Artificial Intelligence for Medical Diagnosis

KT of AI developments for research infrastructures

CERN – Technology Department

Luigi Serio – Principal Investigator

Ioannis Stathopoulos – Doctoral student

Roman Stoklasa – Fellow

Bernardo Camajori Tedeschini – Master of Science student

HELLENIC REPUBLIC
National and Kapodistrian University of Athens
EST. 1837

POLITECNICO
MILANO 1863

Outline

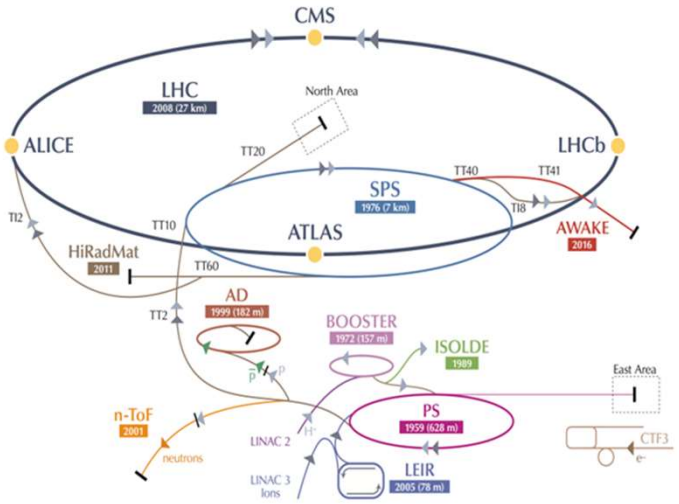
- Introduction
- Background work and developments
- Potential of applications to the medical field
- Proposed architecture
- Proof of concept
- First deliverables and results
- Future work

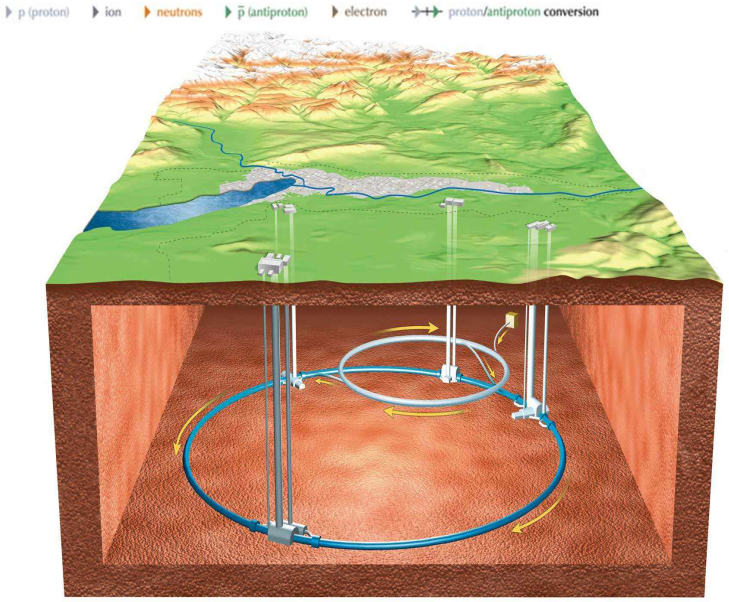
Outline

- Introduction
- Background work and developments
- Potential of applications to the medical field
- Proposed architecture
- Proof of concept
- First deliverables and results
- Future work

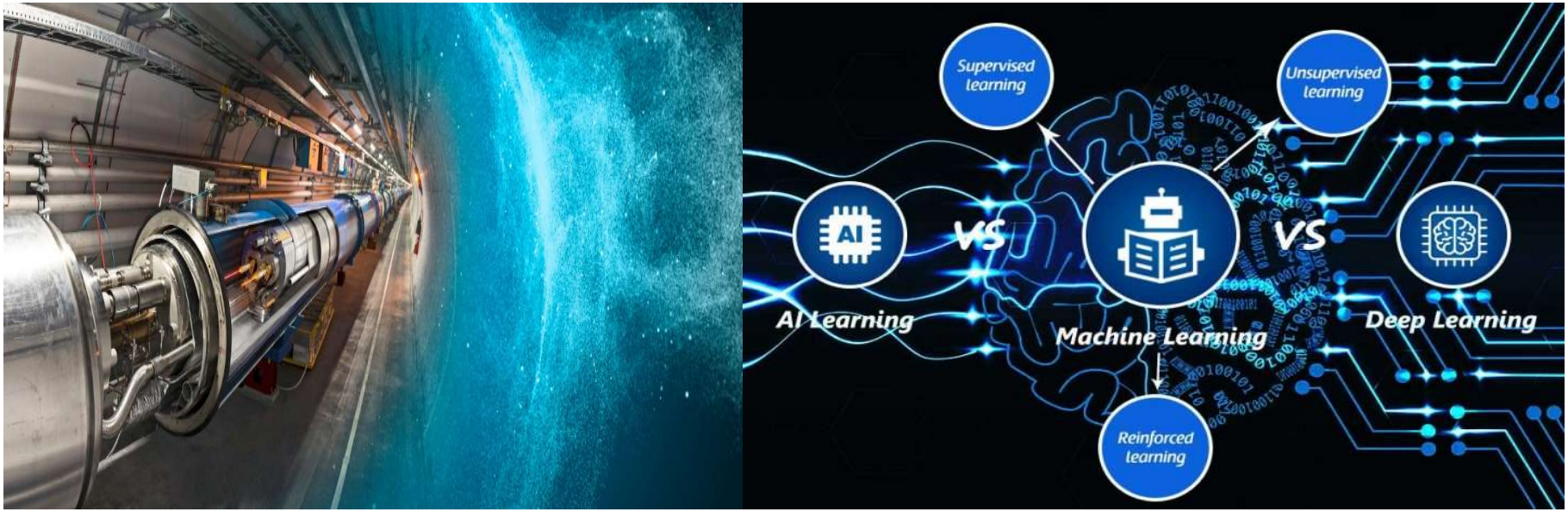
World's largest Particle Physics Laboratory (1954)

- Push forward the frontiers of knowledge
  - e.g. the secrets of the Big Bang …
  - what was the matter like within
  - the first moments of the Universe's existence?





