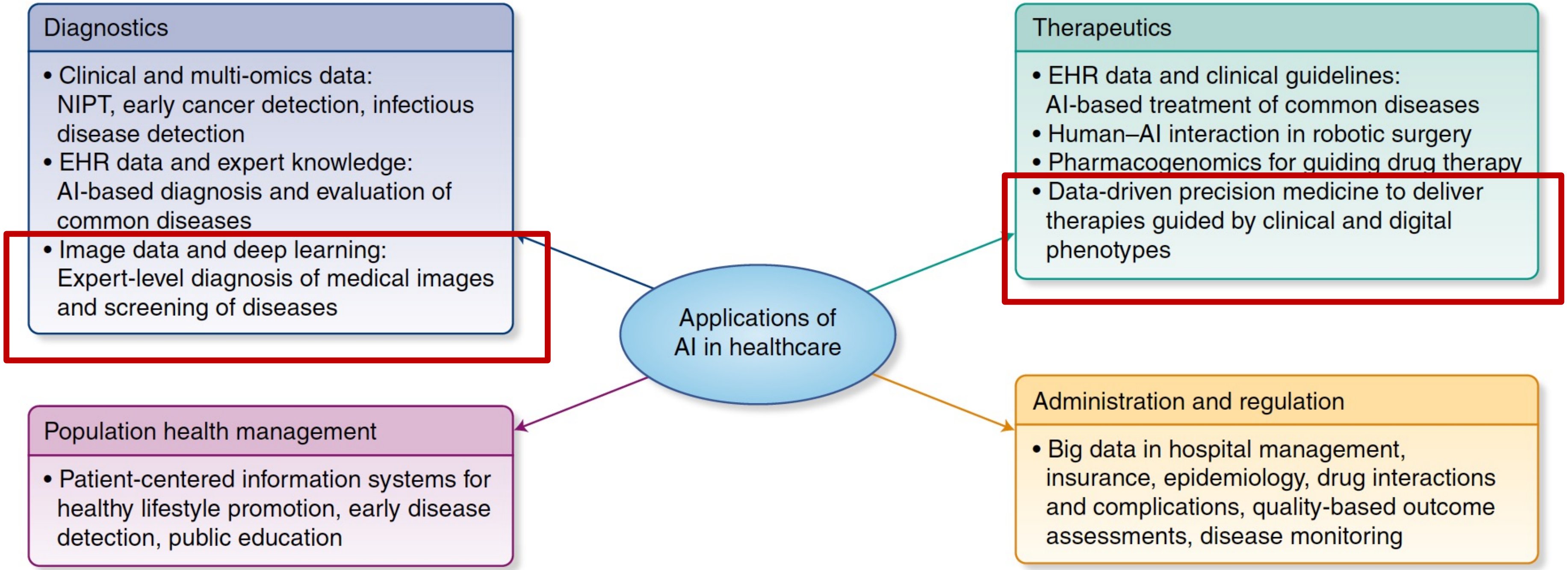
Machine Learning for Applications in Medical Physics

Alessandra Retico
Istituto Nazionale di Fisica Nucleare
Sezione di Pisa

alessandra.retico@pi.infn.it

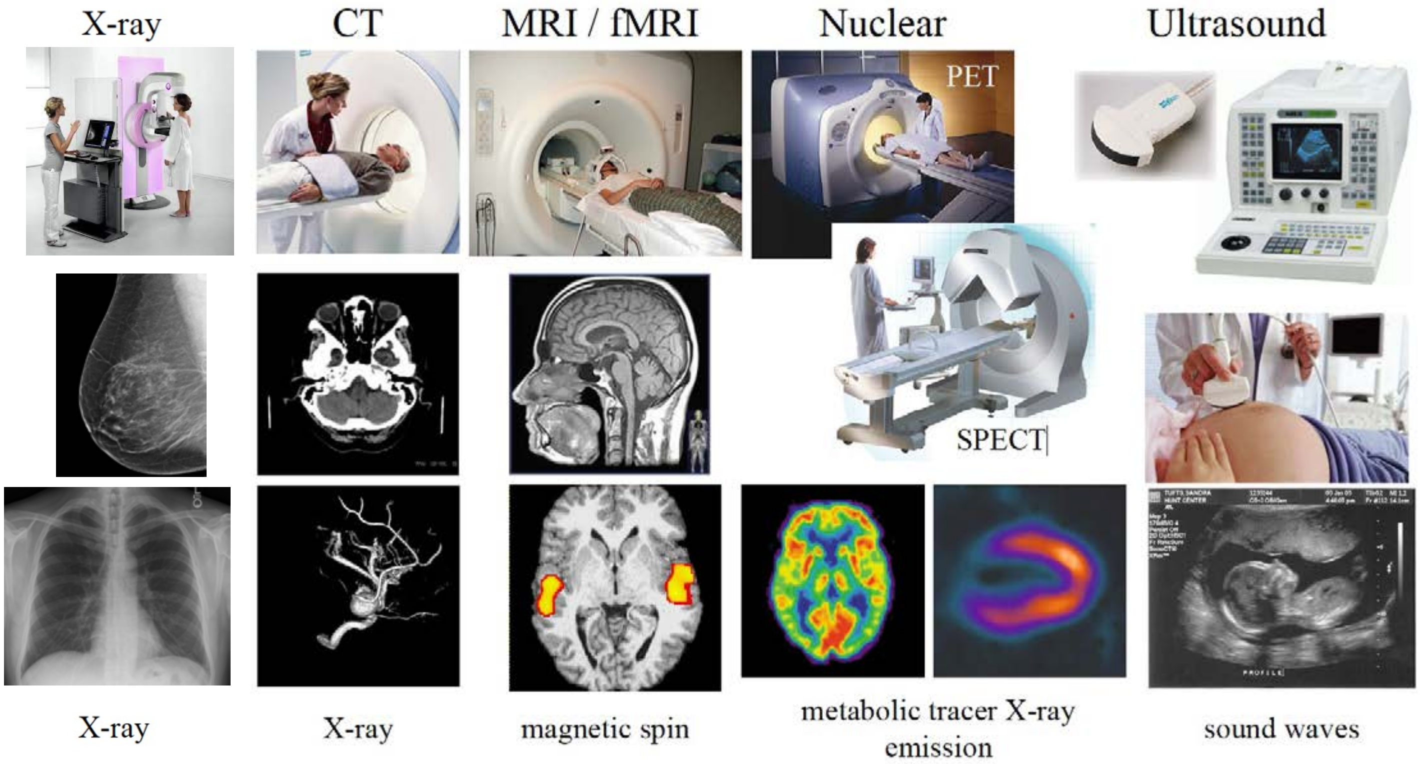
Artificial Intelligence applications in Healthcare



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

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

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



Medical images are more than pictures!!!

Decision Support Systems (DSS) for Detection/Diagnosis

Image processing and analysis techniques can help:

- Improve image visualization
- Detect abnormalities in diagnostic images (lesions, etc.)
- Follow up pathological conditions (growth rate of lesions)
- Assessment of treatment efficacy

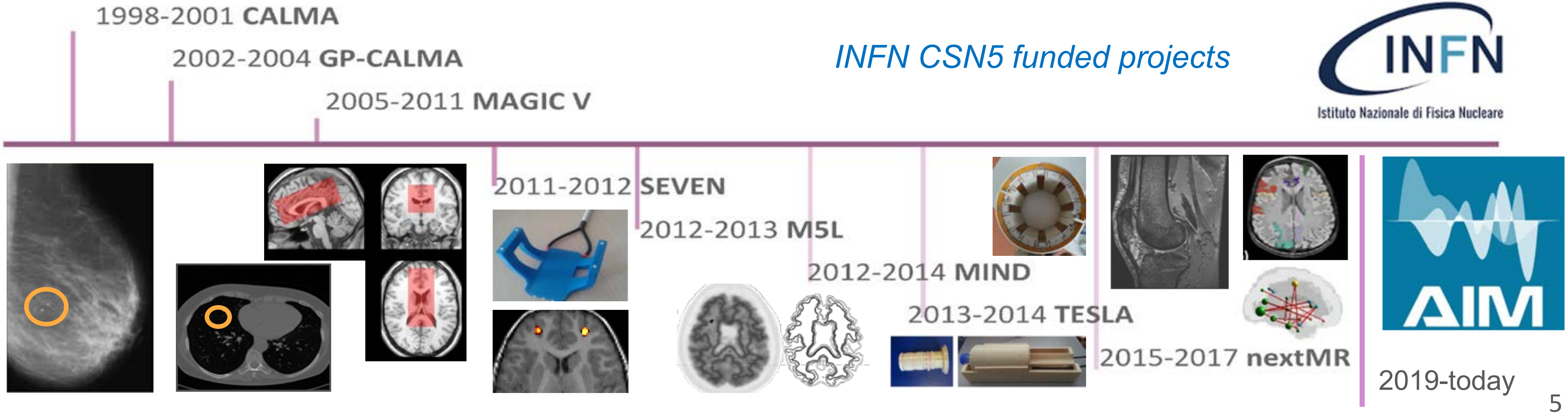
The aim is to assist clinicians in their tasks, not to replace them, through the design and development of:

