2020 SAGES POSTER

Automated operative phase identification in peroral endoscopic myotomy

Thomas M. Ward[1,2] · Daniel A. Hashimoto[1,2] · Yutong Ban[1,2,4] · David W. Rattner[2] · Haruhiro Inoue[3] · Keith D. Lillemoe[2] · Daniela L. Rus[4] · Guy Rosman[1,4] · Ozanan R. Meireles[1,2]

Received: 9 April 2020 / Accepted: 16 July 2020 / Published online: 27 July 2020

Abstract

Background Artificial intelligence (AI) and computer vision (CV) have revolutionized image analysis. In surgery, CV applications have focused on surgical phase identification in laparoscopic videos. We proposed to apply CV techniques to identify phases in an endoscopic procedure, peroral endoscopic myotomy (POEM).

Methods POEM videos were collected from Massachusetts General and Showa University Koto Toyosu Hospitals. Videos were labeled by surgeons with the following ground truth phases: (1) Submucosal injection, (2) Mucosotomy, (3) Submucosal tunnel, (4) Myotomy, and (5) Mucosotomy closure. The deep-learning CV model—Convolutional Neural Network (CNN) plus Long Short-Term Memory (LSTM)—was trained on 30 videos to create POEMNet. We then used POEMNet to identify operative phases in the remaining 20 videos. The model's performance was compared to surgeon annotated ground truth.

Results POEMNet's overall phase identification accuracy was 87.6% (95% CI 87.4–87.9%). When evaluated on a per-phase basis, the model performed well, with mean unweighted and prevalence-weighted F1 scores of 0.766 and 0.875, respectively. The model performed best with longer phases, with 70.6% accuracy for phases that had a duration under 5 min and 88.3% accuracy for longer phases.

Discussion A deep-learning-based approach to CV, previously successful in laparoscopic video phase identification, translates well to endoscopic procedures. With continued refinements, AI could contribute to intra-operative decision-support systems and post-operative risk prediction.

Keywords Computer vision · Deep learning · Endoscopy · Artificial intelligence · Phase identification · Phase segmentation

We are in the midst of an artificial intelligence (AI) and computer vision (CV) revolution. CV, a sub-field of AI, teaches computers to not only "see" images but understand their contents [1]. In 2012, Krizhevsky designed a novel neural network architecture that could attain human-like image comprehension with readily available computer hardware, thereby democratizing CV and sparking the CV revolution [2]. Since this discovery, the medical field has seen numerous CV applications, with algorithms published that perform at levels similar to pathologists [3], radiologists [4], and dermatologists [5].

CV in surgery, though, has seen relatively fewer advances, which stems from the magnitude of information in surgical video and difficulties in teaching AI algorithms surgical workflow. Current efforts in the surgical community focus on automated phase identification in surgical videos. Laparoscopic surgeries, due to their readily available video feed and stable field-of-view, lend themselves to CV analysis, with work done in cholecystectomy (86.7% accuracy [6]), sleeve gastrectomy (85.6% accuracy [7]), and sigmoid colectomy (91.9% accuracy [8]).

We sought to apply CV techniques to identify surgical phases for the first time on an endoscopic procedure. As our

✉ Thomas M. Ward
tmward@mgh.harvard.edu

[1] Surgical AI and Innovation Laboratory, Massachusetts General Hospital, 15 Parkman St., WAC 460, Boston, MA 02114, USA

[2] Department of Surgery, Massachusetts General Hospital, Boston, MA, USA

[3] Digestive Disease Center, Showa University Koto Toyosu Hospital, Tokyo, Japan

[4] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

target procedure, we selected peroral endoscopic myotomy (POEM), a minimally invasive endoscopic treatment for esophageal achalasia.

Materials and methods

Institutional approval

This study's protocol was reviewed and approved by the Partners Healthcare Institutional Review Board (Protocol No: 2015P001161). Written patient consent for use of the videos for research purposes was obtained prior to any procedures being performed.

Dataset

We collected 35 videos from Massachusetts General Hospital (Boston, USA) and 15 videos from Showa University Koto Toyosu Hospital (Tokyo, Japan) for a total of 50 POEM videos. Videos were processed and de-identified with the FFmpeg software [9].

