

Research Article

AI-Triaged 3D Pathology to Improve Detection of Esophageal Neoplasia While Reducing Pathologist Workloads

Lindsey A. Erion Barner^a, Gan Gao^a, Deepti M. Reddi^b, Lydia Lan^{a,c}, Wynn Burke^{b,d}, Faisal Mahmood^{e,f,g}, William M. Grady^h, Jonathan T.C. Liu^{a,b,i,*}

^a Department of Mechanical Engineering, University of Washington, Seattle, Washington; ^b Department of Laboratory Medicine & Pathology, University of Washington School of Medicine, Seattle, Washington; ^c Department of Biology, University of Washington, Seattle, Washington; ^d Department of Medicine (Gastroenterology Division), University of Washington School of Medicine, Seattle, Washington; ^e Department of Pathology, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts; ^f Cancer Program, Broad Institute of Harvard and MIT, Cambridge, Massachusetts; ^g Harvard Data Science Initiative, Harvard University, Cambridge, Massachusetts; ^h Clinical Research Division, Fred Hutchinson Cancer Research Center, Seattle, Washington; ⁱ Department of Bioengineering, University of Washington, Seattle, Washington

ARTICLE INFO

Article history:

Received 25 March 2023

Revised 25 July 2023

Accepted 25 August 2023

Available online 31 August 2023

Keywords:

3D pathology
light-sheet microscopy
deep learning
Barrett's esophagus
dysplasia
esophageal adenocarcinoma

ABSTRACT

Early detection of esophageal neoplasia via evaluation of endoscopic surveillance biopsies is the key to maximizing survival for patients with Barrett's esophagus, but it is hampered by the sampling limitations of conventional slide-based histopathology. Comprehensive evaluation of whole biopsies with three-dimensional (3D) pathology may improve early detection of malignancies, but large 3D pathology data sets are tedious for pathologists to analyze. Here, we present a deep learning-based method to automatically identify the most critical 2D image sections within 3D pathology data sets for pathologists to review. Our method first generates a 3D heatmap of neoplastic risk for each biopsy, then classifies all 2D image sections within the 3D data set in order of neoplastic risk. In a clinical validation study, we diagnose esophageal biopsies with AI-triaged 3D pathology (3 images per biopsy) vs standard slide-based histopathology (16 images per biopsy) and show that our method improves detection sensitivity while reducing pathologist workloads.

© 2023 United States & Canadian Academy of Pathology. Published by Elsevier Inc. All rights reserved.

Introduction

The incidence of esophageal adenocarcinoma (EAC) has increased more than 10-fold in Western populations over the last several decades.¹ Most EAC patients are diagnosed with late-stage or distant disease for which the 5-year survival rate is <20%.^{2–4} EAC is believed to arise from Barrett's esophagus (BE), a complication of gastrointestinal reflux disease that results in specialized intestinal metaplasia.^{5,6} BE typically progresses to low-grade dysplasia (LGD) and

then high-grade dysplasia (HGD) before finally advancing to EAC. Although the risk of progression to EAC is low for patients with BE (0.3% per year), the risk of progression to EAC increases for patients who develop LGD and HGD, for whom the likelihood increases to 2% and >6% per year, respectively.^{7–11} Therefore, patients with BE are recommended to receive periodic endoscopic screening for neoplasia (defined here as dysplasia or EAC) to enable early detection and early intervention, which is the key to maximizing patient outcomes.^{12–14} During these screening procedures, 4-quadrant random biopsies are obtained via endoscopic forceps at ~1 to 2 cm increments along the segments of the esophagus containing BE as well as of suspicious areas seen with endoscopy, as shown in Figure 1 (ie, Seattle protocol).⁴ In addition to the sampling limitations of random biopsies, the gold-standard method for evaluating these

These authors contributed equally: Lindsey A. Erion Barner and Gan Gao.

* Corresponding author.

E-mail address: jonliu@uw.edu (J.T.C. Liu).

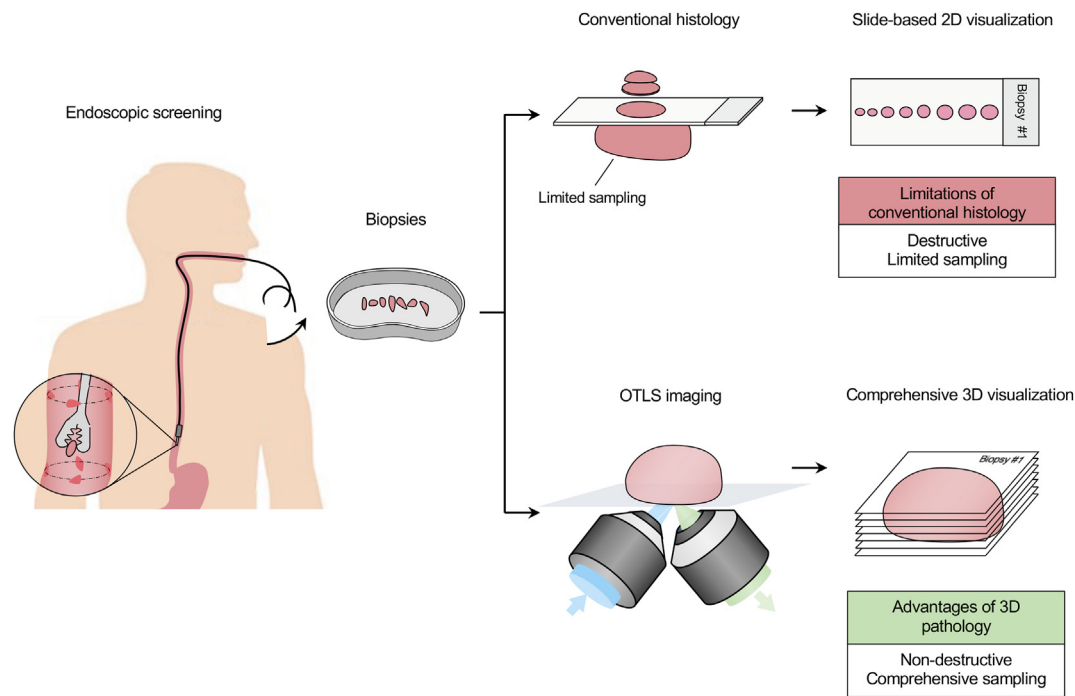


Figure 1.

