



DYNAMIC MANUSCRIPT

# Real-time detection of active bleeding in laparoscopic colectomy using artificial intelligence

Kenta Horita<sup>1</sup> · Koya Hida<sup>1</sup> · Yoshiro Itatani<sup>1</sup> · Haruku Fujita<sup>1</sup> · Yu Hidaka<sup>2</sup> · Goshiro Yamamoto<sup>3</sup> · Masaaki Ito<sup>4,5</sup> · Kazutaka Obama<sup>1</sup>

Received: 12 February 2024 / Accepted: 20 April 2024

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

## Abstract

**Background** Most intraoperative adverse events (iAEs) result from surgeons' errors, and bleeding is the majority of iAEs. Recognizing active bleeding timely is important to ensure safe surgery, and artificial intelligence (AI) has great potential for detecting active bleeding and providing real-time surgical support. This study aimed to develop a real-time AI model to detect active intraoperative bleeding.

**Methods** We extracted 27 surgical videos from a nationwide multi-institutional surgical video database in Japan and divided them at the patient level into three sets: training ( $n=21$ ), validation ( $n=3$ ), and testing ( $n=3$ ). We subsequently extracted the bleeding scenes and labeled distinctively active bleeding and blood pooling frame by frame. We used pre-trained YOLOv7<sub>6w</sub> and developed a model to learn both active bleeding and blood pooling. The Average Precision at an Intersection over Union threshold of 0.5 (AP.50) for active bleeding and frames per second (FPS) were quantified. In addition, we conducted two 5-point Likert scales (5 = Excellent, 4 = Good, 3 = Fair, 2 = Poor, and 1 = Fail) questionnaires about sensitivity (the sensitivity score) and number of overdetection areas (the overdetection score) to investigate the surgeons' assessment.

**Results** We annotated 34,117 images of 254 bleeding events. The AP.50 for active bleeding in the developed model was 0.574 and the FPS was 48.5. Twenty surgeons answered two questionnaires, indicating a sensitivity score of 4.92 and an overdetection score of 4.62 for the model.

**Conclusions** We developed an AI model to detect active bleeding, achieving real-time processing speed. Our AI model can be used to provide real-time surgical support.

---

✉ Koya Hida  
hidakoya@kuhp.kyoto-u.ac.jp

<sup>1</sup> Department of Surgery, Kyoto University Graduate School of Medicine, 54 Shogoin-Kawahara-Cho, Sakyo-Ku, Kyoto 606-8507, Japan

<sup>2</sup> Department of Biomedical Statistics and Bioinformatics, Kyoto University Graduate School of Medicine, Kyoto, Japan

<sup>3</sup> Division of Medical Information Technology and Administration Planning, Kyoto University, Kyoto, Japan

<sup>4</sup> Surgical Device Innovation Office, National Cancer Center Hospital East, Chiba, Japan

<sup>5</sup> Department of Colorectal Surgery, National Cancer Center Hospital East, Chiba, Japan

## Graphical abstract

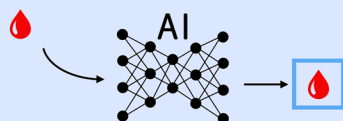
# Real-time detection of active bleeding in laparoscopic colectomy using artificial intelligence.

## Aim



- To develop an AI model for monitoring the entire surgical view constantly and detecting active bleeding in laparoscopic colectomy

## Methods



- Training 26,581 images from 205 bleeding scenes
- 20 surgeons assessed the AI model by five-point Likert scales (from 1=Fail to 5=Excellent)

## Results



AP.50 was 0.574



48.5 frames per second



a sensitivity score was 4.92  
an over-detection score was 4.62

## Conclusions

Our AI model can detect active bleeding with a real-time processing speed, and the clinical performance was satisfactory.

AP.50 indicates the Average Precision at an intersection over union threshold of 0.5.  
Over-detection score reflects number of over-detection areas.

**Keywords** Artificial intelligence · Bleeding · Object detection · Surgery

Most intraoperative adverse events (iAEs) occur due to surgeons' technical errors and inappropriate judgment. iAEs have a significant impact on the patient's postoperative course and have been shown to increase 30-day mortality, 30-day morbidity, postoperative length of stay [1], and hospital costs [2]. Recent developments in the field of minimally invasive surgery (MIS), including laparoscopic and robotic surgeries, have made it easier to record and analyze surgical procedures, including iAEs. Therefore, many surgical video assessment methods by human evaluators have been developed [3–9], and these methods have revealed that bleeding is a major iAE [5–8].

Bleeding has a direct impact on patients, and greater blood loss has been shown to cause worse postoperative outcomes [10, 11]. In addition, we experimentally understand that bleeding has indirect negative effects and sometimes leads to additional severe iAEs as suggested by Heinrich's pyramid [9]. For example, when bleeding flows into and stains loose connective tissues during left colorectal surgery, surgeons may misidentify the correct dissection plane and injure the autonomic nerve and ureter. The recognition and appropriate control of even small amounts of bleeding are important to ensure safe surgery. However, bleeding caused by too much retraction or assistant surgeons may arise apart from the dissection

area. Therefore, bleeding can arise at the edge of the surgical view or be hidden behind instruments, preventing its detection timely. To assist surgeons in monitoring the entire surgical view constantly and detecting bleeding, we have used artificial intelligence (AI) and computer vision (CV) technologies.

Over the past decade, following the advent of deep-learning algorithms, AI and CV technologies have rapidly developed, with their use spreading to several medical fields, including radiological diagnosis [12], dermatology [13], and gastrointestinal endoscopy [14, 15]. In surgical videos of MIS, many AI models have been reported for anatomy identification [16–22], instrument identification [23, 24], surgical phase recognition [25, 26], and surgical skill assessment [27–30]. However, AI and CV models that focus on iAEs, including bleeding, have rarely been studied. A few models [31–33] aimed to classify pixels into blood or non-blood pixels, but these were not designed to distinguish whether the blood pixels were from sites of active bleeding. Such real-time differentiation of active bleeding from blood pooling is essential for surgical support.

