Proceedings of the 2023 Design of Medical Devices Conference
DMD2023
April 17-19, 2023, Minneapolis, MN, USA

DMD2023-1691

AI BASED GLAND DETECTION IN BARRETT'S ESOPHAGUS USING OPTICAL COHERENCE TOMOGRAPHY FOR CAPSULE ENDOSCOPY DEVICE

Jieun Lee
Department of Medicine, GIH Division, Mayo Clinic, Rochester, MN, USA

Vaishnavi K. Modi
Department of Medicine, GIH Division, Mayo Clinic, Rochester, MN, USA

Renisha Redij
Department of Medicine, GIH Division, Mayo Clinic, Rochester, MN, USA

Srikanth Gadam
Department of Radiology, Mayo Clinic, Rochester, MN, USA

Keerthy Gopalakrishnan
Department of Medicine, Mayo Clinic, Rochester, MN, USA

Anjali Rajagopal
Department of Medicine, Mayo Clinic, Rochester, MN, USA

Cadman L. Leggett
Department of Medicine, GIH Division, Mayo Clinic, Rochester, MN, USA

Shivaram P. Arunachalam
Department of Medicine, GIH Division, Mayo Clinic, Rochester, MN, USA

ABSTRACT

Optical coherence tomography (OCT) is an advanced imaging modality to detect Barrett's esophagus *(BE) dysplasia, providing widefield, cross-sectional imaging and microscopic resolution. BE dysplasia is characterized under OCT by the presence and number of glandular structures with atypical morphology. Accurate detection and interpretation of BE glands under OCT is essential to detect dysplastic lesions. Object Detection using deep learning has the potential to identify glands from OCT images. We developed a YOLO model to identify the presence of glands in BE tissue. The YOLOv4 object detector was trained on a custom BE dataset of 30 patients with confirmed BE who underwent OCT imaging, of which 222 OCT images included at least one gland. Our model identified glands with a high average precision of 88.79% on the test dataset. We showed that the developed model is robust to rotation, brightness, and blur in images. We have implemented an object detection model to identify glands from OCT images with promising results accurately. This model has the potential to improve the diagnosis and surveillance of BE by eliminating human error and missed dysplastic lesions adaptable for capsule endoscopy applications.*

Keywords: Barrett's esophagus, Optical coherence tomography, identification of glands, Object detection, You Only Look Once (YOLO), capsule endoscopy

1. INTRODUCTION

Barrett's esophagus (BE) is a pre-malignant condition that arises as a consequence of chronic gastroesophageal reflux leading to the replacement of normal stratified squamous epithelium (SE) of the distal esophagus by columnar mucosa containing goblet cells [1]. Esophagogastroduodenoscopy (EGD) using a systematic biopsy protocol is the gold-standard for screening and surveillance of BE. The risk of progression of BE is estimated by the highest degree of dysplasia identified on endoscopic biopsies. Mucosal abnormalities associated with BE dysplasia can be subtle and sometimes missed under routine surveillance due to sampling error. Advanced imaging modalities such as optical coherence tomography (OCT) can be used to enhance the detection of dysplasia during BE surveillance. Endoscopic OCT systems use principles of interferometry to generate cross-sectional images of the BE epithelium at microscopic level resolution. Barrett's epithelium contains glands that are readily identified under OCT as hyporeflective structures. Features of BE dysplasia under OCT includes the presence and number of glands with atypical morphology. Capsule endoscopy for BE detection is gaining attention which offers a new platform to integrate endoscopic OCT technology to enhance diagnosis [2].

With the advancement of image processing and machine learning techniques, the ability to computer-aided analysis in BE for the detection of esophageal disease using OCT is now within

reach. Recently, multiple studies have been done to identify specific OCT image features within the esophagus, including tissue surface hyperreflectivity, layering, and hyporeflective structures using machine learning and deep learning techniques [3-5], and demonstrated high diagnostic accuracy. However, accurate detection and interpretation of BE glands under OCT are yet to have cooperated much in this field. Chao Zhou et al. [6] stated that these buried glands are likely underappreciated with the current surveillance protocol because areas of neosquamous epithelium do not routinely undergo biopsy in clinical practice. The authors showed a higher prevalence of glands and a significantly lower number of buried glands per patient in the pre-radiofrequency ablation group than post-one.

