CIBM
Center for Biomedical Imaging

CIBM Annual Symposium 2024
Forum Rolex Learning Center, EPFL, Lausanne Switzerland | 7th November 2024
20th Anniversary

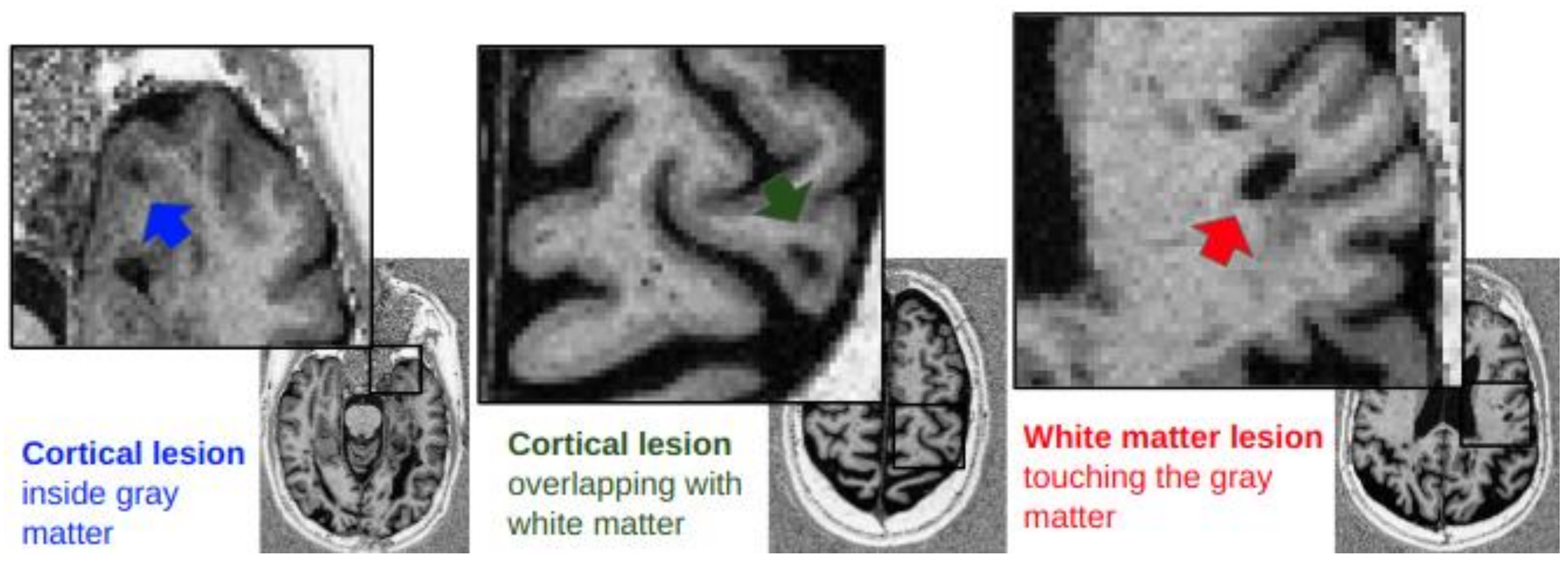
Interpretability of Uncertainty: Exploring Cortical Lesion Segmentation in Multiple Sclerosis

Nataliia Molchanova[1,2,3], Alessandro Cagol[4,5], Pedro M. Gordaliza[1,3], Mario Ocampo-Pineda[4], Po-Jui Lu[4], Xinjie Chen[4], Matthias Weigel[4], Adrien Depeursinge[1,2], Henning Müller[2,6], Cristina Granziera[4], Meritxell Bach Cuadra[1,3]

[1]University of Lausanne and Lausanne University Hospital, Switzerland, [2] University of Applied Sciences Western Switzerland (HES-SO), Switzerland, [3] CIBM Center for Biomedical Imaging, Switzerland, [4] University Hospital and University Basel, Switzerland, [5] University of Genova, Italy, [6] University of Geneva, Switzerland

BACKGROUND

- **Reliability of AI Predictions:** Uncertainty quantification (UQ) measures "untrustworthiness" in AI predictions, crucial for medical applications.
- **Link to Prediction Errors:** High uncertainty often signals higher error likelihood, making UQ a key tool for quality assessment without ground-truth labels.
- **Need for UQ Interpretation:** UQ is applied in various tasks, but understanding its values for deeper insights remains underexplored.


