THE ROLE OF ARTIFICIAL INTELLIGENCE IN VOCAL CORD ULTRASOUND INTERPRETATION

William Sebelik-Lassiter[1]; Lauren Ottenstein, MS, CCC-SLP[2,3]; Merry Sebelik, MD, FACS, MAMSE[2,3]
[1]Milwaukee School of Engineering, [2]Emory University, [3]Winship Cancer Institute

INTRODUCTION

Vocal cord paralysis (VCP) occurs as the result of anatomic or neurologic injury. Thyroid and parathyroid surgery is the most common mechanism. Good clinical practice supports evaluation of vocal cord function before and after thyroid surgery.[1]
The widely accepted examination method is fiberoptic laryngoscopy, which requires expensive equipment, training, lengthy cleaning, as well as moderate patient discomfort.
Vocal cord ultrasound (VCUS) has been suggested as an alternative to laryngoscopy but is operator-dependent and may be difficult to interpret in some patients.[2,3] Previous research has shown that human user-to-user variability in ability to acquire and interpret the ultrasound images is quite high.[4] In this research, we created an artificial intelligence (AI) model that can differentiate vocal cord motion from non-motion, by computer vision analysis of ultrasound images.[5] **FIGURE 1**

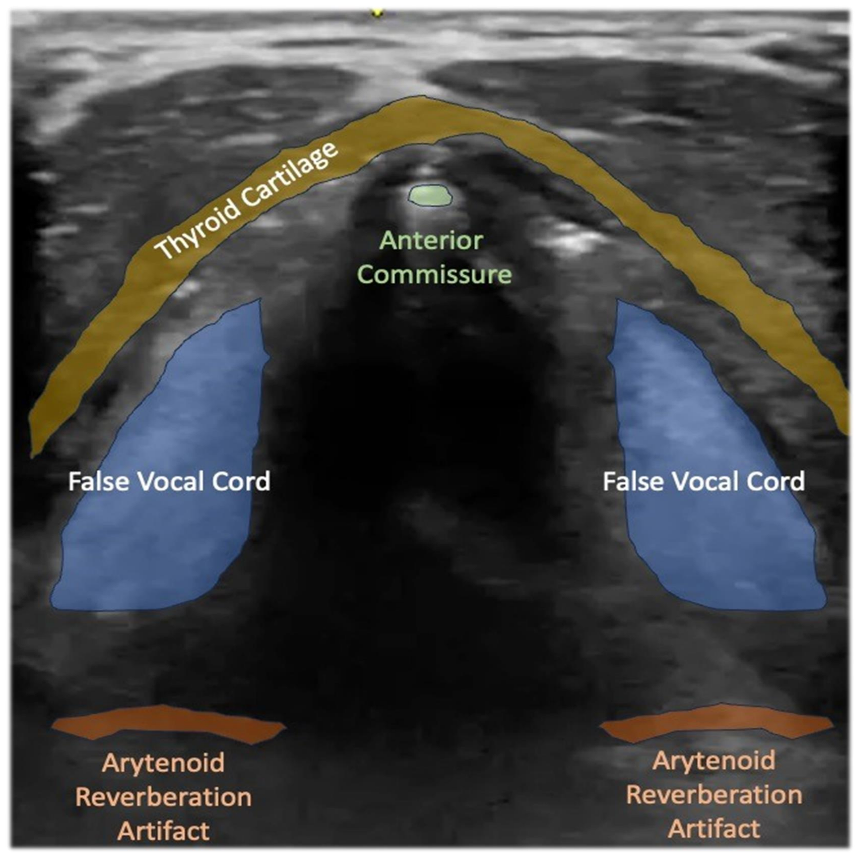

FIGURE 1. Annotated ultrasound image of vocal cords, transverse view.

