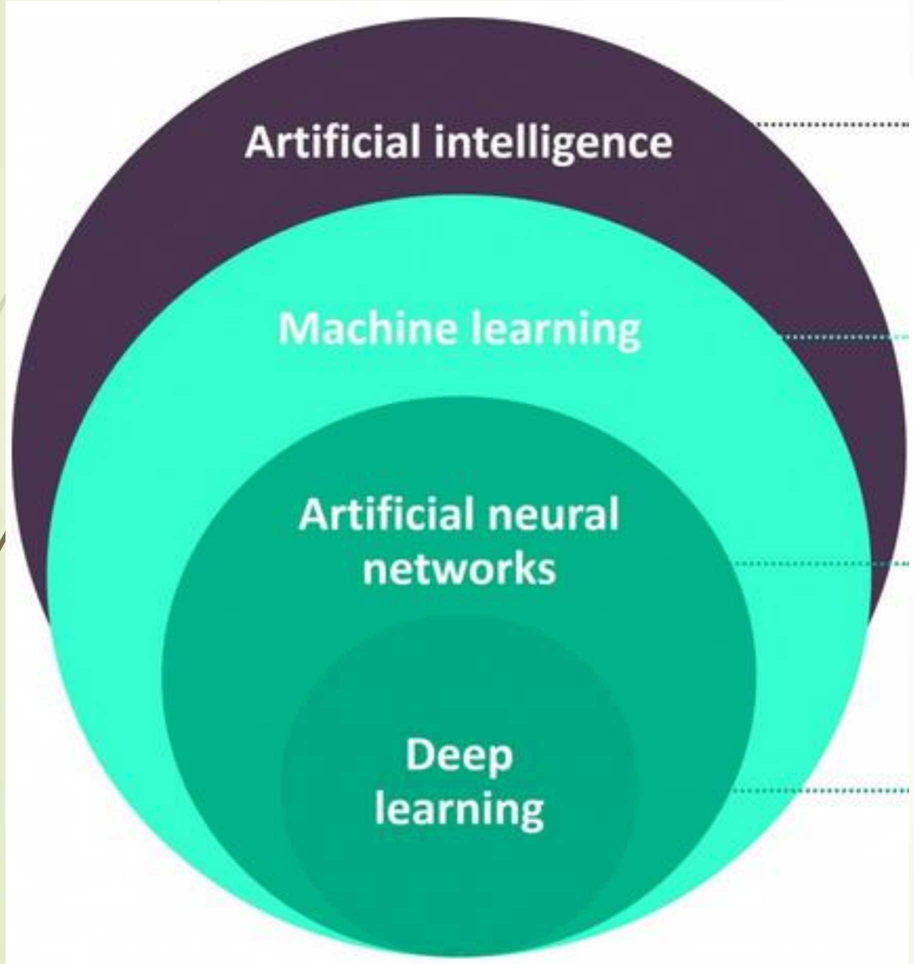
Overview of deep learning for medical Imaging

Teacher: Piersimone Marrocchesi

Phd student: Stefano Truzzi

What is Deep Learning



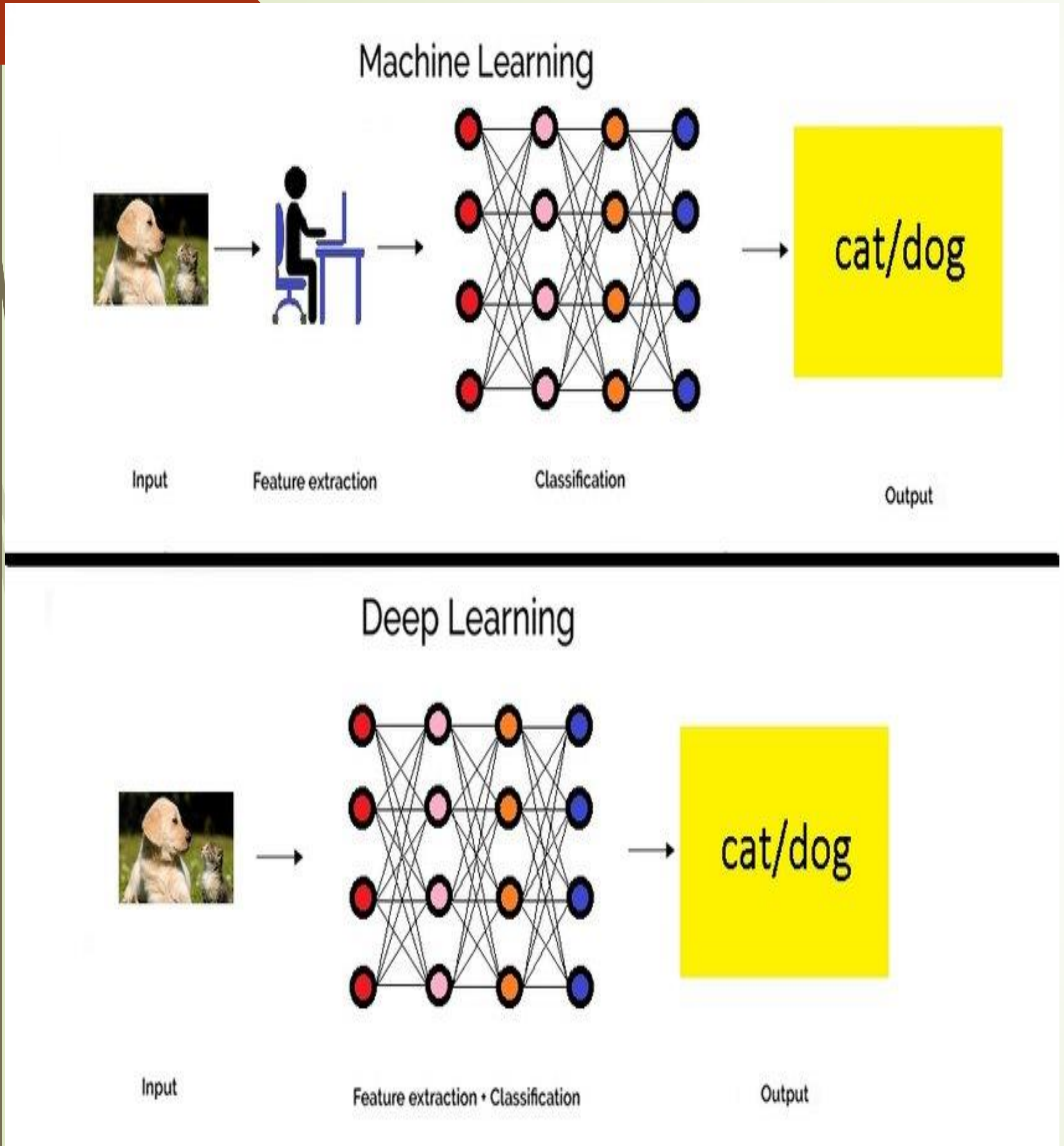
AI: simulation of human intelligence processes by machines

ML: is the study of computer algorithms that can improve through experience and by the use of data. In general are used for: **Classification or Regression**

ANN and DL are often used interchageably.
Are **Brain-inspired** Machine Learning Methods

The difference is in the Network architecture:
ANN has only 3 layers Input 1 Hidden Output.
DL has many hidden layers

