An overview of Machine Learning in Medicine and Medical Physics

Alessandra Retico
Istituto Nazionale di Fisica Nucleare
Sezione di Pisa
alessandra.retico@pi.infn.it

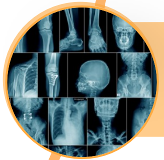
ML_INFN Hackathon, Pisa, November 13-16, 2023

Outline


Artificial Intelligence in Medicine: focus on Medical Imaging


Segmentation, Classification, Radiomics, Predictive Models


Open issues and challenges


Examples and case studies from real-world datasets

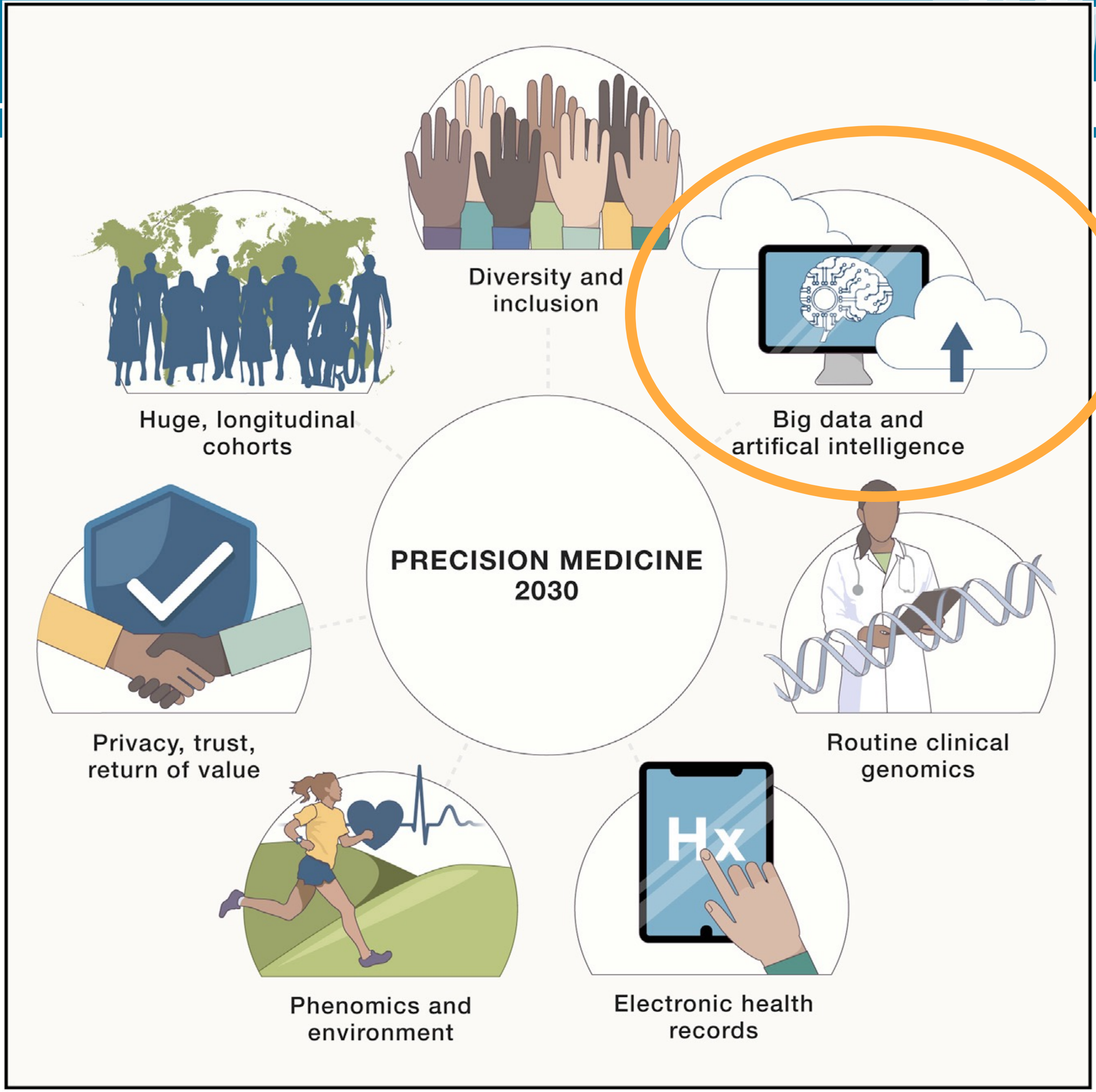

Future goals and perspectives

The goal: Precision Medicine

Precision medicine promises improved health by accounting for individual variability in genes, environment, and lifestyle.

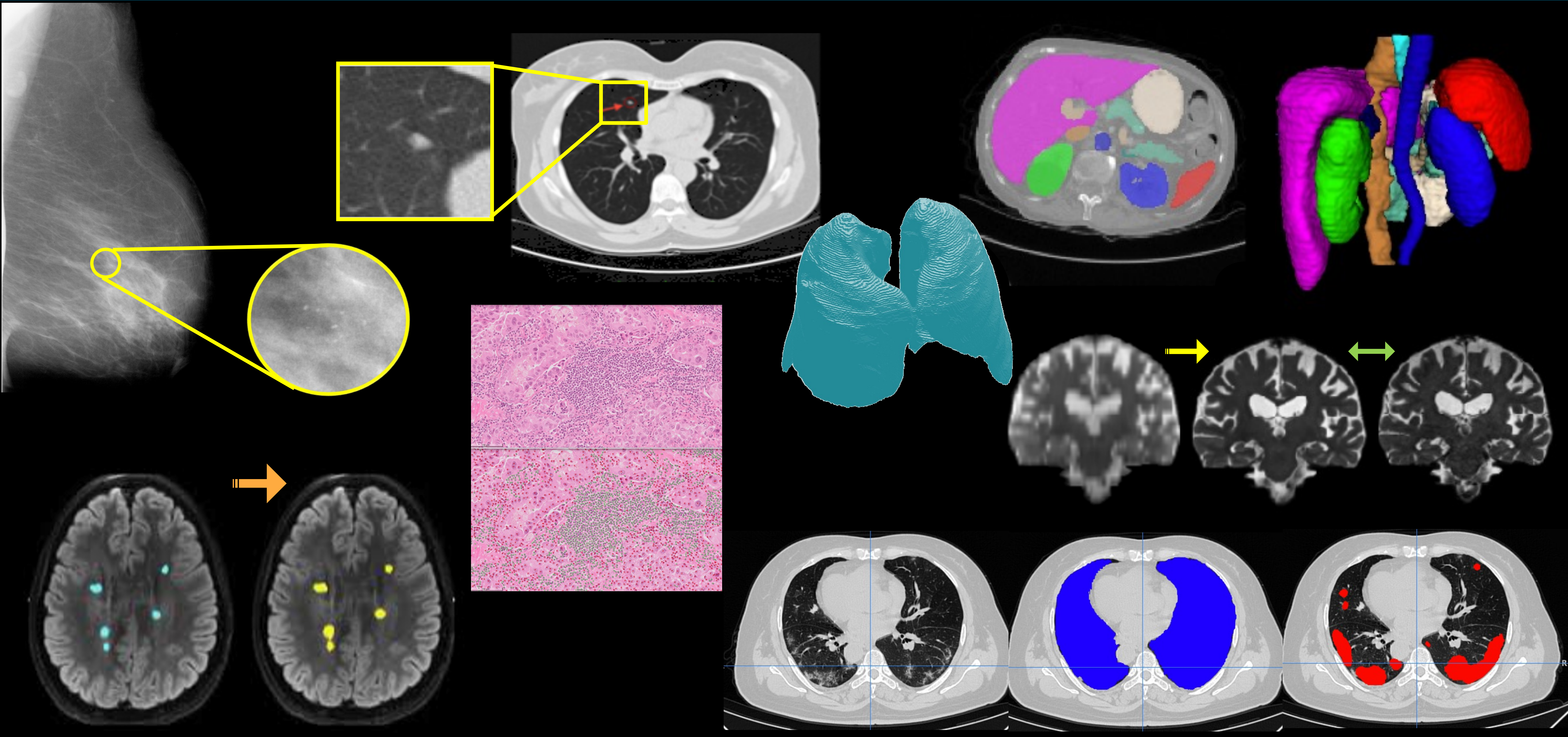
Precision medicine will continue to transform healthcare in the coming decade as it expands in key areas:
- huge cohorts,
- artificial intelligence (AI),
- routine clinical genomics,
- phenomics and environment,

returning value across diverse populations.

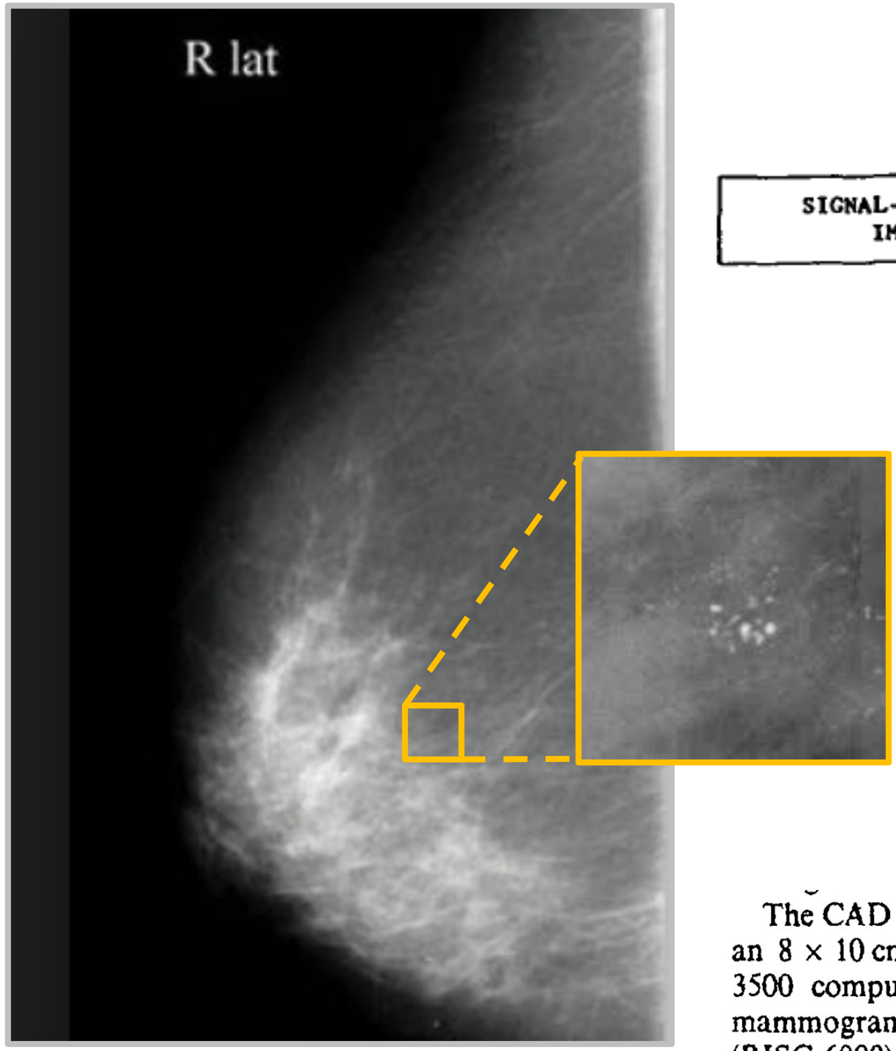


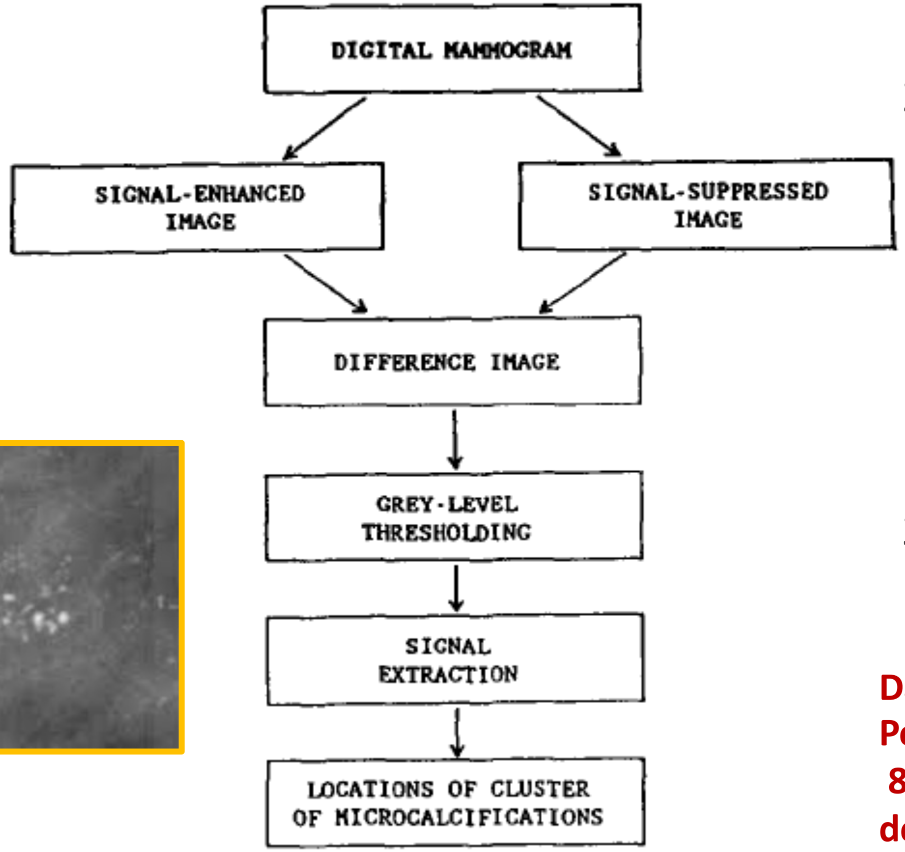
[Denny and Collins, Precision medicine in 2030—seven ways to transform healthcare. Cell 2021;184:1415–9. https://doi.org/10.1016/j.cell.2021.01.015.]