Computer Aided Detection/Diagnosis (CAD) systems
or Decision Support Systems (DSS)

Historical overview

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

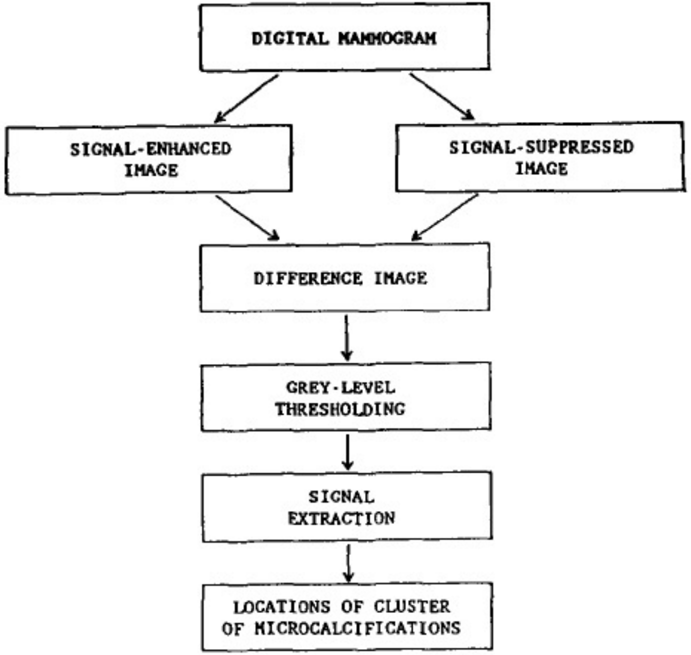
- In the ’90 - Old-fashion systems (rule-based)
  - Since 2000 - Hand-crafted feature and Machine Learning classification (Radiomics and ML)
    - Since 2016 – Deep-Learning image classification

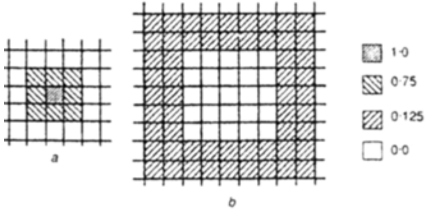


Old-fashion CAD systems

Analysis pipeline for automated detection of microcalcification clusters on mammograms

“Old-fashion” decision-making systems were based on explicitly coding of specific rules



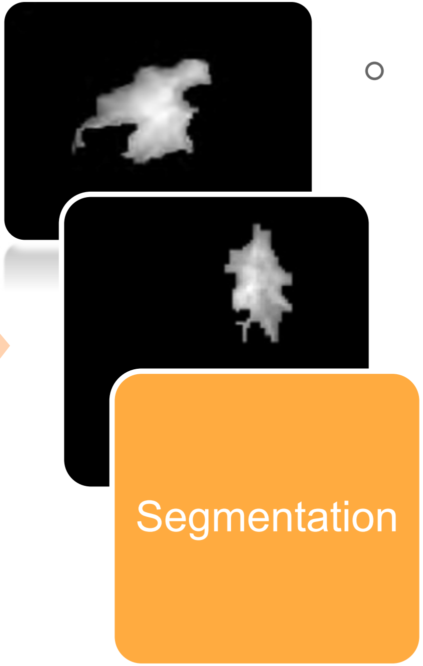


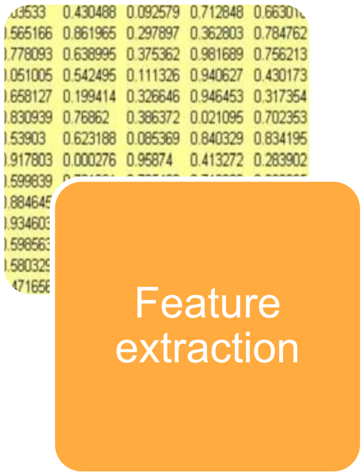
Kernels used for (a) enhancement filter and (b) suppression filter

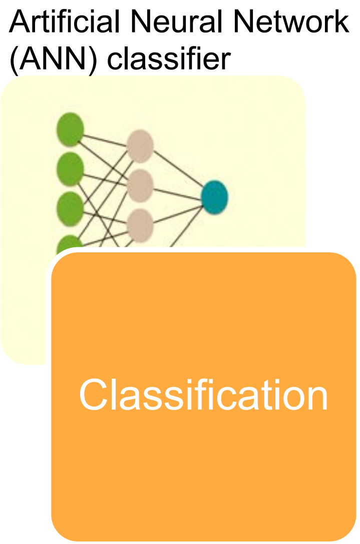
[Chan HP et al., “Image feature analysis and computer-aided diagnosis in digital radiography. 1. Automated detection of microcalcifications in mammography,” Med. Phys. 14, 538–548, 1987]

Hand-crafted feature + Machine Learning classification

- Semiautomated lesion segmentation of mass lesions in mammograms
- Automated classification into benign/malignant category