Advanced tools to support maintenance and operation

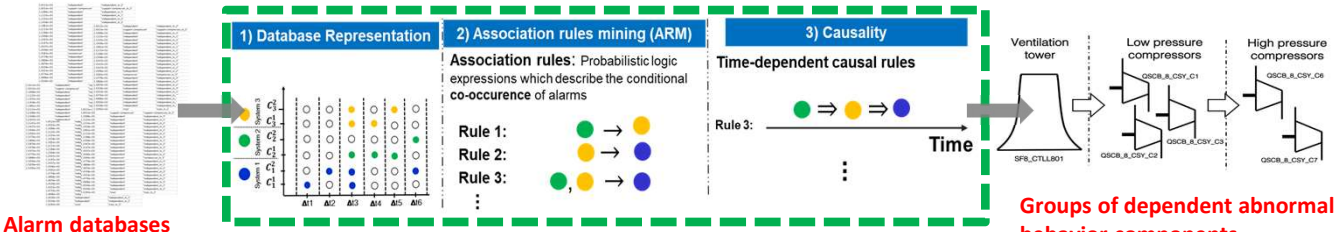


- CERN Complex Technical Infrastructure System of Systems with very stringent requirements in terms of availability and reliability
- Quality control, faults analysis, prevention, prediction and mitigation
- Complex fault trees, systems dependencies, risks and failures propagation, data and images analysis and interpretation
- Artificial Intelligence based systems to support and complement experts and operators in design, operation and maintenance

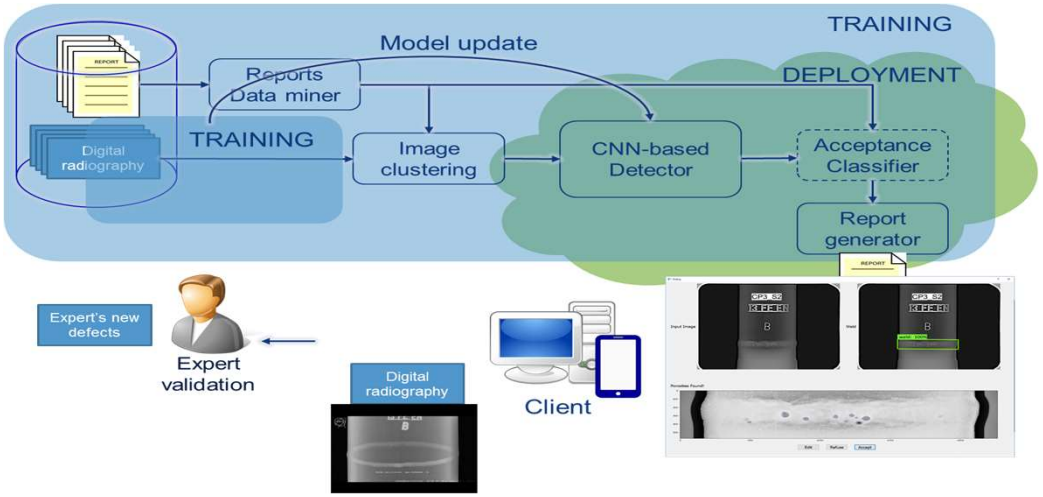
Outline

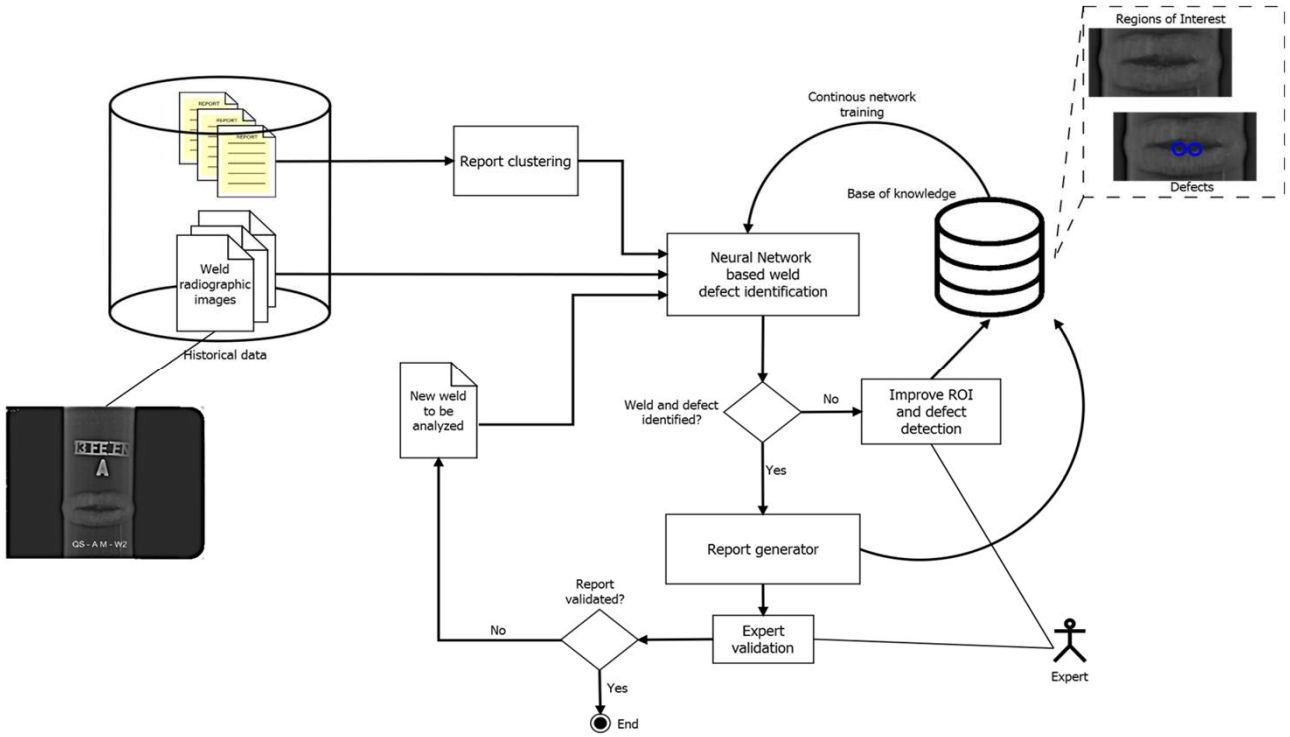
- Introduction
- Background work and developments
- Potential of applications to the medical field
- Proposed architecture
- Proof of concept
- First deliverables and results
- Future work

- Developed data-driven tools capable of discovering dependencies and abnormal behaviours:



- Developed automatic analysis of digital images based on CNN for radiography autonomous defects detection

