Evaluating Pancreas and Pancreatic Ductal Adenocarcinoma Tumor Segmentation Methods

Anya Ranavat[1], Adeel Siddiqi[2], Amanda Warren[3], Dr. Rahul Gomes[*]

1 University of Texas at Austin, Statistics and Data Science, 2 University of Central Florida, Computer Science, 3 Brigham Young University, Bioinformatics,

* University of Wisconsin – Eau Claire, Computer Science

Introduction

Pancreatic ductal adenocarcinoma (PDAC)

- **Deadliest forms** of pancreatic cancer
- Exceptionally **low 5-year survival rates**
- Represents **more than 90%** of pancreatic cancer cases
- Tumors are small, deeply located in the abdomen, and blend in with other soft tissues

PDAC's **high mortality rate** stems from its **difficult detection** by clinicians, which often means it's **diagnosed too late** for most treatments to be effective.

Importance of AI Segmentation Models

- Highlight tumor regions with **high accuracy** to enhance PDAC detection
- Enables clinicians to **catch PDAC sooner,** when therapies are most effective
- Validated reliable outputs **bolster confidence** for both clinicians and patients
- Rapid automated segmentations **accelerates review and decision-making** in busy clinical workflow

This project aims to identify which model architectures are most effective for **accurately detecting tumor signature**.

Dataset

Medical Segmentation Decathlon Pancreas Task dataset (MSD) comprises **226 three-dimensional, contrast-enhanced CT volumes** and corresponding segmentation masks obtained from PDAC patients in the portal-venous phase. Each volume features 512 × 512 pixel in-plane resolution with slice counts varying across scans, reflecting differences in patient anatomy and acquisition protocols. The scans were acquired at Memorial Sloan Kettering Cancer Center in New York include annotations for background, healthy pancreas, and PDAC tumor regions.

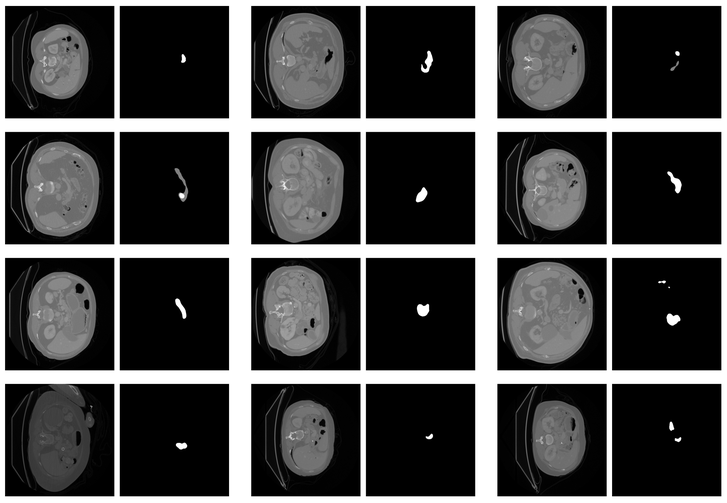


Figure 1. Example middle slice for image-mask pairs in the MSD dataset

Methods

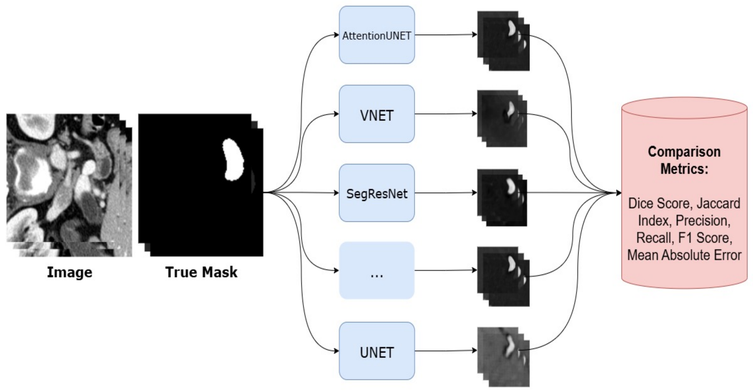


Figure 2. Visual Representation of Model Comparison Process

Dataset Preprocessing

Each image–mask volume pair was cropped to the contiguous **span of slices containing pancreatic tissue or tumor**, with a 20-slice margin added on either side to preserve contextual information. Patient cases were then randomly allocated into a **70 % training set and a 30 % testing set** to facilitate model development and unbiased evaluation.

