AI Enhanced Surgeons' Faster Responses To Intraoperative Bleeding In Laparoscopic Surgery: A Preclinical Randomized Controlled Trial

Authors: Jing Zhang, He Cai, Junjie Zhang, Zixin Chen, Yang Zhou, Ran Hu, Ming Tang, Lingli Wu, Songlin Cai, Jianzheng Jie, Rui Shi, Wenjing Gong, Bing Peng, Xin Wang, Zixin Chen

Date: 2026-05-09T01:15:26+00:00

Abstract

Objective: To develop and validate an artificial intelligence (AI) model for real-time detection and tracking of intraoperative bleeding sites in laparoscopic surgery, and evaluate whether AI assistance could enhance surgeons' responses to intraoperative bleeding.Background: Intraoperative bleeding during laparoscopic surgery is a critical complication that requires timely detection and handling to ensure surgical safety.Methods: From January to September 2023, laparoscopic surgical videos were retrospectively collected from multiple Chinese medical centers. Senior surgeons marked bleeding events, extracted related clips, and manually labeled bleeding frames to train the AI model, evaluated by F1-score, mean absolute error (MAE), and mean distance error (MDE). A preclinical randomized controlled trial was conducted. Eligible surgeons were randomly assigned to either the experimental group or the control group. The primary outcome was reaction time, defined as the time elapsed from the appearance of a bleeding site to the participant' s first successful click on that site.Results: We collected 257 laparoscopic surgical videos from 20 medical centers. The training, validation, and test sets included 145 (56.42%), 55 (21.40%), and 57 (22.18%) videos, respectively. The model achieved an F1-score of 0.860 (95% CI, 0.838–0.878). In addition, the MAE was 3.315 (95% CI, 3.051–3.614), and the MDE was 0.087 (95% CI, 0.086–0.088). The experimental group had shorter reaction time (2.31s vs. 3.13s, $P < 0.001$) and lower negative emotion score (1.68 vs. 2.95, $P < 0.001$).Conclusion: The AI model has high accuracy; AI assistance shortens surgeons' reaction time and alleviates negative emotional burden.

Full Text

Preamble

AI Enhanced Surgeons' Faster Responses To Intraoperative Bleeding In Laparoscopic

Surgery: A Preclinical Randomized Controlled Trial

Jing Zhang※, MD1,2; He Cai※, MD1,2; Junjie Zhang, MM1,2; Zixin Chen, MD1,2; Yang Zhou, BS3; Ran Hu, MM1,2; Ming Tang, MM1,2; Lingli Wu, MD1,2; Songlin Cai, MM1,2; Jianzheng Jie , MD4; Rui Shi, MD5; WenJing Gong ,MD6; Bing Peng*, MD, PhD, FACS1,2; Xin Wang*, MD1,2.

1. Division of Pancreatic Surgery, Department of General Surgery, West China Hospital of Sichuan University, Chengdu, China.
2. West China School of Medicine, West China Hospital of Sichuan University, Chengdu, China.
3. ChengDu Withai Innovations Technology Company, Chengdu, China.
4. Department of Colorectal Surgery, China-Japan Friendship Hospital, Beijing, China.
5. Information Center, West China Hospital, Sichuan University, Chengdu, China.
6. Department of Colorectal Surgery, Zhejiang Provincial People' s Hospital, Hangzhou, China.

※Jing Zhang and He Cai are co-first authors.

*Corresponding author: Bing Peng and Xin Wang are co-corresponding authors.

Abstract

Objective: To develop and validate an artificial intelligence (AI) model for real-time detection

and tracking of intraoperative bleeding sites in laparoscopic surgery, and evaluate whether AI

assistance could enhance surgeons' responses to intraoperative bleeding.

Background: Intraoperative bleeding during laparoscopic surgery is a critical complication

that requires timely detection and handling to ensure surgical safety.

Methods: From January to September 2023, laparoscopic surgical videos were retrospectively

collected from multiple Chinese medical centers. Senior surgeons marked bleeding events,

extracted related clips, and manually labeled bleeding frames to train the AI model, evaluated

by F1-score, mean absolute error (MAE), and mean distance error (MDE). A preclinical

randomized controlled trial was conducted. Eligible surgeons were randomly assigned to either

the experimental group or the control group. The primary outcome was reaction time, defined

as the time elapsed from the appearance of a bleeding site to the participant' s first successful

click on that site.

Results: We collected 257 laparoscopic surgical videos from 20 medical centers. The training,

validation, and test sets included 145 (56.42%), 55 (21.40%), and 57 (22.18%) videos,

respectively. The model achieved an F1-score of 0.860 (95% CI, 0.838–0.878). In addition, the

MAE was 3.315 (95% CI, 3.051–3.614), and the MDE was 0.087 (95% CI, 0.086–0.088). The

experimental group had shorter reaction time (2.31s vs. 3.13s, P <

0.

1) and lower negative

emotion score (1.68 vs. 2.95, P < 0.001).

Conclusion: The AI model has high accuracy; AI assistance shortens surgeons' reaction time

and alleviates negative emotional burden.

Key words: AI; Laparoscopic surgery; Intraoperative bleeding; Detection and tracking; Surgeon

response.

Trial registration: Chinese Clinical Trial Registry Identifier: ChiCTR2600118185.

Introduction

Over the past several decades, laparoscopic surgery has become a cornerstone of modern surgical

practice due to its minimal invasiveness and improvements in postoperative recovery1, 2.

Nevertheless, intraoperative bleeding remains a major challenge that can compromise the safety and

efficacy of fully laparoscopic procedures. A substantial body of evidence indicates that greater

intraoperative blood loss is associated with higher rates of postoperative complications, mortality,

conversion to open surgery, and blood transfusion3-5. Collectively, these findings underscore the

critical importance of timely and effective control of intraoperative bleeding during laparoscopic

surgery.

