Multi-Phase Mandible-Anchored Automated Segmentation of Oropharyngeal GTVs in FDG-PET/CT

Dominic LaBella[1][0000-0003-1713-9538], Keshav Jha[2] [0009-0003-2163-1290], Kevin Goff[1][0000-0001-5862-0219], Esther Yu[1], Jared Robbins[1][0000-0003-1724-5242]

[1] Department of Radiation Oncology, Duke University Medical Center, Durham, NC 27710, USA
[2] Duke University, Durham, NC 27710, USA

Abstract. Accurate delineation of gross tumor volumes (GTVs) in oropharyngeal carcinoma remains central to radiotherapy (RT) planning, and recent advances in automated segmentation are beginning to influence multiple aspects of clinical workflow. However, automated segmentation of head and neck primary tumors and nodal metastases in multi-modal FDG-PET/CT is challenging due to anatomical complexity and varying image resolution. We describe the DLaBella29 team's approach for Task 1 of the MICCAI 2025 HECKTOR challenge, which involves a multi-phase deep learning pipeline for GTVp and GTVn segmentation. Our method leverages the MONAI Auto3DSeg framework with a 3D SegResNet backbone, trained on co-registered PET/CT scans using a cross-validation strategy (3/7/15 folds across Phases). Key innovations include a mandible-anchored region-of-interest cropping strategy derived from automated mandible segmentation to focus the model on the oropharyngeal region and improve efficiency. A multi-phase segmentation pipeline was employed: an initial Phase infers GTVp and GTVn in the focused anatomical region-of-interest, and a second Phase independently refines these predictions. Post-processing ensues to merge GTVp and GTVn outputs to the original image coordinates. On the HECKTOR 2025 test set, our more limited and resource-constrained algorithm achieved a primary tumor Dice Similarity Coefficient (DSC) of 0.5289, an aggregated nodal DSC of 0.6156, and a GTVn detection F1-score of 0.5561, indicating fair performance in the challenge. This paper details the clinical context, methodology, cross-validation and testing results, and the implications of PET/CT-guided automated segmentation for oropharyngeal carcinoma.

Keywords: Oropharyngeal Cancer, FDG-PET/CT, Automated Segmentation, MONAI, Auto3DSeg, SegResNet, Radiotherapy Planning, Dose De-escalation

1 Introduction

Head and neck cancers are a major global health burden, motivating continuous efforts to optimize radiation therapy (RT) effectiveness while minimizing treatment morbidity [1]. Studies have shown that intensity-modulated RT with deliberate organs-at-risk (OAR) sparing techniques achieved toxicity reduction without compromising local control, underscoring the importance of anatomy-guided dose sculpting and target delineation [2–4].

Fludeoxyglucose positron emission tomography (FDG PET) and computed tomography (CT) are now integral across the head and neck cancer (HNC) care continuum, from staging and RT planning to response assessment, and quantitative PET signals have been linked to prognosis and treatment tailoring [5,6]. PET-derived and CT-derived radiomics features capture complementary aspects of tumor biology relevant to local control modeling; in comparative studies, PET-based models often provide more reliable risk estimates for adverse outcomes [7]. Multi-feature radiomics strategies have also been shown to stratify failure risk in HNC, reinforcing the promise of image-derived biomarkers for individualized therapy [8].

These advances have motivated selective de-escalation strategies in HPV-associated oropharyngeal squamous cell carcinoma (OPSCC), designed to reduce late effects without compromising disease control. Prospective data indicate that thoughtfully implemented de-escalation can maintain excellent oncologic outcomes while improving quality of life [9]. Building on this, a recent phase II trial used mid-treatment FDG-PET response as a biomarker to selectively reduce RT dose in early-stage HPV-positive OPSCC, reporting feasibility with encouraging disease-control and toxicity profiles [10]. Together, these data frame FDG-PET as both a planning tool and a dynamic response indicator to guide patient-specific de-intensification.

Adaptive radiation therapy (ART) provides a paradigm for customized daily radiotherapy. By accommodating day-to-day anatomic and volumetric variation, ART can improve target coverage while reducing dose to OARs which may ultimately improve outcomes [11–13]. ART workflows require frequent on-treatment imaging, deformable mapping, and rapid plan adaptation; their clinical scalability increasingly depends on accurate, automated contouring of both gross tumor volumes (GTVs) and OARs to shorten decision cycles [14]. In this context, daily deep-learning automated segmentation of GTVs, ideally integrated with routine image guidance, could enable response-adapted thresholds and ART triggers while maintaining consistency in normal-tissue delineation [14].

