Artificial intelligence system for detecting superficial laryngopharyngeal cancer with high efficiency of deep learning

Artificial intelligence system for detecting superficial laryngopharyngeal cancer with high efficiency of deep learning

Atsushi Inaba[1),2)], M.D., Keisuke Hori[1)], M.D., Yusuke Yoda[1),3)], M.D., Hiroaki Ikematsu[1),4)], M.D., Hiroaki Takano[3)], B.Arch., Hiroki Matsuzaki[3)], M.E., Yoshiki Watanabe[5)], M.E., Nobuyoshi Takeshita[3)], M.D., Toshifumi Tomioka[7)], M.D., Genichiro Ishii[2),6)], M.D., Satoshi Fujii[6)], M.D., Ryuichi Hayashi[7)], M.D., Tomonori Yano[1), 3)], M.D.

[1)]Department of Gastroenterology and Endoscopy, National Cancer Center Hospital East, Kashiwa, Chiba, Japan

[2)]Course of Advanced Clinical Research of Cancer, Juntendo University Graduate School of Medicine, 2-1-1 Hongo, Bunkyo-ku, Tokyo, Japan

[3)]Medical Device Innovation Center, National Cancer Center Hospital East, Kashiwa, Chiba, Japan

[4)]Division of Science and Technology for Endoscopy, Exploratory Oncology Research & Clinical Trial Center, National Cancer Center East, Kashiwa, Chiba, Japan

[5)]Department of Medical Information, National Cancer Center Hospital East, Kashiwa, Chiba, Japan

[6)]Division of Pathology, Exploratory Oncology Research & Clinical Trial Center, National Cancer Center East, Kashiwa, Chiba, Japan

[7)]Department of Head and Neck Surgery, National Cancer Center Hospital East, Kashiwa, Chiba, Japan

Corresponding Author:

Keisuke Hori, M.D.

Department of Gastroenterology and Endoscopy, National Cancer Center Hospital East, 6-5-1, Kashiwanoha, Kashiwa, Chiba, 277-8577, Japan.

E-mail: khori@east.ncc.go.jp

Office telephone number: 04-7133-1111

Office fax number: 04-7134-6928

Financial support: This study was supported by The National Cancer Center Research and Development Fund (29-A-10). The sponsor had no role in the design of the study, data collection, analysis, and interpretation, writing of the manuscript, or decision to submit for publication. The sponsor had no access to the raw data. All authors had access to the raw data. The corresponding author had full access to all of the data and has the

final responsibility to submit for publication.

Short title: Artificial intelligence model for SLPC detection

Keywords: Superficial laryngopharyngeal cancer, Artificial intelligence, Object detection, Endoscopy, Narrow band imaging,

Conflict of interest: All authors declared no conflict of interest.

Author contributions: Atsushi Inaba, Keisuke Hori, and Tomonori Yano contributed to data collection, study design, drafting, and manuscript writing. Hiroaki Takano, Hiroki Matsuzaki, Yoshiki Watanabe, and Nobuyoshi Takeshita contributed to data analysis and interpretation. Yusuke Yoda, Hiroaki Ikematsu, and Toshifumi Tomioka contributed to data collection and interpretation. Satoshi Fujii, Genichiro Ishii, and Ryuichi Hayashi: contributed to data collection and assembly of data. Tomonori Yano gave final approval of the manuscript. All authors contributed to the revision of the manuscript and approved the final draft submitted.

ABSTRACT

Background: There are no published reports evaluating the ability of artificial intelligence (AI) in the endoscopic diagnosis of superficial laryngopharyngeal cancer (SLPC). We presented our newly developed diagnostic AI model for SLPC detection.

Methods: We used RetinaNet for object detection. SLPC and normal laryngopharyngeal mucosal images obtained from narrow-band imaging were used for the learning and validation datasets. Each independent dataset comprised 400 SLPC and 800 normal mucosal images. The diagnostic AI model was constructed stage-wise and evaluated at each learning stage using validation datasets.

Results: In the validation datasets (100 SLPC cases), the median tumor size was 13.2 mm; flat/elevated/depressed types were found in 77/21/2 cases. Sensitivity, specificity, and accuracy improved each time a learning image was added and were 95.5%, 98.4%, and 97.3%, respectively, after learning all SLPC and normal mucosal images.

Conclusions: The novel AI model is helpful for detection of laryngopharyngeal cancer at an early stage.

INTRODUCTION

Laryngopharyngeal cancer (LPC) is the 13th most common cancer worldwide, with estimated 480,000 new cases annually.[1] Most LPC cases are detected at the advanced stage, which require invasive treatment (e.g., complete laryngectomy) that reduces patients' quality of life including loss of speech or swallowing difficulties postoperatively.[2, 3] Although early cancer detection offers the best prognosis, there is no effective mass screening program for LPC, unlike gastrointestinal cancer.[4, 5] Recent studies have reported that narrow-band imaging (NBI) endoscopy is useful for detecting superficial LPC (SLPC).[6-10] Moreover, advances in endoscopic treatment in recent years such as in endoscopic mucosal resection, endoscopic submucosal dissection, and endoscopic laryngo-pharyngeal surgery (ELPS) have allowed less invasive treatment of early stage cancers.[11-12] Therefore, gastroenterologists have come to perform laryngopharyngeal observation with regular esophagogastroduodenoscopy (EGD) using NBI for high-risk patients especially those with esophageal squamous cell carcinoma (SCC).[13-15] However, laryngopharyngeal region observation is difficult due to pharyngeal reflexes such as coughing and vomiting, and anatomical complexity. Moreover, inexperienced gastrointestinal physicians find it difficult to diagnose SLPC accurately,

and biopsy procedures in this region require technical proficiency compared with those in the gastrointestinal tract[16]. Therefore, a real-time diagnosis supporting the system may be helpful for endoscopists in SLPC detection.

