Automated Video-Based Assessment of Surgical Meta-competencies Using Computer Vision

*Joshua A. Villarreal, M.D.[1]; *Josiah Aklilu, B.S.[2]; Xiaohan Wang, Ph.D.[2]; Charlotte Egeland, M.D., Ph.D.[3]; Anita Rau, Ph.D.[2]; Elaine Sui, M.S.[4]; Vanessa Palter, M.D., Ph.D.[5]; Eric Sorenson, M.D.[6]; Roger Bohn, Ph.D.[7]; Teodor Grantcharov, M.D.,Ph.D.[2]; Jeffrey K. Jopling, M.D.[8] ; Serena Yeung-Levy, Ph.D.[2,4] (*Co-first authors)

[1]Department of Surgery, Stanford University; [2]Department of Biomedical Data Science, Stanford University; [3]Department of Transplantation and Digestive Diseases, Rigshospitalet; [4]Department of Computer Science, Stanford University; [5]Surgical Safety Technologies; [6]Department of Surgery, Intermountain Health; [7]University of California San Diego; [8]Department of Surgery, Johns Hopkins School of Medicine

Background

Traditional assessments of operative performance based on video review relies on direct human observation and feedback, which poses a logistical challenge in busy training programs. Prior studies have demonstrated the promise of AI in assessing surgical skill but fall short in facilitating actionable feedback to improve technical skill in laparoscopy[1,2,3]. Our study aims to digitize a video-based assessment (VBA) by developing computer vision algorithms aligned with validated surgical meta-competency (MC) skill domains[4], offering a streamlined approach to assess skill and facilitate feedback.

Methods

Multi-institutional retrospective study included a video dataset of 319 IRB-approved deidentified laparoscopic cholecystectomies (LC). We used the SAGES annotations for AI framework[5], and internally validated a VBA rubric for LC based on validated universal MC for surgical skills[4]. Ratings covered five MC domains, namely tissue handling (TH), psychomotor skills (PS), efficiency (Eff), dissection quality (DQ), and exposure quality (EQ). Meta-competency ratings were assigned on a 5-point likert scale. A hierarchical task analysis for LC was conducted to clearly define the annotations of four intracorporeal steps for MC review. We leveraged the visual features of a robust vision foundation model (DINOv2)[6] and temporal convolutional neural network to differentiate between a binarized composite of low (1-4) and high (5) levels of operative performance according to the labeled meta-competency domains.

Preliminary Results

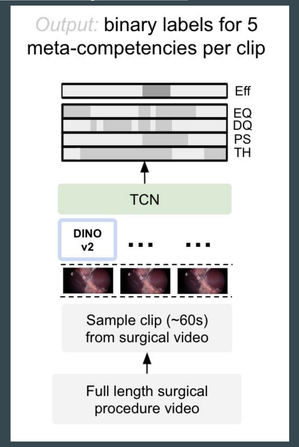


Figure 1: Computer vision model schema. *Abrev: Eff, efficiency; EQ, exposure quality; DQ, dissection quality; PS, psychomotor skills; TH, tissue handling; TCN, temporal convolution network*

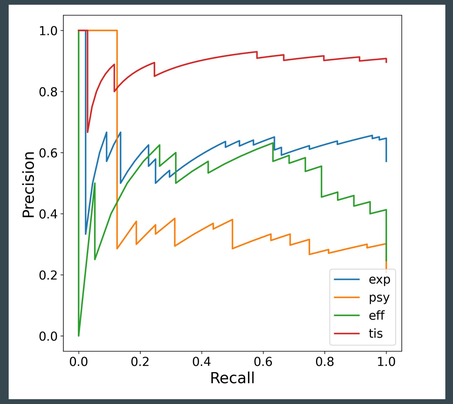


Figure 3: Precision-Recall curve for the Ligation & Division step.

Labeled Datasets	No. clips (Unique proc.)	Step in Cholecystectomy	No. clips (%)
Dataset 1	759 (139)	Exposing Gallbladder	406 (14)
Dataset 2	620 (77)	Dissection of Hepatocystic Triangle	1584 (55)
Dataset 3	649 (80)	Ligation and Division	384 (14)
Dataset 4	210 (23)	Gallbladder Removal	488 (17)
		Total clips	*2862*

Table 1: Meta-competency laparoscopic video dataset characteristics

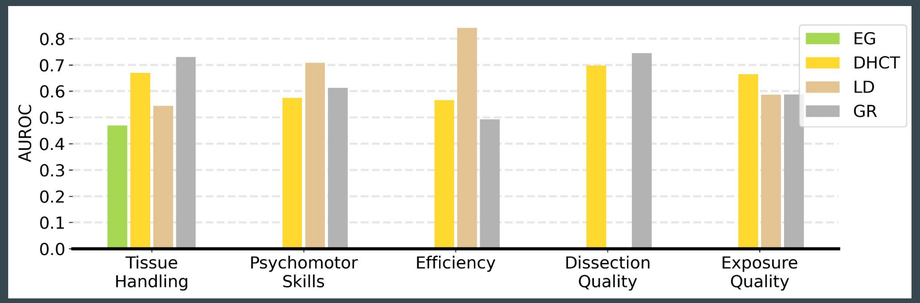


Figure 2: Computer vision performance metrics across 15 MC classifier models. *EG , Exposing Gallbladder; DHCT, Dissection of Hepatocystic Triangle; LD, Ligation and Division; GR, Gallbladder Removal. Note: LD step does not have ratings for dissection quality per VBA rubric.*



Figure 4: Qualitative example of model output for artificial surgical video review using a sliding window approach to identifying skill level.

Conclusions

This study is the first to utilize holistic, vision-only features of surgical video clips to predict skill MC. We have shown that automating VBA scores using computer vision is feasible. Shorter duration video clips enhanced annotation consistency and model performance compared to traditional step or procedural level assessments. These automated assessment could enable more meaningful feedback at both step and operation levels, supporting objective and scalable performance assessments in surgical training.

In future work, we plan to undertake a prospective pilot study having surgical trainees interact with these surgical performance models to provide user feedback and examine impact on technical skill acquisition.

References

1. Ryder CY, Mott NM, Gross CL, Anidi C, Shigut L, Bidwell SS, et al. Using Artificial Intelligence to Gauge Competency on a Novel Laparoscopic Training System. J Surg Educ. 2024 Feb;81(2):267–74.
2. Azari DP, Frasier LL, Quamme SRP, Greenberg CC, Pugh CM, Greenberg JA, et al. Modeling Surgical Technical Skill Using Expert Assessment for Automated Computer Rating. Annals of Surgery. 2019 Mar;269(3):574–81.
3. Igaki T, Kitaguchi D, Matsuzaki H, Nakajima K, Kojima S, Hasegawa H, et al. Automatic Surgical Skill Assessment System Based on Concordance of Standardized Surgical Field Development Using Artificial Intelligence. JAMA Surg. 2023 Jun 7;e231131.
4. Igaki T, Takenaka S, Watanabe Y, Kojima S, Nakajima K, Takabe Y, et al. Universal meta-competencies of operative performances: a literature review and qualitative synthesis. Surg Endosc. 2023 Feb;37(2):835–45.
5. Meireles OR, Rosman G, Altieri MS, Carin L, Hager G, Madani A, et al. SAGES consensus recommendations on an annotation framework for surgical video. Surg Endosc. 2021 Sep 1;35(9):4918–29.
6. Oquab M, Darcet T, Moutakanni T, Vo HV, Szafraniec M, Khalidov V, et al. DINOv2: Learning Robust Visual Features without Supervision. Transactions on Machine Learning Research. 2024