

AsymmetryNet: A Clinically Inspired Asymmetry Attention Model for Predicting HPV Status in Oropharyngeal Squamous Cell Carcinoma on Computed Tomography

John D. Mayfield, M.D. Ph.D. M.Sc.

radres2019@gmail.com

Massachusetts General Hospital, Harvard Medical School, Athinoula A. Martinos Center for Biomedical Imaging <https://orcid.org/0000-0001-5063-6160>

Elly Arizono, M.D.

Massachusetts General Hospital, Harvard Medical School

Karen Buch, M.D.

Massachusetts General Hospital, Harvard Medical School <https://orcid.org/0000-0001-9579-8921>

Osamu Sakai, M.D. Ph.D.

Massachusetts General Hospital, Harvard Medical School, Massachusetts Eye and Ear
<https://orcid.org/0000-0003-3462-5116>

Research Article

Keywords:

Posted Date: February 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-8979142/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

AsymmetryNet: A Clinically Inspired Asymmetry Attention Model for Predicting HPV Status in Oropharyngeal Squamous Cell Carcinoma on Computed Tomography

John D. Mayfield, M.D., Ph.D., M.Sc.^{1,2,3}, Elly Arizono, M.D.^{1,2}, Karen Buch, M.D.^{1,2}, and Osamu Sakai, M.D., Ph.D.^{1,2,4}

¹*Massachusetts General Hospital, Boston, MA, USA*

²*Harvard Medical School, Boston, MA, USA*

³*Athinoula A. Martinos Center for Biomedical Imaging, Boston, MA, USA*

⁴*Massachusetts Eye and Ear, Boston, MA, USA*

February 25, 2026

Abstract

Accurate, non-invasive HPV status prediction in oropharyngeal squamous cell carcinoma (OPSCC) is essential for risk stratification and treatment planning, yet current imaging methods lack sufficient sensitivity. We developed a novel framework combining 3D Masked Autoencoder (MAE) self-supervised learning (SSL) pretraining on a Swin Transformer encoder with 33 slice-aggregated asymmetry features (airway effacement, mass effect, midline shift, hypodensity metrics). In a multi-institutional cohort ($n = 173$) using five-fold cross-validation and holdout testing, the combined model achieved CV AUC 0.909 (95% CI 0.850–0.957) and holdout AUC 0.828 (95% CI 0.643–0.964), outperforming ablations (SSL-only: 0.703/0.636; asymmetry-only: 0.545/0.588). At a high-sensitivity threshold, it reached 100% sensitivity and 100% NPV. This establishes a new benchmark for imaging-based HPV prediction in OPSCC without segmentation, enabling reliable non-invasive risk assessment.

1 Introduction

Oropharyngeal squamous cell carcinoma (OPSCC) is one of the fastest rising cancers in high-income countries, with marked increases reported in the United Kingdom, United States, Europe, New Zealand, and parts of Asia.[1–3] Globally, about 33–45% of OPSCC cases are related to the human papillomavirus (HPV), with wide geographic variation: 1–11% in regions of India/Brazil versus 70–85% in Sweden, New Zealand, and Lebanon.[1, 4] A 2023 meta-analysis estimated 42% HPV prevalence in OPSCC overall [1], and a separate American Society of Clinical Oncology (ASCO) criteria meta-analysis found 44.8% HPV-driven OPSCC (6009 patients, 11 countries).[3] Disease burden is highest in North America, Northern Europe, and Oceania, where HPV-positive OPSCC has increased by more than 200% over recent decades and now predominates HPV-negative disease.[1, 5, 6] As HPV-associated OPSCC now accounts for roughly one-third to one-half of OPSCC worldwide and is rapidly increasing, diagnostic delay can lead to more advanced stages at presentation with worse survival in HPV-positive disease.[5]

Recent advances in cancer therapies, including immunotherapy specific to patients with HPV-positive OPSCC, continue to be developed.[7] Quantitative imaging biomarkers have been developed for the detection of various features of OPSCC including predictions of HPV status, extranodal extension of tumors, and other biological and molecular features as a way of providing supportive imaging-based risk stratification, primarily diagnosed using contrast-enhanced computed tomography (CT).[8]

With the growing utility of deep learning methods in medical imaging, there is an opportunity for earlier detection, facilitated specialist referral, improved prognosis, and treatment optimization.[6, 9–12] The current literature includes studies analyzing a variety of biomarkers ranging from histopathology to medical imaging, many of which use radiomic feature data.[13–18] More recently, AI tools centered on further developing an automatic detection of these imaging biomarkers has been pursued in recent years with the hope for providing earlier detection, facilitated specialist referral, improved prognosis, and treatment optimization.[2, 10, 19–22] Many of these models rely on manual segmentation of the lesion or surrogates such as the gross tumor volume (GTV) from prior radiotherapy planning.[23]

The study expands upon the current body of literature, utilizing a novel method of directing attention to the site of potential pathology as separate attention heads driven by clinically inspired asymmetry features that parallel radiologists’ search patterns, including aerodigestive asymmetries, airway effacement, fat plane disruption, and lymph node hypoattenuation, collectively fed into a self-supervised learning (SSL) architecture. These attention networks helped identify key slices which have a higher likelihood of pathology, thus improving model efficiency without segmentation.[24–26] To simulate practical, real-world application, we evaluated this SSL architecture on a holdout external test set across patients identified from outpatient referrals.

Our hypothesis: Asymmetry features can direct an SSL to identify pathology in unlabeled, non-segmented CT volumes, thus directing more clinically impactful HPV status prediction.

2 Results

2.1 Model Development and Pretraining

We pretrained a novel asymmetry-guided attention network using SSL on contrast-enhanced CT volumes from the The Cancer Imaging Archive (TCIA) OPC-Radiomics [27], incorporating auxiliary tasks for airway segmentation and lymph node laterality detection. The self-supervised architecture leveraged the inherent bilateral symmetry of the head and neck anatomy, training the encoder to reconstruct masked regions. The model was subsequently fine-tuned for HPV status prediction on a multi-institutional dataset comprising 173 patients (122 HPV+, 51 HPV-) from four academic medical centers, Massachusetts General Hospital (MGH), Brigham and Women’s Hospital (BWH), Dana Farber Cancer Institute (DFCI), and Massachusetts Eye and Ear (MEE). A select set of patients from affiliated outpatient centers was selected as a holdout test set to demonstrate utility in a community setting, consisting of 35 patients (25 HPV+, 10 HPV-). For training the SSL (Figure 1), cropped 3D neck CT volumes are patch-embedded ($4 \times 4 \times 4$ patches $\rightarrow$ 96 channels, $4 \times$ spatial down sampling) before entering the Swin Transformer [28] 3D encoder pretrained via 3D Masked Autoencoder (MAE) [29]. During self-supervised pretraining, 75% of patches are masked and reconstructed by a lightweight decoder, which was discarded after pretraining (Figure 1A). The encoder learns hierarchical representations, as shown by example z-scored activation maps at each stage (early stages capture low-level edges and textures). Later stages show higher-level anatomical and semantic patterns). In downstream HPV classification, global average pooling yields a 768-dimensional feature vector. This vector feeds directly into a Multilayer Perceptron (MLP) head for the SSL Encoder-only ablation or is concatenated with 33 hand-crafted asymmetry features before the MLP in the Combined model. All paths produce sigmoid probabilities for binary HPV status (Figure 1B).

2.2 Cross-Validation Performance

Model performance was evaluated using 5-fold stratified cross-validation on the training set (N=138; 97 HPV+, 41 HPV-), with an ensemble of all fold models used for final inference. The self-supervised pretrained model achieved an overall cross-validation AUC of 0.909, with individual fold AUCs ranging from 0.690 to 1.000 (mean $\pm$ SD: 0.916 ± 0.121). The observed fold variance reflected the challenge of limited HPV- cases (8-9 per validation fold), a known difficulty in OPSCC cohorts where HPV positivity predominates (Figure 2A).