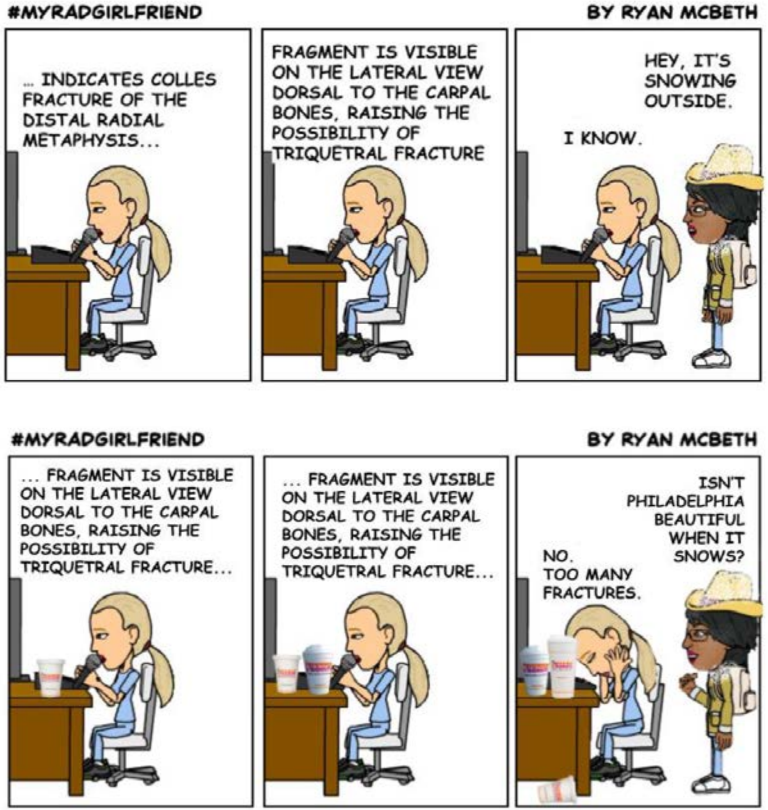




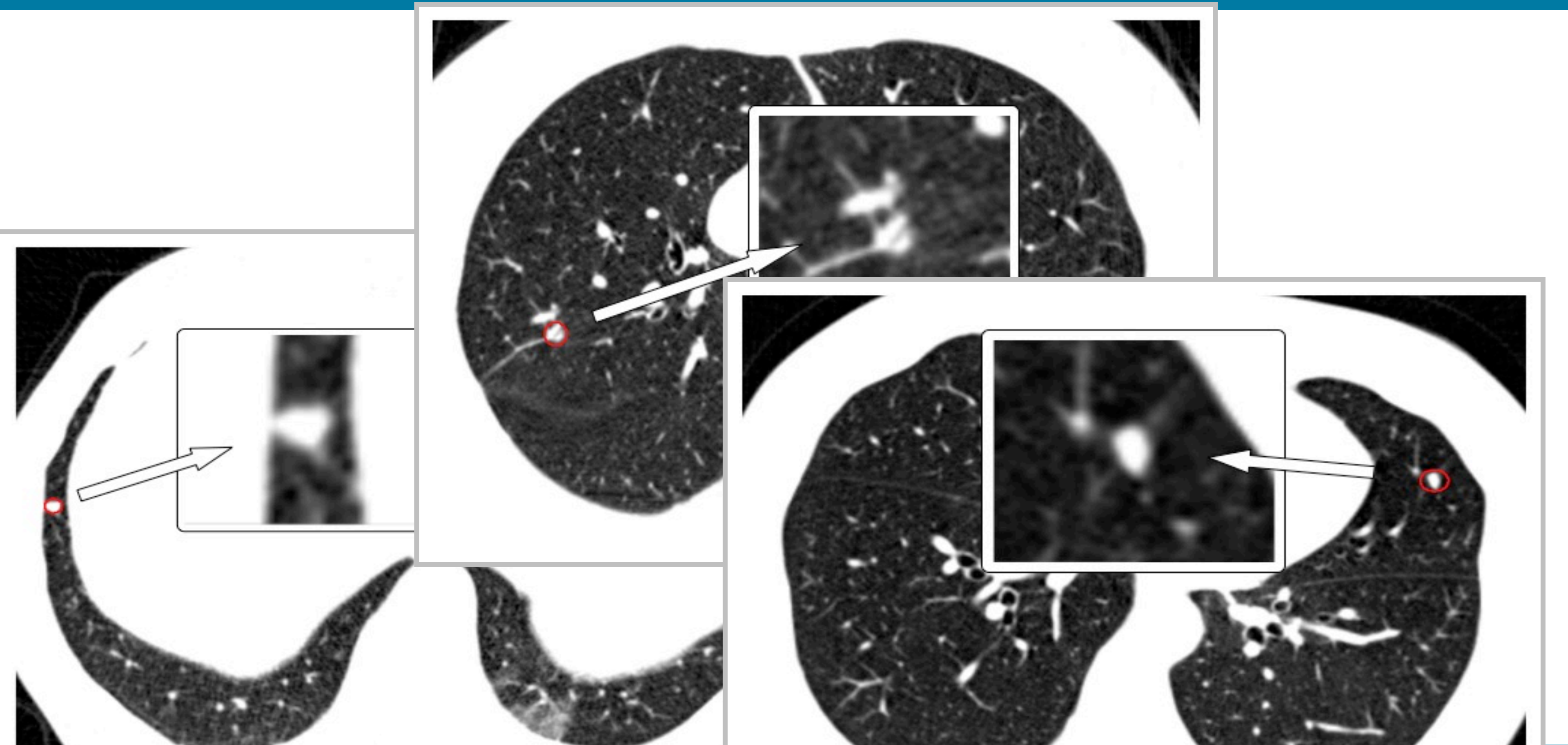
[Delogu P., Fantacci M.E., Kasae P., Retico, A., Computers in Biology and Medicine 37 (2007) 1479 – 1491]

Why using machine-learning techniques?

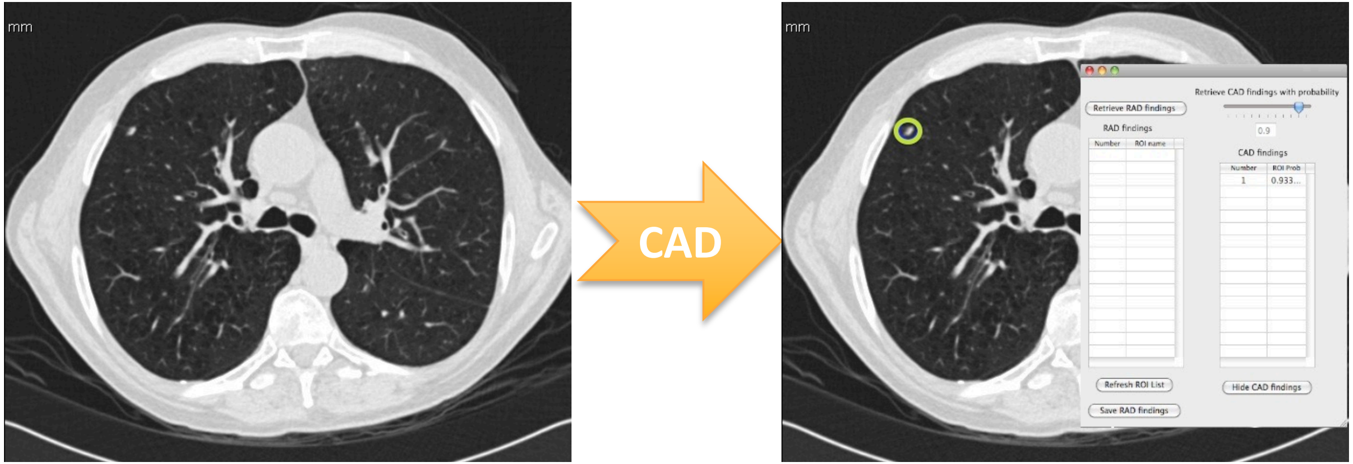
- Some tasks cannot be defined well, except by examples.
- Relationships and correlations can be hidden within large amounts of data.
    - Machine Learning/Data Mining may be able to find these relationships.
- The amount of knowledge available about certain tasks might be too large for explicit encoding by humans (e.g., medical diagnostic).
    - ***Learning from examples***

