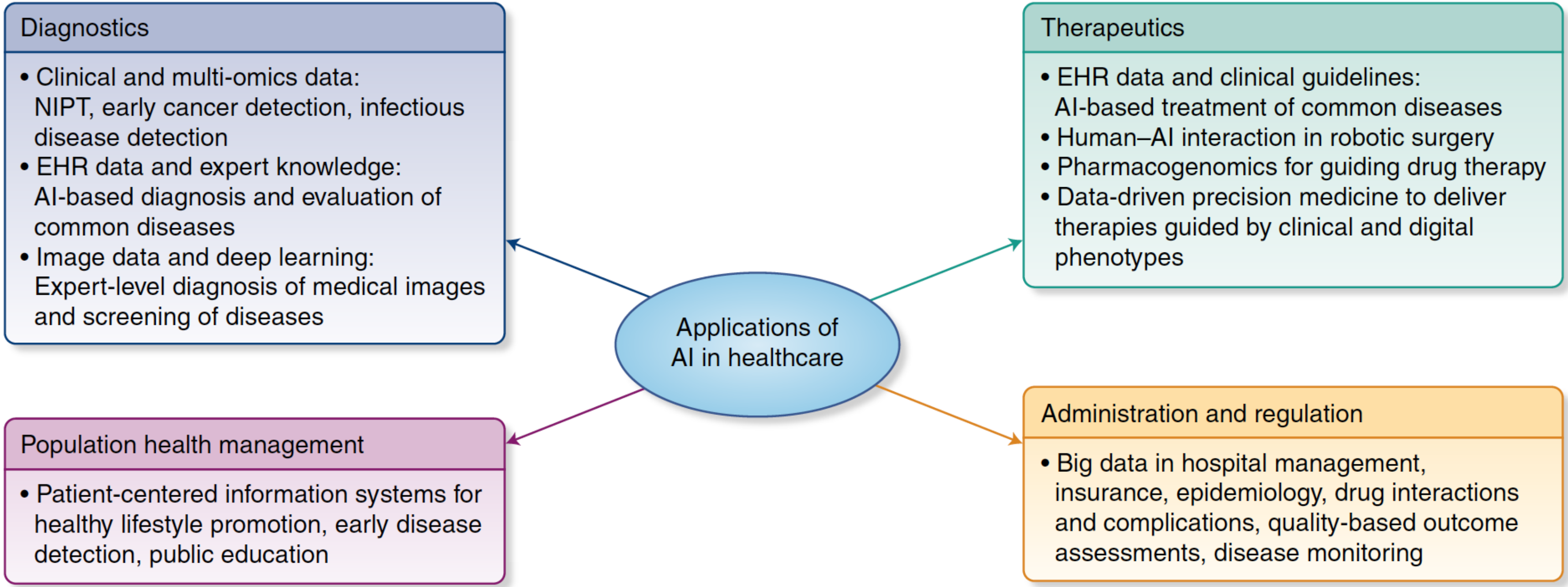
Machine Learning for Applications in Medical Physics

Piernicola Oliva
Università di Sassari & INFN Cagliari

oliva@uniss.it

Artificial Intelligence applications in Healthcare

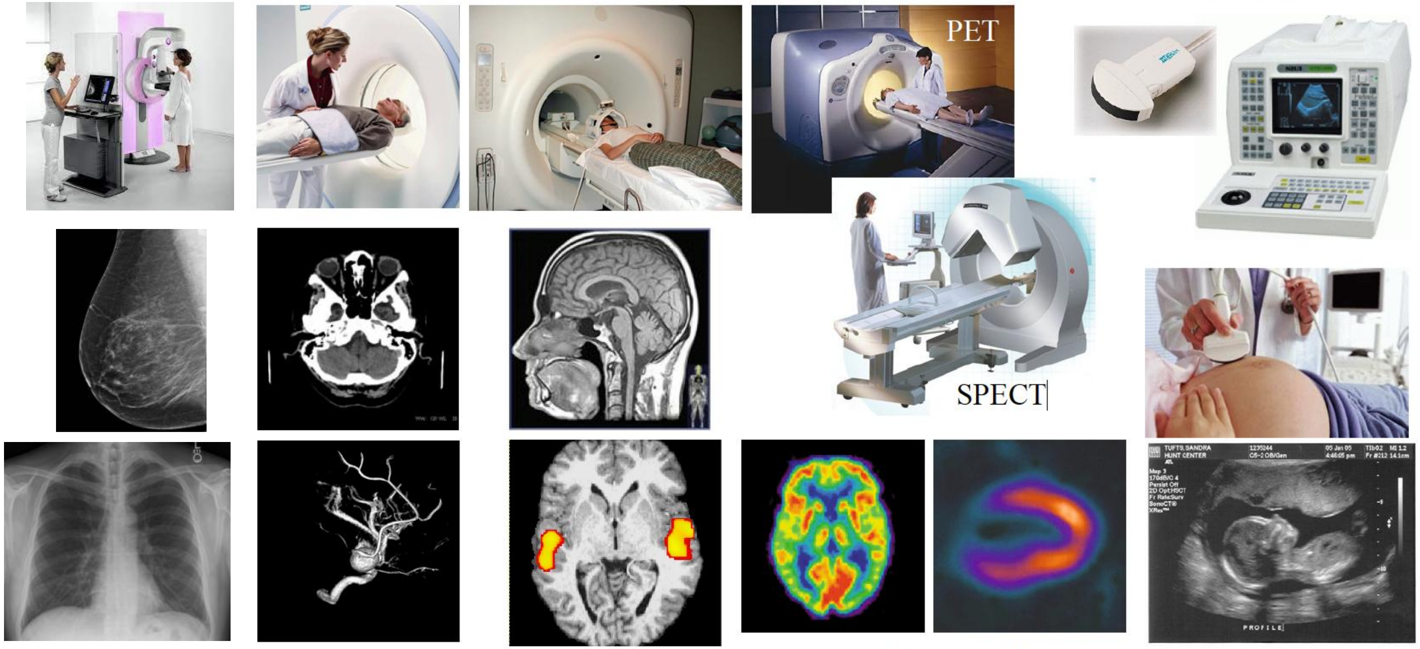


Legend: HER, Electronic Health Records; NIPT, noninvasive prenatal test

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

Medical Imaging: there are many techniques based on different physical principles

X-ray | CT | MRI / fMRI | Nuclear | Ultrasound



X-ray | X-ray | magnetic spin | metabolic tracer X-ray emission | sound waves

Medical images are more than pictures!!!

Decision Support Systems (DSS) for Detection/Diagnosis

Image processing and analysis techniques can help:

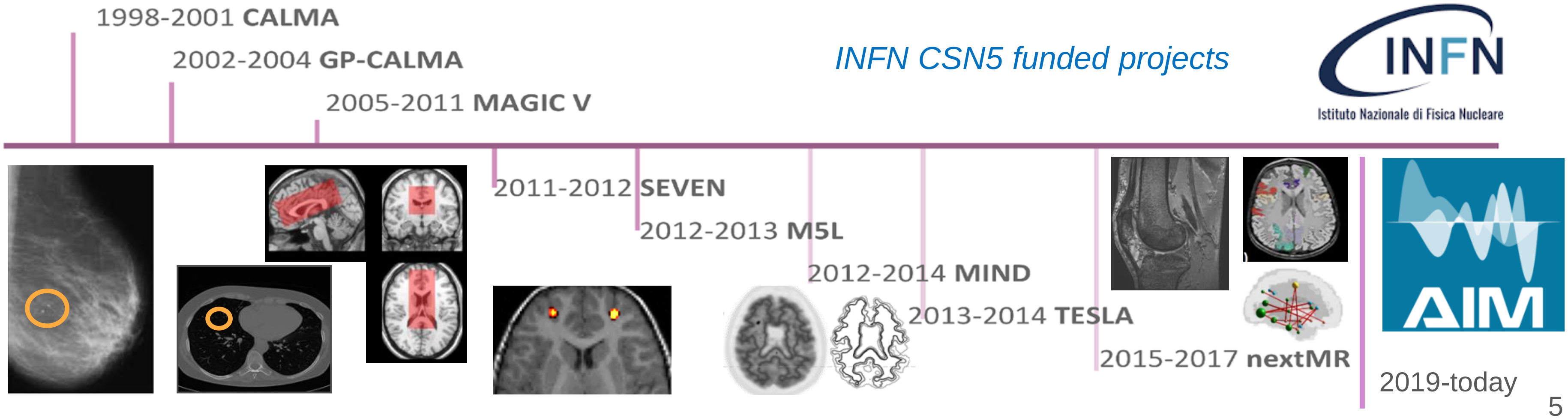
- to improve image visualization
- to detect abnormalities in diagnostic images (lesions, etc.)
- to follow up pathological conditions (growth rate of lesions)
- to evaluate the efficacy of treatment

Computer Aided Detection/Diagnosis (CAD) systems
or Decision Support Systems (DSS)
are developed to assist clinicians in their tasks, not to replace them!

