


## Original Article

## Improved classification and localization approach to small bowel capsule endoscopy using convolutional neural network

Yunseob Hwang,<sup>1,2</sup> Han Hee Lee,<sup>4,5</sup> Chunghyun Park,<sup>1</sup> Bayu Adhi Tama,<sup>1</sup>  Jin Su Kim,<sup>4</sup> Dae Young Cheung,<sup>4</sup> Woo Chul Chung,<sup>4</sup> Young-Seok Cho,<sup>4</sup> Kang-Moon Lee,<sup>4</sup> Myung-Gyu Choi,<sup>4</sup> Seungchul Lee<sup>1,2,3</sup> and Bo-In Lee<sup>4</sup>

<sup>1</sup>Department of Mechanical Engineering, <sup>2</sup>Postech-Catholic Biomedical Engineering Institute, <sup>3</sup>Graduate School of Artificial Intelligence, Pohang University of Science and Technology (POSTECH), Pohang, <sup>4</sup>Division of Gastroenterology, Department of Internal Medicine and <sup>5</sup>Postech-Catholic Biomedical Engineering Institute, College of Medicine, The Catholic University of Korea, Seoul, Korea

**Background:** Although great advances in artificial intelligence for interpreting small bowel capsule endoscopy (SBCE) images have been made in recent years, its practical use is still limited. The aim of this study was to develop a more practical convolutional neural network (CNN) algorithm for the automatic detection of various small bowel lesions.

**Methods:** A total of 7556 images were collected for the training dataset from 526 SBCE videos. Abnormal images were classified into two categories: hemorrhagic lesions (red spot/angioectasia/active bleeding) and ulcerative lesions (erosion/ulcer/stricture). A CNN algorithm based on VGGNet was trained in two different ways: the combined model (hemorrhagic and ulcerative lesions trained separately) and the binary model (all abnormal images trained without discrimination). The detected lesions were visualized using a gradient class activation map (Grad-CAM). The two models were validated using 5,760 independent images taken at two other academic hospitals.

**Results:** Both the combined and binary models acquired high accuracy for lesion detection, and the difference between the two models was not significant (96.83% vs 96.62%,  $P = 0.122$ ). However, the combined model showed higher sensitivity (97.61% vs 95.07%,  $P < 0.001$ ) and higher accuracy for individual lesions from the hemorrhagic and ulcerative categories than the binary model. The combined model also revealed more accurate localization of the culprit area on images evaluated by the Grad-CAM.

**Conclusions:** Diagnostic sensitivity and classification of small bowel lesions using a convolutional neural network are improved by the independent training for hemorrhagic and ulcerative lesions. Grad-CAM is highly effective in localizing the lesions.

**Key words:** artificial intelligence, capsule endoscopy, computer-assisted diagnosis, deep learning, gastrointestinal hemorrhage

## INTRODUCTION

SMALL BOWEL CAPSULE endoscopy (SBCE) was first introduced in 2000 and has become the leading

**Corresponding:** Seungchul Lee, Department of Mechanical Engineering and Graduate School of Artificial Intelligence, Pohang University of Science and Technology (POSTECH), 5th Engineering Building, 77 Chengam-Ro, Nam-Gu Pohang, Gyeongbuk 37673, Korea. Email: seunglee@postech.ac.kr Bo-In Lee, Division of Gastroenterology, Department of Internal Medicine, Seoul St. Mary's Hospital, College of Medicine, The Catholic University of Korea, 222 Banpo-daero, Seocho-gu, Seoul 137-701, Korea. Email: gidoc4u@gmail.com

The authors Yunseob Hwang and Han Hee Lee contributed equally to this work as the first authors.

Received 20 April 2020; accepted 1 July 2020.

diagnostic tool for small bowel lesions.<sup>1</sup> However, it has an obvious drawback: interpreting SBCE images is time consuming and labor intensive. Each examination generates 50,000–100,000 images, which means that the reading time usually exceeds an hour.<sup>2–4</sup> Moreover, that extended analysis tends to have low accuracy because lesions appear in only a few frames, which strains the limits of human interpretation. Although several software enhancements, including the Suspected Blood Indicator (Given Imaging, Yokneam, Israel), Quick-View (Medtronic, Minneapolis, MN, USA), and Express View (Capsovision, Inc., Saratoga, CA, USA), attempt to overcome these drawbacks, their performance is insufficient for use in clinical practice because of their low sensitivity and limited accuracy.<sup>4–6</sup>

Computer-aided diagnosis studies using deep learning have produced remarkable results in various fields.<sup>7,8</sup> Particularly, there has been a rapid increase in medical image-based diagnosis using convolutional neural networks (CNNs), which already exceed humans in general image recognition.<sup>9</sup> CNN-based diagnosis processes have been suggested for breast cancer histology, skin cancer, and gastric cancer.<sup>10–12</sup>

Recent studies have suggested the promise of artificial intelligence (AI) using CNN in the interpretation of SBCE images.<sup>13–16</sup> However, the development of a CNN-based reading system suitable for daily use still faces important limitations. First, the regions on SBCE images that contain abnormalities must be manually annotated by a human to train a CNN algorithm.<sup>14</sup> Second, even if a CNN algorithm could adequately divide SBCE images into normal and abnormal, no existing CNN can classify the abnormalities as ulcers or angioectasias.<sup>13</sup>

In this study, we developed a CNN-based AI model using gradient-weighted class activation mapping (Grad-CAM) to classify hemorrhagic and ulcerative lesions on SBCE images without manual annotation, and we validated the model by comparing it with conventional models that can only differentiate normal and abnormal images.