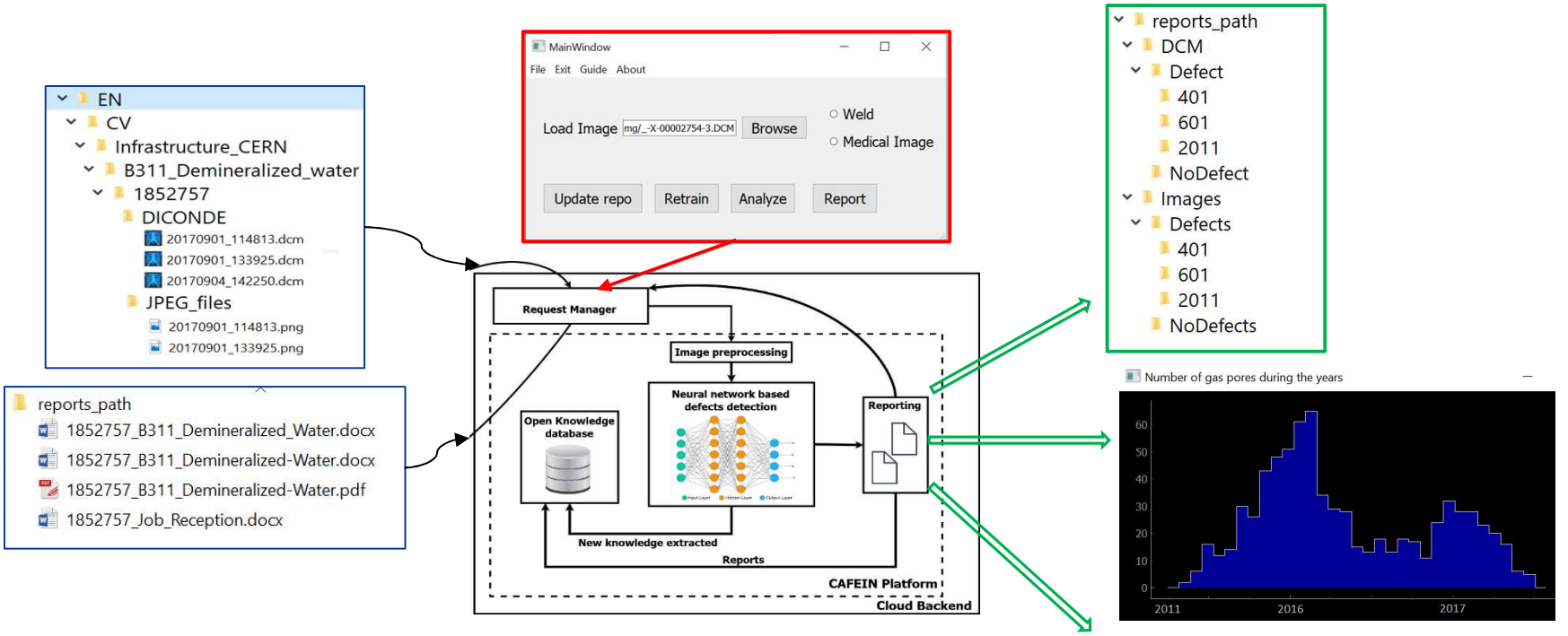


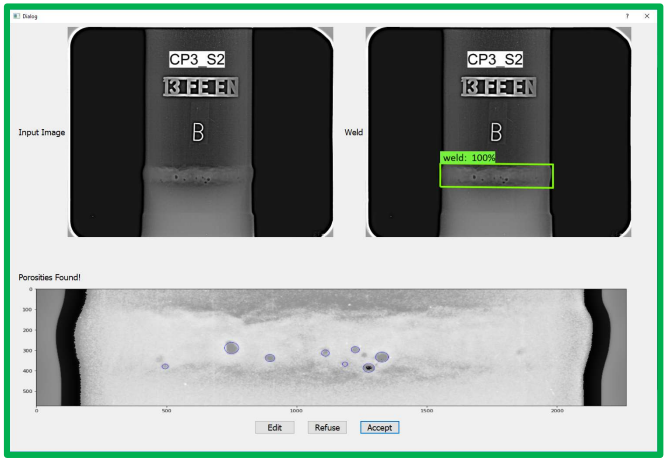


- Framework for welds and X-ray images and data
  - » Exploiting artificial intelligence, machine learning, and cloud-based techniques

- Efficient and error-reducing detection of defects and abnormalities but also the continuous optimization of welding and NDT techniques, machines and assessment criteria:
  - ✓ combine welds, X-ray machines and operator's data
  - ✓ provide support for quick and accurate diagnosis
  - ✓ improve accuracy, reliability, early detection
  - ✓ establish X-ray machines, welds and components database (images, data and findings)

Application development and results





The developed CNN-based detector (by category):

- AP with IoU* of 50 % for **weld** detection = 97 %
- AP with IoU* of 50 % for **pore** detection = 79 %

The evaluation set considered is composed of 450 images

*Intersection over Union

Outline

- Introduction
- Background work and developments
- Potential of applications to the medical field
- Proposed architecture
- Proof of concept
- First deliverables and results
- Future work

Potential extension of the application to the medical field

- Computer-Aided deFEcts and anomaly detection, Identification and classificatioN system for diagnostic imaging, industrial, radiology and medical data, exploiting artificial intelligence, machine learning, federated learning, distributed computing and cloud-based techniques

- Efficient and error-reducing detection of diseases, illnesses and injuries but also the continuous optimization of imaging and radiology techniques, tools and assessment criteria:

  - combine imaging with data
  - provide support for quick and accurate analysis diagnosis
  - improve accuracy, reliability, early detection
  - establish industrial and medical database (images, data and findings)

Market potential

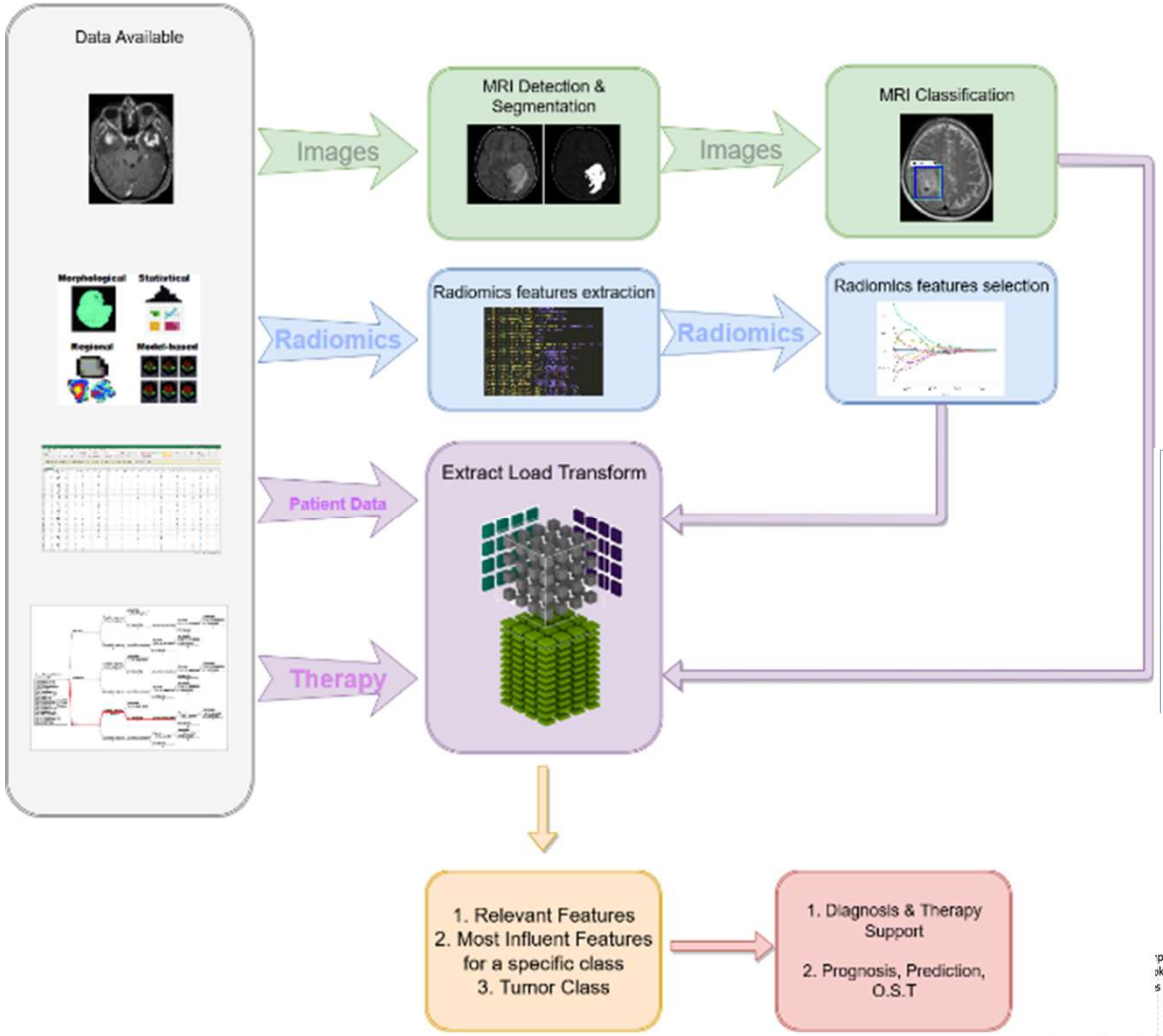
- Field of application:
  - Semi-automated analysis and modelling of medical data & images
  - Diagnosis and treatments based on multiple features and data beyond human perception
  - Edge devices and mobile technologies to enable early and fully remote and autonomous medical diagnosis in remote areas and harsh environments
  - Federated learning and distributed computing to ensure privacy for a wide and safe international collaboration as well as access to diagnostic models in remote areas
- Competing technologies:
  - No existing technologies ready to use in the field
  - High cost, no tailoring for clinical needs, black boxes, lack flexibility and no privacy preserving
- Envisioned innovation potential:
  - No competitive technologies
  - Potential commercial partners, user community:
    - Major imaging and diagnostic devices manufacturers, medical data centers, World Health Organisation and disease analysis centers, hospitals, developing countries