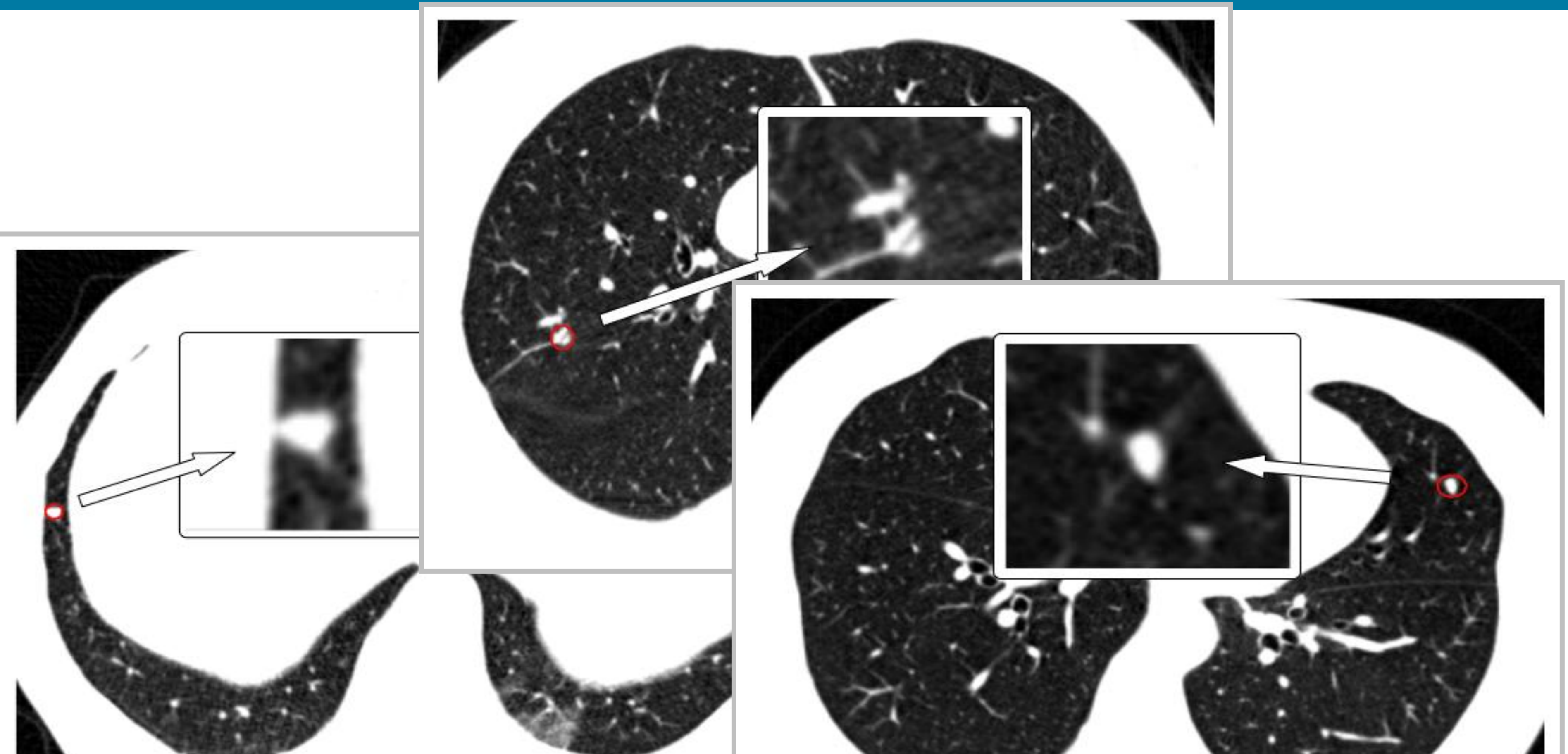
Historical overview

Artificial Intelligence (AI) methods used in the development of DSS:

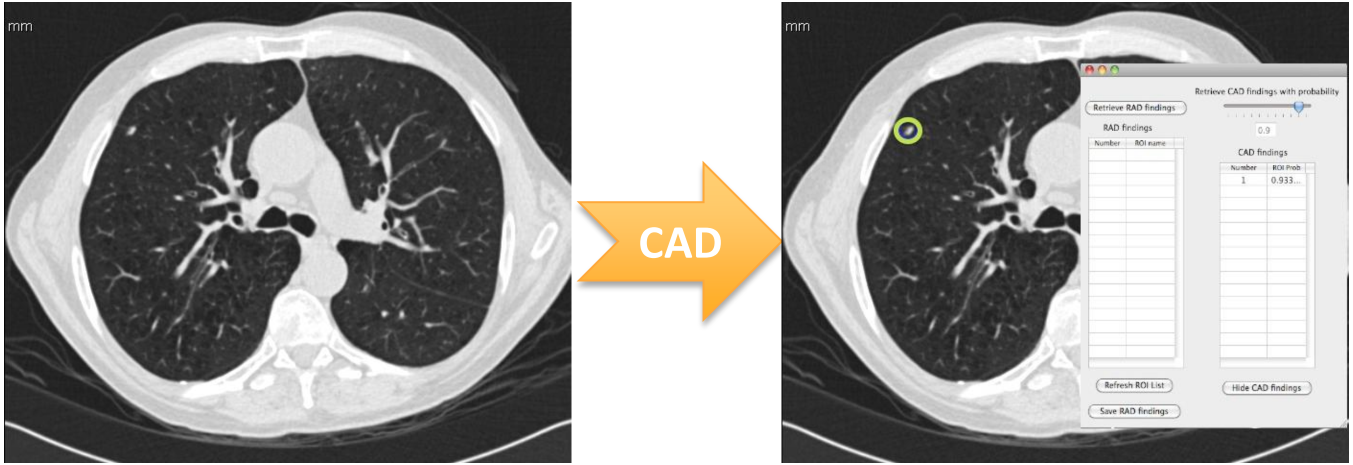
- In the 90s - Old-fashion systems (rule-based)
  - Since the 2000s - Hand-crafted feature and Machine Learning classification (Radiomics and ML)
    - Since 2015 – Deep-Learning image classification