Intraoperative bleeding events are often sudden and difficult to predict, posing two major challenges

in hemostasis. First, once bleeding occurs, delayed control can rapidly lead to blood accumulation

that obscures the bleeding source, thereby complicating visualization and subsequent hemostasis 6-

8. Moreover, the abrupt onset of bleeding imposes acute psychological stress, which may disrupt

attentional focus and transiently impair operative performance 9-11. Accordingly, the ability to

rapidly direct attention to the bleeding site and achieve immediate initial control —most commonly

through compression—is critical, as it establishes favorable conditions for definitive hemostatic

management7. Currently, surgery remains a largely experience-based discipline rooted in an

apprenticeship model, and substantial inter-individual variability among surgeons contributes to

wide differences in the time required to achieve effective bleeding control9, 11,

12. Therefore, there is

substantial clinical and practical value in developing standardized technologies or adjunctive tools

to mitigate inter-surgeon variability in hemostatic efficiency and to enhance the precision and

consistency of bleeding control.

With recent advances in artificial intelligence (AI), AI-based systems have begun to play an

increasingly important role in time-critical decision-making during emergency medical events. For

example, AI systems can rapidly identify occult hemorrhagic foci and fracture sites on CT images,

providing emergency physicians with real-time diagnostic support and shortening the time to

treatment decisions13. By extension, similar AI-assisted decision-making paradigms are becoming

increasingly feasible for intraoperative adverse events. In the surgical setting, AI has the potential

to identify key intraoperative events and provide real-time visual or contextual guidance, enabling

more consistent and standardized management of emergent situations such as intraoperative

bleeding14-16.

Therefore, this study aimed to determine whether AI assistance could enable surgeons to respond

more rapidly to intraoperative bleeding during laparoscopic surgery. First, a high-precision, real-

time AI model was developed for detecting and tracking intraoperative bleeding sites. Second, a

preclinical randomized controlled trial (RCT) was conducted to evaluate whether AI-assisted visual

guidance could shorten surgeons' hemostatic response times while alleviating psychological stress

during bleeding events.

Methods

Study design

This study adopted a phased approach comprising model development, tool deployment, and

preclinical validation (Fig. 1 [Figure 1: see original paper]). First, an AI model was trained to detect and track intraoperative

bleeding sites. Second, the model was integrated into an interactive AI-assisted tool. The tool was

designed to approximate real surgical scenarios and incorporated task-based, gamified interactions.

The tool is accessible at: https://app.surgsmart.com/login?from=/game. Finally, a preclinical RCT

based on this tool was conducted to evaluate whether AI assistance could improve surgeons'

performance in responding to intraoperative bleeding. The study was approved by the Ethics

Committee of West China Hospital, Sichuan University (No. 2025-2346). This study was registered

with the Chinese Clinical Trial Registry under registration number ChiCTR2600118185, with a

registration date of February 3,

2026. The study was reported in accordance with the CONSORT-

AI guidelines17, a complete CONSORT checklist is available in the supplemental materials

(Appendix 1).

Participants and AI Model Development

Participants were eligible if they were surgeons aged 20–50 years whose daily clinical practice

primarily involved laparoscopic surgery and who were proficient in operating electronic devices

required for interactive computer-based tasks. The core development process of the AI algorithm

for detecting and tracking bleeding sites in laparoscopic surgery included data annotation and the

construction of a multi-task deep learning framework (Appendix

2. Supplementary Methods). Data

annotation was performed by senior surgeons (XW and HC), who determined the onset and end

times of bleeding events, established video subsets of bleeding or non-bleeding clips, and conducted

frame-by-frame point annotation of bleeding sites using Label Studio. Based on the annotated data,

a multi-task deep learning framework with a hierarchical spatiotemporal architecture was

constructed to jointly perform bleeding presence classification, frame-level bleeding detection, and

spatial localization. Detection performance was enhanced by integrating multi-scale spatiotemporal

features, and an L1 loss function was applied to optimize bleeding sites localization. An AI-assisted

visual tool integrating the algorithm was subsequently developed, allowing activation or

deactivation of AI-assisted functions. The tool recorded performance data in real time and provided

immediate feedback, enabling quantitative comparison of surgeons' performance between AI-

assisted and non-assisted conditions (Fig. 1b). For detailed methodologies, refer to the Appendix 2.

Randomization

Participants were stratified into junior and senior surgeon cohorts according to their laparoscopic

surgical experience, with the cutoff set at 3 years ($\leq$ 3 years vs. $>$ 3 years). Within each stratum,

block randomization with a block size of 4 was implemented, and participants were assigned in a

1:1 ratio to either the experimental or control group via a computer-generated random sequence (Fig.

1c). A detailed description of the sample size calculation is provided in the Supplementary Methods.

Interventions

Participants in both groups were required to identify and click on bleeding sites across 20

laparoscopic surgical video clips. For the experimental group, participants completed the task with

the aid of AI-generated bounding box prompts to provide visual guidance. In contrast, participants

in the control group performed the task via a standard video interface devoid of AI-derived visual

cues. All other experimental conditions were held constant between the two groups, including the

sequence of video clips, hardware configuration, and network environment.

Outcomes

The primary outcome was reaction time, defined as the interval from the appearance of a bleeding

site in the video to the participant' s first successful click on that site. Secondary outcomes included

the following measures: (1) total successful hemostasis events, defined as the cumulative number

of bleeding sites successfully identified and clicked during the task; (2) number of click attempts,

defined as the total number of clicks made within a single video clip before successful identification

of a bleeding site; and (3) distance error, defined as a normalized relative deviation value obtained

by dividing the pixel linear distance between the click site and the bleeding site by 6% of the video' s

long side. All outcome measures were automatically recorded and calculated by the interactive tool

during task performance. In addition, participants'negative emotional responses were assessed

using a 5-point Likert scale administered before, during, and after the validation process, with results

summarized as a Negative Emotion Score (NES).

