Original Article | **Open Access**

Artificial Intelligence Surgery

Park *et al. Art Int Surg.* 2026;6:710-24 **DOI:**10.20517/ais.2026.91

Hierarchical quantification of operative workflow in laparoscopic left lateral sectionectomy using step-task-action triplets and automated recognition

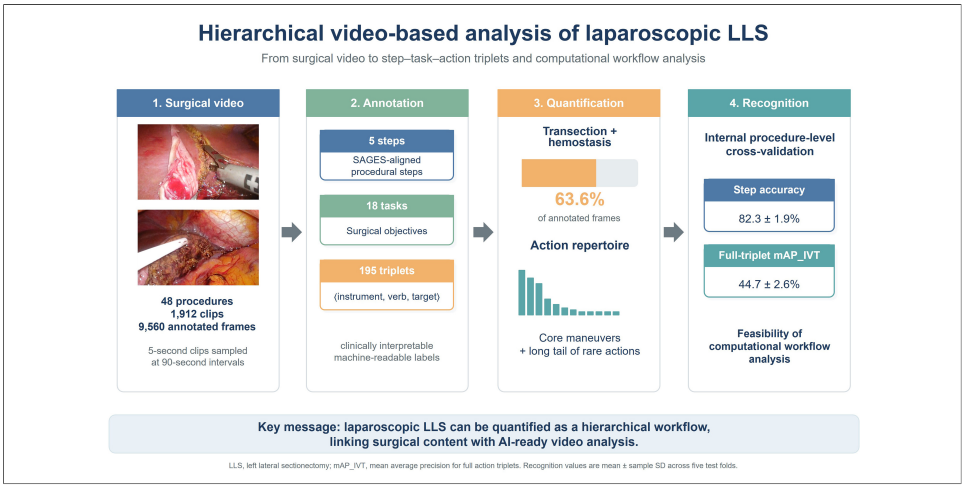
Han Sang Park[1,#], Yongjun Jeon[2,3,#], Namkee Oh[1,2], Kanggil Park[2,3], Jongmin Shin[2], Kyu-Hwan Jung[2,3], Jinsoo Rhu[1], Jongman Kim[1], Gyu-Seong Choi[1]

Keywords:
Surgical video analysis, computer vision, artificial intelligence, surgical data science, liver surgery, workflow annotation, action recognition

Citation: Park HS, Jeon Y, Oh N, Park K, Shin J, Jung KH, Rhu J, Kim J, Choi GS. Hierarchical quantification of operative workflow in laparoscopic left lateral sectionectomy using step-task-action triplets and automated recognition. *Art Int Surg.* 2026;6:710-24. https://dx.doi.org/10.20517/ais.2026.91

Received: 31 Aug 2026 **First Decision:** 14 Oct 2026 **Revised:** 28 Oct 2026 **Accepted:** 5 Nov 2026 **Published:** 19 Nov 2026

Academic Editor: Andrew Gumbs
Copy Editor: Tong Wang
Production Editor: Tong Wang



Abstract

Aim: Laparoscopic left lateral sectionectomy (LLS) is an index procedure for learning laparoscopic hepatectomy, but quantitative descriptions of its fine-grained workflow remain limited. This study characterized its hierarchical workflow and evaluated automated recognition from surgical video.

Methods: Forty-eight laparoscopic LLS procedures with archived surgical videos were retrospectively analyzed. Videos were sampled into 5-s clips at 90-s intervals, yielding 9,560 frames at 1-s resolution. A hierarchy aligned with the Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) framework comprised 5 steps, 18 tasks, and frame-level action triplets <instrument, verb, target> (IVT). Automated recognition used five-fold cross-validation with procedure-level splits.

[1]Department of Surgery, Samsung Medical Center, Seoul 06351, Republic of Korea.
[2]Clinical Robotics and Embodied AI Research Center, Smart Healthcare Research Institute, Research Institute for Future Medicine, Samsung Medical Center, Seoul 06351, Republic of Korea.
[3]Department of Medical Device Management and Research, Samsung Advanced Institute for Health Sciences & Technology (SAIHST), Sungkyunkwan University, Seoul 06351, Republic of Korea.
[#]These authors contributed equally to this work.

Correspondence to: Prof. Namkee Oh, Prof. Gyu-Seong Choi, Department of Surgery, Samsung Medical Center, Seoul 06351, Republic of Korea. E-mail: ngnyou@gmail.com; med9370@gmail.com

Results: Based on recorded step labels, parenchymal transection accounted for 52.5% of annotated frames and post-transection hemostasis for 11.1%. The 195 unique action triplets showed a long-tail distribution: the five most frequent accounted for 42.1% of all occurrences, whereas more than half occurred fewer than 50 times. Step recognition achieved an accuracy of 82.3% ± 1.9%. Task recognition achieved an accuracy of 62.6% ± 4.5% and a macro-averaged Jaccard index of 38.7% ± 4.2%. Mean average precision was 44.7% ± 2.6% for full action triplets and 67.5% ± 2.3% for instruments.

Conclusion: Laparoscopic LLS showed a transection-dominant workflow in the sampled video annotations, with frequent core maneuvers and infrequent, context-specific actions. Hierarchical annotation supports quantification of operative content and automated workflow recognition; its potential applications in education, benchmarking, and outcome-linked analysis require further validation.

INTRODUCTION

Laparoscopic left lateral sectionectomy (LLS) is widely regarded as an index procedure for surgeons learning laparoscopic hepatectomy, providing structured exposure to hepatic mobilization, vascular control, parenchymal transection, and hemostasis[1-3]. Despite this established role, quantitative descriptions of its operative composition at the level of individual technical actions remain limited.

Surgical workflow analysis addresses this gap by decomposing operative video into structured, interpretable components[4-6]. The Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) framework recommends a hierarchical representation of steps, tasks, and actions (instrument-tissue interactions), enabling standardized quantification of operative content[7]. Automated phase and action recognition has been investigated across different procedures[8-11]. However, much of this work has focused on laparoscopic cholecystectomy[8,10,11], and its applicability to more complex procedures remains uncertain, particularly with respect to scalable hierarchical annotation and fine-grained action-level descriptions. Previous studies have demonstrated automated workflow recognition in liver surgery[12,13].

By enabling quantitative representation of operative procedures, surgical workflow analysis may support objective comparison across cases and institutions. Benchmarking has largely focused on perioperative outcomes rather than intraoperative processes[14], limiting insight into how results are achieved. Without structured workflow representation, intraoperative processes remain difficult to quantify and compare, constraining process-level benchmarking and generalizable models of surgical performance[7].

This study aimed to characterize the hierarchical workflow of laparoscopic LLS using a SAGES-aligned framework and to evaluate its capacity for automated recognition across multiple levels, establishing a framework for workflow quantification with potential applications in surgical education and benchmarking.

