Uncertainty quantification in Medical image analysis

Nataliia Molchanova

Sources of uncertainty

Aleatoric
irreducible data noise

Distributional shift

Label uncertainty

Epistemic
limited data/model knowledge

Training stochasticity

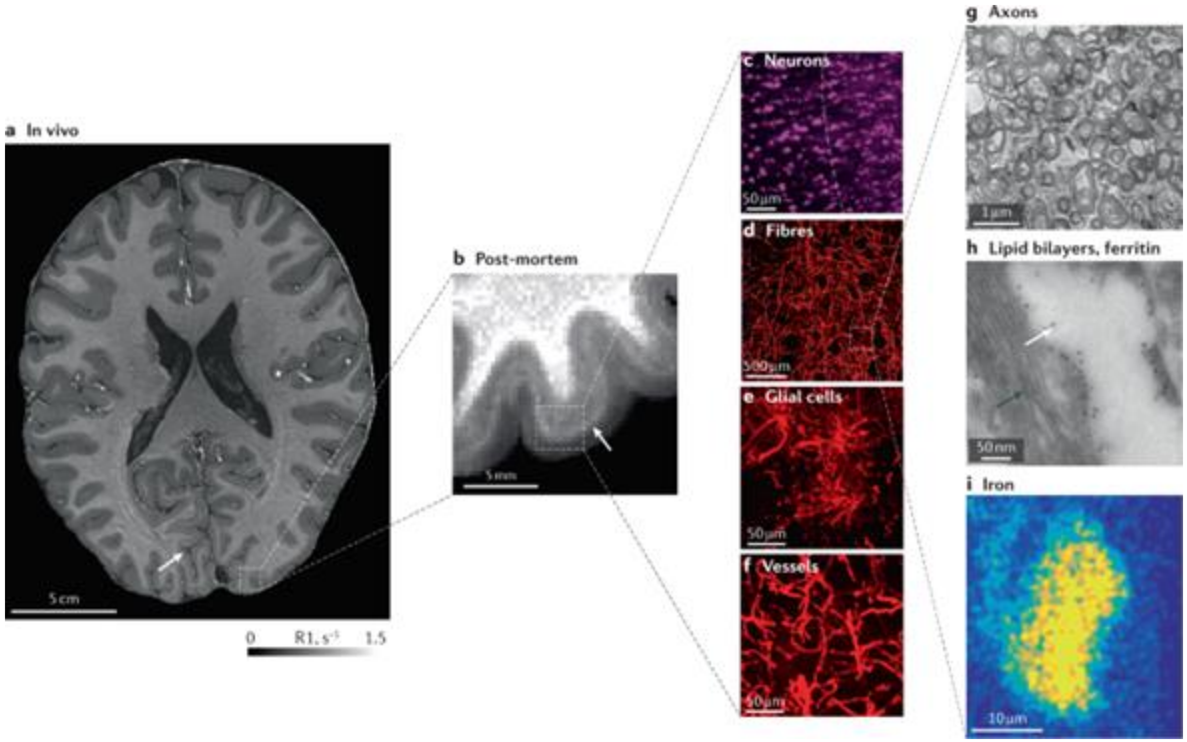
Model specification

Malinin, PhD Thesis, 2019; Hüllermeier & Waegeman, Machine Learning, 2021; Wimmer et al., UAI'23; Gawlikowski et al., Artif. Intell. Rev., 2024

Medical Imaging Context

Image-related noise

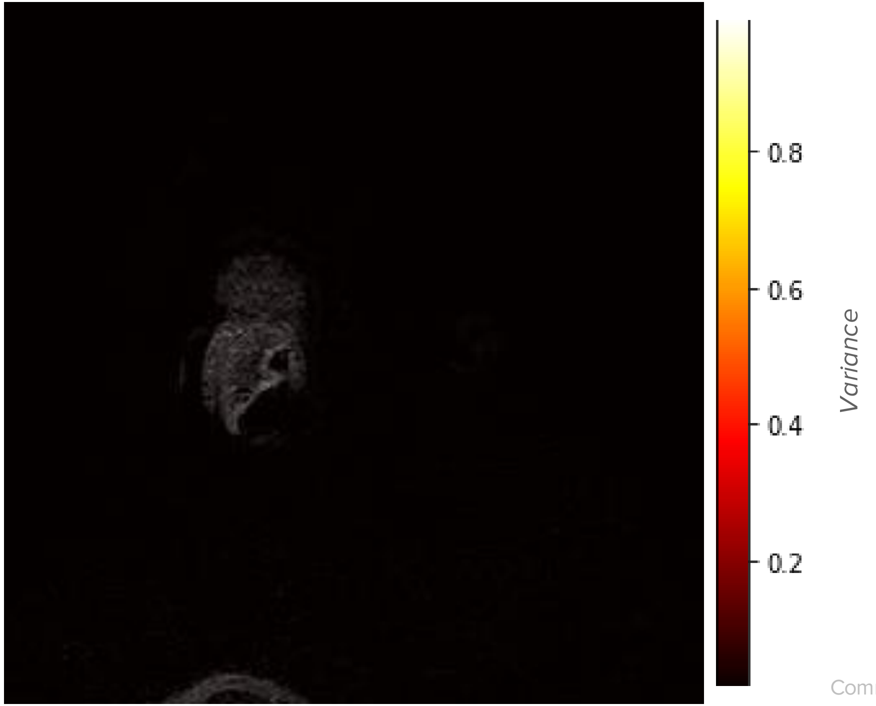
- Limited resolution
- Hardware limitations
- Artifacts
- Processing