Statistical Analysis

For evaluation of the bleeding site detection and tracking model, sensitivity (recall), specificity,

precision, F1-score, area under the receiver operating characteristic curve (ROC), mean absolute

error (MAE), and mean distance error (MDE) were used. The calculation of these metrics is

described in detail in the Supplementary Methods. All statistical analyses were performed using R

software (version 4.5.1; R Foundation for Statistical Computing, Vienna, Austria). Descriptive

statistics were reported as mean with standard deviation, median with interquartile range, or

frequency with percentage as appropriate. Continuous variables were compared using independent

t-tests or Mann–Whitney U tests, as appropriate, while categorical variables were analyzed using

chi-square tests. Prespecified subgroup analyses were performed to explore consistency of effects

across surgeon experience levels. A two-sided $P < 0.05$ was considered statistically significant, with

Bonferroni correction applied for multiple comparisons.

Results

Performance Results of Model Development

of bleeding sites. The dataset comprised 162,421 bleeding frames extracted from 5,662 annotated

bleeding clips across 257 laparoscopic surgical videos. Among these bleeding frames, 90,730

(55.86%), 41,483 (25.54%), and 30,208 (18.60%) were assigned to the training, validation, and test

sets, respectively. For bleeding clip classification, the model achieved a sensitivity (recall) of 0.882

(95% CI, 0.862–0.900), specificity of 0.964 (95% CI, 0.959–0.969), precision of 0.838 (95% CI,

0.815–0.858), and an F1-score of 0.860 (95% CI, 0.838–0.878). The area under the ROC was 0.985

(95% CI, 0.980–0.

990) (Appendix 3). For temporal detection and spatial localization of bleeding

sites, the MAE was 3.315 (95% CI, 3.051–3.614), and the MDE was 0.087 (95% CI, 0.086–0.088).

Participants Baseline Characteristics

Baseline characteristics of participants in the preclinical validation are presented in Appendix

4. A

total of 77 participants were enrolled, including 38 (49.35%) in the experimental group and 39

(50.65%) in the control group. The cohort included 58 (75.32%) men and 19 (24.68%) women, with

a mean age of 30.09 ± 5.37 years; 38 participants (49.35%) were classified as senior surgeons. In

addition, 36 participants (46.75%) reported little to no prior gaming experience, and 34 (44.16%)

reported no prior experience with serious adverse consequences resulting from delayed hemostasis

during laparoscopic surgery. Notably, none of the participants experienced physical discomfort

severe enough to require task interruption or adjustment.

Participants' Performance in Preclinical Validation

group, the experimental group demonstrated a shorter mean reaction time (2.31 s vs. 3.13 s), fewer

mean click attempts (2.36 vs. 3.36), a greater number of successful hemostasis events (18 vs. 15),

and a higher proportion of hemostasis achieved within 2 seconds (52% vs. 35%). All these

differences were statistically significant ($P < 0.001$). However, there was no statistically significant

difference in mean distance error between the two groups (0.64 vs. 0.65, $P = 0.967$). Subgroup

analyses demonstrated that, compared with the control group, the experimental group exhibited

significantly shorter mean reaction times in most subgroups ($P < 0.05$). These results are presented

in Fig. 2 [Figure 2: see original paper].

Negative Emotional Load

After completion of the task, the experimental group exhibited a significantly lower mean NES than

the control group (1.68 vs. 2.95, P < 0.001). In addition, the mean change in NES was significantly

lower in the experimental group than in the control group (−0.03 vs. 1.13, P = 0.003). Detailed

results are presented in Appendix

5. Temporal trends in mean NES among all participants are

illustrated in Fig. 3 [Figure 3: see original paper]. Overall, participants' NES increased initially and then decreased over the course

of the testing process, with NES levels in the experimental group remaining consistently lower than

those in the control group throughout.

Discussion

As noted above, intraoperative bleeding remains one of the most critical adverse events during

laparoscopic surgery, because even a short delay in the initial response can rapidly lead to blood

accumulation, impaired visualization, and increased procedural complexity. The first response to

bleeding—typically involving visual identification and immediate compression—plays a pivotal

role in preventing further deterioration of the operative field and creating favorable conditions for

definitive hemostasis. However, our previous studies have shown that the time required to initiate

effective bleeding control varies substantially among surgeons, reflecting differences in experience,

situational awareness, and stress response. Such inter-individual variability highlights the current

lack of tools to support surgeons at the very onset of bleeding events. In this study, we developed

an AI-based model for frame-level detection and tracking of bleeding sites in laparoscopic surgery.

We further integrated this model into an interactive, task-based visual assistance tool designed to

direct surgeons' attention during the critical moments of bleeding onset. To rigorously evaluate its

potential benefits, we conducted a RCT in a preclinical setting to assess the performance of the tool.

In summary, we developed an AI model for frame-level detection and tracking of bleeding sites, and

our findings suggest that the AI-based visual assistance tool can support surgeons during urgent and

stressful phases of intraoperative bleeding by improving responsiveness, task efficiency, and

psychological stability. Together, these preclinical results indicate that the proposed AI model has

the potential to assist surgeons in timely and accurate identification of intraoperative bleeding events,

while reducing the impact of adverse intraoperative conditions.

As intraoperative bleeding is one of the most critical adverse events during minimally invasive

surgery, extensive efforts have been devoted to its automated detection and analysis over the past

decade. Early studies primarily relied on pixel-level color information, combining RGB- or HSV-

based features with static or dynamic thresholding strategies to identify blood pixels or estimate

bleeding burden18-21. These approaches demonstrated the feasibility of extracting bleeding-related

visual signals from surgical videos; however, they are inherently sensitive to background tissue color,

illumination variation, specular reflection, smoke, and camera motion, thereby limiting their

robustness in complex real-world surgical scenes. With the rise of machine learning and deep

learning, research shifted toward data-driven models capable of learning higher-level visual

representations. Convolutional neural network–based methods enabled automatic detection of

bleeding frames or regions22, showing better generalization than traditional feature-based methods.