To address this unmet need, the present study aimed to develop an AI and CV model that could detect active bleeding in the surgical field.

## Materials and methods

### Study design

The study protocol was reviewed and approved by the Ethics Committee of Kyoto University Graduate School and Faculty of Medicine (approval number: R-3614). This study was conducted in accordance with the Checklist for Artificial Intelligence in Medical Imaging (CLAIM) [34] and conformed to the provisions of the Declaration of Helsinki in 1964 [35] (revised in Brazil in 2013).

We further accessed a nationwide surgical video database containing videos of 3558 cases taken in 71 institutions in Japan. This database was supported by the Japan Agency for Medical Research and Development (AMED) under Grant Number JP21he2102001h0003. Written informed consent was obtained from all patients. All data were completely anonymized before being accessed.

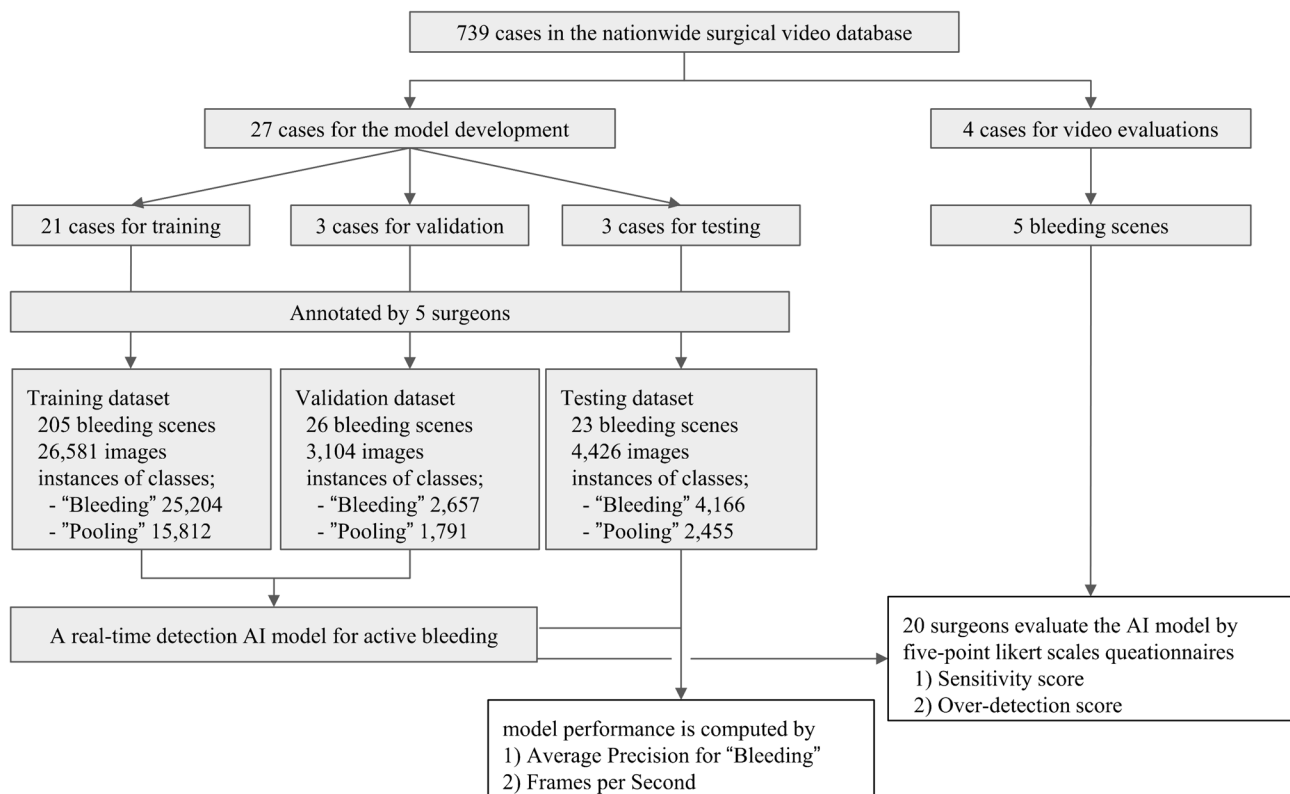
### Dataset

We randomly extracted 27 videos of 739 laparoscopic sigmoidectomies and high anterior resection videos from a nationwide database to develop AI models, maintaining

an equal distribution of the three scope systems (Olympus, KARL STORZ, and Stryker). We divided the 27 videos into three groups: 21 for training, 3 for validation, and 3 for testing. We estimated that more than 160 bleeding scenes were required for training from a previous study [14] and determined that 21 videos were sufficient for this purpose according to our pilot study. The data split in this study at the patient level prevented contamination of the validation and test data into training data and ensured no data leakage. The flow of the data splitting is shown in Fig. 1.

### Annotation

Before annotation, five surgeons, including the first author, established annotation protocols and trained video editing and image labeling for 6 months. In the first step of annotation, bleeding scenes were extracted at 1920×1080 resolution at 30 frames per second (fps) from the entire surgical video and the surgical phases of the bleeding scenes were recorded according to the definition of a previous study [25]. The inclusion criteria for bleeding scenes were defined as scenes where active bleeding was clearly observed, and the exclusion criteria were any scenes determined as poor-quality views due to surgical smoke, bleu lens, or halation.