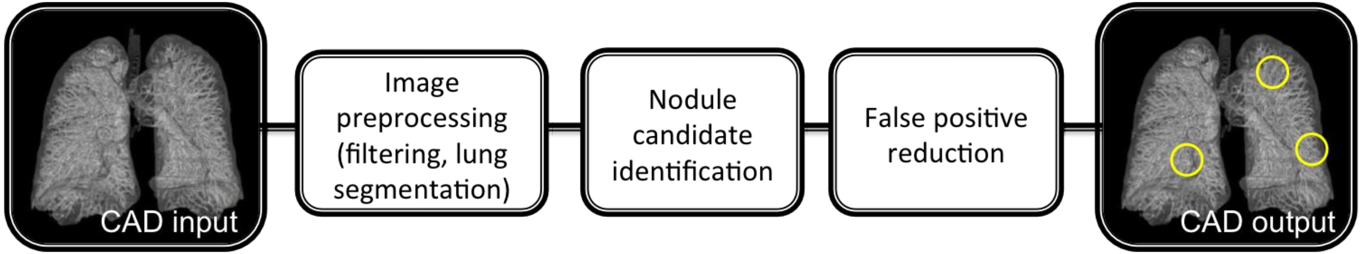


Automated detection of lung nodules in CT images



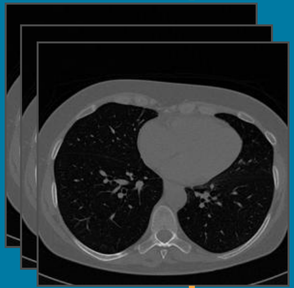
CAD system for lung nodule detection

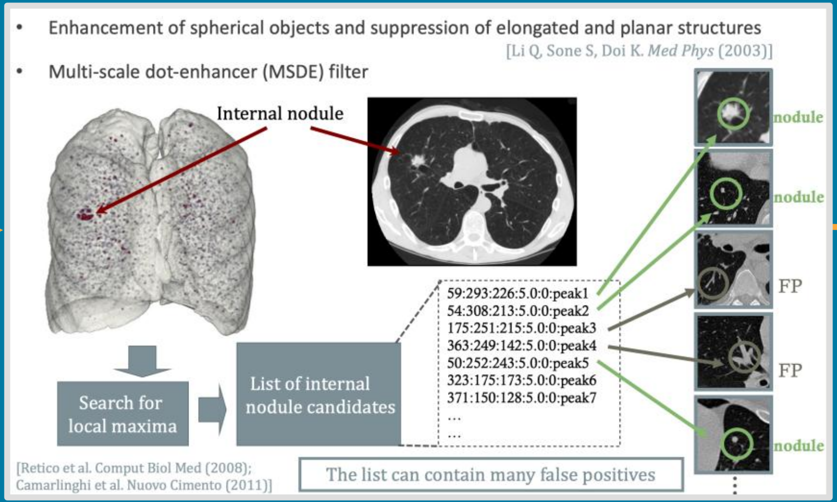


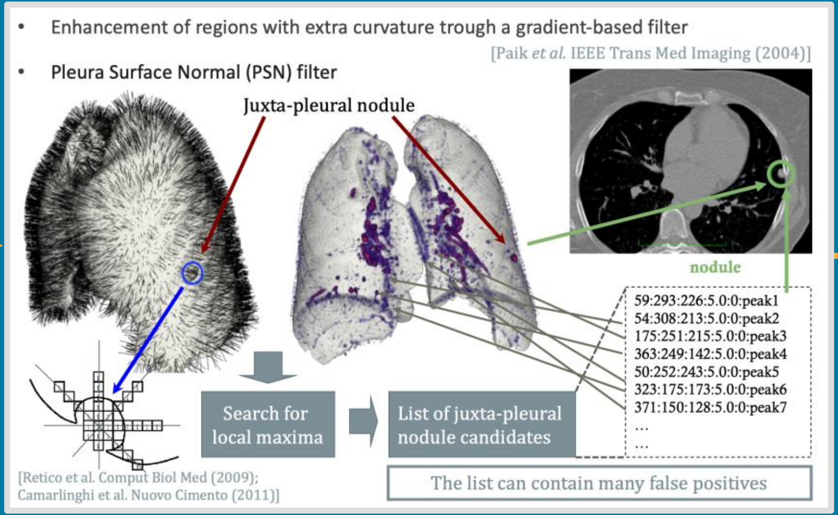


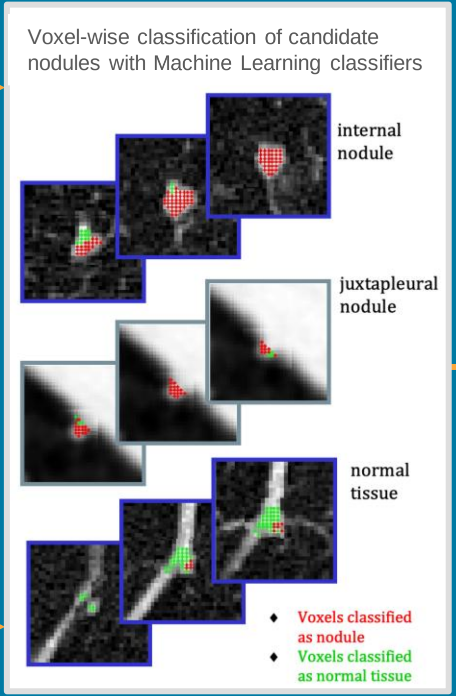
A **majority criterion** is adopted to assign candidates to either the “nodule” or the “healthy tissue” class

3D input

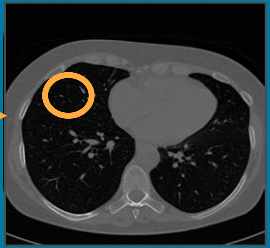








output

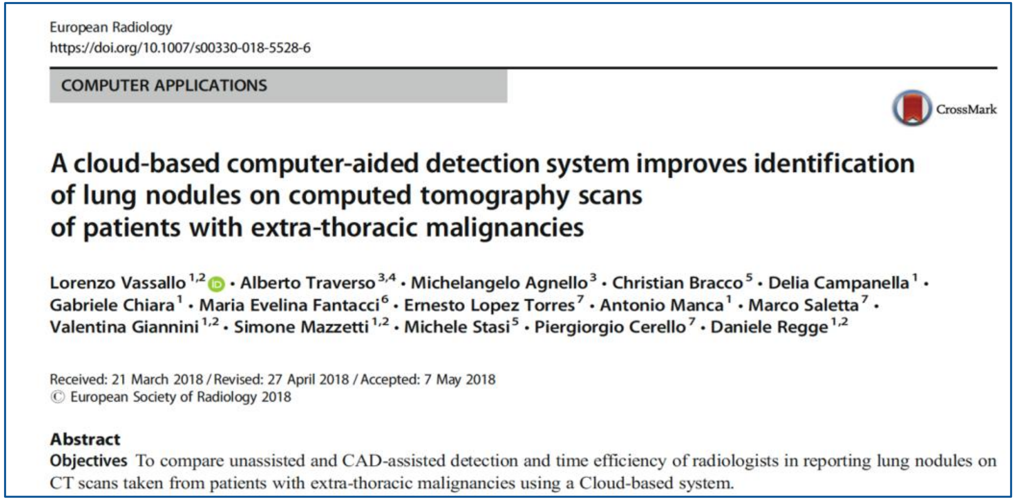


MAGIC-5 and ML5 INFN projects [2005-2010]

The system was developed in collaboration with:
- Azienda Ospedaliera Universitaria Pisana (AOUP) and the Radiology Dep. of Pisa University
- Bracco Imaging S.p. A.

M5L lung CAD on-demand

Lung nodule detection SW developed by INFN MAGIC-5 and M5L projects
→ laboratory performance: **80%** sensitivity to nodules @ **5 FP/exam**
→ **clinical validation**


European Radiology
https://doi.org/10.1007/s00330-018-5528-6

COMPUTER APPLICATIONS

CrossMark

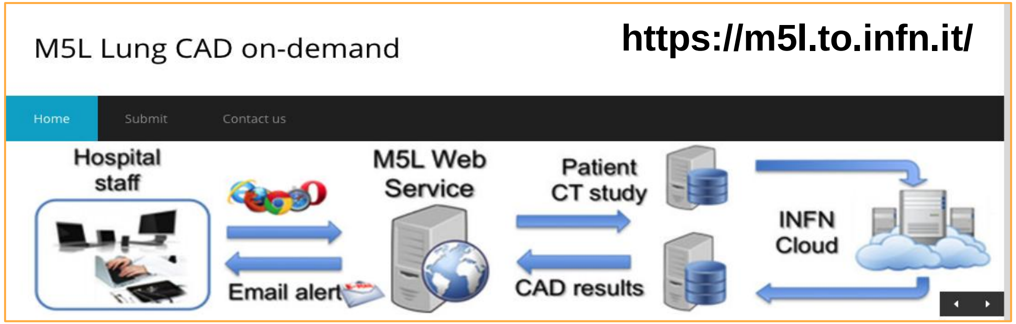
A cloud-based computer-aided detection system improves identification of lung nodules on computed tomography scans of patients with extra-thoracic malignancies

Lorenzo Vassallo[1,2] · Alberto Traverso[3,4] · Michelangelo Agnello[3] · Christian Bracco[5] · Delia Campanella[1] · Gabriele Chiara[1] · Maria Evelina Fantacci[6] · Ernesto Lopez Torres[7] · Antonio Manca[1] · Marco Saletta[7] · Valentina Giannini[1,2] · Simone Mazzetti[1,2] · Michele Stasi[5] · Piergiorgio Cerello[7] · Daniele Regge[1,2]

Received: 21 March 2018 / Revised: 27 April 2018 / Accepted: 7 May 2018

Abstract
Objectives To compare unassisted and CAD-assisted detection and time efficiency of radiologists in reporting lung nodules on CT scans taken from patients with extra-thoracic malignancies using a Cloud-based system.