Artificial Intelligence (AI) in Medical Image Analysis



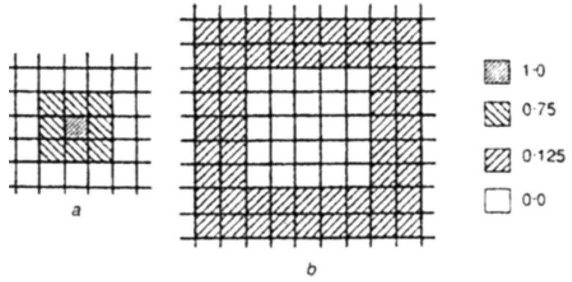
Old-fashion rule-based automated decision systems





The CAD scheme takes approximately 13 min to analyse an 8 × 10 cm region of a mammogram on a DEC VAX 3500 computer. Preliminary results indicate that a full mammogram can be analysed in less than 40 s on an IBM (RISC 6000) Powerstation 560.

1) Image digitization
2) Image filtering



Kernels used for (a) enhancement filter and (b) suppression filter

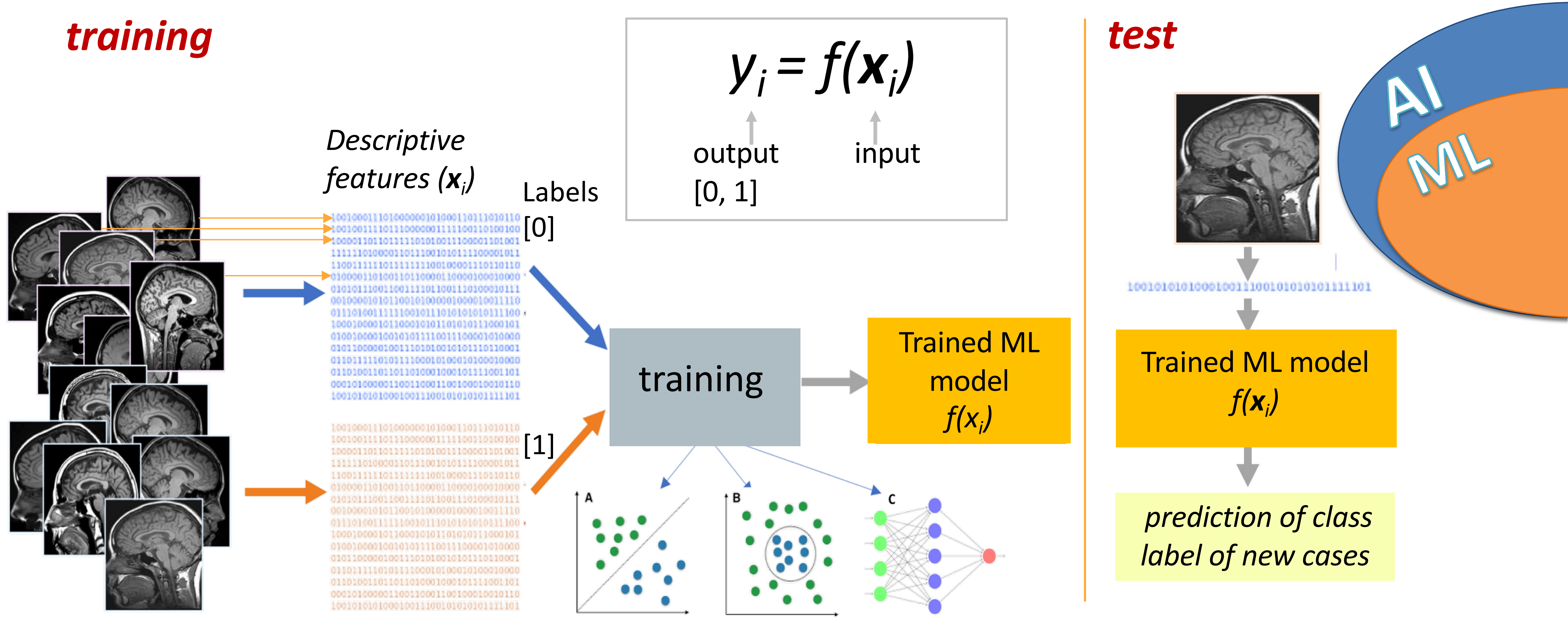
3) Decisional systems based on the codification of a series of rules

Dataset: 78 mammograms
Performance:
85% sensitivity @ 2 false positive (FP) detection per mammogram

AI

[Chan HP et al., "*Image feature analysis and computer-aided diagnosis in digital radiography. 1. Automated detection of microcalcifications in mammography,*" Med. Phys. 14, 538–548, 1987]

Machine Learning (ML) models

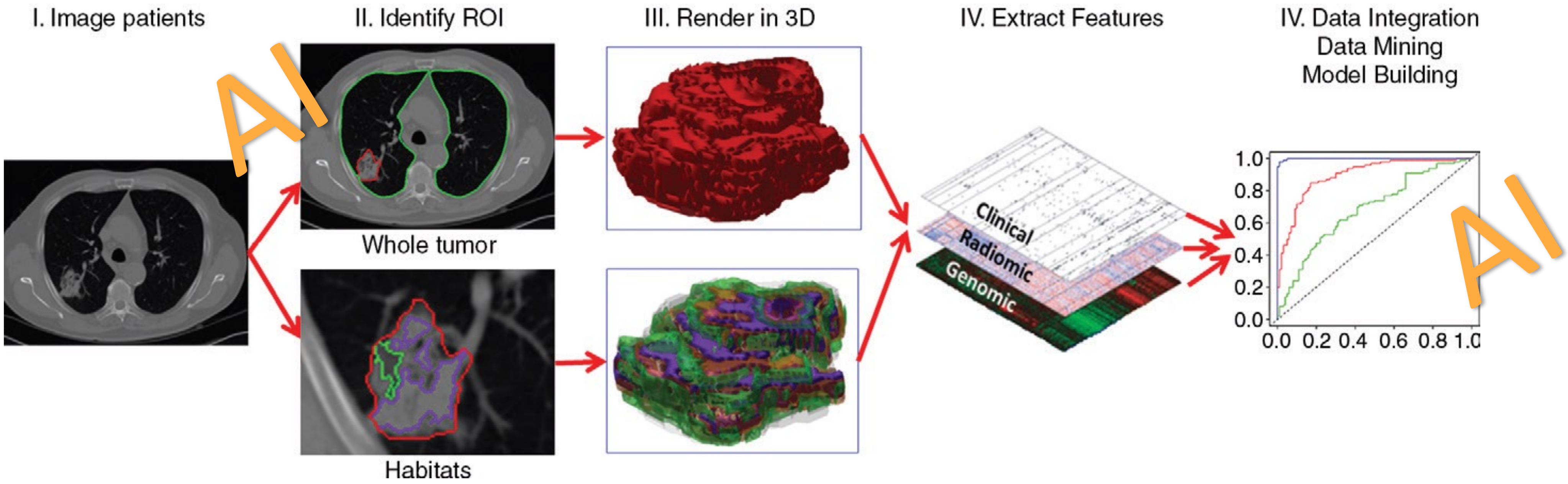


Radiomics and Machine Learning (ML)

Radiomics

extracting quantitative features from medical images

Gillies, R. J., Kinahan, P. E. & Hricak, H. Radiomics: Images Are More than Pictures, They Are Data. *Radiology* **278**, 563–577 (2015).



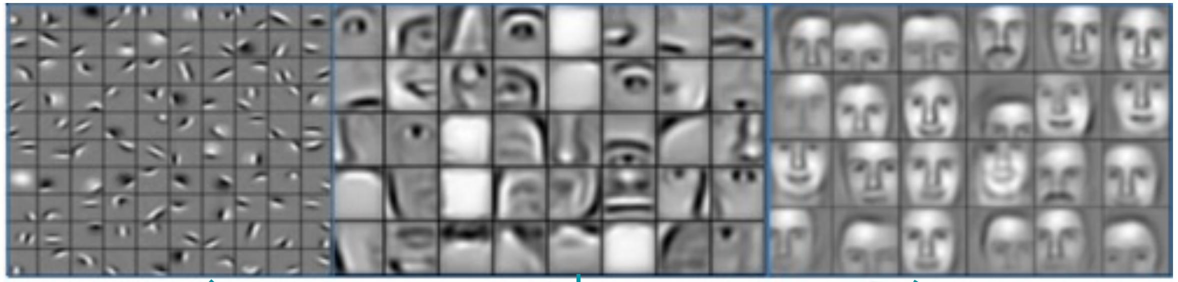
Radiomics and **AI**, in particular **ML** and **DL**, allow us to develop **predictive models** of patients' diagnosis, prognosis, prediction of treatment efficacy or any other outcome of interest

Convolutional Neural Networks (CNN)

https://towardsdatascience.com

- 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 brain seems to do.

Convolutional Neural Networks (CNN) learn in multiple levels of representation, corresponding to different levels of abstraction



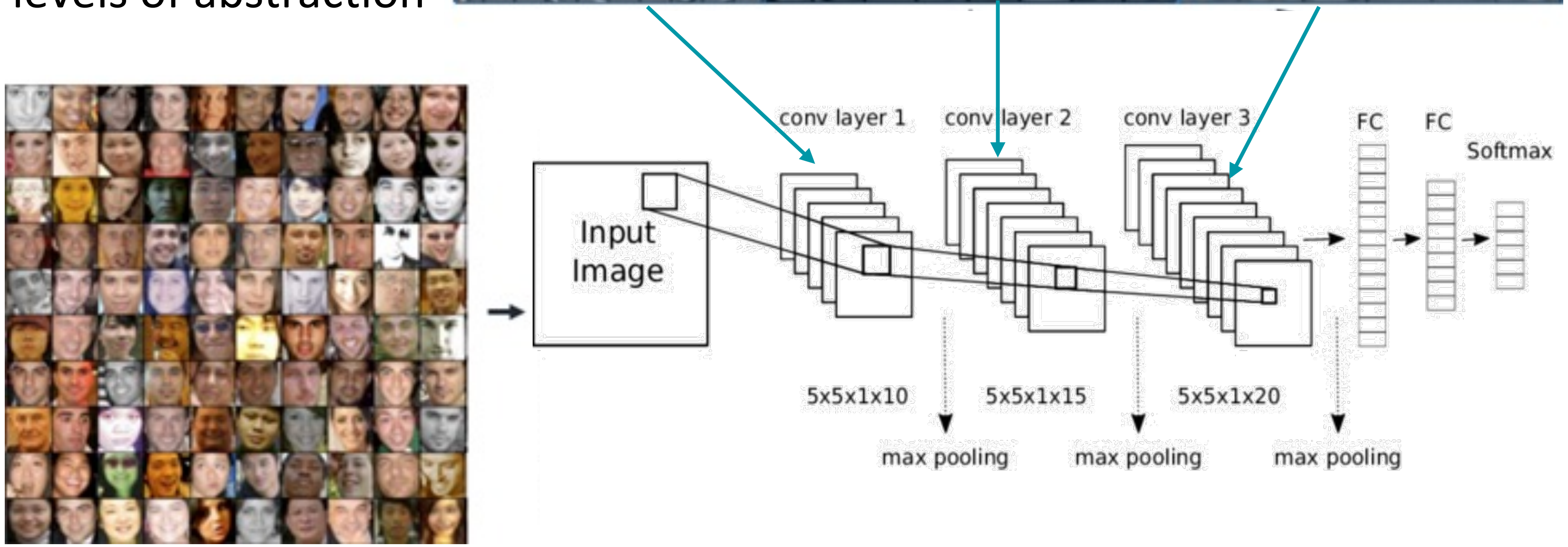
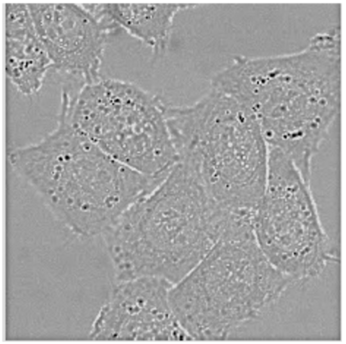
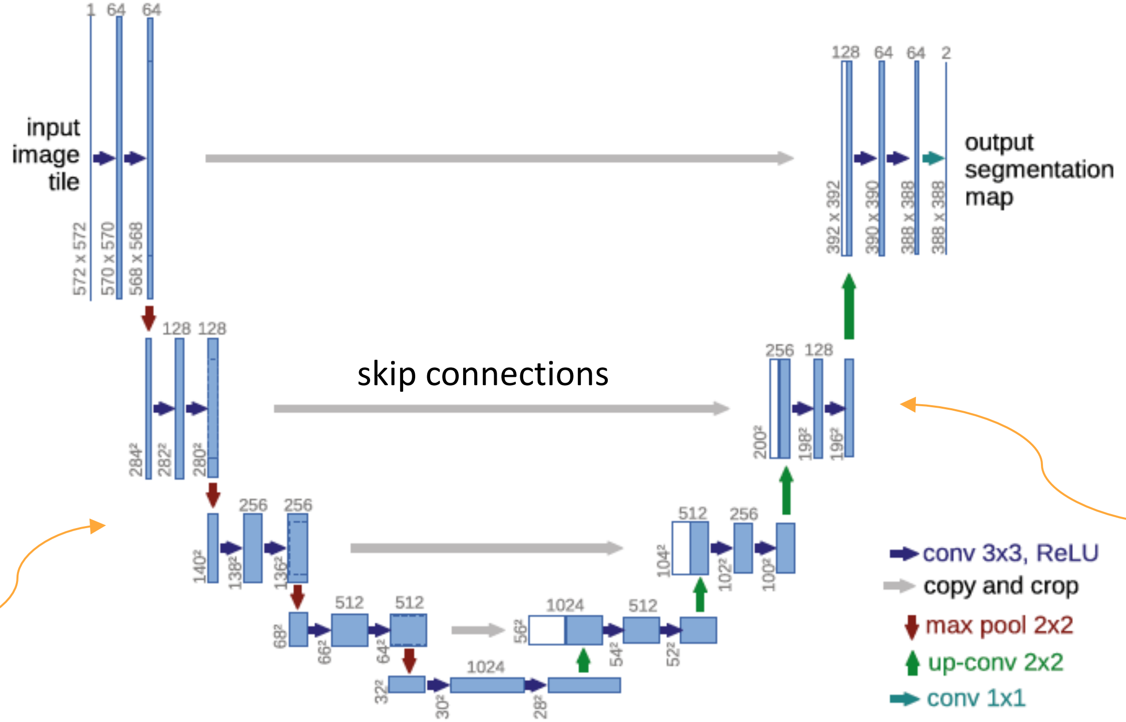