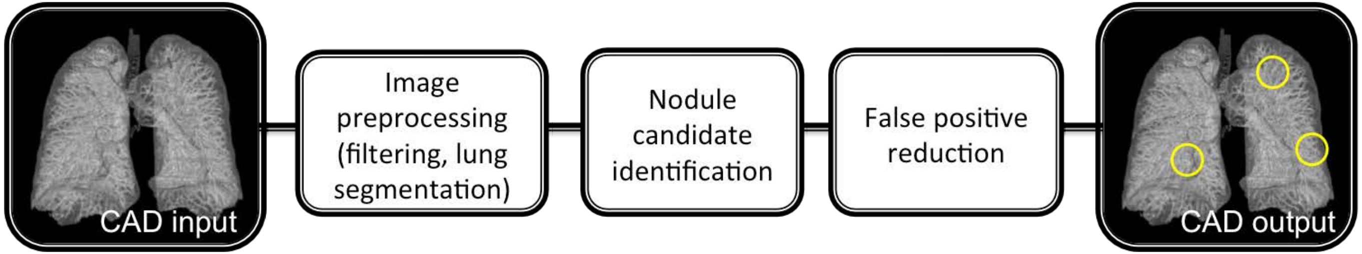


Automated detection of lung nodules in CT images

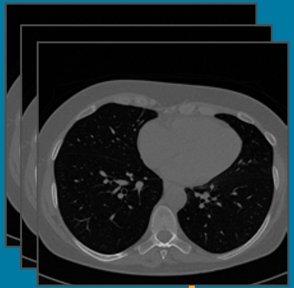


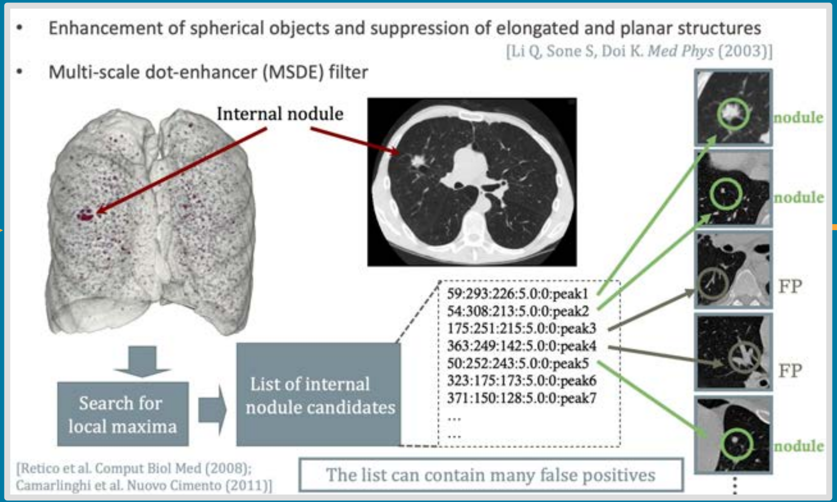
CAD system for lung nodule detection

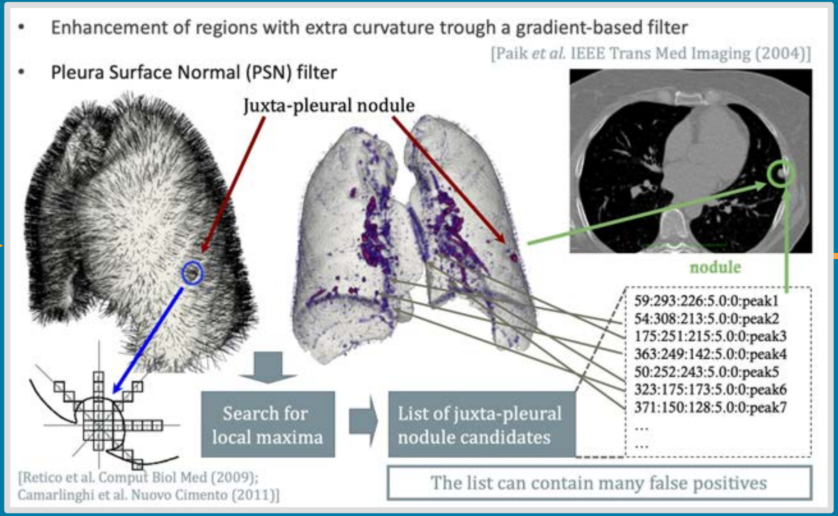


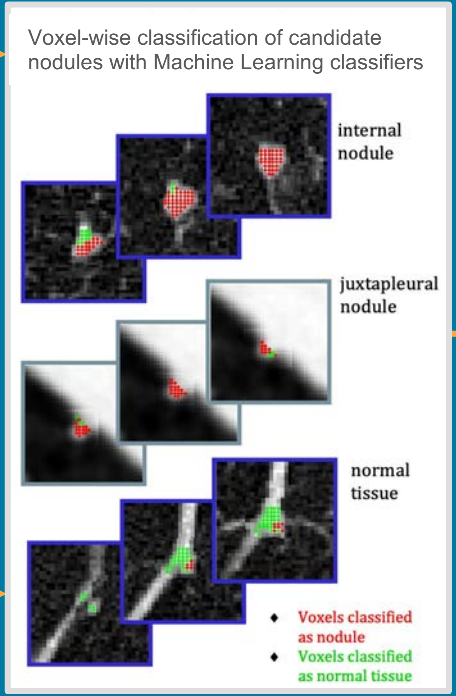


3D input

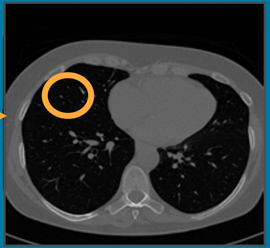








output

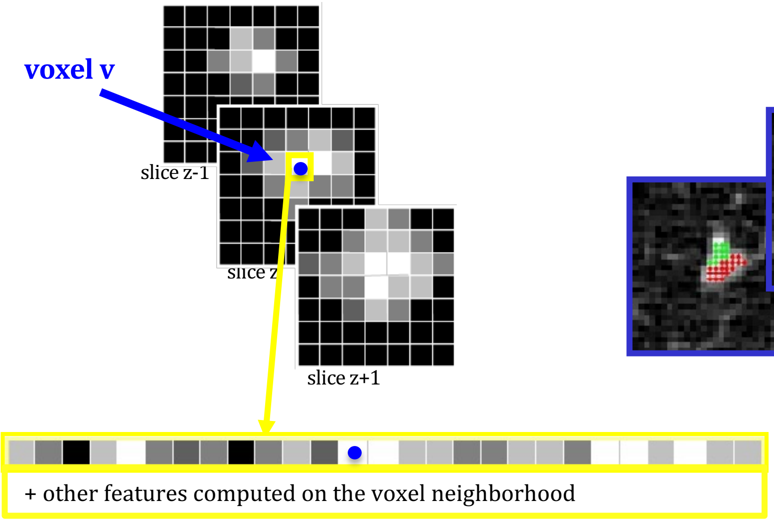


MAGIC-5 and ML5
INFN projects
[2005-2010]

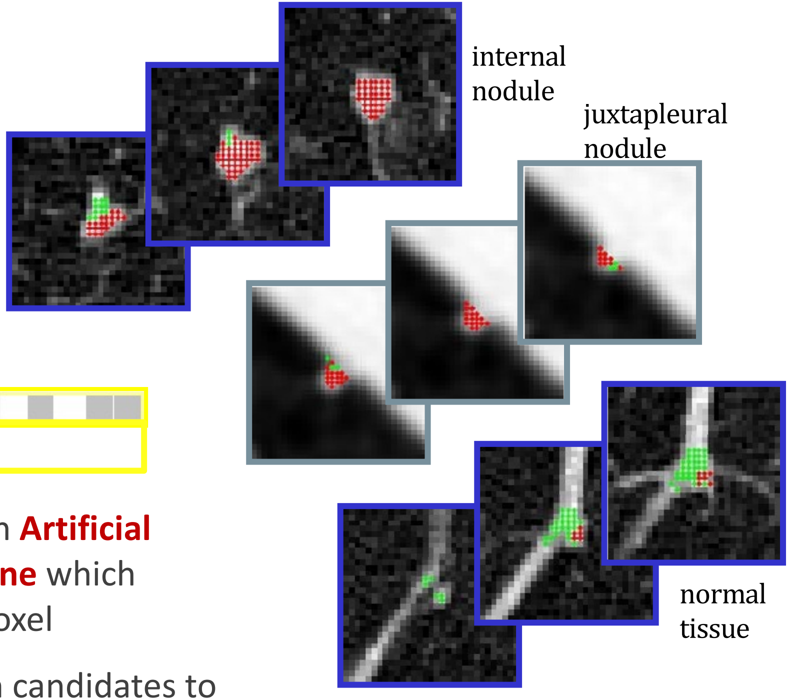
The system was developed in collaboration with:
- Azienda Ospedaliera Universitaria Pisana (AOUP) and the Radiology Dep. of Pisa University
- Bracco Imaging S.p. A.

Example: voxel-based nodule characterization

- Each voxel v of a nodule candidate is described in terms of a vector of features



- Voxels classified as nodule
- Voxels classified as normal tissue



Supervised classification

Each vector of features is analyzed by an **Artificial Neural Network/Support Vector Machine** which assigns the class membership to each voxel

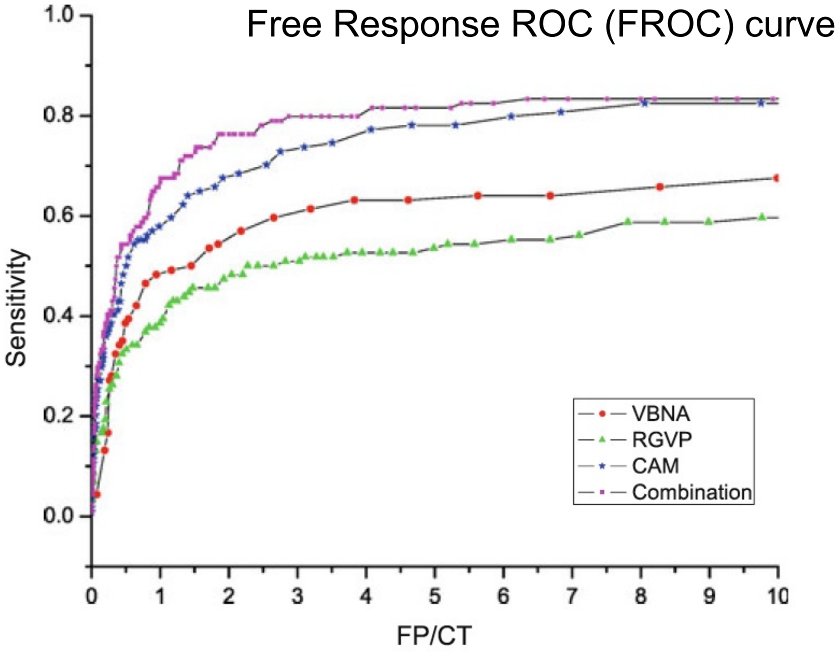
A majority criterion is adopted to assign candidates to either the “nodule” or the “healthy tissue” class