Training Transformations

- **Orientation:** Rotate all scans into standard RAS coordinates
- **Intensity normalization:** Clamp Hounsfield units in images to [-100, 300] and normalize values linearly to [0, 1]
- **ROI:** Center crop (or pad) volumes from (512 x 512 x n) to (128 x 128 x 64)
- **Patch sampling:** For each scan, extract six random (64 x 64 x 64) patches – three containing pancreas or tumor and 3 containing no pancreas or tumor

Testing Transformations

- Apply the same orientation, intensity normalization, and resizing as in training, while omitting random patch sampling

Proposed Models

- UNET
  - Classic 2D encoder–decoder with skip connections—serves as our baseline for capturing multi-scale features.
- SegResNet
  - 3D residual UNet variant that uses volumetric convolutions and identity mappings to improve gradient flow.
- SegResNetDS
  - Builds on SegResNet by adding deep-supervision at intermediate decoder layers, refining feature learning at multiple scales.
- AttentionUNET
  - Integrates attention gates into the UNet architecture to automatically weight and highlight relevant tumor regions.
- UNETR
  - Combines a CNN encoder with transformer-based bottleneck (VIT) for long-range context modeling, capturing anatomical structures
- VNET
  - Fully 3D volumetric network with residual blocks and Dice-optimized loss, designed for dense medical image segmentation

Results

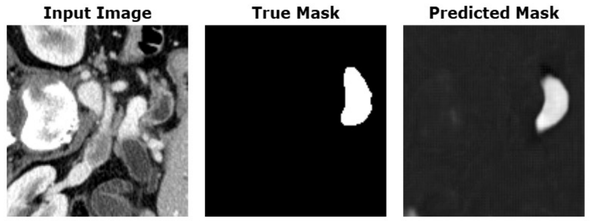
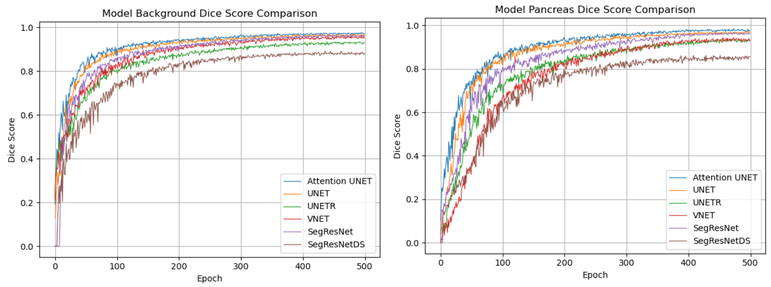


Figure 3. Single Slice Visualization of Data Utilized in Training Models



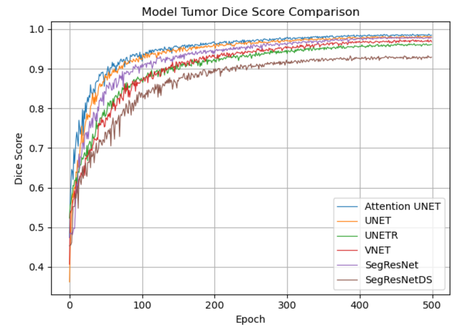


Figure 4. Dice Score of Each Proposed Model

Model	Best Background Score	Best Pancreas Score	Best Tumor Score	Best Tumor Score Epoch
Attention UNET	**0.976**	**0.984**	**0.988**	**474**
UNET	**0.969**	**0.972**	**0.983**	**481**
UNETR	**0.934**	**0.938**	**0.963**	**476**
VNET	**0.961**	**0.943**	**0.973**	**495**
SegResNet	**0.966**	**0.966**	**0.98**	**479**
SegResNetDS	**0.892**	**0.862**	**0.933**	**495**

Figure 5. Best Dice Score Value Comparison of Each Proposed Model

Future Work

- Investigate which **architectural choices, training strategies, or data characteristics** cause certain segmentation models to outperform others
- Training on a **larger and more diverse collection of scans** is necessary to better assess performance and generalizability
- **Extract the feature matrix** from the bottleneck layer of the best performing segmentation model for use in **training an autoencoder model**

Acknowledgments

This research was funded by the National Science Foundation (NSF) Research Experience for Undergraduates (REU) grant OAC-2447779. We would also like to thank the Office of Research and Sponsored Programs (ORSP) at UW-Eau Claire for student support. The computational resources of this study was 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.