



Published in final edited form as:

Laryngoscope. 2025 March ; 135(3): 1046–1053. doi:10.1002/lary.31812.

Development of Machine Learning Copilot to Assist Novices in Learning Flexible Laryngoscopy

Mattea E Miller, MD^a, Dan Witte, MSIS^b, Ioan Lina, MD^a, Jonathan Walsh, MD^a, Anaïs Rameau, MD MPhil MS FACS^c, Nasir I. Bhatti, MBBS, MD^a

^aDepartment of Otolaryngology-Head and Neck Surgery, Johns Hopkins School of Medicine

^bPerceptron Health, Inc.

^cSean Parker Institute for the Voice, Department of Otolaryngology - Head and Neck Surgery, Weill Cornell Medical College, New York, NY

Abstract

Objectives: Here we describe the development and pilot testing of the first artificial intelligence (AI) software “copilot” to help train novices to competently perform Flexible Fiberoptic Laryngoscopy (FFL) on a mannikin and improve their uptake of FFL skills.

Methods: Supervised machine learning, was used to develop an image classifier model, dubbed the “anatomical region classifier”, responsible for predicting the location of camera in the upper aerodigestive tract and an object detection model, dubbed the “anatomical structure detector”, responsible for locating and identifying key anatomical structures in images. Training data was collected by performing FFL on an AirSim Combo Bronchi X mannikin (United Kingdom, TruCorp Ltd) using an Ambu aScope 4 RhinoLaryngo Slim connected to an Ambu® aView™ 2 Advance Displaying Unit (Ballerup, Ambu A/S). Medical students were prospectively recruited to try the FFL copilot and rate its ease of use and self- rate their skills with and without the copilot.

Results: This model classified anatomical regions with an overall accuracy of 91.9% on the validation set and 80.1% on the test set. The model detected anatomical structures with overall mean average precision of 0.642. Through various optimizations we were able to run the AI Copilot at approximately 28 frames per second (FPS), which is imperceptible from real time and nearly matches the video frame rate of 30 FPS. Sixty-four novice medical students were recruited for feedback on the copilot. While 90.9% strongly agreed/agreed that the AI copilot was easy to use, their self-rating of FFL skills following use of the copilot were overall equivocal to their self-rating without the co-pilot.

Conclusion(s): The AI copilot tracked successful capture of diagnosable views of key anatomical structures effectively guiding users through FFL to ensure all anatomical structures are sufficiently captured. This tool has the potential to assist novices in efficiently gaining competence in FFL.

Graphical Abstract

Address correspondence to: Nasir Bhatti, MBBS, MD, Johns Hopkins University School of Medicine, 601 N. Caroline Street 6th floor, Baltimore, MD 21287. nbhatti1@jhmi.edu.

This AI copilot tracks successful capture of diagnosable views of key anatomical structures effectively guiding users through Flexible Laryngoscopy to ensure all anatomical structures are sufficiently captured. This tool has the potential to assist novices in efficiently gaining competence in this procedure.

Keywords

Flexible Laryngoscopy; AI; Machine Learning; Competence; Surgical Education

Introduction

It has been well documented that most otolaryngologists practice within densely populated metropolitan area.^{1–5} Two thirds of US counties lack a practicing otolaryngology-head and neck surgery (OHNS) specialist with their density per county being associated with the highest education and income quartiles.^{4–9} Using the CMS Provider Utilization and Payment Data Physician and Other Supplier PUF dataset Davis et. al. observed a direct association between access to otolaryngologists able to perform flexible fiberoptic laryngoscopy (FFL) based on billing codes and number of Medicare enrollees further demonstrating the disproportionate clustering of otolaryngologists in urban areas, consistent with prior studies.^{6,9–11} The lack of access to OHNS among low-income rural areas further exacerbates health care disparities across the US. Telemedicine has been a solution proposed to alleviate some of the socioeconomic inequity but has several limitations within OHNS, namely being unable to remotely perform FFL as part of a physical exam.

FFL is one of the most common procedures performed by otolaryngologists and is an early core competency that OHNS residents are expected to acquire. It allows for detailed examination of the sinonasal cavity, nasopharynx pharynx and larynx, and is crucial in the evaluation of patients with voice, swallowing and airway symptoms, patients needing airway management such as awake fiberoptic nasotracheal intubation, and those with head and neck cancer.¹² There are several steps to a successful FFL, including: identification of key structures, successful navigation of patient's anatomy, patient comfort/coaching, and visual capture of each anatomical area of interest which have varied difficulty dependent on patient anatomy, pathology, and learner experience. Recently, there has been interest within other specialties to train residents and Advance Practice Providers (APPs) to perform FFL. For instance, Price et. al. demonstrated that within radiation oncology a simulation-based training (SBT) workshop on FFL for radiation oncology residents was both feasible and increased confidence in head and neck anatomy and FFL procedural skills.¹³