Three surgeons annotated the videos with the operative phases (Table 1), outlined in Inoue et al. [10], to create the "ground truth" information from which our AI model would learn. To assess inter-annotator reliability and agreement, each surgeon annotated 14 identical videos. We set a Krippendorff's alpha greater than 0.800 to indicate sufficient inter-annotator reliability to allow for pooling of multiple surgeons' annotations into a single training set for the model [11]. We compared levels of inter-annotator agreement, on a per-step and overall basis, with Fleiss' kappa, using kappa ranges of 0–0.20, 0.21–0.40, 0.41–0.60, 0.61–0.80, and 0.81–1.00 to indicate poor, slight, fair, moderate, substantial, and almost perfect agreement, respectively [12]. All calculations were performed using R version 3.6.1 [13] with the *irr* package [14], treating the annotation for each video second as an independent observation. After confirmation of annotation similarity, each surgeon annotated a subset of the 50 total videos.

Model architecture

Automated surgical phase identification AI models receive individual video frames and attempt to classify them as one of a pre-learned set of phases. Similar to our prior work on laparoscopic sleeve gastrectomy [7], we based the architecture of our AI model on a Convolutional Neural Network (CNN) visual model (ResNet [15]) combined with a Long Short-Term Memory (LSTM) temporal model (Fig. 1). As their names imply, the visual model attempts to classify each frame of a video based on visual features alone, while the temporal model considers data about the temporal order in which frames have appeared. We implemented our model with the PyTorch library [16].

For the AI model to learn to identify POEM phases, we first trained it on 30 randomly selected videos. We down-sampled the original videos to one image per second. These images, paired with their annotated surgical phase ground truths, were given to the model as the training set. The model self-adjusted the internal parameters in its neural network until it could generate a consistent and correct phase identification for each second, thereby creating the "trained" model: POEMNet.

We tested POEMNet on the remaining 20 videos. As with our training set, we down-sampled the original videos to one image per second. Each second's image was given first to the visual model then to the model's temporal "memory" component (LSTM) to generate a likeliest phase identified from the visual cues combined with its knowledge of prior seen images (Fig. 1). As a proof-of-concept for offline applications, a forward–backward Hidden Markov Model (HMM) was applied in post-processing to improve performance on phases with a short duration.

Evaluation metrics

We compared POEMNet's phase identifications to the surgeon-labeled phases to evaluate performance on an overall, per-phase, and per-duration, basis. Calculations were performed with the *caret* package [17] and graphics generated with *ggplot2* [18] in R 3.6.1. We computed the following metrics:

$$\text{Accuracy} = \frac{\text{Correctly identified frames}}{\text{Total number of frames for a time duration}}$$

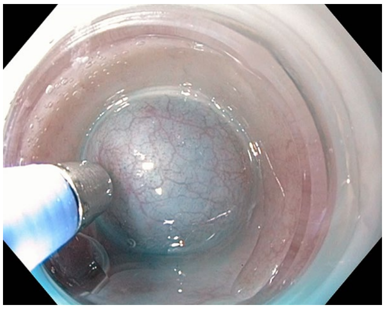
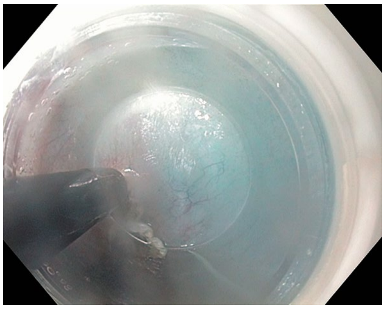
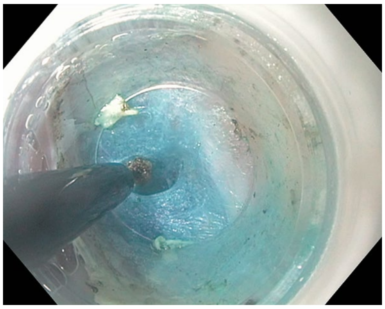
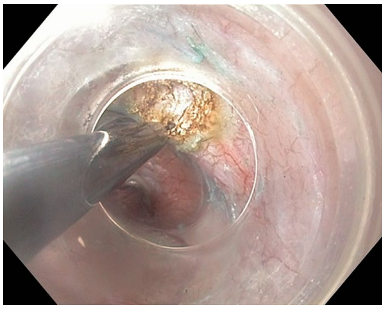
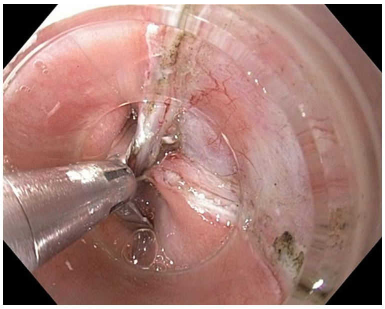
$$\text{Precision} = \frac{\text{Correctly identified frames for a phase}}{\text{Total number of frames identified for a phase}}$$

