

AFRICAI Summer School

MODEL DEVELOPMENT 2: MODEL-CENTRIC BEST PRACTICES, PITFALLS, AND OPEN ACCESS INFRASTRUCTURES



Dr. ir. Martijn P. A. Starmans, PhD

*PostDoc AI for Integrated Diagnostics in Oncology
Department of Radiology and Nuclear Medicine &
Department of Pathology, Erasmus MC, Rotterdam, NL*

✉ m.starmans@erasmusmc.nl

<https://mstarmans91.github.io/>

EURO BIOIMAGING

eosc cancer

EUCAN
IMAGE

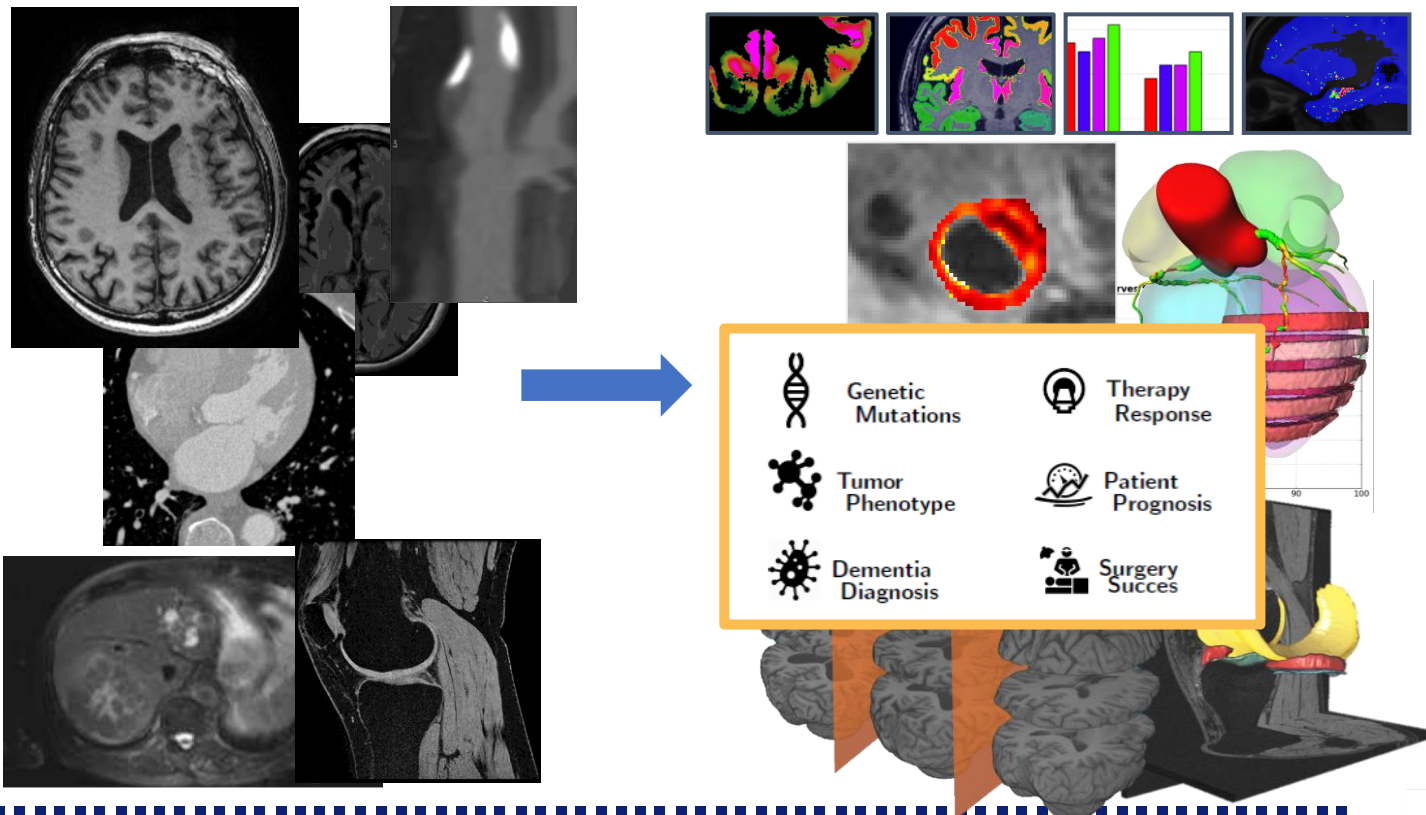
radioval EUCAIM

Erasmus MC
University Medical Center Rotterdam

Erasmus

Erasmus University Medical Center





Programme

Model Development 2: Model-centric best practices, pitfalls, and open-access infrastructures

