学位論文

Usefulness of an artificial intelligence system for the
detection of esophageal squamous cell carcinoma
evaluated with videos simulating overlooking situation
(見落とし想定動画で検証した食道扁平上皮癌の検出における人工知能システムの有用性)

脇　幸太郎

Kotaro　Waki

指導教員

佐々木　裕　前教授

熊本大学大学院　医学教育部博士課程医学専攻　消化器内科学

田中　靖人　教授

熊本大学大学院　医学教育部博士課程医学専攻　消化器内科学

2023 年度

学　位　論　文

論文題名　: **Usefulness of an artificial intelligence system for the detection of esophageal squamous cell carcinoma evaluated with videos simulating overlooking situation**

（見落とし想定動画で検証した食道扁平上皮癌の検出における人工知能システムの有用性）

著 者 名　:　脇　幸太郎

Kotaro　Waki

指導教員名　:　佐々木 裕 前教授

熊本大学大学院 医学教育部博士課程医学専攻 消化器内科学

田中　靖人　教授

熊本大学大学院 医学教育部博士課程医学専攻 消化器内科学

審査委員名　:

消化器外科学担当教授	馬場 秀夫
放射線診断学担当教授	平井 俊範
耳鼻咽喉科・頭頸部外科学担当教授	折田 頼尚
呼吸器内科学担当教授	坂上　拓郎

2023年度

Digestive Endoscopy 2021; **33**: 1101–1109 doi: 10.1111/den.13934

Original Article

Usefulness of an artificial intelligence system for the detection of esophageal squamous cell carcinoma evaluated with videos simulating overlooking situation

Kotaro Waki,[1,2] Ryu Ishihara,[1] Yusuke Kato,[4] Ayaka Shoji,[1] Takahiro Inoue,[1] Katsunori Matsueda,[1] Muneaki Miyake,[1] Yusaku Shimamoto,[1] Hiromu Fukuda,[1] Noriko Matsuura,[1] Yoichiro Ono,[7] Kenshi Yao,[7] Satoru Hashimoto,[8] Shuji Terai,[8] Masayasu Ohmori,[9] Kyosuke Tanaka,[10] Motohiko Kato,[6] Takashi Shono,[3] Hideaki Miyamoto,[2] Yasuhito Tanaka[2] and Tomohiro Tada[4,5,11]

[1]*Department of Gastrointestinal Oncology, Osaka International Cancer Institute, Osaka,* [2]*Department of Gastroenterology and Hepatology, Faculty of Life Sciences, Kumamoto University,* [3]*Department of Gastroenterology and Hepatology, Kumamoto Chuo Hospital, Kumamoto,* [4]*AI Medical Service Inc,* [5]*Department of Surgical Oncology, Graduate School of Medicine, The University of Tokyo,* [6]*Division of Gastroenterology and Hepatology, Department of Internal Medicine, Keio University School of Medicine, Tokyo,* [7]*Department of Gastroenterology, Fukuoka University Chikushi Hospital, Fukuoka,* [8]*Division of Gastroenterology and Hepatology, Graduate School of Medical and Dental Sciences, Niigata University, Niigata,* [9]*Department of Gastroenterology, Okayama University Hospital, Okayama,* [10]*Department of Endoscopic Medicine, Mie University Hospital, Mie and* [11]*Tada Tomohiro Institute of Gastroenterology and Proctology, Saitama, Japan*

Objectives: Artificial intelligence (AI) systems have shown favorable performance in the detection of esophageal squamous cell carcinoma (ESCC). However, previous studies were limited by the quality of their validation methods. In this study, we evaluated the performance of an AI system with videos simulating situations in which ESCC has been overlooked.

Methods: We used 17,336 images from 1376 superficial ESCCs and 1461 images from 196 noncancerous and normal esophagi to construct the AI system. To record validation videos, the endoscope was passed through the esophagus at a constant speed without focusing on the lesion to simulate situations in which ESCC has been missed. Validation videos were evaluated by the AI system and 21 endoscopists.

Results: We prepared 100 video datasets, including 50 superficial ESCCs, 22 noncancerous lesions, and 28 normal esophagi. The AI system had sensitivity of 85.7% (54 of 63 ESCCs) and specificity of 40%. Initial evaluation by endoscopists conducted with plain video (without AI support) had average sensitivity of 75.0% (47.3 of 63 ESCC) and specificity of 91.4%. Subsequent evaluation by endoscopists was conducted with AI assistance, which improved their sensitivity to 77.7% ($P = 0.00696$) without changing their specificity (91.6%, $P = 0.756$).