## METHODS

### Datasets and data preprocessing

WE BEGAN WITH 526 SBCE videos (PillCam SB2 or SB3 system, Medtronic) consecutively recorded in Seoul St. Mary's Hospital, Korea, from May 2007 to May 2019. Those videos had been performed in accordance with the indications normally applied in real clinical practice. From this video dataset, two expert endoscopists (HHL and BIL) extracted 3778 well-focused and high-quality still images containing pathologic findings for CNN training. An equal number of normal images were also selected to prevent data imbalance problems that causes detection bias of the CNN model. Pathologic findings were labeled using two categories: hemorrhagic lesions (red spot, angioectasia, and active bleeding) and ulcerative lesions (erosion, ulcer, and stricture) (Fig. 1). If an image possessed both findings, it was classified into both categories. This study was reviewed and approved by the Institutional Review Board (XC19REDI0034K).

An additional 5760 images (960 abnormal and 4800 normal images) were collected from 162 videos taken at two other academic hospitals (Yeouido St. Mary's Hospital and St. Vincent Hospital, Korea) from July 2017 to September 2019 for the validation dataset. We selected all images showing hemorrhagic or ulcerative lesions to avoid selection bias.

Figure 2 shows the workflow for the development of our CNN-based AI model. To develop significant attributes for lesion detection using SBCE images, two preprocessing steps were conducted, as described in a previous study.<sup>17</sup> First, the frame information for the SBCE images, such as patient name, timestamp, version of SBCE, and elapsed time, were removed. Second, some unnecessary parts were cropped, such as the window frame of the SBCE image (Fig. S1).

Data were augmented to eight folds by rotating the image by 90 degrees three times and flipping each rotated image horizontally. As a result of that augmentation, 30,224 abnormal images (11,776 hemorrhagic lesions and 18,448 ulcerative lesions) and 30,224 normal images were used to train CNN models.

### Proposed CNN and Grad-CAM

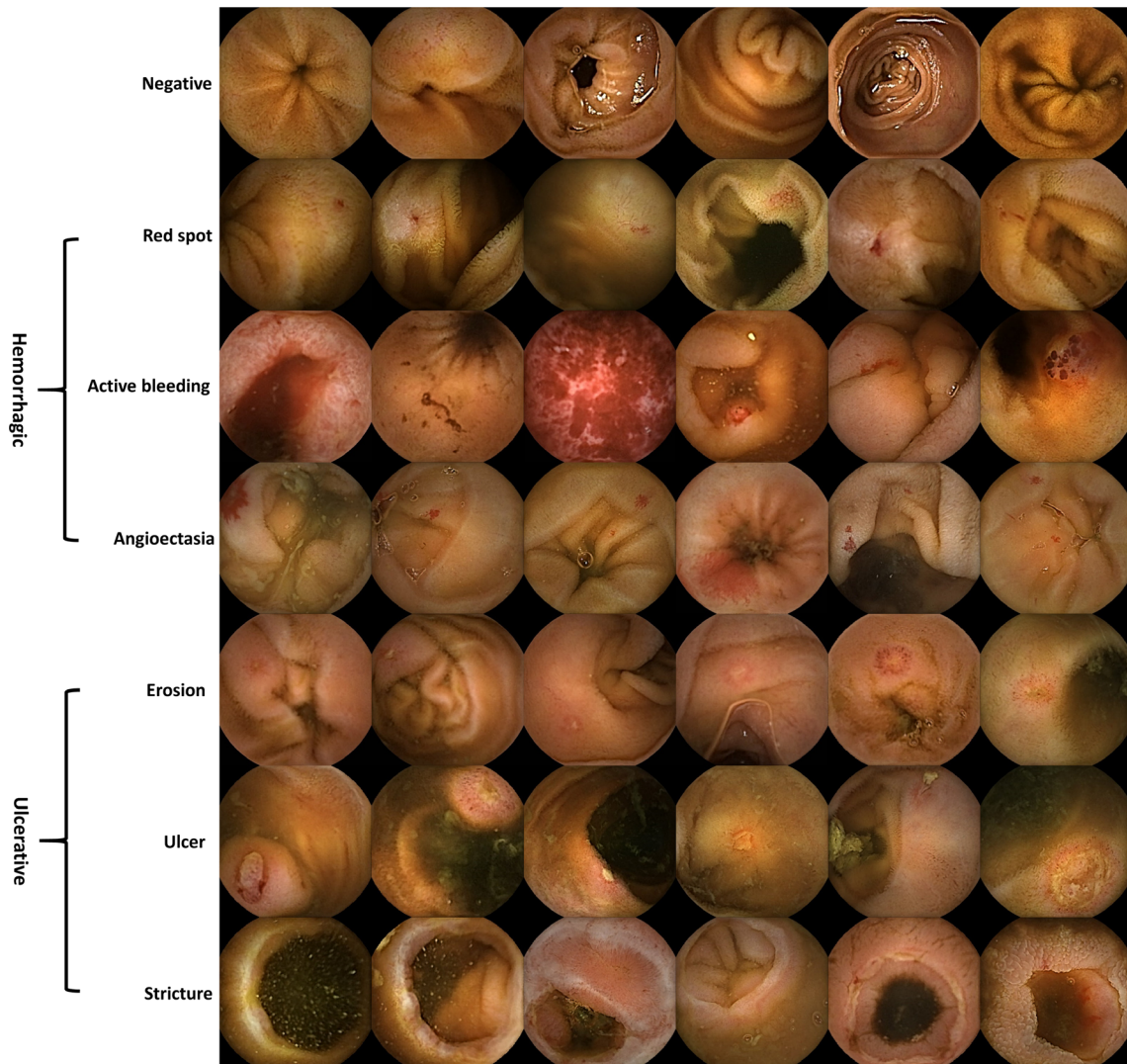
We designed a new CNN based on VGGNet, which is one of the most representative CNNs, as the backbone for detecting small bowel lesions.<sup>18</sup> Figure 3a shows the entire architecture of the CNN on the image scale. The algorithm contains fourteen convolutional layers and seven max-pooling layers, two fully connected layers, and an output layer. As shown in Figure 3b, the convolutional blocks of the proposed algorithm each contain two convolutional layers with a 3 by 3 kernel and one max-pooling layer with a 2 by 2 kernel. The channel of the kernels in the convolution layers of each block is equal.

