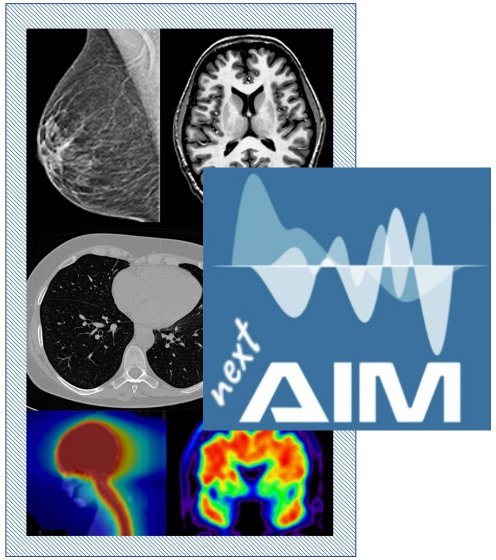
Workshop sul Calcolo nell'INFN

Palau 20-24 Maggio 2024

Open challenges of Artificial Intelligence applied to medical imaging: the **nextAIM** project



Camilla Scapicchio

INFN - Sezione di Pisa

CSN5 Ricerca Tecnologica

Outline

- Artificial Intelligence in medical imaging
- Open challenges
- Computing challenges
- The NextAIM project
- COVID-19 applications
- Overview of study cases
- Conclusions

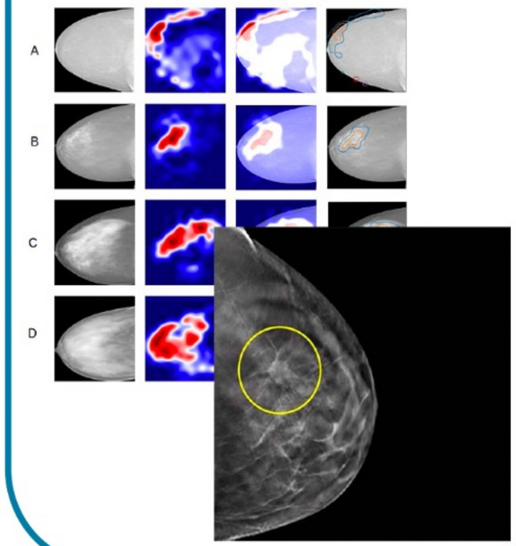
Artificial Intelligence in medical imaging

Medical images are not simple images, they reflect various physical properties of the body and the underlying pathophysiology. They can be converted into meaningful and **mineable data** through a quantification process.

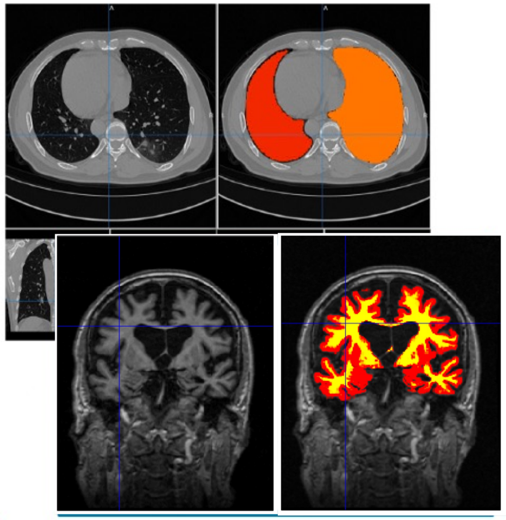
These quantitative data are **not easily interpretable by the human** mind without risk of error.

AI steps in, offering the ability to process vast amounts of data and extract meaningful insights through complex algorithms.

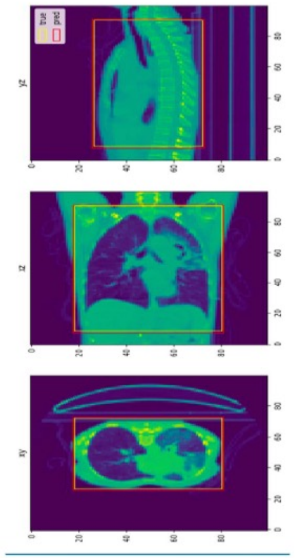
Classification



Segmentation

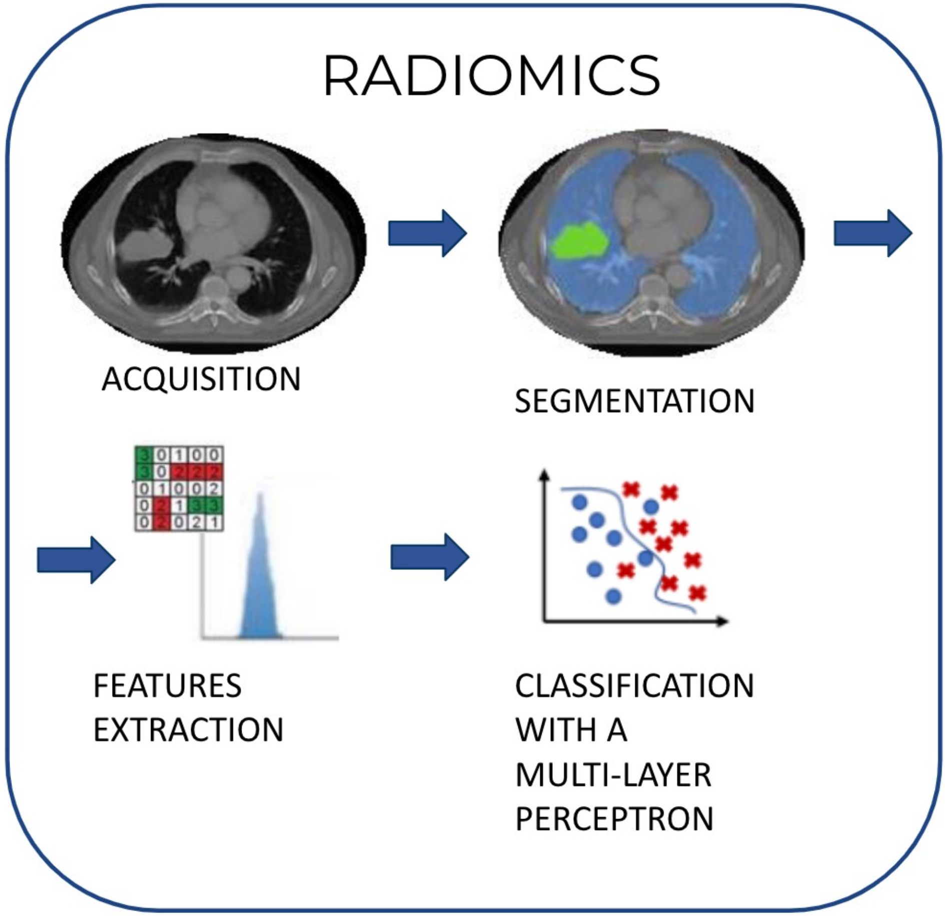


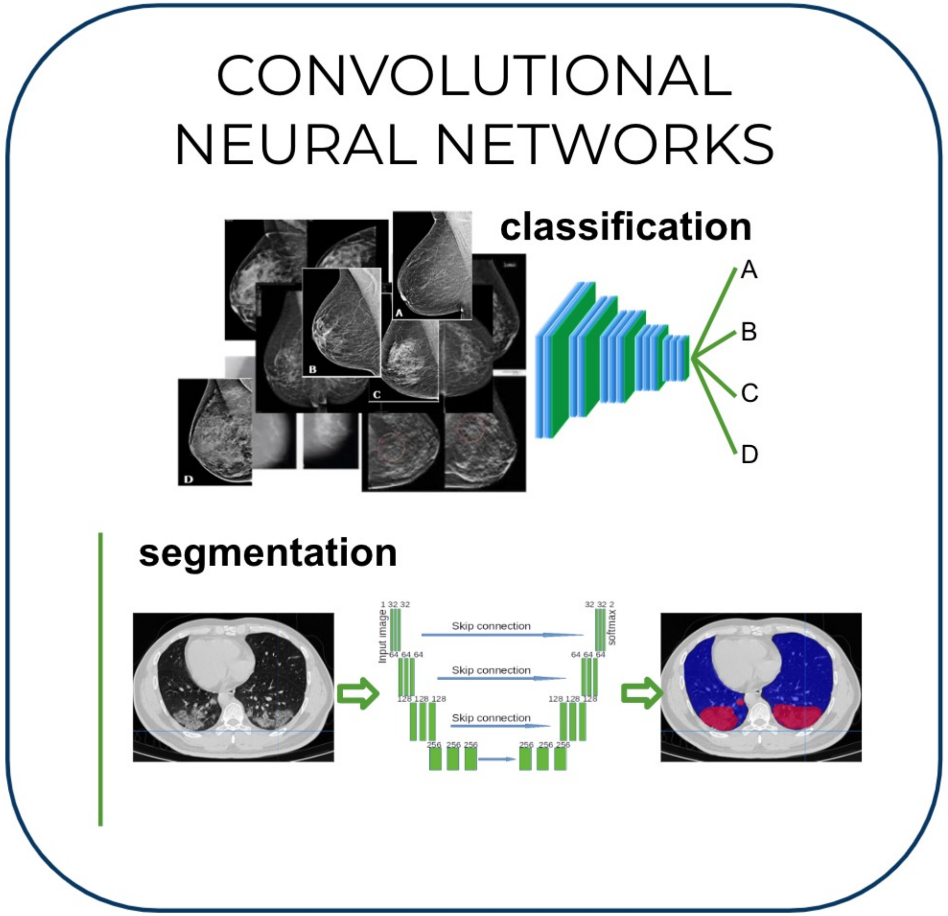
Regression



Artificial Intelligence in medical imaging

Two main paradigms: **(Radiomics** and **Machine Learning)** vs **(Deep Learning)**





The nextAIM project

Artificial Intelligence in Medicine (AIM): next *steps*

focus on no-so-big data and explainable techniques

Main goal: to take steps towards developing robust and explainable AI algorithms and validating them on realistic use cases in the medical field

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

Involves 13 INFN groups

Resp. Nazionale: A. Retico (PI)

All activities rely on a long-standing collaboration with Italian medical centers (hospitals and IRCCS) and with international consortia for data sharing.

Resp. Locali:

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)
Padova (A. Zucchetta)
Pavia (A. Lascialfari)
Pisa (M.E. Fantacci)

WP1 **Challenge I: no-so-big data**

Strategies for efficient learning with limited data samples.

Evaluation of robustness and reliability of trained models.

WP2 **Challenge II: explainable AI (XAI)**

Make AI results understandable to humans.

Which image/data features were relevant to make a decision?

WP3 **Applications to real-world data samples**

Practical medical data analysis use cases on available samples (public data, private collections, integration of both), where the analysis approaches a, b or both are implemented, and the challenges I, II or both are encountered.

Implementation, test and validation in collaboration with colleagues working in Clinical context

WP4 **Computing resources and SW repository organization**

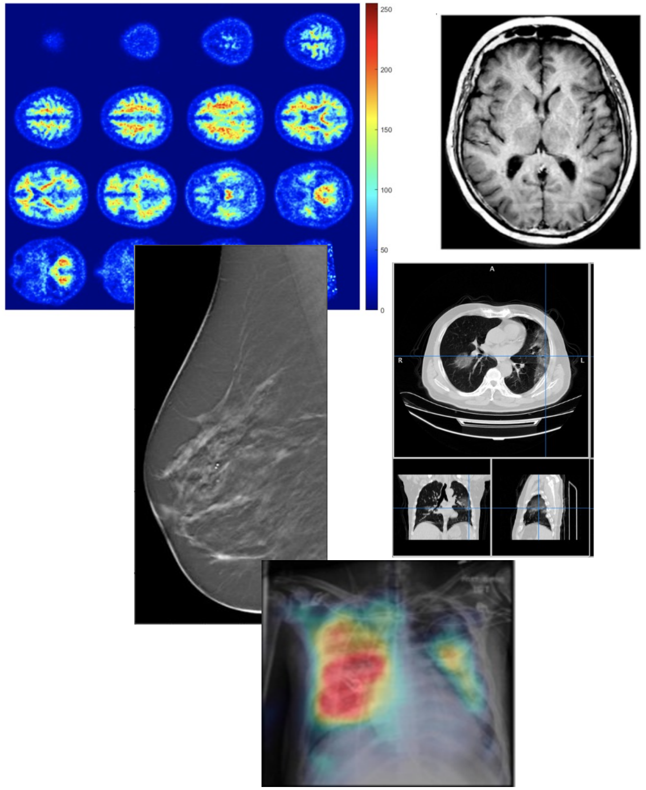
(ReCaS, IBiSCo, INFN-Cloud + risorse HW locali)

WP5 **Exploitation of research results and communication**

(connection with AIFM, conferences, publications)

The nextAIM project

Main topics



Radiomics in Digital Breast Tomosynthesis (DBT)
Super-Resolution in Medical Imaging
Radiomics in prostate cancer
Radiomics and DL in tcMRgFUS
Nuclear Imaging Quantification and Radiomics
Connectivity in functional MRI and EEG
Radiomics and Deep Learning analysis of CT and patients' data in COVID-19
Radiomics and ML-segmentation on Facio-Scapulo-Humeral dystrophy (FSHD), lung and liver tumor
ML on Imaging data of 10B uptake tracks and dose monitoring by Compton cameras
Artificial intelligence for monitoring RT response in soft-tissue sarcomas
Machine Learning techniques for cardiological applications
Application of NLP techniques to clinical notes towards the automated reading of instrumental data

Open challenges….(just a few)

Improving the robustness and generalization ability of AI models… even with limited data availability

Quantity of data — Difficulties in data collection and curation, limited availability of annotated data, privacy regulations, and restrictions on data sharing

Standardization and harmonization — Lack of general guidelines to unify and align data of different sources and characteristics.