Previous work has shown that a learning curve exists for novices learning to become competent in performing FFL using a manikins model with a mean of six attempts required to achieve competency¹⁴ based on a validated checklist. This learning curve historically has been addressed by having residents perform this procedure on patients needing evaluation. However, this approach risks discomfort, nasal bleeding, and mucosal injury to the patient.¹² These risks are greatest before competence is achieved by the learner. Simulation based training (SBT) has been proposed as a tool for novices to achieve competence in a variety of

otolaryngologic procedures, and both trainers and trainees have ranked FFL as a skill where SBT would be highly useful.¹⁵

Artificial Intelligence (AI) applications are rapidly being investigated within the field of OHNS and AI-based simulation tools may be able to bridge the gap between learning to perform FFL during SBT and performing FFL on a patient.¹⁶ Here we describe the development and prospective pilot testing of machine learning (ML) software “copilot” that uses a pretrained convolutional neural network for image processing of diagnostic laryngoscopy to help train novice medical students to competently perform FFL on a mannikin and improve their uptake of FFL skills.

Methods

This study was approved by the Johns Hopkins Institutional Review Board: IRB00343343.

Objective

In defining the requirements of the AI Copilot a team of two experienced otolaryngologists (NB and AR) determined that when performed on a simplified model (AirSim Combo Bronchi X mannikin [Figure 1] (United Kingdom, TruCorp Ltd) a basic FFL procedure consisted of:

- Entering the nasal cavity and navigating the nasal passage to reach the nasopharynx
- Visualizing the soft palate
- Visualizing the epiglottis and vallecula
- Visualizing the vocal folds
- Withdrawing the scope

A computer scientist then translated these high-level requirements to specific capabilities the AI Copilot would need:

- Identifying the optimal scoping path in the nasal passage
- Identifying the scope’s location
- Highlighting key anatomical structures
- Providing real-time feedback and navigational cues

AI Copilot architecture

To develop the AI Copilot, we used supervised machine learning, in which neural networks learn to predict output labels from human labeled data. The AI copilot consisted of two key machine learning components:

1. An image classifier model, dubbed the “anatomical region classifier”, responsible for predicting the location of camera in the upper airway.

2. An object detection model, dubbed the “anatomical structure detector”, responsible for locating and identifying key anatomical structures in images.

The outputs of these models were then filtered and at times averaged to reduce noise in the system. These outputs were then incorporated as inputs into the logic of a larger system that kept track of the state of the procedure and the camera location. Based on these inputs and the state of the system, instructions and cues were provided to the user via an overlay on the video feed.

To run the AI Copilot, the live video feed from an Ambu aView 2 Advance was transferred to a computer via an HDMI connection and a video capture card. Running a local FastAPI web server¹⁷, the computer read in the video feed, processed it, and sent the modified feed along with metadata to a web browser through a webhook. This new video feed was displayed on the computer monitor, allowing the Machine Learning Copilot to provide real-time cues to the user overlaid on top of the raw FFL video. See **Video 1** for a representation of the cues provided.

Data Collection

The training data was collected by performing FFL on an AirSim Combo Bronchi X mannikin [Figure 1] (United Kingdom, TruCorp Ltd) using an Ambu aScope 4 RhinoLaryngo Slim [Figure 1] connected to an Ambu aView 2 Advance Displaying Unit (Ballerup, Ambu A/S) [Figure 1]. A representative video of the AI copilot annotating the FFL in real time can be seen in **Video 1**.

Data for training both models were selected from a dataset that was generated by performing 20 flexible laryngoscopies on the mannikin. All scoping videos were performed by a single, experienced otolaryngologist on an Ambu aScope 4 RhinoLaryngo Slim [Figure 1]. Various levels of expertise were simulated when recording videos to ensure a variety of angles and views were captured. The FFL AI copilot was trained on the left nasal cavity only to keep the machine learning consistent, and so there was no confusion or misclassification of structures based on sidedness.

Anatomical Region Classifier

The anatomical region classifier was used to identify the physical location of the camera at the time of image capture. The classes used were Larynx, Nasal Cavity, Nasopharynx, Oropharynx, Out of Body, and Whiteout. Videos were split into image frames and each image was assigned to a class by a trained graduate student. The classes were defined in the manikin model as follows:

- Out of body: Any image from outside of the mannikin.
- Nasal cavity: The scope had entered the mannikin but not yet past the posterior nasal septum.
- Nasopharynx: Between the distal most end of the septum but not yet past a clear delineation in the material where the velum would be in a human.

- Oropharynx: Past the velum delineation but not yet past the superior border of the epiglottis.
- Larynx: Past the superior edge of the epiglottis but had not yet passed through the vocal folds to the trachea.
- Whiteout: If the scope was too close to the mucosa such that the image was 80% red or whited out and a human observer could not reasonably determine where the scope was currently located without information from previous frames.

In total there were 56,262 images in the anatomical region dataset. They were broken down into 42,971 training images, 3,596 validation images, and 9,695 testing images. See Table 1 for breakdown of the class splits. To prevent data leakage due to the high correlation between consecutive frames, each video and all its images were assigned exclusively to a single split: train, validation, or test.