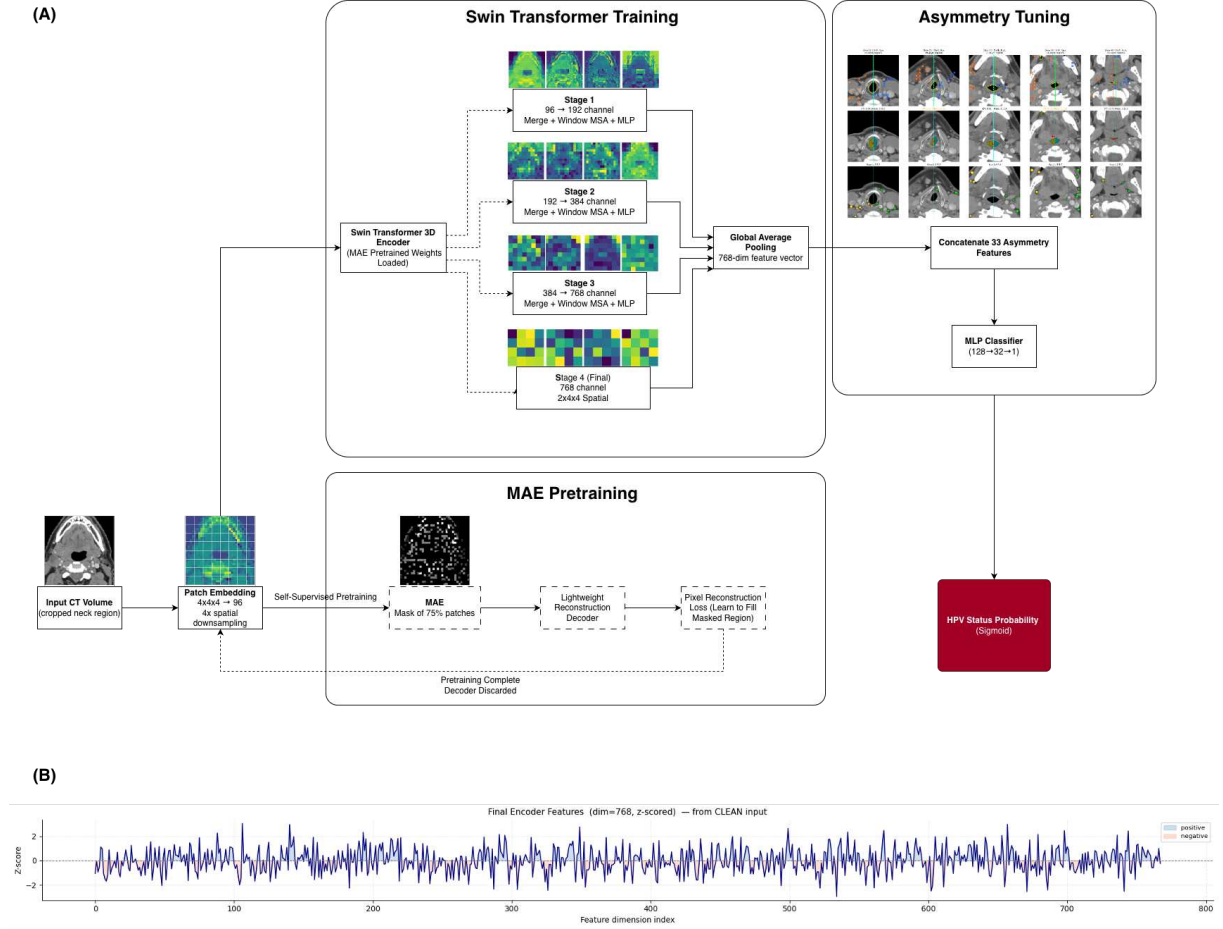


Figure 1: Model architecture and learned representations in the MAE-pretrained Swin Transformer 3D for HPV status prediction. (A) Self-Supervised AsymmetryNet Architecture. (B) Final Encoder Features (dim=768, z-scored).

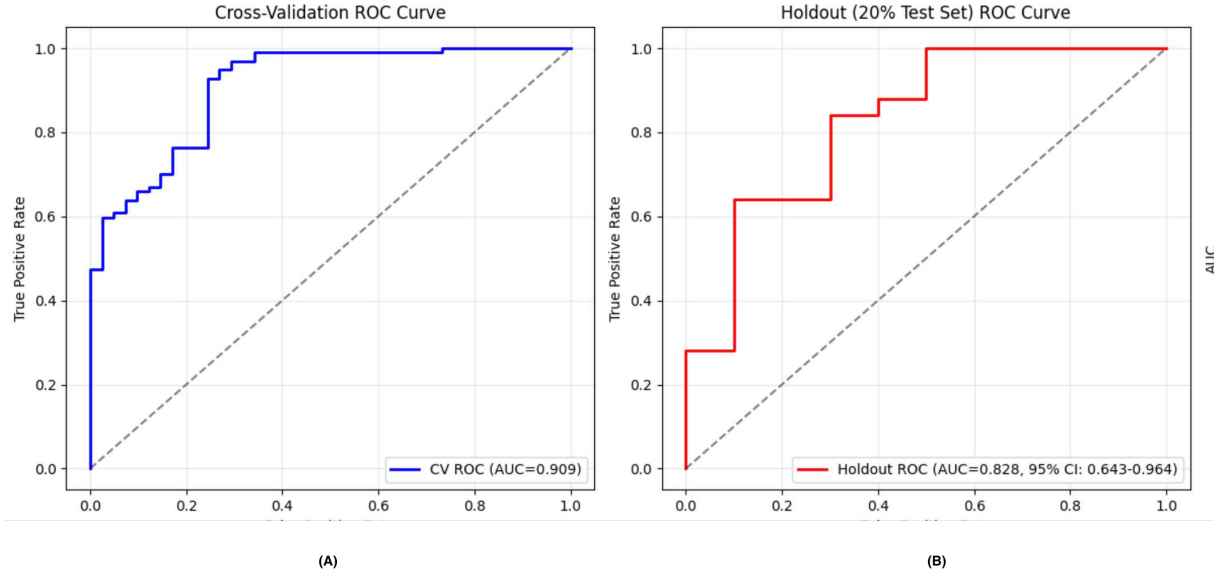


Figure 2: Cross-Validation Performance. (A) Receiver operating characteristic (ROC) curve demonstrating AUC of 0.909 with optimal operating point marked. (B) Probability distribution histogram showing separation between HPV+ (red) and HPV- (blue) predictions.

Probability distribution analysis revealed appropriate separation between classes, with mean predicted probability of 0.644 for HPV+ cases versus 0.592 for HPV- cases (Figure 2B). At the cross-validation-derived threshold of 0.612, optimized to achieve greater than 95% sensitivity, the model achieved 96.9% sensitivity (94/97 HPV+ detected), 70.7% specificity (29/41 HPV- correctly identified), 88.7% positive predictive value (PPV), and 90.6% negative predictive value (NPV), missing only 3 HPV+ cases. Calibration analysis revealed a Brier score of 0.194 and expected calibration error (ECE) of 0.223, indicating moderate calibration performance that informed subsequent threshold selection for clinical deployment.

2.3 Holdout Validation and Threshold Calibration

The final model was evaluated on a held-out test set (N=35; 25 HPV+, 10 HPV-) comprising patients not used during any phase of model development. Holdout AUC was 0.828 (95% CI: 0.643-0.964 via 1,000 bootstrap resamples), demonstrating maintained discriminative ability on unseen data (Figure 3A).

Critically, we observed a calibration shift between training and test distributions. The CV-derived threshold of 0.612, which achieved 97% sensitivity during cross-validation, yielded only 60% sensitivity (15/25) on the holdout set, missing 10 HPV+ cases despite achieving 90% specificity. This phenomenon, visualized in Figure 3B, reflects the narrower probability distribution in holdout predictions (range: 0.536–0.667) compared to cross-validation (range: 0.497–0.747), likely attributable to the ensemble averaging effect and distributional differences between training and test cohorts.

2.4 Clinical Operating Point Selection

To identify clinically actionable thresholds, we performed systematic threshold optimization on the holdout set across the full range of sensitivity-specificity trade-offs (Table 1).

For screening applications where the clinical cost of missing an HPV+ case (potential undertreatment, delayed appropriate therapy) substantially exceeds the cost of additional confirmatory testing, a threshold of 0.561 achieved 100% sensitivity (25/25 HPV+ detected), 50% specificity (5/10 HPV- correctly identified), 83% PPV, and 100% NPV (Figure 4A). At this operating point, no HPV+ cases were missed, though 50% of HPV- patients would be flagged for confirmatory p16 immunohistochemistry or HPV PCR testing.