«Multicenter»

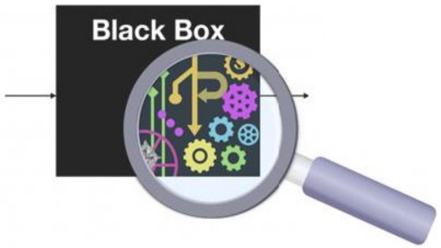
Open challenges….(just a few)

Improving the robustness and generalization ability of AI models… even with limited data availability

Quantity of data Difficulties in data collection and curation, limited availability of annotated data, privacy regulations, and restrictions on data sharing

Standardization and harmonization Lack of general guidelines to unify and align data of different sources and characteristics. «Multicenter»

Developing explainable AI methods



Explainability Explaining the decision process and the reason behind the prediction

Computing challenges

Hardware – Software trade-off

Balancing computational capabilities of hardware with software optimization in the development of AI models

1. **Hardware:** The type of hardware (CPU, GPU, TPU) can have a significant impact on model performance.

2. **Software**: Software and model specifications may vary depending on hardware.

Computing challenges

Hardware – Software trade-off

Balancing computational capabilities of hardware with software optimization in the development of AI models

1. **Hardware**: The type of hardware (CPU, GPU, TPU) can have a significant impact on model performance.

2. **Software**: Software and model specifications may vary depending on hardware.

TensorFlow PyTorch — Offer support on a variety of hardware, but may require different configurations to optimize performance on specific hardware

Computing challenges

Hardware – Software trade-off

Balancing computational capabilities of hardware with software optimization in the development of AI models

1. **Hardware**: The type of hardware (CPU, GPU, TPU) can have a significant impact on model performance.

2. **Software**: Software and model specifications may vary depending on hardware.

TensorFlow PyTorch Offer support on a variety of hardware, but may require different configurations to optimize performance on specific hardware

Packages conflict and virtual environments :

When hardware architecture is changed or software components are upgraded, conflicts can arise between packages and virtual environments must be used for model development.

Available facilities and tricks

Local Resources

Pisa: our machine (no queue!)

- **3 Nvidia V100** (2 16GB and 1 32GB VRAM);
- **4 Nvidia K80** with 8 GB VRAM;
- accessible with AAI credentials;
- to use this resources ask me.

Pavia: our machine (no queue!)

- 2 Nvidia RTX3090 (24GB)
- 1 Nvidia RTX3060 (12GB)
- to use this resources ask me*

National Resources

ML-INFN:

1 Nvidia Tesla T4 16 GB VRAM

ReCaS:

- 1 node with:
  3 Nvidia A100 with 40GB VRAM;
  256 core;
  2 TB RAM;
  100 TB storage.

AI INFN
14 partizioni MIG di A100 da 10GB ciascuna e 4 GPU RTX5000

Available facilities and tricks

Local Resources

Pisa: our machine (no queue!)

- **3 Nvidia V100** (2 16GB and 1 32GB VRAM);
- **4 Nvidia K80** with 8 GB VRAM;
- accessible with AAI credentials;
- to use this resources ask me.

Pavia: our machine (no queue!)

- 2 Nvidia RTX3090 (24GB)
- 1 Nvidia RTX3060 (12GB)
- to use this resources ask me*

National Resources

ML-INFN:

1 Nvidia Tesla T4 16 GB VRAM

ReCaS:

- 1 node with:
  3 Nvidia A100 with 40GB VRAM;
  256 core;
  2 TB RAM;
  100 TB storage.

AI INFN
14 partizioni MIG di A100 da 10GB ciascuna e 4 GPU RTX5000

See presentation by Rosa Petrini on AI-INFN

Available facilities and tricks

Local Resources

Pisa: our machine (no queue!)
- **3 Nvidia V100** (2 16GB and 1 32GB VRAM);
- **4 Nvidia K80** with 8 GB VRAM;
- accessible with AAI credentials;
- to use this resources ask me.

Pavia: our machine (no queue!)
- 2 Nvidia RTX3090 (24GB)
- 1 Nvidia RTX3060 (12GB)
- to use this resources ask me*

National Resources

ML-INFN:

1 Nvidia Tesla T4 16 GB VRAM

ReCaS:
- 1 node with:
  3 Nvidia A100 with 40GB VRAM;
  256 core;
  2 TB RAM;
  100 TB storage.

AI INFN
14 partizioni MIG di A100 da 10GB ciascuna e 4 GPU RTX5000

See presentation by Rosa Petrini on AI-INFN

Tricks we use to compensate for the memory constraints:

- **Image size**: Images are often down sampled to a lower resolution
- **Batch size**: Batch sizes are often reduced to one or two images
- **Patching**: Images are often subsampled into overlapping patches
- **Model Complexity**: Reductions in the number of feature maps and/or layers are often necessary
- **Model Parallelism**: Models may be distributed across several compute nodes in a parallel fashion

Available facilities and tricks

Local Resources

Pisa: our machine (no queue!)
- **3 Nvidia V100** (2 16GB and 1 32GB VRAM);
- **4 Nvidia K80** with 8 GB VRAM;
- accessible with AAI credentials;
- to use this resources ask me.

Pavia: our machine (no queue!)
- 2 Nvidia RTX3090 (24GB)
- 1 Nvidia RTX3060 (12GB)
- to use this resources ask me*

National Resources

ML-INFN:

1 Nvidia Tesla T4 16 GB VRAM

ReCaS:
- 1 node with:
  3 Nvidia A100 with 40GB VRAM;
  256 core;
  2 TB RAM;
  100 TB storage.

AI INFN
14 partizioni MIG di A100 da 10GB ciascuna e 4 GPU RTX5000

See presentation by Rosa Petrini on AI-INFN

Tricks we use to compensate for the memory constraints:

- **Image size**: Images are often down sampled to a lower resolution
- **Batch size**: Batch sizes are often reduced to one or two images
- **Patching**: Images are often subsampled into overlapping patches
- **Model Complexity**: Reductions in the number of feature maps and/or layers are often necessary
- **Model Parallelism**: Models may be distributed across several compute nodes in a parallel fashion

These tricks were not created to obtain better models — they are instead necessary workarounds for hardware limitations.

Researchers would prefer to use the full-resolution image without having to account for hyperparameters such as batch size, model complexity, or subsampling!!!

Computing challenges

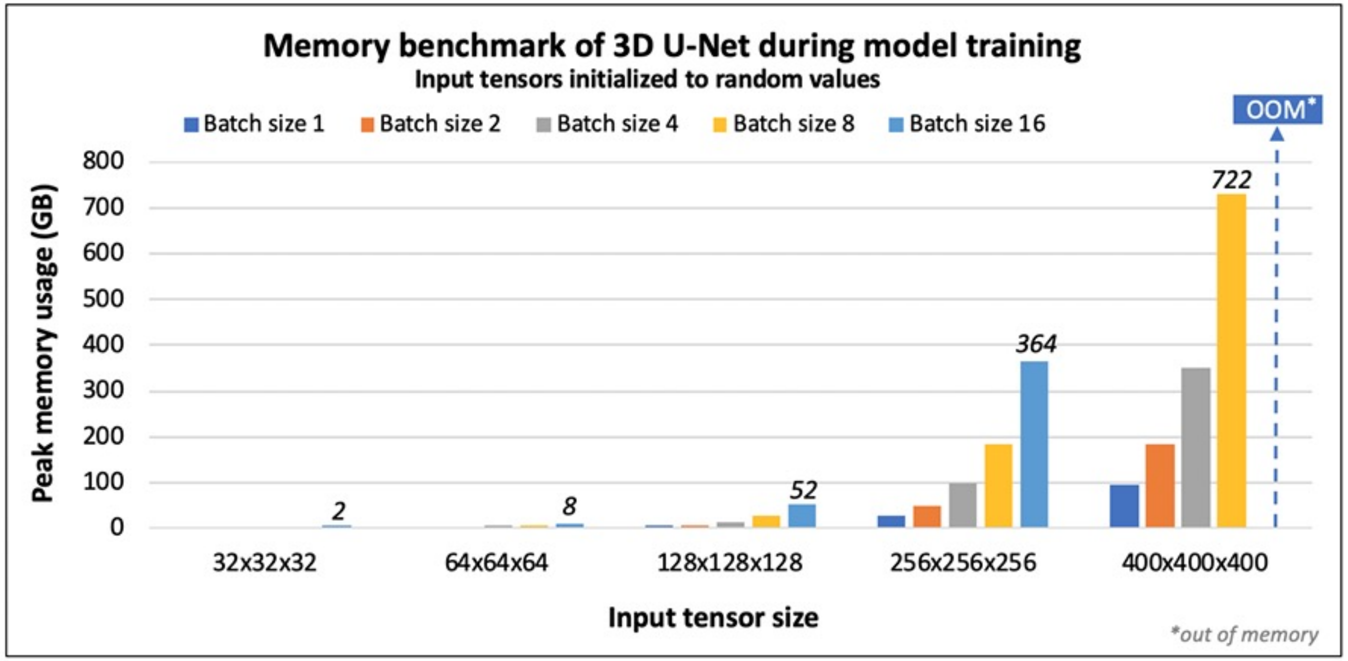
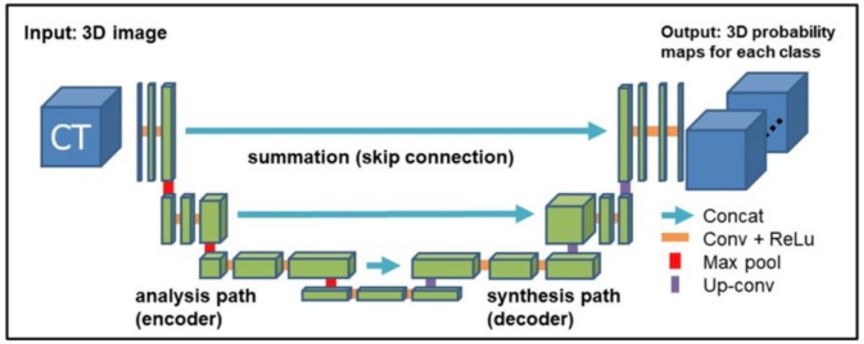


*Figure 3. Benchmarking the memory usage of 3D U-Net model-training over various input tensors sizes on an Intel Xeon Scalable Processor-based server with **1.5 TB system***

https://dl.dell.com/manuals/common/dellemc_overcoming_memory_bottleneck_ai_healthcare.pdf



Example augmented dataset size: 6GB
Example trained model: nnunet: 806 GB

High processing requirements of medical data analysis may be addressed with hardware accelerators (**GPU**s).
These models' **memory footprint** is not solely due to trainable parameters (also several million), but also to the model's **activation maps**, which are a function of the size of the input to the network.

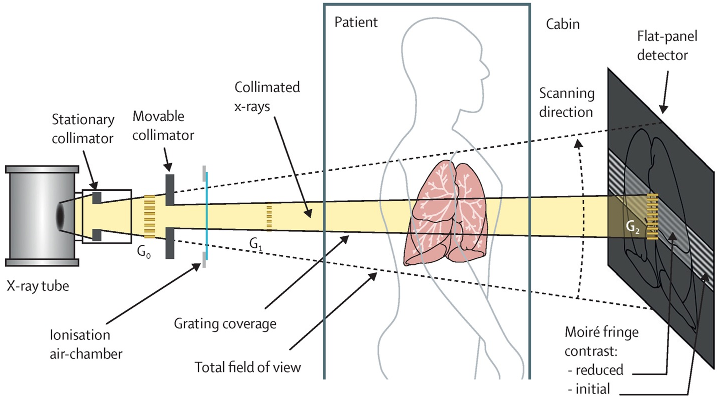
Models that use **large batch, high resolution, high dimensional image inputs** often require more memory than the accelerator card can accommodate.