In addition, we adopted the Grad-CAM to visualize the region contributing to the CNN algorithm's decisions as a heat map.<sup>19</sup> As shown in Figure 3c, the Grad-CAM was extracted from the feature maps (the output of the convolutional and pooling layers) about the abnormal class and then resized to the preprocessed SBCE image size (512 × 512) and overlapped onto the original input image.

### CNN training: combined and binary models

By using two training protocols for the CNN presented above, we developed two different AI models: the combined model and the binary model (Fig. 2).

Figure 3d illustrates all the CNN models and their outputs for a single SBCE image. The combined model was obtained by combining two basal models, each of which was trained to detect an individual pathology type, e.g., either hemorrhagic or ulcerative pathology. The NH model was developed to detect hemorrhagic lesions by training it with only hemorrhagic lesion images and normal images, and the NU model was trained to detect ulcerative lesions by including only ulcerative lesion images and normal images



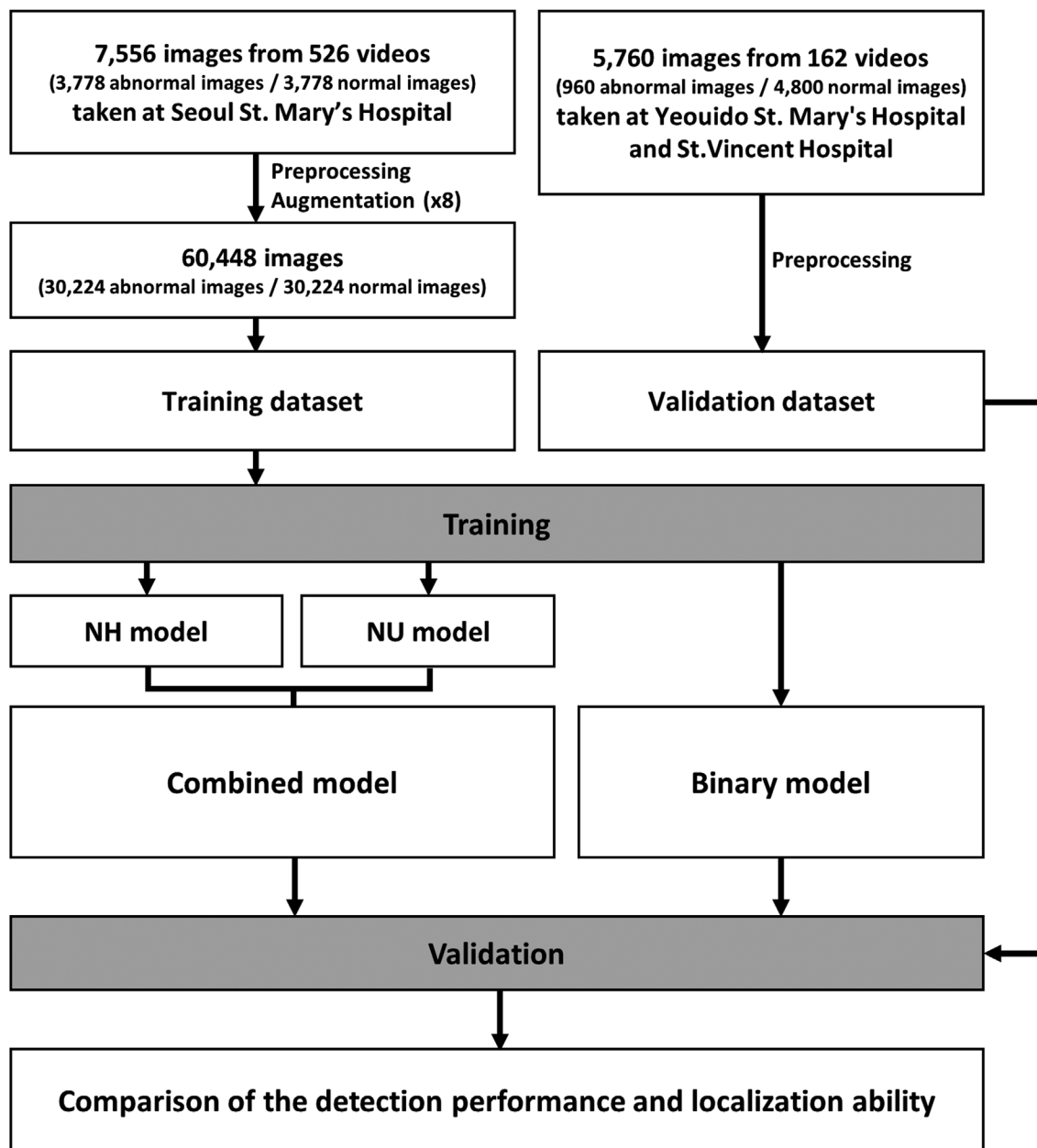
**Figure 1** Typical small bowel capsule endoscopy images containing pathological findings.

in the training dataset. Each model yielded a prediction probability for each class and a Grad-CAM corresponding to the abnormal class. The combined model thus predicted an input image as normal only if both basal models predicted the image as normal. In addition, the combined model could infer two prediction probabilities from the NH model and the NU model. Moreover, the Grad-CAM of the combined model was obtained by combining the Grad-CAMs of the two basal models.

Like the conventional model for binary classification, the binary model was trained to detect all pathological images as abnormal without distinguishing the types of lesions. Likewise, the binary model also yielded a prediction probability belonging to each class and a Grad-CAM.

