MELBA Journal
Machine Learning for Biomedical Imaging

SCOPE-HN: A Segmentation-based Collection of OroPharyngeal Structures Using Flexible Endoscopy for Head and Neck Cancers

Nikita **Bedi** [1], Anita **Rau** [2], Alberto **Paderno** [3], Hlu **Vang** [1], Yoon Kyoung **So** [1], F. Christopher **Holsinger** [1]

1 Division of Head & Neck Surgery, Department of Otolaryngology, Stanford University
2 Department of Biomedical Data Science, Stanford University
3 Department of Otolaryngology–Head and Neck Surgery, Humanitas Research Hospital

Abstract

Oropharyngeal cancer (OPC) is one of the few cancers with a rising incidence, driven primarily by increasing rates of human papillomavirus (HPV)-associated disease. Early detection and accurate delineation of OPC are critical for diagnosis, treatment planning and improving outcomes but remain challenging due to anatomical complexity and variability in clinical expertise. We present the **SCOPE-HN dataset**, a curated collection of annotated endoscopic images from 106 patients with histologically confirmed OPC, collected at Stanford between 2015 and 2023. The dataset consists of 942 RGB images extracted from diagnostic nasopharyngolaryngoscopy videos, with pixel-level annotations performed by expert Head and Neck surgeons. Annotations include 12 semantic classes representing tumor, normal anatomical structures, and common endoscopic artifacts. The SCOPE-HN dataset is publicly available at scope-hn.stanford.edu or bit.ly/SCOPE-HN under the Stanford University Dataset Research Use Agreement license.

Keywords

Oropharyngeal cancer, endoscopy, segmentation, dataset, annotation, machine learning, UADT

Article informations

https://doi.org/https://doi.org/10.59275/j.melba.2025-51gc

Volume 3, Received: 2025-06, Published 2025-12

Check for updates

Corresponding author: nbedi@stanford.edu

Special issue: MICCAI Open Data Special Issue 2025

Guest editors: Martijn Starmans, Apostolia Tsirikoglou, Lidia Garrucho Moras, Kaouther Mouheb

1. Introduction

Head and neck cancers (HNCs) comprise a heterogeneous group of malignancies of the oral cavity, oropharynx, larynx, and related structures. Globally, the incidence of oropharyngeal carcinoma (OPC) has risen substantially, with many countries reporting annual increases exceeding 3%. Faraji et al. (2019). In the United States, OPC is now the most common HPV-associated cancer in the United States Zumsteg et al. (2023). Early and accurate diagnosis of OPC is critical to improve patient outcomes, facilitate organ-preserving treatment, and support risk-adaptive strategies. However, several anatomical and clinical factors make early OPC detection challenging. The oropharynx is anatomically complex, and tumors often present as subtle mucosal irregularities or submucosal lesions which are difficult to identify through physical exam or radiologic imaging. Cross-sectional imaging such as CT, MRI, and PET, while essential for staging, often lacks the resolution to detect early mucosal disease or locate unknown primaries, particularly in HPV-positive cases. This places transnasal flexible endoscopy at the forefront of diagnostic investigation, offering direct, real-time visualization and guiding targeted biopsies Elrefaey et al. (2014). Despite its importance, endoscopic assessment remains subjective and varies considerably by operator experience, lighting conditions, and the presence of visual artifacts Watanabe et al. (2009). These limitations have accelerated interest in computer-aided diagnostic (CAD) systems using deep learning and semantic segmentation to standardize and enhance lesion detection during endoscopy. To address this gap, we introduce **SCOPE-HN**: an open-access dataset of 942 endoscopic frames extracted from 106 flexible nasopharyngolaryngoscopy videos of patients with confirmed OPC. Each image is annotated with pixel-level labels across 12 semantic classes—including tumor, mucosal landmarks, and artifacts—by clinical experts.