Conclusions: Our AI system had high sensitivity for the detection of ESCC. As a support tool, the system has the potential to enhance detection of ESCC without reducing specificity. (UMIN000039645)

Key words: artificial intelligence, esophageal squamous cell carcinoma

INTRODUCTION

ESOPHAGEAL CANCER IS the seventh most common cancer and the sixth most common cause of cancer-related mortality worldwide.[1] The most common esophageal cancer in Asian countries, including Japan, is esophageal squamous cell carcinoma (ESCC).[2] The prognosis of patients with advanced ESCC is poor but improvable if the cancer is detected early.[3–7]

Esophageal squamous cell carcinoma is often diagnosed late, as superficial ESCCs are asymptomatic and changes in the mucosa are subtle.[8] Furthermore, the role of white-light imaging (WLI) endoscopy in detection is limited because of its low sensitivity.[9] In contrast, narrow-band imaging (NBI), a form of equipment-based image-enhanced endoscopy, is useful.[10] A meta-analysis that compared the detection ability of NBI and Lugol chromoendoscopy, then considered the

Corresponding: Ryu Ishihara, Department of Gastrointestinal Oncology, Osaka International Cancer Institute, 3-1-69 Otemae, Chuo-ku, Osaka 541-8567, Japan. Email: ryu1486@gmail.com
Received 3 December 2020; accepted 21 January 2021.

reference standard for diagnosing superficial ESCC, showed that NBI had better performance in diagnosing ESCC.[11] Accordingly, NBI is currently the standard for detecting superficial ESCC. However, in clinical practice, NBI's performance at detecting superficial ESCCs varies between endoscopists, and inexperienced endoscopists' detection sensitivity was found to be significantly lower than that of experts (53% vs. 100%, $P < 0.001$).[12]

Artificial intelligence systems have shown favorable performance in ESCC detection.[13–16] Their performance was initially evaluated using still images[13,14] and then using video images.[15,16] The video images used for evaluation were slow observations of the esophagus that simulated detailed inspection of high-risk populations for esophageal cancers, and they were usually taken from the front and edited for the evaluations. However, in practice, quick observation of the esophagus is usually conducted for average- or low-risk populations, and consequently, early-stage esophageal cancers may be overlooked. To evaluate the AI system's performance as a support tool, it should be evaluated in more realistic situations.

In this study, we prepared NBI and blue laser imaging (BLI) videos strictly to simulate the situation of overlooking ESCC. The AI system's performance was compared with that of endoscopists of varying experience.

METHODS

Training dataset and image annotation

WE DEVELOPED A deep learning-based AI system to detect superficial ESCCs. It was trained with endoscopic images taken via daily esophagogastroduodenoscopy. We gathered endoscopic still and video images of pathologically proven superficial ESCCs captured at Osaka International Cancer Institute, Fukuoka University Chikushi Hospital, and Niigata University Hospital in December 2005–June 2019. We also gathered images of noncancerous lesions and normal esophagi taken at Osaka International Cancer Institute in January 2009–June 2019. Noncancerous lesions included pathologically or endoscopically diagnosed esophagitis, submucosal tumor, vascular abnormality, glycogenic acanthosis, atypical epithelium (EP), and intraepithelial neoplasm. As in our previous study,[16] still images extracted from the videos were used to obtain cancer images with diverse shooting conditions (e.g., various distances, angles, and focus).

The endoscopic procedures were conducted using the following equipment: GIF-RQ260Z, GIF-FQ260Z, GIF-Q240Z, GIF-H290Z, GIF-HQ290, GIF-H260Z, GIF-XP290N, GIF-Q260J, or GIF-H290 endoscopes (Olympus, Tokyo, Japan) with the CV260 (Olympus), EVIS LUCERA CV-260/CLV-260, or EVIS LUCERA ELITE CV-290/CLV-290SL (Olympus Medical Systems) video processors; or EG-L590ZW, EG-L600ZW, or EG-L600ZW7 endoscopes (Fujifilm Co., Tokyo, Japan) and the LASEREO video endoscopic system (Fujifilm Co.). Observations using the LASEREO system used BLI, which provides images similar to those of NBI.[11,17,18]

After extracting still images from the videos, our construction dataset included 17,336 images from 1567 pathologically proven superficial ESCCs and 1461 images from 196 noncancerous lesions and normal esophagi (Fig. 1). These images included those captured by non-magnifying endoscopy (non-ME) with WLI, NBI, and BLI.

As in our previous study,[16] the images were annotated manually by delineating the boundaries and filling in the areas containing the ESCC or other disease conditions. Annotation was conducted by eight endoscopists, and all annotated images were reconfirmed by a board-certified trainer at the Japan Gastroenterological Endoscopy Society (R.I.).

Creation of validation videos

In the present study, we took videos simulating quick inspection with NBI/BLI and WLI, respectively, and created independent validation datasets. The videos were taken during esophagogastroduodenoscopy of consecutive subjects with or without superficial ESCCs by eight