Community challenges have been pivotal in stress-testing these technologies on a large scale. The HEad and neCK TumOR (HECKTOR) series, initiated at MICCAI, standardized multicenter FDG-PET/CT datasets and performed objective evaluation to benchmark automatic tumor segmentation [15,16]. The MICCAI 2021 edition combined primary-tumor segmentation with progression-free survival prediction, accelerating translation of robust PET/CT pipelines [17]. More recently, HEad and neCK TumOR Lesion Segmentation, Diagnosis and Prognosis Using Multimodal Data Fourth Edition Challenge (HECKTOR25) extends this agenda on the Grand-Challenge platform, focusing on multimodal PET/CT lesion segmentation (including primary and nodal disease) with containerized submissions and standardized metrics to ensure fair, reproducible comparison across methods [15,16]. This iteration emphasizes clinical applicability through large, multi-institutional data and harmonized evaluation protocols [15,16].

We present *Multi-Phase Mandible-Anchored Automated Segmentation of Oropharyngeal GTVs in FDG-PET/CT*, an algorithm designed for HECKTOR25 that aims to produce anatomically consistent, high-fidelity delineations suitable for PET-guided de-escalation studies and ART workflows. Using a

CT-based mandible segmentation to define a stable oropharyngeal anchor, our approach crops to a biologically and anatomically focused field-of-view; a subsequent multimodal PET/CT model performs joint coarse GTVp and GTVn segmentation, followed by dedicated refinement models for primary and nodal disease, with fusion back to canonical space. This design is intended to: (i) improve localization stability across heterogeneous acquisitions, (ii) reduce silent failure modes in high-complexity neck anatomy, and (iii) furnish daily-ready GTV contours that can be paired with automated OAR delineation to support ART decision-making. The method is implemented with open-source components (MONAI, Auto3DSeg, and SegResNet) and evaluated in the HECKTOR25 framework; methodological details and analyses are provided in the subsequent sections [18,19].

2 Methods

2.1 Dataset Selection

The training set consisted of co-registered CT and FDG-PET scans from patients with head and neck cancer, with expert contours for the primary gross tumor volume (GTVp) and any pathologic gross lymph nodes (GTVn); the data were provided as part of the HEad and neCK TumOR Lesion Segmentation, Diagnosis and Prognosis Using Multimodal Data Fourth Edition Challenge (HECKTOR25) [15,16]. Every CT and FDG-PET case included the patient's head and neck within the field of view, and a subset of cases also included the chest, abdomen, and pelvis. All CT and FDG-PET image pairs were already registered, as provided by the challenge, but we resampled all images to a consistent and common sampling of 0.9766 × 0.9766 × 1.5 mm. We preserved the original images origin and direction metadata such that predictions made on resampled and cropped images could be transformed back to the original CT space for submission without loss of spatial correspondence.

Additionally, we generated a synthetic reference standard mandible label for all cases in the training set. The synthetic reference standard mandible label was generated using TotalSegmentator on each case's CT to first segment major anatomical structures, from which we extracted the mandible (jaw bone) mask [20]. TotalSegmentator is a robust, pretrained model capable of segmenting over 100 structures on CT, and upon qualitative review, it provided a reliable mandible label in the training set data [20]. We did not manually modify any mandible label within the training set, as this would have created an unallowable expert annotated private dataset for training, contradictory to challenge rules.

2.2 Mandible Segmentation and Cropping Strategy

A key aspect of our method is focusing the Phase 1 segmentation task on a specific anatomical region-of-interest (aROI) around the mandible-anchor-point, since the oropharyngeal primary tumors in this dataset typically occur in anatomical proximity to the mandible (e.g. tonsillar region, base of tongue) and involved lymph nodes are often in levels II-III of the neck. We therefore hypothesized that using the mandible-anchor-point as a landmark could define a cropped volume that still contains all relevant targets (GTVp and GTVn) while excluding irrelevant areas (e.g. chest, abdomen) that increase complexity and inference time.

To define a reproducible anatomical landmark for Phase 1 cropping, we utilized the aforementioned synthetic mandible label. From this mandible label, we identified the most anterior midline voxel, which corresponds to the anterior symphyseal point of the mandible. This voxel, termed the mandible-anchor-point, was chosen because (i) it is highly consistent across subjects and (ii) it lies immediately anterior to the oropharynx, providing a stable reference for downstream aROI placement.

To ensure that the Phase 1 cropped field-of-view always encompassed the full aROI, we compute, for every training subject, the 3D Euclidean distance between the mandible-anchor-point and the minimum/maximum coordinates of all expert-annotated GTVp and GTVn contours. These distances (left/right, anterior/posterior, cranial/caudal) were aggregated across the 682 training cases. We then selected fixed, symmetric safety margins that exceeded the largest observed GTV extent in each direction. The resulting expansions (**Table 1**) defined a standardized mandible-anchored aROI that covered all primary tumors and nodal levels II-III without including irrelevant thoracic or abdominal structures.