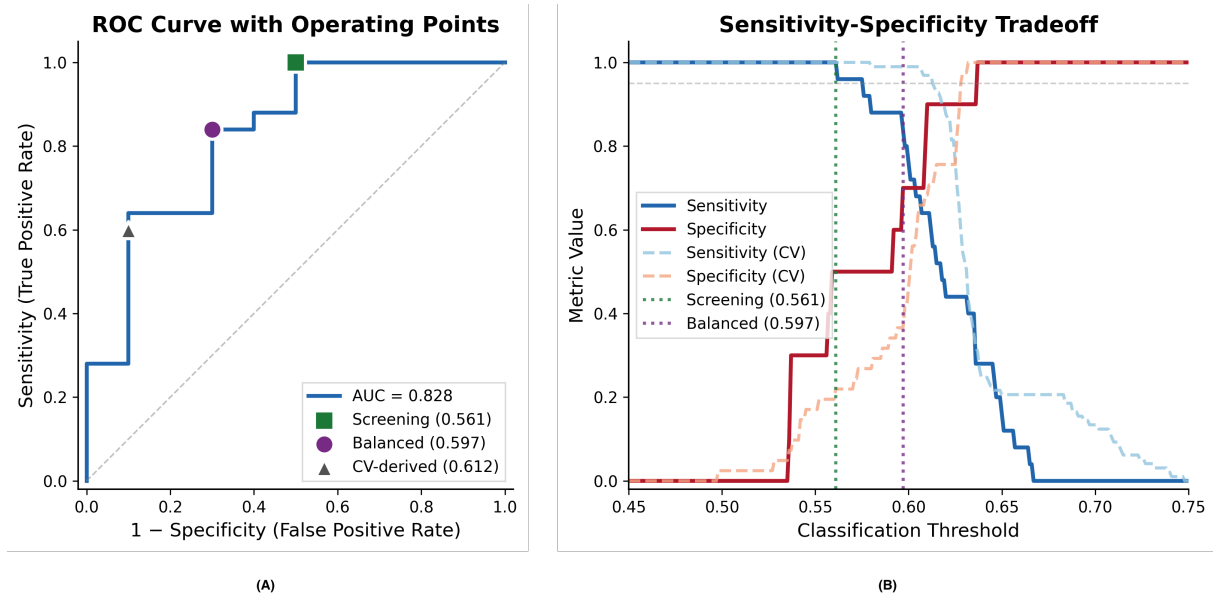


Figure 3: Calibration Shift Between Training and Holdout. (A) Holdout ROC curve (AUC = 0.828, 95% CI: 0.643–0.964) with three operating points marked. (B) Sensitivity (blue) and specificity (red) as a function of classification threshold for holdout (solid lines) versus cross-validation (dashed lines).

A high-sensitivity threshold of 0.575 maintained 96% sensitivity (24/25) with identical 50% specificity, missing only 1 HPV+ case while achieving 83% NPV. For applications requiring balanced performance, a threshold of 0.597 optimized via Youden’s J statistic (sensitivity + specificity - 1) yielded 84% sensitivity (21/25), 70% specificity (7/10), 88% PPV, and 64% NPV, missing 4 HPV+ cases (Figure 4B).

Table 1: Clinical Operating Points for HPV Screening (Holdout Set, $N = 35$)

Operating Point	Threshold	Sensitivity	Specificity	PPV	NPV
Screening	0.561	100% (25/25)	50% (5/10)	83%	100%
High Sensitivity	0.575	96% (24/25)	50% (5/10)	83%	83%
Balanced (Youden)	0.597	84% (21/25)	70% (7/10)	88%	64%
CV-derived	0.612	60% (15/25)	90% (9/10)	94%	47%

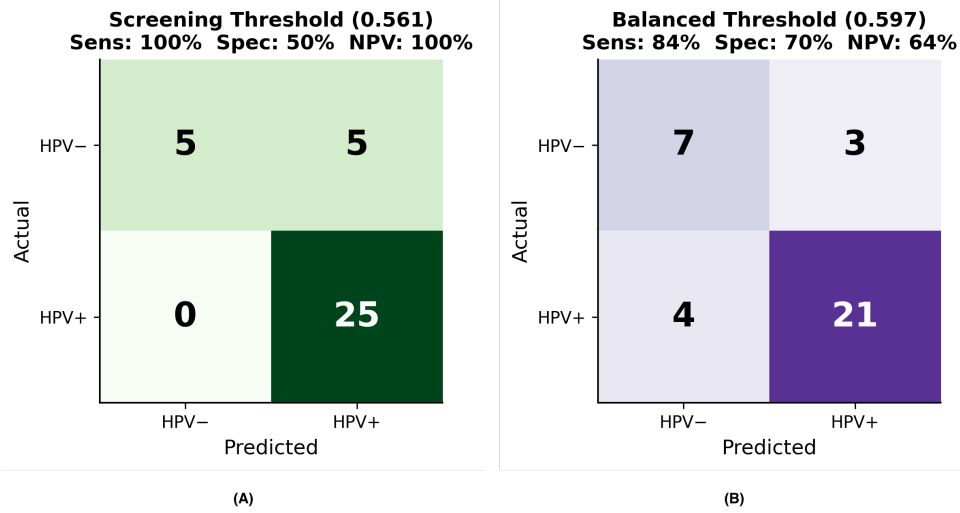


Figure 4: Confusion Matrices at Clinical Operating Points. Classification performance at selected thresholds on holdout set (N=35). (A) Screening threshold (0.561): optimized for maximum sensitivity, achieving 100% detection of HPV+ cases with 50% specificity. (B) Balanced threshold (0.597): optimized via Youden’s J statistic, achieving 84% sensitivity and 70% specificity. Cell values indicate patient counts; color intensity reflects magnitude.n

2.5 Comparison to Supervised Learning

To quantify the contribution of self-supervised pretraining, we trained a fully supervised model with identical architecture but random weight initialization (i.e., no pretraining). Both models incorporated the same asymmetry-derived features and underwent identical fine-tuning procedures.

The supervised approach achieved cross-validation AUC of 0.818 (95% CI: 0.737-0.898), substantially lower than the SSL-pretrained model (AUC 0.909). More critically, the supervised model failed to generalize: holdout AUC was 0.512 (95% CI: 0.298-0.726), statistically indistinguishable from random chance. There is an inherent generalization gap where SSL achieves 0.828 versus supervised achieving 0.512 on identical holdout data, demonstrating the critical role of self-supervised pretraining in learning transferable representations from limited labeled data, particularly in class-imbalanced scenarios where HPV+ cases predominate (70% prevalence in our cohort).

2.6 Ablation Study

We performed a targeted ablation study using the same patient cohort and preprocessing pipeline as the original model to evaluate the independent contributions of self-supervised learning (SSL) pretraining and hand-crafted airway/soft-tissue asymmetry features to HPV status prediction in oropharyngeal cancer. Five different studies were evaluated (Figure 5): (1) the full combined (SSL + asymmetry) model (SSL-pretrained Swin transformer 3D encoder + 33 asymmetry features, end-to-end trained), (2) SSL encoder-only (frozen SSL encoder + MLP classifier on imaging features alone), (3) asymmetry Only (MLP classifier using only the 33 rotation-invariant asymmetry statistics derived from 3D airway effacement, mass effect, midline shift, and hypo-density analysis), (4) supervised From scratch (randomly initialized Swin Transformer 3D + MLP on concatenated imaging and asymmetry features), and (5) logistic regression of asymmetry features (L2-regularized logistic regression on asymmetry features as a simple non-deep baseline).

Ablation analysis demonstrated that the combined model significantly outperformed all reduced architectures (Table 2). The SSL-pretrained encoder alone achieved AUC 0.703, significantly exceeding the supervised-from-scratch baseline (AUC 0.550, $p < 0.001$), highlighting the critical role of self-supervised pretraining. Asymmetry features alone showed poor discriminative ability (MLP: AUC 0.545; Logistic Regression: AUC 0.413), consistent with the absence of significant feature-label correlations.