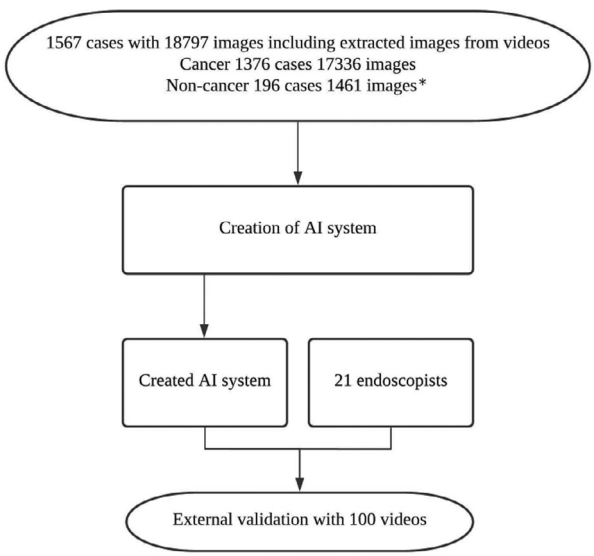

Figure 1 Construction and validation of the artificial intelligence (AI) system. *Five cases used noncancerous mucosal areas of cancer cases.

endoscopists at Osaka International Cancer Institute in October 2019–April 2020. Written informed consent for video acquisition was obtained from all subjects. Subjects with advanced ESCC and histories of surgery, radiation therapy, and endoscopic treatment for ESCC were excluded.

For video acquisition, the endoscope was inserted through the center of the esophageal lumen from the upper esophagus to the esophagogastric junction (EGJ) over 10–15 s and immediately withdrawn at a similar rate (total 20–30 s). During this period, we used only non-ME. To simulate situations in which ESCC had been overlooked, the endoscope was passed through the esophagus at a constant speed without focusing on the lesions or any particular parts. The videos were used directly for validation to prevent bias from editing.

Regarding sample size, we initially estimated a minimum of 40 ESCCs in the validation set to achieve a 95% confidence interval narrower than 20% based on AI systems' sensitivity of 90% in previous study[16]. We prepared as many ESCCs as possible to shrink the 95% confidence interval.

Evaluation of the validation dataset by the AI system and endoscopists

A detailed description of construction of an AI system is shown in Appendix S1. The purpose of this study was to compare the performance of the AI system with that of endoscopists for cancer detection in suspected lesions with quick inspection videos using only non-ME.

The trained neural network generated diagnosis scores of 0–1 for superficial ESCC. The threshold for judging each frame as cancer was ≥0.60. Conclusive diagnosis of ESCC was established when three serial video frames were judged as cancer.

We invited 21 endoscopists from five centers with varying experience levels to interpret the validation dataset. They had been diagnosing gastrointestinal cancers, including superficial ESCCs, in their daily practice. Seven endoscopists had <2 years of experience, six had 3–10 years, and eight had ≥11 years. They were asked to detect lesions suspicious for superficial ESCC. They were not informed about the proportion of cancerous lesions in the dataset.

Evaluation of the validation dataset with the addition of AI assistance

First, 21 endoscopists interpreted the plain videos without AI assistance on both WLI and NBI/BLI, followed by interpreting videos with AI assistance on NBI/BLI about 6 weeks later. Subsequent evaluation by the endoscopists was conducted using two monitors: one for plain videos and one for videos with AI support (indicating the AI system's diagnosis on the monitor, as shown in Videos S1–S4).

Statistical analysis

The main outcome measures were sensitivity and specificity for superficial ESCC. These parameters were calculated as follows: sensitivity = (number of correctly detected superficial ESCCs)/(number of total superficial ESCCs); specificity = (number of correctly diagnosed noncancerous lesions or normal esophagus videos)/(number of total noncancers or normal esophagus videos). Noncancerous lesions or normal videos that showed ≥1 false positive were judged incorrectly. We analyzed the sensitivity of plain viewing and that with AI assistance using the Wilcoxon signed rank sum test and Mann–Whitney *U*-test for continuous variables. The statistical significance threshold was $P < 0.05$. All analyses were performed using the EZR software package, version 1.27 (Saitama Medical Center, Jichi Medical University, Tochigi, Japan).

Ethics approval

This study was approved by the Ethics Committee of Osaka International Cancer Institute (no. 18149-4) and all cooperating institutions and was registered in the University Hospital Medical Network Clinical Trials Registry as number UMIN 000039645.

RESULTS

Details of validation datasets

OF THE VIDEOS, 54 were excluded before evaluation for the following reasons: esophageal observation lasted <15 or >35 s in total or focused on some lesions for more than 1 s. Finally, we prepared 100 video datasets: 50 superficial ESCCs, 22 noncancerous lesions, and 28 normal esophagi (Table 1). Sixty-three ESCCs were included in the 50 videos of ESCCs: one ESCC in 39 videos, two ESCCs in nine videos, and three ESCCs in two videos. All 63 of these lesions were pathologically proven as ESCC (53 by ESD specimens, three by surgical specimens, and seven by biopsy specimens). The 56 ESCCs with pathologic diagnosis of cancer had the following invasion depth: 39 were limited to the EP or lamina propria, 10 were invading the muscularis mucosa, and seven were invading the submucosa.