Table 1. Fixed margin expansions for cropped Phase 1 images from the mandible-anchor-point in the initial resampled uncropped image space. The new aROI was used for Phase 1 training and inference. Note that both -y and +y expansions were in the anatomical posterior direction since no significant amount of GTV were identified anterior to the mandible-anchor-point.

Direction	**Distance (voxels)**	**Distance (mm)**
-x (left)	75	73.3
+x (right)	75	73.3
-y (posterior)	143	139.7
+y (anterior)	-23	-22.5
-z (caudal)	72	108
+z (cranial)	52	78

After expansion from the mandible-anchor-point, this aROI was additionally bounded by the case's CT image dimensions to prevent over-expansion to full-sized image out-of-bound regions. Each case's FDG-PET and CT were cropped to this cropped volume, yielding a significantly smaller volume for subsequent processing. An illustration of this mandible-anchored aROI pipeline is provided in **Figure 1**, which shows adequate coverage of the primary and nodal tumor regions while excluding unrelated anatomy. Importantly, during inference, we retained the original spatial coordinates of the aROI so that any segmentation outputs could later be placed back into the full image grid at the correct location. This geometry-preserving strategy avoids any misalignment when fusing the final GTVp and GTVn masks into the original coordinate system.

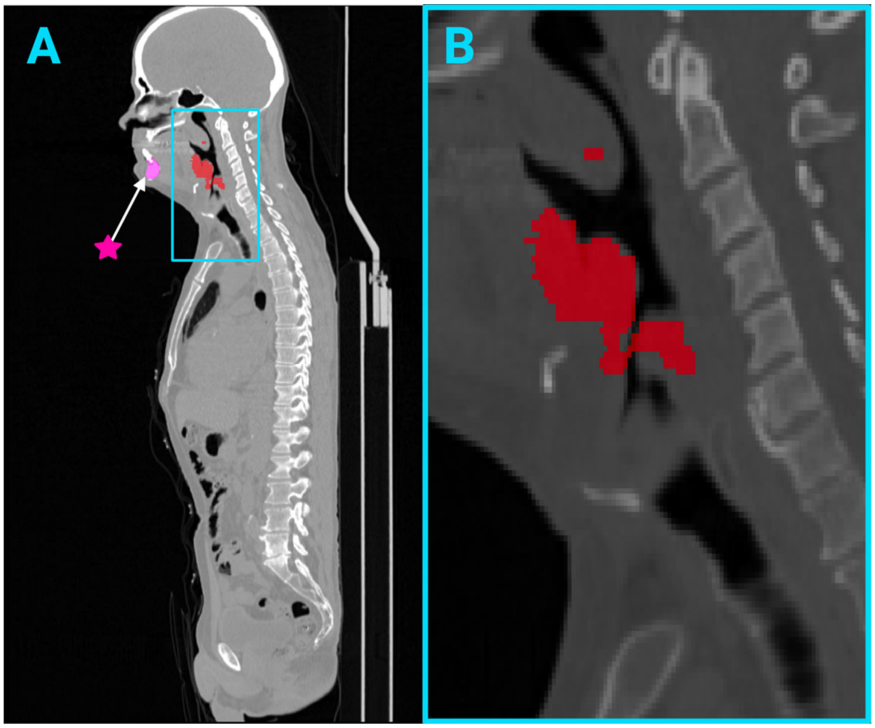

Fig. 1. Mandible-anchored field-of-view selection. (A) Whole-volume sagittal CT (512 × 512 × 568 voxels) resampled to 0.9766 × 0.9766 × 1.5 mm spacing, showing the mandible (purple) and the mandible anchor point (pink star) with expert annotated GTVp (Red) and the mandible-anchored aROI (light-blue box). (B) Corresponding cropped sagittal subvolume (151 × 121 × 125 voxels) centered on the anchor, used for Phase 1 inference.

2.3 Model Architecture, Training, and Inference

We implemented a four-phase Auto3DSeg pipeline using a 3D SegResNet backbone (18,19), that included segmentation of the mandible (Phase 0), coarse segmentation of GTVp and GTVn (Phase 1), and independent refinement stages for GTVp (Phase 2p) and GTVn (Phase 2n). Across all stages, the network used deep supervision with dsdepth of 4, an encoder-decoder with 1-2-2-4-4 convolutional blocks, Dice + Cross-Entropy loss, automatic mixed precision (AMP), AdamW optimization algorithm, and a batch size of 1. Data augmentation included random rotations in 3D, random intensity scaling and random intensity shifting to improve generalization. The CT image served as the canonical reference throughout the pipeline.