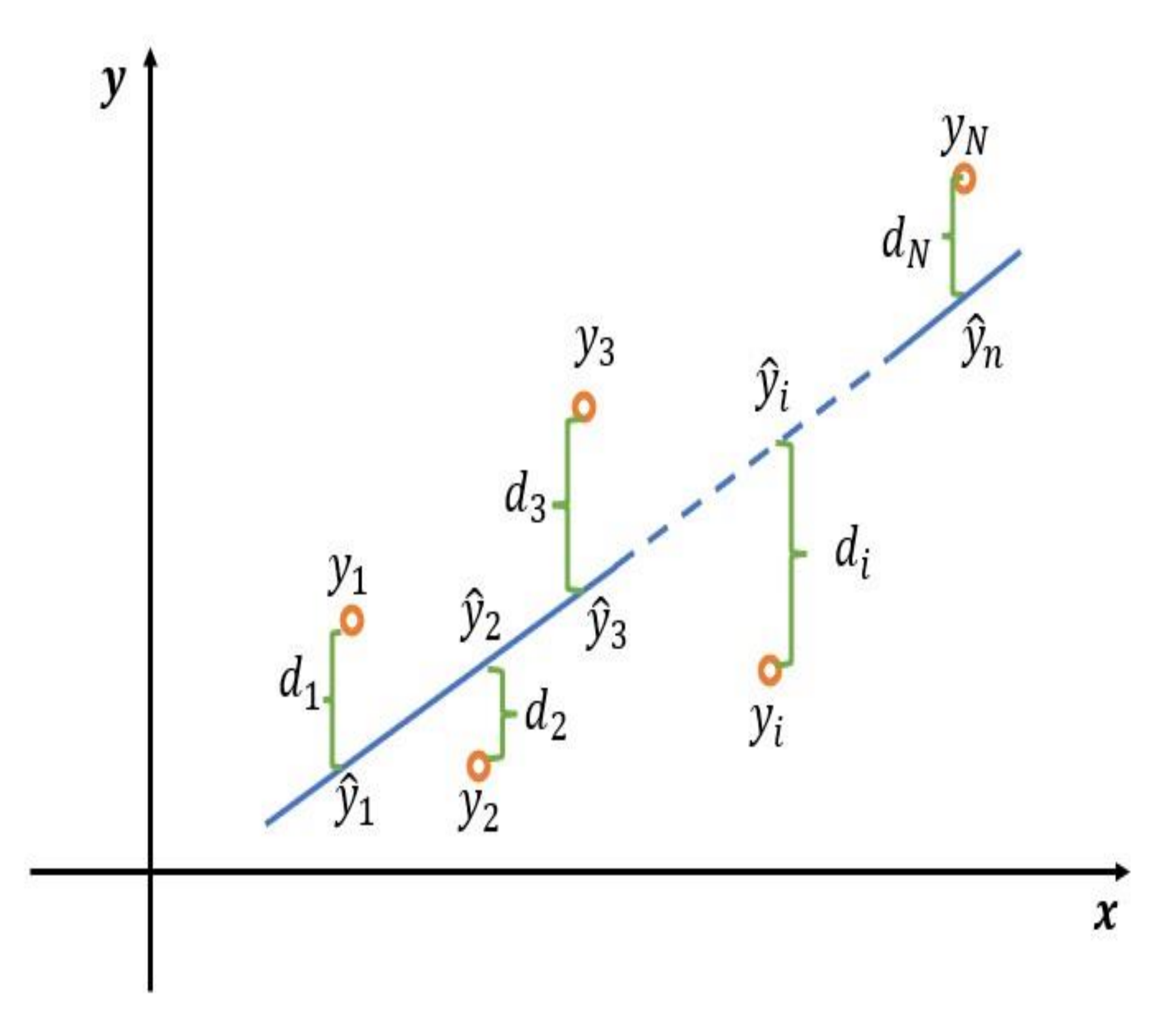
Machine Learning vs Deep Learning



DL +	ML -
Automatic features extraction	human features extraction
Can solve high end problem	divide the main problem in smaller task
High accuracy rate	Low accuracy rat
greater train time required	lesser train time required

DL -	ML +
need big quantity of data	smaller quantity of data neede
need of very high hardware resources	works on old hardware
networks are black box	
difficult to optimize the hyperparameters	
difficult optimize design	It's easiest project a ML program height

Linear Regression



$$\sum_{i=1}^{N}(d_i)^2 = \sum_{i=1}^{N}(\hat{y}(x_i) - y_i)^2$$

$$\hat{y}_i = ax_i + b$$

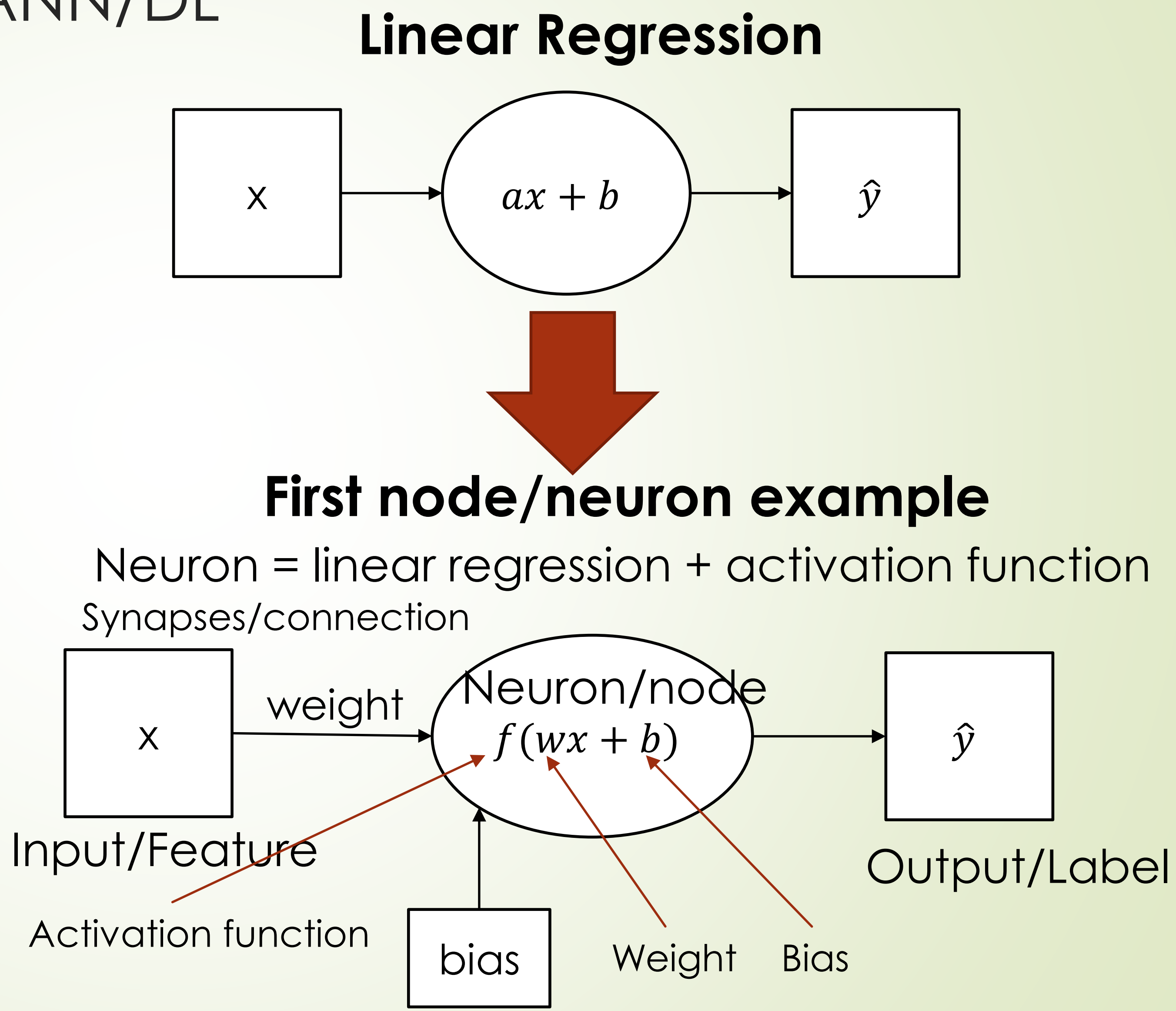
Best parameters value is the parameters value that **minimize the LOSS function**

$$\min_{a,b} \sum_{i=1}^{N}(\hat{y}(x_i) - y_i)^2 \rightarrow \frac{\partial f(a,b)}{\partial(a,b)} = 0$$

$$\begin{cases} a = \dfrac{N\sum x_i y_i - \sum x_i \sum y_i}{N\sum x_i^2 - (\sum x_i)^2} \\ b = \dfrac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{N\sum x_i^2 - (\sum x_i)^2} \end{cases}$$