Table 1 Details of validation datasets

Patient data	
Age, years; median (range)	70 (47–88)
Sex, male/female	80/20
Number of videos, cancer/noncancer/normal	50/22/28
Lesion data	
ESCC	50 videos with 63 lesions
Location, Ut/Mt/Lt/Ae	11/37/15/0
Median tumor size in mm (range)	26 (5–140)
Macroscopic type, 0-I, IIa/0-IIb, IIc	5/58
Depth of tumor†, EP-LPM/MM/SM1/SM2–3	39/10/3/4
Noncancerous lesions and normal esophagus	50 videos
Normal esophagus	28
Vascular abnormalities	13
Reflux esophagitis	7
Papilloma	4
Esophageal varices	2

Ae, abdominal esophagus; EP, epithelium; ESCC, Esophageal squamous cell carcinoma; LPM, lamina propria mucosae; Lt, lower thoracic esophagus; MM, muscularis mucosa; Mt, middle thoracic esophagus; SM, submucosa; Ut, upper thoracic esophagus.
†Seven ESCCs diagnosed by biopsy were excluded from analysis for invasion depth.

Performance of the AI system

On NBI/BLI, the AI system detected 54 of 63 ESCCs (sensitivity, 85.7%; Figs 2–4), and false positives were detected in 30 videos (specificity, 40%) in the validation dataset. Of the 54 ESCCs detected by the AI system, six were submucosal cancers, eight were muscularis mucosa cancers, and 33 were EP/lamina propria cancers. False positive diagnoses were caused by reflux esophagitis and vascular abnormalities in six cases each, and others were caused by bubbles, fluid storage, screen blurring and shadows, or the normal EGJ (Fig. 5). On WLI, the AI system detected 49 of 63 ESCCs (sensitivity: 77.8%), and false positives were detected in 36 videos (specificity, 28%; Table S1). Additional evaluation of specificity using 50 videos is shown in Table S2.

Performance of endoscopists without AI support

The endoscopists' initial evaluation was conducted with plain videos (without AI support). On NBI/BLI, the endoscopists detected 47.3 of 63 ESCCs (sensitivity, 75.0%; range, 58.7–90.5%), and false positives were detected in 4.3 videos (specificity, 91.4%; range, 50.0–100%) in the validation dataset. The endoscopists' overall results are summarized in Table 2. False positive diagnoses were mainly caused by vascular abnormalities, shadows, and reflux esophagitis. On WLI, the endoscopists detected 35.3 of 63 ESCCs (sensitivity, 56.1%; range, 36.5–69.8%), and false positives were detected in 6.1 videos (specificity, 87.7%; range: 50–100%; Table S1).

Performance of endoscopists with AI support

On NBI/BLI, the endoscopists detected 49.0 of 63 ESCCs (sensitivity, 77.7%; range, 61.9%–87.3%), and false positives were detected in 4.2 videos (specificity, 91.6%; range, 62.0%–100%) in the validation dataset. Sensitivity for cancer diagnosis by endoscopists improved by 2.7% with the addition of AI assistance ($P = 0.00696$), and there was no decrease in specificity (91.4–91.6%, $P = 0.756$; Table 2).

The change in each endoscopist's sensitivity between plain viewing and that with AI assistance is shown in Fig. 6. Fifteen of the 21 endoscopists showed sensitivity improvements, two showed no change, and four showed decreased sensitivity. In subgroup analysis, we classified the endoscopists into two groups: those with high and low sensitivity (higher and lower than the AI system, respectively) on plain videos. The 19 endoscopists who had lower sensitivity than the AI system had a median (interquartile range, [range]) of 5 (2–14 [1–22]) years of experience as endoscopists and had conducted 2000 (850–10,000 [30–30,000]) endoscopic procedures. The two endoscopists who had higher sensitivity than the AI system had 4 and 15 years of experience as endoscopists and had conducted 3500 and 15,000 endoscopic procedures, respectively. Although the AI assistance provided no additional sensitivity improvement to the high-sensitivity group ($n = 2$), the low-sensitivity group ($n = 19$) had 3.7% additional improvement ($P = 0.000703$; Tables 3 and 4). There was no decrease in specificity with the addition of AI assistance in either group.

Characteristics of detected and missed lesions in validation dataset

The characteristics of detected and missed lesions on NBI/BLI are shown in Table 5. Nine cancers were missed by the AI system, and their median size by the AI system was significantly smaller than that of detected cancers (15 vs. 30 mm, $P = 0.01$). The AI system did not miss any cancers larger than half of the circumference of the esophagus. Fifteen cancers were missed by more than half of the