Final CAD performance

Combining different CAD methods increases the performance in the identification of pulmonary nodules

Train + validation sets:
- 69 CTs with 138 nodules (96 internal and 42 juxtapleural)

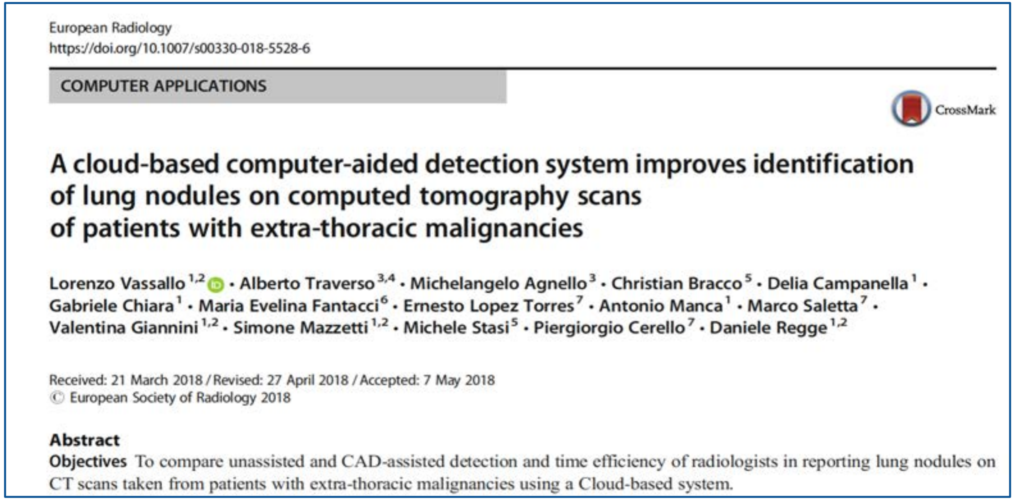
Independent Test set:
- 69 CTs with 114 nodules (95 internal and 19 juxtapleural)



[Camarlinghi et al, Int. Journal of Computer Assisted Radiology and Surgery, IJCARS (2011)]

M5L lung CAD on-demand

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



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

COMPUTER APPLICATIONS

CrossMark

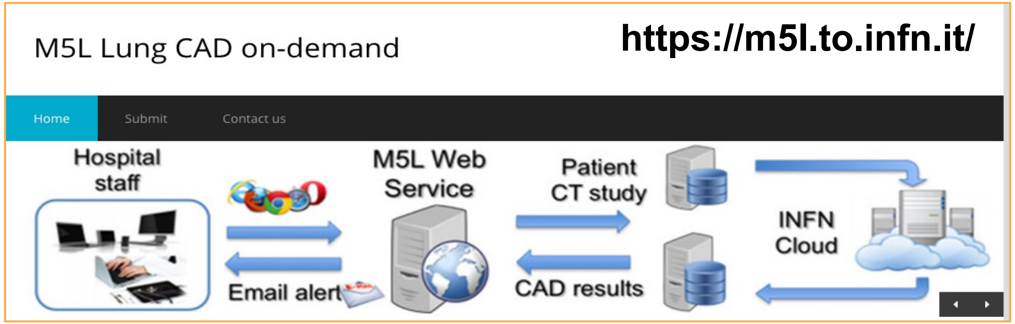
A cloud-based computer-aided detection system improves identification of lung nodules on computed tomography scans of patients with extra-thoracic malignancies

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

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

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

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



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

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

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

artificial intelligence in medicine AIM

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.

Project coordinator: A. Retico

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)

Objectives	Methods	Applications	
AIM1: Multicenter data harmonization	image processing; data filtering; feature extraction; feature classification; data harmonization; image classification; pattern recognition; radiomics; complex network analysis; network models; machine learning; deep learning	Reliability of multicenter studies in neuroimaging and in mammography.	MRI; RX
AIM2: Quantification		Quantification of radiotracer uptake in PET. Quantitative MRI methods.	MRI; PET; omics
AIM3: Predictive models		Radiomics and dosiomics in oncology. ML and DL for disease diagnosis	MRI; RX; RT dose; CT; omics

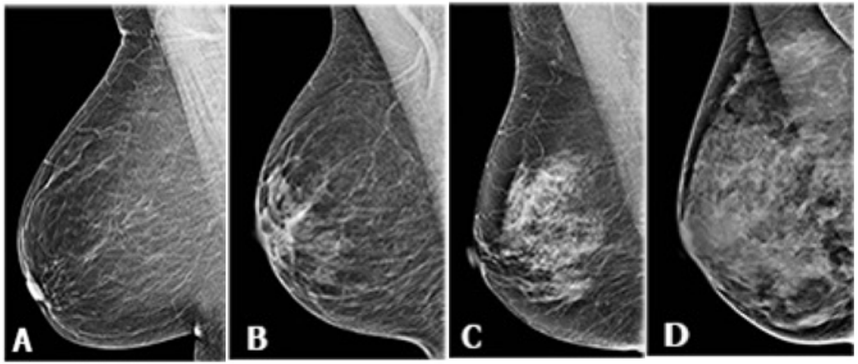
Analysis of lung CT in Covid-19

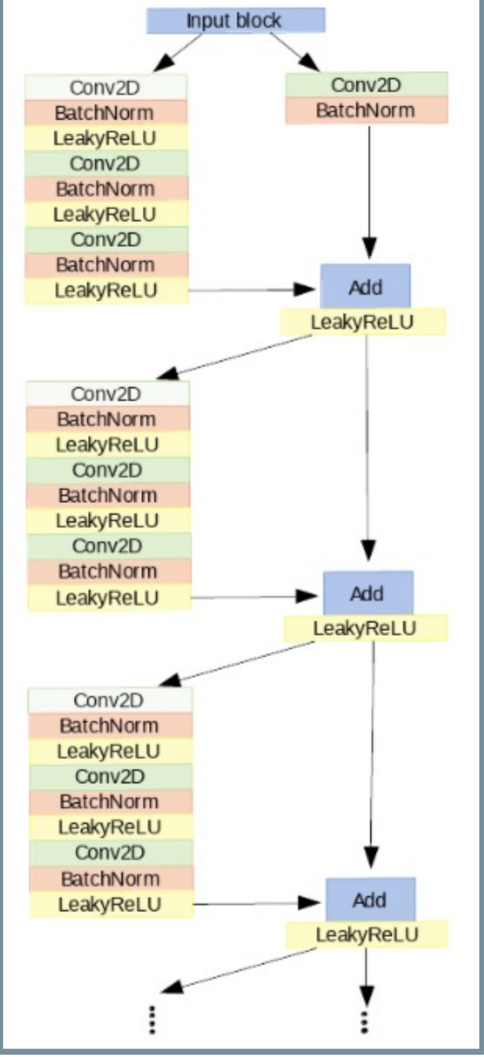
Long-standing collaboration with Italian & European centers (Hospitals / IRCCS) and with international consortia for data sharing

Deep residual CNN for breast density classification

Automated identification of tissue density class (A, B, C, D) with a residual CNN (R-CNN)

Goal: To contribute to the development of a new personalized **dose index** (depending on breast density) for each patient and each mammographic exam.





Dataset: about **2000 digital mammographic exams** collected by Azienda Ospedaliero-Universitaria Pisana (AOUP).