### CNN validation: assessments, metrics, and statistics

Two quantitative assessments were carried out to validate the detection performance of the models. First, accuracy, sensitivity, specificity, negative predictive value (NPV), and positive predictive value (PPV) were calculated to compare the combined model and the binary model. Second, the accuracy of all models, including the individual basal models (NH and NU), was calculated for normal and all individual abnormal lesions. For the qualitative assessment, we first compared the localization ability of the binary model and combined model using their Grad-CAMs with pathological SBCE images. Second, we used the t-stochastic



**Figure 2** Workflow of the study: data processing, training, and validation.

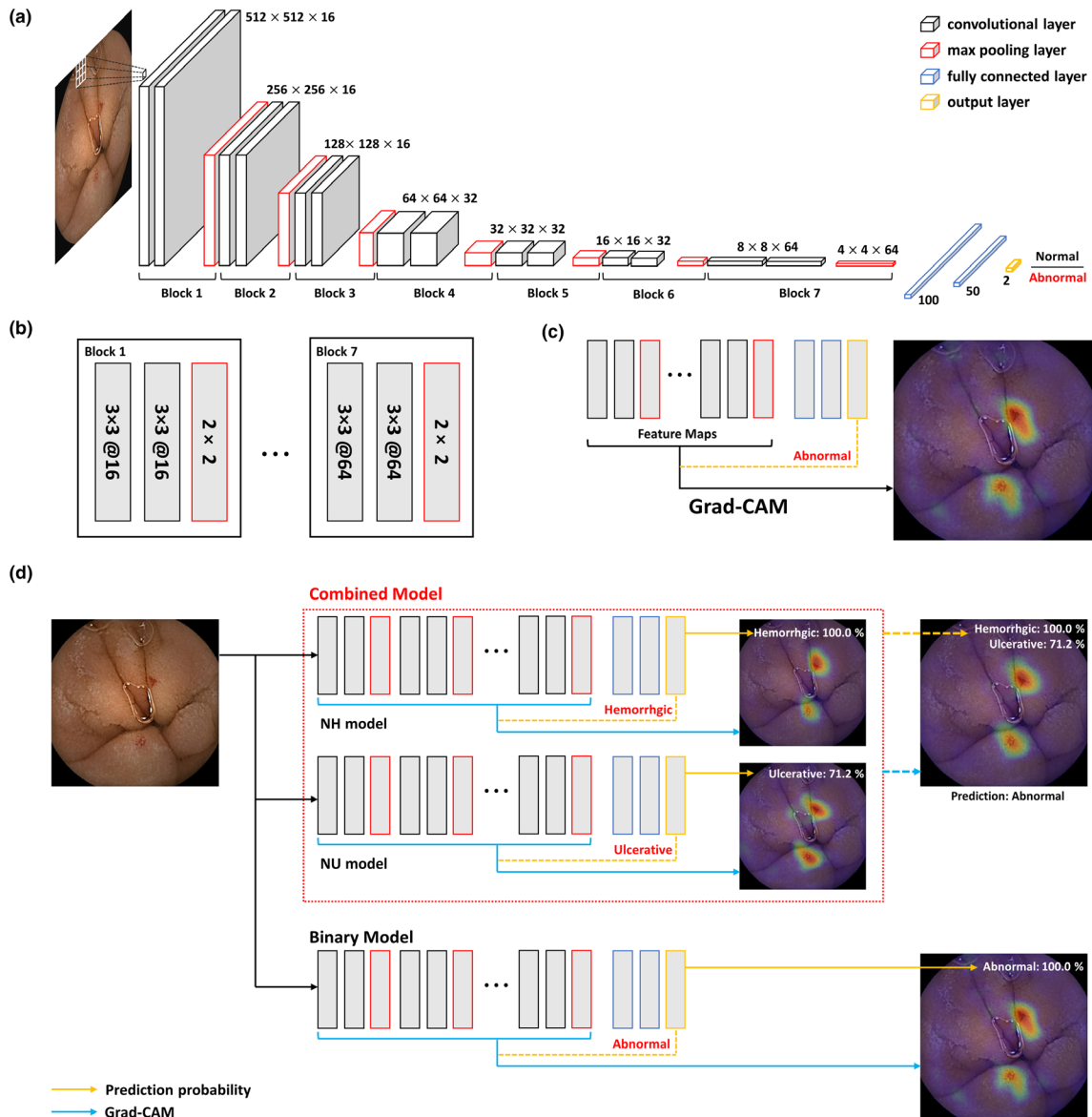
neighbor embedding (t-SNE) algorithm to visualize how well the CNN model distinguished three endoscopic findings (normal, hemorrhagic lesion, and ulcerative lesion).<sup>20</sup>

In addition, the area under the receiver operating characteristics (AUROC) curve was considered to assess the reproducibility of lesion recognition. In this study, G-Mean

score was utilized to seek the optimal balance between true-positive rate (TPR) and false-positive rate (FPR). G-Mean score is calculated as follows.

$$G - \text{Mean} = \sqrt{\text{Sensitivity} * \text{Specificity}}$$

where sensitivity = TPR and specificity = 1–FPR. The confidence intervals (CIs) around the metric values were



**Figure 3** Proposed convolutional neural network (CNN) model. (a) The CNN architecture of the proposed model. The model consists of seven convolutional blocks and two fully connected layers. (b) The structure of the seven convolutional blocks. (c) The process of gradient-weighted class activation mapping (Grad-CAM). Grad-CAM is extracted from feature maps corresponding to the abnormal class. (d) All CNN models and their outputs. The combined model can yield two different prediction probabilities and a Grad-CAM by conjoining the Grad-CAMs of its basal models.

calculated at the 95% confidence level. We also measured the difference in detection performance between the binary model and combined model using paired two-sampled *t*-testing. That test considered a pairwise comparison and rejected the null hypothesis, i.e., the two models performed equivalently, at a significance level of  $\alpha < 0.05$ .