$$\text{Recall} = \frac{\text{Correctly identified frames for a phase}}{\text{Total number of frames for a phase}}$$

$$\text{F-score} = 2 \cdot \frac{\text{Recall} \ \cdot \text{Precision}}{\text{Recall} + \text{Precision}}.$$

Accuracy demonstrates a model's overall performance across the entire test set and allows for performance evaluation of phases with certain intervals (e.g., all phases that are under 30 s in length). Precision (positive predictive value) demonstrates the model's rate of mis-classifications of a phase, while recall (sensitivity) shows its ability to find all the frames of a phase [19].

Table 1 Five phases annotated in the POEM videos with representative images and descriptions of the start and stop points

Phase	Representative image	Description
Submucosal injection		Start: needle touches mucosa End: needle withdraws from mucosa
Mucosotomy		Start: energy device touches mucosa End: mucosal incision finished
Submucosal tunnel		Start: energy device touches submucosa End: no further energy application
Myotomy		Start: energy device first touches muscle End: no further energy application
Mucosotomy closure		Start: first clip touches mucosotomy End: last clip fully applied

We chose start and stop points based on tools touching their target tissue to improve agreement between annotators

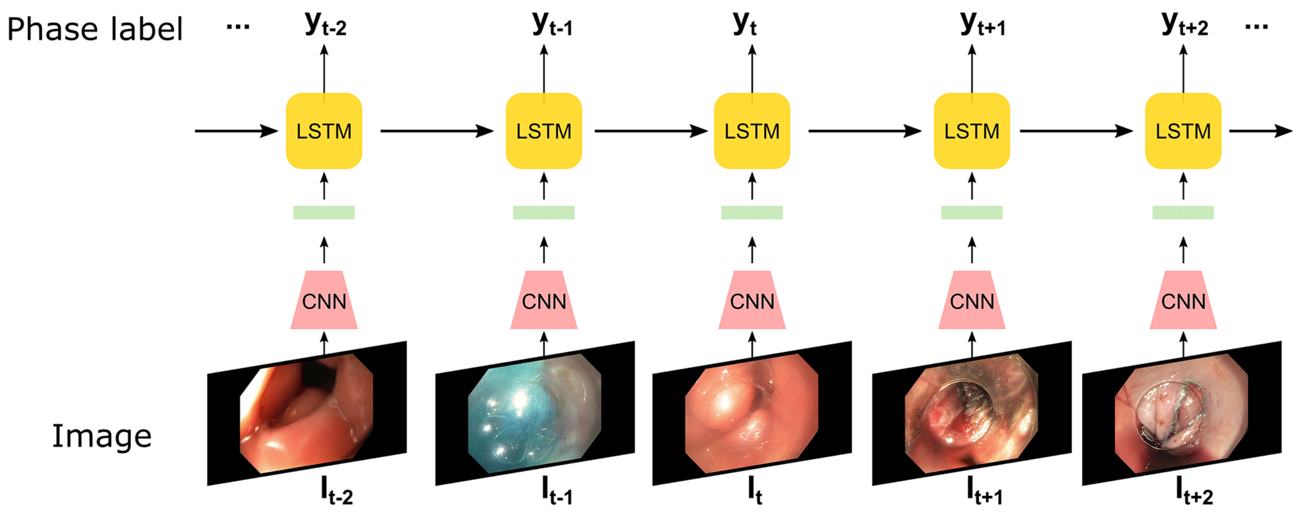


Fig. 1 Simplified graphical representation of POEMNet's architecture. The CNN (visual model) processes each image, generating intermediate outputs (green bars). The LSTM (temporal model) refines these outputs to generate likeliest phase labels. The LSTM carries forward memory of prior images and phase labels across time to better inform phase identification (Color figure online)