METHODS

Study design and setting

This was a single-center, retrospective observational study designed to characterize the hierarchical workflow of laparoscopic LLS and evaluate the feasibility of automated workflow recognition from surgical video. The observational cohort is reported according to the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) guidance, and the model development and evaluation according to the Checklist for Artificial Intelligence in Medical Imaging (CLAIM 2024)[15]. Endoscopic videos were retrieved from a prospectively maintained surgical video archive at a tertiary referral hospital. All procedures were performed between February 2020 and March 2022 by four hepatobiliary surgeons with extensive experience in laparoscopic liver surgery. The study was approved by the Institutional Review Board of Samsung Medical

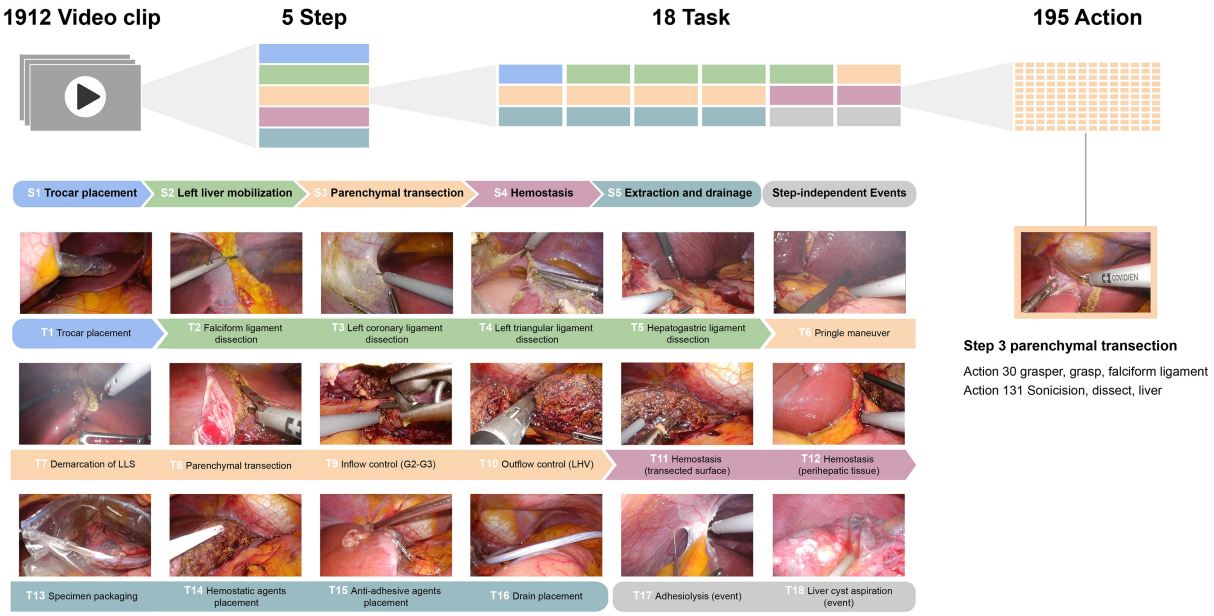


Figure 1. Hierarchical workflow definition of LLS. The procedure was organized into five steps (S1-S5), with 16 tasks (T1-T16) grouped by their primary clinical association. Representative intraoperative images are shown for each task. Adhesiolysis and liver cyst aspiration (T17-T18) were annotated as step-independent events, as they may occur at any stage of the operation regardless of the procedural step. LLS: Laparoscopic left lateral sectionectomy.

Center (IRB No. SMC-2023-04-128-001). All surgical videos were de-identified, and the requirement for individual informed consent was waived by the IRB.

Patients and surgical technique

The study included 48 adult patients who underwent elective laparoscopic LLS during the study period and whose surgical videos were available in the institutional archive. All 48 selected procedures were included in the analysis. Baseline clinical characteristics, operative variables, and postoperative outcomes were collected. All procedures followed the standardized laparoscopic LLS technique used at our institution. After trocar placement, the left lateral section was mobilized by dissecting the falciform, left coronary, and left triangular ligaments, with adhesiolysis or additional mobilization as required. Parenchymal transection was performed using energy devices and staplers, with inflow control of the segment II/III pedicles and division of the left hepatic vein as indicated. Hemostasis of the transection surface was achieved with bipolar coagulation and topical hemostatic agents. The specimen was retrieved in a bag, with drain placement at the surgeon's discretion. This sequence corresponds to the five procedural steps defined below.

Video acquisition and sampling

All procedures were recorded using a standard laparoscopic tower. Video resolution was either 1,920 × 1,080 pixels at 60 frames per second or 1,356 × 728 pixels at 30 frames per second, depending on the recording system. Using FFmpeg, 5-s clips were extracted at 90-s intervals from the beginning of the endoscopic view to the end of the procedure. The 5-s duration provided brief local temporal context, while the 90-s interval balanced sampling coverage across the procedure with the workload of expert annotation. The sampling protocol yielded 1,912 clips, all of which were retained for analysis. Five frames were extracted from each clip at 1-s intervals, yielding 9,560 annotated frames. The resulting dataset was designated LLS48 and represented approximately 2 h 39 min of sampled video.

Hierarchical workflow definition

A three-level hierarchical workflow model aligned with the SAGES consensus framework[7] was adopted, consisting of steps, tasks, and actions [Figure 1 and Supplementary Table 1].

Steps were defined as procedure-specific stages fulfilling distinct clinical goals. Five steps were defined: S1, trocar placement; S2, left liver mobilization; S3, parenchymal transection; S4, hemostasis after transection; and S5, extraction and drainage. The distinction between S3 and S4 reflects the shift from dividing liver parenchyma and controlling vascular structures to securing hemostasis of the transected surface.

Tasks were defined as semantically coherent surgical objectives required to accomplish a procedural step and potentially shared across procedures. Eighteen tasks were defined: 16 tasks (T1-T16) grouped by their primary clinical association and two step-independent events (T17-T18). The 16 grouped tasks were trocar placement (T1); dissection of the falciform ligament (T2); dissection of the left coronary ligament (T3); dissection of the left triangular ligament (T4); dissection of the hepatogastric ligament (T5); Pringle maneuver (T6); demarcation of the left lateral section (T7); parenchymal transection (T8); inflow control of the segment II and III pedicles (T9); outflow control of the left hepatic vein (T10); hemostasis of the transected liver surface (T11); hemostasis of the perihepatic tissue (T12); specimen packaging (T13); placement of hemostatic agents (T14); placement of anti-adhesive agents (T15); and drain placement (T16). Adhesiolysis (T17) and liver cyst aspiration (T18) were annotated as step-independent events because they may arise in response to intraoperative findings rather than as part of a fixed procedural sequence. Although these tasks have no fixed parent step in the descriptive hierarchy, their frames retained the corresponding procedural step labels for five-class step recognition. Both tasks were included as separate classes in the 18-task recognition analysis.

Step labels reflected the current procedural context, whereas task labels identified specific technical objectives, including those occurring across different steps. For example, temporary hemostasis during ongoing transection was labeled as S3 with the corresponding hemostasis task, whereas S4 denoted hemostasis after completion of transection. Task groupings in Figure 1 and Supplementary Table 1 indicate their primary clinical associations and do not determine the recorded step labels. The Pringle maneuver, inflow/outflow control, and demarcation were grouped under S3 because they support transection; in this cohort, the Pringle maneuver occurred exclusively within S3.