Outline

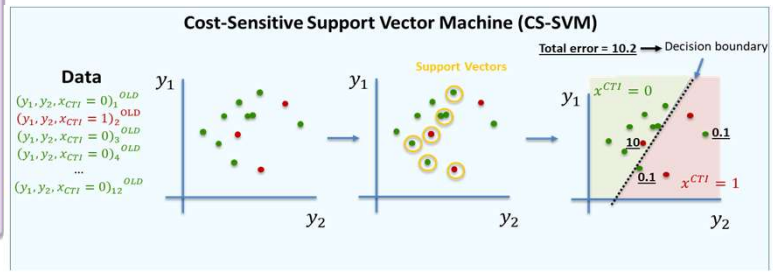
- Introduction
- Background work and developments
- Potential of applications to the medical field
- Proposed architecture
- Proof of concept
- First deliverables and results
- Future work

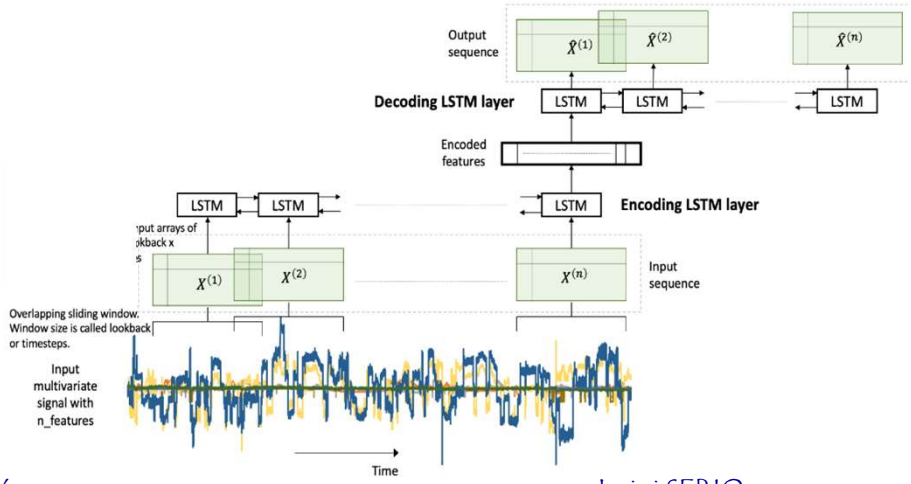
Architecture development proposal 1/2

Computer-Aided Defects and Anomalies Detection, Identification and Classification system for Digital Images and Medical Data



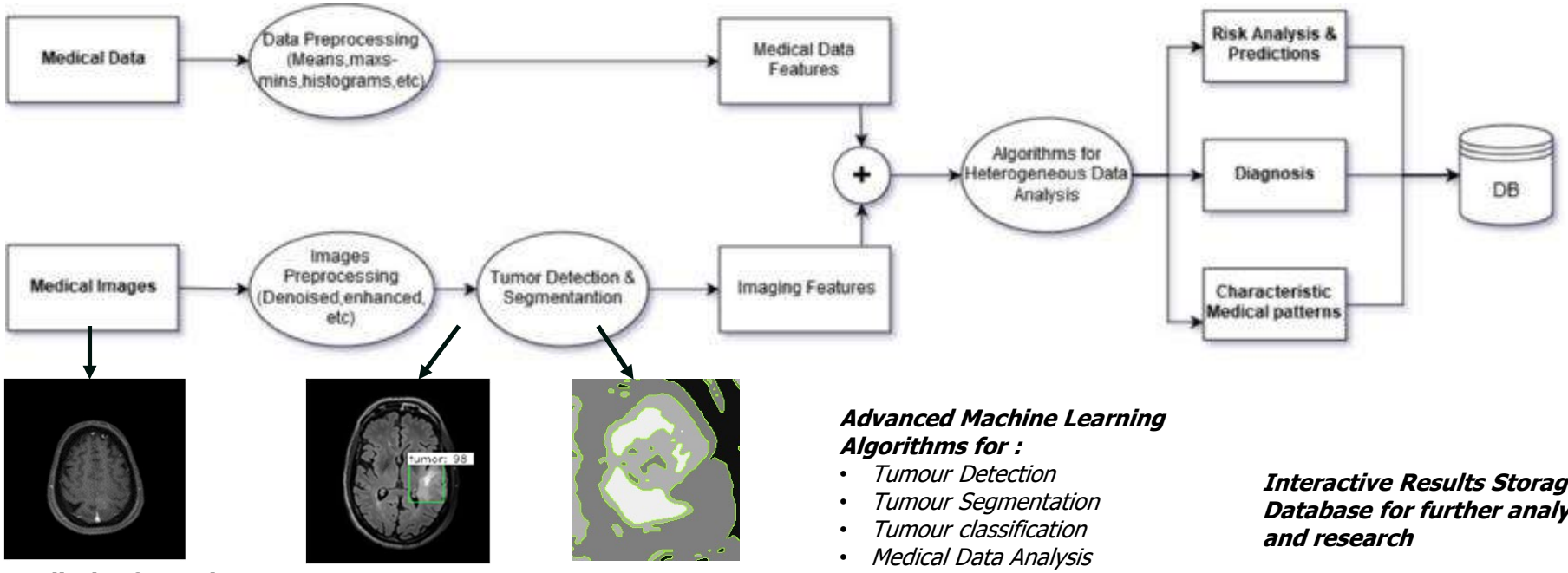
AI algorithms to extract features, patterns and develop models for risk prediction and prevention based on digital images and clinical, therapies and genetic data.