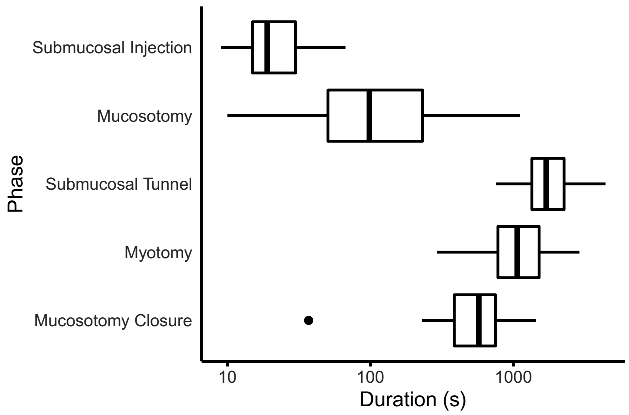


Fig. 2 Distribution of phase duration. Phase duration (seconds) was calculated from the ground truth annotation of all 50 videos in the dataset. Duration was plotted on a logarithmic scale. "Submucosal Injection" and "Submucosal Tunnel" were the shortest and longest phases, with mean lengths of 25.1 ± 14.4 s and 31.8 ± 12.8 min, respectively. The 37 s "Mucosotomy Closure" outlier occurred in a video that ended early after a single clip application

Results

Video information

We collected 50 POEM videos with a mean length of 93.04 min (± 25.98 min standard deviation). The five phases of POEM vary widely in length, from the shortest, "Submucosal Injection," to the longest, "Submucosal Tunnel" (Fig. 2).

Table 2 Inter-annotator agreement, per-phase and overall, across 14 videos assessed by Fleiss' kappa

Phase	κ
Submucosal injection	0.809
Mucosotomy	0.775
Submucosal tunnel	0.839
Myotomy	0.920
Mucosotomy closure	0.932
Overall	0.821

Kappa ranges of 0–0.20, 0.21–0.40, 0.41–0.60, 0.61–0.80, and 0.81–1.00 indicate poor, slight, fair, moderate, substantial, and almost perfect agreement, respectively [12]

Inter-annotator reliability and agreement

The three annotators had sufficient inter-annotator reliability to combine their annotations of different videos into a single dataset, with a Krippendorff's alpha coefficient of 0.882. Overall, there was almost perfect agreement between the three surgeons, with a Fleiss' kappa of 0.821. They almost perfectly agreed across all phases except for "Mucosotomy" (Table 2).

Model results

Overall

POEMNet's overall automated phase identification accuracy was 87.6% (95% CI 87.4–87.9%). Straightforward cases yielded nearly perfect phase identification, while more difficult cases that required repeated tunnel inspection, scope