Building on these advances, subsequent video-based models incorporated temporal or

spatiotemporal mechanisms to improve prediction stability and continuity by explicitly modeling

inter-frame relationships23. In addition, recent SAM-based frameworks24 have demonstrated strong

performance in jointly detecting bleeding regions and source points within short video clips centered

on confirmed bleeding events. By coupling region segmentation and point localization through

bidirectional guidance, these methods achieve improved spatial consistency and robustness under

challenging visual conditions. However, their evaluation is primarily restricted to frame-level

localization within preselected bleeding segments, without assessing bleeding event presence or

onset in continuous surgical videos. In contrast, our approach formulates intraoperative bleeding

analysis as a hierarchical spatiotemporal perception problem aligned with surgeons' immediate

perceptual requirements during bleeding events. By explicitly defining bleeding point localization

as the primary perceptual target and jointly modeling bleeding presence, onset timing, and

temporally consistent point tracking across entire video sequences, our framework advances

bleeding analysis from descriptive localization toward task-oriented visual guidance for rapid

bleeding source identification in simulated settings.

Beyond algorithmic performance, an important contribution of the present study lies in

demonstrating how AI-based visual assistance can translate into measurable improvements in

surgeon perceptual and interactive performance in a preclinical task-based platform. In this

randomized controlled simulation study, AI assistance was consistently associated with faster

reaction times, fewer interaction attempts, and higher overall task efficiency across experimental

conditions, indicating that the system reduced the time and effort required to visually identify and

localize bleeding sites in the simulated interface. Notably, our previous work on bleeding control

efficiency revealed substantial inter-individual variability among surgeons. In an analysis of 634

surgical cases, the time required to achieve initial bleeding control ranged from less than one second

to more than one minute after bleeding onset, underscoring the vulnerability of early hemostatic

management to differences in experience, situational awareness, and stress response. In the present

study, subgroup analyses demonstrated significant performance improvements with AI assistance in

both junior and senior surgeons, suggesting that the system enhances visual localization

responsiveness across experience levels and may reduce inter-individual variability in this specific

perceptual task. Given the preclinical, scenario-based design using a mobile interactive platform,

potential confounding effects of prior gaming experience were also considered. Stratified analyses

demonstrated that AI assistance benefited participants regardless of gaming experience, indicating

that the observed performance improvements were attributable to clinically relevant visual guidance

rather than familiarity with interactive interfaces. Taken together, these findings suggest that AI-

based visual assistance can provide robust, experience-independent support for rapid bleeding point

localization in the early perceptual phase of simulated bleeding events, with potential implications

for promoting more timely and consistent visual identification of bleeding sources in future clinical

translation.

Sudden intraoperative bleeding is not only a technical challenge but also a potent psychological

stressor that can adversely affect surgical performance. Human factors research in surgery and other

high-risk domains has consistently shown that negative activating emotions, such as anxiety and

tension, are closely associated with performance degradation during time-critical tasks25,

26. Acute

stress during bleeding events increases cognitive workload, disrupts attentional control, and impairs

decision-making and motor execution, thereby elevating the risk of delayed or suboptimal initial

hemostatic responses27. In recent years, growing evidence from other clinical domains has suggested

that AI-based assistance can help mitigate psychological burden and cognitive fatigue during

complex diagnostic and decision-making processes28-31. These findings support the notion that a

well-designed AI system can relieve surgeons of perceptual and attentional burdens, acting as an

assistant in clinical practice. However, despite the growing body of relevant evidence, the potential

of AI assistance in alleviating the psychological stress triggered by simulated acute bleeding

scenarios—an emergency characterized by extreme time constraints and heavy emotional burden—

has not yet been systematically investigated in preclinical task-based settings. In this study, a

combination of participants' subjective questionnaire responses and objective measurements of

negative emotions demonstrated that the AI-assisted visual guidance approach effectively reduced

surgeons' psychological stress during simulated bleeding localization tasks. From a subjective

perspective, participants generally reported that the AI-generated prompts for bleeding point

identification were highly practical, and that AI assistance significantly alleviated their operational

pressure in the task context. Objectively, emotional trajectory analysis revealed that surgeons using

AI support exhibited lower levels of negative emotions throughout the task and achieved a more

rapid emotional recovery post-task. Notably, this effect was observed consistently among both

junior and senior surgeons.

Limitations

Important limitations of this preclinical investigation must be acknowledged. The present study

evaluated surgeon performance in a simulated, task-based clicking paradigm rather than in actual

intraoperative hemostasis, which requires instrument manipulation, tissue handling, bimanual

coordination, team communication, and definitive hemostatic procedures. Improved performance

in visual localization and interactive response in this platform should not be equated with improved

clinical bleeding control, surgical performance, or patient safety outcomes. The primary endpoint

of reaction time to successful clicking represents a surrogate marker for early visual perception and

localization rather than real-world hemostatic efficiency. Further validation in high-fidelity surgical

simulators and subsequent prospective clinical studies will be necessary to confirm the translational

value and clinical impact of this AI-assisted system.

Conclusions

This study demonstrates that an AI-based visual assistance system for detection and tracking of

bleeding sites can effectively support early visual localization and perceptual responses to simulated

intraoperative bleeding. By providing accurate visual cues, the system was associated with

improved reaction speed, task efficiency, and response consistency in a preclinical, task-based

clicking platform across surgeons with different experience levels, while also reducing

psychological burden during simulated bleeding tasks. These findings provide important empirical

evidence supporting the potential translational value of AI-based visual assistance for rapid

identification of bleeding sources in precision surgery, with the caveat that clinical benefits for

intraoperative hemostasis and patient safety remain to be validated in real-world surgical settings.

Reference

Ricciardi R, Seshadri-Kreaden U, Yankovsky A et al. The COMPARE Study: Comparing Perioperative

Outcomes of Oncologic Minimally Invasive Laparoscopic, da Vinci Robotic, and Open Procedures: A

Systematic Review and Meta-analysis of the Evidence. Ann Surg 2025; 281 (5): 748-763.

development. J Robot Surg 2024; 19 (1): 12.

