Influence of Data Provenance on Anatomical Representations in Pancreas CT Segmentation Models

[1]Harsha Pillarisetti, [1]Wes Parker, [1]Sophie Vu, [1]Samidha Nageshwar,
[1]Caleb Smock, [1]Landon Dierkes, [1]Will Jerome, [1]Dr. Rahul Gomes, [2]Dr. Sushil K. Garg
[1]Department of Computer Science – UW Eau Claire | [2]Mayo Clinic Health System

Introduction

- Advancements in Deep Learning have made image segmentation useful in quantifying tumor load CT scans.
- Clinical imaging datasets are compiled from sources with varying scanner hardware, imaging software, and acquisition protocols.
- These factors influence the data's provenance (origin and processing history).
- Data provenance influences the anatomical representations learned by segmentation models, even when standard performance metrics appear similar.

Project Goals

- Understand the influence of data provenance on learned anatomical representation in pancreas CT segmentation models.
- Quantify the discrepancies between models using overlap-based and structural metrics and develop disagreement-based analysis as a tool for identifying model bias.

Methodology

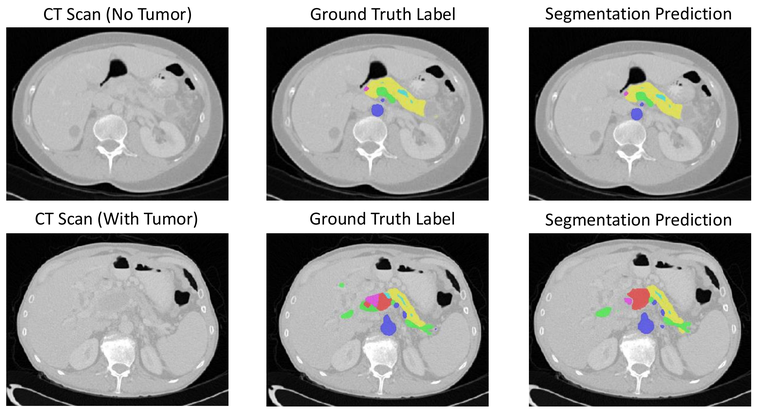


Fig. 1: CT Scan Segmentations of Abdominal Organs

- We used the [1]PANORAMA dataset with 1958 individual CT scans and labels.
- The dataset is split 4 ways by source of PDAC confirmation.
- We trained a separate [2]nnU-Net model for 500 epochs for each dataset.

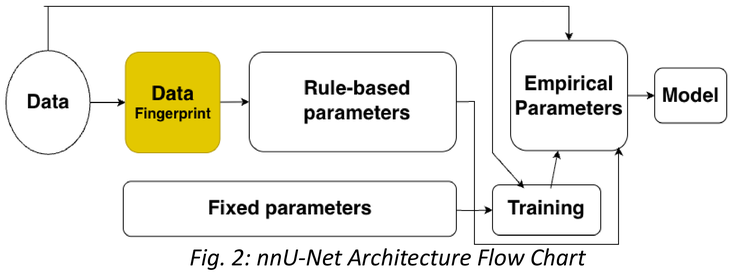


Fig. 2: nnU-Net Architecture Flow Chart

- Each model is trained and validated in-domain (using only its corresponding dataset) and tested cross-domain (on all datasets).