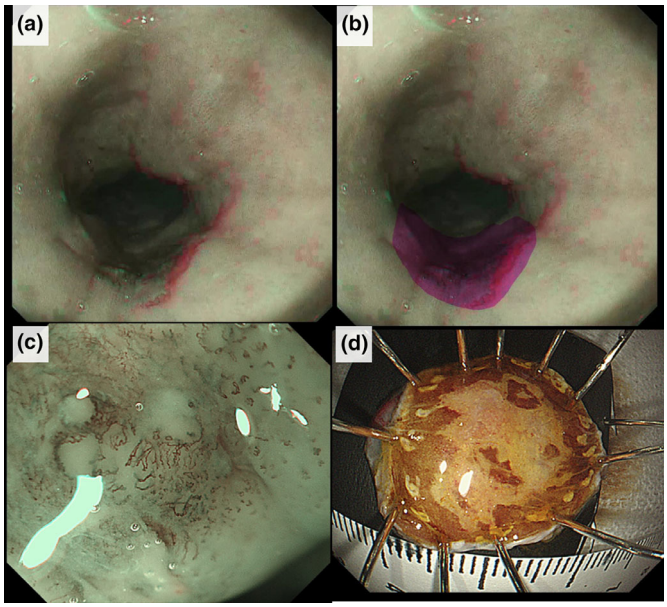

Figure 2 (with Video S1). The lesion was located on the posterior wall of the upper thoracic esophagus. The lesion was detected by the artificial intelligence (AI) system for 1 s. With AI assistance, the proportion of endoscopists who detected the lesion improved from 76.2% (without AI assistance) to 95.2% (with AI assistance). (a) Image of the lesion without AI assistance. (b) Image of the lesion with AI assistance. The lesion is indicated in pink. (c) Magnified image of the lesion with NBI. (d) The resected specimen with iodine staining. Histopathological diagnosis was cancer invading into the lamina propria.

endoscopists. The median size of the cancers that were missed by endoscopists was slightly smaller than that of the detected cancers (20 vs. 27 mm), although the difference was not significant ($P = 0.19$). More than half of the endoscopists missed two cancers larger than half of the circumference of the esophagus.

DISCUSSION

IN THIS STUDY, we examined the detection ability of an AI system using videos simulating situations in which ESCC had been overlooked. The results confirm the AI system's high sensitivity. In addition, the endoscopists' sensitivity improved with AI support without negatively impacting specificity.

We evaluated the AI system using videos simulating the overlooking situation: attention did not focus on any lesions, and the videos were used without editing. For validation, the AI system and the endoscopists attempted to detect any suspicious lesions for cancer in the video images. This approach seems suitable for evaluating the AI system's performance.

The AI system's sensitivity in validation with NBI/BLI was 85.7%, which was significantly higher than the endoscopists' average sensitivity. Therefore, the AI system has the potential to improve endoscopists' sensitivity to the same level. Conversely, the AI system's specificity was only 40%. Approximately half of the 30 false positives were related to findings in the EGJ area. Because the training dataset was dedicated to squamous EP lesions, further training of the system with EGJ images may decrease false positives. As the incidence of esophageal cancer is low and most examinees do not have cancer, high specificity (i.e., accurately diagnosing noncancerous lesions as noncancerous) is important. The endoscopists' specificity was similar with or without AI support. Therefore, our system showed no negative effects on diagnosis when used as a support tool.

In a subgroup analysis for validation with NBI/BLI, two endoscopists with higher sensitivity than the AI system showed no improvement in sensitivity with AI assistance, whereas the other 19 endoscopists with lower sensitivity than the AI system showed a statistically significant improvement. This result suggests that the AI system may be useful as a support tool for the majority of endoscopists.

The sensitivity of four endoscopists decreased with AI support during validation with NBI/BLI. We expect that the main reason for this result is that they thought the AI system had perfect sensitivity for cancer. Accordingly, they changed their diagnoses from cancer to noncancer if the AI system did not detect the lesion. Thus, the endoscopists' sensitivity declined, whereas their specificity improved considerably.

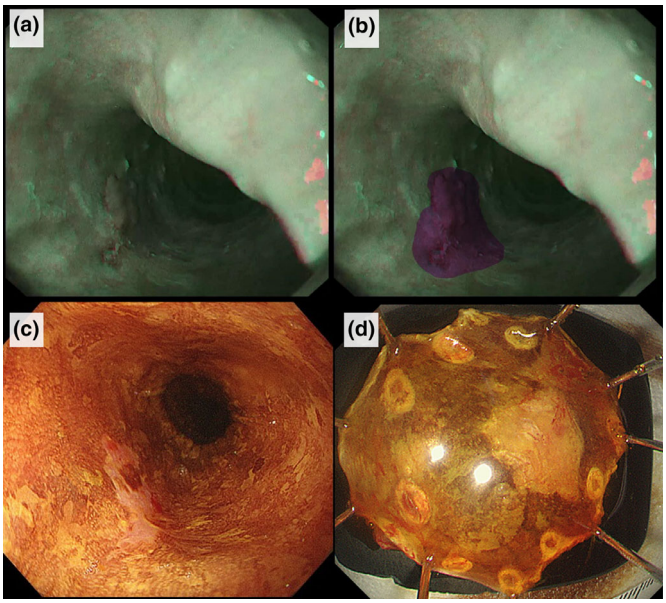

Figure 3 (with Video S2). The case achieved 100% detection rate with AI assistance. The lesion was located on the posterior wall of the middle thoracic esophagus. With AI assistance, the proportion of endoscopists who detected the lesion improved from 90.5% (without AI assistance) to 100% (with AI assistance). (a) Image of the lesion without AI assistance. (b) Image of the lesion with AI assistance. The lesion is indicated in pink. (c) The image of the lesion with Lugol chromoendoscopy. (d) The resected specimen with iodine staining. Histopathological diagnosis was cancer invading into the muscularis mucosa.

