Reg. No: 2025/57/N/NZ7/03738; Principal Investigator: mgr Monika Żaneta Pytlarz

Abstract for General Public

Title: Developing an AI algorithm for longitudinal tracking of post-treatment changes in brain metastases using a hybrid Uncertainty-Aware Explainable (UQ-XAI) architecture accounting for inter- and intra-rater variability in segmentations.

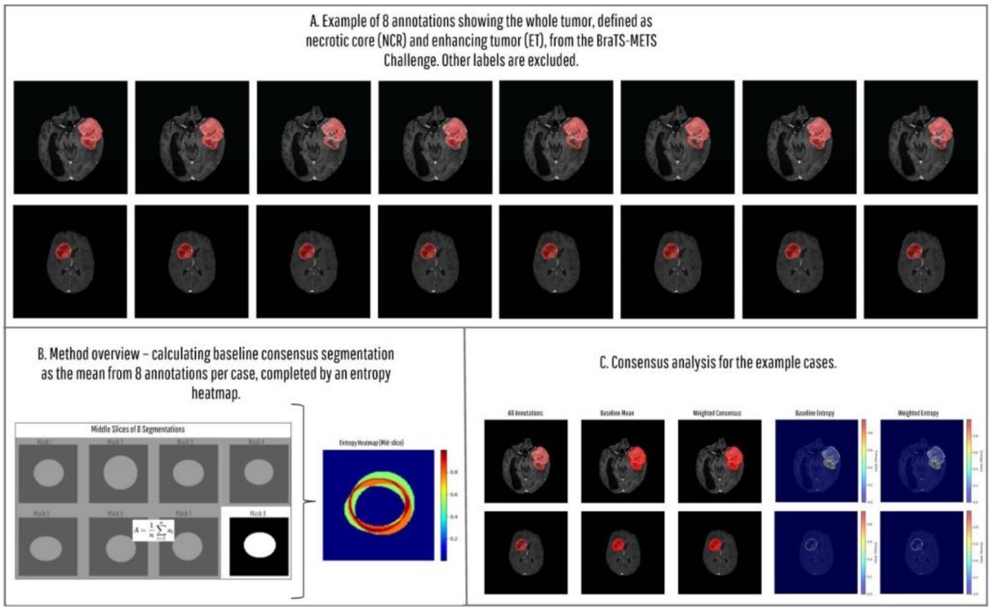
Brain metastases (BM)—tumors that spread to the brain from other parts of the body—are often monitored using MRI scans to assess treatment response. In recent years, artificial intelligence (AI) methods have shown promise in analyzing these scans, particularly in identifying and measuring tumors. However, doctors still mostly rely mainly on simple 2D measurements to evaluate BM. AI models have made progress in volumetric BM segmentation, but small lesions are still hard to detect accurately, and many AI systems are not yet used in hospitals because they work like black boxes—offering little explanation or information about how confident they are in their results.

Additionally, studies raise concerns about the quality of training data annotations, and beyond that, expert radiologists vary in their segmentations because, intrinsically, human interpretation of medical images is a span, not a single fixed ground truth. Inter- and intra-rater variability reflect the complexity and subjectivity of image interpretation, influencing the development of segmentation algorithms. Rather than treating expert disagreement as noise, it should be leveraged to model variability and enhance algorithm robustness and clinical applicability.

Our goal is to create an AI system that not only identifies brain metastases in MRI scans but also clearly communicates how confident it is about its findings and how confident it is about the explanations of the predictions given. This system will combine two key innovations: uncertainty estimation (understanding what the AI is unsure about) and explainability (visualizing why the AI made a certain decision) with the uncertainty of it. We call this approach Uncertainty-Aware Explainable AI (UQ-XAI). The system will be trained to recognize 2 types of uncertainty - those that come from the data, and those that come from the AI model itself (e.g., lack of experience with particular cases).

We are using a dataset of 75 MRI scans, each annotated 8 times by expert radiologists, and our preliminary work (Figure) of establishing a "consensus ground truth" for validation of our AI. By comparing the AI's uncertainty with actual human experts' disagreement, we aim to make its predictions more aligned with clinical reality. We will also explore whether changes in uncertainty over time, for example, if a tumor area becomes harder to define, can work as biomarker of heterogenous treatment response, even independently of whether the lesion is growing or shrinking.

The project has the potential to significantly improve how brain metastases are tracked over time. If successful, our approach will lead to AI tools that are not only more accurate but also more transparent, and trustworthy. Ultimately, in the long term, we aim to integrate this technology into hospital systems (PACS), to equip neuroradiologists with uncertainty-aware explainable volumetric segmentations when making treatment decisions. This would be a major step forward in applying AI safely and effectively in neuro-oncology.


A. Example of 8 annotations showing the whole tumor, defined as necrotic core (NCR) and enhancing tumor (ET), from the BraTS-METS Challenge. Other labels are excluded.

B. Method overview – calculating baseline consensus segmentation as the mean from 8 annotations per case, completed by an entropy heatmap.

C. Consensus analysis for the example cases.