Cortical lesion inside gray matter
Cortical lesion overlapping with white matter
White matter lesion touching the gray matter

AIMS

- This study **targets cortical lesion (CL) segmentation on MP2RAGE MRI scans**, a challenging multiple sclerosis (MS) diagnosis task due to noisy labels and class imbalance.
- **Proposed Interpretability Analysis:** We propose an analysis that uses lesion-scale uncertainty values to provide global model explanations and ensure the sanity of UQ measures, validated through clinical feedback.

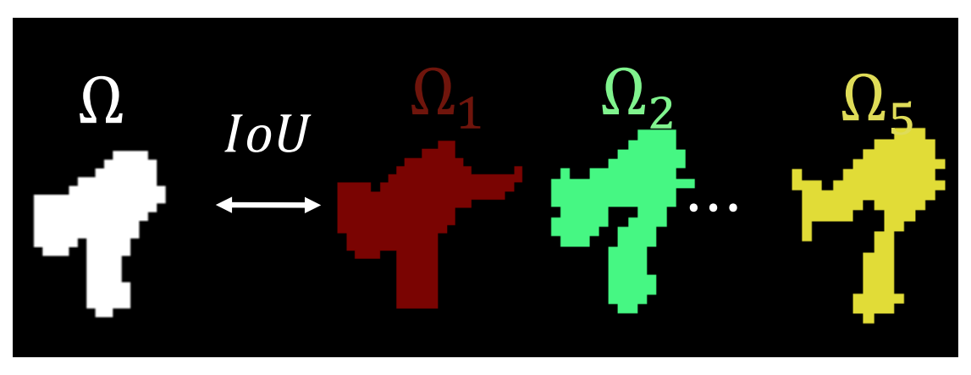
METHODS

We use a private dataset from the University of Basel, Switzerland with 117 MS patients (train:val:test=79:8:30, corresponding to 859:69:302 CLs)

1. **Training DL models capable of UQ:** For the same 3D U-Net architecture, we train a deep ensemble with K=10 members and a single model with the dropout (p=0.1) layers between each layer. K and p tuned based on the uncertainty-robustness trade-off.
2. **Lesion structural uncertainty (LSU):** To measure uncertainty associated with a predicted lesion, we use a structural disagreement between the lesion regions predicted by DE members or MCDP samples

$$LSU = 1 - \frac{1}{K}\sum_{k=1}^{K} IoU(\Omega, \Omega_k)$$

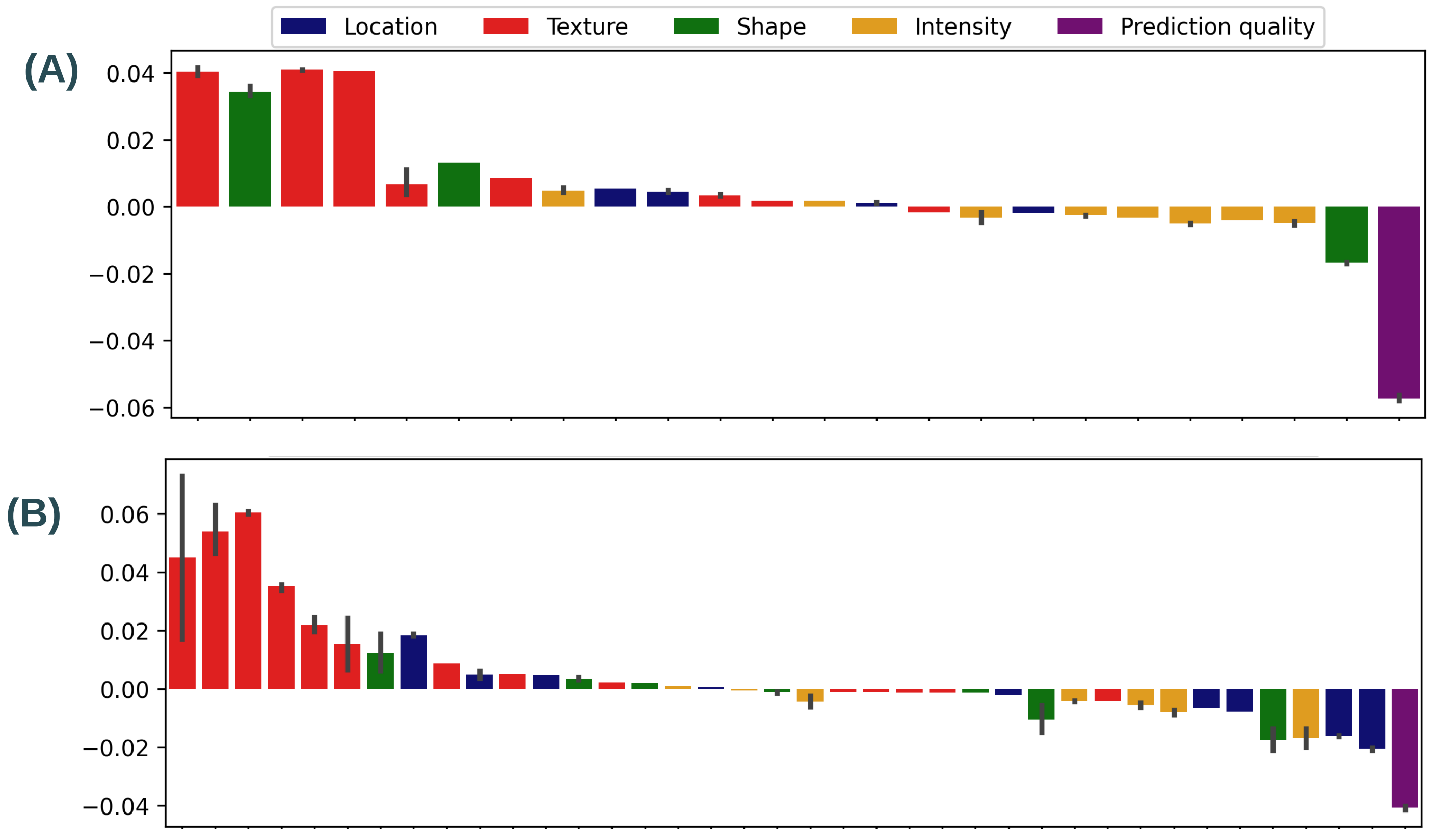
where Ω and Ω_k are lesion regions predicted by the ensemble and its members, respectively.