Gigot JF, Glineur D, Santiago Azagra J et al. Laparoscopic liver resection for malignant liver tumors: preliminary results of a multicenter European study. Ann Surg 2002; 236 (1): 90-97.

Shiba H, Ishida Y, Wakiyama S et al. Negative impact of blood transfusion on recurrence and prognosis of hepatocellular carcinoma after hepatic resection. J Gastrointest Surg 2009; 13 (9): 1636-1642.

Katz SC, Shia J, Liau KH et al. Operative blood loss independently predicts recurrence and survival after resection of hepatocellular carcinoma. Ann Surg 2009; 249 (4): 617-623.

Sunakawa T, Kitaguchi D, Kobayashi S et al. Deep learning-based automatic bleeding recognition during liver resection in laparoscopic hepatectomy. Surg Endosc 2024; 38 (12): 7656-7662.

Szabó DI, Vereczkei A, Papp

A. From gaming to surgery: the influence of digital natives on robotic skills

Abu Hilal M, Underwood T, Taylor MG et al. Bleeding and hemostasis in laparoscopic liver surgery. Surg Endosc 2010; 24 (3): 572-577.

Shin H, Cho JY, Han HS et al. Risk factors and long-term implications of unplanned conversion during

laparoscopic liver resection for hepatocellular carcinoma located in anterolateral liver segments. J Minim

Invasive Surg 2021; 24 (4): 191-199.

performance during a simulated tumor resection task. J Neurosurg 2017; 126 (1): 71-80.

Awtry J, Skinner S, Polazzi S et al. Association Between Surgeon Stress and Major Surgical Complications.

JAMA Surg 2025; 160 (3): 332-340.

Solheim

O. Is aneurysm surgery too exciting for our own good? Acta Neurochir (Wien) 2016; 158 (2):

Bajunaid K, Mullah MA, Winkler-Schwartz A et al. Impact of acute stress on psychomotor bimanual

Maruthappu M, Gilbert BJ, El-Harasis MA et al. The influence of volume and experience on individual surgical performance: a systematic review. Ann Surg 2015; 261 (4): 642-647.

Kiefer J, Kopp M, Ruettinger T et al. Diagnostic Accuracy and Performance Analysis of a Scanner-

Integrated Artificial Intelligence Model for the Detection of Intracranial Hemorrhages in a Traumatology

Emergency Department. Bioengineering (Basel) 2023; 10 (12).

surgery. Nature 2023; 622 (7984): 842-849.

Byerly S, Maurer LR, Mantero A et al. Machine Learning and Artificial Intelligence for Surgical Decision Making. Surg Infect (Larchmt) 2021; 22 (6): 626-634.

Li H, Han Z, Wu H et al. Artificial intelligence in surgery: evolution, trends, and future directions. Int J Surg 2025; 111 (2): 2101-2111.

Vermeulen C, Pagès-Gallego M, Kester L et al. Ultra-fast deep-learned CNS tumour classification during

Liu X, Rivera SC, Moher D et al. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI Extension. Bmj 2020; 370: m3164.

Xu H, Han T, Wang H et al. Detection of blood stains using computer vision-based algorithms and their

association with postoperative outcomes in thoracoscopic lobectomies. Eur J Cardiothorac Surg 2022; 62

bleeding during minimally invasive surgery. Int J Med Robot 2020; 16 (6): 1-9.

Ribbens VJ, Baltus SC, Tan CO et al. Machine learning-based detection of electrosurgical device-induced bleeding in laparoscopic videos of diaphragmatic hernia repair. Surg Endosc 2025; 39 (9): 6230-6238.

Garcia-Martinez A, Vicente-Samper JM, Sabater-Navarro JM. Automatic detection of surgical haemorrhage using computer vision. Artif Intell Med 2017; 78: 55-60.

Rahbar MD, Reisner L, Ying H et al. An entropy-based approach to detect and localize intraoperative

Hua S, Gao J, Wang Z et al. Automatic bleeding detection in laparoscopic surgery based on a faster regionbased convolutional neural network. Ann Transl Med 2022; 10 (10): 546.

Rabbani N, Seve C, Bourdel N et al. Video-based Computer-aided Laparoscopic Bleeding Management:

a Space-time Memory Neural Network with Positional Encoding and Adversarial Domain Adaptation. 5th

International Conference on Medical Imaging with Deep Learning (MIDL).

172. Zurich,

SWITZERLAND; 2022:961-973.

Videos. 2025.

Vedder K, Reissfelder

C. The Stress Level of Surgeons influences Surgical Complications. Chirurgie 2025; 96 (8): 687-689.

Tjonnas MS, Muller S, Vapenstad C et al. Stress responses in surgical trainees during simulation-based training courses in laparoscopy. Bmc Medical Education 2024; 24 (1).

Pei J, Zhou Z, Guo D et al. Synergistic Bleeding Region and Point Detection in Laparoscopic Surgical

Tam A, Bateman S, Buckingham G et al. The effects of stress on surgical performance: a systematic review.

Surgical Endoscopy and Other Interventional Techniques 2025; 39 (1): 77-98.

Fazlollahi AM, Bakhaidar M, Alsayegh A et al. Effect of Artificial Intelligence Tutoring vs Expert Instruction on Learning Simulated Surgical Skills Among Medical Students: A Randomized Clinical Trial.

JAMA Netw Open 2022; 5 (2): e2149008.

Simulation Performance: A Randomized Clinical Trial. JAMA Surg 2025; 160 (9): 993-1003.

Giglio B, Albeloushi A, Alhaj AK et al. Artificial Intelligence-Augmented Human Instruction and Surgical

Pekrun R, Marsh HW, Elliot AJ et al. A three-dimensional taxonomy of achievement emotions. J Pers Soc Psychol 2023; 124 (1): 145-178.

Young JQ, Van Merrienboer J, Durning S et al. Cognitive Load Theory: implications for medical education:

AMEE Guide No.