Endoscopic screening with biopsies evaluated by conventional histology vs nondestructive three-dimensional (3D) pathology. During periodic endoscopic screening of Barrett's esophagus patients, 4-quadrant biopsies are obtained at ~1 to 2 cm increments along the length of the Barrett's esophagus (Seattle protocol). With conventional histology, these biopsies are thinly sectioned, mounted onto glass slides, and stained with hematoxylin and eosin (H&E) to enable microscopic evaluation by pathologists (Fig. 1). Because this process is destructive to the tissue and time-consuming, only a few tissue sections (typically 4 to 16 sections) are processed as H&E slides. The limited amount of each biopsy (~1%) that pathologists view as 2D sections may negatively impact the sensitivity for detecting neoplasia.⁸

biopsies, slide-based histology, severely undersamples the biopsied tissues.¹⁵⁻¹⁷ In conventional histology, biopsy specimens are thinly sectioned, mounted onto glass slides, and stained with hematoxylin and eosin (H&E) to enable microscopic evaluation by pathologists (Fig. 1). Because this process is destructive to the tissue and time-consuming, only a few tissue sections (typically 4 to 16 sections) are processed as H&E slides. The limited amount of each biopsy (~1%) that pathologists view as 2D sections may negatively impact the sensitivity for detecting neoplasia.⁸

We hypothesize that nondestructive three-dimensional (3D) pathology may improve the sensitivity of detecting neoplasia in endoscopic biopsies in comparison to conventional histology because a vastly greater volume of biopsy tissue may be interrogated with this method (Fig. 1). The ability to characterize diagnostically important structures in 3D can also reduce certain ambiguities inherent to 2D histology.¹⁸⁻²⁰ Here, we utilized open-top light-sheet (OTLS) microscopy as a 3D pathology platform, which enables comprehensive examination of diverse clinical specimens.²⁰ By staining intact clinical specimens with a fluorescent analog of H&E and optically clearing the tissues with a solvent-based protocol, 3D pathology data sets generated by OTLS microscopy can be false-colored to mimic the appearance of standard-of-care H&E histology images for convenient assessment by pathologists.^{18,20-24} Additionally, tissue processing for OTLS microscopy utilizes gentle reagents and reversible protocols so that tissues may be submitted for conventional histology after OTLS imaging without negative effects.^{23,25} The value of this technology has been demonstrated for several applications, such as the grading of prostate cancer,^{23,26} multiscale examination of lymph nodes for breast cancer staging,²² and examination of fresh breast tissue in intraoperative settings.^{19,25}

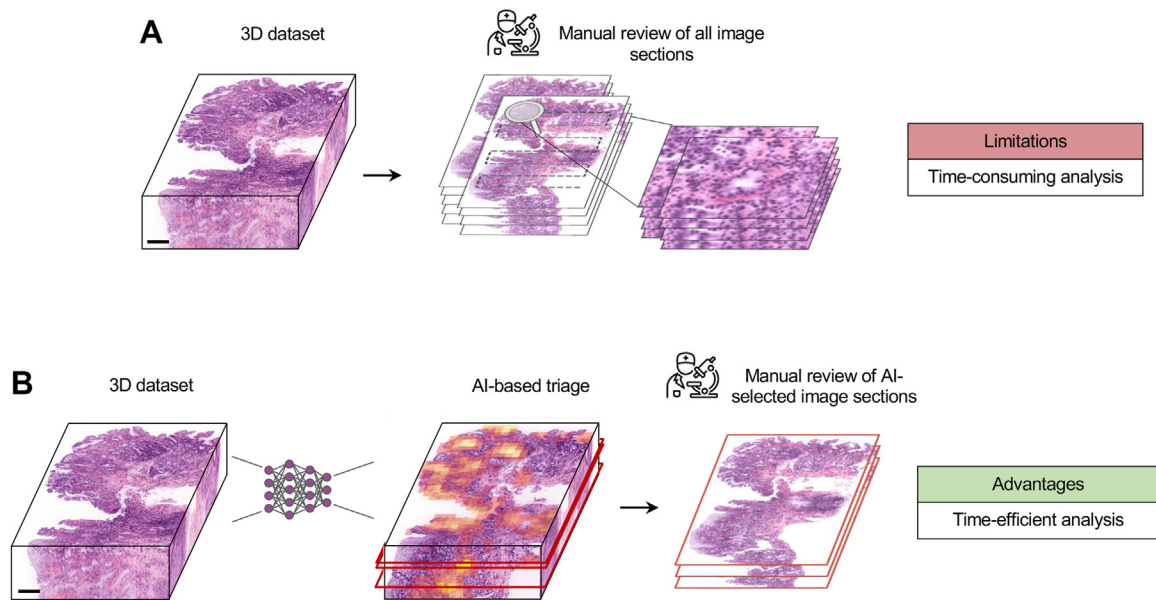
Although OTLS microscopy is a nondestructive approach that enables histologic evaluation of whole biopsies, 3D pathology data

sets can be tedious to assess by human observers (Fig. 2A). Considering that endoscopies alone generate 20 million cases annually in the US for pathologist review,²⁷ and that there is a worsening shortage of pathologists across the globe,²⁸⁻³⁰ diagnostic workflows that take advantage of 3D information without further increasing pathologist workloads would be ideal. To this end, we have developed an AI-based triage method to streamline the diagnosis of 3D pathology data sets while keeping the pathologist in the loop to minimize risk and to facilitate clinical adoption. Our method uses a deep learning algorithm to automatically identify neoplastic regions in small 2D image patches, aggregates these patch-based predictions in 3D, and then uses a random forest classifier (RFC) to select the most important 2D image sections from the 3D data set for pathologist review (Fig. 2B).^{31,32} Importantly, this AI-based triage method screens the entire volume of each biopsy for neoplasia to facilitate improved detection in comparison to conventional histology, which evaluates a modest number of thin sections from only one side of each tissue specimen. We quantify the patch-based and image section-based performance of our triage method and report a preliminary clinical validation study showing that AI-triaged 3D pathology can potentially improve the sensitivity of diagnosing neoplasia in endoscopic biopsies while reducing the workload for pathologists in comparison to standard-of-care histology.

Methods

Tissue Preparation and Open-Top Light-Sheet Imaging

Archived esophageal biopsy and endoscopic mucosal resection specimens were obtained as formalin-fixed paraffin-embedded blocks from the Gastrointestinal Center for Analytic Research and

**Figure 2.**

Motivation and advantages of AI-triaged three-dimensional (3D) pathology. (A) Nondestructive 3D pathology provides comprehensive sampling of biopsies while preserving the tissue for downstream assays/archives. However, the manual evaluation of such large 3D pathology data sets is time-consuming. (B) Alternatively, our AI-based triage method automatically identifies regions most likely to contain neoplasia within a 3D pathology data set (center) and then selects the highest-risk 2D image sections (right) for manual review by a pathologist. Scale bars represent 200 μm .