The purpose of this study was to develop a deep learning model to aid in identifying and characterizing BE glands under OCT, which are transferable to capsule endoscopy applications. In this paper, we apply YOLO to detect glands in Mayo owned OCT dataset, which is newly implemented. The model will predict the 2D bounding boxes around the glands and class label (include gland or not) of detected regions in OCT images.

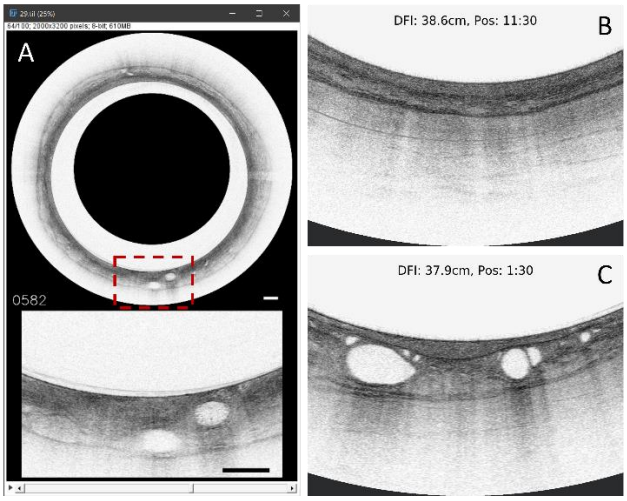

FIGURE 1: REPRESENTIVE EXAMPLES OF VLE OCT IMAGES IN THE ESOPHAGUS. **A**: A COMPLETE CIRCUMFERENTIAL CROSS-SECTIONAL OF THE ESOPHAGUS. REGIONS OF INTEREST (INSIDE THE RED SQURE) IN THE CIRCUMFERENTIAL VIEW ARE MAGNIFIED IN THE SCREEN INSETS. **B**: NORMAL SQUAMOUS EPITHELIUM. **C**: BE WITH ATYPICAL GLANDS BASED ON SIZE AND MORPHOLOGY.

2. MATERIALS AND METHODS

2.1 Volumetric Laser Endomicroscopy (VLE) images

Volumetric laser endomicroscopy (VLE) is a high-speed, second-generation OCT device. The Nvision VLE Imaging System (Nine-Point Medical, Cambridge, MA) used in this study, in conjunction with a balloon-imaging catheter, enables a rapid scan over the entire distal portion (4-6 cm) of the human esophagus [7-9].

VLE images have an axial resolution of 7 mm, a transverse resolution of about 40 mm, and can reach a depth of up to 2 to 3 mm, with microscopic resolution. VLE scans are composed of 1200 subsequential cross-sectional gray-scale images. Cross-sectional transverse and longitudinal views were simultaneously examined by a software interface of VLE scans [10].

The VLE system enables comprehensive imaging of large regions of BE, making it possible to identify abnormal features predictive of dysplasia including glandular structures [10-12].

Representative examples of OCT images from VLE scans are shown in FIGURE 1. A complete circumferential cross-sectional format of the esophagus is shown (FIGURE 1A), and the regions of interest in the circumferential view are magnified in the insets (FIGURE 1B-1C).

OCT Dataset

A multicenter VLE study prospectively enrolled 238 patients with BE at Mayo Clinic, Rochester, MN. Two physicians with expertise in BE annotated 30 patients with confirmed BE based on histopathology of which 222 OCT images were extracted and considered for analysis in this study. This study was approved by the Mayo Clinic Institutional Review Board (IRB).

2.2 Object Detection Model for Identifying Glands

In general, object detection has two sub-tasks: determining the location of an object in an image (localization) and assigning a class to that object (classification). For localization, object detection methods provide 2D bounding boxes around the objects in the image to predict the object's location. They also provide the possible class label and their confidence scores of detected regions.