We believe that this problem can be solved by informing the endoscopists about the AI system's performance or by reconfirmation of the lesions.

Clinical trials are another method by which the AI system's performance could be evaluated. However, ESCC is not a common cancer, and the prevalence of ESCC in prospective studies has been 1.9% (2/105 cases)[19] to 10.8% (36/334 cases)[20] even in high-risk populations. Evaluating the performance of the AI system with a large number of ESCCs would require a long investigative period and be expensive, in contrast to the current study, in which we directly compared the endoscopists' diagnostic performance with that of the AI system.

This study has several limitations. First, the validation videos used in the current study did not include reconfirmation of suspicious lesions because the reconfirmation process contradicts the concept of simulating situations in which ESCC has been overlooked. This limitation is the main reason why the additional sensitivity improvement imparted by the AI assistance was only 2.5%. Because the sensitivity of the AI by itself was 85.7% in validation with NBI/BLI, we expect that further improvement in the sensitivity of AI-assisted diagnosis can be achieved by reviewing lesions that are only detected by the AI system. Second, the proportion of videos containing ESCC was high (50% of the validation dataset). The high rate of validation videos containing ESCC may have increased the endoscopists' attention levels, which may have enhanced diagnostic performance. However, in clinical practice, ESCC is much less common. ESCC's low incidence makes it difficult to maintain high attention during daily endoscopic examinations and may reduce diagnostic performance. Therefore, we expect the detection rate by endoscopists to be lower in clinical endoscopy. The AI system is not affected by such factors and may provide benefits above those seen in this study. Third, sample size of the current study may not be sufficient for detailed statistical analysis. Further study is required in the future.

In conclusion, our AI system has high sensitivity for the detection of ESCC and can increase endoscopists' sensitivity without reducing specificity. We believe that our AI system has the potential to enhance the detection of ESCC.

ACKNOWLEDGMENTS

WE THANK S. Hayashi, M. Kobashi, Y. Kono, H. Kanzaki (Okayama University Hospital), A. Hattori, Y. Umeda, H. Miura, J. Tsuboi, Y. Hamada, M. Katsurahara (Mie University Hospital), K. Inokuchi, K. Iwata, M. Mizutani, Y. Kurosawa (Keio University School of Medicine), H. Iwasaki, K. Noda (Kumamoto chuo hospital),

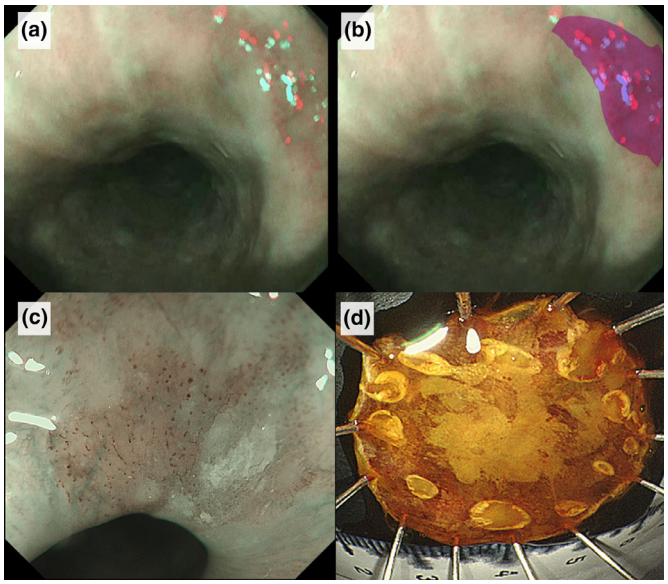

Figure 4 (with Video S3) The hard-to-detect case. The lesion was located on the right wall of the lower thoracic esophagus. The artificial intelligence (AI) system detected the lesion at the time of withdrawal. With AI assistance, the proportion of endoscopists who detected the lesion improved from 19.0% (without AI assistance) to 38.1% (with AI assistance). (a) Image of the lesion without AI assistance. (b) Image of the lesion with AI assistance. The lesion is indicated in pink. (c) Magnified image of the lesion with NBI. (d) The resected specimen with iodine staining. Histopathological diagnosis was cancer invading into the lamina propria.

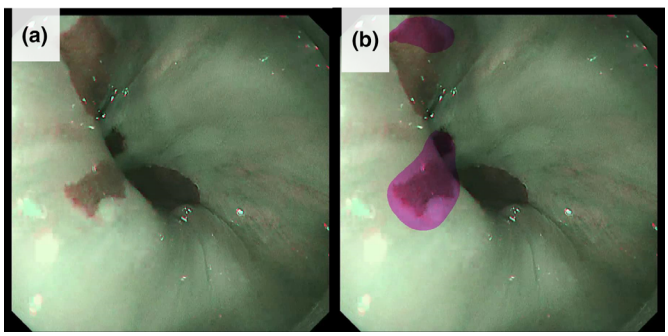

Figure 5 (with Video S4). False positive finding in the esophagogastric junction by the artificial intelligence (AI) system. AI assistance did not mislead the endoscopists because the proportion of endoscopists who detected false positive lesions decreased from 4.8% (without AI assistance) to 0% (with AI assistance). (a) Image of the false positive lesion without AI assistance. (b) Image of the false positive lesion with AI assistance. The false positive lesion is indicated in pink.