Weiskopf et al., Nat. Rev. Phys., 2021

Inter- and intra- rater variability

Example of white matter lesions annotations from 7 raters



Commowick et al., Sci. Rep., 2018

Label-related noise

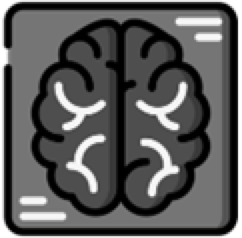
Expertise level

Cognitive load

Visual perception

Low quality

Artifacts

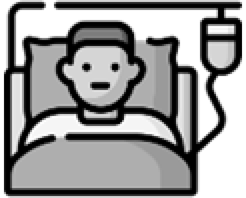


Guidelines ambiguity

Annotation tools

Rare pathologies

Demographic differences



Low data regime

Challenge Name	Task	Medical Imaging Modality	Number of Patients	Countries	Patients Worldwide
KiTS21 Challenge	Kidney Tumor Segmentation	CT	**300**	USA	**400K/year**
MSSEG-1	Multiple Sclerosis Lesion Segmentation	MR	**53**	France	**2.9M**
LiTS	Liver Tumor Segmentation	CT	**130**	Germany	**800K/year**
PROMISE12	Prostate MR Image Segmentation	MR	**50**	United States, Canada, Germany, France, UK	**1.4M/year**