Architecture development proposal 2/2

Computer-Aided Defects and Anomalies Detection, Identification and Classification system for Digital Images and Medical Data



Medical Information Dataset

- *Personal information*
- *Medical history*
- *Laboratory test/results*
- *Medications prescribed*
- *Disabilities, allergies*
- *Genetic Data*
- *Geolocalization Data*

Medical Imaging Dataset

- *Multimodal Magnetic Resonance Images*
- *Brain Tumours*

Advanced Machine Learning Algorithms for :

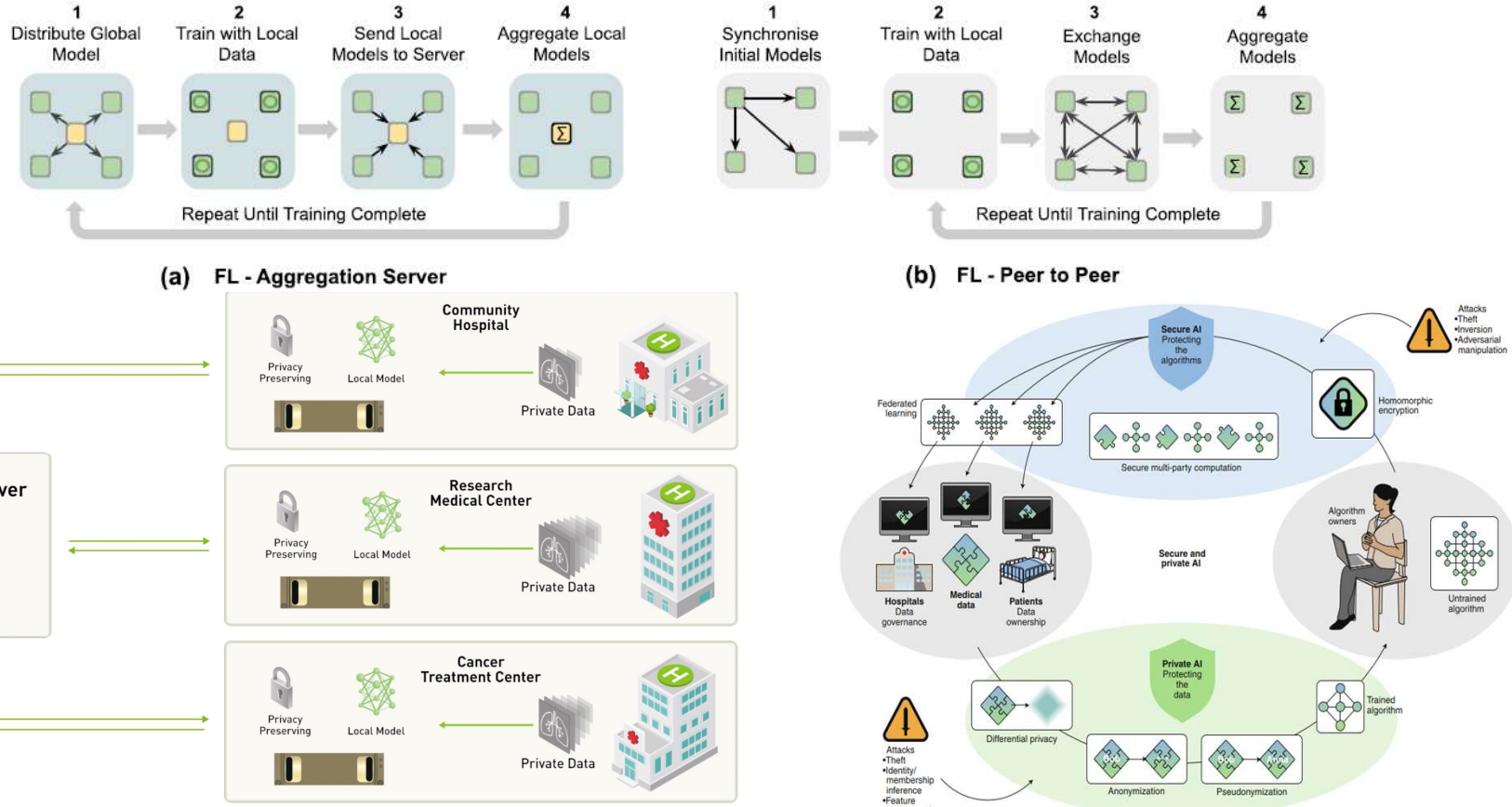
- *Tumour Detection*
- *Tumour Segmentation*
- *Tumour classification*
- *Medical Data Analysis*
- *Integration & Analysis for Heterogeneous Input Data*

Interactive Results Storage Database for further analysis and research

- Efficient and error-reducing detection of diseases, illnesses and injuries
- Continuous optimization of imaging and radiology techniques, tools and assessment criteria:
  - ✓ combine medical imaging with patient's data
  - ✓ provide support for quick and accurate diagnosis
  - ✓ improve accuracy, reliability, early detection
  - ✓ establish medical database (images, data and findings)

Distributed learning of models without sharing data

- Privacy-preservation
- Increase (federate) datasets
- Implement on edge devices and in remote location



Outline

- Introduction
- Background work and developments
- Potential of applications to the medical field
- Proposed architecture
- Proof of concept
- First deliverables and results
- Future work

Materials and Methods

- Cases from a MRI machine of 3-Tesla from the 2nd dept of Radiology of the Medical School of Athens
- 230 Brain MRI examinations containing different pathologies: *Tumors, Multiple Sclerosis , White Matter Hyperintensities, Strokes, Lesions*
- Up to six different MRI modalities
- Several Image resolutions (3D or 2D)

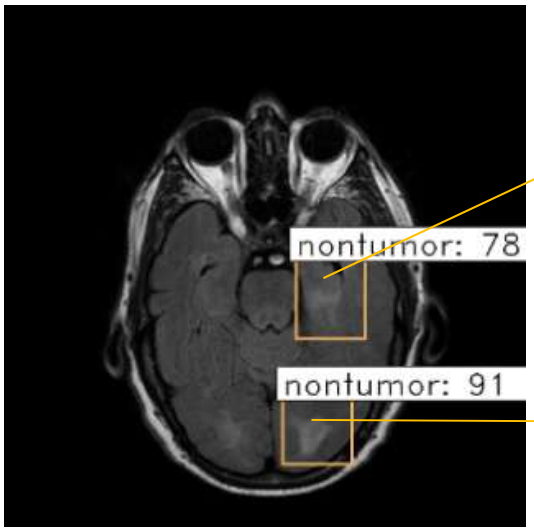
First results and proof of concept 1/2

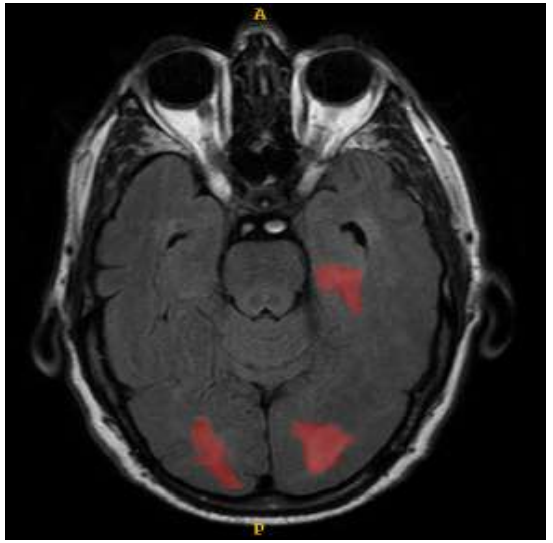
Model's Output with Detection & Classification