T. Fujimoto, D. Arakawa, Y. Hanazono, K. Sakisaka, R. Gushima (Kumamoto University Hospital), for validation. We also thank, Peter Morgan, PhD, from Edanz Group (https://en-author-services.edanz.com/ac), for editing a draft of this manuscript.

CONFLICT OF INTEREST

AUTHOR T.T. IS a shareholder of AI Medical Service Inc. All other authors have no financial relationships relevant to this publication to disclose.

Table 2 Sensitivities and specificities of endoscopists with and without artificial intelligence (AI) assistance

	Plain	AI assist	*P*-value
Sensitivity			
Average	0.750	0.777	0.00696
Range	0.587–0.905	0.619–0.873	-
Specificity			
Average	0.914	0.916	0.756
Range	0.500–1.000	0.620–1.000	-

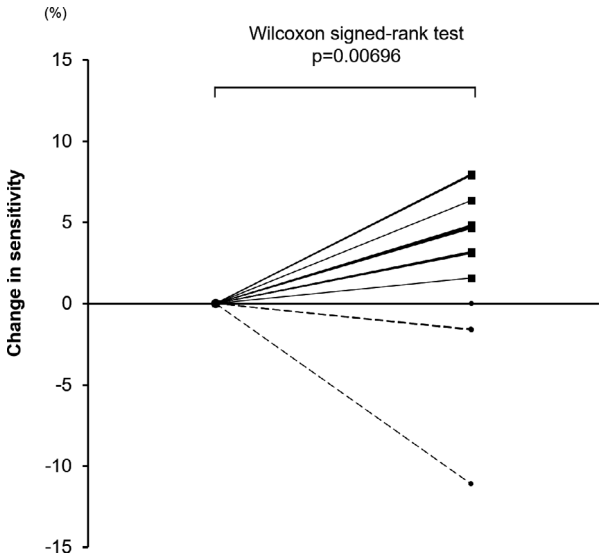

Figure 6 Changes in sensitivity of endoscopists with and without artificial intelligence (AI) assistance. Fifteen of the 21 endoscopists showed an improvement, two showed no change and four showed a decrease.

Table 3 Changes in sensitivity of endoscopists with and without artificial intelligence (AI) assistance (low-sensitivity group; $n = 19$)

	Plain	AI assist	*P*-value
Sensitivity			
Average	0.735	0.772	0.000703
Range	0.587–0.825	0.619–0.873	-
Specificity			
Average	0.914	0.914	0.909
Range	0.500–1.000	0.620–1.000	-

Table 4 Changes in sensitivity of endoscopists with and without artificial intelligence (AI) assistance (high-sensitivity group; $n = 2$)

	Plain	AI assist	*P*-value
Sensitivity			
Average	0.889	0.825	0.5
Range	0.873–0.905	0.794–0.857	-
Specificity			
Average	0.920	0.940	1
Range	0.900–0.940	0.940	-

Table 5 ESCCs detected and missed by the artificial intelligence (AI) system and endoscopists according to validation with NBI/BLI

	AI system		Detected or missed by more than half of endoscopists	
	Missed ($n = 9$)	Detected ($n = 54$)	Missed ($n = 15$)	Detected ($n = 48$)
Location				
Ut	1	10	3	8
Mt	7	30	9	28
Lt	1	14	3	12
Median tumor size	15	30	20	27
(IQR)	(15–20)	(17.3–40)	(15–30)	(15.8–40)
[range]	[5–30]	[7–140]	[5–40]	[7–140]
Macroscopic type				
0-I, IIa	0	5	0	5
0-IIb, IIc	9	49	15	43
Circumferential extent				
<1/2	9	34	13	30
≥1/2	0	20	2	18
Depth of tumor[†]				
EP-LPM	6	33	10	29
MM	2	8	4	6
SM1	1	2	1	2
SM2–3	0	4	0	4

EP, epithelium; ESCC, esophageal squamous cell carcinoma; IQR, interquartile range; LPM, lamina propria mucosae; Lt, lower thoracic esophagus; MM, muscularis mucosa; Mt, middle thoracic esophagus; SM, submucosa; Ut, upper thoracic esophagus.
[†]Seven ESCCs diagnosed by biopsy were excluded from analysis for invasion depth.

FUNDING INFORMATION

NONE.