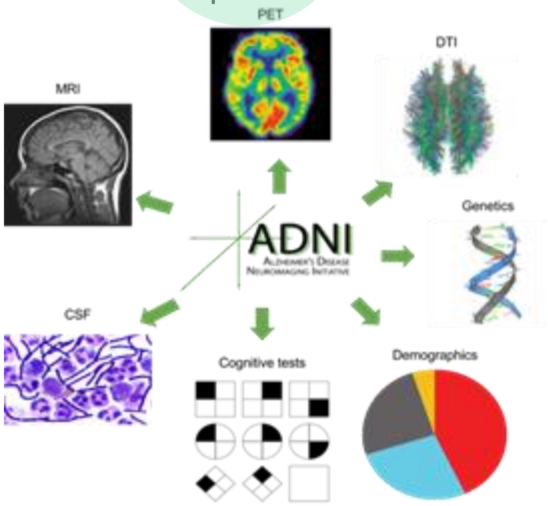
Low data regime

Worldwide prevalence

60M patients

TADPOLE dataset

≈ 1.5k patients

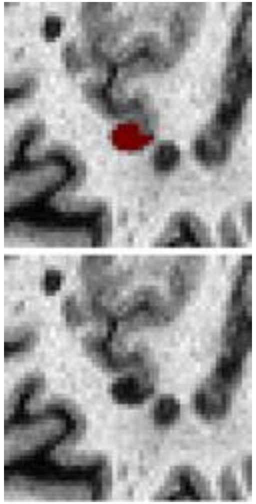


Low data regimes

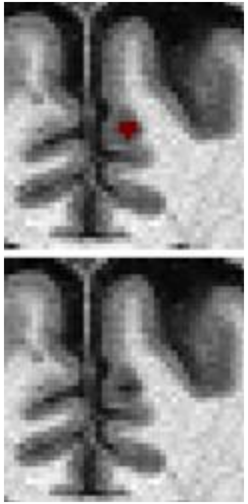
Hidden data biases

Subtypes of cortical brain lesions