Image to label conversion [classification]

Deep learning for image segmentation: U-nets

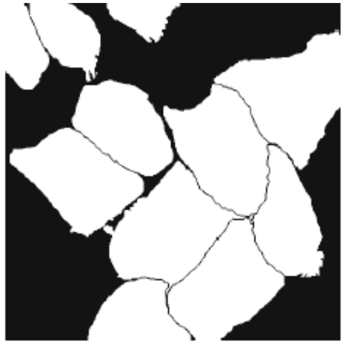


HeLa cells



encoder

decoder



segmentation masks

Image to image conversion [segmentation]

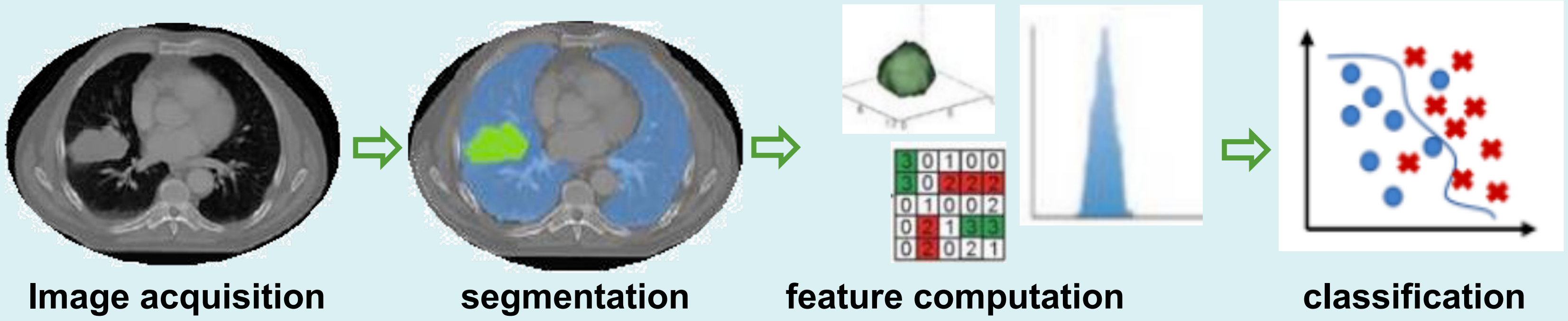
The U-net deep learning network design demonstrated superior capabilities in image segmentation in a large variety of segmentation tasks, including medical images

[Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. Lecture Notes in Computer Science 9351, 234–241 (2015). DOI 10.1007/978-3-319-24574-4 28]

Artificial Intelligence (mainly ML and DL) in Medicine

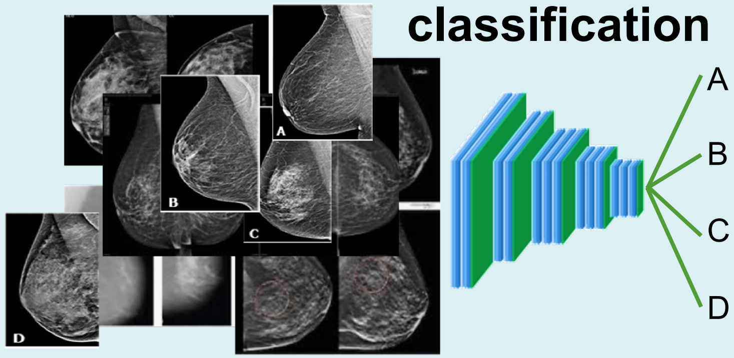
In **medical image analysis** a large variety of approaches based on AI can be developed according to different goals, e.g. image segmentation, image classification, building predictive models based on both images and additional patient information.

Radiomics + Machine Learning

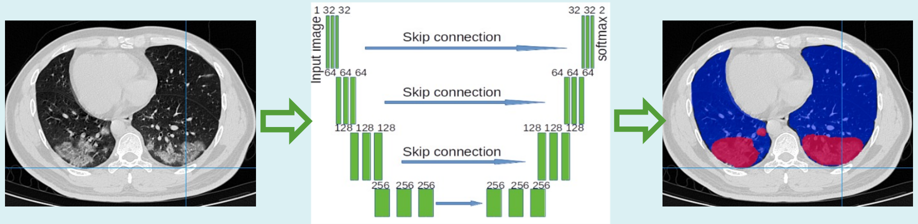


Deep Learning

classification



segmentation



AI in Medical Image Analysis: performance and impact

Comparison between **DL models** and **health-care professionals (HCP)** in the same sample [14 studies/82, different diseases]:

- a sensitivity of **87.0%** with 95% CI [83.0–90.2] for **DL** models and **86.4%** [79.9–91.0] for **HCP**
- a specificity of **92.5%** with 95% CI [85·1–96·4] for **DL** models and **90.5%** [80.6–95.7] for **HCP**

➔ DL models and HCP show **equivalent performance**

[Liu et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: a systematic review and meta-analysis. Lancet Digit Heal 2019;1:e271–97]

Radiologists can guide the introduction of AI into healthcare. They **will not be replaced by AI**, which, in turn will:

- standardize the level of reporting across different clinical centres
- speed up the diagnosis process and allow radiologists to perform more value-added tasks

[Pesapane F, Codari M, Sardanelli F. Artificial intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. Eur Radiol Exp 2018;2]

AI in Medical Image Analysis: current limitations

AI algorithms for medical imaging **must be effectively evaluated** before they are used in clinical practice.
The performance obtained in the R&D stage is difficult to maintain in the clinical use.
➔ Both the generalizability of AI algorithms and the benefits of AI-assisted care relative to conventional care should be proved

[Park SH, Han K, Jang HY, Park JE, Lee J, Kim DW, et al. Methods for Clinical Evaluation of Artificial Intelligence Algorithms for Medical Diagnosis. Radiology 2022:1–12]

It is not enough for AI to efficiently detect image abnormalities/pathological conditions. **AI imaging studies** should be refined to **predict clinically meaningful endpoints**, e.g.: lesion malignancy, need for treatment, patient survival.

[Oren O, Gersh BJ, Bhatt DL. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. Lancet Digit Heal 2020;2:e486–8.]

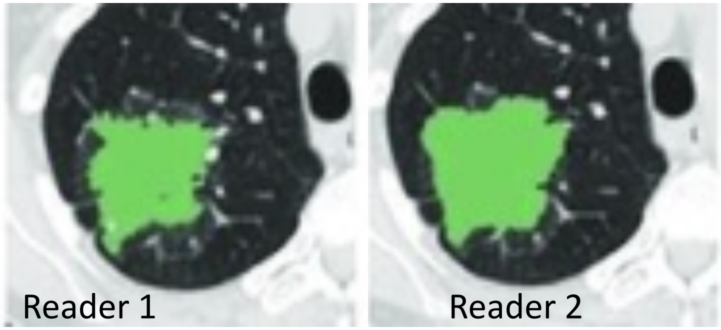
Open issues to address

- **Definition of clinically meaningful endpoints:**
    - A <u>multidisciplinary team</u> is needed to define the objective and collect suitable data accordingly

- **Open technical issues and challenges:**
    - Limited availability of annotated data
    - Mining data from multiple sources
    - Reliability of AI-based systems
    - Explainability (XAI)

Limiting factors for ML training: small annotated datasets

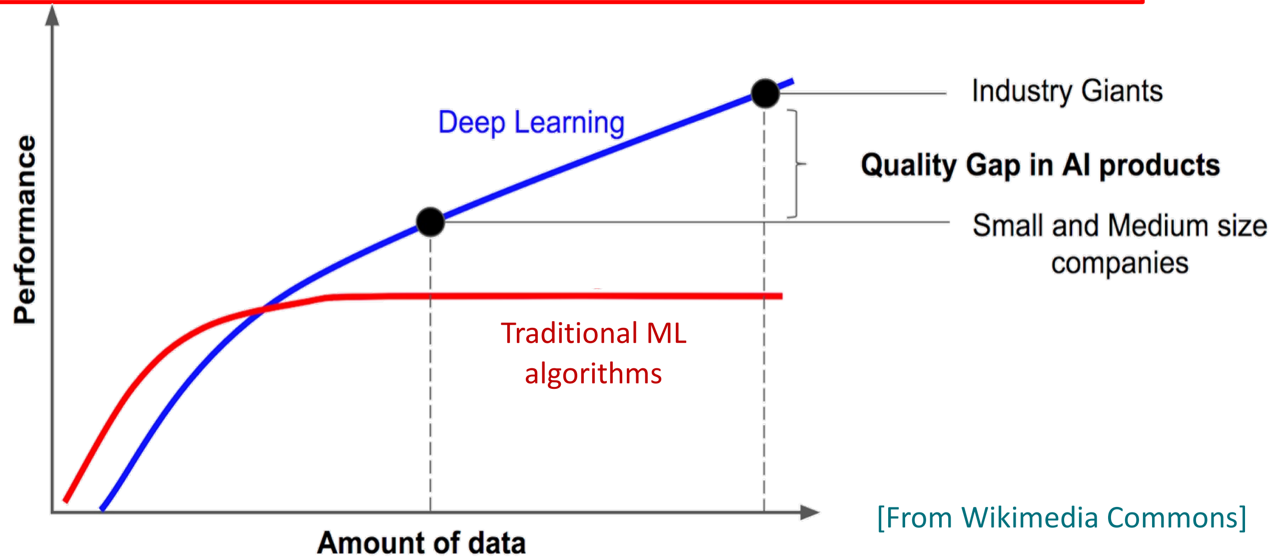
- **Data annotation by human experts is an extremely time-consuming task**, which typically requires:
  - the collection of additional information from other storing systems,
  - expertise in segmenting meaningful regions in images,
  - specific knowledge to assign class labels.
- Moreover, segmentation of organs or lesions (i.e. **voxel-wise annotation**) are affected by inter- and intra-reader variability.



