DOI: doi.org/10.47363/JAICC/ICAIC2025/2025(4)8

International Conference on
Artificial Intelligence and Cybersecurity (ICAIC 2025)

Conference Proceedings November 27-28, 2025 - Japan

Mask R-CNN Based Multiclass Segmentation Model for Endotracheal Intubation Using Video Laryngoscope

Kim Dae Kon

Department of Emergency Medicine, Hanil General Hospital, Seoul, Republic of Korea

Abstract
Endotracheal intubation (ETI) is critical to secure the airway in emergent situations. Although artificial intelligence algorithms are frequently used to analyze medical images, their application to evaluating intraoral structures based on images captured during emergent ETI remains limited. The aim of this study is to develop an artificial intelligence model for segmenting structures in the oral cavity using video laryngoscope (VL) images.

Methods
From 54 VL videos, clinicians manually labeled images that include motion blur, foggy vision, blood, mucus, and vomitus. Anatomical structures of interest included the tongue, epiglottis, vocal cord, and corniculate cartilage. EfficientNet-B5 with DeepLabv3+, EffecientNet-B5 with U-Net, and Configured Mask R-Convolution Neural Network (CNN) were used; EffecientNet-B5 was pretrained on ImageNet. Dice similarity coefficient (DSC) was used to measure the segmentation performance of the model. Accuracy, recall, specificity, and F1 score were used to evaluate the model's performance in targeting the structure from the value of the intersection over union between the ground truth and prediction mask.

Results
The DSC of tongue, epiglottis, vocal cord, and corniculate cartilage obtained from the EfficientNet-B5 with DeepLabv3+, EfficientNet-B5 with U-Net, and Configured Mask R-CNN model were 0.3351/0.7675/0.766/0.6539, 0.0/0.7581/0.7395/0.6906, and 0.1167/0.7677/0.7207/0.57, respectively. Furthermore, the processing speeds (frames per second) of the three models stood at 3, 24, and 32, respectively.

Conclusions
We developed and validated an AI algorithm to segment intraoral structures in images obtained from VL during emergent ETI. This algorithm demonstrated a high performance. The algorithm developed in this study can assist medical providers performing ETI in emergent situations.

Keywords: Biomedical image processing; Intubation; Deep Learning; Convolutional Neural Networks; Image Segmentation