COVID-19 applications

Multi-input CNN for severity prediction on CXR data

AI4COVID Hackathon
(https://ai4covid-hackathon.it/)

CXR
(Chest Radiography)



CRX image

+

Clinical features

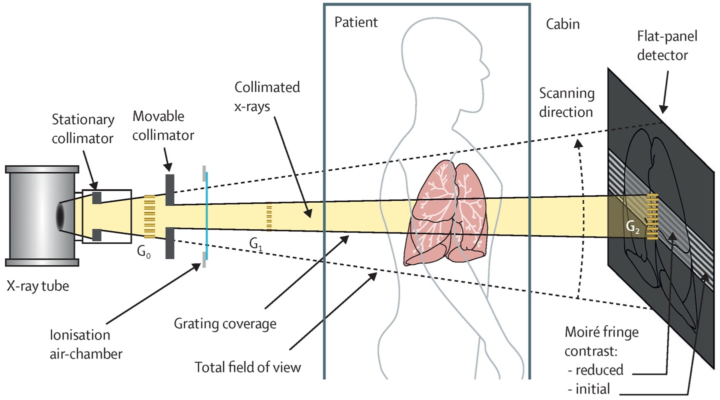
Age
Sex
Body Temperature (°C)
Cough
Dyspnea
WBC
CRP
Fibrinogen
LDH
D-dimer
O2
PaO2
SaO2
pH
Cardiovascular Disease
Respiratory Failure

COVID-19 applications

Multi-input CNN for severity prediction on CXR data

AI4COVID Hackathon
(https://ai4covid-hackathon.it/)

CXR
(Chest Radiography)



CRX image

Clinical features

+

Age
Sex
Body Temperature (°C)
Cough
Dyspnea
WBC
CRP
Fibrinogen
LDH
D-dimer
O2
PaO2
SaO2
pH
Cardiovascular Disease
Respiratory Failure

Conv2D
LeakyReLU
BatchNorm
Conv2D
LeakyReLU
BatchNorm
Add
MaxPooling

Input 1024x1024
Conv2D
Residual Block
Residual Block
Residual Block
Residual Block
Residual Block
Residual Block
GlobalAveragePooling

Clinical Features:
Age, Temp_C, WBC, CRP, LDH, D_dimer, PaO2, SaO2, pH, Sex, Cough, DifficultyInBreathing, RespiratoryFailure

Input 13x1
Dense (8)
LeakyReLU
Dropout

Concatenate
Dense (16)
LeakyReLU
Dense (16)
LeakyReLU
Dense (8)
LeakyReLU
Dense (8)
LeakyReLU
Dense (1)

Hyperparamters

- Training for 100 epochs
- Adam optimizer
- Binary Cross-Entropy loss function
- Accuracy as evaluation metric

COVID-19 applications

Multi-input CNN for severity prediction on CXR data

Training on a Tesla V100 GPU provided by the Computing center of INFN-Pisa

→ Gray Net Model →

Correct gray level

→ 2D U-net →

CXR lung mask

Clinical Features:
Age, Temp_C, WBC, CRP, LDH, D_dimer, PaO2, SaO2, pH, Sex, Cough, DifficultyInBreathing, RespiratoryFailure

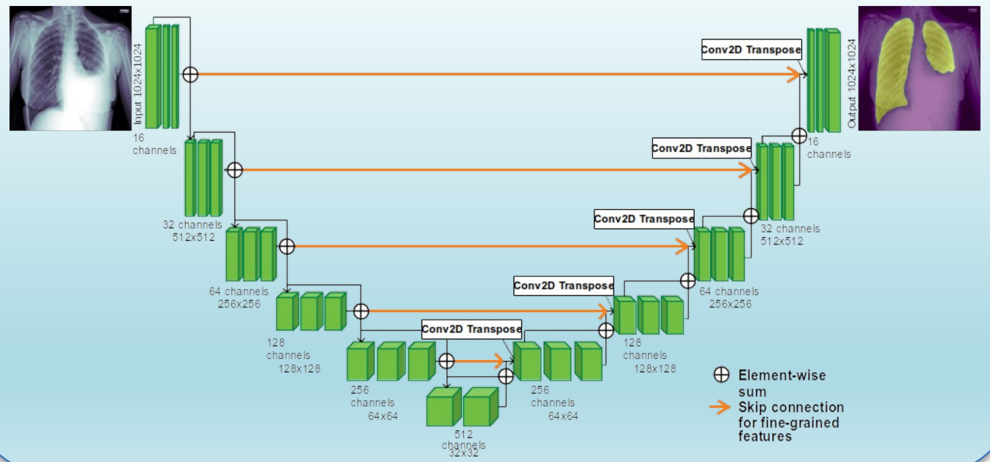
CNN	Metric	Performance
Gray-Level	Accuracy	98.9%
Lung Segmentation	DSC	0.96 ± 0.03
Severity	AUC	84%
	Accuracy	76%
	Sensitivity	77%
	Specificity	76%
	Precision/PPV	67%
	NPV	85%

Hyperparameters
- 300 epochs
- Adam optimizer
- DSC loss

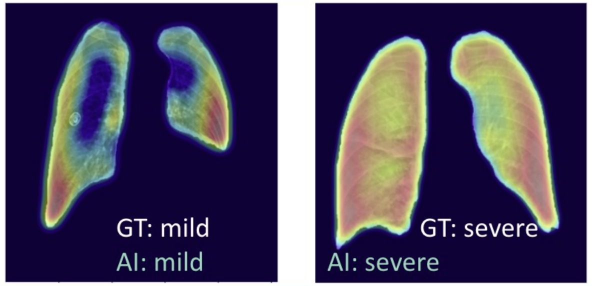
Severity Net Model (Multi-input CNN)

↓

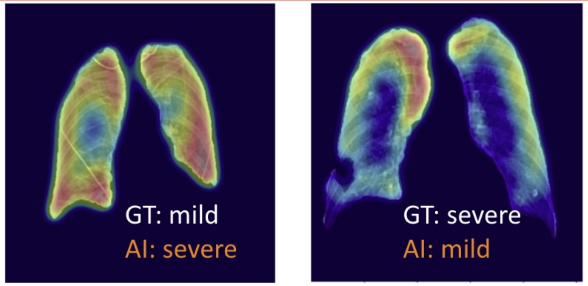
Severe / Mild



Correctly classified by the multi-input CNN



Misclassified by the multi-input CNN



Explainability

Features that can better predict the outcome:
- *PaO2 (Partial pressure of oxygen)*
- *SaO2 (arterial oxygen saturation)*

COVID-19 applications

Multi-input CNN for severity prediction on CXR data

Training on a Tesla V100 GPU provided by the Computing center of INFN-Pisa

→ Gray Net Model →

Correct gray level

→ 2D U-net →

CXR lung mask

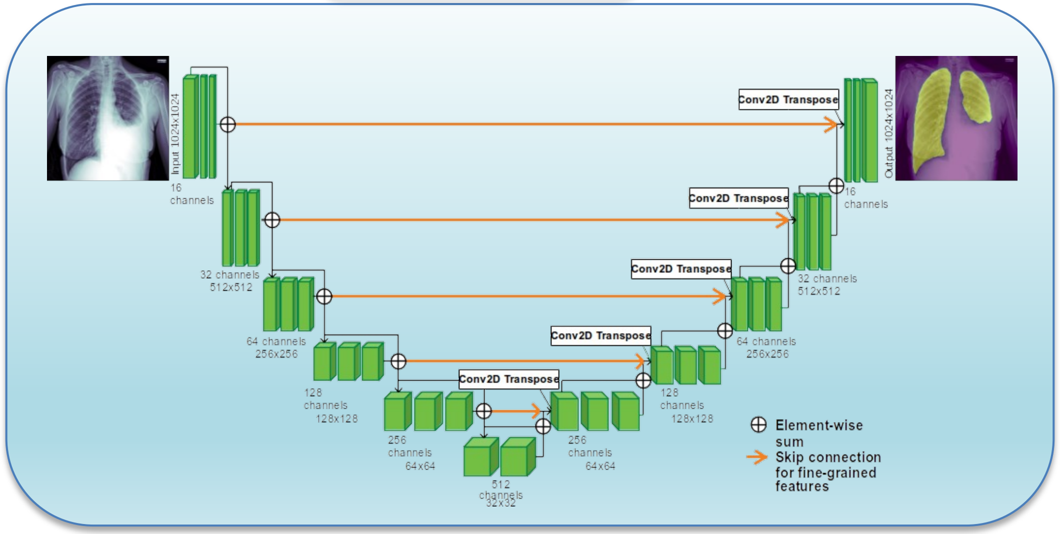
Clinical Features:
Age, Temp_C, WBC, CRP, LDH, D_dimer, PaO2, SaO2, pH, Sex, Cough, DifficultyInBreathing, RespiratoryFailure

Severity Net Model
(Multi-input CNN)

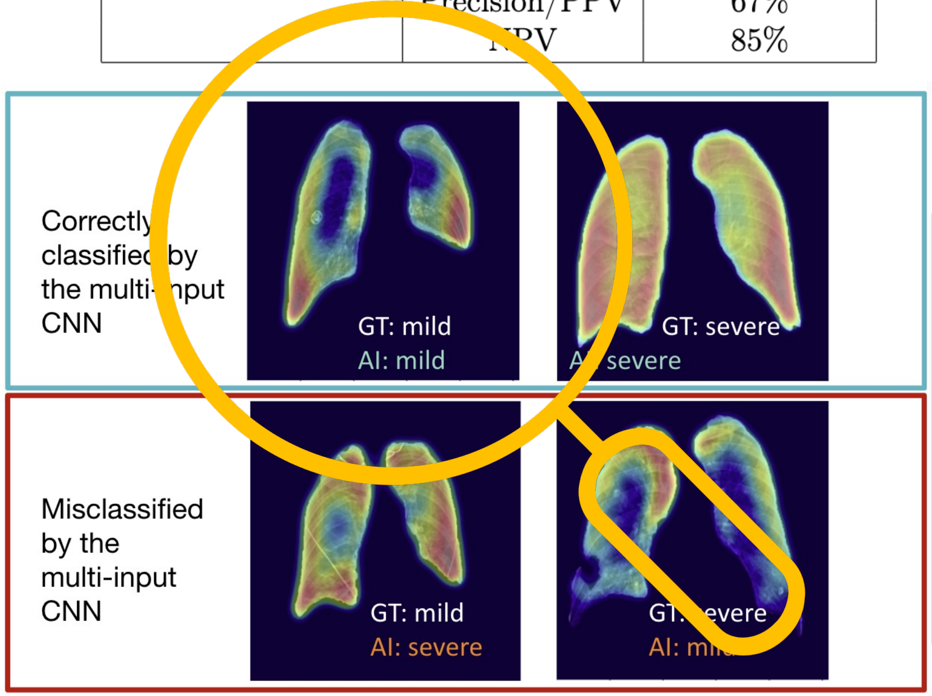
↓

Severe / Mild

Hyperparameters
- 300 epochs
- Adam optimizer
- DSC loss



CNN	Metric	Performance
Gray-Level	Accuracy	98.9%
Lung Segmentation	DSC	0.96 ± 0.03
Severity	AUC	84%
	Accuracy	76%
	Sensitivity	77%
	Specificity	76%
	Precision/PPV	67%
	NPV	85%



Explainability

Features that can better predict the outcome:

- *PaO2 (Partial pressure of oxygen)*
- *SaO2 (arterial oxygen saturation)*

COVID-19 applications

Training on a Tesla V100 GPU provided by the Computing center of INFN-Pisa

Multi-input CNN for severity prediction on CXR data

→ Gray Net Model →

Correct gray level

→ 2D U-net →

CXR lung mask

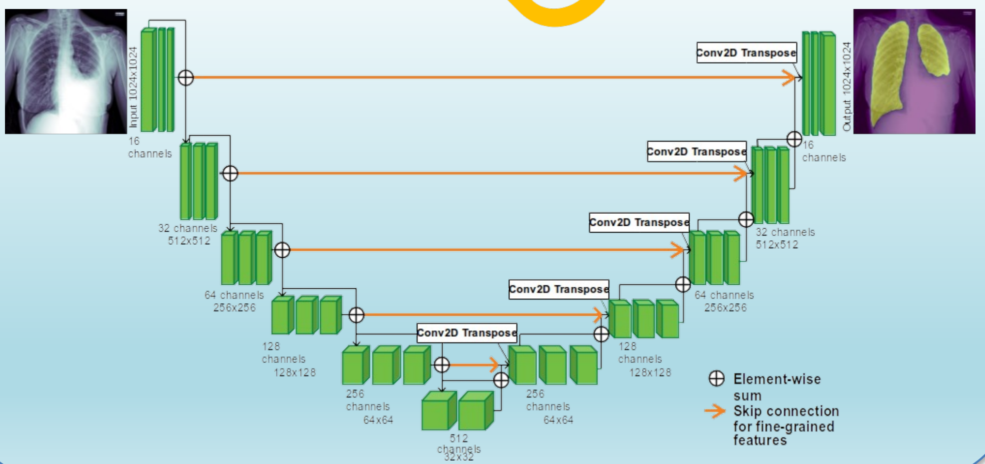
Clinical Features:
Age, Temp_C, WBC, CRP, LDH, D_dimer, PaO2, SaO2, pH, Sex, Cough, DifficultyInBreathing, RespiratoryFailure