Exploratory Science at the University of Washington Medical Center following institutional review board-approved protocols (IR9066). For initial model development, 30 specimens from 11 patients (22 biopsies and 8 endoscopic mucosal resection specimens) were used for algorithm training. An additional 20 biopsy specimens from 10 other patients were set aside as an independent validation cohort. Specimens were deparaffinized by incubating them at 70 °C for 1 hour and then immersing them in Xylene at 65 °C for 48 hours. Specimens were washed twice in 100% ethanol (EtOH) for 1 hour each, incubated in 70% EtOH (30% deionized water) for 1 hour, then stained with a fluorescent analog of H&E by incubating them in 70% EtOH at pH 4 (titrated with hydrochloric acid) with a 1:1000 dilution of TO-PRO-3 and a 1:2000 dilution of Eosin-Y. Finally, the specimens were optically cleared in ethyl cinnamate ($n = 1.56$) for 4 hours.²¹

Esophageal specimens were then placed on a flat index-matched ($n = 1.56$) sample plate (Hivex, 200- μm thick). The whole volume of each biopsy was imaged (0.6-mm average thickness) with a previously reported OTLS microscope using a 20 $\times$ objective (numerical aperture = 0.40).²⁰ A digitally scanned light sheet (numerical aperture ~0.09) was used to illuminate TO-PRO-3 and Eosin-Y at wavelengths of 660 nm and 488 nm, respectively.³³ Fluorescence was collected with a band-pass filter (Semrock FF01-496/LP-25 and Semrock LP02-664RU-25) and 20 $\times$ objective (Nikon CFI60 TU Plan ELWD), then relayed to an sCMOS camera as the specimen was stage-scanned in XYZ. Images were collected with an isotropic sampling pitch of 0.21 $\mu\text{m}/\text{px}$. Data were compressed during imaging by ~10 $\times$ with B3D lossless compression as reported in a study by Balázs et al.³⁴ Examples of false-colored 3D pathology data sets may be found in [Supplementary Movies 1-2](#).

Image Preprocessing

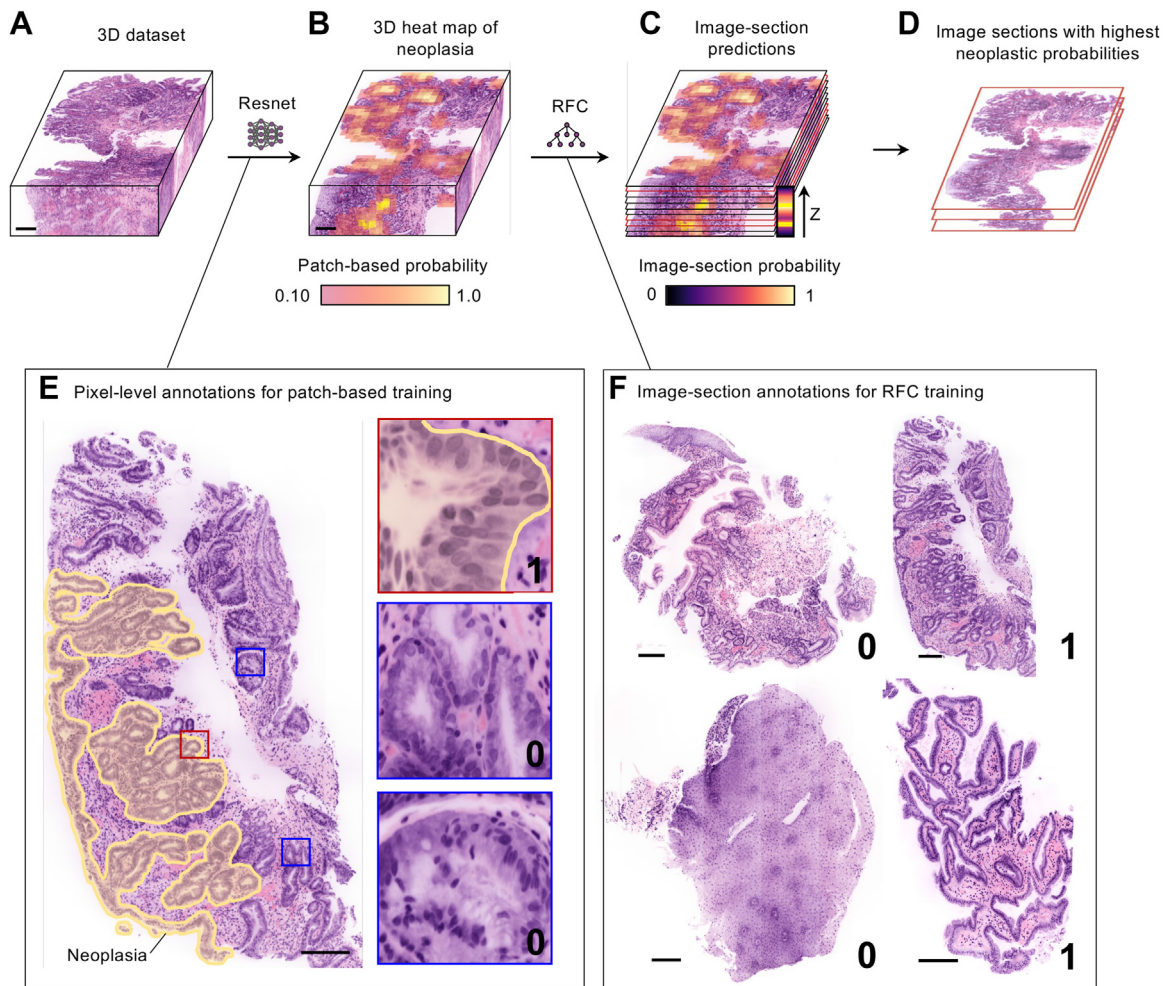
3D pathology data sets of 30 esophageal specimens from 11 patients were processed to facilitate patch-based training of the

deep learning network. One to 2 cross-sectional images (2D) were selected from the 3D pathology data sets of each specimen (43 total image sections). Prior to pathologist review and annotation, these images were false-colored to mimic an H&E-like appearance and saved in a pyramidal TIFF format. Pixel-level annotations were provided by a board-certified pathologist (DMR) to indicate regions of neoplasia (dysplasia or cancer) using the Automated Slide Analysis Platform and recorded in an XML file.

After annotations were prepared, the images were pre-processed for training the deep learning algorithm (ResNet18), which was pretrained on ImageNet.^{31,35} The raw data for both fluorescence channels (TO-PRO-3 and Eosin-Y) were normalized and saved as 2 channels within an RGB image (as per the conventional image format used for patch-based deep learning algorithms). Otsu thresholding was used to segment the tissue boundaries from the background of the image.³⁶ Overlapping patches (512 $\times$ 512 pixels or ~100 $\times$ 100 μm , with 50% overlap between adjacent patches) were extracted from the tissue-containing regions in each 2D image. Finally, each patch was assigned a label of 0 (benign) or 1 (containing neoplasia) based on the pathologist's pixel-level annotations of the images. This procedure generated approximately 393,000 patches (355,800 benign and 37,300 neoplastic) for training.

