Pilot study on surgical procedure labeling using artificial intelligence video analysis

Kan Tanabe, M.D., Ph.D.[1,2]; Kenji Baba, M.D., Ph.D.[1]; Tsubasa Hidaka, B.S.[3]; Rara Deguchi, B.S.[3]; Haruka Eto, B.S.[3]; Mashiho Mukaida, M.S., Ph.D.[3]; Masumi Wada, M.D., Ph.D.[1]; Naoki, Kuroshima, M.D., Ph.D.[1]; Takeru Arima, M.D.[2], Yuta Hirakawa, M.D.[2], Taiki Nakashima, M.D. , Ph.D.[2]; Keiichi Yonemori, M.D., Ph.D.[2]; Shuichi Hokita, M.D., Ph.D.[2]; Yoshikazu Uenosono, M.D., Ph.D.[2]; Noritaka Shigei, M.S., Ph.D.[3]; Satoshi Ono M.S., Ph.D.[3]; Takao Ohtsuka, M.D., Ph.D.[1]

1. Department of Digestive Surgery, Graduate School of Medical and Dental Sciences, Kagoshima University, Kagoshima, Japan
2. Department of Digestive Surgery, Imamura General Hospital, Kagoshima, Japan
3. Department of Information Science and Biomedical Engineering, Graduate School of Science and Engineering, Kagoshima University, Kagoshima, Japan

Disclosures
The authors have no conflicts of interest to disclose.

INTRODUCTION

- ✓ Minimally invasive gastrointestinal surgery, including laparoscopic and robot-assisted techniques, requires mastery of diverse and advanced operative skills. Efficient acquisition of these skills depends on structured feedback; however, conventional assessments often rely on subjective judgment. Artificial intelligence (AI)–based surgical video analysis offers a promising approach to quantify technical performance and enable objective feedback.
- ✓ This study aimed to characterize technical differences between trainees and Endoscopic Surgical Skill Qualification System (ESSQS)–qualified surgeons in laparoscopic right colectomy using manual video annotation, and to evaluate the feasibility of AI-based automatic action classification for objective surgical education.

METHODS

- ✓ Surgical videos of 30 patients who underwent laparoscopic right colectomy at our institution between January 2017 and December 2022 were retrospectively analyzed.
- ✓ The operative segment from superior mesenteric vein exposure to ileocolic vein division was extracted. Videos were divided into 1/60-second frames, and each frame was annotated with one of eight action labels: blunt dissection, sharp dissection, switch (tissue regrasping), hemostasis, cut (vessel cutting), clip (Vessel clipping), separation (with dissecting forceps), or no action.
- ✓ Manual annotation was performed independently by two ESSQS-qualified surgeons.
- ✓ Operative performance was compared between trainees and ESSQS-qualified surgeons.
- ✓ Automatic action classification was attempted using temporal action segmentation models: Multi-Stage Temporal Convolutional Network (MS-TCN)[1], MS-TCN++[2], and an ensemble approach.

1. Y. A. Farha, et al. IEEE/CVF Conf. CVPR, 2019
2. S. Li, et al. IEEE Trans. Pattern Anal. Mach. Intell, 2020

Figure1. Action labels

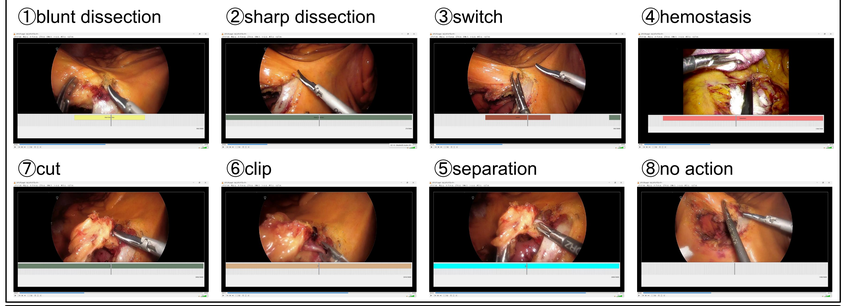


Table1. Surgeon characteristic

	trainees (n=20)	ESSQS surgeons (n=10)	P value
Postgraduate years, median [range]	11 [6-22]	15 [14-26]	0.01

RESULTS

- ✓ Trainees required significantly more time for tissue handling and hemostasis than qualified surgeons.
- ✓ Time-distribution analysis showed trainees exhibited a higher proportion of “no action,” reflecting inefficient or unclassified movements, whereas qualified surgeons demonstrated greater proportions of dissection, particularly sharp dissection. These findings underscore distinct technical differences between groups.
- ✓ The accuracy of AI-based classification was 62.5% with MS-TCN, 62.5% with MS-TCN++, and 63.7% with the ensemble model. Across all models, accuracy was lower for blunt dissection, hemostasis, and tissue handling, therefore, additional refinement of accuracy is necessary for clinical application.