Severity Net Model
(Multi-input CNN)

↓

Severe / Mild

Hyperparameters
- 300 epochs
- Adam optimizer
- DSC loss

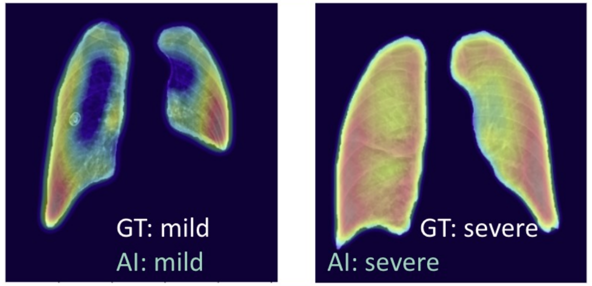


CNN	Metric	Performance
Gray-Level	Accuracy	98.9%
Lung Segmentation	DSC	0.96 ± 0.03
Severity	AUC	84%
	Accuracy	76%
	Sensitivity	77%
	Specificity	76%
	Precision/PPV	67%
	NPV	85%

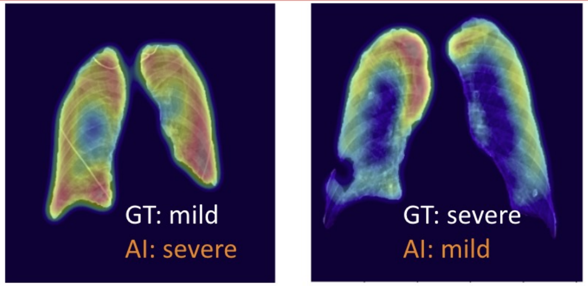
Explainability

Features that can better predict the outcome:
- *PaO2 (Partial pressure of oxygen)*
- *SaO2 (arterial oxygen saturation)*

Correctly classified by the multi-input CNN



Misclassified by the multi-input CNN

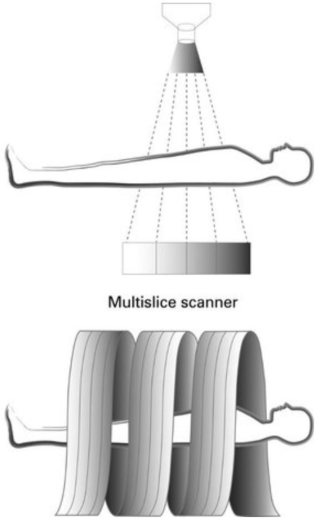


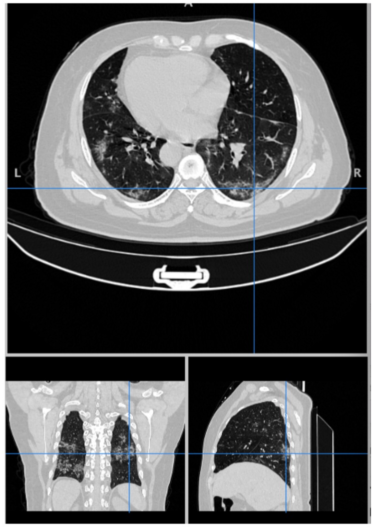
COVID-19 applications

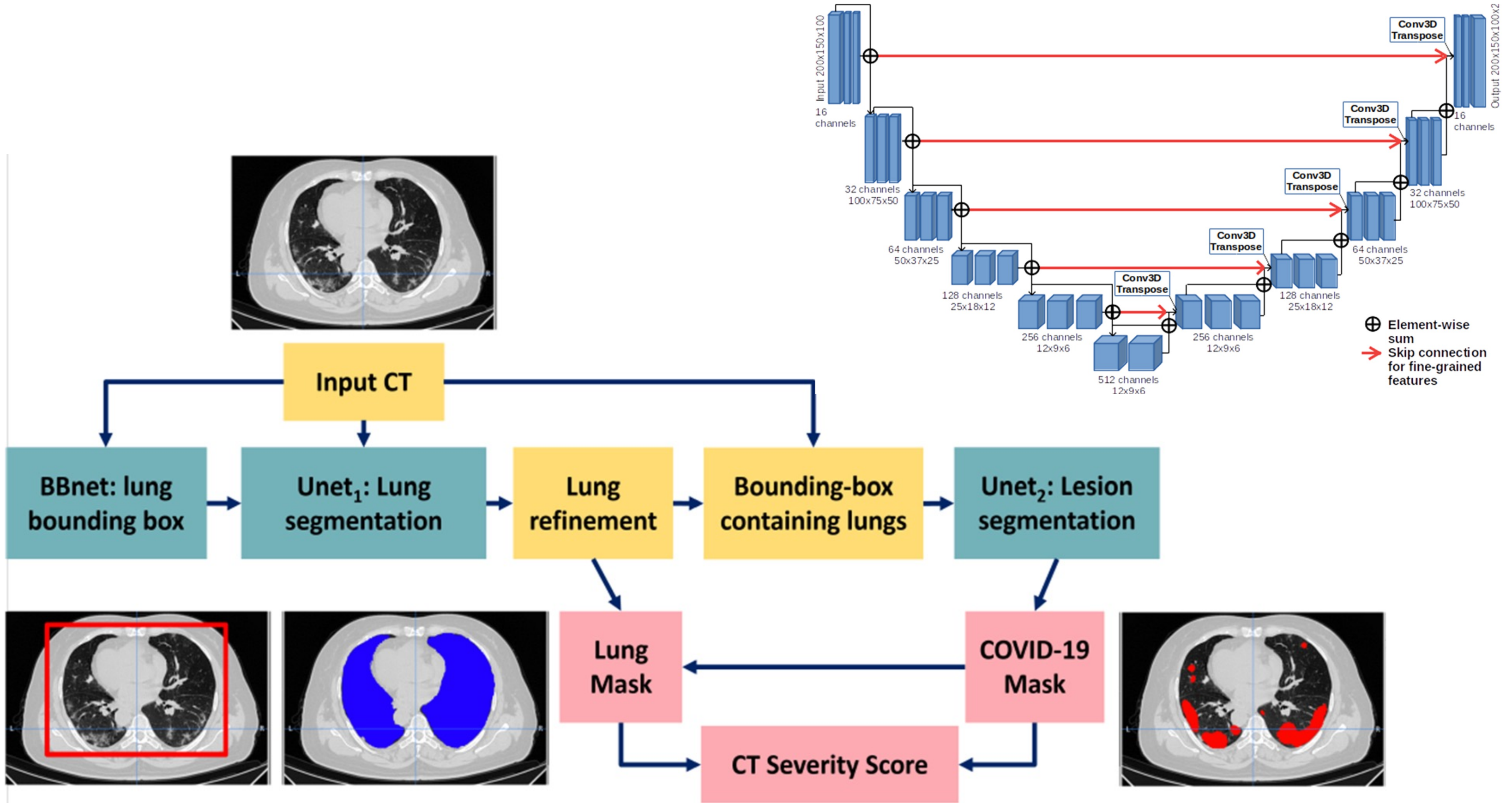
LungQuant

CT
(Computed Tomography)

We trained both the U-nets for 300 epochs on a NVIDIA V100 GPU, provided by the Computing Center of the INFN Division of Pisa;
We also thank the CINECA Italian computing center and the EOS cluster of the Department of Mathematics "F. Casorati" (Pavia).







LungQuant Open Access Repository:
https://www.openaccessrepository.it/record/76937

[F.Lizzi et al. International Journal of Computer Assisted Radiology and Surgery, 2021.]

COVID-19 applications

LungQuant

Total development folder size: **3,5 TB**

Original dataset: **9 GB** (200 patients from challenge dataset) **+**
12 GB (1000 patients from Mosmed dataset)

Saving pre-processing step-by-step - Data augmentation

Fast access storage !

COVID-19 applications

LungQuant

Total development folder size: **3,5 TB**

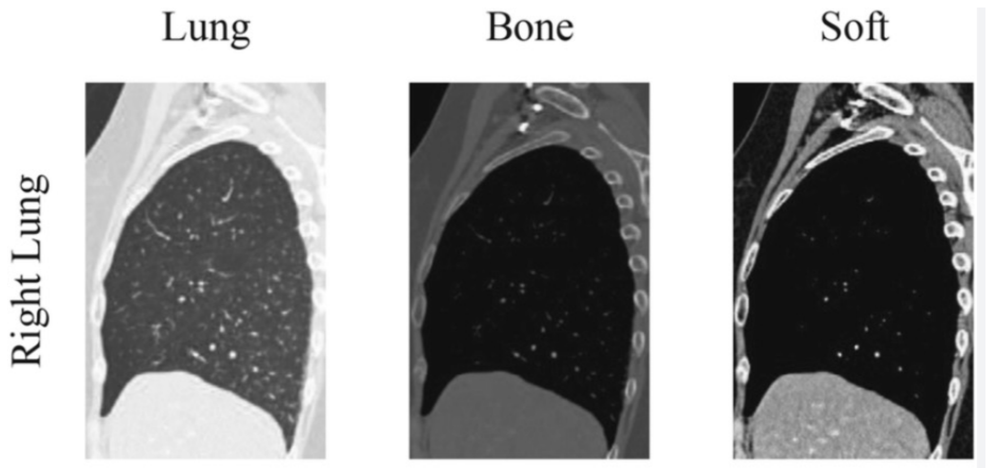
Original dataset: **9 GB** (200 patients from challenge dataset) **+**
12 GB (1000 patients from Mosmed dataset)

Saving pre-processing step-by-step - Data augmentation

Fast access storage !

Private Dataset: **30 GB** (77 patients)

→ More than 1 scan per patient
most of them useless … at the moment!

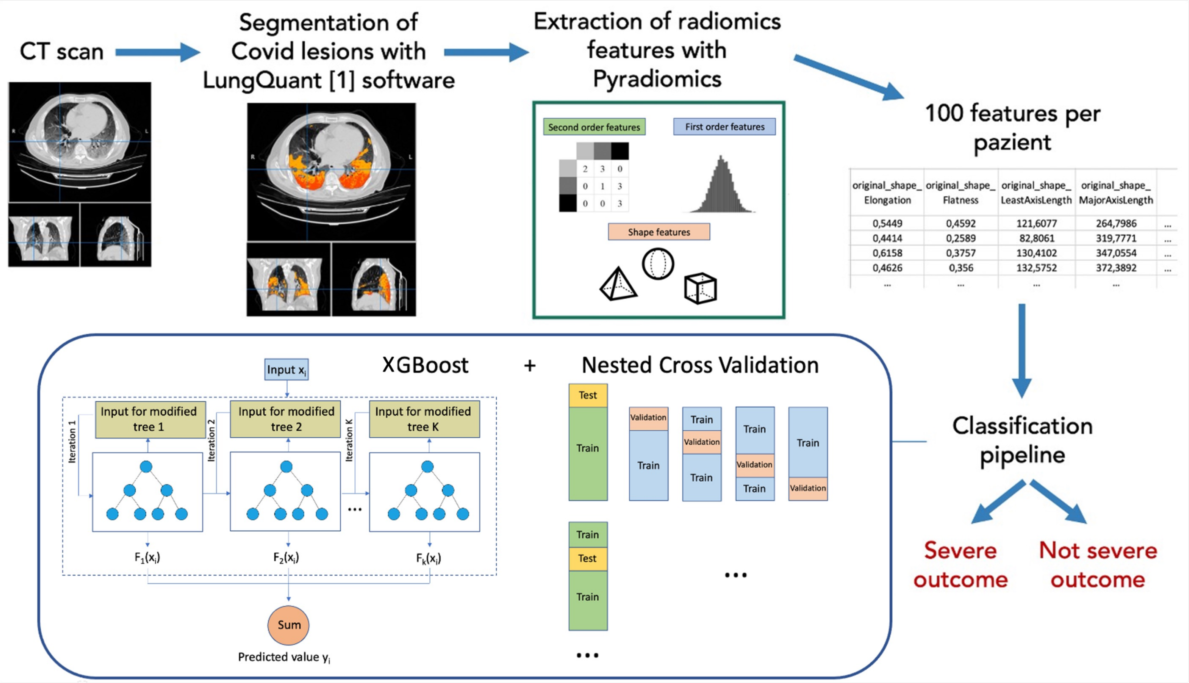


COVID-19 applications

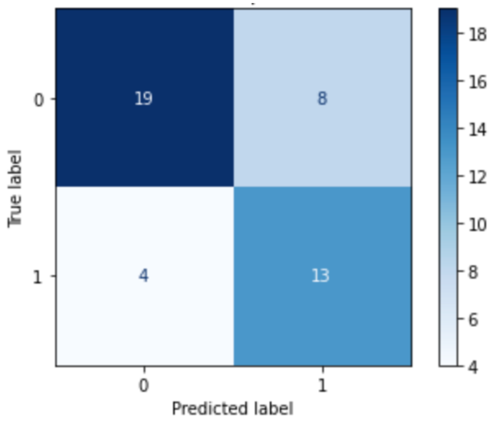
Machine Learning pipeline for severity prediction from radiomics features