Actions represented the finest level of granularity and were defined at the frame level as action triplets < instrument, verb, target> (IVT)[7,11]. This framework captures the fundamental structure of a surgical maneuver - what instrument is being used, what the surgeon is doing, and which entity is acted upon - and corresponds to how operative technique is described clinically. Across all annotated frames, 17 instrument categories, 16 verb categories, and 38 target categories were identified, resulting in 195 observed action triplet combinations. Representative examples include <grasper, grasp, falciform ligament>, <Sonicision, dissect, liver parenchyma>, and <clipper, clip, intrahepatic vein>. The component and triplet label distributions are presented in Supplementary Tables 2 and 3, respectively.

Action annotation was multi-label, using four role-hand slots per frame corresponding to the surgeon's and assistant's left and right hands. Each non-empty slot contained an IVT triplet. Empty slots contributed no triplet occurrence and were distinct from explicit null_verb or null_target labels. Identical triplets appearing in multiple slots were counted once per frame, while different triplets within the same frame contributed separately.

The target denotes the entity acted upon and is not restricted to anatomical structures. Targets therefore include anatomical structures, blood, fluid, adhesions, and devices or materials such as clips and drain tubes. This interpretation is consistent with action-triplet frameworks that include non-anatomical targets[16]. The instrument field retains the original codebook categories, including hemostatic and anti-adhesive agents.

Brand-specific labels include Sonicision (Medtronic, Boulder, CO, USA) and Hem-o-lok (Teleflex Medical, Morrisville, NC, USA).

A null_verb label indicates that an instrument is visible without an identifiable action, and null_target indicates that no target is being acted upon. These states are distinguished from an empty instrument slot. They were retained in the descriptive annotation inventory.

Annotation process

All video clips were annotated by five domain experts, including two hepatobiliary surgeons and three surgical scopists, each with more than 10 years of minimally invasive surgery experience. In the first round, the scopists independently assigned step, task, and action triplet labels using a custom web-based annotation dashboard. In the second round, the surgeons reviewed all annotations, resolved disagreements, and refined class definitions through consensus meetings.

Because this study established a new hierarchical framework, annotation was iterative and class definitions were refined as ambiguous cases were encountered. The original dataset represents a consensus reference derived through structured expert deliberation, comprehensive surgeon review, and alignment with the SAGES hierarchy[7]. Initial ratings made under evolving definitions were not used to estimate agreement for the finalized codebook. An additional annotation review was therefore conducted after the original consensus annotation.

For additional annotation review, the 956 odd-indexed clips among the 1,912 clips were selected. One of the five sampled frames in each selected clip was chosen at random, yielding 956 frames (10.0% of the 9,560 annotated frames). Two additional reviewers independently reviewed the existing frame-level IVT annotations and retained or revised the labels. Each reviewer was assigned 574 frames, including 192 frames assigned to both. Each frame contained four role-hand slots corresponding to the surgeon's and assistant's left and right hands. The original consensus annotations, assigned using the codebook, served as the reference labels. Agreement was assessed both between each reviewer and the reference labels and between the two reviewers. For complete-configuration agreement, the four-slot annotation was treated as one nominal category. A frame with an unassessable slot was excluded from complete-configuration comparisons; slot-level comparisons excluded only that slot. Empty slots and explicit null_verb or null_target labels were retained as distinct states. The finalized codebook was held fixed throughout the additional annotation review.

Automated workflow recognition

Action triplet recognition was performed using the published CurConMix methodology[17] with ImageNet-pretrained Swin Transformer-Base (Swin-B) backbones. CurConMix was developed in part by a co-first author of the present study and was applied here to the LLS48 dataset. Input images were resized to 224 × 224 or 384 × 384 pixels according to the backbone configuration and normalized using ImageNet statistics. Spatial models were optimized using AdamW with an initial learning rate of 2×10^{-4} and a weight decay of 1×10^{-6}. A cosine-annealing scheduler with warm restarts was used with a minimum learning rate of 2×10^{-5}. The batch size was 64 for Swin-B with 224 × 224 inputs and 32 for Swin-B with 384 × 384 inputs.

For LLS48, CurConMix training comprised seven epochs of curriculum-based supervised contrastive pretraining, followed by 20 epochs of teacher training and 40 epochs of student training under a self-distillation (SelfD) framework. The curriculum progressed from target labels to instrument-target pairs and then to full action triplets. Feature-level and input-level mixup used $\alpha = 0.4$, and the supervised contrastive temperature was set to 0.1. Training augmentations included random horizontal flipping, shifts, scaling,

rotation, hue-saturation-value perturbations, and Gaussian noise. Validation images underwent resizing and normalization without random augmentation.

For step- and task-level recognition, the action-triplet-trained CurConMix student encoder was frozen and used to extract frame-level visual features. Separate TeCNO temporal decoders were trained on these features to predict the five step classes and 18 task classes[18]. The temporal decoders were trained for 100 epochs using the Adam optimizer, with a learning rate of 7×10^{-4}, a weight decay of 1×10^{-5}, and a batch size of one sequence. Temporal inputs consisted of sampled observations rather than densely annotated continuous video timelines. Action-triplet recognition results for Rendezvous (RDV)[16], SelfD[19], tail-enhanced representation learning (TERL)[20], and CurConMix[17] are summarized in Supplementary Table 4.

Evaluation used five fixed procedure-level folds, with all clips and frames from each procedure assigned to the same fold. In each run, one fold was held out for testing and excluded from all stages of LLS48-based training, including curriculum-based supervised contrastive pretraining. Fixed-epoch pretraining used the remaining four folds. For subsequent CurConMix teacher and student training and TeCNO training, three folds were used for training and one prespecified fold for validation. Thus, the validation fold was separate from subsequent model fitting but had contributed to pretraining.

Teacher and student checkpoints were selected according to the highest validation triplet average precision (AP). TeCNO checkpoints were selected according to the highest macro AP across classes with positive validation examples. Hyperparameters were fixed across folds. For the revised CurConMix and downstream TeCNO analyses, test-fold results were used only for performance evaluation after checkpoint selection and were not used for hyperparameter tuning. The SelfD-B and TERL-B results retained in Supplementary Table 4 were obtained from earlier experiments in which checkpoints were selected according to their highest test-fold performance; these models were not retrained under the revised protocol.

Outcome measures and statistical analysis

Step-level workflow composition was calculated directly from the recorded frame-level step labels, with all 9,560 frames assigned to one of five steps. Task counts were calculated separately from the recorded task labels. Clinical task groupings were used for presentation only and were not summed to derive step counts. Proportions used all 9,560 annotated frames as the denominator. Action-triplet frequencies were summarized as occurrence counts and proportions of all annotated triplet occurrences, allowing for multiple triplets within a single frame. Continuous clinical variables were summarized as medians with interquartile ranges (IQRs), and categorical variables as counts and percentages.