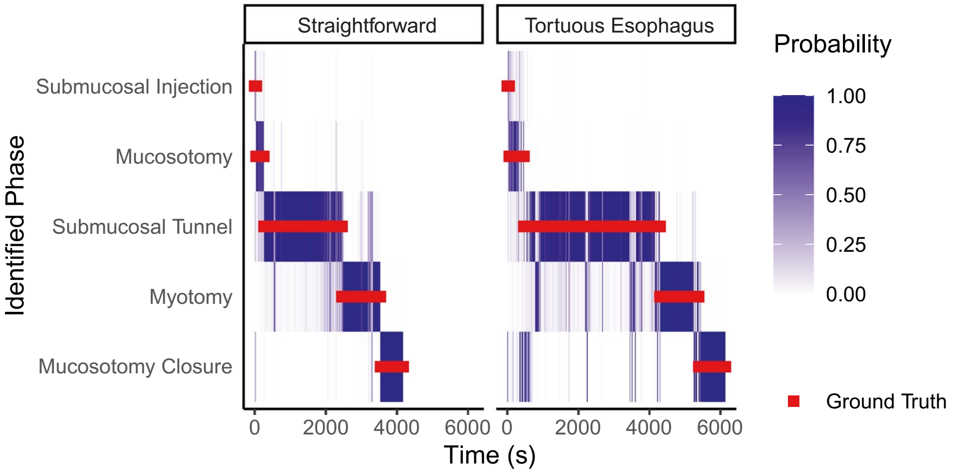


Fig. 3 "Surgical fingerprints" for two representative cases. The "Straightforward" case proceeded with nearly complete alignment between identified phase (blue bar) and ground truth (red line), with an accuracy of 96.0%. The model's certainty is proportional to the bar's darkness (white for low probability of identified phase; dark blue for high probability of identified phase). The "Tortuous Esophagus" case had a difficult "Submucosal Tunnel" creation, leading to repeated inspections, scope cleanings, and a mucosal perforation. POEMNet tried to classify these deviations as other, more visually similar, phases, shown by areas along the ground truth with gaps, signifying no identification for the technically "correct" phase. Accuracy for this video was 84.8% (Color figure online)

Table 3 POEMNet's performance across phases

	Precision	Recall	F1 score	Prevalence
Submucosal injection	0.667	0.361	0.468	0.006
Mucosotomy	0.837	0.602	0.700	0.044
Submucosal tunnel	0.955	0.840	0.894	0.513
Myotomy	0.791	0.945	0.861	0.278
Mucosotomy closure	0.848	0.971	0.906	0.159
Overall (unweighted)	0.820	0.744	0.766	
Overall (weighted)	0.885	0.876	0.875	

The model did well in phase identification across phases. It experienced some difficulties in recall for the shorter phases, "Submucosal Injection" and "Mucosotomy," which together comprised under 5% of all video frames. The overall metrics were calculated from unweighted and prevalence-weighted averages of each phase. Overall accuracy (frames correctly identified/total frames) was 87.6% (95% CI 87.4–87.9%)

cleaning, and repair of mucosal perforations caused the machine to mis-classify phases (Fig. 3).

Per-phase

The model performed well when evaluated on a per-phase basis (Table 3). The longer phases—"Submucosal Tunnel," "Myotomy," and "Mucosotomy Closure"—had the best results, with few mis-classifications. Due to visual similarity, "Submucosal Tunnel" and "Myotomy" had a small amount of shared mis-classifications. The model struggled with the shorter phases—"Submucosal Injection" and "Mucosotomy"—with decreased phase identification performance and increased shared mis-classifications (Fig. 4). The model's difficulty with these phases mirrors their lower inter-annotator agreements (Table 2). The shorter phases comprised a small percentage of total video duration and therefore minimally impacted overall performance.

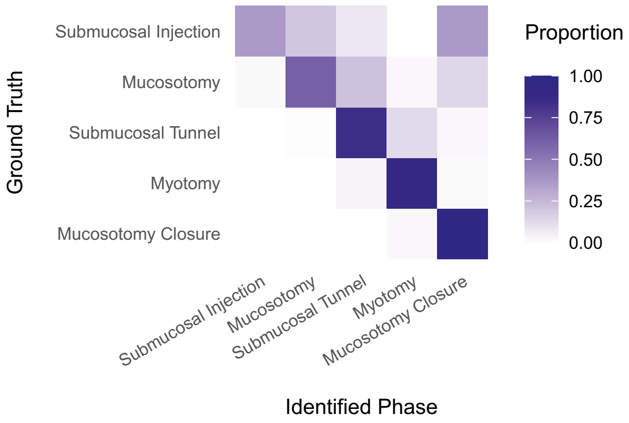


Fig. 4 Confusion matrix (recall) of the model's identified phase for each ground truth phase. For each ground truth phase, the proportion of frames assigned to each identified phase is represented by the color in the heat map (white for low proportion of frames assigned to identified step; dark blue for high proportion of frames assigned to identified step). The short "Submucosal Injection" and "Mucosotomy" phases proved more difficult for accurate classification, while later phases, despite high visual similarity, demonstrated improved performance (Color figure online)

Per-duration

The model performed best on longer duration phases as noted above. Some phases varied widely in duration (Fig. 2), so model performance was additionally analyzed on a phase duration basis. Model performance increased proportional to phase duration, with the largest jumps when the phase duration increased to over 30 s (55.1% to 72.3% accuracy) and over 5 min (77.1% to 95.4%) in length. As a proof-of-concept, we applied a forward–backward HMM, which took both past and future phase information, to improve phase identification. The HMM's filtering boosted the model's phase identification performance of phases with a duration under 60 s (Table 4).