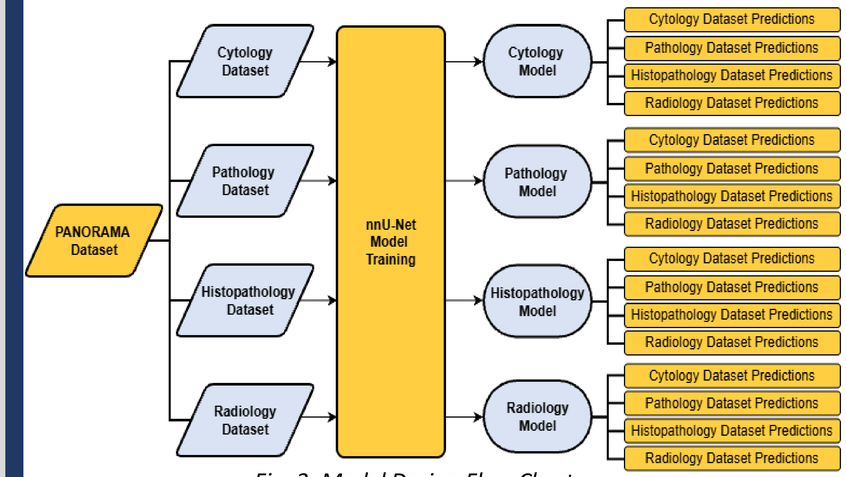


Fig. 3: Model Design Flow Chart

Results

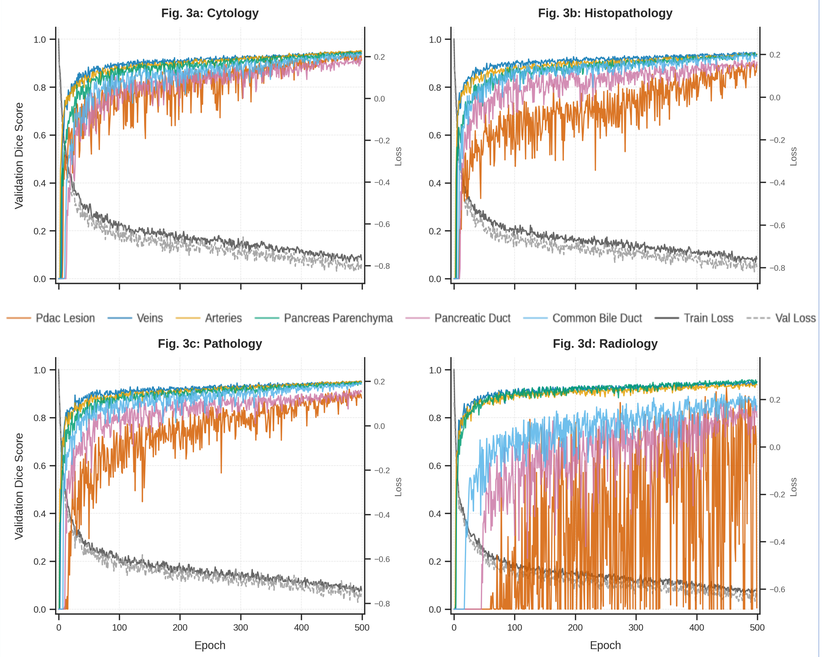


Fig. 4: Training Curves for each model

- Training curves show the accuracy (Dice score) per anatomical structure for each segmentation model.
- Pancreas, arteries, and veins achieve the highest mean Dice scores (93.8% - 95.4%) as well as the lowest standard deviations (3.3% - 4.3%) across all models, demonstrating structural robustness.
- Tumor segmentation is most variable in Radiology (37.5% Dice standard deviation) due to numerous false positive cases.