**Fig. 1** The flowchart of the model development and evaluation

The bleeding scene video clips were then down sampled to 10 fps to reduce annotation costs, and each frame was labeled in bounding box formats using CVAT, a free open-source annotation tool (<https://cvat.org>), by the same five surgeons. Two classes were labeled: “bleeding” and “pooling.” “Bleeding” refers to any active flowing bleeding, whereas “pooling” refers to non-flowing blood pooling with a liquid component or judged by annotators to be potentially misidentified as active bleeding, such as blood residue and blood clots. The aim of adding “pooling” was to indicate explicitly the difference between active bleeding and blood pooling to thereby improve the detection of active bleeding. Although the stained tissue and blood vessels were also red and visually resembled “bleeding” and “pooling,” they were not labeled as the annotation costs were considered too great.

All annotation data were reviewed and verified frame by frame by the first author, and confirmation of the annotation data was performed by the chief surgeon of the colorectal surgery team at Kyoto University Hospital, not the annotator. Examples of annotated images are presented in Figure S1.

## Model development

We selected “YOLOv7” series [36], a state-of-the-art real-time object detector, and used pre-trained “YOLOv7\_w6” because the default input resolution was 1280 and the model size was suitable for our computational resources. The codes and pre-trained weights for fine-tuning were used in MMYOLO [37], an open-source library for object detection. The batch size was set to eight, which was the maximum memory capacity of our GPU, RTX A6000 (NVIDIA Corp., Santa Clara, California, USA), with 48 GB of VRAM, and the anchor box size was optimized using the scripts provided by MMYOLO. Our preliminary study revealed that changes in hyper parameters and data augmentation settings lead to the same or slightly worse results; thus, we used the default hyper parameters and data augmentation settings provided by MMYOLO.

We trained the model for 100 epochs using training datasets including 2 classes, “bleeding” and “pooling,” and selected the best epoch models based on the average precision (AP) for the class “bleeding” on validation datasets. For the model performance evaluation and video assessment by the surgeons, we used the best epoch model. To verify the effect of two class labels, we further trained the model with the same settings using the datasets including only the class “bleeding” (1 class-model) and compared it to the model including the two classes (2 classes-model).

## Evaluation of model performance

To evaluate the model performances, we used the test datasets and calculated the AP for the class “bleeding.” AP is a commonly used performance metric to assess object detection model performance, which was computed according to the method of the Microsoft Common Objects in Context [38]. The AP is scored between 0 and 1, with values closer to 1 being superior. The primary outcome of model performance was  $AP_{50}$ , which indicates that the Intersection over Union (IoU) threshold was set to 0.5.

We further evaluated the inference speed in frames per second (FPS) and set the threshold for real-time object detection to  $\geq 30$ . For the measurement of FPS, we used two types of computer machines: one with a Core i9 10980XE (18 core/36 thread, 3.0 GHz) with 64 GB of RAM and an RTX A6000 (NVIDIA Corp., Santa Clara, California, USA) with 48 GB of VRAM and the other with a Core i5 12,400 (6 core/12 thread, 2.5–4.4 GHz) with 32 GB of RAM and an RTX 3060 (NVIDIA Corp., Santa Clara, California, USA) with 12 GB of VRAM.

## Video assessment by surgeons

Although AP is the gold standard for evaluating object detection models, the association between AP and clinical meaning is unclear, and the interpretation of AP is difficult for surgeons. Therefore, we created two 5-point Likert scales (5 = Excellent, 4 = Good, 3 = Fair, 2 = Poor, and 1 = Fail) to investigate the surgeons’ subjective assessment of our AI models, with reference to a previous study [17]. The first question was “How much did the AI model fail to detect an actual detect active bleeding?” (sensitivity score). The answers were provided on a 5-point scale in 20% increments (from a score of 1 for 80–100% failure to detect active bleeding to a score of 5 for 0–20% failure to detect active bleeding). The second question was “How many active bleeding areas did the AI model over detect?” (overdetection score). The answers were provided on a 5-point scale (a score of 1 for four or more over detection areas, 2 for three, 3 for two, 4 for one, and 5 for no overdetection areas). “Overdetection” was defined as the misrecognition of a blood pooling area as active bleeding lasting for more than 1 s (accumulation if intermittent).

For video evaluation, we extracted five other bleeding scenes from four videos that were not used in the model development and created inference video clips for each model with a confidence score threshold of 0.5. The duration of the inference video clip was approximately 15 s. Twenty surgeons, who did not overlap with the annotators, assessed the video clips and answered two questionnaires.

## Statistical analysis

All statistical analyses were performed using Python (V.3.10.12), and we used the numpy (V1.23.5), pandas (V.1.5.3), and matplotlib (V.3.7.1) libraries. Continuous variables were expressed as the means and standard deviations, and categorical variables were expressed as numbers and percentages.

## Results

### Datasets

The patient characteristics are summarized in Table 1. For the training dataset, we extracted 205 bleeding scenes from 21 videos and annotated a total of 26,581 images, which contained 25,204 instances of “bleeding” and 15,812 instances of “pooling.” We further extracted 26 bleeding scenes from three videos for the validation dataset, from which we annotated 3104 images, and a further 23 bleeding scenes from three videos for the test dataset, from which we annotated 4432 images (Fig. 1). The training dataset contained all the surgical phases (Table S1).

### Model performances and surgeon’s assessments

The  $AP_{50}$  for the class “bleeding” of the 2 classes-model (0.574) was higher than the 1 class-model (0.560). The FPS using RTX A6000 was both over 30; 48.5 (2

classes-model) vs 50.4 (1 class-model), while the FPS using RTX 3060 was 23.3 vs 23.9, respectively (Table 2).