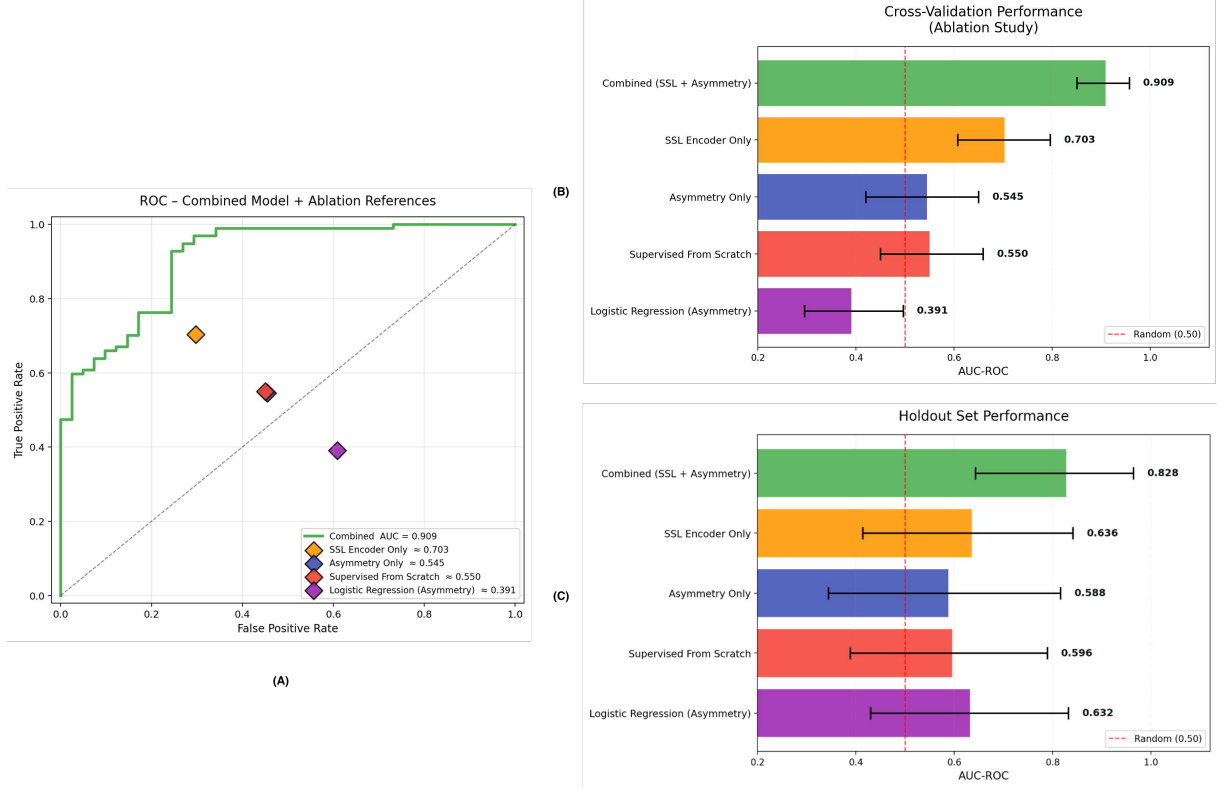


Figure 5: Ablation study results demonstrating the contributions of self-supervised learning (SSL) pre-training and asymmetry features to HPV status prediction in oropharyngeal cancer.

The combined architecture achieved the highest performance (AUC 0.909), significantly outperforming even the SSL encoder alone ($\Delta\text{AUC} +0.207$, $p < 0.001$), suggesting that asymmetry features provide complementary information when integrated with learned representations.

Table 2: Ablation Analysis Results

Model	CV AUC	ΔAUC vs Combined	p -value
Combined (SSL + Asymmetry)	0.909	—	—
SSL Encoder only	0.703	-0.207	< 0.001
Asymmetry only (MLP)	0.545	-0.364	< 0.001
Supervised from Scratch	0.550	-0.360	< 0.001
Logistic Regression (Asymmetry)	0.413	-0.496	< 0.001

3 Discussion

In this study, we introduce a novel asymmetry-driven attention network to identify key pathologies on the initial CT of the neck. This attention network provided objective metrics to a downstream SSL architecture that provided salient features for holdout testing. More importantly, these features helped the model discriminate HPV status at various levels of sensitivity. In comparison to state-of-the-art techniques, our study demonstrated a consistently high level of performance across a diverse test set.

For clinical deployment as a CT-based HPV screening tool, we recommend the threshold of 0.561 based on the following clinical reasoning. In OPSCC, HPV status fundamentally alters prognosis (3-year overall survival 80% for HPV+ vs. 50% for HPV-) and increasingly guides treatment intensity, with

multiple trials evaluating de-escalation protocols for HPV+ disease. Missing an HPV+ case could result in inappropriately aggressive treatment or exclusion from de-escalation trials, while a false-positive result incurs only the modest cost of confirmatory immunohistochemistry.

At the screening threshold, the model functions as a highly sensitive triage tool: patients scoring below 0.561 can be confidently classified as HPV- (NPV 100% in our cohort), while those scoring above proceed to standard confirmatory testing. Applied prospectively, this approach would reduce confirmatory testing burden by approximately 14% (5/35 patients correctly ruled out as HPV-) while maintaining perfect sensitivity for HPV+ detection. External validation on larger, multi-institutional cohorts is warranted before clinical implementation.

The strengths of our study include a non-invasive CT-only screening tool that does not require manual segmentation or large-scale radiomics implementation. Testing on a diverse external dataset demonstrated real-world performance. The self-supervised learning method also surpassed convolutional neural network-based supervised learning.

Limitations of the study include a class imbalance in the holdout testing set (primarily HPV+), the use of a single modality, the lack of prospective validation, and potential regional bias (all centers were located in the greater Boston area). Contrast-enhanced CT was utilized in this study as it is the imaging modality most often used in these patients, except for ultrasound examination of palpable lymph nodes or soft tissue masses of the neck, which would be further evaluated using CT. The regional bias is somewhat attenuated considering that the cases seen at these centers are often referrals from other parts of the country, thus introducing potential selection bias (worse cases than the baseline population).

To mitigate these factors, we are currently planning a prospective observational trial in partnership with otolaryngology. This study would analyze initial CTs in patients suspected of a possible aerodigestive lesion with single-blinded metrics and no control group. Following histopathology and HPV status testing in patients with lesions, the prediction metrics will be compared to the ground-truth HPV status. Projected imaging findings will not impact patient management.

Table 3: Comparison to State-of-the-Art across Imaging Domains. **UZH combined with the HNC1 dataset and tested on HNC2 and HNC3 datasets, precluding true external testing. SSL = Self-Supervised Learning; External Val = External Validation, NA = non-applicable, NR = not reported*

Study	Train <i>N</i>	Holdout <i>N</i>	Train/Val AUC	Holdout AUC	Segmentation?	External Test?
Our Method	138	35	0.91	0.83	No	Yes
La Greca Saint-Estevan (2022) [30]	640	152	0.88	0.84	Yes	Yes*
Fanizzi (2024) [31]	499	92	0.76	0.74	Yes	Yes
Woo (2023) [32]	88	20	0.78	0.52	Yes	Yes
Fazelpour (2023) [10]	393	93	0.80	NA	Yes	No
Fujima (2020) [11]	96	30	0.83	NA	Yes	No
Qiu (2024) [12]	450	49	0.80	NA	Yes	No

4 Methods

4.1 Patient Population and Data Partitions

This study utilized retrospective patient imaging data from two sources: The Cancer Imaging Archive (TCIA) OPC-Radiomics dataset (N=606) from a single institution between 2005 and 2010, and a cohort of 187 patients (Figure 6) in accordance with the Declaration of Helsinki and the local institutional review board (IRB 2024P000504). The only exclusion criteria were for missing imaging data in the TCIA set (n=64) and severe motion artifact that obscured the aerodigestive tract in the fine-tuning and

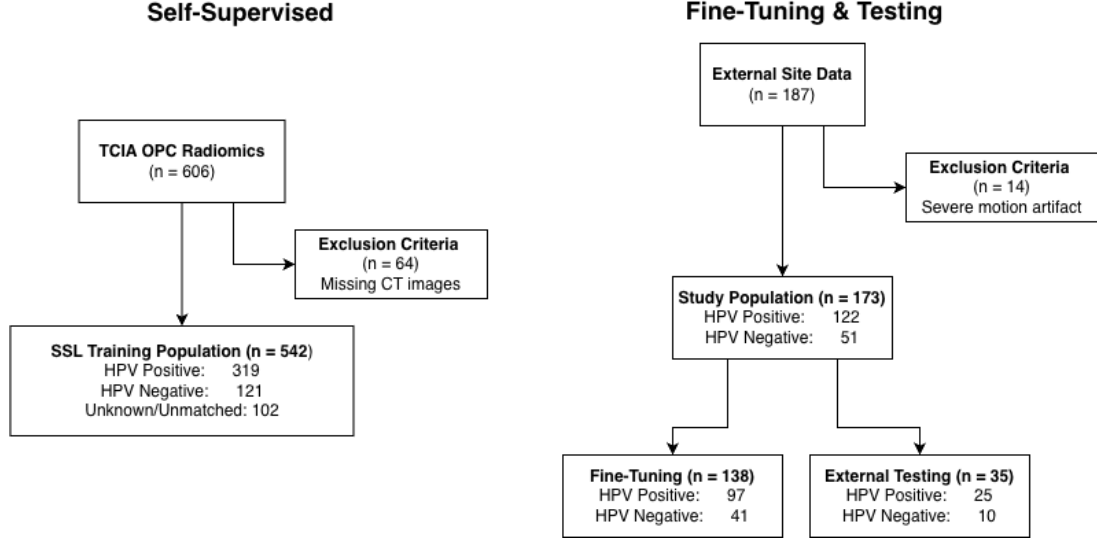


Figure 6: Flowchart of Patient Selection and Partitioning. OPC = oropharyngeal cancer, TCIA = The Cancer Imaging Archive, CT = computed tomography, HPV = human papillomavirus.

testing sets (n=14). Unlike other studies, metal streak artifact from dental amalgam/artifact was not a basis for exclusion as we wanted to evaluate in a real-world setting.

