Received: 17 March 2024 | Revised: 14 May 2024 | Accepted: 23 May 2024

DOI: 10.1002/alr.23384

IFAR: International Forum of Allergy & Rhinology

RESEARCH NOTE

Enhancing nasal endoscopy: Classification, detection, and segmentation of anatomic landmarks using a convolutional neural network

Vinayak Ganeshan MD[1] | **Jonathan Bidwell PhD**[1] | **Dipesh Gyawali MS**[1] | **Thinh S. Nguyen MD**[1] | **Jonathan Morse BS**[1] | **Madeline P. Smith MD**[2,3] | **Blair M. Barton MD**[1,2,3] | **Edward D. McCoul MD, MPH**[1,2,3]

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

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

[3]Department of Otolaryngology, Tulane University School of Medicine, 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 Number: EE220702/EE220802

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

Key points

- A convolutional neural network (CNN)-based model can accurately localize and segment turbinates in images obtained during nasal endoscopy (NE).
- This model represents a starting point for algorithms that comprehensively interpret NE findings.

1 | INTRODUCTION

The nasal cavity is characterized by protrusions and involutions, which can pose substantial challenges to the interpretation of the endoscopic examination.[1] The inferior turbinate (IT) and middle turbinate (MT) are key features of interest on nasal endoscopy (NE) because they are host to or adjacent to a range of pathology.[2] Prior studies have demonstrated significant inter-operator variability when classifying anatomy and interpreting pathology.[3,4] Architectural heterogeneity may further complicate recognition. These factors can cause even skilled clinicians to inadvertently miss subtle signs of sinonasal disease.

Recent developments in machine learning, specifically applying convolutional neural networks (CNNs), have shown excellent performance in tasks such as detecting and segmenting structures in video feeds.[5] There has been limited application of CNN models to endoscopic procedures in rhinology.[6–8] We sought to apply a CNN to identify and delineate these landmarks in the nasal cavity during NE.

