The multi-dimensional aspect of uncertainty

Michel Dojat

The multi-dimensional aspect of uncertainty

Michel DOJAT

Research Director Inserm
Deputy Scientific Director Inria

michel.dojat@inria.fr

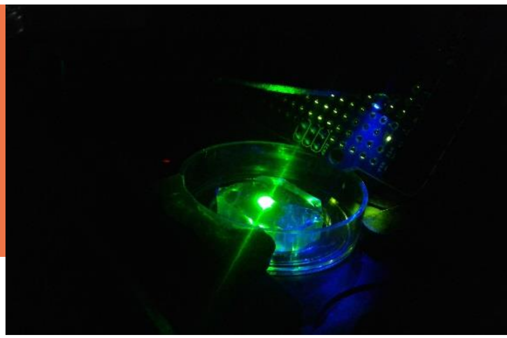
I declare relationships with the company Pixyl (pixyl.ai)

My main collaborators

Benjamin Lambert

Florence Forbes

RESEARCH TOPICS AND EXPERIMENTAL APPROACHES



/ Fundamental neurosciences

Cytoskeleton, Intracellular traffic, Synaptic plasticity, Mechanisms studied in normal and pathological conditions (neurobiological diseases, neurodegenerative diseases, myopathies)

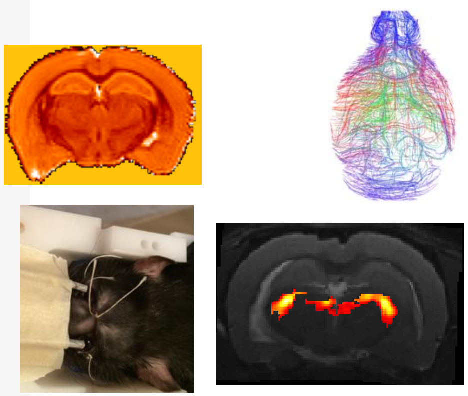
/ Pre-clinical and clinical research

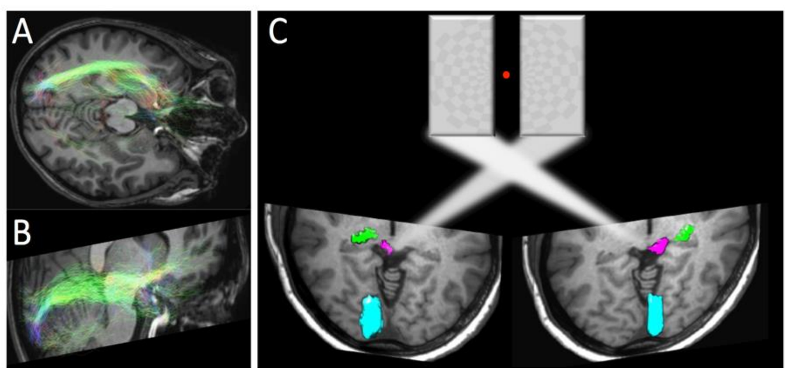
Developing tools and concepts

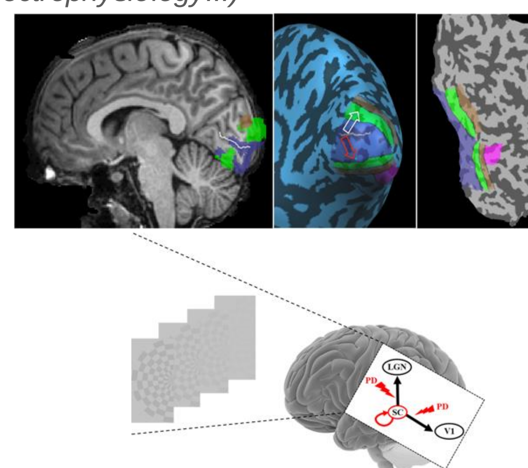
Close links with networks such as GREEN, Neuropsynov, NeuroCoG...

/ Innovative technologies and treatments

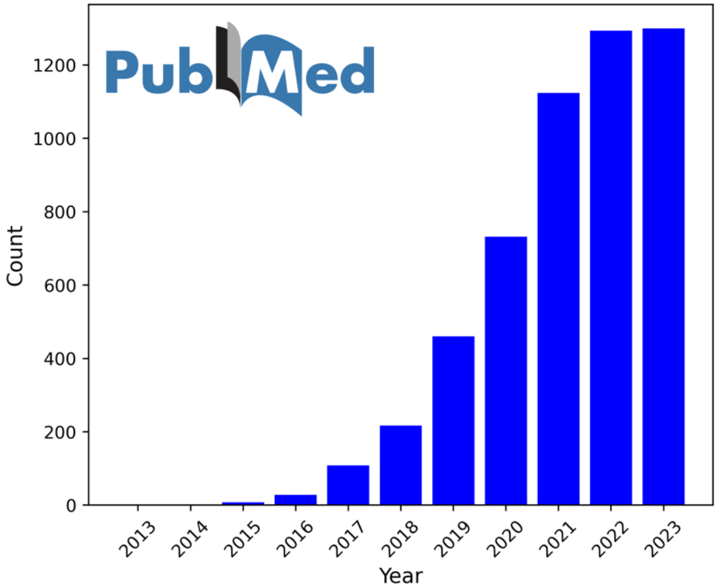
Multidisciplinary approaches including human social sciences, and methodological developments (optogenetics, reconstruction of neural networks, electrophysiology...)



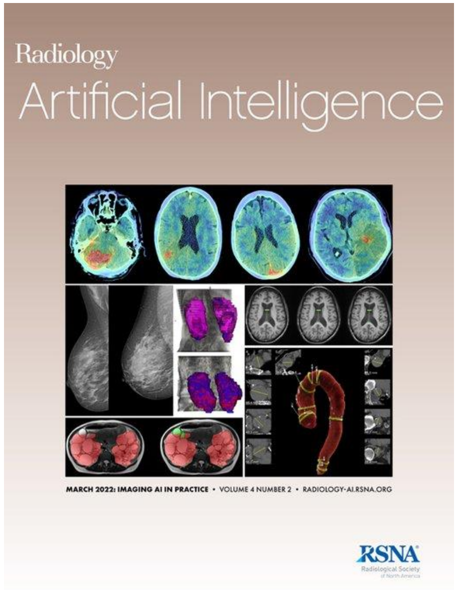




Automatic analysis of medical images

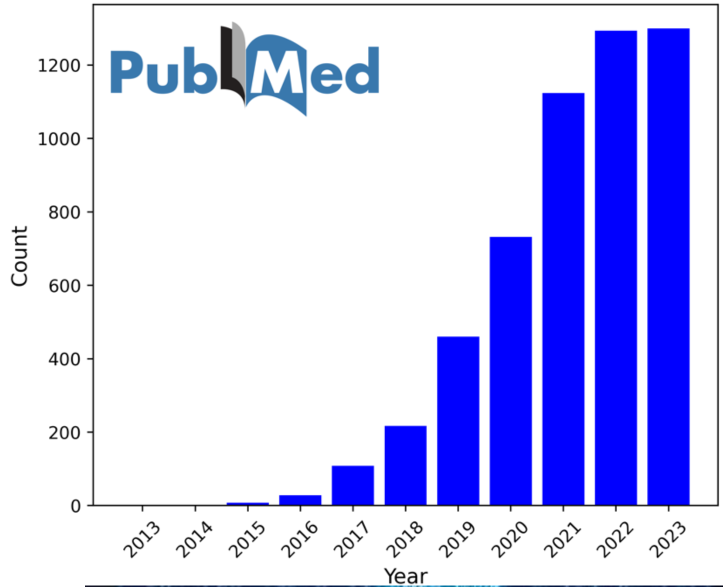


MIDL
Paris 2024

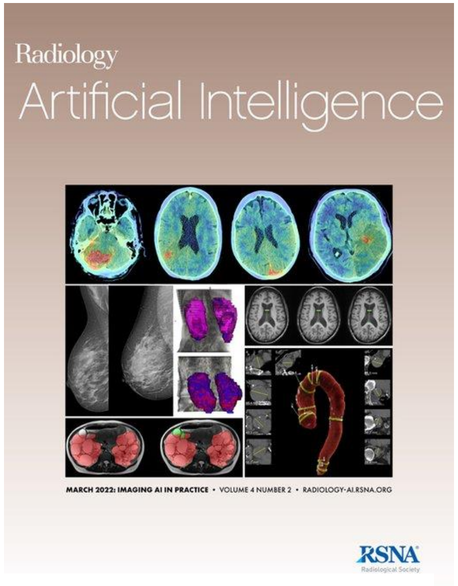


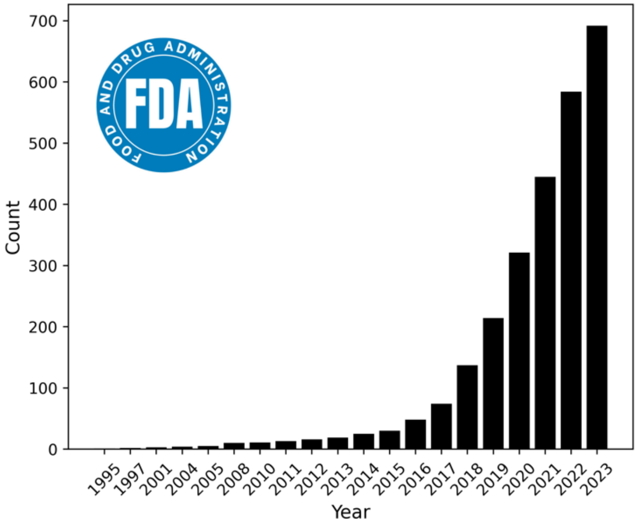
Kw: « Deep Learning » and « Medical Image Analysis »

Automatic analysis of medical images



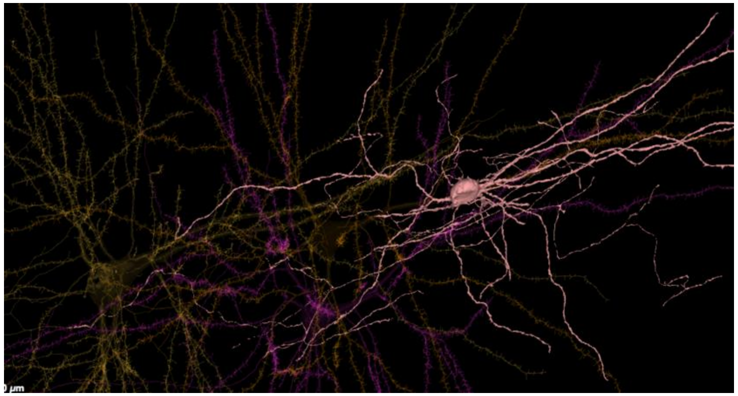
MIDL
Paris 2024



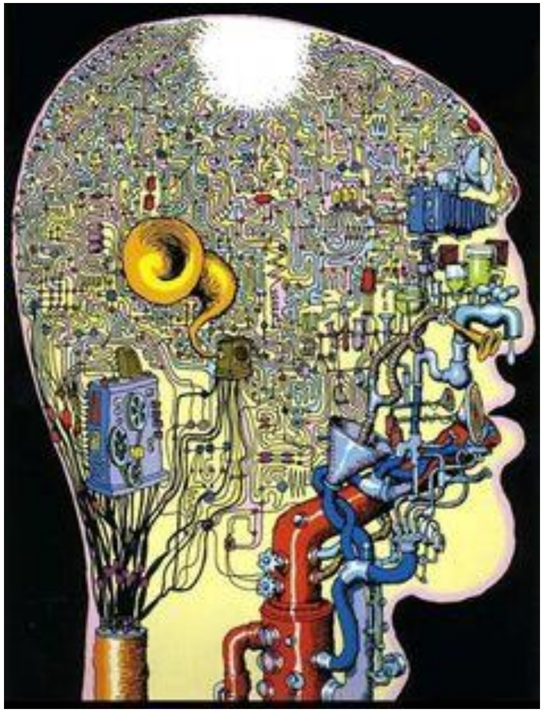


The curse of the Black-Box

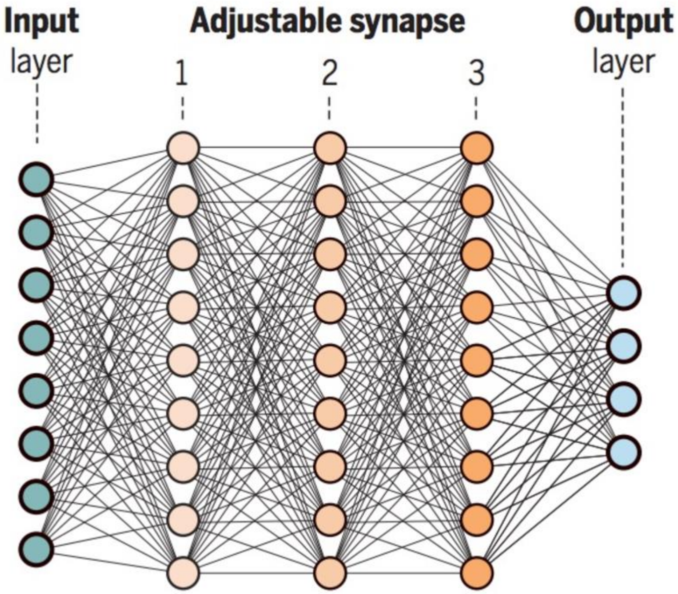
The curse of the Black-Box



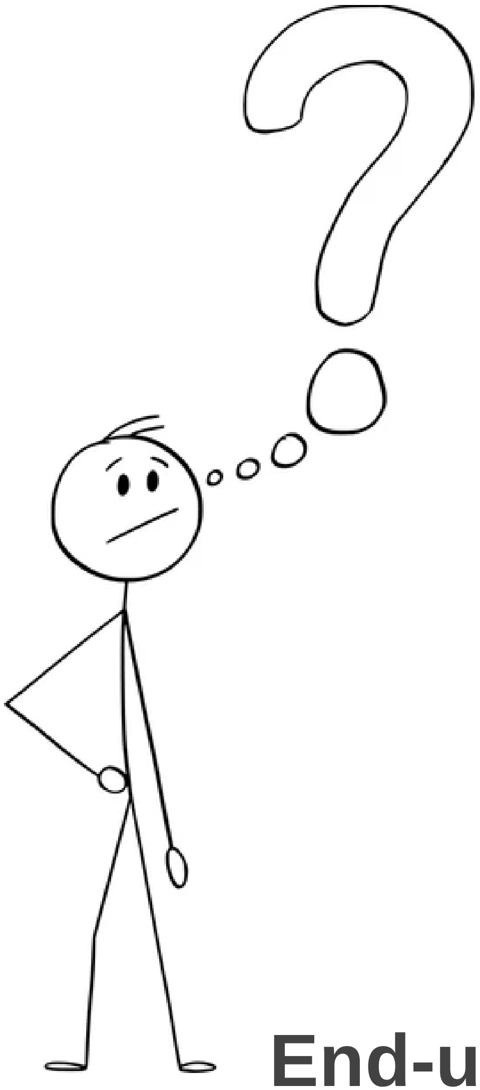
[Shapson-Coe et al. Science 384, 635 (2024)]

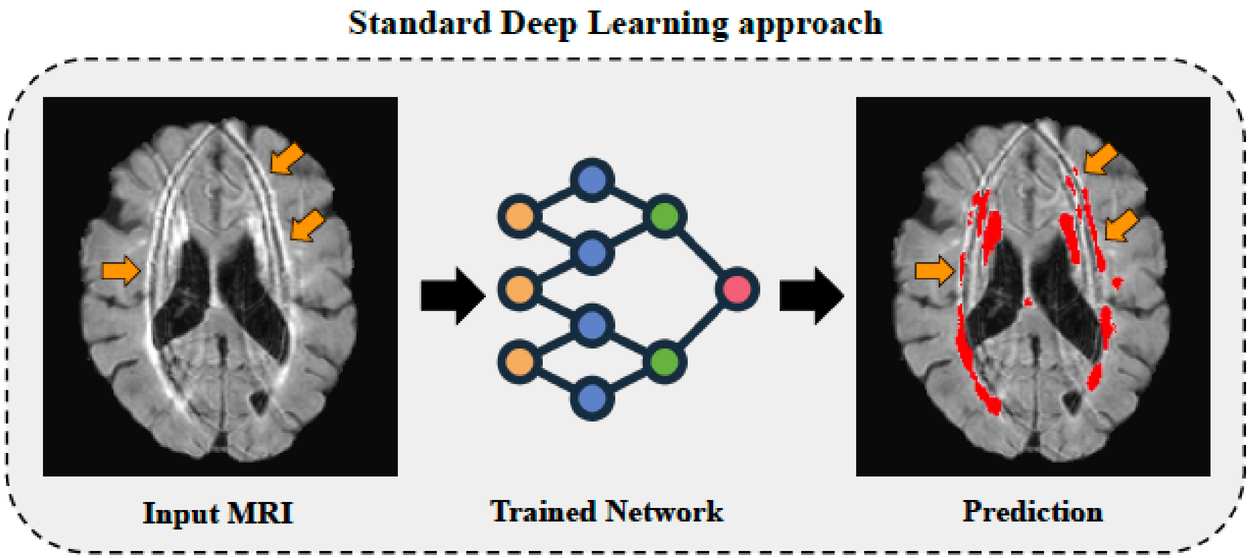


[Crumb, Says 1967 Num 1]