An important issue in ML model training in medical domains are the **small datasets**

Performance of ML algorithms vs. sample size

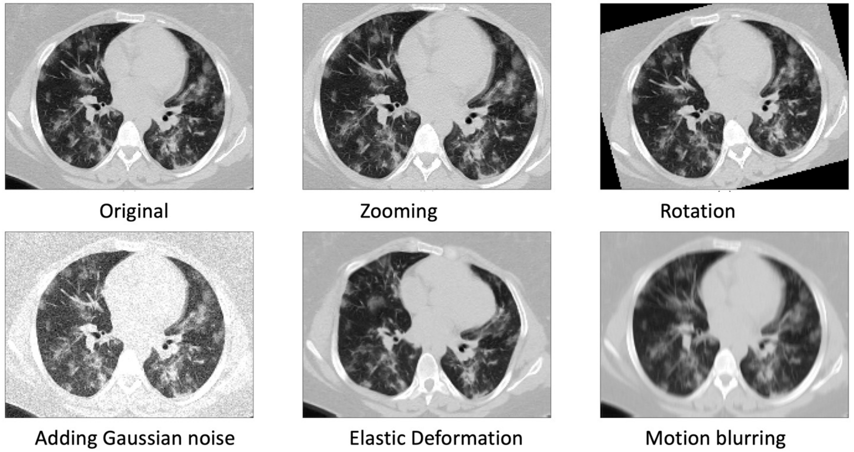
- Traditional ML models can perform even better than DL ones for small sample sizes
- DL models outperform traditional ones in case of large data samples



[From Wikimedia Commons]

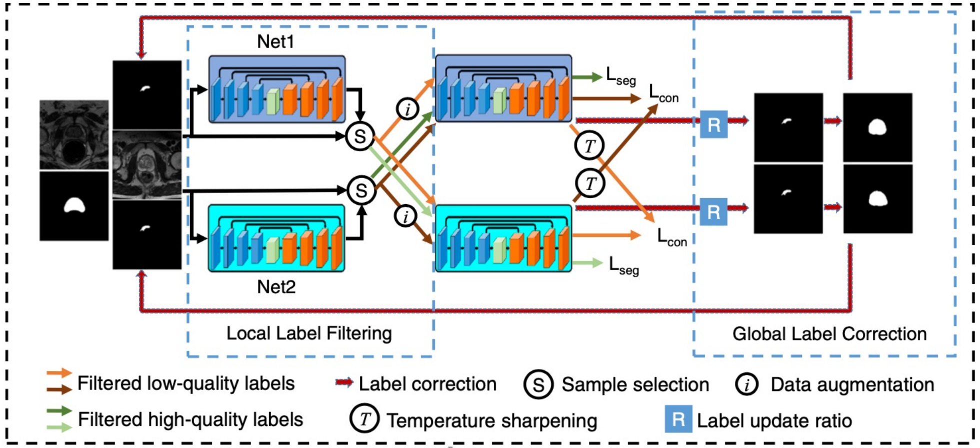
Strategies to mitigate the “small data” problem

➔ Data augmentation with traditional techniques



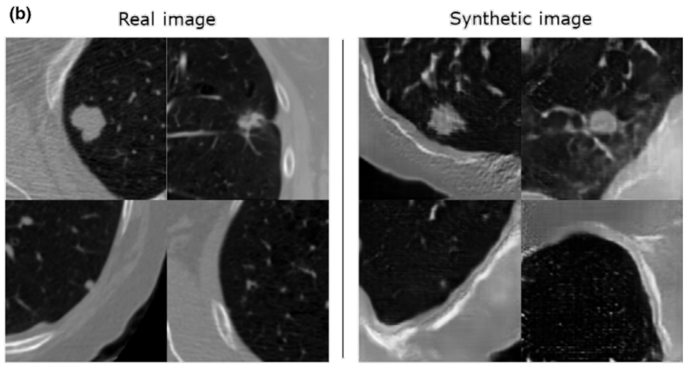
[Lizzi F et al, Quantification of pulmonary involvement in COVID-19 pneumonia…, *IJCARS,* 17(2), 229–237 (2022)]

➔ Automated/semi-automated annotation



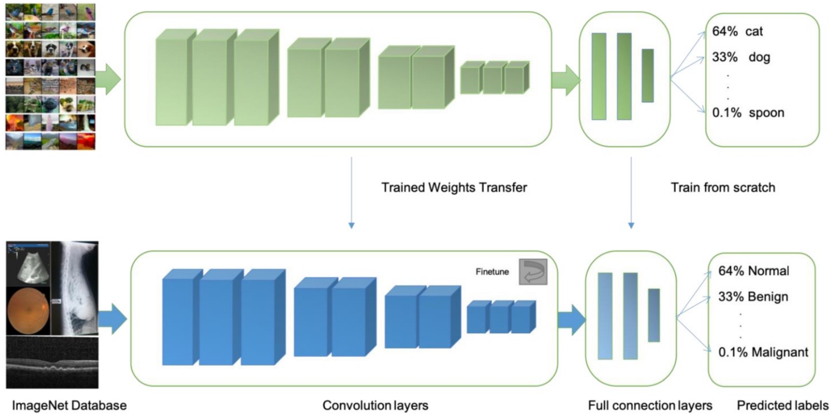
[Wang et al, Annotation-efficient deep learning for automatic medical image segmentation. *Nature Communications*, *12*(1), 1–13 (2021)]

➔ Data augmentation via synthetic data generation



[Chlap P et al, A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, *65*(5), 545–563 (2021)]

➔ Transfer learning



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

[Piffer S, Ubaldi L, Tangaro S, Retico A and Talamonti C, Tackling the small data problem in medical image classification with artificial intelligence: a systematic review, *under review*]

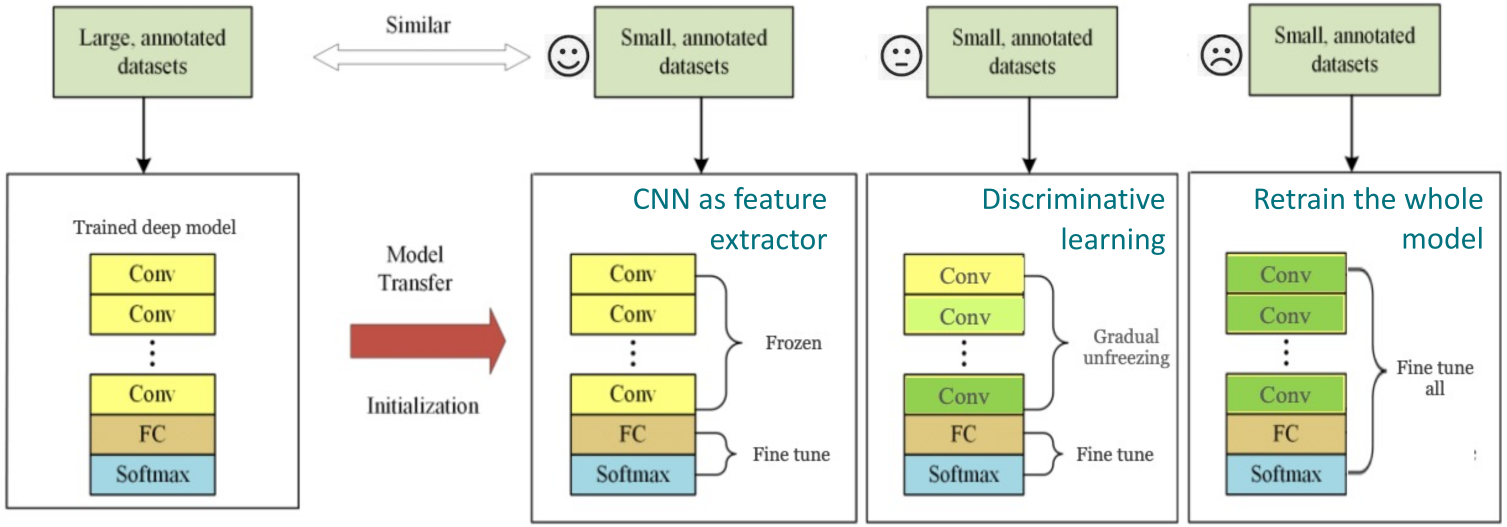
Transfer learning (TL)

Different TL approaches can be implemented:

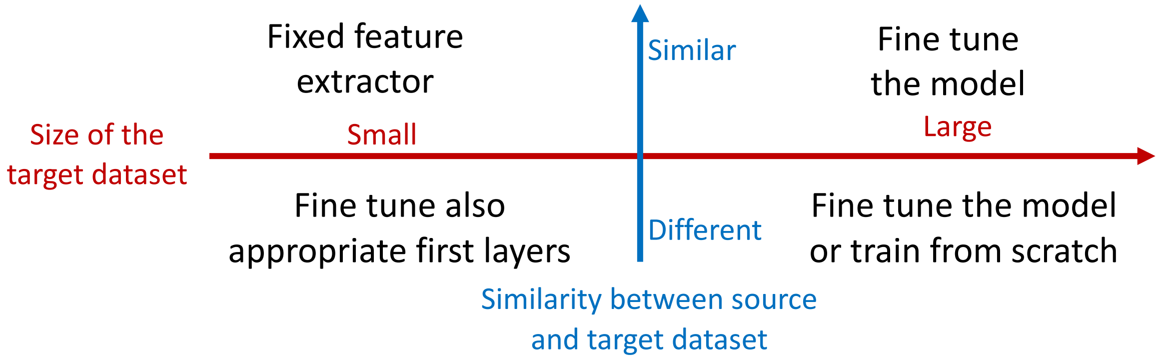
- **CNN as feature extractor**
  All but the last feed-forward layer(s) of the network are frozen. The only weights that are trained are those in the last layers.

- **Discriminative learning rates with gradual unfreezing**
  The first layers of a network typically learn general features (e.g., lines, circles, colors, etc.). Thus, the weights in those layers should be changed less than the weights of the downstream layers which are more specialized in the target task.

- **Fine tune all CNN simultaneously**
  None of the weights are frozen. The pretrained network is used as a starting point.

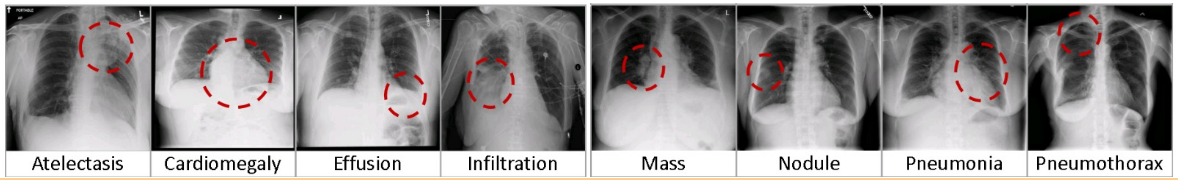


Similarity between source and target datasets, and target dataset size matter



Transfer learning (TL)

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

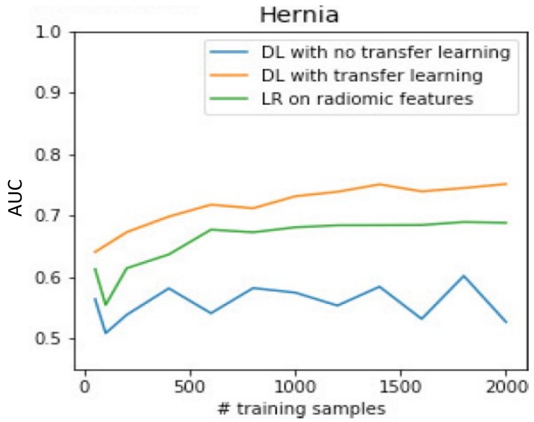


Comparison of three different TL methods, using DenseNet121, and different training dataset sizes and different classification tasks.