For automated recognition, class-specific AP was computed for instruments, verbs, targets, instrument-verb pairs, instrument-target pairs, and full triplets. Class-averaged results are denoted mAP_I, mAP_V, mAP_T, mAP_IV, mAP_IT, and mAP_IVT, respectively. mAP_IVT served as the primary action-recognition metric. Classes containing null_verb and/or null_target were excluded from mAP_IVT averaging. Frames containing null annotations were retained for evaluation of the remaining triplet classes. Null annotations were retained in the descriptive inventory and annotation-agreement analysis. Within each test fold, class-specific AP_IVT was computed separately for each procedure and averaged across procedures containing positive examples of that class. mAP_IVT was then calculated as the unweighted mean across eligible non-null classes with positive examples in that fold.

Step- and task-level recognition were evaluated across all five step classes and all 18 task classes, respectively, including the two step-independent tasks, adhesiolysis and liver cyst aspiration. Each frame was assigned to the class with the highest softmax probability. Within each test fold, frames were pooled across procedures,

Table 1. Patient demographics, operative characteristics, and postoperative outcomes of the 48 laparoscopic left lateral sectionectomy procedures included in the study

Category	Variable	Value
Patient characteristics	Age, years	61.0 (55.0-65.2)
	Sex (male/female)	32/16 (66.7%/33.3%)
	Body mass index, kg/m^2	24.7 (23.0-26.6)
	ASA physical status ≥ III	9 (18.8%)
	Indication for surgery, malignant	37 (77.1%)
Operative characteristics	Operative time, minutes	89.0 (74.2-110.5)
	Estimated blood loss, mL	100.0 (50.0-150.0)
	Use of Pringle maneuver	34 (70.8%)
	Length of hospital stay, days	6.0 (5.0-6.0)
Postoperative outcomes	Major complications (Clavien-Dindo ≥ III)	0 (0.0%)
	30-day mortality	0 (0.0%)

Continuous variables are reported as median with IQR, and categorical variables as number (%). ASA: American Society of Anesthesiologists; IQR: interquartile range.

and accuracy was calculated as the proportion of correctly classified frames. Class-specific precision, recall, and Jaccard index were calculated as TP/(TP + FP), TP/(TP + FN), and TP/(TP + FP + FN), respectively, where TP, FP, and FN denote true-positive, false-positive, and false-negative frame counts. These metrics were macro-averaged with equal weights across all predefined classes, with undefined class-specific values set to zero. Results were summarized as the mean ± sample standard deviation across the five equally weighted test folds.

For annotation agreement, exact agreement and unweighted Cohen's κ were calculated for the complete four-slot configurations, for individual IVT slots, both between the two reviewers and between each reviewer and the codebook-based consensus reference annotations. Percentile 95% confidence intervals (CIs) were obtained from 5,000 surgical-case cluster bootstrap resamples (seed 42), retaining all observations from each resampled case. A sensitivity analysis excluded slots marked as empty by both reviewers. Agreement on the binary decision to retain or modify the original annotation was analyzed separately from agreement on the submitted IVT labels.

RESULTS

Patient and operative characteristics

Forty-eight laparoscopic LLS procedures were included. Baseline characteristics, operative variables, and postoperative outcomes are summarized in Table 1. The indication was malignancy in 37 cases (77.1%). Median operative time was 89 min (IQR 74.2-110.5), and median estimated blood loss was 100 mL (IQR 50.0-150.0). Pringle maneuver was applied in 34 cases (70.8%). No major postoperative complications (Clavien-Dindo grade ≥ III) or deaths within 30 days were observed.

Workflow composition: the procedural landscape of laparoscopic LLS

Across 9,560 annotated frames, the recorded step labels identified parenchymal transection as the dominant procedural step, accounting for 5,015 frames (52.5%). Left liver mobilization accounted for 2,030 frames (21.2%), followed by hemostasis after transection (1,065 frames, 11.1%), extraction and drainage (845 frames, 8.8%), and trocar placement (605 frames, 6.3%) [Figure 2 and Supplementary Table 1]. Together, parenchymal transection and post-transection hemostasis accounted for 6,080 frames (63.6%).

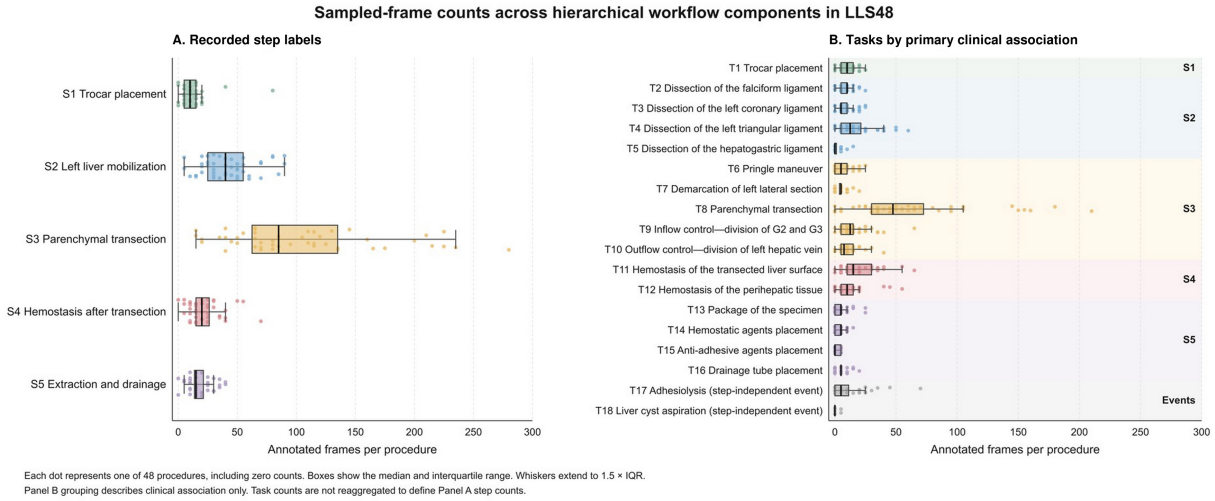


Figure 2. Hierarchical distribution of workflow components in LLS. (A) Distribution of sampled-frame counts calculated directly from the recorded step labels for all 48 laparoscopic LLS procedures. Each frame contributes to exactly one of the five steps, including frames annotated with step-independent tasks. Parenchymal transection (S3) is the dominant step, followed by left liver mobilization (S2) and hemostasis after transection (S4); (B) Distribution of sampled-frame counts for the 18 tasks, arranged by their primary clinical association. This arrangement is descriptive only. Tasks can occur in different procedural contexts, and their grouped counts are not used to calculate the step counts in panel A. Adhesiolysis and liver cyst aspiration are displayed as step-independent events. Each dot represents one procedure, with all 48 procedures included for each step or task, including zero counts. Boxes show the median and IQR, and whiskers reach the most extreme observations within 1.5 times the IQR. Frames were sampled at 1-s intervals within 5-s clips extracted every 90 s. Counts describe sampled annotations, not operative durations or proportions of operative time. LLS: Laparoscopic left lateral sectionectomy; IQR: interquartile range.