All training and inference ran on a single NVIDIA RTX 4090. **Table 2** shows varying training parameters and metrics identified during each phase's training. We trained using cross-validation for each phase, and we evaluated performance on the held-out fold using Dice Similarity Coefficient (DSC) for both GTVp and GTVn. The number of folds used for each Phase was determined based on the amount of estimated available training and inference time due to real-world time constraints. Grand-Challenge only allows 10 minutes for inference for each case using an NVIDIA T4 GPU with 16 GB of available VRAM, of which 8 GB is shared amongst multiple processes [15].

Phase 0: Mandible (1 fold; CT-only).
For the mandible model, we trained a single-fold SegResNet on CT volumes resampled to 3 mm isotropic resolution, using the synthetic reference standard mandible labels. Training used an input tensor spatial dimension region-of-interest (ROI) size of 192 × 192 × 64 voxels for 19 epochs, and with a learning rate of 4e-4. This larger 3 mm resampling was selected due to the ease of mandible segmentation owing to its larger anatomical size and high contrast relative to nearby anatomy. Because the structure served only as an approximate guide for subsequent Phase window placement, no additional training beyond 19 epochs was performed. At inference, the mandible is predicted from the full RAS-canonical CT image, and the anterior mandible voxel is used as an anchor to compute the Phase 1 crop window.

Phase 1: Coarse GTVp + GTVn (3 folds; CT + PET).
For the coarse tumor model (joint GTVp and GTVn), we trained three folds using cross-validation on CT-PET pairs that are resampled to 0.9766 × 0.9766 × 1.5 mm and cropped using the mandible-anchored Phase 1 cropped subvolume, with training ROI size of 128 × 112 × 112 voxels, a learning rate of 4e-4, training for 137 epochs. At inference, each of the three folds is applied to the same RAS-canonical crop and fused with MultiLabel STAPLE to produce a consensus Phase 1 GTVp and GTVn label map in canonical space [21].

Phase 2p: Focused GTVp (15 folds; CT + PET).
For primary tumor refinement, we trained using a dataset including a tight GTVp cropped image by taking the Phase 1 GTVp bounding box and expanding it by 10 mm in x/y/z directions (**Figure 2**). We used a ROI size of 48 × 48 × 48 voxels, resampled to 0.9766 × 0.9766 × 1.5 mm, used a learning rate to 2e-4, and trained for 137 epochs. We trained 15 folds using cross-validation, and at inference we ran all fold's models on the focused crop region and fused them with MultiLabel STAPLE; the fused GTVp then replaced the Stage 1 tumor prediction inside the focused region, with tumor label assignment prioritized over surrounding structures, before reintegration onto the full canonical grid.

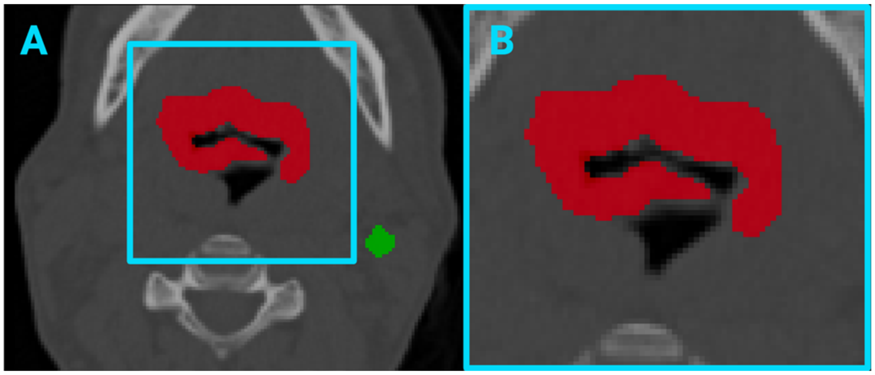

Fig. 2. GTVp-focused field-of-view refinement. Axial CT (A) from the Phase 1 mandible-anchored crop showing expert-annotated GTVp (red) and GTVn (green); the light-blue box denotes the Phase-2p, GTVp-centered aROI. Corresponding cropped axial subvolume (B) restricted to this aROI, generated by applying a 10 mm isotropic expansion to the GTVp bounding boxes, used for Phase 2p inference.

For nodal refinement, we computed 26-connected components of Stage 1 GTVn, and for each instance GTVn lesion we expanded the bounding box by 10 mm in x/y/z directions (**Figure 3**). We then repeated that process for all instance GTVn for each case. We used a ROI size of 48 × 48 × 48 voxels, resampled to 0.9766 × 0.9766 × 1.5 mm, used a learning rate to 2e-4, and trained for 137 epochs. We trained 7 folds using cross-validation, and, at inference, ensembled them with MultiLabel STAPLE per component; nodes were inserted without overriding tumor, and non-confirmed node voxels within the component's aROI were cleared before the result was fused back to the canonical grid.

