Quantifying Uncertainty in a Tumor Segmentation Model

Claire Chu[1], Sara Colando[2], Dhruba Nandi[3], Xavier Serrano[4]

[1]University of North Carolina, [2]Pomona College, [3]Indian Statistical Institute, [4]Swarthmore College

Background

- Radiologists segment tumors from MRI scans to determine treatment plans such as surgical resection or radiation therapy.
- Neural networks can streamline the segmentation process to ensure ideal tumor removal and reduce the burden on radiologists.
- These black-box models currently lack explainability, which leads to a lack of trust from different end users like physicians and patients.

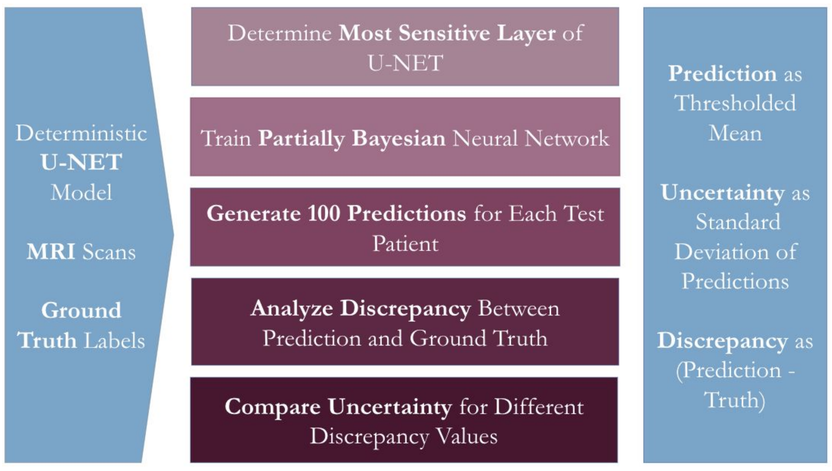
Project Goal & Research Question

Uncertainty Quantification **communicates** to stakeholders: **(a)** if and when they should **trust** model predictions. **(b)** how **fair** these predictions are on sample-wide and patient-specific cases. Therefore, Uncertainty Quantification enhances a model's transparency by **exposing** a model's properties to various stakeholders to **better understand**, **improve**, and **contest** the model's predictions.

Goal: Quantify model uncertainty by using a partially bayesian neural network to communicate where the model is uncertain of its prediction of a pixel being classified as "tumor" or "non-tumor."

Research Questions: Where is this model failing, and how is it failing to properly segment the tumor? In what cases is the model certain but still making mistakes in tumor segmentation?

Methods



Methods

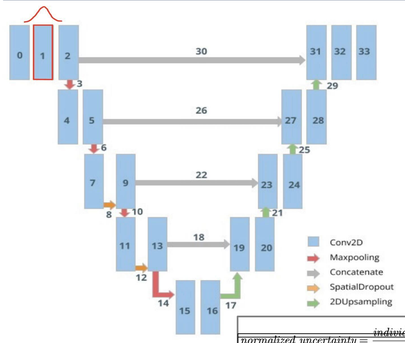


Figure 1: **UNET model** with layers labeled.

Layer 1: encoder_1a uses Bayesian Conv2DFlipout weights from the tensorflow package with a normal prior distribution centered at 0 with a variance of 100.

The weights for each prediction are generated via sampling 100 times from the push-forward posterior distribution. Any pixel with a mean less than 0.5 is classified as "no-tumor"; a pixel with a mean at or above 0.5 is classified as "tumor", The standard deviation quantifies the model's uncertainty. The uncertainty is normalized across the sample with formula below:

$$normalized\ uncertainty = \frac{individual\ pixel\ uncertainty - minimum\ uncertainty}{range\ of\ uncertainty} = \frac{individual\ pixel\ uncertainty - 0}{0.4925}$$

Results

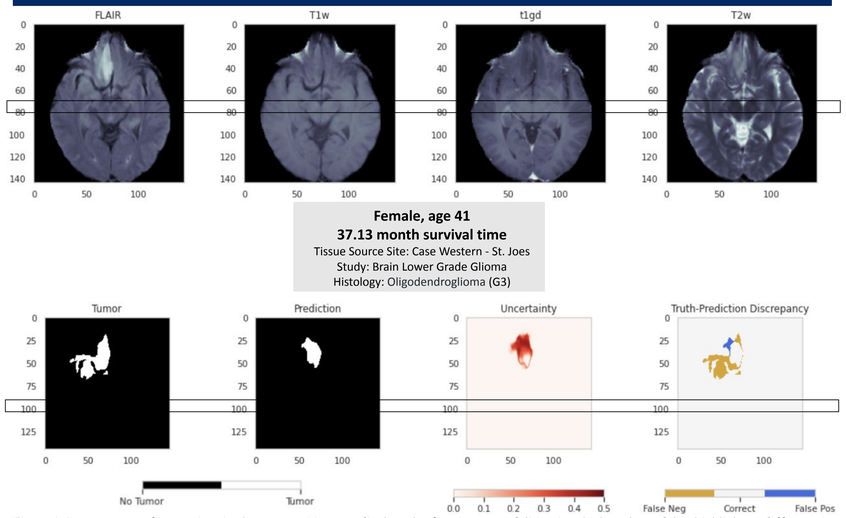


Figure 2: Image output for a patient in the test set. Top row displays the four input modalities, in which each modality highlights a different tissue type. In the bottom row, the leftmost plot is the radiologist-segmented tumor, or the "ground truth." The right three images are model-generated outputs. The prediction plot is generated by thresholding the mean of the samples at 0.5. The uncertainty plot depicts the standard deviation of the samples. Lastly, the truth-prediction discrepancy plot portrays the difference between the ground truth prediction and thresholded mean.