86. Med Teach 2014; 36 (5): 371-384.

are used to train the model via AI algorithms;

b. tool deployment: The two images show the interface during validation, where the blue circles represent the click cursors and the green boxes denote the AI-assisted prompt boxes;

c. preclinical validation: flow chart of study participants enrollment for preclinical randomized controlled trials.

history of bad events: dark blue square blocks represent the estimated values of the standardized mean difference (SMD); the length of the line segment indicates the width of the confidence interval; reference line (zero line): an SMD of 0 indicates no significant difference between the two groups.

E: experimental group; C: control group; SD: standard deviation.

b. subgroup analysis of senior surgeons;

c. subgroup analysis of junior surgeons. Groups were color - coded (see key). The y - axis represents the scores of the emotional scale. **: $P < 0.01$;** *: $P <$

0.

1. All statistical results were based on the baseline (before) emotional scores of different groups.

Error bars indicate the standard error.

Dataset characteristics for model development. Training

Validation

data sets

data sets

Cases of Surgery (n, %)

145 (56.42)

55 (21.40)

Number of Clips (n, %)

13280 (56.55)

4949 (21.07)

Cases of Surgery (n, %)

127 (54.98)

50 (21.65)

Number of Clips (n, %)

3062 (54.08)

1494 (26.39)

Number of frames (n, %)

90730 (55.86)

41483 (25.54)

AUC (95% CI)

Sensitivity/Recall (95% CI)

Specificity (95% CI)

Precision (95% CI)

F1 score (95% CI)

Non-Bleeding

Bleeding

Performance of bleeding clips detection by AI model.

Performance of the AI model in detecting frame-level bleeding clips.

MAE * (95%

CI) Performance of the AI model in detecting the location of bleeding sites.

MDE† (95% CI)

*MAE (Mean Absolute Error) is defined as follows: For a 4-second video clip with 25 FPS, 32 frames are extracted at intervals of 2 frames

predicts a bleeding frame (e.g., Frame 51, corresponding to the 17th position among the 32 extracted frames) and the annotated bleeding fra position among the 32 extracted frames), the absolute error is calculated as

1. MAE is the average of all absolute errors across the test set.

†MDE (Mean Distance Error) refers to the Euclidean distance between two points. With the top-left corner of the image as the origin of co

coordinates of the bleeding point are (0.6,

0.

4) and the annotated coordinates are (0.5, 0.5); the distance error is the Euclidean distance betw

Experimental Group

Control group

(AI-assisted)

(No AI-assisted)

(n=38)

(n=39)

Average reaction time (s)

2.31 (1.97-3.01)

3.13 (2.48-3.78)

Average distance error

0.64 (0.61-0.67)

0.65 (0.60-0.68)

Average click attempts (n)

2.36 (2.01-3.00)

3.36 (2.86-4.69)

18 (16-19)

15 (14-16)

52% (40%-65%)

35% (25%-45%)

Total hemostasis success times (n) Success rate of hemostasis within 2 second

Data in this table are presented as median (interquartile range), with values of $P < 0.05$ shown in bold.

Appendix 1 CONSORT 2025 checklist. Appendix 2 Supplementary Methods.

Appendix 3 Appendix Figure

1. The ROC curve of AI-based bleeding sites detection model.

Appendix 4 Baseline characteristics for the participants in preclinical validation.

Appendix 5 NES of participants in the experimental and control groups at different stages during the preclinical validation.

Appendix 1 CONSORT 2025 checklist of information to include when reporting a randomised trial* Section / Topic

CONSORT 2025 checklist item description

Reported on page no.

Identification as a randomised trial

Structured summary of the trial design, methods, results, and conclusions

Trial registration

Name of trial registry, identifying number (with URL) and date of registration

Protocol and statistical analysis plan

Where the trial protocol and statistical analysis plan can be accessed

Data sharing

Where and how the individual de-identified participant data (including data dictionary), statistical code and any other materials can be accessed

Funding and conflicts of interest

Sources of funding and other support (e.g., supply of drugs), and role of funders in the design, conduct, analysis and reporting of the trial

Financial and other conflicts of interest of the manuscript authors

Background and rationale

Scientific background and rationale

Objectives

Specific objectives related to benefits and harms

Details of patient or public involvement in the design, conduct and reporting of the trial

Trial design

Description of trial design including type of trial (e.g., parallel group, crossover), allocation ratio, and framework (e.g., superiority, equivalence, non-inferiority, exploratory)

Changes to trial protocol

Important changes to the trial after it commenced including any outcomes or analyses that were not

Title and abstract Title

abstract

structured

Open science

Methods

Patient involvement

public

prespecified, with reason Trial setting

Settings (e.g., community, hospital) and locations (e.g., countries, sites) where the trial was conducted

Eligibility criteria for participants

If applicable, eligibility criteria for sites and for individuals delivering the interventions (e.g., surgeons, physiotherapists)

Intervention and comparator with sufficient details to allow replication. If relevant, where additional materials describing the intervention and comparator (e.g., intervention manual) can be accessed

Outcomes

Pre-specified primary and secondary outcomes, including the specific measurement variable (e.g., systolic blood pressure), analysis metric (e.g., change from baseline, final value, time to event), method of aggregation (e.g., median, proportion), and time point for each outcome

Harms

How harms were defined and assessed (e.g., systematically, non-systematically)

Sample size

How sample size was determined, including all assumptions supporting the sample size calculation

Explanation of any interim analyses and stopping guidelines

Who generated the random allocation sequence and the method used

Type of randomisation and details of any restriction (e.g., stratification, blocking and block size)

Allocation concealment mechanism

Mechanism used to implement the random allocation sequence (e.g., central computer/telephone; sequentially numbered, opaque, sealed containers), describing any steps to conceal the sequence until interventions were assigned

Implementation

Whether the personnel who enrolled and those who assigned participants to the interventions had access to the random allocation sequence