Dense/Non-dense	Left (%)	Right (%)	All (%)	BI-RADS	Left (%)	Right (%)	All (%)
Accuracy	84.4	88.8	89.4	Accuracy	73.3	76.7	77.3
Recall	82.3	89.9	90.0	Recall	72.1	79.2	77.1
Precision	85.5	87.7	88.9	Precision	76.6	75.2	78.6

[Lizzi F. et al., Residual convolutional neural networks to automatically extract significant breast density features. vol. 1089. Springer International Publishing; 2019]

RADIOMA

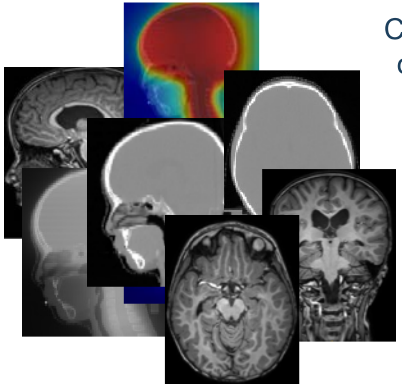
AZIENDA OSPEDALIERO UNIVERSITARIA PISANA

INFN Istituto Nazionale di Fisica Nucleare

FONDAZIONE PISA

Radiomics and Machine Learning to predict patients’ outcome

- Radiomic features are analyzed with Machine Learning methods to develop predictive models of diagnosis, prognosis or treatment outcome.
- For example: Predictive models of outcome of Radiotherapy Treatment (RT) based on Radiomics and Dosiomic features



Computation of features on multimodal images (MRI, CT, RT dose distribution)

Arrays of features

Predittive model based on machine learning

Prediction of treatment outcome

in collaborazione con

UNIVERSITÀ DEGLI STUDI FIRENZE · SST Servizio Sanitario della Toscana · aouc Azienda Ospedaliero Universitaria Careggi · MEYER con cura e con passione

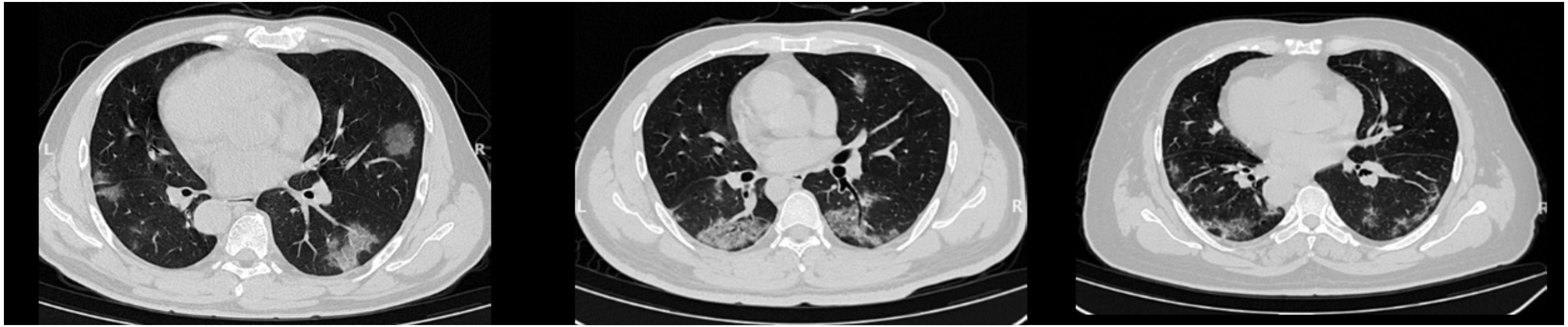
Talamonti C et al., *Radiomic and dosiomic profiling of paediatric Medulloblastoma tumours treated with Intensity Modulated Radiation Therapy*, Proc VIMABI2019

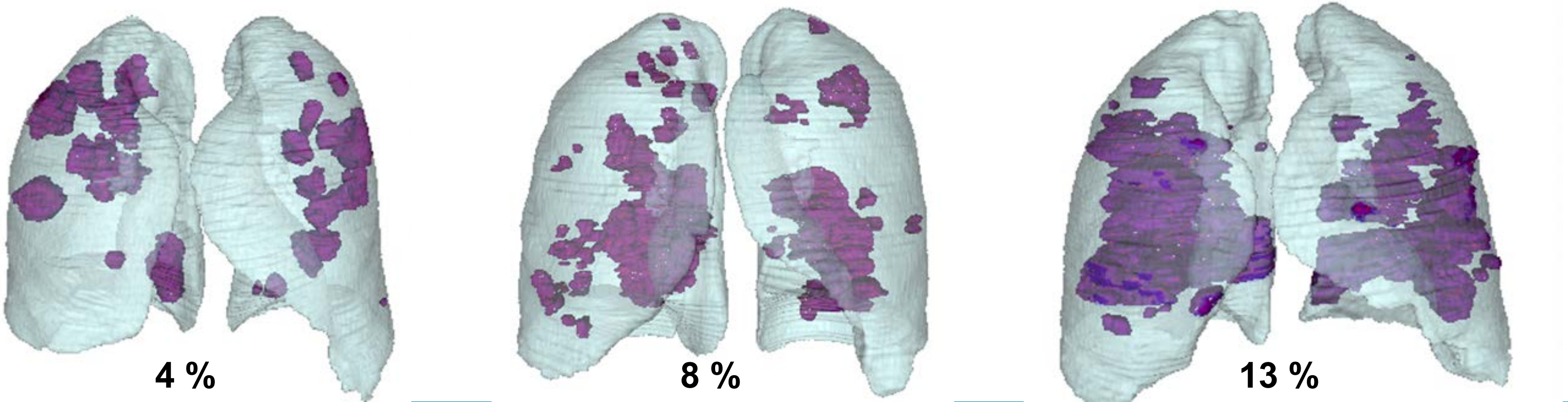
The AIM working group on lung CT analysis (AIM-Covid19-WG)

<u>Objective</u>: Automatic quantification of lung involvement on CT scans.

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

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

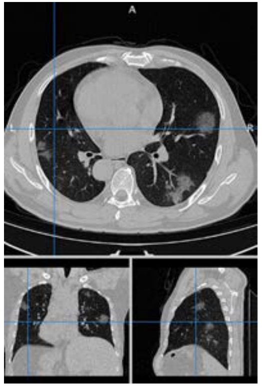




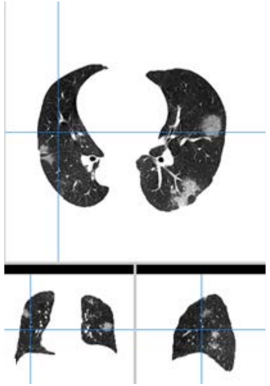
Alessandra Retico

Steps for the automatic quantification of lung involvement in CT scans

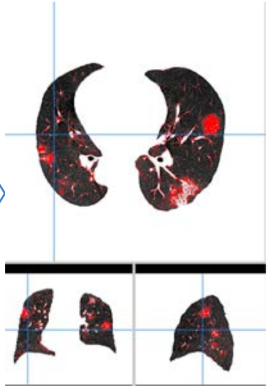
- AIM1: Multicenter Data Harmonization
- AIM2: Quantification
- AIM3: Predictive Models



Lung volume segmentation



Quantification of lung parenchyma affected by ground glass (GG) opacities



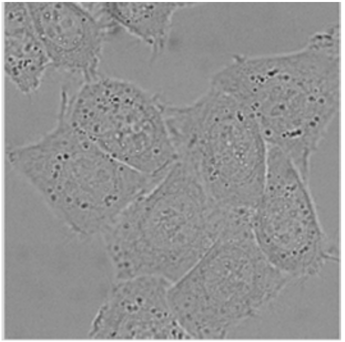
Classical algorithms for lung segmentation fail when lung appearance is strongly affected by interstitial pneumonia