A pixel is classified as **False Negative** if the model identifies it as non-tumor when the true label is tumor.
A pixel is classified as **False Positive** if the model identifies it as tumor when the true label is non-tumor.

Male, age 67, 7.69 month survival time
Tissue Source Site: Thomas Jefferson University
Study: Lower Brain Grade Glioma
Histology: Astrocytoma (G3)

Female, age 70, 5.32 month survival time
Tissue Source Site: Case Western St. Joes
Study: Lower Brain Grade Glioma
Histology: Astrocytoma (G3)

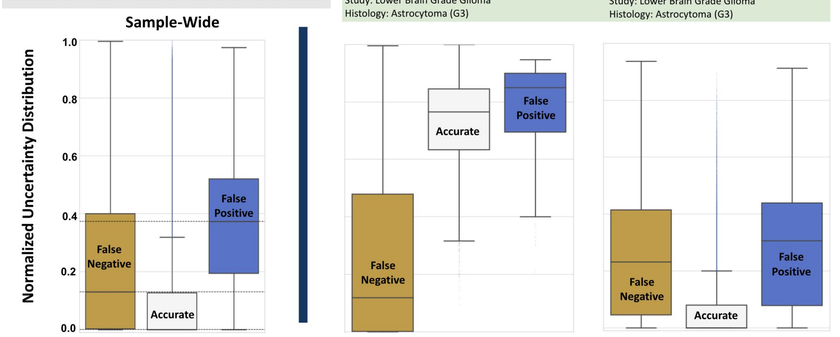


Figure 3: Box plots of the normalized uncertainty distribution across different discrepancy values (from left to right: false negative, accurate, false positive). The leftmost set of box plots represents the entire sample (sample-wide). The right two sets of box plots illustrate two patients in the test set with similar clinical features (e.g. same tumor study, histology, and grade) but different normalized uncertainty distributions, grouped by discrepancy.

Discussion

The results presented in Figures 2 and 3 depict **common patterns in the model's uncertainty**. Below are patterns and possible hypotheses.

From Figure 2:

- Highest uncertainty is found in boundary regions of the model's predicted tumor location.
  - Hypothesis: The model is most uncertain about its pixel classification near predicted tumor boundaries.
- The false negative and false positive pixels cluster in groups.

From Figure 3:

- Generally, there is higher certainty for accurately classified pixels.
- There is also greater certainty for false negatives than false positives.
  - Hypothesis: The model is more confident about classifying a pixel as "non-tumor" than "tumor."
- The sample-wide trends do not necessarily hold at patient-level.
  - Two example patients have highly similar clinical information but their normalized uncertainty distributions vary greatly.

Future Work

- Collaborating with clinicians to better understand **why model fails in specific brain regions**, and why false positive and false negative results tend to cluster.
- **Comparing** model performance and uncertainty levels **across various subsets** (e.g., different tumor histologies, tissue source sites, patient sex, vital status, etc.).
- Investigating the **implications** of the different kinds of **model failure on clinical outcomes**. Investigating what kind of model failure (i.e., false positives or false negatives) is considered more dangerous by clinicians.
- Investigating **why** the model produces **different errors** on patients with **similar clinical information.**

Acknowledgements & References

We would like to thank Dr. Nikola Banovic and Snehal Prabhudesai for their phenomenal mentorship this summer. We would also like to thank Dr. Bhramar Mukherjee and the rest of the Big Data Summer Institute for the opportunity to conduct this research.

Bhatt, Umang, Javier Antorn, Yunfeng Zhang, Q. Vera Liao, Prasanna Sattigeri, Riccardo Fogliato, Gabrielle Melançon, et al. 2021. "Uncertainty as a Form of Transparency: Measuring, Communicating, and Using Uncertainty." In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, 401–13. AIES '21. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/3461702.3462571.

Prabhudesai, Snehal, Nicholas Wang, Vinayak Ahluwalia, Xun Huan, Jayapalli Bapuraj, Nikola Banovic, and Arvind Rao. 2021. "Stratification by Tumor Grade Groups in a Holistic Evaluation of Machine Learning for Brain Tumor Segmentation." *Frontiers in Neuroscience* 15 (October). https://doi.org/10.3389/fnins.2021.740353.

Snehal Prabhudesai, Dingkun Guo, Jeremiah Hauth. 2022. "Partially Bayesian Neural Networks: Low-Cost Bayesian Uncertainty Quantification for Deep Learning in Medical Image Segmentation."