## RESULTS

### Detection performance

**T**ABLE 1 compares the performance of the combined and binary models. In terms of detection accuracy, both the combined and binary models showed high accuracy

without a significant difference between them (96.83% vs 96.62%,  $P = 0.122$ ). The combined model showed higher sensitivity (97.61% vs 95.07%,  $P < 0.001$ ) and NPV (97.58% vs 95.23%,  $P < 0.001$ ) than the binary model. On the contrary, the binary model showed higher specificity (98.18% vs 96.04%,  $P < 0.001$ ) and PPV (98.12% vs 96.11%,  $P < 0.001$ ) than the combined model. The AUROCs for lesion detection were measured to compare the combined and binary models, and no significant difference was found (0.9957; 95% CI, 0.9951–0.9963 and 0.9956; 95% CI, 0.9948–0.9964, respectively). However, considering the optimal cutoff obtained by G-Mean score, it was clear that the combined model showed a higher sensitivity than the binary model. (Fig. 4).

Table 2 shows the accuracy of all models in identifying abnormal small bowel lesions. As expected, the basal models achieved low accuracy for unseen pathological lesions, i.e., the NH model poorly predicted the probability of ulcerative lesions. Nonetheless, those models did show better than 80% accuracy for unseen lesions. By combining those models, the combined model showed the best accuracy for all lesions: 97.48% for red spots, 99.77% for active bleeding, 97.15% for erosions, 96.86% for ulcers, and 97.58% for stricture. On the other hand, the binary model yielded the best accuracy for normal images at 98.21%. For angioectasia, all models achieved 100.00% accuracy.

### Lesion localization using Grad-CAM

We explored the performance of the combined and binary models in terms of their lesion localization ability. Figure 5 shows the Grad-CAMs of the two models with corresponding SBCE images. We compared each model's localization ability on 12 different lesions. The culprit area on the images was localized more accurately by the combined model. In addition, the combined model was able to show the probabilities for hemorrhagic and ulcerative lesions, whereas the binary model could show only the probability of an abnormal lesion.

### Distribution of three endoscopic findings

Figure 6 shows the results from the t-SNE algorithm corresponding to the features extracted from the CNN models. The results contain three independent clusters that represent the three classes, i.e., normal, hemorrhagic, and ulcerative. On the right side of the figure, red dots (c, e) within a red cluster show a special case of hemorrhagic pathology (active bleeding). In the center of the figure, some red dots (b) belonging to the hemorrhagic class are mixed up with blue dots belonging to the ulcerative class (g), which indicates that several ulcers or erosions were accompanied by hemorrhagic features. This finding accords with the fact that in real clinical practice most ulcers and erosions are accompanied by hemorrhagic features in SBCE images. On the left side of the figure, green dots (a, d, f) within a green cluster show examples from the normal class.

### False-negative and false-positive cases

Despite their high accuracy and sensitivity, both models had false-negative cases (Fig. S2). In our experiments, the binary model had more false-negative cases than the combined model: 47 vs 23, respectively.

There were also false-positive cases (Fig. S3). The combined and binary models had 190 and 87 false-positive cases, respectively.

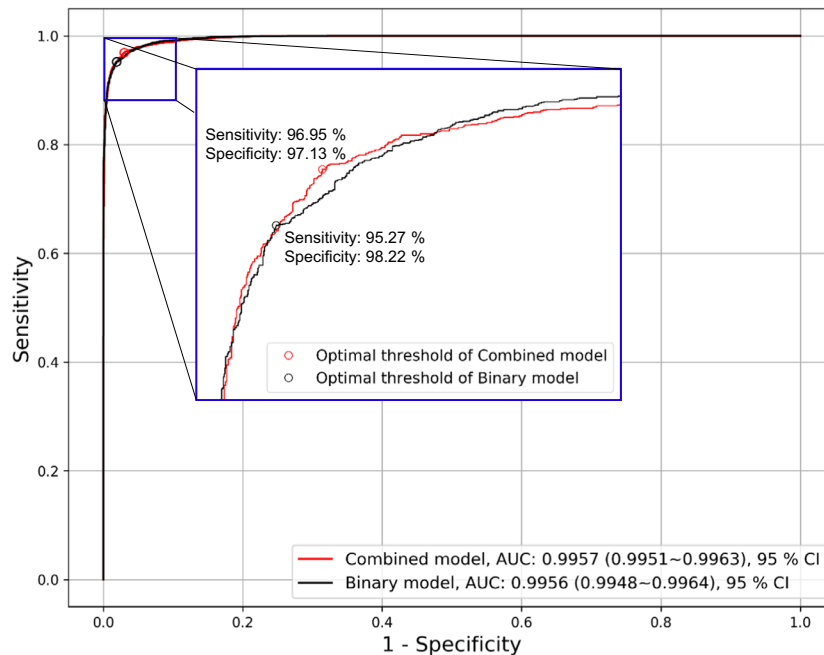
## DISCUSSION

IN THIS STUDY, we developed a novel CNN-based AI algorithm that can classify SBCE images even without manual annotation for a certain abnormal area. We trained this algorithm in two different ways, one was to identify abnormal images without discrimination (the binary model), and the other was to separately identify hemorrhagic and ulcerative lesions and then combine the results (the combined model). Both models showed satisfactory levels of accuracy in detecting abnormal images. The combined

**Table 1** Detection performance comparison between the combined and binary models

Assessment	Combined model, % (95% CI)	Binary model, % (95% CI)	P-value
Accuracy	96.83 (97.22–97.92)	96.62 (96.19–97.06)	0.122
Sensitivity	97.61 (97.08–98.14)	95.07 (94.35–95.79)	<0.001
Specificity	96.04 (95.19–96.89)	98.18 (97.71–98.64)	<0.001
NPV	97.58 (97.05–98.11)	95.23 (95.13–95.89)	<0.001
PPV	96.11 (95.31–96.92)	98.12 (97.64–98.60)	<0.001

The values of these five assessments were calculated by 10-fold cross validation.



