Automated Surgical Target Volumes for Enhanced Neck Dissection Planning in Head and Neck Squamous Cell Carcinoma

Jintong Li[1], Jose O. Romero[1], Jerome M. Tullo[1], Travis T. Byrum[1], Daniel Shen[1], Joseph A. Paydarfar[13], David A. Pastel[4], Indrani Bhattacharya[2]
[1]Geisel School of Medicine & [2]Department of Biomedical Data Science at Dartmouth, [3]Section of Otolaryngology & [4]Department of Radiology, Dartmouth Hitchcock Medical Center

Background

- Head and neck squamous cell carcinoma (HNSCC) causes high global morbidity and mortality, with around 890,000 cases annually.

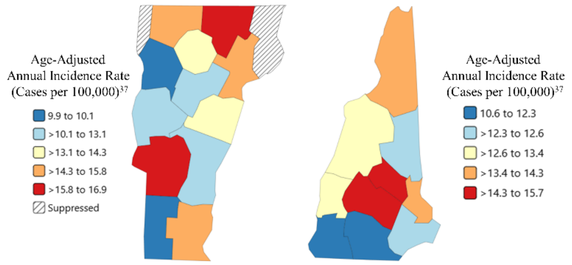
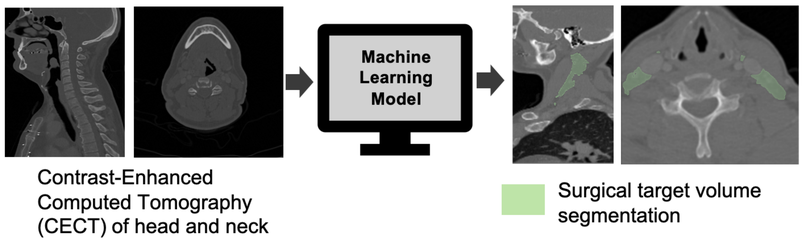


Fig. 1: Incidence rates for oral cavity and pharyngeal cancer (all stages) in Vermont (left) and New Hampshire (right) between 2017-2021

- **Data limitations:** No public labeled datasets exist for surgical target volumes (STVs) in HNSCC, and manual annotation is time-consuming (~1–2 hours per case), making **semi-supervised learning** essential.
- **Need for innovation:** Existing machine learning (ML) models are radiotherapy-focused and include non-resectable structures, missing smaller nodes and surrounding fibrofatty tissue—the true surgical target.

Objective

Aim 1: To develop a machine learning (ML) method for automated segmentation of lymphatic surgical target volumes (STVs)



Aim 2: To evaluate the model's performance and clinical utility.

