



Original Investigation | Surgery

Development and Validation of a 3-Dimensional Convolutional Neural Network for Automatic Surgical Skill Assessment Based on Spatiotemporal Video Analysis

Daichi Kitaguchi, MD; Nobuyoshi Takeshita, MD, PhD; Hiroki Matsuzaki, MS; Takahiro Igaki, MD; Hiro Hasegawa, MD; Masaaki Ito, MD, PhD

Abstract

IMPORTANCE A high level of surgical skill is essential to prevent intraoperative problems. One important aspect of surgical education is surgical skill assessment, with pertinent feedback facilitating efficient skill acquisition by novices.

OBJECTIVES To develop a 3-dimensional (3-D) convolutional neural network (CNN) model for automatic surgical skill assessment and to evaluate the performance of the model in classification tasks by using laparoscopic colorectal surgical videos.

DESIGN, SETTING, AND PARTICIPANTS This prognostic study used surgical videos acquired prior to 2017. In total, 650 laparoscopic colorectal surgical videos were provided for study purposes by the Japan Society for Endoscopic Surgery, and 74 were randomly extracted. Every video had highly reliable scores based on the Endoscopic Surgical Skill Qualification System (ESSQS, range 1-100, with higher scores indicating greater surgical skill) established by the society. Data were analyzed June to December 2020.

MAIN OUTCOMES AND MEASURES From the groups with scores less than the difference between the mean and 2 SDs, within the range spanning the mean and 1 SD, and greater than the sum of the mean and 2 SDs, 17, 26, and 31 videos, respectively, were randomly extracted. In total, 1480 video clips with a length of 40 seconds each were extracted for each surgical step (medial mobilization, lateral mobilization, inferior mesenteric artery transection, and mesorectal transection) and separated into 1184 training sets and 296 test sets. Automatic surgical skill classification was performed based on spatiotemporal video analysis using the fully automated 3-D CNN model, and classification accuracies and screening accuracies for the groups with scores less than the mean minus 2 SDs and greater than the mean plus 2 SDs were calculated.

RESULTS The mean (SD) ESSQS score of all 650 intraoperative videos was 66.2 (8.6) points and for the 74 videos used in the study, 67.6 (16.1) points. The proposed 3-D CNN model automatically classified video clips into groups with scores less than the mean minus 2 SDs, within 1 SD of the mean, and greater than the mean plus 2 SDs with a mean (SD) accuracy of 75.0% (6.3%). The highest accuracy was 83.8% for the inferior mesenteric artery transection. The model also screened for the group with scores less than the mean minus 2 SDs with 94.1% sensitivity and 96.5% specificity and for group with greater than the mean plus 2 SDs with 87.1% sensitivity and 86.0% specificity.

CONCLUSIONS AND RELEVANCE The results of this prognostic study showed that the proposed 3-D CNN model classified laparoscopic colorectal surgical videos with sufficient accuracy to be used for screening groups with scores greater than the mean plus 2 SDs and less than the mean minus 2 SDs. The proposed approach was fully automatic and easy to use for various types of surgery, and no

(continued)

Key Points

Question Is it possible to apply deep learning-based spatiotemporal video analysis using a 3-dimensional convolutional neural network to automate surgical skill assessment?

Findings In this prognostic study, 1480 video clips from 74 laparoscopic colorectal surgical videos with reliable skill assessment scores were extracted and assigned 4:1 to training and test sets. The proposed model automatically classified video clips into 3 different score groups with 75% accuracy.

Meaning A 3-dimensional convolutional neural network, which analyzes videos instead of static images, has the potential to be applied to automatic surgical skill assessment.

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

Open Access. This is an open access article distributed under the terms of the CC-BY-NC-ND License.

Abstract (continued)

special annotations or kinetics data extraction were required, indicating that this approach warrants further development for application to automatic surgical skill assessment.

JAMA Network Open. 2021;4(8):e2120786. doi:10.1001/jamanetworkopen.2021.20786

Introduction

A high level of surgical skill is essential to prevent intraoperative problems, such as bleeding or tissue injury, and a significant association between surgical skill and postoperative outcomes has been shown, especially in the field of laparoscopic gastrointestinal surgery.^{1,2} Therefore, novice surgeons must be comprehensively trained to increase patient safety and to avoid adverse surgical outcomes. One important aspect of surgical education is surgical skill assessment, in which pertinent feedback facilitates efficient skill acquisition by novices.

However, surgical skill assessment remains challenging. Currently, surgical training programs aim to objectively evaluate the basic surgical skills of trainees by using tools such as the Objective Structured Assessment of Technical Skills³ and the Global Operative Assessment of Laparoscopic Skills⁴; however, these tools rely on the observations and judgments of individuals, which are inevitably associated with subjectivity and bias. Furthermore, those methods require the time and resources of an expert surgeon or trained rater. Therefore, automating the process of surgical skill evaluation may be necessary, and thus far, automation approaches have shown promise. However, although numerous efforts have been made to develop automatic surgical skill assessments (ASSAs), most of these methods require specific types of kinetics data, such as tool motion tracking,⁵⁻⁷ hand motion tracking,^{8,9} eye motion tracking,^{10,11} and muscle contraction analysis,¹² and sensors or markers are required for data extraction.