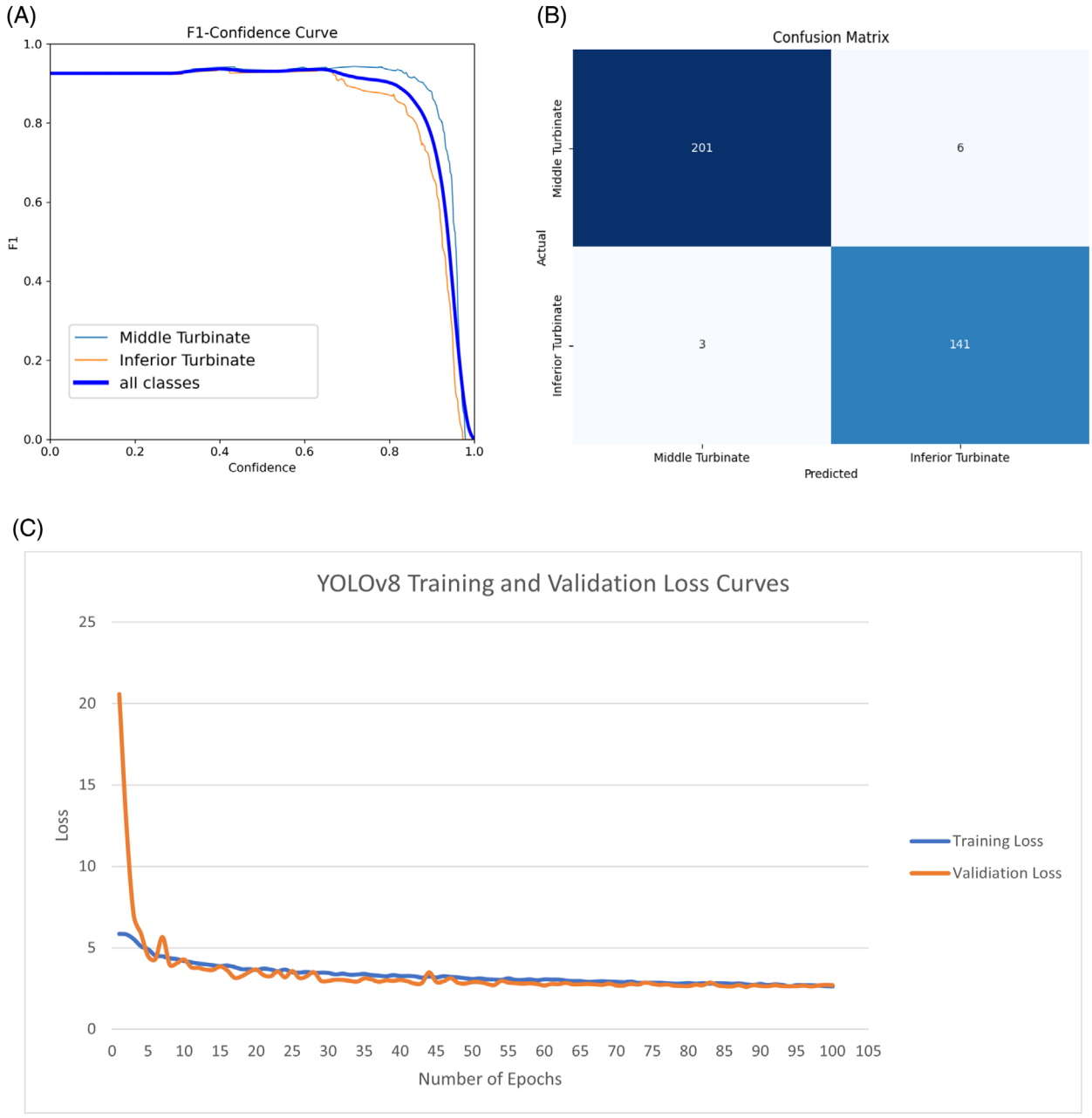

FIGURE 1 (A) F1-confidence curve for the fifth model, with a peak F1-score of 94.1% at the confidence threshold of 60%. (B) Confusion matrix showing performance of the model on the fifth fold validation set. (C) Learning curve of the model as measured by the total loss on the fifth fold training and validation set.

2 | METHODS

Images were obtained by the senior author during NE examinations for the evaluation of sinonasal disease at Ochsner Medical Center in New Orleans, LA, between 2014 and 2023. Images were acquired using a 4-mm flexible digital endoscope (Karl Storz) with a resolution of 1024 by 768 pixels, saved in JPEG format. Images eligible for labeling met the following criteria: (1) the surface of the IT/MT was clearly identifiable, and (2) the image had no distortion, clouding, obstruction, or artifact. We randomly selected 2111 images to manually segment the IT and/or MT. Segmentation masks were drawn by three physicians using the open-source CocoAnnotator labeling software (https://github.com/jsbroks/coco-annotator). This dataset was randomly divided into training, validation, and testing subsets with an 80/15/5 split.

The open-source YOLOv8 (https://github.com/ultralytics/ultralytics) object detection model (Ultralytics) was configured to perform three tasks: classify whether a turbinate was present, detect its location, and apply a segmentation mask delineating its borders.[9]