SCOPE-HN is the first dataset of its kind focused on OPC and is intended to catalyze AI development in segmentation, classification, and endoscopic scene understanding. Hosted via Stanford AIMI, it aims to foster reproducibility, benchmarking, and innovation in AI-augmented OPC detection and care.

2. Related Work

The growing interest in artificial intelligence (AI) for head and neck oncology has spurred the development of various datasets to facilitate model training and evaluation. However, the majority of publicly available datasets for head and neck cancers (HNCs) are primarily radiology-based, offering cross-sectional imaging data such as computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET). Notable examples include the HECKTOR dataset, which provides PET/CT imaging with tumor annotations for head and neck squamous cell carcinoma, and various collections within The Cancer Imaging Archive (TCIA), offering large repositories of CT and PET scans with associated segmentations for anatomical structures and tumors Laves et al. (2020). While these datasets are invaluable for tasks like tumor localization, volume delineation, and treatment planning in a radiological context, they inherently lack the surface-level detail and mucosal visibility critical for the detection of subtle or early-stage oropharyngeal carcinoma (OPC). Endoscopic imaging, which directly visualizes the mucosal surface, offers unique insights that cannot be captured by these modalities.

In the realm of endoscopic imaging datasets, efforts have largely focused on other anatomical regions or broader categories of the upper aerodigestive tract (UADT). For instance, numerous datasets exist for gastrointestinal (GI) endoscopy, such as Kvasir-SEG for polyp segmentation in colonoscopy, which has significantly advanced AI research in that field Jha et al. (2019). In the UADT, some datasets offer images of laryngeal anatomy, vocal cord lesions, or general pathologies, often supporting tasks like classification or semantic segmentation of broad areas like "vocal folds" or "pathology" Sampieri et al. (2023); Laves et al. (2020); Wilmes et al. (2025). For example, studies have provided datasets for laryngeal lesion assessment using contact endoscopy and narrow-band imaging Wilmes et al. (2025); Paderno et al. (2021), and some work has explored semantic segmentation in laryngeal surgery videos Wilmes et al. (2025). However, these datasets typically do not focus specifically on the oropharynx, a distinct anatomical subsite with unique challenges related to its complex mucosal folds, lymphatic tissue (e.g., tonsils, base of tongue), and varied tumor presentations. The subtle, often submucosal spread of HPV-associated OPC, in particular, makes its endoscopic detection and delineation uniquely challenging and under-represented in existing data resources.

The lack of publicly available, expertly annotated endoscopic datasets specific to the oropharynx has created a major bottleneck in developing and validating advanced AI tools in this domain. Thus, current research efforts rely on private, institution-specific datasets, hindering reproducibility, comparative studies, and the broader application of developed algorithms Paderno et al. (2024); Wilmes et al. (2025). This limitation is particularly pronounced for pixel-level semantic segmentation, which demands meticulous and time-consuming expert annotation.

SCOPE-HN addresses this gap, being the first open-access dataset designed specifically for oropharyngeal cancer endoscopy. This novel resource distinguishes itself by offering detailed, pixel-level semantic segmentation annotations across 12 clinically relevant classes for a large collection of nasopharyngoscopy images of patients with histologically confirmed OPC. By providing high-quality, comprehensive data for this under-represented anatomical region, SCOPE-HN aims to serve as a foundational benchmark, enabling the development of more accurate, robust, and clinically applicable AI models for OPC diagnosis, surveillance, and real-time endoscopic guidance.

3. Material and Methods

3.1 Data Acquisition

The dataset utilized in this study consists of videos selected from **106 unique patients** undergoing flexible oropharyngoscopy at Stanford University between 2015 and 2023. All patients had histologically confirmed OPC. The endoscopic videos were acquired using either the ENF-VH/V3a or the ENF-V/VQ Olympus flexible videolaryngoscope fiberoptic endoscope. All video captures were performed by the same experienced Head and Neck surgeon prior to the initiation of any OPC treatment, ensuring consistency in acquisition technique. The videos and images are captured in 480p or greater.