The region classifier used the Resnet18 convolutional neural network model pretrained on ImageNet and was finetuned on the new dataset designed for anatomical region classification within the manikin¹⁸. The final linear layer of the Resnet18 was replaced with a new linear layer with six outputs for the six anatomical regions. Models were trained using a sweep of hyperparameters of frozen layers, learning rate, image transforms, and batch size.

Anatomical structure detector

The anatomical structure object detector is used to identify key anatomical structures in the manikin and place a bounding box around each structure. The anatomical structures labeled were Inferior turbinate, Middle turbinate, Uvula, Vallecula, Epiglottis and Vocal Folds. In addition, the path we desired for the user to take in the nasal cavity was labeled into 2 separate classes, one for the path leading up to the middle turbinate (Path 1), and one for the path after passing the proximal end of the inferior turbinate (Path 2). Images were labeled with bounding boxes by trained graduate students using Label Studio, a data annotation tool¹⁹. Some of the structures did not have apparent boundaries with clear delineations, which led to noisy bounding box labels. A confusion matrix was generated for the anatomical region classifier which can be seen in [Figure 2].

The anatomical structure detector used a YoloV7 model that was finetuned on a dataset made of 11,337 images from a subset of 16 videos and evaluated against 3,096 images from 4 videos²⁰. No validation set was used due to the time intensive nature of labeling the data. See Table 2 for breakdown of the class splits. Models were trained using a hyperparameter sweep of image size, learning rate, image transforms, and batch size. Since multiple structures could be identified within a single frame, predicting the maximum likelihood class was not viable when integrating the anatomical structure detector into the AI Copilot. Instead, each class had its own confidence threshold that was hand-tuned based on human judgement from interacting with the complete system.

The performance of the model was measured using mean average precision (mAP). Mean average precision is a standard metric used in object detection and has the benefit of

balancing precision and recall. An intersection over union threshold of 0.5 was used to calculate the mAP since the ground truth bounding box labels for some classes were noisy.

The AI copilot was pilot tested prospectively by having 64 medical students naïve to FFL use the AI copilot to perform FFL on the AirSim Combo Bronchi X mannikin [Figure 1] (United Kingdom, TruCorp Ltd). After using the FFL, anonymous surveys were handed out to the medical students, and they were then asked to rate the ease of use of machine learning copilot during FFL, as well as self-rate their FFL skills with and without the Copilot, both on a 5-point Likert Scale. Figure 3 shows percentages of students who answered within each Likert category (strongly disagree, somewhat disagree, neutral, somewhat agree, strongly agree). Descriptive statistics were used to analyze medical student responses to ease of use of the tool, and subjective skill set before and after use of the tool. This was a proof-of-concept study to test the feasibility of the AI copilot. The authors plan to do a formal study evaluating the impact of the AI copilot on novice learners.

Results

Anatomical region classifier

The best performing model had 6 frozen layers and was trained with a batch size of 128 for 4 epochs with a learning rate of .0002. This model achieved an overall accuracy of 91.9% on the validation set and 80.1% on the test set. A breakdown of the per class metrics can be found in Table 3. From the confusion matrix [Figure 2], it was apparent that the whiteout class was most frequently incorrectly classified as the nasal cavity where most of the whiteouts happened in the training data. Anatomical regions were most likely to be misclassified as one of their adjacent regions since the transition point from one region to another is not discrete.

Anatomical structure detector

The best performing model was trained for 80 epochs using a batch size of 128, and a learning rate of .001. This model achieved an overall mean average precision of 0.642. A breakdown of the per class metrics can be found in Table 4. Large variability was seen in the mAP across classes. With strong delineations, structures like the vocal folds and epiglottis likely had more consistency and less noise in their bounding box labeling, and therefore were easier for the model to learn.

AI Copilot compute performance

Initially, the AI Copilot was designed to run on a computer with a graphics processing unit (GPU), but we felt this limited the potential environments in which this might eventually be used. To make the software more broadly usable we focused on making it work on a MacBook Pro M1. Through various optimizations we were able to run the AI Copilot at approximately 28 frames per second (FPS), which is imperceptible from real time and nearly matches the video frame rate of 30 FPS.

As seen in Figure 3, 90.9% of medical students naïve to FFL strongly agreed/agreed that the AI copilot was easy to use. However, as illustrated on Figure 4, medical students' self-rating

of FFL skills following use of AI copilot were overall equivocal to their self-rating without the co-pilot.

Discussion

Though in their infancy within medicine AI applications are rapidly expanding. In particular, the use of machine learning for real time image processing during procedures and in surgery shows great potential to provide guidance and navigation to trainees in various specialties, as well as attending surgeons in need of feedback on their skills.^{21,22} Within OHNS, tele-otoscopy augmented with machine learning algorithms has become possible and shows promise in aiding health care providers outside of OHNS in performing otoscopic analysis.^{23–26} By guiding providers less experienced with otoscopy in correct otoscope placement combined with image processing, users have an assistive tool in identifying ear pathology which in turn can help in making timely referral to Otolaryngologists when needed.^{23–25} Similar to the work being done with tele-otoscopy, the AI copilot presented in this study uses a pretrained convolutional neural network for image processing of laryngoscopy, which can be used to help novices improve their uptake of FFL skills and competently perform FFL.

