next_AIM: overview and challenges in medical image analysis

Piernicola Oliva
Università di Sassari
INFN - Sezione di Cagliari
oliva@uniss.it

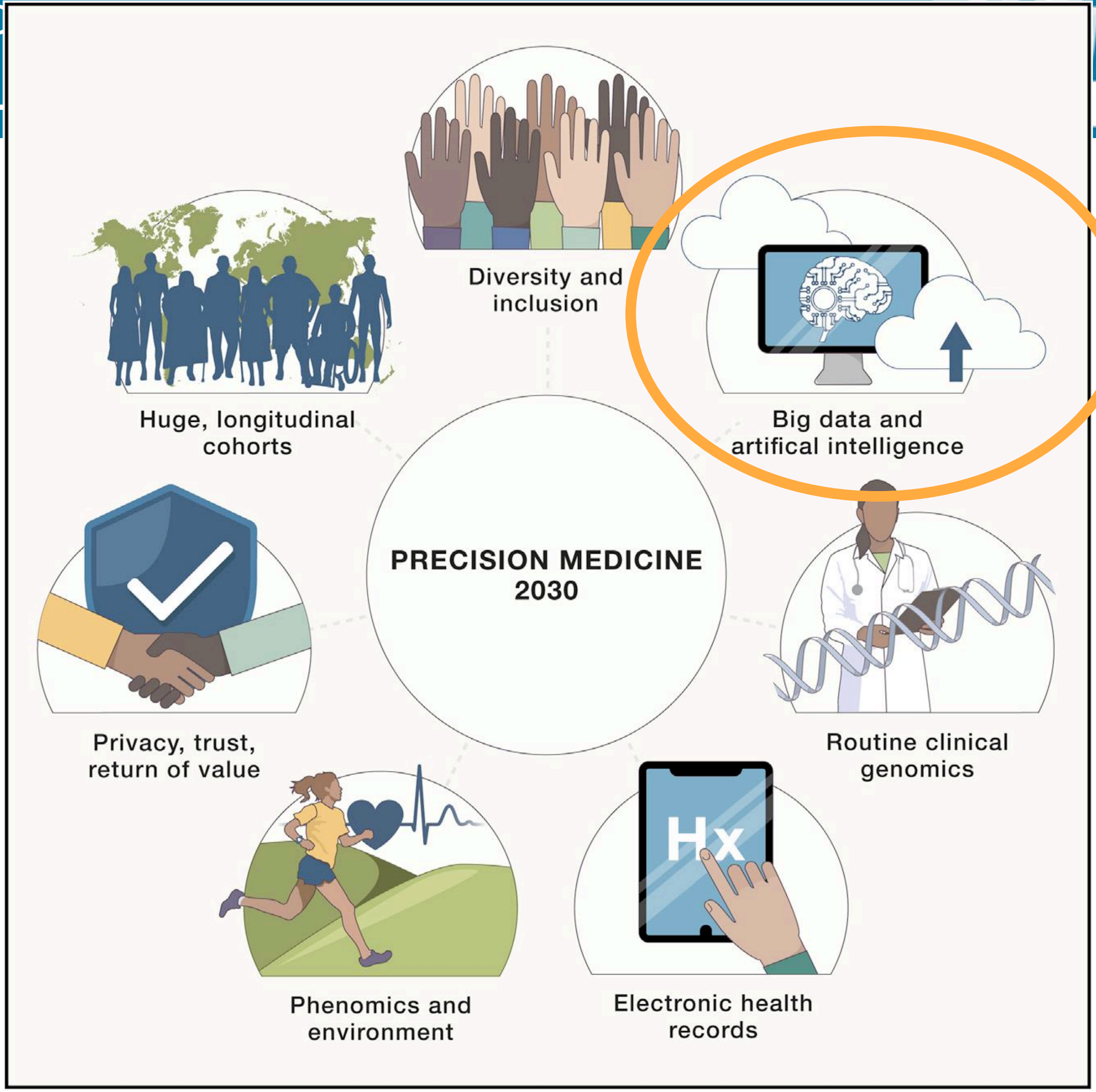
Computing@CSN5: applications and innovations at INFN
Bari, October 14-16, 2024

The Opportunity of Precision Medic

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

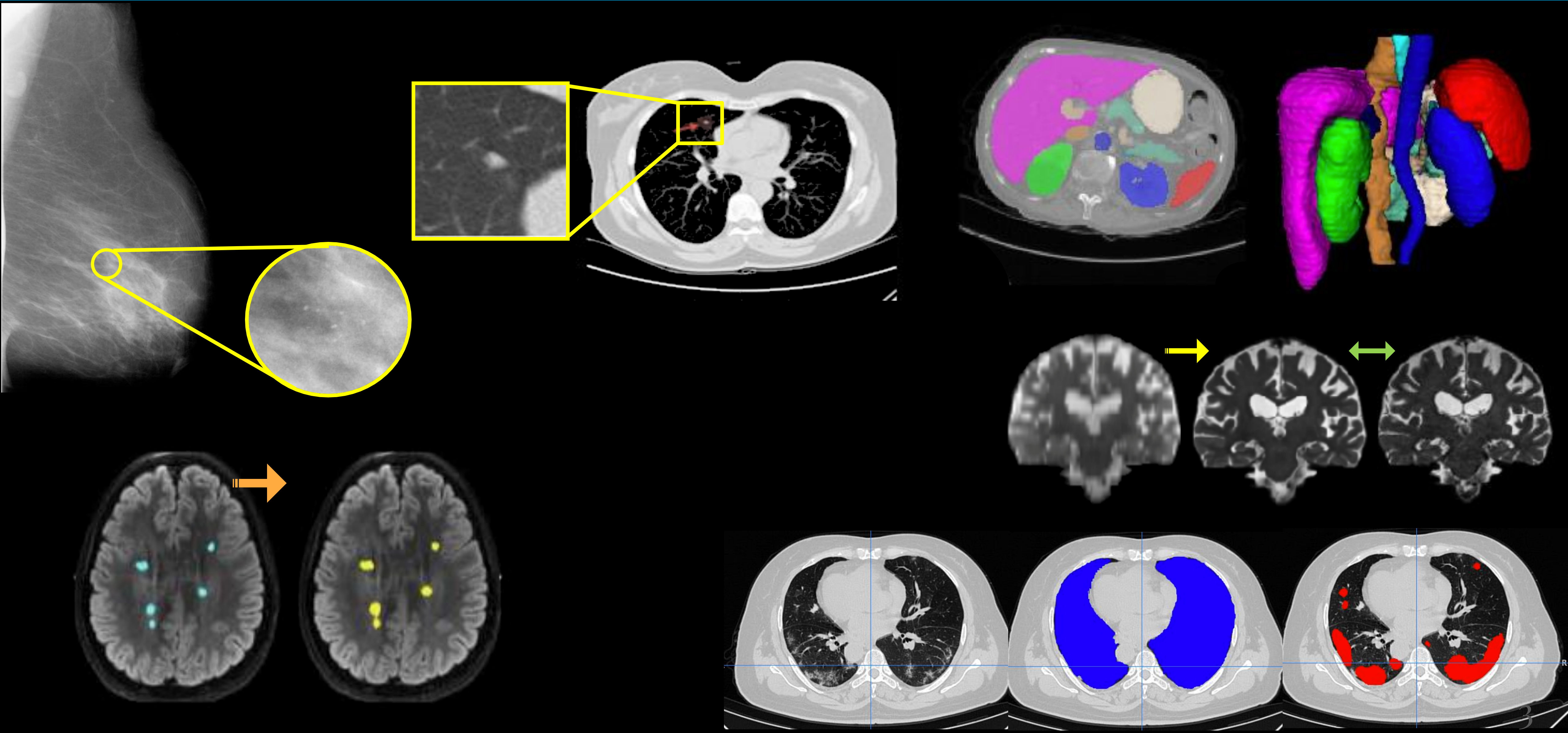
Precision medicine will continue to transform healthcare in the coming decade as it expands in key areas:
- huge cohorts,
- routine clinical genomics,
- phenomics and environment,
- artificial intelligence (AI),