ML-INFN / AI-INFN

Acquisition site (Site ID)	Total number of cases	Severe cases	Not-Severe cases
Florence (FI)	100	50	50
Milan (MI)	160	62	98
Palermo (PA)	78	30	48
Pavia (PV)	25	7	18
Pisa (PI)	69	24	45



roc_auc	accuracy	precision	recall
0.85	0.80	0.70	0.82



Analysis of significant features

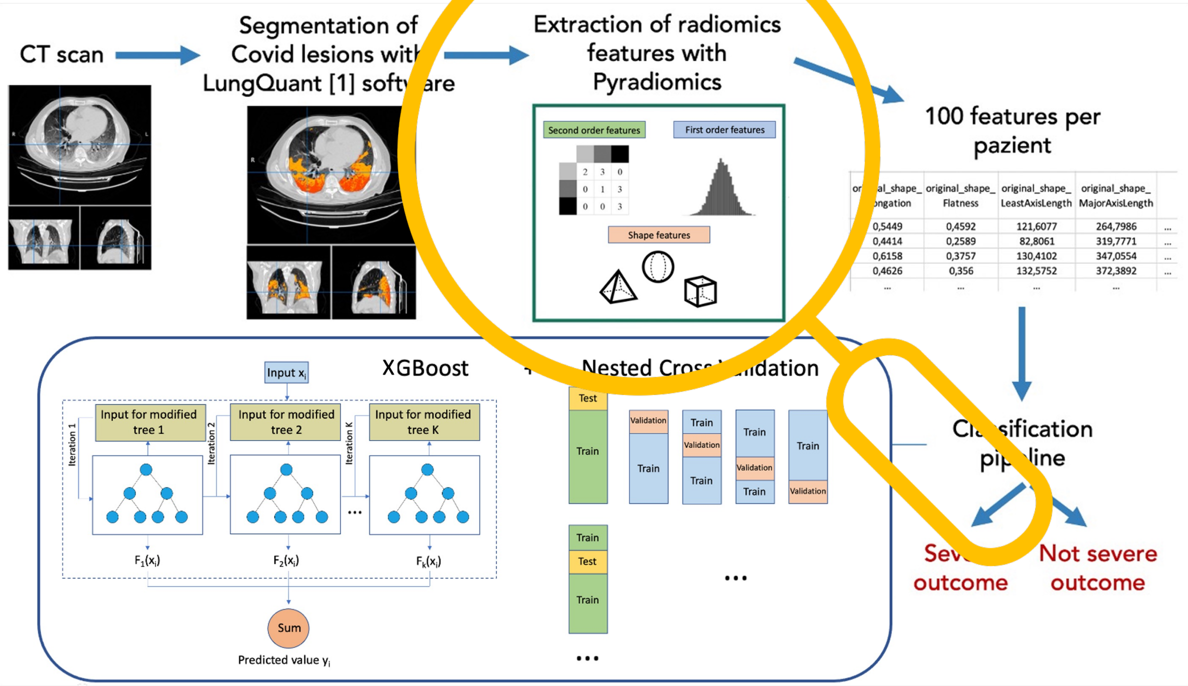
- mRMR
- Feature Importance
- Mutual Information

COVID-19 applications

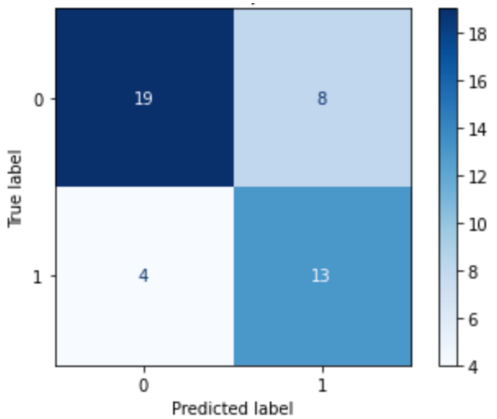
Machine Learning pipeline for severity prediction from radiomics features

ML-INFN / AI-INFN

Acquisition site (Site ID)	Total number of cases	Severe cases	Not-Severe cases
Florence (FI)	100	50	50
Milan (MI)	160	62	98
Palermo (PA)	78	30	48
Pavia (PV)	25	7	18
Pisa (PI)	69	24	45



roc_auc	accuracy	precision	recall
0.85	0.80	0.70	0.82

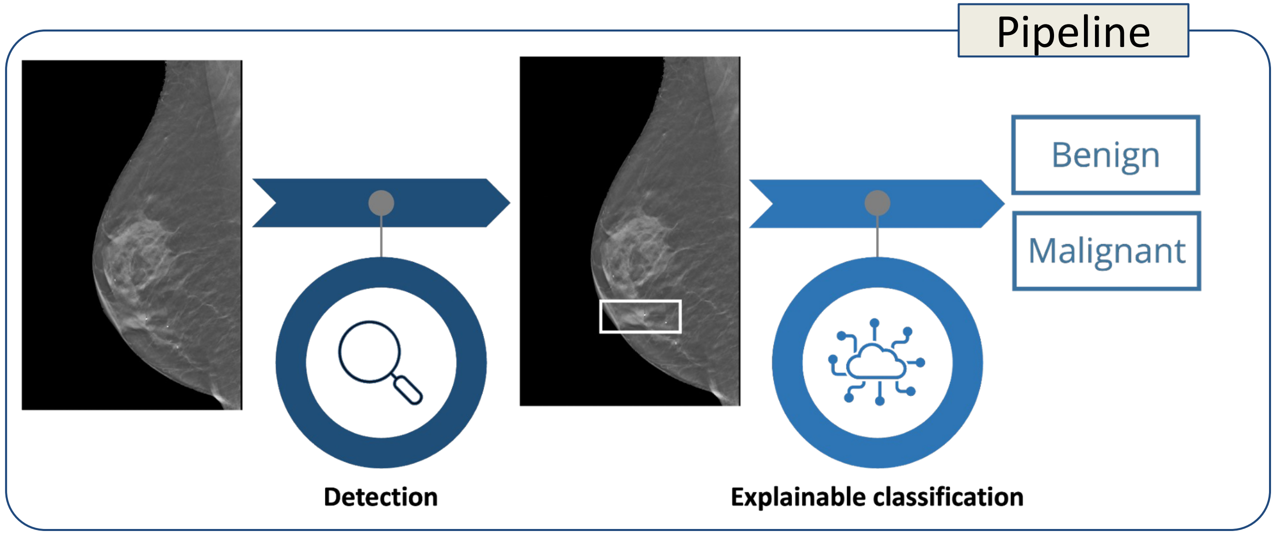


Analysis of significant features

- mRMR
- Feature Importance
- Mutual Information

Overview of study cases

ProtoPNet on breast imaging




On the Applicability of Prototypical Part Learning in Medical Images: Breast Masses Classification Using ProtoPNet

Gianluca Carloni[1,2], Andrea Berti[1,2(✉)], Chiara Iacconi[3], Maria Antonietta Pascali[1], and Sara Colantonio[1]

See the work by Andrea Berti in the poster session

Detection:

- State-of-the-art DL architectures:
  - *YOLOv5* vs *YOLOv8*
- Challenging task
- Recall up to 0.79 on the test set

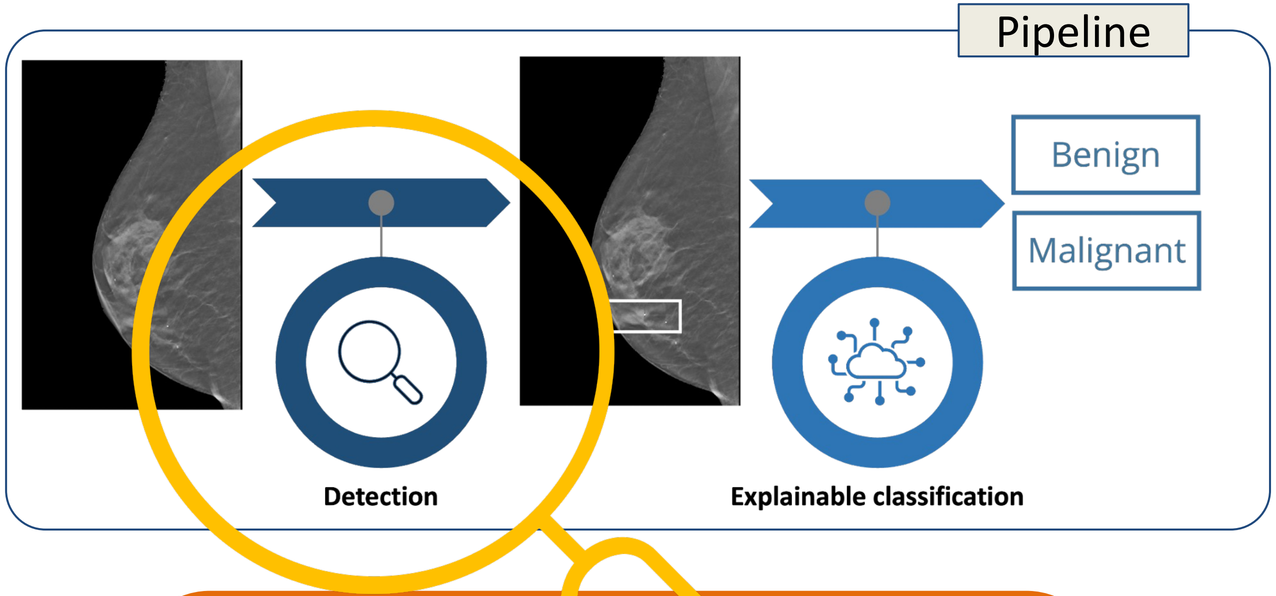
Explainable Classification:

- Classification of benign vs malignant mass crops
- Explainable NN vs Standard CNN:
  - *ProtoPNet* vs *ResNet18*
- Comparable performance: Accuracy 80% c.a.

Training on a Nvidia Ampere A100 GPU 64 GB, 512 GB RAM, provided by Cineca Leonardo.
Training on Nvidia V100 provided by the Computing center of INFN-Pisa.
Training on Nvidia A100 GPU 40 GB, provided by the AI@Edge cluster of ISTI-CNR.

Overview of study cases

ProtoPNet on breast imaging



Training on a Nvidia Ampere A100 GPU 64 GB, 512 GB RAM, provided by Cineca Leonardo.
Training on Nvidia V100 provided by the Computing center of INFN-Pisa.
Training on Nvidia A100 GPU 40 GB, provided by the AI@Edge cluster of ISTI-CNR.