Assisted reading improves nodule detection by +7% in the per-patient analysis



MAGIC-5 and M5L project leader:
P. Cerello, INFN, Turin

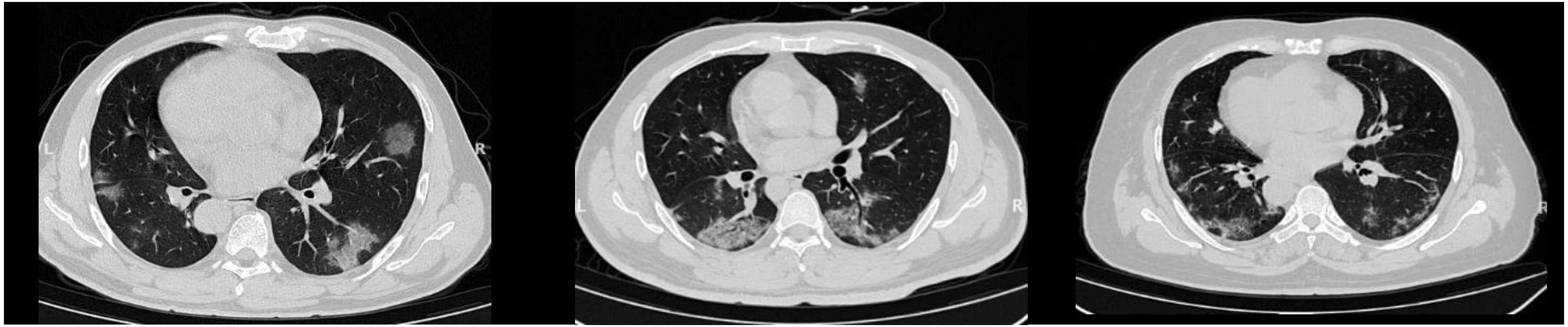
Collaboration with Candiolo Cancer Institute-FPO, IRCCS and Univ. of Turin

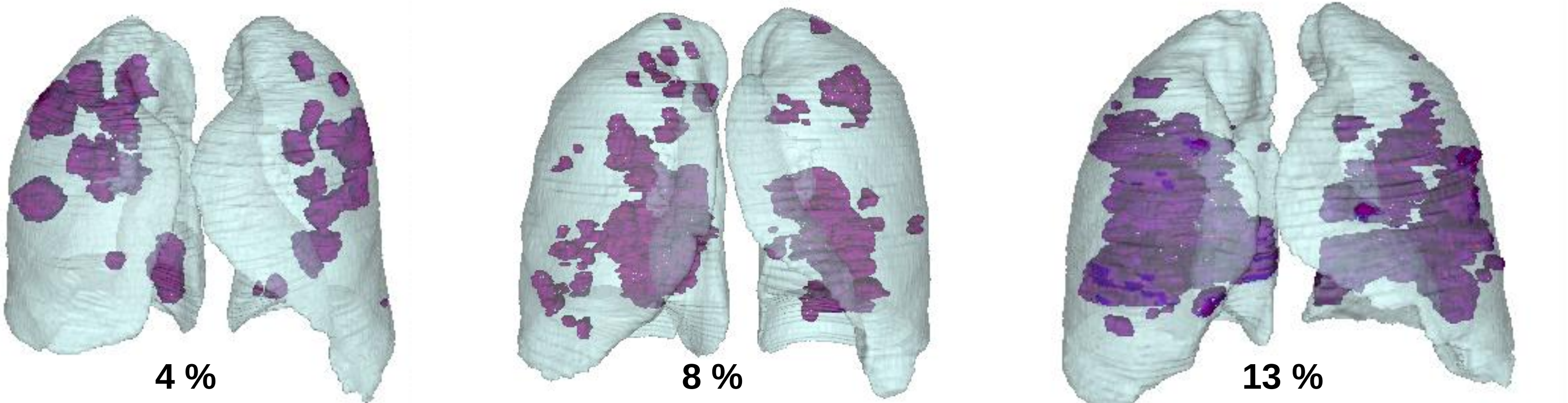
The AIM working group on lung CT analysis (AIM-Covid19-WG)

Objective: Automatic quantification of lung involvement on CT scans.

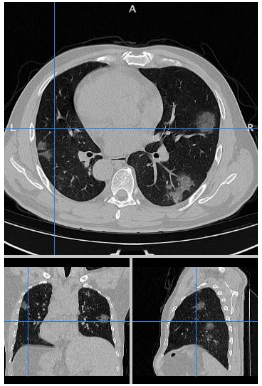
An index of severity of lung involvement has been defined [Yang, Radiology, 2020]: **CT-Severity Score (CT-SS)**

CT-SS= **1** (<5%), **2** (5%-25%), **3** (25%-50%), **4** (50%-75%), **5** (>75%)

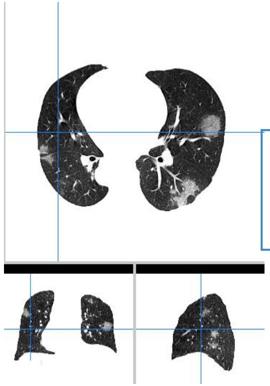




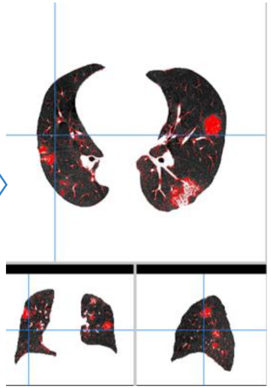
Steps for the automatic quantification of lung involvement in CT scans



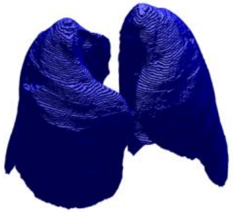
Lung volume segmentation



Quantification of lung parenchyma affected by COVID-19 lesions



Classical algorithms for lung segmentation fail when lung appearance is strongly affected by COVID-19 lesions



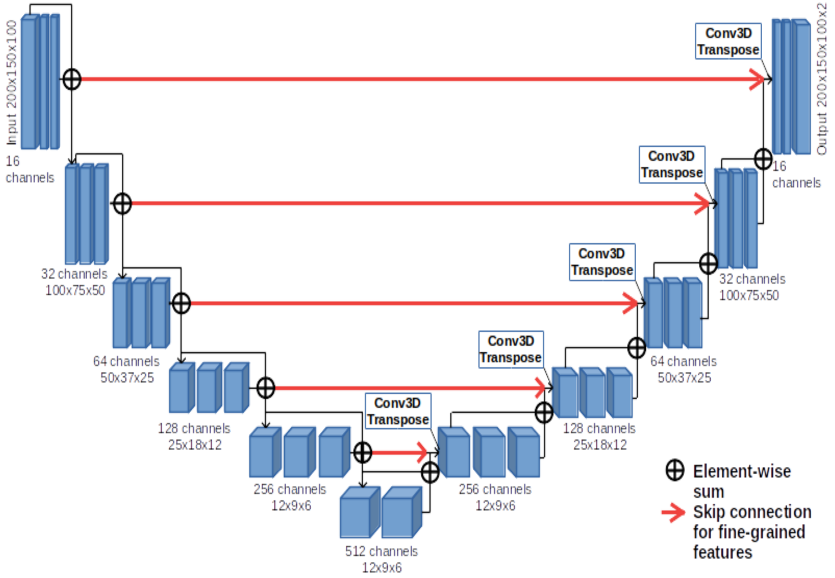
==> Deep learning segmentation methods need thousands of annotated cases to be "transferred" to accomplish this task

- Quantitative information on the amount of Covid-19 related lesions and their distribution, possibly combined with clinical and epidemiological patient's information, may be relevant to set up **predictive models for patients' stratification, prognosis prediction**, etc.
- Even only pure quantification modules, once properly validated, could be valuable tools for clinicians to set up large-scale population studies based on Radiomics

Network architecture and available datasets

Input (3D, 16-bit data): CT data resampled to 200x150x100 arrays

Target: 200x150x100 arrays; 2-bit data



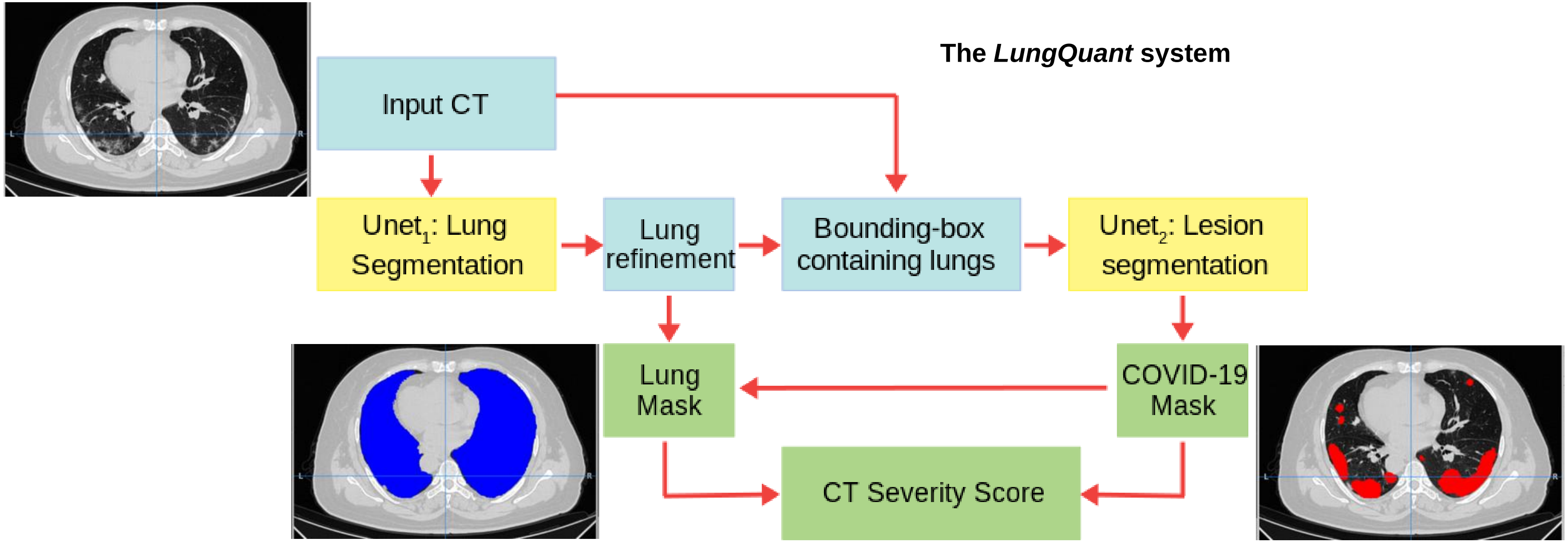
The **U-Net architecture** is outperforming other methods in most segmentation tasks
about 17 M trainable parameters