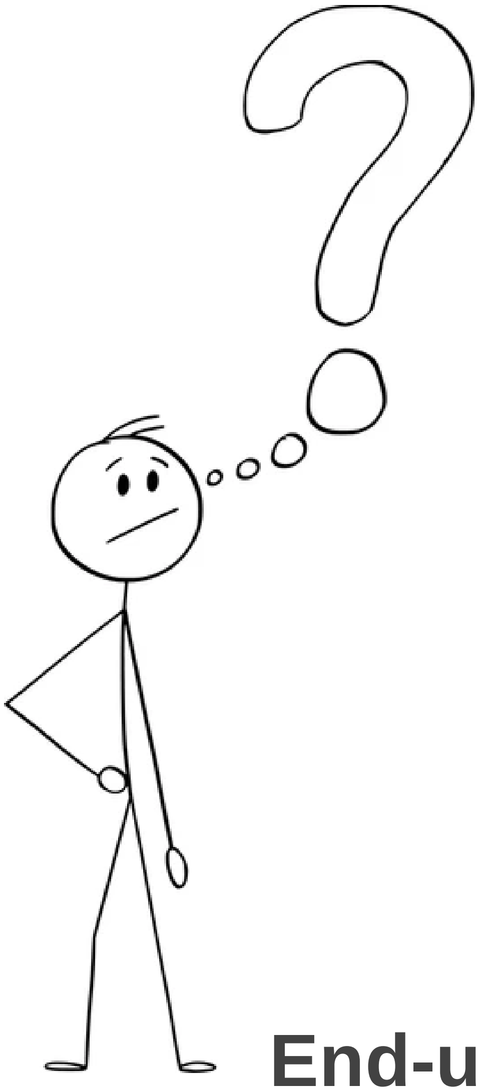
The curse of the Black-Box





End-user

The curse of the Black-Box

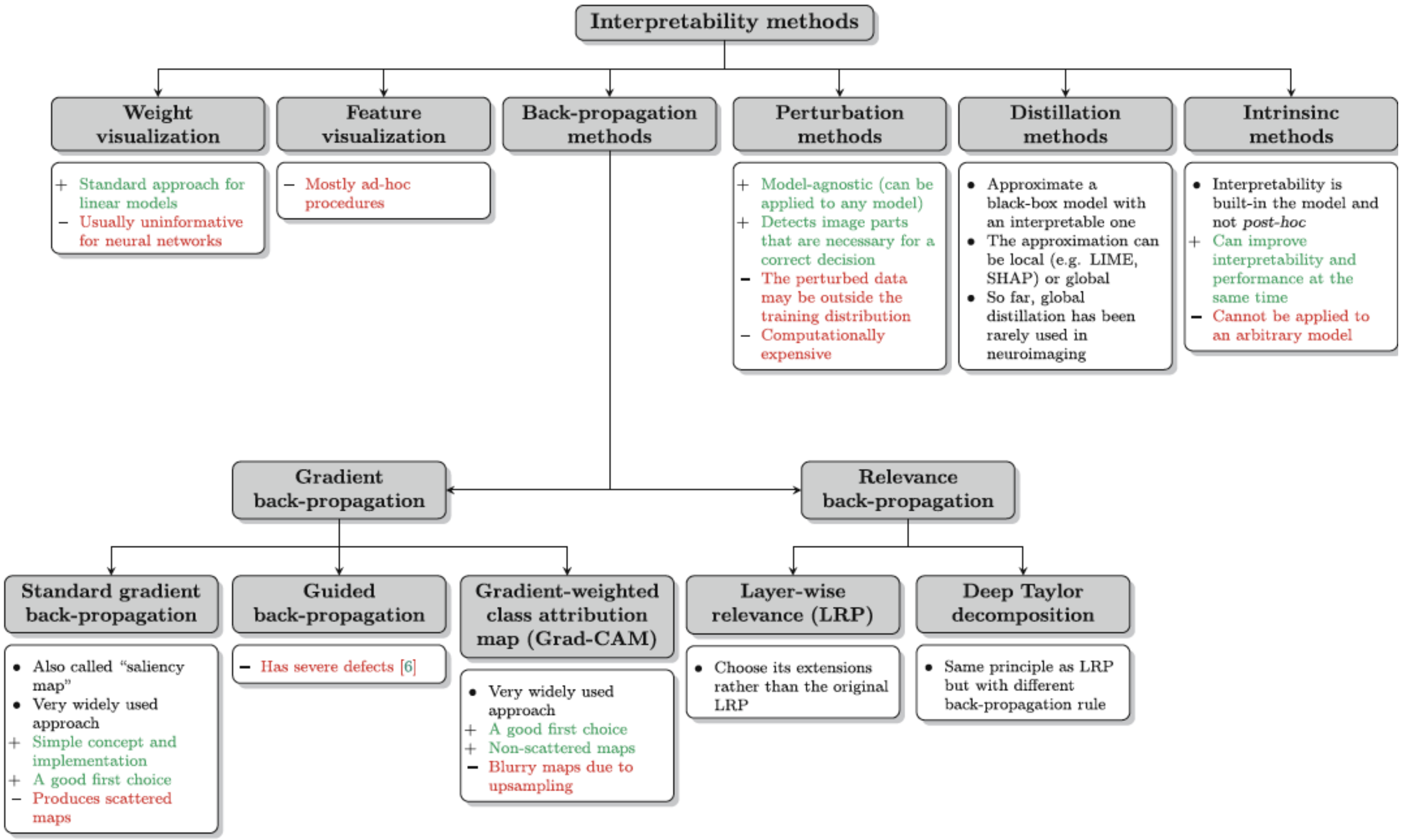


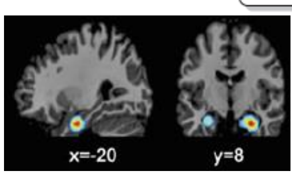
Explainability (XAI)
Interpretability
Understandability
[Erasmus et al 2021 Philosophy & Technology]

For who?
About what?
At which level?
…

End-user

Useful approaches : but mainly for developers …

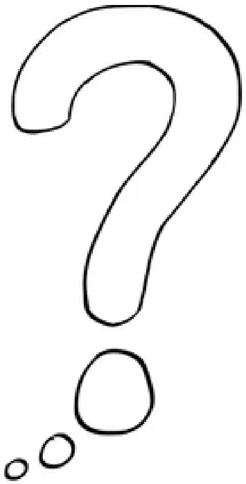




AD vs Crtl

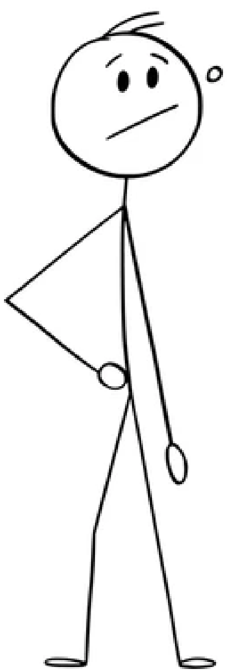
[Thibau-Sutre et al Neuromethods vol 197 2023]

The curse of the Black-Box



~~Explainability (XAI)~~
~~Interpretability~~
~~Understanding~~

For who?
About what?
At which level?
…

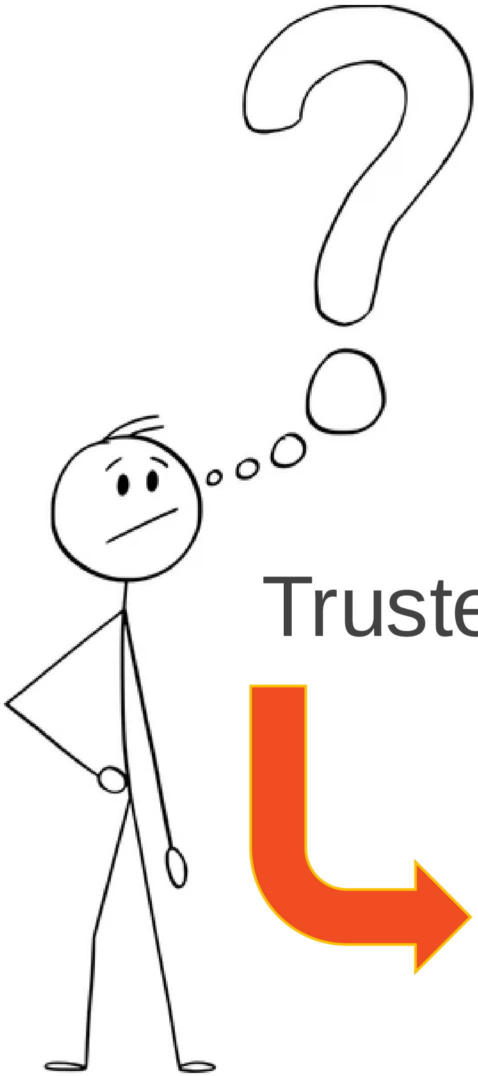


To Improve Confidence
To trust in AI

Rigorous Validation
Usage conditions
Adverse effects

….

The curse of the Black-Box



~~Explainability~~
~~Interpretabilty~~

For who?
About what?
What contents?
...

Trusted AI

G. Hinton [In response to how do we trust systems?] ***You should regulate them based on how they perform.***

Uncertainty Quantification in NN decision

Uncertainty: A keypoint for AI researchers-I

ERC survey (2023): the use of AI for data analysis and processing

- 1034 ERC (/14829)
- Kw: AI in Title, abstract and kw

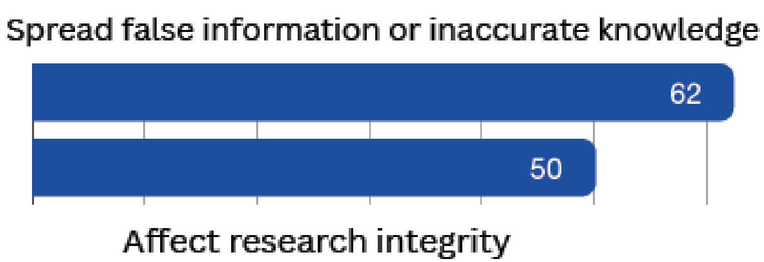
« the current **underdevelopment of uncertainty quantification**, that is, the assessment of the reliability of models and simulations, and also concerns over the transparency of AI systems. »

https://erc.europa.eu/sites/default/files/2023-12/AI_in_science.pdf

Uncertainty: A keypoint for AI researchers-II

ERC survey (2023): the use of AI for data analysis and processing
- 1034 ERC (/14829)
- Kw: AI in Title, abstract and kw

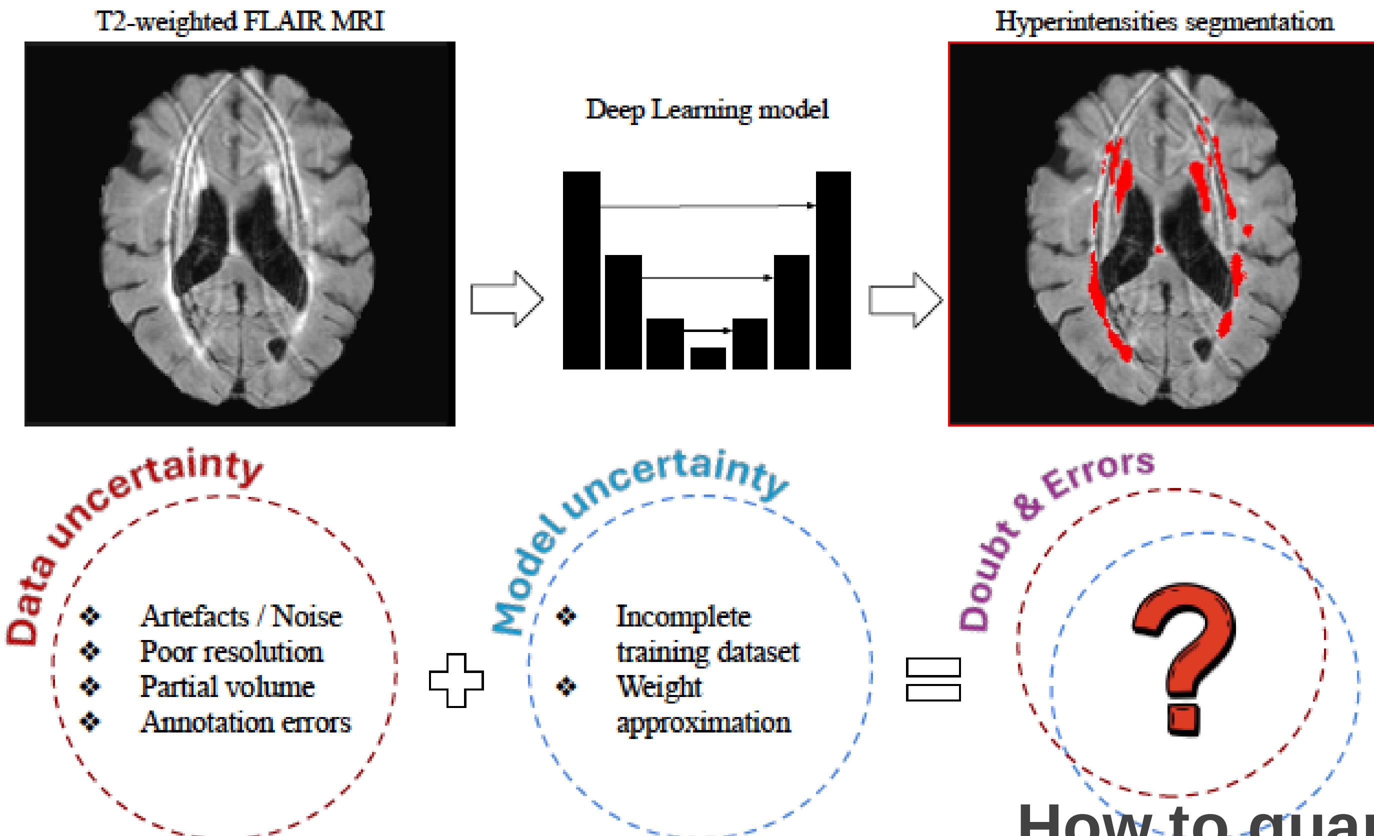
« the current **underdevelopment of uncertainty quantification**, that is, the assessment of the reliability of models and simulations, and also concerns over the transparency of AI systems. »



https://www.certain-trust.eu

https://erc.europa.eu/sites/default/files/2023-12/AI_in_science.pdf

Where is uncertainty (UC) hidden?



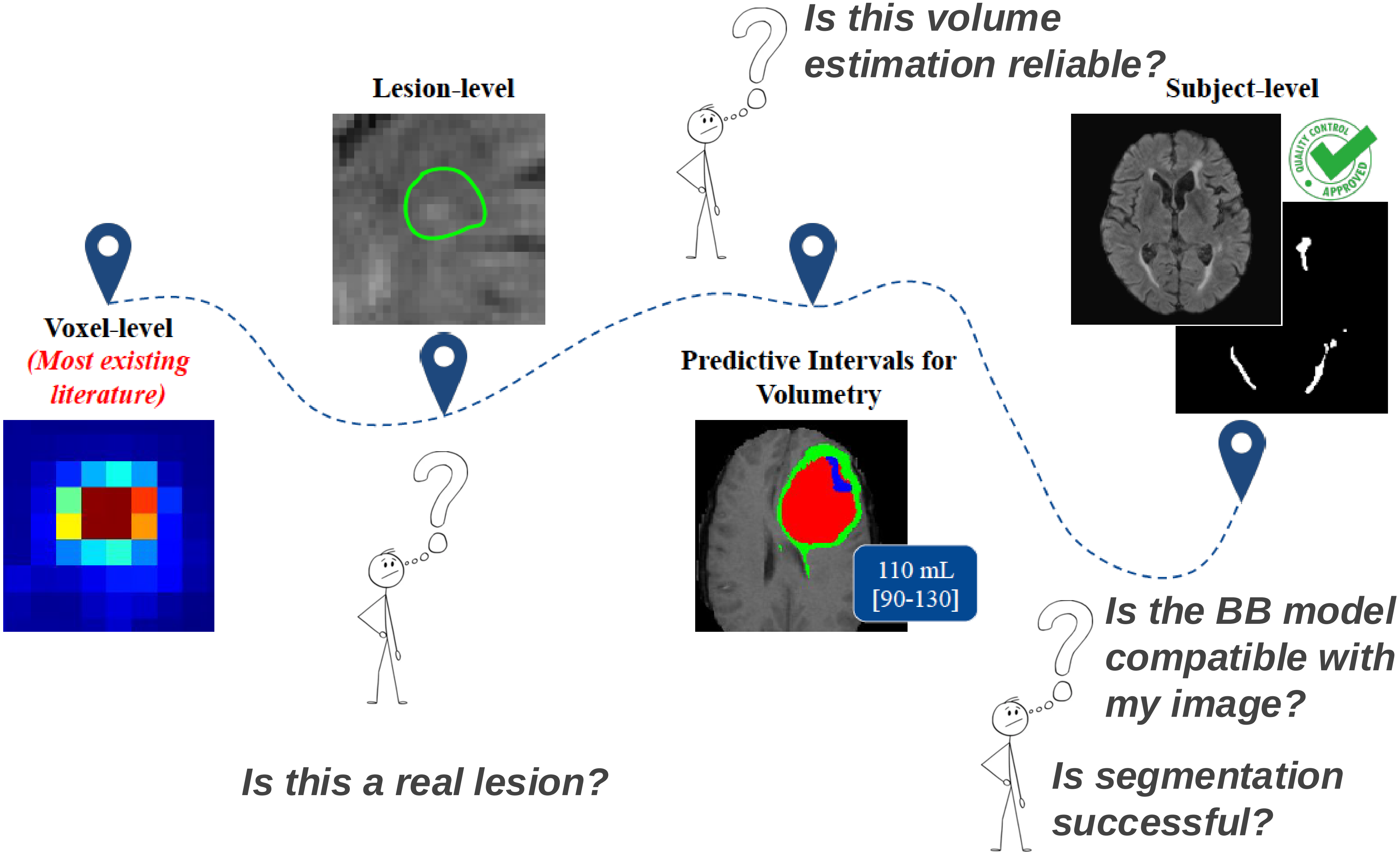
How to quantify UC?