On the Applicability of Prototypical Part Learning in Medical Images: Breast Masses Classification Using ProtoPNet

Gianluca Carloni[1,2], Andrea Berti[1,2(✉)], Chiara Iacconi[3], Maria Antonietta Pascali[1], and Sara Colantonio[1]

Detection:

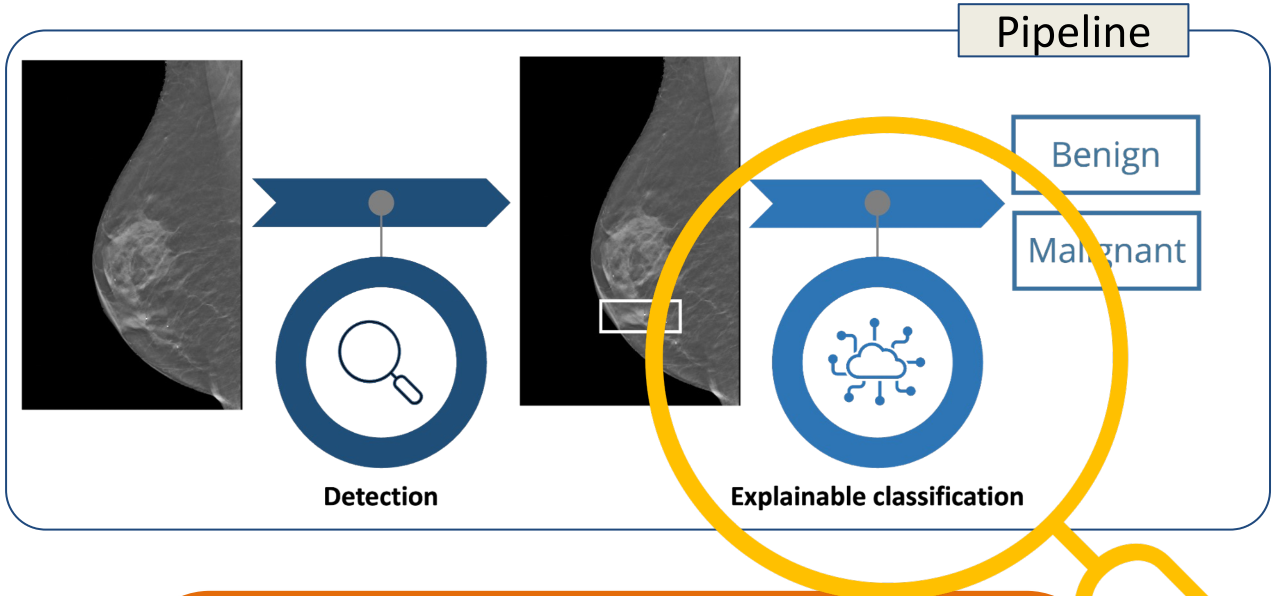
- State-of-the-art DL architectures:
  - *YOLOv5* vs *YOLOv8*
- Challenging task
- Recall up to 0.79 on the test set

Explainable Classification:

- Classification of benign vs malignant mass crops
- Explainable NN vs Standard CNN:
  - *ProtoPNet* vs *ResNet18*
- Comparable performance: Accuracy 80% c.a.

Overview of study cases

ProtoPNet on breast imaging



Training on a Nvidia Ampere A100 GPU 64 GB, 512 GB RAM, provided by Cineca Leonardo.
Training on Nvidia V100 provided by the Computing center of INFN-Pisa.
Training on Nvidia A100 GPU 40 GB, provided by the AI@Edge cluster of ISTI-CNR.

On the Applicability of Prototypical Part Learning in Medical Images: Breast Masses Classification Using ProtoPNet

Gianluca Carloni[1,2], Andrea Berti[1,2(✉)], Chiara Iacconi[3], Maria Antonietta Pascali[1], and Sara Colantonio[1]

See the work by Andrea Berti in the poster session

Detection:

- State-of-the-art DL architectures:
  - *YOLOv5* vs *YOLOv8*
- Challenging task
- Recall up to 0.79 on the test set

Explainable Classification:

- Classification of benign vs malignant mass crops
- Explainable NN vs Standard CNN:
  - *ProtoPNet* vs *ResNet18*
- Comparable performance: Accuracy 80% c.a.

Synergies with PNRR and other projects on AI in medicine

PNRR - CN1: ICSC Centro Nazionale HPC

- Partner in Spoke 8 - WP5 - *Development of clinical machine learning algorithms for EHRs and omics data (including radiomics)*

PNRR PE1 Partenariato Esteso su AI: FAIR - Future AI Research

- Coinvolgimento in Spoke 8 - *Pervasive AI*

PNRR THE Tuscany Health Ecosystem

- Partner in Spoke 1 - *Advanced radiotherapies and diagnostics in oncology*

Regione Toscana Bando Assegni di Ricerca

Progetto *FAIR-AIM*

Piano Operativo Salute (POS) **Traiettoria 2 - eHealth, diagnostica avanzata, medical device e mini invasività**

Titolo: *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 (TELE-NEURART)*

PNRR - InTrEPID

- *In vivo 3D dosimetry in radiotherapy Treatments with EPID*

Overview of study cases

NLP-based approach to convert free-text radiological reports into structured reports

Francesca Lizzi, Sara Saponaro, Leonardo Ubaldi, Irene Minetti, Sandro Ubbiali, Giovanni Ferrando, Marco Fruscione, "*Natural Language Processing Transformer-based system for the translation of COVID-19 CT Free-Text Reports into Structured Radiological Reports: the role of data quality*", EuSoMII 2023

Training on Nvidia V100(32GB VRAM) provided by the Computing center of INFN-Pisa

Aim: Train and test a multilingual pre-trained **Text-to-Text Transformer (mT5 algorithm)** for the translation of free text report into structured radiological reports.

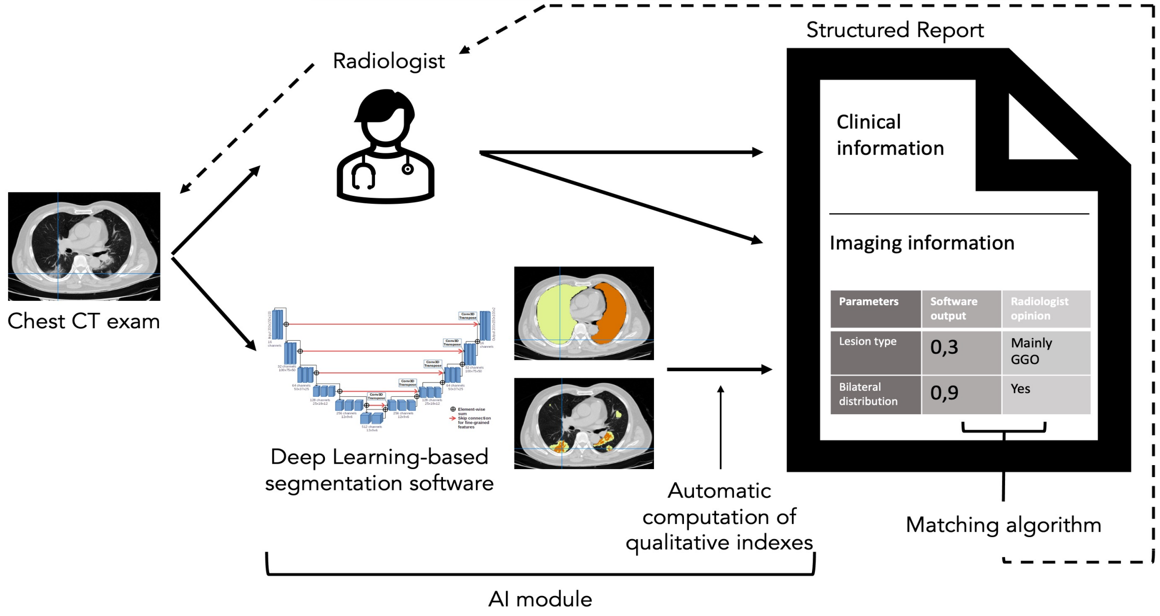
Material & Methods:

The model was trained using three different datasets:

- COVID-19 CT scan reports from Azienda Ospedaliera Universitaria Pisa: **163** training cases, 20 validation cases, 20 test cases.
- COVID-19 CT scan reports from USL3: **123** training cases, 16 validation cases, 20 test cases
- Reports of Echocardiography: **110** training cases, 20 validation cases, 20 test cases

Results:

	Mean Accuracy	Balanced accuracy	F1 score
Covid Pisa	0.97 ± 0.08	0.96 ± 0.09	0.97 ± 0.09
Covid USL3	0.79 ± 0.17	0.58 ± 0.22	0.75 ± 0.18
Echocardiography	0.72 ± 0.26	0.48 ± 0.21	0.67 ± 0.28



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." (2023). Proceedings.

Overview of study cases

NLP-based approach to convert free-text radiological reports into structured reports

Francesca Lizzi, Sara Saponaro, Leonardo Ubaldi, Irene Minetti, Sandro Ubbiali, Giovanni Ferrando, Marco Fruscione, "*Natural Language Processing Transformer-based system for the translation of COVID-19 CT Free-Text Reports into Structured Radiological Reports: the role of data quality*", EuSoMII 2023

Aim: Train and test a multilingual pre-trained **Text-to-Text Transformer (mT5 algorithm)** for the translation of free text report into structured radiological reports.

13 Billions of parameters – 13GB

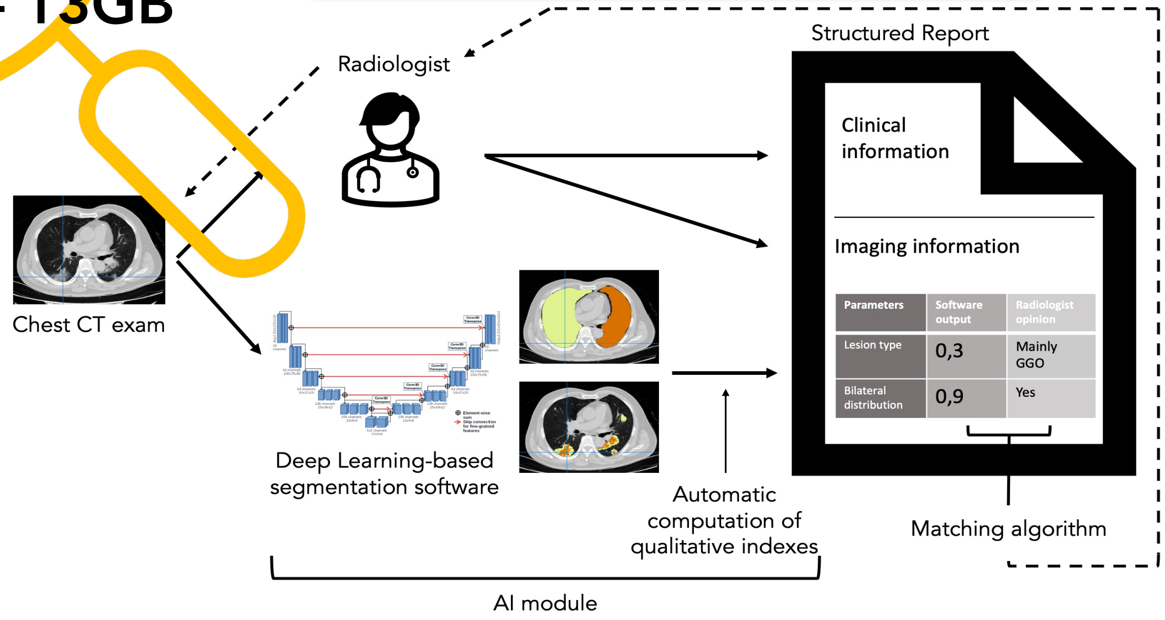
Material & Methods:
The model was trained using three different datasets:

- COVID-19 CT scan reports from Azienda Ospedaliera Universitaria Pisa: **163** training cases, 20 validation cases, 20 test cases.
- COVID-19 CT scan reports from USL3: **123** training cases, 16 validation cases, 20 test cases
- Reports of Echocardiography: **110** training cases, 20 validation cases, 20 test cases

Results:

	Mean Accuracy	Balanced accuracy	F1 score
Covid Pisa	0.97 ± 0.08	0.96 ± 0.09	0.97 ± 0.09
Covid USL3	0.79 ± 0.17	0.58 ± 0.22	0.75 ± 0.18
Echocardiography	0.72 ± 0.26	0.48 ± 0.21	0.67 ± 0.28

Training on Nvidia V100(32GB VRAM) provided by the Computing center of INFN-Pisa



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." (2023). Proceedings.

Overview of study cases

NLP-based approach to convert free-text radiological reports into structured reports

Francesca Lizzi, Sara Saponaro, Leonardo Ubaldi, Irene Minetti, Sandro Ubbiali, Giovanni Ferrando, Marco Fruscione, "*Natural Language Processing Transformer-based system for the translation of COVID-19 CT Free-Text Reports into Structured Radiological Reports: the role of data quality*", EuSoMII 2023

Aim: Train and test a multilingual pre-trained **Text-to-Text Transformer (mT5 algorithm)** for the translation of free text report into structured radiological reports.