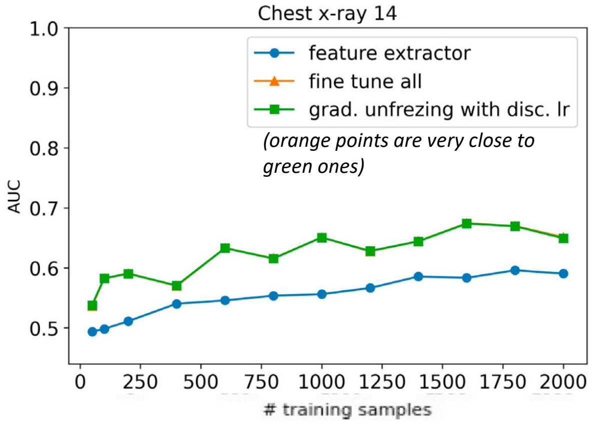
Results:

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

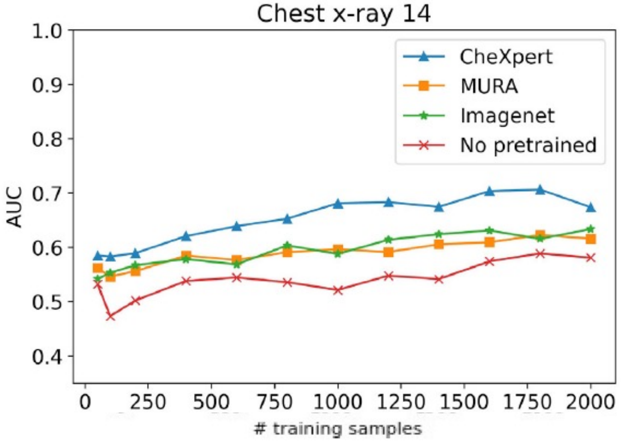
Traditional ML vs DL (w and w/o TL)



Different TL methods



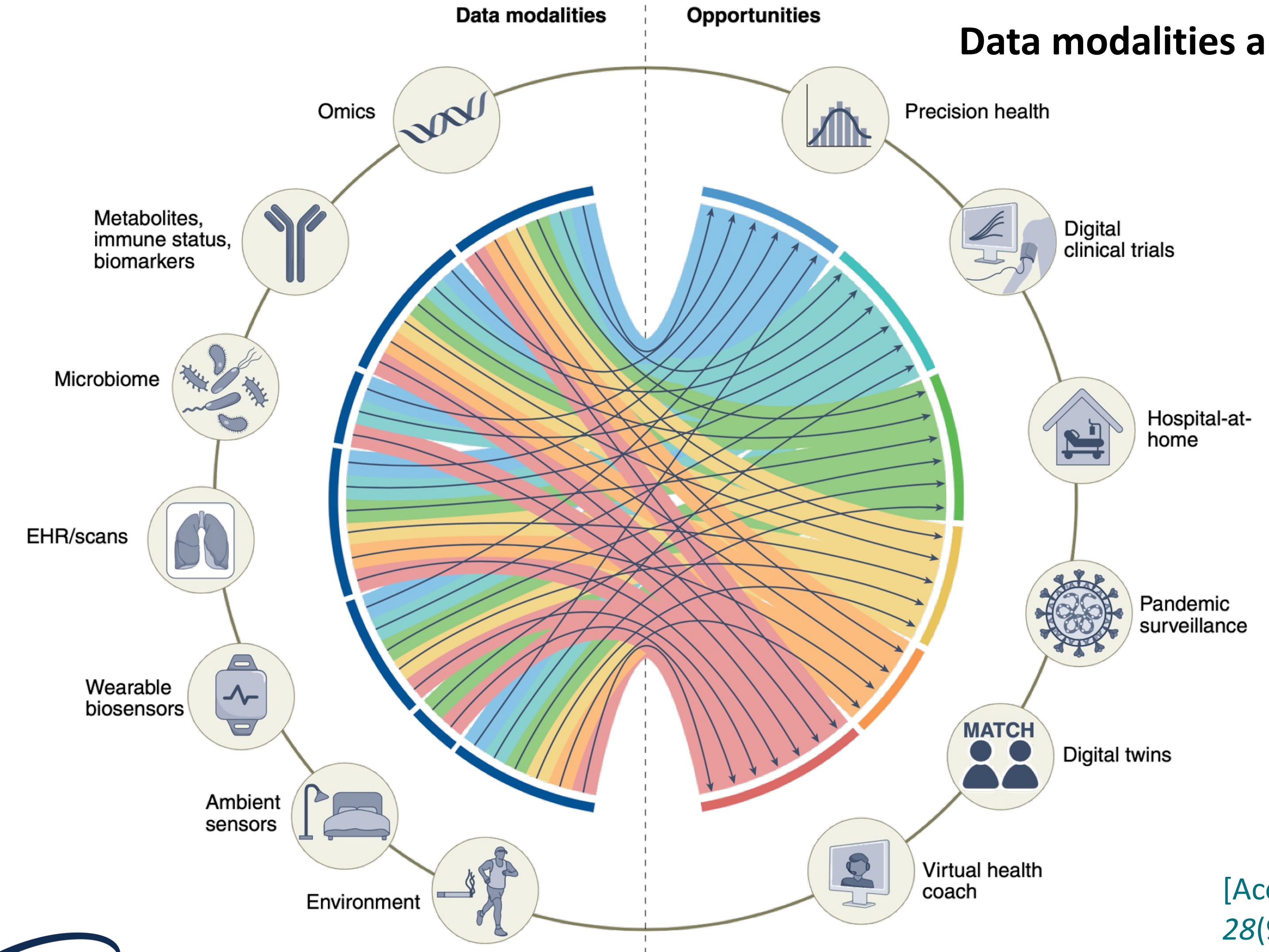
Similarity between source and target datasets



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)]

Integration of data from multiple sources



Data modalities and opportunities for multimodal biomedical AI

The integration of the complementary information encoded in omics data, electronic health records (HER), imaging data is expected:

- to increase the understanding of human health and disease conditions
- to allow personalized preventive, diagnostic and therapeutic strategies

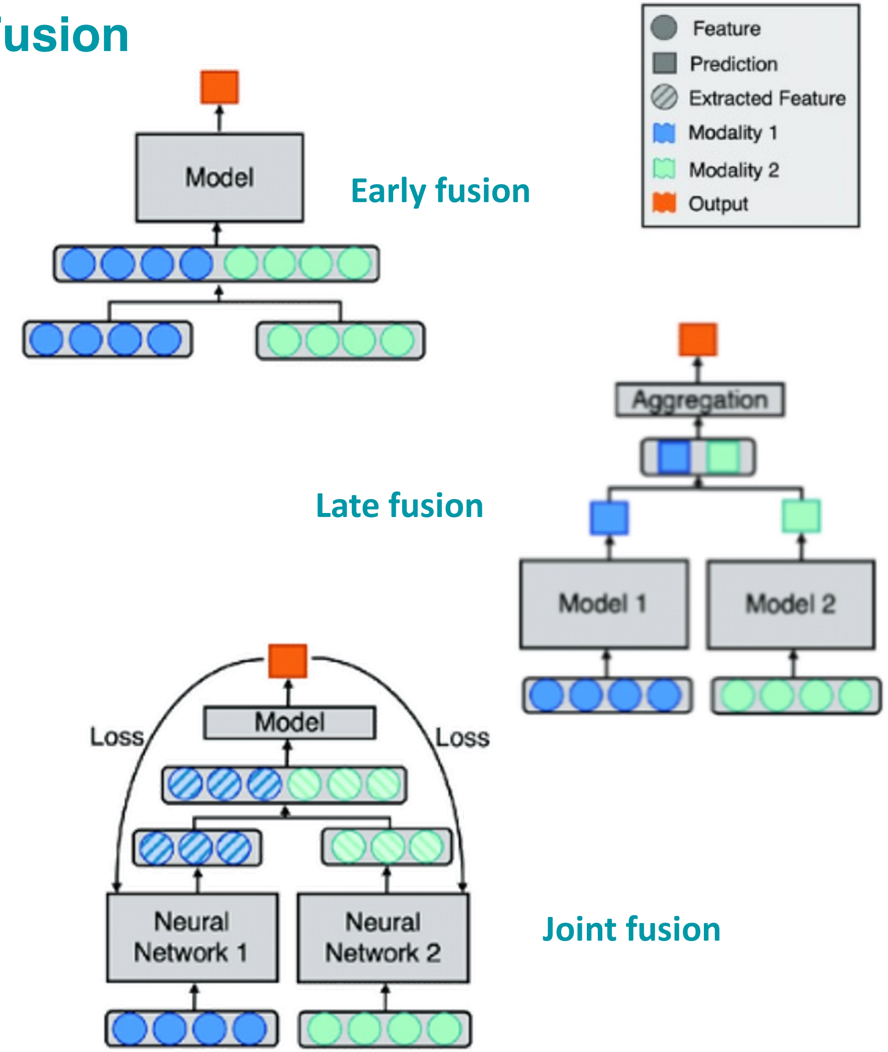
UK Biobank (started in 2006) enrolled more than 500k participants to follow for 30 years.

The integration of these very distinct types of data remains a challenge.

[Acosta et al (2022). Multimodal biomedical AI. *Nature Medicine*, *28*(9), 1773–1784. https://doi.org/10.1038/s41591-022-01981-2]

Multimodal learning

Data from different modalities should be combined → **Multimodal Fusion**



Early fusion:

- It is the simplest approach. Input modalities or features are concatenated before any processing.

Late fusion:

- Separate models are trained for each modality and the output probabilities are combined at the end. It a simple and robust approach, but any possible information encoded in the interaction between data modalities is missing.

Joint fusion:

- The representations of the different modalities are co-learned and combined during the training process. It allows for modality-specific preprocessing and also capturing the interaction between data modalities.

Missing values in multimodal analysis

A high proportion of missing data may affect multimodal data collections.
Simply excluding patients with missing data:

- reduces the dataset dimensionality
- may lead to a selection bias

FILE_ID	AGE	FIQ	DX GROUP	L Mean Thickness	T Mean Thickness	L Cortex Vol	R Cortex Vol
Caltech_0051457	22.9	107	1	2.56	2.57	321132	317005
Caltech_0051458	39.2	93	1	2.65	NaN	266419	266456
Caltech_0051459	22.8	106	1	2.71	2.73	307157	310540
Caltech_0051461	37.7	99	1	2.59	2.61	263933	269417
Caltech_0051464	20.9	101	1	2.76	2.78	381572	394085
Caltech_0051472	17.5	125	1	NaN	2.77	345432	345834
Caltech_0051474	20.9	100	1	NaN	2.63	298059	309524
CMU_a_0050654	24	95	1	2.68	2.71	287010	287815
CMU_a_0050659	27	109	-1	2.72	2.72	330376	330325
CMU_a_0050660	25	NaN	-1	2.74	2.74	279281	284870
CMU_a_0050663	21	101	-1	2.63	2.67	292910	296389
CMU_a_0050664	21	109	-1	2.58	2.58	262753	261800
CMU_a_0050665	33	109	-1	2.55	2.57	237432	237280
CMU_a_0050666	31	107	-1	2.59	2.56	315076	311384
CMU_a_0050668	25	NaN	-1	2.65	2.65	NaN	250974
CMU_b_0050643	21	123	1	2.65	2.66	257398	NaN
CMU_b_0050645	20	124	1	2.58	2.59	264307	260833
CMU_b_0050651	39	116	1	2.47	2.45	306868	306173
KKI_0050814	8.46	108	-1	2.79	2.81	351159	355034
KKI_0050815	10.6	105	1	2.46	2.62	NaN	219226
KKI_0050816	9.73	119	-1	2.61	2.69	271211	275917
KKI_0050817	9.97	NaN	-1	2.61	2.60	298531	299597
KKI_0050818	11.8	98	-1	2.67	2.69	277694	286714
KKI_0050819	9.71	101	-1	2.61	2.62	296599	295511
KKI_0050821	11.2	114	-1	2.75	2.75	288554	288993
KKI_0050822	12.4	98	-1	2.74	2.72	287403	289102
KKI_0050823	11.4	120	1	2.86	2.83	314385	310731

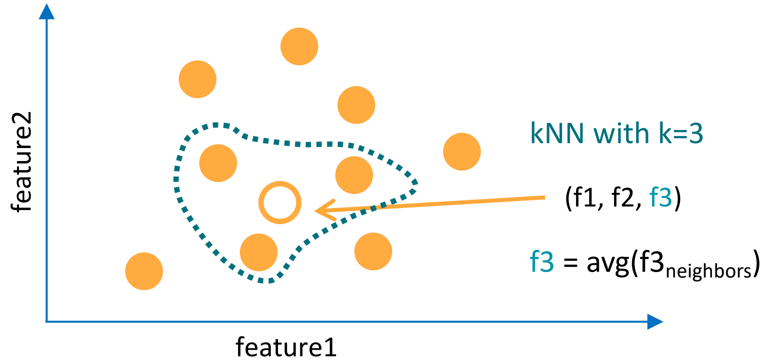
data imputation →

FILE_ID	AGE	FIQ	DX GROUP	L Mean Thickness	T Mean Thickness	L Cortex Vol	R Cortex Vol
Caltech_0051457	22.9	107	1	2.56	2.57	321132	317005
Caltech_0051458	39.2	93	1	2.65	2.63	266419	266456
Caltech_0051459	22.8	106	1	2.71	2.73	307157	310540
Caltech_0051461	37.7	99	1	2.59	2.61	263933	269417
Caltech_0051464	20.9	101	1	2.76	2.78	381572	394085
Caltech_0051472	17.5	125	1	2.83	2.77	345432	345834
Caltech_0051474	20.9	100	1	2.62	2.63	298059	309524
CMU_a_0050654	24	95	1	2.68	2.71	287010	287815
CMU_a_0050659	27	109	-1	2.72	2.72	330376	330325
CMU_a_0050660	25	115	-1	2.74	2.74	279281	284870
CMU_a_0050663	21	101	-1	2.63	2.67	292910	296389
CMU_a_0050664	21	109	-1	2.58	2.58	262753	261800
CMU_a_0050665	33	109	-1	2.55	2.57	237432	237280
CMU_a_0050666	31	107	-1	2.59	2.56	315076	311384
CMU_a_0050668	25	110	-1	2.65	2.65	250916	250974
CMU_b_0050643	21	123	1	2.65	2.66	257398	256875
CMU_b_0050645	20	124	1	2.58	2.59	264307	260833
CMU_b_0050651	39	116	1	2.47	2.45	306868	306173
KKI_0050814	8.46	108	-1	2.79	2.81	351159	355034
KKI_0050815	10.62	105	1	2.46	2.62	207923	219226
KKI_0050816	9.73	119	-1	2.61	2.69	271211	275917
KKI_0050817	9.97	119	-1	2.61	2.60	298531	299597
KKI_0050818	11.79	98	-1	2.67	2.69	277694	286714
KKI_0050819	9.71	101	-1	2.61	2.62	296599	295511
KKI_0050821	11.17	114	-1	2.75	2.75	288554	288993
KKI_0050822	12.43	98	-1	2.74	2.72	287403	289102
KKI_0050823	11.37	120	1	2.86	2.83	314385	310731

Data imputation techniques

Infer missing values from the existing part of the data.