We used only **public datasets** with annotations (in part collected for other clinical purposes)

DATASETS	Clinical motivation	Number of cases	Lung mask	Lesion mask	CT-SS
COVID-19-Challenge [1]	COVID-19 pandemic	**199**	**No**	**Yes**	**No**
MosMed [2]	COVID-19 pandemic	1110	**Yes,** only for **91** CTs (made in house)	**Yes,** only for **50** CTs	**Yes**
TCIA-Plethora [3]	Lung/pleura diseases	**402**	**Yes**	**No**	**No**
TCIA-LCTSC Lung segmentation [3]	Lung cancer	**60**	**Yes**	**No**	**No**
COVID-19-CT-Seg Benchmark [4]	COVID-19 pandemic	**10**	**Yes**	**Yes**	**Yes**

[1] https://covid-segmentation.grand-challenge.org/
[2] https://mosmed.ai/
[3] https://www.cancerimagingarchive.net/
[4] https://zenodo.org/record/3757476

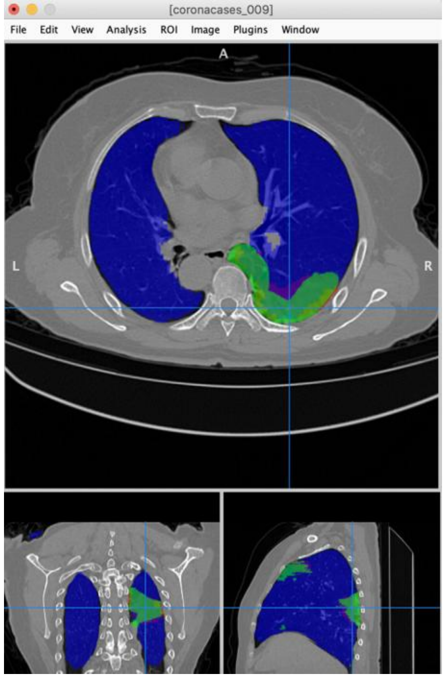
LungQuant: a sequence of two U-nets to segment lungs and COVID-19 lesions on CT scans



The *LungQuant* system

[Lizzi, F. *et al* (2021). Making data big for a deep-learning analysis: Aggregation of public COVID-19 datasets of lung computed tomography scans. *Proceedings of the 10th International Conference on Data Science, Technology and Applications, DATA 2021*, (Data), 316–321. https://doi.org/10.5220/0010584403160321]
[Lizzi, F., Agosti, A., Brero, F., Cabini, R. F., Fantacci, M. E., Figini, S., … Retico, A. (2021). Quantification of pulmonary involvement in COVID-19 pneumonia by means of a cascade of two U-nets: training and assessment on multiple datasets using different annotation criteria. *IJCARS,* https://link.springer.com/article/10.1007/s11548-021-02501-2]

The *LungQuant* system performance

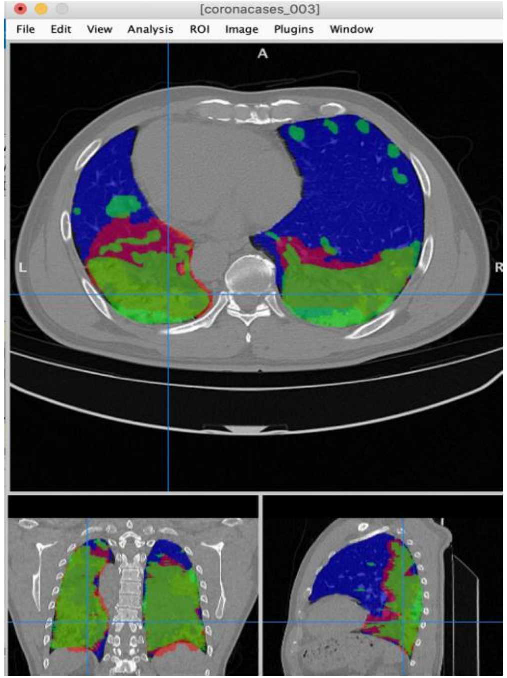


Test on the COVID-19-CT-Seg benchmark set of 10 fully annotated CT scans

best

Blue: U-net lung mask
Red: U-net lesion mask
Green: reference lesion segmentation

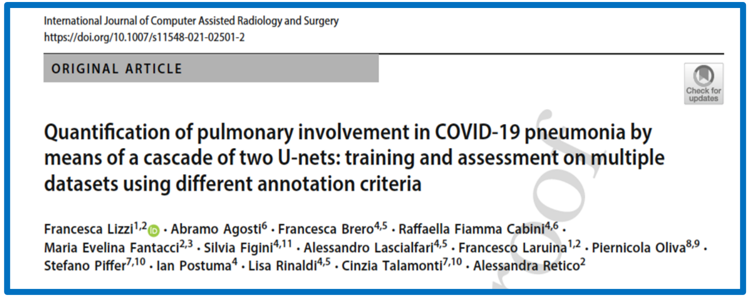
worst



Dice coefficients:

$$\mathrm{Dice}_{metric} = \frac{2 \cdot |M_{true} \cap M_{predict}|}{|M_{true}| + |M_{pred}|}$$

0.95 ± 0.01 for lung segmentation
0.66 ± 0.13 for lesion segmentation


International Journal of Computer Assisted Radiology and Surgery
https://doi.org/10.1007/s11548-021-02501-2

ORIGINAL ARTICLE

Check for updates

Quantification of pulmonary involvement in COVID-19 pneumonia by means of a cascade of two U-nets: training and assessment on multiple datasets using different annotation criteria

Francesca Lizzi[1,2] · Abramo Agosti[6] · Francesca Brero[4,5] · Raffaella Fiamma Cabini[4,6] · Maria Evelina Fantacci[2,3] · Silvia Figini[4,11] · Alessandro Lascialfari[4,5] · Francesco Laruina[1,2] · Piernicola Oliva[8,9] · Stefano Piffer[7,10] · Ian Postuma[4] · Lisa Rinaldi[4,5] · Cinzia Talamonti[7,10] · Alessandra Retico[2]

F. Lizzi et al. IJCARS,
doi: 10.1007/s11548-021-02501-2

Clinical validation:
Scapicchio C. *et* al.,A multicenter evaluation of a deep learning software (LungQuant) for lung parenchyma characterization in COVID-19 pneumonia, European Radiology Experimental, (2023) 7:18

Deep Learning vs. traditional Machine Learning approaches

- Deep Neural Networks are replacing traditional handcrafted feature extraction + ML approaches in many Medical Physics applications
  - **Pros:**
    - **No prior selection of problem-related features $\Longrightarrow$ no loss of information**
  - **Cons:**
    - **Larger and larger samples of annotated data are needed to train the models**
    - **Deep Neural Networks are black boxes: which image features are relevant for making a decision?**

Data augmentation

Model interpretability, explainable AI

Mandatory in medical applications

Critical aspects of DL use in medical image analysis

Problems with clinical data

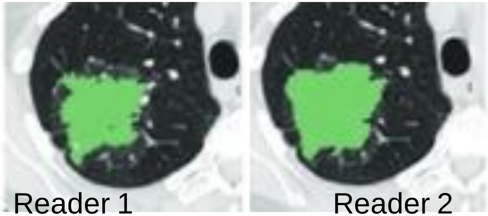
- Annotation of the dataset (ground truth)
- Inadequate dataset size
    - Appropriate size for DL/ML training
    - Sampling bias
    - Unknown dimension
    - Batch effect

Problems of the software

- Reliability (out of the lab)
- Explainability of the results

The “true label” problem

- Data need to be annotated!
- Data annotation by human experts is an extremely time-consuming task, which may require:
    - the collection of additional information stored in other data storing systems,
    - expertise in segmenting meaningful regions in images,
    - specific knowledge to assign class labels.