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



Artificial Intelligence (mainly ML and DL) in Medical Imaging

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

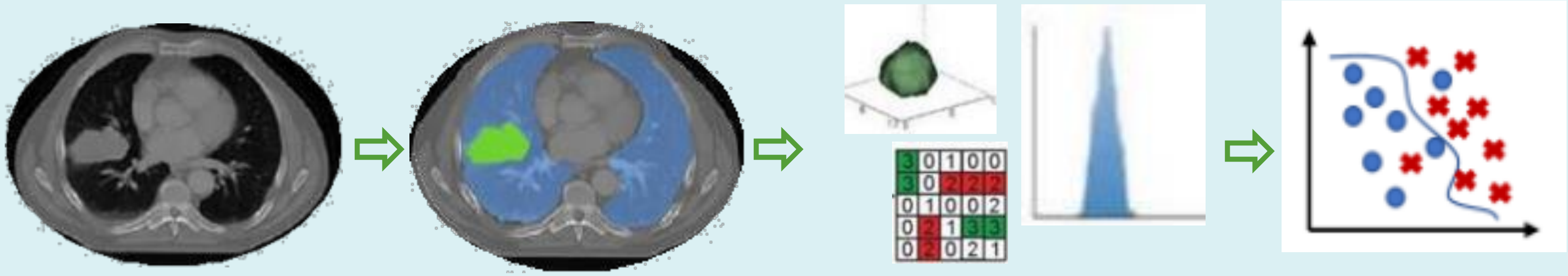
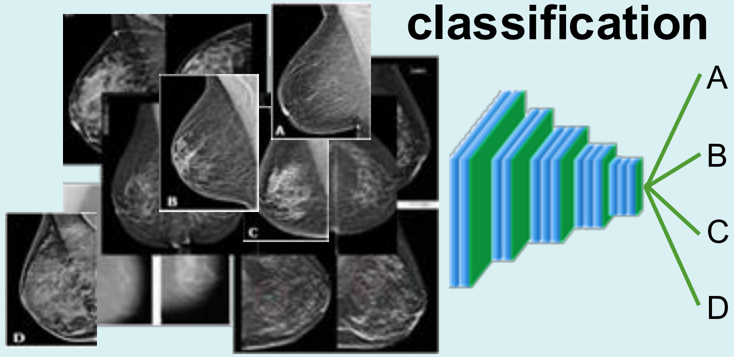


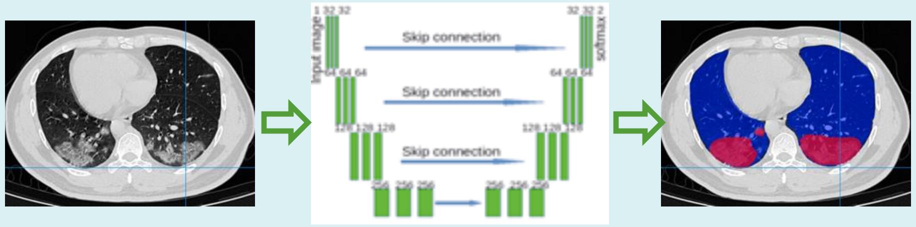
Image acquisition | segmentation | feature computation | classification

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.

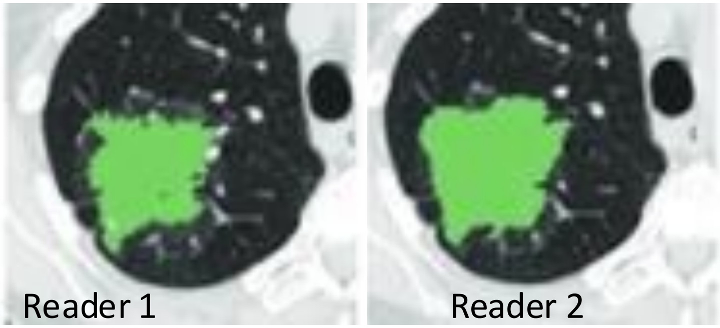
[Oren O, Gersh BJ, Bhatt DL. Artificial intelligence in medical imaging: switching from radiographic pathological data to clinically meaningful endpoints. Lancet Digit Heal 2020;2:e486–8.]

Open issues to address

- **Definition of clinically meaningful endpoints:**
  - A multidisciplinary team 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

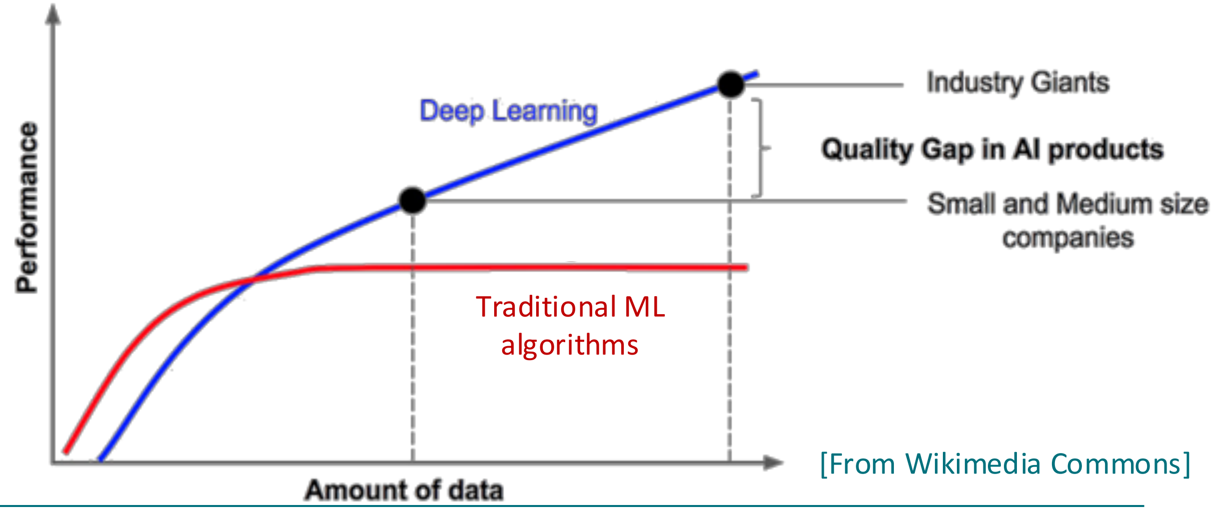
- **Data annotation by human experts is an extremely time-consuming task**, which typically requires:
  - the collection of additional information from other storing systems,
  - expertise in segmenting meaningful regions in images,
  - specific knowledge to assign class labels.
- Moreover, segmentation 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