**Figure 4** Receiver operating characteristic (ROC) curves and the area under the ROC curve values for the combined and binary models.

**Table 2** Accuracy of two basal models (NH and NU models), the combined model, and the binary model according to small bowel abnormal lesions

	NH model	NU model	Combined model	Binary model
N –	97.57%	97.44%	95.79%	98.21%
H Red spot	96.22%	81.29%	97.48%	92.81%
Angioectasia	100.00%	100.00%	100.00%	100.00%
Active bleeding	99.77%	95.45%	99.77%	98.60%
U Erosion	87.87%	96.07%	97.15%	95.33%
Ulcer	84.29%	95.90%	96.86%	94.40%
Stricture	81.16%	97.10%	97.58%	94.69%

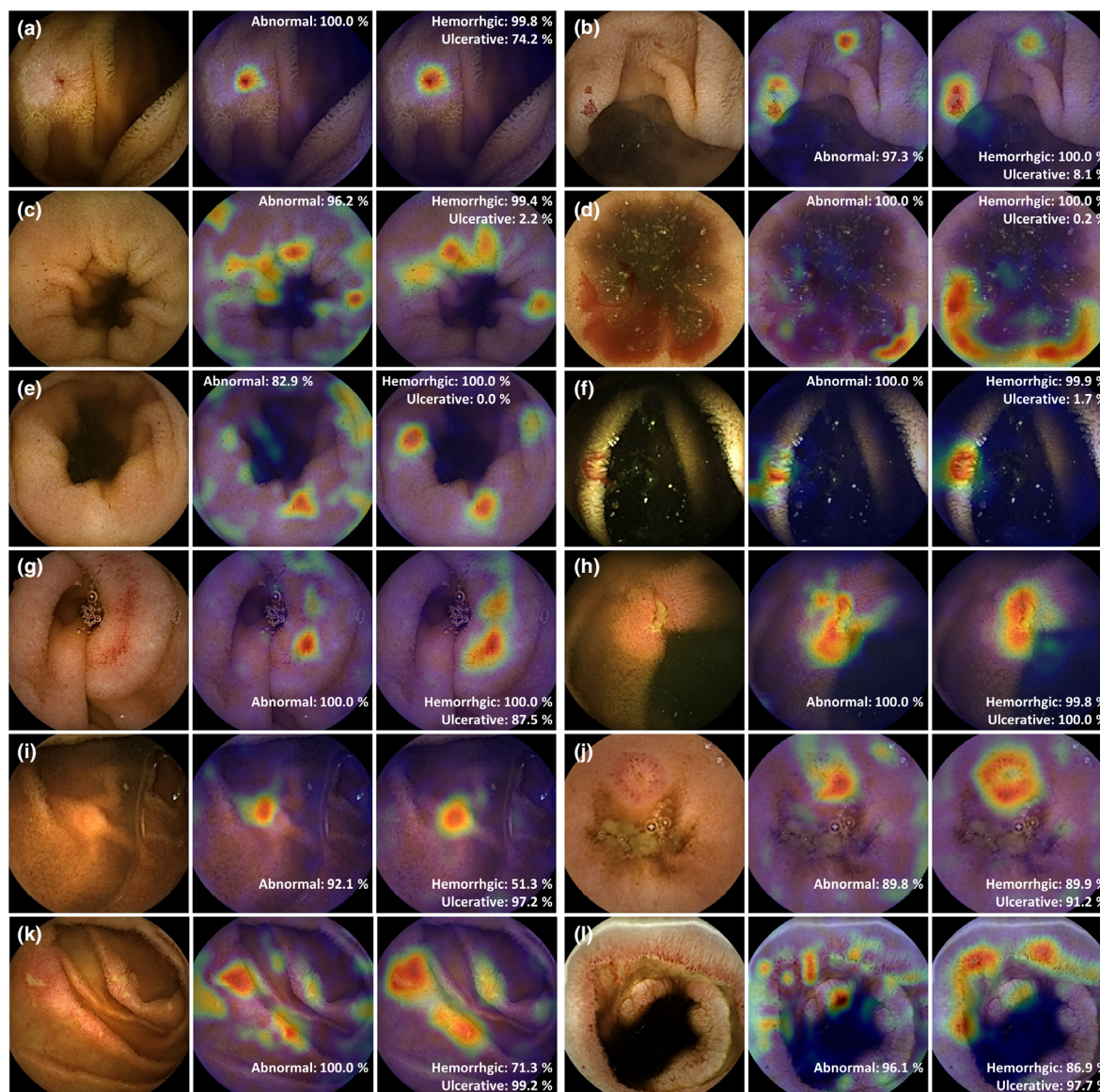
H, Hemorrhagic; N, Normal; U, Ulcerative.

model showed higher sensitivity and localized the abnormal lesions more accurately in the Grad-CAM than the binary model.