YOLO is a high-speed and accurate framework [13]. YOLO provides the predicted bounding box coordinates and class probabilities (class confidence scores). For every input image, Yolo splits the image into cells using a square grid (typically a 19 x 19 grid). The model predicts bounding boxes for each grid and calculates their box confidence scores, reflecting how likely the box contains an object. To make final predictions, we keep those with high box confidence scores (greater than 0.25) and high Intersection over Union (IoU) of the ground truth and the predicted bounding box (typically greater than 0.5) as our final predictions. Each prediction box's final class confidence score is computed as the box confidence score is multiplied by the conditional class probability. Conditional class probability is the probability that the detected object belongs to a particular class.

The overall workflow of our object detection model using Yolov4 for identifying atypical glands is as follows. Firstly, we extract cross-sectional OCT images and their magnified insets from multiple longitudinal views. Images including glands are annotated with their locations and class labels in inset images when there are irregular and/or dilated glands using LabelImg, a free, open-source tool for graphically labeling images. With the pre-trained model on the COCO database [14], we set config and train our custom YOLOv4 object detector for glands using Mayo owned OCT dataset. Among multiple the trained weights for our detector, which are generated every few iterations, we reload the

best weight with the highest accuracy (mean Average Precision, mAP) on the validation set and finally make inferences on test images. We chose YOLOv4 due to its outstanding performance with high speed.

2.3 Performance Metrics in Object Detection

The performance of the object detection model was evaluated using accuracy metrics: Precision, Recall, Precision-Recall Curve, and mean Average Precision (mAP).

Precision and recall are described in Equations (1) and (2). Like binary classification problems, true positive (TP), false positive (FP), false negative (FN), and true negative (TN) are considered.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

For each prediction, the precision would be calculated using accumulated TP and accumulated FP. The recall is calculated using accumulated TP as the numerator and the number of ground truth objects as the denominator. The plotting of the precision-recall values (i.e., precision-recall curve) summarizes the trade-off in errors between the true positive rate (TPR) and the positive predictive value (PPV) for a predictive model using different probability thresholds.

AP describes the area under the precision-recall curve. We can compute the AP for each class and average them (i.e., the mean average precision (mAP)). The formulas of AP and mAP are shown in Equations (3) and (4), respectively.

$$AP = \int_0^1 Precision(Recall)\, dRecall \quad (3)$$

$$mAP = \frac{\sum_{i=1}^{n} AP_i}{n} \quad (4)$$

To smooth out the zig-zag pattern before calculating the AP, interpolation methods such as the 11-point interpolation were applied.

3. RESULTS AND DISCUSSION

To implement the object detection model for identifying glands in OCT images, we have used the following datasets and settings. A total of 222 OCT images were considered. This dataset has only one Class having glands or not. We split the data set into 143 (65%) for training, 34 (15%) for validation, and 45 (20%) for testing the model. Each image was resized to 416 x 416 pixels. The training dataset was augmented by using rotation, brightness, and blur three times.

We implemented the YOLOv4 Darknet object detector trained on custom Barrett's dataset from the pre-trained model. Our model with the best-trained weight achieving the highest mAP on the validation set subsequently made inferences on a total of 165 glands through 45 images. The batch size and subdivision of 64 and 64 were used for deep learning. The model was trained for 2000 training steps.

A proposed bounding box is considered TP if IoU ≥ 0.5 with the ground truth, and its box confidence score of at least 25% were needed. Under this study for finding glands, there is no difference between AP and mAP.

3.1 Performance of Object Detection model for Identifying Glands

Our detection model showed high precision and recall scores for test sets. The location of glands was appropriately recognized, along with the correct labels.

For the test set, a total of 184 predictions through 45 images were detected with high performance (0.8587, 0.9576) precision and recall, respectively. In FIGURE 2, the precision-recall curve with a high AP of 88.79% is shown. TP was 158, FP was 26, and 7 glands were missed (FN).

