JGH Open Open Access An open access journal of gastroenterology and hepatology

JGHF Journal of Gastroenterology and Hepatology Foundation

doi:10.1002/jgh3.70030

ORIGINAL ARTICLE

Exploring vision transformers for classifying early Barrett's dysplasia in endoscopic images: A pilot study on white-light and narrow-band imaging

Jin L Tan,[*,†] Dileepa Pitawela,[‡] Mohamed A Chinnaratha,[*,†] Andrawus Beany,[*] Enrik J Aguila,[*,§] Hsiang-Ting Chen,[‡] Gustavo Carneiro[¶] and Rajvinder Singh[*,†]

[*]Department of Gastroenterology and Hepatology, Lyell McEwin Hospital, SA Health, Elizabeth Vale, [†]Faculty of Health and Medical Sciences, [‡]School of Computer and Mathematical Sciences, The University of Adelaide, Adelaide, South Australia, Australia, [§]Institute of Digestive and Liver Diseases, St. Luke's Medical Center Global City, Taguig, Philippines and [¶]Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK

Key words
artificial intelligence, Barrett's esophagus, dysplasia, esophageal adenocarcinoma, vision transformer.

Accepted for publication 4 September 2024.

Correspondence
Dr Jin L Tan, Department of Gastroenterology, Lyell McEwin Hospital, Haydown Road, Elizabeth Vale, SA 5112, Australia. Email: jin.tan@adelaide.edu.au

Declaration of conflict of interest: None.
Author contribution: Jin L Tan, Dileepa Pitawela, Mohamed A Chinnaratha, Hsiang-Ting Chen, Gustavo Carneiro, Rajvinder Singh contributed to conceptualization, analysis, interpretation of data, writing, and review of the manuscript. Jin L Tan, Andrawus Beany, Enrik J Aguila, Mohamed A Chinnaratha, and Rajvinder Singh contributed to the data classification and validation. Dileepa Pitawela, Hsiang-Ting Chen, Gustavo Carneiro, and Jin L Tan contributed to the training of artificial intelligence models.

Abstract

Background and Aim: Various deep learning models, based on convolutional neural network (CNN), have been shown to improve the detection of early esophageal neoplasia in Barrett's esophagus. Vision transformer (ViT), derived from natural language processing, has emerged as the new state-of-the-art for image recognition, outperforming predecessors such as CNN. This pilot study explores the use of ViT to classify the presence or absence of early esophageal neoplasia in endoscopic images of Barrett's esophagus.
Methods: A BO dataset of 1918 images of Barrett's esophagus from 267 unique patients was used. The images were classified as dysplastic (D-BO) or non-dysplastic (ND-BO). A pretrained vision transformer model, ViTBase16, was used to develop our classifier models. Three ViT models were developed for comparison based on imaging modality: white-light imaging (WLI), narrow-band imaging (NBI), and combined modalities. Performance of each model was evaluated based on accuracy, sensitivity, specificity, confusion matrices, and receiver operating characteristic curves.
Results: The ViT models demonstrated the following performance: WLI-ViT (Accuracy: 92%, Sensitivity: 82%, Specificity: 95%), NBI-ViT (Accuracy: 99%, Sensitivity: 97%, Specificity: 99%), and combined modalities-ViT (Accuracy: 93%, Sensitivity: 87%, Specificity: 95%). Combined modalities-ViT showed greater accuracy (94% *vs* 90%) and sensitivity (80% *vs* 70%) compared with WLI-ViT when classifying WLI images on a subgroup testing set.
Conclusion: ViT exhibited high accuracy in classifying the presence or absence of EON in endoscopic images of Barrett's esophagus. ViT has the potential to be widely applicable to other endoscopic diagnoses of gastrointestinal diseases.

Introduction

Barrett's esophagus (BO) is a metaplastic alteration of the normal esophagus epithelium, predisposing to the development of esophageal adenocarcinoma (OAC).[1] International guidelines recommend endoscopic surveillance to detect early esophageal neoplasia (EON), the presence of dysplasia or intramucosal OAC, at a curable stage before its progression to advanced OAC for which the 5-year overall survival rate is at a dismal 17%.[2–4] Despite advanced imaging modalities such as narrow-band imaging (NBI), an average of 25% of EON remained missed.[5]