Convolutional neural networks (CNNs) have become the dominant approach to solve problems in the field of computer vision. A CNN is a type of deep learning model with a remarkable ability to learn high-level concepts from training data, and it can mimic the human vision system. Convolutional neural networks have shown state-of-the-art performance in many computer vision tasks and have been gradually used in surgical fields, including in recognition tasks for steps, tools, and organs during laparoscopic surgery.¹³⁻¹⁶ Because CNN can automatically extract kinetics data during surgery without the use of any sensors or markers, it has also been applied to ASSA¹⁷⁻²¹; however, ASSA based on kinetics data requires a time-consuming annotation process to construct the training data set. In addition, because this ASSA method is applied mainly to target actions, such as suturing and knot tying, it is uncertain whether its results truly reflect practical surgical skills.

Recently, 3-dimensional (3-D) CNN models that capture both spatial and temporal dimensions have been proposed and used for human action recognition tasks on videos.^{22,23} The use of 3-D CNN enables the analysis of laparoscopic surgical videos as dynamic images instead of as static images; therefore, 3-D CNN can be applied to ASSA using the full information obtained from laparoscopic surgical videos. This assessment process is considered convincing because it also involves manual assessment based not only on kinetics data but also on full video information.

The objectives of the present study were to develop a 3-D CNN model for ASSA of laparoscopic colorectal surgery using intraoperative videos and to evaluate the automatic skill classification performance of the proposed model. To our knowledge, this is the first effort to develop ASSA for actual surgery using a 3-D CNN-based approach. We also evaluated whether the classification results could be used for automatically screening groups by mean scores less than -2 SDs and greater than 2 SDs.

Methods

This study followed the reporting guidelines of the Standards for Quality Improvement Reporting Excellence (SQUIRE) and the Standards for Reporting of Diagnostic Accuracy (STARD). The protocol for this study was reviewed and approved by the Ethics Committee of the National Cancer Center Hospital East. This study conforms to the provisions of the Declaration of Helsinki, 1964 (as revised in Brazil in 2013).²⁴ Informed consent was obtained from all participants in the form of an opt-out choice, in accordance with the Good Clinical Practice guidelines of the Ministry of Health and Welfare of Japan. No one received compensation or was offered any incentive for participating in this study.

Video Data Set

The video data set included 650 intraoperative videos of laparoscopic colorectal surgery (sigmoid-colon resection or high anterior resection) obtained from the Japan Society for Endoscopic Surgery (JSES). The JSES established an Endoscopic Surgical Skill Qualification System (ESSQS) and started examinations in 2004.²⁵ Since then, nonedited videos have been submitted every year by candidates for examination, and the submissions are assessed by 2 judges in a double-blind manner, with strict criteria for the qualification of candidates. Therefore, every video in the data set had highly reliable scores for surgical skill assessment. In the ESSQS, evaluation criteria are divided into 2 categories: common criteria, which correspond to basic endoscopic techniques that are commonly used for all procedures, and organ-specific criteria, which correspond to special endoscopic surgical techniques for individual organs. These criteria are allotted 60 and 40 points, respectively. The evaluation is focused on surgical techniques and camera work, and a total score of 70 points is designated the minimum for qualification. The details of the common and procedure-specific criteria are given in eTable 1 and eTable 2 in the Supplement, respectively. The videos were submitted to the JSES from different parts of Japan between 2016 and 2017, and the JSES did not disclose any information on patients, surgeons, or intraoperative and postoperative outcomes.

Preprocessing

After manually eliminating scenes unrelated to the surgeon's skill, such as extracorporeal scenes and exposure scenes, the following 4 surgical steps were targeted in the present study: medial mobilization, lateral mobilization, inferior mesenteric artery (IMA) transection, and mesorectal transection. These steps are described in the Box. The learning process of the proposed 3-D CNN model is illustrated in Figure 1. Initially, each video was split into 8-second video fragments, each of which was further split into 80-frame consecutive static images and input into the model. Finally, 5 consecutive 8-second video fragments were aggregated and analyzed as 40-second video clips. Each static image was down-sampled from a resolution of 1280 × 720 pixels to a resolution of 224 × 224 pixels for preprocessing.

Because predicting the exact ESSQS score or a pass or fail classification was expected to be difficult even with a 3-D CNN, we focused on screening for groups with extremely low scores or with extremely high scores in this study. Based on ESSQS scores (range, 0-100 points), 650 intraoperative videos were divided into the following 3 score groups: scores less than the difference between the mean and 2 SDs (approximated as <50 points), scores within the range spanning the mean plus or minus the SD (approximated as 55-75 points), and scores greater than the sum of the mean and 2 SDs (approximated as >80 points). For 650 collected videos, the numbers of cases in the groups with scores less than the mean minus 2 SDs and scores greater than the mean plus 2 SDs were markedly fewer than in the group with scores within 1 SD of the mean. Because interclass imbalance can increase the difficulty of a classification task, we limited the number of videos used in this study to 74. Thus, from the groups with scores less than the mean minus 2 SDs, within 1 SD of the mean, and greater than the mean plus 2 SDs, 17, 26, and 31 videos were randomly extracted, respectively. Furthermore, from every step in every video, five 40-second video clips were randomly extracted. In total, 1480 video clips were used in this study.