Discussion

This study demonstrates that a deep-learning approach to automated operative phase identification, previously employed with laparoscopic video, translates well to endoscopic procedures with 87.6% accuracy. We obtained this accuracy with a video dataset from two institutions in different countries and a training set of only 30 videos, unlike prior studies that all used single-institution datasets and larger training sets [7, 8, 20].

Phase identification in POEM has immediate applications. It could offer real-time updates on the procedure's progress for optimization of operating room and endoscopy suite workflow. In addition, it could automate indexing of cases for surgical libraries. Clinical decision-support systems could incorporate our "Surgical fingerprint" technology to elicit a "call a friend" prompt if portions of the case start to become complicated, as illustrated in Fig. 3. These prompts could be configurable based on operator experience, with surgeons and residents learning new procedures having lower thresholds for an assistance prompt.

Table 4 Per-duration model accuracy

Duration (s)	Accuracy (POEMNet)	Accuracy (POEMNet + HMM)
1–30	0.418	0.551
31–60	0.643	0.723
61–300	0.760	0.771
301–600	0.936	0.954
> 600	0.878	0.900

Individual phases from each video were grouped based on length, and then accuracy (total number of seconds correctly identified/total number of seconds) was calculated in aggregate. POEMNet's accuracy increased proportionally to phase duration. We additionally assessed accuracy applying a forward–backward Hidden Markov Model (HMM) filter, which took into account past and future phase information. The HMM improved performance for phases with a duration under 60 s

Our current work is limited by the annotation granularity. Our annotation schema, with its five labels, only captures big-picture operative phases, which does not account for the smaller sub-components that, in total, comprise a phase. "Submucosal Tunnel," for example, could be split into "Introduction of endoscope into submucosal space," "Dissection of submucosal layer," "Submucosal inspection of tunnel," and "Intraluminal inspection of tunnel." POEMNet's lower accuracy in more complicated cases, like the Tortuous Esophagus in Fig. 3, shows the limitations of a non-granular annotation schema: it attempted to classify the non-standard deviations from a typical "Submucosal Tunnel" (e.g., closure of inadvertent mucosotomy) as more visually similar (and "surgically" correct) phase labels ("Mucosotomy Closure"). Training a machine to identify these smaller sub-components would greatly increase understanding of the operation and afford more opportunity to identify deviation phases and errors. The more granular annotation structure would create a foundation for advanced warning prior to adverse events (like mucosal perforation in POEM) and provide meaningful intra-operative findings to guide post-operative risk prediction.

Increased annotation granularity presents difficulties for current deep-learning technology because it leads to shorter temporal segments. POEMNet's overall performance was excellent, but it struggled with phases less than 1 min in length. To our knowledge, CV model performance, on a per-phase-duration basis, has not previously been reported; other studies have only analyzed overall and per-phase performance. Some of the difficulty in the detection of short duration phases stems from inherent training issues. First, there is simply less training data available. Second, training optimizes for overall accuracy, which means the neural network self-tunes to perform best on longer duration phases to best optimize its score and complete training. To mitigate the training issues, we were able to boost the short-phase identification performance with HMM post-processing (Table 4). This HMM approach refines phase identification for a certain second with information from past and future seconds of the video, so it is only useful for offline applications after a completed procedure.