Aleatoric UC

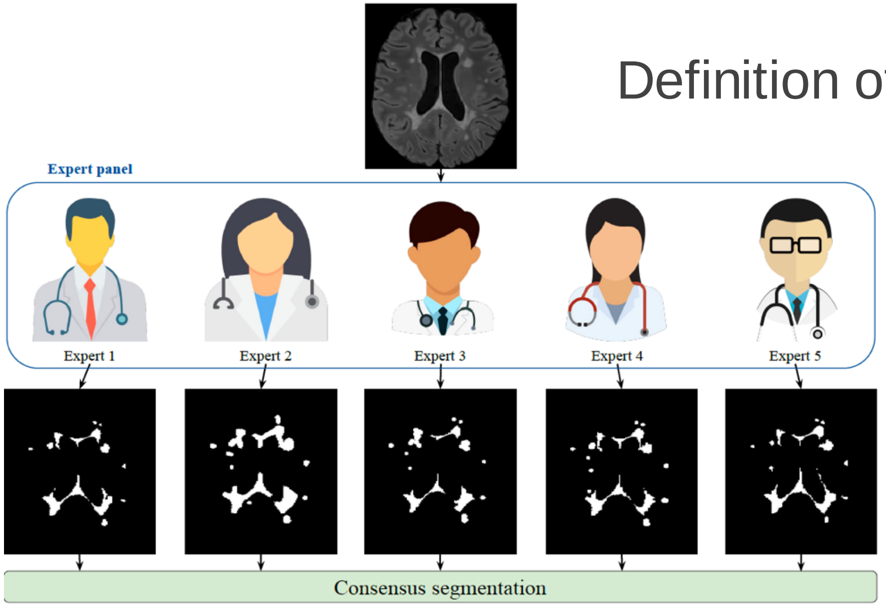
Epistemic UC

[Kendall & Gal 2017 Adv Neural Inf Process Syst]

The multi-dimensional aspect of uncertainty

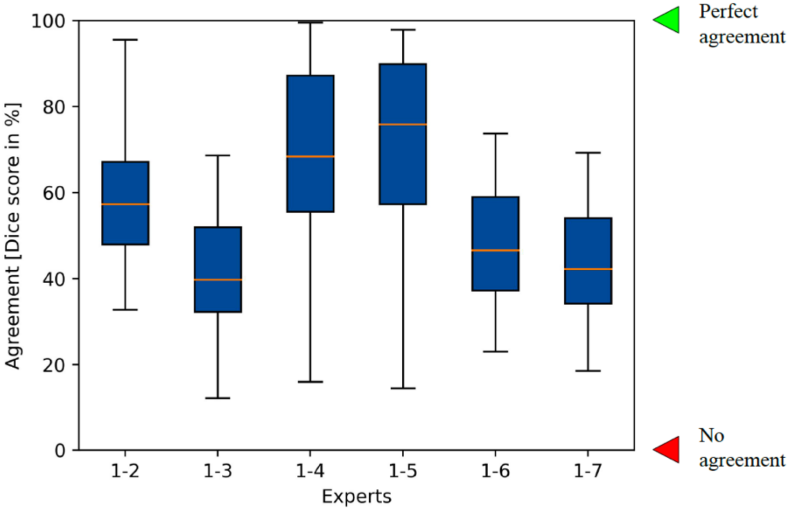


Consensus meeting



Definition of ground truth

Inter-expert variability



[Commowick et al NeuroImage 2021]

Existing works

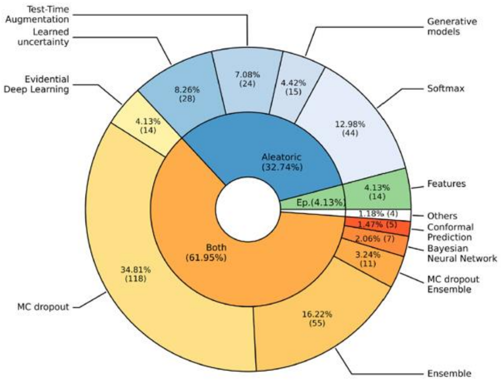
218 papers reviewed
KW: Deep Learning + Medical Image + Classification + Segmentation + Uncertainty

↓

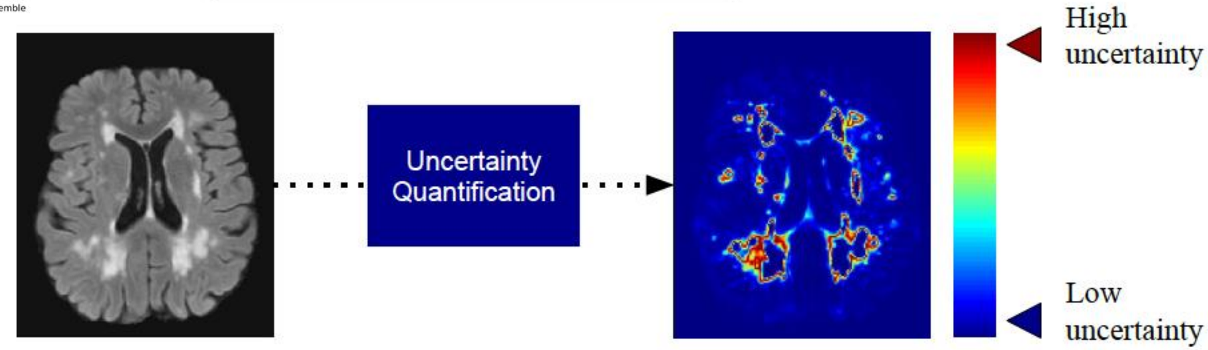
11 frameworks identified

↓

9/11 provide voxel-level uncertainty estimates

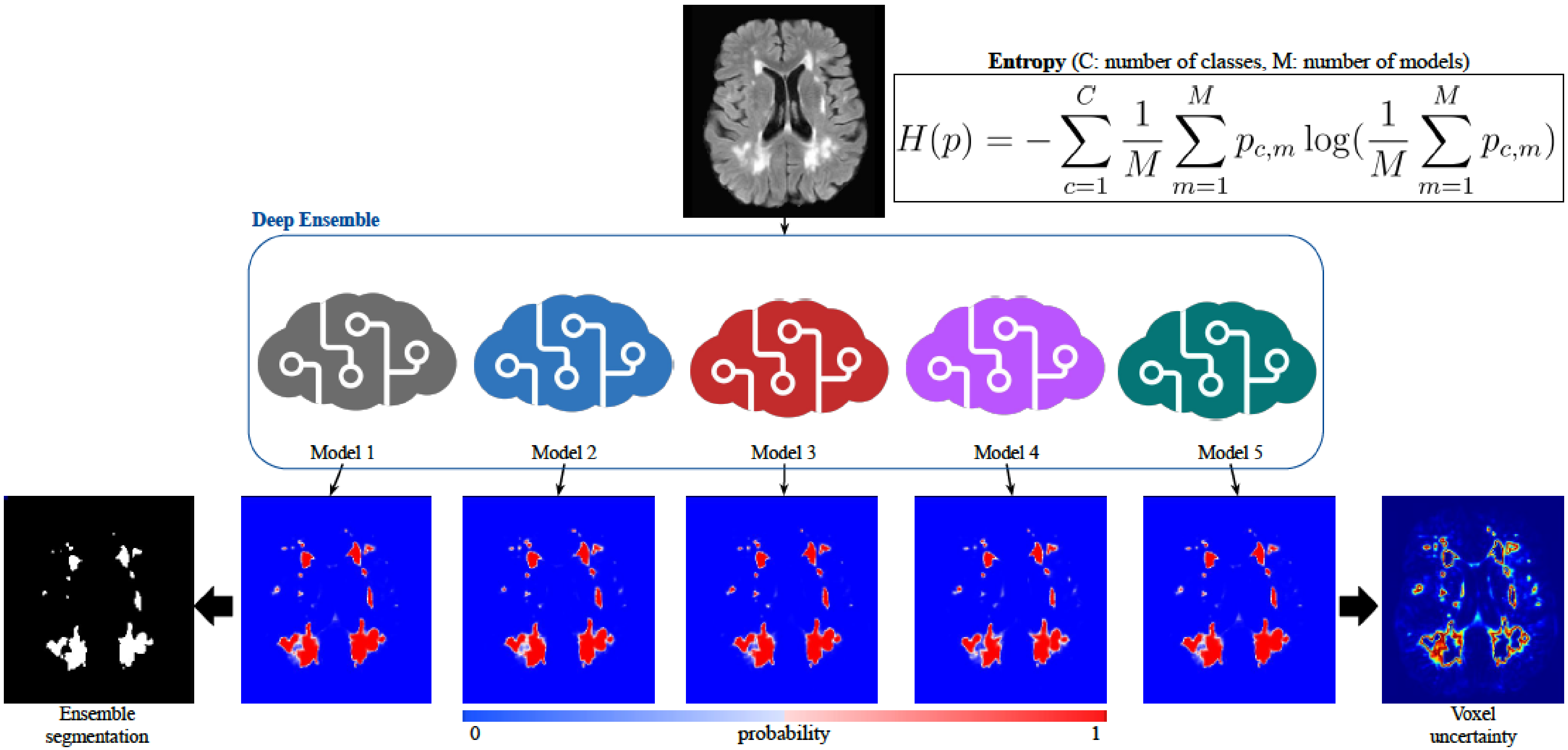


SEP lesions
in a FLAIR image

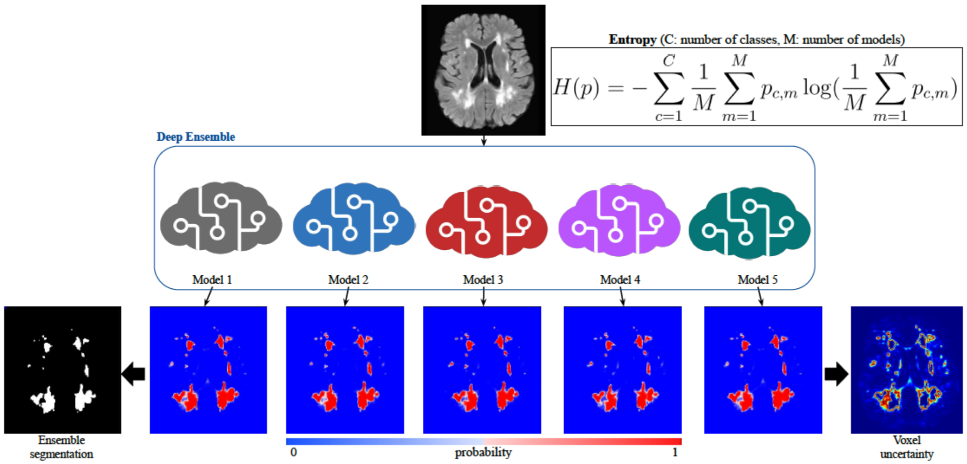


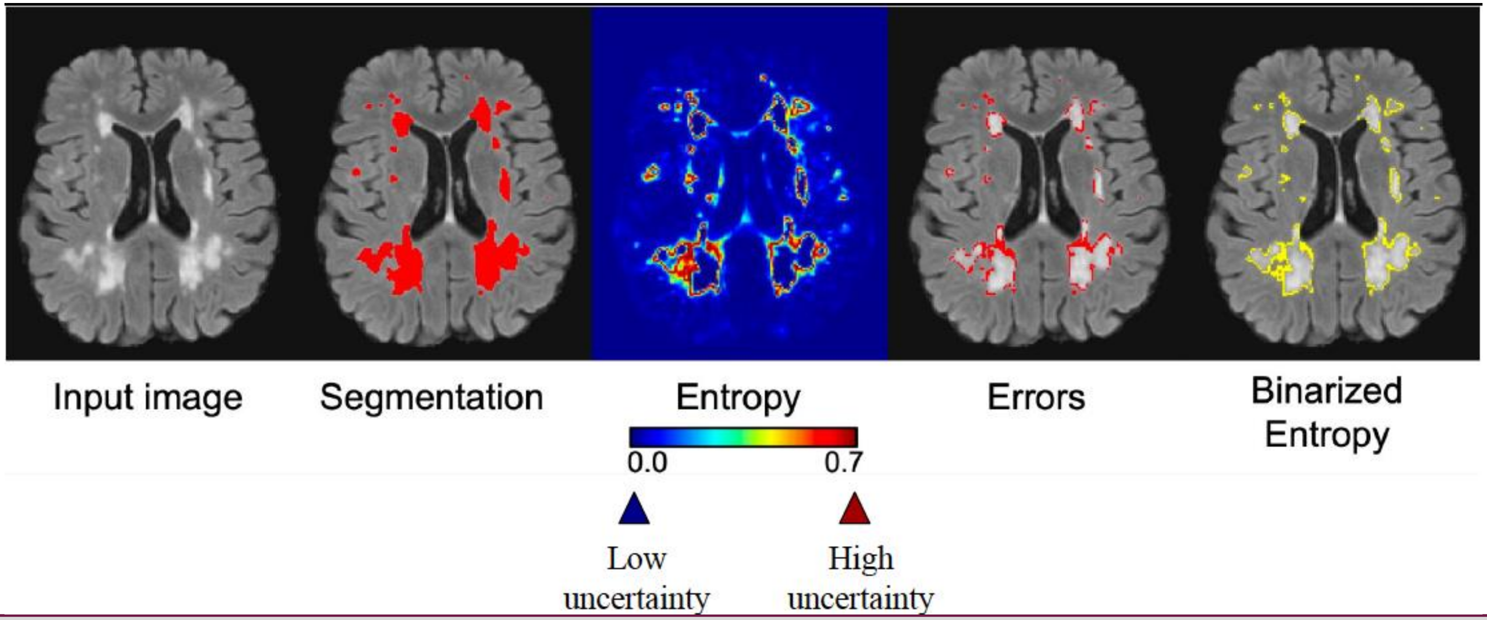
[Lambert et al. AIM 2024.
Trustworthy clinical AI solutions: A unified review of uncertainty quantification in Deep Learning models for medical image analysis]

Voxel level UQ: Ensembling



Voxel level UQ: Ensembling

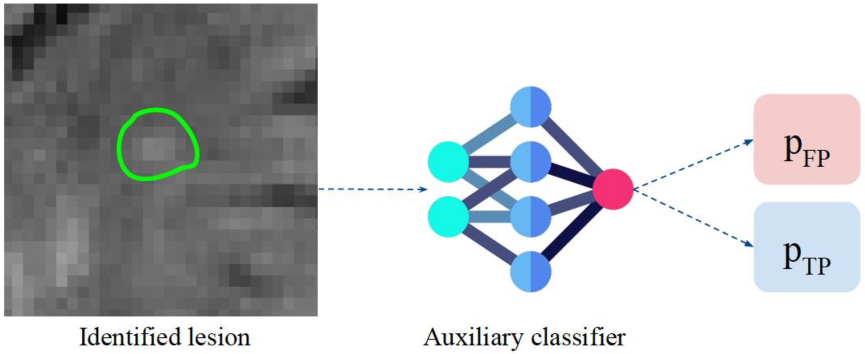




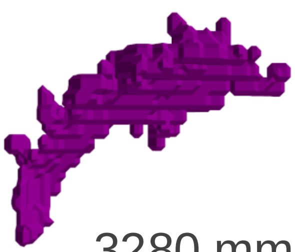
Lesion level UQ

Goal: Identify FP

Estimate P_{FP}: proba that the lesion is FP using an auxillary classifier



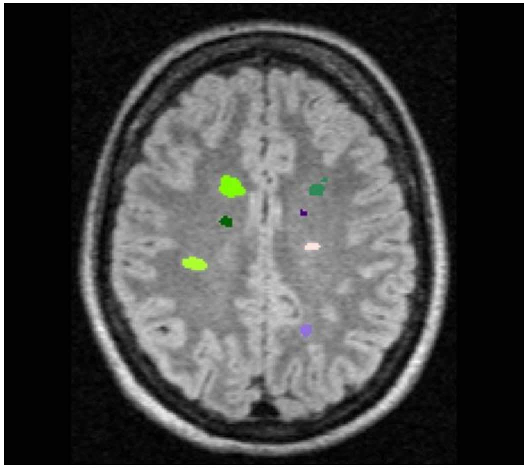
Note: Lesion are highly variable **in shape**