Agenda

Time	Mon. 11 Sept	Tue. 12 Sept	Wed. 13 Sept	Thu. 14 Sept	Fri. 15 Sept	Sat. 16 Sept
8.30-10:00	Welcome (Karim, Jihad) Presentation MICCAI & AFRICAI (Nassir, Karim)	From an idea to a MICCAI paper Paper 1 (Mohammad)	From an idea to a MICCAI paper Paper 2 (Anees)	New trends in AI (Victor)	MICCAI writing Part 2 Methods (Sandrine)	Final presentations by the participants
10h00-10h30	Break	Break	Break	Break	Break	Break
10h30-12h00	What is a MICCAI Paper? (Scope, structure, review process)	Developing medical AI in Africa (Florent)	Design Thinking in AI (Islem)	MICCAI writing Part 1 Introduction (Karim)	MICCAI writing Part 3 Results (Martijn)	Feedback by mentors + Awards
12h00-13h30	Lunch	Lunch	Lunch	Lunch	Lunch	Lunch
13h30-15h00	Pitch presentations by the Participants	Model development Part 1: Data (Apostolia)	Model development Part 2: Models (Martijn)	Model development Part 3: Evaluation (Martin)	Discussion of next steps until 2024 (All)	Free
15h00-15h30	Break	Break	Break	Break	Break	
15h30-17h30	Practical session with mentors	Practical session with mentors	Practical session with mentors	Practical session with mentors	Practical session with mentors	

Machine Learning

Framework

Platform

Library

Deep Learning

Framework

Platform

Library

Tool

Reinforcement Learning

Programming

Education

Lineage

Relational DB

Store & Format

Versioning

Operations

Feature Engineering

Stream Processing

SQL Engine

Visualization

Pipeline Management

Labeling & Annotation

Governance

Inference

Federated Learning

Training

Parameter

Format & Interface

Marketplace

Workflow

Benchmarking

Tool

Explainability

Adversarial

Bias & Fairness

Distributed Computing

Interface

Security & Privacy

Natural Language Processing

Notebook Environment

EDL

SOAJS

Spark

EMR

Intel

OpenStack

Security

Interface

Security & Privacy

Natural Language Processing

Notebook Environment

Security & Privacy

Natural Language Processing

Notebook Environment

Goals

You will know:

- Which AI toolboxes are available
- How to use the MONAI AI toolbox: (**main aim of practice today**)
 - To make a data loader for your data
 - To design and train a deep learning model
- Which monitoring tools you can use to track the progress of your models
- Which processing environments you can use
- How you can benefit from pre-trained models
- Which open-access infrastructures you can use



AI toolboxes

Focus: Python

Arguably, three main AI toolboxes:

- Tensorflow
- PyTorch
- Keras (Tensorflow or PyTorch backend)

All AI in general: what is most suitable for medical imaging?

Top 10 Python Libraries



Pandas

Data analysis and manipulation



NumPy

Mathematical functions



Matplotlib

Data visualisations



SeaBorn

Data visualisations



Tensorflow

Machine Learning



Keras

Deep Learning



SciPy

Scientific computing



PyTorch

Machine Learning



Scrapy

Web crawling



SQLModel

Interact with SQL databases

MONAI+

Medical Open Network for Artificial Intelligence

Core v1.1

Label v0.6

Deploy App SDK v0.5.1

1,000,000+ downloads and counting

Contributors



dkfz.



NHS
Guy's and St Thomas'
NHS Foundation Trust



kitware



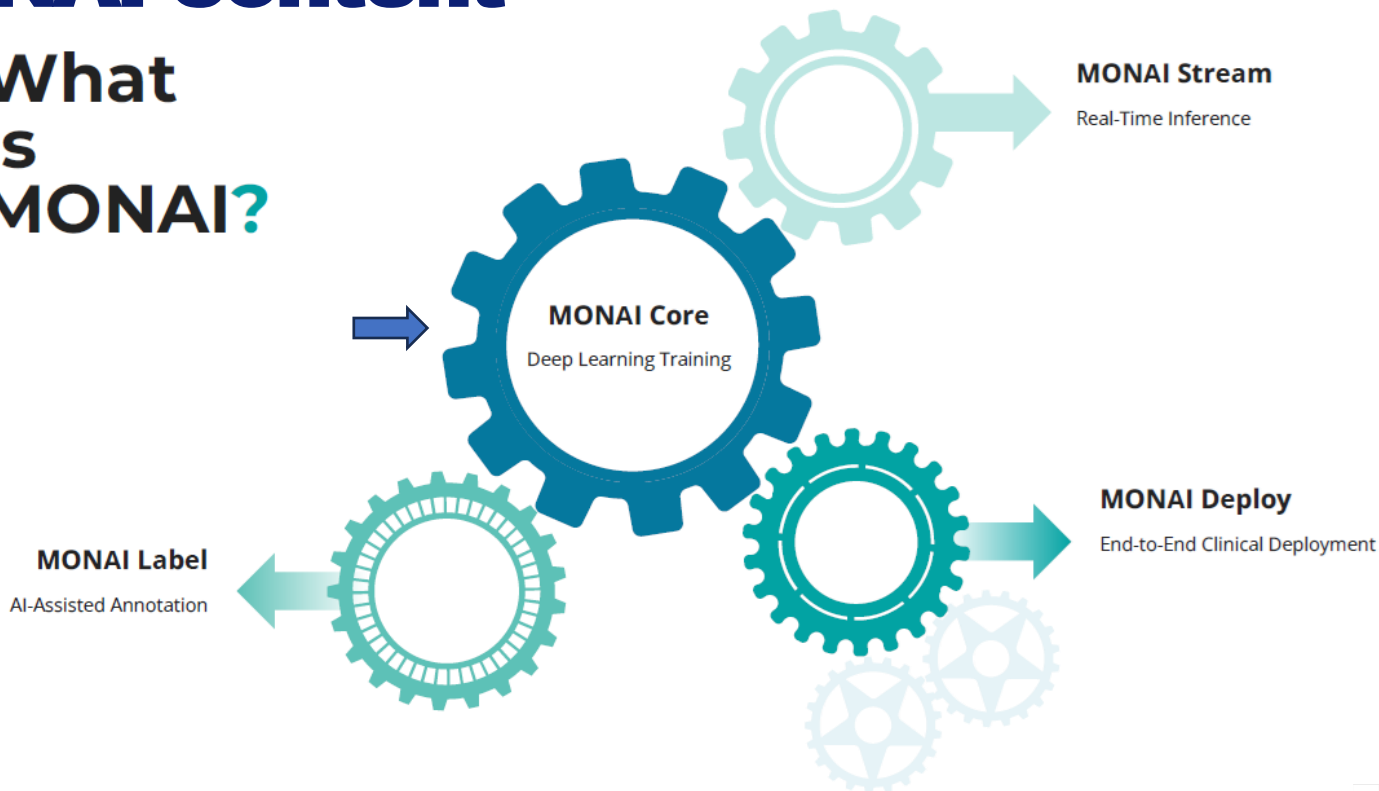
**2021, 2023: MICCAI/MIDL – MONAI
bootcamp**