Segmented ROI

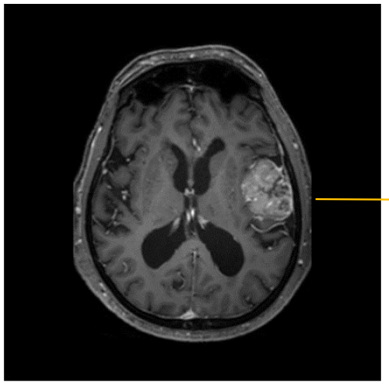
Contoured ROI

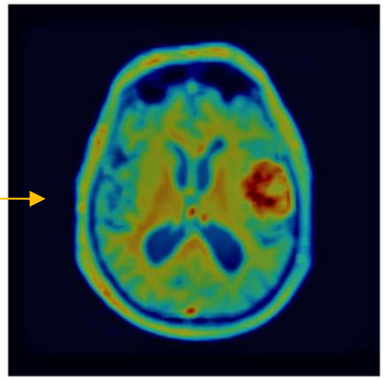
Image labeled by the Radiologist





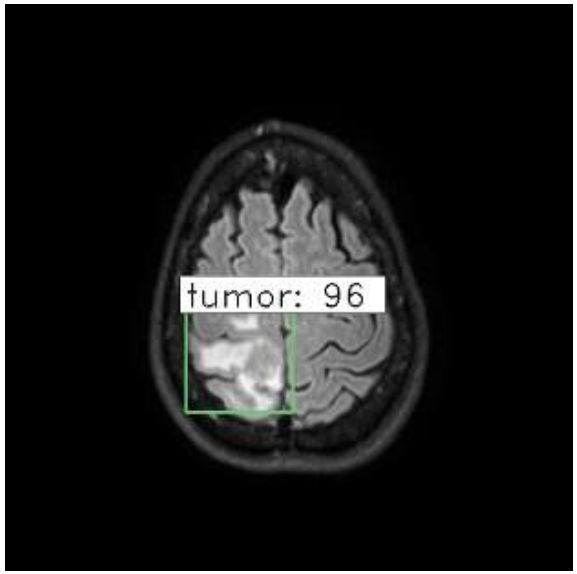
Activation Map





First results and proof of concept 2/2

Model's Output with Detection & Classification



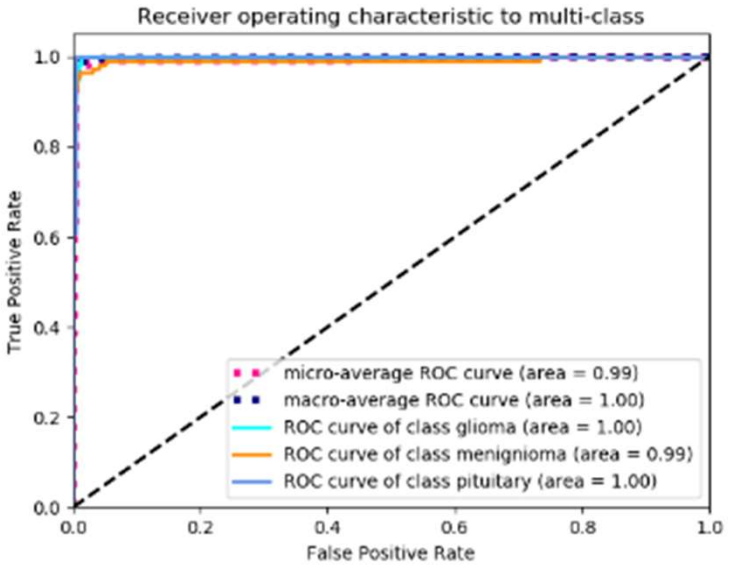
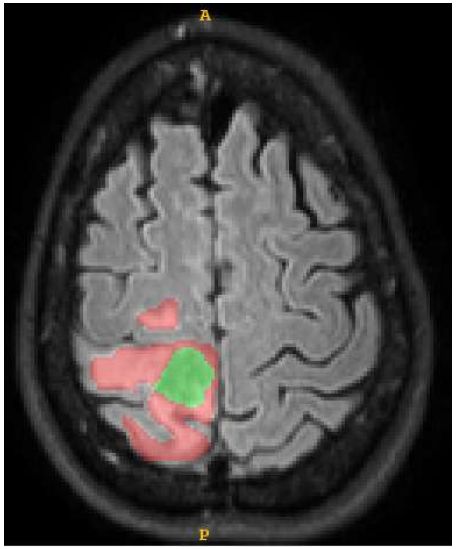


Image labeled by the Radiologist

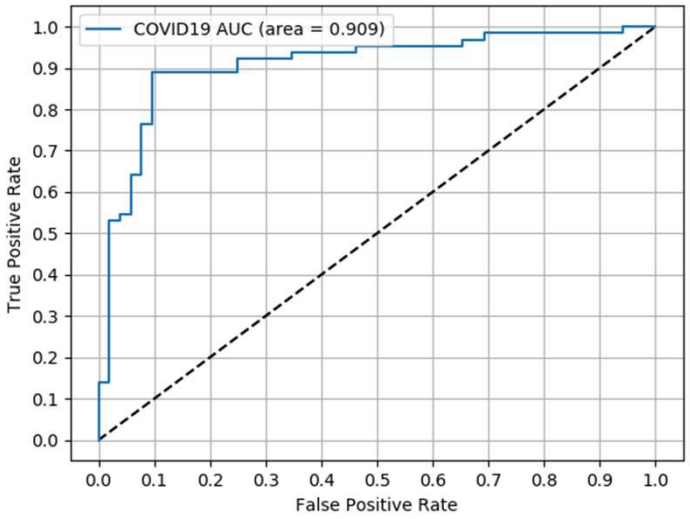


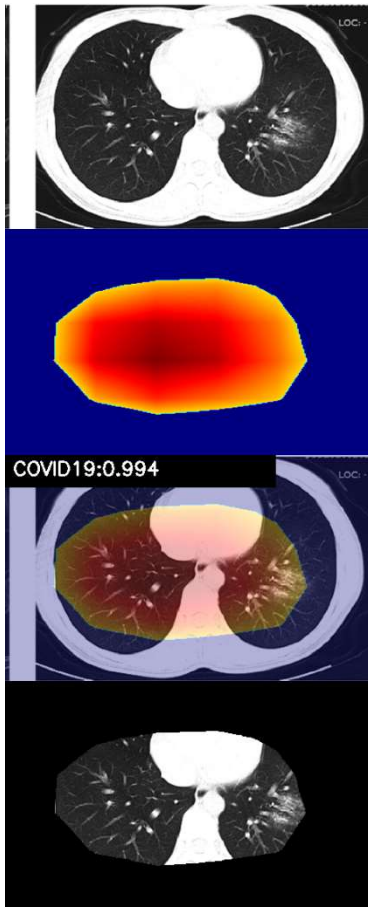
Results based on the training on initial 230 multimodal MR Images

- Detection accuracy: 84 %
- Classification accuracy: above 93 % (image based)
- Classification accuracy: 85 % (radiomics based)

Robustness and adaptability of the tool ->Testing on a completely different clinical case: Lung CT scans – COVID diagnosis

	precision	recall	f1-score	test	train
Non COVID	0.85	0.90	0.88	52	345
COVID	0.92	0.88	0.90	64	285
accuracy			0.89	116	