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

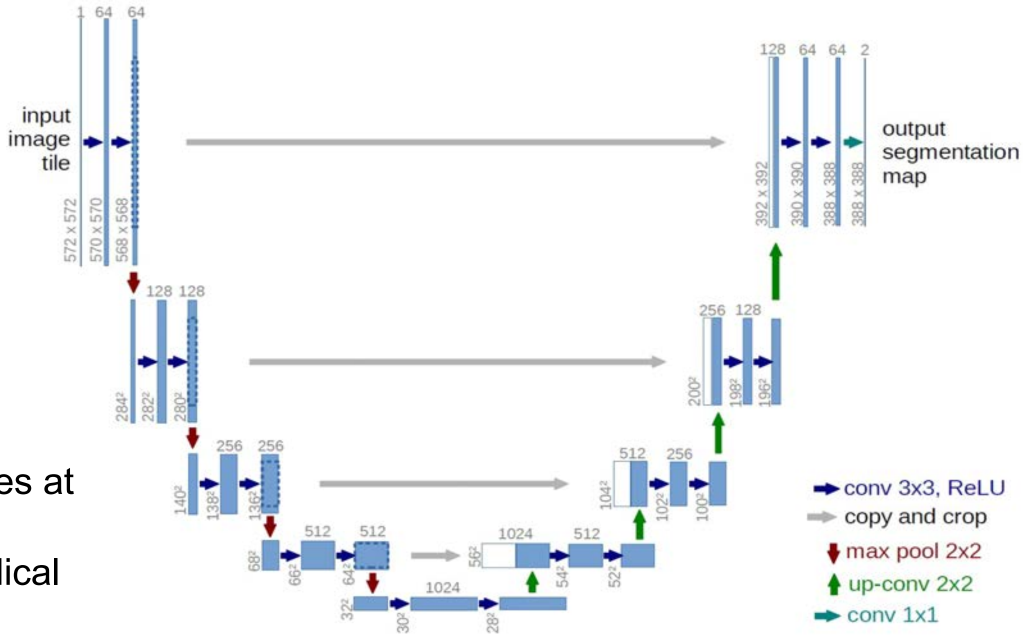
- Quantitative information on the amount of Covid-19 related lesions and their distribution, possibly combined with clinical and epidemiological patient's information, may be relevant to set up predictive models for patients' stratification, prognosis prediction, etc.
- Even only pure quantification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

U-Net: Convolutional Networks for Biomedical Image Segmentation

[O. Ronneberger et al., U-Net: Convolutional Networks for Biomedical Image Segmentation, Medical Image Computing and Computer-Assisted Intervention (MICCAI), Springer, LNCS, Vol.9351: 234--241, 2015]



HeLa cells



segmentation masks

U-net won two segmentation challenges at the International Symposium on Biomedical Imaging (ISBI) 2015

The U-Net architecture is outperforming other methods in most segmentation tasks

https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/

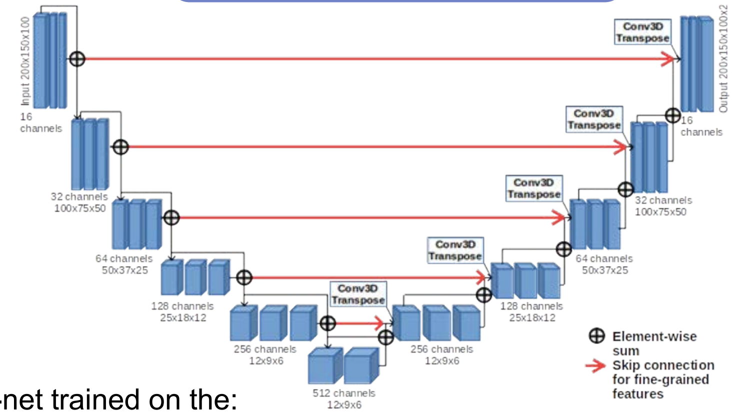
U-net$_1$ for lung segmentation

Windowing
[-1000, 1000] HU

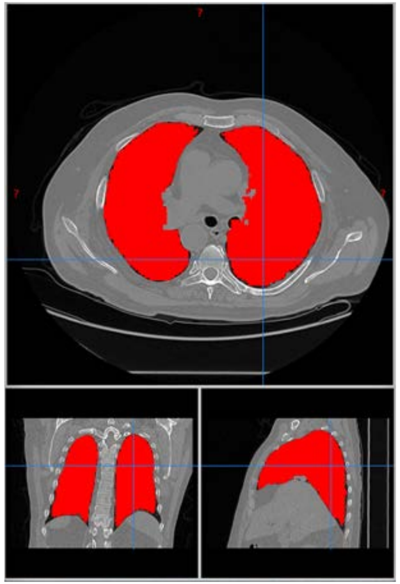
Resample to
200x150x100 array

U-net training and validation

Lung mask prediction



U-net trained on the:
- TCIA PleThora dataset (402 annotated CTs with lung masks)
- TCIA Lung CT Segmentation Challenge (60 CT, taken for dosimetry)
- MosMed CT0 subset (91 CT with normal lung, in-house lung segmentation SW to generate the reference lung masks)