Erasmus MC



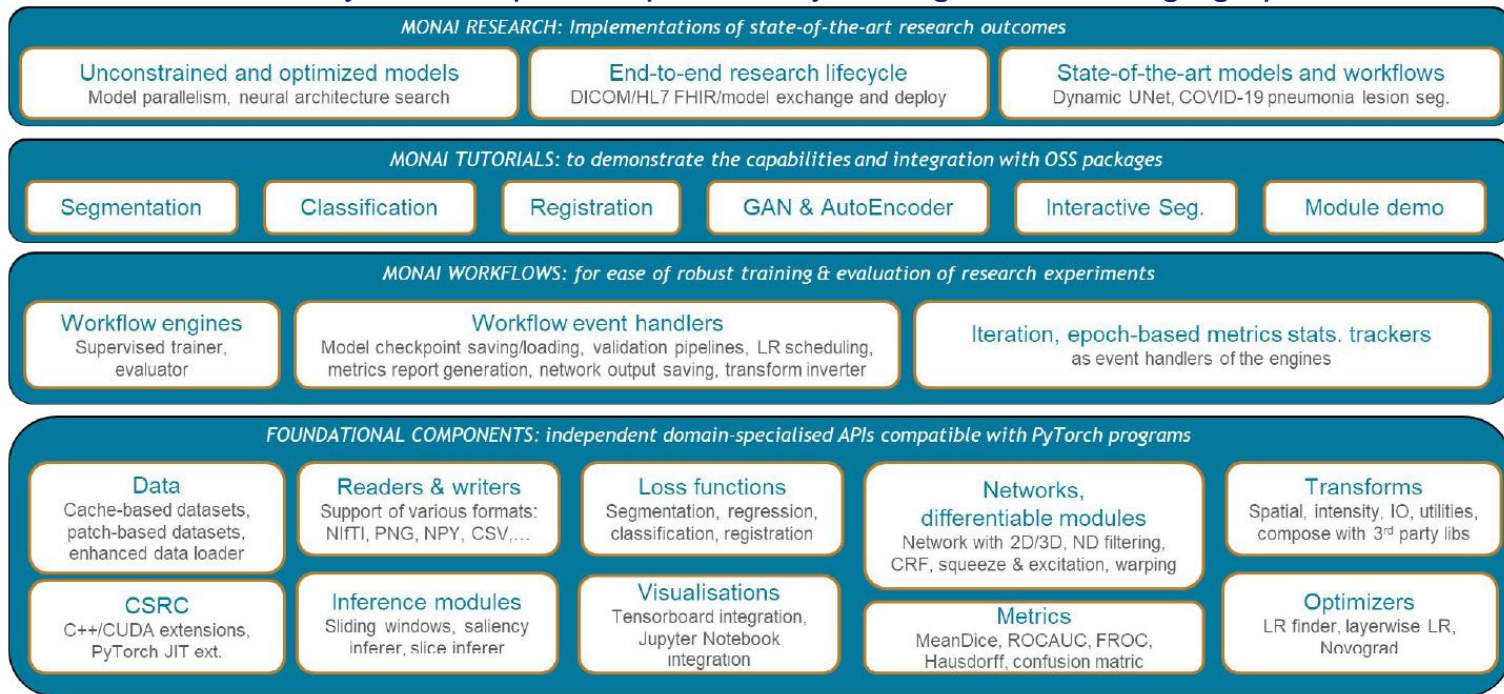
MONAI content

What
is
MONAI?