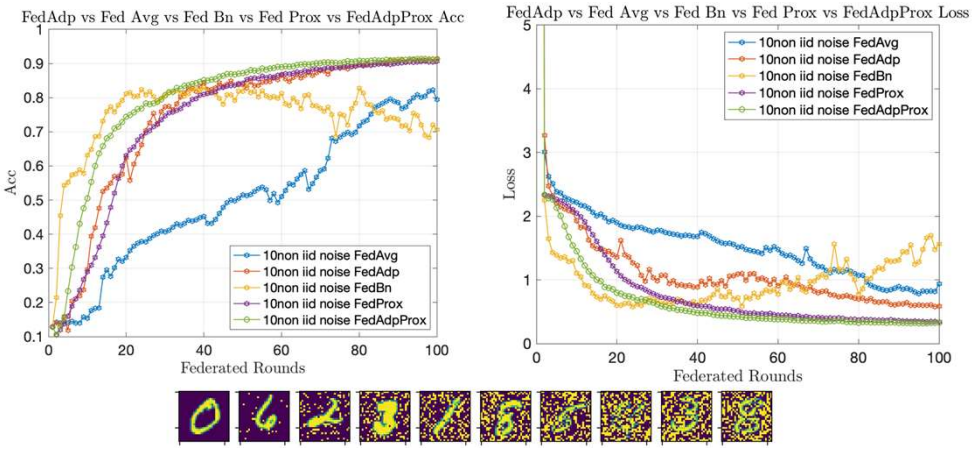
Attention map

CAFEIN Federated Learning proof of concept

Computer-Aided Defects and Anomalies Detection, Identification and Classification system for Digital Images and Medical Data

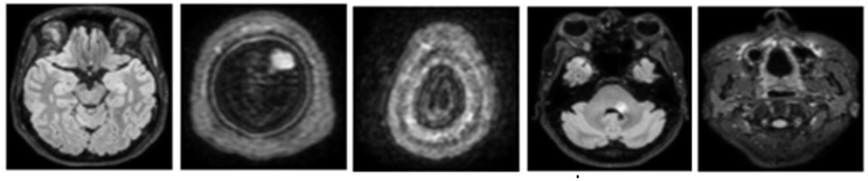
Model testing:

- Demonstration on simple CNN, then increasing model complexity and layers
- Successful reproduction of the performance of centralized learning above 98 %
- Disjoint classes divided to different clients with random distribution & partial datasets per client
- Introduction of noise and offsets in the datasets as well as integration of new clients



Ongoing proof of concept on real medical dataset

- Divided into multiple "institutions"
  - Slices from the same examination ends in the same institution
  - Examinations assigned randomly to institutions
  - Variants: rand 1 / 3 / 5



	classification accuracy %
centralized	99.2
rand-1	99.1
rand-1 - from scratch	98.0

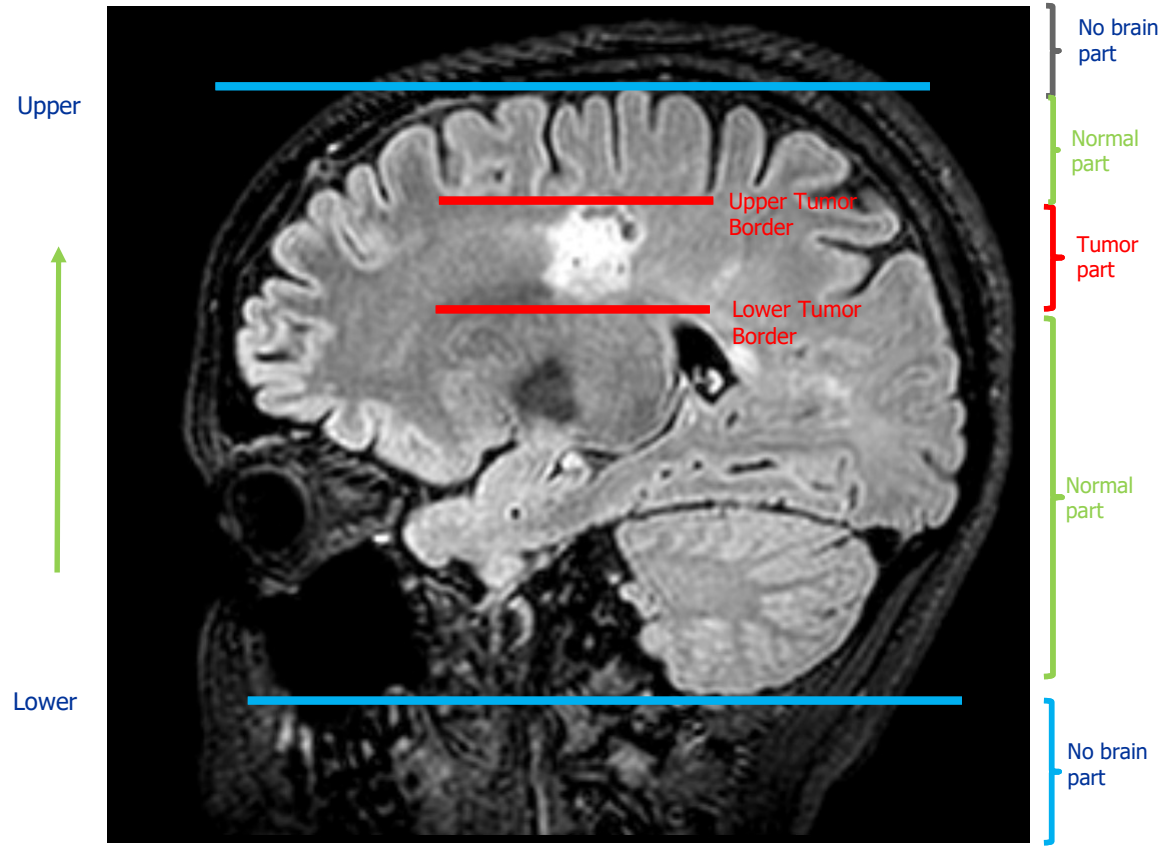
rand-3 @ 3 cpr	99.3
rand-5 @ 3 cpr	99.1
rand-5 @ 5 cpr	99.1

Outline

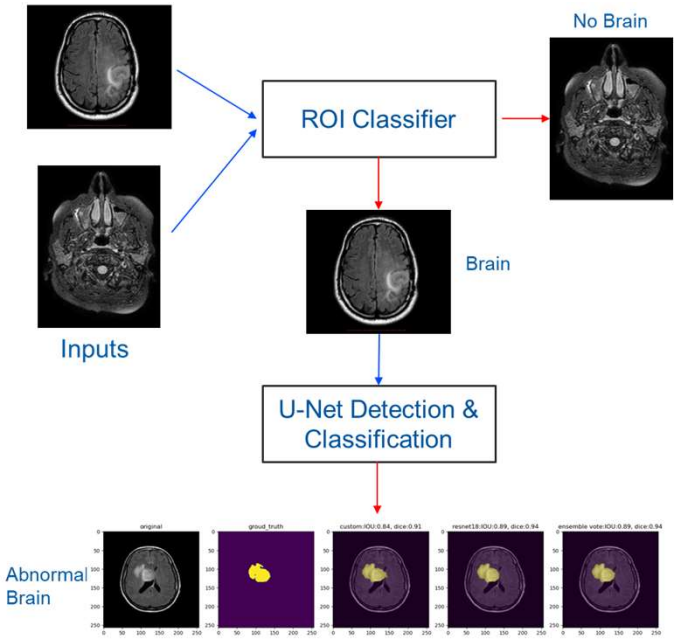
- Introduction
- Background work and developments
- Potential of applications to the medical field
- Proposed architecture
- Proof of concept
- First deliverables and results
- Future work