The surgeons’ questionnaires revealed that the 2 classes-model had a slightly higher sensitivity score of 4.92 (2 classes-model) vs 4.88 (1 class-model). Further, the 2 classes-model had slightly more over detection areas, with over detection scores of 4.62 (2 classes-model) vs 4.75 (1 class-model) (Table 3 and Fig. 2). The inference video clips used for the surgeon’s assessments are shown in Videos S1–5 and screenshots of the video clips are shown in Fig. 3. Our model misrecognized residual blood as active bleeding (Fig. 3a and b); however, it could detect active bleeding from the onset (Fig. 3b) or find active bleeding hidden under the instruments (Fig. 3e).

**Table 2** Average precision (AP) and frames per second (FPS) at test dataset

|                 | $AP_{50}$ for bleeding | $AP_{50}$ for pooling | FPS with RTX A6000 | FPS with RTX 3060 |
|-----------------|------------------------|-----------------------|--------------------|-------------------|
| 2 classes-model | 0.574                  | 0.068                 | 48.5               | 23.3              |
| 1 class-model   | 0.560                  | –                     | 50.4               | 23.9              |

**Table 1** Patient characteristics

|                           | Training ( $N=21$ ) | Validation ( $N=3$ ) | Test ( $N=3$ ) |
|---------------------------|---------------------|----------------------|----------------|
| Age (years)*              | 65.7 (12.6)         | 67.7 (13.1)          | 68.0 (9.2)     |
| Sex ratio (M: F)          | 12: 9               | 3: 0                 | 1: 2           |
| BMI*                      | 24.3 (5.3)          | 25.5 (2.6)           | 19.1 (0.7)     |
| Abdominal surgery history | 7 (33.3)            | 0 (0)                | 3 (100)        |
| Sigmoidectomy/HAR         | 19 / 2              | 2 / 1                | 3 / 0          |
| cStage                    |                     |                      |                |
| I                         | 4 (19.0)            | 1 (33.3)             | 0 (0)          |
| II                        | 6 (28.6)            | 2 (66.6)             | 0 (0)          |
| III                       | 11 (52.4)           | 0 (0)                | 3 (100)        |
| IV                        | 0 (0)               | 0 (0)                | 0 (0)          |
| Operator                  |                     |                      |                |
| Years*                    | 12.0 (9.7)          | 10.0 (3.0)           | 10.3 (12.9)    |
| JSES qualified surgeon    | 6 (28.6)            | 0 (0)                | 1 (33.3)       |
| Scopist                   |                     |                      |                |
| Years*                    | 6.2 (4.9)           | 5.7 (7.2)            | 4.0 (6.9)      |
| Surgical time (min)*      | 212.6 (70.3)        | 216.3 (77.0)         | 204.3 (68.1)   |
| Blood loss (g) *          | 7.9 (11.8)          | 16.7 (28.9)          | 1.0 (1.7)      |

Values are  $n$  or  $n$  (%) unless indicated otherwise; values are \*mean (s.d.). HAR: High antrectomy. JSES: Japan Society for Endoscopic Surgery

**Table 3** Video assessment by 20 surgeons

(a) Q1: How much did the AI model fail to detect an actual active bleeding? (sensitivity score): mean (s.d.)

| Video clips       | No. 1       | No. 2       | No. 3       | No. 4       | No. 5       | Total       |
|-------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Two classes-model | 4.90 (0.45) | 4.95 (0.22) | 4.95 (0.22) | 4.90 (0.31) | 5.00 (0.00) | 4.94 (0.28) |
| One class-model   | 4.90 (0.45) | 4.95 (0.22) | 4.95 (0.22) | 4.65 (0.59) | 4.95 (0.22) | 4.88 (0.38) |

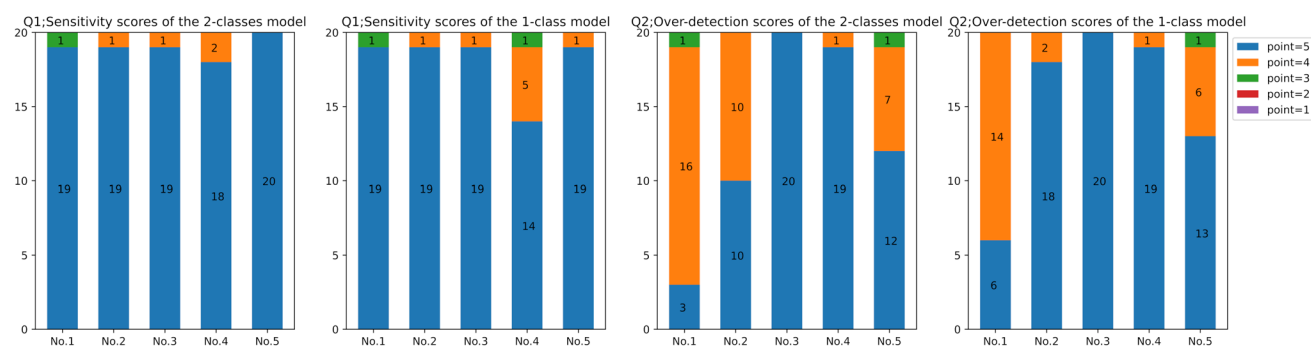
(b) Q2: How many active bleeding areas did the AI model over detect? (overdetection score): mean (s.d.)

| Video clips       | No. 1       | No. 2       | No. 3       | No. 4       | No. 5       | Total       |
|-------------------|-------------|-------------|-------------|-------------|-------------|-------------|
| Two classes-model | 4.10 (0.45) | 4.50 (0.51) | 5.00 (0.00) | 4.95 (0.22) | 4.55 (0.60) | 4.62 (0.53) |
| One class-model   | 4.30 (0.47) | 4.90 (0.31) | 5.00 (0.00) | 4.95 (0.22) | 4.60 (0.60) | 4.75 (0.46) |