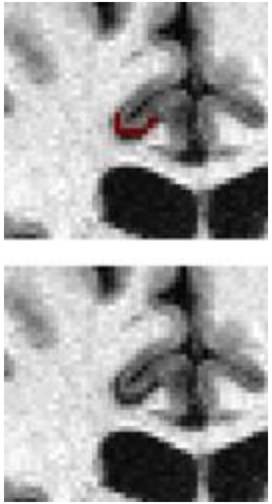
Leukocortical 60%



Intracortical 30%



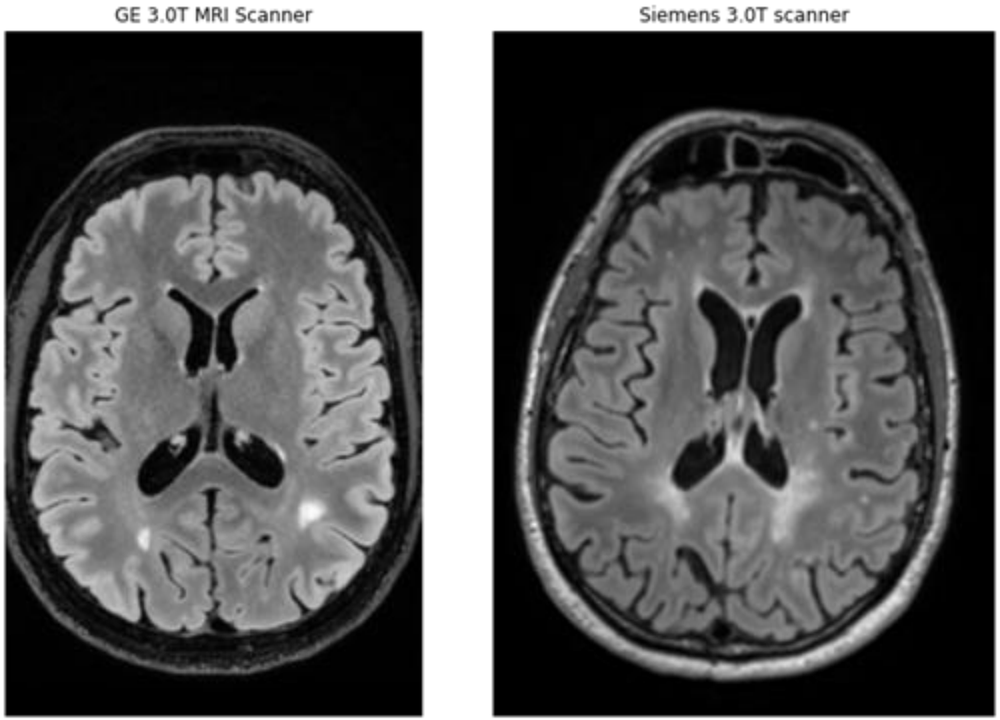
Subpial <5%



Molchanova et al., MICCAI iMIMIC, 2024

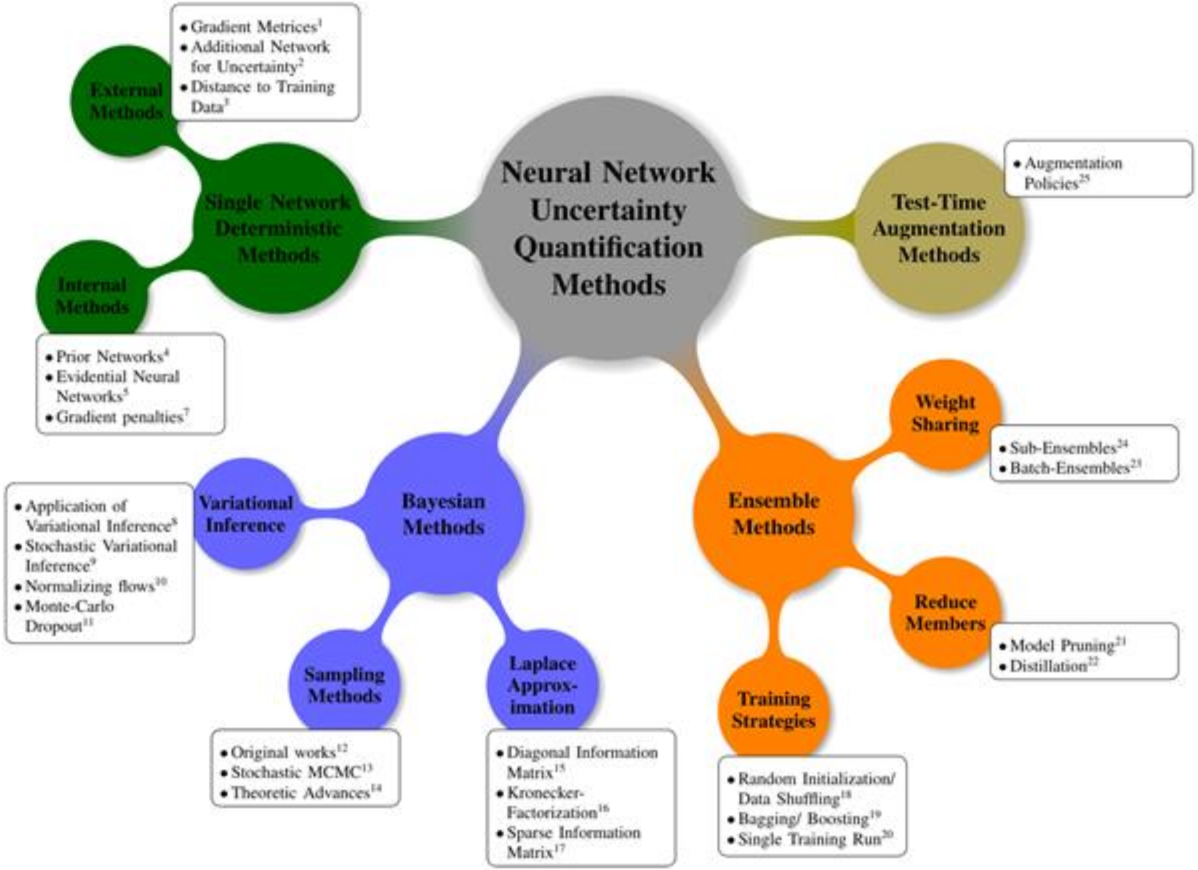
Distributional shifts

- Covariate (input data)
- Label
- Concept drift (reality)