4.2 Image Preprocessing and Anonymization

Patient privacy was protected through multiple layers of anonymization. DICOM headers were stripped of all protected health information (PHI). Defacing was applied using a deep learning-based facial anonymization tool to remove identifiable facial structures (eyes, nose, ears, jawline) while preserving oropharyngeal anatomy. Volumes were then cropped to the oropharyngeal soft tissues, further excluding peripheral identifiable anatomy (neck, shoulders, skull base). All processing occurred on secure institutional servers with access restricted to authorized personnel. The study complied with HIPAA regulations and institutional data use agreements, with no re-identification risk.

Raw CT volumes were preprocessed to standardize input. Volumes were cropped to the oropharyngeal region of interest using an automated bounding box derived from airway centroid calculation to create a localized volume at original resolution to maintain visualization of surrounding structures. The resultant volume consisted of $60 \times 128 \times 128$ voxels, eliminating peripheral structures (neck, shoulders, skull base) that could introduce confounding signals or privacy risks. Cropping ensured consistent spatial context across cases and minimized irrelevant anatomical variation.

Volumes were then resampled to isotropic 1 mm^3 resolution, normalized to Hounsfield units (HU) range $[-1000, 1000]$, clipped, and scaled to $[0, 1]$ for model input. No contrast enhancement, intensity augmentation, or slice-level processing was applied during preprocessing to prevent artificial leakage or bias.

4.3 Asymmetry Feature Extraction

We developed a self-supervised learning framework employing a 3D Swin Transformer as the backbone within a Masked Autoencoder (MAE) architecture to learn robust representations from contrast-enhanced CT volumes of oropharyngeal squamous cell carcinoma (OPSCC). Input volumes are randomly masked with a fixed ratio of 75% of the 3D patches. The encoder consists of a patch embedding layer ($4 \times 4 \times 4$ non-overlapping patches, initial embedding dimension 96) followed by four stages of Swin Transformer blocks (depths $[2, 2, 6, 2]$, window size $4 \times 4 \times 4$, shifted-window attention, relative position bias, and stochastic depth) with progressive patch merging which produces a compact global representa-

tion. A lightweight decoder reconstructs the masked input patches, while two auxiliary regression heads simultaneously predict per-slice airway asymmetry metrics (effacement, mass effect, midline shift, hybrid score) and lymph-node-related asymmetry features (left/right hypodense counts and asymmetry) derived from heuristic detectors. These detectors identify airway asymmetry via midline optimization and bilateral air-volume comparison, and soft-tissue hypodensity asymmetry via connected-component analysis of hypoattenuating regions within soft-tissue Hounsfield Unit (HU) windows. The composite loss combines masked reconstruction MSE (weighted by slice-wise hybrid asymmetry score, boost factor 5.0), airway prediction MSE, and lymph-node prediction MSE. Strong data augmentations (intensity shift/scale, Gaussian noise, random flips, 90° axial rotations, spatial jitter with crop-and-resize, Gaussian blur, and cutout) are applied to generate positive pairs for monitoring representation collapse via cosine similarity of encoder features between augmented views of the same volume, computed periodically during training. Early stopping and checkpointing are guided by both reconstruction loss and cosine similarity trends to prevent feature collapse and overfitting.

4.4 Self-Supervised Learning (SSL) Pretraining

The MAE was pretrained on unlabeled head and neck CT volumes from the TCIA repository using a 3D Swin Transformer encoder. 75% of patches were masked, and the model was trained to reconstruct the original volume. Auxiliary heads predicted airway and lymph asymmetry maps (4×60 and 3×60 outputs for z-slices). Training used AdamW (lr=1e-4), MSE loss, and 100 epochs. This phase learned robust, asymmetry-aware representations without HPV labels, reducing reliance on labeled data imbalance. Cosine similarity was monitored throughout pretraining to identify any potential overfitting or model collapse. Final cosine similarity was 0.363 ± 0.164 .

4.5 Evaluation and Statistical Analysis

Performance was assessed using AUC, sensitivity, specificity, PPV, and NPV at predefined sensitivity targets (90%, 95%, 99%) chosen via validation sets within each fold. Final thresholds were the median across folds to avoid optimistic bias. ROC and PR curves were generated. The Mass General Brigham (MGB) external set was evaluated once using the final model ensemble (mean probabilities across folds).

4.6 Data Leakage and Bias Mitigation Strategies

To rigorously prevent data leakage, we implemented fold-specific processing. Asymmetry feature extraction and normalization (StandardScalar) were performed exclusively on training data. For patient-level separation, five-fold cross-validation ensured no patient appeared in both training and validation folds. The MGB external cohort was held out entirely during all training, tuning, and threshold selection. No metrics from validation/test data influenced feature computation or normalization. Additionally, there was no test set tuning. All thresholds and hyperparameters were fixed from internal cross-validation with no adjustment on the external test set.

Bias mitigation included (1) Class imbalance: Weighted random sampling over-sampled HPV-negative cases during training, (2) Site/institutional bias: External multi-institutional validation reduced reliance on single-site patterns, (3) Selection bias: Predefined exclusion criteria removed poor-quality cases before training but kept studies with metal streak artifact from dental amalgam/hardware, and (4) Algorithmic bias: SSL pretraining on diverse unlabeled TCIA volumes reduced dependence on labeled data biases.

4.7 Large Language Model Usage

Anthropic Claude Sonnet 4.5/4.6 was used for debugging of code and checking accuracy of pseudocode. No Large Language Models (LLMs) were used in the authorship of the manuscript.

Acknowledgments

The authors gratefully acknowledge the use of imaging data from The Cancer Imaging Archive (TCIA), which provided the medical imaging used for model pretraining. Specifically, data were accessed from the OPC-Radiomics collection [27]. All data use complies with TCIA’s Data Usage Policies and Restrictions. We thank the data contributors, curators, and the National Cancer Institute (NCI) for making these resources publicly available. The authors are also grateful for support from the MGH Ralph Schlaeger Foundation 2025 Fellowship Award which provided funding for the necessary machine learning hardware.

Data and Code Availability

All fine-tuning and holdout testing data is proprietary to Mass General Brigham. TCIA OPC-Radiomics data is available by permission, as above. The model was pretrained using a custom 3D Swin Transformer MAE implementation available at <https://huggingface.co/jdmayfield/AsymmetryNet> [33].

References

- [1] Carlander, A. F. et al. A Contemporary Systematic Review on Repartition of HPV-Positivity in Oropharyngeal Cancer Worldwide. *Viruses* **13**, 912 (2021).
- [2] Klein, S. et al. Deep Learning Predicts HPV Association in Oropharyngeal Squamous Cell Carcinomas and Identifies Patients with a Favorable Prognosis Using Regular H&E Stains. *Clin. Cancer Res.* **27**, 1131–1138 (2021).
- [3] Chun, W. B. et al. Sex and survival in oropharyngeal squamous cell carcinoma: A systematic review and meta-analysis. *Eur. Ann. Otorhinolaryngol. Head Neck Dis.* **143**, 30–39 (2026).
- [4] Skovvang, A. et al. Geographic variation in HPV prevalence in oropharyngeal cancer. (2021).
- [5] Lechner, M., Liu, J., Masterson, L. & Fenton, T. R. HPV-associated oropharyngeal cancer: epidemiology, molecular biology and clinical management. *Nat. Rev. Clin. Oncol.* **19**, 306–327 (2022).
- [6] Fujita, A. et al. Difference Between HPV-Positive and HPV-Negative Non-Oropharyngeal Head and Neck Cancer: Texture Analysis Features on CT. *J. Comput. Assist. Tomogr.* **40**, 43–47 (2016).
- [7] Khoo, A. et al. Human papilloma virus positive oropharyngeal squamous cell carcinoma and the immune system: pathogenesis, immunotherapy and future perspectives. *Int. J. Mol. Sci.* **25**, 1245 (2024).
- [8] Farrell, D. et al. Quantitative Imaging Advances in HPV-Positive Oropharyngeal Carcinoma. *Cancers (Basel)* **18**, 567 (2026).
- [9] Thakur, G. K. et al. Deep learning approaches for medical image analysis and diagnosis. *Cureus* **16**, e59507 (2024).
- [10] Fazelpour, S. et al. Multiparametric machine learning algorithm for human papillomavirus status and survival prediction in oropharyngeal cancer patients. *Head Neck* **45**, 2882–2892 (2023).
- [11] Fujima, N. et al. Prediction of the human papillomavirus status in patients with oropharyngeal squamous cell carcinoma by FDG-PET imaging dataset using deep learning analysis: A hypothesis-generating study. *Eur. J. Radiol.* **126**, 108936 (2020).
- [12] Qiu, E. et al. Fully automated 3D machine learning model for HPV status characterization in oropharyngeal squamous cell carcinomas based on CT images. *Am. J. Otolaryngol.* **45**, 104357 (2024).
- [13] Araújo, A. L. D. et al. Radiomic-Based Machine Learning Classifiers for HPV Status Prediction in Oropharyngeal Cancer: A Systematic Review and Meta-Analysis. *Diagnostics (Basel)* **16**, 345 (2025).
- [14] Suh, C. H. et al. Oropharyngeal squamous cell carcinoma: radiomic machine-learning classifiers from multiparametric MR images for determination of HPV infection status. *Sci. Rep.* **10**, 17525 (2020).
- [15] Kim, S. I., Kang, J. W., Eun, Y.-G. & Lee, Y. C. Prediction of survival in oropharyngeal squamous cell carcinoma using machine learning algorithms: A study based on the surveillance, epidemiology, and end results database. *Front. Oncol.* **12**, 974678 (2022).
- [16] Varricchio, S. et al. A machine learning approach to predict HPV positivity of oropharyngeal squamous cell carcinoma. *Pathologica* **116**, 379–389 (2024).