MONAI Core

MONAI is based on PyTorch, expands upon that by adding medical imaging-specific solutions:



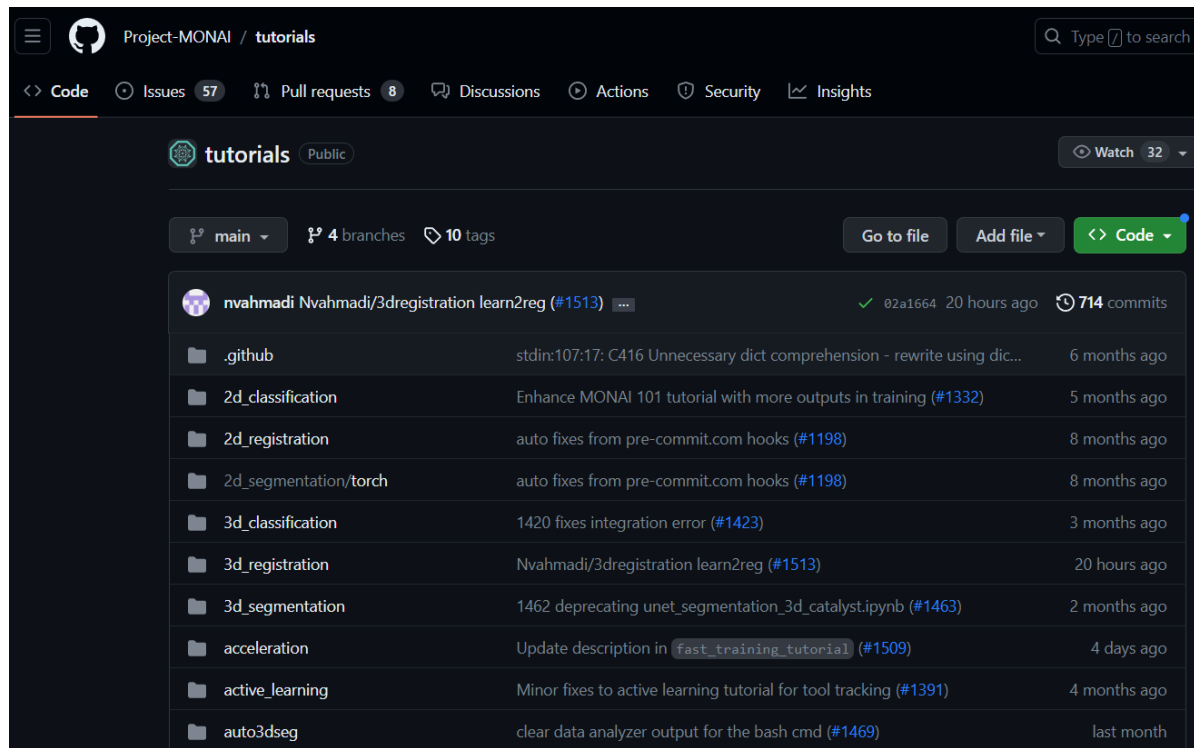
MONAI Core

MONAI is based on PyTorch, expands upon that by adding medical imaging-specific solutions:

- ✓ Easy to use (lot of functionality implemented already for you)
- ✓ Easy to learn (lot of tutorials)
- ✓ Efficient (training about 12x speedup compared with native PyTorch implementation)
- ✓ Scalable (lot of newly proposed (MICCAI) solutions developed in MONAI)
- ✓ Sustainable (live community)

MONAI Core tutorials

<https://github.com/Project-MONAI/tutorials> <https://docs.monai.io/en/stable/index.html>



The screenshot shows the GitHub repository page for 'Project-MONAI / tutorials'. The repository is public and has 32 watchers. It features a table of files and folders with their commit history and timestamps. The files listed include .github, 2d_classification, 2d_registration, 2d_segmentation/torch, 3d_classification, 3d_registration, 3d_segmentation, acceleration, active_learning, and auto3dseg. Each entry shows the commit message and the time since the last commit.

File/Folder	Commit Message	Time
.github	std:in:107:17: C416 Unnecessary dict comprehension - rewrite using dic...	6 months ago
2d_classification	Enhance MONAI 101 tutorial with more outputs in training (#1332)	5 months ago
2d_registration	auto fixes from pre-commit.com hooks (#1198)	8 months ago
2d_segmentation/torch	auto fixes from pre-commit.com hooks (#1198)	8 months ago
3d_classification	1420 fixes integration error (#1423)	3 months ago
3d_registration	Nvahmadi/3dregistration learn2reg (#1513)	20 hours ago
3d_segmentation	1462 deprecating unet_segmentation_3d_catalyst.ipynb (#1463)	2 months ago
acceleration	Update description in fast_training_tutorial (#1509)	4 days ago
active_learning	Minor fixes to active learning tutorial for tool tracking (#1391)	4 months ago
auto3dseg	clear data analyzer output for the bash cmd (#1469)	last month