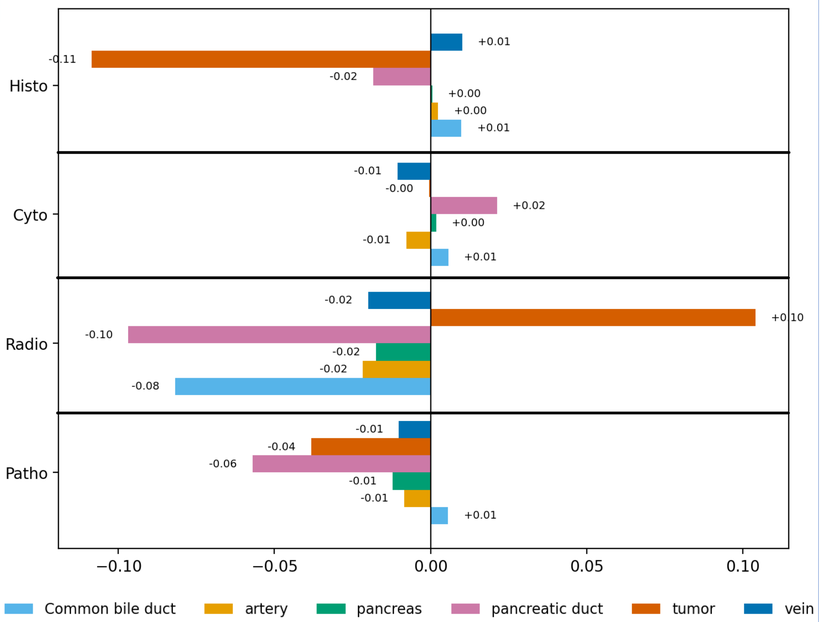


Fig. 5: Signed Dice Drop by Target Dataset and Class

- Dice Drop calculated as the mean Dice score of the out-of-domain models minus the in-domain model's own Dice score for a given anatomical structure within a dataset.
- Bars on the right of the chart are anatomical structures within a dataset for which out-of-domain models perform better on average than the in-domain model.
- Smaller structures are most negatively affected by domain transfer

Results

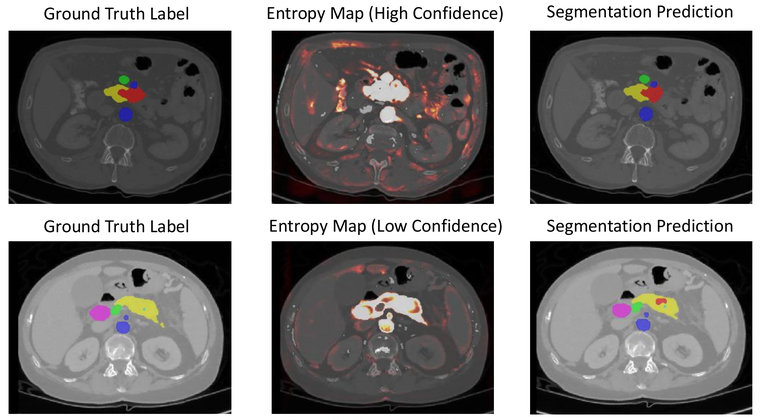


Fig. 6: Entropy Maps

- Confidence and entropy track closely across all models, indicating model confusion is typically only between two segmentation classes.
- Tumor and pancreatic duct classes are confused with pancreas in 71% and 50% of confused voxel cases respectively.
- For all other structures, confusion is almost exclusively with background.

Structure	Confused With	Rate
Tumor	Pancreas	**71.0%**
Pancreatic duct	Pancreas	**50.0%**
Common bile duct	Tumor	**14.9%**
Pancreatic duct	Tumor	**9.9%**
Vein / Artery	(background only)	**0%**

Table 1: In-domain cases where top confusion class was not background

Key Findings

- Pancreas, arteries, and veins show the most resistance to differences in data provenance.
- Tumors and pancreatic ducts are the most source-sensitive structures.
- Some cross-domain models outperform in-domain baselines, indicating beneficial transfer.
- Data provenance most impacts the segmentation of small, inconsistently-visible structures.
- Low cross-domain Dice drop shows models can generalize across sources, suggesting shared anatomical representations of larger structures.
- Greatest obstacle to PDAC segmentation is the boundary between lesion and pancreas.

Future Work

- Develop voxel-based normalization (scaling counts to [0,1]) to enable fair structural comparison across scan types with differing resolutions and acquisition protocols.
- Determine ideal data provenance for experimentation with synthetic image generation using PANORAMA data

References

[1] Roth, H., Farag, A., Turkbey, E. B., Lu, L., Liu, J., & Summers, R. M. (2016). Data From Pancreas-CT (Version 2) [Data set]. The Cancer Imaging Archive.
[2] Isensee, F., Jaeger, P.F., Kohl, S.A.A. *et al.* nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods* **18**, 203–211 (2021).

Acknowledgements

This research project was funded by the National Science Foundation (NSF)-Research Experience for Undergraduates OAC-2447779. The computational resources of the study were provided by the Blugold Center for High-Performance Computing under NSF grant CNS-1920220

We thank the Office of Research and Sponsored Programs for supporting this research, and Learning & Technology Services for printing this poster.