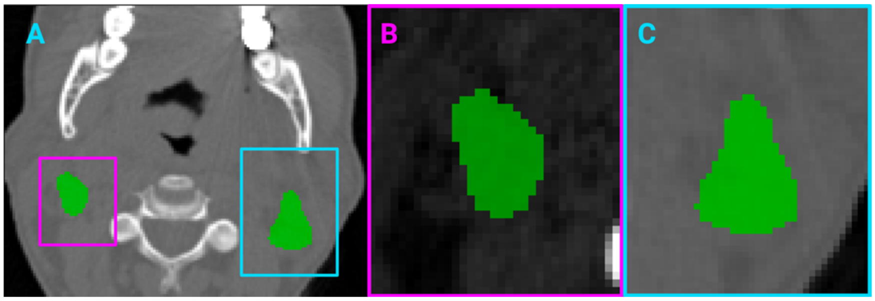

Fig. 3. GTVn-focused field-of-view refinement. Axial CT (A) from the Phase 1 mandible-anchored crop showing expert-annotated GTVn (green) with candidate nodal aROIs (pink and light-blue boxes). Corresponding instance-specific cropped axial subvolumes (B-C) centered on each node, generated by applying a 10 mm isotropic expansion to the instance GTVn bounding boxes, used for Phase 2n inference.

Finalization.

After Phase 2 fusions, the canonical GTVp and GTVn label map were resampled back to the original input CT geometry (matching size, spacing, origin, and direction), small components (< 150 voxels) of instance GTVp and instance GTVn lesions were removed, and a single CT-native label MetaImage Header File was saved per case to an output folder as seen in **Figure 4**.

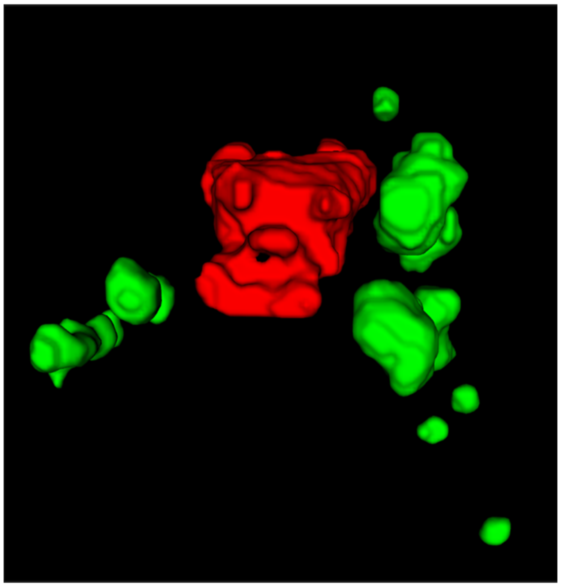

Fig. 4. 3D volume of final segmentation for GTVp (Red) and GTVn (Green).

Table 2. Summary of each phase's training ROI size, learning rate, observed VRAM used for training, and approximate training time per fold.

Stage	ROI (resampled voxels)	Initial Learning Rate	VRAM Used	Epochs per Fold	Training Time per Fold
Phase 0 (Mandible, 1 fold)	[192, 192, 64]	0.0004	≈ 4 GB	19	5.8 hours
Phase 1 (GTVp + GTVn, 3 folds)	[128, 112, 112]	0.0004	≈ 12 GB	300	8.5 hours
Phase 2p (GTVp focus, 15 folds)	[48, 48, 48]	0.0002	≈ 4 GB	137	0.9 hours
Phase 2n (GTVn focus, 7 folds)	[48, 48, 48]	0.0002	≈ 4 GB	200	1.5 hours

3 Results

3.1 Data

The training set included 682 co-registered CT and FDG-PET scans from patients with head and neck cancer, with expert contours for the GTVp and any GTVn if present. A subset of 73 cases also included the chest, abdomen, and pelvis.

3.2 Cross-Validation

We report the cross-validation results on the training sets for phases 0, 1, 2p, and 2n. The average cross-validation results are presented in **Table 3**, demonstrating consistent segmentation accuracy across folds. These results provided an estimate of generalization performance prior to final test submission.

Table 3. Cross-validation DSC results at the best performing epoch during training for GTVp and GTVn during each of the phases 0, 1, 2p, and 2n.

