Development of a Deep Learning Model for Whole Slide Image Analysis in the Histologic Diagnosis of Dysplastic Barrett's Esophagus

Shahriar Faghani, MD, Postdoc Research Fellow, Mayo Clinic, Radiology Department

Chamil Codipilly, MD; David Vogelsang, MD; Mana Moassefi, MD; Pouria Rouzrokh, MD, MPH; Bardia Khosravi, MD, MPH; Siddharth Agarwal, MBBS; Lovekirat Dhaliwal, MBBS; David A. Katzka, MD; Catherine Hagen, MD; Jason Lewis, MD; Cadman L. Leggett, MD; Bradley J. Erickson, MD; PhD; Prasad G. Iyer, MD

Introduction

Barrett's esophagus (BE) is the only known precursor lesion of esophageal adenocarcinoma (EAC), and progression occurs via the development of dysplasia. While endoscopic eradication therapy is recommended for high grade dysplasia (HGD), controversy exists regarding optimal management of low-grade dysplasia (LGD). Poor interobserver agreement amongst expert pathologists results in a significant proportion of community-based diagnoses of LGD being downgraded by expert gastrointestinal (GI) pathologists. The presence of confirmed LGD increases the risk of malignant transformation, and the inability to reliably diagnose LGD results in unnecessary procedures. Therefore, improving LGD diagnostic capabilities is critical. We developed a deep learning model leveraging whole slide image (WSI) processing, object detection, and classification to grade BE dysplasia.

Hypothesis

An ensemble of object detection and classifier can automatically diagnose the grade of BE dysplasia on a WSI.

Methods

We obtained non-dysplastic BE (NDBE), LGD, and HGD histology slides (hematoxylin and eosin stained) from our institution's clinical pathology archive from 1992-2020. These slides were digitized (Aperio AT2 Scanner, Leica Biosystems, Buffalo Grove, IL) and reviewed by at least 2 expert

gastrointestinal pathologists. Subsequently, the digitized images were annotated with areas of NDBE, LGD, or HGD by the pathologists utilizing Aperio ImageScope (Sausalito, CA).

WSIs at 40x magnification were converted to non-overlapping tiles of 1280x1280 pixels. We then converted slide annotations into object detection compatible formats. A “You Look Only Once” (YOLO) version 5 m6 model was trained to identify regions of interest and perform first-pass prediction for degree of dysplasia (Figure 1). Next, we fed the identified crops of each tile into a previously developed ResNet101 model adept at 3 class (NDBE, LGD, & HGD) classification. The whole slide prediction was based on the highest grade of dysplasia predicted by the YOLOv5 and ResNet101 models.

Results

We included 542 slides: 164 NDBE, 226 LGD, and 152 HGD slides as determined by pathologist review. The training set consisted of 368 slides (8,596 tiles), the validation set 104 slides (1,946 tiles), and the test set 70 slides (840 tiles; Figure 1). This ensemble approach lead to high sensitivity, specificity, and F1 score (a measure of precision and recall) for the diagnosis of LGD and HGD (Table 1).

Conclusion

We report development of a deep learning model (structurally composed of two distinct models) that accurately diagnoses BE dysplasia from WSI analysis representing a significant improvement over current diagnostic capabilities. Further refinement and external validation are required prior to widespread adoption in clinical practice.

Figures

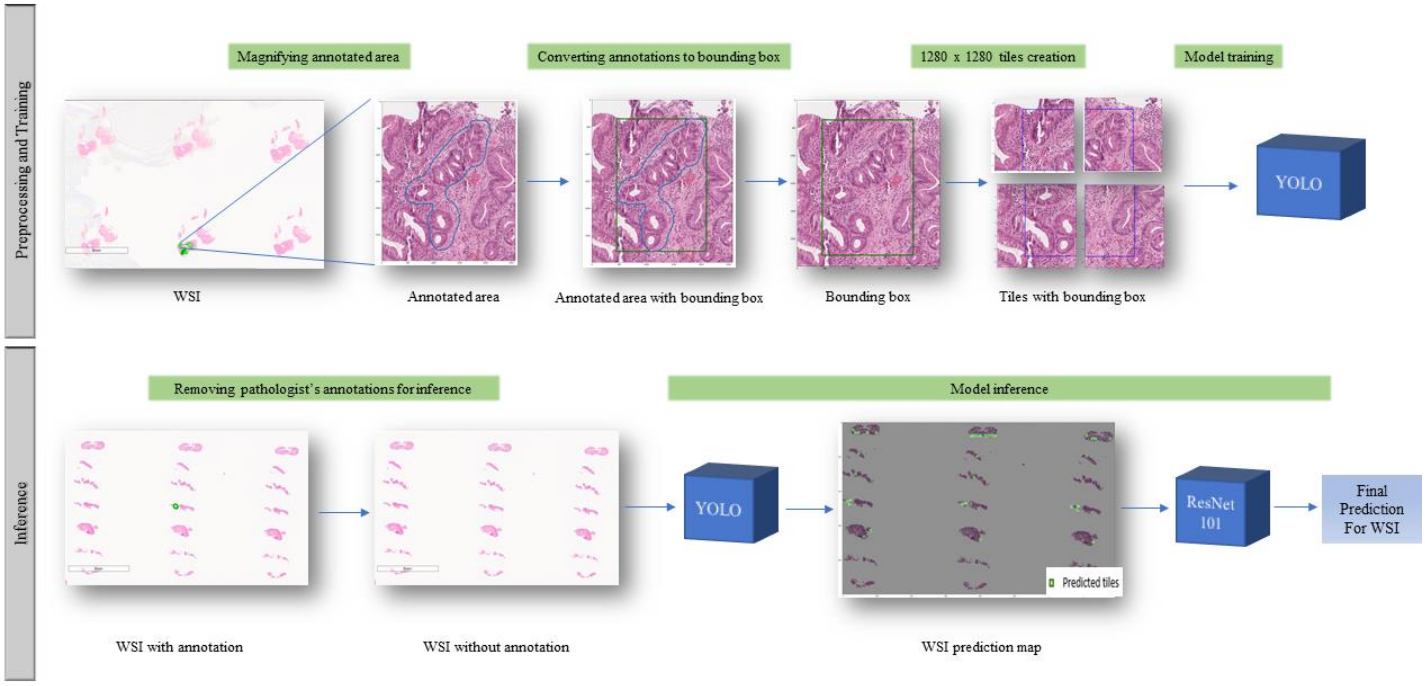

Figure 1. Illustration of Preprocessing, Training, and Inference Process.

	Sensitivity	Specificity	PPV	NPV	F1 Score
NDBE (N = 18)	94.4%	96.2%	89.5%	98.0%	0.92
LGD (N=33)	81.8.%	97.3%	96.4%	85.7%	0.89
HGD (N = 19)	94.7%	90.2%	78.3%	97.9%	0.86

Table 1. Accuracy of YOLOv5/ResNet 101 models on Whole Slide Imaging for Prediction of BE Dysplasia grade in the Test Set

Keywords

Applications; Artificial Intelligence; Imaging Research

SIIM22 Presentation

Abdomen & Neuroradiology Machine Learning Research Abstract | 1014 | In-Person