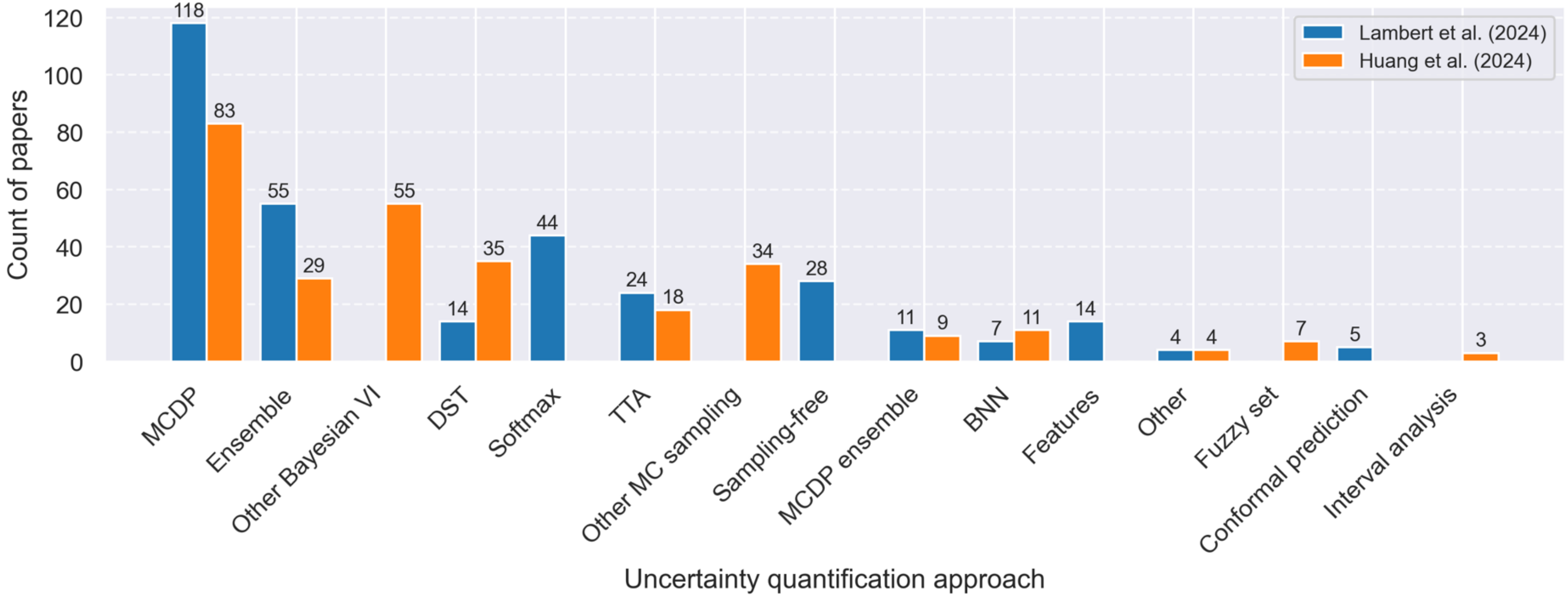
Uncertainty Quantification Methods

A complex landscape of UQ methods in DL…



Gawlikowski et al., Artif. Intell. Rev., 2024

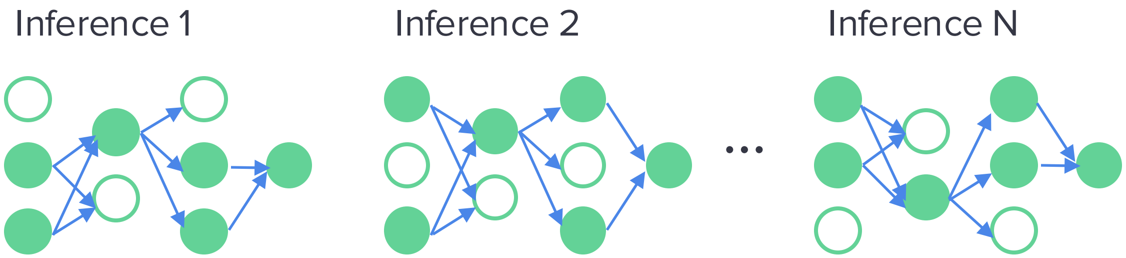
Uncertainty quantification in MIA



Lambert et al., Artif. Intell. Med., 2024; Huang et al., Med. Image Anal., 2024

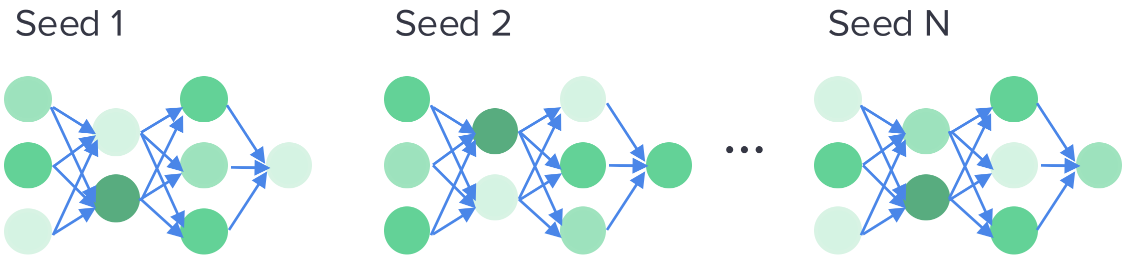
Monte Carlo dropout

Dropout during the inference time induces a distribution over the weights and biases of the network



Deep ensemble

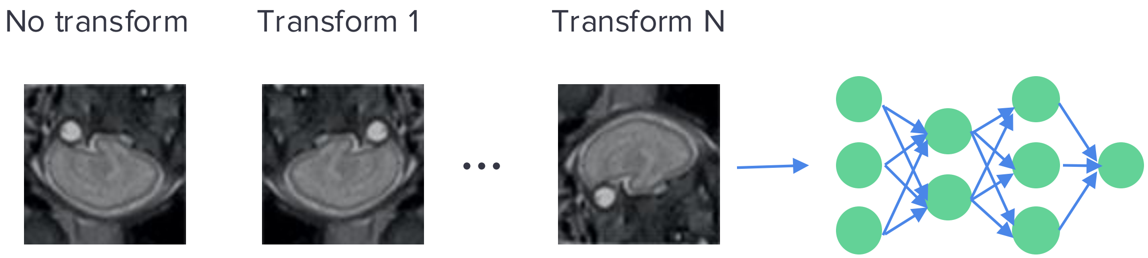
Train N identical neural networks with different random seeds



Lakshminarayanan et al., Neurips, 2017

Test-time augmentation

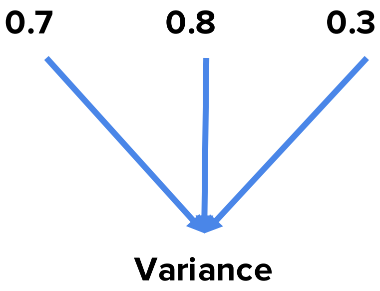
Perform several inferences with the same input, but transformed using an invertible transformation