From Regression to ANN (and DL) simple example of ANN/DL

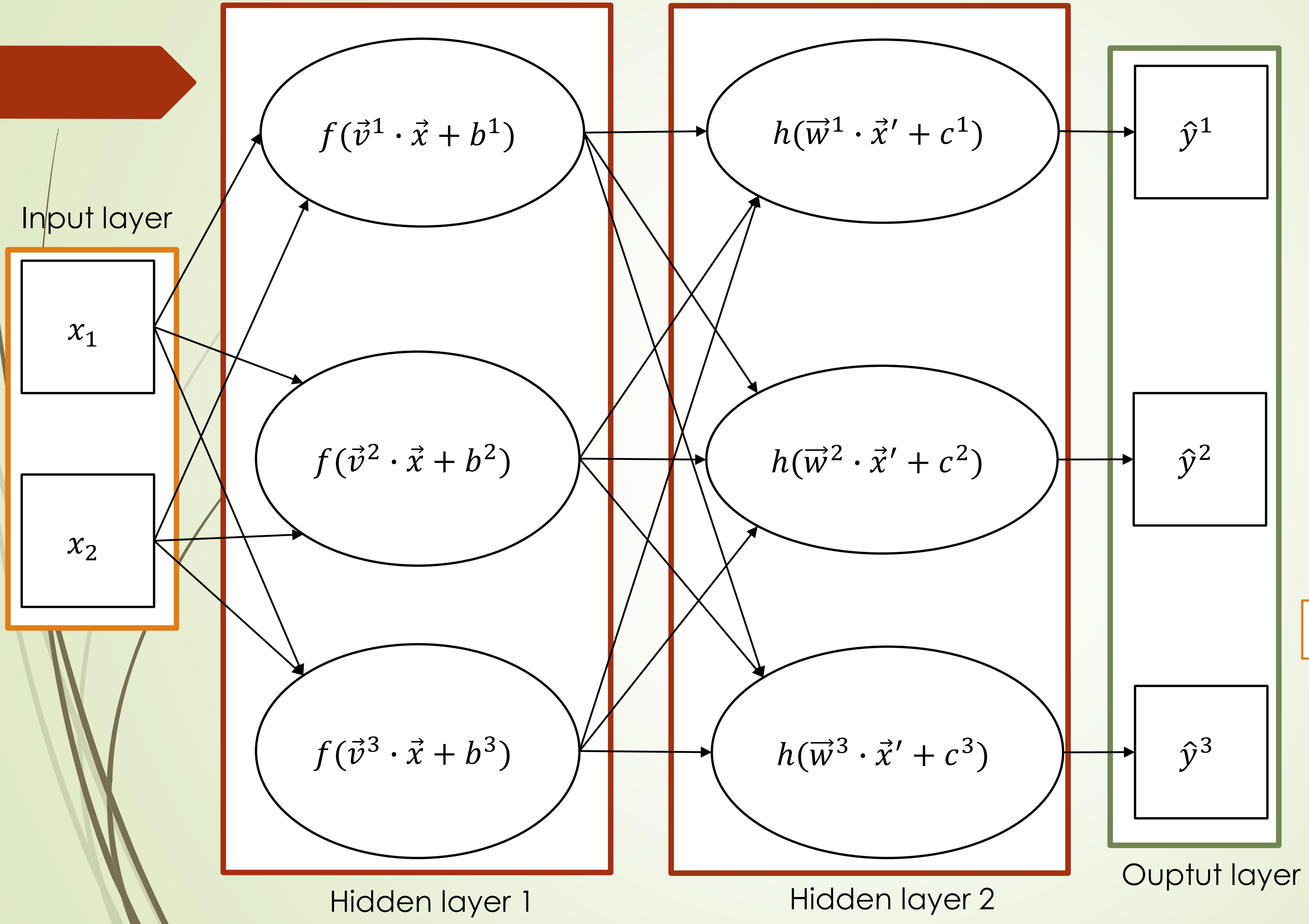
- The **Activation function** introduces **non-linearity**
- **Neurons/Nodes** are connected into a neural network
- Intelligence arises from **connection**
- Generalization from **scalar** (one node) to **vector** (network)



ANN (DL) simple example

Example:

- 2 hidden layers of 3 neurons each
- input layer with 2 input parameters
- Output layer with 3 label output



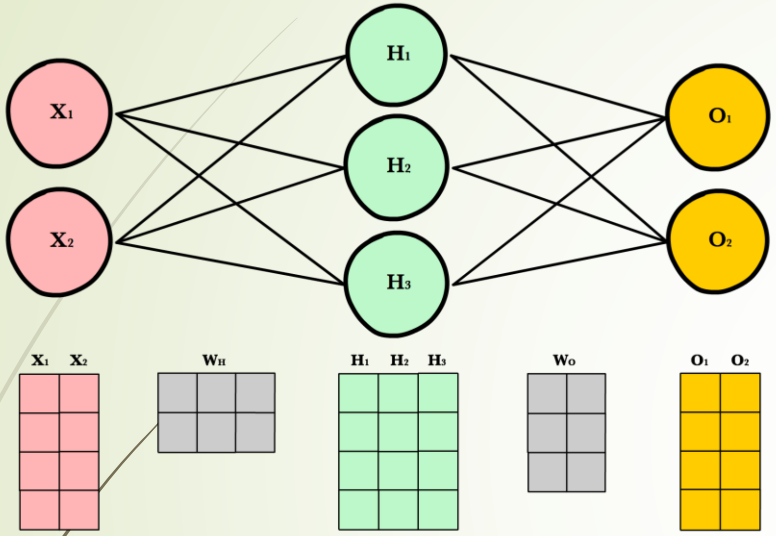
$$x'^i = f\left(\sum_{k=1}^{N} v_k^i x_k + b^i\right)$$

$$\hat{y}^n = h\left(\sum_{i=1}^{M} w_i^n x_i' + c^n\right)$$

ANN more in general

$$x'^{i} = f\left(\sum_{k=1}^{N} v_k^i x_k + b^i\right)$$

$$\hat{y}^n = h\left(\sum_{i=1}^{M} w_i^n x_i' + c^n\right)$$

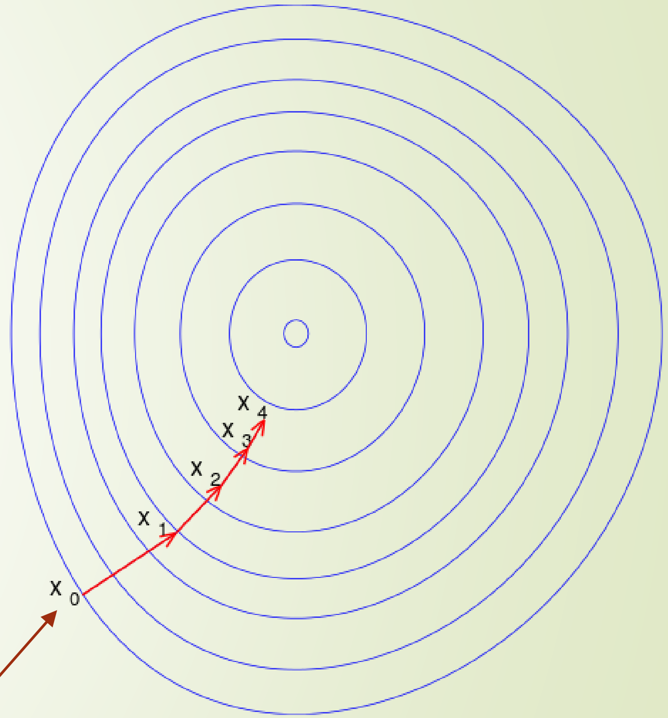


*The picture is an example the biases are not indicated

$\boldsymbol{W}$: weigths matrix
b: biases matrix
x: input vector
σ: activation function
L: Layer number

$$f(\vec{x}; \not{W}; \not{B}) = \sigma(W^L \sigma(W^{L-1} \ldots \sigma(W^0 \vec{x} + b^0) + b^{L-1}) + b^L)$$

- The first layer is the internal in the equation.
- The point is: **MINIMIZE** the $\mathbf{LOSS}[\boldsymbol{Y} - \boldsymbol{f}(\vec{\boldsymbol{x}}; \not{\boldsymbol{W}}; \not{\boldsymbol{B}})]$ respect ($\not{W}$)
- The most used techniques are based on the **Gradient Descent (GD)**
- **GD** is an iterative method that consist in take repeated steps (learning rate $\boldsymbol{l_r}$) in the opposite direction of the gradient