- [17] Ren, J., Yuan, Y., Qi, M. & Tao, X. Machine learning-based CT texture analysis to predict HPV status in oropharyngeal squamous cell carcinoma: comparison of 2D and 3D segmentation. *Eur. Radiol.* **30**, 6858–6866 (2020).
- [18] Buch, K. et al. Using texture analysis to determine human papillomavirus status of oropharyngeal squamous cell carcinomas on CT. *AJNR Am. J. Neuroradiol.* **36**, 1343–1348 (2015).
- [19] Diaz Badilla, E. et al. Predictive analysis of clinical features for HPV status in oropharynx squamous cell carcinoma: A machine learning approach with explainability. *Comput. Methods Programs Biomed. Update* **7**, 100170 (2025).
- [20] Jarwal, A. et al. A deep learning method for classification of HNSCC and HPV patients using single-cell transcriptomics. *Front. Mol. Biosci.* **11**, 1395721 (2024).
- [21] Hue, J., Valinciute, Z., Thavaraj, S. & Veschini, L. Multifactorial estimation of clinical outcome in HPV-associated oropharyngeal squamous cell carcinoma via automated image analysis of routine diagnostic H&E slides and neural network modelling. *Oral Oncol.* **141**, 106399 (2023).
- [22] Migliorelli, A. et al. Role of artificial intelligence in human papillomavirus status prediction for oropharyngeal cancer: A scoping review. *Cancers (Basel)* **16**, 789 (2024).
- [23] Kuijter, K. M. et al. Comparable performance between automatic and manual laryngeal and hypopharyngeal gross tumor volume delineations validated with pathology. *Int. J. Radiat. Oncol. Biol. Phys.* **122**, 186–193 (2025).
- [24] Curtin, H. D. Anatomy, imaging, and pathology of the larynx. In: *Head and neck imaging 1905–2039* (Elsevier, 2011).
- [25] Forghani, R., Smoker, W. R. K. & Curtin, H. D. Pathology of the oral region. In: *Head and neck imaging 1643–1748* (Elsevier, 2011).
- [26] Som, P. M. & Brandwein-Gensler, M. S. Lymph nodes of the neck. In: *Head and neck imaging 2287–2383* (Elsevier, 2011).
- [27] Grossberg, A. et al. HNSCC. *The Cancer Imaging Archive* <https://doi.org/10.7937/k9/tcia.2020.a8sh-7363> (2020).
- [28] Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S. & Guo, B. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 10012–10022 (2021).
- [29] He, K., Chen, X., Xie, S., Li, Y., Dollár, P. & Girshick, R. Masked Autoencoders Are Scalable Vision Learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15979–15988 (2022).
- [30] La Greca Saint-Estevan, A. et al. A 2.5D convolutional neural network for HPV prediction in advanced oropharyngeal cancer. *Comput. Biol. Med.* **142**, 105215 (2022).
- [31] Fanizzi, A. et al. Explainable prediction model for the human papillomavirus status in patients with oropharyngeal squamous cell carcinoma using CNN on CT images. *Sci. Rep.* **14**, 14276 (2024).
- [32] Woo, C. et al. Development and Testing of a Machine Learning Model Using 18F-Fluorodeoxyglucose PET/CT-Derived Metabolic Parameters to Classify Human Papillomavirus Status in Oropharyngeal Squamous Carcinoma. *Korean J. Radiol.* **24**, 51–61 (2023).
- [33] Mayfield, J. AsymmetryNet: 3D Swin Transformer MAE for OPSCC CT Pretraining. Hugging Face model repository <https://huggingface.co/jdmayfield/AsymmetryNet> (2026).

Supplementary Materials

S1 Ablation Analysis Design

Deep models shared the same Swin Transformer 3D architecture (patch size $4 \times 4 \times 4$, depths [2,2,6,2], heads [3,6,12,24], embed dim 96) and were trained with 5-fold stratified cross-validation, label smoothing ($\alpha=0.1$), AdamW (lr $5e-4$, weight decay 0.1), batch size 4, gradient clipping, ReduceLROnPlateau scheduling, early stopping, and mild Gaussian noise augmentation. Holdout predictions used ensemble averaging across CV folds. Performance was assessed primarily via AUC-ROC (with 95% bootstrap percentile CIs, $n=1,000$) on both CV and holdout sets, supplemented by accuracy, balanced accuracy, sensitivity, specificity, PPV, NPV, and F1 score. Statistical significance of AUC differences between the combined model and each ablation arm was evaluated using DeLong’s test for correlated ROCs and non-parametric bootstrap testing ($n=2,000$ resamples).

S2 SSL Fine-Tuning and Model Architecture

The pretrained Swin3D encoder was frozen and integrated into the HPVClassifier model, concatenating encoder features (dim=768) with asymmetry features (dim=33). The classifier consisted of: Linear(801 $\rightarrow$ 256) $\rightarrow$ ReLU $\rightarrow$ Dropout(0.5) $\rightarrow$ Linear(256 $\rightarrow$ 64) $\rightarrow$ ReLU $\rightarrow$ Dropout(0.3) $\rightarrow$ Linear(64 $\rightarrow$ 1).

Fine-tuning was performed using 5-fold stratified cross-validation, with weighted random sampling to address HPV+ imbalance (pos_weight = negative/positive ratio). Loss was BCEWithLogitsLoss with pos_weight. Early stopping was implemented after 30 epochs to avoid overfitting.

Algorithm 1 Asymmetry Feature Extraction per Axial Slice

Require: Axial slice z from preprocessed contrast-enhanced CT volume

Ensure: Slice-level asymmetry feature vector $A_z \in \mathbb{R}^{33}$

- 1: Detect airway centerline and bilateral contours $C_L(z)$, $C_R(z)$
 - 2: Compute airway effacement score:
3:
$$e_z = 1 - \frac{\min(|C_L(z)|, |C_R(z)|)}{\max(|C_L(z)|, |C_R(z)|)}$$
 - 4: Compute midline shift distance:
5:
$$m_z = |\text{centroid}_x(C_L(z)) - \text{centroid}_x(C_R(z))|$$
 - 6: Segment hypoattenuating regions in soft-tissue HU window $[-200, 200]$
 - 7: Compute lymph-node hypoattenuation asymmetry:
8:
$$l_z = |\text{count}_{\text{hypo, left}}(z) - \text{count}_{\text{hypo, right}}(z)|$$
 - 9: Extract additional asymmetries (mass effect, fat-plane disruption, etc.)
 - 10: Aggregate 33 rotation-invariant statistics into feature vector A_z
 - 11: **return** A_z
-

S3 Asymmetry Metrics and Saliency

Figure S2. displays five high-ranking asymmetric slices (from superior to inferior), arranged in three rows: Row 1 — original image with overlaid true midline (solid lime), algorithm-detected midline (dashed cyan/red when shifted >5 px), detected asymmetric soft-tissue regions (orange = left $>$ right density, blue = right $>$ left density), and airway contour (yellow); Row 2 — focused airway evaluation showing left (yellow fill/contour) and right (cyan fill/contour) airway segmentation, centroid deviation (red cross + arrow when relevant), midline, and quantitative effacement metrics (Eff = effacement ratio, Mass = mass effect score); Row 3 — hypodense region detection within soft-tissue windows (yellow =

Algorithm 2 Hybrid Asymmetry Score Computation and Slice Ranking

Require: Set of per-slice asymmetry feature vectors $\{A_z\}_{z=1}^D$ **Ensure:** Ranked list of key pathology slices $S = [z_{(1)}, \dots, z_{(k)}]$

- 1: Initialize empty list of hybrid scores $H \leftarrow []$
 - 2: **for** each axial slice $z = 1$ to D **do** ▷ Compute hybrid score for each slice
 - 3: $h_z \leftarrow w_e \cdot e_z + w_m \cdot m_z + w_l \cdot l_z + \sum w_i \cdot A_z[i]$
 - 4: Append h_z to H
 - 5: **end for**
 - 6: Sort slices in descending order of hybrid scores:
 - 7: $S \leftarrow \text{argsort}_{\downarrow}(H)$ ▷ Descending sort indices
 - 8: **return** top- k slices $S[1 : k]$
-

left-sided, lime/green = right-sided) with airway midline reference. Top asymmetric slices exhibit rightward midline shift (4–10 px), airway effacement predominantly on the left (Eff 0.07–0.24), right-sided airway displacement/hypodensity, multiple right-predominant hypodense soft-tissue foci, and elevated hybrid asymmetry scores. Orange boxed annotations and green dots highlight regions of greatest density deviation and hypodensity, respectively.