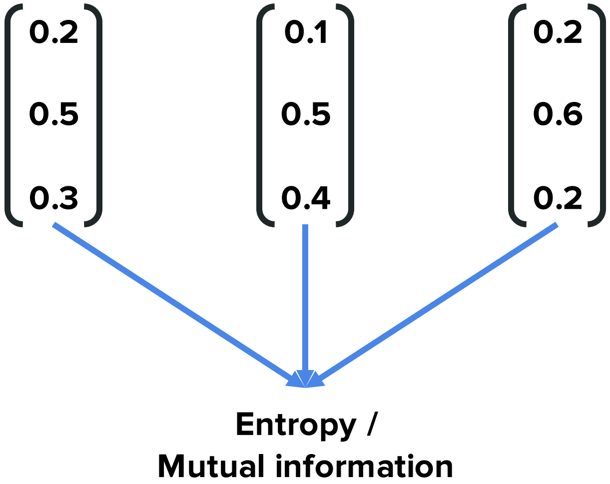
Wang et al., Neurocomputing, 2019

Uncertainty measures

Regression

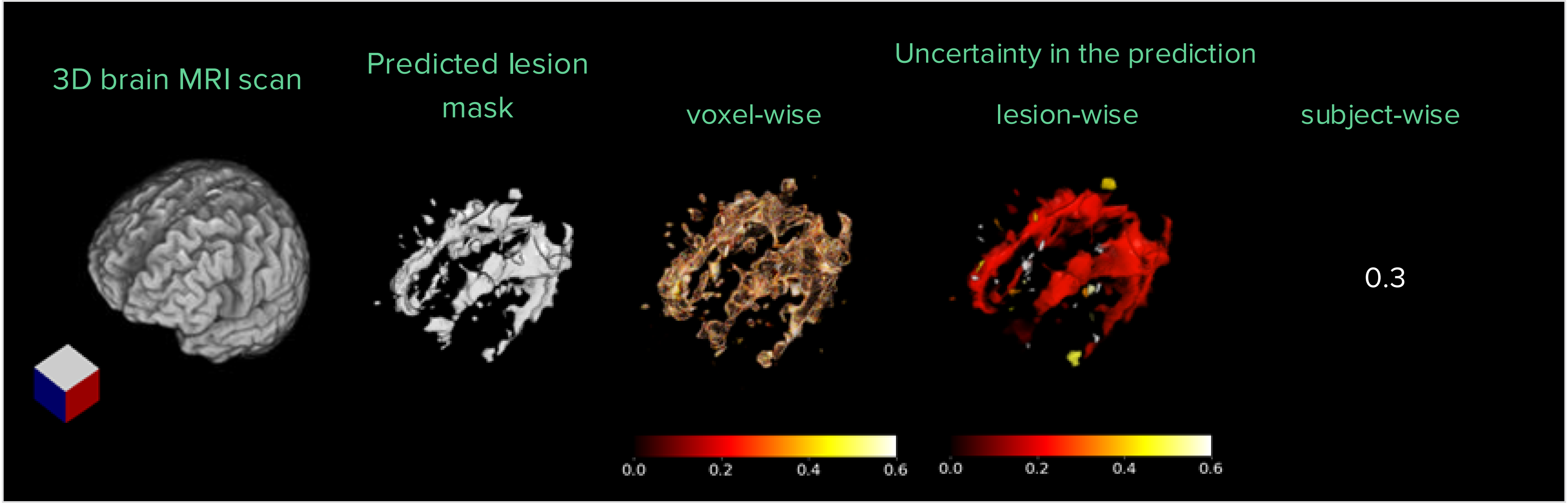


Classification



Uncertainty measures

Reconstruction and segmentation have structured outputs



Roy et al., NeuroImage, 2019; Nair et al., Med. Image Anal., 2020; Molchanova et al., Comput. Biol. Med., 2025

Evaluation

Generic evaluation strategy

No ground truth uncertainty in MIA (only in toy ML tasks)

Higher uncertainty values = higher likelihood of error

Calibration measures (e.g., expected calibration error)