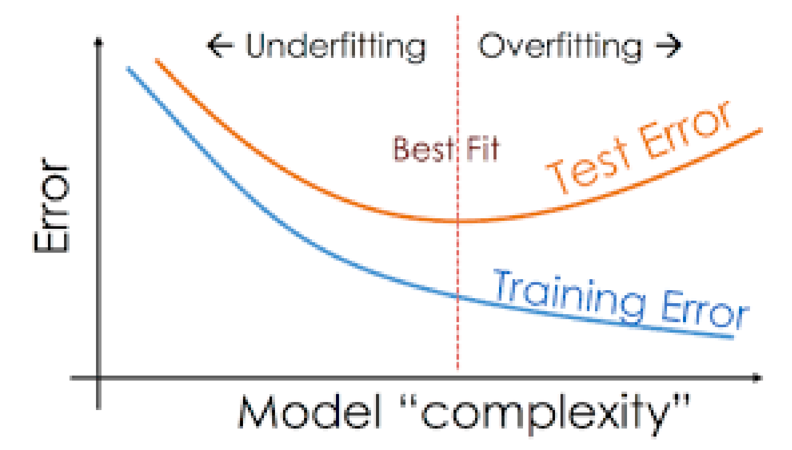


$$x_{n+1} = x_n - l_r \nabla(x_n)$$

Step size

Training and Test

- A **training** data set is a **data set of examples** used during the learning process and is **used to fit the parameters** (e.g., weights).
- A **test** data set is a data set that is **independent of the training data set**, but that follows the **same probability distribution** as the training data set.
- After the training phase the model is tested using the test data to see if the model give a **general result**



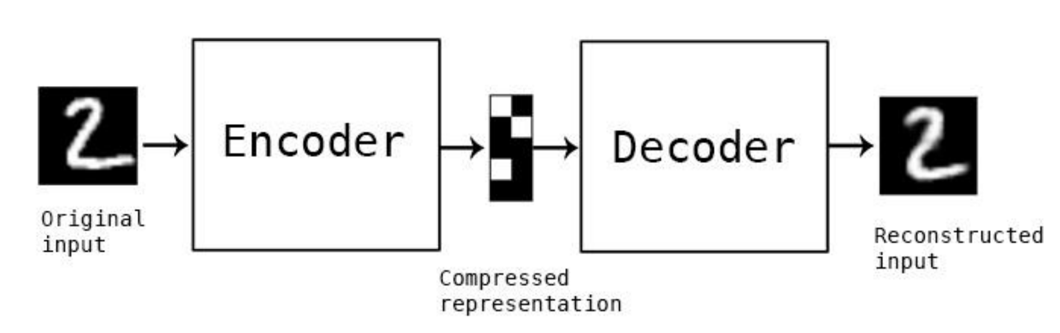
Under fitting: the model can **learn more**

Over fitting: the model learn "**too perfect"** the features of the training dataset.

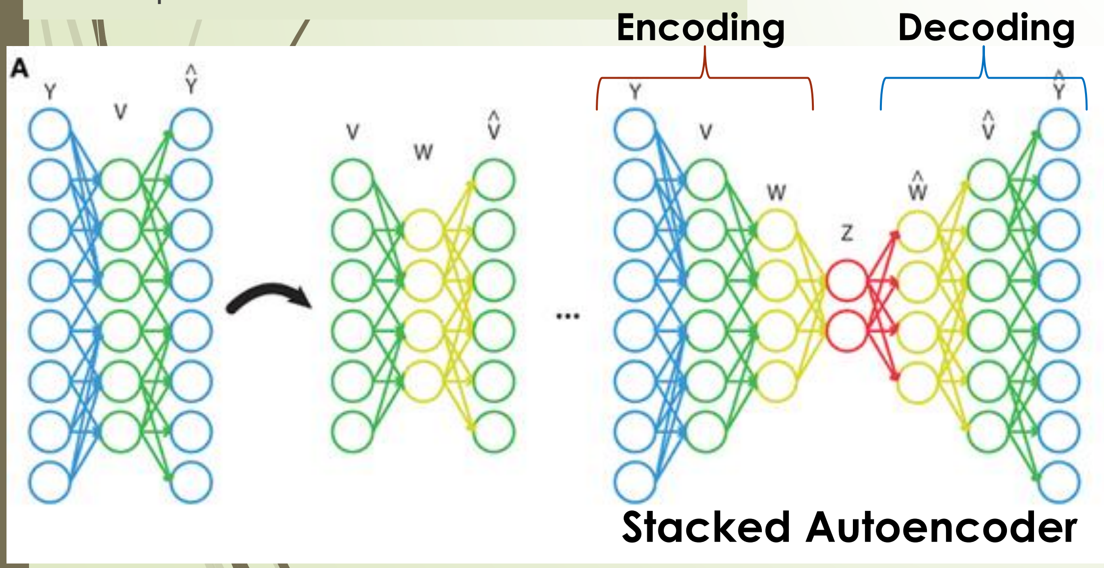
DL algorithm used for analysing medical imaging

Unsupervised Learning

Autoencoder



- **Autoencoder** is **ANN** that Encode (scompose) data and Decode (re-compose) the data to learn a different representation of data

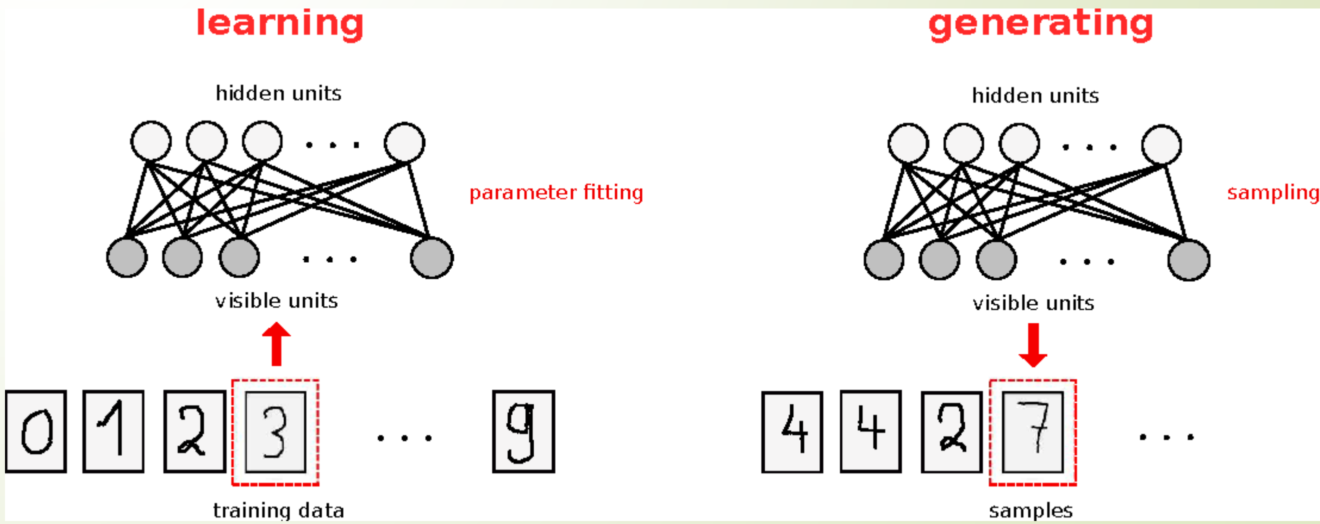


Typical uses:

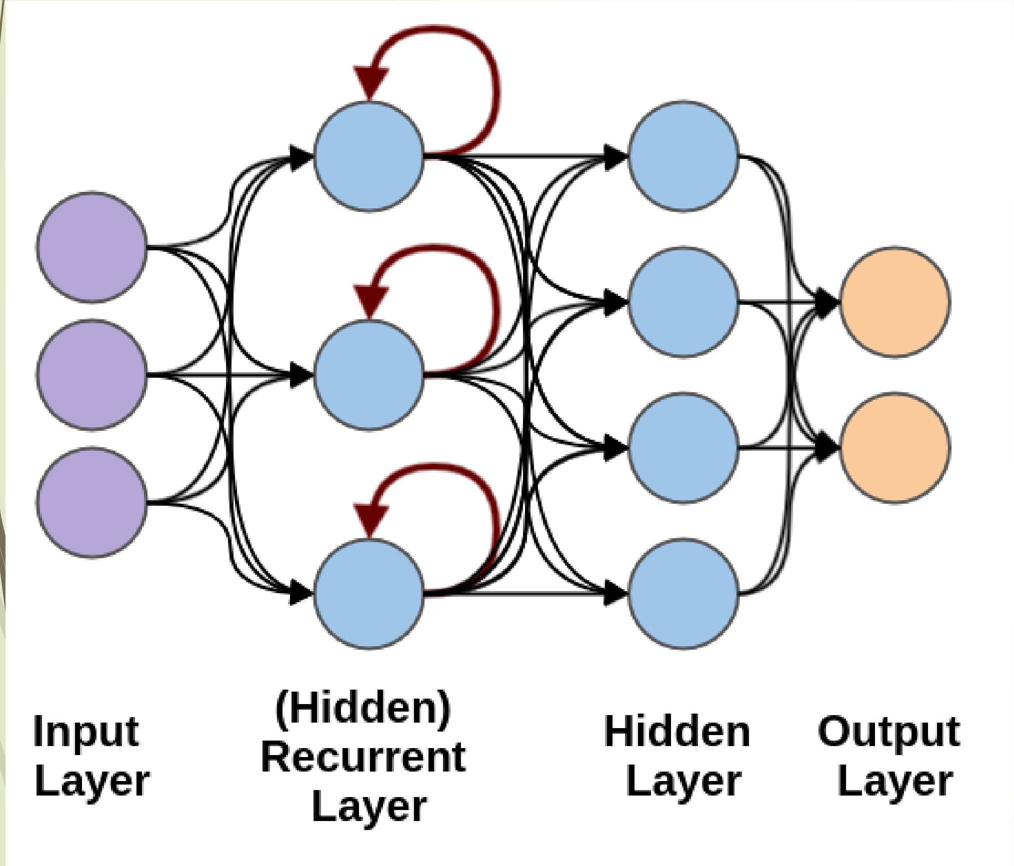
- Dimensionality Reduction
- Information Retrieval (with Dimensional reduction)
- Anomaly detection

Restricted Boltzmann Machine (RBM)

- **RBM are GENERATIVE Stochastics ANN** that can probability **distrubution** over its set of inputs.
- Represent the data with statistical functions and not with a recomposition of the original data
- Similar purpose of Autoencoder
- Can be **stacked** like Autoencoder (**Deep Belief Networks DBNs**)



Recurrent Neural Network (RNN) (Supervised)



- **RNN** use **Recurrent layer** a layer of nodes with feedback connection that can
- the input and output can be of varying length
- Some typical use:
  - Machine translation, Speech recognition, Speech synthesis, Grammar learning , Handwriting recognition that are part of **Natural Learning Processing (NLP)**
  - Rhythm learning and Music composition
- Heavy suffering of **Vanishing Gradient***
- **Long short-term Memory (LSTMs)** are a sort of upgrading of **RNN** that solve in part the **Vanishing Gradient*** problems

*is a problem related to gradient descent. We don't deal with it here

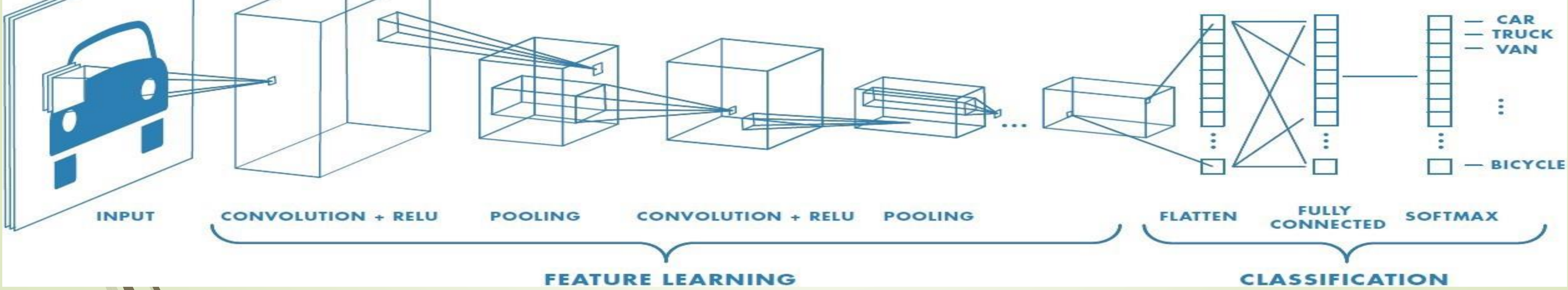
Convolutional Neural Network (CNN) (supervised)

Common uses:

- image and video recognition
- image classification,
- image segmentation
- medical image analysis

- **Convolutional Layer:** shared-weight architecture of the convolution kernels or filters that slide along input image and produce for output a **features maps**
- **Pooling Layer:** reduce the dimensions of the features maps (ex takes maximum)
- **Fully connected:** attached to CNN
- **Most used** in literature
- **MultistreamCNN:** consist in attach two or more CNN like in figure
  - For multi-scale images (1th CNN simplified global (**downscale**) representation in addition + 2th high resolution local info)
  - 2.5 D classification (multiple angled patches from 3D space)

CNN example



https://www.cs.ryerson.ca/~aharley/vis/conv/flat.html

Deep learning uses in medical imaging

Classification

Image/medical Exam classification (bigger eg. organ)

Objective is to classify if the exam is **negative or positive (2 classes classification)**

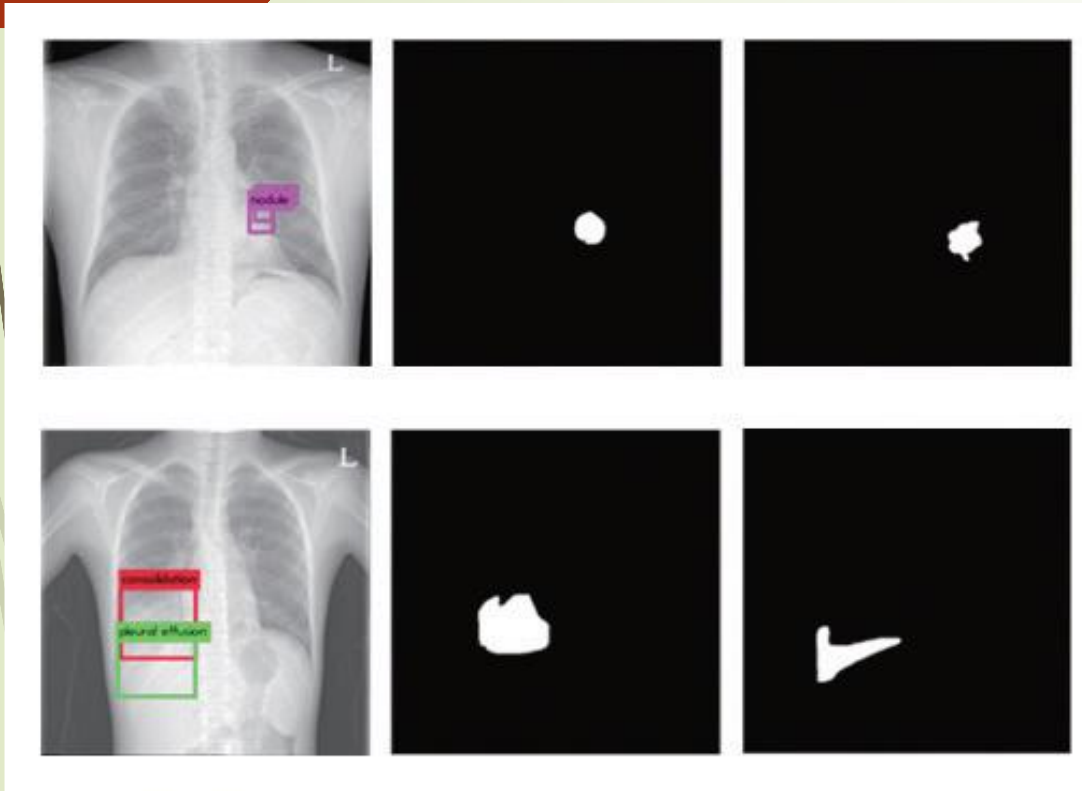
- Input image(s) (exams)
- **Output** single diagnostic variable (is disease? Yes or No)
- Every exam is sample
- CNN are the standard

Object or lesion classification (smaller eg. Cancer cells)

Find if you have a lesion and which type of lesion you have

- Often 2 info required to be accurate
  - **local lesion**
  - **global lesion**
- **Multistream CNN** where each stream work with a different resolution image.
- Not important the position only the diagnosis

Detection



Preliminary results of lesion detection on chest radiographs, by using faster Regional-CNN architecture.
Each result set is composed of 3 rows. First column shows faster R-CNN results, and ground truth lesion mask is delineated by radiologists in second column. Automatic description is provided in third column.