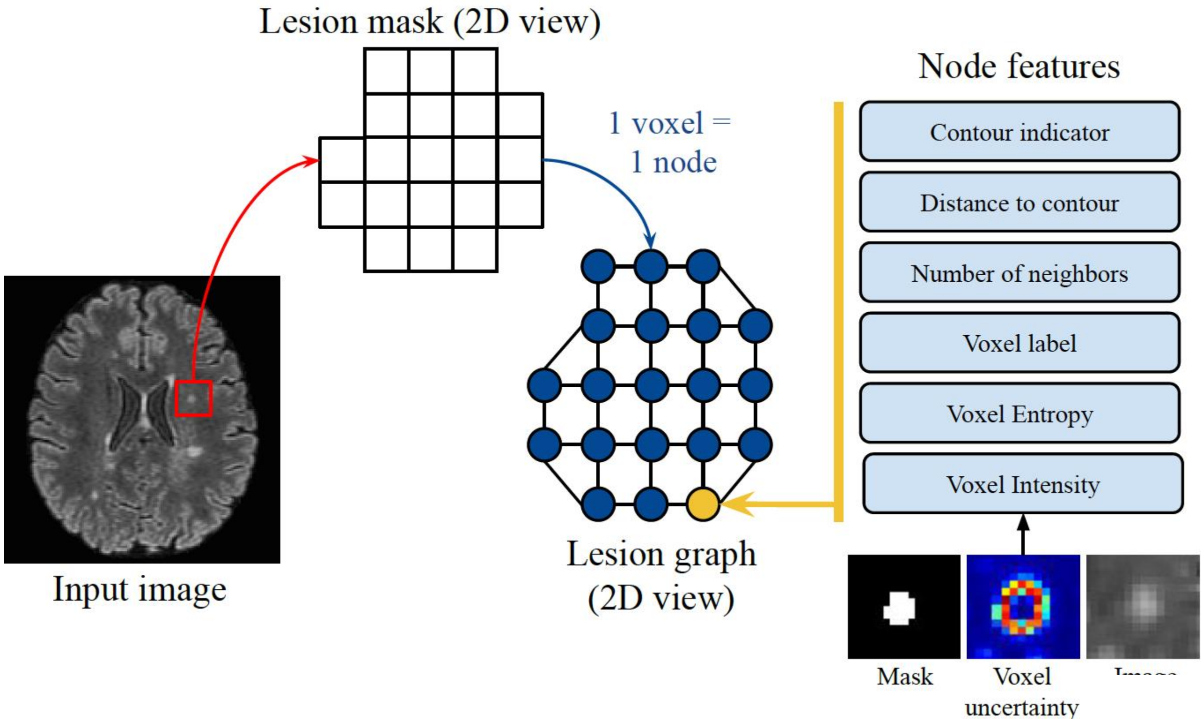
3280 mm^3

71 mm^3

and **position**

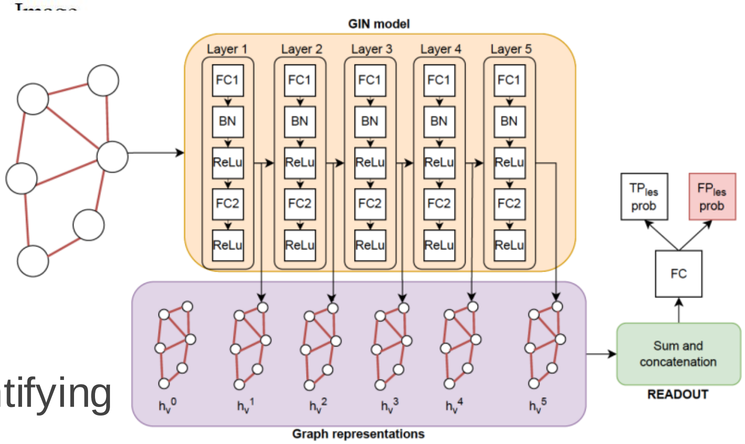


Lesions as Graphs



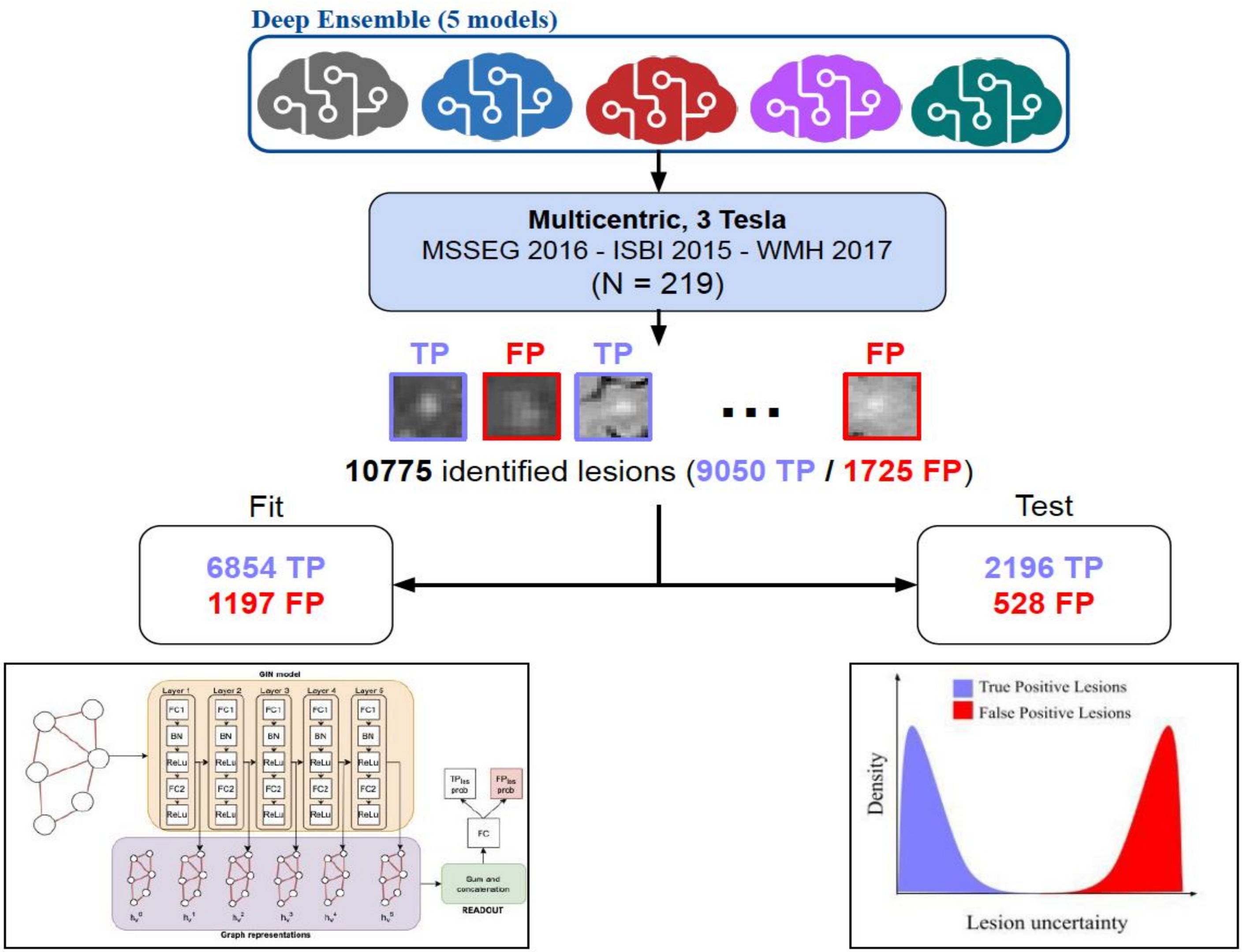
Graph classification

Training to separate
True and False Positive lesions



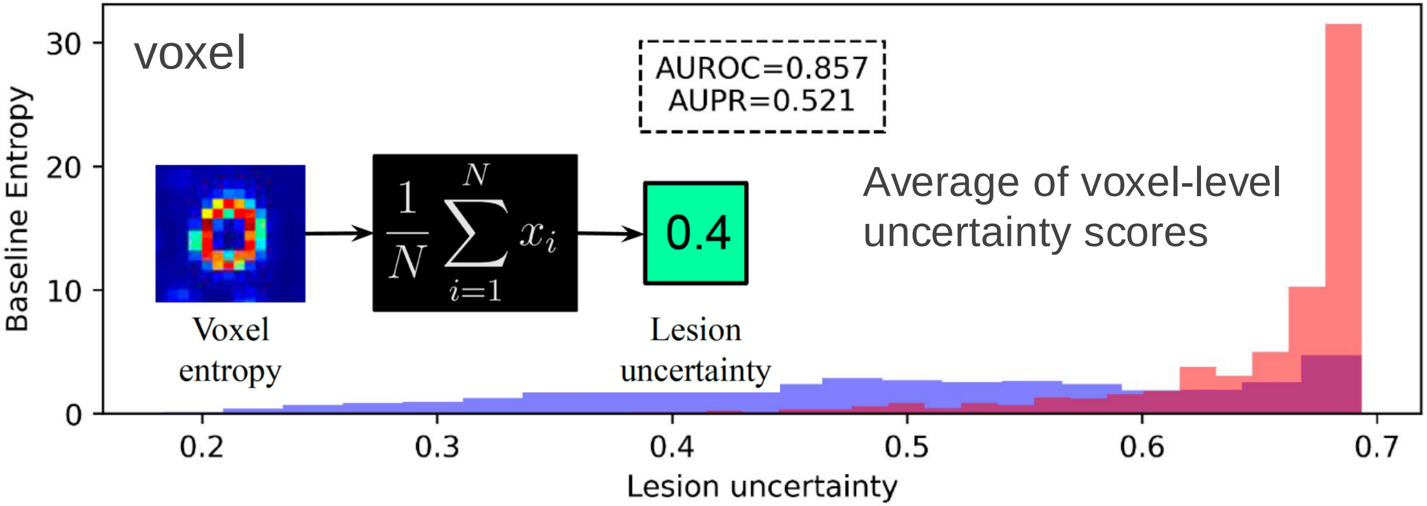
Lambert B et al. Beyond Voxel Prediction Uncertainty: Identifying brain lesions you can trust 2022, LNCS 13611 pp. 61-70.

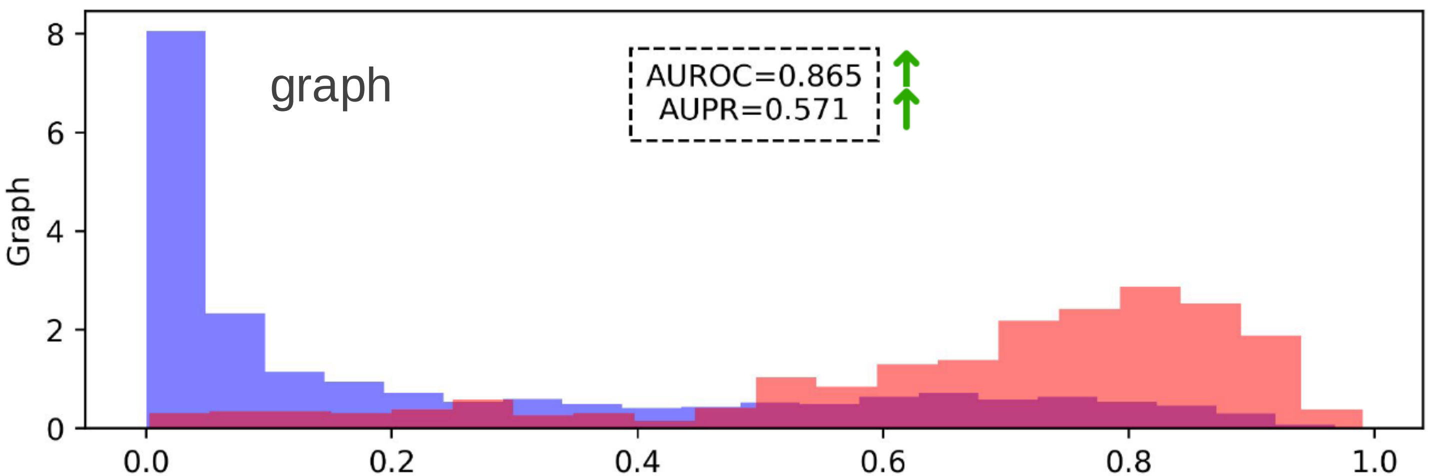
The Graph Isomorphism Network



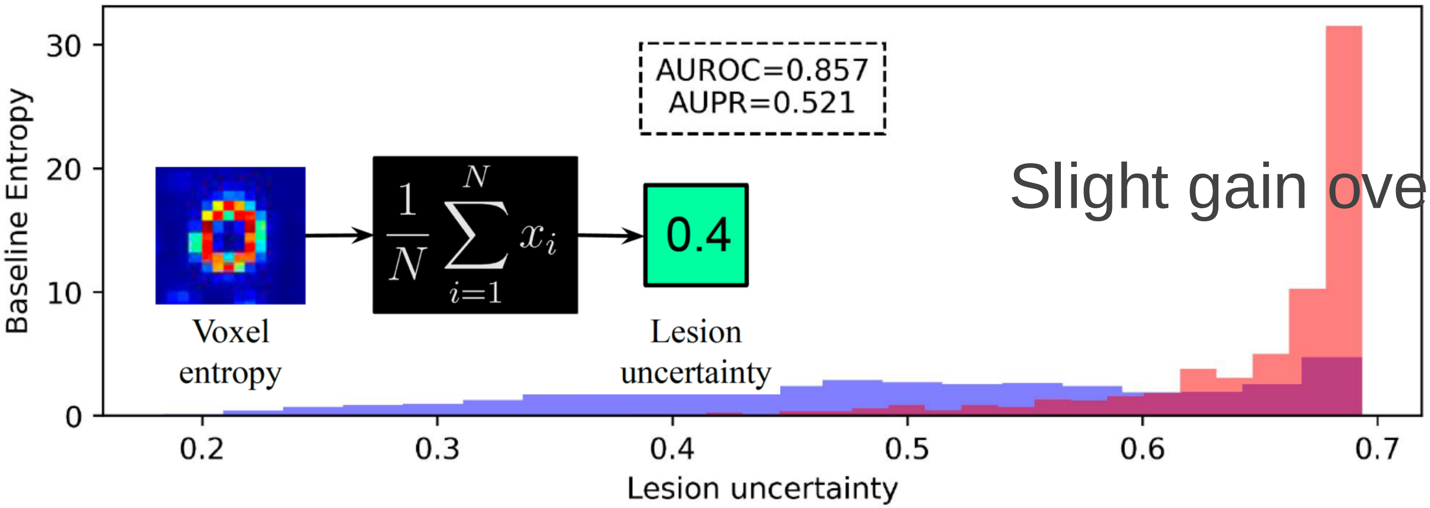
FC: Fully-connected layer. BN: Batch Normalization. ReLU: Rectified Linear Unit. Parameters: 26 700

The Graph Isomorphism Network



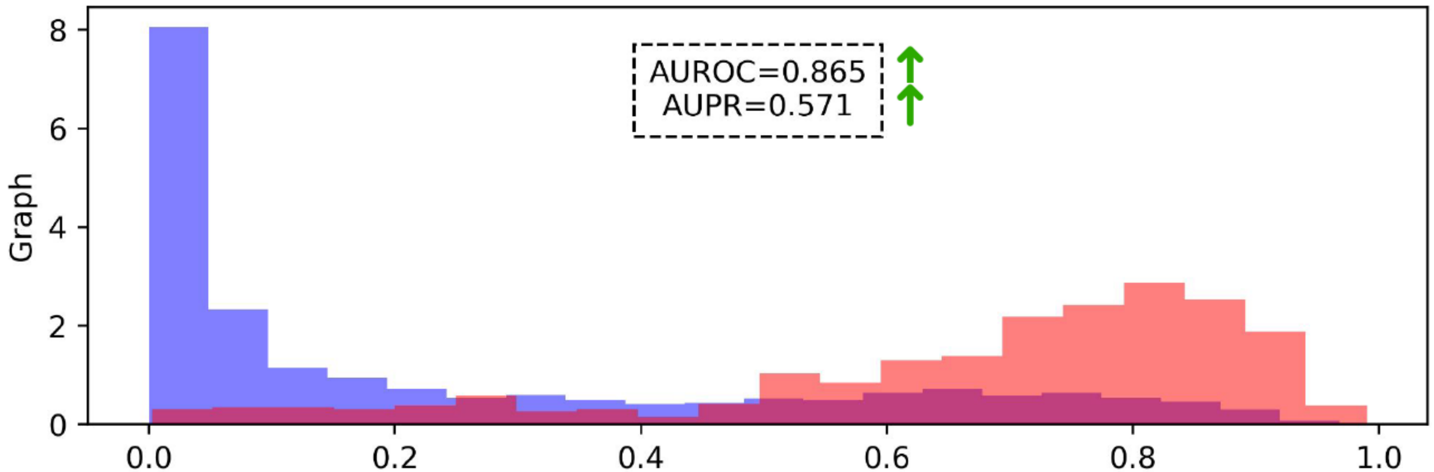


The Graph Isomorphism Network



Slight gain over the baseline

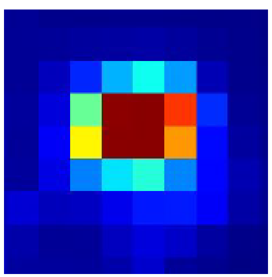
Score easily interpretable



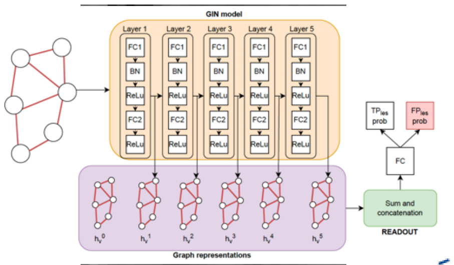
The multi-dimensional aspect of uncertainty

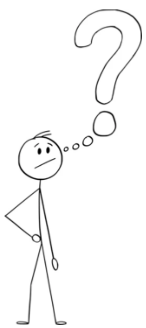
Voxel-level

(Most existing literature)

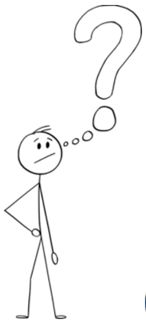


Lesion-level



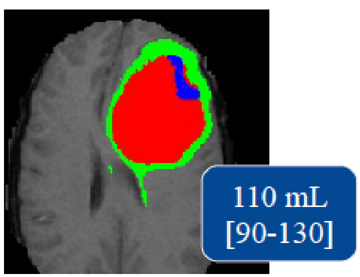


Is this a real lesion?