Box. Descriptions of Surgical Steps in Laparoscopic Colorectal Surgery

- Medial mobilization:
 - Approach to mesocolon from medial side
 - Total mesorectal excision on right or posterior side
- Lateral mobilization:
 - Approach to mesocolon from lateral side
 - Total mesorectal excision on left side
- Inferior mesenteric artery transection:
 - Approach to inferior mesenteric artery root for dissection and transection
- Transection of mesorectum:
 - Approach to mesorectum for transection toward rectal transection line

Three-Dimensional CNN Model

We used a 3-D CNN model called the Inception-v1 I3D 2-stream model (RGB [red, green, and blue] plus optical flow)²³ for the automatic surgical skill classification task. The model was pretrained on the ImageNet data set²⁶ and the Kinetics data set.²⁷ ImageNet contains more than 14 million manually annotated images in more than 20 000 typical categories, such as balloon and strawberry, with several hundred images per category. Kinetics, one of the largest human-action video data sets available, consists of 400 action classes, with at least 400 video clips per class.

Computer Specification

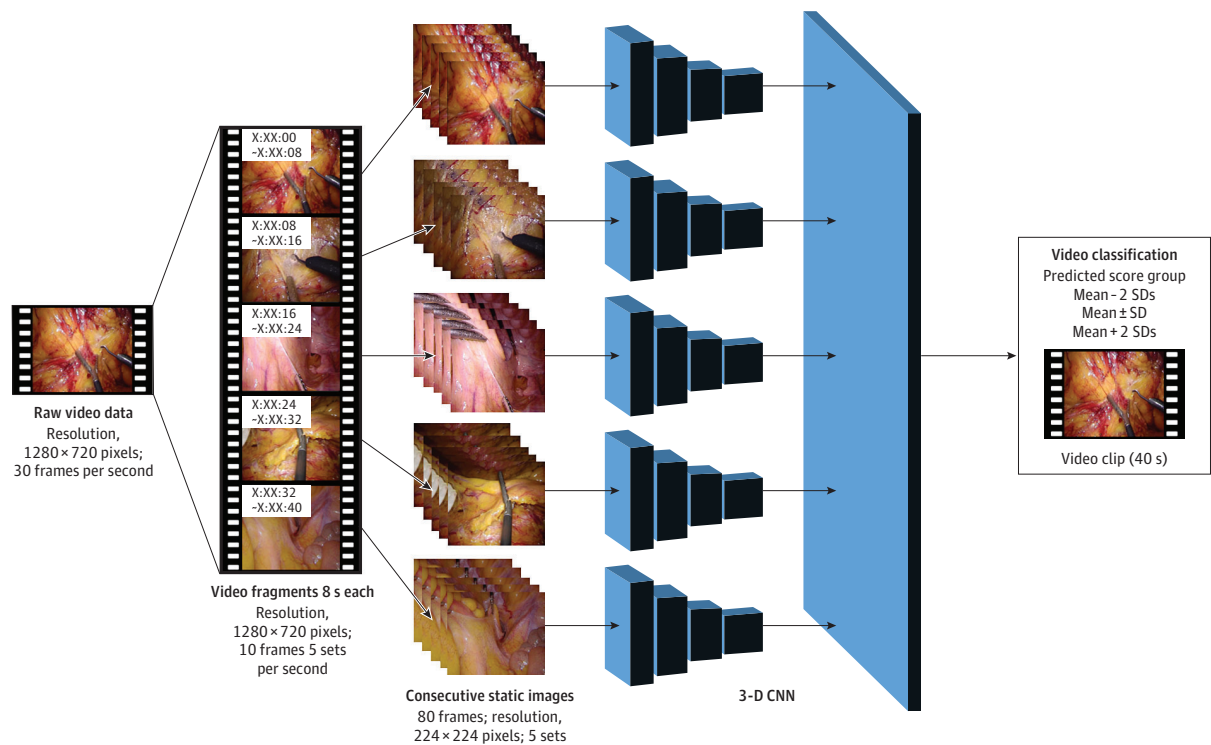
A computer equipped with an Nvidia Quadro GP 100 graphics processing unit with 16 GB of video RAM (Nvidia) and an Intel Xeon E5-1620 v4 central processing unit at 3.50 GHz with 32 GB of RAM was used for model training. All modeling procedures were performed using source code written in Python 3.6 (Python Software Foundation) based on an open-source surgical skill assessment package for 3-D CNN.²⁸

Evaluation

After model training, automatic surgical skill classification was performed for each video clip in the test set, and the output (ie, prediction) was any of the following labels: "<Ave – 2 SD score," "Ave ± SD score," and ">Ave + 2 SD score." Classification accuracy was calculated for each surgical step as (true positive + true negative) / (true positive + false positive + false negative + true negative).