FIGURE 3-4 shows representative examples of object detection predictions. In FIGURE 3, all the glands are predicted precisely. All the bottom figures (green bounding boxes) indicate the true location of the glands. All the top figures (magenta boxes) indicate the predicted glands. Both bulky and small-sized glands are accurately identified. All glands from images with rotation, brightness, and blur are detected with precision.

In FIGURE 4A and 4D, the model predicted a single large gland in the center as two separate glands. The model does not detect small glands due to inadequate intensity and unclear shape, as shown in FIGURE 4B and 4E. Partial glands present in the boundary of the image are not considered ground truth, but our model identifies some partial glands, as shown in FIGURE 4C and 4F.

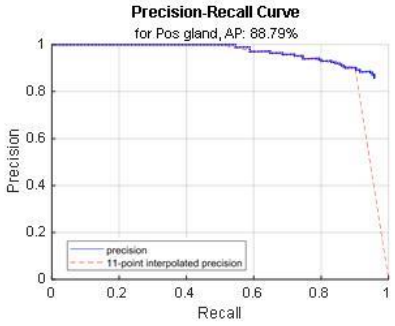

FIGURE 2: PRECISION-RECALL CURVE OF THE OBJECT DETECTION MODEL FOR IDENTIFYING GLANDS ON THE TEST SET. TO SMOOTH OUT THE ZIG-ZAG PATTERNS OF PRECISION-RECALL CURVE (BLUE LINE), THE 11-POINT INTERPOLATION OF THE SMOOTHED PRECISION-RECALL CURVE (RED DASHED LINE) IS APPLIED BEFORE CALCULATING THE AVERAGE PRECISION (AP). OUR MODEL HAS A HIGH AP OF 88.79%. AP DESCRIBES THE AREA UNDER THE PRECISION-RECALL CURVE.

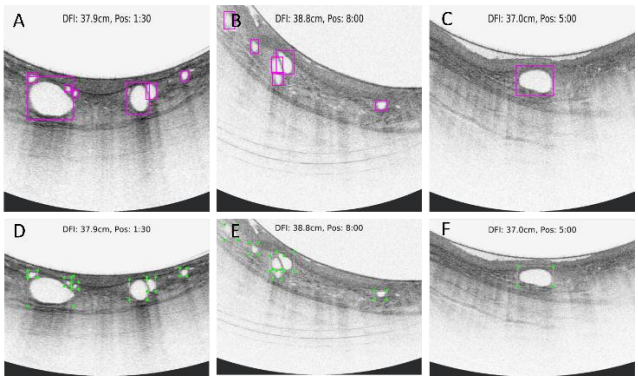

FIGURE 3: REPRESENTATIVE EXAMPLES OF THE PREDICTIONS PROVIDED FROM OUR DETECTION MODEL. **A-C**, TOP IMAGES WITH MEGENTA BOUNDING BOXES ARE THE PREDICTION RESULTS. **D-F**, BOTTOM IMAGES WITH GREEN BOUNDING BOXES ARE THE GROUND TRUTHS. ALL THREE CASES HAVE PERFECT ACCURACY.