The AI copilot includes an image capture guidance system as well as a procedure checklist and quality check system that tracks successful capture of diagnosable views of key anatomical structures effectively guiding users through FFL to ensure all anatomical structures are sufficiently captured. The mAPs in this study are within the same general range of performance and are comparable to similar models focused on classifying structural laryngeal lesions for laryngeal cancer detection.^{27,28} This tool has the potential to assist novices in becoming competent to perform FFL and may be a helpful adjunct to assist junior otolaryngology residents in gaining competence with this procedure. Numerous studies have shown that skills learned during simulation-based training sessions are transferable clinical practice.²⁹ With the addition of this AI copilot to simulation based training there may be added benefit for new trainees as they start learning early on in residency. Additionally, there is a recognized need for training in FFL within other specialties who take care of patients with pathologies in the upper aerodigestive tract, such as Emergency Medicine physicians, anesthesiologists, radiation oncologists, and advance practice providers.¹³ In specialties where FFL is not used regularly, a copilot tool as presented here could be of potential benefit to guide physicians and providers in performing FFL. Given the recognized disparity in otolaryngologic care within rural communities, expanding number of providers who are competent in FFL has the potential to play a role in increasing access to vital otolaryngologic care.^{1–8,10,11}

Future applications of this AI copilot are not limited to training providers to be competent performing FFL. Recently, a deep learning tool for automated classification of vocal fold polyps during FFL has been shown to distinguish healthy vocal folds from vocal folds with polyps providing potentially a key resource in the development of computer aided disease recognition systems within the larynx.³⁰ We envision the copilot could be used in conjunction with automated classification tools to guide providers in resource limited settings both with performance of FFL and diagnosis of disease. If concerning pathology

was detected, patients could then be referred on to the specialist care. Though early in the development of such technology, these advances in the application of machine learning AI tools to FFL have promise in expanding access to vital patient care for those living in areas with little to no access to otolaryngologic care.

Though this technology shows promise to decrease the gap in disparities in access to otolaryngologic care, it is not without limitations. The machine learning copilot was only tested in manikin with non-variable anatomy and only one modality of video data capture was used, limiting the generalizability of the copilot to real clinical scenarios. In the future, the AI co-pilot will need to be trained on clinical data with various FFL devices to prior to testing it on actual clinical FFLs. The copilot should be trained on a robust set of patient data and evaluated in patients to determine efficacy in clinical practice. Additionally, machine learning tools are susceptible to the biases in datasets they are trained on and can even amplify bias in healthcare.^{16,31–34} Ultimately, we foresee this AI copilot as a tool that can assist a provider during FFL, though the responsibility of performing competent FFL remains with the practitioners performing the exam. Precision of the anatomical structure detector was limited, which likely reflects the overall limited size of the dataset. Finally, medical students did not perceive an improvement of skills with the use of the AI co-pilot, though they reported ease of use of the software, indicating feasibility. Future studies should compare the AI copilot to traditional teaching methods for teaching novices to perform FFL and better tease out factors that may improve usefulness of the AI copilot in training.

Conclusion(s):

We described the development and pilot testing of the first AI copilot to help train novices to competently perform FFL on a mannikin. The AI copilot tracked successful capture of diagnosable views of key anatomical structures effectively guiding users through FFL to ensure all anatomical structures are sufficiently captured. This tool has the potential to assist novices in efficiently gaining competence in FFL.

Funding and Conflict of Interests:

This work was funded by NIH # 1R41EB034186-01. Data analysis done by Perceptron Health, Inc., New York, NY, 10044. Anaïs Rameau was supported by a Paul B. Beeson Emerging Leaders Career Development Award in Aging (K76 AG079040) from the National Institute on Aging and by the Bridge2AI award (OT2 OD032720) from the NIH Common Fund. Anaïs Rameau owns equity of Perceptron Health, Inc. She is a medical advisor for Sound Health Systems, Inc.