The demographic and clinical characteristics of the patient cohort are summarized in Table 1. The dataset includes a diverse representation of T-classifications and tumor subsites, reflecting the heterogeneity of OPC presentation.

From each patient's video, we systematically extracted between 8 and 10 representative image frames. This selection process aimed to provide diverse views of the identified tumors and the surrounding anatomical tissue within the oropharynx, capturing a variety tumor type (exophytic vs. endophytic, e.g.) and anatomical contexts. In total, this process yielded **942 high-resolution images** that form the basis of our annotated dataset.

Table 1: Patient Demographic and Clinical Characteristics (N=106)

Characteristic	Value
Age	
Mean (SD)	63.8 (9.78)
Median [Min, Max]	64.0 [38.0, 88.0]
Race	
Asian	5 (4.7%)
Black	1 (0.9%)
Hispanic/Latino	10 (9.4%)
Other	2 (1.9%)
Pacific Islander	1 (0.9%)
White	87 (82.1%)
Ethnicity	
Hispanic/Latino	11 (10.4%)
Non-Hispanic	91 (85.8%)
Unknown	4 (3.8%)
Gender	
Female	15 (14.2%)
Male	91 (85.8%)
T Classification	
T1	27 (25.5%)
T2	46 (43.4%)
T3	16 (15.1%)
T4	17 (16.0%)
Laterality	
Left	51 (48.1%)
Midline	1 (0.9%)
Right	54 (50.9%)
Tumor Subsite	
Base of Tongue	55 (51.9%)
Glossopharyngeal Sulcus	7 (6.6%)
Overlapping Sites of the Oropharynx	4 (3.8%)
Palatine Tonsil	36 (34.0%)
Vallecula	4 (3.8%)

3.2 Annotation Protocol

Image annotation was performed using the Hasty.AI annotation platform Hasty.ai and CloudFactory (2025). An extensive multi-stage annotation protocol was implemented to ensure high-quality, pixel-level labels. Two board-certified Head and Neck Surgeons performed the primary image annotations, delineating specific anatomical structures and tumor regions. A third independent surgeon subsequently verified these initial annotations, providing an additional layer of clinical review and quality control. Finally, a non-clinician expert performed a further check for potential labeling errors, leveraging Hasty.AI's manual review tool to identify and correct any inconsistencies. The annotations provide pixel-level class labels for each image. The full list of the **12 clinically relevant classes** and their respective occurrences within the dataset is detailed in **Table 2**. Given the subtle and often submucosal nature of OPC, annotators were specifically encouraged to utilize information from the entire video sequence when inferring tumor boundaries for individual frames, as these are often best discerned from dynamic endoscopic views. Figure 1 illustrates the overall annotation pipeline, while Figure 2 presents example images alongside their corresponding expert annotations.

Table 2: List of classes and their occurrence in the dataset.

Class	# of annotated instances
Tumor	**953**
Epiglottis (normal)	853
Lympho-Epithelium (normal "tonsil" tissue)	731
Pharyngeal mucosa (no lymphoepithelium)	921
Secretions & obstructions	162
Shadow & poorly lit area	28
Digital objects (Text, etc)	619
Vocal cord (normal)	628
Black background	1311
Uvula/Soft Palate (normal)	327
Spectral Reflectance	56
Pyriform sinus / hypopharynx (normal)	680

Although quantitative inter-observer agreement metrics were not calculated, the multi-tiered expert review protocol ensures a high degree of annotation reliability and reflects best practices for clinical segmentation tasks.