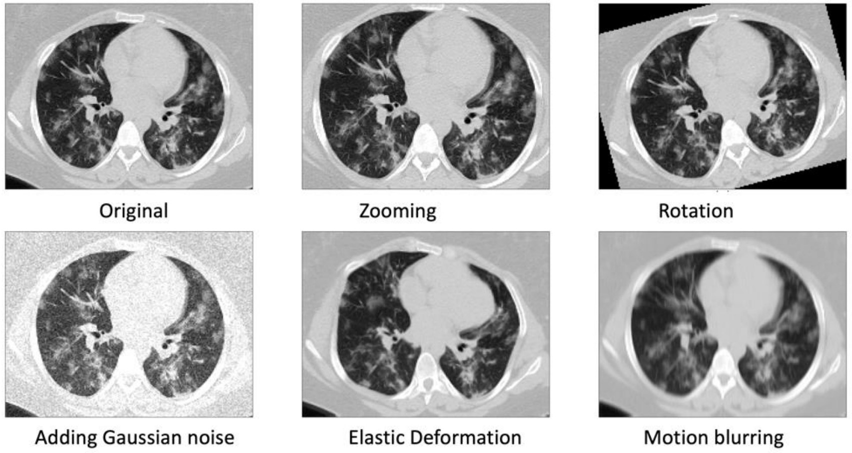
- Traditional ML models can perform even better than DL ones for small sample sizes
- DL models outperform traditional ones in case of large data samples



[From Wikimedia Commons]

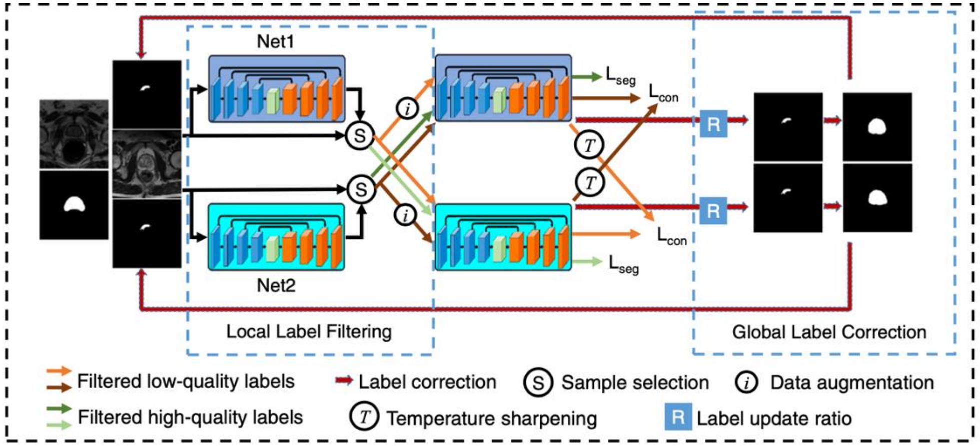
Strategies to mitigate the "small data" problem

Data augmentation with traditional techniques



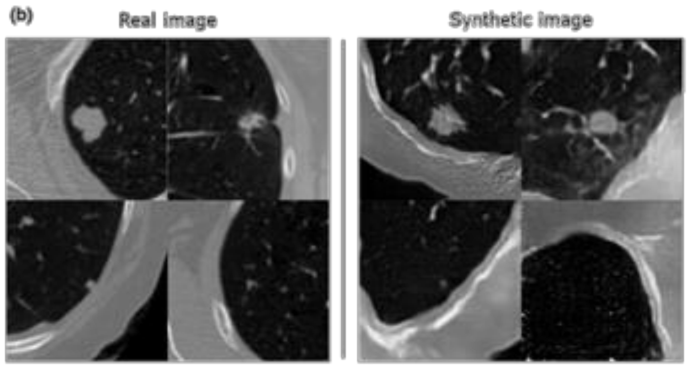
[Lizzi F et al, Quantification of pulmonary involvement in COVID-19 pneumonia…, *IJCARS*, 17(2), 229–237 (2022)]

Automated/semi-automated annotation



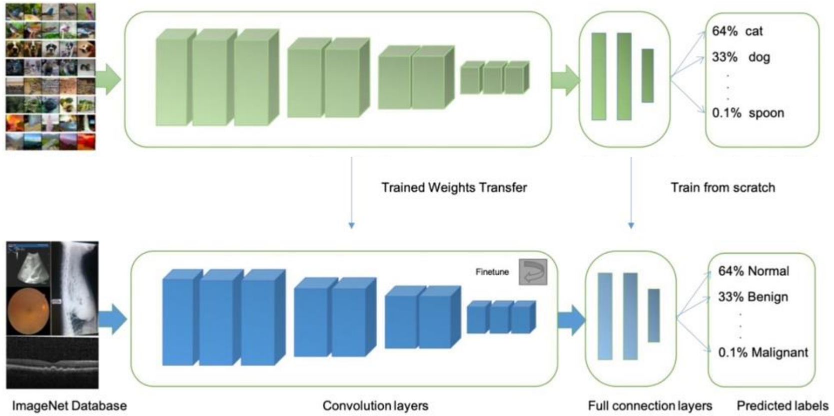
[Wang et al, Annotation-efficient deep learning for automatic medical image segmentation. *Nature Communications*, *12*(1), 1–13 (2021)]

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, 2024 Prog. Biomed. Eng. 6 032001 DOI 10.1088/2516-1091/ad525b]

Integration of data from multiple sources

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

Early fusion:

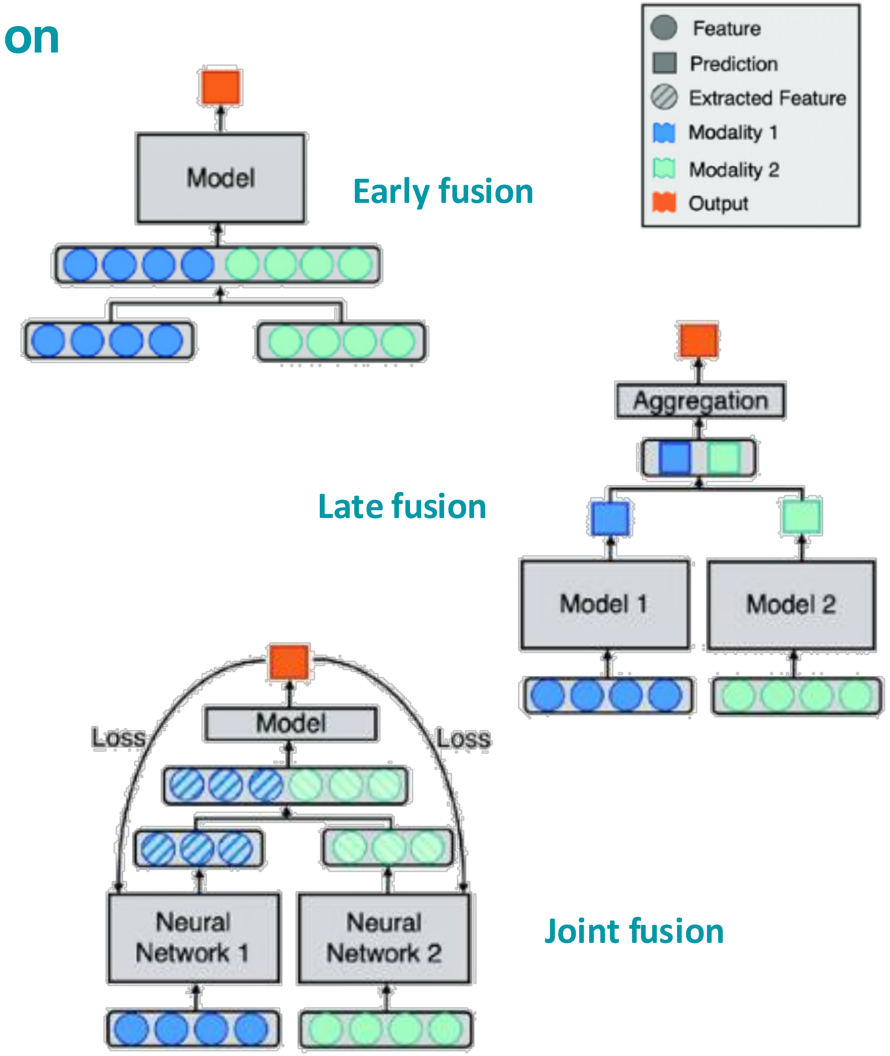
- It is the simplest approach. Input modalities or features are concatenated before any processing.