Training of Patch-Based Algorithm

Once OTLS data sets were prepared for network training, a deep learning network (ResNet18) was trained to label patches extracted from the 2D image sections as benign or neoplastic³¹ using cross-validation resampling.³⁷ Image transformations were applied to patches at random to improve model performance and reduce overfitting: image rotations, vertical and horizontal flips, brightness and contrast adjustments, and saturation adjustments. We used 15-fold cross-validation to train the model and estimate its overall performance. For each fold, all image patches from 28

**Figure 3.**

Prediction pipeline for AI-triaged three-dimensional (3D) pathology. (A) For each 3D pathology data set, a deep learning algorithm is used to identify neoplastic image patches. These predictions are volumetrically aggregated, resulting in a (B) 3D heatmap of the average predicted probability of each patch containing neoplasia. (C) A random forest classifier then predicts the probability that each 2D image section within the heatmap contains neoplasia. (D) The 3 highest-ranked 2D sections are then identified for pathologist review. (E) The training data used for the ResNet model consists of 43 image sections with pixel-level annotations for neoplasia (left). These images are split into $100 \times 100 \mu\text{m}$ patches (insets) where each patch is assigned a ground-truth label of 0 (benign) or 1 (neoplastic). (F) To train the random forest classifier, 43 2D image sections are assigned a ground-truth label of 0 if they are entirely benign, or 1 if they contain any amount of neoplasia. Scale bars represent $200 \mu\text{m}$.

specimens were used for training, and patches from the remaining 2 specimens were held out for testing. The model was trained over 20 epochs (hard stop) with a batch size of 64 for each of the 15 folds. A fully connected dropout layer was added to the model to reduce overfitting.³⁸ The model was optimized with stochastic gradient descent with a learning rate of 0.001 and momentum of 0.9. Training took approximately 9 hours for each fold on a workstation equipped with a Nvidia TITAN Xp graphics card, 128 GB of RAM, and an Intel Xeon processor (E5-1620 v4 3.5GHz 4 core).

For each 2D image section, overlapping patch-based predictions were aggregated (overlapping patch regions were averaged) to create a probability heatmap, for which the intensity value of any given patch (maximum value of 1.0) represents the predicted probability of that image patch containing neoplasia. With 50% overlap between all adjacent patches, the intensity value of each patch in the heatmap was therefore the average of 4 overlapping patch predictions (except for patches at the boundary of the tissue). Predictions were generated and aggregated over all 2D image sections in each 3D specimen (~3 hours for a 1 mm^3

biopsy), resulting in a 3D heatmap predicting the presence of neoplasia within the whole specimen (Fig. 3A, B).

Training of Image-Based Random Forest Classifier

The RFC was trained to discriminate between benign and neoplastic 2D image sections based on the probability heatmaps generated by the patch-based classifier. A set of 3 “hand-crafted” features extracted from the heatmaps served as inputs to the RFC: the maximum predicted probability of neoplasia, the number of patches for which P value was $>.10$, and the image noise (SD of the patch values).³² To train the classifier for this task, we used the heatmaps corresponding to the 43 annotated images generated during cross-validation testing of the patch-based algorithm. The ground-truth label for each image was 0 or 1 according to DMR’s annotations, in which images that contained any amount of neoplasia were assigned a label of 1, and entirely benign images were assigned a label of 0. We used 15-fold cross-validation to

train and evaluate the RFC's performance with the same train-test splits described for the patch-based classifier.

After training the RFC, we used the classifier to generate predictions on all 2D image sections within each 3D pathology data set. The output for each image section was a single value (maximum of 1.0) corresponding to the probability that the image section contained neoplasia (Fig. 3C). We sorted the sections based on the probability of containing neoplasia and then identified the top 3 image sections for manual review by a pathologist (Fig. 3D).

Results

Patch-Based and Image-Based Classification Results

We first demonstrate the performance of the patch-based deep learning classifier. We show several examples of 2D heatmaps generated by the patch-based classifier along with corresponding ground-truth annotations (regions encircled by black lines) in Figure 4A, B. We also show examples of true-positive and true-negative patch-based predictions from these examples (insets on the left), as well as the patch-based receiver operating characteristic (ROC) curve for each of the 2D cross-sectional images. Several additional examples of generated 2D and 3D heatmaps, ground-truth annotations, and patch-based predictions may be found in [Supplementary Figure 1](#) and [Supplementary Movie 3](#). Additionally, we use principal component analysis to visualize the model's predictions and confirmed how the algorithm distinguishes between benign and neoplastic image patches in feature space (Fig. 4C).³⁹ Two principal components were fitted to the 512-element feature vectors generated by the model for each patch before the final fully connected layer of the ResNet model. Visualization of all predictions in this feature space suggests that the neoplastic patches are well-distinguished from the benign patches. See [Supplementary Figure 2](#) for additional visualization of these patches from different regions of the principal component analysis plot. Finally, the model's performance was estimated with 15-fold cross-validation. As shown in Figure 4D, the overall performance was benchmarked by computing ROC curves for all 15-folds applied to 43 2D image sections from 30 specimens. The average ROC curve is also shown ($AUC = 0.89$, $\sigma = .02$; 95% CI, 0.85-0.92) where the 95% CI is estimated by bootstrapping performance across all cross-validation folds as described in the study by Tsamardinos et al.⁴⁰ where $B = 10,000$. On average, the algorithm identifies neoplastic regions with 90% patch-based sensitivity and 71% patch-based specificity (Fig. 4E), which is deemed adequate in this study for a triage algorithm to screen for the presence of neoplasia in thousands of image patches from hundreds of image sections per 3D pathology data set. In [Supplementary Figure 3A, B](#), we visualize the calibration of the patch-based algorithm,^{41,42} as well as cross-entropy loss and ROC-AUC performance over all 20 epochs during training and testing.

To demonstrate the performance of the image-based classifier, we benchmarked the RFC's performance in discriminating between image sections that are entirely benign vs containing any amount of neoplasia. The average ROC curve for the performance of this image-based classifier is shown in Figure 4E, with an $AUC = 0.92$ ($\sigma = 0.048$; 95% CI, 0.80-0.98). Selection of the most-optimal probability threshold yields an overall 2D image-based sensitivity of 87% and specificity of 73%. We examine the calibration of the image-based classifier in [Supplementary Figure 3C](#).

