Received: 28 May 2024 | Revised: 10 September 2024 | Accepted: 11 September 2024

DOI: 10.1002/alr.23458

IFAR: International Forum of Allergy & Rhinology

RESEARCH NOTE

Real-time augmentation of diagnostic nasal endoscopy video using AI-enabled edge computing

Jonathan Bidwell PhD[1] | **Dipesh Gyawali MS**[1] | **Jonathan Morse**[1] | **Vinayak Ganeshan MD**[1] | **Thinh Nguyen MD**[1] | **Edward D. McCoul MD, MPH**[1,2]

[1]Department of Otorhinolaryngology, Ochsner Health, New Orleans, Louisiana, USA

[2]Ochsner Clinical School, University of Queensland, New Orleans, Louisiana, USA

Correspondence
Edward D. McCoul, Department of Otorhinolaryngology, Ochsner Medical Center, 1514 Jefferson Highway, New Orleans, LA 70121, USA.
Email: emccoul@gmail.com

Funding information
Eye, Ear, Nose, and Throat Foundation, Grant/Award Numbers: EE220702, EE220802

KEYWORDS
artificial intelligence, deep learning, machine learning, nasal endoscopy, neural network

KEY POINTS

- AI-enabled augmentation of nasal endoscopy video images is feasible in the clinical setting.
- Edge computing hardware can interface with existing nasal endoscopy equipment.
- Real-time AI performance can achieve an acceptable balance of accuracy and efficiency.

1 | INTRODUCTION

The process of performing nasal endoscopy (NE) has remained largely unchanged since the introduction of digitally integrated endoscopes over 15 years ago.[1] Artificial intelligence (AI) models, particularly deep convolutional neural networks (D-CNN) in machine learning (ML), excel in interpreting medical video, positioning them as potential tools for advancing NE.[2] However, their computational complexity often requires them to run on central servers, raising privacy issues that hinder their adoption in medicine.[2] The application of edge computing—broadly defined as transferring computation capabilities away from servers to local environments—is a promising solution to these issues.[3] By utilizing ML models running locally, healthcare systems can apply these technologies to improve patient care while ensuring the security of their data. Similar applications of the YOLO network for real-time detection in endoscopy, such as in nasopharyngeal carcinoma detection, have shown the feasibility and effectiveness of such approaches.[4]

In this study, we sought to develop and evaluate the performance of a D-CNN, which augmented NE by interpreting the video feed synchronously during the clinical examination.

2 | METHODS

2.1 | Model

The AI-augmented NE instrument utilized the YOLOv8 (You Only Look Once version 8, Ultralytics) architecture, selected for its real-time object detection.[5] We generated