S4 Supplemental Figures

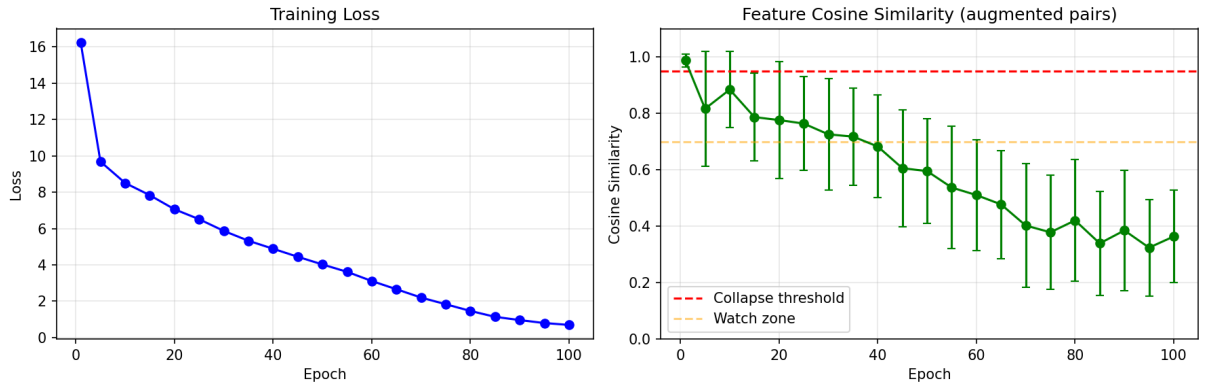
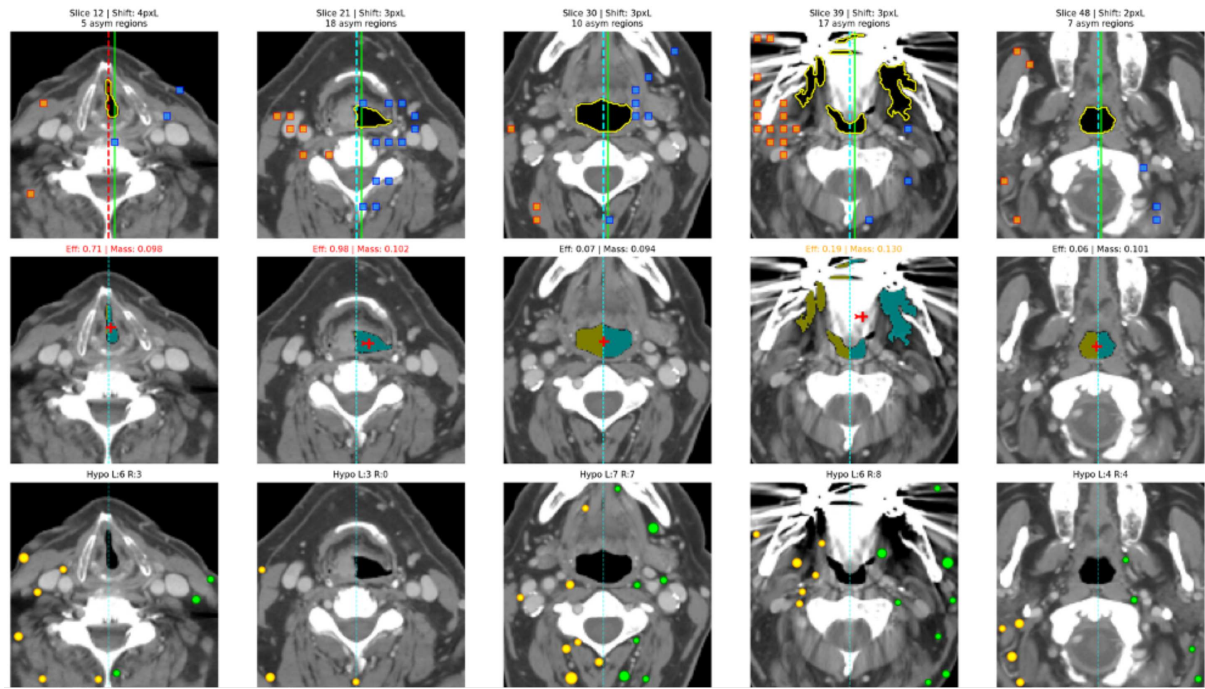
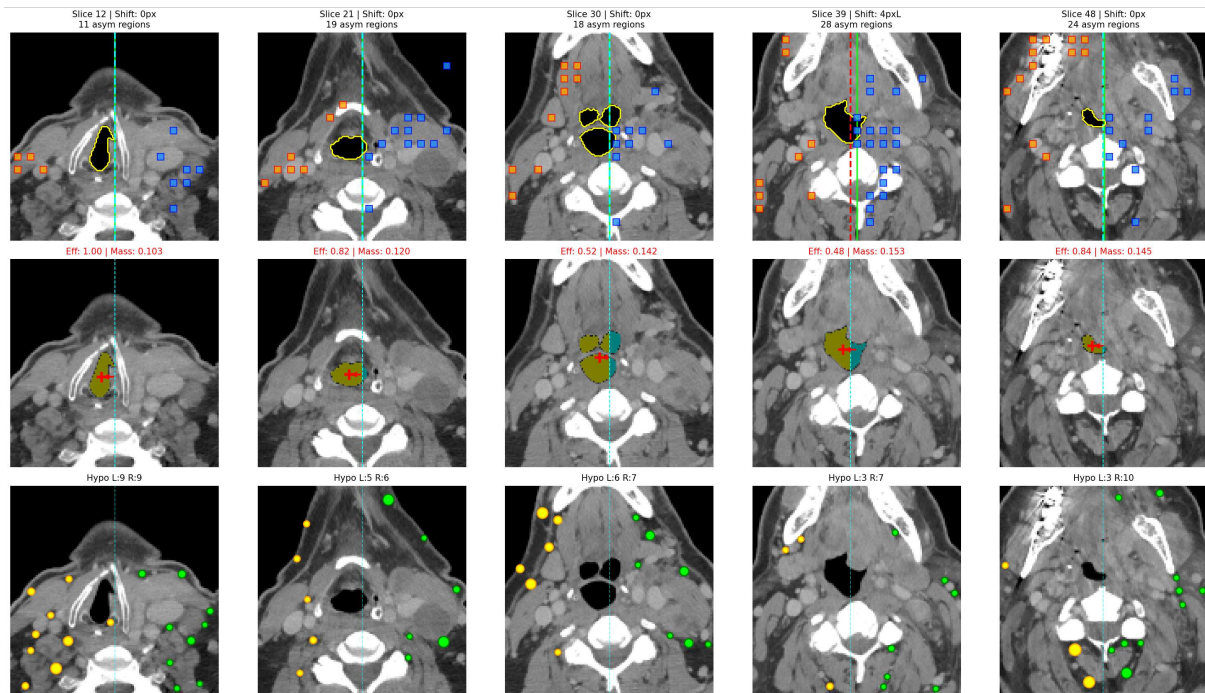


Figure S1: Training Loss and Feature Cosine Similarity for SSL Training.



(A)



(B)

Figure S2: Saliency Asymmetry Maps. (A) HPV-positive case. Top row: Soft tissue density symmetries are shown (orange=right; blue=left) with asymmetry regional counts listed per slice, saved as a feature. Midline asymmetry is measured across entire volume (global asymmetry). Middle row: Airway effacement and midline shift (red cross hatch) at each level. Features are saved as float point (e.g., effacement or Eff=0.98, mass effect or Mass=0.102). Bottom row: Fat planes and soft tissue hypodensities (HU -200 to 200) are quantified and measured for asymmetry (yellow = right; green = left). Cumulative asymmetries are saved as a feature. (B) HPV-positive case. Textbook example of an oropharyngeal lesion with internally cystic lymphadenopathy, collectively contributing to mass effect and partial effacement of the airway. Soft tissue hypodensity maps demonstrate fat plane effacement and nodal asymmetry.