Organ, region and landmark localization (bigger eg. organ)

Object or lesion detection (smaller eg. Cancer cells)

classification+localization

Localization of Anatomical object in space/time (organ landmark)

- Important for **segmentation** or in the clinical workflow.
- **parsing of 3D** volumes.
- several work treat the **3D space** as a **composition of 2D orthogonal** planes (slices)
- To use pre-trained **(transfer learning)** various studies threat the **localization task as a classification task**
- Few study use direct 3D localization (**too complex**)

localization and identification of small lesions in the full image space.

- Very similar to Object or lesion classification
- CNNs used for pixel (or voxel 3D pixel) classification. After the object position is calculated.
- Difference with only classification
  - The individual pixels are more important
  - **Class Unbalance** towards non-object class because each pixel is evaluated
  - Easy discriminate non-object
  - The network is too influenced by non-object

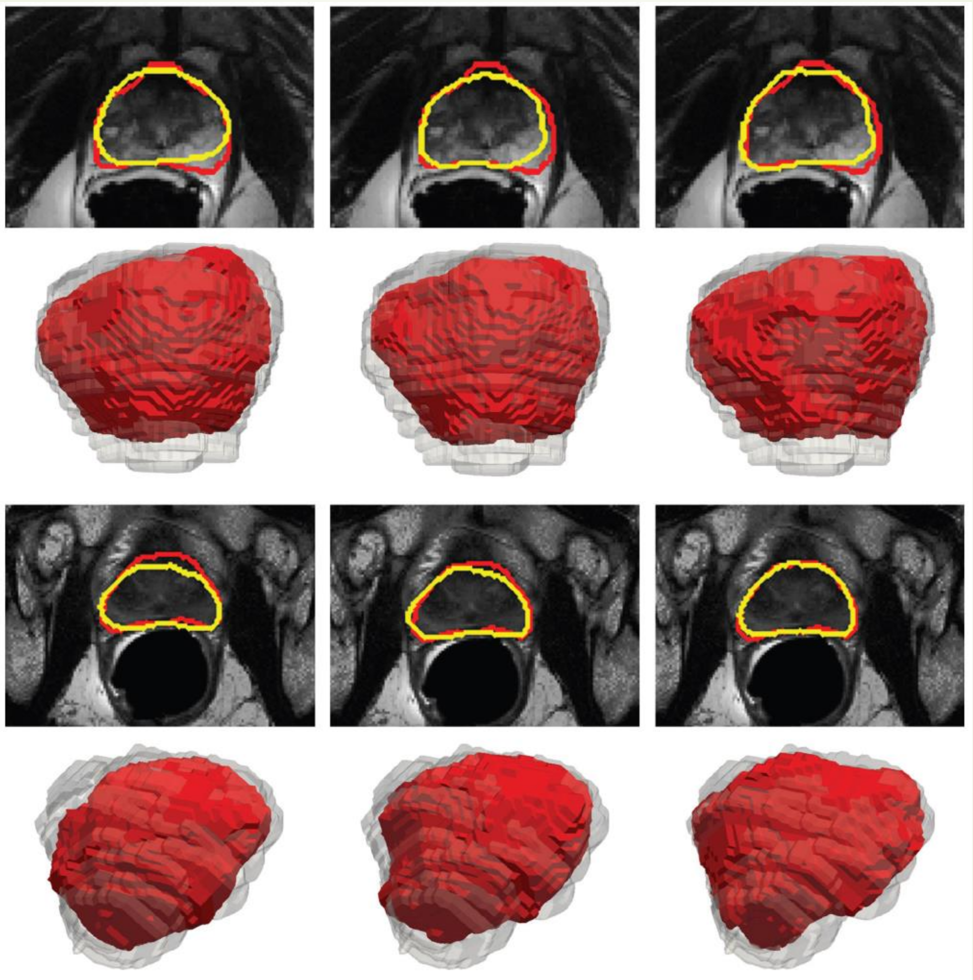
Segmentation

Organ and substructure segmentation

Find set of **voxels** defining **the contour or the interior of the object(s)** (organs or other)

- Most common subject of paper and widest variety in methodology (**CNN** and various **RNN**)
- One Famous CNN architecture for this is **U-Net**
- Most work in litereature

a Intensity b Hand-designed features c SAE-learned



Typical prostate segmentation results of two different patients produced by three different representations. **Red** contours
indicate **manual** ground-truth segmentations, and **yellow** contours indicate **automatic** segmentations

Lesion Segmentation

Combines organs and substructure segmentation

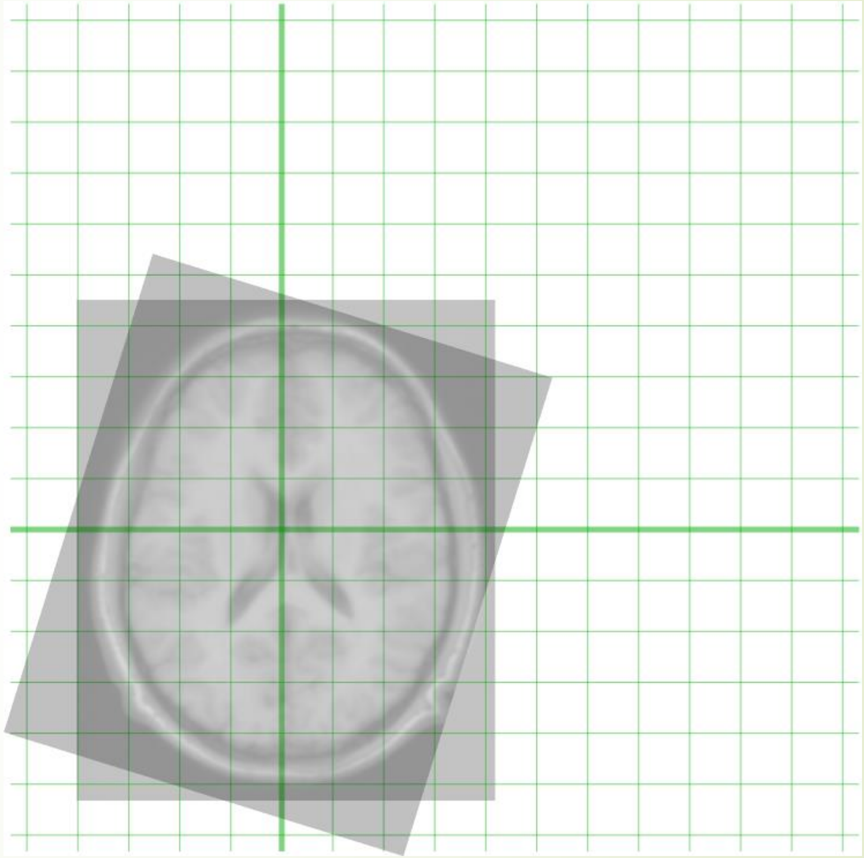
- multi- stream networks with different scales or non-uniformly sampled patches are used
- Also the **U-Net**
- **Class-unbalance** problems (too much non-lesion) same as lesion detection

Registration

Spatial aligment of medical images coordinate transform is cal- culated from one medical image to another

Two general strategies:

- Using DL networks to **estimate a similarity measure** for two images to drive an iterative optimization strategy
- **directly predict** transformation parameters using deep **regression networks**
- **Not many** papers