Comparison with the rater disagreement

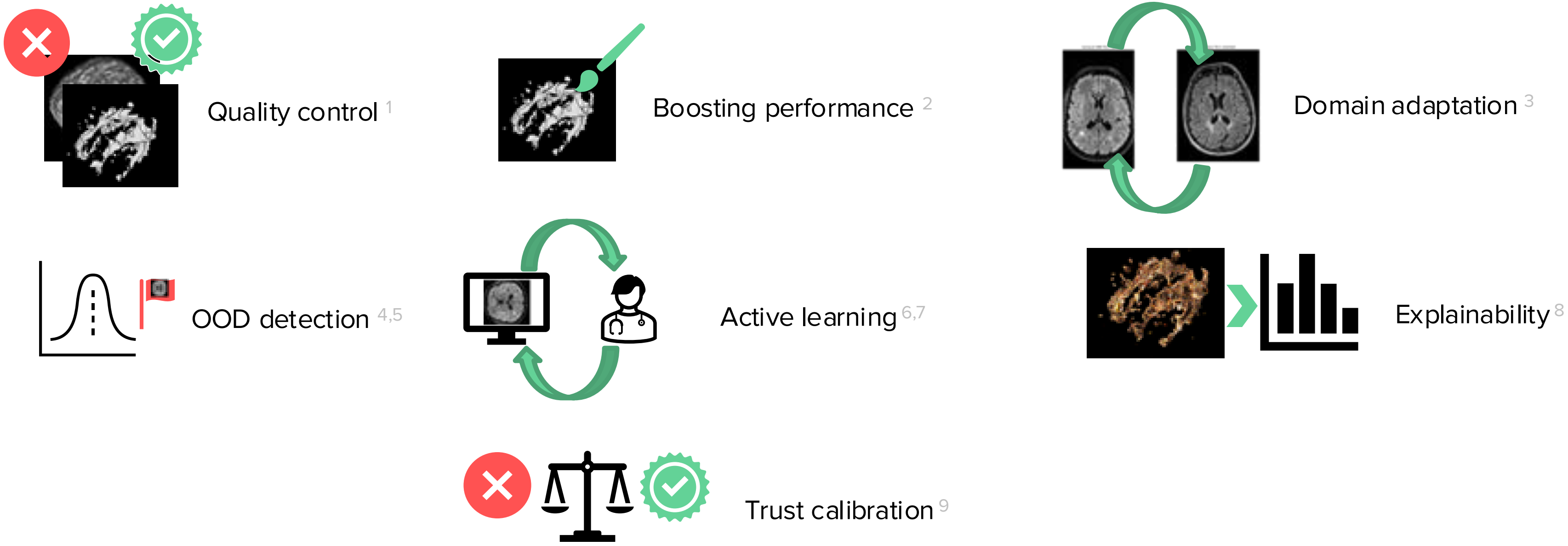
Joint uncertainty-robustness evaluation (e.g., error retention curve analysis)

Visual assessment

...

QU-BraTS. Mehta et al., MELBA, 2022; *Shifts 2.0*. Malinin et al., Arxiv, 2022; *QUBIQ*. Li et al., Arxiv, 2024

Application-specific evaluation



[1] Roy et al., NeuroImage, 2019; [2] Nair et al., Med. Image Anal., 2020; [3] Xia et al., Med. Image Anal., 2020; [4] Linmans et al., Med. Image Anal., 2023; [5] Hong et al., Arxiv, 2024; [6] Budd et al., Med. Image Anal., 2021; [7] Wang et al., Med. Image Anal., 2023; [8] Molchanova et al., MICCAI iMIMIC, 2024; [9] Evans et al., FGCS, 2022

Human-AI interaction

- Technical advancements are not sufficient for the integration [1]
- Several studies evaluated physician perception [2-4]
- Uncertainty should contribute to better trust calibration [2]
- Uncertainty is not an intuitive notion for physicians

MSxplain workshop at Lausanne University Hospital. Photo by Wen Zhan and Delphine Ribes.

[1]Kohn et al., Front. Psychol., 2021; [2]Evans et al., Future Gener. Comput. Syst., 2022; [3]Izegbua et al., PLOS, 2024; [4]Manz et al., J. Pathol.: Clin. Res., 2025

Final remarks

Final remarks

- Importance of UQ in healthcare and medical imaging
  - Acknowledged by EU Trustworthy AI and FUTURE-AI Guidelines, and human-AI interaction studies [1,2]
- Large models require lightweight robust methods
  - Compact ensembles
  - Sampling-free methods
  - Conformal prediction
- Segmentation and reconstruction are as usually lagging behind
  - Computational complexity and difficult evaluation
- Different methods communicate different “uncertainty”
  - E.g., values from DE and TTA do not necessarily communicate the same information
- Evaluation of UQ quality is still tricky even despite the consensus
  - Explaining high uncertainty values
  - Human perception evaluation

[1]EU Trustworthy AI Guidelines (2019); [2]Lekadir et al., BMJ (2025)

Q&A

UQinMIA Cookbook

Tutorial materials

YouTube Hands-on

References

Andrey Malinin. PhD Thesis. 2019

Eyke Hüllermeier and Willem Waegeman. "Aleatoric and epistemic uncertainty in machine learning: an introduction to concepts and methods". In: Machine Learning 110.3 (Mar. 2021), pp. 457–506. doi: 10.1007/s10994-021-05946-3.

Lisa Wimmer et al. «Quantifying Aleatoric and Epistemic Uncertainty in Machine Learning: Are Conditional Entropy and Mutual Information Appropriate Measures?» UAI'23.

Jakob Gawlikowski et al. "A survey of uncertainty in deep neural networks". In: Artificial Intelligence Review (July 2023), pp. 1–77. doi:10.1007/s10462-023-10562-9.