References

1. GlobalData Plc. The Complexities of Physician Supply and Demand: Projections From 2021 to 2036. Published online March 2024. Accessed April 9, 2024. <chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.aamc.org/media/75236/download?attachment>
2. Williams TE, Satiani B, Thomas A, Ellison EC. The impending shortage and the estimated cost of training the future surgical workforce. *Ann Surg*. 2009;250(4):590–597. doi:10.1097/SLA.0b013e3181b6c90b [PubMed: 19730238]
3. Pillsbury HC, Cannon CR, Sedory Holzer SE, et al. The workforce in otolaryngology-head and neck surgery: moving into the next millennium. *Otolaryngol Head Neck Surg*. 2000;123(3):341–356. doi:10.1067/mhn.2000.109761 [PubMed: 10964321]

4. Gadkaree SK, McCarty JC, Siu J, et al. Variation in the Geographic Distribution of the Otolaryngology Workforce: A National Geospatial Analysis. *Otolaryngol Head Neck Surg.* 2020;162(5):649–657. doi:10.1177/0194599820908860 [PubMed: 32122214]
5. Cass LM, Smith JB. The Current State of the Otolaryngology Workforce. *Otolaryngologic Clinics of North America.* 2020;53(5):915–926. doi:10.1016/j.otc.2020.05.016 [PubMed: 32682535]
6. Vickery TW, Weterings R, Cabrera-Muffly C. Geographic Distribution of Otolaryngologists in the United States. *Ear Nose Throat J.* 2016;95(6):218–223. doi:10.1177/014556131609500607 [PubMed: 27304439]
7. Lango MN, Handorf E, Arjmand E. The geographic distribution of the otolaryngology workforce in the United States. *The Laryngoscope.* 2017;127(1):95–101. doi:10.1002/lary.26188 [PubMed: 27774588]
8. Hur K, Gibbons J, Finch BK. Geographic Heterogeneity in Otolaryngology Medicare New Patient Visits. *Otolaryngol--head neck surg.* 2020;162(6):860–866. doi:10.1177/0194599820913495 [PubMed: 32204646]
9. Davis RJ, Exilus S, Best S, Willink A, Akst LM. The Geographic Distribution of Videolaryngostroboscopy in the United States. *Journal of Voice.* 2023;37(5):798.e7–798.e14. doi:10.1016/j.jvoice.2021.05.002
10. Winters R, Pou A, Friedlander P. A “Medical Mission” at Home: The Needs of Rural America in Terms of Otolaryngology Care: A “Medical Mission” at Home. *The Journal of Rural Health.* 2011;27(3):297–301. doi:10.1111/j.1748-0361.2010.00343.x [PubMed: 21729157]
11. Urban MJ, Wojcik C, Eggerstedt M, Jagasia AJ. Rural–Urban Disparities in Otolaryngology: The State of Illinois. *The Laryngoscope.* 2021;131(1). doi:10.1002/lary.28652
12. Alvi S, Harsha P. Flexible Nasopharyngoscopy. In: StatPearls. StatPearls Publishing; 2024. Accessed April 9, 2024. <http://www.ncbi.nlm.nih.gov/books/NBK539740/>
13. Price JG, Spiegel DY, Yoo DS, et al. Development and Implementation of an Educational Simulation Workshop in Fiberoptic Laryngoscopy for Radiation Oncology Residents. *Int J Radiat Oncol Biol Phys.* 2020;108(3):615–619. doi:10.1016/j.ijrobp.2020.05.009 [PubMed: 32417408]
14. Laeeq K, Pandian V, Skinner M, et al. Learning curve for competency in flexible laryngoscopy. *Laryngoscope.* 2010;120(10):1950–1953. doi:10.1002/lary.21063 [PubMed: 20824777]
15. Lüscher M, Nayahangan LJ, Thrane JF, Danstrup CS, Konge L, Andersen SAW. Simulation-based surgical training needs in otorhinolaryngology. *Dan Med J.* 2022;69(4):A08210630. [PubMed: 35319453]
16. Alter IL, Chan K, Lechien J, Rameau A. An introduction to machine learning and generative artificial intelligence for otolaryngologists-head and neck surgeons: a narrative review. *Eur Arch Otorhinolaryngol.* Published online February 23, 2024. doi:10.1007/s00405-024-08512-4
17. Ramirez S FastAPI. <https://fastapi.tiangolo.com>
18. He K, Zhang X, Ren S, Sun J. Deep Residual Learning for Image Recognition. *CoRR.* 2015;abs/1512.03385. <http://arxiv.org/abs/1512.03385>
19. Tkachenko M, Malyuk M, Holmanyuk A, Liubimov N. {Label Studio}: Data labeling software. Published online 2022 2020. <https://github.com/heartexlabs/label-studio>
20. Wang CY, Liao HYM, Yeh IH. Designing Network Design Strategies Through Gradient Path Analysis. *Journal of Information Science and Engineering.* Published online 2023. doi:10.48550/arXiv.2211.04800 Focus to learn more
21. Mascagni P, Alapatt D, Urade T, et al. A Computer Vision Platform to Automatically Locate Critical Events in Surgical Videos: Documenting Safety in Laparoscopic Cholecystectomy. *Annals of Surgery.* 2021;274(1):e93–e95. doi:10.1097/SLA.0000000000004736 [PubMed: 33417329]
22. Laplante S, Namazi B, Kiani P, et al. Validation of an artificial intelligence platform for the guidance of safe laparoscopic cholecystectomy. *Surg Endosc.* 2023;37(3):2260–2268. doi:10.1007/s00464-022-09439-9 [PubMed: 35918549]
23. Ezzibdeh R, Munjal T, Ahmad I, Valdez TA. Artificial intelligence and tele-otoscopy: A window into the future of pediatric otology. *International Journal of Pediatric Otorhinolaryngology.* 2022;160:111229. doi:10.1016/j.ijporl.2022.111229 [PubMed: 35816971]