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.

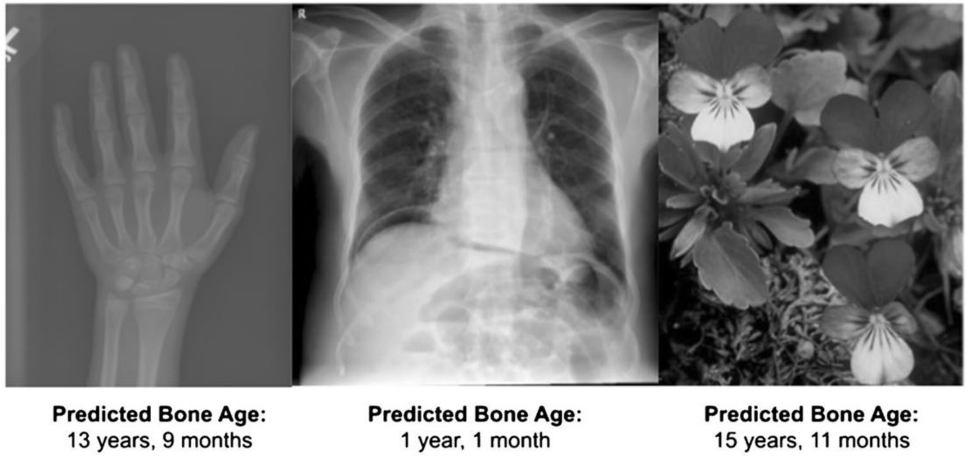


Reliability of AI systems

- 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]

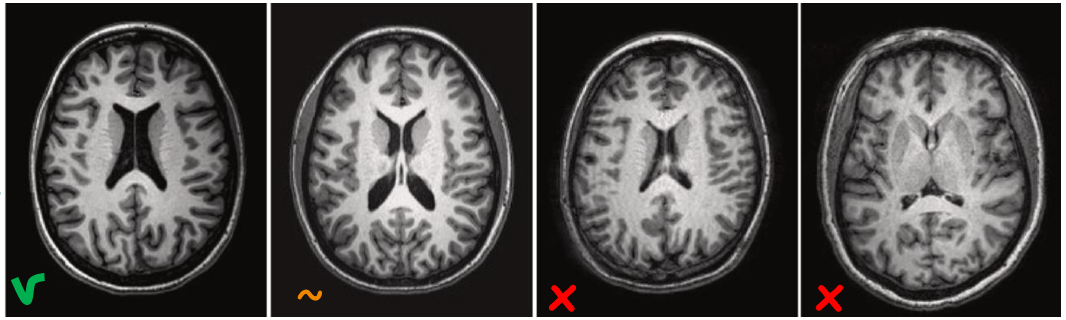
Outputs of a CNN trained to predict bone age from RX of left hands



- 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

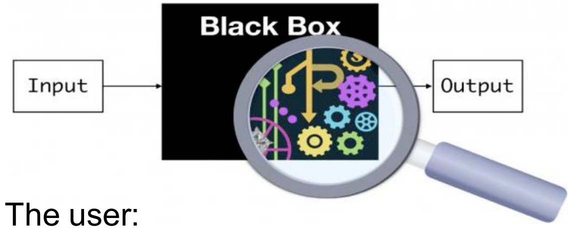
[Fantini et al. (2021). Automatic MR image quality evaluation using a Deep CNN: A reference-free method to rate motion artifacts in neuroimaging. Computerized Medical Imaging and Graphics, 90, 101897. https://doi.org/10.1016/j.compmedimag.2021.101897]

Motion-free vs motion corrupted images



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

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

https://www.pi.infn.it/aim/

[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)

Connections with external entities and other projects

- **Collaboration with Italian Association of Medical Physicists (AIFM)**
  - A 5-year research framework agreement has been renewed in January 2024 between INFN and the Italian Association of Medical Physicists (AIFM). This agreement foresees the collaboration between the parties for the synergistic realization of common research objectives in the healthcare field.
  - **A. Retico, M.E. Fantacci**, **M. Marrale** and **C. Lenardi** are members of the Research Committee of AIFM
  - **C. Talamonti** coordinates the AI working group of AIFM
  - **A. Lascialfari** is the chair of the Research Committee of AIFM
  - **C. Lenardi** is in the Executive Board of AIFM
  - Example of joint activities
    - Organization of **joint webinars INFN-AIFM** on research topics related to the use of AI in the medical field, https://fisicamedica.it/formazione/agenda-eventi-formativi/come-affrontare-insieme-le-sfide-dellia/
    - Participation of AIFM medical Physicists to INFN projects
    - Joint organization of workshop, training opportunities and dissemination events (e.g. Bright-Nights)

- **Collaboration with other associations**
  - **A. Chincarini** is a member of the Executive Board of European Alzheimer's Disease Consortium (**EADC**, https://eadc.online/) and a member of the Neurological Study Committee of the Italian Association of Nuclear Medicine (**AIMN**, https://aimn.it/)
  - **C. Testa** is a member of the Executive Board of the Italian Chapter of the International Society of Magnetic Resonance in Medicine (**AIRMM**, https://www.ismrm.it/it/airmm/)

Connections with external entities and other projects

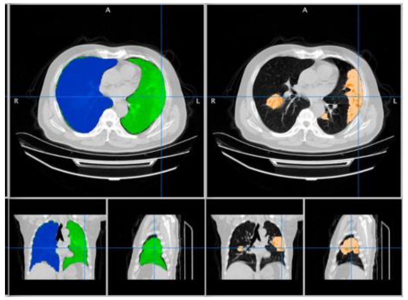
- **PNRR**
  - M4C2 – CN **ICSC – Centro Nazionale di Ricerca in High Performance Computing, Big Data and Quantum Computing** – Spoke 8 – *In silico Medicine and Omics data*
  - M4C2 – PE1 **FAIR - Future Artificial Intelligence Research** – Spoke 8 – *Pervasive AI*
  - M4C2 – PE6 **HEAL - Health Extended ALliance for Innovative Therapies, Advanced Lab-research, and Integrated Approaches of Precision Medicine** – Spoke 2
  - M4C2 – ECS **Tuscany Health Ecosystem (THE)** – Spoke 1 and Spoke 4
  - M4C2 – ECS **Robotics and AI in SociEty (RAISE)**
  - M6C2 – POC - *Predictive tools for precision medicine in prodromal stages of neurodegeneration: quantification of molecular imaging and integration with other biomarkers*
- **Ministry of Health**
  - Piano Operativo Salute (POS) - T2 - **Rete TELENEURART** – *Rete Pediatrica per il tele-monitoraggio e la tele-riabilitazione dei disturbi e delle disabilità del neurosviluppo tramite l'individuazione e l'analisi di biomarker digitali, identificati tramite intelligenza artificiale*
  - RF2021 (PNRR) – *Probing neuroinflammation in the prodromal stages of alpha-synucleinopathies. A multimodal neuroimaging, neurophysiological, neuropsychological proof-of-concept study*