Blinding

Who was blinded after assignment to interventions (e.g., participants, care providers, outcome assessors, data analysts)

If blinded, how blinding was achieved and description of the similarity of interventions

Statistical methods used to compare groups for primary and secondary outcomes, including harms

Eligibility criteria Intervention comparator

Randomisation: Sequence generation

Statistical methods

Definition of who is included in each analysis (e.g., all randomised participants), and in which group

How missing data were handled in the analysis

Methods for any additional analyses (e.g., subgroup and sensitivity analyses), distinguishing prespecified from post-hoc

For each group, the numbers of participants who were randomly assigned, received intended intervention, and were analysed for the primary outcome

For each group, losses and exclusions after randomisation, together with reasons

Dates defining the periods of recruitment and follow-up for outcomes of benefits and harms

If relevant, why the trial ended or was stopped

Intervention and comparator as they were actually administered (e.g., where appropriate, who delivered the intervention/comparator, how participants adhered, whether they were delivered as intended [fidelity])

Concomitant care received during the trial for each group

Baseline data

A table showing baseline demographic and clinical characteristics for each group

Numbers analysed, outcomes and estimation

For each primary and secondary outcome, by group: the number of participants included in the analysis the number of participants with available data at the outcome time point result for each group, and the estimated effect size and its precision (such as 95% confidence interval) for binary outcomes, presentation of both absolute and relative effect size

Harms

All harms or unintended events in each group

Ancillary analyses

Any other analyses performed, including subgroup and sensitivity analyses, distinguishing prespecified from post-hoc

Interpretation consistent with results, balancing benefits and harms, and considering other relevant

Results

Participant flow, including flow diagram Recruitment Intervention comparator delivery

Discussion

Interpretation

evidence Limitations

Trial limitations, addressing sources of potential bias, imprecision, generalisability, and, if relevant, multiplicity of analyses

*We strongly recommend reading this statement in conjunction with the CONSORT 2025 Explanation and Elaboration and/or the CONSORT 2025 Expanded Checklist for important clarifications on all the items. We also recommend reading relevant CONSORT extensions. See www.consort-spirit.org.

Citation: Hopewell S, Chan AW, Collins GS, Hróbjartsson A, Moher D, Schulz KF, et al. CONSORT 2025 Statement: updated guideline for reporting randomised trials. BMJ. 2025; 388:e081123. https://dx.doi.org/10.1136/bmj-2024-081123.

Appendix 2 Supplementary Methods Development of an AI-Based Model and Visual Assistive Tool For AI model development, laparoscopic surgical videos collected from 20 medical centers between January 6, 2023, and September 1, 2023, were used. The inclusion criteria for the videos were defined as follows: a resolution of no less than 720$\times$560 pixels and full normal playback capability.

All videos were retrospectively collected and anonymized prior to analysis to ensure data security and privacy protection. To prevent data leakage between subsets, videos were split on a per-video basis into training, validation, and test sets at a ratio of 6:2:2.

During data annotation, two experienced surgeons (HC and XW) independently labeled the onset and offset of active bleeding events in each video. Discrepancies were resolved by consultation with a third senior surgeon to reach a consensus.

After temporal annotation, a 4-second clip (2 seconds before and 2 seconds after event onset) was extracted for each bleeding event to form the bleeding-event subset. During non-bleeding periods of each video, up to 100 non-overlapping 4-second clips were randomly sampled to form a nonbleeding control subset. For all frames in the bleeding-event subset, Label Studio was used for frame-by-frame point annotation of bleeding site locations. These annotations provided spatiotemporal ground truth for model training, including bleeding site localization and temporal tracking.

Based on the annotation information of the aforementioned video frames, a multi-task deep learning framework was constructed for detection of bleeding frames and tracking of bleeding sites in laparoscopic surgery videos. The framework adopted a hierarchical spatiotemporal architecture in which video segments were processed through a shared backbone network to jointly perform three tasks within a unified model: bleeding presence classification, frame-level bleeding detection, and spatial localization of bleeding sites. The architecture was designed to balance spatial localization precision and temporal consistency by integrating feature representations at multiple spatial and temporal scales, thereby enabling preservation of fine-grained visual details while incorporating higher-level semantic context. For loss function design, task-specific loss functions were employed, with an L1 loss used for optimization of bleeding site spatial localization.

To evaluate the potential utility of the proposed model in supporting surgeons during operative tasks, a surgical visual assistance tool was developed and deployed in a preclinical setting. The primary design objective of the tool was to integrate the bleeding detection and tracking algorithm into a task-oriented interactive interface. Functionally, the algorithm was embedded within the core system of the tool, with an option allowing users to activate or deactivate AI-assisted visual guidance. When activated, the system was configured to analyze surgical videos in real time and highlight detected bleeding sites using a hollow green bounding box. When deactivated, the system displayed the surgical videos without any AI-generated visual prompts. In addition, the tool incorporated an automated data recording and feedback module. This module recorded response time and number of clicks, compared each click location with ground truth bleeding site annotations using spatial matching, and provided immediate feedback to users. This design enabled direct comparison of surgeons' task performance under AI-assisted and non-assisted conditions, facilitating objective evaluation of the algorithm' s supportive role in bleeding management.

Model Performance Evaluation

Model performance was evaluated using sensitivity (recall), specificity, precision, F1-score, and the area under the receiver operating characteristic (ROC) curve. These metrics were defined as follows:

Sensitivity/ Recall = TP/ (TP+FN) Specificity = TN/ (FP+TN) Precision = TP/ (TP+FP) F1 = (2×Precision×Recall)/ (Precision +Recall) Here, TP, FP, TN, and FN represent the numbers of true positive, false positive, true negative, and false negative pixels, respectively. All these indicators range from 0 to 1, with higher values indicating better performance.