CONCLUSIONS

- Vocal cord ultrasound (VCUS) is a non-invasive, cost-effective, and accurate alternative to flexible laryngoscopy for diagnosing vocal cord paralysis (VCP).
- Machine learning has the potential to reduce operator-dependence and provide consistent, accurate analysis of ultrasound imaging.
- High accuracy can be achieved with a **YOLOv8** model for vocal cord detection as well as a novel CNN (**VIPRnet**) for VCP classification.
- Synthetic Generation of VCP frames alongside augmentation can expand our training set significantly and improve model robustness despite not fully replicating complex pathologies.
- This project serves as a promising proof of concept for machine learning-assisted analysis of Vocal Cord Ultrasound (VCUS).
- Our techniques lay the foundation for screening tools which could be deployed in rural/low-resource healthcare settings,[6] telehealth and remote diagnostics, as well as being a more cost-effective and less invasive method of perioperative vocal cord motion documentation.

FIGURE 2. QSS: VCUS quality scoring scale
Normal=able to see at least one of the following: arytenoids same level, contraction of false cords equally, symmetric abduction/adduction, false or true cords equal in length
Abnormal=arytenoids not level, obvious false cord contraction on one side only, no clearly visible symmetric abduction/adduction, unequal length of vocal cords

RESULTS (FIGURES 3-6)

- Both **YOLOv8** and **VIPRnet** achieved effective convergence despite differences in training durations and batch sizes.
- Both models showed high validation accuracy, suggesting promise in differentiating healthy VCUS images from synthetic paralysis images.
- **VIPRnet** performed better than pretrained **YOLO** on the classification task.

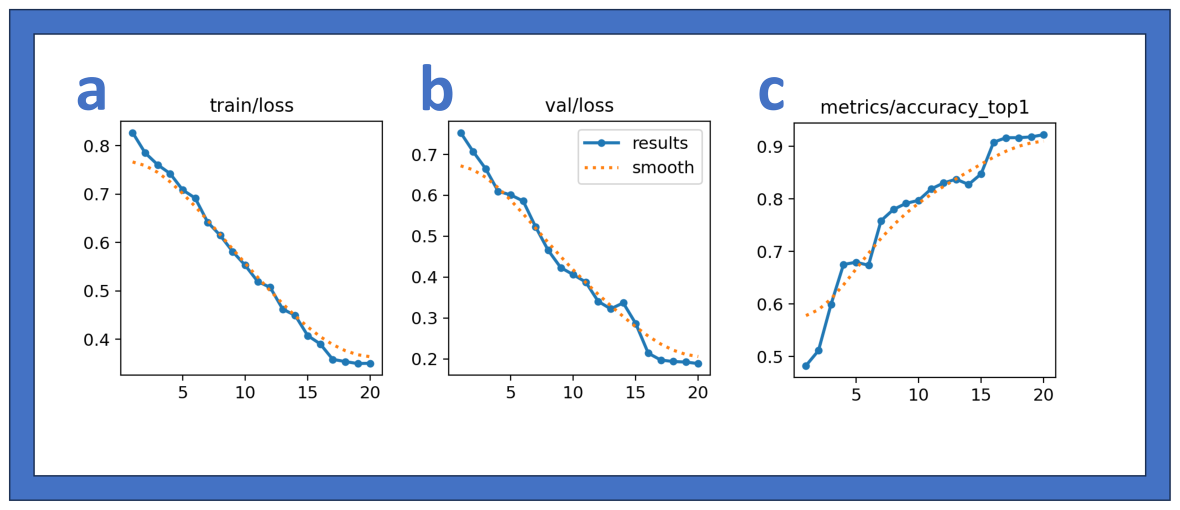

FIGURE 3: YOLO Training Results
YOLOv8n-cls converged reliably over **20** epochs, with training and validation loss steadily decreasing. Indicating successful generalization (a, b). Top-1 accuracy improved from **0.48** to **0.92**, showing strong learning of class-discriminative features (c).