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

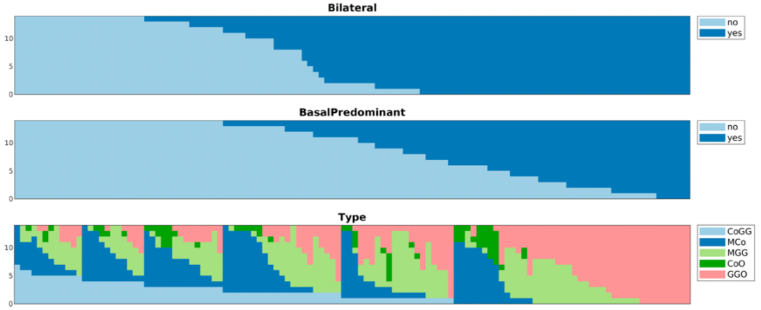
$V_{Consolidation} / V_{Lesion}$

0: unilateral
1: bilateral

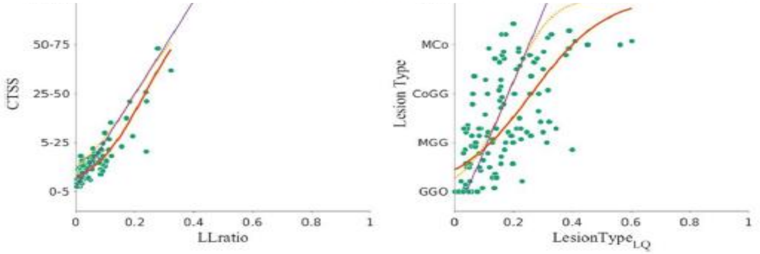
0: basal
100: apical

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:

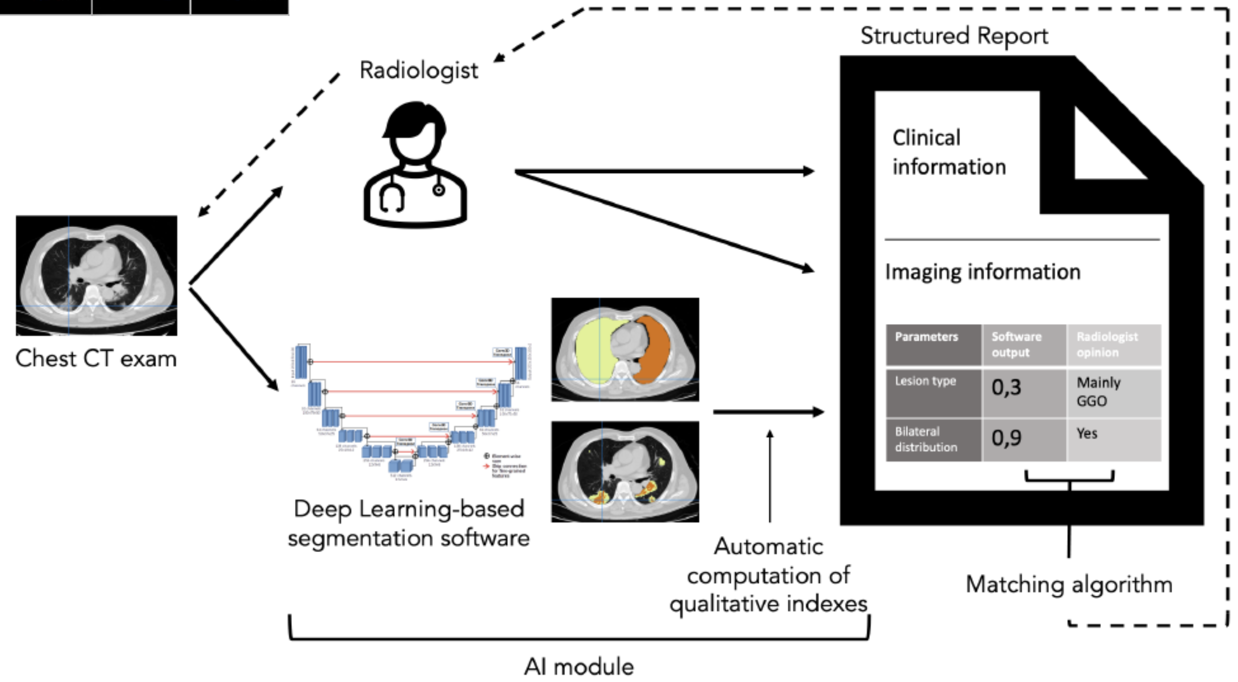
- a poor agreement among the opinions of radiologists



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



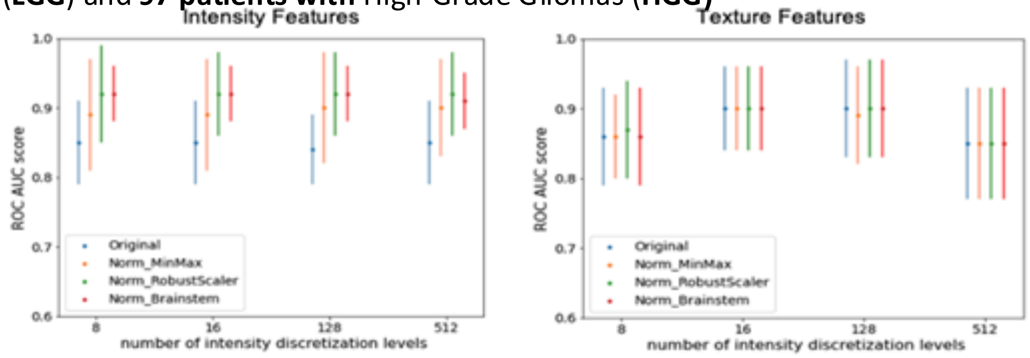
[Chincarini A, Scapicchio C et al A multicenter evaluation of the LungQuant software for lung parenchyma characterization in COVID-19 pneumonia, European Radiology Experimental, https://doi.org/10.1186/s41747-023-00334-z]



[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.]

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

Multiparametric MRI scans (T1, T1-Gd, T2, FLAIR) of **61 patients with** Low-Grade Gliomas (**LGG**) and **97 patients with** High-Grade Gliomas (**HGG**)



