Thieme

Automatic detection of different types of small-bowel lesions on capsule endoscopy images using a newly developed deep convolutional neural network

Authors

Keita Otani[1], Ayako Nakada[2], Yusuke Kurose[1,3], Ryota Niikura[2], Atsuo Yamada[2], Tomonori Aoki[2], Hiroyoshi Nakanishi[4], Hisashi Doyama[4], Kenkei Hasatani[5], Tetsuya Sumiyoshi[6], Masaru Kitsuregawa[7,8], Tatsuya Harada[3,9,10], Kazuhiko Koike[2]

Institutions

1 Department of Mechano-Informatics, Graduate School of Information Science and Technology, The University of Tokyo, Tokyo, Japan
2 Department of Gastroenterology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
3 Center for Advanced Intelligence Project, RIKEN, Tokyo, Japan
4 Department of Gastroenterology, Ishikawa Prefectural Central Hospital, Kanazawa-shi, Ishikawa, Japan
5 Department of Gastroenterology, Fukui Prefectural Hospital, Fukui-shi, Fukui, Japan
6 The Center for Digestive Disease, Tonan Hospital, Sapporo-shi, Hokkaido, Japan
7 Institute of Industrial Science, The University of Tokyo, Tokyo, Japan
8 National Institute of Informatics, Tokyo, Japan
9 Research Center for Advanced Science and Technology, The University of Tokyo, Tokyo, Japan
10 Research Center for Medical Bigdata, National Institute of Informatics, Tokyo, Japan

submitted 25.12.2019
accepted after revision 14.4.2020

Bibliography
DOI https://doi.org/10.1055/a-1167-8157
Published online: 2020 | Endoscopy

ISSN 0013-726X

Corresponding author
Ryota Niikura, MD, Department of Gastroenterology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan
niikura-dky@umin.ac.jp

Tab. 1s – 3s, Fig. 1s – 3s
Online content viewable at:
https://doi.org/10.1055/a-1167-8157

ABSTRACT

Background Previous computer-aided detection systems for diagnosing lesions in images from wireless capsule endoscopy (WCE) have been limited to a single type of small-bowel lesion. We developed a new artificial intelligence (AI) system able to diagnose multiple types of lesions, including erosions and ulcers, vascular lesions, and tumors.

Methods We trained the deep neural network system RetinaNet on a data set of 167 patients, which consisted of images of 398 erosions and ulcers, 538 vascular lesions, 4590 tumors, and 34 437 normal tissues. We calculated the mean area under the receiver operating characteristic curve (AUC) for each lesion type using five-fold stratified cross-validation.

Results The mean age of the patients was 63.6 years; 92 were men. The mean AUCs of the AI system were 0.996 (95 %CI 0.992 – 0.999) for erosions and ulcers, 0.950 (95 %CI 0.923 – 0.978) for vascular lesions, and 0.950 (95 %CI 0.913 – 0.988) for tumors.

Conclusion We developed and validated a new computer-aided diagnosis system for multiclass diagnosis of small-bowel lesions in WCE images.

Introduction

Wireless capsule endoscopy (WCE) allows direct observation of the gastrointestinal (GI) tract and captures 100 000 images per examination [1]. However, lesions may be visible in only a few of the 100 000 images. Even expert endoscopists require many hours to evaluate WCE images, and the accuracy of diagnoses may differ among endoscopists and hospitals. Therefore, analytical methods based on artificial intelligence (AI) are needed to improve the quality of WCE diagnoses.

Previous studies have investigated the diagnostic accuracy of AI in WCE [2–4]. For erosions and ulcers, values of 99 % for the area under the receiver operating characteristic curve (AUC) have been reported. Similar results have been reported for vascular lesions. In addition, one study evaluated AI diagnostic systems using a classification system capable of analyzing multiple types of lesions including erosions, vascular lesions, and tumors [5]. However, that study focused on systems that are capable of classifying lesions, rather than on systems that perform object detection. Object-detection networks can identify the precise location of lesions. Furthermore, the object-detection system used in previous studies, the Single-Shot Multi-Box Detector (SSD), is limited by its inability to be trained using normal images. Therefore, we modified RetinaNet, an object-detection network, to learn normal images in order to develop and validate a new model for the diagnosis of all types of lesions of the small intestine.

Methods

Study sample and preparation of the image set

We performed a retrospective study using a database that contained WCE images captured from 2009 to 2019 at the University of Tokyo Hospital. All WCE procedures used the PillCam SB2 or SB3 capsule endoscope (Medtronic, Minneapolis, Minnesota, USA) and were performed after patients had fasted for 12 hours. Oral simethicone (40 mg) was administered before WCE examinations [6, 7].

From the database, we extracted a case group of 156 patients with erosions and ulcers, angioectasias, or tumors. We also randomly extracted a control group of 11 patients and randomly extracted normal images from these patients. We used images from these patients to develop a data set consisting of 398 images of erosions and ulcers, 538 images of angioectasias, 4590 images of tumors, and 34 437 normal images (▶ **Fig. 1**). We defined vascular lesions as angioectasias and venous malformations; we defined tumors as polyps, nodules, masses, and submucosal tumors. Two expert endoscopists (A.N. and R. N.) manually annotated all erosions and ulcers, vascular lesions, and tumors with bounding boxes (gold-standard boxes).

In the external validation study, we retrospectively analyzed 288 patients who underwent WCE at Ishikawa Prefectural Central Hospital, Fukui Prefectural Hospital, and Tonan Hospital. We randomly extracted images of erosions, vascular ulcers, tumor lesions, and normal tissue to create a balanced data set, as in the internal cross-validation data set. This study was approved by the Ethics Committee of the University of Tokyo (No. 12016-[1]).

Development of the AI-based diagnostic system

To construct an AI-based diagnostic system, we used the deep neural network architectures of RetinaNet and SSD. SSD was also used by Aoki et al. [2]. Small objects have weak feature representation in SSD, which makes them difficult to detect [8]. In contrast, RetinaNet uses strong multiscale image features to detect small objects. Therefore, its ability to detect small objects is particularly good [9]. The RetinaNet network architecture uses a Feature Pyramid Network backbone on top of a feedforward ResNet architecture to generate a rich, multiscale convolutional feature pyramid. To this backbone, RetinaNet attaches two subnetworks, one for classifying anchor boxes and one for regressing from anchor boxes to ground-truth object boxes.

We trained RetinaNet and SSD models to identify areas within the bounding boxes as erosions and ulcers, vascular lesions, or tumors, and areas outside the bounding boxes as the background. The input image size was 512 × 512. Learning was performed by penalizing the incorrect output and minimizing this penalty. Lesion detection differs from general object detection in that the boundaries of the detection targets are ambiguous. Therefore, the penalty was relaxed to allow some positional shifting of output boxes.

Outcome measures and statistics

The outcome was per-lesion image analysis, which was defined as per-image diagnosis of erosions and ulcers, vascular lesions, and tumors. The definitions of model accuracy were the presence of any overlap between the AI-drawn bounding boxes and the gold-standard boxes. We used five-fold stratified cross-validation to balance the lesion ratios to test the model (▶ **Fig. 1**). When generating the internal and external validation sets, random sampling was performed to avoid bias that caused RetinaNet or SSD to exhibit falsely high performance.

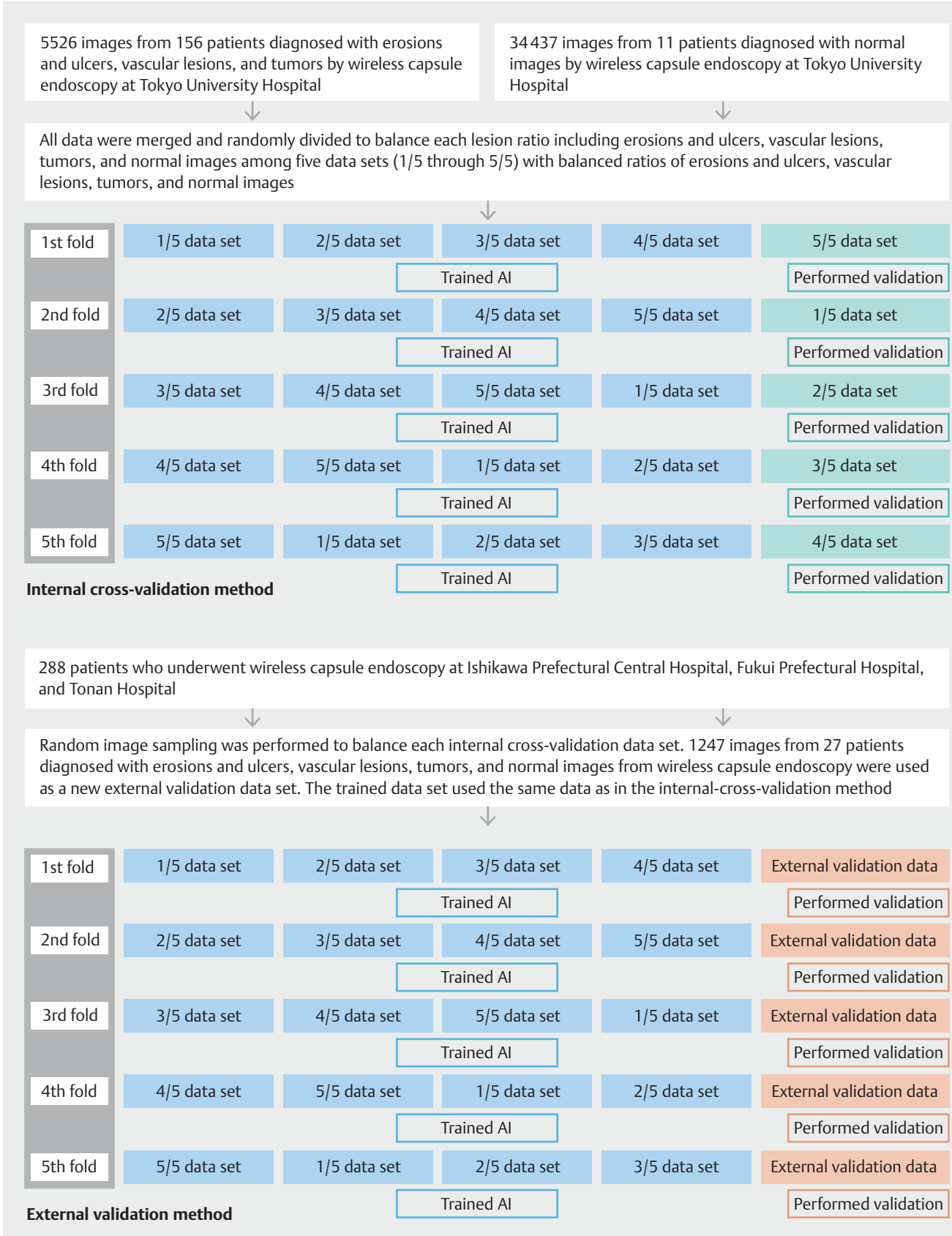
We plotted the receiver operating characteristic curve and estimated the AUC, sensitivity, specificity, and accuracy of the AI detection system for each lesion image for each probability score cutoff and Youden index. We compared the AUCs of the RetinaNet and SSD systems using the Delong test [10]. *P* values < 0.05 were considered indicative of statistical significance. Statistical analyses were performed using SAS software v.9.4 (SAS Institute, Cary, North Carolina, USA).

Results

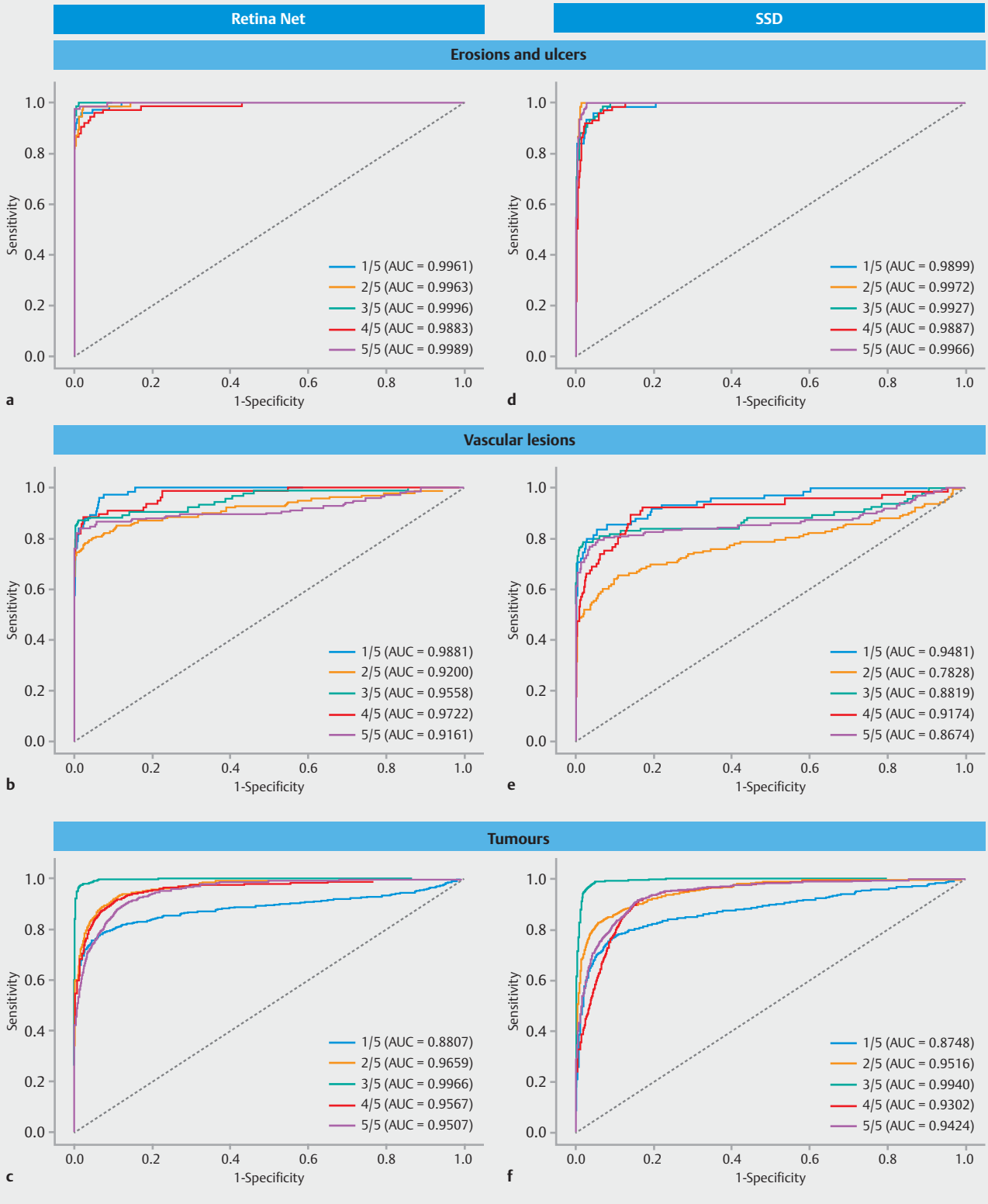
Derivation and internal validation cohort

The characteristics of the patients are shown in ▶ **Table 1**. The mean age of the patients was 63.6 years, and 92 were men. The most common indication for WCE was obscure GI bleeding. Each cross-validation set consisted of between 7047 and 9034 images from 27 to 42 patients (**Table 1s**, see online-only Supplementary Material). The lesion ratios were well balanced among the cross-validation sets.

The mean AUCs of the RetinaNet model were 0.996 (95 % confidence interval [CI] 0.992 – 1.000) for detecting erosions and ulcers, 0.950 (95 %CI 0.923 – 0.978) for vascular lesions, and 0.950 (95 %CI 0.913 – 0.988) for tumors (▶ **Fig. 2**). Compar-



▶ **Fig. 1** Study flow diagram.



▶ **Fig. 2** Receiver operating characteristic curves of the RetinaNet (**a – c**) and SSD (**d – f**) systems used in the internal validation cohort to detect: **a, d** erosions and ulcers; **b, e** vascular lesions; **c, f** tumors.

▶ **Table 1** Characteristics of the 167 patients who made up the training data set.

Patients with pathological images	**n = 156**
Mean age (SD), years	63.5 (16.2)
Sex, male, n (%)	87 (55.8)
Indication for capsule endoscopy, n (%)	
▪ Occult OGIB	37 (23.7)
▪ Overt previous OGIB	53 (34.0)
▪ Overt ongoing OGIB	8 (5.1)
▪ Tumor surveillance	45 (28.8)
▪ Inflammatory bowel disease	4 (2.6)
▪ Abdominal symptoms	9 (5.8)
Patients with normal images	**n = 11**
Mean age (SD), years	65.7 (13.0)
Sex, male, n (%)	5 (45.5)
Indication for capsule endoscopy (%)	
▪ Occult OGIB	6 (54.5)
▪ Overt previous OGIB	4 (36.4)
▪ Overt ongoing OGIB	1 (9.1)

OGIB, obscure gastrointestinal bleeding.

ed with the SSD model, the AUCs of the RetinaNet model were significantly higher for the detection of erosions and ulcers, vascular lesions, and tumors (**Fig. 1s**). The sensitivity, specificity, and accuracy for each probability score cutoff (0.1 to 0.9) and Youden index value are shown in **Tables 2s** and **3s**.

External validation cohort

We analyzed data from 40 patients in the external validation cohort (▶ **Fig. 1**). The lesion ratios were well balanced between the external and internal validation cohorts (**Table 1s**). The mean AUCs of the RetinaNet model were 0.928 (95 %CI 0.919 – 0.937) for detection of erosions and ulcers, 0.884 (95 % CI 0.874 – 0.893) for vascular lesions, and 0.902 (95 %CI 0.873 – 0.930) for tumors (**Fig. 2s**).

Discussion

We developed and validated a superior new AI-based diagnostic system using RetinaNet to detect all types of small-bowel lesions in WCE images. Our AI system had an AUC value of 0.95 for tumors, which are particularly difficult to detect with an AI system.

Compared with SSD-based AI systems [2, 3], including our own SSD system, the RetinaNet AI system had significantly superior diagnostic ability and higher AUC values. In addition, the RetinaNet system identified fewer false-negative images compared with our SSD system (**Fig. 3s**). We suggest two reasons for this difference. First, RetinaNet uses the Feature Pyramid Network, which creates a multiscale feature pyramid in which strong features are used to detect small objects that SSD is unable to identify. Second, our approach used a relatively loose penalty method to lessen the imbalance between the background and target objects, which is suitable for assessment of small data sets.

Our RetinaNet system also has several limitations. First, we missed two images from a patient in the erosions and ulcers group, as well as one image from a patient with vascular lesions. These images included small or large bubbles and residuals that may have made the lesions difficult to identify. We also may have missed a vascular lesion (**Fig. 3s, image c**) because its image was similar to a normal type of venous dilation, which may have made it difficult to detect.

Second, our RetinaNet system may have been overfitted, with respect to our training data set, which would limit its diagnostic accuracy in other data sets. The number of training images was limited; indeed, the small number of normal images was from only 11 patients. The external validation results were inferior to those of the internal cross-validation. To reduce overfitting, the RetinaNet model must be trained using a larger number of lesion images and various numbers of normal images, including residuals, to further improve the diagnostic accuracy of the RetinaNet system.

Third, a video-based study to compare RetinaNet with expert endoscopists is needed prior to the use of RetinaNet in daily clinical practice; a randomized controlled trial is also needed. Notably, we have the technology necessary to evaluate video images using RetinaNet and are planning to conduct a randomized controlled trial.

In conclusion, we developed and validated a new RetinaNet system for multiclass diagnosis of lesions in WCE images. If this high quality diagnostic system is further improved, it could be a valuable diagnostic tool for endoscopists in daily clinical practice.

Acknowledgments

The authors received a grants from ICT infrastructure establishment and implementation of artificial intelligence for clinical and medical research from the Japan Agency of Medical Research and Development AMED (20lk1010036s0702) and from JST AIP Acceleration Research Grant Number JPMJCR20U3. We thank T. K. for assistance with the statistical analysis.

Competing interests

The authors declare that they have no conflict of interest.

References

[1] Iddan G, Meron G, Glukhovsky A et al. Wireless capsule endoscopy. Nature 2000; 405: 417

[2] Aoki T, Yamada A, Aoyama K et al. Automatic detection of erosions and ulcerations in wireless capsule endoscopy images based on a deep convolutional neural network. Gastrointest Endosc 2019; 89: 357–363

[3] Tsuboi A, Oka S, Aoyama K et al. Artificial intelligence using a convolutional neural network for automatic detection of small-bowel angioectasia in capsule endoscopy images. Dig Endosc 2020; 32: 382–390

[4] Zhou T, Han G, Li BN et al. Quantitative analysis of patients with celiac disease by video capsule endoscopy: a deep learning method. Comput Biol Med 2017; 85: 1–6

[5] Ding Z, Shi H, Zhang H et al. Gastroenterologist-level identification of small-bowel diseases and normal variants by capsule endoscopy using a deep-learning model. Gastroenterology 2019; 157: 1044–1054

[6] Albert J, Gobel CM, Lesske J et al. Simethicone for small bowel preparation for capsule endoscopy: a systematic, single-blinded, controlled study. Gastrointest Endosc 2004; 59: 487–491

[7] Niikura R, Yamada A, Maki K et al. Associations between drugs and small-bowel mucosal bleeding: multicenter capsule-endoscopy study. Dig Endosc 2018; 30: 79–89

[8] Liu W, Anguelov D, Erhan D et al. SSD: Single Shot MultiBox Detector. In: Leibe B, Matas J, Sebe N, Welling M eds. Computer Vision – ECCV 2016. Proceedings of the 14th European Conference on Computer Vision; 2016 Oct 11–14. Amsterdam, The Netherlands: doi:10.1007/978-3-319-46448-0_2

[9] Lin T, Goyal P, Girshick R et al. Focal loss for dense object detection. In: Fangyu L, Shuaipeng L, Liqiang Z et al. eds. IEEE International Conference on Computer Vision (ICCV) 2017. Proceedings of the 16th ICCV; 2017 Oct 22–29. Venice, Italy: 2980–2988

[10] DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics 1988; 44: 837–845