To evaluate the screening performance of the proposed model on the group with scores less than the mean minus 2 SDs, sensitivity and specificity were calculated for each condition based on the number of "<Ave – 2 SD score" outputs. For example, if the threshold value was set at 2, the cases with a predicted output of "<Ave – 2 SD score" in more than 2 surgical steps were judged as

Figure 1. Learning Process of the Proposed 3-Dimensional Convolutional Neural Network (3-D CNN) Model



Mean – 2 SDs indicates the group with scores less than the mean minus 2 SDs; Mean ± SD, scores within 1 SD of the mean; and Mean + 2 SDs, scores greater than the mean plus 2 SDs.

belonging to the group with scores less than the mean minus 2 SDs. The area under the receiver operating characteristic (AUROC) curve was then calculated. Similarly, the sensitivity, specificity, and AUROC curve for screening the group with scores greater than the mean plus 2 SDs were also calculated.

Validation

The data set was divided into training and test data sets and validated using the leave-one-supertrial-out scheme.²⁹ In this validation scheme, 1 of every 5 video clips was extracted and included in the test data set, resulting in 1184 and 296 video clips in the training and test data sets, respectively. None of the video clips included in the training data set were included in the test data set, and to reduce the similarity, consecutive video clips had an interval of at least 20 seconds and were not continuous.

Statistical Analysis

All data and statistical analyses were conducted from June to December 2020. Analyses were performed using EZR, version 1.41 (Saitama Medical Center, Jichi Medical University), which is a graphical user interface for R (The R Foundation for Statistical Computing).

Results

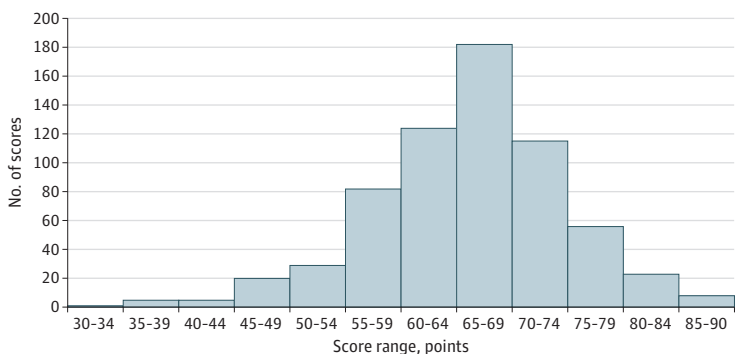
Surgical Skill Assessment Score

The mean (SD) ESSQS score of all 650 intraoperative videos was 66.2 (8.6) points, and the exact value was 48.9 points for the group with scores less than the mean minus 2 SDs, 57.6 to 74.9 points for the group with scores within 1 SD of the mean, and 83.5 points for the group with scores greater than the mean plus 2 SDs. The ESSQS score distribution for the entire cohort is shown in **Figure 2**. The mean (SD) scores were as follows: 67.6 (16.1) points for the 74 intraoperative videos used in this study, 42.8 (5.1) points for the 17 videos in the group with scores less than the mean minus 2 SDs, 65.1 (1.4) points for the 26 videos in the group with scores within 1 SD, and 83.3 (3.0) points for the 31 videos in the group with scores greater than the mean plus 2 SDs.

Automatic Surgical Skill Classification

The results of the 3-D CNN-based automatic surgical skill classification are shown in **Figure 3** as a confusion matrix. The mean (SD) accuracy of classification into the 3 different score groups was 75.0% (6.3%). Furthermore, the classification accuracies were 73.0% for medial mobilization, 74.3% for lateral mobilization, 83.8% for IMA transection, and 68.9% for mesorectal transection.

Figure 2. Endoscopic Surgical Skill Qualification System Score Distribution



Distribution of scores for intraoperative videos of laparoscopic colorectal surgery submitted to the Japan Society for Endoscopic Surgery from 2016 to 2017.

Automatic Screening for Score Groups Beyond 2 SDs

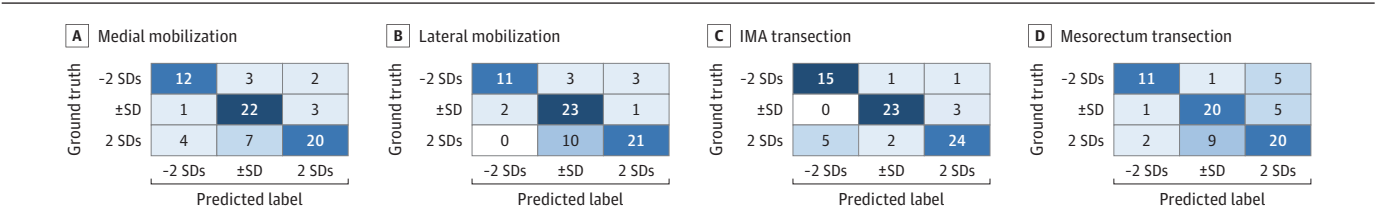
The receiver operating characteristic curves for the screening of the groups with scores less than the mean minus 2 SDs and greater than the mean plus 2 SDs are shown in **Figure 4**. For the group with scores less than the mean minus 2 SDs, the sensitivity of the screening was 94.1% and specificity was 96.5% when the threshold value was 2 (ie, when the label “<Ave – 2 SD score” was output for 2 or more surgical steps), and the intraoperative video was judged as belonging to the group with scores with less than the mean minus 2 SDs. The sensitivity and specificity were 100% and 77.2%, respectively, when the threshold value was 1; 76.5% and 100% when the threshold value was 3; and 17.6% and 100% when the threshold value was 4. The AUROC for the screening of the group with scores less than the mean minus 2 SDs was 0.989.