MONAI 2023 Bootcamp

<https://github.com/Project-MONAI/monai-bootcamp>

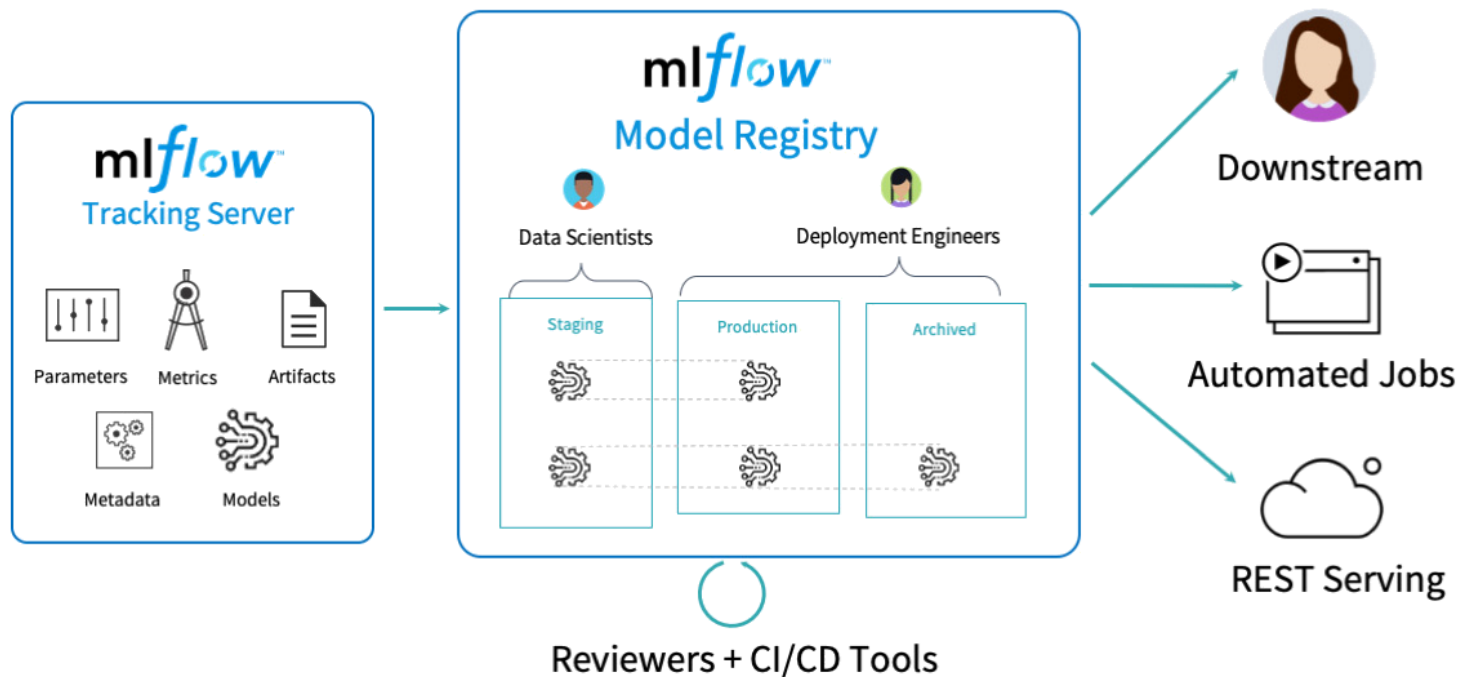
<https://www.youtube.com/playlist?list=PLtoSVSQ2XzyAJAGzaHF0nUIkav0BnxhrJ>

Erasmus MC




Monitoring



When training many models, how to keep an overview -> Monitoring!





Monitoring



When training many models, how to keep an overview -> Monitoring!


 Experiments Models GitHub Docs

Experiments  



Search Experiments

Default  

mlflow-demo  


mlflow-demo 

Share

 Track machine learning training runs in an experiment. [Learn more](#) 


Experiment ID: 1

► Description [Edit](#)

 Refresh



Compare

Delete


Download CSV 


↓ Start Time

All time

Columns

Only show differences 






 metrics.rmse < 1 and params.model = "tree"