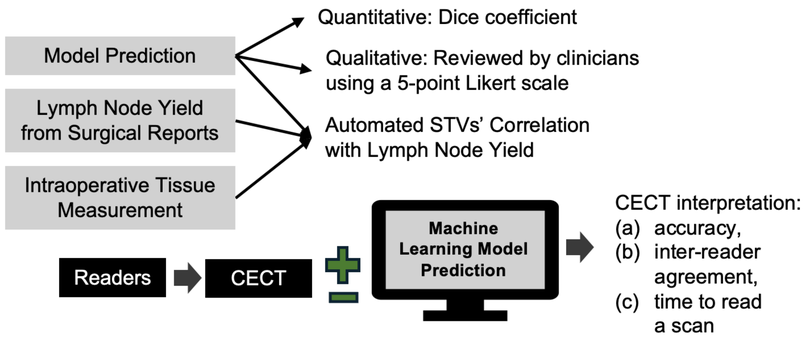


Fig. 2: Proposed specific aims

Method

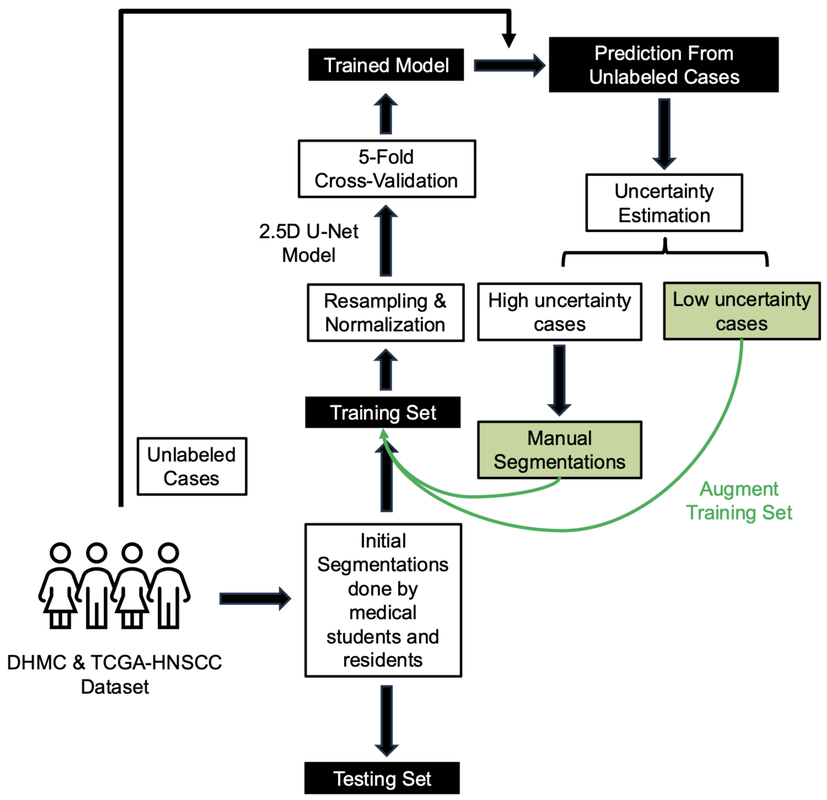


Fig 3: The proposed system integrates hybrid uncertainty-guided active learning with semi-supervised learning to generate accurate, generalizable predictions from limited labeled and abundant unlabeled data.

- As an initial step, we trained a **2.5D U-Net model** to automatically generate STVs, using 40 contrast-enhanced CT scans from DHMC and the TCGA-HNSCC dataset.

Result

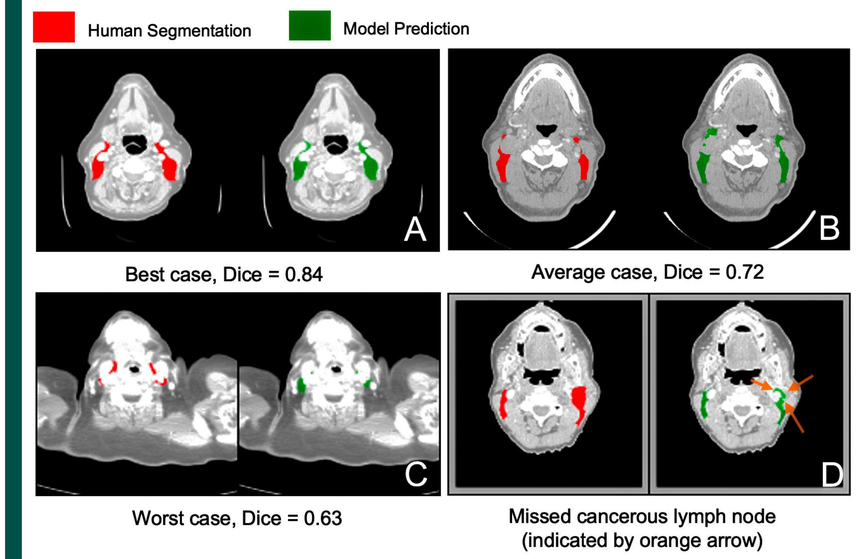


Fig. 4: Preliminary results from our baseline model show encouraging surgical target volume segmentation on a held-out test set. (A-B) The model segmentation (green) closely follows the human segmentation (red), despite non-perfect Dice scores. (C-D) However, the model can under-segment and sometimes miss cancerous nodes.

- We found that manual STVs correlated strongly with:
  1. Lymphadenectomy specimen mass (r = 0.88)
  2. Postoperative lymph node counts (r = 0.66)
- On an independent test set of 10 DHMC + 10 TCGA cases, the model achieved a Dice score of **0.72 ± 0.06**, showing promising early results.
- The trained model is **under-predicting** volumes, and we envision the performance to improve when larger training sets are included using our proposed hybrid uncertainty-guided active and semi-supervised learning framework.

Future Steps

- **Model refinement**: Use **active and semi-supervised learning** with Monte Carlo dropout to select high-uncertainty cases for labeling and low-uncertainty cases for pseudo-labeling.
- **Qualitative evaluation**: Dr. Pastel will rate 40 segmentations using a 5-point Likert scale for accuracy, edit effort, and time saved.
- **Clinical utility**: Perform **multi-reader** studies comparing manual vs. automated STVs for accuracy, efficiency, variability, and clinical confidence.

Acknowledgement

- We gratefully acknowledge the support of the Prouty Pilot Research Grant to DAP, JAP, and IB, and the Pressman Fellowship to JOR, which made this work possible.