At the task level, parenchymal transection was most frequent (2,910 frames), followed by hemostasis of the transected liver surface (935 frames), left triangular ligament dissection (775 frames), inflow control involving division of the segment II and III pedicles (680 frames), and hemostasis of perihepatic tissue (540 frames). Outflow control of the left hepatic vein accounted for 470 frames, whereas tasks such as hepatogastric ligament dissection (75 frames) and liver cyst aspiration (10 frames) were rare [Figure 2 and Supplementary Table 1]. The tasks grouped by primary clinical association with S3 included parenchymal transection, inflow and outflow control, the Pringle maneuver (310 frames), and demarcation (240 frames). Adhesiolysis, annotated as a step-independent event, was observed across 450 frames.

At the action level, 17 instrument categories, 16 verb categories, and 38 target categories yielded 195 unique action triplets IVT [Supplementary Tables 2 and 3]. The five most common triplets were <suction, retract, liver parenchyma> (1,855 occurrences), <Sonicision, dissect, liver parenchyma> (1,710), <grasper, retract, liver> (1,590), <grasper, grasp, round ligament> (1,100), and <suction, aspirate, blood> (885). These five triplets accounted for 7,140 of 16,950 total triplet occurrences (42.1%), reflecting the multi-label annotation in which several triplets may occur in one frame. The distribution showed a pronounced long-tail pattern [Figure 3 and Supplementary Table 3]: more than half of all triplets occurred fewer than 50 times, including context-specific maneuvers involving structures such as the aberrant left hepatic artery, liver cysts, or hepatogastric ligament.

Of the 16,950 triplet occurrences, 2,320 (13.7%) belonged to 16 classes containing null_verb and/or null_target. These comprised 2,250 occurrences across 15 classes containing both null_verb and null_target and 70 occurrences of <anti-adhesive agents, place, null_target>. These 16 classes were retained in the descriptive inventory but excluded from mAP_IVT averaging.

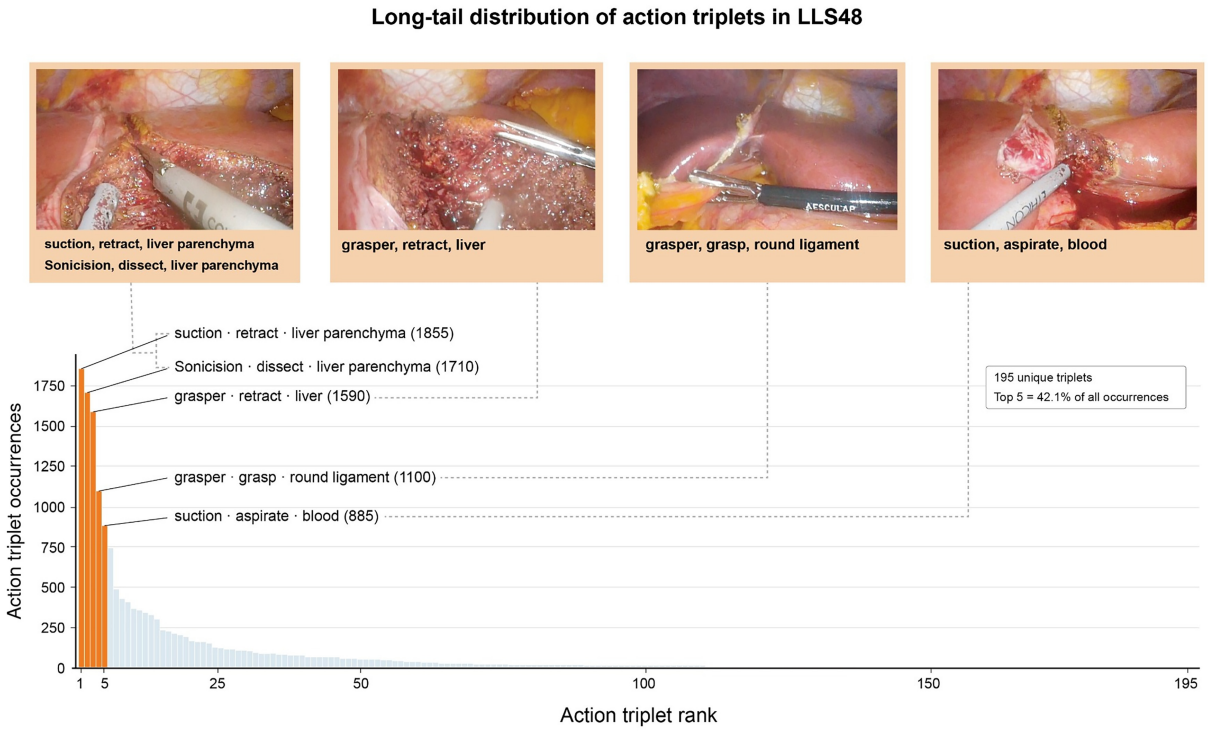


Figure 3. Long-tail distribution of action triplets in LLS. The x-axis ranks 195 unique action triplets (<instrument, verb, target>) by decreasing frequency, and the y-axis shows their annotated occurrence counts (total *n* = 16,950). Each distinct triplet is counted once per annotated frame; different triplets within the same frame contribute separately. The five most frequent triplets account for 42.1% of all occurrences and are highlighted with representative intraoperative images. More than half of the triplets occurred fewer than 50 times, demonstrating a pronounced long-tail distribution. These counts reflect sampled annotations and do not represent proportions of operative time. LLS: Laparoscopic left lateral sectionectomy.

Agreement after independent annotation review

The overlapping subset comprised 192 frames from 10 surgical cases, of which 191 had fully assessable four-slot annotations from both reviewers. Complete-configuration agreement was 86.9% (166/191; κ = 0.867, 95%CI: 0.818-0.921). Slot-level IVT agreement was 96.3% (739/767; κ = 0.947, 95%CI: 0.926-0.969). After excluding slots marked as empty by both reviewers, agreement was 92.0% (324/352; κ = 0.917, 95%CI: 0.882-0.951).

Complete-configuration agreement with the codebook-based consensus reference annotations was 93.7% (538/574; κ = 0.936, 95%CI: 0.912-0.960) for reviewer A and 93.2% (534/573; κ = 0.931, 95%CI: 0.910-0.950) for reviewer B. Agreement between reviewers on retaining versus modifying the original annotations was 88.0% (168/191; κ = 0.144, 95%CI: -0.029-0.349). This binary decision is distinct from agreement on the submitted IVT configuration.

Automated workflow recognition

Automated step recognition achieved a frame-level accuracy of 82.3% ± 1.9% across five-fold cross-validation [Table 2]. Task-level recognition achieved an accuracy of 62.6% ± 4.5% and a macro-averaged Jaccard index of 38.7% ± 4.2%. The task-level confusion matrix showed frequent confusion between parenchymal transection and inflow/outflow control [Supplementary Figure 1].