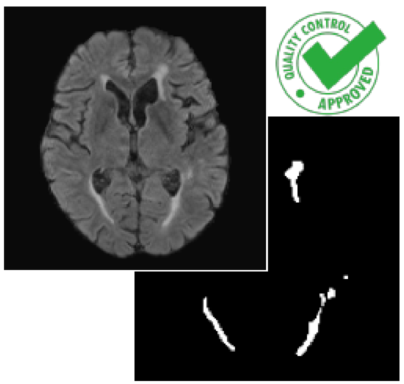


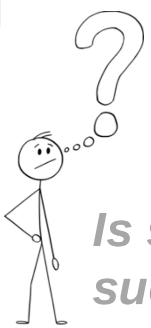
Is this volume estimation reliable?

Predictive Intervals for Volumetry



Subject-level





Is the BB model compatible with my image?

Is segmentation succesful?

Predictive intervals in volumetry

Definition

- $X \in \mathbb{R}^{N-1}$ are estimates of the true volumes $Y \in \mathbb{R}^{N-1}$, obtained from the segmentation.
- A predictive interval $\Gamma_\alpha(X)$ is a range of values intended to encompass Y with a specified degree of confidence $1-\alpha$ (e.g. 90%, 95%), so that $P(Y \in \Gamma_\alpha(X)) \geq 1-\alpha$

Sampling-based approaches

- Sample a set of estimated volumes $X_1, \ldots, X_K$ for the given image.
- Estimate the mean $\mu(X)$ and standard deviation $\sigma(X)$.
- Assuming $Y|X \sim \mathcal{N}(\mu(X), \sigma(X))$:

$$\Gamma_\alpha(X) = [\mu(X) - z\sigma(X), \mu(X) + z\sigma(X)] \quad (1)$$

Predictive intervals in volumetry

Definition

- $X \in \mathbb{R}^{N-1}$ are estimates of the true volumes $Y \in \mathbb{R}^{N-1}$, obtained from the segmentation.
- A predictive interval $\Gamma_\alpha(X)$ is a range of values intended to encompass Y with a specified degree of confidence $1-\alpha$ (e.g. 90%, 95%), so that $P(Y \in \Gamma_\alpha(X)) \geq 1-\alpha$

Sampling-based approaches

- Sample a set of estimated volumes $X_1, ..., X_K$ for the given image.
- Estimate the mean $\mu(X)$ and standard deviation $\sigma(X)$.
- Assuming $Y|X \sim \mathcal{N}(\mu(X), \sigma(X))$

$\Gamma_\alpha(X) = [\mu(X) - z\sigma(X), \mu(X) + z\sigma($

Limitations

- Inference time, due to the sampling procedure.
- The normality assumption, which may not always hold.
- Lack of flexibility, as intervals are symmetrical by design.

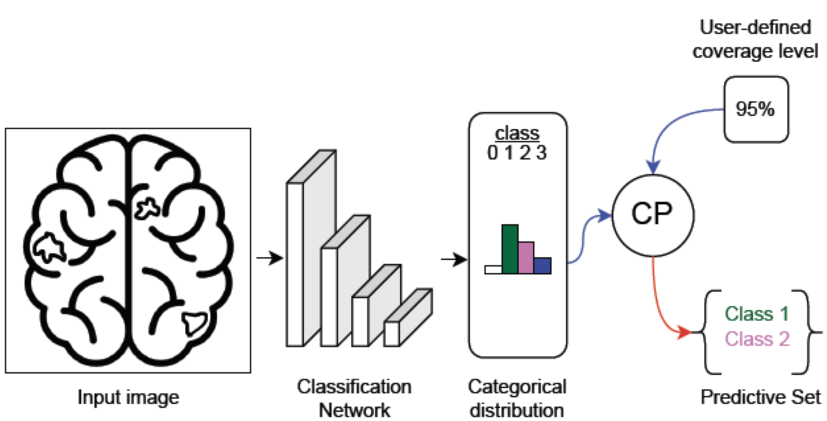
Predictive intervals in volumetry

Definition

- $X \in \mathbb{R}^{N-1}$ are estimates of the true volumes $Y \in \mathbb{R}^{N-1}$, obtained from the segmentation.
- A predictive interval $\Gamma_\alpha(X)$ is a range of values intended to encompass Y with a specified degree of confidence $1-\alpha$ (e.g. 90%, 95%), so that $P(Y \in \Gamma_\alpha(X)) \geq 1-\alpha$

Sampling-based approaches

- Sample a set of estimated volumes $X_1, ..., X_K$ for the given image.
- Estimate the mean $\mu(X)$ and standard deviation $\sigma(X)$.



Direct approaches

- Directly estimate the quantiles $\hat{t}_{\alpha/2}(X)$ and $\hat{t}_{1-\alpha/2}(X)$.
- The PI is computed as:

$$\Gamma_\alpha(X) = [\hat{t}_{\alpha/2}(X), \hat{t}_{1-\alpha/2}(X)] \qquad (2)$$

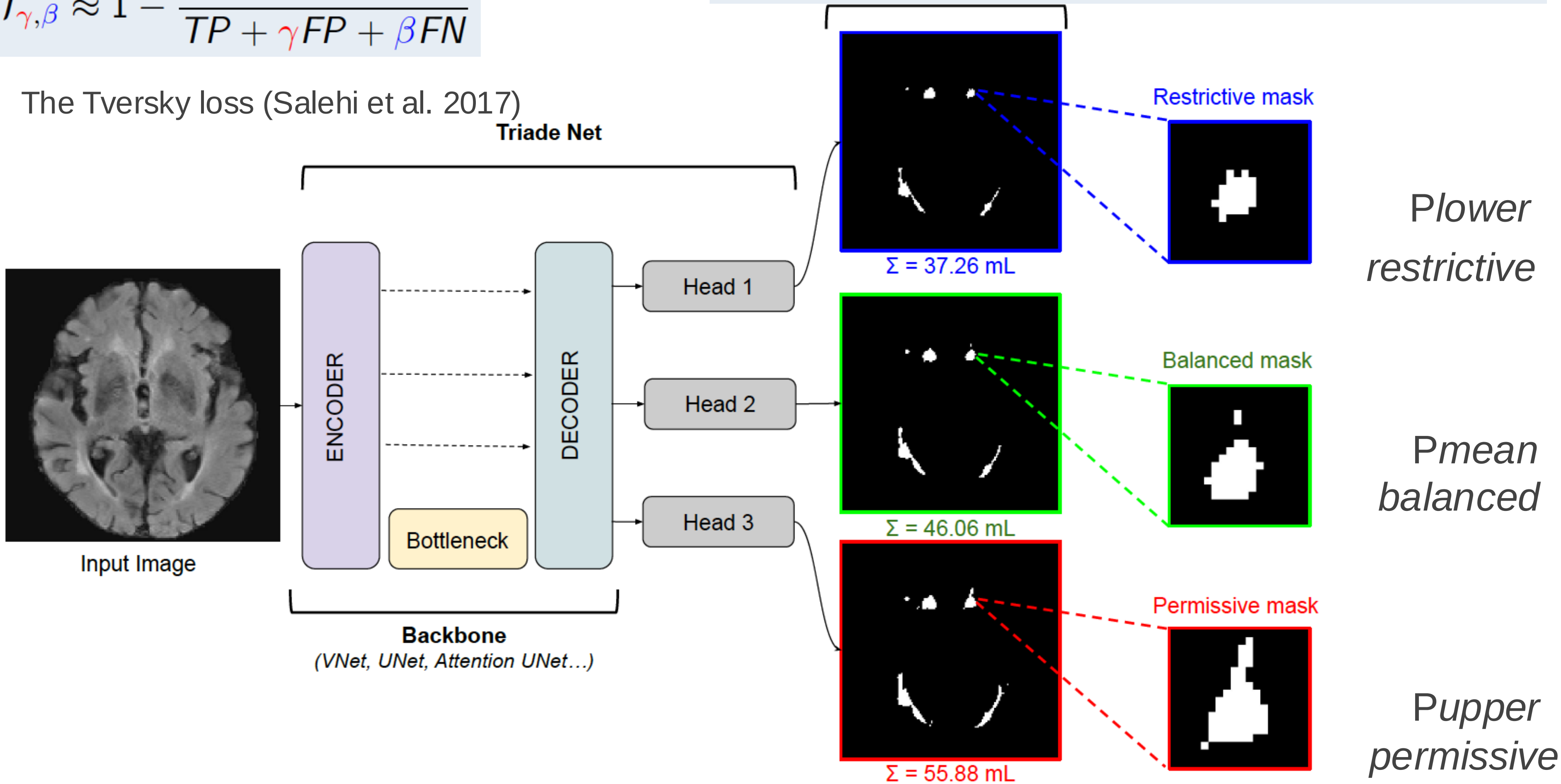
Conformal prediction

The TriadNet approach

$$\mathcal{T}_{\gamma,\beta} \approx 1 - \frac{TP}{TP + \gamma FP + \beta FN}$$

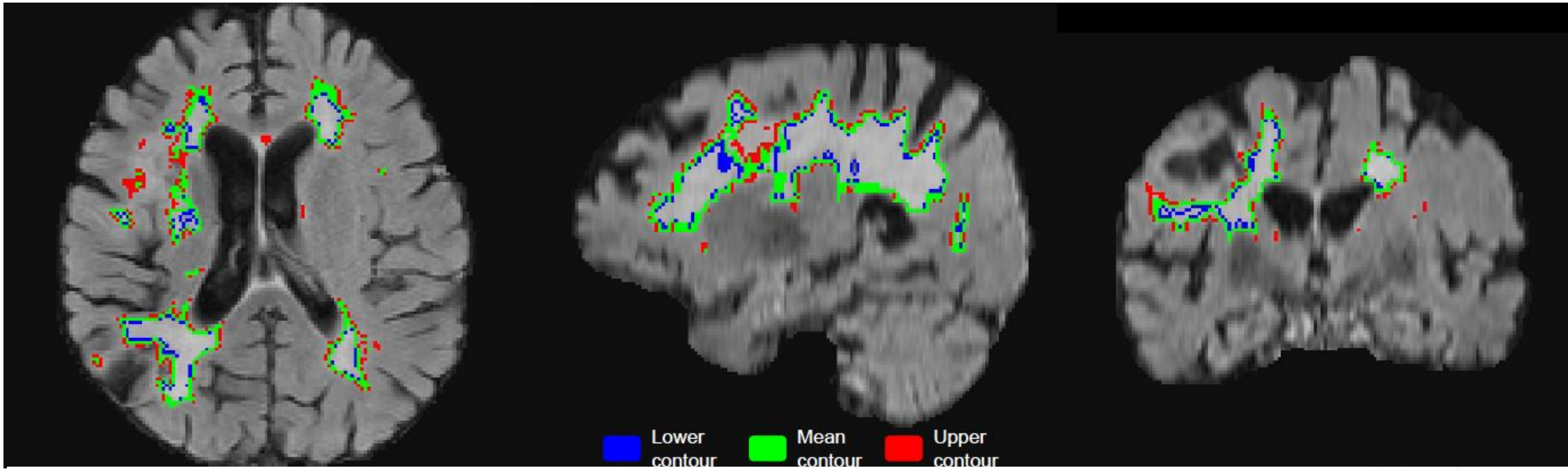
The Tversky loss (Salehi et al. 2017)

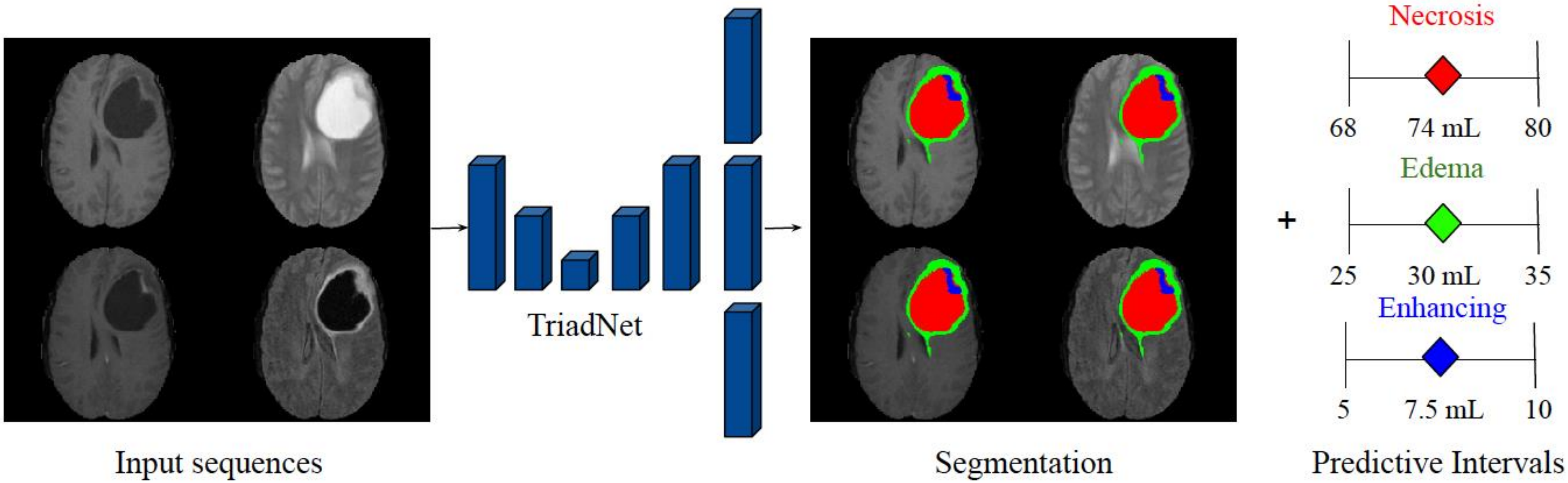
$$\text{TriadLoss} = \mathcal{T}_{0.8,0.2}(p_{lower,y}) + \mathcal{T}_{0.5,0.5}(p_{mean,y}) + \mathcal{T}_{0.2,0.8}(p_{upper,y})$$



[B. Lambert et al. (2023). “TriadNet: Sampling-Free Predictive Intervals for Lesional Volume in 3D Brain MR Images”. In: UNSURE 2023, LNCS 14291, pp. 32–41]

TriadNet predictions

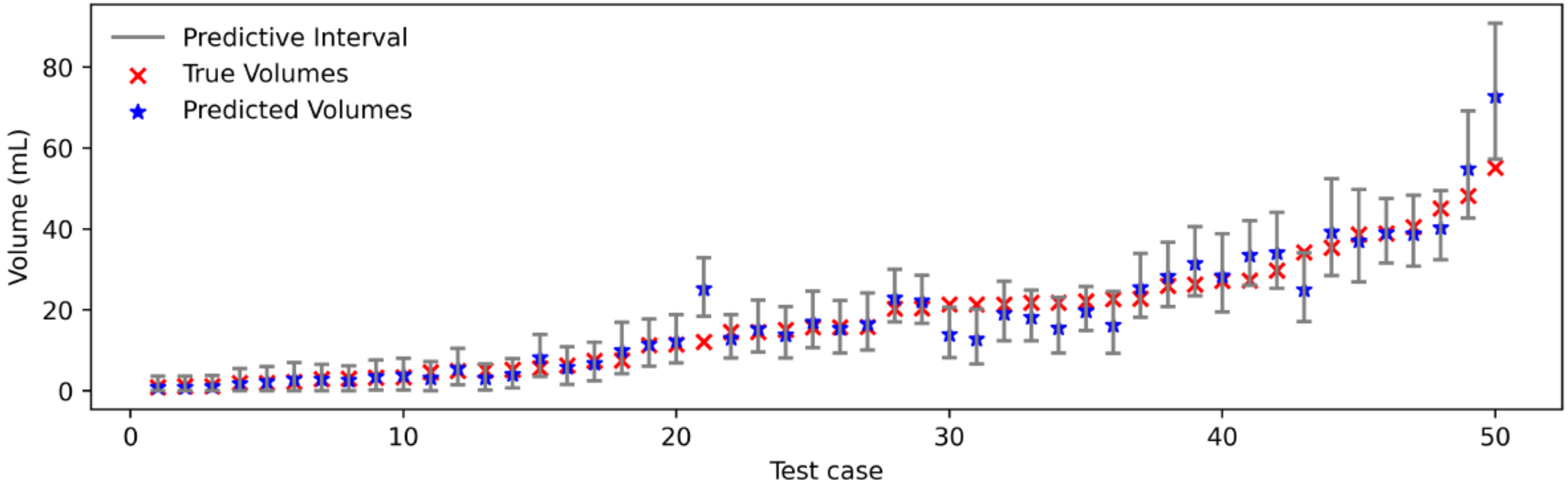




TriadNet predictions

Need for calibration

- 120 subjects for training, 40 for calibration and 50 for in-distribution testing. (Multicentric - 3 Tesla: MSSEG 2016 / WMH 2017 / ISBI 2015)
- Intervals calibrated for a target coverage of 90%.
- Metrics (bootstrapping, $M = 15000$):
  - Mean Average Error: 3.08 ± 0.46 mL
  - Empirical Coverage: $92.06 \pm 5.34\%$



Tumor volume estimation

Glioblastoma

- 679 subjects for training, 227 for calibration, and 227 for testing (BraTS 2023 dataset)
- Intervals calibrated for a target coverage of 90%.