24. Habib A, Kajbafzadeh M, Hasan Z, et al. Artificial intelligence to classify ear disease from otoscopy: A systematic review and meta-analysis. *Clinical Otolaryngology*. 2022;47(3):401–413. doi:10.1111/coa.13925 [PubMed: 35253378]
25. Habib AR, Wong E, Sacks R, Singh N. Artificial intelligence to detect tympanic membrane perforations. *J Laryngol Otol*. 2020;134(4):311–315. doi:10.1017/S0022215120000717 [PubMed: 32238202]
26. You E, Lin V, Mijovic T, Eskander A, Crowson MG. Artificial Intelligence Applications in Otology: A State of the Art Review. *Otolaryngol--head neck surg*. 2020;163(6):1123–1133. doi:10.1177/0194599820931804 [PubMed: 32513061]
27. Bur AM, Zhang T, Chen X, et al. Interpretable Computer Vision to Detect and Classify Structural Laryngeal Lesions in Digital Flexible Laryngoscopic Images. *Otolaryngol--head neck surg*. 2023;169(6):1564–1572. doi:10.1002/ohn.411 [PubMed: 37350279]
28. Mamidi IS, Dunham ME, Adkins LK, McWhorter AJ, Fang Z, Banh BT. Laryngeal Cancer Screening During Flexible Video Laryngoscopy Using Large Computer Vision Models. *Ann Otol Rhinol Laryngol*. 2024;133(8):720–728. doi:10.1177/00034894241253376 [PubMed: 38755974]
29. Spiliotis AE, Spiliotis PM, Palios IM. Transferability of Simulation-Based Training in Laparoscopic Surgeries: A Systematic Review. *Minim Invasive Surg*. 2020;2020:5879485. doi:10.1155/2020/5879485 [PubMed: 32908700]
30. Yao P, Witte D, German A, et al. A deep learning pipeline for automated classification of vocal fold polyps in flexible laryngoscopy. *Eur Arch Otorhinolaryngol*. 2024;281(4):2055–2062. doi:10.1007/s00405-023-08190-8 [PubMed: 37695363]
31. Aggarwal P, Papay FA. Artificial intelligence image recognition of melanoma and basal cell carcinoma in racially diverse populations. *Journal of Dermatological Treatment*. 2022;33(4):2257–2262. doi:10.1080/09546634.2021.1944970 [PubMed: 34154489]
32. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447–453. doi:10.1126/science.aax2342 [PubMed: 31649194]
33. Jain A, Brooks JR, Alford CC, et al. Awareness of Racial and Ethnic Bias and Potential Solutions to Address Bias With Use of Health Care Algorithms. *JAMA Health Forum*. 2023;4(6):e231197. doi:10.1001/jamahealthforum.2023.1197 [PubMed: 37266959]
34. Vokinger KN, Feuerriegel S, Kesselheim AS. Mitigating bias in machine learning for medicine. *Commun Med (Lond)*. 2021;1:25. doi:10.1038/s43856-021-00028-w [PubMed: 34522916]



A: Ambu aView 2 Advance
(source getimage.ashx)



B: Ambu aScope 4 RhinoLaryngo Slim
(source getimage.ashx)



C: AirSim Combo Bronchi X
(source AirSim-Advance-X.jpg)

Figure 1:
Set up for Pilot Testing of Machine Learning Co-Pilot. Panel A: Ambu aView 2 Advance;
Panel B: Ambu aScope 4 RhinoLaryngo Slim; Panel C: AirSim Combo Bronchi X.