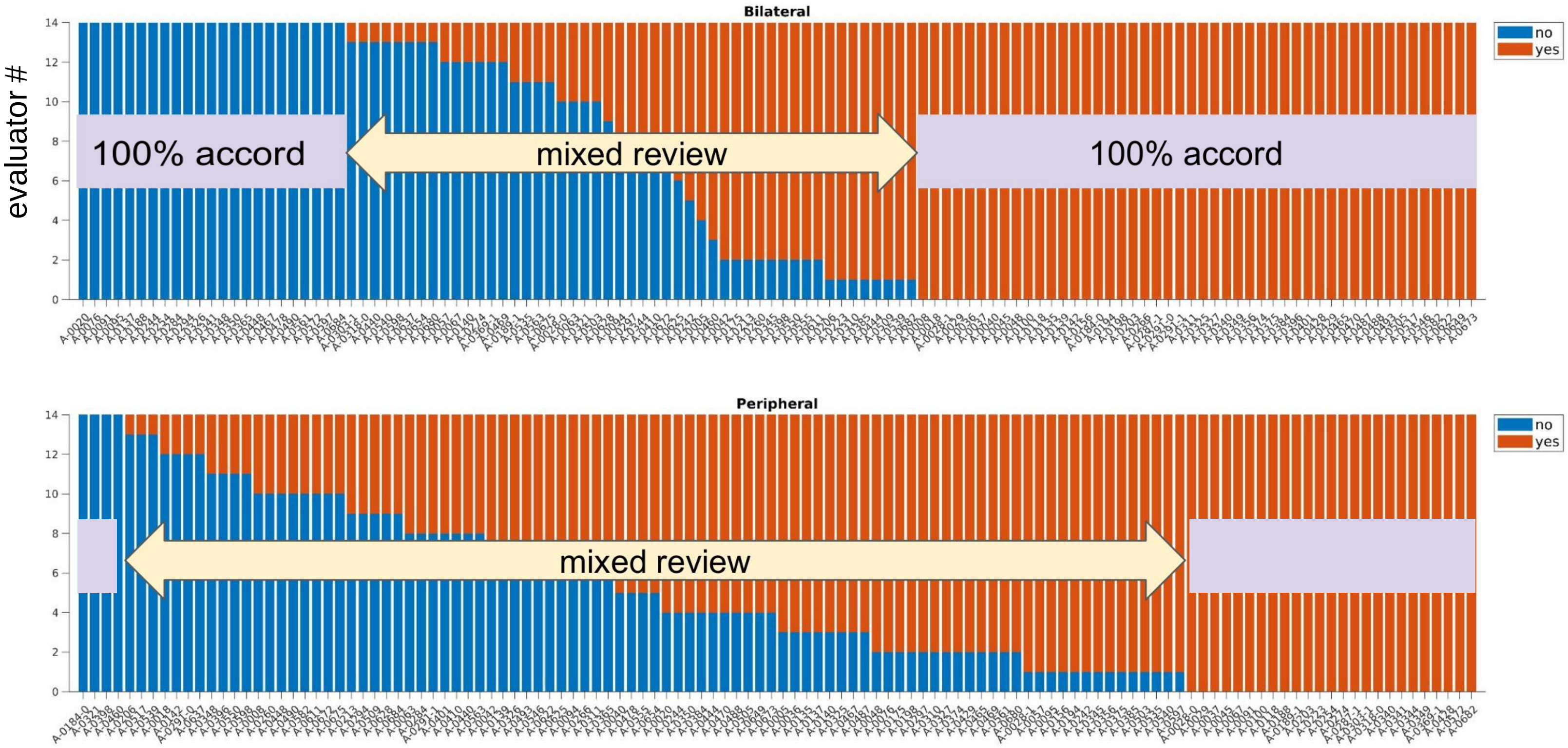


- In the medical imaging field, segmentation of organs or lesions can be affected by inter- and intra-reader variability.

- Datasets are often evaluated by **only one human expert**

- Gathering data and annotations from many sources increases the heterogeneity of the sample

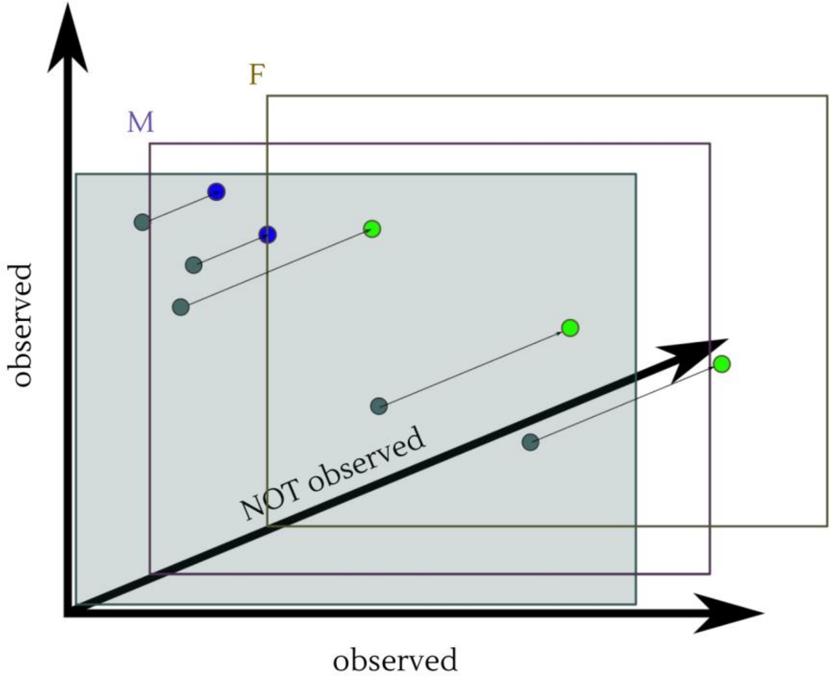
The “true label” problem: an example from COVID-19



Scapicchio C. *et* al.,A multicenter evaluation of a deep learning software (LungQuant) for lung parenchyma characterization in COVID-19 pneumonia, European Radiology Experimental, (2023) 7:18 https://doi.org/10.1186/s41747-023-00334-z

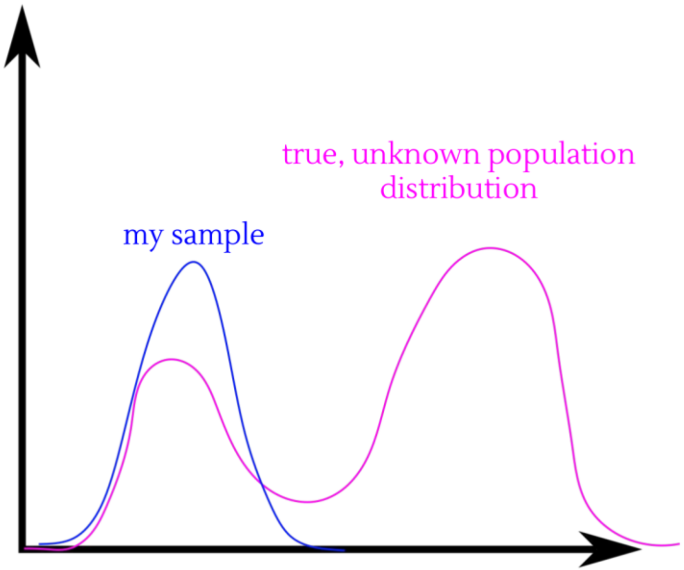
The “unobserved dimensions” problem

- there are several unobserved variables with relevant implication in the data (if they were observed)
- rules learned on the dataset are not trustworthy
    - Examples:
        - Gender
        - Ethnicity
        - Comorbidities



The sampling bias

- Also for a defined pathology, significant differences may occur in in-patient statistics both among nations and within centers
- Several factors affect these differences, which are difficult to control, in particular in retrospective studies:
    - Regional differences in population
    - Different acquisition systems and procedures
    - Small size of the datasets
- Multicentric datasets may help to reduce this problem



Multicentric dataset in Autism Spectrum Disorders

Autism spectrum disorder (ASD)

- ASD is a heterogeneous neurodevelopmental condition with a consistently high prevalence worldwide.
- Early diagnosis is crucial for intervention
- ML techniques have been widely used on MRI data, with the goal of identifying the main brain areas involved and consequently facilitating the diagnostic process.
- In this field, large datasets are often obtained by collecting images from different centers

Dataset

ABIDE
Autism Brain Imaging Data Exchange

- The Autism Brain Imaging Data Exchange (ABIDE)
- Public dataset, 24 collection centers
- MRI, structural and functional
- Retrospectively collected data
- More than 2000 subjects (equally divided between ASD and TD)
- Ages: 5-64 years

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

Harmonization of multicenter data in the study of Autism Spectrum disorders (ASD)

Caltech_51456_mprage.nii

Caltech_51463_mprage.nii

CMU_50642_mprage.nii

CMU_50649_mprage.nii

KKI_50778_mprage.nii

NYU_50957_mprage.nii

Data gathered by different scanner and/or acquisition systems encode the site "signature", which can confound ML algorithms and hide subtle information of interest.