These struggles speak to the architecture of current models, which mainly rely on visual features with some incorporation of prior temporal events to help guide phase identification. Unfortunately, "temporal models," like LSTM, are limited in their ability to handle temporal relations. Their limited memory (usually only under a minute of information) leads to identification errors through memory bias. If the past minute was all a single phase, they tend to oversmooth their identifications, predicting future seconds to be the same as prior ones. This bias will produce improved results, if the phases are long and continuous with few transitions, but will fail to predict short duration phases. Their limited memory also fails to capture long-term understanding of the procedure, which can lead to predictions of phases that happened well in the past due to "forgetting" that they already occurred. Complete understanding of a surgery at a certain time necessitates knowledge of all the procedure's preceding events and the ability to look at different time scales. Neural network models need novel designs to afford this introspection of temporal surgical dependencies. There have been initial efforts in this area, though they do not generalize to all types of procedures and require hand-crafted knowledge of operative phase workflow to limit incorrect identification of already performed phases [21].

Annotation granularity will also present challenges for annotation consistency. The machine depends on its inputs to correctly learn. Building upon our previous work [7], we analyzed inter-annotator agreement and reliability for independently annotated data. We ensured that our three surgeons had sufficient inter-annotator reliability to train a CV model with pooled annotations, a pre-training validation step that, to our knowledge, has not previously been reported. Overall, our three annotators agreed nicely, with near-perfect agreement across most phases. The shorter phases, "Submucosal Injection" and "Mucosotomy," did experience lower agreement, which gave differing inputs from which the machine learned. These inconsistent inputs could account for the model's difficulties with phase identification in these shorter phases. Similarly, an annotator disagreement source centered around the transition from "Mucosotomy" to "Submucosal Tunnel," which reflected in the model's mis-classification of some of the former phase as the latter (Fig. 4). Improved annotation granularity requires reproducible annotation standards. The Society of American Gastrointestinal and Endoscopic Surgeons (SAGES) has recently spear-headed the effort to produce surgical video annotation standards, bringing together an international cross-disciplinary group for a Video Annotation Conference in February 2020.

With improved model design and these new annotation standards, we hope that short-phase identification will one day become a reality. AI that can perform granular phase identification promises a collective surgical consciousness, based on myriad videos and surgeons' experiences, that can provide intra- and post-operative adverse event prevention and guide surgeons and endoscopists to performing better, and more importantly safer, procedures across the world.

Acknowledgements This work was supported by a 2017 research award from the Natural Orifice Surgery Consortium for Assessment and Research (NOSCAR). Daniel Hashimoto was partly funded for this work by NIH Grant T32DK007754-16A1 and the MGH Edward D. Churchill Research Fellowship. The authors thank Caitlin Stafford, CCRP, for her assistance and support in research management.

Compliance with ethical standards

Disclosures Drs. Ban, Hashimoto, Meireles, Rosman, Rus, and Ward receive research support from Olympus Corporation. Dr. Hashimoto is a consultant for Verily Life Sciences, Johnson & Johnson Institute, Worrell Inc, Mosaic Research Management, and Gerson Lehrman Group. Dr. Rattner is a consultant for the Olympus Corporation. Dr. Meireles is a consultant for the Olympus Corporation and Medtronic. Dr. Inoue is a consultant for the Olympus and Top Corporations and received research support from the Olympus Corporation and Takeda Pharmaceutical Company. Drs. Rosman and Rus receive research support from Toyota Research Institute (TRI). Dr. Lillemoe has no conflicts of interest nor financial ties to disclose.