- Basic approaches:
  - For numerical features:
    - replacing missing values with the mean/median of the non-missing values in a column
    - replacing with a «0» or constant value
  - For categorical features:
    - replacing missing values with the most frequent values within each column
- ML-based approaches (multivariate approach):
  - **k Nearest Neighbors** classification (k-NN): it looks at the nearest observations in the training sample and imputes the missing value from that of the neighbours



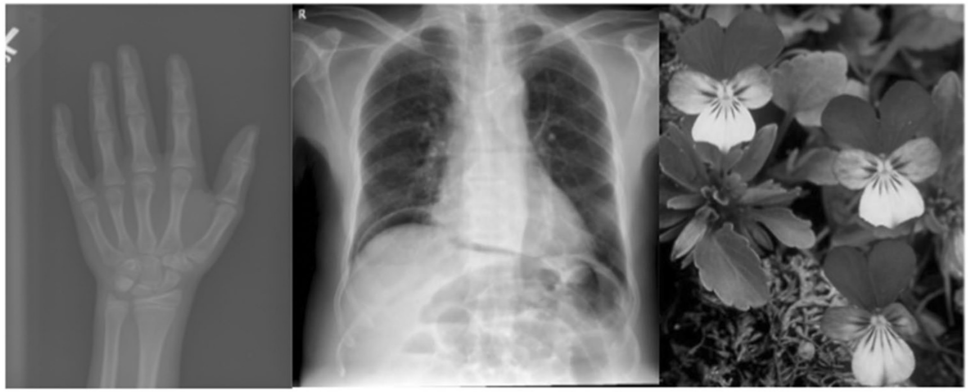
[Cismondi et al, Missing data in medical databases: Impute, delete or classify? Artif Intell Med 2013;58:63–72. https://doi.org/10.1016/j.artmed.2013.01.003.

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

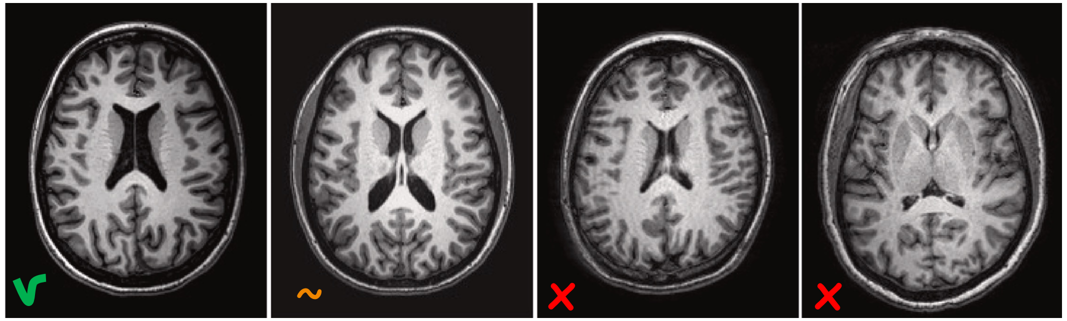


Predicted Bone Age: 13 years, 9 months	Predicted Bone Age: 1 year, 1 month	Predicted Bone Age: 15 years, 11 months

- 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

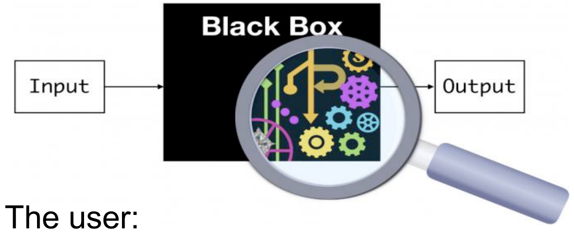
[Fantini et al. (2021). Automatic MR image quality evaluation using a Deep CNN: A reference-free method to rate motion artifacts in neuroimaging. Computerized Medical Imaging and Graphics, 90, 101897. https://doi.org/10.1016/j.compmedimag.2021.101897]

Motion-free vs motion corrupted images



The need for AI explainability (XAI)

AI-based Decision Support Systems (DSS) nowadays are almost completely "opaque"



ML systems nowadays

Training Data → ML/DL → Decision function → DSS output → User

Task

User

The user:

- does not understand the motivation why a certain output is given
- does not know whether the DSS succeeded/failed
- does not know when to trust the DSS
- does not know why the DSS failed, thus how it can be improved

The goal is to make them more "transparent"

Future XAI systems

Training Data → new ML/DL algorithms → Explainable Model | Explanation Interface ⇄ User

Task

User

The user:

- understands the motivation why a certain output is given
- knows whether the DSS succeeded/failed
- knows when to trust the DSS
- knows why the DSS failed, thus, how to improve it

20–22 Nov 2023 - Università di Milano Bicocca
https://indico.cern.ch/event/1312529/

The Artificial Intelligence in Medicine (AIM) INFN Project

[INFN, CSN5, 2019-2021]

Principal Invest.: A. Retico
Research Units:
- Bari (S. Tangaro)
- Bologna (D. Remondini)
- Cagliari (P. Oliva)
- Catania (M. Marrale)
- Firenze (C. Talamonti)
- Genova (A. Chincarini)
- Lab. Naz. Sud (G. Russo)
- Milano (C. Lenardi)
- Napoli (G. Mettivier)
- Pavia (A. Lascialfari)
- Pisa (M.E. Fantacci)

Artificial Intelligence to become the next revolution in **medical diagnostics** and **therapy**.

- New image processing and data analysis strategies, including radiomics approaches, need to be developed and extensively validated.

AIM's objectives

- AIM1: Multicenter data harmonization
- AIM2: Quantification
- AIM3: Predictive models

next_AIM's objectives

- **WP1** – Challenge I: no-so-big data
- **WP2** – Challenge II: explainable AI (XAI)
- **WP3** – Applications to real-world data samples
- **WP4** – Computing resources and SW organization
- **WP5** – Exploitation of results and Dissemination

Long-standing collaboration with Italian centers (hospitals / IRCCS) and with international consortia for data sharing

[INFN, CSN5, 2022-2024]

Principal Invest.: A. Retico
Research Units:
- Bari (S. Tangaro)
- Bologna (D. Remondini)
- Cagliari (P. Oliva)
- Catania (M. Marrale)
- Ferrara (G. Paternò)
- Firenze (C. Talamonti)
- Genova (A. Chincarini)
- Lab. Naz. Sud (G. Russo)
- Milano (C. Lenardi)
- Napoli (G. Mettivier)
- Pavia (A. Lascialfari)
- Padova (A. Zucchetta)
- Pisa (M.E. Fantacci)

The Artificial Intelligence in Medicine (AIM) INFN-CSN5 Project

Researchers from INFN divisions and University Departments collaborate closely with Radiologists, Clinicians and Medical Physicists in Clinical Centers to develop innovative solutions based on data mining and AI

Clinical partners

- IRCCS S. Martino (GE)
- IRCCS Stella Maris (PI)
- IRCCS Gaslini (GE)
- IRCCS Centro S. G. di Dio (BS)
- IRCCS G.Paolo II (BA)
- IRCCS Mondino (PV)
- IRCCS SDN (NA)
- IRCCS IRST Meldola (FC)
- IRCCS Bellaria (BO)
- IRCCS S. Orsola (BO)
- IMAGO7
- Azienda Osp. Univ. Pisana (PI)
- Azienda Osp. Univ. Careggi (FI)
- Osp. Pediatrico Meyer (FI)
- Ospedale Cardarelli (NA)
- Azienda Sanitaria Cuneo 1 (CN)
- IFO-Ist. Naz. Tumori Regina Elena (RM)
- ASST Niguarda (MI)
- Policlinico di Bari
- Policlinico di Palermo
- Policlinico Univ. di Napoli
- Policlinico San Matteo (PV)

EU / consortia

- EADC (EU)
- EDLBC (EU)
- ADNI (US)
- ABIDE (EU/US)
- ENIGMA (WW)

Scientific associations

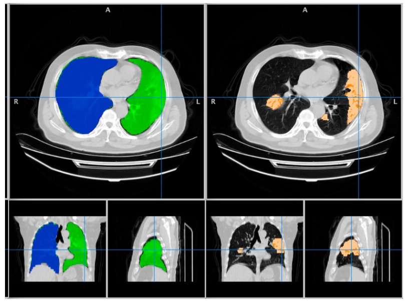
- Italian Association of Medical Physics (AIFM)

LungQuant: SW tool for lesion detection and structured reporting

[https://www.openaccessrepository.it/record/76937]

LungQuant

[Lizzi F et al 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 2022;17:229–37. doi.org/10.1007/s11548-021-02501-2.

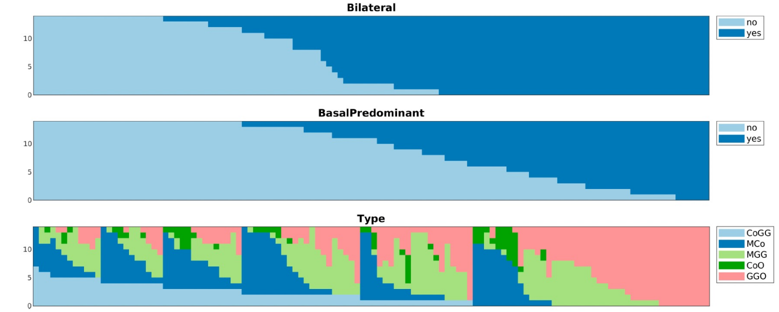


ID	LESION_TYPE_INDEX	BILATERAL_INDEX	BASAL_INDEX
A-0037	0,137	0,447	37
A-0311	0,198	0,041	61
A-0291_0	0,224	0,193	31
A-0327	0,292	0,351	60

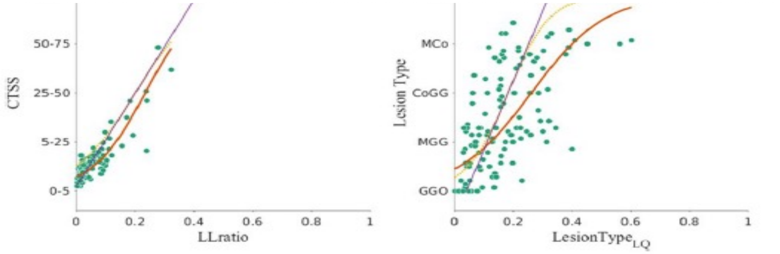
$V_{Consolidation} / V_{Lesion}$ 0: unilateral 1: bilateral → 0: basal → 100: apical

The validation of the LungQuant software output against the qualitative assessment of 14 radiologists from 5 University Hospitals (Pisa, Pavia, Firenze, Palermo, Milano) has shown:

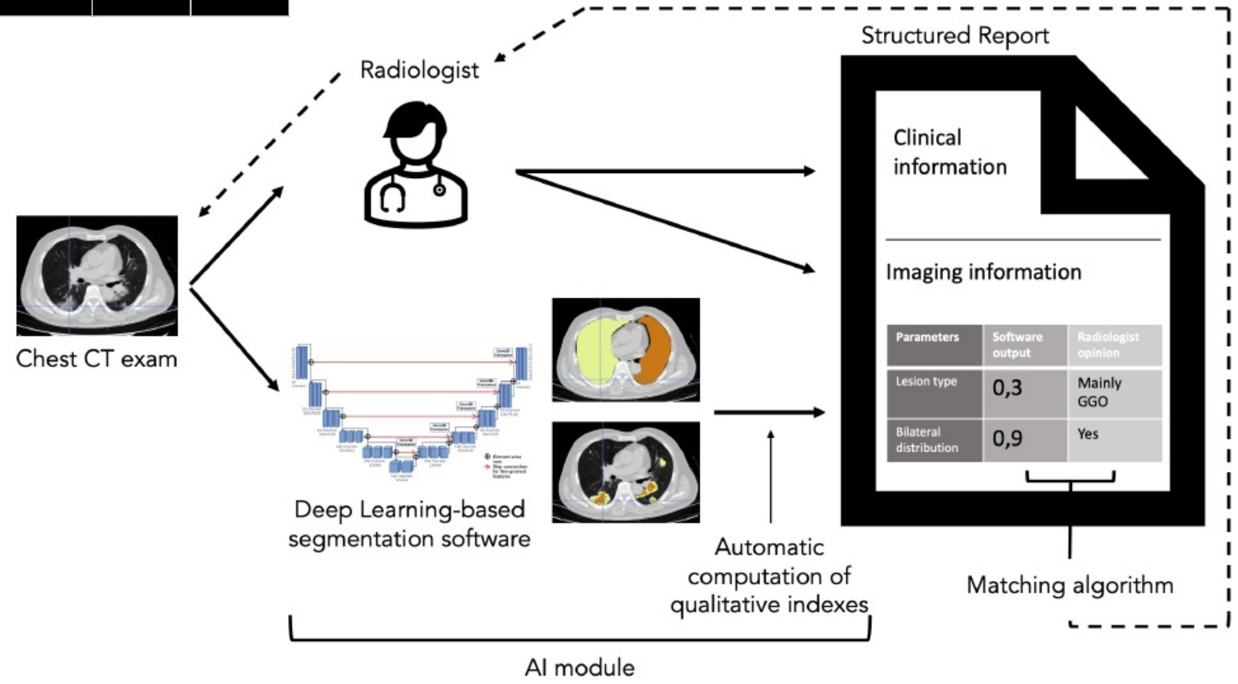
- a poor agreement among the opinions of radiologists



- a good correlation between average radiologists' opinions and the equivalent software output metrics



[Chincarini A, Scapicchio C et al A multicenter evaluation of the LungQuant software for lung parenchyma characterization in COVID-19 pneumonia, European Radiology Experimental, https://doi.org/10.1186/s41747-023-00334-z]



[Scapicchio C, et al. Integration of a Deep Learning-Based Module for the Quantification of Imaging Features into the Filling-in Process of the Radiological Structured Report. Int. Jt. Conf. Biomed. Eng. Syst. Technol., SCITEPRESS 2023, p. 663–70. https://doi.org/10.5220/0011921900003414.]