ABIDE

Autism Brain Imaging Data Exchange (2200 MRI scans, 40 acquisition sites)

AUC	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

ML classifiers can easily distinguish brain features of subjects from site A vs. site B (AUC ~1), whereas barely distinguish ASD vs. controls (AUC~0.6).

Artificial Intelligence in Medicine 108 (2020) 101926

Contents lists available at ScienceDirect

Artificial Intelligence In Medicine

journal homepage: www.elsevier.com/locate/artmed

ELSEVIER

Dealing with confounders and outliers in classification medical studies: The Autism Spectrum Disorders case study

Elisa Ferrari[a,*], Paolo Bosco[b], Sara Calderoni[b,c], Piernicola Oliva[d,e], Letizia Palumbo[f], Giovanna Spera[f], Maria Evelina Fantacci[f,g], Alessandra Retico[f]

How to mitigate site effects?

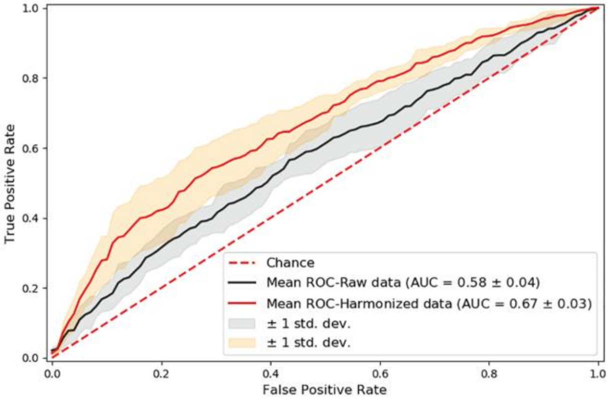
The site contribution to can be modelled and discarded, while keeping interesting data dependencies (e.g. on age and sex)

$$Y^*_{ijk} = (Y_{ijk} - f_k\,(x_{ij}, z_{ij}, w_{ij}) - g^*_{ik}) \,/\, d^*_{ik} + f_k\,(x_{ij}, z_{ij}, w_{ij})$$

non-linear function of age, sex, TIV

location

scale

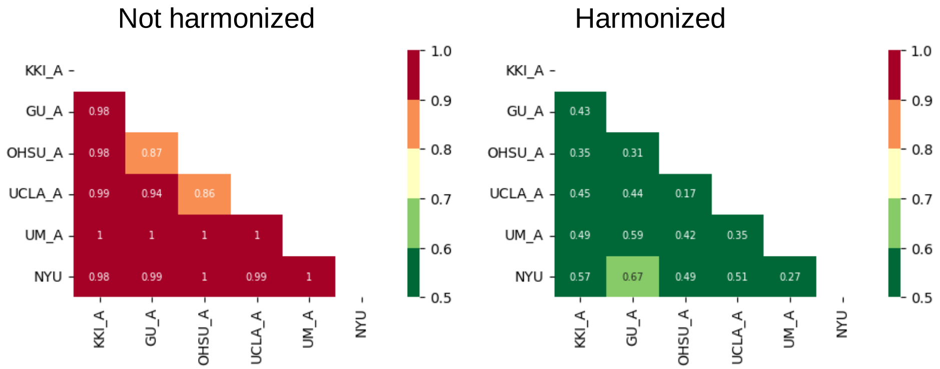


The case vs. control separation ability of the ML classifiers is significantly improved

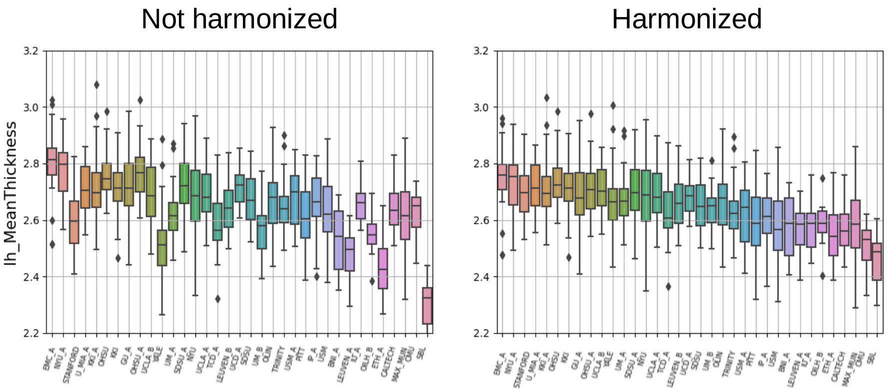
S. Saponaro, A. Giuliano, R. Bellotti, A. Lombardi, S.Tangaro, P. Oliva, S. Calderoni, A. Retico, Multi-site harmonization of MRI data uncovers machine-learning discrimination capability in barely separable populations: An example from the ABIDE dataset, NeuroImage: Clinical 35 (2022) 103082

Harmonization

Site identification



Age dependence



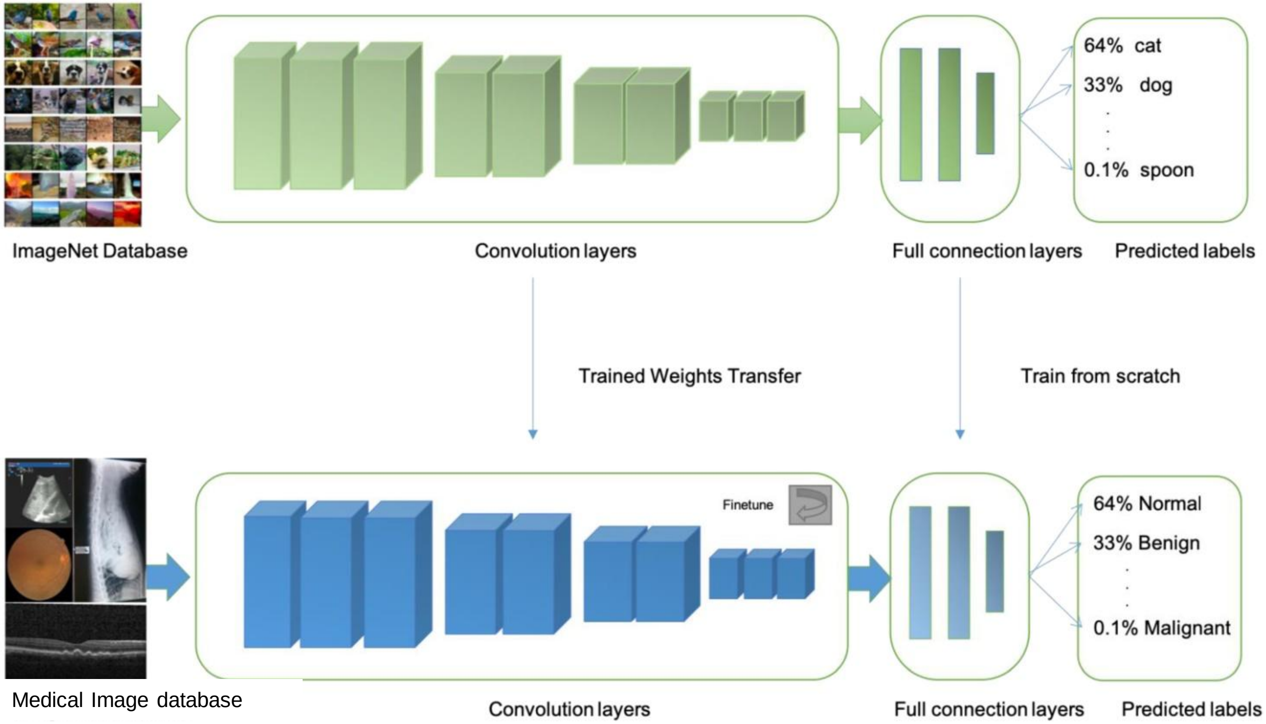
Sites are sorted by increasing average age

Limited availability of annotated data: Transfer learning

In case of **small datasets**
[i.e. when # of training examples << # of trainable parameters]
we can avoid training DL models from scratch and take advantage of the knowledge already acquired on other data and/or in other tasks

- **Transfer Learning**

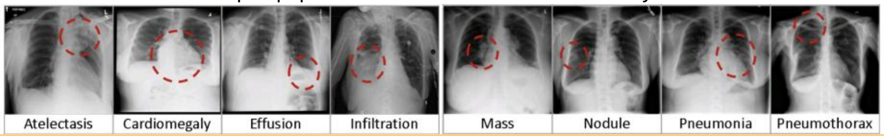
DenseNet121, ResNet50, Inception are widely used pretrained Deep Neural Networks.
Typically, they are trained on ImageNet



[Xu et al. Current status and future trends of clinical diagnoses via image-based deep learning. *Theranostics*, *9*(25), 7556–7565 (2019)]

Transfer learning (TL)

https://paperswithcode.com/dataset/chestx-ray14



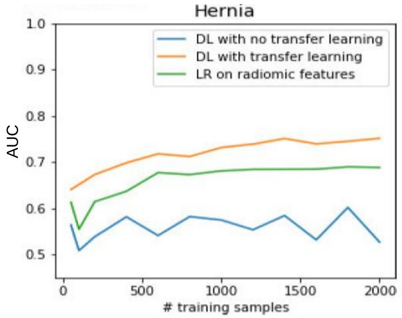
…

Comparison of three different TL methods, using DenseNet121, and different training dataset sizes and different classification tasks.