Two 5-point Likert scales (5 = Excellent, 4 = Good, 3 = Fair, 2 = Poor, and 1 = Fail) questionnaires

(a) The answers are provided on a 5-point scale in 20% increments (from a score of 1 for 80–100% failure to detect active bleeding to a score of 5 for 0–20% failure to detect active bleeding)

(b) A score of 1 for four or more overdetection areas, 2 for three, 3 for two, 4 for one, and 5 for no overdetection areas

**Fig. 2** The point distributions of the 5-point Likert scales (5 = Excellent, 4 = Good, 3 = Fair, 2 = Poor, 1 = Fail) questionnaires. Q1 is “How much did the AI model fail to detect an actual detect active

bleeding?” (sensitivity score). Q2 is “How many active bleeding areas did the AI model over detect?” (overdetection score)

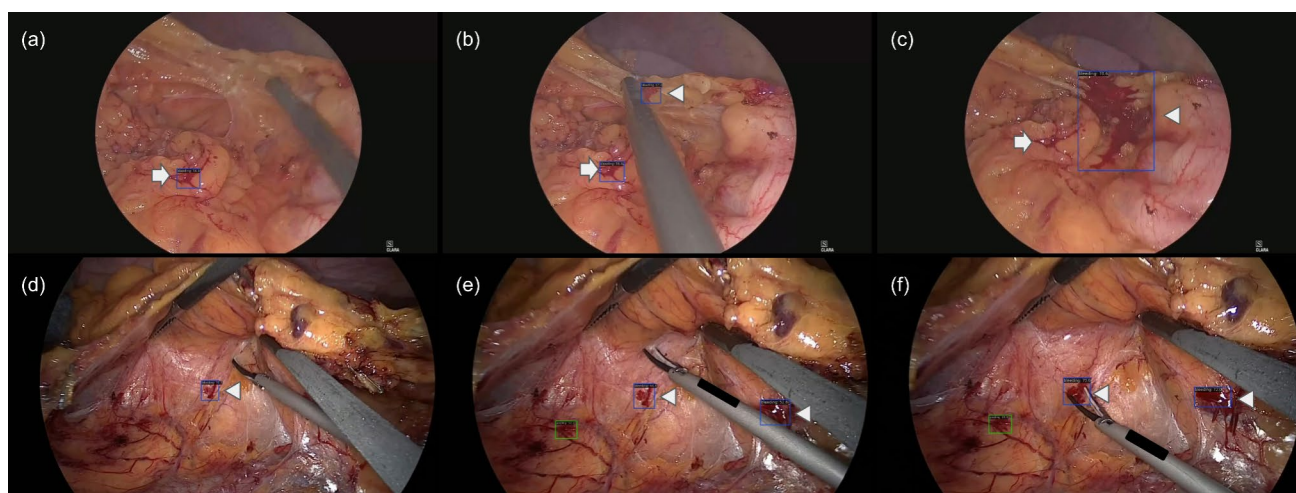
## Discussion

In this study, we successfully developed a real-time AI model that could detect active bleeding during laparoscopic colectomy. The AP for active bleeding was 0.574 for the 2 classes-model and 0.560 for the 1 class-model, with FPS values of 48.5 and 50.4, respectively. The clinical performance was appropriately assessed by surgeons watching the video clips, and notably, both models had very high sensitivity; the sensitivity score was 4.94 for the 2 classes-model and 4.88 for the 1 class-model. Therefore, our model offers real-time surgical support by monitoring active bleeding.

The AP of our models was lower than that for detecting surgical instruments [23, 24], but higher than that of the models for detecting anatomical structures [18, 22]. AI and CV models can easily detect surgical instruments as they are relatively large and have textures that are clearly different from those of the human tissues in the background,

such as fat and the gastrointestinal tract. However, bleeding and anatomical structures occupy only a small portion of the surgical field with ambiguous boundaries, and their recognition requires expertise. In addition, the fluidity of bleeding results in the creation of indefinite shapes. To overcome these difficulties, we annotated a total of 34,117 images, which is more than 10 times larger than those of previous studies for the detection [18, 22] or segmentation [16, 17, 20, 21] of anatomical structures. This extensive annotation data may have allowed our models to achieve a higher AP.

Our 2 classes-model had a slightly higher AP and sensitivity score than that of the 1 class-model, while the overdetection score of the 2 classes-model was somewhat lower than that of the 1 class-model. Namikawa et al. previously reported that learning about gastric ulcers, not only gastric cancer, improved the positive predictive value of their AI model for detecting gastric cancer [39]. Thus, we expected that the 2 classes-model would reduce false positives, resulting in a higher overdetection score. However,



**Fig. 3** Screenshots of inference videos of 2 classes-model, which was assessed by surgeons. Blue bounding box is the class “bleeding,” active bleeding. Green bounding box is the class “pooling,” blood pooling (e.g., blood residue and blood clots). White arrowhead represents active bleeding areas and white arrow represents blood residual areas. **a–c** are sequences of the inference video no.1 (see Video S1),

and the AI model can detect active bleeding from the onset (white arrowhead in **b**). On the contrary, the model misrecognized a blood residual area as active bleeding (white arrow in **a** and **b**). **d–f** are sequences of the inference video no. 5 (see Video S5), and the AI model can find it hidden under instruments (white arrowhead in **e**) (Color figure online)

this approach showed no expected improvement in this study and increased annotation costs. There are two reasons for this finding. First, active bleeding is obviously brighter than blood pooling, and learning the class “bleeding” alone was sufficient to recognize active bleeding. In fact, annotators can indicate active bleeding areas using still images to some extent, focusing on subtle differences in color tone and texture. Second, our annotation protocol of “pooling” was not appropriate. We defined the class “pooling” as blood residue and blood clots. However, there would have been countless blood residues and blood clots in the surgical field if small and trivial ones were included as well. We could not label everything because of our limited annotation resources, and therefore labeled only relatively large ones. Thus, the definition of the class “pooling” was inconsistent, and our model did not learn the class “pooling” well, which reflected the very low AP for the class “pooling” (0.068).