Phase	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10	Fold 11	Fold 12	Fold 13	Fold 14	Fold 15	Average
0 (Mandible)	0.85															0.85
1 (GTVp)	0.72	0.72	0.75													0.73
1 (GTVn)	0.69	0.72	0.72													0.71
2p (GTVp)	0.84	0.85	0.85	0.83	0.85	0.84	0.83	0.85	0.83	0.83	0.83	0.85	0.84	0.84	0.82	0.83
2n (GTVn)	0.82	0.81	0.82	0.81	0.82	0.81	0.81									0.81

3.3 Sanity Check Set and Testing Set Performance

Our final submitted algorithm was evaluated by the HECKTOR25 organizers on the held-out 2025 Sanity Check set and Testing set, comprising approximately 3 and 450 patients, respectively (15). The DLaBella29 team's method achieved a Sanity Check Phase inference average DSC of 0.87 for GTVp, an aggregate DSC of 0.75 for GTVn, and an Aggregate F1 of 0.67 for GTVn. This Sanity Check Phase inference utilized all 4 phases (0, 1, 2p, and 2n) as described in the methods section.

Because of shared VRAM constraints on the Grand-Challenge platform, inference on the 450 testing cases was limited to the coarse Phase 1 predictions, without refinement from Phase 2p or Phase 2n. Under these conditions, the testing set mean DSC was 0.5289 for GTVp, with an aggregate mean DSC of 0.6156 and aggregate F1 of 0.5561 for GTVn. By definition of the challenge, the GTVn aggregate DSC was computed by combining all nodal lesions per case into a single mask, such that the score reflected summed overlap between predicted and true nodal volumes. The F1-score evaluated lesion-level detection performance by balancing sensitivity and precision of nodal detection. An F1 of 0.5561 indicates that just over half of the nodal metastases were correctly identified with relatively few false positives, a reasonable result given the difficulty of detecting small nodes.

Qualitative internal review of inference of the public training cases suggested robust performance for large, FDG-avid primary tumors, often delineating them nearly as well as human experts. Performance on very small lymph nodes was more variable; some sub-centimeter nodes with low PET uptake were missed by the model, whereas others were detected but with slightly underestimated volume. Finally, mandible-anchored cropping significantly reduced inference time and did not exclude any tumors in the training cases, as all GTVs fell within the chosen aROI margins.

4 Discussion

We developed and evaluated a four-phase, anatomy-guided pipeline for automatic segmentation of oropharyngeal primary tumors and nodal disease on paired FDG-PET/CT. The key design choice was to anchor the coarse segmentation to the mandible (Phase 0) and to constrain learning and inference to anatomically plausible aROI through fixed, empirically chosen expansion margins around the mandible-anchor-point. This anatomically grounded cropping substantially reduced memory footprint and computation time, while maintaining coverage of all training tumors. Building on this, we used a coarse-to-fine progression: a joint GTVp and GTVn model for coarse localization (Phase 1), followed by focused refinements using tight, task-specific aROIs including GTVp centered crops (Phase 2p) and lesion-wise nodal crops derived from 26-connected components (Phase 2n). The phased design sought to improve performance over a single-phase model.

4.1 Cross-validation Performance

The mandible model achieved a mean cross-validation DSC of 0.85 after only 19 training epochs, enabling robust and fast aROI definition for subsequent phases. Coarse tumor segmentation (Phase 1) produced mean cross-validation DSCs of 0.73 for GTVp and 0.71 for GTVn. Focused refinements improved both tasks: GTVp (Phase 2p) increased to 0.84 across 15 folds, and GTVn (Phase 2n) increased to 0.81 across 7 folds (**Table 2**). These improvements are consistent with the intended role of Phase 2 as a high-precision stage that fine-tunes instance detections by inferring on smaller cropped and anatomically informed aROIs, after Phase 1 has addressed full-image localization.

4.2 Held-out Evaluations

On the HECKTOR25 Sanity Check set, our full pipeline (all four phases) achieved an average DSC of 0.87 for GTVp and an aggregate DSC of 0.75 for GTVn, with an aggregate F1 of 0.67 for nodal presence/absence. In contrast, Testing set results were lower: 0.5289 DSC for GTVp, 0.6156 aggregate DSC for GTVn, and 0.5561 aggregate F1 for nodal detection. The principal driver of this generalization gap was a practical constraint: due to shared-GPU VRAM limits on the Grand-Challenge inference platform when processing the 450 case Testing set, we were required to run only the coarse model

(Phase 1) during testing. As a result, none of the fine-tuning steps from Phase 2p and Phase 2n were applied at testing time, which is hypothetically consistent with the observed drop in performance. However, due to the smaller sample size of the Sanity Check set, a legitimate statistical comparison of the single-phase vs four-phase algorithms cannot be made. Further testing should be considered to compare the performance of the single-phase algorithm compared to the four-phase algorithm.