13 Billions of parameters – 13GB

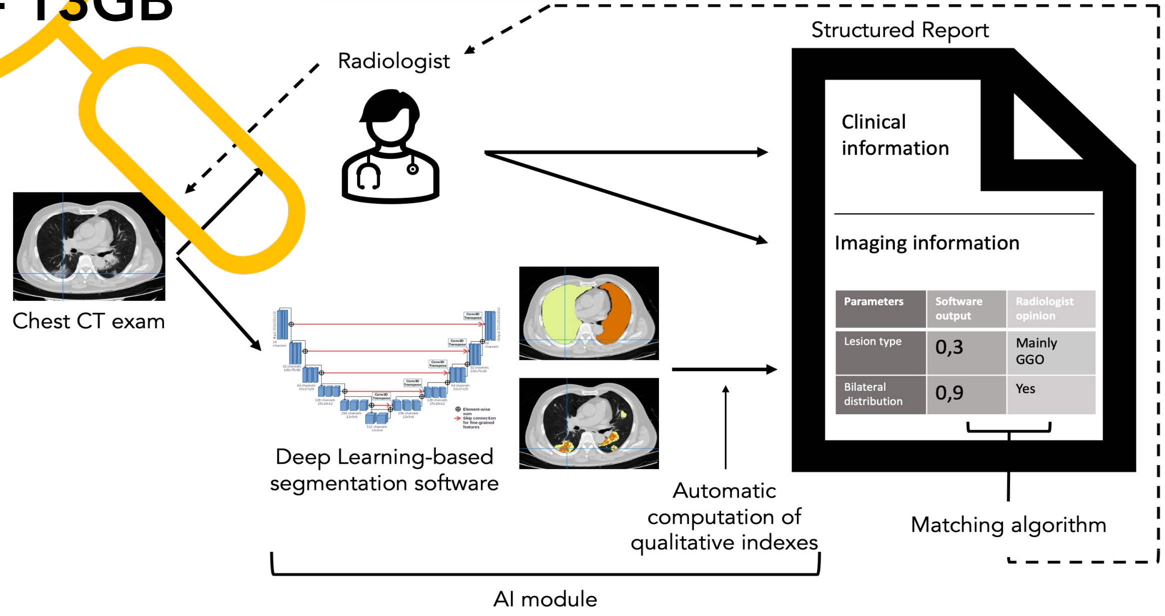
Material & Methods:
The model was trained using three different datasets:

- COVID-19 CT scan reports from Azienda Ospedaliera Universitaria Pisa: **163** training cases, 20 validation cases, 20 test cases.
- COVID-19 CT scan reports from USL3: **123** training cases, 16 validation cases, 20 test cases
- Reports of Echocardiography: **110** training cases, 20 validation cases, 20 test cases

Results:

	Mean Accuracy	Balanced accuracy	F1 score
Covid Pisa	0.97 ± 0.08	0.96 ± 0.09	0.97 ± 0.09
Covid USL3	0.79 ± 0.17	0.58 ± 0.22	0.75 ± 0.18
Echocardiography	0.72 ± 0.26	0.48 ± 0.21	0.67 ± 0.28

Training on Nvidia V100(32GB VRAM) provided by the Computing center of INFN-Pisa



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." (2023). Proceedings.

Computing challenges

- Healthcare data sets often consist of large, multi-dimensional modalities. Deep Learning (DL) models require both high accuracy and high confidence levels to be useful in clinical practice. Researchers employ advanced hardware and software to speed up both data- and computation-intensive process.

- Moreover, in dealing with medical data acquired on patients, **secure storage** is needed, meaning storage services based on certified security standards (e.g. the information security standard ISO 27001) and GDPR compliant.

Computing challenges

- Healthcare data sets often consist of large, multi-dimensional modalities. Deep Learning (DL) models require both high accuracy and high confidence levels to be useful in clinical practice. Researchers employ advanced hardware and software to speed up both data- and computation-intensive process.

- Moreover, in dealing with medical data acquired on patients, **secure storage** is needed, meaning storage services based on certified security standards (e.g. the information security standard ISO 27001) and GDPR compliant.

Setting up an open-source XNAT IT platform for medical imaging research

See presentation by Antonino Formuso in the Infrastrutture ICT session

Computing challenges

- Healthcare data sets often consist of large, multi-dimensional modalities. Deep Learning (DL) models require both high accuracy and high confidence levels to be useful in clinical practice. Researchers employ advanced hardware and software to speed up both data- and computation-intensive process.

- Moreover, in dealing with medical data acquired on patients, **secure storage** is needed, meaning storage services based on certified security standards (e.g. the information security standard ISO 27001) and GDPR compliant.

Setting up an open-source XNAT IT platform for medical imaging research

See presentation by Antonino Formuso in the Infrastrutture ICT session

Medical image database	AI-based image analysis	Monte-Carlo simulations
↓	↓	↓
HDD/SDD with fast access	GPU with large VRAM	Many CPUs with parallelization

- Considerable **computing resources** (e.g., volumetric data and large models)
- Computational capability, also in terms of **VRAM** (a bottleneck in training the models and finding the best possible model design).

Acknowledgments

INFN groups involved in nextAIM:

Bari - Bologna - Cagliari - Catania - Ferrara - Firenze - Genova - Lab. Naz. Sud - Milano - Napoli - Padova - Pavia - Pisa

Staff of the Data Center @ INFN Division of Pisa.

next AIM

IN SUPREMÆ DIGNITATIS 1343

UNIVERSITÀ DI PISA

INFN

Istituto Nazionale di Fisica Nucleare

FAIR Future Artificial Intelligence Research

ICSC

Centro Nazionale di Ricerca in HPC, Big Data and Quantum Computing

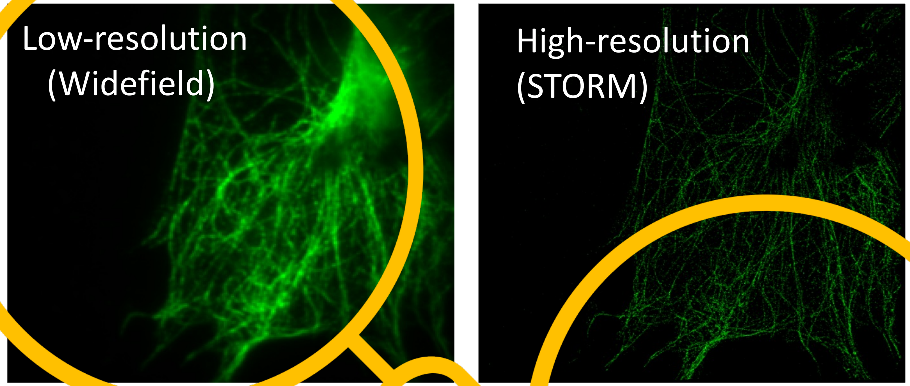
THE Tuscany Health Ecosystem

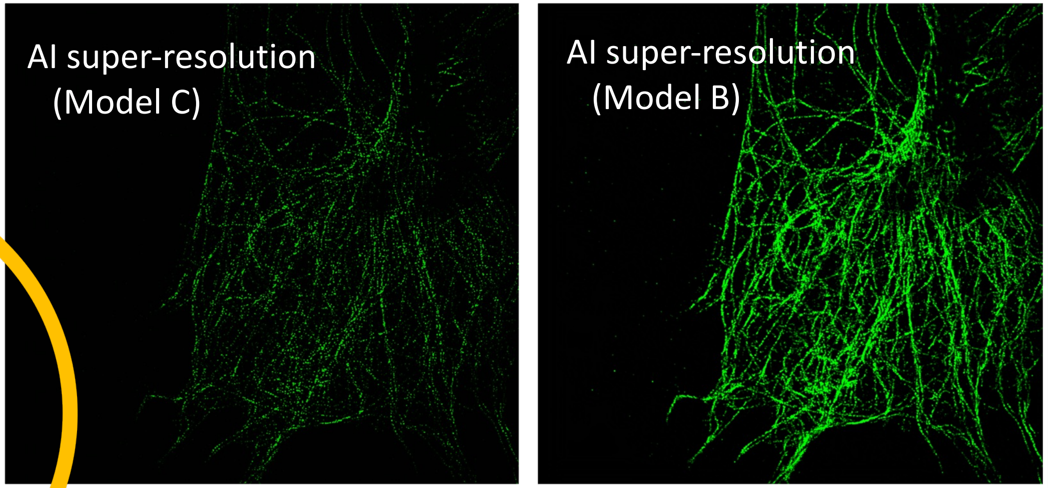
contact: camilla.scapicchio@pi.infn.it

Overview of study cases

A generative DL approach for super-resolution microscopy

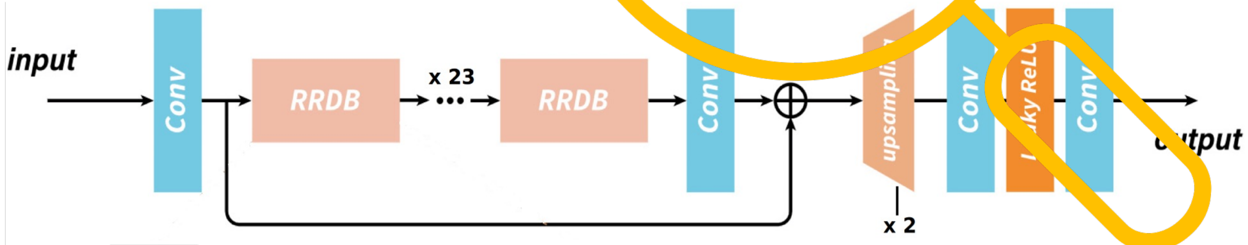
THE Tuscany Health Ecosystem





See the work by Simone Lossano in the poster session

Enhanced Super-Resolution GAN (ESRGAN)

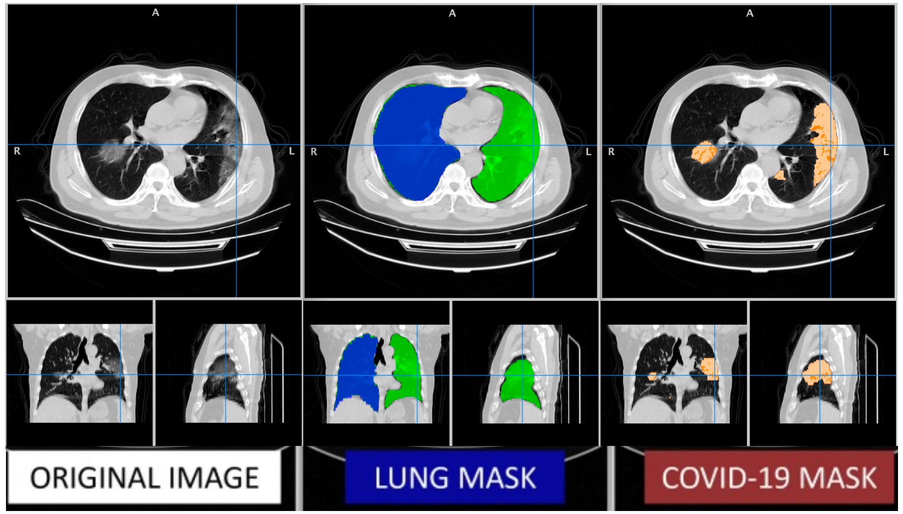


Model	Test set	PSNR	SSIM	Model	Test set	PSNR	SSIM
A	BioSR	26 ±2	0.92 ±0.03	B	α-Tub-1	23 ±3	0.8 ±0.2
B	α-Tub-2	20 ±1	0.8 ±0.1	B	β-Tub-1	14 ±2	0.4 ±0.2
B	β-Tub-2	16 ±1	0.7 ±0.1	C	α-Tub-2	24 ±1	0.8 ±0.1
C	β-Tub-1	17 ±1	0.7 ±0.2	C	β-Tub-2	19 ±1	0.7 ±0.1

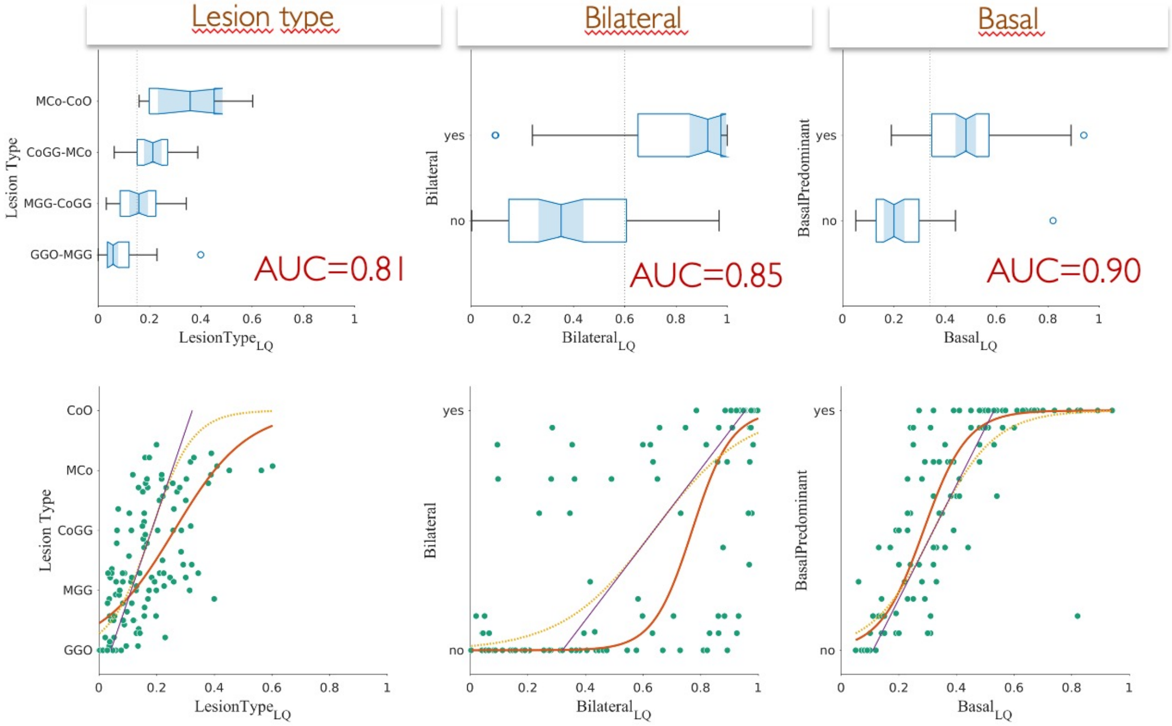
6x 10 cores Intel Xeon E5-2640v4 @2.40 GHz, 1x NVIDIA Tesla V100 with 16/32 Gb VRAM and 64 Gb RAM, provided by the Computing Center of the Pisa Section of INFN

COVID-19 applications

LungQuant – A multicenter evaluation



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



A good correlation between average radiologists' opinions and software output metrics.