A recent systematic review and meta-analysis demonstrated that artificial intelligence (AI) can improve the accuracy of detecting EON, with high accuracy, sensitivity and specificity of 0.94 (95% CI: 0.92–0.96), 90.3% (95% CI: 87.1–92.7%) and 84.4% (95% CI: 80.2–87.9%), respectively among the included studies.[6] These results are above the threshold targets (sensitivity of more than 90%, specificity of more than 80%) recommended by the American Society for Gastrointestinal Endoscopy (ASGE).[7] Current published AI models for Barrett's surveillance employ a popular AI method known as convolutional neural networks (CNNs) to train their models.[8–14] CNN extracts feature methodically from pixels of an image by undergoing repeated convolution and pooling processes to achieve accurate performance in disease detection and diagnosis.[15]

Vision transformer (ViT) has recently emerged as the current state-of-the-art in machine learning for image recognition.[16] First, an image is split into patches (also known as tokens) before being fed into the Transformer Encoder. The encoder has an

"attention-based" network that assigns weights to tokens based on their key features in relation to the image. An image is subsequently classified based on the calculated probabilities of the image's tokens.[16] ViT models have been shown to outperform CNN models by fourfold in terms of computational efficiency and accuracy.[16,17] Given the novelty of ViT, it has not yet been used to train AI models for Barrett's surveillance. Therefore, we embarked on the development of ViT-based models to explore their performances at detecting EON in BO using both white-light imaging (WLI) and NBI.

Methods

Ethics. This research was approved by the Human Research Ethics Committee (Central Adelaide Local Health Network reference number: 16179).

Setting. Study data were obtained from a BO database from the Lyell McEwin Hospital (LMH), a large tertiary referral centr for the management of BO in Australia, spanning the period from 2008 to 2022. The study was performed in collaboration with the Australian Institute for Machine Learning at the University of Adelaide, Australia, and the Centre for Vision, Speech and Signal Processing at the University of Surrey, United Kingdom.

Phase 1: Development and classification BO dataset. Endoscopic images were retrospectively collected and de-identified from sequential patients in the BO database, consisting of 23 197 endoscopic images from 550 patients. Endoscopy procedures were performed on a wide range of Olympus endoscopes (Models: GIF-160, GIF-Q160, GIF-Q160Z, GIF-H180, GIF-H180J, GIF-XP180N, GIF-Q180, GIF-HQ190, GIF-EZ1500).

Inclusion criteria: Images of the esophagus and gastroesophageal junction containing Barrett's esophagus were included, with biopsies confirming BO as defined by the presence of intestinal metaplasia, with the presence of mucin-containing goblet cells.[2] The BO dataset was classified based on imaging modality (WLI or NBI) and the presence of EON, with dysplastic BO and non-dysplastic BO labelled as D-BO and ND-BO, respectively.

Exclusion criteria: Images of Barett's esophagus containing low-grade or indeterminate dysplasia were excluded from the study due to significant interobserver variability between pathologists. This decision was made due to the challenges in establishing a consistent histological ground truth for such cases.

To determine data ground truth, images were selected by three gastroenterology fellows (J.T., A.B., and E.A.) and verified by two expert gastroenterologists (R.S. and M.A.C.). Where there was interobserver disagreement, a consensus was achieved by discussion between expert gastroenterologists (R.S. and M.A.C.). Respective histopathology reports were obtained to confirm the presence (dysplastic) or absence (nondysplastic) of neoplasia. An independent histopathology assessment was sought where there were equivocal results, and consensus was reached after discussion by two pathologists.

Phase 2: Training of AI models

Defining training and testing dataset. The BO dataset was split to ensure that images in the testing dataset (10%) were from unique patients, independent of the training dataset (90%), as shown in Figure 1. Three datasets were defined for comparison based on modalities: WLI, NBI, and combined (WLI and NBI). We included a combined modalities dataset to evaluate whether vision transformers (ViT) can handle mixed modalities effectively, as WLI and NBI images differ significantly in appearance. WLI provides a broad view with natural color representation, while NBI enhances vascular and mucosal patterns using specific wavelengths of light.