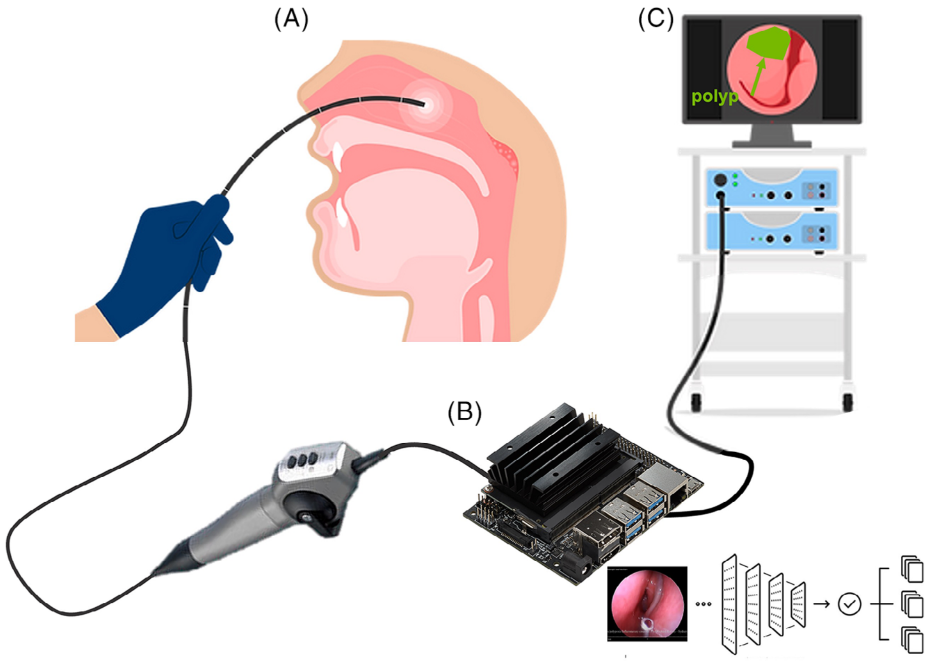

FIGURE 1 Diagram the illustrating integration of our system into a nasal endoscopy (NE) examination clinical workflow. (A) The clinician uses an endoscopic camera to capture images of the nasal cavity. (B) The video feed is transmitted to the AI system, where the machine learning model analyzes the data in real-time. (C) The processed video feed, with the model's predictions overlaid, is then displayed on a monitor, providing augmented visual feedback during the examination.

five distinct variants (nano, small, medium, large, and extra-large) of the YOLOv8 model, each tailored to accommodate varying computational constraints. Traditionally, larger models predict more accurate segmentations at the cost of slower inference time; we evaluated all five variants to find one suitable for deployment on a computer running locally in the endoscopy suite.

2.2 | Dataset and training

We used the Karl Storz Flex Videoendoscope 3.0 × 51.5, equipped with photo capture buttons, to perform NE. Images were taken during routine exams over a 10-year period starting in 2014. Two medical students manually segmented the inferior and middle turbinates from a subset of images, achieving an 80% F1 score inter-rater agreement with an ENT specialist. This dataset, detailed in our prior study,[6] included 2111 labeled images of both normal and diseased states, excluding patients with turbinate resection, and was split into training, validation, and testing subsets at a ratio of 80:15:5. We applied the same transfer learning methods as previously described but omitted these specific results for brevity.[6]

2.3 | Hardware

The Nvidia Jetson Orin Nano Platform (Nvidia) was chosen for its graphics processing unit (GPU) acceleration capabilities, enabling us to run an ML model locally using compute unified device architecture (CUDA) for efficient real-time processing. We housed the nano embedded computer in a plastic enclosure designed and fabricated using Autodesk Fusion360 (Autodesk) and Bambu Labs Slicer (Bambu Lab) with a Bambu Lab P1S 3D Printer. A Telepak (TP-100) streamed video to the Nano enclosure via an HDMI cable, enabling real-time display of the processed video on a separate screen beside the Telepak. Figure 1 shows the workflow from the nasal endoscope camera to the AI-augmented video feed.

2.4 | Testing

The senior author tested the augmented TP-100 system by performing NE on a volunteer. We assessed the inference time of each model variant and recorded videos for each examination.

3 | RESULTS

The models each had comparable accuracy, centered at 87%, despite variations in model size. The smallest model, YOLOv8-nano, provided predictions every 25.6 ms on the Nano computer, enabling the interpretation of 30 Hz video input from the TP-100. We evaluated images using intersection over union criteria: true positives where predicted

TABLE 1 YOLOv8 model variants, parameters, performance, and inference time.

Model	Parameters (million)	Accuracy	Precision	Recall	F1 score	Inference time (ms)
YOLOv8-nano	3.4	86.0%	91.5%	93.0%	92.0%	25.6
YOLOv8-small	11.8	87.3%	92.0%	93.5%	92.5%	46.7
YOLOv8-medium	27.3	88.5%	93.5%	94.0%	93.5%	77.7
YOLOv8-large	46.0	87.6%	92.0%	94.0%	93.0%	127.5
YOLOv8-extra-large	71.8	88.1%	92.5%	94.5%	93.5%	177.2

Note: These speed results are from an Nvidia Jetson Orin Nano Platform, CUDA acceleration.

and actual segmentations overlapped more than 50%, false positives where the model predicted a non-existent segmentation, false negatives where the model missed actual segmentations, and true negatives for correctly identified non-segmentations. The average F1 score, balancing precision (accuracy of positive predictions) and recall (ability to find all positive instances), was 92.9%. The latency between endoscope movement and visual updates on an unmodified TP-100 was 12 ms, with a total latency of 37.6 ms for the augmented TP-100 (Table 1).

4 | DISCUSSION

We demonstrated the feasibility of deploying ML in a clinical environment to interpret an NE examination in real time. Prior studies have evaluated the role of ML algorithms in post hoc analysis of NE imagery, but to our knowledge, this is the first evaluation of software running synchronously during a clinical examination.[7,8] The AI model performance metrics (F1 score, precision, and recall) each surpassed the conventional inter-rater agreement of 0.8, demonstrating the potential to provide immediate clinical decision support[9] (Table 1).