Recent CNN studies using SBCE images have suggested that CNNs could be powerful and efficient in detecting lesions and abnormalities in the small bowel. Aoki *et al.* developed a CNN algorithm for erosions and ulcers.<sup>14</sup> Their model located lesions accurately within the bounding box and yielded 88.2% sensitivity and 90.9% specificity. In addition, Leenhardt *et al.*<sup>15</sup> explored a segmentation

approach and demonstrated a CNN model that could precisely localize lesions. The CNNs tested by both Aoki *et al.* and Leenhardt *et al.* required a bounding box or pixel-level annotation on each image for lesion detection. Such a labeling task is too costly to adopt in clinical practice. On the other hand, Ding *et al.*<sup>13</sup> compared a CNN-based auxiliary reading of about 5000 videos with conventional readings by 20 gastroenterologists and verified a diagnostic performance of close to 100% sensitivity. They emphasized the value of CNN-based auxiliary reading as a time-saving tool that reduced the reading time. However, they did not notice the significance of its localization ability, which is another potential of CNN and CAM.

To the best of our knowledge, no one has previously attempted to localize small bowel lesions without manual annotation or combined two separate models to detect different types of small bowel lesions. Our study proposes a combined approach that unites two CNN models to detect hemorrhagic and ulcerative lesions separately. Our proposed method has merits in four aspects: sensitivity, localization (reliability), multi-lesion detection, and expandability. First, sensitivity is critical in lesion detection because it indicates how well the model can detect abnormalities with no omissions. The proposed approach enables us to boost the sensitivity of the model without revising it. Second, unlike a binary model that extracts a single feature set to differentiate



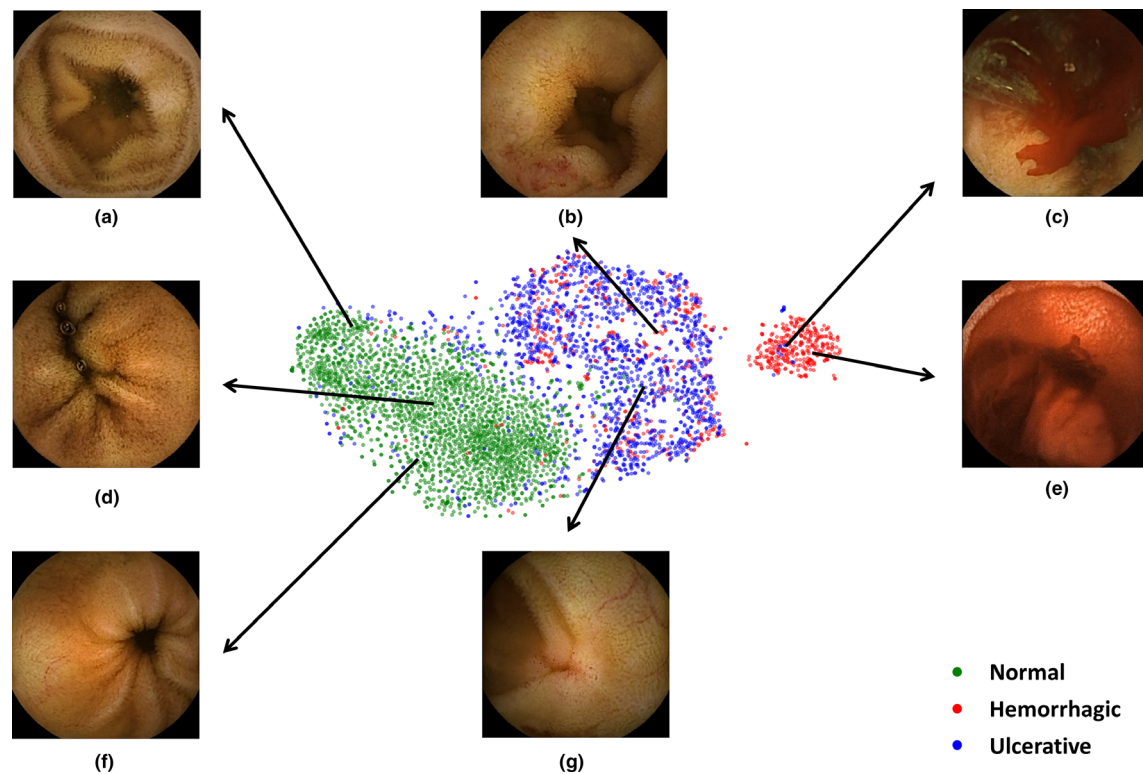
**Figure 5** Comparison of the lesion localization ability of the combined and binary models using gradient-weighted class activation mapping (Grad-CAM). (a) Red spot. (b) Angioectasia. (c) Red spot. (d) Active bleeding. (e) Red spot. (f) Active bleeding. (g) Red spot. (h) Ulcer. (i) Erosion. (j) Erosion. (k) Ulcer. (l) Stricture (from the left: original capsule endoscopy image, Grad-CAM of the binary model, and Grad-CAM of the combined model).

between normal and abnormal classes, the proposed model considers two feature sets to detect hemorrhagic and ulcerative lesions. As shown in Figure 5, the proposed model provides higher detection precision in localizing the lesions than the binary model. Third, because hemorrhagic and ulcerative lesions can be found together in SBCE images, a model is needed to distinguish them intelligently. Unlike a multi-class classifier that takes into account the ranking probability that both classes are present, the proposed model

shows two separate probabilities for hemorrhagic and ulcerative lesions in a single SBCE image. Fourth, the conventional binary model must be re-trained entirely from scratch to detect different types of lesions (e.g., a polyp or tumor) as abnormal. On the other hand, the proposed method would only need to train additional models and then combine them with the existing basal NH and NU models.

We compared the localization ability of the combined model with that of the binary model and found two





**Figure 6** Visualization of three different pathology findings using the t-stochastic neighbor embedding algorithm.

important results. First, our AI model could locate abnormalities without requiring manual annotation such as a bounding box or segmentation label. Second, the localization ability of our model is associated with its reliability because the CAM indirectly provides information about how much the region contributes to the CNN's decision.