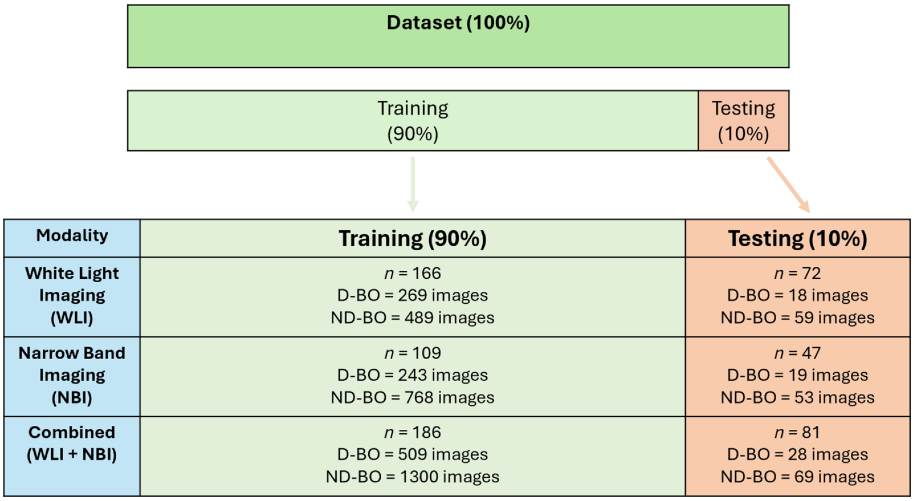
Model architectures. A pretrained Vision Transformer model, ViTBase16, was used as the base model. ViTBase 16 is equipped with 12 transformer layers and approximately 86 million learnable parameters.[16] This model was pretrained on ImageNet1K dataset, a collection of 1.2 million labeled real-world nonmedical images, enabling deep learning networks to acquire fundamental discriminative features, which can be subsequently used to enhance model performance and efficiency.[18]

Data augmentation. Data augmentation was performed using offline and online methods to expand the dataset to improve the generalizability of our model. During the offline augmentation phase, the class with a smaller number of images in each dataset was augmented using random horizontal or vertical flip and random affine transformations. Additionally, a set of online augmentations was applied to further enhance the variability of the training data. These augmentations included color jitter, random histogram equalization, random sharpness adjustments, random image resizing, random rotation, and random perspective transformations. Each online augmentation technique was applied with specific probabilities of occurrence, valuing 0.1, 0.2, 0.2, 0.3, 0.2, and 0.2, respectively.

Training. The ViTBase16 model underwent an extension of its classification head by incorporating three additional linear layers. These layers consisted of 512, 128, and 2 neurons, respectively, to reduce dimensionality for binary classification and accompanied by rectified linear unit (ReLU) activations and dropout layers in between. During the fine-tuning process, the classification head and the 12th encoder layer of the ViT model were fine-tuned for 50, 40, and 80 epochs on WLI, NBI, and combined modalities datasets, respectively.[19] Among the various combinations tested, we observed that employing Cross Entropy loss and Adam optimizer with a learning rate of 0.001 and a batch size of 64 yielded better results. All the experiments were conducted on an NVIDIA GeForce RTX 4090 GPU having 24 GB memory.

Outcomes and statistical analysis. The primary outcome of this study is to assess the accuracy of using ViT models to classify the presence or absence of EON on the following imaging modalities: WLI, NBI, and combined modalities.

The secondary outcome is to compare the performance of the combined modalities model against WLI and NBI models individually. For this comparison, the WLI and NBI subgroup testing datasets were filtered to ensure that they do not include



Modality	Training (90%)	Testing (10%)
White Light Imaging (WLI)	*n* = 166 D-BO = 269 images ND-BO = 489 images	*n* = 72 D-BO = 18 images ND-BO = 59 images
Narrow Band Imaging (NBI)	*n* = 109 D-BO = 243 images ND-BO = 768 images	*n* = 47 D-BO = 19 images ND-BO = 53 images
Combined (WLI + NBI)	*n* = 186 D-BO = 509 images ND-BO = 1300 images	*n* = 81 D-BO = 28 images ND-BO = 69 images

Figure 1 Defining training and testing dataset.

endoscopic images of patients used in the training datasets for both the single and combined modalities models.

Diagnostic accuracy for each model will be determined by sensitivity, specificity, precision, confusion matrix, and receiver operating characteristic (ROC) curve analysis. Precision measures the accuracy of the positive predictions made by the AI models. It is defined as the ratio of true positives to the total number of true positives predictions. The ROC curve analysis plots the true-positive rate (sensitivity) against the false-positive rate (1-specificity), showing the trade-off between the true-positive rate and false-positive rate, to evaluate the performance of the AI models.

Results

Characteristics of BO dataset. The flowchart of BO database development and classification is summarized in Figure 2. From a total of 23 197 endoscopic images from 550 unique patients, 1918 images of Barrett's esophagus from 267 unique patients, including 835 WLI images and 1083 NBI images, were clinically validated and classified accordingly. The WLI dataset contains 287 D-BO images and 548 ND-BO images from 238 unique patients. The NBI dataset contains 262 D-BO images and 821 ND-BO images from 156 unique patients.

Training and testing datasets. The WLI training dataset consisted of 269 D-BO and 489 ND-BO images from 166 unique patients. The WLI testing dataset comprised 18 D-BO and 59 ND-BO images from 72 unique patients.

NBI training dataset consisted of 243 D-BO and 768 ND-BO images from 109 unique patients. The NBI testing dataset comprised 19 D-BO and 53 ND-BO images from 47 unique patients.

The combined modality training dataset consisted of 268 WLI/D-BO, 491 WLI/ND-BO, 241 NBI/D-BO, and

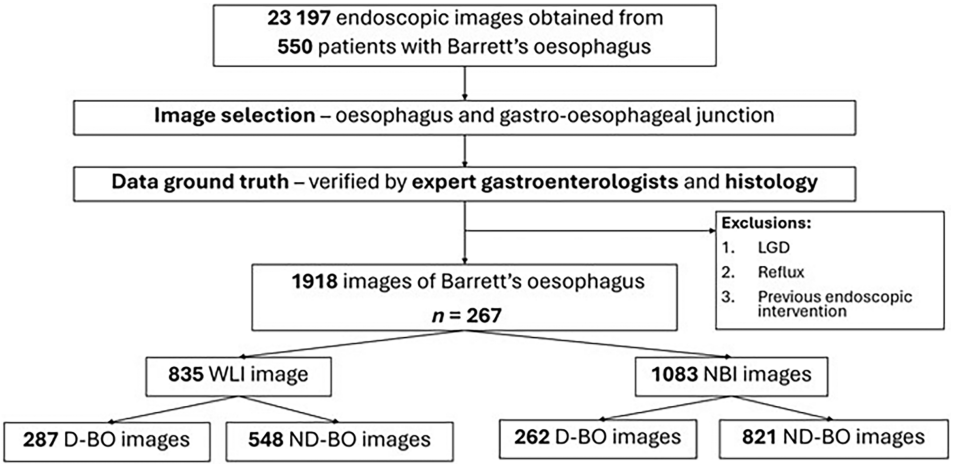


Figure 2 Flowchart of BO database development and classification.

Table 1 Performance of vision transformer models by imaging modalities: white-light imaging, narrow-band imaging, and combined modalities

	Accuracy	Sensitivity	Specificity
White-light imaging	92%	78%	97%
Narrow-band imaging	99%	94%	100%
Combined modalities	93%	86%	96%

809 NBI/ND-BO images from 186 unique patients. The combined modality testing dataset comprised 19 WLI/D-BO, 57 WLI/ND-BO, 9 NBI/D-BO, and 12 NBI/ND-BO images from 81 unique patients.

Primary outcomes. Table 1 summarizes the performance of the three ViT models for the following three comparator datasets: WLI, NBI, and combined modalities.

White-light imaging-ViT. The WLI-ViT model demonstrated the following performance: accuracy of 92%, sensitivity of 78%, and specificity of 97%. Confusion matrices and receiving operating curves of both models are shown in Figures 3 and 4, respectively.

Narrow-band imaging-ViT. The NBI-ViT model demonstrated the following performance: accuracy of 99%, sensitivity of 94%, and specificity of 100%. Confusion matrices and receiving operating curves of both models are shown in Figures 3 and 4, respectively.

Combined modalities-ViT. The combined modalities-ViT model demonstrated the following performance: accuracy of 93%, sensitivity of 86%, and specificity of 96%. Confusion matrices and receiving operating curves of both models are shown in Figures 3 and 4, respectively. Table 2 shows examples of true positives, true negatives, false positives, and false negatives.

Secondary outcomes

Combined modalities-ViT versus *WLI-VIT.* The WLI subgroup testing dataset for this comparison comprised 54 images from 51 unique patients: 10 D-BO images from 7 patients and 44 ND-BO images from 44 patients. The WLI-ViT model demonstrated the following performance: accuracy of 90%, sensitivity of 70%, and specificity of 95%. The combined modalities-ViT model demonstrated the following performance: accuracy of 94%, sensitivity of 80%, and specificity of 98%.

Combined modalities-ViT versus *NBI-ViT.* The NBI subgroup testing dataset for this comparison comprised 18 images from 16 unique patients: 8 D-BO images from 6 patients and 10 ND-BO images from 10 patients. The NBI-ViT model demonstrated the following performance: accuracy of 100%,

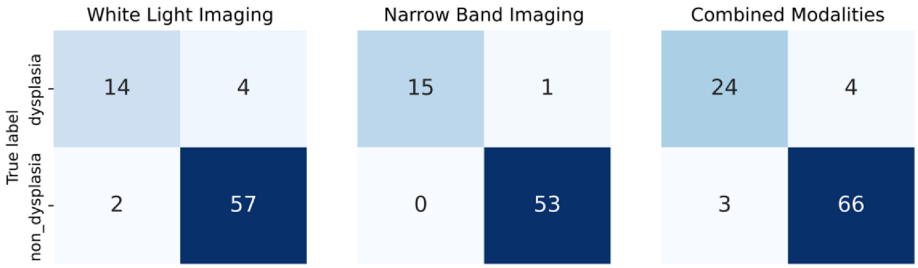


Figure 3 Performance of vision transformer models—confusion matrices.

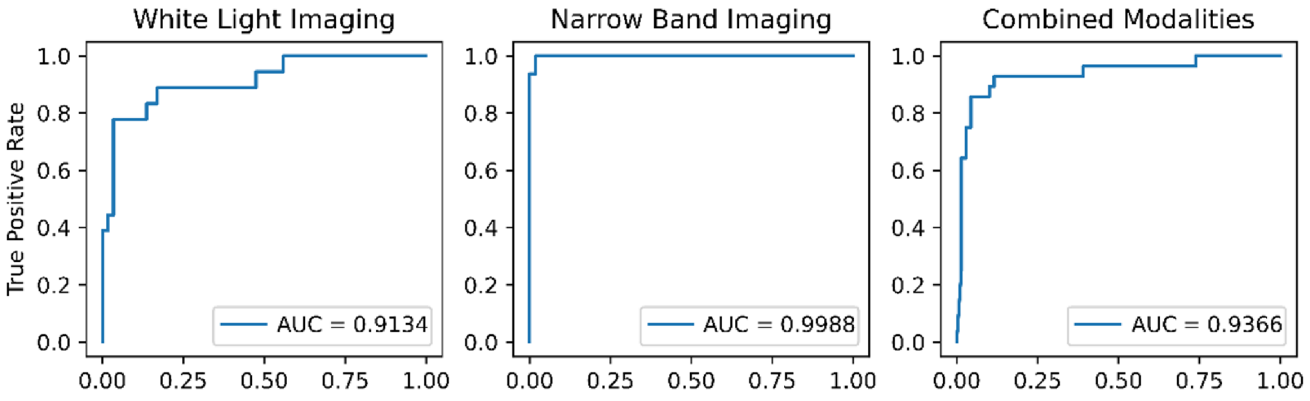
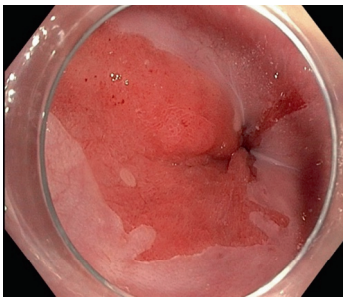
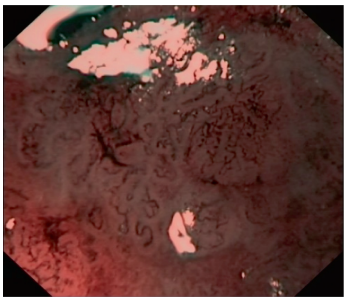
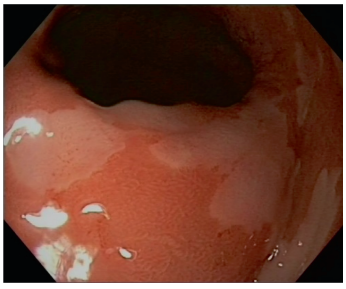
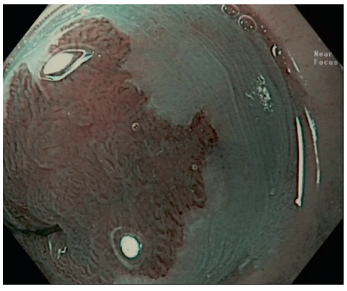
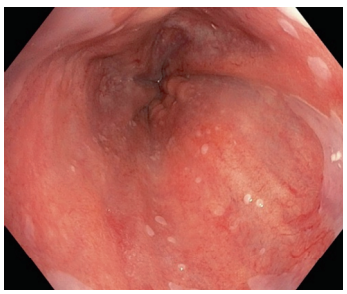
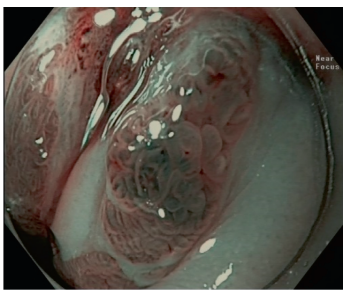
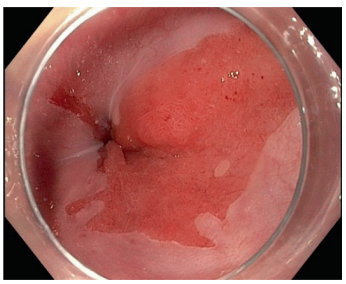
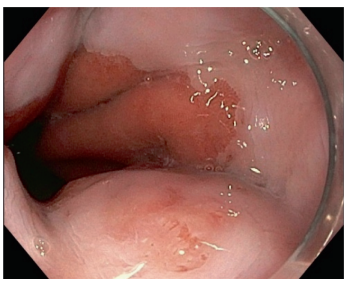