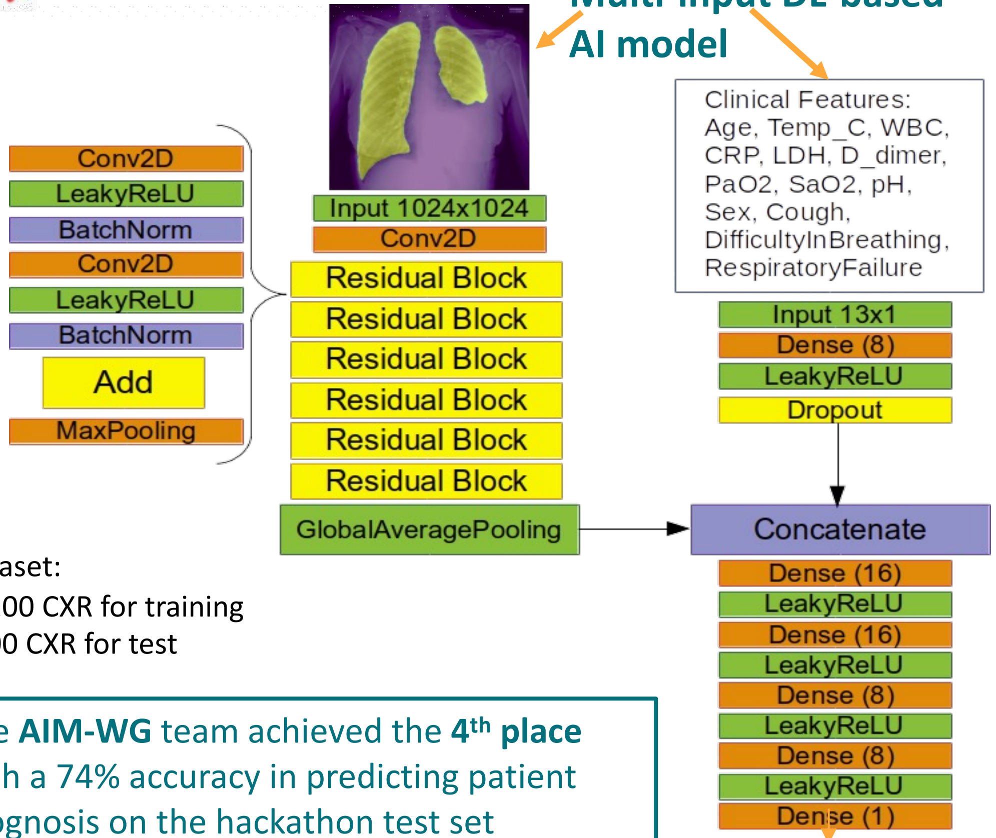
Prediction of COVID-19 severity: the covidcxr hackathon

https://ai4covid-hackathon.it/

Challenge on chest X-ray and clinical data of patients with COVID-19 pneumonia

covidcxr - Hackathon

Multi-input DL-based AI model

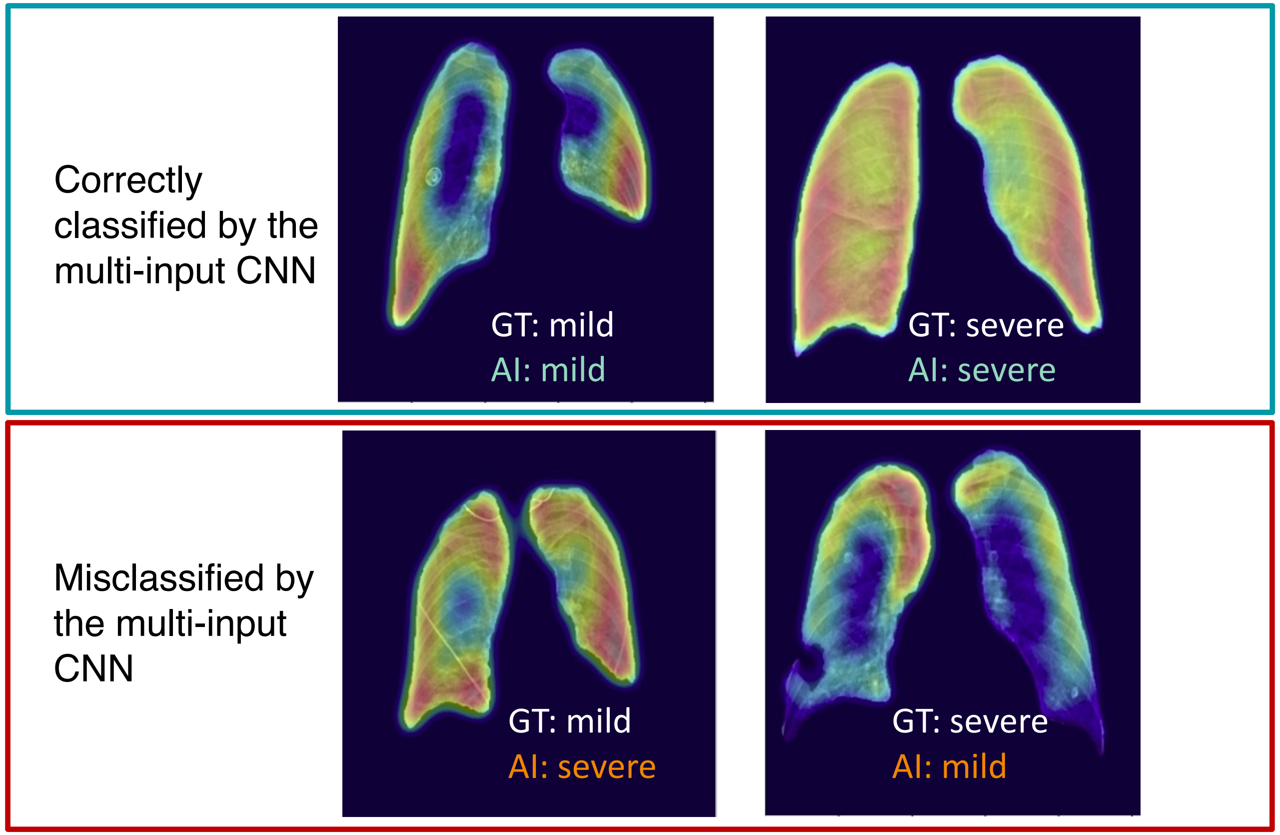


Outcome prediction:
(severe/not severe)

Dataset:
~1100 CXR for training
~500 CXR for test

The **AIM-WG** team achieved the **4th place** with a 74% accuracy in predicting patient prognosis on the hackathon test set

Explainability: The grad-CAM technique produced saliency maps, which indicate whether the classifier is looking at the right parts of the image when assigning a certain class label



[Lizzi et al, Fully automated deep learning based system for COVID-19 patient outcome prediction, *Intelligence-based medicine*, under review]

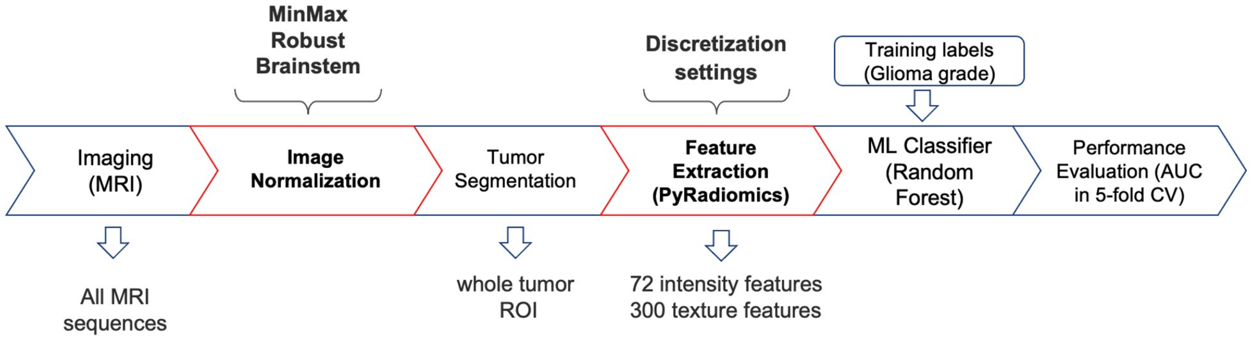
Predictive model to discriminate low-grade vs. high-grade gliomas

Evaluation of the robustness of radiomic features in multiparametric MRI and its impact on predictive value of AI models

Multiparametric MRI scans (T1, T1-Gd, T2, FLAIR) of:

- **61 patients with** Low-Grade Gliomas (**LGG**)
- **97 patients with** High-Grade Gliomas (**HGG**)

The analysis pipeline



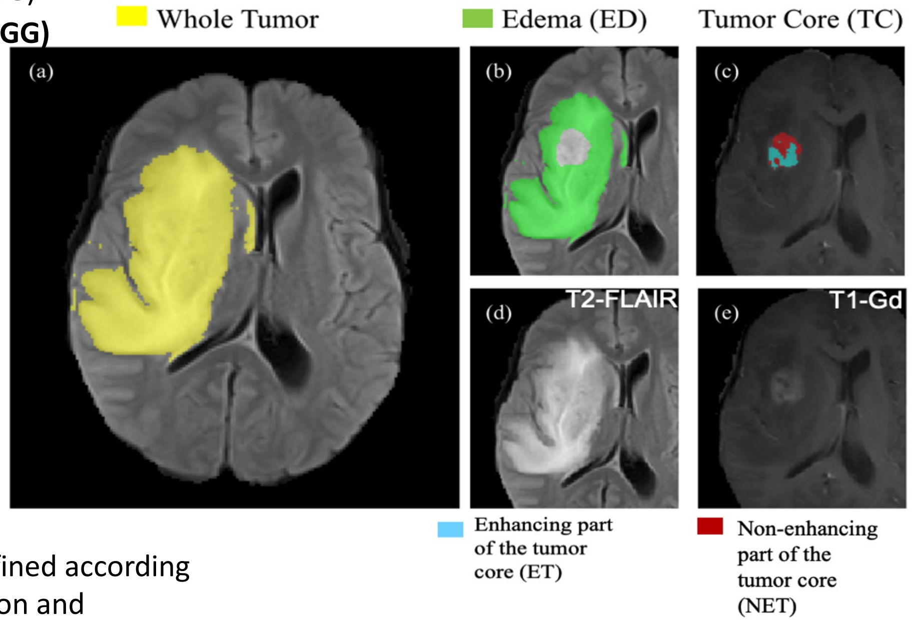
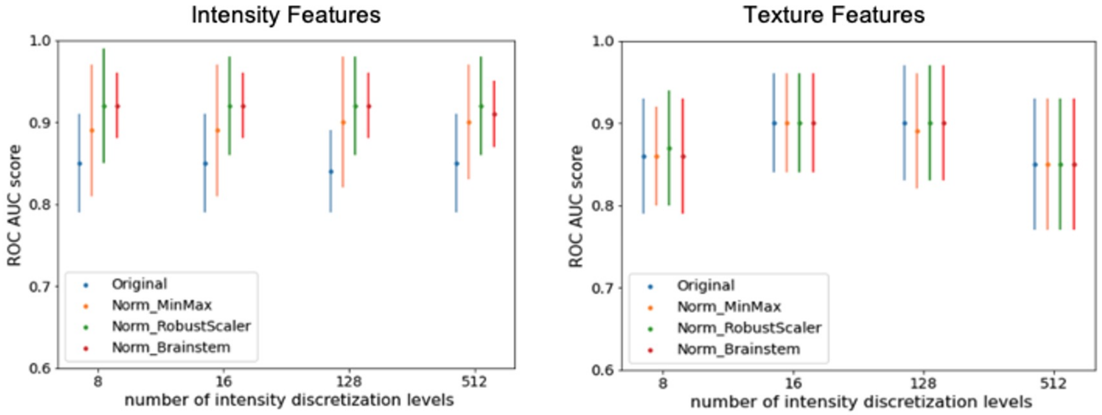


Image normalization and intensity discretization have an impact on the performance of ML classifiers based on radiomic features.