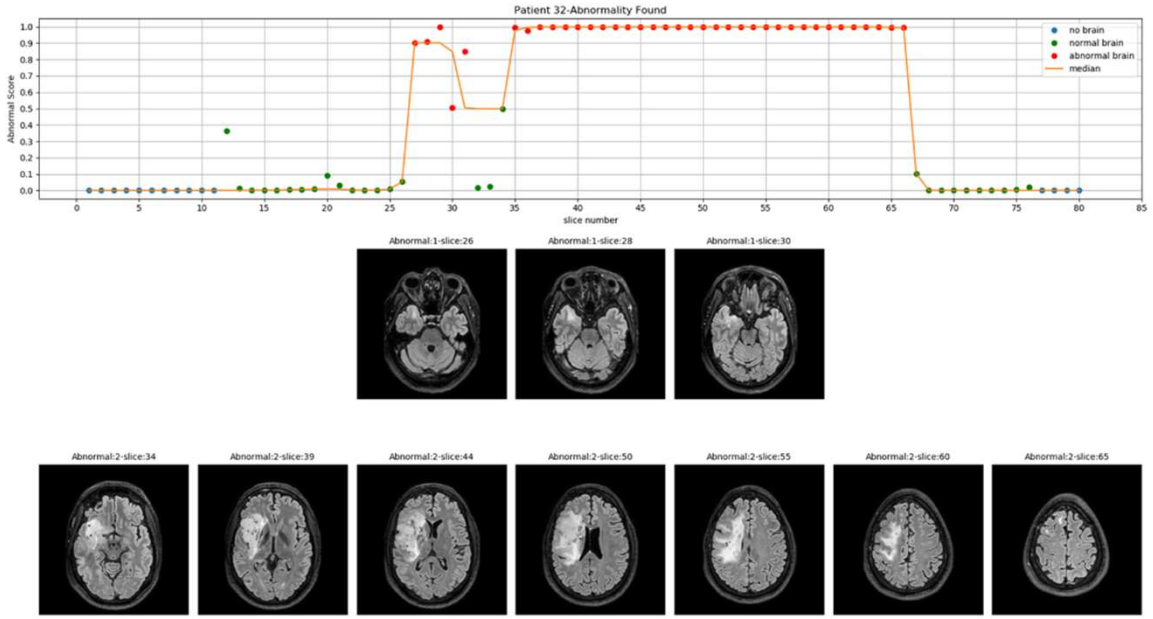
Clinical application based on developments

- Typical examination comes as consecutive 2D slices. Only **a small amount** contains useful information
- Need of a tool for automated and optimized **screening** to save time and use efficiently clinical resources
- Models investigated:
  - VGG16 + 2 dense layer
  - 2D U-Net
  - 3D multi-modal U-Net



Screening tool for MRI images pathologies detection with identification of relevant images

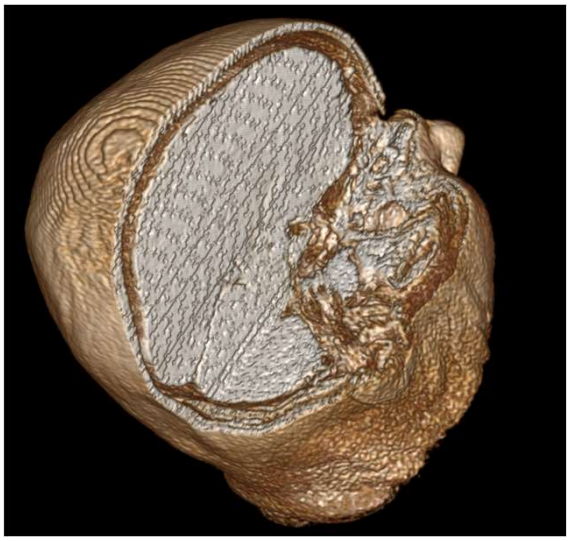


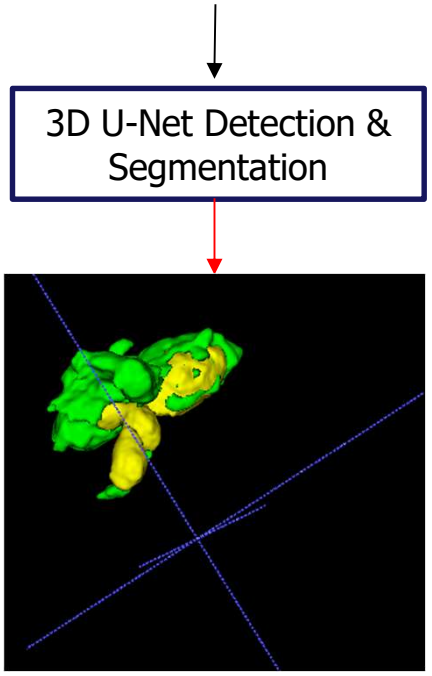


- Per-slice accuracy above 97 %
- Average Dice coeff.: 0.89 - Average tumor area coverage: 0.90
- Parametric detection thresholding

Screening, detection and segmentation

- 3D (Volume) based - Multimodal
- Detection & Segmentation
- **3D multi-modal U-Net-based models**
- *Training and testing with*
  - *Public dataset*
  - *Medical School of Athens*
- Preliminary Average Dice coef: 0.89
- *To be tested and evaluated in a clinical environment at Radiology department of the Medical School of Athens*





Outline

- Introduction
- Background work and developments
- Potential of applications to the medical field
- Proposed architecture
- Proof of concept
- First deliverables and results
- Future work

Future work

- ✓ Implementation of developed tools in
  - ✓ Federated learning and distributed computing environment
  - ✓ On edge devices

  (several institutes interested – seeking funding)

- ✓ Extension to medical diagnosis and treatment (initial dataset from Athens)

- ✓ Extension to prevention and treatment (initial dataset from Athens)

EPIC study:

'More than 521 000 study participants enrolled from 23 centres in 10 western European countries. Detailed information on diet, lifestyle characteristics, anthropometric measurements, and medical history was collected at recruitment (1992—1999).'

'Biological samples including plasma, serum, leukocytes, and erythrocytes were also collected at baseline from 387 889 individuals and are stored at the International Agency for Research on Cancer – World Health Organization (IARC-WHO) and mirrored at EPIC collaborating centres. Overall, the EPIC biorepositories host more than 9 million aliquots, constituting one of the largest biobanks in the world for biochemical and genetic investigations on cancer and other chronic diseases. Follow-up measures of lifestyle exposures have been collected and centralized at IARC in 2019. '

1. extract features, patterns and develop descriptive and predicting models;
2. risk prediction and prevention based on heterogeneous clinical data and guidelines;
3. improve and complement the diagnosis and medical imaging analysis;
4. ensure the full privacy of the data via federated learning and distributed computing.