References

1. Hashimoto DA, Rosman G, Rus D, Meireles OR (2018) Artificial intelligence in surgery: promises and perils. Ann Surg 268(1):70–76. https://doi.org/10.1097/SLA.0000000000002693
2. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. In: Pereira F, Burges CJC, Bottou L, Weinberger KQ (eds) Advances in neural information processing systems, vol 25. Curran Associates Inc, New York, pp 1097–1105
3. Hollon TC, Pandian B, Adapa AR, Urias E, Save AV, Khalsa SSS, Eichberg DG, D'Amico RS, Farooq ZU, Lewis S, Petridis PD, Marie T, Shah AH, Garton HJL, Maher CO, Heth JA, McKean EL, Sullivan SE, Hervey-Jumper SL, Patil PG, Thompson BG, Sagher O, McKhann GM, Komotar RJ, Ivan ME, Snuderl M, Otten ML, Johnson TD, Sisti MB, Bruce JN, Muraszko KM, Trautman J, Freudiger CW, Canoll P, Lee H, Camelo-Piragua S, Orringer DA (2020) Near real-time intraoperative brain tumor diagnosis using stimulated Raman histology and deep neural networks. Nat Med. https://doi.org/10.1038/s41591-019-0715-9
4. Rajpurkar P, Irvin J, Zhu K, Yang B, Mehta H, Duan T, Ding D, Bagul A, Langlotz C, Shpanskaya K, Lungren MP, Ng AY (2017) CheXNet: radiologist-level pneumonia detection on chest X-rays with deep learning. arXiv:1711.05225 [cs, stat]
5. Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118. https://doi.org/10.1038/nature21056
6. Yengera G, Mutter D, Marescaux J, Padoy N (2018) Less is more: surgical phase recognition with less annotations through self-supervised pre-training of CNN-LSTM networks. arXiv:1805.08569 [cs]
7. Hashimoto Daniel A, Rosman Guy R, Witkowski Elan J, Stafford Caitlin W, Navarette-Welton Allison D, Rattner David L, Lillemoe Keith R, Rus Daniela R, Meireles Ozanan R (2019) Computer vision analysis of intraoperative video: automated recognition of operative steps in laparoscopic sleeve gastrectomy. Ann Surg 270(3):414–421. https://doi.org/10.1097/SLA.0000000000003460
8. Kitaguchi D, Takeshita N, Matsuzaki H, Takano H, Owada Y, Enomoto T, Oda T, Miura H, Yamanashi T, Watanabe M, Sato D, Sugomori Y, Hara S, Ito M (2019) Real-time automatic surgical phase recognition in laparoscopic sigmoidectomy using the convolutional neural network-based deep learning approach. Surg Endosc. https://doi.org/10.1007/s00464-019-07281-0
9. Fabrice Bellard: FFmpeg. http://ffmpeg.org/about.html
10. Inoue H, Minami H, Kobayashi Y, Sato Y, Kaga M, Suzuki M, Satodate H, Odaka N, Itoh H, Kudo S (2010) Peroral

endoscopic myotomy (POEM) for esophageal achalasia. Endoscopy 42(04):265–271. https://doi.org/10.1055/s-0029-1244080

11. Krippendorff K (2004) Content analysis: an introduction to its methodology, 2nd edn. Sage, Thousand Oaks
12. Landis JR, Koch GG (1977) The measurement of observer agreement for categorical data. Biometrics 33(1):159–174. https://doi.org/10.2307/2529310
13. R Core Team (2019) R: a language and environment for statistical computing. R Foundationfor Statistical Computing, Vienna
14. Gamer M, Lemon J, Singh IFP (2019) Irr: various coefficients of interrater reliability and agreement. CRAN
15. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), pp 770–778. https://doi.org/10.1109/cvpr.2016.90
16. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison A, Kopf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S (2019) PyTorch: an imperative style, high-performance deep learning library. In: Wallach H, Larochelle H, Beygelzimer A, Alche-Buc F, Fox E, Garnett R (eds) Advances in neural information processing systems, vol 32. Curran Associates Inc, New York, pp 8024–8035
17. Kuhn M (2020) Caret: classification and regression training. CRAN
18. Wickham H (2016) Ggplot2: elegant graphics for data analysis. Springer, New York
19. Padoy N, Blum T, Ahmadi SA, Feussner H, Berger MO, Navab N (2012) Statistical modeling and recognition of surgical workflow. Med Image Anal 16(3):632–641. https://doi.org/10.1016/j.media.2010.10.001
20. Twinanda AP, Shehata S, Mutter D, Marescaux J, de Mathelin M, Padoy N (2017) EndoNet: a deep architecture for recognition tasks on laparoscopic videos. IEEE Trans Med Imaging 36(1):86–97. https://doi.org/10.1109/TMI.2016.2593957
21. Jin Y, Dou Q, Chen H, Yu L, Qin J, Fu CW, Heng PA (2018) SV-RCNet: workflow recognition from surgical videos using recurrent convolutional network. IEEE Trans Med Imaging 37(5):1114–1126. https://doi.org/10.1109/TMI.2017.2787657