For action triplet recognition, CurConMix achieved a mean average precision (mAP) of 44.7% ± 2.6% for full <instrument, verb, target> triplets (mAP_IVT) [Table 2]. Component-level mAP was highest for instruments (67.5% ± 2.3%), followed by verbs (64.4 ± 2.9%) and targets (52.8 ± 3.4%), with lower values for pairwise and

Table 2. Performance of hierarchical workflow recognition at the step, task, and action levels using features extracted from the CurConMix framework

Level	Metric	Value (%)
Step	Accuracy	82.3 ± 1.9
	Macro recall	79.8 ± 3.9
	Macro precision	80.2 ± 3.7
	Macro Jaccard	65.5 ± 4.0
Task	Accuracy	62.6 ± 4.5
	Macro recall	55.1 ± 3.2
	Macro precision	54.1 ± 5.1
	Macro Jaccard	38.7 ± 4.2
Action (Triplet)	mAP_I	67.5 ± 2.3
	mAP_V	64.4 ± 2.9
	mAP_T	52.8 ± 3.4
	mAP_IV	46.9 ± 2.9
	mAP_IT	47.1 ± 2.4
	mAP_IVT	44.7 ± 2.6

Results are reported as mean ± sample standard deviation across five equally weighted test folds. For step and task recognition, sampled frames within each test fold were pooled across procedures; accuracy is the proportion of correctly classified frames. Step- and task-level precision, recall, and Jaccard index are macro-averaged across all five step classes and all 18 task classes, respectively. Task classes include adhesiolysis and liver cyst aspiration. Undefined class-specific values were set to zero. Action-level performance is reported as mAP for I, V, T, IV, IT, and full triplets (IVT). Classes containing null_verb and/or null_target were excluded from mAP_IVT averaging. Action results are for CurConMix with a Swin Transformer-Base backbone and 224 × 224 inputs; step and task predictions used TeCNO decoders with features from the frozen CurConMix student encoder. mAP: Mean average precision; I: instruments; V: verbs; T: targets; IV: instrument-verb pairs; IT: instrument-target pairs; IVT: < instrument, verb, target>.

full triplet combinations. Results from the additional CurConMix analysis and the previously evaluated comparison methods are summarized in Supplementary Table 4. Frequency-stratified mAP_IVT was 31.1% ± 4.5% for non-null triplet classes observed in fewer than 50 annotated frames and 59.9% ± 3.8% for those observed in at least 50 frames [Supplementary Table 5].

DISCUSSION

This study presents a SAGES-aligned step-task-action characterization of laparoscopic LLS, integrating step- and task-level labels with frame-level action triplets across 48 procedures. The main findings are that LLS is dominated by parenchymal transection and post-transection hemostasis, that action-level activity comprises both a core repertoire of frequent maneuvers and a long tail of rare, context-specific actions, and that the resulting annotations support automated recognition in internal procedure-level cross-validation. These findings provide an initial quantitative description of operative content in a laparoscopic liver procedure widely used for training and benchmarking, and demonstrate that clinically interpretable workflow structure can be captured at multiple levels of granularity.

These findings have implications for how LLS is understood and taught. Based on recorded step labels, parenchymal transection accounted for 52.5% of annotated frames, rising to 63.6% when combined with post-transection hemostasis, emphasizing that the procedure is predominantly a parenchymal division and hemostasis operation rather than a simple mobilization procedure. Here, “transection-dominant” refers to the composition of sampled annotated frames, and the reported percentages should not be interpreted as proportions of operative time. These results support the view that transection and hemostasis should remain central components of LLS training, not only because of their technical difficulty[21], but also because they

constitute the majority of sampled operative activity. Grouping tasks by their primary clinical association with S3 highlights multiple technical objectives, including parenchymal division, inflow and outflow control, the Pringle maneuver, and demarcation, revealing a composite structure not apparent from conventional operative reports. Adhesiolysis as a step-independent event further illustrates that standardized procedures incorporate situational adaptations that a useful annotation framework must accommodate. At the action level, frequent triplets may define core competencies, whereas more than half of the 195 triplets occurred fewer than 50 times and represented context-specific maneuvers. This long-tail distribution is a recognized challenge in surgical action recognition, limiting both learning efficiency and model generalization[20]. It may also reflect why laparoscopic liver resection has a prolonged learning curve[22,23], as mastery requires fluency in common maneuvers and exposure to infrequent situations. Although the present study did not evaluate training outcomes, annotated video may support structured review of both common and rare maneuvers.

These findings also extend surgical workflow analysis beyond its current methodological base. The field has been developed predominantly in laparoscopic cholecystectomy, supported by public datasets such as Cholec80 and CholecT45, and generalization to longer, more complex operations has remained uncertain[8,10,11,16,20,24]. Liu *et al.* established a multilevel framework for robotic LLS using instrument-action-object triplets to describe 26 activity categories, with annotation focused on principal instrument activities[12]. Building on this work, the present study integrates a procedure-specific hierarchy of five steps and 18 tasks with simultaneous multi-label IVT annotation of surgeon and assistant actions in laparoscopic LLS. This integration captures co-occurring technical actions within their procedural context and supports both workflow composition analysis and automated recognition across all three levels. The annotation inventory comprises 195 observed triplet classes, with separate component- and combination-level evaluation. The studies also differ in temporal coverage: Liu *et al.* used temporal annotation across recorded endoscopic videos, whereas LLS48 uses interval-based frame sampling[12]. These differences in surgical setting, annotation scope, sampling, and evaluation procedures preclude direct comparison of recognition performance.

LLS was selected because it is an index procedure for surgeons learning laparoscopic hepatectomy and is comparatively standardized within liver surgery[1-3]. This standardization is useful for initial framework development because step, task, and action definitions can be established consistently. However, the findings should not be assumed to generalize directly to major hepatectomy, repeat liver resection, or procedures with difficult anatomy or severe adhesions, where workflow variability, step-boundary ambiguity, and rare action triplets may be greater. LLS should therefore be viewed as a starting point for hierarchical workflow modeling before extension to more variable hepatobiliary procedures.

The automated recognition results provide a complementary perspective. Step-level accuracy of 82.3% suggests that the five procedural steps correspond to visually distinguishable operative states in internal cross-validation. Although slightly lower than phase recognition accuracies commonly exceeding 90% in laparoscopic cholecystectomy[10], this may reflect less discrete step boundaries in LLS, where multiple surgical objectives can occur within one step. The lower task-level accuracy (62.6%) may reflect visual and semantic overlap among tasks within the same step, particularly parenchymal transection and inflow/outflow control. These tasks may use similar instruments on adjacent structures and differ mainly in technical objectives. Interval-based sampling may further limit the temporal context available for distinguishing these tasks. Denser annotation of continuous videos and models incorporating broader procedural context may improve task discrimination and warrant future evaluation. The performance gradient across action components, from instruments to verbs, targets, and full triplets, parallels surgical learning, in which instrument recognition is simpler than interpreting what is being done, to which structure, and why. These findings support the feasibility of hierarchical modeling, but external validation and continuous-video evaluation are

needed before real-time assistance, institutional benchmarking, or clinical deployment. The results in Supplementary Table 4 are descriptive because checkpoint-selection conditions differed among methods. In particular, the retained SelfD-B and TERL-B results used test-fold performance for checkpoint selection, which may yield optimistic estimates. These results therefore do not establish the relative superiority of any model.