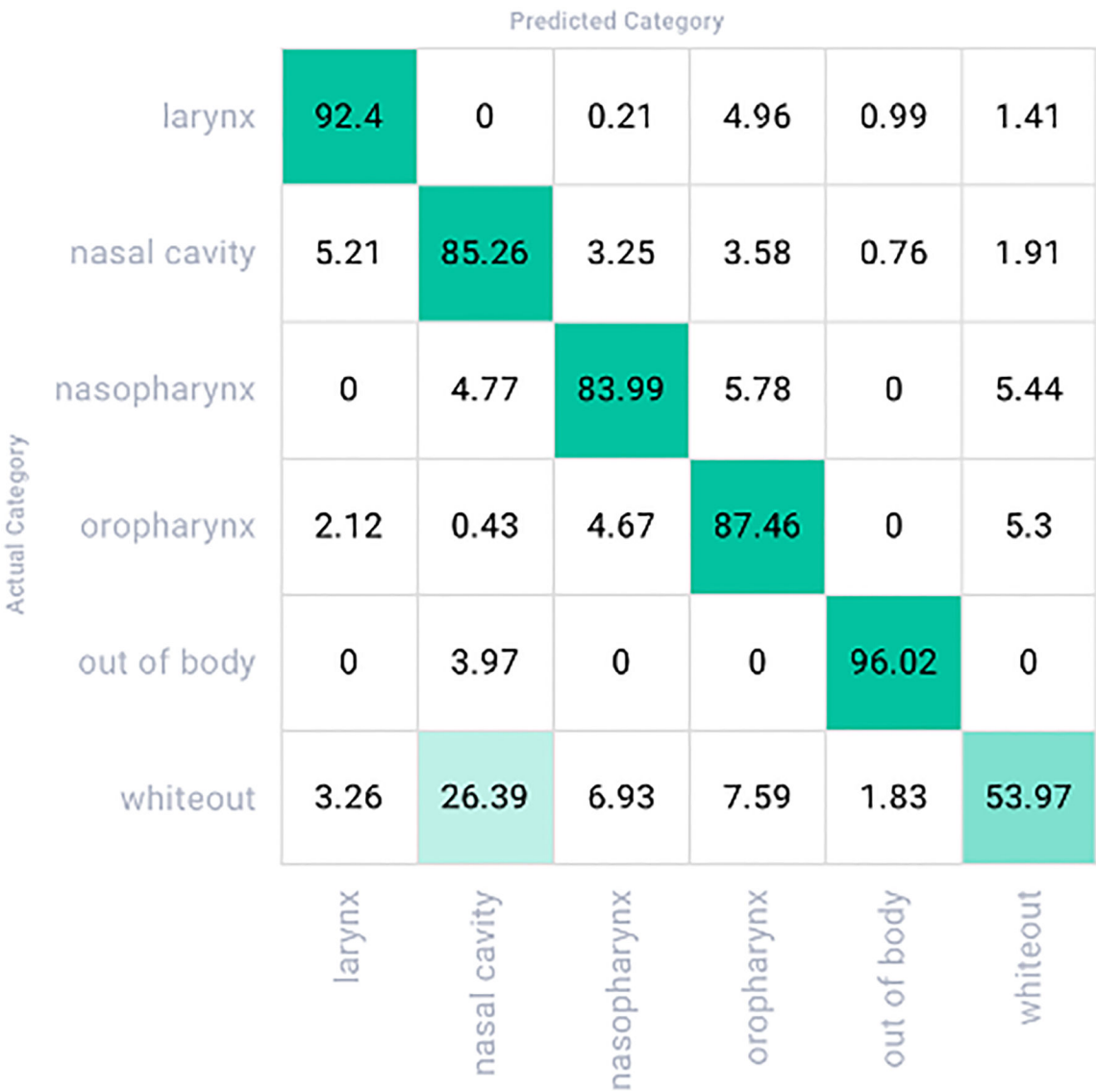


Figure 2:
Confusion matrix for the anatomical region classifier

"I thought the tool was easy to use"