In recent years, there have been several reports about the impact of artificial intelligence (AI) for the detection or precise diagnosis of esophageal cancer, gastric cancer, and colon polyps in the endoscopy field.[17-19] However, there are no published reports evaluating the ability of AI in the endoscopic diagnosis of SLPC. In the context of medical imaging, deep learning can become a powerful supportive tool that can interpret medical images based on an accumulated set of unique algorithms. Quite recently, evidence of a faster lesion detection approach is being reported due to advances in algorithms and metastasis learning in glomerular lesion and gastric cancer detection.[20, 21] This study aimed to construct and evaluate an AI model for SLPC detection by using a highly efficient learning system.

MATERIALS AND METHODS

Preparation of learning and validation image datasets

For this single-center retrospective study, we obtained EGD images via NBI from

October 2010 to June 2017 from the National Cancer Center Hospital East to develop an algorithm for SLPC detection.

To prepare the learning dataset, we collected 400 images of SLPC that were histologically proven to be SCC from 85 patients (100 cases), of whom preoperative examinations were performed prior to ELPS, and 800 images of the normal laryngopharyngeal mucosa from 100 patients (100 cases).

To evaluate the diagnostic accuracy of the constructed AI model, we prepared an independent validation dataset. We collected 400 images of SLPC that were histologically confirmed to be SCC from 89 patients (100 cases), for whom screening was performed, and 800 images of the normal laryngopharyngeal mucosa from 100 patients (100 cases).

All SLPC cases included at least four non-magnified or moderately magnified narrow-band images of sufficient quality, excluding poor quality images resulting from halation, blur, defocus, mucus, and poor insufflation of the air. All normal cases included the soft palate, oropharynx posterior wall, hypopharynx posterior wall, piriform sinus, and glottis region that were taken by non-magnified NBI. We extracted four images from each SLPC case and eight images from each normal case. We collected learning images from October 2010 to March 2016 and validation images from April 2016 to June 2017 (Fig. 1).

All images were captured using standard endoscopes (GIF-H290Z, GIF-H290, GIF-

H260Z, GIF-H260, GIF-Q260; Olympus Medical Systems, Co, Ltd, Tokyo, Japan) and standard endoscope video systems (EVIS LUCERA CV-260/CLV-260, EVIS LUCERA ELITE CV-290/CLV-290SL; Olympus Medical System).

All SLPC images were marked manually using blue rectangular frames by one author (A.I.), who was a non-expert endoscopist. Moreover, these images were checked by another author (K.H.), who was a well-experienced endoscopist with 17 years of experience and had performed approximately 3000 examinations at a high-volume cancer center.

Convolutional neural network algorithm construction

To construct an AI-based diagnostic system, we used the RetinaNet network architecture, which is among the latest deep learning algorithms for object detection with a high accuracy.[22] RetinaNet is a single, unified network composed of a backbone network and two task-specific subnetworks. The backbone is responsible for computing a convolutional feature map over an entire input image and is called the ResNet-50 network architecture.[23] The first and second subnets perform convolutional object classification on the backbone's output and the convolutional bounding box regression, respectively. The two subnetworks feature a simple design that we propose specifically for one-stage,

dense detection. RetinaNet uses a Feature Pyramid Network (FPN) as its backbone network, and FPN enables more accurate recognition of location information and more accurate object detection. There is a class imbalance, which is the primary obstacle preventing the one-stage object detectors from surpassing the top-performing, two-stage methods. To address this, RetinaNet is equipped with a focal loss that applies a modulation term to the cross-entropy loss to focus on hard negative examples.[22]

Outcome measures of artificial intelligence diagnosis

After constructing the diagnostic AI model using the learning dataset, we evaluated the performance using the validation dataset. The diagnostic AI model was constructed stage-wise, by gradually increasing the number of learning images (100-200-300-400 images of the SLPC and 200-400-600-800 images of normal laryngopharyngeal mucosa). The validation for the detection was conducted at each point of the learning stage by using the validation dataset consisting of 400 SLPC and 800 normal laryngopharyngeal mucosal images (Fig. 1).

In the validation stage, when the diagnostic AI model detected SLPC from the input data of the validation images, a red rectangular frame was displayed on the endoscopic image to surround the lesion that was predicted as SLPC (Fig. 2).

To evaluate the diagnostic performance of the diagnostic AI model for SLPC detection, we used the following definitions.

- We used intersection over union (IoU) to determine whether AI was able to identify SLPC correctly. IoU is computed as the ratio of the area of overlap divided by the area of union. In our case, IoU was the area of the common part between the red rectangle (AI diagnosed as cancer) and blue rectangle (an endoscopist manually marked the cancer location) divided by the union of the region of the red rectangle and the region of the blue rectangle (Fig. 2). We judged that AI was able to correctly diagnose SLPC only when IoU shows recording of 0.40 or more.
- For an image with normal laryngopharyngeal mucosa, when the diagnostic AI model identified noncancerous lesions as cancer, it was judged as a false-positive result.
- Contrarily, for an image with SLPC findings, when the diagnostic AI model identified the SLPC lesions with IoU < 0.40 or as completely noncancerous, it was judged as a false-negative result.

Statistical analysis

All continuous variables are expressed as medians with ranges. A two-by-two table was

constructed using the predicted and actual outcomes to calculate the different domains of sensitivity, specificity, accuracy, and the receiver operating characteristic (ROC) curve, in the diagnostic test. We conducted statistical analyses using Fisher's exact test. The sensitivity, specificity, accuracy, positive predictive value (PPV), and negative predictive value (NPV), false-positive rate, and false-negative rate of the diagnostic AI model to detect SLPC were calculated as follows:

Sensitivity (%) = Result A / Test A × 100

Specificity (%) = Result B / Test B × 100

Accuracy (%) = (Result A+Result B) / (Test A+Test B) × 100

PPV (%) = Detected number of correct SLPC lesions/Number of lesions that were diagnosed as SLPC by the diagnostic AI model × 100

NPV (%) = Detected number of correct normal mucosal lesions/Number of lesions that were diagnosed as normal mucosal lesions by the diagnostic AI model × 100

False-positive rate (%) = 100 – Specificity (%)

False-negative rate (%) = 100 – Sensitivity (%)

Result A: Number of images in which the AI model correctly detected SLPC with an IoU recording of 0.40 or more.