Search

Filter

Clear

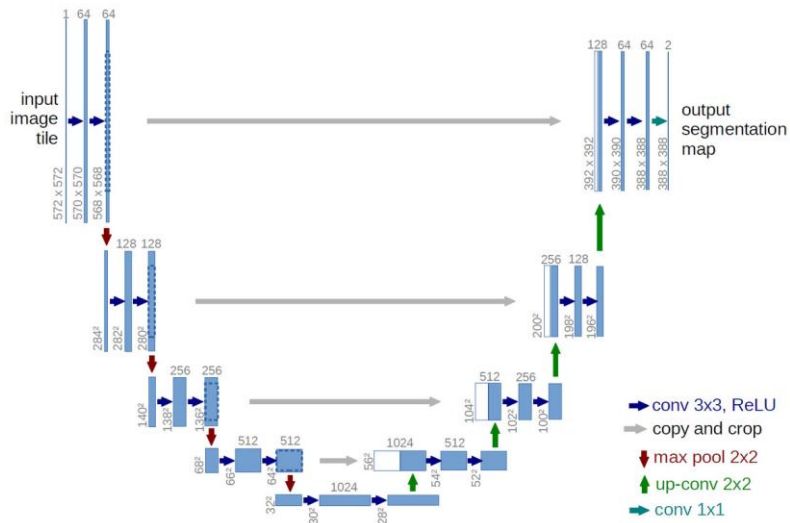
Showing 5 matching runs

								Metrics	Parameters
<input type="checkbox"/>	↓ Start Time	Duration	Run Name	User	Source	Version	Models	accuracy	depth
<input type="checkbox"/>	✓ 2 minutes ago	9.4s	run_4	KayJan	C:\Program	-	 sklearn	1	20
<input type="checkbox"/>	✓ 2 minutes ago	11.9s	run_3	KayJan	C:\Program	-	 sklearn	1	10
<input type="checkbox"/>	✓ 2 minutes ago	12.4s	run_2	KayJan	C:\Program	-	 sklearn	0.967	5
<input type="checkbox"/>	✓ 3 minutes ago	16.8s	run_1	KayJan	C:\Program	-	 sklearn	0.967	2
<input type="checkbox"/>	✓ 3 minutes ago	10.7s	run_0	KayJan	C:\Program	-	 sklearn	0.633	1

Load more

Hyperparameter Optimization

Which configurations of network give the best performance?

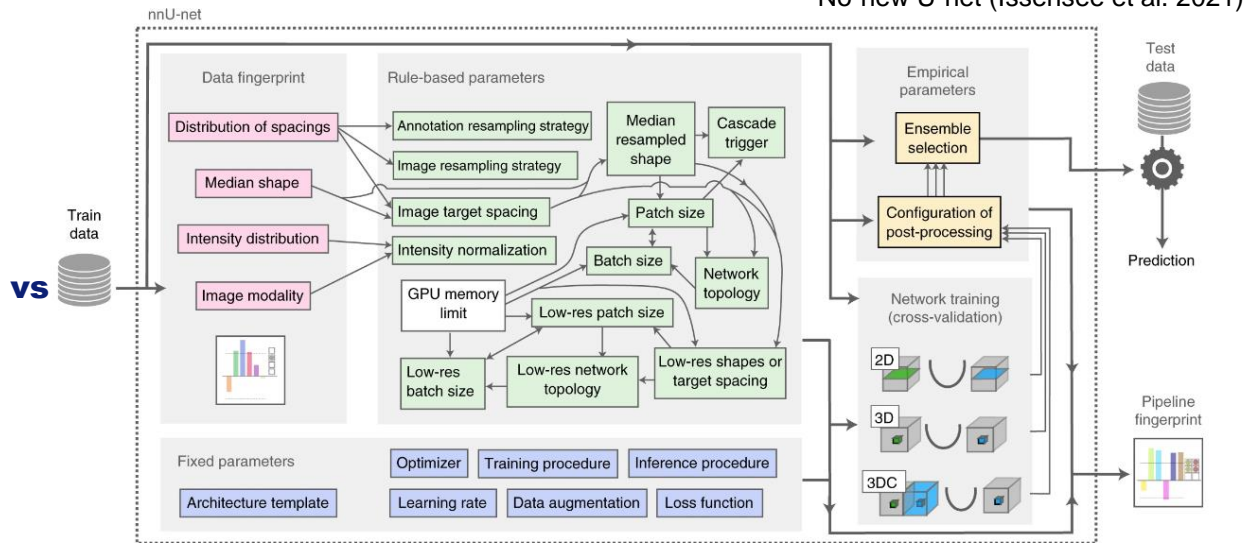
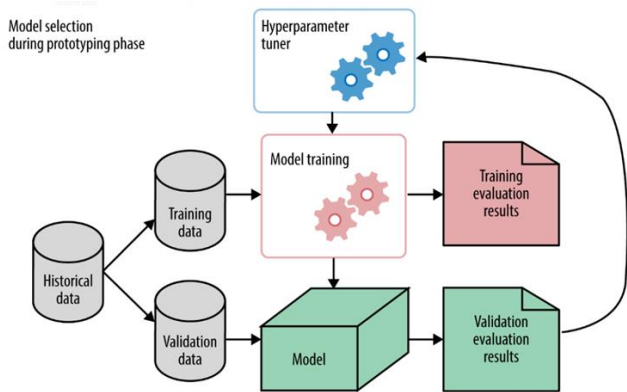


U-net (Ronneberger et al. 2015)

Parameter	Values
Min. Kernels	16, 32, 64, 128
Kernel Size	1, 3, 5, 7, 9
Activations	sigmoid, tanh, relu, elu, PReLU, LeakyReLU, ThresholdedReLU
Initializers	zeros, ones, glorot_normal, he_normal
Regularizers	l1, l2, l1_l2
Dropout Rate	uniform distribution over $[0, 1]$
Learning Rate	uniform distribution over $[10^{-4}, 1]$

Hyperparameter Optimization

Optimization versus logical motivation (versus experimental observation)



Never use test dataset for anything except testing, also not hyperparameter tuning!