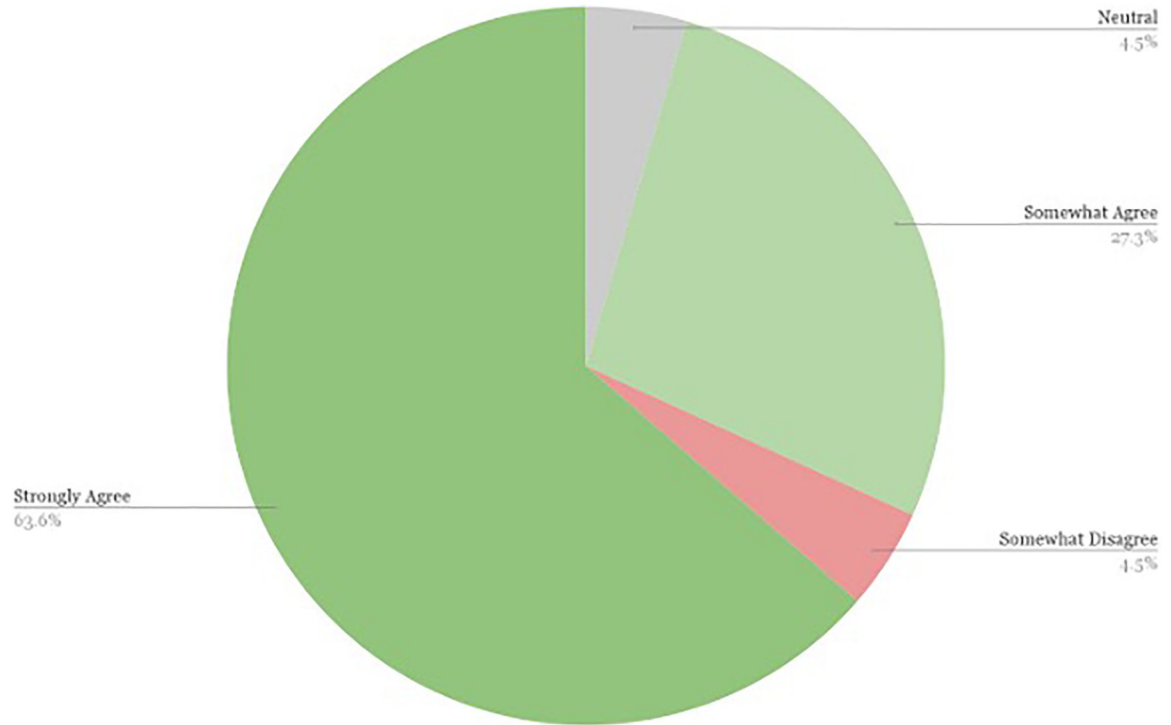


Figure 3:
Medical Student Rating of Ease of Use of Machine Learning Copilot during Flexible
Fiberoptic Laryngoscopy

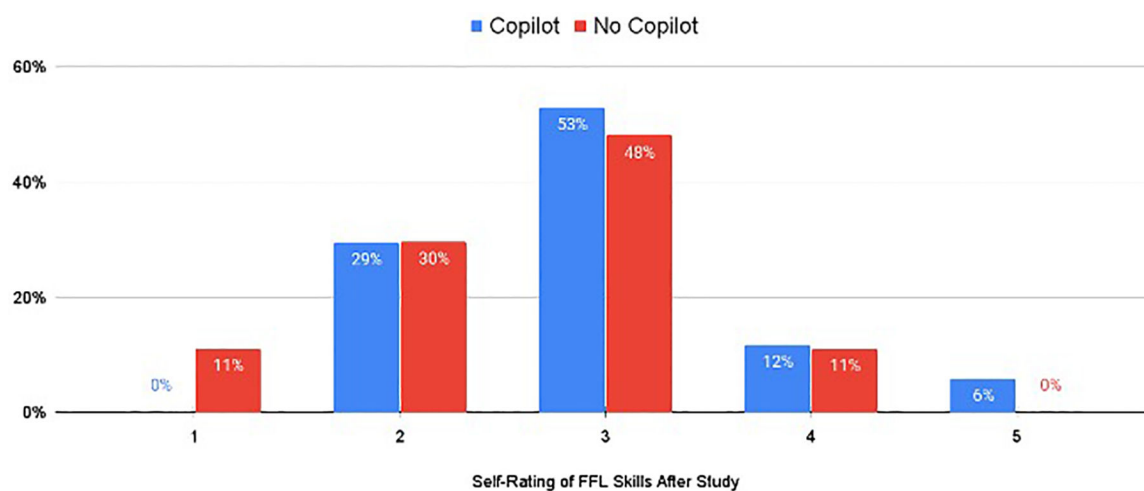


Figure 4:
Medical Student Self-Rating of Flexible Fiberoptic Laryngoscopy Skills with and without Machine Learning Copilot

Embedded Video 1:

Machine Learning Copilot Annotating Flexible Fiberoptic Laryngoscopy in Real time on an AirSim Combo Bronchi X mannikin

Table 1:

Region Classification Dataset Breakdown

Class	Train Images	Validation Images	Test Images
Out of body	730 (2%)	145 (4%)	277 (3%)
Nasal cavity	10,817 (25%)	1780 (49%)	2,762 (28%)
Nasopharynx	7,294 (17%)	749 (21%)	1,487 (15%)
Oropharynx	8,848 (21%)	685 (19%)	2,074 (21%)
Larynx	8,949 (21%)	127 (4%)	1,409 (15%)
Whiteout	6,333 (15%)	110 (3%)	1,686 (17%)
Total	42,971	3,596	9,695

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2:

Anatomical Object Detection Dataset Breakdown

Class	Train Images	Test Images
Epiglottis	1,262 (8%)	597 (15%)
Uvula	1,630 (11%)	327 (8%)
Vallecula	825 (6%)	221 (5%)
Vocal Folds	3,604 (24%)	1,263 (31%)
Inferior Turbinate	1,196 (8%)	298 (7%)
Path 1	1,719 (12%)	420 (10%)
Path 2	4,648 (31%)	961 (24%)
Total	14,884	4,087

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 3:

Region Classifier Performance Metrics

Class	Validation	Test
Overall accuracy	91.9%	81.4%
Out of body precision	.845	.801
Out of body recall	.979	.960
Nasal cavity precision	.944	.815
Nasal cavity recall	.972	.853
Nasopharynx precision	.901	.803
Nasopharynx recall	.788	.840
Oropharynx precision	.914	.826
Oropharynx recall	.920	.875
Larynx precision	.914	.843
Larynx recall	1.0	.924
Whiteout precision	.718	.775
Whiteout recall	.763	.539

Table 4:

Anatomical Structure Detector Performance Metrics

Class	Test Mean Average Precision
Epiglottis	.878
Uvula	.88
Vallecula	.76
Vocal Folds	.937
Inferior Turbinate	.189
Path 1	.269
Path 2	.548
Overall	.642

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript