87913_Auto_Edited.docx

Retrospective Study

An artificial intelligence system for the detection of Barrett's esophagus

Barrett's esophagus AI prediction model

Ming-Chang Tsai, Hsu-Heng Yen, Hui-Yu Tsai, Yu-Kai Huang, Yu-Sin Luo, Edy Kornelius, Wen-Wei Sung, Chun-Che Lin, Ming-Hseng Tseng, Chi-Chih Wang

Abstract

BACKGROUND

Barrett's esophagus (BE), which has increased in prevalence worldwide, is a precursor for esophageal adenocarcinoma (EAC). Although there is a gap in the detection rates between endoscopic BE and histological BE in current research, we trained our artificial intelligence (AI) system with images of endoscopic BE and tested the system with images of histological BE.

AIM

We assess whether an AI system can aid in the detection of BE in our setting.

METHODS

Endoscopic narrow-band imaging (NBI) was collected from Chung Shan Medical University Hospital and Changhua Christian Hospital, resulting in 724 cases, with 86 patients having pathological results. Three senior endoscopists, who were instructing physicians of the Digestive Endoscopy Society of Taiwan, independently annotated the images in the development set to determine whether each image was classified as an

endoscopic BE. The test set consisted of 160 endoscopic images of 86 cases with histological results.

RESULTS

Six pre-trained models were compared, and EfficientNetV2B2 (accuracy: 0.85) was selected as the backbone architecture for further evaluation due to better accuracy results. In the final test, the AI system correctly identified 66 of 70 cases of BE and 85 of 90 cases without BE, resulting in an accuracy of 94.37%.

CONCLUSION

Our AI system, which was trained by NBI of endoscopic BE, can adequately predict endoscopic images of histological BE. The accuracy (ACC), sensitivity (SEN), and specificity (SPE) are 94.37%, 94.29%, and 94.44%, respectively.

Key Words: Barrett's esophagus; Artificial intelligence system; Endoscopy; Narrow-band imaging; Gastroesophageal reflux disease

Tsai MC, Yen HH, Tsai HY, Huang YK, Luo YS, Kornelius E, Sung WW, Lin CC, Tseng MH, Wang CC. An artificial intelligence system for the detection of Barrett's esophagus. *World J Gastroenterol* 2023; In press

Core Tip:

The prevalence of BE diagnosed by endoscopy significantly differs from BE diagnosed by histology (7.8% vs. 1.3%). Current research showed that image-enhanced endoscopy can only increase the detection ability for dysplasia lesions in BE. Our AI prediction system, which was trained by endoscopic BE images with the Olympus NBI system, still provided good prediction results for images of histological BE. The ACC, SEN, and SPE are 94.37%, 94.29%, and 94.44%, respectively, in the final test, which

implies that endoscopic BE images have characteristics similar to images of histological BE.

INTRODUCTION

Barrett's esophagus (BE), which is characterized by a columnar lined esophagus with documented intestinal metaplasia[1, 2], is a precursor for esophageal adenocarcinoma (EAC)[3, 4]. The prevalence of BE has been increasing in western countries[5, 6] for decades, and this trend was also recently noticed observed in Asian countries[7]. The annual incidence of EAC is approximately 0.3% in non-dysplasia BE[8], while the annual incidence of EAC increases to 0.76% in low-grade dysplasia BE[9] and to 6% in high-grade dysplasia BE[10]. Although the importance of the detection of BE in patients is recognized by endoscopists worldwide, the prevalence of BE diagnosed by endoscopy significantly differs from BE diagnosed by histology[7] (7.8% vs. 1.3%).

The methods of biopsy in Seattle protocol[11] have been under debate, as random four-quadrant biopsies at 2-cm intervals are time-consuming and intolerable under some situations. A recent study showed that targeted biopsy using image-enhanced endoscopy and molecular biomarkers can increase the detection ability for dysplasia lesions in BE[12]. The above issues imply that there is a large degree of inter-observer disagreement in the recognition and biopsy methods of BE. One previous study showed that the prevalence of histological BE is 2.6% in a health examination center data[13] but that the utilization of biopsy during endoscopy is far less in daily practice in Taiwan. We propose that an artificial intelligence (AI) system should be helpful for promoting awareness of BE in daily endoscopic practice.

AI systems are already applied in many fields of modern medicine, which started with diabetic retina observation[14, 15], diabetes mellitus care[16], and the detection and classification of Alzheimer's disease[17] and later extended to the fields of gastroenterology and endoscopy. AI systems can help detect and differentiate the classification of colon polyps[18, 19], gastric cancer[20], neoplastic and non-neoplastic small bowel lesions[21], pancreatic lesions[22, 23], BE[24-26], and EAC[24, 26, 27]. Despite recent

improvements in the AI prediction of BE and EAC, the accuracy of such models has remained approximately 90%; one of the insurmountable problems is that there are insufficient images of histologically proven BE to train AI models. Our healthcare system faced the same dilemma, and we tried to conquer this problem by establishing a relationship between images of endoscopic BE and images of histological BE.

We trained our AI model with images of endoscopic BE and endoscopic images from patients without BE and then tested our AI prediction system using cases both with and without histologically proven BE. We also compared the outcomes of our prediction system with data from similar recent studies.

MATERIALS AND METHODS

Data acquisition and preparation

Patient data were collected from Chung Shan Medical University Hospital and Changhua Christian Hospital, resulting in 724 cases, with 86 patients having complete histological results. The collection of clinical data was reviewed and approved by the institutional review board (IRB) with IRB number CS1–20075 and conducted under IRB regulations to ensure the rights and welfare of the participants.

The images utilized in this study were obtained from the hospitals' endoscopy systems, which were operated by professional physicians. The narrow-band imaging (NBI) mode in the Olympus endoscopy system was selected for capturing images. We collected both endoscopic BE images and histological BE images and corresponding image data from patients without BE to establish a study dataset.

Data annotation

The development set of this study consisted of 964 images. Three senior endoscopists who were instructing physicians of the Digestive Endoscopy Society of Taiwan independently annotated the images in the development set to determine whether each image was classified as endoscopic BE. After a majority vote among the three senior endoscopists, the final classification was determined. The test set consisted of 160 images of 86 cases, and the categorization was based on the histological results.

Dataset

The development set consisted of 704 images without BE and 260 endoscopic BE images. To avoid overfitting, the image dataset was divided using the holdout method, with 80% of the data employed as the training set and 20% of the data employed as the validation set for model training and development. The test set, which consists of complete histological result images, was used to evaluate the predictive performance of the model. The test set included 90 images without BE and 70 histological BE images. These details are shown in Table 1.

Model design

This section describes the deep learning model architecture utilized in the study to accomplish the binary classification task for BE. First, the image size was resized to 224 × 224 for inputting into the model. To address the issue of class imbalance in the dataset, this study employed a resampling technique to augment the data from the minority class. Data class balancing was performed only on the training set. The original training dataset consisted of 771 images, of which 563 images did not have BE, and 208 images showed endoscopic BE. After data augmentation, the training dataset expanded to 1,187 images, of which 563 did not show BE, and 624 images showed endoscopic BE. The effects of data augmentation are listed in the supplementary table.

In this study, a transfer learning technique was applied for model training. Six pre-trained models were compared, including EfficientNetV2B1, EfficientNetV2B2, EfficientNetV2B3, DenseNet201, ResNet50, and VGG16. Ultimately, EfficientNetV2B2 was selected as the backbone architecture for further evaluation due to its better prediction accuracy.

The deep neural network in this study employed the sigmoid function as the activation function in the output layer. An early stopping technique was utilized to monitor the maximization of validation accuracy, which allowed the model to stop training at a suitable convergence point. After completion of the training, the best model weights were saved. To further enhance the model's generalization ability, the development set was repeatedly sampled to create multiple training sets for model

training and validation, with the aim of developing an optimal AI model. A flow chart of the study design and data acquisition is shown in Figure 1.

Classification performance evaluation

The evaluation metrics applied in this study include accuracy (ACC), sensitivity (SEN), and specificity (SPE), which determine the classification performance of the model. ACC indicates the classification accuracy of the model, SEN reveals the proportion of actual positive samples correctly identified as positive, and SPE indicates the proportion of actual negative samples correctly identified as negative.

RESULTS

We collected endoscopic images of the gastroesophageal junction in a total of 724 cases, with 86 patients having complete histological results. There were 771 original images in the training set, including 563 images that did not show BE and 208 images that showed endoscopic BE (26.98% of the total number of images). A total of 193 original images were included in the validation set, including 141 images without BE and 52 images with endoscopic BE (26.94% of the validation images). Due to the imbalanced distribution in these two categories and the limited number of endoscopic BE images, we augmented the BE images in the training set to obtain a total of 624 images of endoscopic BE, which provided 52.57% of the images in the training set. These baseline characteristics are shown in Table 1.

Comparisons of different pre-trained models

We compared the prediction outcomes in different pre-trained models, which included EfficientNetV2B1, EfficientNetV2B2, EfficientNetV2B3, ResNet50, DenseNet201, and VGG16. ACC was comparable in the training process of most of the pre-trained models, with the exception of VGG16, which had an ACC of only 26.98%. The same trends were observed in the validation set, with an ACC of 78.76%, 84.97%, 78.24%, 82.90%, 77.72%, and 26.94% for EfficientNetV2B1, EfficientNetV2B2, EfficientNetV2B3, ResNet50, DenseNet201, and VGG16, respectively. For final testing, EfficientNetV2B2 achieved the best ACC of 85%, which was superior to the other pre-

trained models; hence, we chose EfficientNetV2B2 as the training model for our BE prediction system. Detailed data on SEN, SPE, and ACC for the pre-trained model comparisons are shown in Table 2.

Results of AI image-based model performance

For the training set, two images were incorrectly classified as having no BE among 208 images of endoscopic BE. For images that showed no BE, two images were incorrectly classified as having BE, while the other 561 images were correctly classified as having BE. Overall, ACC, SEN, and SPE were 99.48%, 99.04%, and 99.64%, respectively.

For the validation set, two images were incorrectly classified as having no BE among 52 images of endoscopic BE, while three images were incorrectly classified as having BE among 141 non-BE images. Overall, ACC, SEN, and SPE were 97.41%, 96.15%, and 97.87%, respectively.

For the test set, four images were incorrectly classified as having no BE among 70 images of histological BE, and five images were incorrectly classified as having BE among 90 non-BE images (as proven by histology) in the final test. The AI BE prediction system had an ACC of 94.37%, a SEN of 94.29%, and a SPE of 94.44%. Detailed information is provided in Table 3 and Figure 2.

Representative images of BE with successful detection (panel A), no BE with successful detection (panel B), BE with false detection (panel C), and no BE with false detection (panel D) are demonstrated in Figure 3.

DISCUSSION

We used only NBI images from the Olympus endoscopy system in our study instead of white light endoscopy images because our previous data showed the superiority of NBI images over white light endoscopy images for the differentiation of GERD classification[28]. However, some studies have shown that NBI images do not display differences in detecting BE esophagus compared to white light endoscopy images[29, 30], while another recent convincing study showed benefits in neoplasia

detection with an NBI system[12]. The concept that targeted biopsies using NBI enable the detection of more dysplastic areas and thus reduce the number of biopsies required is widely accepted by the Japan Esophageal Society[12, 31]. As a result, we exclusively utilized NBI still images in our AI prediction system training.

Our study is unique in that training of the pre-trained model and selected AI prediction model was performed with images of endoscopic BE rather than images of histological BE. The AI prediction system was still able to maintain a high ACC in the final prediction test that used images of histological BE. This finding implies that the images of endoscopic BE have characteristics similar to those of histological BE in our study. The final prediction results using the EfficientNetV2B2 training model showed that SEN, SPE, and ACC were 94.29%, 94.44%, and 94.37%, respectively. Previous data have highlighted that ACC is currently unsatisfied in AI systems of BE endoscopic interpretation[32], but the AI prediction of early cancer or dysplasia were impressive. Our database employed larger real-case numbers and more original images than previous studies and showed improved SEN, SPE, and ACC. We believe that the data augmentation technique limited the ACC in the final test results due to the large number of training images with similar detailed characteristics.

The most similar recent studies focused on BE neoplasia detection. Ebigbo's 2020 pilot study[33] demonstrated an ACC of 89.9% using only 14 test patients, while Hussein's 2022 study[26] improved the ACC to 93% with 118 patients. Another study by Abdelrahim in 2023, which utilized real-time video for the identification of BE dysplasia, showed a similar ACC of 92% with 75 cases.

Our study selected 771 original images from 579 patients as the training set, 193 original images from 145 patients as the validation set, and 160 images from 86 patients as the test set; the final ACC was 94.37%. Detailed information on these results is provided in Table 4. We focused on building an AI reminder system that helps endoscopists decide whether to perform a biopsy when BE is suspected instead of a system that reminds endoscopists which BE lesions may already have dysplastic changes.

Limitations

Due to the lack of awareness regarding the diagnosis of BE, cases of histological BE are relatively rare in central Taiwan. First, we initiated this study and collected endoscopic images of the lower esophagus and esophagogastric junction in two medical centers in central Taiwan. The endoscopic BE dataset was developed using the Delphi method, which involved blind voting on endoscopic images by three endoscopists, instead of histological BE images, which can lead to selection bias due to the previously known difference in prevalence between endoscopic BE and histologic BE[7] (7.8% vs. 1.3%). Second, the method of biopsy was not recorded in all cases, which meant that some of the cases involved random biopsies, while other cases were NBI-targeted biopsies. Third, the images in our database were obtained from an Olympus endoscopic system with lower esophageal still NBI pictures. It is not known whether our system can adequately perform for white light or other endoscopic systems. The main limitations were attributed to the limited cases and limited original images in our study and the limited number of cases of histological BE. However, our AI prediction system, which was trained by endoscopic BE images, still provided good prediction results for images of histological BE.

CONCLUSION

Our AI prediction system can provide good prediction results for images of histological BE obtained with Olympus NBI after training with images of endoscopic BE.

ARTICLE HIGHLIGHTS

Research perspectives

Our result implies that images of endoscopic BE share similar characteristics with images of histological BE even in the perspectives of AI system. The gap from endoscopic BE to histological BE maybe comes from biopsy or sampling bias. This

opinion needs further prospective studies to confirm. Meanwhile, a better AI prediction system for endoscopic video BE detection is an ongoing task in the near future.

Research conclusions

Our AI prediction system can provide good prediction results after training with images of endoscopic BE.

Research results

EfficientNetV2B2 (accuracy: 0.85) was selected as the backbone architecture from six training model due to better accuracy result. In the final test, the AI system obtained 94.37%, 94.29%, and 94.44%, in accuracy, sensitivity, and specificity respectively.

Research methods

Endoscopic narrow-band images of 724 cases, were collected from two medical centers at central Taiwan, with 86 patients having pathological results. Images of endoscopic BE was classified using independent annotation by three senior endoscopists, who were instructing physicians of the Digestive Endoscopy Society of Taiwan. The test set consisted of 160 endoscopic images in 86 histological BE cases.

Research objectives

To build up an AI system for the detection of endoscopic images of histological BE.

Research motivation

We want to train an artificial intelligence (AI) system to identified images of BE under endoscopic environments.

Research background

The prevalence of endoscopic Barrett's esophagus (BE) differs significantly from histological BE. We believe the endoscopic characteristics are similar with endoscopic BE and histological BE.

87913_Auto_Edited.docx

ORIGINALITY REPORT

2%

SIMILARITY INDEX

PRIMARY SOURCES

1 Shaojie Li, Yifan Zhang, Xuan Yang. "Semi-supervised Cardiac MRI Segmentation Based on Generative Adversarial Network and Variational Auto-Encoder", 2021 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), 2021
Crossref
17 words — 1%

2 Yan Chen, Song Yu, Ke Ma, Shuangyuan Huang, Guofeng Li, Siqi Cai, Longhan Xie. "A Continuous Estimation Model of Upper Limb Joint Angles by Using Surface Electromyography and Deep Learning Method", IEEE Access, 2019
Crossref
17 words — 1%

3 Seiji Saito, Kazuki Ohno, Jun Sese, Kanako Sugawara, Hitoshi Sakuraba. "Prediction of the clinical phenotype of Fabry disease based on protein sequential and structural information", Journal of Human Genetics, 2010
Crossref
14 words — < 1%

4 id.nii.ac.jp
Internet
14 words — < 1%

EXCLUDE QUOTES	ON	EXCLUDE SOURCES	< 12 WORDS
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	< 12 WORDS