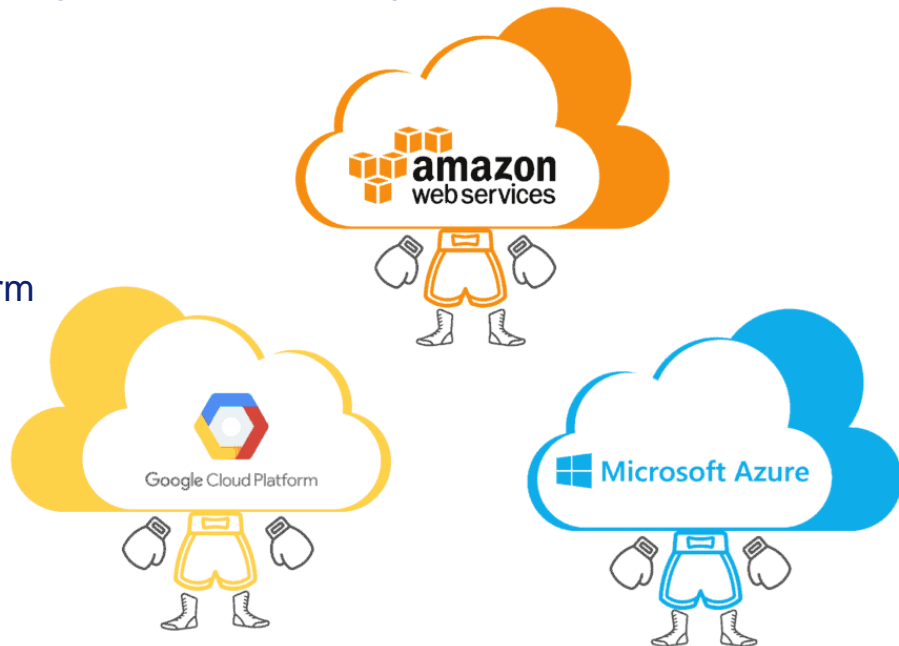
Processing Environments

How to get access to (GPU) resources for processing -> cloud processing environments?

- Google Colab: Free, but limited resources

Three “big” players, but:

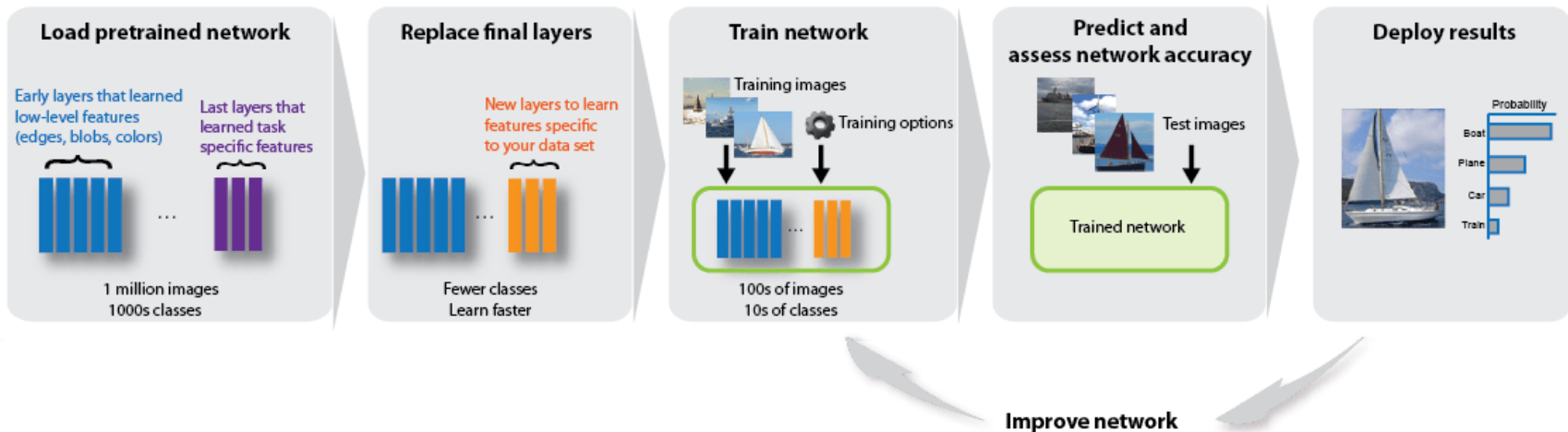
- Pricing per / hour minute
- Hospital may not allow data on a cloud platform



Pre-trained models

Initialize model with weights learned from another dataset:

Reuse Pretrained Network



Potential improvement in efficiency and performance, but no guarantees

- Pretrain on large public datasets, finetune on African datasets?

Open Access Infrastructures

How to combine all these separate tools in an infrastructure?