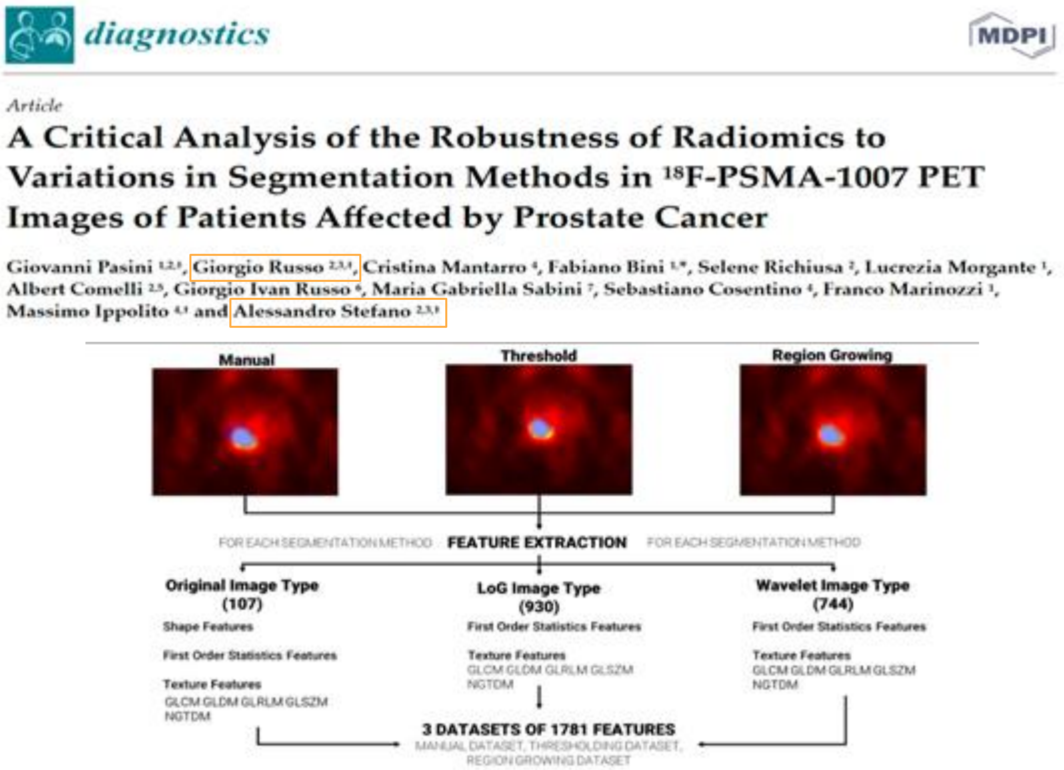
Modality	Raw feature Set (372 Features for all modalities)	MRI-reliable feature Set (372 Features) [Norm_Brainstem] (bin counts = 128)
T1	0.73 ± 0.05	0.69 ± 0.04
T1-Gd	0.89 ± 0.05	0.93 ± 0.05
T2	0.76 ± 0.08	0.75 ± 0.06
T2 FLAIR	0.76 ± 0.08	0.76 ± 0.06
All sequences	0.88 ± 0.08	0.93 ± 0.05

Conclusions

- The **complementary information of multi-parametric MRI** has to be taken into account

- The **image preprocessing** step (normalization and intensity discretization) is relevant for radiomic and ML analysis

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


diagnostics MDPI

Article

A Critical Analysis of the Robustness of Radiomics to Variations in Segmentation Methods in ^{18}F-PSMA-1007 PET Images of Patients Affected by Prostate Cancer

Giovanni Pasini [1,2,†], Giorgio Russo [2,3,‡], Cristina Mantarro [4], Fabiano Bini [1,*], Selene Richiusa [2], Lucrezia Morgante [1], Albert Comelli [2,5], Giorgio Ivan Russo [6], Maria Gabriella Sabini [7], Sebastiano Cosentino [4], Franco Marinozzi [1], Massimo Ippolito [4,‡] and Alessandro Stefano [2,3,‡]

RESULTS:

1. Shape feature class demonstrated the **least robustness**, while the GLCM feature class exhibited the **highest robustness**.
   Furthermore, segmentation methods **significantly impacted** feature selection.
2. **High performance** was achieved using region growing and DA to **discriminate** between low-risk and high-risk prostate patients.

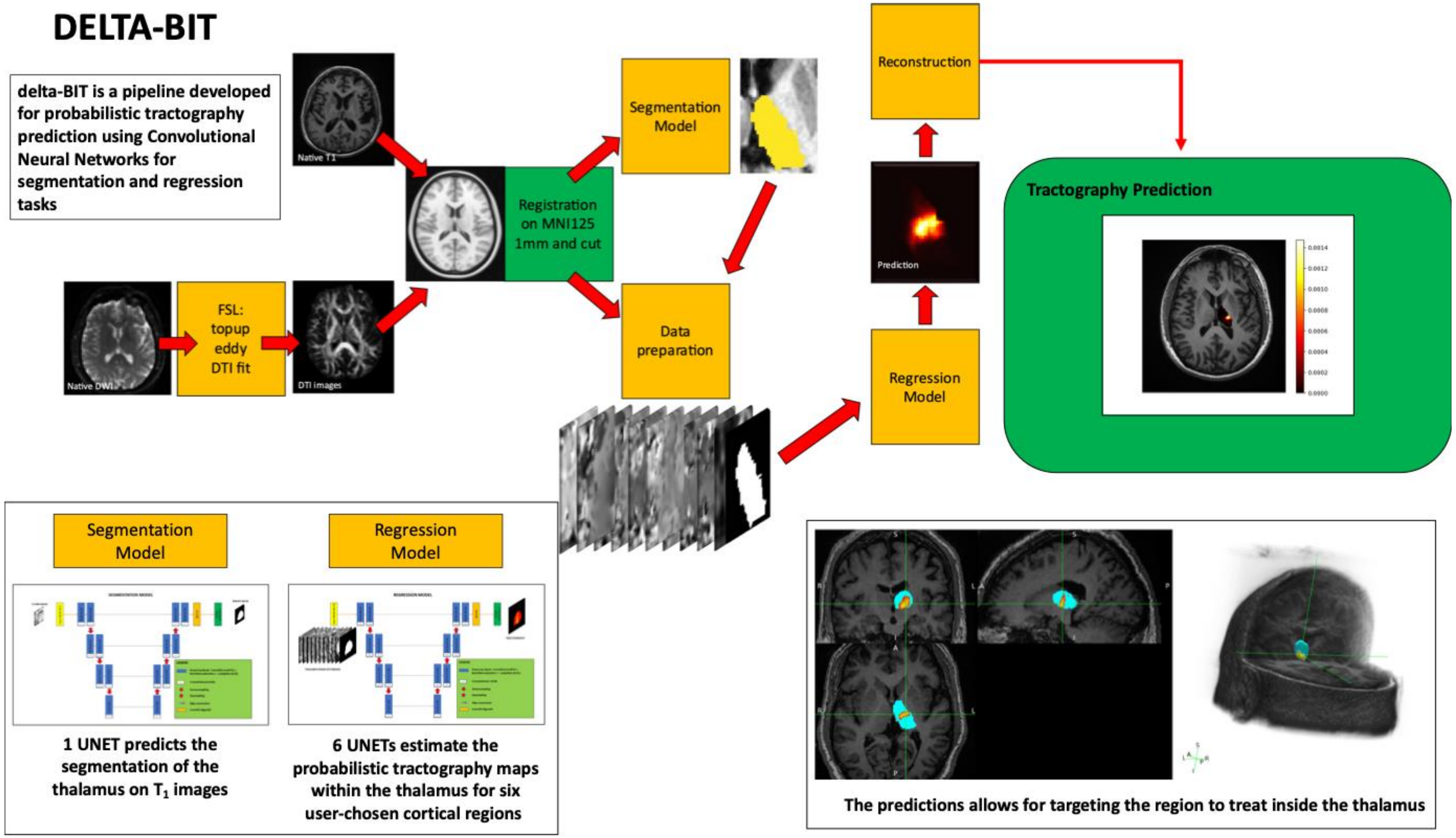
Deep Learning models in tcMRgFUS

Development of deep learning models for identification of the target for thermal ablation by means of transcranial-MR guided Focused Ultrasound Surgery for patients with essential tremor

Patiens with essential tremor

Transcranial-MR guided Focused Ultrasound Surgery (tcMRgFUS)

Romeo at al. Submitted to NeuroImage
Software available on Github
https://github.com/mromeo1992/delta-BIT