Random forest (RF) classification

- target: LGG vs HGG discrimination
- features: **MRI-reliable features** defined according to the most appropriate normalization and discretization settings.

Conclusions

- The **complementary information of multi-parametric MRI** has to be taken into account
- The **image preprocessing** step is relevant for radiomic and ML analysis

Modality	**Raw feature Set (372 Features for all modalities)**	**MRI-reliable feature Set (372 Features) [Norm_Brainstem] (bin counts = 128)**
T1	0.73 ± 0.05	0.69 ± 0.04
T1-Gd	0.89 ± 0.05	0.93 ± 0.05
T2	0.76 ± 0.08	0.75 ± 0.06
T2 FLAIR	0.76 ± 0.08	0.76 ± 0.06
All sequences	0.88 ± 0.08	0.93 ± 0.05

Ubaldi L, Saponaro S, Giuliano A, Talamonti C, Retico A. Deriving quantitative information from multiparametric MRI via Radiomics: Evaluation of the robustness and predictive value of radiomic features in the discrimination of low-grade versus high-grade gliomas with machine learning. Phys Medica 2023;107:102538, https://doi.org/10.1016/j.ejmp.2023.102538

Joint fusion approach to exploit both structural and functional data

Brain imaging features of ~1400 subjects

- **sMRI –** The Freesurfer *recon-all* pipeline has been implemented to extract 221 structural features for each subject
- **rs-fMRI –** The CPAC processing pipeline for fMRI data has been implemented:
  - The Harvard-Oxford atlas has been used, thus generating 103 temporal series for each subject
  - The functional connectivity matrix has been computed for each subject implementing the Pearson correlation, thus obtaining 5253 functional features for each subject

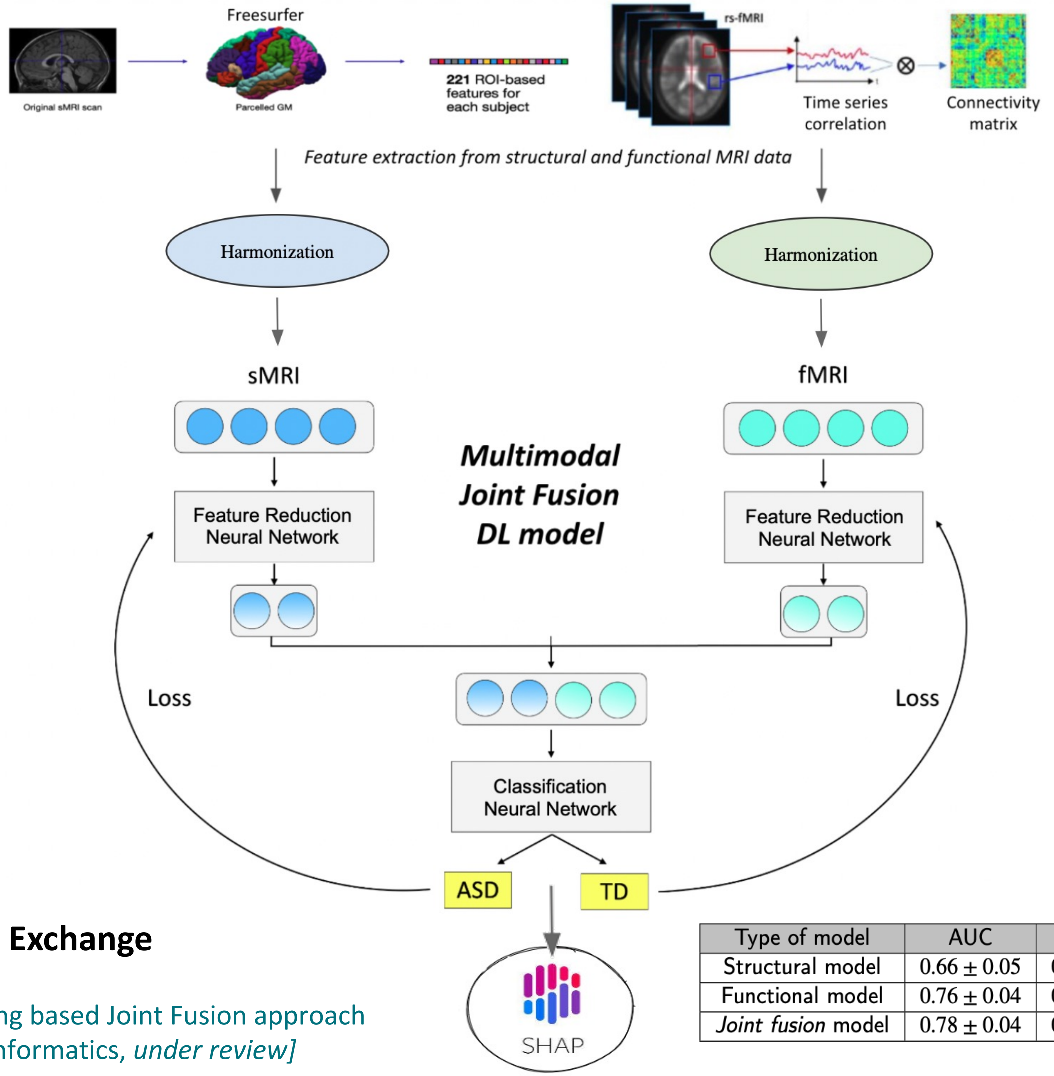
Joint fusion approach:

- The Feature Reduction and the Feature Classification Neural Networks are trained using a single cost function, thus the most meaningful features for the classification are extracted
- The model was trained with 150 epochs within a 10-fold cross validation scheme

Explainability framework:

- SHpley Additive exPlanations (**SHAP**)

ABIDE
Autism Brain Imaging Data Exchange

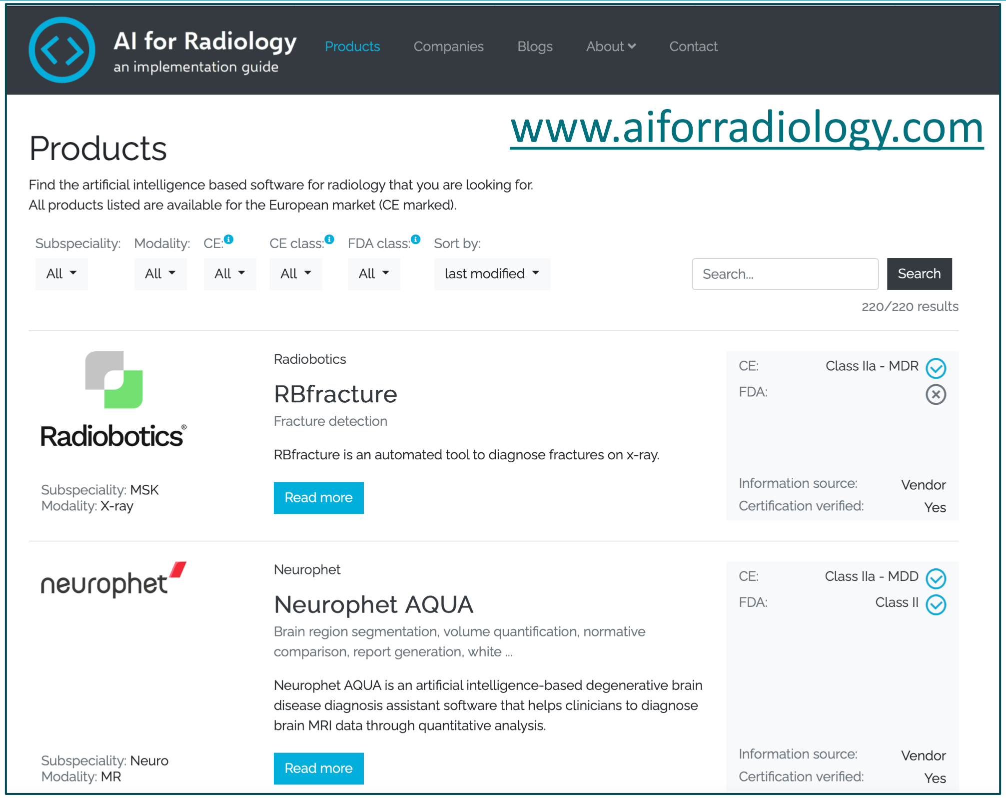


Type of model	AUC	Accuracy
Structural model	0.66 ± 0.05	0.75 ± 0.08
Functional model	0.76 ± 0.04	0.83 ± 0.12
Joint fusion model	0.78 ± 0.04	0.85 ± 0.12

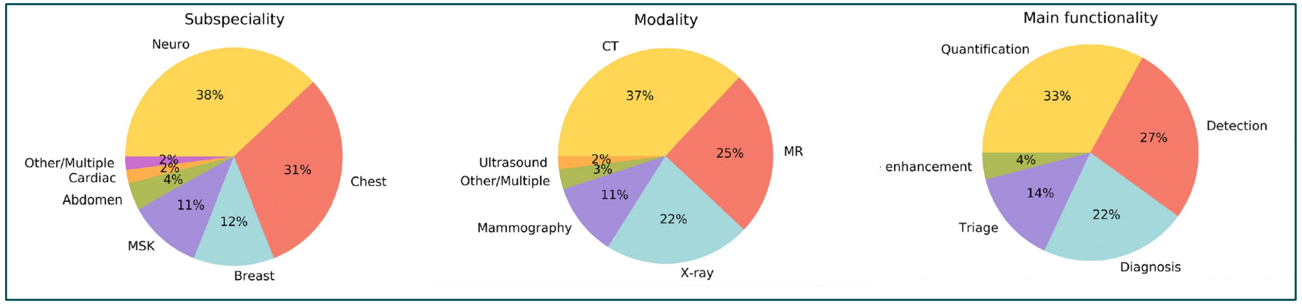
[Saponaro S, Lizzi F, Serra G, Mainas F, Oliva P, Giuliano A, Calderoni S, Retico A. Deep Learning based Joint Fusion approach to exploit anatomical and functional brain information in Autism Spectrum Disorders, Brain Informatics, *under review*]

CE marked AI-based tools available on the market

- A review of **100 CE marked software products** was carried out and made available online
- An extensive bibliographic research on the scientific evidence of the validity of these products has highlighted that:
    - For 64 products out of 100, no evidence of efficacy has been published in a peer-reviewed journal.
    - Only 18 products out of 100 demonstrated a relevant (potential) clinical impact with studies on: impact on diagnostic thinking, on the patient's diagnostic/therapeutic pathway or on costs.



[van Leeuwen, K. G., Schalekamp, S., Rutten, M. J. C. M., van Ginneken, B., & de Rooij, M. (2021). Artificial intelligence in radiology: 100 commercially available products and their scientific evidence. *European Radiology*, *31*(6), 3797–3804. https://doi.org/10.1007/s00330-021-07892-z]



Perspectives and conclusions

AI will continue to **improve healthcare** to promote **precision medicine**

- AI-based tools can assist clinicians in:
  - Making automated interpretation of medical images (prioritization of patients, second opinion)
  - Speeding up clinical work by automated contouring/annotating/ reporting findings
  - Detecting diseases at an early stage

However:

«… we are far better at collating and storing such data, than we are at data analysis.»

[Acosta et al (2022). Multimodal biomedical AI. *Nature Medicine*, *28*(9), 1773–1784. https://doi.org/10.1038/s41591-022-01981-2]

- **In the future, AI systems should be:**
  - Capable of exploiting **multi-modal** information
  - **Reliable**
  - **Explainable**
- A dedicated **multidisciplinary effort** is needed to develop **trustworthy AI systems**

Thank you for your kind attention!

Research partly supported by: Artificial Intelligence in Medicine **next_AIM**, https://www.pi.infn.it/aim project funded by INFN-CSN5; **FAIR-AIM** project funded by Tuscany Government (POR FSE 2014-2020); PNRR - M4C2 - Partenariato Esteso "**FAIR - Future Artificial Intelligence Research**" - Spoke 8, and PNRR - M4C2 - Centro Nazionale "**ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing"** - Spoke 8, funded by the European Commission under the NextGeneration EU programme.

Contact: alessandra.retico@pi.infn.it
INFN, Pisa Division
https://www.pi.infn.it/aim/