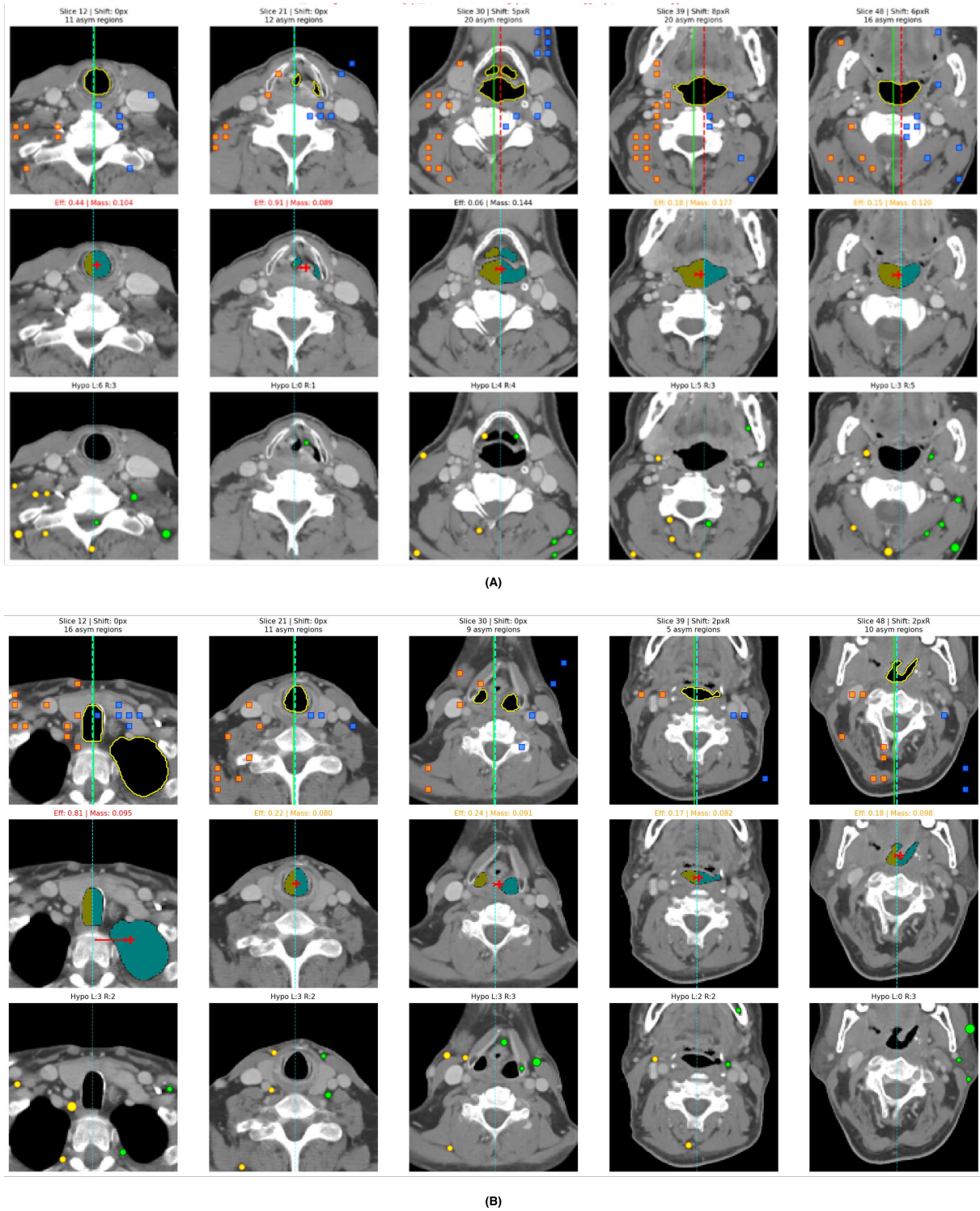


Figure S3: Asymmetry maps from misclassified cases. (A) HPV-negative case. Mild rotational asymmetry created artifactual mass effect and soft tissue asymmetry scores. Effacement of the airway at the glottis was identified, but summed features led to misclassification of the case as a false positive. (B) HPV-positive case. Mild rotational asymmetry may have partially obscured focal asymmetry in the right glossotonsillar sulcus. Airway effacement was still calculated as borderline elevated (yellow text); however, the collective metrics yielded a sub-threshold probability for HPV positivity. Of note, there is artifactual inclusion of the left lung apex as part of the airway creating erroneous mass effect metrics, despite exclusion of the top and bottom 0.15 of the image in asymmetry calculations.

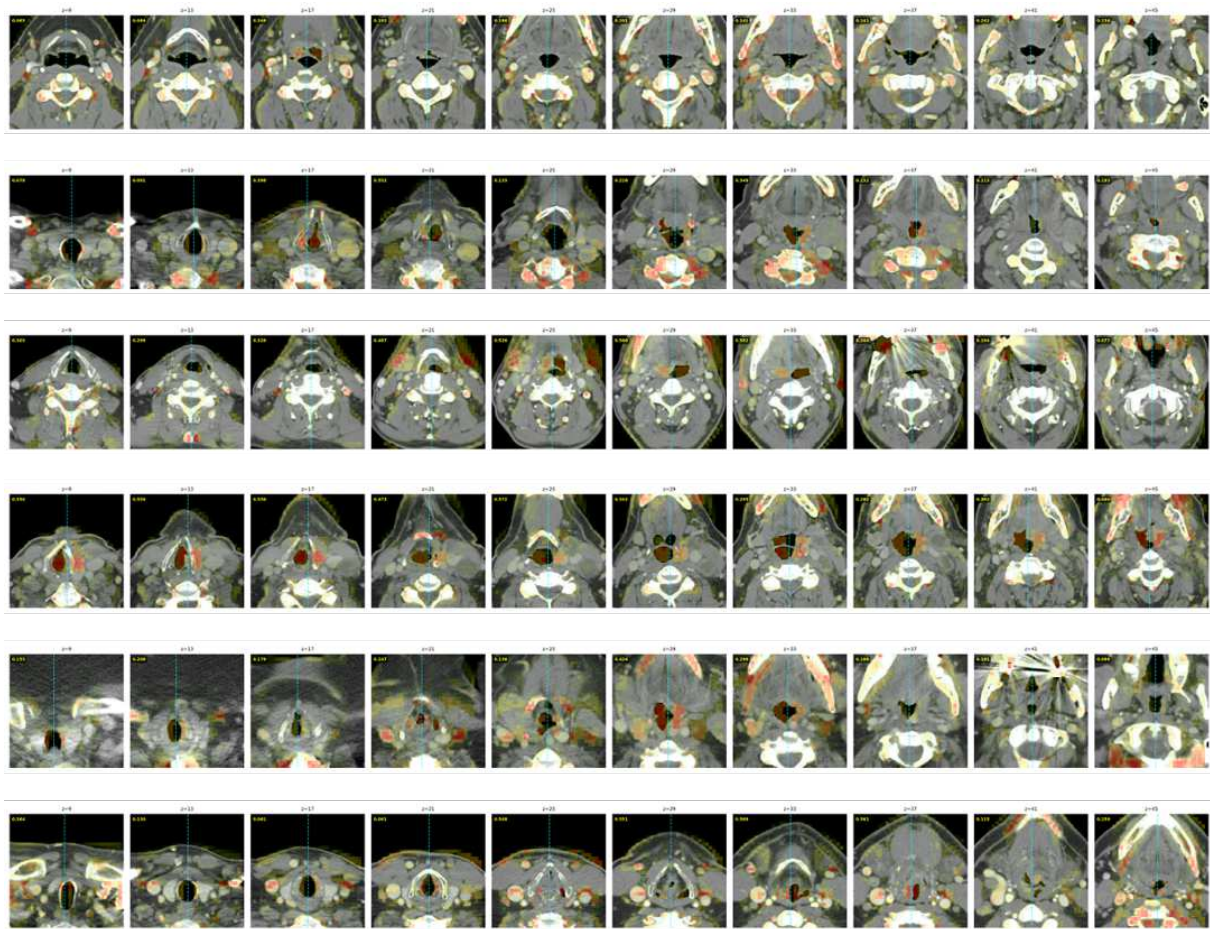


Figure S4: Symmetry Attention Maps. Multiple slice saliency mapping of asymmetry weights applied to the CT volumes at the classifier layer.