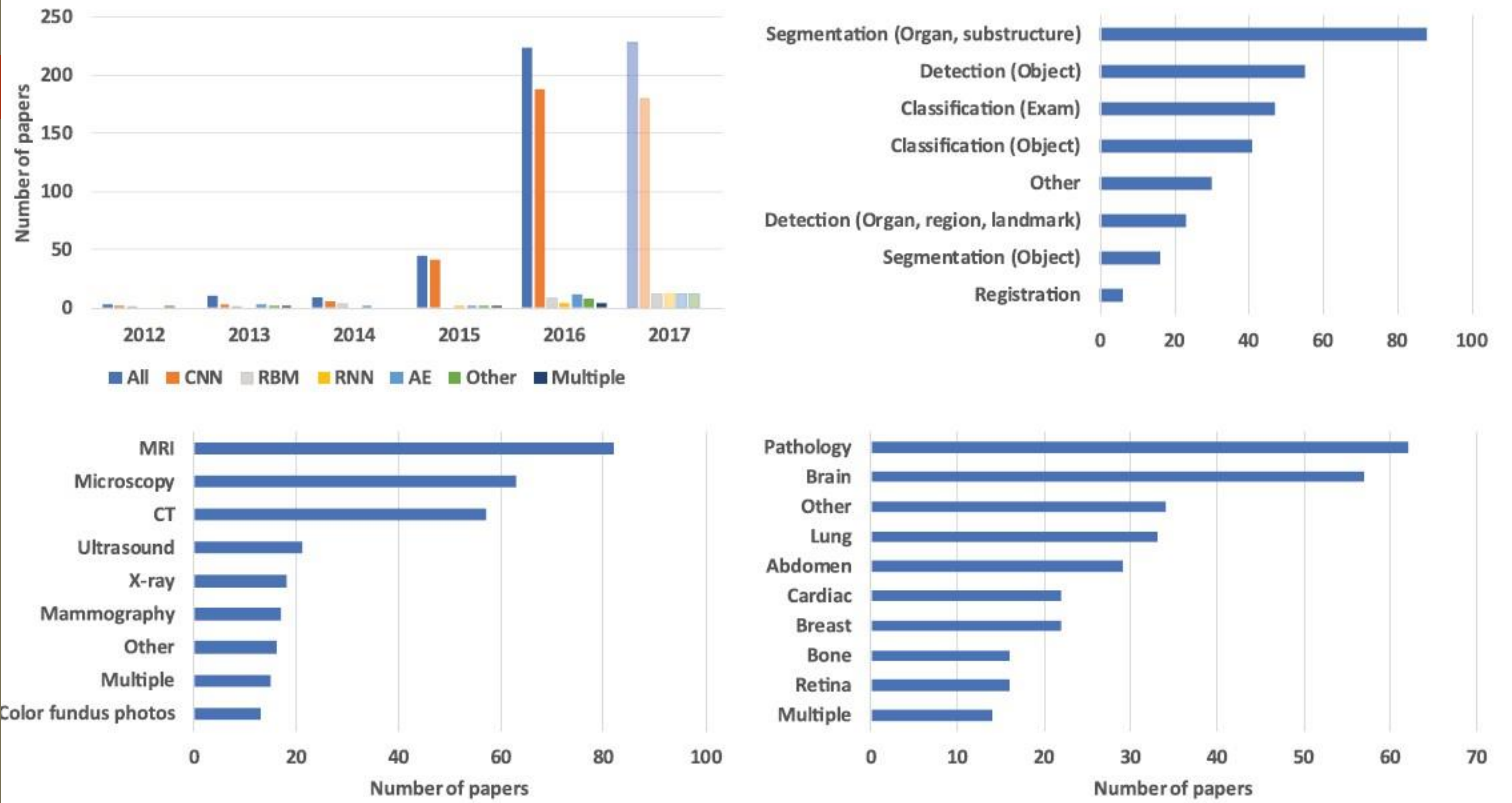
Registration of two MRI images of the brain

Problems with ANN with medical imaging

- Small Dataset (Data augmentation and Transfer learning)
  - Data privacy
  - Data common standard seems necessary
  - Legal problems connected to wrong diagnosis
- Human error classification for supervised (most used) networks

Final comment

- The major references article I used is from 2017
- Show that CNN is the most famous Neural Network (based on number of articles) (next slide)
- I speak only about the most common standard network, various other combinations are possible
- The ANN algorithms study is growly a lot not only in medical field



Number of papers
250
200
150
100
50
0
2012
2013
2014
2015
2016
2017
All
CNN
RBM
RNN
AE
Other
Multiple
Segmentation (Organ, substructure)
Detection (Object)
Classification (Exam)
Classification (Object)
Other
Detection (Organ, region, landmark)
Segmentation (Object)
Registration
0
20
40
60
80
100
MRI
Microscopy
CT
Ultrasound
X-ray
Mammography
Other
Multiple
Color fundus photos
0
20
40
60
80
100
Number of papers
Pathology
Brain
Other
Lung
Abdomen
Cardiac
Breast
Bone
Retina
Multiple
0
10
20
30
40
50
60
70
Number of papers

Thanks for the attention

References

- Introduction to Deep Learning with Keras. Lisa Benato and other Infieri 2019
- A survey of deep learning in medical image analysis. Geert Litjens and others. Diagnostic Image Analysis Group, Radound University Medical Center, Nijmegen, The Netherlands. 2017
- Comparing two classes of end-to-end machine-learning models in lung nodule detection and classification: MTANNs vs. CNNs. Nima Tajbakhsh and Kenji Suzuki. 2017

Backup

Transfer learning

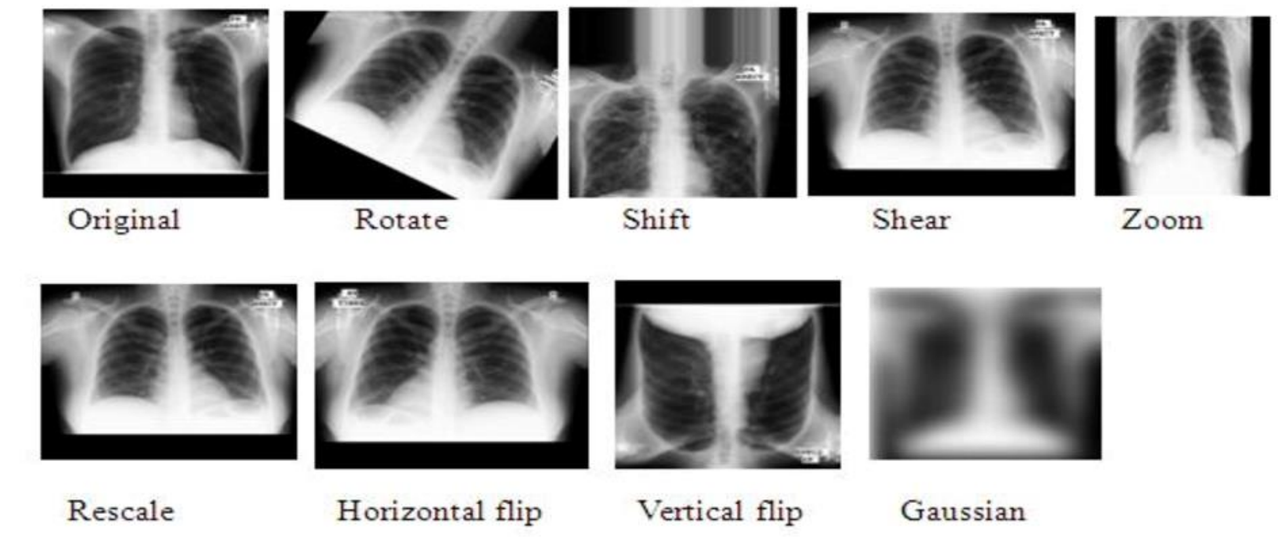
- Use a pre-trained model (a saved network) that was previously **trained** on a **large dataset**.
- Reduce training time
- Helps with the small dataset

The methods:

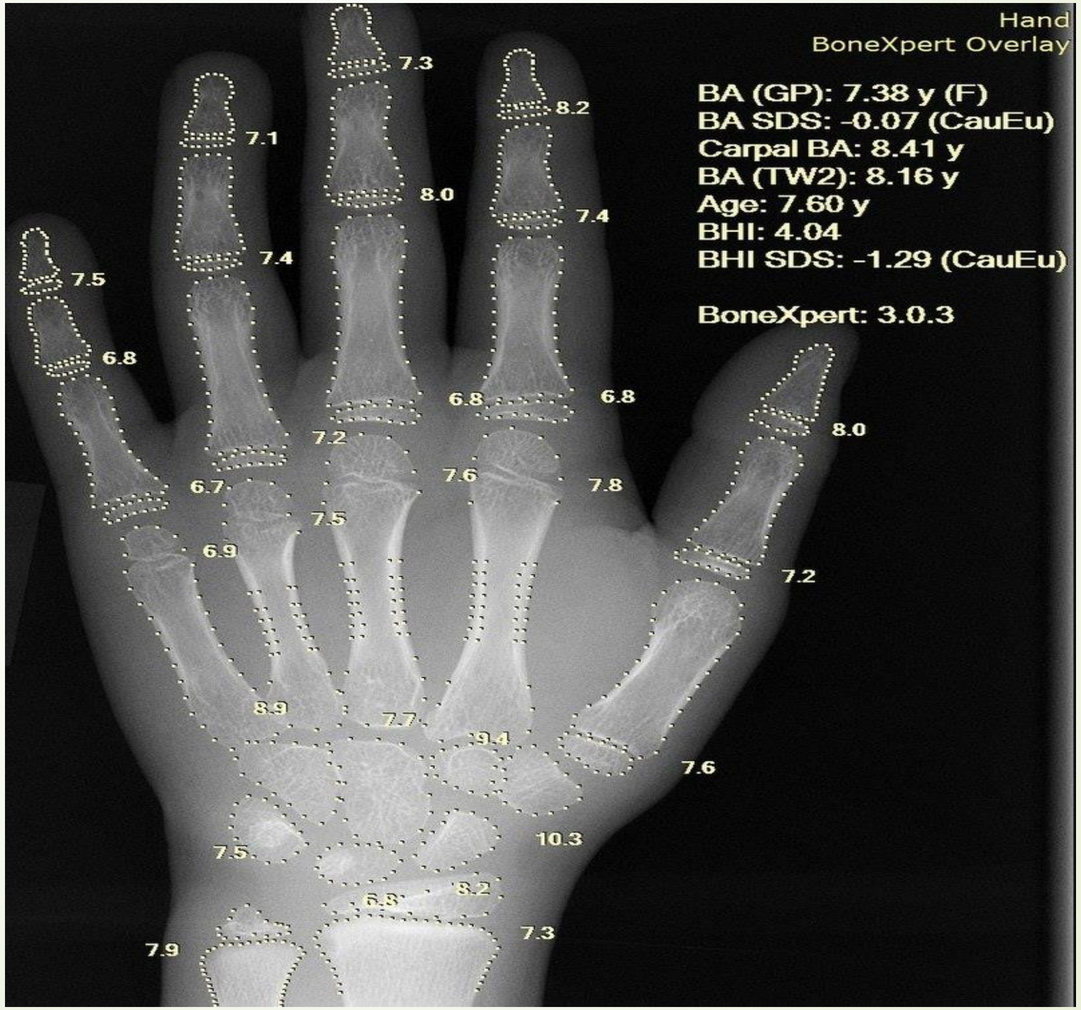
- **Feature extractor: FIXED** pretrained-weights. Attach and train only a classification layer at the end of the network.
- **Fine tuning: NOT/PARTIALLY FIXED** pretrained-weights. Attach and train a classification layer and all or a part of the old-weights

Data Augmentation

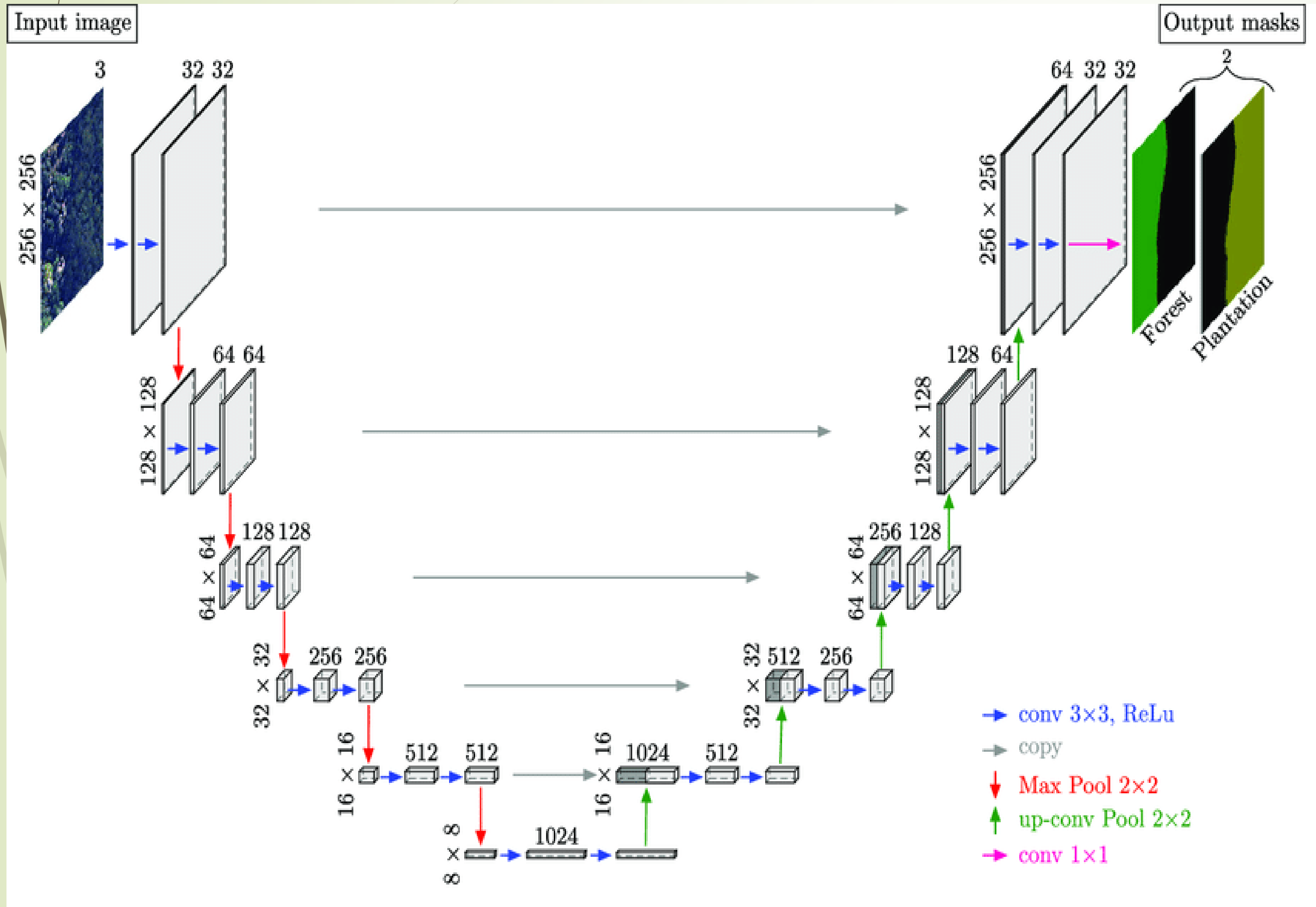
- increase the training dataset
- Create other images based on the original
  - Rotation
  - Translation
  - Flip
  - Etc…



Computer assisted diagnostic (CAD) (ML)



U-NET



- Type of Fully CNN (only CNN withouth Fully connected layers)
- Segmentation of a 512 × 512 image takes less than a second on a modern GPU
- **Downsampling** CNN to have the images decomposed in small pieces. The data of each Convolutional layer level are used for output.
- **upsampling** to increase the resolution of output for merge all the features
- Is like to use "**big**" and "**small**" features

Weight and Biases

- **Weights** control the signal (or the **strength of the connection**) between two neurons. In other words, a weight decides how much influence the input will have on the output.
- **Biases**, which are **constant**, are an additional input into the next layer that will always have the value of 1. Bias units are not influenced by the previous layer (they do not have any incoming connections) but they do have outgoing connections with their own weights. The bias unit **guarantees that even when all the inputs are zeros there will still be an activation in the neuron**.