REFERENCES

1 Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018; **68**: 394–424.
2 Rustgi AK, El-Serag HB. Esophageal carcinoma. *N Engl J Med* 2014; **371**: 2499–509.
3 Kodama M, Kakegawa T. Treatment of superficial cancer of the esophagus: A summary of responses to a questionnaire on superficial cancer of the esophagus in Japan. *Surgery* 1998; **123**: 432–9.
4 Shimizu Y, Tsukagoshi H, Fujita M, Hosokawa M, Kato M, Asaka M. Long-term outcome after endoscopic mucosal resection in patients with esophageal squamous cell carcinoma invading the muscularis mucosae or deeper. *Gastrointest Endosc* 2002; **56**: 387–90.
5 Katada C, Muto M, Momma K *et al*. Clinical outcome after endoscopic mucosal resection for esophageal squamous cell carcinoma invading the muscularis mucosae–A multicenter retrospective cohort study. *Endoscopy* 2007; **39**: 779–83.
6 Yamamoto S, Ishihara R, Motoori M *et al*. Comparison between definitive chemoradiotherapy and esophagectomy in patients with clinical stage I esophageal squamous cell carcinoma. *Am J Gastroenterol* 2011; **106**: 1048–54.
7 Yamashina T, Ishihara R, Nagai K *et al*. Long-term outcome and metastatic risk after endoscopic resection of superficial esophageal squamous cell carcinoma. *Am J Gastroenterol* 2013; **108**: 544–51.
8 Tachimori Y, Ozawa S, Numasaki H *et al*. Comprehensive registry of esophageal cancer in Japan, 2012. *Esophagus* 2019; **16**: 221–45.
9 Hashimoto CL, Iriya K, Baba ER *et al*. Lugol's dye spray chromoendoscopy establishes early diagnosis of esophageal cancer in patients with primary head and neck cancer. *Am J Gastroenterol* 2005; **100**: 275–82.
10 Muto M, Minashi K, Yano T *et al*. Early detection of superficial squamous cell carcinoma in the head and neck region and esophagus by narrow band imaging: A multicenter randomized controlled trial. *J Clin Oncol* 2010; **28**: 1566–72.
11 Morita FH, Bernardo WM, Ide E *et al*. Narrow band imaging versus lugol chromoendoscopy to diagnose squamous cell carcinoma of the esophagus: A systematic review and meta-analysis. *BMC Cancer* 2017; **17**: 54.
12 Ishihara R, Takeuchi Y, Chatani R *et al*. Prospective evaluation of narrow-band imaging endoscopy for screening of esophageal squamous mucosal high-grade neoplasia in experienced and less experienced endoscopists. *Dis Esophagus* 2010; **23**: 480–6.
13 Horie Y, Yoshio T, Aoyama K *et al*. Diagnostic outcomes of esophageal cancer by artificial intelligence using convolutional neural networks. *Gastrointest Endosc* 2019; **89**: 25–32.
14 Ohmori M, Ishihara R, Aoyama K *et al*. Endoscopic detection and differentiation of esophageal lesions using a deep neural network. *Gastrointest Endosc* 2020; **91**: 301–9.e1.
15 Guo L, Xiao X, Wu C *et al*. Real-time automated diagnosis of precancerous lesions and early esophageal squamous cell carcinoma using a deep learning model (with videos). *Gastrointest Endosc* 2020; **91**: 41–51.
16 Fukuda H, Ishihara R, Kato Y *et al*. Comparison of performances of artificial intelligence versus expert endoscopists for real-time assisted diagnosis of esophageal squamous cell carcinoma (with video). *Gastrointest Endosc* 2020; **92**: 848–55.
17 Kaneko K, Oono Y, Yano T *et al*. Effect of novel bright image enhanced endoscopy using blue laser imaging (BLI). *Endosc Int Open* 2014; **2**: E212–9.
18 Tomie A, Dohi O, Yagi N *et al*. Blue laser imaging-bright improves endoscopic recognition of superficial esophageal squamous cell carcinoma. *Gastroenterol Res Pract* 2016; **2016**: 6140854.
19 Kawai T, Takagi Y, Yamamoto K *et al*. Narrow-band imaging on screening of esophageal lesions using an ultrathin transnasal endoscopy. *J Gastroenterol Hepatol* 2012; **27**(Suppl 3): 34–9.
20 Gruner M, Denis A, Masliah C *et al*. Narrow-band imaging versus Lugol chromoendoscopy for esophageal squamous cell cancer screening in normal endoscopic practice: Randomized controlled trial. *Endoscopy*. Published online: 22 Jul 2020; DOI: 10.1055/a-1224-6822.

SUPPORTING INFORMATION

ADDITIONAL SUPPORTING INFORMATION may be found in the online version of this article at the publisher's web site.

Appendix S1 Construction of an AI system.

Table S1 Sensitivities and specificities of the AI system and endoscopists at diagnosis on the basis of white-light imaging.

Table S2 Details on dataset for additional validation of noncancerous lesions and normal esophagus by AI system and specificity results.

Video S1 (Figure 2 case). The lesion on the posterior wall of the upper esophagus. The lesion was detected by the AI system for 1 s.

Video S2 (Figure 3 case). The case achieved 100% detection rate with AI assistance. The lesion was located on the posterior wall of the middle thoracic esophagus.

Video S3 (Figure 4 case). The hard-to-detect case. The lesion was located on the right wall of the lower thoracic esophagus.

Video S4 (Figure 5 case). False positive finding in the esophagogastric junction by the AI system.