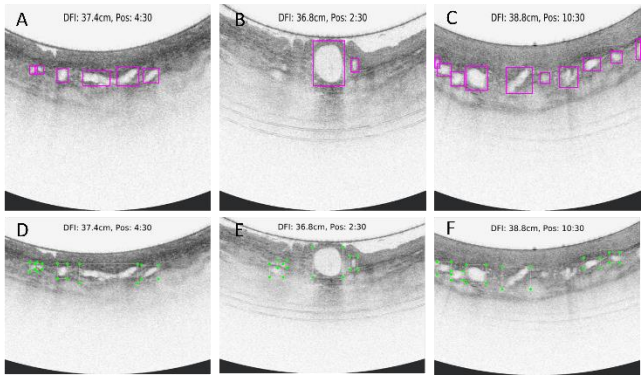

FIGURE 4: REPRESENTATIVE EXAMPLES OF PREDICTIONS PROVIDED FROM OUR DETECTION MODEL. **A-C**, TOP IMAGES WITH MAGENTA BOUNDING BOXES ARE THE PREDICTION RESULTS. **D-F**, BOTTOM IMAGES WITH GREEN BOUNDING BOXES ARE THE GROUND TRHTUS. **A AND D (LEFT)**, A LONG FLAT GLAND IN THE MIDDLE OF THE IMAGE IS PREDICTED AS TWO SEPARATE OBJECTS. **B AND E (MIDDLE)**, THERE ARE 2 TP AND 2 MISSED GLAND (FN). **C AND F (RIGHT)**, PARTIAL GLANDS IN BOTH LEFT AND RIGHT BOUNDARIES OF IMAGES ARE CONSIDERED AS GLANDS IN PREDICTIONS, BUT NOT ANNOTATED IN GROUND TRUTH (FP CASES). *TP*, TRUE POSITIVE; *FP*, FALSE POSITIVE; *FN*, FALSE NEGATIVE; *TN*, TRUE NEGATIVE.

3.2 Discussion

We have implemented an object detection model to identify glands from OCT images accurately. The key findings are as follows: 1) the model has a high AP of 88.79% (no difference between AP and mAP in this study), indicating precise identification of glands from the test dataset. 2) the developed model is robust to images' rotation, brightness, and blur. 3) Partial glands (too small size or located at the edge of images) that are not counted as actual glands by humans are well detected in some cases.

Our model performs robustly in cases of rotation, brightness, and blur. Also, it works well to identify partial/split glands. In some cases, it misclassified glands in dim images and glands with highly irregular shapes. In the future, additional information such as texture (intensity, patterns), shape and size of the gland, and accurate adjudication of partial glands will be used to optimize the object detection model.

We design a model in an automated, non-laborious process. We used the minimum number of images, and then we will use it in the remaining patients so that this model can annotate the class label of images. Using a pre-trained model can compensate for the quality of the object detector when using a small number of images. Saleh Shahinfar, et al. [15] suggested that the minimum number of image data for training is around 150-500 per class. It is still debatable, but our results with the closely sufficient number of images (143 images) showed precise identification for glands.

Due to subtle imaging features, the interpretation of OCT images is complex and time-consuming. To address this, deep learning algorithm is promising with accuracy matching human assessment when detecting the BE gland in endoscopic OCT images. This automated gland detection method helps clinicians to locate and recognize the risk of BE in endoscopic examinations.

In the future, real-time gland identification and analysis models can be implemented with capsule endoscopy application. More features, such as object size and shape, must be included to improve the accuracy of detections during BE surveillance. Also, our model must be tested on an independent patient cohort from different institutions.

4. CONCLUSION

An accurate object detector for identifying irregular glands can assist in the endoscopic resection or ablation of lesions and aid in the early detection of dysplasia. We have implemented a deep learning object detection model to identify glands from OCT images with promising results accurately. This model has the potential to improve the diagnosis and surveillance of BE by eliminating human error and missed dysplastic lesions, as well as with potential integration with capsule endoscopic devices.

ACKNOWLEDGEMENTS

This work was supported by the GIH Division for the GIH Artificial Intelligence Laboratory (GAIL), Department of Medicine, Mayo Clinic, Rochester, MN USA.

REFERENCES

[1] Spechler, Stuart Jon, and Raj K. Goyal. "Barrett's esophagus." *New England Journal of Medicine* 315, no. 6 (1986): 362-371.

[2] Beg, Sabina, Tim Card, Samantha Warburton, Imdadur Rahman, Emilie Wilkes, Jonathan White, and Krish Ragunath. "Diagnosis of Barrett's esophagus and esophageal varices using a magnetically assisted capsule endoscopy system."

Gastrointestinal Endoscopy 91, no. 4 (2020): 773-781.

[3] Wang, Zhao, Hsiang-Chieh Lee, Osman O. Ahsen, Kaicheng Liang, Marisa Figueiredo, Qin Huang, James G. Fujimoto, and Hiroshi Mashimo. "Computer-Aided Analysis of Gland-Like Subsurface Hyposcattering Structures in Barrett's Esophagus Using Optical Coherence Tomography." *Applied Sciences* 8, no. 12 (2018): 2420.