- No need to re-invent the wheel, use and learn from open access infrastructures!

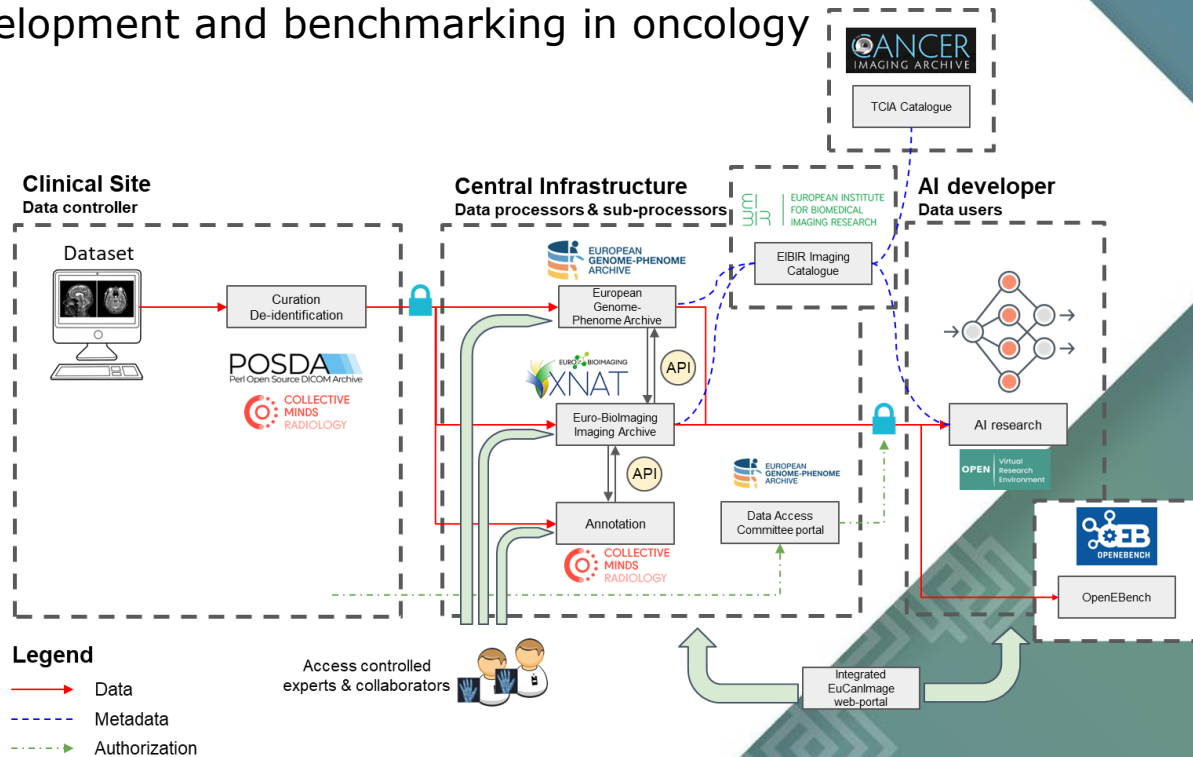




EUCAN
IMAGE

The EuCanImage Platform

- Build a **federated**, **GDPR-compliant**, **scalable** and **FAIR** cancer imaging platform linked to biological and health repositories for integrated multi-scale AI development and benchmarking in oncology



Programming Time!

Goal: get familiar with MONAI, and optionally start applying it to your own data.

Notebooks:

- 2- custom-DL-PyTorch-TorchIO-[optional].ipynb
- 2- TorchIO_MONAI_PyTorch_Lightning-[optional].ipynb
- **3- MONAI.ipynb (main notebook)**
- 4-MONAI_MLFlow-[optional].ipynb
- 7- Pretrained-Models-MONAI_Model_Zoo-[optional].ipynb



Suggestion: check the Readme, start with the main notebook. Afterwards, start running MONAI on your own data (see main notebook), follow other notebooks, or other MONAI tutorials (see Readme).



Programming Time!

Goal: get familiar with MONAI, and optionally start applying it to your own data

Timeline

- 08.30 – 08.45: Introduction in central Zoom call
- 08.45 – 09.50: Practice session in Zoom breakout rooms
- 09.50 – 10.00: Wrap-up and discussion in central Zoom call

Note: you can pick your own breakout room this time!



AFRICAI Summer School

MODEL DEVELOPMENT 2: MODEL-CENTRIC BEST PRACTICES, PITFALLS, AND OPEN ACCESS INFRASTRUCTURES



Dr. ir. Martijn P. A. Starmans, PhD

*PostDoc AI for Integrated Diagnostics in Oncology
Department of Radiology and Nuclear Medicine &
Department of Pathology, Erasmus MC, Rotterdam, NL*

✉ m.starmans@erasmusmc.nl

<https://mstarmans91.github.io/>

EURO BIOIMAGING

eosc cancer

EUCAN
IMAGE

radioval EUCAIM

Erasmus MC
University Medical Center Rotterdam

Erasmus