4.3 Interpretation and Implications.

The results support three practical conclusions. First, anatomy-anchored cropping is an effective and reproducible way to reduce search space and computation without sacrificing coverage; our fixed margins around a mandible-derived anchor included every training GTV and significantly reduced inference time. Second, a coarse-to-fine pipeline with lesion-wise refinement is important for head-and-neck disease, where the class imbalance and anatomical clutter make single-pass models particularly susceptible to false positives and false negatives in nodal basins and to boundary under-segmentation along mucosal surfaces. Third, inference configuration matters: disabling the refinement phases for operational reasons (here, platform VRAM) predictably degrades external performance, highlighting that system-level constraints can be as consequential as model choice when translating to large, real-world cohorts.

4.4 Strengths and Limitations.

Strengths include (i) a transparent, modular design that cleanly separates localization and refinement; (ii) explicit lesion-wise handling of nodal disease using connected-component crops; (iii) efficiency compatible with a single RTX 4090 (with modest per-fold training times); and (iv) no dependence on external training datasets. Limitations include (i) reliance on accurate mandible segmentation for aROI definition; (ii) potential miss or truncation of very distant nodal disease if encountered in distributions different from the training set; and (iii) over-filtering risk from post-processing for tiny reference standard lesions. In addition, the Testing set constraint to Phase 1 only means the reported external numbers reflect our coarse model's ceiling rather than the whole four-phase algorithm's true capacity.

4.5 Future Directions.

Several extensions are straightforward. The refinement phases could be made inference-robust under strict VRAM by reducing network size, inference optimization, and reducing the number of folds utilized so that Phases 2p and 2n can run even on shared VRAM hardware. Addition of test-time augmentation may stabilize boundaries and small-lesion recall. Exploring multi-anchor aROIs (e.g., mandible plus hyoid or soft-palate landmarks) could reduce the risk of missing atypically located disease. Finally, adding uncertainty quantification would alert clinicians when outputs are unreliable (e.g., out-of-distribution scans).

4.6 Clinical Perspective.

In its intended multiphase configuration, our pipeline provides initial contours rapidly and with high DSC performance for typical, FDG-avid primaries and many nodal presentations, making it well-suited for draft-contouring workflows that incorporate expert verification. The coarse-only Testing set findings underscore that deployment details (hardware, memory budget, and batching strategy) are central to achieving clinically acceptable performance at scale. While challenges remain in perfecting automated head and neck tumor segmentation, our results show the potential to speed up and make more consistent radiotherapy contours, which could help enable adaptive treatment strategies in OPSCC.

5 Conclusion

In summary, we have demonstrated a multi-phase, mandible-anchored Auto3DSeg approach for automated GTV segmentation in oropharyngeal cancer PET/CT scans. By integrating anatomical knowledge with deep learning SegResNet networks, our algorithm effectively segmented primary and nodal tumors and achieved fair performance in the HECKTOR 2025 challenge. The approach emphasizes preserving geometric fidelity and leveraging multi-modal imaging. While challenges remain in perfecting automated head and neck tumor segmentation, our results show the potential to speed up and make more consistent radiotherapy contours, which could help enable adaptive treatment strategies in OPSCC. Continued improvements and validation on larger cohorts will be necessary, but automated segmentation is poised to become a valuable assistant in the radiation oncologist's toolkit, particularly as we move toward more personalized and adaptive therapy paradigms.

Data Availability. Our code for all pre-processing, training, inference, and post-processing are available as open-source at **GitHub: dlabella29/HECKTOR25_Auto3DSeg**.

Acknowledgments. The authors thank the organizers of the HECKTOR 2025 challenge for providing the dataset and a collaborative platform. We also acknowledge MONAI and the Auto3DSeg developers for open-source tools that enabled this work.