[4] Fonollà, Roger, Thom Scheeve, Maarten R. Struyvenberg, Wouter L. Curvers, Albert J. de Groof, Fons van der Sommen, Erik J. Schoon, Jacques JGHM Bergman, and Peter HN de With. "Ensemble of deep convolutional neural networks for classification of early Barrett's neoplasia using volumetric laser endomicroscopy." *Applied Sciences* 9, no. 11 (2019): 2183.

[5] van der Putten, Joost, Maarten Struyvenberg, Jeroen de Groof, Thom Scheeve, Wouter Curvers, Erik Schoon, Jacques JGHM Bergman, Peter HN de With, and Fons van der Sommen. "Deep principal dimension encoding for the classification of early neoplasia in Barrett's Esophagus with volumetric laser endomicroscopy." *Computerized Medical Imaging and Graphics* 80 (2020): 101701.

[6] Zhou, Chao, Tsung-Han Tsai, Hsiang-Chieh Lee, Tejas Kirtane, Marisa Figueiredo, Yuankai K. Tao, Osman O. Ahsen et al. "Characterization of buried glands before and after radiofrequency ablation by using 3-dimensional optical coherence tomography (with videos)." *Gastrointestinal endoscopy* 76, no. 1 (2012): 32-40.

[7] Yun, Seok H., Guillermo J. Tearney, Benjamin J. Vakoc, Milen Shishkov, Wang Y. Oh, Adrien E. Desjardins, Melissa J. Suter et al. "Comprehensive volumetric optical microscopy in vivo." *Nature medicine* 12, no. 12 (2006): 1429-1433.

[8] Yun, Seok-Hyun, Guillermo J. Tearney, Johannes F. de Boer, N. Iftimia, and Brett E. Bouma. "High-speed optical frequency-domain imaging." *Optics express* 11, no. 22 (2003): 2953-2963.

[9] Wolfsen, Herbert C., Prateek Sharma, Michael B. Wallace, Cadman Leggett, Guillermo Tearney, and Kenneth K. Wang. "Safety and feasibility of volumetric laser endomicroscopy in patients with Barrett's esophagus (with videos)." *Gastrointestinal endoscopy* 82, no. 4 (2015): 631-640.

[10] Leggett, Cadman L., Emmanuel C. Gorospe, Daniel K. Chan, Prasuna Muppa, Victoria Owens, Thomas C. Smyrk, Marlys Anderson, Lori S. Lutzke, Guillermo Tearney, and Kenneth K. Wang. "Comparative diagnostic performance of volumetric laser endomicroscopy and confocal laser endomicroscopy in the detection of dysplasia associated with Barrett's esophagus." *Gastrointestinal endoscopy* 83, no. 5 (2016): 880-888.

[11] Swager, Anne-Fré, Guillermo J. Tearney, Cadman L. Leggett, Martijn GH van Oijen, Sybren L. Meijer, Bas L. Weusten, Wouter L. Curvers, and Jacques JGHM Bergman. "Identification of volumetric laser endomicroscopy features predictive for early neoplasia in Barrett's esophagus using high-quality histological correlation." *Gastrointestinal endoscopy* 85, no. 5 (2017): 918-926.

[12] Struyvenberg, Maarten R., Albert J. de Groof, Roger Fonollà, Fons van der Sommen, Peter HN de With, Erik J. Schoon, Bas LAM Weusten et al. "Prospective development and validation of a volumetric laser endomicroscopy computer algorithm for detection of Barrett's neoplasia." *Gastrointestinal endoscopy* 93, no. 4 (2021): 871-879.

[13] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[14] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." *arXiv preprint arXiv:2004.10934* (2020).

[15] Shahinfar, Saleh, Paul Meek, and Greg Falzon. ""How many images do I need?" Understanding how sample size per class affects deep learning model performance metrics for balanced designs in autonomous wildlife monitoring." *Ecological Informatics* 57 (2020): 101085.