Figure 4 Performance of vision transformer models—receiver operating characteristic (ROC) curves.

Table 2 Combined modalities-ViT model classification: examples of true positive, true negative, false positive, and false negative

Combined modalities-ViT model		
True positive		
True negative		
False positive		
False negative		

sensitivity of 100%, and specificity of 100%. The combined modalities-ViT model demonstrated the following performance: accuracy of 94%, sensitivity of 100%, and specificity of 90%.

Discussion

All three ViT models exhibited high accuracies at differentiating between D-BO and ND-BO. The performances are comparable

with the currently published results of existing models for EON detection, shown in a recent meta-analysis where the overall accuracy of all included studies was 0.94 (95% CI: 0.92–0.96). The pooled sensitivity and specificity were 90.3% (95% CI: 87.1–92.7%) and 84.4% (95% CI: 80.2–87.9%), respectively.[6]

Currently, most published models are predominantly trained on single imaging modality, with majority being WLI models.[6,14] In clinical practice, the combination of WLI and NBI has been shown to improve diagnostic accuracies to detect EON compared with WLI only with random biopsies.[20,21] Consequently, there is a need for AI models to support both imaging modalities. By combining images of NBI and WLI during the training process, the combined modalities-ViT model was shown to have a higher accuracy (94% *vs* 90%) and sensitivity (80% *vs* 70%) at classifying WLI images compared with the WLI-ViT model, which was trained on WLI images alone. This could be attributed to additional characteristics provided by NBI images, such as the details of mucosal pit patterns and vasculature that are not readily visible on WLI, or merely due to the larger number of training data alone in the combined modalities training dataset. Conversely, combined modalities-ViT appeared to have a lower accuracy and sensitivity at classifying the NBI subgroup testing dataset compared with NBI-ViT. However, the small size of the subgroup NBI testing dataset for this comparison limits the generalizability of these results.

Our NBI-ViT model exhibited a high sensitivity of 94%, representing the best reported results in the literature, to the best of our knowledge. NBI is widely employed in Barrett's surveillance, enhancing the detection of early dysplasia.[22] Existing classification systems such as the Nottingham Classification and the Asia-Pacific Barrett's Consortium Classification necessitate extensive training and experience at specialized tertiary centers, a gap now addressed by AI models.[23] With such high sensitivities, random biopsies could be averted in the future, a potentially practice changing advancement in BO surveillance.

This study has three major strengths. First, this is the largest BO dataset in the Asia-Pacific region, consisting of 1918 pre-augmented images derived from 267 unique patients, all of which were validated by experts with histological correlation. This dataset included a large number of unique patients compared with other published studies, therefore allowing us to perform a pilot exploratory study on ViT as a classifier tool to detect EON.[6] In a study by de Groof *et al*., their AI model was trained using images from 509 unique patients. However, only 95 of those patients' images were validated by experts with histological correlation, while the remaining images were solely validated by experts.[10] Another study reported training their AI model with over a million BE images, but these images were augmented and originated from only 161 patients.[8] To ensure the generalizability of our AI models, we included images from a diverse range of endoscopes spanning over a decade, adding to the breadth and longevity of our data collection.

Second, this study is the first to utilize ViT in training a classifier model for EON detection. Additionally, this study represents one of the earlier attempts to train ViT models on gastrointestinal endoscopic images. Our findings are promising, as they indicate the potential to combine diverse datasets of other endoscopic GI lesions, into a larger model using ViT, while maintaining high accuracy.