One of the advantages of our deployed model is its ability to work without user input; clinicians already struggle with convoluted user interfaces that can lead to alert fatigue, resulting in missed diagnoses or treatment opportunities.[4] In contrast, our physically unobtrusive system makes predictions without alarms or additional configurations, directly addressing real-world rhinologic clinical practice.

Limitations of this study include using static images for model training, testing on a single volunteer, and hardware constraints. Training on static images may introduce bias compared to continuous video. Testing on a single volunteer limits generalizability. Future studies will include broader pilot testing to address this. The Nano hardware, with 4 GB RAM and 15 W power, limits the model variants deployed. Future hardware upgrades may enable more advanced models to run at comparable speeds, that is, greater than 30 frames per second.

Our findings illustrate AI's growing role in improving clinical examinations and diagnoses. Incorporating semi-supervised learning techniques such as pseudo-labeling and entropy minimization could enhance model accuracy, particularly when dealing with limited annotated data.[10] Future research should focus on developing more robust ML models to achieve the eventual goal of producing semi-autonomous endoscopic clinical decision support.

5 | CONCLUSION

We have demonstrated the successful application of a CNN using edge computing principles to interpret the NE examination video feed. This study provides a starting point for further exploration into the role of AI in the clinical workflow, heralding benefits for efficiency and patient care.

ACKNOWLEDGMENTS

This work was supported by a research grant from the Eye, Ear, Nose, and Throat Foundation (EE220702/EE220802).

CONFLICT OF INTEREST STATEMENT

Edward McCoul is a consultant for 3D Matrix, Advanced Rx, Optinose, Sanofi, Stryker, and Zsquare. The remaining authors declare they have no conflicts of interest.

ORCID

Vinayak Ganeshan MD https://orcid.org/0009-0007-6573-6106
Edward D. McCoul MD, MPH https://orcid.org/0000-0003-1812-2105

REFERENCES

1. Kennedy DW. Technical innovations and the evolution of endoscopic sinus surgery. *Ann Otol Rhinol Laryngol*. 2006;115(_suppl9):3-12. doi:10.1177/00034894061150S902
2. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347-1358. doi:10.1056/NEJMra1814259
3. Chen J, Ran X. Deep learning with edge computing: a review. *Proc IEEE*. 2019;107(8):1655-1674. doi:10.1109/JPROC.2019.2921977

4. Oppenheim MI, Vidal C, Velasco FT, et al. Impact of a computerized alert during physician order entry on medication dosing in patients with renal impairment. In: *Proceedings of the AMIA Annual Symposium*. 2002. p. 577-581.
5. Jocher G, Chaurasia A, Qiu J. *YOLO by Ultralytics*. 2023. https://ultralytics.com
6. Ganeshan V, Bidwell J, Gyawali D, et al. Enhancing nasal endoscopy: classification, detection, and segmentation of anatomic landmarks using a convolutional neural network. *Int Forum Allergy Rhinol*. 2024;14(9):1521-1524. doi:10.1002/alr.23384
7. Yui R, Takahashi M, Noda K, et al. Preoperative prediction of sinonasal papilloma by artificial intelligence using nasal video endoscopy: a retrospective study. *Sci Rep*. 2023;13(1):12439. doi:10.21203/rs.3.rs-2819365/v1
8. Girdler B, Moon H, Bae MR, Ryu SS, Bae J, Yu MS. Feasibility of a deep learning-based algorithm for automated detection and classification of nasal polyps and inverted papillomas on nasal endoscopic images. *Int Forum Allergy Rhinol*. 2021;11(12):1637-1646. doi:10.1002/alr.22854
9. Taha AA, Hanbury A. Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. *BMC Med Imaging*. 2015;15(1):29. doi:10.1186/s12880-015-0068-x
10. Jiao R, Zhang Y, Ding L, et al. Learning with limited annotations: a survey on deep semi-supervised learning for medical image segmentation. *Comp Biol Med*. 2024;169:107840. doi:10.1016/j.compbiomed.2023.107840

How to cite this article: Bidwell J, Gyawali D, Morse J, Ganeshan V, Nguyen T, McCoul ED. Real-time augmentation of diagnostic nasal endoscopy video using AI-enabled edge computing. *Int Forum Allergy Rhinol*. 2024;11111. https://doi.org/10.1002/alr.23458