4. Data Format and Access

4.1 File Types and Organization

The **SCOPE-HN** dataset is publicly available at `scope-hn.stanford.edu` or `bit.ly/SCOPE-HN`, under the Stanford University Dataset Research Use Agreement license. The dataset includes endoscopic still images, corresponding pixel-wise segmentation masks, and associated metadata. The file structure and formats are as follows:

- **Images:** All endoscopic frames are stored as high-resolution RGB images in `.jpeg` or `.png` format.
- **Annotations:** Each image is paired with a corresponding segmentation mask in `.png` format, where each pixel label corresponds to one of the 12 predefined classes. The masks use unique integer values to represent each semantic class.
- **Metadata:** A separate `.csv` file is included that contains

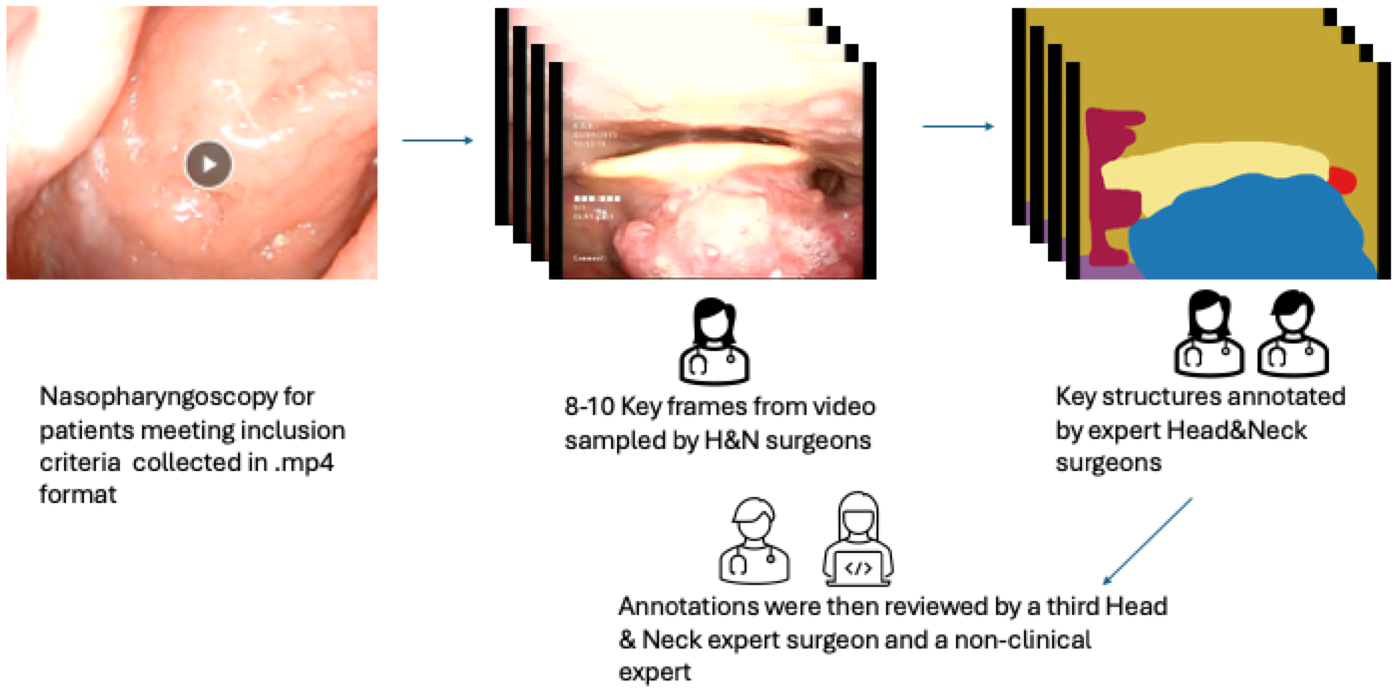

Figure 1: Overview of the data acquisition pipeline: Videos are collected during oropharyngeal endoscopy. Individual frames are extracted and manually annotated. A secondary expert review ensures annotation quality.

de-identified metadata for each patient, including:

- Age (continuous)
- Sex (Male/Female)
- Race and Ethnicity
- T Classification (T1–T4)
- Tumor laterality and subsite

- **ReadMe and Class Map:** A `README.md` file provides usage instructions, and a `class_index.json` file maps label indices to their corresponding semantic class names.

A subset (104) of the de-identified original diagnostic nasopharyngolaryngoscopy videos will be made available in `.mp4` format pending final institutional review, which is currently underway. All videos have been manually redacted using bounding boxes to obscure any identifying facial features or text, and audio has been removed to ensure full de-identification in compliance with HIPAA standards.

5. Potential Use Cases and Applications

The SCOPE-HN dataset is designed to serve as a foundational resource for the development and evaluation of machine learning models in endoscopic oropharyngeal imaging. Potential applications include:

- **Semantic segmentation of tumors and anatomical structures**, enabling real-time tissue identification during flexible endoscopy.
- **Tumor classification and subtype differentiation** using patient-level metadata and image-derived features.
- **Benchmarking computer vision models** for OPC detection in comparison to human expert performance.
- **Model robustness evaluation** across a diverse patient cohort with variation in tumor stage, subsite, and visual appearance.

Preliminary Model Validation

To demonstrate the utility of SCOPE-HN for AI applications, we trained a semantic segmentation model using the FCBFormer architecture, a vision transformer designed for efficient and accurate medical image segmentation. This preliminary model was trained on pixel-level annotations from the dataset to segment oropharyngeal tumors and anatomical structures. Results indicate strong potential for automated analysis of endoscopic images. These findings were presented as a poster at the American Head and Neck Society (AHNS) Annual Meeting 2025 Rau et al. (2025).

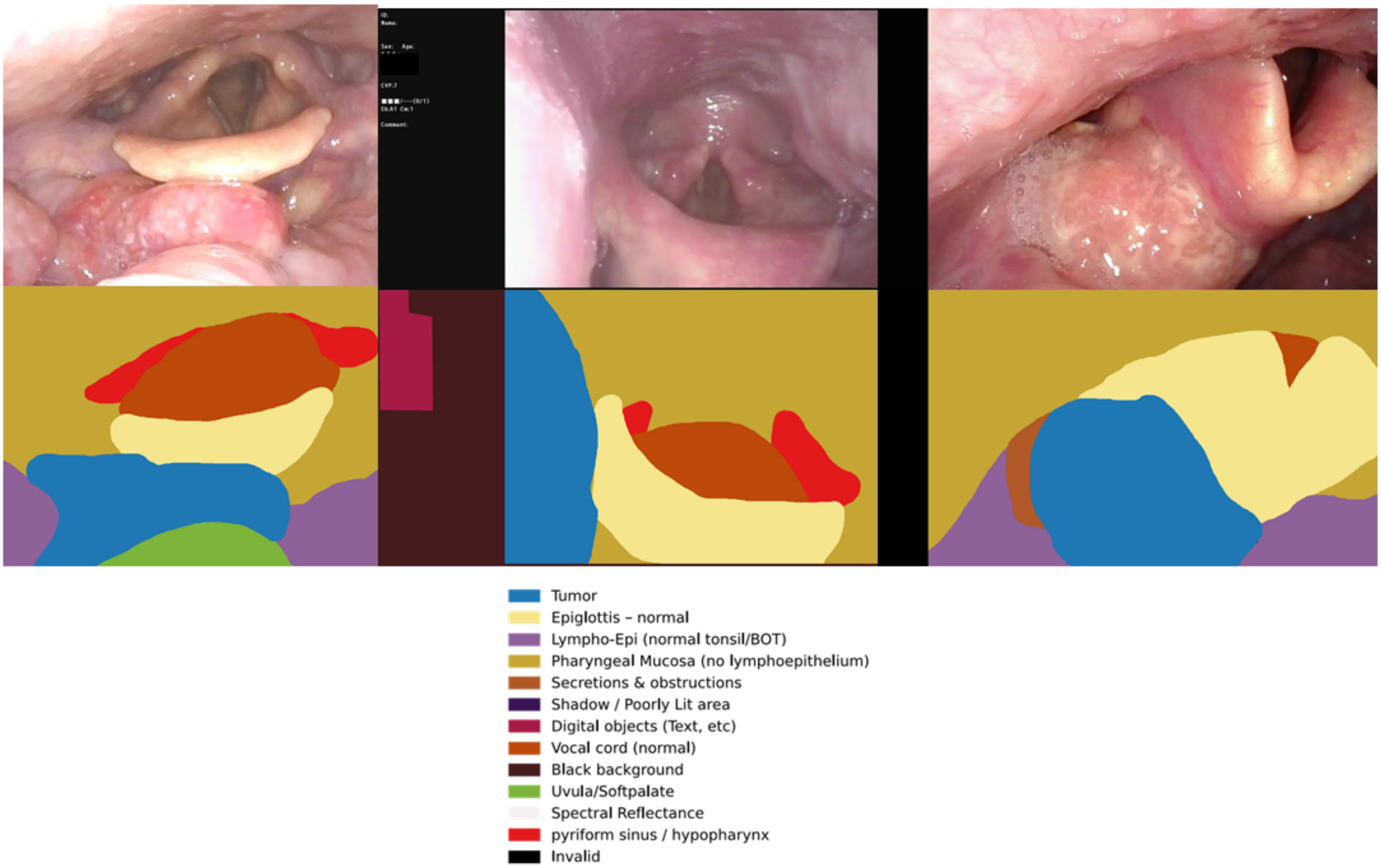

Figure 2: Sample images and annotations: Each row shows images from the same patient. The right-hand legend maps semantic classes to colors in the segmentation masks.

6. Ethics and Data Availability

This study was conducted under Institutional Review Board (IRB) approval from Stanford University (IRB protocol #56217). All patient data were de-identified in accordance with HIPAA guidelines prior to analysis. Only patients with histologically confirmed oropharyngeal squamous cell carcinoma were included. No protected health information (PHI) is present in the dataset.

The SCOPE-HN dataset is publicly available at `scope-hn.stanford.edu` or `bit.ly/SCOPE-HN` under the Stanford University Dataset Research Use Agreement license, which permits research use with proper attribution. Commercial use requires separate licensing and approval from the dataset custodians.

7. Author Contribution

NB led manuscript writing and curated the dataset. AP, YS conducted clinical annotation and FCH, AR helped with validation. FCH also conceived the study and supervised all aspects of data acquisition and clinical protocol. All authors contributed to manuscript drafting and approved the final version.

8. Competing Interests

The authors declare no competing interests.

9. Acknowledgements

We thank the patients and clinical teams at Stanford Health-Care for their contributions to this work. The authors acknowledge support from the Isackson Family Foundation, the Stanford Head and Neck Surgery Research Fund. We also thank the Hasty.ai team for providing access to their annotation platform.

References

S. Elrefaey, M. A. Massaro, S. Chiocca, F. Chiesa, and M. Ansarin. HPV in oropharyngeal cancer: the basics to know in clinical practice. *Acta Otorhinolaryngologica Italica*, 34(5):299, October 2014. URL `https://pmc.ncbi.nlm.nih.gov/articles/PMC4299160/`.

Farhoud Faraji, Eleni M. Rettig, Hua-Ling Tsai, Margueritta El Asmar, Nicholas Fung, David W. Eisele, and Carole Fakhry. The prevalence of human papillomavirus in oropharyngeal cancer is increasing regardless of sex or race, and the influence of sex and race on survival is modified by human papillomavirus tumor status. *Cancer*, 125(5):761–769, March 2019. ISSN 1097-0142. .

Hasty.ai and CloudFactory. Hasty.ai: Ai-assisted annotation platform. https://hasty.ai/, 2025. Accessed 2023–2025.