Table2. Operative performance

	Procedure time analysis			Time-distribution analysis		
	trainees (n=20) Average seconds	ESSQS surgeons (n=10) Average seconds	P value	trainees (n=20) Average(%)	ESSQS surgeons (n=10) Average(%)	P value
Dissection	76.7	118.1	0.08	16.8	27.1	0.02
Blunt dissection	27.4	32.1	0.27	6.6	7.9	0.15
Sharp dissection	49.3	86.0	0.07	10.2	19.2	0.02
Switch	33.9	21.0	0.04	7.0	5.6	0.14
Hemostasis	25.1	7.2	0.02	5.3	2.2	0.09
Separation	46.9	40.4	0.32	9.9	11.2	0.33
Clip	37.3	29.2	0.09	9.1	8.9	0.44
Cut	24.3	21.2	0.28	5.4	6.2	0.21
No action	232.4	147.0	0.03	46.4	38.7	<0.01
Total	476.5	384.2	0.14	100	100	-

Figure2. Time-distribution analysis

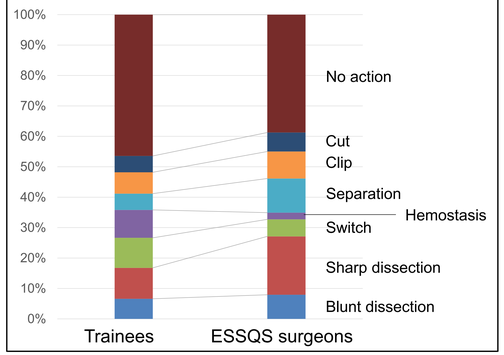
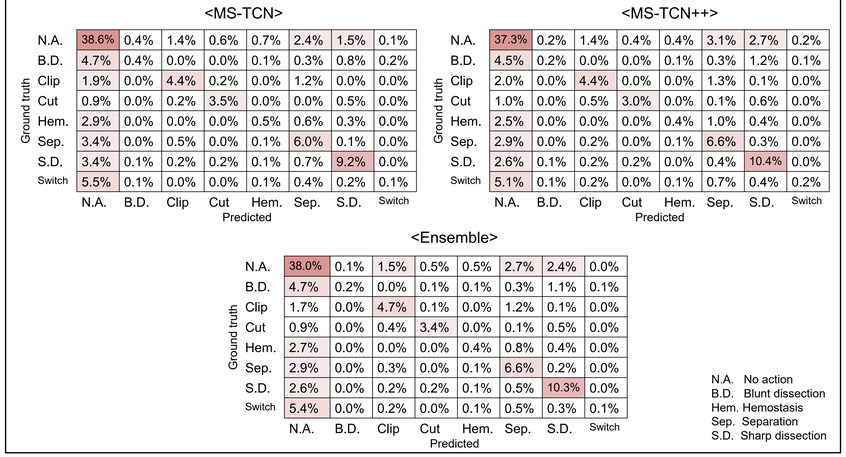


Table3. AI Model Performance

	Accuracy ± SD (%)
MS-TCN	62.54 ± 8.09
MS-TCN++	62.53 ± 8.55
Ensemble	63.70 ± 8.33

Figure3. Confusion matrix



DISCUSSIONS

- ✓ Previous studies have shown that, compared with trainees, expert surgeons exhibit characteristic performance patterns, including shorter movement distances of the non-dominant instrument, a higher proportion of simultaneous bilateral instrument movements, shorter incision times, and reduced time required for surgical field exposure[3,4].
- ✓ In this study, we evaluated surgical performance not only by analyzing instrument motion but also by incorporating action labeling. This framework may enable a more comprehensive assessment of surgical technique and provide meaningful feedback for surgical education.
- ✓ Previous study on surgical video analysis have shown that action recognition is more challenging than surgical phase recognition or instrument detection[5]. Therefore, extracting discriminative features of surgical actions may contribute to improving classification accuracy.

3. Erlend Fagertun Hofstad, et al. Surg Endosc, 2013
4. Sato K, et al. Langenbeck's Archives of Surgery, 2025
5. Wagner M, et al. Med Image Anal, 2023

CONCLUSIONS

- ✓ Manual surgical video analysis enabled objective and quantitative identification of technical differences between trainees and qualified surgeons in laparoscopic right colectomy. Such analyses can provide actionable feedback for training and contribute to standardized assessment frameworks.
- ✓ Although current AI-based classification accuracy remains modest, further model refinement holds promise for automated, objective feedback systems. Integration of manual and AI-driven approaches may advance surgical education by enabling reproducible, quantitative skill evaluation and supporting individualized training strategies.