3. **Lesion feature engineering:** Each predicted lesion is characterized by intensity, texture, shape, location in the brain, lesion prediction quality (IoU)
4. **Uncertainty regression model:** An ElasticNet model is used to explain uncertainty values in terms of relevant lesion features. The coefficients of the linear model are interpreted as feature importance. Analysis is repeated 10 times with different random seeds to assess the standard error.

RESULTS

Feature importance: Coefficients of a linear ElasticNet model explaining variability in **DE (A) and MCDP (B)**



f$_1$ - small dependence low gray emphasis
f$_2$ - surface to volume ratio
f$_3$ - small dependence emphasis
f$_4$ - small dependence high gray emphasis
f$_5$ - max. diameter
f$_6$ and f$_7$ - temporal and occipital left lobes
f$_8$ - mean abs. deviation
f$_9$ - 90 percentile
f$_{10}$ - sphericity
f$_{11}$ - small area low gray level emphasis
f$_{12,13}$ - frontal and parietal right lobes

- **Prediction quality and lesion features explain uncertainty.** Prediction quality accounts for a great portion of the uncertainty variability, but lesion features (e.g., texture, shape, location) add more insight, highlighting complex factors beyond prediction errors.
- **Features related to high uncertainty**: i) complex textures, heterogeneous intensity patterns; ii) spiculated or elongated lesions; iii) peripheral lesion locations

Coefficient of determination R2 (↑) of ElasticNet model explaining uncertainty. Features used to fit the linear models: only prediction quality (Only IoU); all features except for the prediction quality (No IoU); all features (All).

	Cross validation (train set)			Test set		
	Only IoU	No IoU	All	Only IoU	No IoU	All
DE	0.520±0.006	0.598±0.004	0.661±0.004	0.431±0.001	0.512±0.002	0.632±0.004
MCDP	0.393±0.006	0.589±0.014	0.604±0.013	0.261±0.003	0.425±0.013	0.494±0.004

CONCLUSION

- **Clinical feedback confirms findings.** Medical doctors reviewed the lesion examples (on the left) and confirmed that the described high-uncertainty lesions as (smaller, irregularly shaped, and with complex texture blending into surroundings) are the hardest to visually identify and annotate.
- **Unexplained uncertainty** could be attributed to the non-linear relationships, lack of lesion surrounding characterisation features, noise in the data or in the UQ itself.

UNIL | Université de Lausanne
CHUV
Hes·so VALAIS WALLIS
Universität Basel
UNIVERSITÉ DE GENÈVE
UNIVERSITÀ DEGLI STUDI DI GENOVA

Financial support: Hasler Foundation (Responsible AI program)

HASLERSTIFTUNG