Concerning the interpretability in clinical practice, CNNs have an important limitation, despite the remarkable performance shown here. Deep learning methods, such as CNNs, have been a “black-box” model, in which the inputs and outputs are observable, but how the internal algorithm works is not.<sup>21</sup> When two CNN models offer similar performance and detect similar lesions, it is important to determine which one is more reliable. In this study, we considered the CNN model with better localization ability to be more reliable. Therefore, we compared localization ability and confirmed that the combined model was more reliable, in addition to achieving higher accuracy. Explainable and interpretable AI models have been attracting attention recently to overcome this limitation.

Nonetheless, this study has several limitations. First, the 3778 pathological SBCE still images used in model training are not enough to ensure that this algorithm is suitable for practical clinic use. A well-designed study using full SBCE

videos of a large number of cases should be conducted to verify the clinical accuracy of the proposed model before the CNN-based combined model is used in clinical practice. Second, because abnormal images such as a tumor or diverticulum were rare, it was difficult to train on them. However, our models could detect some cases of small bowel tumors when they bled or had erosions or ulcers at their surface. If sufficient image acquisition were achieved for such lesions, individual models for them could be trained and merged into the existing combined model. Third, we could not compare our models with existing software enhancements such as Quick-View because our CNN algorithm was developed and validated based on still images not videos. We are planning to compare them in our next prospective validation study.

In conclusion, we developed and validated CNN-based AI models with Grad-CAM, and they distinguished abnormal small bowel images with high accuracy and sensitivity. The proposed combined model not only provides effective detection but also classifies each type of lesion separately. Furthermore, this model demonstrates an improved localization ability compared with the binary model, providing a more reliable lesion detection approach for real-world clinical practice.

## CONFLICT OF INTERESTS

**A**UTHORS DECLARE NO conflicts of interest for this article.

## FUNDING INFORMATION

**T**HIS RESEARCH WAS supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Science & ICT (NRF-2018M3A9E8021507). This work was also supported by the Po-Ca Networking Groups funded by The Postech-Catholic Biomedical Engineering Institute (PCBMI) (No. 5-2019-B0001-00118).

## REFERENCES

- Iddan G, Meron G, Glukhovskiy A, Swain P. Wireless capsule endoscopy. *Nature* 2000; **405**: 417.
- Lewis BS, Swain P. Capsule endoscopy in the evaluation of patients with suspected small intestinal bleeding: Results of a pilot study. *Gastrointest Endosc* 2002; **56**: 349–53.
- Hartmann D, Schmidt H, Bolz G *et al.* A prospective two-center study comparing wireless capsule endoscopy with intraoperative enteroscopy in patients with obscure GI bleeding. *Gastrointest Endosc* 2005; **61**: 826–32.
- Saurin JC, Lapalus MG, Cholet F *et al.* Can we shorten the small-bowel capsule reading time with the “Quick-view” image detection system? *Dig Liver Dis* 2012; **44**: 477–81.
- Yung DE, Sykes C, Koulaouzidis A. The validity of suspected blood indicator software in capsule endoscopy: A systematic review and meta-analysis. *Expert Rev Gastroenterol Hepatol* 2017; **11**: 43–51.
- Saurin JC, Jacob P, Heyries L *et al.* Multicenter prospective evaluation of the express view reading mode for small-bowel capsule endoscopy studies. *Endosc Int Open* 2018; **6**: E616–21.
- Cheng JZ, Ni D, Chou YH *et al.* Computer-aided diagnosis with deep learning architecture: Applications to breast lesions in US images and pulmonary nodules in CT scans. *Sci Rep* 2016; **6**: 24454.
- Kooi T, Litjens G, van Ginneken B *et al.* Large scale deep learning for computer aided detection of mammographic lesions. *Med Image Anal* 2017; **35**: 303–12.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: He K, Zhang X, Ren S, Sun J (eds). *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016 Jun 27–30. Las Vegas, NV: IEEE, 2016; 770–8.
- Araújo T, Aresta G, Castro E *et al.* Classification of breast cancer histology images using Convolutional Neural Networks. *PLoS One* 2017; **12**: e0177544.
- Esteva A, Kuprel B, Novoa RA *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 2017; **542**: 115–8.
- 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**: 653–60.
- 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**: 1044–54.e5.
- 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**: 357–63.e2.
- Leenhardt R, Vasseur P, Li C *et al.* A neural network algorithm for detection of GI angiectasia during small-bowel capsule endoscopy. *Gastrointest Endosc* 2019; **89**: 189–94.
- Alaskar H, Hussain A, Al-Aseem N, Liatsis P, Al-Jumeily D. Application of convolutional neural networks for automated ulcer detection in wireless capsule endoscopy images. *Sensors* 2019; **19**: 1265.
- Fan S, Xu L, Fan Y, Wei K, Li L. Computer-aided detection of small intestinal ulcer and erosion in wireless capsule endoscopy images. *Phys Med Biol* 2018; **63**: 165001.
- Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv* 2014; 1409.1556.
- Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D. Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (eds). *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*; 2017 Oct 22–29. Venice, Italy: IEEE, 2017; 618–26.
- van der Maaten L, Hinton G. Visualizing data using t-SNE. *J Mach Learn Res* 2008; **9**: 2579–605.
- LeCun Y, Bengio Y, Hinton G. Deep learning. *Nature* 2015; **521**: 436–44.

## SUPPORTING INFORMATION

**A**DDITIONAL SUPPORTING INFORMATION may be found in the online version of this article at the publisher’s web site.

**Figure S1** The results of data preprocessing (top) and data augmentation (bottom).

**Figure S2** False-negative images from the binary and combined models: (a) four of the 47 false-negative images in the binary model, (b) four of the 23 false-negative images in the combined model.

**Figure S3** Various false-positive images that were misdiagnosed in both of the binary and combined models: (a) wrinkles. (b) vascular structures. (c) bubbles.