Surgeons’ assessments showed a very high sensitivity score, which is desirable for surgical support. Surgeons and AI should work together, not independently, and when considering that missing active bleeding is critical, the collaboration of high-sensitivity AI with surgeons could greatly contribute to achieving the goal of safe surgery. Collaboration between humans and AI has been reported to be effective in endoscopic polyp detection [40, 41], and the same effects are expected during surgery. In particular, less experienced surgeons will benefit more from collaboration.

To further improve the performance of the proposed model, we propose three approaches. First, we could gather additional annotated data; this approach is simple and the

most robust but would require expensive annotation costs. Following this, unsupervised learning, semi-supervised learning, and self-supervised learning can be used to reduce annotation costs. The second approach would be to utilize temporal and interframe features, such as optical flow, long short-term memory (LSTM), tracking, and 3D convolutional neural networks (3D-CNN), which have been reported to be effective in surgical skill assessment models [27, 28, 30]. Temporal and interframe features may enable models to deal with motion information (e.g., spreading or spurting); however, the inference speed will be significantly slower. The third is the use of vision transformer models [42], which have recently garnered significant attention. In our preliminary study, we trained the “deformable DETR” [43], a vision transformer model, but the AP of this model was slightly lower, and the FPS was less than 40% than that of the “YOLOv7\_w6” model. Vision transformer models have evolved at a tremendous rate, and models with higher accuracy and speed are expected to emerge in the next few years. These three approaches will improve AP, but these improvements may be difficult for surgeons to notice, considering the high evaluation scores in the surgeons’ assessment in the present study.

This study had several limitations. First, bleeding scenes were extracted using only two procedures: laparoscopic sigmoidectomy and high anterior resection. Hence, our model performances for other surgical procedures such as right hemicolectomy and gastrectomy were not secure because the detection performances of the AI and CV models were significantly affected by the background. Although bleeding is

common and observed in all surgical procedures, additional validation is required to generalize our model to other procedures. Second, the performance of our model may not be sufficiently robust when detecting very massive or spurting bleeding as these situations were not included in the training dataset. Indeed, we could not find any cases of severe bleeding in our datasets, possibly due to a selection bias in which surgical videos including severe bleeding could not be provided in the nationwide database. Third, there is a time lag associated with display speed, which was not investigated in this study. This time lag arises from the fact that the display speed in practical settings depends not only on the FPS of the model, but also on the data transmission speed. Although an FPS > 30 is a necessary condition for real-time applications, this time lag should be measured during actual use in the operating room, and its acceptability must be assessed by surgeons. Finally, the clinical impact of our models on actual surgical outcomes, such as decreased bleeding, organ injury, and better postoperative courses, remains unclear, and a prospective assessment will be necessary to evaluate their practical use as a next step.

In conclusion, we developed an AI model that can detect active bleeding with a real-time processing speed. In addition, we recruited surgeons to assess the validity of the models and found that their clinical performance was satisfactory. Thus, our AI model can provide real-time surgical support and contribute to assisting in improving the safety of surgeries.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s00464-024-10874-z>.

**Acknowledgements** We would like to thank Yoshiharu Sakai, Tomoaki Okada, and all the staff at the National Cancer Center Hospital East for the creation of the nationwide surgical video database; Yusuke Fujii, Kentaro Goto, and Yoshihisa Umamoto for preparing annotation data; Chang Liu for technical support; all the medical staff of the Department of Surgery at Kyoto University Hospital for answering the questionnaires; and Yugo Matsui and Editage ([www.editage.jp](http://www.editage.jp)) for English language editing.

**Funding** This study was supported by the Japan Agency for Medical Research and Development (AMED) under Grant Number JP21he2102001h0003 and Senko Medical Instrument.

## Declarations

**Disclosures** Kenta Horita, Koya Hida, Yoshiro Itatani, Haruku Fujita, Yu Hidaka, Goshiro Yamamoto, Masaaki Ito, and Kazutaka Obama have no conflicts of interest or financial ties to disclose