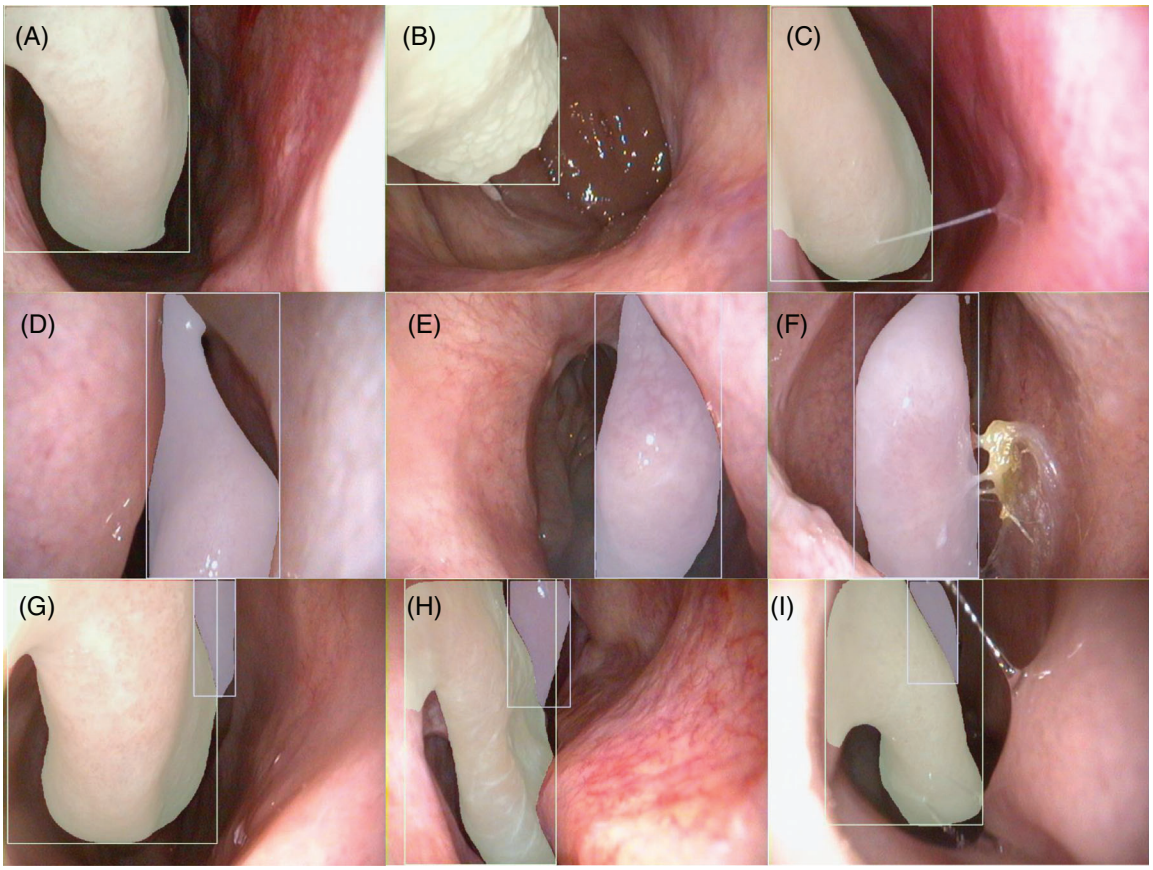

FIGURE 2 Model predictions of inferior and middle turbinates. (A–C) Segmentation masks and bounding boxes drawn over images of the inferior turbinate taken during nasal endoscopy (NE). (D–F) Segmentation masks and bounding boxes drawn over images of the middle turbinate taken during NE. (G–I) Segmentation masks and bounding boxes drawn over images of both the inferior and middle turbinates taken during NE. Note the model performs accurately in the presence of surgical changes (E), crusting (F), septal deviation (H), and mucus strands (C and I).

The design of YOLOv8 allows it to perform these tasks efficiently and simultaneously, permitting its use in real time. Another feature of the YOLOv8 model we employed is its ability to use the background of labeled images as true negatives, allowing the model to make classifications without providing unlabeled images. Transfer learning was applied to YOLOv8 using NE images to fine-tune its performance. Stochastic gradient descent was selected as the learning algorithm, weights were adjusted through backpropagation, and binary cross-entropy was selected for the loss function. Training was set to halt when performance on the validation set stalled for 15 consecutive epochs and fivefold cross-validation was employed to prevent overfitting. Hyperparameters were manually selected. The initial learning rate, momentum, and decay rate were set to 1e-2, 0.937, and 5e-4, respectively. Training utilized PyTorch 2.2.1 on a computer equipped with an NVIDIA RTX 5500 GPU with GPU acceleration using CUDA 11.8 (Nvidia).

When YOLOv8 detects a turbinate, it localizes it with a bounding box and applies a segmentation mask. The model makes a prediction if it is at least 60% confident that its label reflects the ground truth, optimizing for the highest F1 score. Each model was evaluated on its respective validation subset and the average accuracy, precision, and recall values were recorded. A prediction was defined as accurate if the correct label was applied to an image with an intersection-over-union score of at least 0.5. The Python packages *matplotlib* and *seaborn* were utilized to construct confusion matrices and learning curves.

This study was approved by the Institutional Review Board of Ochsner Health (IRB# 2022.165).

3 | RESULTS

The model successfully identified the IT and MT with an average accuracy of 91.5%. The model's average precision and recall were 92.5% and 93.8%, respectively. At the confidence threshold of 60%, the average F1-score was 93.1%. The F1 confidence curve, confusion matrix, and learning

curves are presented for the model trained on the fifth fold (Figure 1A–C). The model demonstrated reduction in loss until epoch 101, at which point the stopping mechanism terminated the learning process (Figure 1C). This model's performance is representative of the other models trained in all five folds.