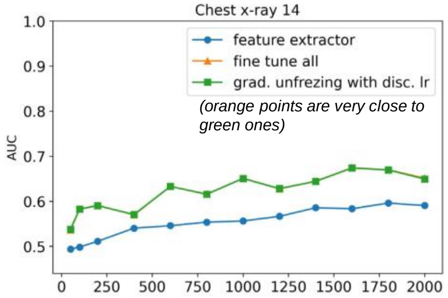
Results:

- Traditional ML can perform better that DL for small datasets; if DL is used, TL performs better.
- Fine-tune performs better than feature extractor
- Features learned may not be as general as currently believed:
  - TL from models trained on similar images from different anatomical site is equivalent to using ImageNet
- TL is useful for small datasets (N < 2000)

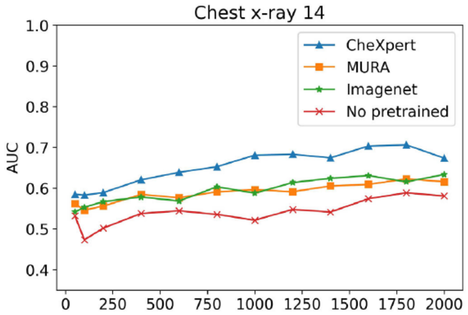
Traditional ML vs DL (w and w/o TL)



Different TL methods



Similarity between source and target datasets



CheXpert: Chest X-ray images
MURA: Musculoskeletal RX images (elbow, finger, forearm, hand, humerus, shoulder, and wrist)
ImageNet: natural images

[Romero et al. Targeted transfer learning to improve performance in small medical physics datasets. *Medical Physics*, *47*(12), 6246–6256 (2020)]

Limited availability of annotated data: Data augmentation

Synthetic data generation with GAN

Generative adversarial networks (GAN) can generate plausible images via the adversarial training of a generator **G** and a discriminator **D**.

- Adversarial training refers to the competition between the two networks **G** and **D**.
- An equilibrium is eventually reached, where the generator can approximate data from the target data distribution and the discriminator predicts ‘real’ or ‘generated’ for its input data with 50% probability.
- Realistic **synthetic data** can be generated by the generator via sampling the fixed distribution p(z) for data augmentation.

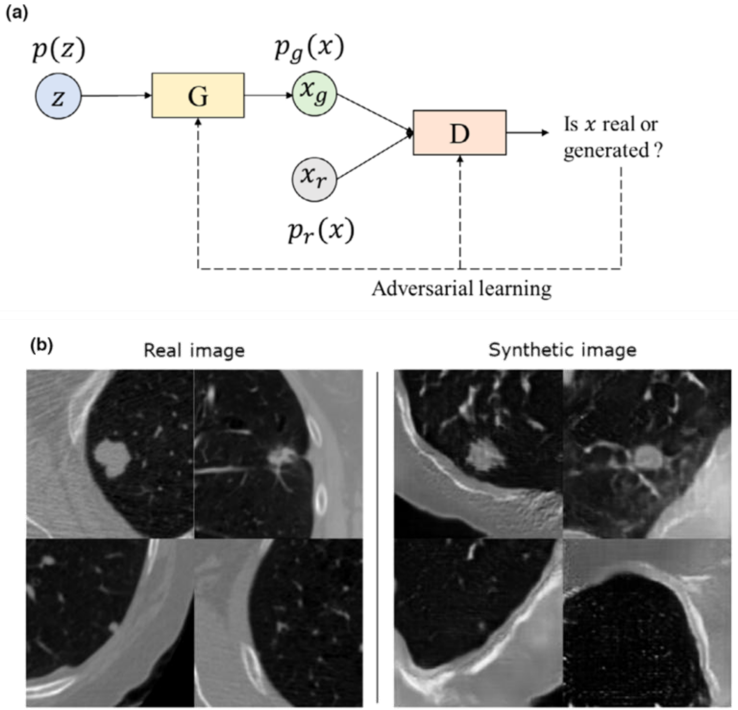


Fig. 5. (a) The diagram of a basic GAN. (b) Real CT images from the LIDC lung nodule dataset[12] and synthetic images generated by a GAN network.

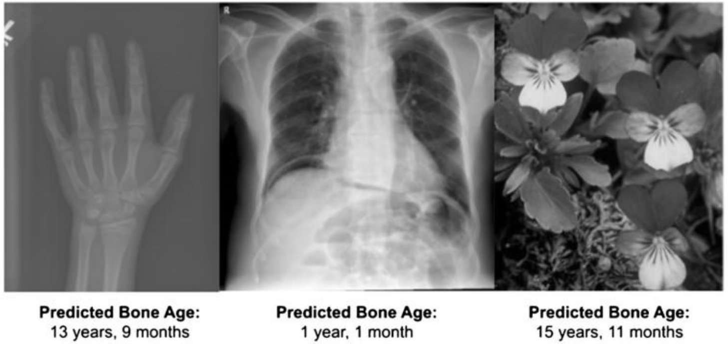
[Chlap, P., Min, H., Vandenberg, N., Dowling, J., Holloway, L., & Haworth, A. (2021). A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, *65*(5), 545–563. https://doi.org/10.1111/1754-9485.13261]

Reliability of AI systems

- What happens when an AI algorithm trained for a specific task is executed on “inappropriate input data”?
    - Typically, it provides its prediction!!!

[Yi et al (2022). Can AI distinguish a bone radiograph from photos of flowers or cars? Evaluation of bone age deep learning model on inappropriate data inputs. *Skeletal Radiology*, *51*(2), 401–406. https://doi.org/10.1007/s00256-021-03880-y]

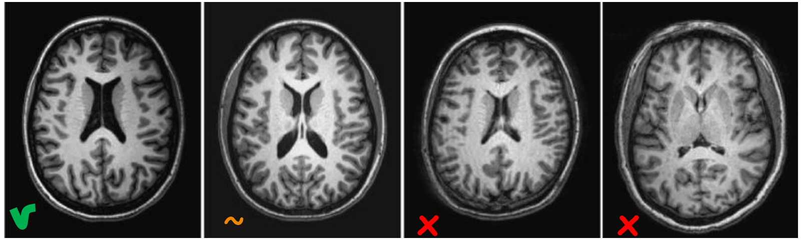
Outputs of a CNN trained to predict bone age from RX of left hands



- To avoid feeding an AI algorithm with a wrong input:
    - Image type/quality can be evaluated by another AI algorithm, and possibly discarded if not appropriate

[Fantini et al. (2021). Automatic MR image quality evaluation using a Deep CNN: A reference-free method to rate motion artifacts in neuroimaging. Computerized Medical Imaging and Graphics, 90, 101897. https://doi.org/10.1016/j.compmedimag.2021.101897]

Motion-free vs motion corrupted images



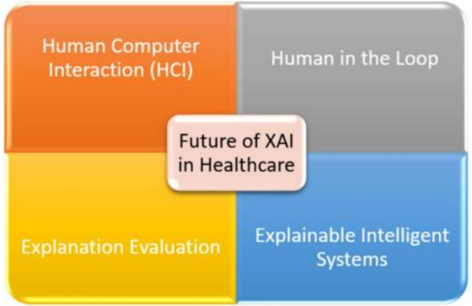
Explainability

Trusting the algorithm

- AI systems are often seen as objective and unbiased
- their complexity and technical nature can make them seem more credible and trustworthy
- success in other scientific fields

This is unacceptable

- **For scientists**
  - Lack of critical thinking
  - Needs to understand cause-effect relationship
- **In clinical practice**
  - For the same reasons!
  - Ethical (and legal) issues in providing diagnosis by a back-box system



Reliable XAI is still an open field…

Conclusions

- Medical imaging daily produces an incredible amount of digital information which is not fully exploited neither for diagnosis/therapy nor for research!

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

- The development of AI-based clinical DSS has multiple levels of complexity, thus it requires multidisciplinary skills

  - There is still lot of room to make original contributions in this field of research!

Thank you for your attention!

Contact: oliva@uniss.it
Università di Sassari
INFN, Sezione di Cagliari