Test A: Actual number of SLPC images in the validation dataset

Result B: Number of images in which the AI model correctly diagnosed the normal mucosa.

Test B: Actual number of normal mucosal images in the validation dataset

Pathological diagnosis

Macroscopic types of SLPC lesions were assessed according to the Japanese Classification of Head and Neck Cancer, edited by the Japan Society for Head and Neck Cancer. Histopathologically, SLPC is defined as tumor cell infiltration remaining in the subepithelial layer and not extending to the proper muscle layer.[24]

Ethics

This study used a retrospective design, and patient's personal information were removed. The protocol was approved by the Institutional Review Board of the National Cancer Center East (2017-090).

RESULTS

Characteristics of patients and lesions in the learning dataset

The median tumor size was 13 mm, and all cases were of the superficial type. Of the 85 patients with SLPC lesions, 73 had 1 lesion each, 10 had 2 lesions each, 1 had 3 lesions each, and 1 had 4 lesions each, totaling 100 SLPC cases. Regarding the macroscopic types, 0-IIa, 0-IIb, and 0-IIc types were found in 7, 93, and 0 cases, respectively. Twenty-six cases had treatment history for another LPC, and five of the patients with these cases had a history of receiving radiotherapy. All cases were histopathologically diagnosed as SCC (Table 1).

Characteristics of patients and lesions in the validation dataset

The median tumor size was 12 mm, and all cases were of the superficial type. Of the 89 patients with SLPC lesions, 79 had 1 lesion each, 9 had 2 lesions each, and 1 had 3 lesions each, totaling 100 SLPC cases. Regarding the macroscopic types, 0-IIa, 0-IIb, and 0-IIc types were found in 21, 77, and 2 cases, respectively. Nine patients with these cases had treatment history for another LPC and seven of these had a history of receiving radiotherapy. All cases were histopathologically diagnosed as SCC (Table 1).

Trends in sensitivity, specificity, and accuracy rates on an image basis at each

learning stage

The sensitivity, specificity, and accuracy rates at each learning stage improved each time a learning image of the SLPC was added (100-image SLPC: 97.8%, 66.4%, and 76.8%; 200-image SLPC: 99.3%, 70.0%, and 79.8%; 300-image SLPC: 99.3%, 69.6%, and 79.5%; 400-image SLPC: 98.5%, 83.6%, and 88.6%, respectively). Further improvement in specificity was confirmed through learning stages 4-8, which included the normal mucosal images. Ultimately, the AI diagnosis of the images after learning 400 SLPC and 800 normal mucosal images also had a sensitivity of 95.5%, specificity of 98.4%, and accuracy of 97.3% (Table 2).

Artificial intelligence diagnosis for each image

After learning 400 SLPC and 800 normal mucosal images, the diagnostic AI model correctly diagnosed 382 out of the 400 SLPC images. Moreover, in the normal mucosal images, the diagnostic AI model correctly diagnosed 787 out of the 800 normal mucosal images (Table 3). The diagnostic AI model showed a PPV of 96.7% (382/395 images), and NPV of 98.0% (787/803 images). The examples of images that AI correctly diagnosed as SLPC are shown in Fig. 3.

Learning curve on sensitivity, specificity, and accuracy in each learning step (SLPC 100-200-300-400 and 400 SLPC+800 normal mucosal images) and ROC curve after learning all images (400 SLPC+800 normal mucosal images) were created based on these results (Fig. 4). At the SLPC image-only learning stage, the sensitivity was high, but the specificity was low. However, the specificity improved with additional learning of the normal mucosal images. Finally, the accuracy was reported as 97.4% and area under the ROC curve (AUC) was 0.98.

Examination of the effect of the clinical characteristics on the sensitivity of diagnostic artificial intelligence model

The sensitivity of diagnostic AI model by clinical characteristics are shown in Table 4.

We statistically analyzed difference in sensitivity for some clinical factors (tumor size, tumor macroscopic type, history of radiotherapy, endoscopy system, and endoscopic magnification) that might affect the sensitivity of diagnostic AI model (Table 5).

In tumor size, we examined the difference in the sensitivity in 2 groups: tumors less than 10 mm and tumors larger than 10 mm. The sensitivity for tumors less than 10 mm was 93.5%, and those larger than 10 mm was 97.0% (p value = 0.141). With tumor macroscopic type, we examined the difference in the sensitivity in 3 groups (0-IIa, 0-IIb,

and 0-IIc). The sensitivity for 0-IIa was 92.9%, 0-IIb was 96.8%, and 0-IIc was 75.0% (p value = 0.01). When the significance level was adjusted using the Bonferroni method, no significant difference was found among the 3 groups. For the location of the tumor, we calculated the sensitivity of diagnostic AI model by dividing it into the 8 sections in the laryngopharyngeal region. The diagnostic AI model showed high sensitivity regardless of the location.

There were 7 patients (28 images) with a history of receiving radiotherapy. The diagnostic AI model correctly detected 27 out of the 28 SLPC images, and the sensitivity was 96.4%. The sensitivity of the SLPC images where the patients never received radiotherapy was 95.4% (p value = 1.000).

The sensitivity of the SLPC images obtained with the EVIS LUSERA CV-260/CLV260 and that obtained with the EVIS LUCERA ELITE CV-290/CLV-290SL were 92.5% and 98.5%, respectively (p value = 0.006).

Of the 100 SLPC cases and the 400 images, 134 non-magnified SLPC images were collected from 76 cases, and 266 magnified SLPC images were collected from 93 cases. The diagnostic AI model correctly detected 261 out of the 266 magnified SLPC images (98.5%), and 121 out of 134 non-magnified SLPC images (90.3%), respectively (p value = 0.001).

Causes of false positives and false negatives

The causes of false positives and false negatives are shown in Table 6. The diagnostic AI model diagnosed 13 out of the 800 normal mucosal images as cancer, and the false-positive rate was 1.6%. The most frequent cause of a false positive was the hard palate, accounting for 69% (9/13) of all false positives. Several normal structures also caused false positives, including the tongue (2/13), posterior wall of the oropharynx (1/13), and the saliva (1/13) (Fig. 5).

Contrarily, the diagnostic AI model diagnosed 18 out of the 400 SLPC images as non-cancer, and the false-negative rate was 4.5%. Nearly half of the false-negative images were due to the small difference in color tone in the surrounding normal mucosa. The other common cause of a false negative was the blood vessel pattern, which appeared as dots that were obscured by halation. Moreover, the diagnostic AI model could not accurately identify the lesion depicted in white tone in NBI (Fig. 5).

DISCUSSION

In this study, we established an algorithm of AI diagnosis for SLPC. To the best of our knowledge, this is the first report that evaluated the ability of AI to detect SLPC in the

field of endoscopy. Deep learning allows computational models that are composed of multiple processing layers to learn representations of data with multiple levels of abstraction.[25] However, a large amount of learning data is required to construct a detection model of neoplastic diseases by a convolutional neural network . Given that SLPC has a low detection frequency and few opportunities to obtain suitable images, it was necessary to construct a highly efficient learning model. In our study, we trained a deep learning model using a small number of images for learning, which showed a highly efficient learning scheme. One reason that enabled efficient learning was the use of NBI images that could easily identify the lesions. Another reason was the use of RetinaNet, which is one of the most recent algorithms. Recently, pre-trained AI by the ImageNet has showed favorable results for detecting neoplastic lesions in gastrointestinal endoscopy.[17-19] In recent years, lesion detection for gastrointestinal neoplasms with a single-stage detector such as Single shot multibox detector (SSD) and You Only Look Once (YOLO) has been reported.[26-28] Although these detectors showed a high detection rate, their accuracy was low. Therefore, previous reports using SSD required a large number of images for training, and the false-positive rate was high.[17, 18] RetinaNet was used for object detection and enabled one-stage detection with a higher accuracy than the previous detector using one-stage and two-stage methods, by adjusting the class imbalance and adopting focal loss.[22]

In our report, the combination of the use of NBI images, pre-trained AI on the ImageNet, and state-of-the-art algorisms revealed that the limited number of learning images provides favorable results.

In our study, instead of having the AI model learn all images at one time as described in previous reports, a fixed number of images were learned gradually and evaluated at each learning stage.[17-19] Using this method, we examined the transition of the diagnostic ability of AI and evaluated the sensitivity, specificity, and accuracy rates at each stage. In our system, when AI learned only the SLPC images, although high sensitivity could be achieved, the specificity was low and the false-positive rate was high. However, when learning a normal mucosal image in addition to the SLPC image, the false-positive rate decreased and high specificity was achieved. In our study, we also aimed to create an AI model with almost a similar sensitivity and lesser false positives as a well-experienced endoscopist. Our AI model achieved a sensitivity of 95.5% on the image-based analysis, and it was 100% on a patient-based analysis.

In addition, learning of the normal mucosal images enhanced the creation of an AI model with a low false positive rate for the normal mucosa. The image-based false-positive rate was also very low at 1.6%. Based on this result, we were able to show the possibility of creating a high accuracy AI model for detecting SLPC using a limited number of learning

images. There is a potential to expand this method to train AI in diagnosing other laryngopharyngeal lesions efficiently with few images. Clinical application should be possible in the future. This high-performing AI algorithm will contribute not only to helping general endoscopists to detect SLPC, but also to encourage diagnostic confirmation of the endoscopists, which could reduce unnecessary biopsies of laryngopharyngeal lesion.

We statistically analyzed difference in sensitivity for some clinical factors (tumor size, tumor macroscopic type, history of radiotherapy, endoscopy system, and endoscopic magnification). The diagnostic AI model showed a high sensitivity regardless of the tumor size and the macroscopic type. In this study SLPC that occurred after radiotherapy were also included, but our AI model showed high sensitivity for these lesions. There was the significant difference in the sensitivity for lesions between the endoscopy system in the present study. EVIS LUCERA ELITE CV-290 / CLV-290SL is the latest equipment and will be widespread and high brightness images can be obtained. Therefore, high brightness images captured using latest system will be larger in the future, and further improvement in the diagnostic performance of AI model can be expected. Moreover, our AI model tended to make it difficult to recognize lesions in non-magnified images. The diagnostic AI model could not detect correctly 13 non-magnified SLPC images (11 cases).

In 7 of the 11 cases, the diagnostic AI model could not be detected for all non-magnified SLPC images (1 to 4). However, in all these 7 cases, in addition to the magnified SLPC images of the same lesion, the diagnostic AI model detected SLPC correctly. Hence, the sensitivity of SLPC was improved by using the magnified endoscopic images. Although the sensitivity in the non-magnified endoscopic images was lower than in the magnified endoscopic images; however, the sensitivity was 90% or more.

Although we were able to develop such a high-performing AI algorithm, we had 13 false-positive and 18 false-negative images. In the false-positive images, AI tends to identify SLPC when the background mucosa of the lesion exhibits a distinct brownish area, or when expanded dotted blood vessels or vessels running in a loop shape are seen. When one of these characteristics is seen on the normal mucosa, the model may diagnose this as cancer. In the false-negative images, in areas prone to being in the shadow of the epiglottis and arcuate folds, such as pear-shaped depressions and annular posterior portions, AI could not identify the lesions appearing as brownish areas hidden behind the shadows and those that appear dark. Even if the lesion is clearly imaged, AI tends to be unable to identify those with NBI appearing white or those with a small color difference from the surrounding mucosa.

There are several limitations to our study. First, we only verified a small number of

images. Although efficient learning using few images was possible, the number of images used for learning was only 1,200, which is small compared to the number of images evaluated in existing reports (8,428-13,584 learning images).[17, 18] However, since SLPC is infrequent, it is difficult to collect a large number of lesion images. Moreover, we validated the model using 100 SLPC cases, which was a sufficient number of cases for conditions considered as rare cancer. There is no significant difference in the number of evaluation images compared with past reports (1,118-2,269 validation images).[17, 18] Second, we only used high-quality endoscopic still images for the learning and validation images. We should consider validating the model using video images for constructing real-time AI that can be put to practical use at the clinical stage. Third, our study dataset consisted of non-magnified and moderately magnified NBI images. Although magnification may affect the diagnostic accuracy, the accuracy rate of SLPC was very high in the non-magnified NBI images. In the regular per oral EGD for high risk patients with esophageal squamous cell carcinoma, it is common to use the magnified observations in combination with NBI for laryngopharyngeal region, magnified endoscopic images can be used to build significant AI model in clinical for gastroenterologists.[16] However, it is necessary to additionally train using non-magnified images in order to expand the ability of this AI model for otolaryngologists. Fourth, this

was a single-center retrospective study. However, since the verification was conducted using consecutive cases, the results were considered to be reliable. In the future, if it is possible to verify the images collected at multiple facilities, this AI model's diagnostic accuracy and reliability will be improved.

In conclusion, we constructed and evaluated a diagnostic AI model for SLPC in this study. This AI model was constructed for the precise endoscopic detection of SLPC using a small number of learning datasets of NBI images. The AI model can be a potential tool to assist endoscopists and otorhinolaryngologists in the diagnosis of laryngopharyngeal lesions without needing to perform biopsy. In the future, this AI model may facilitate early treatment and contribute to the improvement in the prognosis of LPC.

REFERENCES

1. GLOBOCAN. Estimated cancer incidence, mortality and prevalence worldwide in 2018. International Agency for Research on Cancer– World Health Organization. [http://gco.iarc.fr/]. Accessed January 14, 2019.

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2016. CA Cancer J Clin 2016;66:7-30.

3. Ferlay J, Soerjomataram I, Dikshit R, et al. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 2015;136:E359-E386.

4. Hamashima C, Shabana M, Okada K, Okamoto M, Osaki Y. Mortality reduction from gastric cancer by endoscopic and radiographic screening. Cancer Sci 2015;106:1744-1749.

5. Shaukat A, Mongin SJ, Geisser MS, et al. Long-term mortality after screening for

colorectal cancer. N Engl J Med 2013;369:1106-1114.

6. Muto M, Nakane M, Katada C, et al. Squamous cell carcinoma in situ at oropharyngeal and hypopharyngeal mucosal sites. Cancer 2004;101:1375-1381.

7. Muto M, Katada C, Sano Y, Yoshida S. Narrow band imaging: a new diagnostic approach to visualize angiogenesis in superficial neoplasia. Clin Gastroenterol Hepatol 2005;3:S16-S20.

8. Katada C, Nakayama M, Tanabe S, et al. Narrow band imaging for detecting superficial oral squamous cell carcinoma. Laryngoscope 2007;117:1596-1599.

9. Nonaka S, Saito Y. Endoscopic diagnosis of pharyngeal carcinoma by NBI. Endoscopy 2008;40:347-351.

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;28:1566-1572.

11. Muto M, Satake H, Yano T, et al. Long-term outcome of transoral organ-preserving pharyngeal endoscopic resection for superficial pharyngeal cancer. Gastrointest Endosc 2011;74:477-484.

12. Satake H, Yano T, Muto M, et al. Clinical outcome after endoscopic resection for superficial pharyngeal squamous cell carcinoma invading the subepithelial layer. Endoscopy 2015;47:11-18.

13. Shibuya H, Wakita T, Nakagawa T, Fukuda H, Yasumoto M. The relation between an esophageal cancer and associated cancers in adjacent organs. Cancer 1995;76:101-105.

14. Begg CB, Zhang ZF, Sun M, Herr HW, Schantz SP. Methodology for evaluating the incidence of second primary cancers with application to smoking-related cancers from the surveillance, epidemiology, and end results (SEER) program. Am J Epidemiol 1995;142:653-665.

15. Matsubara T, Yamada K, Nakagawa A. Risk of second primary malignancy after

esophagectomy for squamous cell carcinoma of the thoracic esophagus. J Clin Oncol 2003;21:4336-4341.

16. Naoki O, Hiroyuki M, Yoichi Y, et al. Skill-up Study of Systemic Endoscopic Examination Technique Using Narrow Band Imaging of the Head and Neck Region of Patients With Esophageal Squamous Cell Carcinoma: Prospective Multicenter Study. Dig Endosc, 2019; 31 (6), 653-661.

17. 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.

18. Hirasawa T, Aoyama K, Tanimoto T, et al. Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images. Gastric Cancer 2018;21:653-660.

19. Mori Y, Kudo SE, Misawa M, et al. Real-time use of artificial intelligence in identification of diminutive polyps during colonoscopy: a prospective study. Ann Intern

Med 2018;169:357-366.

20. Kawazoe Y, Shimamoto K, Yamaguchi R, et al. Faster R-CNN-based glomerular detection in multistained human whole slide images. J Imaging 2018;4:91.

21. Zhu Y, Wang QC, Xu MD, et al. Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy. Gastrointest Endosc 2019;89:806-815.e1.

22. Lin TY, Goyal P, Girshick R, He K, Dollár P. Focal loss for dense object detection. arXiv preprint arXiv:1708.02002, 2017.

23. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In CVPR, 2016.

24. Hayashi R, Fujii S, Watanabe A, et al. General rules for clinical studies on head and neck cancer. 6th ed. In: Japan Society for Head and Neck Cancer, ed. Tokyo: Kanehara Shuppan; 2018. p 65-66.

25. LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436-444.

26. Fu CY, Liu W, Ranga A, Tyagi A, Berg AC. DSSD: deconvolutional single shot detector. arXiv:1701.06659, 2016.

27. Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector. In ECCV, 2016.

28. Joseph R, Santosh D, Ross G, Farhadi A. You only look once: unified, real-time object detection. arXiv:1506.02640v5 [cs.CV], 2016.

Supplementary material (Table)

We presented supplementary table to show P value adjustment using Bonferroni method to tumor macroscopic type.

Supplementary material (Video images)

Realtime detection for SLPC video images using an established model by RetinaNet architecture. The SLPC taken with the non-magnified and moderately magnified NBI was well identified in real time. SLPC could be identified well not only in endoscopic images of NBI captured by the Olympus Medical System, but also through the blue laser imaging (BLI) and linked color imaging (LCI) captured by the FUJIFILM Medical System.

Table 1. Background of the learning and validation SLPC datasets

Patient characteristics	**Learning dataset (n=85)**	**Validation dataset (n=89)**
Sex, male/female	82/3	83/6
Median age, year (range)	69 (47-83)	67 (46-87)
Number of lesions per patient		
One/two/three/four	73/10/1/1	79/9/1/0
SLPC characteristics	**Learning dataset (n=100)**	**Validation dataset (n=100)**
Median tumor size, mm (range)	13 (3-50)	12 (3-40)
0-10 mm/11-20 mm/21 mm,	31/45/24	42/52/6
Tumor location		
Soft palate/tonsillar pillars/epiglottis	2/1/12	4/0/7
Oropharynx lateral wall/posterior wall	9/6	3/14
Piriform sinus left/right	21/39	26/30
Hypopharynx posterior wall/postcricoid area	5/5	8/8
Tumor macroscopic type		
0-IIa/0-IIb/0-IIc	7/93/0	21/77/2

Endoscopy system		
EVIS LUCERA CV-260/CLV-260	63	50
EVIS LUCERA ELITE CV-290/CLV-290SL	37	50
History of treatment for head and neck area		
CRT/RT/CT/ESD/ELPS/OPE/NONE	5/0/1/0/18/2/74	5/2/0/0/2/2/89
Histopathology		
SCC/others	100/0	100/0

CRT: chemoradiotherapy, RT: radiotherapy, CT: chemotherapy, ESD: endoscopic submucosal dissection, ELPS: endoscopic laryngo-pharyngeal surgery, OPE: operation, SCC: squamous cell carcinoma

Table 2. Results of the AI diagnosis at each learning stage

Models	Learning datasets			Results		
	SLPC	Normal mucosa	Total images	Accuracy	Sensitivity	Specificity
RetinaNet	100	-	-	76.83%	97.75%	66.38%

200	-	-	79.75%	99.25%	70.00%
300	-	-	79.50%	99.25%	69.62%
400	-	-	88.58%	98.50%	83.62%
100	200	300	94.75%	93.00%	95.62%
200	400	600	96.00%	95.25%	96.38%
300	600	900	96.58%	96.25%	96.75%
400	800	1200	97.25%	95.50%	98.38%

AI: artificial intelligence, SLPC: superficial laryngopharyngeal cancer

Table 3. Results of AI diagnosis at the final learning stage

AI diagnosis for each image	**Results of biopsy specimen**	
	Cancer	Non-cancer
Cancer	382	13
Non-cancer	18	787
Total	400	800

AI: artificial intelligence

Table 4. The sensitivity of diagnostic AI model by clinical characteristics

	The sensitivity for each SLPC image (%)
Tumor size (Median;12mm)	
0-10mm (42 cases)	157/168 (93.5%)
11-20mm (52 cases)	201/208 (96.6%)
21mm- (6 cases)	24/24 (100%)
Tumor location	
Soft palate (4 cases)	16/16 (100%)
Tonsillar pillars (0 cases)	N/A
Epiglottis (7 cases)	24/28 (85.7%)
Oropharynx lateral wall (14 cases)	12/12 (100%)
Piriform sinus left (26 cases)	101/104 (97.1%)
Piriform sinus right (30 cases)	115/120 (95.8%)
Hypopharynx posterior wall (8 cases)	31/32 (96.9%)
Postcricoid area (8 cases)	32/32 (100%)
Tumor macroscopic type	
0-IIa (21 cases)	78/84 (92.9%)
0-IIb (77 cases)	298/308 (96.8%)
0-IIc (2 cases)	6/8 (75.0%)
History of radiotherapy (CRT or only RT)	
Present (7cases)	27/28 (96.4%)
Absent (93 cases)	355/372 (95.4%)
Endoscopy system	
EVIS LUCERA CV-260/CLV-260 (50 cases)	185/200 (92.5%)
EVIS LUCERA ELITE CV-290/CLV-290SL (50 cases)	197/200 (98.5%)
Endoscopic magnification	
Without magnification	121/134 (90.3%)
With magnification	261/266 (98.5%)

CRT: chemoradiotherapy, RT: radiotherapy

Table 5. The effect of clinical characteristics on the sensitivity of diagnostic AI model

	The sensitivity for each SLPC image (%)	P value
Tumor size		
10mm or less	157/168 (93.5%)	P=0.141
11mm or more	225/232 (97.0%)	
Tumor macroscopic type		
0-IIa	78/84 (92.9%)	
0-IIb	298/308 (96.8%)	P=0.01*
0-IIc	6/8 (75.0%)	
History of radiotherapy (CRT or only RT)		
Present (7cases)	27/28 (96.4%)	P=1.000
Absent (93 cases)	355/372 (95.4%)	
Endoscopy system		
EVIS LUCERA CV-260/CLV-260	185/200 (92.5%)	P=0.006*
EVIS LUCERA ELITE CV-290/CLV-290SL	197/200 (98.5%)	
Endoscopic magnification		
Without magnification	121/134 (90.3%)	P=0.001*
With magnification	261/266 (98.5%)	

CRT: chemoradiotherapy, RT: radiotherapy

Table 6. Causes for false positives and false negatives in the AI diagnosis

Cause for false positive (n=13)	**No. of images (%)**
Hard palate	9 (69)
Tongue	2 (15)

Posterior wall of oropharynx	1 (8)
Saliva	1 (8)

Cause for false negative (n=18)	**No. of images (%)**
White tone at NBI	5 (28)
Obscure by shadow	4 (22)
Obscure vessel pattern appearing as dots	3 (17)
The same color as the surrounding mucosa	3 (17)
Normal blood vessel running in the SLPC lesion	2 (11)
Only a piece of lesion	1 (5)

AI: artificial intelligence, NBI: narrow band imaging, SLPC: superficial laryngopharyngeal cancer

Supplementary table: P value adjustment using Bonferroni method to tumor macroscopic type

	0-IIa	0-IIb
0-IIb	0.367	
0-IIc	0.427	0.098

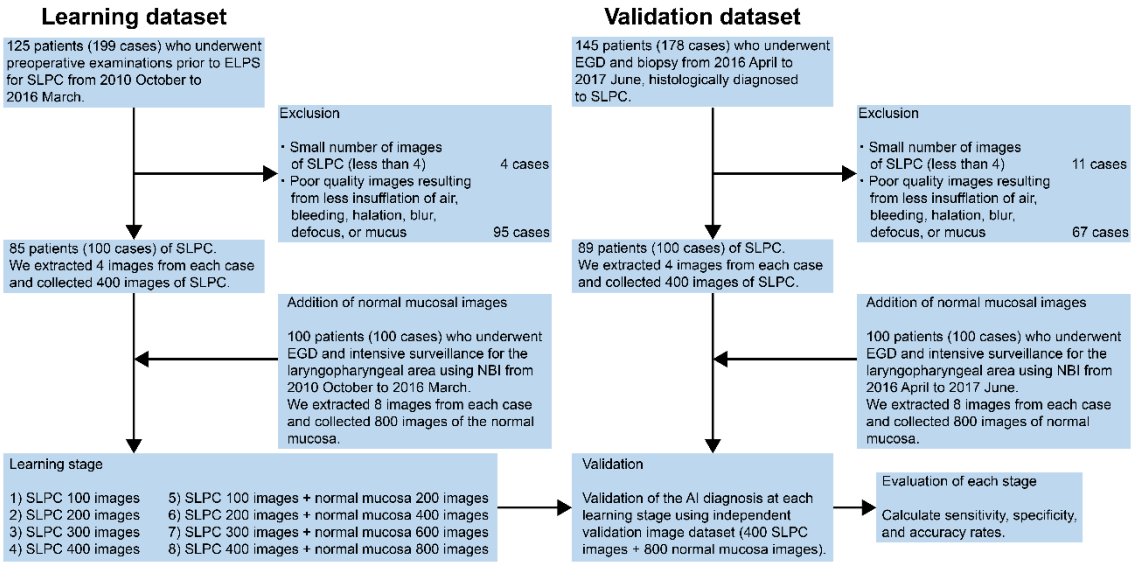

Figure 1. Flow chart depicting patient selection for making datasets and learning and validation schemes in the study

In each learning stage (1 to 8), a test was performed using the image for the validation dataset to calculate the sensitivity, specificity, and accuracy rates.

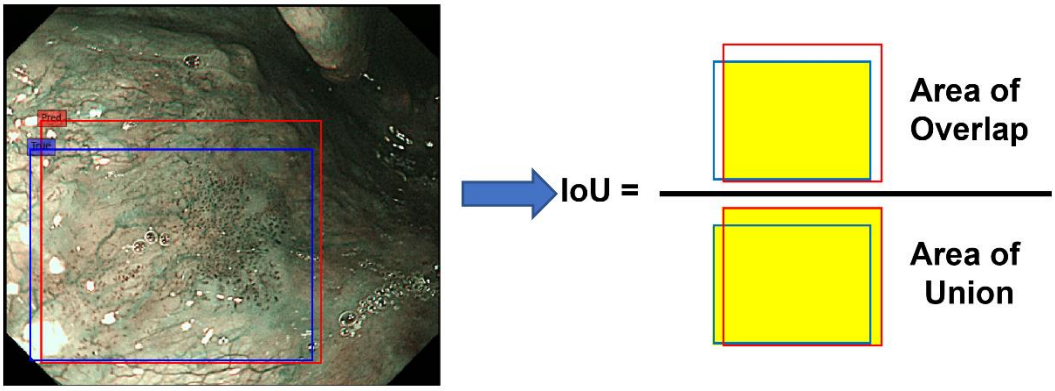

Figure 2. Evaluation of whether AI was correctly diagnosed.

The red rectangle is marked by the AI as a predicted SLPC lesion. The blue rectangle is the location of SLPC, which is manually marked by the endoscopist. Intersection over the union (IoU) is the area of overlap divided by area of union. We judged that AI was able to correctly diagnose SLPC when SLPC was in the red rectangle and IoU was ≥0.40.

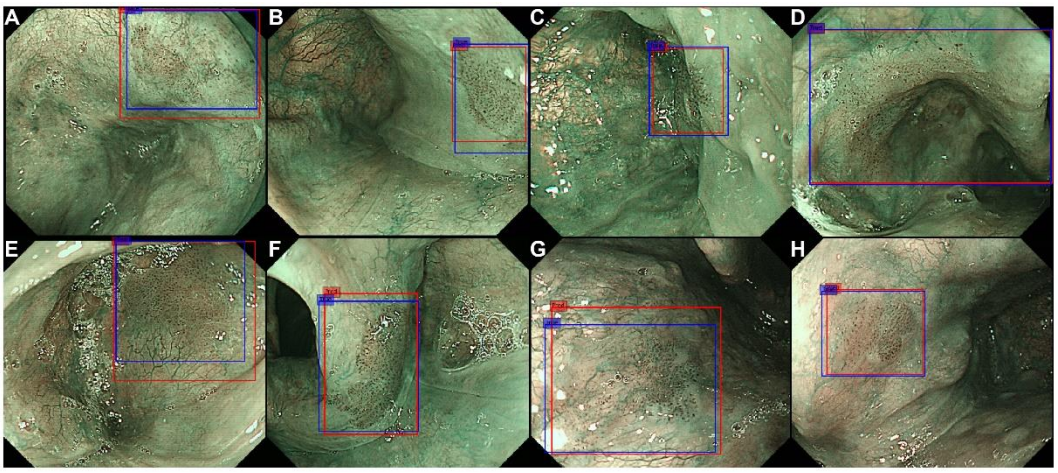

Figure 3. Examples of images that AI correctly diagnosed as SLPC.

The red rectangular frame was marked by the CNN as a predicted lesion and it indicates the extent of a suspected SLPC lesion. An endoscopist manually marked the location of the cancer using a blue rectangular frame. **A.** 0-IIb at the right piriform (IoU=0.85). **B.** 0-IIa at the left piriform (IoU=0.81). **C.** 0-IIb at the left piriform (IoU=0.85). **D.** 0-IIc at the left piriform (IoU=0.97). **E.** 0-IIb at the right piriform (IoU=0.78). **F.** 0-IIb at the postcricoid area (IoU= 0.87). **G.** 0-IIb at the hypopharynx posterior wall (IoU=0.82). **H.** 0-IIb at the postcricoid area (IoU=0.91). For all cases exceeding the IoU of 0.4, we judged that the AI diagnosed the lesion as cancer correctly.

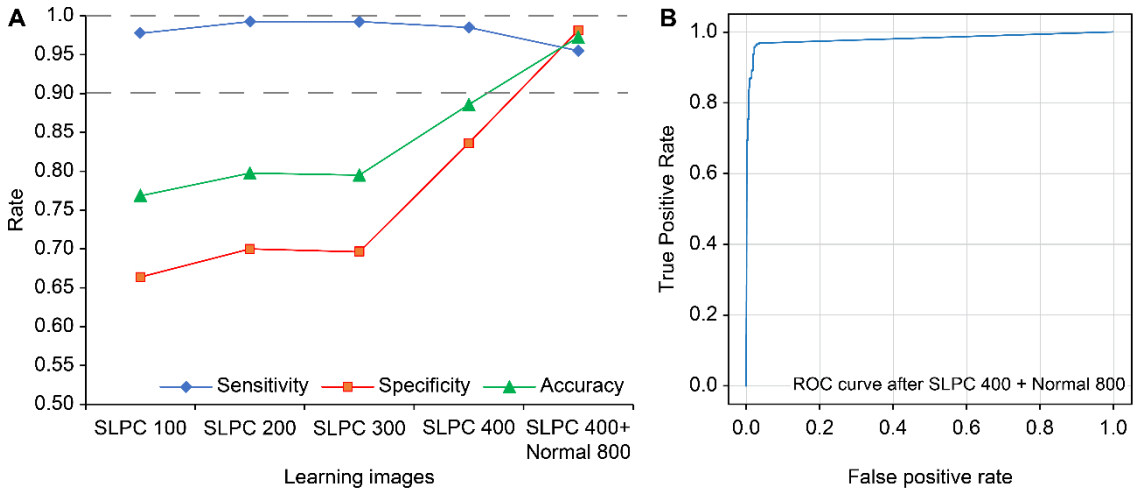

Figure 4. Learning curve of the AI model for detecting SLPC

A. Learning curve on sensitivity, specificity, and accuracy in each learning step (SLPC 100-200-300-400 and 400 SLPC + 800 normal mucosal images). The diagnostic

specificity and accuracy of AI improved according to learned normal images. **B.** Receiver-operating characteristic (ROC) curve after learning 400 SLPC and 800 normal mucosal images. Area under the ROC curve (AUC) = 098.

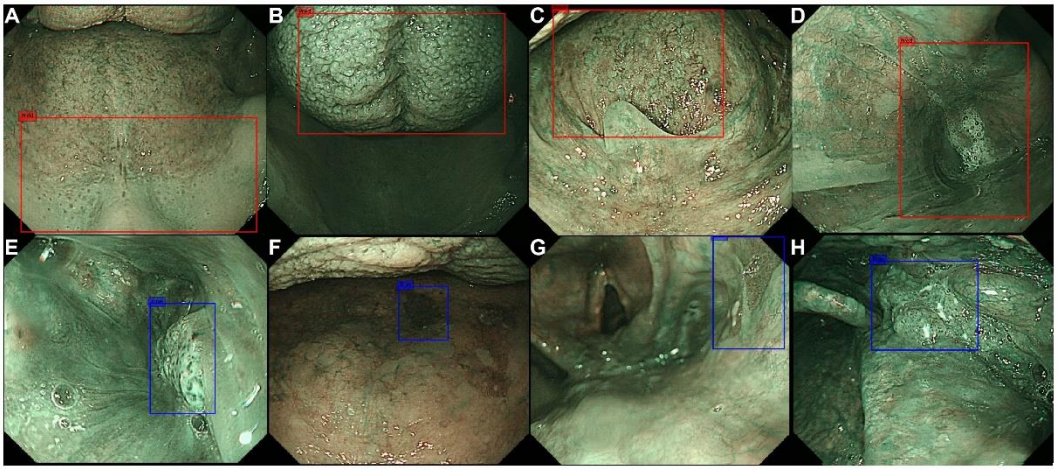

Figure 5. Examples of false-positive and false-negative images.

The images from **A to D** were normal mucosa, which were AI diagnosed as cancer. The red squares indicate areas that were misdiagnosed as cancer, which are as follows: **A.** hard palate, **B.** tongue, **C.** posterior wall of oropharynx, and **D.** saliva. The images from **E to H** were those that AI could not correctly diagnose as SLPC. The blue squares were marked by the endoscopist as cancer. Moreover, the estimated causes were as follows: **E.** white tone at NBI, **F.** obscure by shade. **G.** only a piece of the lesion, and **H.** the same color as the surrounding mucosa.