Similarly, for the group with scores greater than the mean plus 2 SDs, the screening sensitivity was 87.1% and the screening specificity was 86.0% when the threshold value was 2 (ie, when the label “>Ave + 2 SD score” was output for 2 or more surgical steps), and the intraoperative video was judged as belonging to the group with scores greater than the mean plus 2 SDs. The sensitivity and specificity were 96.8% and 60.5%, respectively, when the threshold value was 1; 61.3% and 100% when the threshold value was 3; and 29.0% and 100% when the threshold value was 4. The AUROC for the screening of the group with scores greater than the mean plus 2 SDs was 0.934.

Discussion

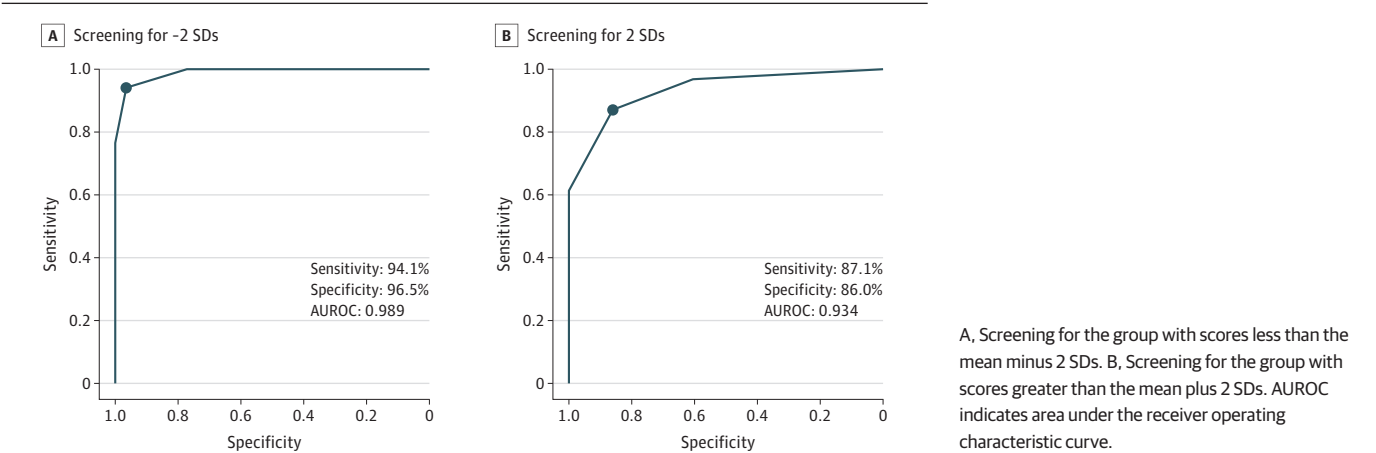
In this study, we successfully developed a 3-D CNN-based automatic surgical skill classification model and showed that the proposed model automatically classified intraoperative videos into groups with scores less than the mean minus 2 SDs, within 1 SD of the mean, and greater than the mean plus 2 SDs with a mean (SD) accuracy of 75.0% (6.3%). The highest accuracy was 83.8%, which was for IMA

Figure 3. Confusion Matrix of Results of 3-Dimensional Convolutional Neural Network-Based Automatic Surgical Skill Classification



-2 SDs indicates the group with scores less than the mean minus 2 SDs; ±SD, scores within 1 SD of the mean; 2 SDs, scores greater than the mean plus 2 SDs; IMA, inferior mesenteric artery.

Figure 4. Receiver Operating Characteristic Curves for Screening



A, Screening for the group with scores less than the mean minus 2 SDs. B, Screening for the group with scores greater than the mean plus 2 SDs. AUROC indicates area under the receiver operating characteristic curve.

transection. Furthermore, based on the classification results, the proposed model could screen for the group with scores less than the mean minus 2 SDs with 94.1% sensitivity and 96.5% specificity, and for the group with scores greater than the mean plus 2 SD with 87.1% sensitivity and 86.0% specificity. These results suggest that 3-D CNN may be applied to not only human-action recognition but also ASSA.

Numerous studies have thus far been conducted in the field of computer vision on human-action recognition using video data sets, and state-of-the-art accuracy is being updated with new studies reported frequently³⁰⁻³³; however, these technologies have not been widely applied to the field of surgery. To our knowledge, this is the first study to use 3-D CNN for actual laparoscopic surgical skill assessment.