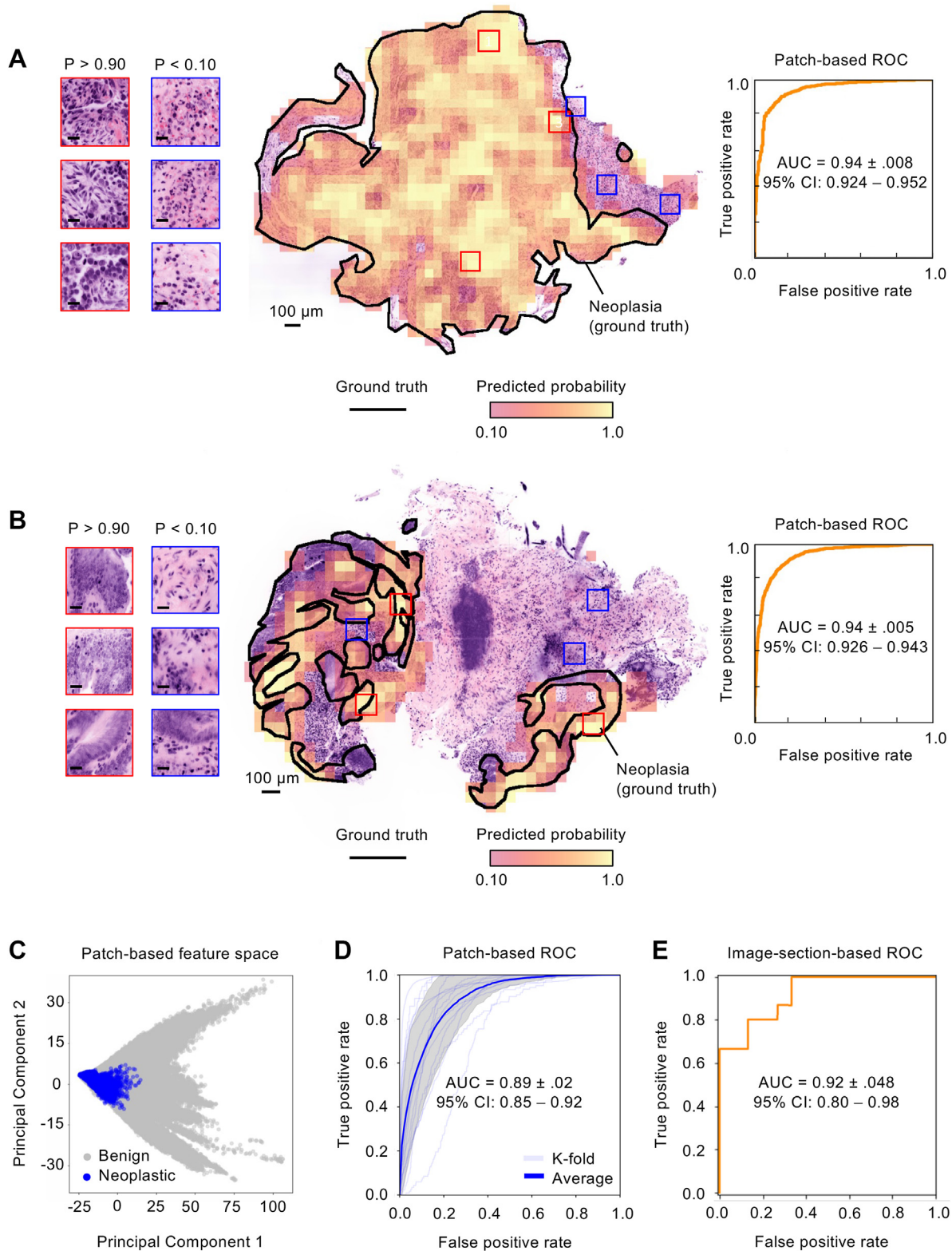
Preliminary Clinical Validation Study

To explore the clinical value of this triage method, we compared our AI-triaged 3D pathology method with gold-standard 2D histology using an independent validation cohort of 20 endoscopic biopsies from 10 patients. As shown in Figure 5A, a board-certified gastrointestinal pathologist first diagnosed each biopsy based on only the top 3 image sections identified with our computational triage method. After AI-triaged 3D pathology was performed, conventional histology sections were obtained from the same biopsies. Here, ~16 physical tissue sections were viewed by the pathologists, per standard practice at our institution. A washout period of at least 2 weeks was implemented between the time in which the AI-triaged image sections were viewed and the standard histology sections were viewed.⁴³ Of the 20 biopsies evaluated, 3 biopsies diagnosed as benign by conventional histology were found to contain neoplasia based on AI-triaged 3D pathology, as shown in the Table. Two examples are shown in Figure 5B. In the first example (biopsy 13), image sections identified by AI-triaged 3D pathology showed hallmarks of HGD, such as fused and crowded glands (green box), large nuclei (yellow box), prominent nucleoli (green arrowhead), and mitoses (yellow arrowhead). However, normal architectural features of basally oriented epithelial cells with mucin caps (blue box) are observed with gold-standard histopathology. In another example (biopsy 15), regions of HGD characterized by focal areas of fused glands (green box), nucleoli (green arrowhead), and mitosis (yellow arrowhead) are seen in the AI-triaged image sections but are not seen in any of the conventional histology sections.

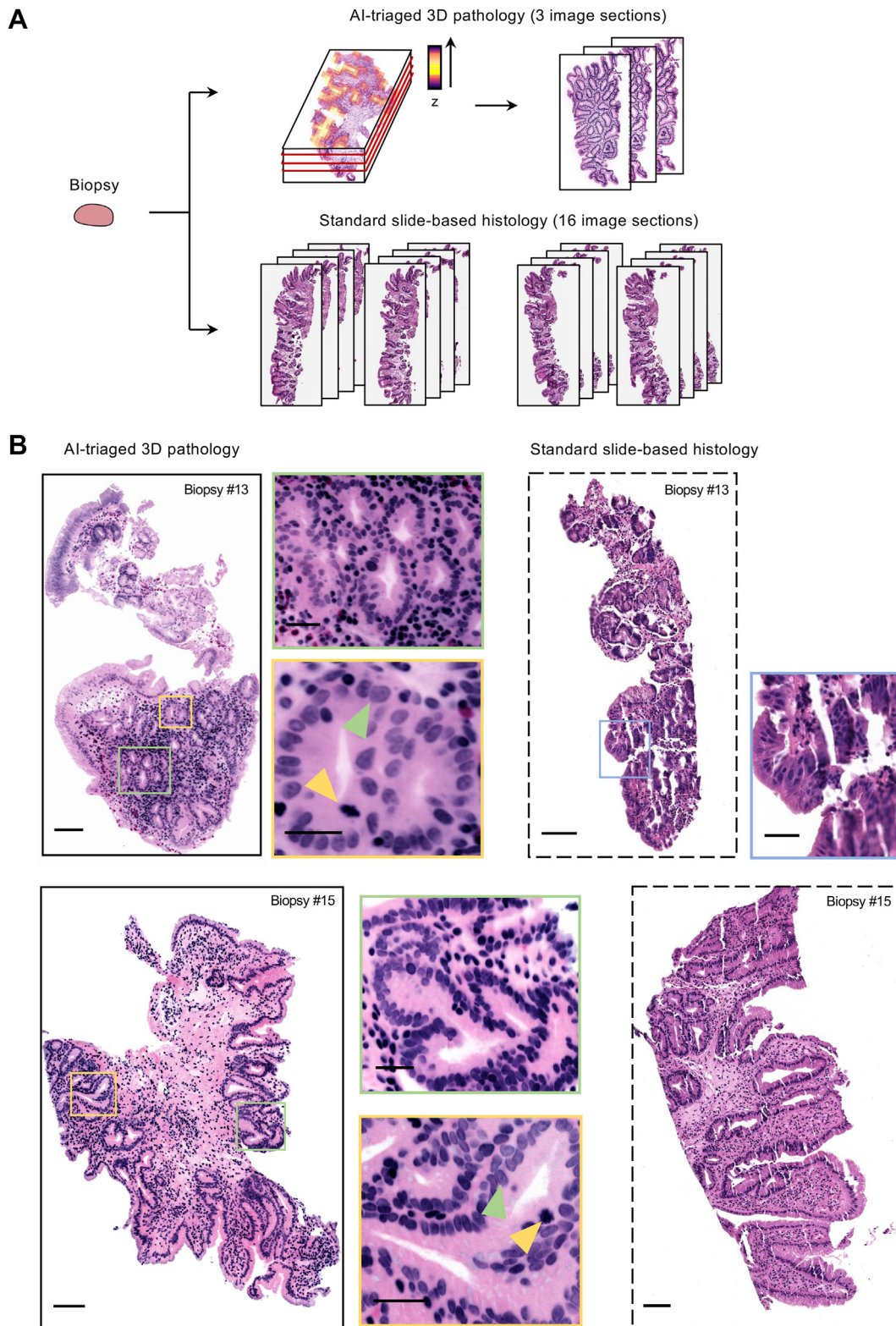
In addition to the 3 biopsies that were diagnostically upgraded from benign BE to neoplastic (biopsies 11, 13, and 15), a fourth biopsy diagnosed as LGD by conventional histology (biopsy 2) was upgraded to HGD with AI-triaged 3D pathology. There were no examples of diagnostic downgrading based on AI-triaged 3D pathology. Note that in these images, certain artifacts inherent to 2D histology, such as cracks, folds, and regions of poor staining quality, are eliminated with nondestructive 3D pathology.