Disclosures of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Siegel RL, Giaquinto AN, Jemal A. Cancer statistics, 2024. CA Cancer J Clin. 2024 Jan 17;74(1):12–49.
2. Chajon E, Lafond C, Louvel G, Castelli J, Williaume D, Henry O, et al. Salivary gland-sparing other than parotid-sparing in definitive head-and-neck intensity-modulated radiotherapy does not seem to jeopardize local control. Radiat Oncol. 2013 May 30;8:132.
3. Petkar I, Rooney K, Roe JWG, Patterson JM, Bernstein D, Tyler JM, et al. DARS: a phase III randomised multicentre study of dysphagia- optimised intensity- modulated radiotherapy (Do-IMRT) versus standard intensity- modulated radiotherapy (S-IMRT) in head and neck cancer. BMC Cancer. 2016 Oct 6;16(1):770.
4. Nutting CM, Morden JP, Harrington KJ, Urbano TG, Bhide SA, Clark C, et al. Parotid-sparing intensity modulated versus conventional radiotherapy in head and neck cancer (PARSPORT): a phase 3 multicentre randomised controlled trial. Lancet Oncol. 2011 Feb;12(2):127–36.
5. Castelli J, Depeursinge A, Ndoh V, Prior JO, Ozsahin M, Devillers A, et al. A PET-based nomogram for oropharyngeal cancers. Eur J Cancer. 2017 Apr;75:222–30.
6. Mowery YM, Vergalasova I, Rushing CN, Choudhury KR, Niedzwiecki D, Wu Q, et al. Early 18F-FDG-PET Response During Radiation Therapy for HPV-Related Oropharyngeal Cancer May Predict Disease Recurrence. International Journal of Radiation Oncology*Biology*Physics. 2020 Nov;108(4):969–76.

7. Bogowicz M, Leijenaar RTH, Tanadini-Lang S, Riesterer O, Pruschy M, Studer G, et al. Post-radiochemotherapy PET radiomics in head and neck cancer – The influence of radiomics implementation on the reproducibility of local control tumor models. Radiotherapy and Oncology. 2017 Dec;125(3):385–91.
8. Vallières M, Kay-Rivest E, Perrin LJ, Liem X, Furstoss C, Aerts HJWL, et al. Radiomics strategies for risk assessment of tumour failure in head-and-neck cancer. Sci Rep. 2017 Aug 31;7(1):10117.
9. Chen AM. De-escalated radiation for human papillomavirus virus-related oropharyngeal cancer: evolving paradigms and future strategies. Front Oncol. 2023 Jul 27;13.
10. Regan SN, Rosen BS, Suresh K, Cao Y, Aryal MP, Allen SG, et al. FDG-PET-based Selective De-escalation of Radiotherapy for HPV-Related Oropharynx Cancer: Results from a Phase II Trial. International Journal of Radiation Oncology*Biology*Physics. 2024 Apr;118(5):e11.
11. Courtney PT, L. Santoso M, Savjani RR, K. Reddy V, Chai-Ho W, Velez Velez MA, et al. A phase II study of personalized ultrafractionated stereotactic adaptive radiotherapy for palliative head and neck cancer treatment (PULS-Pal): a single-arm clinical trial protocol. BMC Cancer. 2024 Dec 21;24(1):1564.
12. Atienza C, Shepard A, Uzomah U, Rajan SK, Anderson CM, Katzer J, et al. Preliminary experience using MR-guided adaptive radiotherapy in head and neck cancer. Front Oncol. 2024 Nov 8;14.
13. Avkshtol V, Meng B, Shen C, Choi BS, Okoroafor C, Moon D, et al. Early Experience of Online Adaptive Radiation Therapy for Definitive Radiation of Patients With Head and Neck Cancer. Adv Radiat Oncol. 2023;8(5):101256.
14. Heukelom J, Fuller CD. Head and Neck Cancer Adaptive Radiation Therapy (ART): Conceptual Considerations for the Informed Clinician. Semin Radiat Oncol. 2019 Jul;29(3):258–73.
15. HECKTOR25 [Internet]. [cited 2025 Sep 8]. Available from: hecktor25.grand-challenge.org
16. Saeed N, Hassan S, Hardan S, Aly A, Taratynova D, Nawaz U, et al. A Multimodal and Multicentric Head and Neck Cancer Dataset for Segmentation, Diagnosis and Outcome Prediction. 2025 Sep 20;
17. Andrearczyk V, Oreiller V, Boughdad S, Rest CC Le, Elhalawani H, Jreige M, et al. Overview of the HECKTOR Challenge at MICCAI 2021: Automatic Head and Neck Tumor Segmentation and Outcome Prediction in PET/CT Images. 2022 Feb 17;
18. The MONAI Consortium: Project MONAI. Zenodo. 2020. https://doi.org/10.5281/zenodo.4323059.
19. Myronenko A. 3D MRI brain tumor segmentation using autoencoder regularization. 2018 Oct 27; Available from: http://arxiv.org/abs/1810.11654
20. Wasserthal J, Breit HC, Meyer MT, Pradella M, Hinck D, Sauter AW, et al. TotalSegmentator: Robust Segmentation of 104 Anatomic Structures in CT Images. Radiol Artif Intell. 2023 Sep 1;5(5).
21. Warfield SK, Zou KH, Wells WM. Simultaneous Truth and Performance Level Estimation (STAPLE): An Algorithm for the Validation of Image Segmentation. IEEE Trans Med Imaging. 2004 Jul;23(7):903–21.