The most significant feature of the proposed approach using a 3-D CNN is that intraoperative videos may be analyzed as dynamic images instead of static images. In most previous studies, because information obtained from static images was limited, kinematic features of tracked hand or instrument movements had to be extracted to increase the available information.^{17,18} However, when we actually evaluate surgical skill, we do not necessarily observe only the surgeon's hand or instrument. The ASSA approaches based only on video data, without using kinetics data, have also been studied in recent years.^{20,28,33} Those approaches used 2-D CNN and attempted to extract not only spatial features but also temporal features. However, in those approaches, each analysis was performed on static images, and surgical skill assessment based on static image analysis is considered difficult. In addition, those studies were performed on video data sets consisting only of basic surgical tasks, including suturing, needle passing, and knot tying; therefore, it is still unclear whether those approaches could be applied to actual surgical skill assessment. By contrast, the proposed approach for ASSA based on 3-D CNN considers information obtained from both entire images and surgical flow. Spatiotemporal video analysis is quite similar to the manual skill assessment process and is considered to be convincing as an ASSA approach.

The results of the present study suggest that ASSA using 3-D CNN is suitable for the IMA transection surgical step. In left colectomy and rectal surgery, IMA root dissection and transection procedures are crucial because they may cause major bleeding; in addition, these procedures may affect curability in lymph-node dissection. Because many tiny vessels are present around the IMA root, oozing occurs easily with rough procedures; consequently, the surgical area tends to become wet and reddish. Because meticulous procedures are required in this surgical step, it is considered to truly reflect a surgeon's skill, specifically in terms of dexterity and awkwardness. However, any hypotheses about the suitability of ASSA using 3-D CNN for the IMA transection surgical step should be verified in future studies.

In the field of deep learning-based computer vision, technology innovations have been reported frequently, and new CNN models are actively being developed. Therefore, further improvements in the accuracy and generalizability can be expected with advances in deep learning-based computer vision. In future studies, by using videos with score information for each evaluation item (eg, respect for tissue, time and motion, and instrument handling) and by performing classification tasks for each evaluation item, surgeons may obtain classification results for each evaluation item, which may lead to constructive feedback and to the further development of the ASSA.

Strengths and Limitations

As mentioned previously, although the classification accuracy is associated with the characteristics of the surgical step, one of the strengths of the proposed ASSA method is that it can eliminate time-consuming annotation processes. In most research on ASSA using computer vision, the areas or tips of surgical instruments have to be annotated to extract the kinetics data of surgical instruments.^{17,18} Such methods could be significantly affected by the recognition accuracy of surgical instruments. By contrast, the proposed approach may be applied as long as intraoperative videos have information about surgeons' experience or reliable scores. Furthermore, the proposed approach may easily be

applied to other types of surgery, regardless of the surgical field or type of instruments used during the surgery. Since 2004, the JSES has used video examinations based on ESSQS,²⁵ and only 20% to 30% of ESSQS examinees pass each year. Currently, less than 10% of general surgeons in Japan are ESSQS-certified.^{2,34} With regard to the reliability of the scores used in this study, the video examination is based on anonymized unedited random video reviews and scoring by 2 or 3 expert laparoscopic surgeons designated by the JSES; therefore, the scores are considered highly reliable.

This study has a few limitations. First, the validation scheme leave-one-supertrial-out was the method used to evaluate the performance of the deep learning model for ASSA. To reduce the similarity between video clips, consecutive clips had an interval of at least 20 seconds and were not continuous. However, video clips extracted from the same surgery were included in both the training and test data sets. Therefore, the generalizability of the proposed model for unseen surgical procedures and surgeons could not be reliably established in this study. To show generalizability to a real-world setting, further validation is needed using a data set consisting of completely new intraoperative videos of surgical procedures performed by surgeons that the model was not trained on. Second, this is a proof-of-concept study, and practical application was not examined. With a classification accuracy of 75.0%, it would be unacceptable to use the proposed model for ASSA in a real-world setting. In the JSES, many expert surgeons are tasked with reviewing multiple surgical videos every year, which is a time-consuming and labor-intensive task. To reduce their burden, we envision the development of an automatic screening system for surgical skill as a feasible and valuable next step. Third, although the video data set used in this study contained multi-institutional intraoperative videos, all videos were collected from Japanese hospitals, and the ESSQS by the JSES is not widely used worldwide for surgical skill assessment. Further validation using an international video data set and other surgical skill assessment tools is desirable.

Conclusions

This study showed the feasibility of applying a deep learning-based 3-D CNN model to ASSA. In an automatic classification task categorizing ESSQS scores into 3 groups, the accuracy of the proposed model was 75.0%. Although the accuracy was insufficient for practical applications, the screening performance for groups with scores less than the mean minus 2 SDs and greater than the mean plus 2 SDs was quite high, and the model appeared effective in detecting very poor and good surgical skills. Further improvement is needed to consider the practical application of such a model.

The proposed approach is fully automatic and requires no special annotation, extraction of kinetics data, or sensors. Furthermore, it can be easily used in various types of surgery because its scalability and adaptability are quite high. Therefore, the proposed method warrants further development for future use in the field of surgical skill assessment.

ARTICLE INFORMATION

Accepted for Publication: May 29, 2021.