Harmonization of multicentric datasets

ABIDE Dataset

The Autism Brain Imaging Data Exchange

Public dataset, 24 collection centers

Retrospectively collected data

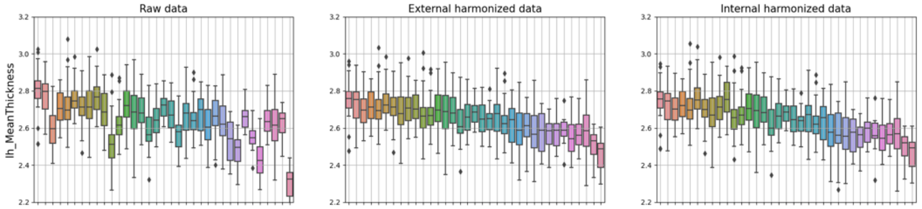
More than 2000 subjects (equally divided between ASD and TD)

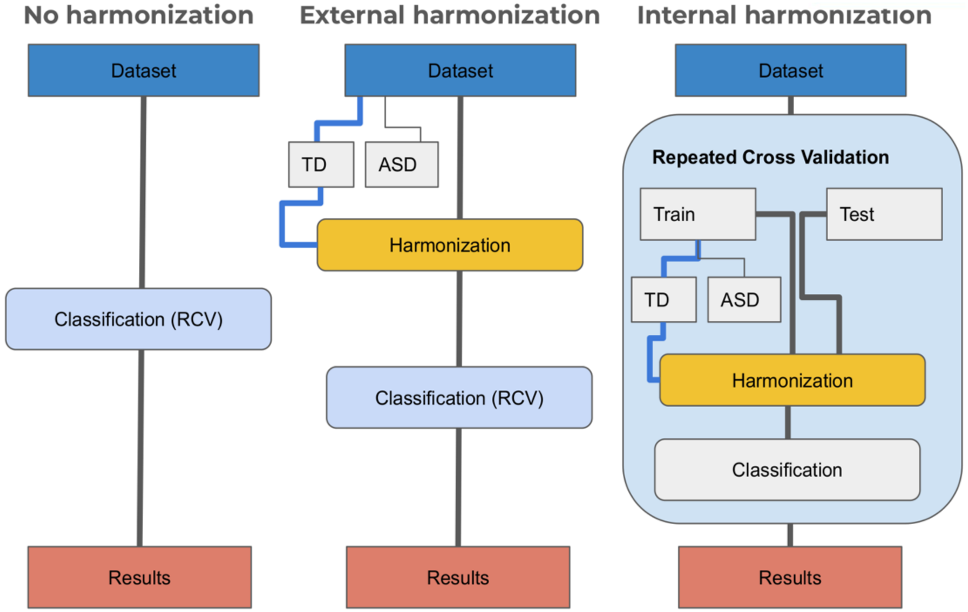
Ages: 5-64 years

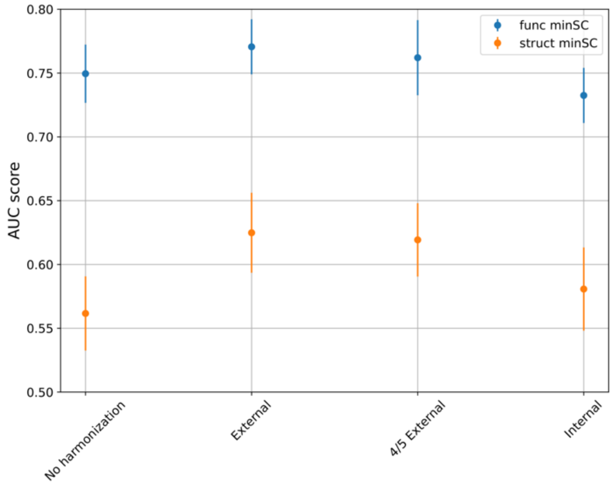
http://fcon_1000.projects.nitrc.org/indi/abide/

ABIDE
Autism Brain Imaging Data Exchange

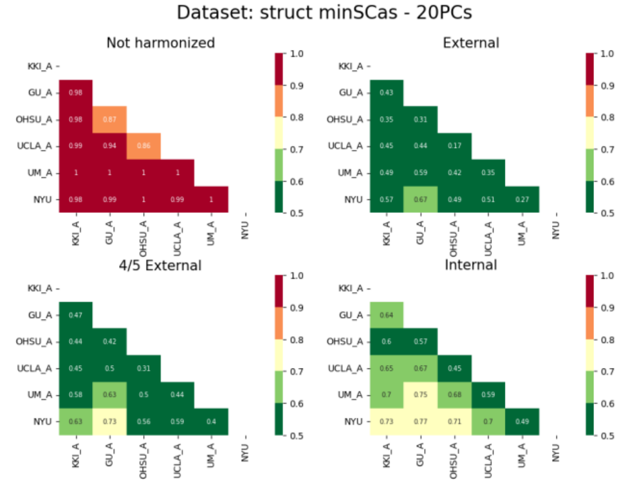
$$y_{ijv}^{\mathrm{ComBat}} = \frac{y_{ijv} - \widehat{\alpha}_v - \mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}_v - \gamma^*_{iv}}{\delta^*_{iv}} + \widehat{\alpha}_v + \mathbf{X}_{ij}\widehat{\boldsymbol{\beta}}_v$$







Sites Distinguishability



- The <u>harmonization strategy affects the classification results</u>
- The important features <u>depend on the harmonization scheme</u>.
- The most methodologically correct approach is to apply harmonization within the cross-validation scheme, using only the subjects from the training set to calculate the harmonization model parameters

Serra, G., Mainas, F., Golosio, B., Retico, A., Oliva, P., Effect of data harmonization of multicentric dataset in ASD/TD classification, Brain Informatics, 2023, 10(1), 32 DOI https://doi.org/10.1186/s40708-023-00210-x

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:

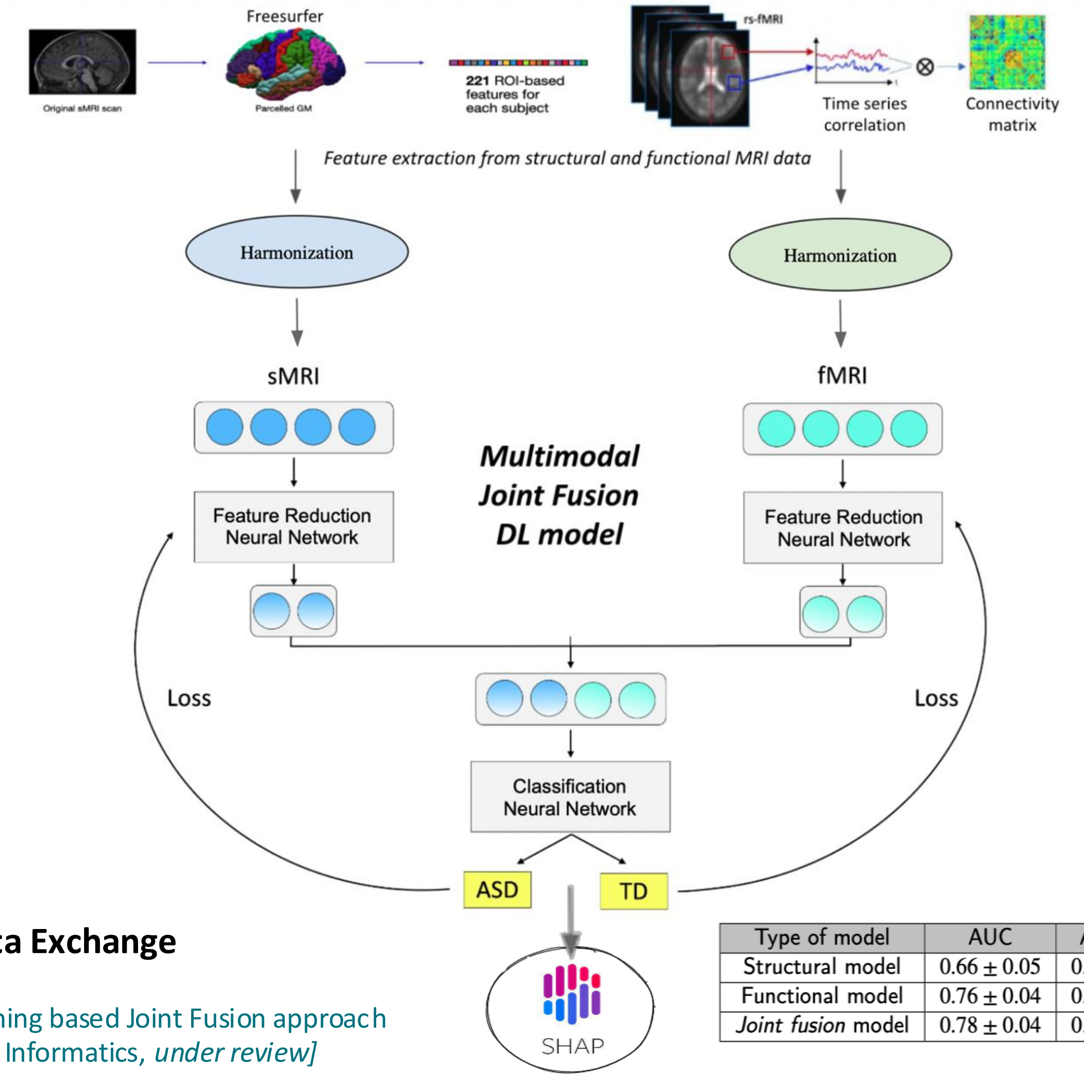
- The Feature Reduction and the Feature Classification Neural Networks are trained using a single cost function, thus the most meaningful features for the classification are extracted
- The model was trained with 150 epochs within a 10-fold cross validation scheme