Third, this study presents stronger evidence for the effectiveness of a combined modalities model. Previously, only one study by Hashimoto *et al*. had trained such a model, but it was a pilot study based on images from only 65 patients.[24] Not only does the combined modality model improves sensitivity and overall accuracy on WLI images, but also its ability to seamlessly switch between WLI and NBI during endoscopy can enhance clinical workflow, as endoscopists toggle between these modes.

Our study has several limitations that should be taken into consideration. Compared with other published models, our ViT models were trained exclusively on still images, without incorporating videos.[8,24,25] However, the primary objective of this study is to explore the use of ViT as a novel classifier tool, therefore using still images will reduce potential confounders due to reduced image quality of frames derived from endoscopic videos. Nonetheless, this pilot model will serve as a foundation for developing a larger ViT model, trained on both images and videos from various imaging modalities derived from multiple centers.

Second, we excluded patients with indeterminate and low-grade dysplasia from our datasets due to their interobserver variability in pathological diagnosis.[26,27] Additionally, low-grade and indeterminate dysplasia may not be visually detectable using WLI and, to a lesser degree, NBI.[20] One possible solution would be to include these lesions in our future models, but their diagnosis would need to be verified by at least two pathologists. However, the challenge of detecting nonvisible indeterminate and low-grade dysplasia remains. Fortunately, the incidence of developing high-grade dysplasia or EAC remains low at 1.73 cases per 100 person-years (95% confidence interval 0.99–2.47).[28]

Third, our AI models have not been trained to perform image localization or segmentation. These functions would be particularly useful for nonexpert endoscopists to assist with identifying dysplasia with real-time "bounding boxes" or "heat maps." However, this limitation could be overcome with the birth of MedSAM, a state-of-the-art segmentation model trained on 1.5 million images across 10 imaging modalities and over 30 cancer types, which can be incorporated to our future models to automatically perform segmentation tasks or to improve the efficiency our image or video annotation.[29]

Conclusions

ViT has demonstrated high accuracy in classifying the presence or absence of EON in endoscopic images of Barrett's esophagus. Combined modalities-ViT model, trained on both NBI and WLI images, showed greater accuracy and sensitivity than the WLI-ViT model. As large Barrett's models are developed in the future or when combined with other gastrointestinal pathologies, ViT may offer a significant advantage in terms of scalability and efficiency.

Acknowledgment

Open access publishing facilitated by The University of Adelaide, as part of the Wiley - The University of Adelaide agreement via the Council of Australian University Librarians.