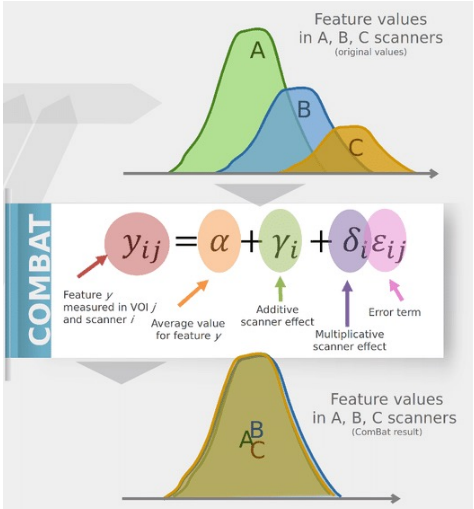
[Scapicchio C., et al. "A multicenter evaluation of a deep learning software (LungQuant) for lung parenchyma characterization in COVID-19 pneumonia." *European Radiology Experimental* (2023). https://doi.org/10.1186/s41747-023-00334-z]

COVID-19 applications

Machine Learning pipeline for severity prediction from radiomics features

Evaluation of the acquisition site effect

	roc_auc	roc_auc
ARTIFICIAL NEURAL NETWORK	0.92 ± 0.04	0.64 ± 0.08
LOGISTIC REGRESSION	0.87 ± 0.03	0.52 ± 0.06



ComBat harmonization

gldm DependenceNonUniformity
gldm GrayLevelNonUniformity
glrlm GrayLevelNonUniformity
glrlm RunLenghtNonUniformity
glszm GrayLevelNonUniformity

The same Non-Uniformity features are significant before and after data harmonization

Overview of study cases

Case-Control classification in ASD on MRI

ABIDE
Autism Brain Imaging Data Exchange

ML-INFN/AI-INFN CENTRO DI CALCOLO PISA

Brain imaging features

•**sMRI –** 221 structural features (morphometric) extracted with the Freesurfer analysis package

•**rs-fMRI –** 5253 functional features for each subject with the CPAC processing pipeline

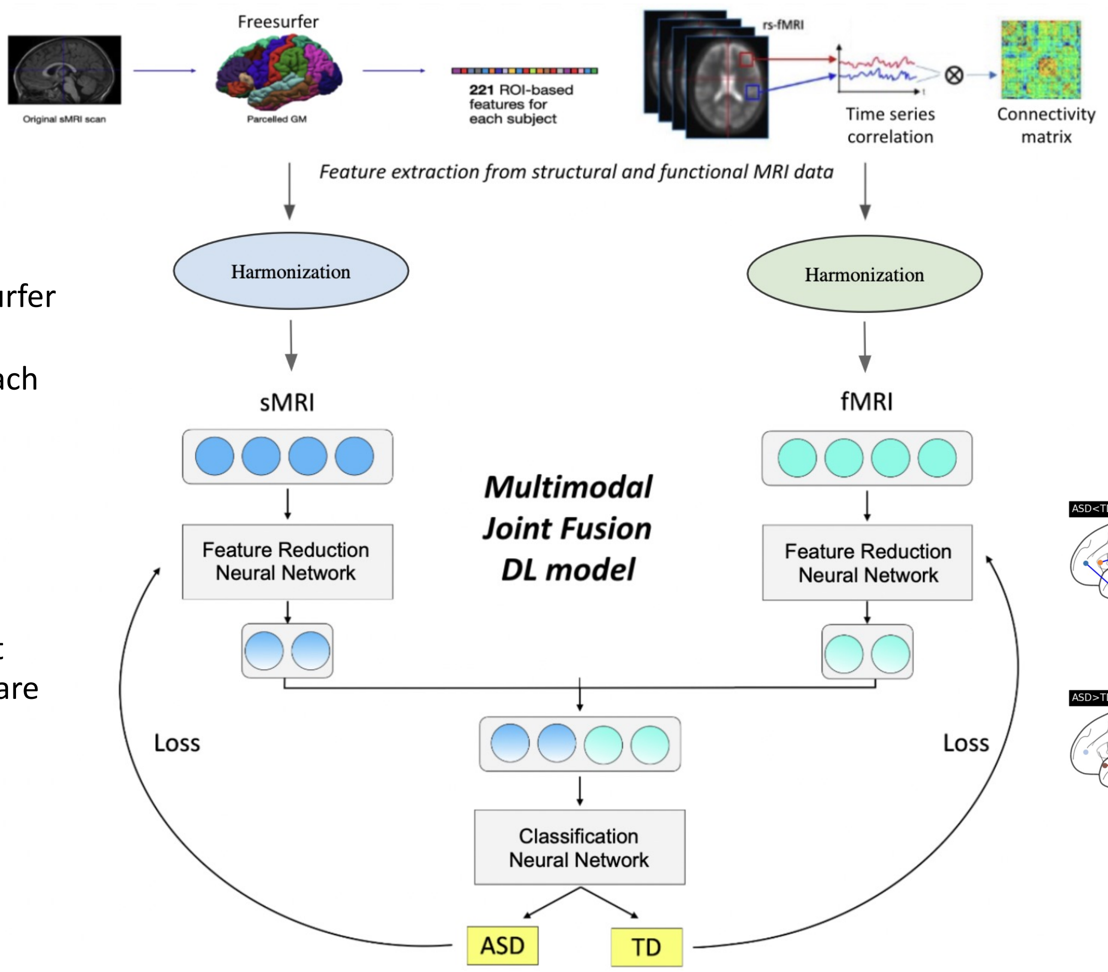
Joint fusion approach:

•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

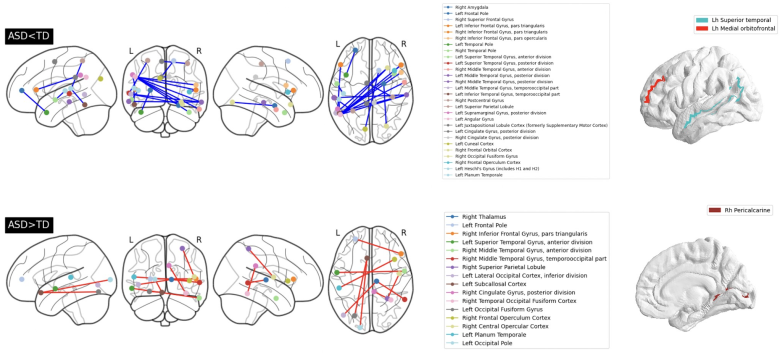
•The model was trained with 150 epochs within a 10-fold cross validation scheme

Explainability framework:

•SHpley Additive exPlanations (**SHAP**)



Brain changes tend to occur in individuals with ASD in regions belonging to the Default Mode Network and the Social Brain.

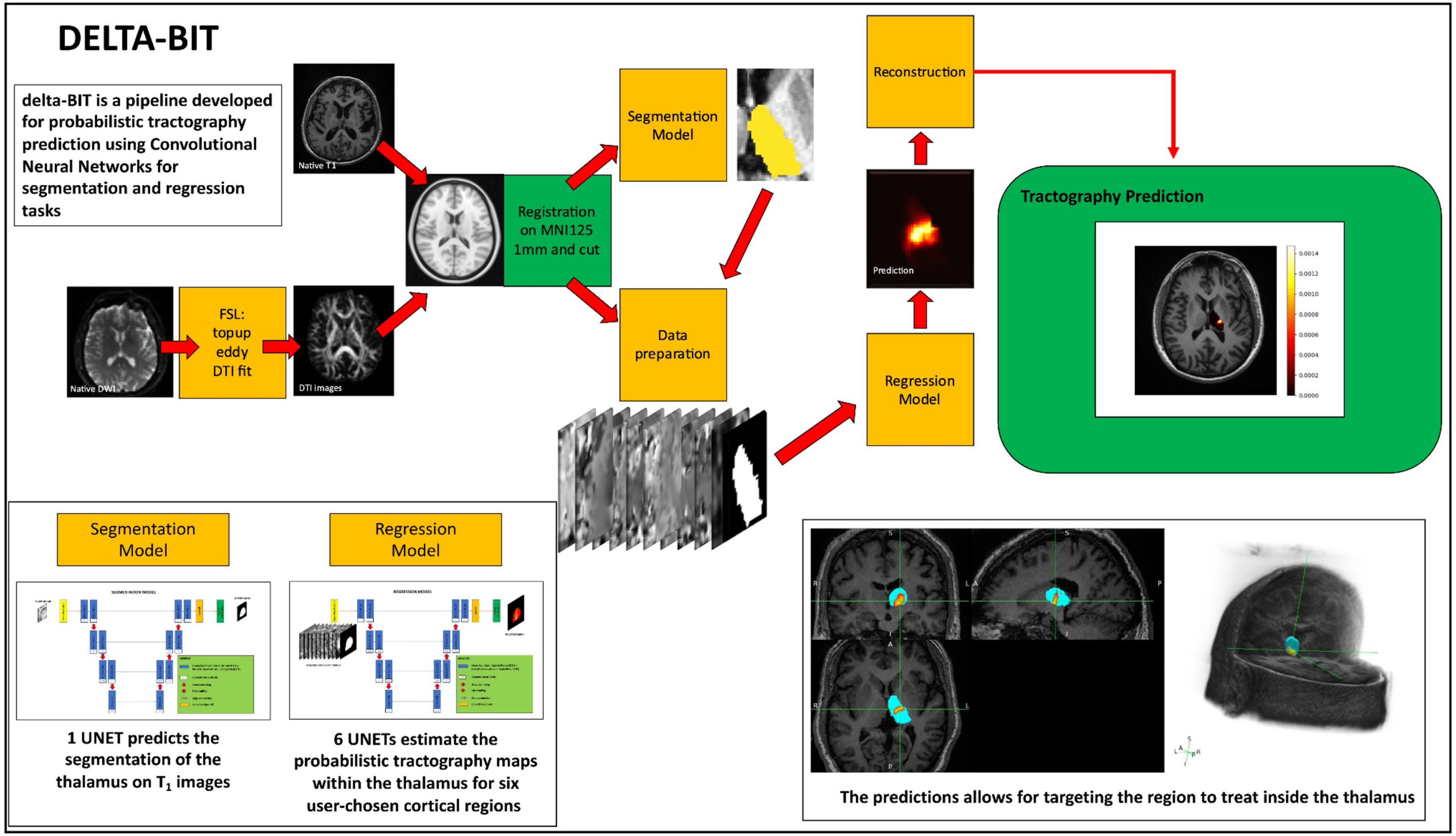


Type of model	AUC	Accuracy
Structural model	0.66 ± 0.05	0.75 ± 0.08
Functional model	0.76 ± 0.04	0.83 ± 0.12
Joint fusion model	0.78 ± 0.04	0.85 ± 0.12

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*, https://doi.org/10.1186/s40708-023-00217-4

Overview of study cases

Deep-learning Local TrActography for BraIn Targeting



Training for 2000 epochs for each tractograpy on a NVIDIA RTX A5000 GPU with 24GB of memory (University of Palermo). It takes about of 50 hours to complete the training of one regression model and 11 to train the segmentation model.

Public Repository:
https://github.com/mromeo1992/delta-BIT

M. Romeo, C. Gagliardo, G. Collura, E. Bruno, M. C. D'Oca, M. Midiri, F. Lizzi, I Postuma, A. Lascialfari, A. Retico, M. Marrale. DeLTA-BIT: an open-source probabilistic tractography-based deep learning framework for thalamic targeting in functional neurological disorders. Submitted to Neuroimage