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*]

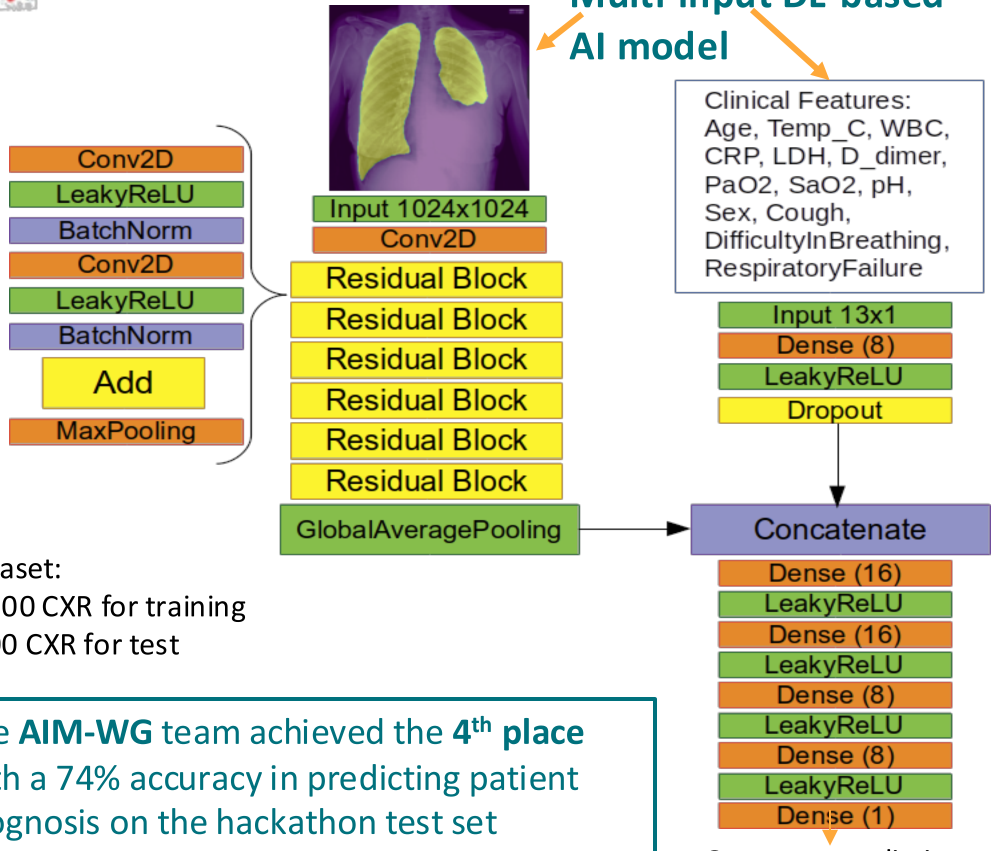
Prediction of COVID-19 severity: the covidcxr hackathon

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)

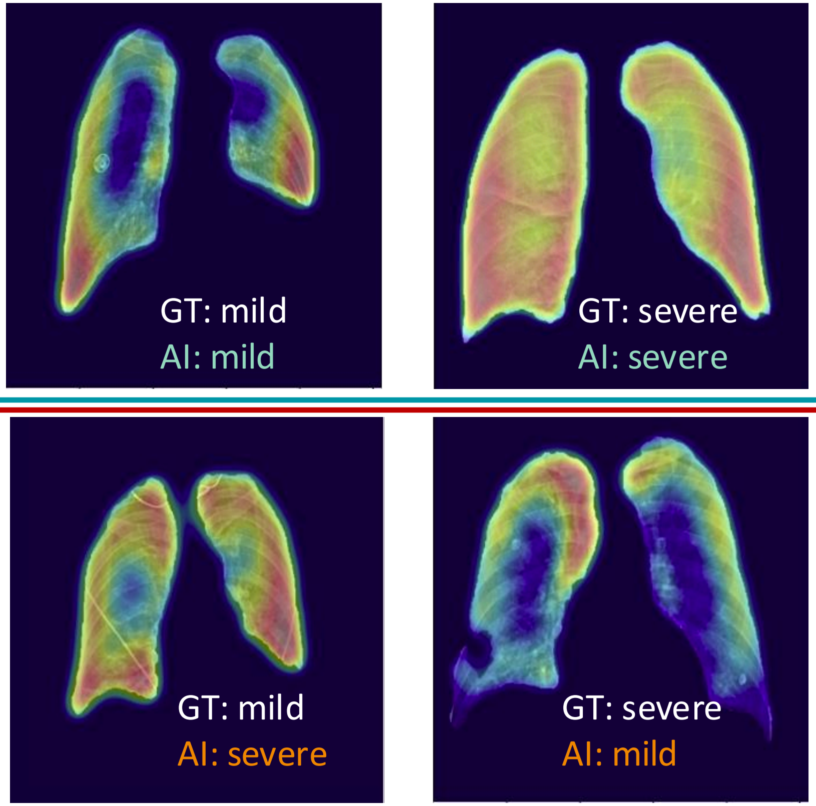
Dataset:
~1100 CXR for training
~500 CXR for test

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

Correctly classified by the multi-input CNN

Misclassified by the multi-input CNN



Lizzi F, Brero F, Fantacci ME, Lascialfari A, Paternò G, Postuma I, Oliva P, Scapicchio C, Retico A (2024). A multi-input deep learning model to classify COVID-19 pneumonia severity from imaging and clinical data. IWBBIO, 1–12.

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, further steps will be necessary to achieve precision medicine:
  - Large annotated datasets
  - Ability to merge different types of information
  - Relevant clinical endpoints

- **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!

Contact: oliva@uniss.it
University of Sassari
INFN, Cagliari Division
https://www.pi.infn.it/aim/