U-net$_2$ for lesion segmentation

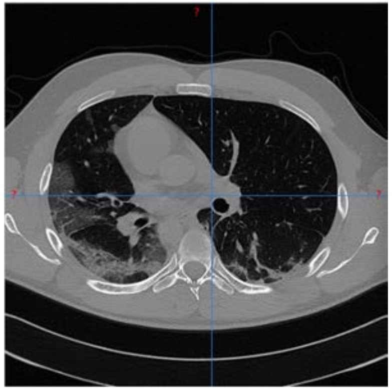
definition of the lung bounding box

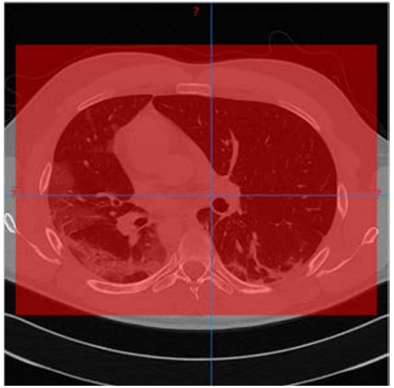
Windowing
[-1000, 300] HU

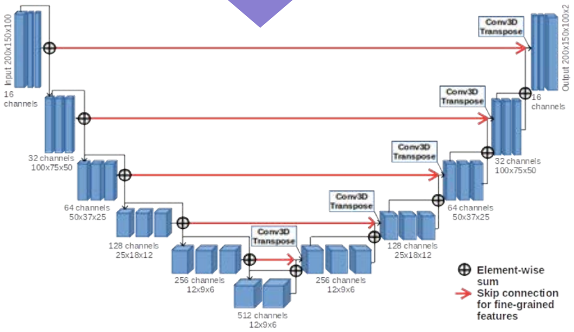
Resampling to 200x150x100 array

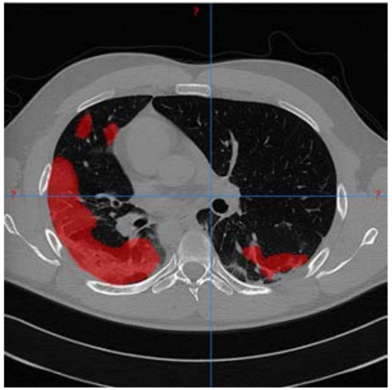
U-net training and validation

GG mask prediction



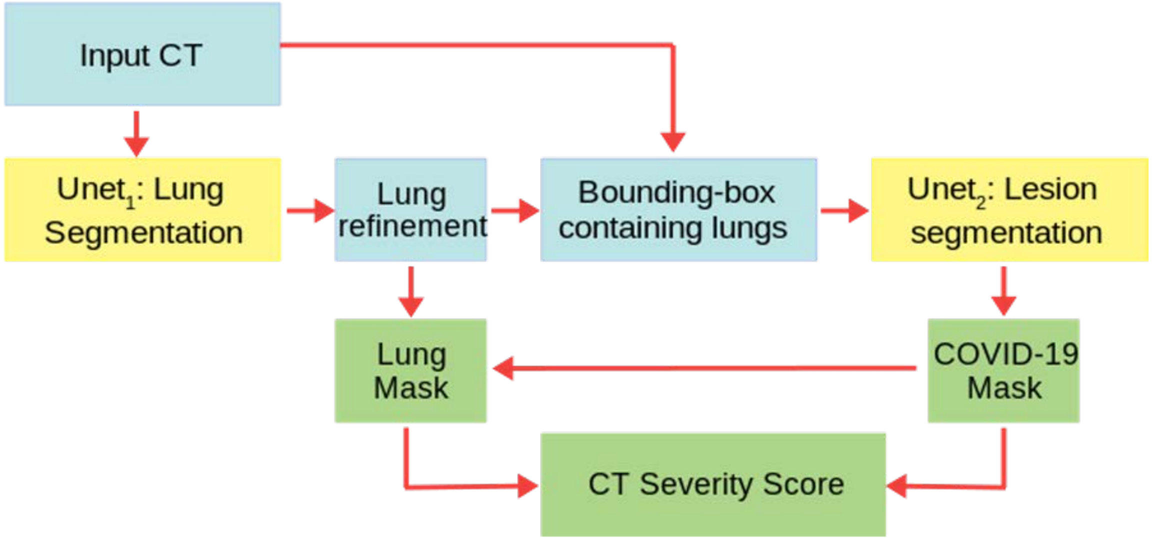




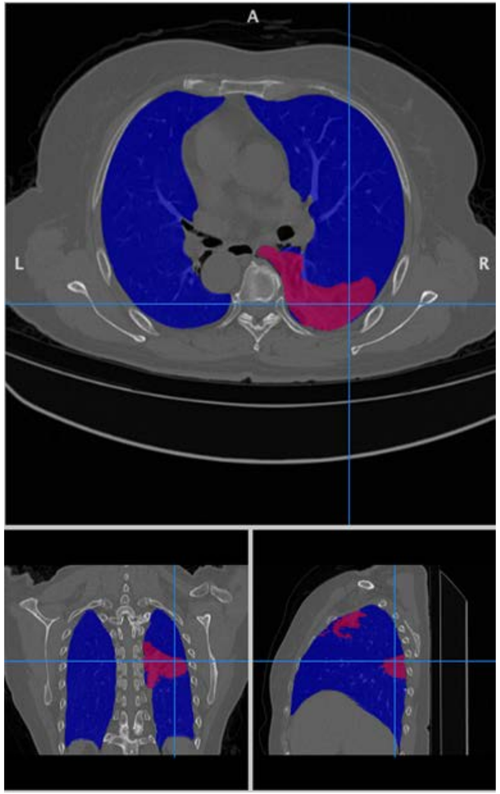


U-net trained on: the https://covid-segmentation.grand-challenge.org/ dataset (199 annotated CTs with COVID lesions, MICCAI endorsed event) and MosMed CT-1 (50 annotated CTs with COVID lesions)

LungQuant: a Deep-Learning based quantification system



LungQuant.py output



- U-nets were trained on (limited!) publicly available datasets
- Computing resources available at INFN-Pisa, CINECA, EOS cluster of Department of Mathematics at Univ. of Pavia have been exploited
  - GPUs with at list 16 GB of RAM were necessary
  - Each run required ~12h to complete 100 epochs
  - ... Test to improve the performances still in progress!

LungQuant : training details and cross-validation scheme

Loss functions:

Unet_1: $$\text{Dice}_{loss} = 1 - \frac{2 \cdot |M_{true} \cap M_{pred}|}{|M_{true}| + |M_{pred}|}$$

Unet_2: $$L = Dice_{loss} + CE_{weighted}$$

$$CE_{weighted} = w(x) \sum_{x \in \Omega} log(M_{true}(x) \cdot M_{pred}(x))$$

Evaluation metric:

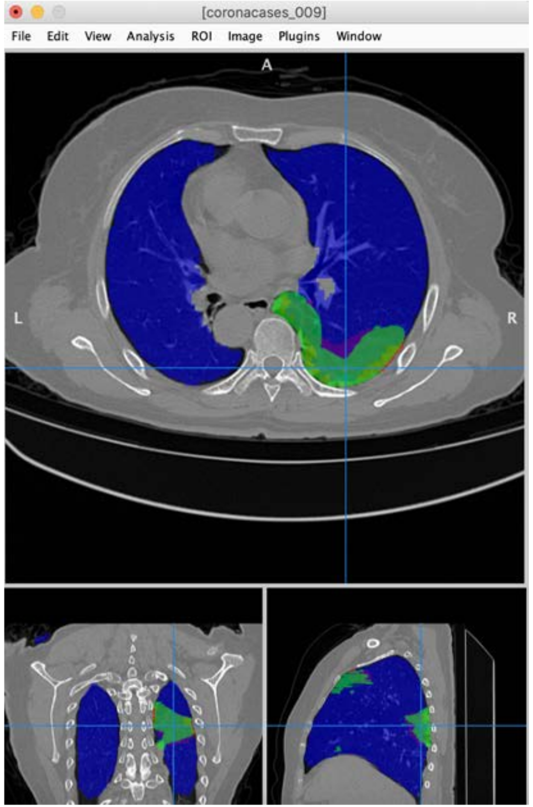
Dice: $$\text{Dice}_{metric} = \frac{2 \cdot |M_{true} \cap M_{predict}|}{|M_{true}| + |M_{pred}|}$$

We trained both the U-net for 300 epochs and we chose the epoch with the best validation metric to evaluate the performance on the test set.

U-net$_1$	train	validation	independent test
Plethora MosMed (91 CT-0) LCTSC	319 55 36	40 18 12	40 18 12
COVID-19-CT-Seg	/	/	10

U-net$_2^{90\%}$	train	validation	independent test
COVID-19 challenge MosMed (50 CT-1)	179 45	20 5	/ /
COVID-19-CT-Seg	/	/	10

The *LungQuant* system performance

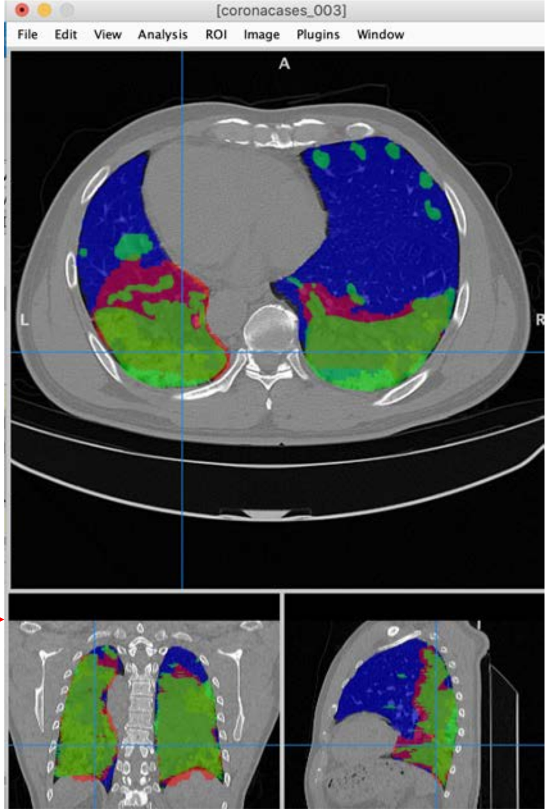


LungQuant system validation on the fully annotated (lungs and lesions) public benchmark dataset **COVID-19-CT-Seg**, https://doi.org/10.5281/zenodo.3757476 which is limited to 10 cases

Lung segmentation (Dice coefficient)	Infection segmentation (Dice coefficient)
0.95 ± 0.01	0.66 ± 0.13

best ←

→ worst



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

F. Lizzi 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, under review on International Journal of Computer Assisted Radiology and Surgery (IJCARS), https://arxiv.org/abs/2105.02566

Deep Learning vs. traditional Machine Learning approaches

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

Data augmentation (flip, rotate, scale images to augment data sets)

Model interpretability, explainable AI

Mandatory in medical applications

Conclusions

- Medical imaging daily produces an incredible amount of digital information which is not fully exploited neither for diagnosis/therapy nor for research!
- Clinicians need to be supported by reliable, effective and easy-to-use DSS (including those based on ML/DL) for diagnosing and monitoring a wide range of diseases
- The development of AI-based clinical DSS has multiple levels of complexity, thus it requires multidisciplinary skills

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

Thank you for your attention!

Contact: alessandra.retico@pi.infn.it
INFN, Sezione di Pisa