Discussion

Nondestructive 3D pathology enables comprehensive visualization of whole biopsies, which may improve detection sensitivity and facilitate earlier intervention, thereby maximizing patient outcomes. The tissue preparation protocols used in this study are straightforward and scalable, obviating the need for highly skilled histotechnologists. Previous studies with our 3D pathology methods have also shown that these methods are reversible (ie, archived tissues can be restored as formalin-fixed paraffin-embedded blocks after imaging) and that tissue morphology and molecular expression are unchanged.^{18,19,21,23,25} Therefore, the incorporation of 3D pathology into clinical practice can be cost effective and low risk. However, a challenge with 3D pathology is that the massive data sets are tedious for pathologists to comprehensively analyze. To address this challenge, we have implemented an AI-based triage method to identify the most critical 2D image sections within each 3D data set for pathologists to review. With deep learning-based triage of 3D pathology data sets, we show the potential to improve diagnostic sensitivity in comparison to gold-standard 2D histology while reducing pathologist workloads (ie, the number of images that must be viewed) compared with standard histopathology practice. We believe that in clinical practice, our method will be synergistic with endoscopic screening tools under development, such as

**Figure 4.**

Patch-based and image-based classifier performance. (A, B) Probability heatmaps generated by the patch-based deep learning algorithm during cross-validation testing are overlaid onto 2D image sections from three-dimensional (3D) pathology data sets of biopsy specimens from the annotated training set. Each heatmap value indicates the probability of that patch containing neoplasia. The heatmaps are overlaid onto their respective hematoxylin and eosin false-colored 2D images, and the ground-truth annotations of the neoplastic regions are also shown (regions encircled by black lines). On the left, a few examples of true-positive (red outline) and true-negative (blue outline) patch-based predictions are shown. On the right, patch-based ROC curves are also shown for each image along with AUC values ($\pm \sigma$) and 95% CIs. Scale bars represent 20 μ m unless otherwise indicated. (C) Principal component analysis was performed to visualize the model's patch-based predictions in feature space. This visualization suggests that the neoplastic patches (blue) are well-distinguished from the benign patches (gray). (D) ROC curves are plotted for each cross-validation fold for the patch-based deep-learning predictions, as well as for the average of all 15 folds (dark blue). The SD of the ROC curve is shaded in gray. (E) An ROC curve is shown for 2D image section-based predictions (random forest classifier) averaged across all 15 cross-validation folds. ROC, receiver operating characteristic.

**Figure 5.**

Preliminary clinical validation study. (A) Nondestructive three-dimensional (3D) pathology data sets were obtained of 20 intact biopsies. Our AI algorithm identified the 3 image sections with the highest probability of containing neoplasia for pathologist review. The same biopsies were submitted for standard hematoxylin and eosin histology, where ~16 physical tissue sections per biopsy were reviewed. (B) Examples are shown for which AI-triaged 3D pathology upgraded the diagnosis compared with conventional 2D histology. Regions of high-grade dysplasia can be identified in the AI-triaged image sections in biopsy #13 and biopsy #15, as characterized by fused and crowded glands (green boxes), large nuclei (yellow boxes), prominent nucleoli (green arrowhead), and mitoses (yellow arrowhead). However, in the conventional histology sections, architectural features showing basally oriented epithelial cells with mucin caps (blue box) are consistent with a diagnosis of benign Barrett's esophagus. Scale bars are 100 μm for black boxes and 30 μm for all other boxes.

Table

An independent validation cohort of 20 endoscopic biopsies evaluated with AI-triaged 3D pathology vs conventional 2D histology

Biopsy ID	Conventional 2D histology diagnosis	AI-triaged 3D pathology diagnosis
1	Benign	Benign
2	Neoplasia (LGD)	Neoplasia (HGD)
3	Benign	Benign
4	Benign	Benign
5	Benign	Benign
6	Benign	Benign
7	Benign	Benign
8	Benign	Benign
9	Benign	Benign
10	Benign	Benign
11	Benign	Neoplasia (LGD)
12	Benign	Benign
13	Benign	Neoplasia (HGD)
14	Benign	Benign
15	Benign	Neoplasia (HGD)
16	Benign	Benign
17	Benign	Benign
18	Benign	Benign
19	Benign	Benign
20	Benign	Benign

Biopsies 11, 13, and 15 are examples for which AI-triaged 3D pathology upgraded the diagnosis from benign (based on conventional histology) to neoplasia.

3D, three-dimensional; HGD, high-grade dysplasia; LGD, low-grade dysplasia.

optical coherence tomography, Cytosponge-TFF3, and hyperspectral imaging,^{44–49} which will enable the acquisition of targeted biopsies (as opposed to random biopsies) that may be comprehensively assessed by 3D pathology.

As machine learning and deep learning-based approaches are rapidly advanced, developing methods that are interpretable will encourage clinical translation. Our multistep process facilitates interpretability by providing a quantitative heatmap of high-risk regions in the whole 3D volume before selecting the most critical 2D image sections for pathologist review. This intermediate checkpoint provides pathologists with the opportunity to review the tissue regions and pathological structures that are ranked with the highest neoplastic probability, which will improve their confidence in the algorithm.

Although our method enables improved diagnostic sensitivity in comparison to conventional histology, the RFC's task in the second stage of our computational pipeline is particularly challenging as it must learn to recognize and correctly prioritize various manifestations of neoplasia. For example, it must evaluate whether image sections with large diffuse regions of neoplasia should be ranked as more important than image sections with small focal areas of neoplasia. To facilitate learning of these distinctions, we chose features (see Methods) that would account for some of these variables as inputs to the RFC. For example, the “maximum predicted probability” feature enables the RFC to consider image sections with small focal areas of neoplasia, whereas the “number of patches for which $P > .10$ ” feature allows the RFC to consider image sections with large regions of neoplasia. More extensive hand-crafted or deep learning-based feature engineering approaches could enable improved selection of the most important image sections for pathologist review.