Published: August 13, 2021. doi:10.1001/jamanetworkopen.2021.20786

Open Access: This is an open access article distributed under the terms of the [CC-BY-NC-ND License](#). © 2021 Kitaguchi D et al. *JAMA Network Open*.

Corresponding Author: Masaaki Ito, MD, PhD, Surgical Device Innovation Office, National Cancer Center Hospital East, 6-5-1, Kashiwanoha, Kashiwa, Chiba 277-8577, Japan (maito@east.ncc.go.jp).

Author Affiliations: Surgical Device Innovation Office, National Cancer Center Hospital East, Kashiwanoha, Kashiwa, Chiba, Japan (Kitaguchi, Takeshita, Matsuzaki, Igaki, Hasegawa, Ito); Department of Colorectal Surgery, National Cancer Center Hospital East, Kashiwanoha, Kashiwa, Chiba, Japan (Kitaguchi, Takeshita, Igaki, Hasegawa, Ito).

Author Contributions: Drs Kitaguchi and Ito had full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Kitaguchi, Takeshita, Hasegawa, Ito.

Acquisition, analysis, or interpretation of data: Kitaguchi, Takeshita, Matsuzaki, Igaki, Ito.

Drafting of the manuscript: Kitaguchi, Igaki.

Critical revision of the manuscript for important intellectual content: Takeshita, Matsuzaki, Hasegawa, Ito.

Statistical analysis: Kitaguchi, Igaki, Ito.

Obtained funding: Takeshita.

Administrative, technical, or material support: Kitaguchi, Matsuzaki, Hasegawa, Ito.

Supervision: Takeshita, Hasegawa, Ito.

Conflict of Interest Disclosures: None reported.

Funding/Support: This study was supported by the Japan Society for Endoscopic Surgery.

Role of the Funder/Sponsor: The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

REFERENCES

1. Birkmeyer JD, Finks JF, O'Reilly A, et al; Michigan Bariatric Surgery Collaborative. Surgical skill and complication rates after bariatric surgery. *N Engl J Med*. 2013;369(15):1434-1442. doi:10.1056/NEJMsa1300625
2. Ichikawa N, Homma S, Funakoshi T, et al. Impact of technically qualified surgeons on laparoscopic colorectal resection outcomes: results of a propensity score-matching analysis. *BJS Open*. 2020;4(3):486-498. doi:10.1002/bjs5.50263
3. Martin JA, Regehr G, Reznick R, et al. Objective Structured Assessment of Technical Skill (OSATS) for surgical residents. *Br J Surg*. 1997;84(2):273-278.
4. Vassiliou MC, Feldman LS, Andrew CG, et al. A global assessment tool for evaluation of intraoperative laparoscopic skills. *Am J Surg*. 2005;190(1):107-113. doi:10.1016/j.amjsurg.2005.04.004
5. Estrada S, O'Malley M, Duran C, et al. On the development of objective metrics for surgical skills evaluation based on tool motion. In: *Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE; 2014:3144-3149.
6. Hofstad EF, Våpenstad C, Chmarra MK, Langø T, Kuhry E, Mårvik R. A study of psychomotor skills in minimally invasive surgery: what differentiates expert and nonexpert performance. *Surg Endosc*. 2013;27(3):854-863. doi:10.1007/s00464-012-2524-9
7. Brown JDO, O'Brien CE, Leung SC, Dumon KR, Lee DI, Kuchenbecker KJ. Using contact forces and robot arm accelerations to automatically rate surgeon skill at peg transfer. *IEEE Trans Biomed Eng*. 2017;64(9):2263-2275. doi:10.1109/TBME.2016.2634861
8. Zirkle M, Roberson DW, Leuwer R, Dubrowski A. Using a virtual reality temporal bone simulator to assess otolaryngology trainees. *Laryngoscope*. 2007;117(2):258-263. doi:10.1097/01.mlg.0000248246.09498.b4
9. Watson RA. Quantification of surgical technique using an inertial measurement unit. *Simul Healthc*. 2013;8(3):162-165. doi:10.1097/SIH.0b013e318277803a
10. Snaineh S, Seales B. Minimally invasive surgery skills assessment using multiple synchronized sensors. In: *Proceedings of the 2015 IEEE International Symposium on Signal Processing and Information Technology (ISSPIT)*. IEEE; 2016:314-319.
11. Ahmidi N, Hager GD, Ishii L, et al. Robotic path planning for surgeon skill evaluation in minimally-invasive sinus surgery. *Med Image Comput Comput Assist Interv*. 2012;15(pt 1):471-478. doi:10.1007/978-3-642-33415-3_58
12. Cavallo F, Pietrabissa A, Megali G, et al. Proficiency assessment of gesture analysis in laparoscopy by means of the surgeon's musculo-skeleton model. *Ann Surg*. 2012;255(2):394-398. doi:10.1097/SLA.0b013e318238350e
13. Hashimoto DA, Rosman G, Witkowski ER, et al. Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. *Ann Surg*. 2019;270(3):414-421. doi:10.1097/SLA.0000000000003460
14. Kitaguchi D, Takeshita N, Matsuzaki H, et al. Real-time automatic surgical phase recognition in laparoscopic sigmoidectomy using the convolutional neural network-based deep learning approach. *Surg Endosc*. 2020;34(11):4924-4931. doi:10.1007/s00464-019-07281-0
15. Kitaguchi D, Takeshita N, Matsuzaki H, et al. Automated laparoscopic colorectal surgery workflow recognition using artificial intelligence: experimental research. *Int J Surg*. 2020;79:88-94. doi:10.1016/j.ijsu.2020.05.015