Necrosis:

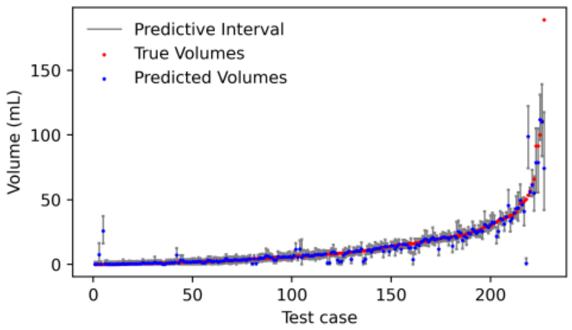
- MAE: 3.10 ± 0.46mL
- Coverage: $90.78 \pm 2.71\%$

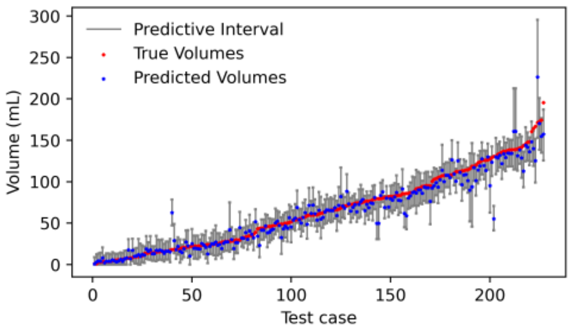
Edema:

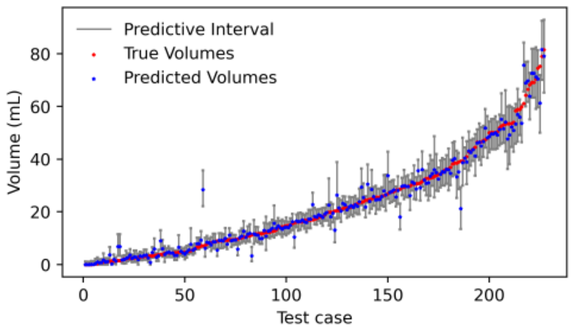
- MAE: 8.22 ± 0.57mL
- Coverage: $90.76 \pm 2.70\%$

Enhancing tumor:

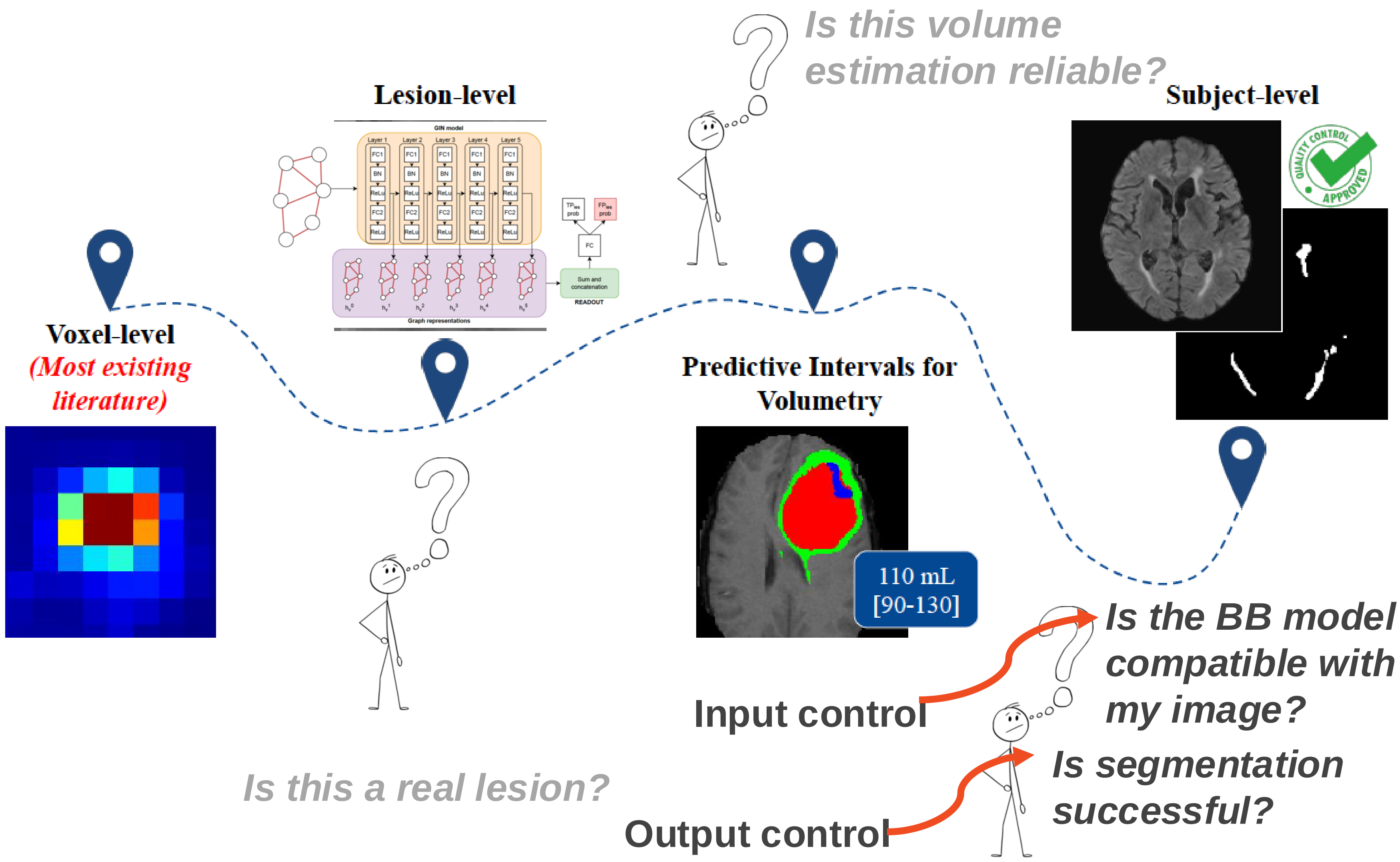
- MAE: 1.73 ± 0.19mL
- Coverage: $90.79 \pm 2.71\%$



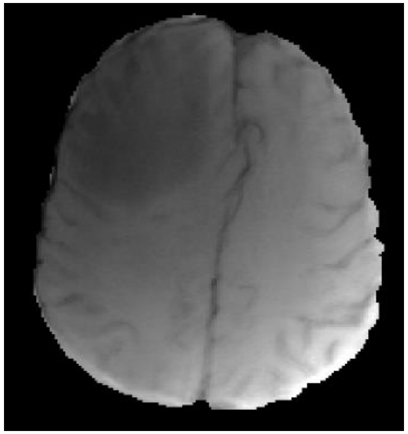




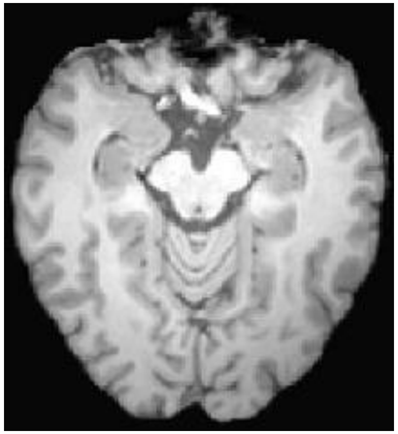
The multi-dimensional aspect of uncertainty