YOLO

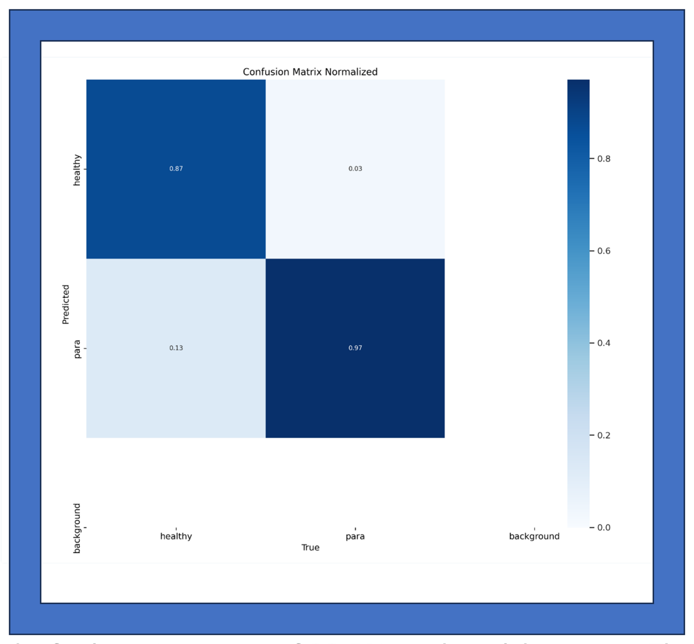

FIGURE 4: YOLO Accuracy
The classification model achieved a final top-1 accuracy of **92.3**% on the validation set. The model correctly identified **87**% of healthy and **97**% of paralyzed cases. Most misclassifications were healthy cords classified as paralyzed.

VIPRnet

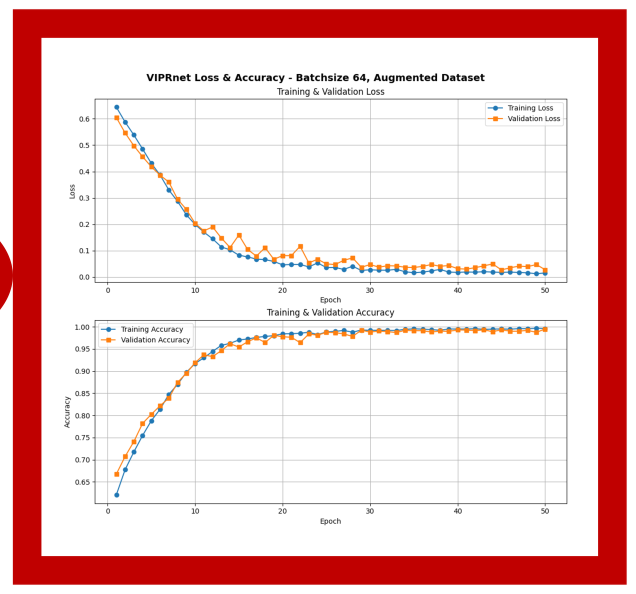

FIGURE 5: VIPRnet Training Results
Trained over **50** epochs with a batch size of **64**. Gradual reduction in both training and validation loss. Minimal overfitting, with a small gap between training and validation metrics. Final epoch: ~**99.9**% validation accuracy, showing strong performance with synthetic data.

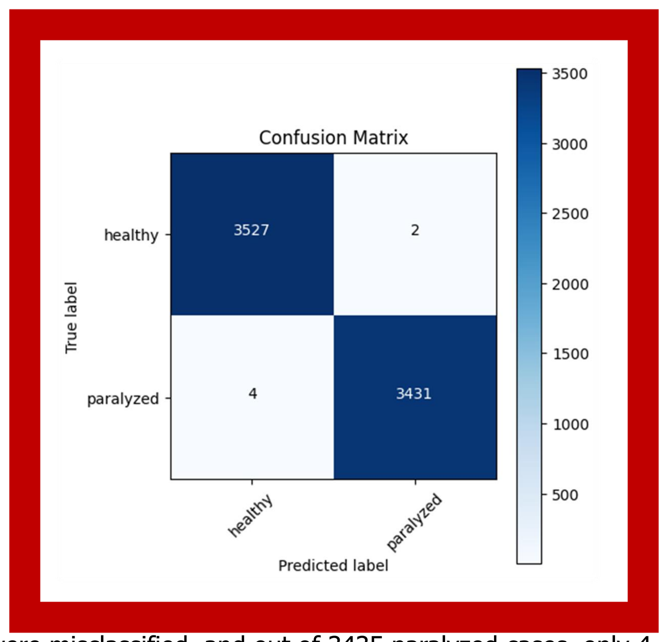

FIGURE 6: VIPRnet Accuracy
Out of 3529 healthy cases, only 2 were misclassified, and out of 3435 paralyzed cases, only 4 were misclassified. The model shows excellent separation between healthy and paralyzed cords.