In addition, mean absolute error (MAE) was defined as the mean absolute difference between the bleeding frames predicted by the model and the manually annotated ground-truth bleeding frames in uniformly sampled video clips. Mean distance error (MDE) was defined as the Euclidean distance between the coordinates of bleeding sites predicted by the model and the corresponding manually annotated ground-truth coordinates. All image coordinates were normalized, with the top-left corner defined as the origin.

Randomization and Blinding Participants were stratified by surgical experience and randomly assigned to study groups using block randomization. The random allocation sequence was generated by a professional statistician using R software and stored in sealed, opaque envelopes to ensure allocation concealment.

Participant enrollment and group assignment were performed by a dedicated research assistant(PX) who was not involved in subsequent data collection or statistical analysis. Group allocation was determined by drawing a sealed envelope immediately after each participant was formally enrolled.

Owing to the visually apparent nature of the AI-assisted visual guidance, blinding of participants, surgeons, investigators, and outcome assessors was not feasible; accordingly, no blinding was implemented in this trial. All personnel involved in the intervention, data collection, and statistical analysis remained aware of group assignments throughout the study.

Sample Size Calculation The sample size was determined based on results from a preliminary pilot experiment. In the pilot experiment, the mean reaction time was 1.99 seconds in the experimental group and 2.18 seconds in the control group, with a pooled standard deviation of

0.

1. Using an independent two-sample ttest (two-tailed $\alpha = 0.05$, power = 0.80), an initial sample size of 67 participants was estimated.

Allowing for an anticipated attrition rate of 15%, the final planned sample size was 77 participants, ensuring balanced representation of junior and senior surgeons.

Appendix Figure Methods. Workflow of the AI-based bleeding sites detection model for laparoscopic surgical videos: This diagram outlines the full pipeline

of data processing, annotation, and model construction for identifying bleeding events (frames and locations) in surgical video footage:

1. Data Collection & Splitting: Surgical videos meeting inclusion criteria (resolution $\geq$ 720$\times$560, no patient privacy issues) are collected, then split into training, validation, and test sets at a 6:2:2 ratio;
2. Video & Frame Annotation: Two surgeons independently annotated bleeding events in the videos,

extracted 4-second video clips of bleeding events, and conducted frame-by-frame labeling of hemorrhage points in each of these clips.

3. Model Input Preparation: Annotated video clips are processed via 3D patch embedding to generate spatiotemporal features;
4. Model Architecture & Outputs: Features are fed into a hierarchical network (Hiera Stage 1–4) with 3D convolution modules. The network outputs three results via dedicated heads: Classification Head: Classifies clips as "Bleeding"or "Non-bleeding"; Detection Head: Identifies bleeding frames within clips; Location Head: Predicts the 2D coordinates (x,

y) of bleeding points in frames.

Appendix 3 Appendix Figure

1. The ROC curve of AI-based bleeding sites detection model: AUC was 0.985 (95% CI: 0.980–0.990)

Appendix 4 Table Baseline characteristics for the participants in preclinical validation.

Overall

Experimental Group

Control group

(AI-assisted)

(No AI-assisted)

(n=38)

(n=39)

25 (65.79)

33 (92.31)

(n=77) Gender Men (n, %)

58 (75.32)

Women (n, %)

19 (24.68)

13 (34.21)

6 (15.39)

30.09$±$5.37

30.71$±$6.04

29.49$±$4.03

Senior surgeon (n, %)

38 (49.35)

19 (50.00)

19 (48.72)

Junior surgeon (n, %)

39 (50.65)

19 (50.00)

20 (51.28)

Almost never play (<0.5 hours/week)

36 (46.75)

18 (47.37)

18 (46.15)

Occasional play (0.5-2 hours/week)

21 (27.27)

12 (31.58)

9 (23.08)

Frequent play (>2 hours/week)

20 ((25.97)

8 (21.05)

12 (30.77)

Age (year) (mean±standard deviation) Seniority*

Previous gaming experience (n, %)

Monthly laparoscopic surgery volume (n, %) < 5 cases

31 (40.26)

13 (34.21)

18 (46.15)

5-10 cases

21 (27.27)

12 (31.58)

9 (23.08)

11-20 cases

10 (12.99)

6 (15.79)

4 (10.26)

> 20 cases

15 (19.48)

7 (18.42)

8 (20.51)

Never

34 (44.16)

20 (52.63)

15 (38.46)

1-2 times

36 (46.75)

15 (39.47)

20 (51.28)

3-5 times

5 (6.49)

2 (5.26)

3 (7.69)

5 times

2 (2.60)

1 (2.63)

1 (2.56)

Bad events experience (n, %)

Physical discomfort during the task (n, %)

58 (75.32)

32 (84.21)

27 (69.23)

Mild discomfort†

19 (24.68)

6 (15.79)

12 (30.77)

Severe discomfort

0 (0)

0 (0)

0 (0)

*Seniority: grouping was performed according to the duration of laparoscopic surgery experience. The junior surgeon group (laparoscopic surgery experience $\leq$ 3 years) and the senior surgeon group (laparoscopic surgery experience $>$ 3 years).

Monthly laparoscopic surgery volume: number of participants per surgical volume category.

+ Bad events experience: participants with a prior history of severe consequences resulting from delayed hemostasis during laparoscopic surgery. †Mild discomfort: such discomfort did not necessitate halting the task for adjustments.

Appendix 5

Table NES of participants in the experimental and control groups at different stages during the preclinical validation.

Experimental Group

Control group

(AI-assisted)

(No AI-assisted)

(n=38)

(n=39)

NES before validation

1.71$\pm$1.01

1.92$\pm$1.20

NES during validation

2.68$\pm$1.28

3.08$±$1.06

NES after validation

1.68$±$0.84

2.95$±$1.05

<0.001

NES during minus before validation

0.97$±$1.41

1.15$±$1.18

NES after minus before validation

-0.03$±$1.37

1.03$±$1.65

NES: negative emotion scores. Data in this table are presented as mean±standard deviation, with values of P < 0.05 shown in bold.

P value

Note: Figure translations are in progress. See original paper for figures.