Debesh Jha, Pia H. Smedsrud, Michael A. Riegler, Pål Halvorsen, Thomas de Lange, Dag Johansen, and Håvard D. Johansen. Kvasir-SEG: A Segmented Polyp Dataset, November 2019. URL http://arxiv.org/abs/1911.07069. arXiv:1911.07069 [eess] version: 1.

Max-Heinrich Laves, Jens Bicker, Lüder A. Kahrs, and Tobias Ortmaier. A Dataset of Laryngeal Endoscopic Images with Comparative Study on Convolution Neural Network Based Semantic Segmentation, September 2020. URL http://arxiv.org/abs/1807.06081. arXiv:1807.06081 [cs].

Alberto Paderno, Cesare Piazza, Francesca Del Bon, Davide Lancini, Stefano Tanagli, Alberto Deganello, Giorgio Peretti, Elena De Momi, Ilaria Patrini, Michela Ruperti, Leonardo S. Mattos, and Sara Moccia. Deep Learning for Automatic Segmentation of Oral and Oropharyngeal Cancer Using Narrow Band Imaging: Preliminary Experience in a Clinical Perspective. *Front Oncol*, 11:626602, 2021. ISSN 2234-943X. .

Alberto Paderno, Anita Rau, Nikita Bedi, Paolo Bossi, Giuseppe Mercante, Cesare Piazza, and F.C̃hristopher Holsinger. Computer vision foundation models in endoscopy: Proof of concept in oropharyngeal cancer. *The Laryngoscope*, 134(11):4535–4541, 2024. .

Anita Rau, Nikita Bedi, Alberto Paderno, Yoon K. So, Masanori Teshima, Hlu Vang, Serena Yeung-Levy, and F. Christopher Holsinger. Refining the oropharyngeal exam using computer vision: Automated segmentation of tumors during laryngopharyngoscopy. Poster presented at the American Head and Neck Society (AHNS) Annual Meeting, 2025. Program No. B017. https://www.ahns.info/meetings/abstracts/2025-abstract-presentations/?programnumber=B017.

Claudio Sampieri, Chiara Baldini, Muhammad Adeel Azam, Sara Moccia, Leonardo S. Mattos, Isabel Vilaseca, Giorgio Peretti, and Alessandro Ioppi. Artificial Intelligence for Upper Aerodigestive Tract Endoscopy and Laryngoscopy: A Guide for Physicians and State-of-the-Art Review. *Otolaryngol Head Neck Surg*, 169(4):811–829, October 2023. ISSN 1097-6817. . Number: 4.

Akihito Watanabe, Masanobu Taniguchi, Hitoshi Tsujie, Masao Hosokawa, Masahiro Fujita, and Shigeyuki Sasaki. The value of narrow band imaging for early detection of laryngeal cancer. *Eur Arch Otorhinolaryngol*, 266(7): 1017–1023, July 2009. ISSN 1434-4726. . URL https://doi.org/10.1007/s00405-008-0835-1. Number: 7.

Celine M. L. H. Wilmes, Arsen Goril BSc, Henri A. M. Marres, David J. Wellenstein, and Guido B. van den Broek. A Systematic Review of the Clinical Impact of Implementing Artificial Intelligence in Upper Aerodigestive Tract Endoscopy. *Head & Neck*, n/a(n/a), 2025. ISSN 1097-0347. . URL https://onlinelibrary.wiley.com/doi/abs/10.1002/hed.28213. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hed.28213.

Zachary S Zumsteg, Michael Luu, Philip S Rosenberg, Julia K Elrod, Freddie Bray, Salvatore Vaccarella, Christopher Gay, Diana J Lu, Michelle M Chen, Anil K Chaturvedi, and Marc T Goodman. Global epidemiologic patterns of oropharyngeal cancer incidence trends. *JNCI: Journal of the National Cancer Institute*, 115(12): 1544–1554, December 2023. ISSN 0027-8874. . URL https://doi.org/10.1093/jnci/djad169.