To facilitate the adoption of 3D pathology into clinical settings, this study explores a diagnostic paradigm that keeps the pathologist “in the loop” to minimize risk. Furthermore, due to the

severe workload and staffing shortages faced by pathology laboratories, we sought to design a time-efficient diagnostic workflow that requires minimal retraining of the pathologist. This motivated our choice to provide pathologists with a small number of 2D H&E false-colored image sections, which they are accustomed to viewing, rather than large image stacks or videos of 3D data. It should be noted, however, that pathologist review of 3D data may enhance diagnostic accuracy, especially for structures that can be ambiguous when viewed as 2D images.^{22,23,26} Therefore, it is worth exploring the display of high-priority localized 3D image regions as image stacks or volumetric renderings. Future studies should also explore algorithms that classify neoplastic regions based on 3D chunks instead of 2D patches, which could improve the accuracy of deep learning models through the incorporation of novel 3D features.²⁶ Finally, future studies may benefit from confirming the 3D pathology diagnosis of neoplasia with immunofluorescent labeling or a panel of pathologists.⁵⁰

One of the limitations of our method is that tedious pixel-level annotations are required to train the patch-based deep learning algorithm. Recent advances in weakly supervised learning facilitate direct image-based predictions after training with image-based or patient-based labels,^{39,51,52} as well as the ability to generate attention maps that offer a similar level of interpretability to pathologists.⁵³ However, it can be challenging to achieve accurate performance with such single-step methods with relatively small training data sets, as used in this study. Future studies are needed with larger training set sizes to explore the performance for various strategies of AI-triaged 3D pathology. Ideally, this work would facilitate improved detection sensitivity of neoplasia (>95%) in comparison to this study (~90%). More importantly, prospective studies are needed to demonstrate that these methods can lead to improved long-term outcomes for patients with BE.

In conclusion, we have developed a nondestructive 3D pathology workflow to enable comprehensive histologic evaluation of whole biopsies and a deep learning-based computational triage method to identify the most important 2D image sections for time-efficient pathologist review. Our results suggest that AI-triaged 3D pathology has the potential to improve diagnostic accuracy while reducing pathologist workloads.

Author Contributions

L.B. and J.L. conceived of the study. L.B., D.R., W.G., and J.L. designed the study. L.B., G.G., and L.L. prepared the tissues for imaging and collected the data. D.R. annotated all images for training and the clinical validation study. L.B. and G.G. performed data analysis and created the figures. G.G. designed and orchestrated the clinical validation study. D.R., W.B., F.M., W.G., and J.L. oversaw the study, made interpretations of the data, and contributed to the manuscript. L.B., G.G., and J.L. wrote the initial draft of the manuscript. L.B. and G.G. contributed equally to this work.

Data Availability

The data that supports the findings of this study were generated with specimens provided by the Gastrointestinal Center for Analytic Research and Exploratory Development (GiCaRes) at UW under a confidentiality pledge. An example data set is available online at <https://github.com/lindseyerionbarner/AI-3Dpath>.

Additional data may be made available from the corresponding authors upon request following these regulations and the establishment of a data use agreement (DUA). Custom code and example data sets are available online at <https://github.com/lindseyerionbarner/AI-3Dpath>.

Funding

These studies were supported by funding from the National Institutes of Health (NIH), including R01EB031002 and R01CA268207 (to J.T.C.L.), and U01CA152756, R01CA220004, U2CCA271902, U54CA163060, U01CA182940 (to W.M.G.). Funding is also provided by the National Science Foundation (NSF) 1934292 HDR: I-DIRSE-FW (to J.T.C.L.) and a NSF Graduate Research Fellowship DGE-1762114 (to L.B.). Additional funding was provided by the Prevent Cancer Foundation, Cottrell Family Fund, Evergreen Fund, and Listwin Foundation (to W.M.G.). These studies are also supported by GiCaRes from UW Departments of Medicine (Gastroenterology Division) and Lab Medicine & Pathology.

Declaration of Competing Interest

J.T.C. Liu is a cofounder, equity holder, and board member of Alpenglow Biosciences, Inc. W. M. Grady is a scientific advisory board member for Freenome, Guardant Health, and SEngine and consultant for DiaCarta, Nephron, Guidepoint, and GLG. He is an investigator in a clinical trial sponsored by Janssen Pharmaceuticals and receives research support from Tempus and LucidDx. The other authors declare that they have no conflicts of interest.

Ethics Approval and Consent to Participate

All procedures were done in accordance with IRB-approved protocols (IR9066).

Supplementary Material

The online version contains supplementary material available at <https://doi.org/10.1016/j.modpat.2023.100322>