DLL trained for Gioblastoma detection on T1w

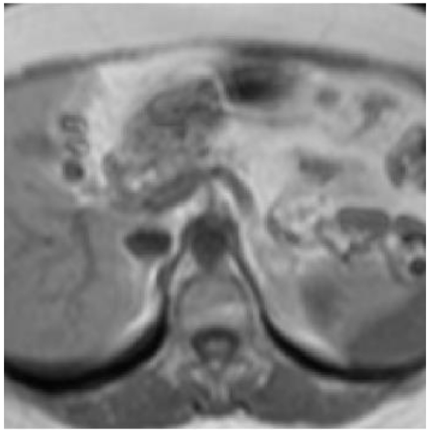


Artefacted T1w



Healthy subject

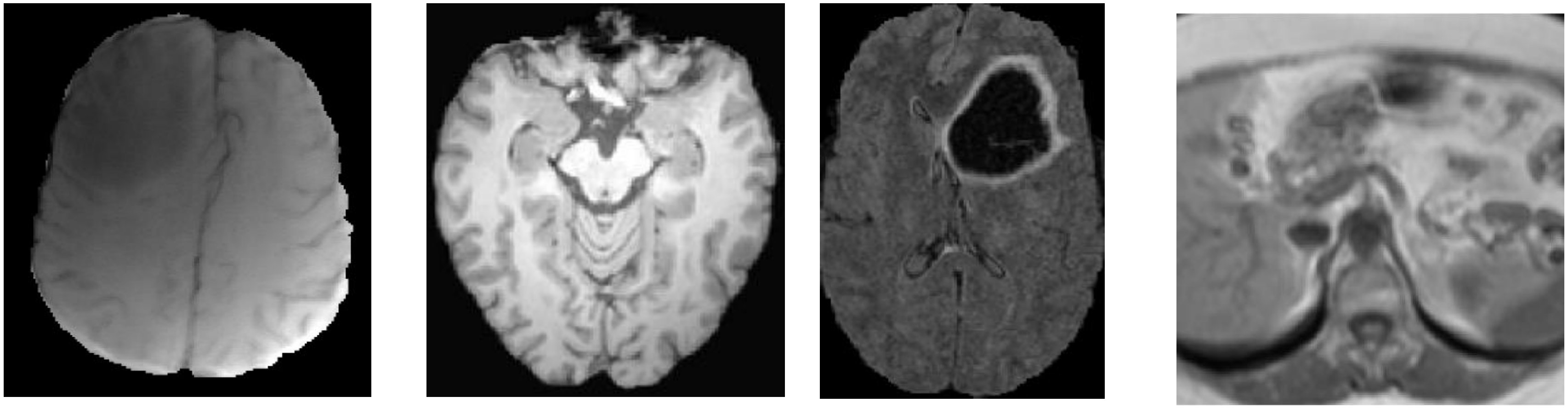
FLAIR



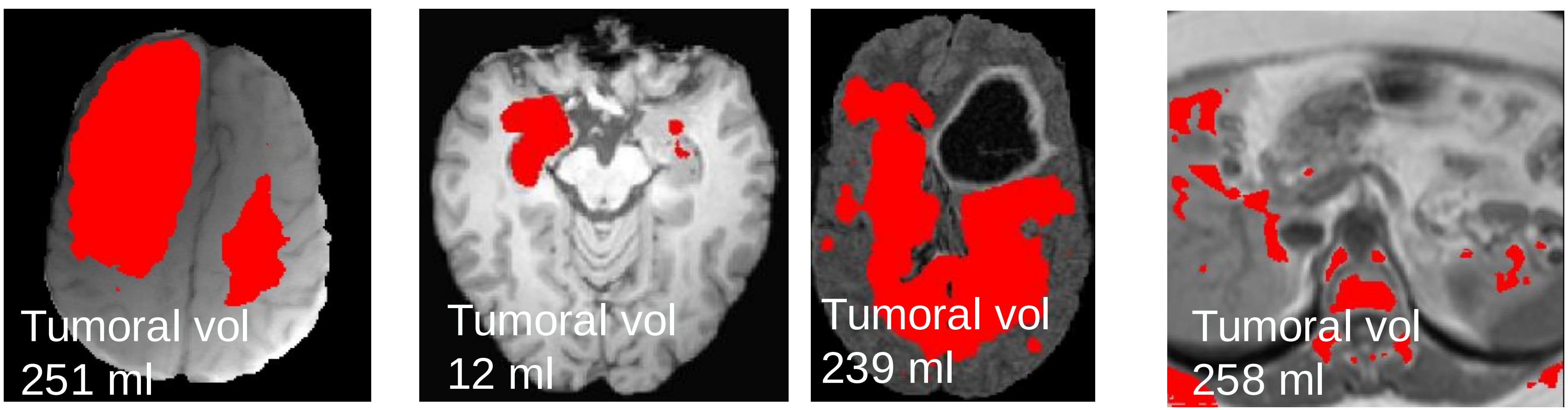
Abdominal T1w

Know-it-all

DLL trained for Gioblastoma segmentation on T1w

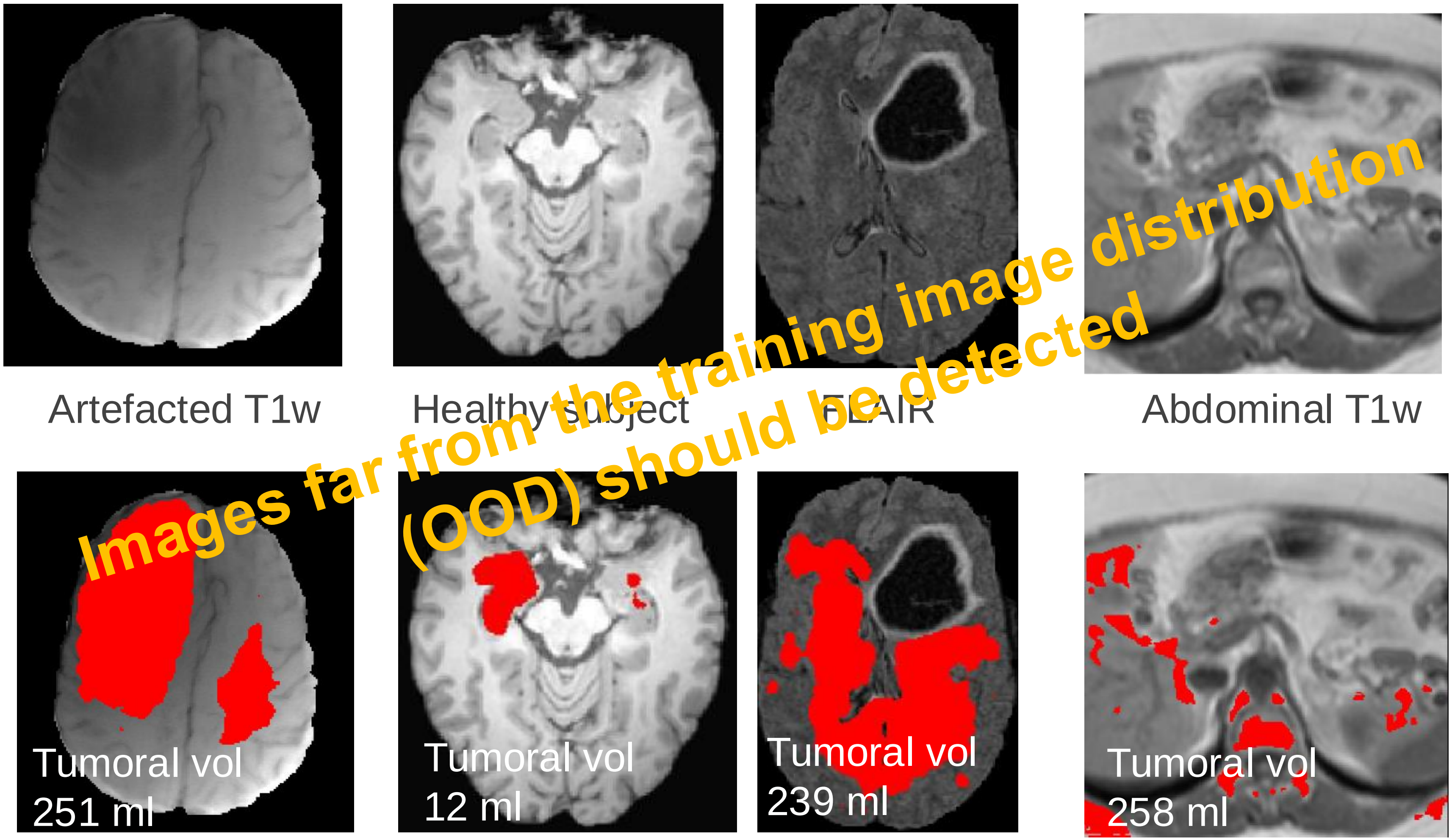


Artefacted T1w | Healthy subject | FLAIR | Abdominal T1w

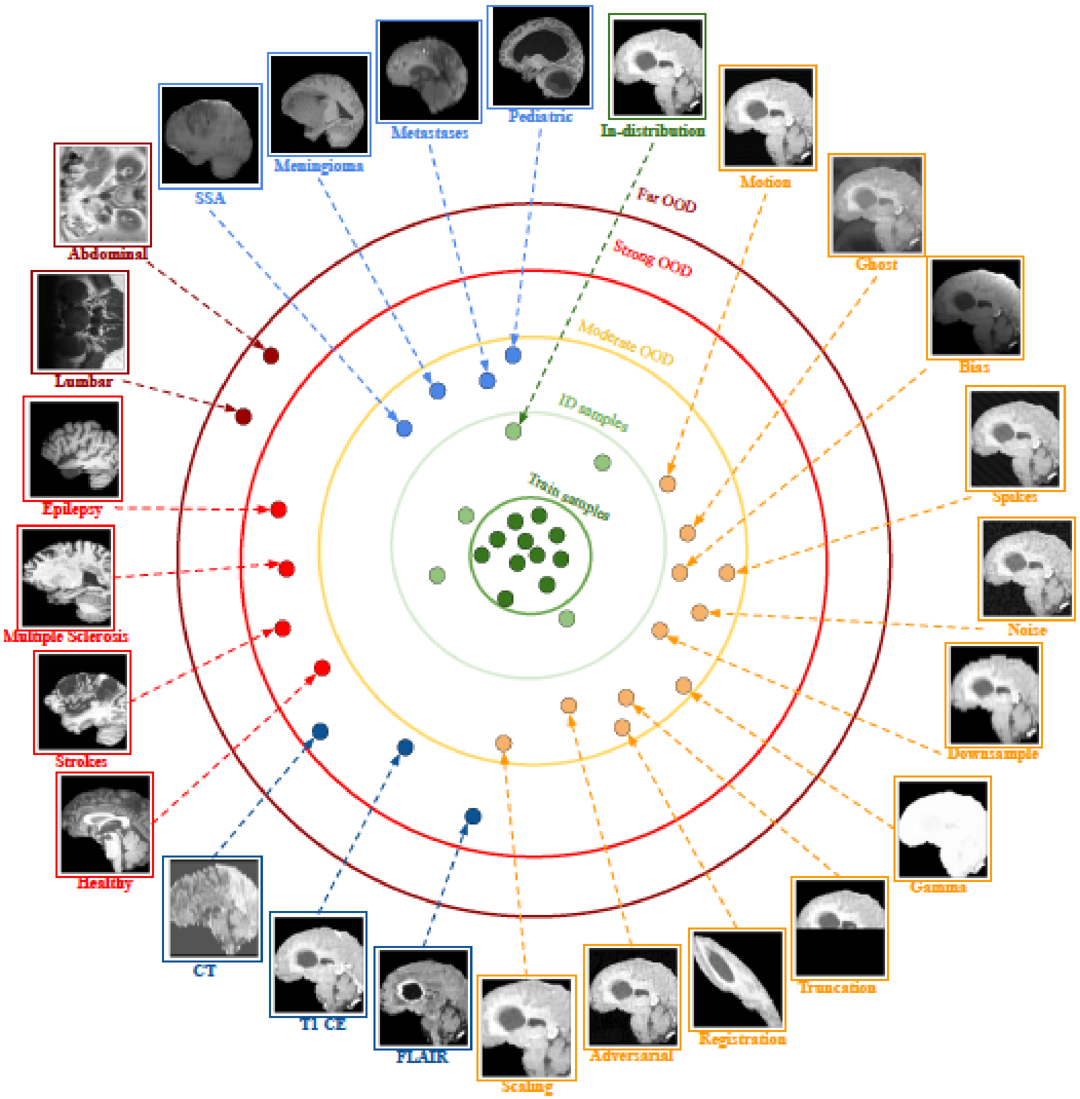


Input Quality Control

DLL trained for Gioblastoma segmentation on T1w



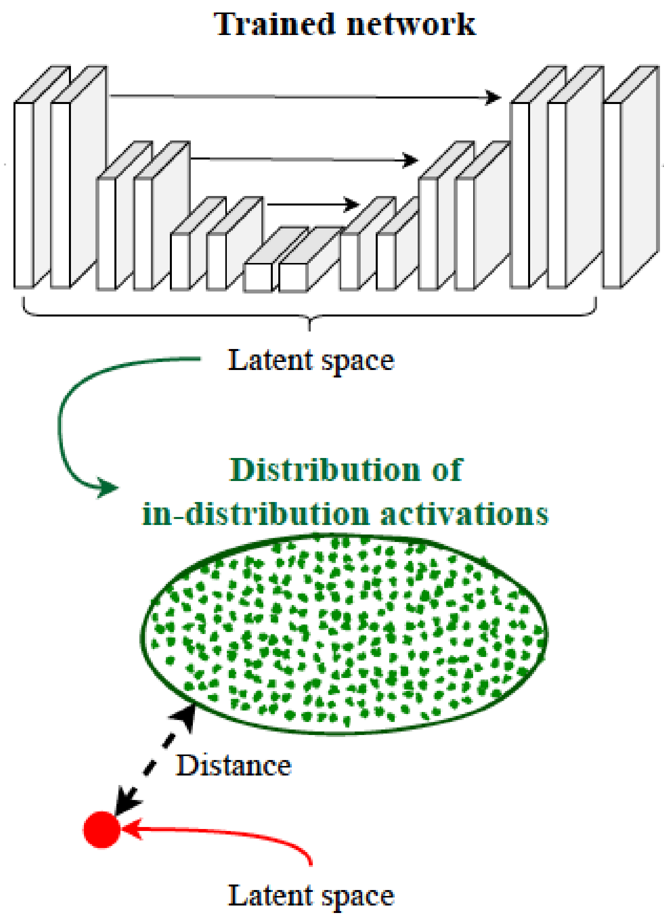
Why an image is OOD?



- In-distribution ↔ training distribution (T1 MRI of Adult glioblastoma patients)
- Are **out-of-distribution**:
  - Images corrupted with artifacts.
  - Shifts in the imaged population.
  - Shifts in image modality.
  - Diseases not present in the training set.
  - Incorrect organs.

Input Quality Control

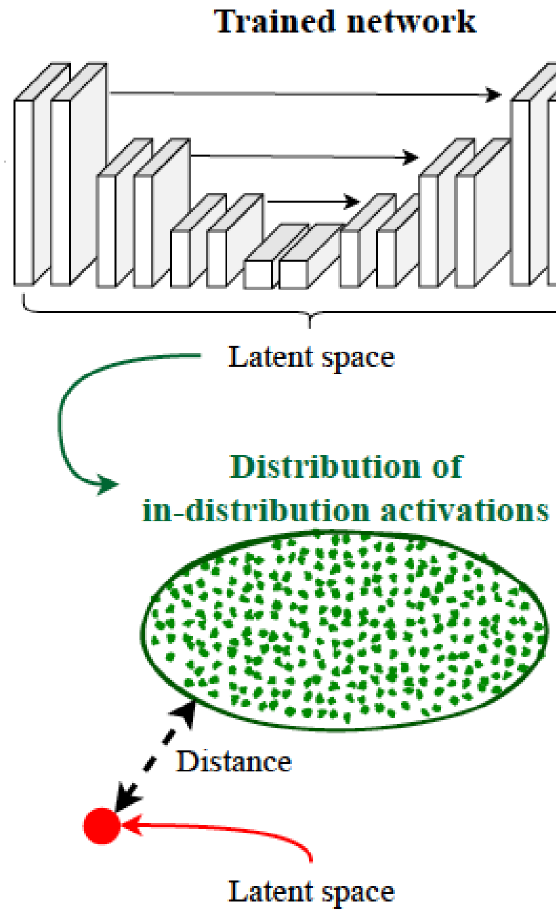
The latent-space distance



OOD activation

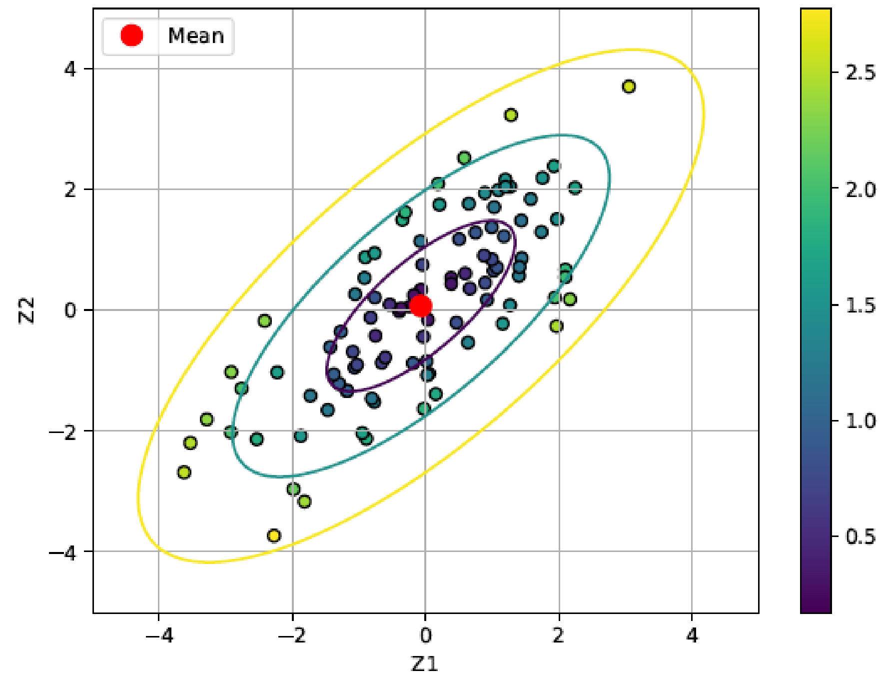
Input Quality Control

The latent-space distance



multivariate Gaussian distribution

OOD activation



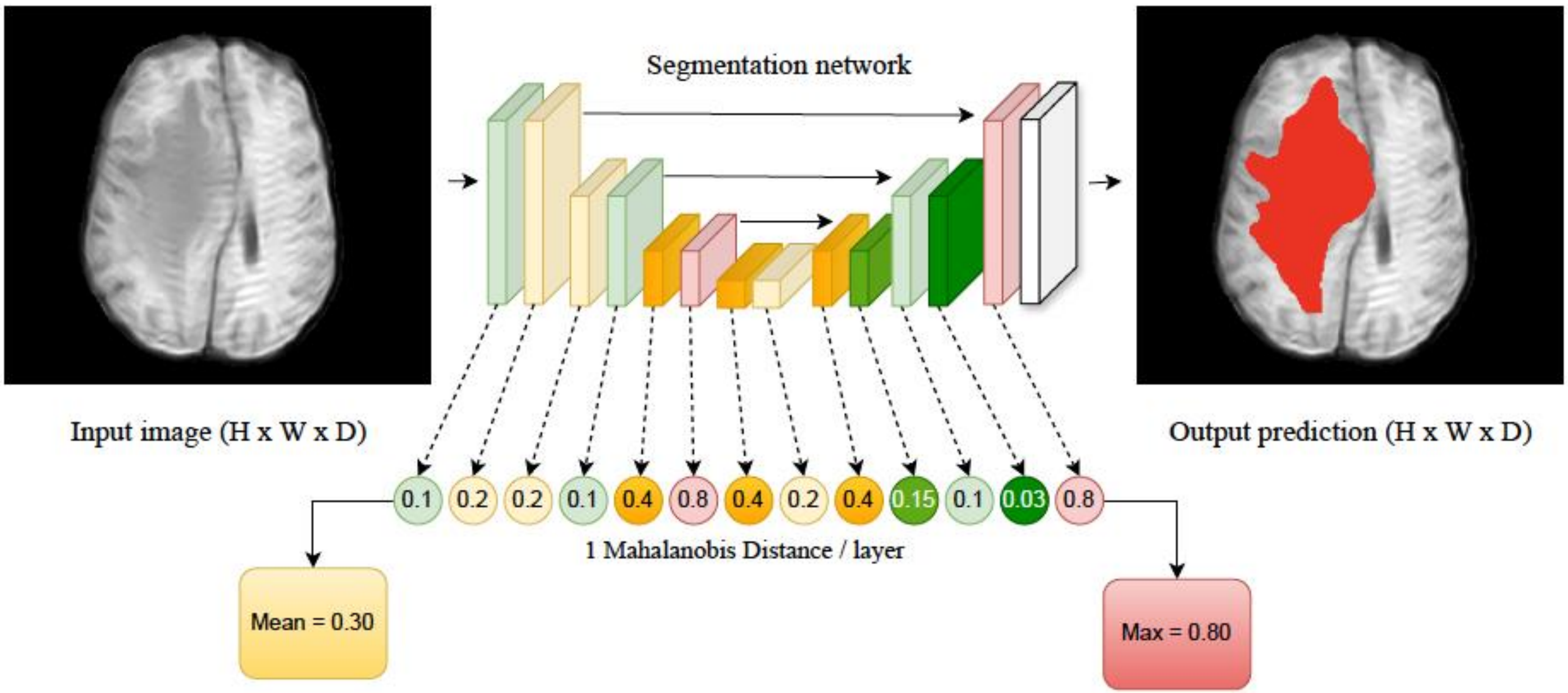
Computed with Mahalanobis distance (MD)

$$\mu = \frac{1}{N}\sum_{i=1}^{N} x_i$$

$$\Sigma = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)(x_i - \mu)^T$$

$$\mathrm{MD}(x_{test}; \mu, \Sigma) = (x_{test} - \mu)^T \Sigma^{-1} (x_{test} - \mu)$$

A multi-layer aggregation of MD

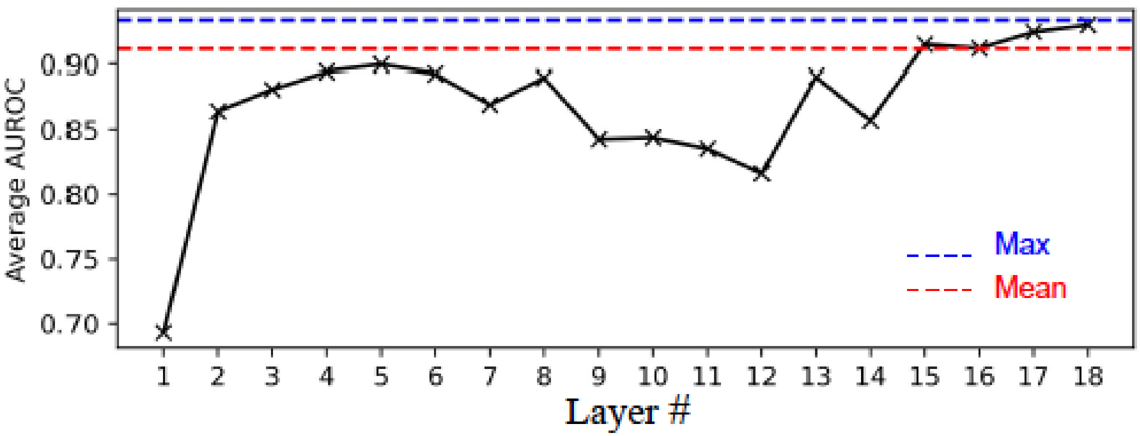


[B. Lambert et al. (2023). “Multi-layer Aggregation as a key to feature-based OOD detection”. In: UNSURE 2023, Held in Conjunction with MICCAI 2023. LNCS 14291, pp. 104–114]

Input Quality Control

Brats: 876 subjects for training,
30 for validation,
227 for in-distribution testing

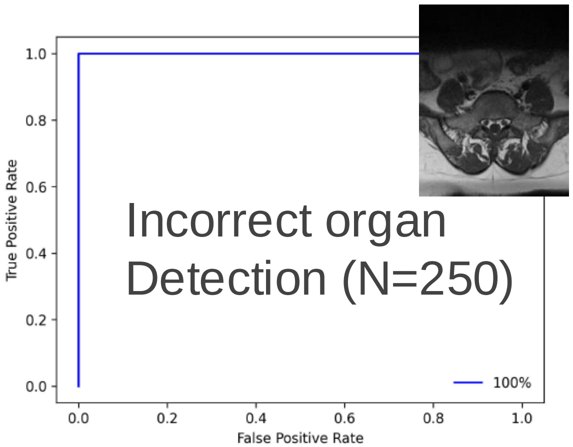
Dynamic U-Net

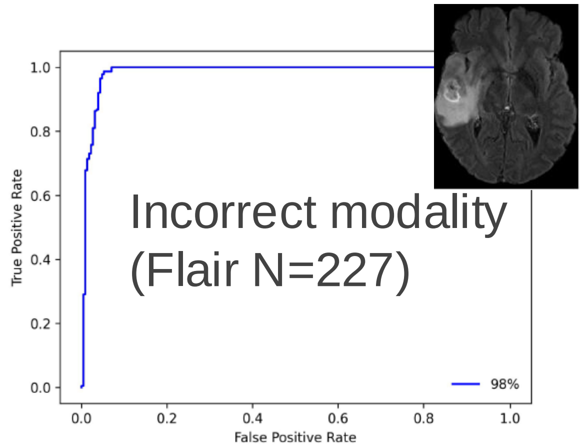


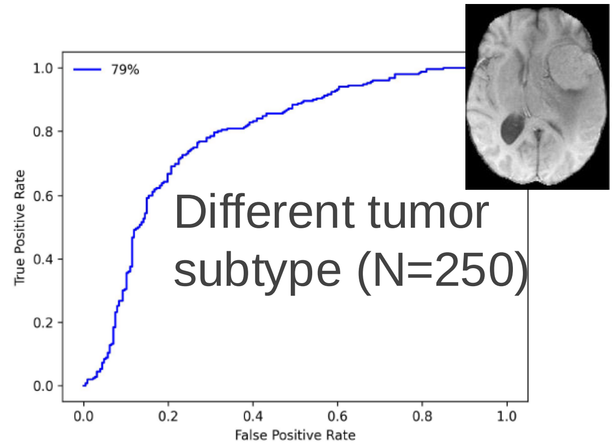
The optimal layer for OOD detection depends on the segmentation architecture.

Input Quality Control

The multi-layer scores (Mean and Max) provides high detection accuracy.





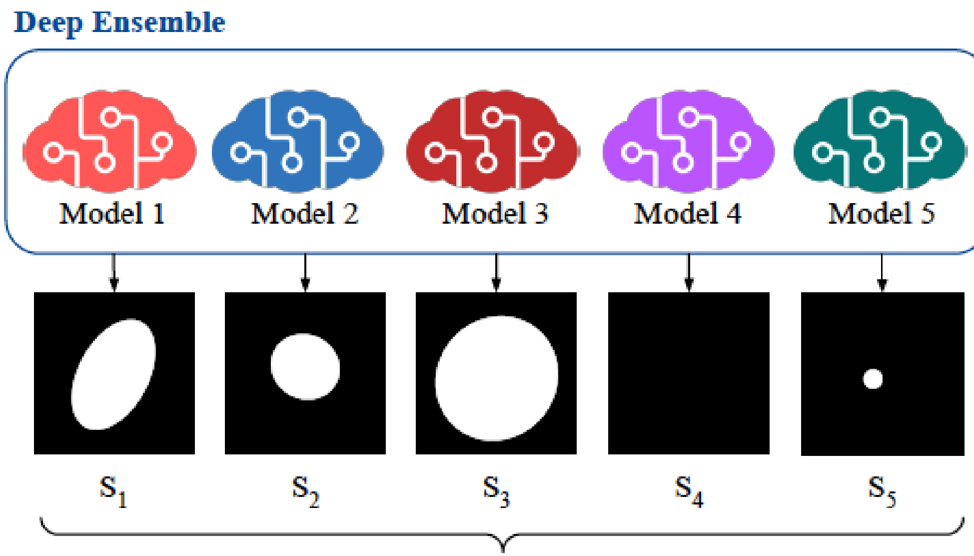


Latent-space distances efficient in detecting images far from the training distribution.