16. Kitaguchi D, Takeshita N, Matsuzaki H, et al. Computer-assisted real-time automatic prostate segmentation during TaTME: a single-center feasibility study. *Surg Endosc*. 2021;35(6):2493-2499. doi:10.1007/s00464-020-07659-5
17. Law H, Ghani K, Deng J. Surgeon technical skill assessment using computer vision based analysis. In: *Proceedings of the 2nd Machine Learning for Healthcare Conference (PMLR)*. ML Research Press; 2017;68:88-99.
18. Azari DP, Frasier LL, Quamme SRP, et al. Modeling surgical technical skill using expert assessment for automated computer rating. *Ann Surg*. 2019;269(3):574-581. doi:10.1097/SLA.0000000000002478
19. Gao Y, Vedula SS, Lee GI, Lee MR, Khudanpur S, Hager GD. Query-by-example surgical activity detection. *Int J Comput Assist Radiol Surg*. 2016;11(6):987-996. doi:10.1007/s11548-016-1386-3
20. Tao L, Zappella L, Hager GD, et al. Surgical gesture segmentation and recognition. *Med Image Comput Comput Assist Interv*. 2013;16(pt 3):339-346. doi:10.1007/978-3-642-40760-4_43
21. Forestier G, Petitjean F, Senin P, et al. Surgical motion analysis using discriminative interpretable patterns. *Artif Intell Med*. 2018;91:3-11. doi:10.1016/j.artmed.2018.08.002
22. Ji S, Yang M, Yu K. 3D convolutional neural networks for human action recognition. *IEEE Trans Pattern Anal Mach Intell*. 2013;35(1):221-231. doi:10.1109/TPAMI.2012.59
23. Carreira J, Zisserman A. Quo vadis, action recognition? a new model and the kinetics dataset. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. CVPR; 2017:4724-4733.
24. World Medical Association. World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects. *JAMA*. 2013;310(20):2191-2194. doi:10.1001/jama.2013.281053
25. Mori T, Kimura T, Kitajima M. Skill accreditation system for laparoscopic gastroenterologic surgeons in Japan. *Minim Invasive Ther Allied Technol*. 2010;19(1):18-23. doi:10.3109/13645700903492969
26. Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis*. 2015;115:211-252. doi:10.1007/s11263-015-0816-y
27. Kay W, Carreira J, Simonyan K, et al. The kinetics human action video dataset. *arXiv*. 2017;1705.06950. Preprint posted online May 19, 2017. Accessed July 13, 2021. <https://arxiv.org/pdf/1705.06950.pdf>
28. Funke I, Mees ST, Weitz J, Speidel S. Video-based surgical skill assessment using 3D convolutional neural networks. *Int J Comput Assist Radiol Surg*. 2019;14(7):1217-1225. doi:10.1007/s11548-019-01995-1
29. Tu Z, Li H, Zhang D, Dauwels J, Li B, Yuan J. Action-stage emphasized spatio-temporal VLAD for video action recognition. *IEEE Trans Image Process*. 2019;28(6):2799-2812. doi:10.1109/TIP.2018.2890749
30. Xu B, Ye H, Zheng Y, Wang H, Luwang T, Jiang YG. Dense dilated network for video action recognition. *IEEE Trans Image Process*. 2019;28(10):4941-4953. doi:10.1109/TIP.2019.2917283
31. Liu Y, Lu Z, Li J, Yang T, Yao C. Deep image-to-video adaptation and fusion networks for action recognition. *IEEE Trans Image Process*. 2019;29:3168-3182. doi:10.1109/TIP.2019.2957930
32. Wei H, Jafari R, Kehtarnavaz N. Fusion of video and inertial sensing for deep learning-based human action recognition. *Sensors (Basel)*. 2019;19(17):3680. doi:10.3390/s19173680
33. Khalid S, Goldenberg M, Grantcharov T, Taati B, Rudzicz F. Evaluation of deep learning models for identifying surgical actions and measuring performance. *JAMA Netw Open*. 2020;3(3):e201664. doi:10.1001/jamanetworkopen.2020.1664
34. Inomata M, Shiroshita H, Uchida H, et al. Current status of endoscopic surgery in Japan: the 14th National Survey of Endoscopic Surgery by the Japan Society for Endoscopic Surgery. *Asian J Endosc Surg*. 2020;13(1):7-18. doi:10.1111/ases.12768

SUPPLEMENT.

eTable 1. Four Categories of Common Criteria (60 Points)

eTable 2. Procedure-Specific Criteria for Sigmoidectomy (2 Points Allotted for Each Item)