Recognition of infrequent triplets may benefit from targeted collection of rare actions, augmentation that preserves action labels, and class-balanced sampling or loss weighting. SelfD and contrastive learning focused on rare classes offer further strategies for addressing imbalance[19,20]. CurConMix already incorporates contrastive pretraining, SelfD, and mixup[17], but its contribution to rare-triplet recognition in LLS48 was not assessed separately. Future studies should evaluate these strategies through ablation experiments and report performance stratified by class frequency.

For future scalability, step-level recognition may support selective retrieval of operative phases from video archives and comparison of workflow composition across cases. More broadly, automated annotation could extend this framework from small manually annotated datasets to larger cohorts, although recognition performance and external validation must improve for reliable multi-center deployment. While benchmark studies quantify surgical outcomes at the population level, this approach may help characterize the intraoperative processes underlying those outcomes. However, the present study did not establish associations between workflow patterns and clinical outcomes, technical performance, or learning curves; future studies should prospectively evaluate whether workflow features are associated with perioperative outcomes, operative efficiency, surgeon experience, and quality indicators[25,26]. Surgical education, inter-institutional benchmarking, outcome-linked analysis, and clinical deployment remain potential future applications of the framework and were not validated in this study.

Recent perspectives on artificial intelligence (AI) agents and automated knowledge extraction highlight the value of multimodal integration, structured representations, and explicit validation[27,28]. Our step-task labels and IVT triplets could provide a structured representation for future multimodal agents linking operative video with clinical records and procedural knowledge, potentially supporting workflow interpretation and clinical decision support. Such integration was not evaluated in this study and would require validation on continuous surgical recordings, external evaluation, and clinician oversight.

Several limitations should be acknowledged. First, this was a single-center retrospective study of 48 procedures performed by four experienced hepatobiliary surgeons, without external validation; workflow patterns may differ across institutions, devices, patient populations, and operative techniques. Multicenter external validation is therefore needed before clinical use[24]. Future institution-specific deployment could use local video processing within secure institutional infrastructure to reduce external transfer of sensitive surgical data, together with controlled access, local performance validation, and ongoing monitoring[29]. Second, because LLS is a relatively standardized index procedure, generalizability to more complex hepatobiliary procedures remains uncertain and requires separate evaluation. Third, the study was not designed to link workflow features with surgical performance, perioperative outcomes, complications, learning curves, efficiency, or quality indicators, and the low event rate, including the absence of major postoperative complications and 30-day mortality, precluded meaningful outcome-linked analysis. Fourth, interval-based clip sampling may miss brief actions and transitions and cannot reconstruct continuous workflow dynamics or exact step durations. The reported proportions therefore describe sampled workflow composition rather than direct estimates of operative time. Fifth, the original consensus labels were visible during the additional review, which may have introduced anchoring and influenced agreement estimates. Reproducibility under blinded *de novo* annotation therefore remains to be established. Although the overall

review covered 10.0% of annotated frames, the inter-reviewer comparison involved 192 shared frames from 10 surgical cases, limiting generalizability and the precision of cluster-bootstrap CIs. The additional review did not establish step- or task-level inter-rater reliability. Sixth, surgeon identifiers were not consistently available, precluding surgeon-stratified analysis. Finally, rare action triplets remain underrepresented, limiting recognition of infrequent but potentially important maneuvers; multi-center data pooling and continuous-video annotation may help address this imbalance.

Conclusion

Based on sampled video annotations, laparoscopic LLS can be represented as a hierarchical, quantifiable workflow dominated by parenchymal transection and hemostasis, with substantial technical diversity at the action level. Step-task-action annotation provides a structured basis for describing operative content, and automated recognition supports the feasibility of computational workflow analysis, although multi-center validation, continuous-video evaluation, blinded annotation-reproducibility assessment, and outcome-linked studies are required before broader application.

DECLARATIONS

Authors' contributions

Contributed to the conception and design of the study: Park HS, Jeon Y, Oh N, Choi GS
Contributed to data acquisition and surgical video curation: Park HS, Shin J, Jung KH, Rhu J, Kim J, Choi GS, Oh N
Contributed to workflow framework development and annotation: Park HS, Jeon Y, Shin J, Jung KH, Oh N
Contributed to model development and computational analysis: Jeon Y, Park K
Contributed to data interpretation and clinical interpretation of the findings: Park HS, Jeon Y, Oh N, Rhu J, Kim J, Choi GS
Drafted the manuscript: Park HS, Jeon Y
Critically revised the manuscript for important intellectual content: Oh N, Park K, Shin J, Jung KH, Rhu J, Kim J, Choi GS
Supervised the study: Oh N, Choi GS
All authors read and approved the final manuscript.

Availability of data and materials

The annotation ontology and label scheme for LLS48 are described in the Methods and Supplementary Tables 1-3 and are provided as part of this publication. Public release of the surgical videos, other underlying study data, and analysis code is planned, subject to institutional and ethical approvals, including IRB review where applicable. Requests for access may be directed to the corresponding author and will be considered under the same approval requirements.

AI and AI-assisted tools statement

Not applicable.

Financial support and sponsorship

This research was supported by grants from the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea, including the Korean ARPA-H Project (Grant No. RS-2025-25424639), the AURORA Lab Project (Grant No. RS-2026-25503221), and the FLAME-ARK Project (Grant No. RS-2025-02223382).

The funding sources had no role in the study design, data collection, data analysis, data interpretation, writing of the manuscript, or the decision to submit the article for publication. The authors had full access to all data and had final responsibility for the decision to submit the manuscript.

Conflicts of interest

All authors declared that there are no conflicts of interest.

Ethical approval and consent to participate

The study was approved by the Institutional Review Board of Samsung Medical Center (IRB No. SMC-2023-04-128-001). The requirement for individual informed consent was waived owing to the retrospective nature of the study and the use of de-identified surgical video data.

Consent for publication

Not applicable.

Copyright

© The Author(s) 2026.