FUTURE DIRECTIONS

- **Source** more true paralysis data
- **Apply greater hyperparameter tuning** and test alternate classification architectures to boost model performance
- **Investigate** additional preprocessing strategies to further enhance diagnostic accuracy
- **Utilize our YOLO/VIPRnet detection models** to automatically identify and label a larger number of frames, increasing the volume and diversity of training images for future datasets.
- **Incorporate alternative ML based** VCP classification techniques, as described by other research groups in literature
- **Deploy** our process onto an ultrasound device to streamline the integration of automatic vocal cord ultrasound analysis into clinical workflows, while also including machine-learned image recognition (image trigger) in recognition of the human learning curve for adequate image acquisition.
- **Expand** clinical training opportunities in image acquisition to all related specialties (surgeons, endocrinologists, sonographers, speech pathologists).[7]

METHODS AND MATERIALS

De-identified VCUS images were obtained from healthy volunteers with clinically normal voice function. VCUS images from larynges known to have unilateral VCP were obtained from de-identified teaching files. Images were only used for CNN training if judged to be interpretable by a VCUS quality scoring scale (QSS), developed by the authors.[4] **FIGURE 2** The images were segmented and cropped using a fine-tuned YOLOv8 model to reduce input noise.
From the dataset, two different convolutional neural network (CNN) architectures were tested to classify VCUS images as healthy or paralyzed.

The first model we evaluated, **YOLOv8n-cls** (Ultralytics), was pre-trained on large-scale image datasets and then fine-tuned on our augmented VCUS dataset. The second was a simple custom model architecture, trained from scratch, which we dubbed **VIPRnet**. Each model was evaluated for training results and accuracy.

CONTACT

Merry Sebelik, MD, FACS, MAMSE
Emory University
Division of Endocrine Head & Neck Surgery
msebeli@emory.edu
404-778-3381

William Sebelik-Lassiter
Milwaukee School of Engineering
Department of Biomedical Engineering
lassiterw@msoe.edu
251-402-6242

M Sebelik Contact

Need Contact QR for other authors

Need QR link to abstract

REFERENCES

1. Sinclair CF, Bumpous JM, Haugen BR, Chala A, Meltzer D, Miller BS, Tolley NS, Shin JJ, Woodson G, Randolph GW. Laryngeal examination in thyroid and parathyroid surgery: An American Head and Neck Society consensus statement: AHNS Consensus Statement. Head Neck. 2016 Jun;38(6):811-9.
2. Kandil E, Deniwar A, Noureldine SI, Hammad AY, Mohamed H, Al-Qurayshi Z, Tufano RP. Assessment of Vocal Fold Function Using Transcutaneous Laryngeal Ultrasonography and Flexible Laryngoscopy. JAMA Otolaryngol Head Neck Surg. 2016 Jan;142(1):74-8.
3. Noel JE, Orloff LA, Sung K. Laryngeal Evaluation during the COVID-19 Pandemic: Transcervical Laryngeal Ultrasonography. Otolaryngol Head Neck Surg. 2020 Jul;163(1):51-53.
4. Deshpande A, Tey CS, Chanani N, Landry A, Raymond M, Sebelik M, Shashidharan S, Wolf M, Raol N. The utility of handheld ultrasound as a point-of-care screening tool to assess vocal fold impairment following congenital heart surgery. Int J Pediatr Otorhinolaryngol. 2021 Sep;148:110825.
5. Sebelik-Lassiter W, Schubert E, Alliyu M, Robbins Q, Olatunji E, Barry M. Machine Learning-Assisted Vocal Cord Ultrasound Examination: Project VIPR. Presented at Midwest Instructional Computer Symposium April 2025 (First Place Winning Project).
6. Sebelik M, Robbins A, Iyiewuare H, Dhingra J. Vocal Cord Ultrasound Before and After Thyroid Surgery in a Resource-Constrained Setting. Podium Presentation at the American Institute of Ultrasound in Medicine Annual Meeting, April 2025.
7. Sebelik-Lassiter W, Ottenstein L, Sebelik M. Vocal Cord Ultrasound for the Speech and Language Pathologist. Technical Training Course submitted to American Speech and Hearing Association Annual Meeting, November 2025.