References

1 Barreda BF, Sanchez LJ, Misad NO *et al*. Barrett's esophagus. *Rev. Gastroenterol. Peru*. 2002; **22**: 45–68.
2 Fitzgerald RC, di Pietro M, Ragunath K *et al*. British Society of Gastroenterology guidelines on the diagnosis and management of Barrett's oesophagus. *Gut*. 2014; **63**: 7–42.
3 Weusten B, Bisschops R, Coron E *et al*. Endoscopic management of Barrett's esophagus: European Society of Gastrointestinal Endoscopy (ESGE) Position Statement. *Endoscopy*. 2017; **49**: 191–8.
4 Rustgi AK, El-Serag HB. Esophageal carcinoma. *N. Engl. J. Med.* 2014; **371**: 2499–509.
5 Schölvinck DW, van der Meulen K, Bergman JJGHM, Weusten BLAM. Detection of lesions in dysplastic Barrett's esophagus by community and expert endoscopists. *Endoscopy*. 2017; **49**: 113–20.
6 Tan JL, Chinnaratha MA, Woodman R *et al*. Diagnostic accuracy of artificial intelligence (AI) to detect early neoplasia in Barrett's esophagus: a non-comparative systematic review and meta-analysis. *Front. Med. (Lausanne)*. 2022; **9**: 890720.
7 Sharma P, Savides TJ, Canto MI *et al*. The American Society for Gastrointestinal Endoscopy PIVI (Preservation and Incorporation of Valuable Endoscopic Innovations) on imaging in Barrett's esophagus. *Gastrointest. Endosc.* 2012; **76**: 252–4.
8 Abdelrahim M, Saiko M, Maeda N *et al*. Development and validation of artificial neural networks model for detection of Barrett's neoplasia: a multicenter pragmatic nonrandomized trial (with video). *Gastrointest. Endosc.* 2023; **97**: 422–34.
9 de Groof AJ, Struyvenberg MR, Fockens KN *et al*. Deep learning algorithm detection of Barrett's neoplasia with high accuracy during live endoscopic procedures: a pilot study (with video). *Gastrointest. Endosc.* 2020; **91**: 1242–50.
10 de Groof AJ, Struyvenberg MR, van der Putten J *et al*. Deep-learning system detects neoplasia in patients with Barrett's esophagus with higher accuracy than endoscopists in a multistep training and validation study with benchmarking. *Gastroenterology*. 2020; **158**: 915–929.e4.
11 Hussein M, Puyal JG-B, Brandao P *et al*. O30 deep neural network for the detection of early neoplasia in Barrett's oesophagus. *Gut*. 2021; **70**: A17.1.
12 Struyvenberg MR, de Groof AJ, van der Putten J *et al*. A computer-assisted algorithm for narrow-band imaging-based tissue characterization in Barrett's esophagus. *Gastrointest. Endosc.* 2021; **93**: 89–98.
13 Hussein M, Lines D, González-Bueno Puyal J *et al*. Computer-aided characterization of early cancer in Barrett's esophagus on i-scan magnification imaging: a multicenter international study. *Gastrointest. Endosc.* 2023; **97**: 646–54.
14 Fockens KN, Jong MR, Jukema JB *et al*. A deep learning system for detection of early Barrett's neoplasia: a model development and validation study. *Lancet Digit Health*. 2023; **5**: e905–16.
15 Gao J, Jiang Q, Zhou B, Chen D. Convolutional neural networks for computer-aided detection or diagnosis in medical image analysis: an overview. *Math. Biosci. Eng*. 2019; **16**: 6536–61.
16 Dosovitskiy A, Beyer L, Kolesnikov A *et al*. *An image is worth 16×16 words: transformers for image recognition at scale*. arXiv. 2020.
17 Willemink MJ, Roth HR, Sandfort V. Toward foundational deep learning models for medical imaging in the new era of transformer networks. *Radiol. Artif. Intell.* 2022; **4**: e210284.
18 Russakovsky O, Deng J, Su H *et al*. *ImageNet large scale visual recognition challenge*. arXiv. 2014.
19 Kingma DP, Ba J. *Adam: a method for stochastic optimization*. arXiv preprint arXiv:14126980. 2014.
20 Sharma P, Hawes RH, Bansal A *et al*. Standard endoscopy with random biopsies versus narrow band imaging targeted biopsies in Barrett's oesophagus: a prospective, international, randomised controlled trial. *Gut*. 2013; **62**: 15–21.
21 Hajelssedig OE, Zorron Cheng Tao PL, Thompson JY *et al*. Diagnostic accuracy of narrow-band imaging endoscopy with targeted biopsies compared with standard endoscopy with random biopsies in patients with Barrett's esophagus: a systematic review and meta-analysis. *J. Gastroenterol. Hepatol.* 2021; **36**: 2659–71.
22 Sharma P, Bansal A, Mathur S *et al*. The utility of a novel narrow band imaging endoscopy system in patients with Barrett's esophagus. *Gastrointest. Endosc.* 2006; **64**: 167–75.
23 Chiam KH, Shin SH, Choi KC, Leiria F, Militz M, Singh R. Current status of mucosal imaging with narrow-band imaging in the esophagus. *Gut Liver*. 2021; **15**: 492–9.
24 Hashimoto R, Requa J, Dao T *et al*. Artificial intelligence using convolutional neural networks for real-time detection of early esophageal neoplasia in Barrett's esophagus (with video). *Gastrointest. Endosc.* 2020; **91**: 1264–1271.e1.
25 Samarasena JB, Konda VJ, Trindade AJ *et al*. Id: 3522405 detection of early esophageal neoplasia in Barrett's esophagus using real time artificial intelligence: a multicenter external video validation study. *Gastrointest. Endosc.* 2021; **93**: AB195.
26 Kerkhof M, van Dekken H, Steyerberg EW *et al*. Grading of dysplasia in Barrett's oesophagus: substantial interobserver variation between general and gastrointestinal pathologists. *Histopathology*. 2007; **50**: 920–7.
27 Vennalaganti P, Kanakadandi V, Goldblum JR *et al*. Discordance among pathologists in the United States and Europe in diagnosis of low-grade dysplasia for patients with Barrett's esophagus. *Gastroenterology*. 2017; **152**: 564–570.e4.
28 Tan JL, Heng K, Chinnaratha MA, Bulamu NB, Kaambwa B, Singh R. Incidence rates of Barrett's esophagus and esophageal adenocarcinoma: a systematic review and meta-analysis. *iGIE*. 2024; **3**: 92–103.e3.
29 Ma J, He Y, Li F, Han L, You C, Wang B. Segment anything in medical images. *Nat. Commun.* 2024; **15**: 654.