Output Quality Control

Goal: estimate the true segmentation accuracy.

How: Measure the segmentation variability among models.

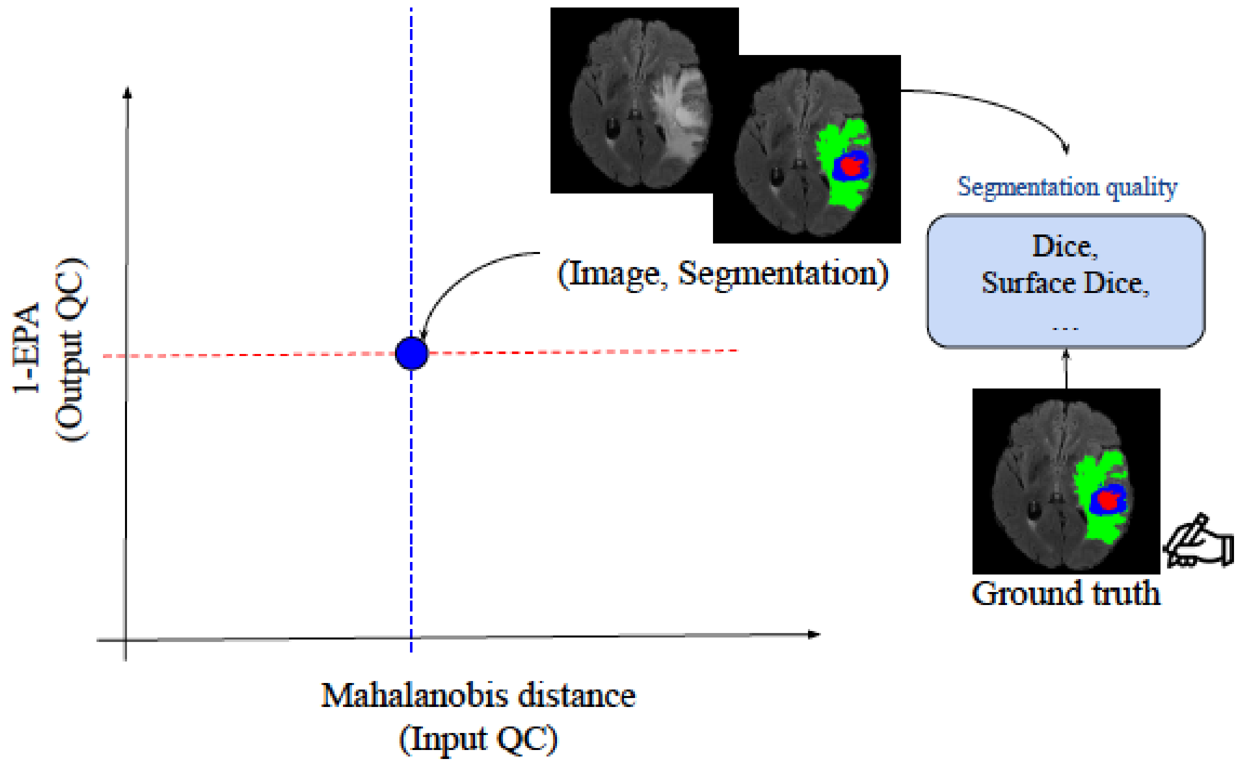


We note S_k the individual segmentations and MV the majority vote segmentation

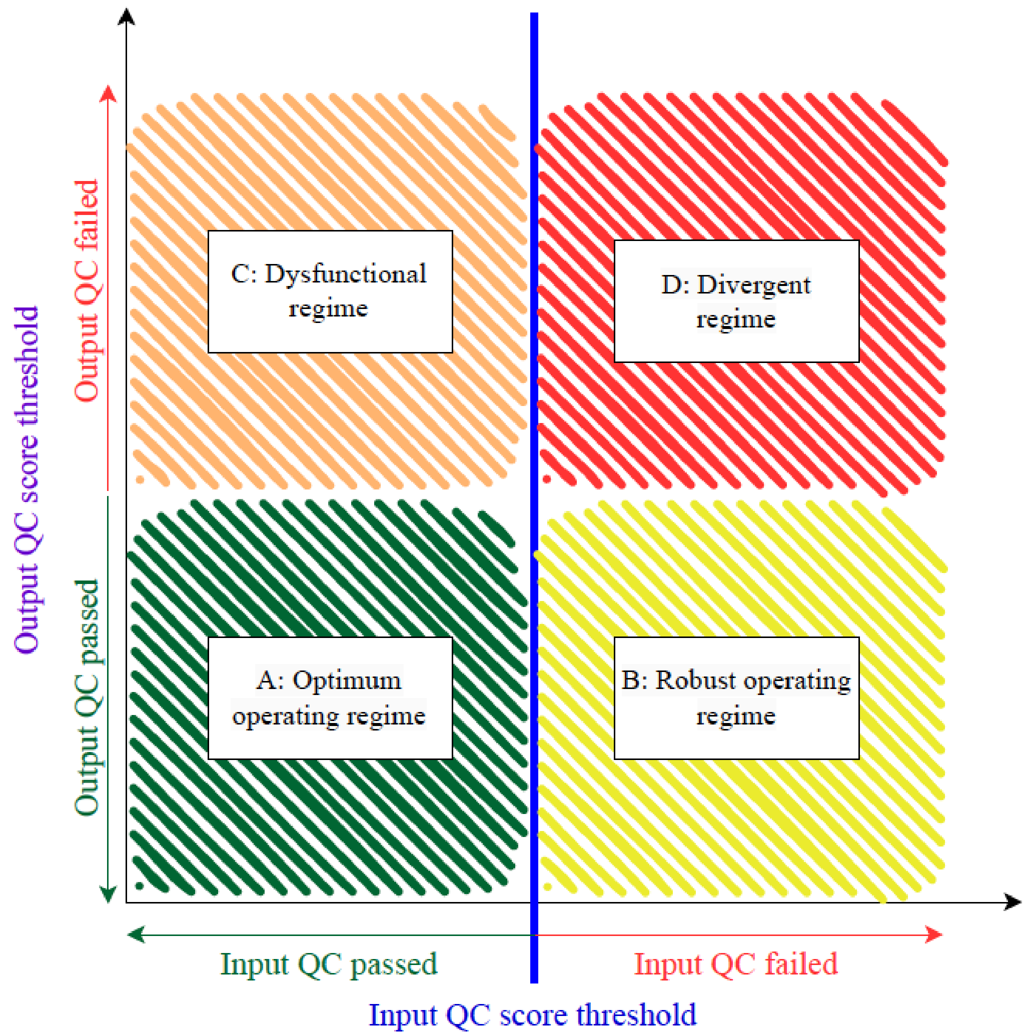
Ensemble Prediction Agreement (EPA): $\text{EPA} = \frac{1}{K}\sum_{i=1}^{K} \text{Dice}(S_k, MV)$

[B. Lambert et al. (2024) "From Out-of-distribution detection to Quality Control". In: Trustworthy AI in Medical Imaging, MICCAI book series]

Unified Input-Output control



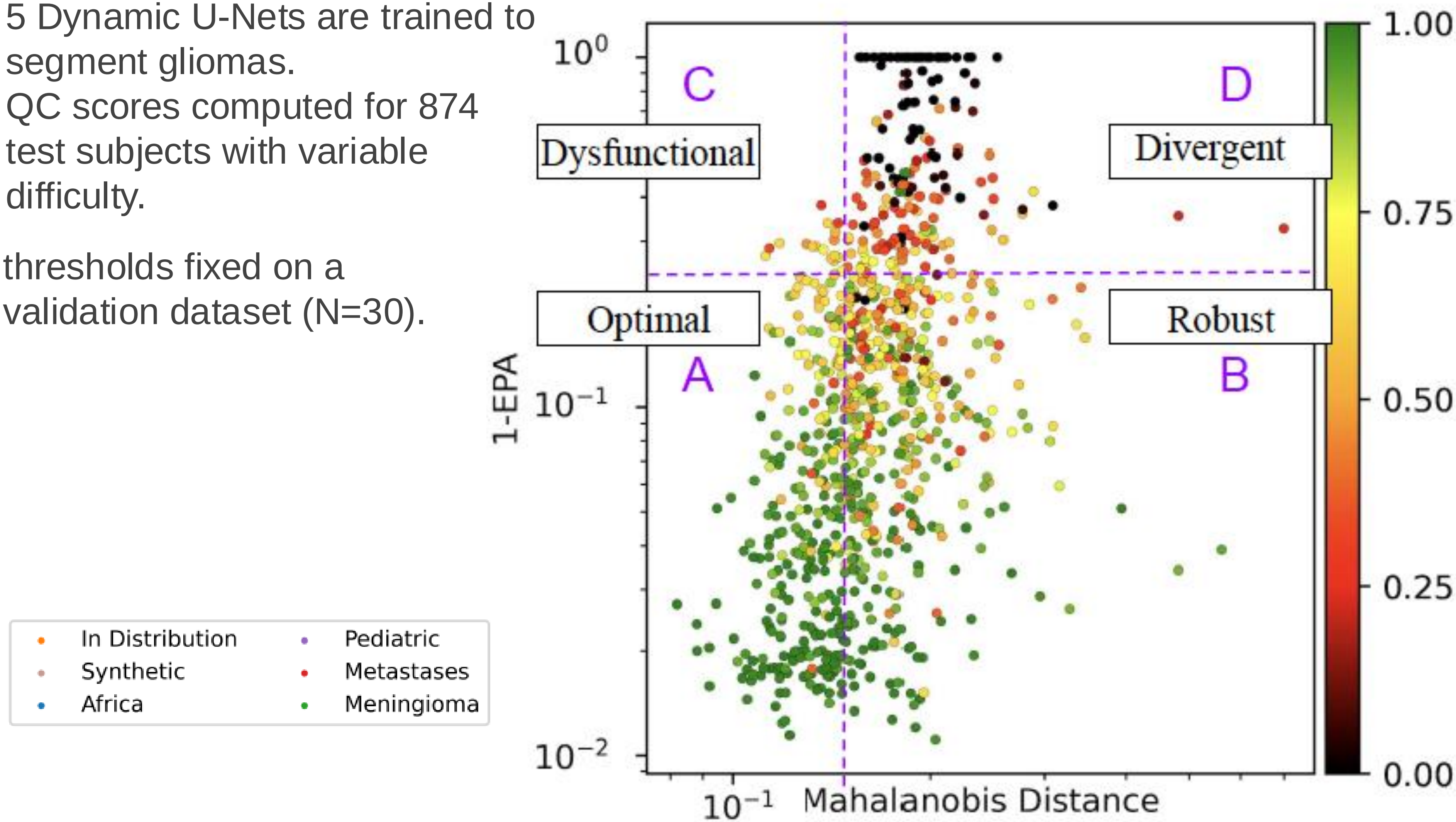
Unified Input-Output control



Unified Input-Output control

5 Dynamic U-Nets are trained to segment gliomas.
QC scores computed for 874 test subjects with variable difficulty.

thresholds fixed on a validation dataset (N=30).



Take home messages

- ✓ Confidence is central for AI deployment
- ✓ Uncertainty quantification improves user's confidence
- ✓ Uncertainty is multidimensional

Lesion uncertainty scores
Predictive volume intervals
Unified input & output controls

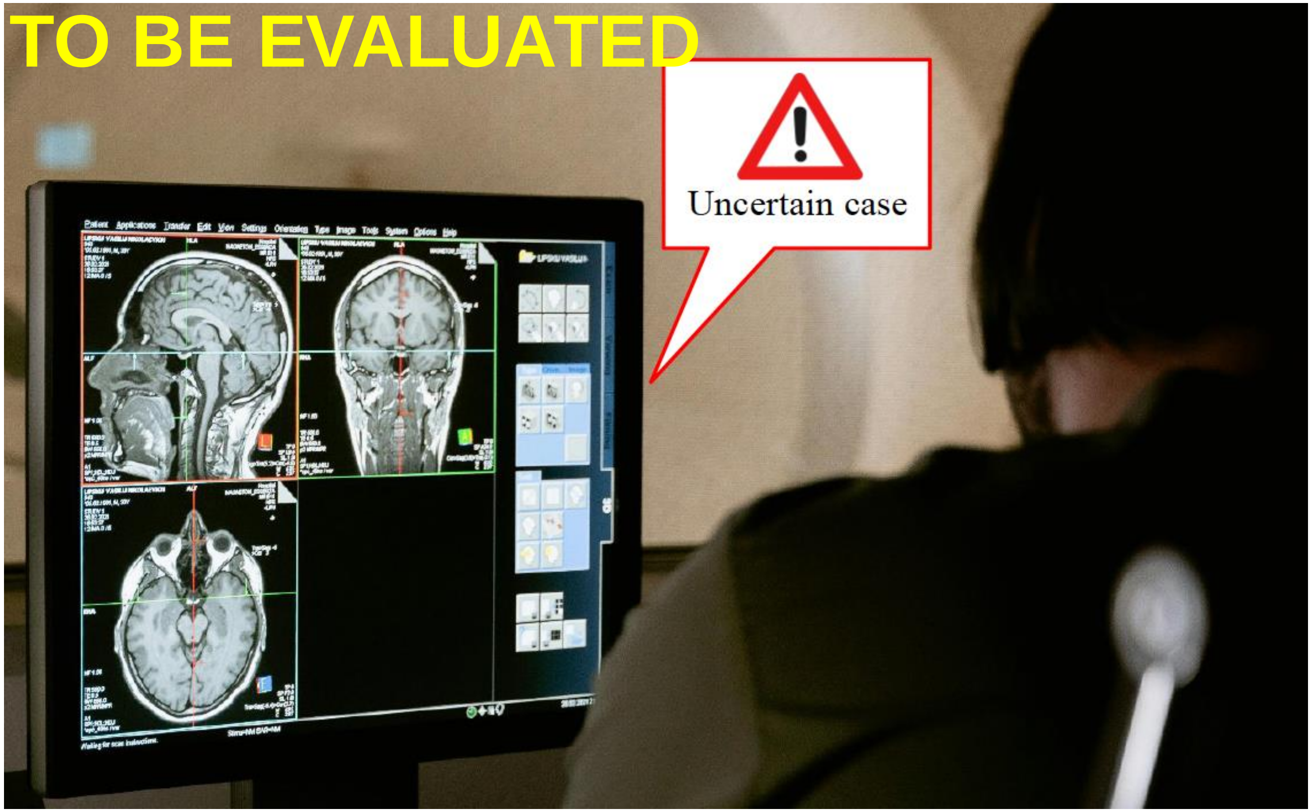
uncertainty quantification
for trusted AI

for AI penetration in clinical routine

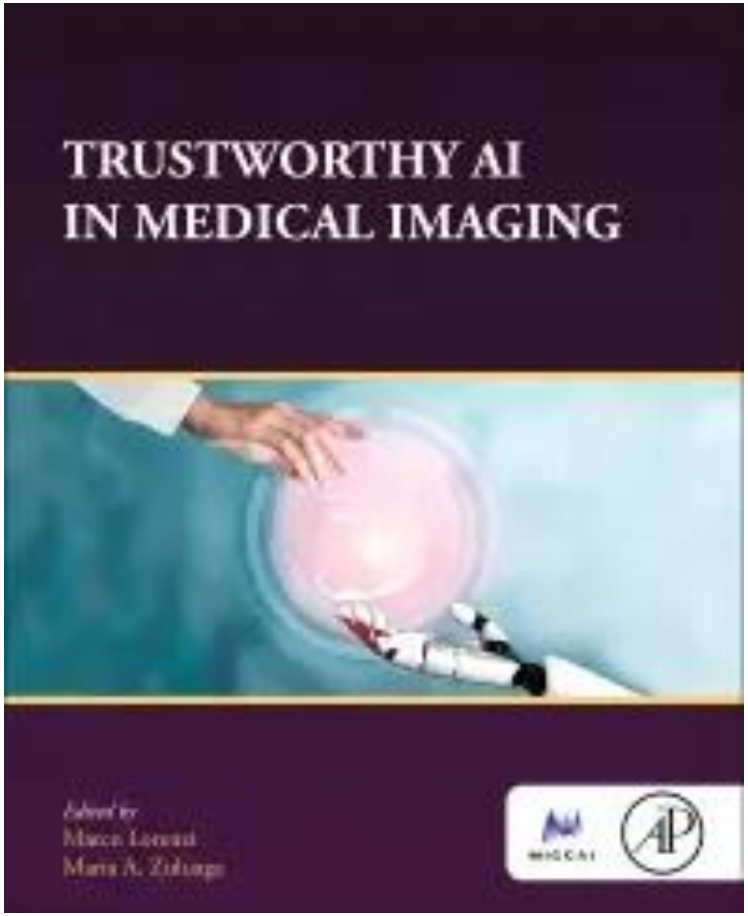
NEXT STEP

What is the added-value
in clinical routine applications?

TO BE EVALUATED



Trustworthy AI in Medical Imaging



Trustworthy AI in Medical Imaging

- 1st Edition - **December 1, 2024**
- Editors: Marco Lorenzi, Maria A Zuluaga
- eBook ISBN: 9780443237607

Section 1 – Robustness

Section 2 - Validation, Transparency and Reproducibility

Section 3 – Bias and Fairness

Section 4 - Explainability, Interpretability and Causality

Section 5 - Privacy-preserving ML

Section 6 - Collaborative Learning

Section 7 - Beyond the Technical Aspects


The End