Nikolaus Weiskopf et al. «Quantitative magnetic resonance imaging of brain anatomy and in vivo histology». Nature Reviews Physics (2021)

Oliver Commowick et al. «Objective Evaluation of Multiple Sclerosis Lesion Segmentation using a Data Management and Processing Infrastructure.» Nature Portfolio (2018)

References

Nataliia Molchanova et al. “Interpretability of Uncertainty: Exploring Cortical Lesion Segmentation in Multiple Sclerosis”. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2024 Workshops. Ed. by M. Emre Celebi, Mauricio Reyes, Zhen Chen, and Xiaoxiao Li. Cham: Springer Nature Switzerland, 2025, pp. 89–98. isbn: 978-3-031-77610-6.

Benjamin Lambert et al. “Trustworthy clinical AI solutions: A unified review of uncertainty quantification in Deep Learning models for medical image analysis”. In: Artificial Intelligence in Medicine 150 (Apr. 2024), p. 102830. issn: 0933-3657. doi: 10.1016/j.artmed.2024.102830.

Ling Huang et al. “A review of uncertainty quantification in medical image analysis: Probabilistic and non-probabilistic methods”. In: Medical Image Analysis 97 (Oct. 2024), p. 103223. issn: 1361-8415. doi: 10.1016/j.media.2024.103223.

Yarin Gal and Zoubin Ghahramani. “Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning”. In: Proceedings of The 33rd International Conference on Machine Learning. Ed. by Maria Florina Balcan and Kilian Q. Weinberger. Vol. 48. Proceedings of Machine Learning Research. New York, New York, USA: PMLR, June 2016, pp. 1050–1059.

Balaji Lakshminarayanan et al. “Simple and Scalable Predictive Uncertainty Estimation using Deep Ensembles”. In: Advances in Neural Information Processing Systems. Ed. by I. Guyon et al. Vol. 30. Curran Associates, Inc., 2017, pp. 1–12.

Guotai Wang et al. “Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks”. In: Neurocomputing 338 (Apr. 2019), pp. 34–45. issn: 0925-2312. doi: 10.1016/j.neucom.2019.01.103.

References

Abhijit Guha Roy et al. Bayesian QuickNAT: Model uncertainty in deep whole-brain segmentation for structure-wise quality control. In: NeuroImage 195 (2019), pp. 11–22. issn: 1053-8119. doi: 10.1016/j.neuroimage.2019.03.042.

Tanya Nair et al. Exploring Uncertainty Measures in Deep Networks for Multiple Sclerosis Lesion Detection and Segmentation. en. arXiv:1808.01200 [cs]. Oct. 2018.

Nataliia Molchanova et al. "Structural-based uncertainty in deep learning across anatomical scales: Analysis in white matter lesion segmentation". In: Computers in Biology and Medicine 184 (Jan. 2025), p. 109336. issn: 0010-4825. doi: 10.1016/j.compbiomed.2024.109336.

Theodore Evans et al. The explainability paradox: Challenges for xAI in digital pathology. In: Future Generation Computer Systems 133 (Aug. 2022), pp. 281–296. doi: 10.1016/j.future.2022.03.009.

Raghav Mehta et al. QU-BRATS: MICCAI BRATS 2020 Challenge on Quantifying Uncertainty in Brain Tumor segmentation – Analysis of ranking scores and benchmarking results. In: The Journal of Machine Learning for Biomedical Imaging 1.August 2022 (Aug. 2022), pp. 1–54. doi: 10.59275/j. melba.2022- 354b.

References

Andrey Malinin et al. Shifts 2.0: Extending The Dataset of Real Distributional Shifts. arXiv:2206.15407 [cs]. Sept. 2022b. doi: 10.48550/arXiv.2206.15407.

Hongwei Bran Li et al. QUBIQ: Uncertainty Quantification for Biomedical Image Segmentation Challenge. arXiv:2405.18435. June 2024. doi: 10.48550/arXiv.2405.18435.

Yingda Xia et al. Uncertainty-aware multi-view co-training for semi-supervised medical image segmentation and domain adaptation. Medical Image Analysis. Oct. 2020. doi:10.1016/j.media.2020.101766.

Jasper Linmans et al. Predictive uncertainty estimation for out-of-distribution detection in digital pathology. Medical Image Analysis. Jan. 2023. doi:10.1016/j.media.2022.102655.

References

Zesheng Hong et al. Out-of-distribution Detection in Medical Image Analysis: A survey. arXiv:2404.18279. doi:10.48550/arXiv.2404.18279.

Samuel Budd et al. A survey on active learning and human-in-the-loop deep learning for medical image analysis. In: Medical Image Analysis. July 2021. doi:10.1016/j.media.2021.102062.

Hoarang Wang et al. A comprehensive survey on deep active learning in medical image analysis. In: Medical Image Analysis. July 2024. doi:10.1016/j.media.2024.103201.

EU. Ethics guidelines for trustworthy AI | Shaping Europe's digital future. Tech. rep. 2019. url: https://digital- strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.