Figure 2 depicts segmentation predictions on images containing both ITs and MTs. These images demonstrate that the model can distinguish the outline of the turbinate from shadowing, overlying pathology, and adjacent mucosa.

4 | DISCUSSION

We have demonstrated successful application of YOLOv8 to classify, detect, and segment anatomic landmarks encountered during NE. Our model had an overall accuracy of 91.5% and an F1-score of 93.1%, indicating excellent segmentation performance. Our dataset contains images of labeled turbinates taken from multiple angles with and without comorbid pathology, which contributes to our model's ability to identify these landmarks reliably as seen in Figure 2.

Previous work has investigated a role for CNNs in identifying polyps, inverted papillomas, and nasopharyngeal carcinomas on NE[6–8] and the role of CNNs in radiographic imaging to identify sinonasal pathology.[10] However, this is the first application of CNNs toward identifying and localizing anatomical landmarks on NE images. This study may be broadly beneficial for trainees and nonspecialists with the interpretation of NE, especially in patients with sinonasal disease.

This study has several limitations. The dataset, composed exclusively of images captured at a single center, might predispose the model to a selection bias. To enhance representativeness, future work may include an expanded training dataset developed by multiple providers. The endoscope camera's limited resolution may have further reduced the model's precision. This could be addressed by either upscaling the resolution of the existing dataset using supersampling or by utilizing higher quality video endoscopy equipment.

ACKNOWLEDGEMENTS

This work was supported by Eye, Ear, Nose, and Throat Foundation, Grant Number: EE220702/EE220802.

Open access publishing facilitated by The University of Queensland, as part of the Wiley - The University of Queensland agreement via the Council of Australian University Librarians.

CONFLICT OF INTEREST STATEMENT

Edward McCoul is a consultant for 3D Matrix, Advanced Rx, Optinose, Sanofi, Stryker, and Zsquare. The other authors have conflicts of interest.

ORCID

Vinayak Ganeshan MD https://orcid.org/0009-0007-6573-6106

Blair M. Barton MD https://orcid.org/0000-0001-8931-8581

Edward D. McCoul MD, MPH https://orcid.org/0000-0003-1812-2105

REFERENCES

1. Gupta AK, Bansal S, Sahini D. Anatomy and its variations for endoscopic sinus surgery. *Clin Rhinol An Int J*. 2012;5(2):55-62.
2. Brunner JP, Jawad BA, McCoul ED. Polypoid change of the middle turbinate and paranasal sinus polyposis are distinct entities. *Otolaryngol–Head Neck Surg*. 2017;157(3):519-523. doi:10.1177/0194599817711887
3. Raithatha R, Anand VK, Mace JC, et al. Interrater agreement of nasal endoscopy for chronic rhinosinusitis. *Int Forum Allergy Rhinol*. 2012;2(2):144-150. doi:10.1002/alr.21009
4. McCoul ED, Smith TL, Mace JC, et al. Interrater agreement of nasal endoscopy in patients with a prior history of endoscopic sinus surgery. *Int Forum Allergy Rhinol*. 2012;2(6):453-459. doi:10.1002/alr.21058
5. Rajkomar A, Dean J, Kohane I. Machine learning in medicine. *N Engl J Med*. 2019;380(14):1347-1358. doi:10.1056/NEJMra1814259
6. 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
7. Xu J, Wang J, Bian X, et al. Deep learning for nasopharyngeal carcinoma identification using both white light and narrow-band imaging endoscopy. *Laryngoscope*. 2022;132(5):999-1007. doi:10.1002/lary.29894
8. He Z, Zhang K, Zhao N, et al. Deep learning for real-time detection of nasopharyngeal carcinoma during nasopharyngeal endoscopy. *iScience*. 2023;26(10):107463. doi:10.1016/j.isci.2023.107463
9. Jocher G, Ayush C, Qiu J, Ultralytics YOLO. 2023. Accessed March 8, 2023. https://ultralytics.com
10. Liu GS, Yang A, Kim D, et al. Deep learning classification of inverted papilloma malignant transformation using 3D convolutional neural networks and magnetic resonance imaging. *Int Forum Allergy Rhinol*. 2022;12(8):1025-1033. doi:10.1002/alr.22958

How to cite this article: 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;14:1521–1524. https://doi.org/10.1002/alr.23384