## References

- Bohnen JD, Mavros MN, Ramly EP, Chang Y, Yeh DD, Lee J, de Moya M, King DR, Fagenholz PJ, Butler K, Velmahos GC, Kaafarani HMA (2017) Intraoperative adverse events in abdominal surgery: what happens in the operating room does not stay in the operating room. *Ann Surg* 265:1119–1125
- Garbens A, Goldenberg M, Wallis CJD, Tricco A, Grantcharov TP (2018) The cost of intraoperative adverse events in abdominal and pelvic surgery: a systematic review. *Am J Surg* 215:163–170
- Vassiliou MC, Feldman LS, Andrew CG, Bergman S, Leffondré K, Stanbridge D, Fried GM (2005) A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg* 190:107–113
- Champagne BJ, Steele SR, Hendren SK, Bakaki PM, Roberts PL, Delaney CP, Brady JT, MacRae HM (2017) The American Society of colon and rectal surgeons assessment tool for performance of laparoscopic colectomy. *Dis Colon Rectum* 60:738–744
- Miskovic D, Ni M, Wyles SM, Parvaiz A, Hanna GB (2012) Observational clinical human reliability analysis (OCHRA) for competency assessment in laparoscopic colorectal surgery at the specialist level. *Surg Endosc* 26:796–803
- Foster JD, Miskovic D, Allison AS, Conti JA, Ockrim J, Cooper EJ, Hanna GB, Francis NK (2016) Application of objective clinical human reliability analysis (OCHRA) in assessment of technical performance in laparoscopic rectal cancer surgery. *Tech Coloproctol* 20:361–367
- Van Rutte P, Nienhuijs S, Jakimowicz J, Van Montfort G (2017) Identification of technical errors and hazard zones in sleeve gastrectomy using OCHRA. *Surg Endosc* 31:561–566
- Jung JJ, Jüni P, Gee DW, Zak Y, Cheverie J, Yoo JS, Morton JM, Grantcharov T (2020) Development and evaluation of a novel instrument to measure severity of intraoperative events using video data. *Ann Surg* 272:220–226
- Bonrath EM, Zevin B, Dedy NJ, Grantcharov TP (2013) Error rating tool to identify and analyse technical errors and events in laparoscopic surgery. *Br J Surg* 100:1080–1088
- Wu WC, Smith TS, Henderson WG, Eaton CB, Poses RM, Uttley G, Mor V, Sharma SC, Vezeridis M, Khuri SF, Friedmann PD (2010) Operative blood loss, blood transfusion, and 30-day mortality in older patients after major noncardiac surgery. *Ann Surg* 252:11–17
- Roshanov PS, Eikelboom JW, Sessler DI, Kearon C, Guyatt GH, Crowther M, Tandon V, Borges FK, Lamy A, Whitlock R, Bickard BM, Szczeklik W, Panju M, Spence J, Garg AX, McGillion M, VanHelder T, Kavsak PA, de Beer J, Winemaker M, Le Manach Y, Sheth T, Pinthus JH, Siegal D, Thabane L, Simunovic MRI, Mizera R, Ribas S, Devereaux PJ (2021) Bleeding Independently associated with Mortality after noncardiac Surgery (BIMS): an international prospective cohort study establishing diagnostic criteria and prognostic importance. *Br J Anaesth* 126:163–171
- McKinney SM, Sieniek M, Godbole V, Godwin J, Antropova N, Ashrafi H, Back T, Chesus M, Corrado GS, Darzi A, Etemadi M, Garcia-Vicente F, Gilbert FJ, Halling-Brown M, Hassabis D, Jansen S, Karthikesalingam A, Kelly CJ, King D, Ledsam JR, Melnick D, Mostofi H, Peng L, Reicher JJ, Romera-Paredes B, Sidebottom R, Suleyman M, Tse D, Young KC, De Fauw J, Shetty S (2020) International evaluation of an AI system for breast cancer screening. *Nature* 577:89–94
- Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542:115–118
- Misawa M, Kudo SE, Mori Y, Cho T, Kataoka S, Yamauchi A, Ogawa Y, Maeda Y, Takeda K, Ichimasa K, Nakamura H, Yagawa Y, Toyoshima N, Ogata N, Kudo T, Hisayuki T, Hayashi T, Wakamura K, Baba T, Ishida F, Itoh H, Roth H, Oda M, Mori K (2018) Artificial intelligence-assisted polyp detection for colonoscopy: initial experience. *Gastroenterology* 154:2027–2029.e2023
- Hirasawa T, Aoyama K, Tanimoto T, Ishihara S, Shichijo S, Ozawa T, Ohnishi T, Fujishiro M, Matsuo K, Fujisaki J, Tada T (2018) Application of artificial intelligence using a convolutional neural