Supplementary Materials

Supplementary Materials

REFERENCES

1. Coelho FF, Kruger JAP, Fonseca GM, et al. Laparoscopic liver resection: experience based guidelines. *World J Gastrointest Surg.* 2016;8:5-26. DOI PubMed PMC
2. Ciria R, Cherqui D, Geller DA, Briceno J, Wakabayashi G. Comparative short-term benefits of laparoscopic liver resection: 9000 cases and climbing. *Ann Surg.* 2016;263:761-77. DOI PubMed
3. Hasegawa Y, Nitta H, Sasaki A, et al. Laparoscopic left lateral sectionectomy as a training procedure for surgeons learning laparoscopic hepatectomy. *J Hepatobiliary Pancreat Sci.* 2013;20:525-30. DOI PubMed
4. Blum T, Feusner H, Navab N. Modeling and segmentation of surgical workflow from laparoscopic video. In: Jiang T, Navab N, Pluim JPW, Viergever MA, Editors. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2010. Berlin, Heidelberg: Springer; 2010. pp. 400-7. DOI PubMed
5. Lo BPL, Darzi A, Yang G. Episode classification for the analysis of tissue/instrument interaction with multiple visual cues. In: Ellis RE, Peters TM, Editors. Medical Image Computing and Computer-Assisted Intervention - MICCAI 2003. Berlin, Heidelberg: Springer; 2003. pp. 230-7. DOI
6. Padoy N, Blum T, Ahmadi SA, Feussner H, Berger MO, Navab N. Statistical modeling and recognition of surgical workflow. *Med Image Anal.* 2012;16:632-41. DOI PubMed
7. Meireles OR, Rosman G, Altieri MS, et al. ; the SAGES Video Annotation for AI Working Groups. SAGES consensus recommendations on an annotation framework for surgical video. *Surg Endosc.* 2021;35:4918-29. DOI PubMed
8. Twinanda AP, Shehata S, Mutter D, Marescaux J, de Mathelin M, Padoy N. EndoNet: a deep architecture for recognition tasks on laparoscopic videos. *IEEE Trans Med Imaging.* 2017;36:86-97. DOI PubMed
9. Zisimopoulos O, Flouty E, Luengo I, et al. DeepPhase: surgical phase recognition in CATARACTS videos. In: Frangi AF, Schnabel JA, Davatzikos C, Alberola-López C, Fichtinger G, Editors. Medical Image Computing and Computer Assisted Intervention - MICCAI 2018. Cham: Springer; 2018. pp. 265-72. DOI
10. Yi F, Jiang T. Hard frame detection and online mapping for surgical phase recognition. In: Shen D, Liu T, Peters TM, Staib LH, Essert C, Zhou S, Yap P, Khan A, Editors. Medical Image Computing and Computer Assisted Intervention - MICCAI 2019. Cham: Springer; 2019. pp. 449-57. DOI
11. Nwoye CI, Gonzalez C, Yu T, et al. Recognition of instrument-tissue interactions in endoscopic videos via action triplets. In: Martel AL, Abolmaesumi P, Stoyanov D, et al., Editors. Medical Image Computing and Computer Assisted Intervention - MICCAI 2020. Cham: Springer; 2020. pp. 364-74. DOI
12. Liu Y, Zhao S, Zhang G, et al. Multilevel effective surgical workflow recognition in robotic left lateral sectionectomy with deep learning: experimental research. *Int J Surg.* 2023;109:2941-52. DOI PubMed PMC
13. Sasaki K, Ito M, Kobayashi S, et al. Automated surgical workflow identification by artificial intelligence in laparoscopic hepatectomy: experimental research. *Int J Surg.* 2022;105:106856. DOI PubMed
14. Goh BKP, Han HS, Chen KH, et al. ; International Robotic and Laparoscopic Liver Resection Study Group Investigators. Defining global benchmarks for laparoscopic liver resections: an international multicenter study. *Ann Surg.* 2023;277:e839-48. DOI PubMed
15. Tejani AS, Klontzas ME, Gatti AA, et al. ; CLAIM 2024 Update Panel. Checklist for artificial intelligence in medical imaging (CLAIM): 2024 update. *Radiol Artif Intell.* 2024;6:e240300. DOI PubMed PMC
16. Nwoye CI, Yu T, Gonzalez C, et al. Rendezvous: attention mechanisms for the recognition of surgical action triplets in endoscopic videos. *Med Image Anal.* 2022;78:102433. DOI PubMed

17. Jeon Y, Shin J, Park S, et al. CurConMix: a curriculum contrastive learning framework for enhancing surgical action triplet recognition. In: Gee JC, Alexander DC, Hong J, et al., Editors. Medical Image Computing and Computer Assisted Intervention – MICCAI 2025. Cham: Springer Nature Switzerland; 2025. pp. 149-58. DOI
18. Czempiel T, Paschali M, Keicher M, et al. TeCNO: Surgical Phase Recognition with Multi-stage Temporal Convolutional Networks. In: Martel AL, Abolmaesumi P, Stoyanov D, et al., Editors. Medical Image Computing and Computer Assisted Intervention - MICCAI 2020. Cham: Springer; 2020. pp. 343-52. DOI
19. Yamlahi A, Tran TN, Godau P, et al. Self-distillation for surgical action recognition. In: Greenspan H, Madabhushi A, Mousavi P, et al., Editors. Medical Image Computing and Computer Assisted Intervention - MICCAI 2023. Cham: Springer; 2023. pp. 637-46. DOI
20. Gui S, Wang Z. Tail-enhanced representation learning for surgical triplet recognition. In: Linguraru MG, Dou Q, Feragen A, et al., Editors. Medical Image Computing and Computer Assisted Intervention - MICCAI 2024. Cham: Springer; 2024. pp. 689-99. DOI
21. Krenzien F, Schöning W, Brunnbauer P, et al. The ILLS laparoscopic liver surgery fellow skills curriculum. *Ann Surg.* 2020;272:786-92. DOI PubMed
22. Kluger MD, Vigano L, Barroso R, Cherqui D. The learning curve in laparoscopic major liver resection. *J Hepatobiliary Pancreat Sci.* 2013;20:131-6. DOI PubMed
23. Vigano L, Laurent A, Tayar C, Tomatis M, Ponti A, Cherqui D. The learning curve in laparoscopic liver resection: improved feasibility and reproducibility. *Ann Surg.* 2009;250:772-82. DOI PubMed
24. Lavanchy JL, Ramesh S, Dall'Alba D, et al. Challenges in multi-centric generalization: phase and step recognition in Roux-en-Y gastric bypass surgery. *Int J Comput Assist Radiol Surg.* 2024;19:2249-57. DOI PubMed PMC
25. Balvardi S, Kammili A, Hanson M, et al. The association between video-based assessment of intraoperative technical performance and patient outcomes: a systematic review. *Surg Endosc.* 2022;36:7938-48. DOI PubMed
26. Aklilu JG, Sun MW, Goel S, et al. Artificial intelligence identifies factors associated with blood loss and surgical experience in cholecystectomy. *NEJM AI.* 2024;1:AIoa2300088. DOI PubMed
27. Xie M, Shen C, Zhu R, Lv C. Computational Nutrition 2.0: an algorithmic frontier for AI agents. *AI Agent.* 2026;2:19. DOI
28. Chen W, Xu H, Cheng D. From unstructured to structured: opportunities, risks and engineering practices for LLM-based data extraction in chemistry and materials. *AI Agent.* 2026;2:12. DOI
29. Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat Mach Intell.* 2020;2:305-11. DOI

Disclaimer/Publisher's Note: All statements, opinions, and data contained in this publication are solely those of the individual author(s) and contributor(s) and do not necessarily reflect those of OAE and/or the editor(s). OAE and/or the editor(s) disclaim any responsibility for harm to persons or property resulting from the use of any ideas, methods, instructions, or products mentioned in the content.

© The Author(s) 2026. Open Access This article is licensed under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, sharing, adaptation, distribution and reproduction in any medium or format, for any purpose, even commercially, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.