References

- Vining P, Birdas TJ. Management of clinical T2N0 esophageal cancer: a review. *J Thorac Dis*. 2019;11(suppl 12):S1629–S1632.
- Smyth EC, Lagergren J, Fitzgerald RC, et al. Oesophageal cancer. *Nat Rev Dis Primers*. 2017;3:17048.
- Joseph A, Raja S, Kamath S, et al. Esophageal adenocarcinoma: a dire need for early detection and treatment. *Cleve Clin J Med*. 2022;89:269–279.
- Qumseya B, Sultan S, Bain P, et al. ASGE guideline on screening and surveillance of Barrett's esophagus. *Gastrointest Endosc*. 2019;90:335–359.e2.
- Killcoyne S, Fitzgerald RC. Evolution and progression of Barrett's oesophagus to oesophageal cancer. *Nat Rev Cancer*. 2021;21:731–741.
- Kambhampati S, Tieu AH, Luber B, Wang H, Meltzer SJ. Risk factors for progression of Barrett's esophagus to high grade dysplasia and esophageal adenocarcinoma. *Sci Rep*. 2020;10:4899.
- Whitson MJ, Falk GW. Predictors of progression to high-grade dysplasia or adenocarcinoma in Barrett's esophagus. *Gastroenterol Clin North Am*. 2015;44:299–315.
- Hvid-Jensen F, Pedersen L, Drewes AM, Sorensen HT, Funch-Jensen P. Incidence of adenocarcinoma among patients with Barrett's esophagus. *N Engl J Med*. 2011;365:1375–1383.
- Spechler SJ. Barrett esophagus and risk of esophageal cancer: A clinical review. *JAMA*. 2013;310:627–636.
- Lekakos L, Karidis NP, Dimitroulis D, Tsigiris C, Kouraklis G, Nikiteas N. Barrett's esophagus with high-grade dysplasia: Focus on current treatment options. *World J Gastroenterol*. 2011;17:4174–4183.
- Shaheen NJ. Risk of cancer in patients with Barrett esophagus. *Gastroenterol Hepatol (N Y)*. 2019;15:688–690.
- Berry MF. Esophageal cancer: Staging system and guidelines for staging and treatment. *J Thorac Dis*. 2014;6(suppl 3):S289–S297.
- van Munster S, Nieuwenhuis E, Weusten BLAM, et al. Long-term outcomes after endoscopic treatment for Barrett's neoplasia with radiofrequency ablation ± endoscopic resection: Results from the national Dutch database in a 10-year period. *Gut*. 2022;71:265–276.
- Reed MF, Tolis G, Edil BH, et al. Surgical treatment of esophageal high-grade dysplasia. *Ann Thorac Surg*. 2005;79:1110–1115.
- Mathew A, Gerkin R, Ramirez FC. Investigating biopsy sampling error in a large Barrett's esophagus (BE) cohort. *Gastrointest Endosc*. 2007;65.
- Kayser K, Schultz H, Goldmann T, Görtler J, Kayser G, Vollmer E. Theory of sampling and its application in tissue based diagnosis. *Diagn Pathol*. 2009;4:6.
- Adami B, Eckardt VF, Paulini K. Sampling error and observer variation in the interpretation of esophageal biopsies. *Digestion*. 1979;19:404–410.
- Reddi DM, Barner LA, Burke W, Gao G, Grady WM, Liu JTC. Nondestructive 3D pathology image atlas of Barrett esophagus with open-top light-sheet microscopy. *Arch Pathol Lab Med*. Published online January 3, 2023. <https://doi.org/10.5858/arpa.2022-0133-0A>
- Glaser AK, Reder NP, Chen Y, et al. Light-sheet microscopy for slide-free non-destructive pathology of large clinical specimens. *Nat Biomed Eng*. 2017;1:1–10.
- Barner LA, Glaser AK, Huang H, True LD, Liu JTC. Multi-resolution open-top light-sheet microscopy to enable efficient 3D pathology workflows. *Biomed Opt Express*. 2020;11:6605–6619.
- Glaser AK, Reder N, Chen Y, et al. Multi-immersion open-top light-sheet microscope for high-throughput imaging of cleared tissues. *Nat Commun*. 2019;10:2781.
- Barner LA, Glaser AK, Mao C, et al. Multiresolution nondestructive 3D pathology of whole lymph nodes for breast cancer staging. *J Biomed Opt*. 2022;27:1–13.
- Reder NP, Glaser AK, McCarty EF, Chen Y, True LD, Liu JTC. Open-top light-sheet microscopy image atlas of prostate core needle biopsies. *Arch Pathol Lab Med*. 2019;143:1069–1075.
- Serafin R, Xie W, Glaser AK, Liu JTC. FalseColor-Python: a rapid intensity-leveling and digital-staining package for fluorescence-based slide-free digital pathology. *PLoS One*. 2020;15, e0233198.
- Chen Y, Xie W, Glaser AK, et al. Rapid pathology of lumpectomy margins with open-top light-sheet (OTLS) microscopy. *Biomed Opt Express*. 2019;10:1257–1272.
- Xie W, Reder NP, Koyuncu C, et al. Prostate cancer risk stratification via nondestructive 3D pathology with deep learning–assisted gland analysis. *Cancer Res*. 2022;82:334–345.
- Leffler DA, Kheraj R, Garud S, et al. The incidence and cost of unexpected hospital use after scheduled outpatient endoscopy. *Arch Intern Med*. 2010;170:1752–1757.
- Metter DM, Colgan TJ, Leung ST, Timmons CF, Park JY. Trends in the US and Canadian pathologist workforces from 2007 to 2017. *JAMA Netw. Open*. 2019;2, e194337.
- Jajosky RP, Jajosky AN, Kleven DT, Singh G. Fewer seniors from United States allopathic medical schools are filling pathology residency positions in the Main Residency Match, 2008–2017. *Hum Pathol*. 2018;73:26–32.
- Lundberg GD. How many pathologists does the United States need? *JAMA Netw. Open*. 2019;2:e194308.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. *Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit*. 2016:770–778.
- Wang D, Khosla A, Gargeya R, Irshad H, Beck AH. Deep Learning for Identifying Metastatic Breast Cancer. *arXiv*. 2016:1–6.
- Keller PJ, Schmidt AD, Wittbrodt J, Stelzer EHK. Digital scanned laser light-sheet fluorescence microscopy (DSLM) of zebrafish and drosophila embryonic development. *Cold Spring Harb Protoc*. 2011;2011:1235–1243.
- Balázs B, Deschamps J, Albert M, Ries J, Hufnagel L. A real-time compression library for microscopy images. *bioRxiv*. 2017.
- Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. ImageNet: a large-scale hierarchical image database. *IEEE Conf Comput Vis Pattern Recognit*. 2009:248–255.
- Otsu N. A threshold selection method from gray-level histograms. *IEEE Trans Syst Man Cybern*. 1979;9:62–66.
- Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. *Int Jt Conf Artif Intell*. 1995:1137–1143.
- Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *J Mach Learn Res*. 2014;15:1929–1958.
- Lu MY, Williamson DFK, Chen TY, Chen RJ, Barbieri M, Mahmood F. Data-efficient and weakly supervised computational pathology on whole-slide images. *Nat Biomed Eng*. 2021;5:555–570.
- Tsamardinos I, Greisdou E, Borboudakis G. Bootstrapping the out-of-sample predictions for efficient and accurate cross-validation. *Mach Learn*. 2018;107:1895–1922.

41. Guo C, Pleiss G, Sun Y, Weinberger KQ. On calibration of modern neural networks. *Proc 34th Int Conf Mach Learn*. 2017;70:2130–2143.
42. Lu MY, Chen TY, Williamson DFK, et al. AI-based pathology predicts origins for cancers of unknown primary. *Nature*. 2021;594:106–110.
43. Pantanowitz L, Sinard JH, Henricks WH, et al. Validating whole slide imaging for diagnostic purposes in pathology: guideline from the College of American Pathologists Pathology and Laboratory Quality Center. *Arch Pathol Lab Med*. 2013;137:1710–1722.
44. Liang K, Ahsen OO, Lee HC, et al. Volumetric mapping of Barrett's esophagus and dysplasia with en face optical coherence tomography tethered capsule. *Am J Gastroenterol*. 2016;111:1664–1666.
45. Kaul V. Optical coherence tomography for Barrett esophagus. *Gastroenterol Hepatol (N Y)*. 2018;14:253–255.
46. Fitzgerald RC, di Pietro M, O'Donovan M, et al. Cytosponge-trefoil factor 3 versus usual care to identify Barrett's oesophagus in a primary care setting: a multi-centre, pragmatic, randomised controlled trial. *Lancet*. 2020;396:333–344.
47. Pilonis ND, Killcoyne S, Tan WK, et al. Use of a Cytosponge biomarker panel to prioritise endoscopic Barrett's oesophagus surveillance: a cross-sectional study followed by a real-world prospective pilot. *Lancet Oncol*. 2022;23:270–278.
48. Yoon J, Joseph J, Waterhouse DJ, et al. A clinically translatable hyperspectral endoscopy (HySE) system for imaging the gastrointestinal tract. *Nat Commun*. 2019;10:1902.
49. Lu G, Fei B. Medical hyperspectral imaging: a review. *J Biomed Opt*. 2014;1:10901.
50. Polkowski W, Van Lanschot JJ, Ten Kate FJ, et al. The value of p53 and Ki67 as markers for tumour progression in the Barrett's dysplasia-carcinoma sequence. *Surg Oncol*. 1995;4:163–171.
51. Chen RJ, Lu MY, Williamson DFK, et al. Pan-cancer integrative histology-genomic analysis via multimodal deep learning. *Cancer Cell*. 2022;40:865–878.e6.
52. Campanella G, Hanna MG, Geneslaw L, et al. Clinical-grade computational pathology using weakly supervised deep learning on whole slide images. *Nat Med*. 2019;25:1301–1309.
53. Ilse M, Tomczak JM, Welling M. Attention-based deep multiple instance learning. *35th Int Conf Mach Learn*. 2018;5:3376–3391.