- network for detecting gastric cancer in endoscopic images. *Gastric Cancer* 21:653–660
16. Madani A, Namazi B, Altieri MS, Hashimoto DA, Rivera AM, Pucher PH, Navarrete-Welton A, Sankaranarayanan G, Brunt LM, Okrainec A, Alseidi A (2020) Artificial intelligence for intraoperative guidance: using semantic segmentation to identify surgical anatomy during laparoscopic cholecystectomy. *Ann Surg*. <https://doi.org/10.1097/SLA.0000000000004594>
  17. Kumazu Y, Kobayashi N, Kitamura N, Rayan E, Neculoiu P, Misumi T, Hojo Y, Nakamura T, Kumamoto T, Kurahashi Y, Ishida Y, Masuda M, Shinohara H (2021) Automated segmentation by deep learning of loose connective tissue fibers to define safe dissection planes in robot-assisted gastrectomy. *Sci Rep* 11:21198
  18. Tokuyasu T, Iwashita Y, Matsunobu Y, Kamiyama T, Ishikake M, Sakaguchi S, Ebe K, Tada K, Endo Y, Etoh T, Nakashima M, Inomata M (2021) Development of an artificial intelligence system using deep learning to indicate anatomical landmarks during laparoscopic cholecystectomy. *Surg Endosc* 35:1651–1658
  19. Kitaguchi D, Takeshita N, Matsuzaki H, Igaki T, Hasegawa H, Kojima S, Mori K, Ito M (2022) Real-time vascular anatomical image navigation for laparoscopic surgery: experimental study. *Surg Endosc* 36:6105–6112
  20. Sato Y, Sese J, Matsuyama T, Onuki M, Mase S, Okuno K, Saito K, Fujiwara N, Hoshino A, Kawada K, Tokunaga M, Kinugasa Y (2022) Preliminary study for developing a navigation system for gastric cancer surgery using artificial intelligence. *Surg Today* 52:1753–1758
  21. Sato K, Fujita T, Matsuzaki H, Takeshita N, Fujiwara H, Mitsunaga S, Kojima T, Mori K, Daiko H (2022) Real-time detection of the recurrent laryngeal nerve in thoracoscopic esophagectomy using artificial intelligence. *Surg Endosc* 36:5531–5539
  22. Takeuchi M, Collins T, Lipps C, Haller M, Uwineza J, Okamoto N, Nkusi R, Marescaux J, Kawakubo H, Kitagawa Y, Gonzalez C, Mutter D, Perretta S, Hostettler A, Dallemagne B (2023) Towards automatic verification of the critical view of the myoeptineal orifice with artificial intelligence. *Surg Endosc* 37:4525–4534
  23. Yamazaki Y, Kanaji S, Matsuda T, Oshikiri T, Nakamura T, Suzuki S, Hiasa Y, Otake Y, Sato Y, Kakeji Y (2020) Automated surgical instrument detection from laparoscopic gastrectomy video images using an open source convolutional neural network platform. *J Am Coll Surg* 230:725–732.e721
  24. Kitaguchi D, Lee Y, Hayashi K, Nakajima K, Kojima S, Hasegawa H, Takeshita N, Mori K, Ito M (2022) Development and validation of a model for laparoscopic colorectal surgical instrument recognition using convolutional neural network-based instance segmentation and videos of laparoscopic procedures. *JAMA Netw Open* 5:e2226265
  25. Kitaguchi D, Takeshita N, Matsuzaki H, Takano H, Owada Y, Enomoto T, Oda T, Miura H, Yamanashi T, Watanabe M, Sato D, Sugomori Y, Hara S, Ito M (2020) Real-time automatic surgical phase recognition in laparoscopic sigmoidectomy using the convolutional neural network-based deep learning approach. *Surg Endosc* 34:4924–4931
  26. Shinozuka K, Turuda S, Fujinaga A, Nakanuma H, Kawamura M, Matsunobu Y, Tanaka Y, Kamiyama T, Ebe K, Endo Y, Etoh T, Inomata M, Tokuyasu T (2022) Artificial intelligence software available for medical devices: surgical phase recognition in laparoscopic cholecystectomy. *Surg Endosc* 36:7444–7452
  27. Funke I, Mees ST, Weitz J, Speidel S (2019) Video-based surgical skill assessment using 3D convolutional neural networks. *Int J Comput Assist Radiol Surg* 14:1217–1225
  28. Kitaguchi D, Takeshita N, Matsuzaki H, Igaki T, Hasegawa H, Ito M (2021) Development and validation of a 3-dimensional convolutional neural network for automatic surgical skill assessment based on spatiotemporal video analysis. *JAMA Netw Open* 4:e2120786
  29. Igaki T, Kitaguchi D, Matsuzaki H, Nakajima K, Kojima S, Hasegawa H, Takeshita N, Kinugasa Y, Ito M (2023) Automatic surgical skill assessment system based on concordance of standardized surgical field development using artificial intelligence. *JAMA Surg* 158:e231131
  30. Kiyasseh D, Ma R, Haque TF, Miles BJ, Wagner C, Donoho DA, Anandkumar A, Hung AJ (2023) A vision transformer for decoding surgeon activity from surgical videos. *Nat Biomed Eng* 7:780–796
  31. Kyungmin J, Bareum C, Songe C, Youngjin M, Jaesoon C (2016) Automatic detection of hemorrhage and surgical instrument in laparoscopic surgery image. *Annu Int Conf IEEE Eng Med Biol Soc* 2016:1260–1263
  32. Garcia-Martinez A, Vicente-Samper JM, Sabater-Navarro JM (2017) Automatic detection of surgical haemorrhage using computer vision. *Artif Intell Med* 78:55–60
  33. Okamoto T, Ohnishi T, Kawahira H, Dergachyava O, Jannin P, Haneishi H (2019) Real-time identification of blood regions for hemostasis support in laparoscopic surgery. *SIVIP* 13:405–412
  34. Mongan J, Moy L, Kahn CE Jr (2020) Checklist for Artificial Intelligence in Medical Imaging (CLAIM): a guide for authors and reviewers. *Radiol Artif Intell* 2:e200029
  35. World Medical Association Declaration of Helsinki (2013) ethical principles for medical research involving human subjects. *JAMA* 310:2191–2194
  36. Wang C-Y, Bochkovskiy A, Liao H-YM (2023) YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp 7464–7475
  37. MMYOLO Contributors (2022) MMYOLO: OpenMMLab YOLO series toolbox and benchmark. <https://github.com/open-mmlab/mmyolo>
  38. Microsoft Common Objects in Context; detection evaluation metrics. <https://cocodataset.org/#detection-eval>
  39. Namikawa K, Hirasawa T, Nakano K, Ikenoyama Y, Ishioka M, Shiroma S, Tokai Y, Yoshimizu S, Horiuchi Y, Ishiyama A, Yoshio T, Tsuchida T, Fujisaki J, Tada T (2020) Artificial intelligence-based diagnostic system classifying gastric cancers and ulcers: comparison between the original and newly developed systems. *Endoscopy* 52:1077–1083
  40. Gong D, Wu L, Zhang J, Mu G, Shen L, Liu J, Wang Z, Zhou W, An P, Huang X, Jiang X, Li Y, Wan X, Hu S, Chen Y, Hu X, Xu Y, Zhu X, Li S, Yao L, He X, Chen D, Huang L, Wei X, Wang X, Yu H (2020) Detection of colorectal adenomas with a real-time computer-aided system (ENDOANGEL): a randomised controlled study. *Lancet Gastroenterol Hepatol* 5:352–361
  41. Wang P, Liu X, Berzin TM, Glissen Brown JR, Liu P, Zhou C, Lei L, Li L, Guo Z, Lei S, Xiong F, Wang H, Song Y, Pan Y, Zhou G (2020) Effect of a deep-learning computer-aided detection system on adenoma detection during colonoscopy (CADE-DB trial): a double-blind randomised study. *Lancet Gastroenterol Hepatol* 5:343–351
  42. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017) Attention is all you need. *Adv Neural Inf Process Syst* 30
  43. Zhu X, Su W, Lu L, Li B, Wang X, Dai J (2020) Deformable DETR: deformable transformers for end-to-end object detection. *arXiv preprint arXiv:201004159*

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.